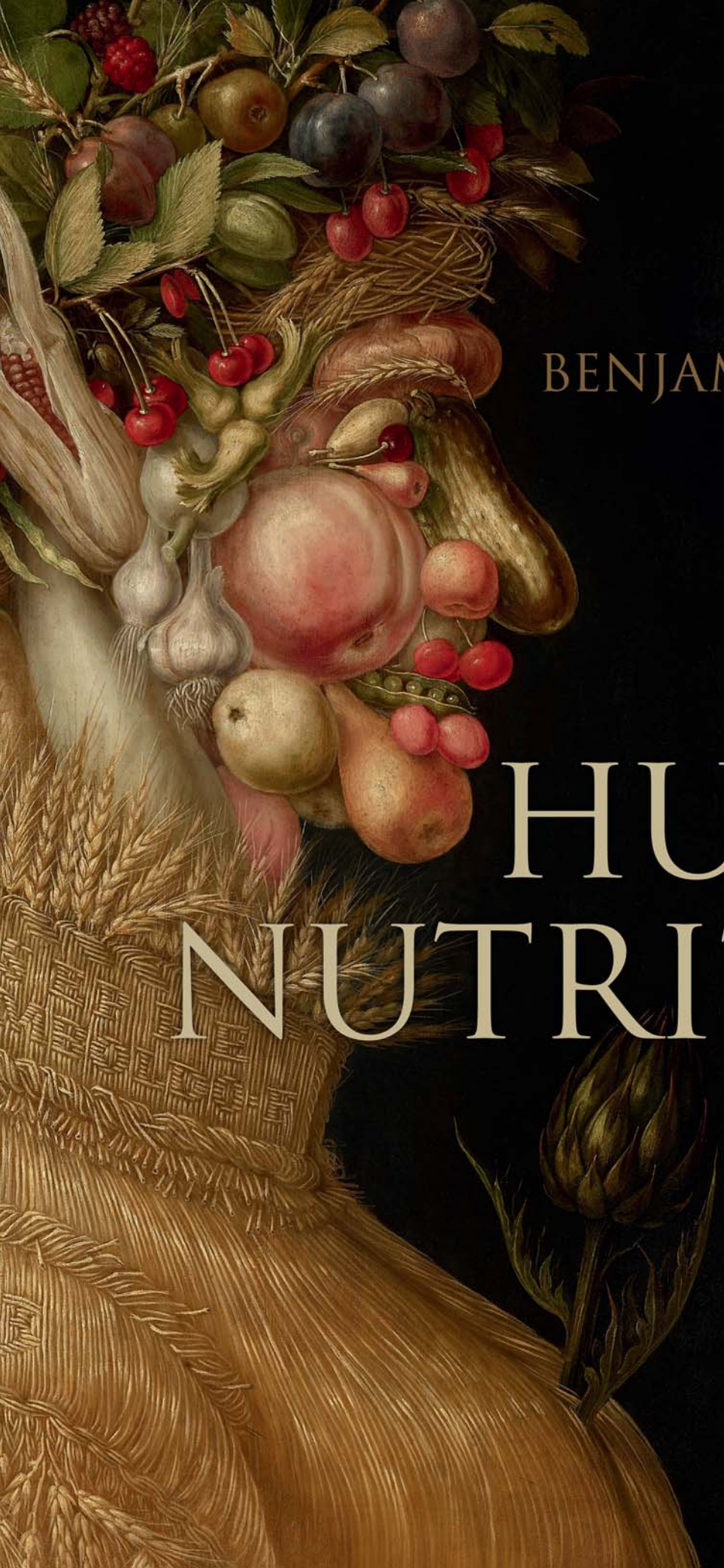


EDITOR IN CHIEF

BENJAMIN CABALLERO

VOLUME ONE

Encyclopedia of HUMAN NUTRITION

FOURTH EDITION

ENCYCLOPEDIA OF HUMAN NUTRITION

FOURTH EDITION

This page intentionally left blank

ENCYCLOPEDIA OF HUMAN NUTRITION

FOURTH EDITION

EDITOR IN CHIEF

Professor Benjamin Caballero,

Emeritus Professor, Bloomberg School of Public Health and School of Medicine,
Johns Hopkins University
Baltimore, MD, United States

VOLUME 1

Section 1: The Foundations of Human Nutrition

Edited by Angel Gil

Honorary President of the Iberoamerican Nutrition Foundation (FINUT),
Emeritus Professor, Department of Biochemistry and Molecular Biology II,
Institute of Nutrition and Food Technology, Centre of Biomedical Research,
University of Granada, Campus de la Salud, Avda. del Conocimiento,
18100 Armilla, Granada, Spain.

ELSEVIER

AMSTERDAM ● BOSTON ● HEIDELBERG ● LONDON ● NEW YORK ● OXFORD
PARIS ● SAN DIEGO ● SAN FRANCISCO ● SINGAPORE ● SYDNEY ● TOKYO
Academic Press is an imprint of Elsevier

ACADEMIC PRESS

Academic Press is an imprint of Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge MA 02139, United States

Copyright © 2023 Elsevier Ltd. All rights reserved

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers may always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN 978-0-12-821848-8

For information on all publications visit our website at http://store.elsevier.com



Publisher: Oliver Walter
Acquisitions Editors: Emma Hayes and Michael Nicholls
Content Project Managers: Richard Berryman and Sindhuja Vadlamani
Associate Content Project Manager: N. Kiruthigadevi
Designer: Matthew Limbert

CONTENTS OF VOLUME 1

Contributors to Volume 1 *ix*

Editor Biographies *xiii*

Preface *xvii*

THE FOUNDATIONS OF HUMAN NUTRITION

Aluminum 1
Robert A Yokel

Amino acids: Chemistry and classification 10
Guoyao Wu

Amino acids: Metabolism 23
Guoyao Wu

Amino acids: Specific functions 36
Guoyao Wu

Antioxidants: Intervention studies 48
Rafael A Casuso, Javier González Gallego, and Jesús Rodríguez Huertas

Antioxidants: Physiology and dietary sources 61
Jesús R Huertas, S Stanner, and E Weichselbaum

Bioavailability 76
L Davidsson and SA Tanumihardjo

Biotin 85
Donald M Mock and Luis Fontana

Caffeine 96
Rafael Urrialde

Calcium 105
V Loria-Kohen, AM López-Sobaler, A Aparicio, and RM Ortega

Carbohydrates: Chemistry and classification 114
Chrysanthi Stylianopoulou

Carbohydrates: Regulation of metabolism 126
Chrysanthi Stylianopoulou

Carbohydrates: Requirements and dietary importance 136
Chrysanthi Stylianopoulou

Carotenoids: Chemistry, sources and physiology 144
E Rodríguez-Rodríguez, LM Bermejo, and RM Ortega

Carotenoids: Health effects 154
Sherry A Tanumihardjo

Choline and phosphatidylcholine 162
Kevin C Klatt

Cholesterol: Sources, absorption, function, and metabolism 175
DJ McNamara

Chromium 181
John B Vincent

Cofactors: Inorganic 189
ED Harris

Copper 200
Peter J Aggett

Dietary fiber: Classification and physiological role 209
Madeline Timm and Joanne Slavin

Energy requirements 217
Susan B Roberts and Rachel E Silver

Fatty acids: Omega-3 polyunsaturated 223
AP Simopoulos

Fatty acids: Omega-6 polyunsaturated 232
TA Mori and JM Hodgson

Fiber intake and resulting health benefits 239
Rebekah Schulz and Joanne Slavin

Folate/folic acid 247
Joshua W Miller

Homocysteine 259
Joshua W Miller

Iodine: Deficiency disorders and prevention programs 268
MB Zimmermann

Iodine: Physiology, dietary sources, and requirements 273
Sonja Y Hess and Elizabeth N Pearce

Iron: Physiology, dietary sources, and requirements 282
B Lönnerdal and O Hernell

Lipids (fats and oils) 291
AH Lichtenstein

Lipoproteins 301
JM Ordovas

Magnesium 309
LH Allen

Manganese 315
CL Keen, JL Ensunsa, B Lönnerdal, and S Zidenberg-Cherr

Niacin and pellagra 322
CJ Bates

Nucleic acids, purine, and pyrimidine nucleotides and nucleosides: Physiology, toxicology, and dietary sources 330
EA Carrey, D Perrett, and HA Simmonds

Pantothenic acid 340
Christopher J Bates and Luis Fontana

Phosphorus 346
Olga P García

Physical activity: Beneficial effects 352
Peter T Katzmarzyk

Potassium 362
LJ Appel

Protein deficiency 368
Rajavel Elango

Protein digestion and bioavailability 375
Daniel Tomé

Proteostasis and turnover 388
DJ Millward

Riboflavin 401
Christopher J Bates and Luis Fontana

Selenium 412
Anna Patricia Kipp

Sodium: Physiology and dietary sources 420
Elena Rho and Harald Seeger

Thiamin: Beriberi 428
DI Thurnham

Thiamin: Physiology 440
DI Thurnham

Ultratrace elements 448
Forrest H Nielsen

Vitamin A: Deficiency and interventions 460
Amanda C Palmer and Keith P West, Jr.

Vitamin A: Physiology, dietary sources and requirements 471
A Catharine Ross

Vitamin B_{12}: Physiology, dietary sources, and requirements 482
R Green

Vitamin B_6 489
PB Mills, EJ Footitt, and PT Clayton

Vitamin C 504
David A Bender and Luis Fontana

Vitamin C: Deficiency states 515
CJ Bates

Vitamin D: Physiology, clinical applications, dietary sources, and requirements 521
MF Holick

Vitamin D: Role in chronic and acute diseases 535
Ana I Álvarez-Mercado, María Dolores Mesa, and Ángel Gil

Vitamin E | metabolism and requirements 545
Maret G Traber

Vitamin E physiology and health effects 556
Jean-Marc Zingg

Vitamin K 569
MK Shea and SL Booth

Zinc: Deficiency disorders and prevention programs 577
SY Hess

Zinc: Physiology, dietary sources, and requirements 584
HC Freake and K Sankavaram

CONTRIBUTORS TO VOLUME 1

Peter J Aggett
University of Central Lancashire, Preston, United Kingdom

LH Allen
University of California, Davis, CA, USA

Ana I Álvarez-Mercado
Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology "José Mataix", Center of Biomedical Research, University of Granada, Granada, Spain; and Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain

A Aparicio
Department of Nutrition and Food Science, Complutense University of Madrid. UCM Research Group VALORNUT-920030, Madrid. Spain

LJ Appel
Johns Hopkins University, Baltimore, MD, United States

Christopher J Bates
MRC Human Nutrition Research, Cambridge, United Kingdom

David A Bender
Department of Structural and Molecular Biology, University College London, London, United Kingdom

LM Bermejo
Department of Nutrition and Foods Science, F. Pharmacy, Complutense University of Madrid, Madrid, Spain

SL Booth
USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA, United States

EA Carrey
UCL Institute of Child Health, London, UK

Rafael A Casuso
Department of Physiology, Institute of Nutrition and Food Technology, Biomedical Research Center, University of Granada, Granada, Spain

PT Clayton
Genetics and Genomic Medicine, Great Ormond Street Institute of Child Health, University College London, London, United Kingdom

L Davidsson
International Atomic Energy Agency (IAEA), Vienna International Centre, Vienna, Austria

Rajavel Elango
Department of Pediatrics, University of British Columbia, BC Children's Hospital Research Institute, School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

JL Ensunsa
University of California, Davis, CA, USA

Luis Fontana
Department of Biochemistry and Molecular Biology 2, School of Pharmacy, Granada, Spain; Institute of Nutrition and Food Technology "José Mataix", Biomedical Research Center, Parque Tecnológico Ciencias de la Salud, Granada, Spain; and Instituto de Investigación Biosanitaria ibs.GRANADA, Complejo Hospitalario Universitario de Granada, Granada, Spain

EJ Footitt
Department of Paediatric Metabolic Medicine, Great Ormond Street Hospital NHS Trust, London, United Kingdom

HC Freake
University of Connecticut, Storrs, CT, USA

Javier González Gallego
Institute of Biomedicine (IBIOMED), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), University of León, León, Spain

Olga P García
School of Natural Sciences, Universidad Autónoma de Querétaro, Querétaro, Mexico

Ángel Gil
Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology "José Mataix", Center of Biomedical Research, University of Granada, Granada, Spain; Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain; and CIBEROBN (Physiopathology of Obesity and Nutrition Network CB12/03/30038), Institute of Health Carlos III (ISCIII), Madrid, Spain

R Green
University of California, Davis, CA, United States

ED Harris
Texas A&M University, College Station, TX, USA

O Hernell
Umeå University, Umeå, Sweden

Sonja Y Hess
Department of Nutrition, University of California Davis, Davis, CA, United States

JM Hodgson
School of Medicine and Pharmacology, University of Western Australia, Perth, WA, Australia

MF Holick
Boston University Medical Center, Boston, MA, United States

Jesús Rodríguez Huertas
Department of Physiology, Institute of Nutrition and Food Technology, Biomedical Research Center, University of Granada, Granada, Spain

Peter T Katzmarzyk
Pennington Biomedical Research Center, Baton Rouge, LA, United States

CL Keen
University of California, Davis, CA, USA

Anna Patricia Kipp
Department of Nutritional Physiology, Institute of Nutritional Sciences, Friedrich Schiller University Jena, Jena, Germany

Kevin C Klatt
Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX, United States; and Center for Precision Environmental Health, Baylor College of Medicine, Houston, TX, United States

AH Lichtenstein
Tufts University, Boston, MA, United States

B Lönnerdal
University of California, Davis, CA, USA

AM López-Sobaler
Department of Nutrition and Food Science, Complutense University of Madrid. UCM Research Group VALORNUT-920030, Madrid. Spain

V Loria-Kohen
Department of Nutrition and Food Science, Complutense University of Madrid. UCM Research Group VALORNUT-920030, Madrid. Spain

DJ McNamara
Eggs for Health Consulting, Washington, DC, USA

María Dolores Mesa
Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology "José Mataix", Center of Biomedical Research, University of Granada, Granada, Spain; and Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain

Joshua W Miller
Department of Nutritional Sciences, School of Environmental and Biological Sciences, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

PB Mills
Genetics and Genomic Medicine, Great Ormond Street Institute of Child Health, University College London, London, United Kingdom

DJ Millward
University of Surrey, Guildford, United Kingdom

Donald M Mock
Departments of Biochemistry and Molecular Biology and Pediatrics, Biomedical Research Building I, University of Arkansas for Medical Sciences, Little Rock, AR, United States

TA Mori
School of Medicine and Pharmacology, University of Western Australia, Perth, WA, Australia

Forrest H Nielsen
Research Nutritionist Consultant, Grand Forks, ND, United States

JM Ordovas
Tufts University, Boston, MA, USA

RM Ortega
Department of Nutrition and Foods Science, F. Pharmacy, Complutense University of Madrid, Madrid, Spain

Amanda C Palmer
Department of International Health, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Elizabeth N Pearce
Section of Endocrinology, Diabetes, and Nutrition, Boston University School of Medicine, Boston, MA, United States

D Perrett
Queen Mary University of London, London, UK

Elena Rho
Division of Nephrology, University Hospital Zurich, Zurich, Switzerland

Susan B Roberts
Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, Medford, MA, United States

E Rodríguez-Rodríguez
Department of Chemistry in Pharmaceutical Sciences, Analytical Chemistry, F. Pharmacy, Complutense University of Madrid, Madrid, Spain

A Catharine Ross
Department of Nutritional Sciences, The Pennsylvania State University, University Park, PA, United States

K Sankavaram
Howard University College of Medicine, Washington, DC, USA

Rebekah Schulz
Department of Food Science and Nutrition, University of Minnesota, St. Paul, MN. United States

Harald Seeger
Division of Nephrology, University Hospital Zurich, Zurich, Switzerland

MK Shea
USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA, United States

Rachel E Silver
Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, Medford, MA, United States

HA Simmonds

AP Simopoulos
The Center for Genetics, Nutrition and Health, Washington, DC, USA

Joanne Slavin
Department of Food Science and Nutrition, University of Minnesota, St. Paul, MN. United States

S Stanner
British Nutrition Foundation, London, United Kingdom

Chrysanthi Stylianopoulou
Public Health Nutritionist, Veria, Greece

Sherry A Tanumihardjo
Department of Nutritional Sciences, University of Wisconsin-Madison, Madison, WI, United States

DI Thurnham
Northern Ireland Center for Food and Health, University of Ulster, Coleraine, United Kingdom

Madeline Timm
Department of Food Science and Nutrition, St. Paul, MN, United States

Daniel Tomé
UMR PNCA, AgroParisTech, INRAE, Université Paris-Saclay, Paris, France

Maret G Traber
Linus Pauling Institute, Oregon State University, Corvallis, OR, United States

Rafael Urrialde
Complutense University of Madrid, San Pablo CEU University of Madrid, Iberoamerican Nutrition Foundation, Granada, Spain

John B Vincent
Department of Chemistry and Biochemistry, The University of Alabama, Tuscaloosa, AL, United States

E Weichselbaum
British Nutrition Foundation, London, United Kingdom

Keith P West Jr.
Department of International Health, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Guoyao Wu
Department of Animal Science and Faculty of Nutrition, Texas A&M University, College Station, TX, United States

Robert A Yokel
Department of Pharmaceutical Sciences, University of Kentucky Academic Medical Center, Lexington, KY, United States

S Zidenberg-Cherr
University of California, Davis, CA, USA

MB Zimmermann
Laboratory of Human Nutrition, Institute of Food, Nutrition and Health, ETH Zurich, Zurich, Switzerland

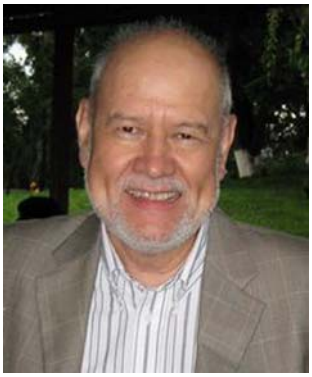
Jean-Marc Zingg
Miller School of Medicine, University of Miami, Miami, FL, United States

This page intentionally left blank

EDITOR BIOGRAPHIES

Editor in Chief

Professor Benjamin Caballero
Emeritus Professor, Bloomberg School of Public Health and School of Medicine, Johns Hopkins University, Baltimore, MD, United States.



Dr. Caballero is Professor Emeritus at the Department of International Health, Bloomberg School of Public Health, with joint appointment at the Department of Pediatrics, School of Medicine, Johns Hopkins University. He obtained his MD from the University of Buenos Aires, Argentina, and his PhD (in neuroendocrine regulation) from MIT, in Cambridge, Massachusetts. He started his academic career at Boston Children's Hospital, Harvard Medical School, and subsequently became the Founding Director of the Center for Human Nutrition at Johns Hopkins University.

Dr. Caballero has focused his research on child nutrition and health in developing countries. In particular, he has explored the combination of undernutrition and overweight that has become increasingly prevalent in low- and middle-income countries.

He is currently a member of the Council of the International Union of Nutritional Sciences. He has served on the Food and Nutrition Board of the US National Academy of Medicine and on a number of expert panels, including the Dietary Reference Intakes Committee, the Expert Panel on Macronutrient Requirements, and the Childhood Obesity Task Force. He was also a member of the U.S. Dietary Guidelines for Americans Advisory Committee, of the Scientific Advisory Board of the Food and Drug Administration, and of advisory committees of the National Institutes of Health and the Department of Agriculture. He is a Fellow of the American Society for Nutrition and of the Royal Society of Medicine (UK), and a member of the Spanish Academy of Nutritional Sciences.

He is the Editor-in-Chief of the *Encyclopedia of Food Sciences and Nutrition*, a 10-volume work on food production, consumption, and biological effects. He is also Editor-in-Chief of the *Encyclopedia of Human Nutrition*, which received the Book of the Year Award from the British Medical Association. His *Guide to Dietary Supplements* summarizes the current scientific basis for the use of mineral and vitamin supplements. He also co-edited a widely used textbook on human nutrition, *Modern Nutrition in Health and Disease*.

Section Editors

Section 1: *The Foundations of Human Nutrition*

Professor Angel Gil

Honorary President of the Iberoamerican Nutrition Foundation (FINUT), Emeritus Professor, Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology, Centre of Biomedical Research, University of Granada, Campus de la Salud, Avda. del Conocimiento, 18100 Armilla, Granada, Spain.



Angel Gil is an Emeritus Professor at the University of Granada and a Full Professor in the Department of Biochemistry and Molecular Biology. He is Doctor Honoris Causa at the Autonomous University of Nuevo León, Monterrey, México, and a visitant Professor of the University of Chile, Santiago de Chile, and University of San Miguel de Tucumán, Argentina. He got a major degree in Biology (1973) and PhD in Biochemistry and Molecular Biology (1978).

Prof. Gil is an internationally recognized authority in Food and Nutrition: His expertise extends from the study of human milk composition to the molecular effects of food bioactive compounds and probiotics and the design and development of novel products for infant and clinical nutrition. He conducted pioneering and innovative research, leading 7 international, 27 national, and more than 50 projects and 120 contracts; he has taught since 1981, supervising more than 50 PhD students.

Prof. Gil has several areas of interest that include evaluating the role of dietary nucleotides in early life and the development of infant nutrition products. Besides, the isolation, identification, and description of the mechanism of action of probiotics and the metabolic, molecular, and genetic factors involved in obesity and the early onset of metabolic syndrome (MS) in childhood; and the design, development, and evaluation of enteral clinical nutrition products. What describes Prof. Gil best is the variety of fields and problems he has faced during his professional carrier and his significant ability to combine his knowledge and expertise in Food Science and Human Biochemistry. This has allowed him to design, develop, innovate, and evaluate exclusive products for Human Nutrition, which are demonstrated in his published articles and his patents' impact.

The multi- and interdisciplinary nature of his work is reflected in the variety of international journals in which he has published 546 articles. Also, he has published 28 books and about 180 book chapters. His five volumes *Treatise of Nutrition*, 3rd Edition, Ed. Medica Panamericana, 2017, with more than 3500 pages, is the "bedside" book for the study of Nutritional Sciences in Spain and all Latin American countries.

He has also been the Chairman of the International Union of Nutritional Sciences (IUNS) 21st International Congress of Nutrition (2013) and the Executive Director of the 23rd International Congress of Nutrition (2017) and has been engaged in the organization of other renowned international congresses. He is a member of prestigious international and national nutrition societies and Honorary President of the Iberoamerican Nutrition Foundation (FINUT), a nonprofit organization promoted by the IUNS, in which the main goal is to contribute to the formation of young scientists in Food and Nutrition in the setting of Iberoamerica. He has received 42 National and International Awards for his contribution to Nutrition and Food Science, among them, the Class Fellow 2022 of the American Society of Nutrition; the Sir David Cuthbertson Lecture Award of the European Society of Clinical Nutrition and Metabolism for scientific achievement in clinical nutrition on 2021; the Award "Granada, City of Science and Innovation" 2021 to the Scientific Career; the Gregorio Marañón Award 2018 to the best Spanish Scientist in the field of Food Science and Nutrition; the Institute Danone Spain Award 2017; the Award of the Spanish Federation of Dairy Industries, 2015; the Nutra Excellence Award 2014, Nutra India Summitt; the UIB Honorary Award of 2013; and the NAOS Strategy Prize 2012, Special recognition for his extensive professional experience in the field of nutrition and obesity, Spanish Ministry of Health, Social Services and Equality (AESAN).

Section 2: *Molecular Mechanisms for the Interaction of Nutrients and Health*

Dr. Noel W. Solomons

Center for the Studies of Sensory Impairment, Aging, and Metabolism (CeSSIAM), Guatemala City, Guatemala, Guatemala



Noel W. Solomons was Assistant and Associate Professor from 1977 to 1984 in the former Department of Nutrition and Food Science of MIT. A resident of Guatemala since 1975, he cofounded the Center for Studies of Sensory Impairment, Aging and Metabolism (CeSSIAM) in 1985, where he has been Scientific Director ever since. He is an Adjunct Professor at Tufts University in both the Friedman School of Nutrition Science and Policy and the Department of Public Health and Community Medicine.

Dr. Solomons received his MD degree from Harvard Medical School and subsequently undertook clinical and research training in infectious diseases and in gastroenterology and clinical nutrition at the Universities of Pennsylvania and University of Chicago, respectively. He was Editor-in-Chief of the *Food and Nutrition Bulletin* from 2016–2020. He has 364 publications indexed on PUBMED. In addition, he has edited 2 books and contributed over 100 articles, reviews, editorials, and commentaries in nonindexed venues and over 50 book chapters. These are dedicated to the scientific and academic interests of his career including: clinical nutrition; human growth and body composition; lactose maldigestion; dietary intake, nutritional status, intestinal absorption, and food fortification related to various micronutrients (vitamins, trace elements, and essential fatty acids); complementary feeding; nutrition in aging and chronic disease; and the interaction of malnutrition and infection. Also, he has supervised doctoral dissertations for 12 PhD candidates from the USA, Canada, Germany, Spain, the UK and the Netherlands through CeSSIAM.

Section 3: *Diet Composition*
Professor Anura Kurpad
Department of Physiology, St John's Medical College, Bengaluru, India



Anura Kurpad works at St. John's Medical College, Bengaluru. He received his MBBS and MD from St. John's, and his PhD from Bangalore University in 1992. He was a Postdoctoral Fellow at the Rowett Research Institute and at Cambridge University, UK, and a visiting Scientist at MIT, Cambridge, United States. He is an elected Fellow of the Royal College of Physicians (London), Fellow of the Indian National Academy of Medical Sciences, Indian Academy of Sciences, and International Union of Nutritional Sciences. His interests are in human, clinical, and public health aspects of nutrition, applied throughout the life cycle. His research focuses on the physiology and clinical aspects of human energy and protein requirements and metabolism, with more recent interests in micronutrient status and metabolism. He is an Associate Editor of *The American Journal of Clinical Nutrition*. He was the President of the Nutrition Society of India from 2012 to 2016 and is the current President of the Asia Pacific Clinical Nutrition Society. He has been, and is, on many national and international advisory bodies, including being the Chair of the Indian Nutrient Requirements Committee.

Section 4: *Disorders Directly Related to Inadequate Nutrient Intake*
Dr. Katherine L. Tucker
Department of Biomedical and Nutritional Sciences, Center for Population Health, University of Massachusetts Lowell, Lowell, MA, United States



Katherine L. Tucker, PhD, is University Distinguished Professor of Nutritional Epidemiology in the Department of Biomedical and Nutritional Sciences, and Director of the Center for Population Health, at the University of Massachusetts Lowell. She holds an adjunct appointment at the University of Massachusetts Medical School. She received her PhD from Cornell University and her undergraduate degree from the University of Connecticut, both in nutritional sciences. Between these degrees, she spent 2 years as a Peace Corps volunteer in the Philippines. Before joining UMass Lowell, she was at the USDA Human Nutrition Research Center on Aging at Tufts University, and McGill University. Dr. Tucker has contributed to more than 450 articles in scientific journals. Her research focuses on dietary intake and risk of chronic disease, including osteoporosis, cognitive decline, obesity, metabolic syndrome, and heart disease, with an emphasis on health disparities. She is the PI of the Boston Puerto Rican Health Study, an ongoing cohort study, to examine the roles of diet, health behaviors, stress, and genetic predisposition in relation to chronic conditions, including heart disease, cognitive decline, and bone health; and is actively involved as a scientific advisor for the NHLBI Jackson Heart Study. She served two terms on the Food and Nutrition Board of the National Academies of Science and Engineering. She is a Fellow of the American Society for Nutrition (ASN), the Gerontological Society of America, and the American Society for Bone Mineral Research and is currently the Editor-in-Chief of *Advances in Nutrition*, the international review journal of the ASN, and Senior Editor of the forthcoming 12th edition of the textbook, *Modern Nutrition in Health and Disease*.

Section 5: *Nutrition in Disease States*
Dr. Paolo M. Suter
Department of Endocrinology, Diabetology and Clinical Nutrition, University Hospital, Zurich, Switzerland.



Paolo M. Suter, MD, MS, is a Professor of Medicine presently affiliated with the Department of Endocrinology, Diabetology and Clinical Nutrition at the University Hospital Zurich (Switzerland). He received his MD at the University of Zurich and an MS in Nutrition at Tufts University (Boston, USA). He specialized in Internal Medicine and was a Faculty Member at the Medical Policlinic of the University Hospital Zurich, where he was directing the well-known Hypertension and Obesity Outpatient Consultation. His research activities focused on vitamin nutriture in the elderly, alcohol metabolism and obesity, blood pressure and hypertension, as well as nutrition and lifestyle in non-communicable disease prevention. During his clinical work he focused on the ideal combination of pharmacological therapy with nonpharmacological strategies and on the therapy and prevention of diseases especially hypertension and obesity. Besides many research publications and review articles he authored a widely used textbook entitled *Checkliste Ernährung* (*Checklist Nutrition*). He served on different national and international boards in the area of his expertise.

Section 6: *Nutrition in the Life Cycle*
Professor Lawrence J. Cheskin
Department of Nutrition and Food Studies, George Mason University, Fairfax, VA, United States
And
Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, United States



Lawrence J. Cheskin, MD, FACP, FTOS, is Professor and Chair of the Department of Nutrition and Food Studies at George Mason University in Fairfax, VA, and an Adjunct Faculty Member of the Department of Medicine at the Johns Hopkins School of Medicine in Baltimore, MD, United States. He received his medical degree from Dartmouth and was a Postdoctoral Fellow in gastroenterology and liver diseases at Yale University-New Haven Hospital. Dr. Cheskin's work is at the intersection of public health nutrition and clinical medicine and has dedicated his career to research, education, and program building in the service of combating obesity, through both treatment and prevention. He founded the Johns Hopkins Weight Management Center, a multidisciplinary clinical research and treatment program. He was also Director of Clinical Research of the Global Obesity Prevention Center at Johns Hopkins and directed its Pilot Studies Core, which evaluated systems-focused proposals worldwide to study such areas as school policies and the built environment in preventing childhood obesity. He is co-PI of the Mason Cohort Study of entering college freshmen and has authored over 240 peer-reviewed journal articles and 8 books. He is committed to mentoring the next generation of public health and clinical scholars dedicated to nutrition and obesity.

Section 7: *Nutrition Epidemiology*
Dr. Manuel Franco
Associate Professor at the Department of Surgery, Medical and Social Sciences, University of Alcalá, Madrid, Spain, and adjunct faculty at the Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States.



Manuel Franco, MD, PhD, Epidemiologist, is an Associate Faculty at the University of Alcalá and an Adjunct Faculty at the Johns Hopkins Bloomberg School of Public Health.

He is the Principal Investigator of the European Research Council Starting Grant Heart Healthy Hoods (http://www.hhhproject.es) studying urban characteristics in relation to eating patterns, physical activity levels, smoking, and alcohol consumption. He is also the PI of the Participatory Project on food in the cities: Photovoice Villaverde (https://youtu.be/VIiFggKzVas).

Manuel Franco trained in Medicine both in Madrid and Berlin. As a Fulbright Scholar he joined the Johns Hopkins Bloomberg School of Public Health Department of Epidemiology to obtain his PhD and Postdoctoral Fellowship in the fields of social epidemiology and urban health.

His work focuses on the prevention of chronic diseases and their major risk factors as nutrition. He has published 94 international peer-reviewed articles and has led 15 studies as PI raising over 2.9 million € in competitive international research bids. His methodological interests include the measurement of urban characteristics related to chronic diseases and nutrition, the use of mixed methods, and the conduction of participatory action research methods.

Section 8: *Global aspects of Human Nutrition*
Professor Daniel J. Hoffman
Department of Nutritional Sciences and director of the Program in International Nutrition at Rutgers University, New Brunswick, NJ, United States



Daniel J. Hoffman, PhD, is a Professor in the Department of Nutritional Sciences and Director of the Program in International Nutrition at Rutgers University, New Brunswick, NJ, United States. He graduated from St. John's University, Collegeville, MN, United States, and received his PhD in Human Nutrition from the Tufts University Friedman School of Nutrition Science and Policy, Boston, MA, United States. For his doctoral training, Dr. Hoffman studied the metabolism of children living in the shantytowns of São Paulo, Brazil. He then completed a Postdoctoral Fellowship at Columbia University, New York, NY, United States, before joining the faculty at Rutgers University.

Dr. Hoffman's research focuses on the long-term impact of undernutrition early in life on body composition and metabolism later in life. This work is complemented by research on sustainable methods to improve diet and growth in lower- and middle-income countries, including Brazil, Mexico, Kenya, South Sudan, and Vietnam. He also directs the Center for Childhood Nutrition Research at the New Jersey Institute for Food, Nutrition, and Health where members conduct interdisciplinary research on childhood growth and development. Dr. Hoffman is an Associate Editor of the *Food and Nutrition Bulletin* and a Section Editor for the *Annals of Human Biology*. His other responsibilities include serving on the editorial board of the *Journal of Nutrition*, serving in leadership positions for the American Society for Nutrition, and advising international and national agencies on matters related to childhood nutrition. Aside from his academic life, Dr. Hoffman is a published documentary photographer whose portfolio includes work on the life of Roma in Europe, urban landscape of São Paulo, Brazil, gentrification of Times Square in New York City, and punk music as the "American Mosaic."

PREFACE

By the middle of last century, the science of nutrition had identified most of the essential nutrients and had provided evidence to propose specific dietary intake recommendations for many diet constituents, with the practical aim of preventing nutrient deficiencies. As chronic, noncommunicable diseases such as cardiovascular disease, cancer, etc., began to emerge as an important causal factor for disability and early death, scientists turned their attention to the potential effects of nutrients and diet patterns on chronic disease risk. Pioneering studies by Burkitt, Keys, Breslow, and others were followed by a large number of studies on the role of dietary patterns and constituents on certain chronic diseases. Many important studies were completed over the second part of the century, providing the evidence to support specific dietary recommendations to reduce disease risk.

The 21st century ushered the next transition in nutrition science, this time centered on the interrelationships between nutrients, dietary patterns, and the human genome. Over the past few decades, advances in our understanding of the human genome and on the molecular tools to explore it have permitted to probe those interactions in increasing detail. In turn, findings from nutrient–gene interaction studies have informed population-wide and clinical and metabolic studies, further advancing our understanding of the effects of diets on human health at the molecular level. This understanding of the links between genotype, phenotype, and nutrient/dietary intake became a key contributor to the emerging area of personalized nutrition/precision medicine.

All those phases of research emphasis, to different degree, continue to exist today and result in a vast, multidisciplinary, ever-expanding amount of information reaching the peer-reviewed literature. This massive amount of information needs to be organized and summarized in a way that makes it accessible to experts, teachers, and, as much as possible, the general public. This has been and continues to be the goal of the *Encyclopedia of Human Nutrition* since its first edition, over 20 years ago.

Such an ambitious task can only be achieved by the collective work of many people. We all have experienced the challenge of writing an article that combines focus and relevance with conciseness, so we are very appreciative of the work of our contributors. Their effort was backed up by an excellent editorial board, which reviewed and provided feedback on every manuscript. Finally, we must acknowledge the outstanding support of the Major Reference Works division at Elsevier. A publication like this *Encyclopedia* has a lot of moving parts, and it is a great privilege to be able to concentrate on the content, knowing that the other parts of the process are in the hands of excellent professionals.

We hope that this book will help satisfy the need for accurate and concise information to the many students and professionals who are committed to use nutrition science as a tool to improve people's quality of life.

Benjamin Caballero, MD, PhD
Editor In Chief

This page intentionally left blank

Aluminum

Robert A. Yokel, Department of Pharmaceutical Sciences, University of Kentucky Academic Medical Center, Lexington, KY, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of R.A. Yokel, Aluminum, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 57–63, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00008-8.

Introduction 1
Properties and natural occurrence 1
Nonfood uses 2
Food uses of aluminum compounds 2
Aluminum in beverages and foods 2
Aluminum in foods from processing, packaging, and storage 3
Exposure and dietary intake 5
Bioavailability and biotransformation 5
Aluminum biokinetics in blood 6
Aluminum tissue deposition and body retention 6
Aluminum excretion 6
Diseases related to aluminum intake 7
Aluminum-induced encephalopathy 7
Aluminum-induced bone disease 7
Aluminum-induced microcytic anemia 8
Evidence for a role in Alzheimer's disease 8
Conclusions 8
References 8
Further reading 9

Key points

- Aluminum is ubiquitous in the environment, beverages and foods
- Foods present the major intake source of aluminum for most humans, ~100-fold more than water
- Governmentally-approved food additives are a major source of aluminum in foods
- The average daily intake of aluminum in foods is ~6 mg, but varies greatly depending on beverages and foods consumed
- Beverage and food preparation and storage in aluminum can increase aluminum content, particularly in acidic or alkaline products or media
- The percentage of aluminum that is taken into the body is very low and most is readily excreted in the urine by people with good renal function

Introduction

Aluminum is a non-essential ubiquitous element, resulting in exposure and some intake into the body from air and many sources humans are exposed to or consume. The major intake source for most people is food. There are many studies quantifying aluminum in various beverages and foods and determining daily intake from them that are summarized herein. This article focuses on the more common exposure sources, their average aluminum level, how the human body handles aluminum exposure and absorbed aluminum, and some adverse effects that have been shown or suggested to result from aluminum exposure and uptake.

Properties and natural occurrence

Aluminum (Al) is a soft (2.5–3 on the Mohs scale), light (2.7 g cm^{-3}, therefore not a heavy metal), ductile, malleable metal. It is the most common metal and the third most common element in the earth's crust (comprising ~ 8%, primarily with silica), and is ubiquitously distributed throughout the environment. It has only one stable isotope, ^{27}Al. It is too reactive to occur in nature as the free metal. Aluminum production from ore requires considerable energy (24 kW h/kg). Recycling requires up to 95% less energy than primary aluminum production. It exists as the trivalent ion (Al^{3+}), therefore does not engage in oxidation–reduction

chemistry. Exposure to water, oxygen, and other oxidants leads to surface formation of Al oxide that provides a 4 nm thick film that has high resistance to corrosion and is virtually insoluble from pH 4.5 to 8.5. In aqueous solution, in the absence of bound ligands, its chemical form (species) is pH-dependent. The Al^{3+} ion with 6 waters of hydration predominates below pH 5.5, as $Al(H_2O)_6^{3+}$. As the pH increases $Al(OH)^{2+}$, $Al(OH)_2^{+}$, $Al(OH)_3$, and $Al(OH)_4^{-}$ predominate at pH 5.5, 6, 6.2, and above 6.2, respectively (Harris et al., 1996). It is least soluble as $Al(OH)_3$ (gel) at pH 6.2. As a Lewis acid (electron pair acceptor, electrophile) it strongly complexes with multidentate carboxylate- and hydroxyl-/keto-containing ligands, for example citrate, through noncovalent interactions, generally involving ionic or electrostatic bonds. Kinetically aluminum ion reactions are slow. The mean residence time of water in $Al(H_2O)_6^{3+}$ is ~1 s. Water replacement by citrate has a complexation half-life of ~1–2 min. The chemical species of aluminum has a great impact on its kinetics and effects. The Al^{3+} and Fe^{3+} ions, having the same valency and similar ionic radii, engage in some of the same chemical and metabolic processes. Aluminum binds particularly strongly to phosphates, giving it the potential to bind with deoxyribonucleic acid, adenosine triphosphate, and many other biomolecules.

In contrast to its abundance in the earth's crust, natural waters contain little dissolved aluminum. Its average concentration in lakes, rivers, ground water, coastal sea water, and open ocean water is ~150, 400, 100, 1 to 7, and 1 μg L^{-1}, respectively. This is a result of the low solubility of some aluminum salts and aluminum hydroxide deposition in sediments. Acidification, e.g., by acid rain and mine runoff, can solubilize aluminum and raise its concentrations in acidified lakes and rivers, producing toxicity to aquatic organisms. Airborne aluminum from natural (soil-derived) and anthropogenic sources ranges from <1 ng/m^3 in remote sites to >1 μg/m^3 in industrialized locations. The World Health Organization (WHO) has a 'practicable level' in finished (drinking) of $\leq$0.1 and $\leq$ 0.2 mg L^{-1} for large and small water treatment facilities (WHO, 2004). The European Union (EU) "Indicator Parameter" directive for aluminum in drinking water is 0.2 mg L^{-1} (EU, 2020). The US Environmental Protection Agency has a secondary, non-enforceable standard of 0.05–0.2 mg L^{-1}, based on esthetic (taste, sight, and smell) properties (US EPA, 2021). Canada set 2.9 mg L^{-1} as the maximum acceptable concentration and 0.1 mg L^{-1} as the operational guidance for total aluminum in drinking water (Health Canada, 2019). The Joint Food and Agriculture Organization (FAO) of the United Nations/WHO Committee on Food Additives established a provisional tolerable weekly intake of 2 mg per kg body weight for aluminum from all aluminum compounds in food including food additives (FAO/WHO, 2011). The Agency for Toxic Substances and Disease Registry Minimal Risk Level for intermediate (15–364 days) oral aluminum intake was set at 1 mg kg^{-1} day^{-1} (ATSDR, 2020). A tolerable weekly intake of 1 mg kg^{-1} body weight was established by the European Food Safety Authority (Aguilar et al., 2008).

Nonfood uses

Aluminum and its compounds are widely utilized. They are used in building and construction, transportation, packaging, containers, water purification, sugar refining, brewing, paper production, glass, ceramics, rubber, abrasives, furnace linings, as wood preservatives, to waterproof textiles, and as a flame retardant. In personal care and medical applications aluminum compounds are in antacids (that are also used as phosphate binders); buffered aspirin products; non-spray and spray antiperspirants; lipsticks as lakes (colorants, salts of water-soluble artificial colors adsorbed onto alumina); skin care products; toothpastes as an abrasive and to reduce dental hypersensitivity; dental rinses; products for dermatitis such as athlete's foot as an astringent and antibacterial; topical products as a protectant against diaper dermatitis and in anorectal disorders; antidiarrheal products; vaginal douches; and anorectal preparations as a keratolytic. It is used an intravesical irrigant to treat urinary bladder hemorrhage and as an adjuvant in vaccines to increase their antigenic properties. Aluminum's mechanism as an adjuvant includes its up-regulation of gene expression related to the innate and adaptive immune response in human monocytes (Vrieling et al., 2020).

Food uses of aluminum compounds

Some United States Food and Drug Administration (FDA) and EU approved aluminum compounds that may be employed as food additives are listed in **Table 1**. These contribute to the high concentrations of aluminum in some of the foods in **Tables 2** and **3**. Other governments also allow the use of aluminum as a food additive.

Aluminum in beverages and foods

Tables 2 and **3** show average values from studies published since 1985 for which there were multiple reported values for each entry in the tables. The entered values are medians. This measure of central tendency was selected because there is considerable variability of reported results from some beverages and foods. This variation may result from poor analytical techniques, inadequate removal of soil and/or contamination, and differences due to climate and soil growing conditions.

Plants that accumulate aluminum deposit it in their leaves. The tea plant accumulates and concentrates aluminum, producing a higher aluminum concentration than in other plant-derived beverages. Addition of the anticaking agent sodium aluminosilicate to nondairy creamer and salt, particularly single-use packets, accounts for their high aluminum concentration. The higher aluminum content of soy-than milk-based infant formula derives from the high aluminum content of beans, including soybeans, compared to cow milk.

Table 1 Some FDA- and EU-approved food additives containing aluminum, or aluminum agents that come in contact with foods, and their uses. The US Code of Federal Regulations title, part, and section numbers are cited in round brackets. The EU No. is in square brackets.

- Aluminum calcium silicate (a.k.a.: calcium aluminum silicate)—Anti-caking agent, ≤2% by weight in table salt (21CFR182.2122) [556].
- Sodium calcium aluminosilicate hydrated (a.k.a.: sodium calcium silicoaluminate)—Anti-caking agent, ≤2% (21CFR182.2729).
- Sodium aluminosilicate (a.k.a.: sodium silicoaluminate)—Anti-caking agent, ≤2% in dried whole eggs and egg yolks and grated cheeses (21CFR182.2727) [554].
- Aluminum ammonium sulfate—Buffer and neutralizing agent (21CFR182.1127) [523].
- Aluminum potassium sulfate—Buffer and neutralizing agent (21CFR182.1129) [522].
- Aluminum sodium sulfate—Buffer and neutralizing agent (21CFR182.1131) [521].
- Sodium aluminum phosphate (SALP)—acidic SALP as a leavening agent in self-rising flours and meals, basic SALP as an emulsifying agent in cheeses. Acidic: (21CFR182.1781) [541] Basic: (21CFR182.1781)
- In lakes, on a substratum of alumina (21CFR82.101, 102, 203, 304, 340, 705, 706) [102, 110, 129, 132, 133].
- Aluminum borate as an antistatic and/or antifogging agent in food-packaging materials, ≤2% by weight of polymer film (21CFR178.3130).
- Aluminum-containing clarifying agents for polypropylene and polypropylene copolymers intended for use in contact with food, ≤0.25% by weight of the copolymer (21CFR178.3295).

Table 2 Aluminum concentrations of beverages and nondairy creamer.

Beverage	*Al concentration (median; as mg L^{-1}, except for non-dairy creamer)*
Beer	0.10
Coffee	0.24
Non-dairy creamer—multiple serving container	26 mg kg^{-1}
Non-dairy creamer—single serving packet	170 mg kg^{-1}
Apple juice	0.40
Orange juice	0.19
Pineapple juice	0.36
Tomato juice	0.88
Wine	0.81
Distilled spirit	0.42
Cola	0.27
Black tea	2.6
Green tea	2.6
Herbal tea	0.16
Tap water	0.045
Mineral and spring water	0.020
Cow milk	0.094
Human milk	0.030
Cow-based infant formula	0.019
Soy-based infant formula	0.76

Even though soil contains 3–10% aluminum, most plants and plant-eating animals contain little aluminum, due to its low oral bioavailability (see **Bioavailability and biotransformation** below). Most unprocessed foods contain <5 mg kg^{-1}. The addition of aluminum as a food additive can greatly increase its food content, illustrated by baking powder, cake and pancake mixes, cakes and pancakes, and cheese containing sodium aluminum phosphate (SALP). Residual dirt may increase the aluminum concentration in lettuce and spinach. Some spices contain considerable aluminum, illustrated by paprika and pepper. Spice aluminum values are expressed based on their dry weight, whereas most other food aluminum concentrations are based on wet weight.

Aluminum in foods from processing, packaging, and storage

Food preparation and storage in contact with aluminum can increase the food's aluminum content. Some examples follow. Coffee brewed in an aluminum pot had about twice as much aluminum as coffee brewed in stainless steel. Tea prepared in an aluminum container had 3.4-fold more aluminum than prepared in glass. Water and milk boiled/cooked in aluminum cookware had more aluminum than when prepared in porcelain, stainless steel, Teflon, or enamel cookware. In one study multiple vegetables cooked in old or new aluminum vessels had 40% and 80% more aluminum than when cooked in stainless steel. Cooking matooke, potatoes, beans, cabbage, dodo, tomatoes, kadhi, sambar, mutton curry, and chicken curry in aluminum increased their aluminum content. However, studies have not consistently shown higher aluminum in foods prepared in new versus old aluminum vessels, perhaps

Table 3 Concentrations of aluminum in foods.

Food	*Al concentration (median; as mg kg^{-1})*
Beef meat	1.2
Pork meat	2.5
Chicken meat	1.0
Bacon	1.0
Ham	1.3
Meat sausages and frankfurters	3.4
Fish	0.18
Cephalopods (octopus, squid, cuttlefish)	3.0
Crustaceans (crabs, lobsters, crayfish, shrimps, prawns)	5.6
Eggs	0.22
Nuts	4.1
Fruits, fresh	1.1
Raisin	10
Tomato	0.5
Corn	1.25
Beans	5.8
Peas	2.0
Mushrooms	4.0
Potato	2.6
Lettuce	5.4
Spinach	24
Wheat flour	5.1
Oats, oatmeal	2.1
Rice	2.1
Baking powder	70
Biscuits	22
Cake mix	445
Cakes (not stated to contain SALP)	6.3
Cakes (that contain SALP)	190
Cookies	8.0
White bread	2.3
Wheat bread	4.6
Breakfast cereal	1.8
Pancakes	179
Pasta	3.7
Cheese, not from sheep or goat milk	2.3
Cheese from goat milk	15
Frozen pizza cheese	415
Restaurant pizza cheese	2.9
Yogurt, not from goat milk	0.30
Yogurt from goat milk	5.0
Soup	1.2
Butter	1.4
Margarine	1.7
Olive oil	0.043
Sugar	1.7
Honey	0.5
Chocolate	9.4
Jellies and jams	4.1
Paprika	92
Pepper	31
Pickles	7.4
Vinegar	0.21
Salt	2.4
Salt—single serving packet	180

due to differences in vessel manufacture and surface coating, extent of use or cleaning before their study, and the studied food. Tomato cooked for 10 min in an aluminum pan had ~60% more aluminum than cooked in a steel pan. Cottage cheese and yoghurt prepared in aluminum containers had more aluminum than when prepared in boron glass or steel containers. Rhubarb boiled for 12 min in a steel pot had 0.8 and in an aluminum pot 13–20 mg L^{-1} aluminum. The aluminum concentration of black currant juice boiled for 90 min in an aluminum vessel increased from 0.05 to 24 mg L^{-1} when sugar was added before boiling and to 56 mg L^{-1} when the sugar was added after it was brought to a boil.

Household aluminum foil is 0.18 mm thick. Beef, mutton, pork, water buffalo, chicken, and turkey baked in aluminum foil had on average a 2.5-fold increase in the aluminum content. Higher temperature (250 vs. 200 vs. 150 °C) contributed more aluminum to beef, mutton, and pork than longer baking time (20 vs. 40 vs. 60 min) whereas longer time, not temperature, increased aluminum in chicken and turkey. Baking and grilling fish in aluminum foil increased its aluminum concentration by five-fold. Potatoes roasted in aluminum foil had 2.5-fold more aluminum than when roasted in stainless steel.

Tap water storage in an aluminum container increased its aluminum concentration. The aluminum concentration of milk stored in an aluminum pan increased ~14-fold. Milk stored in an aluminum-lined container had more aluminum than when stored in plastic or glass. The aluminum concentration in beer stored in aluminum cans was consistently higher (average 60%) than in steel cans or glass bottles, did not increase over time when stored at 5 °C, but increased ~ 10%/month when stored at room temperature. Wine stored for 2 years in an aluminum can had more aluminum than when stored in a glass bottle or steel can. Frozen salmon commercially marketed in aluminum foil had more aluminum than when prepared without. The aluminum concentration in apple, lemon, and orange (but not tropical fruit) juice stored in aluminum cans increased over time and the aluminum concentration in grapefruit juice stored in aluminum cans increased with increased storage temperature. Soft drinks (colas, Fanta orange, and soda water, but not Pocari Sweat or orange squash) stored in aluminum cans had a higher aluminum content than when stored in plastic or glass. The aluminum concentration of lemon, orange, and cola drinks stored in aluminum cans for 12 months increased ~ 1-, 2- and 3-fold monthly above the original aluminum concentration, respectively. Tomato paste stored in steel and aluminum cans increased ~25 and 60% over 2.5 years, respectively. Storage of tomato puree and tamarind in an aluminum pan for 72 h increased its aluminum content 31- and 17-fold, respectively. Addition of salt to the tomato and sugar to the tamarind increased their aluminum content an additional ~15 and 9%, respectively. Aluminum mobilization during food preparation and storage in contact with aluminum is greatest for acidic foods (e.g., rhubarb pH ~ 3.2 and tomato ~ 4.5), consistent with the chemical species of aluminum as a function of pH; solubility increases below and above pH 6.2.

Exposure and dietary intake

Dietary intake provides the greatest percentage of aluminum intake for the typical human. Daily average drinking water consumption is estimated to be 1.4 L. Intake of water with a median aluminum concentration of 0.045 mg L^{-1} (**Table 2**) would provide ~ 0.06 mg Al daily. Ninety studies conducted since 1970 in 30 countries reported dietary aluminum intake based on total diet studies, market basket surveys, dietary records, calculations based on food aluminum levels, and duplicate diets and portions. Median daily aluminum intake of adolescents and adults was 5.1 mg, with 75% of the reported intakes between 2.5 and 13 mg/day. When reported as mg kg^{-1}/week the median intake was 0.6. Reported intakes in China were ~2-fold greater than the US which was ~2-fold greater than Europe and Japan. The greater aluminum intake in the US may be due to a higher utilization of SALP and other approved aluminum food additives in processed foods. Aluminum intake in food is ~100-fold greater than from drinking water (5.1 vs. 0.06 mg daily). As tea contains more aluminum than other beverages, its consumption can contribute 50% of the daily aluminum intake in those who consume considerable amounts of this beverage when other sources do not also have large amounts of aluminum.

Bioavailability and biotransformation

Aluminum bioavailability is low. The main uptake routes are by inhalation, through the gastrointestinal tract, and medical routes such as injections and during dialysis. Absorption from the gastrointestinal tract, lungs, and after underarm (presumably transdermal) application is ~0.1–0.3%, 1.5–2%, and 0.01%, respectively. Aluminum injected in vaccines is slowly absorbed as it dissolves and may ultimately reach 100%. Aluminum absorption from the gastrointestinal tract appears to be primarily in the distal intestine. There is evidence supporting several mechanisms of intestinal aluminum absorption, including paracellular diffusion, an interaction with calcium uptake, and sodium transport processes. Owing to aluminum's ubiquitous presence, which creates measurable levels in all tissues and fluids and potential analytical contamination problems, and its low bioavailability, determination of its pharmacokinetics, including bioavailability, at relevant exposures is very difficult using ^{27}Al. More precise measurements have been made using ^{26}Al. Owing to its high cost and very long half-life (717,000 years), its decay rate is too low for practical application as a radioisotopic tracer. Its quantification by accelerator mass spectrometry is exquisitely sensitive. The limit of detection is ~1,000,000 atoms (4×10^{-17} g), enabling a lower limit of quantitation in blood and urine of ~1×10^{-16} g mL^{-1}. This enables ^{26}Al administration to human subjects of doses that present no significant radiation risk and the conduct of exposure dose-relevant pharmacokinetic studies. Studies conducted in rats have shown oral bioavailability from water, tea, acidic SALP in a food (biscuit), and basic SALP in cheese to be ~0.3%, 0.4%, 0.1%, and 0.2%, respectively (Yokel and Florence, 2008). Studies

in humans showed absorption from drinking water and aluminum hydroxide to be ~0.2% and 0.1%, respectively. Oral aluminum absorption is increased by citrate, and other carboxylic acids to a lesser extent. Oral aluminum bioavailability has also been shown to be greater for more soluble aluminum species and increased by fluoride; low iron, calcium, or sodium status; and uremia. In contrast, silicon-containing compounds appear to reduce its gastrointestinal absorption and enhance its urinary elimination as hydroxyaluminosilicate species. Some results suggest oral aluminum absorption is increased in Alzheimer's disease (AD) and Down's subjects.

Aluminum biokinetics in blood

Median plasma aluminum concentration in several normal human studies was ~3 μg L^{-1}. As this is very low and aluminum is ubiquitously present, much attention must be paid to collection, storage, and blood sample handling to avoid contamination. At equilibrium serum and red blood cell aluminum concentrations are approximately equal. Aluminum's volume of distribution is initially equal to the blood volume, consistent with its equal concentration in serum and blood cells. The volume of distribution then increases over time as it distributes and accumulates in tissues. Within blood plasma, ~91% is bound to transferrin, an iron-transport protein, ~7–8% is bound to citrate, and the remainder to phosphate and hydroxide. Formation of the aluminum transferrin- and aluminum citrate-complexes occurs within minutes. The strength of the binding (stability constant) for Al^{3+} with ligands is 3–4 orders of magnitude lower compared to Fe^{3+}. Consequently, aluminum will not displace iron from transferrin or most other ligands. Transferrin-receptor-mediated endocytosis may mediate aluminum uptake into the brain and other organs. Citrate enhances aluminum distribution out of the blood and into tissues and its renal clearance (by increasing aluminum's filterable fraction), suggesting a mechanism for distribution of aluminum citrate that is different from that mediating aluminum transferrin.

Aluminum tissue deposition and body retention

Aluminum distributes unequally throughout the body in normal and aluminum-intoxicated humans. Because of its low bioavailability by most routes of exposure and its effective clearance from blood by the kidney (see **Aluminum excretion** below), aluminum concentrations in the human are low compared to most exposure sources. The normal human body burden is estimated to be 30–50 mg aluminum. Steady state tissue aluminum concentrations in normal adults are (in mg kg^{-1} wet weight unless stated otherwise): lung: 20, bone: 1–3 mg kg^{-1} dry weight, liver and spleen: 1, kidney: 0.5, heart: 0.45, muscle: 0.4, brain: 0.35, and blood: 0.002. Approximately 50, 25, 11, 10, 3, 1, 0.3, 0.25, and 0.2% of the aluminum body burden is in the skeleton, lung, muscle, skin, liver, brain, heart, kidney, and spleen, respectively. Aluminum localizes at the bone mineralization front and in osteoid. Lung aluminum may be from airborne environmental particles, occupational exposure, and distribution from the blood. Insoluble particles trapped in the lung are very slowly cleared. Blood and tissue concentrations are higher in uremia and higher still in dialysis encephalopathy. Approximately 80% of an intravenous dose of aluminum citrate was excreted within a week by humans, suggesting the remainder was retained within the body, some of which was excreted over a long time. This finding indicates that under conditions of continuous intake, aluminum accumulates in the body, even in subjects with normal renal function. Human brain, bone, other organ, and serum aluminum concentration increase with age. Hair aluminum concentration as an indicator of aluminum body burden has not been validated. Problems include the lack of standardized hair collection procedures.

Aluminum clearance is characterized by multiple half-lives that are estimated in hours, days, and years, suggesting multiple compartments. The initial half-life (distribution from the vascular compartment) is a few hours. The brain, bone, liver, and kidney aluminum half-life in rats was estimated to be 150 to 1635, 520, 430 and 400 days, respectively. Allometric scaling from this animal species with a lifespan of 2 years to the human with an 80-year life span suggests the aluminum half-life in human brain might be decades or even greater than the lifespan. Estimates of the terminal half-life in the human suggest it may be as great as 50 years, attributed to an aluminum depot in bone. The slow release of aluminum from bone, due to the low rate of bone turnover (~10% per year), may account for the prolonged half-life seen in most organs, including the brain, and prolonged elevated urinary aluminum excretion after significant occupational aluminum exposure. Increased bone turnover rates compared to the healthy adult, as seen in children who exhibit high rates of bone growth and turnover or geriatrics, may accelerate bone aluminum release.

Biological monitoring of aluminum exposure can be conducted with urine, which is thought to indicate recent exposure, and plasma or serum, which is thought to better reflect the aluminum body burden and long-term exposure. However, neither indicates the aluminum body burden well, which is better estimated by bone aluminum, the desferrioxamine challenge test, or combined measurement of serum parathyroid hormone and the desferrioxamine test. Desferrioxamine chelates/complexes with aluminum, iron, and other trivalent metals, and increases their urinary elimination.

Aluminum excretion

Kidneys provide the primary route of aluminum elimination, ~95%, presumably by glomerular filtration of aluminum citrate. Humans who consume the typical daily dietary amount of aluminum would be expected to excrete 4–12 mg daily, producing a urinary aluminum concentration of 2–10 mg L^{-1}. However, several industrial studies reported higher values in controls (often

20 mg L^{-1}) and even higher in exposed workers. Reduced or lack of renal function creates the risk of aluminum accumulation and toxicity. Bile (feces) accounts for most of the remaining excreted aluminum, although aluminum is excreted in saliva, sweat, and semen.

Diseases related to aluminum intake

There is no good evidence that aluminum is an essential element for the human or any other plant or animal species. Aluminum toxicity has been extensively reviewed by the WHO, for the US Department of Health and Human Services, and most extensively by a multinational group led by Daniel Krewski (see Further reading). Occupational exposure to high aluminum levels has been associated with lung fibrosis from stamped aluminum powder, asthma from aluminum salts and welding fumes, and increased oxidative stress. Contact dermatitis has been reported. It has been suggested that aluminum can contribute to Crohn's disease. Support was provided by the observation that aluminum increased release of lipopolysaccharide and fragilysin (that contribute to inflammation and diarrhea) by a bacterium (*Bacteroides fragilis*) that resides in the intestine (Alexandrov et al., 2020). In contrast, aluminum altered the expression by cultured human colorectal cells of some cellular oxidative stress and xenobiotic metabolism genes but only at concentrations exceeding those exceeding the exposure from daily intake of 5–10 mg aluminum (Sieg et al., 2020). Neurobehavioral toxicity has been seen in aluminum welders with urine aluminum >100 mg L^{-1} (Yokel and Sjögren, 2021). When hemodialysis was initially extensively used, some patients developed a progressive encephalopathy that was fatal within 6 months (see **Aluminum-induced encephalopathy**, below). Exposure to lower levels of aluminum than those that caused aluminum-induced encephalopathy can produce bone disease (see **Aluminum-induced bone disease**, below) and anemia (see **Aluminum-induced microcytic anemia**, below). It has been suggested that aluminum, from vaccines, is a contributor to autism spectrum disorder. Meta-analyses and reviews of a relationship between autism spectrum disorders and aluminum in hair, urine, and/or blood and vaccination receipt found either no or inconsistent support. A review by a US Institute of Medicine of the National Academies panel concluded the evidence rejected a causal relationship between the measles, mumps, and rubella vaccine and autism (Yokel and Sjögren, 2021).

Premature infants who are fed intravenously, because they do not tolerate oral feeding, are a high-risk group for aluminum toxicity. Parenteral nutrition feeding solutions given intravenously can contain significant aluminum. This is primarily associated with calcium gluconate and phosphate parenteral nutrition components, which complex aluminum from glass storage containers. Use of plastic contains largely avoids this problem. As this feeding solution is given intravenously, therefore 100% bioavailable, to premature infants whose kidneys are not fully matured and therefore less able to excrete the aluminum, they are at risk of sufficient aluminum accumulation to develop metabolic bone disease, cholestatic hepatitis, and reduction of mental development. To address this concern the US FDA adopted a labeling requirement for aluminum in large and small volume parenterals used to prepare total parenteral nutrition solutions.

Studies in mice, rats, rabbits, and dogs have shown aluminum has the potential to produce neurobehavioral toxicity, affect the male reproductive system, produce embryotoxicity, and affect development. Based on these concerns, the FAO/WHO established the 2 mg kg^{-1} body weight as a provisional tolerable weekly intake of all aluminum compounds in food, noted above.

Aluminum-induced encephalopathy

Encephalopathy has been associated with the use of hemo- and peritoneal-dialysis fluids that contain significant amounts of aluminum, high dose administration of oral aluminum salts to renal-impaired humans to bind phosphate in the gastrointestinal tract as insoluble aluminum phosphate, alum instillation into the urinary bladder to stop hemorrhaging, and neurosurgical implants of aluminum-containing biomaterials. Dialysis (associated) encephalopathy (also known as: dialysis dementia) is associated with elevated brain aluminum and serum aluminum >80 mg L^{-1}. Renal dialysis patients are highly susceptible to aluminum accumulation and toxicity from aluminum in dialysis fluids because aluminum can diffuse across the dialysis membrane, it rapidly and very strongly binds to transferrin in the blood, and they lack the primary route of aluminum elimination, renal function. With recognition of the role of aluminum in dialysis-related toxicity, reduction of aluminum exposure guided by the US Association for the Advancement of Medical Instrumentation's recommendation that water used to prepare dialyzate solution contain <10 mg L^{-1}, and use of alternatives to aluminum as an oral phosphate binder, occurrences of aluminum-induced encephalopathy in dialysis patients have become unusual.

Aluminum-induced bone disease

Aluminum-induced low-turnover bone disease is manifest as osteomalacia. Aluminum-induced bone disease in dialysis patients is seen when serum aluminum is > 30 mg L^{-1} and when a stain shows aluminum at 30% of the trabecular bone surface. Aluminum interferes with parathyroid hormone synthesis and function. In osteomalacia, aluminum accumulates at the mineralization front of the bone surface, where it disrupts bone mineralization by inhibiting calcium accretion and osteoblast activity, increases bone resorption, and results in an increase of the nonmineralized bone, leading to painful fractures. Children and adults who received

long-term parenteral nutrition with high aluminum content had bone pain and fractures, and lower bone mass, mineral content, and mineral density than controls.

Aluminum-induced microcytic anemia

A microcytic, hypochromic anemia is associated with elevated plasma aluminum in chronic renal failure patients. Aluminum interferes with the development of hematopoietic progenitor cells and is thought to interfere with heme synthesis. Geophagy contributes to anemia. Up to 5 mg kg^{-1} of the chelator desferrioxamine once or twice a week has been shown to be safe and effective for long-term treatment of aluminum overload, and reduction of aluminum-induced encephalopathy, bone disease, and anemia.

Evidence for a role in Alzheimer's disease

Aluminum has been implicated in the etiology of AD. Hallmark neuropathological signs include neurofibrillary tangles (NFTs), senile plaques (SP), and cerebrovascular amyloid. Early-onset AD usually has a familial link, due to gene mutations which result in increased secretion of neurotoxic amyloid beta (Aβ) protein. No specific gene mutations have been associated with late-onset AD which accounts for ~95% of AD cases. The lack of identified hereditary links for the majority of AD cases suggests environmental factors may interact with other factors to cause this disease. Aluminum is one of the suggested environmental contributors. The genesis of the hypothesis that aluminum plays a role in the etiology of AD was an observation reported in 1965 of neurofibrillary degeneration in rabbit brain after intracerebral aluminum injection, which resembled, but was not identical to, the NFTs of AD. Similarly, the neuropathology in dialysis encephalopathy is different from that seen in AD. The observation of elevated aluminum in post-mortem human AD brain samples reported in 1973 was interpreted as suggesting a role for aluminum in AD. This was followed by many studies, some of which found a few-fold or less increased concentration of aluminum in AD victim brains than in controls, and some which did not. Studies investigating an elevated level of aluminum in AD brain using microprobe techniques which can quantify aluminum within a cell, NFT, or SP, as well as aluminum-selective stains have also produced mixed results. Meta-analysis found significantly elevated levels of aluminum in brain, serum, and cerebrospinal fluid. Elevated brain aluminum in AD does not demonstrate cause and effect. The neuronal degeneration of AD may result in accumulation of metals, such as aluminum. Elevated aluminum may play a role in AD development.

Another approach to address the potential role of aluminum in AD is the epidemiological study of the association between the aluminum concentration in drinking water and AD incidence, comparing geographic regions where drinking water aluminum concentrations differ with AD incidence. The results of many such studies are not consistent. Drinking water is not the major source of aluminum for the general population, as noted in **Exposure and dietary intake**, above. The only published epidemiological study of AD and aluminum consumption from food was too small to draw firm conclusions, although it found a significant odds ratio for higher association of AD with a food category typically high in aluminum (pancakes, waffles, biscuits, muffins cornbread, and corn tortillas).

Aluminum has been shown to produce many effects in the brain that have been suggested as contributors to its neurotoxicant effects, and contributing to AD. These have included promotion of the formation and accumulation of Aβ and aggregation of hyperphosphorylated tau protein, deficits of cortical cholinergic neurotransmission, increased inflammation and oxidative injury, and impaired cognitive function. The controversy of a contributory role of aluminum to AD has not been resolved. As there are many contributing factors to AD, it will be extremely difficult to demonstrate the lack of contribution of any one factor, such as aluminum, to this disease.

Conclusions

Aluminum exposure and uptake is unavoidable. Choices of the products one uses and beverages and foods one consumes and how they are prepared and stored can impact one's aluminum intake. The main source of aluminum exposure for most people is food. Approved food additives and preparation and storage in aluminum increase aluminum content. People who have good renal function are not at risk of any proven adverse effects from typical aluminum exposure. Introduction of aluminum directly into the blood or brain can produce toxicity, particularly in people with reduced or no renal function.

References

Aguilar, F., Autrup, H., Barlow, S., et al., 2008. Safety of aluminium from dietary intake. Scientific opinion of the panel on food additives, flavourings, processing aids and food contact materials (AFC). EFSA J. 754, 1–34.

Alexandrov, P.N., Hill, J.M., Zhao, Y., Bond, T., Taylor, C.M., Percy, M.E., Li, W., Lukiw, W.J., 2020. Aluminum-induced generation of lipopolysaccharide (LPS) from the human gastrointestinal (GI)-tract microbiome-resident *Bacteroides fragilis*. J. Inorg. Biochem. 203, 110886.

ATSDR, 2020. Minimal Risk Levels (MRLs) for Hazardous Substances. https://wwwn.cdc.gov/TSP/MRLS/mrlsListing.aspx.

EU, 2020. Directive (EU) 2020/2184 of the European Parliament and of the Council of 16 December 2020 on the Quality of Water Intended for Human Consumption. https://eur-lex.europa.eu/eli/dir/2020/2184/oj.
FAO/WHO, 2011. Joint FAO/WHO Expert Committee on Food Additives. Summary Report of the Seventy-Fourth Meeting of JECFA, Evaluation of Certain Food Additives and Contaminants, Tolerable or Acceptable Daily Intakes, Other Toxicological Information and Information on Specifications. http://www.fao.org/3/at873e/at873e.pdf.
Harris, W.R., Berthon, G., Day, J.P., Exley, C., Flaten, T.P., Forbes, W.F., Kiss, T., Orvig, C., Zatta, P.F., 1996. Speciation of aluminum in biological systems. J. Toxicol. Environ. Health 48, 543–568.
Health Canada, Water and Air Quality Bureau, Healthy Environments, Consumer Safety Branch, 2019. Guidelines for Canadian Drinking Water Quality Summary Table. https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/ewh-semt/alt_formats/pdf/pubs/water-eau/sum_guide-res_recom/sum_guide-res_recom-eng.pdf.
Sieg, H., Krause, B.-C., Kaestner, C., Böhmert, L., Lichtenstein, D., Tentschert, J., Jungnickel, H., Laux, P., Braeuning, A., Fessard, V., Thünemann, A.F., Luch, A., Lampen, A., 2020. Cellular effects of in vitro-digested aluminum nanomaterials on human intestinal cells. ACS Appl. Nano Mater. 3, 2246–2256.
US EPA, 2021. Secondary Drinking Water Standards: Guidance for Nuisance Chemicals. https://www.epa.gov/sdwa/secondary-drinking-water-standards-guidance-nuisance-chemicals.
Vrieling, H., Kooijman, S., De Ridder, J.W., Thies-Weesie, D.M.E., Soema, P.C., Jiskoot, W., Van Riet, E., Heck, A.J.R., Philipse, A.P., Kersten, G.F.A., Meiring, H.D., Pennings, J.L., Metz, B., 2020. Activation of human monocytes by colloidal aluminum salts. J. Pharmaceut. Sci. 109, 750–760.
WHO, 2004. Guidelines for Drinking-Water Quality Geneva. World Health Organization. https://www.who.int/water_sanitation_health/dwq/GDWQ2004web.pdf.
Yokel, R.A., Florence, R.L., 2008. Aluminum bioavailability from tea infusion. Food Chem. Toxicol. 46, 3659–3663.
Yokel, R.A., Sjögren, B., 2021. Aluminum. In: Nordberg, G., Costa, M. (Eds.), Handbook on the Toxicology of Metals, fifth ed. Elsevier B.V.

Further reading

Agency for Toxic Substances and Disease Registry, 2008. Toxicological Profile for Aluminum. US Department of Health and Human Services, Public Health Service, Atlanta. http://www.atsdr.cdc.gov/ToxProfiles/tp22.pdf.
Klotz, K., Weistenhöfer, W., Neff, F., Hartwig, A., van Thriel, C., Drexler, H., 2017. The health effects of aluminum exposure. Deutsches Ärzteblatt International 114, 653–659. https://pubmed.ncbi.nlm.nih.gov/29034866/.
Krewski, D., Yokel, R.A., Nieboer, E., Borchelt, D., Cohen, J., Harry, J., Kacew, S., Lindsay, J., Mahfouz, A.M., Rondeau, V., 2007. Human health risk assessment for aluminium, aluminium oxide, and aluminium hydroxide. J. Toxicol. Environ. Health B Crit. Rev. 10 (Suppl. 1), 1–269. https://pubmed.ncbi.nlm.nih.gov/18085482/.
Nordic Expert Group for Criteria Documentation of Health Risks from Chemicals and the Dutch Expert Committee on Occupational Safety, 2011. Aluminium and Aluminium Compounds, vol. 46. ARBETE OCH HÄLSA, p. 142. https://gupea.ub.gu.se/handle/2077/28447.
Pennington, J.A.T., 1988. Aluminum content of foods and diets. Food Addit. Contam. 5, 161–232. https://www.tandfonline.com/doi/pdf/10.1080/02652038809373696.
Tietz, T., Lenzner, A., Kolbaum, A.E., et al., 2019. Aggregated aluminium exposure: risk assessment for the general population. Arch. Toxicol. 93, 3503–3521. https://pubmed.ncbi.nlm.nih.gov/31659427/.
World Health Organization, 1997. Environmental Health Criteria 194: Aluminium. World Health Organization, International Programme on Chemical Safety, Geneva. http://www.inchem.org/documents/ehc/ehc/ehc194.htm.
Yokel, R.A., 2012. Aluminum in food—the nature and contribution of food additives. In: El-Samragy, Y. (Ed.), Food Additives. InTech, ISBN 978-953-51-0067-6. http://www.intechopen.com/articles/show/title/aluminum-in-food-the-nature- andcontribution-of-food-additives.

Amino acids: Chemistry and classification

Guoyao Wu, Department of Animal Science and Faculty of Nutrition, Texas A&M University, College Station, TX, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of P.W. Emery, Amino Acids: Chemistry and Classification, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 64–71, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00009-X.

Introduction 10
Nomenclature of AAs 13
Free AAs and Peptide- (or protein)-bound AAs 14
Chemical Properties of AAs 16
Physical appearance, fluorescence, and melting points of crystalline AAs 16
Tastes of crystalline AAs 16
Solubility of AAs in water and organic solvents 18
The Zwitterionic form of AAs 19
Chemical stability of AAs 19
Chemical Reactions of AAs 19
Chemical reactions of the amino group in α-AAs 19
Chemical reactions of the carboxyl group in α-AAs 19
Chemical reactions of the side chain in α-AAs 19
Chemical reactions involving both the amino and carboxyl groups of α-AAs 20
Chemical analyses of AAs 20
Classification of AAs 21
Conclusion and outlook 21
Acknowledgments 22
References 22

Key points

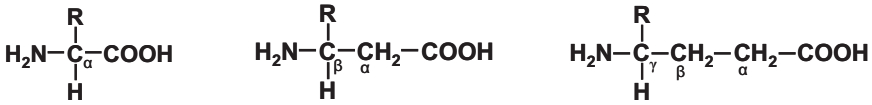
- Amino acids (AAs) are organic molecules containing both amino and acidic groups. They exist in protein and peptides (~97% of total AAs), as well as the free form in humans.
- The configuration of AAs is generally defined in reference to L- and D-glyceraldehyde. The absolute configuration of AAs is also designated according to the R/S system, and includes *cis* and *trans* for those with a ring structure.
- All α-AAs, except for glycine and aminomalonic acid, can have both L- and D-isomers. Except for glycine, proteinogenic (protein-creating) AAs in human tissues are L-isomers.
- All L-AAs plus glycine are the most abundant physiological isomers and account for > 99.9% of total AAs in humans.
- Glycine is the most abundant peptide-bound AA in humans, followed by proline, glutamate, and arginine. Glutamine and alanine are the most and second most abundant free AA in the human plasma, respectively.
- Human liver, skeletal muscle, brain, and small-intestinal mucosa contain high concentrations of free aspartate plus taurine plus glutamine plus glutamate, free glutamine plus taurine plus glutamate, free glutamate plus glutamine, and free taurine plus glutamate, respectively.
- Meat is an abundant source of taurine, carnosine and anserine but plants lack these substances. β-Alanine and 4-hydroxyproline are abundant in meat but is negligible or low in plant-sourced foods.
- Different AAs have different chemical properties (e.g., solubility, stability, melting points, reactions, taste, and electrical charges).
- AAs are classified as neutral, basic, and acidic AAs based on their net charges at neutral pH.
- AAs are designated as aliphatic, sulfur-containing, hydroxylated, amidated, carboxylated, phosphorylated, guanidino, aromatic, imidazole, seleno, and secondary amino acids.

Introduction

Amino acids (AAs) are organic substances containing both amino and acid groups. Usage of the term AA in the English language began in 1898. All natural AAs consist of the elements C, H, N, and O. Some of the natural AAs contain the elements S, P, I, and Se. An acid group in a natural AA can be the carboxyl (-COOH; e.g., alanine), sulfonic acid (-SO_3H; e.g., taurine), phosphoric acid (-PO_4H_2; e.g., phosphoethanolamine), or phosphonic acid (-PO_3H_2; e.g., 2-aminoethylphosphonte) group (Wu, 2022). All

proteinogenic (protein-creating) AAs contain an amino group and an α-carboxyl group, and two of them (aspartate and glutamate) also have a carboxyl group in their side chains (Fig. 1). A 3-letter abbreviation is used to designate an AA, with the first capital letter followed by two lower-case letters (e.g., Leu for leucine). One capital letter denotes an AA in protein sequences (e.g., L for leucine).

The different carbon atoms of AAs are named in sequence according to the Greek alphabet. If the amino group ($-NH_2$) is linked to the α-, ß-, γ-, δ-, or ϵ-carbon of the hydrocarbon chain, the AA is designated an α-, ß-, γ-, δ-, or ϵ-AA, respectively. Examples of natural α-, ß-, γ-, δ-, or ϵ-AAs are alanine, β-alanine, γ-aminobutyric acid [GABA, a neurotransmitter (He and Wu, 2020)], 5-aminovaleric acid, and 6-aminocaproic acid, respectively. Examples of non-proteinogenic AAs and their metabolites in humans (including those produced by gastrointestinal bacteria) are illustrated in Fig. 2. Imino acids (e.g., proline and 4-hydroxyproline) contain both a secondary α-amino (α-imino) group (-NH) and an acid group, and are loosely called AAs in biochemistry and nutrition.



α-Amino Acid β-Amino Acid γ-Amino Acid

There is a rich history of chemical studies of AAs (Meister, 1965; Wu, 2022). Asparagine was the first natural AA discovered by two French chemists L.N. Vauquelin and P.J. Robiquet in 1806, and glycine was the first AA isolated from a protein (i.e., gelatin) by

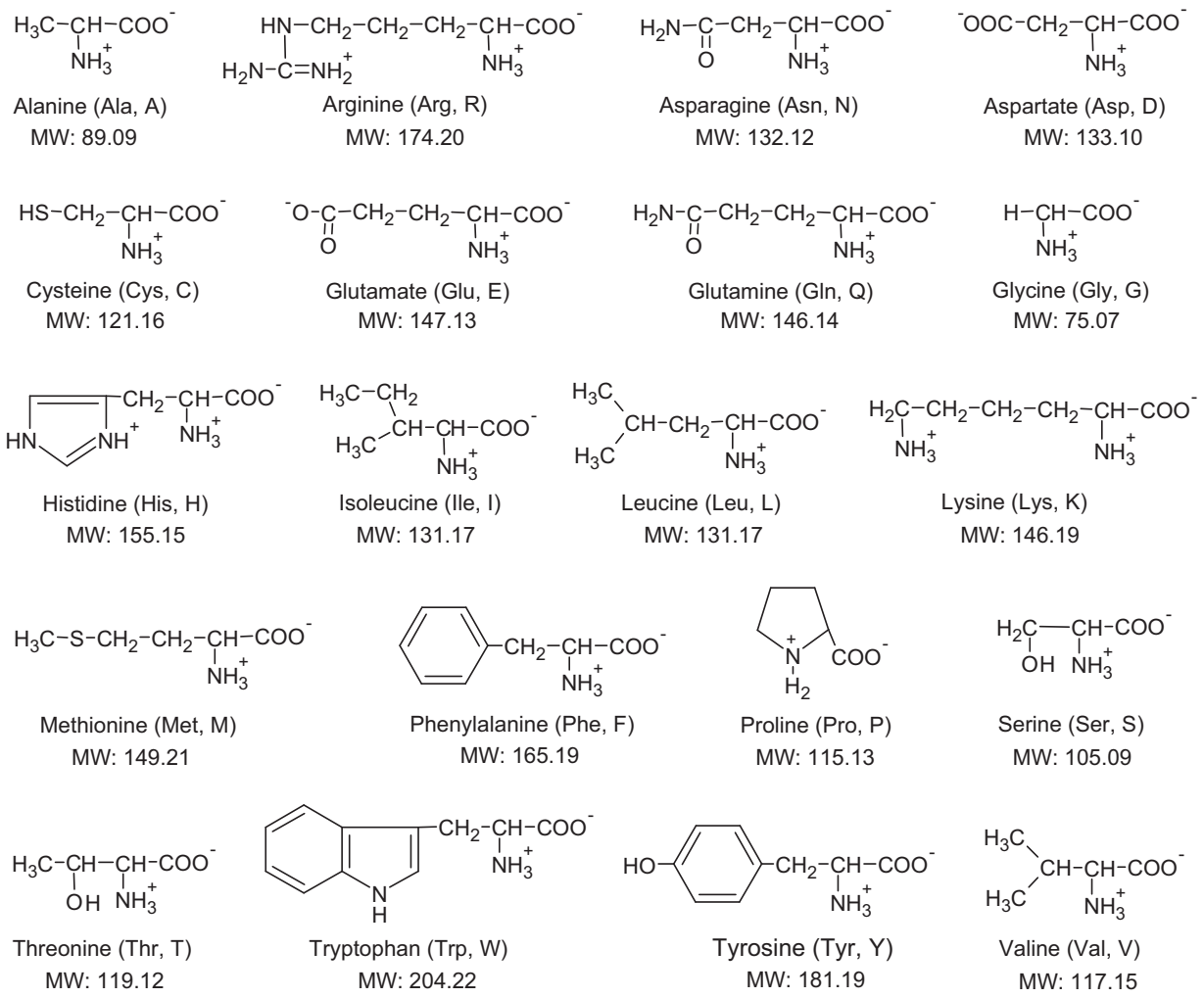


Fig. 1 Chemical structures of proteinogenic amino acids at neutral pH in humans. Except for selenocysteine (a rare, special amino acid in selenoproteins), the 20 canonical proteinogenic amino acids are present in proteins and peptides as well as the free form. Reproduced from Wu (2018) with permission.

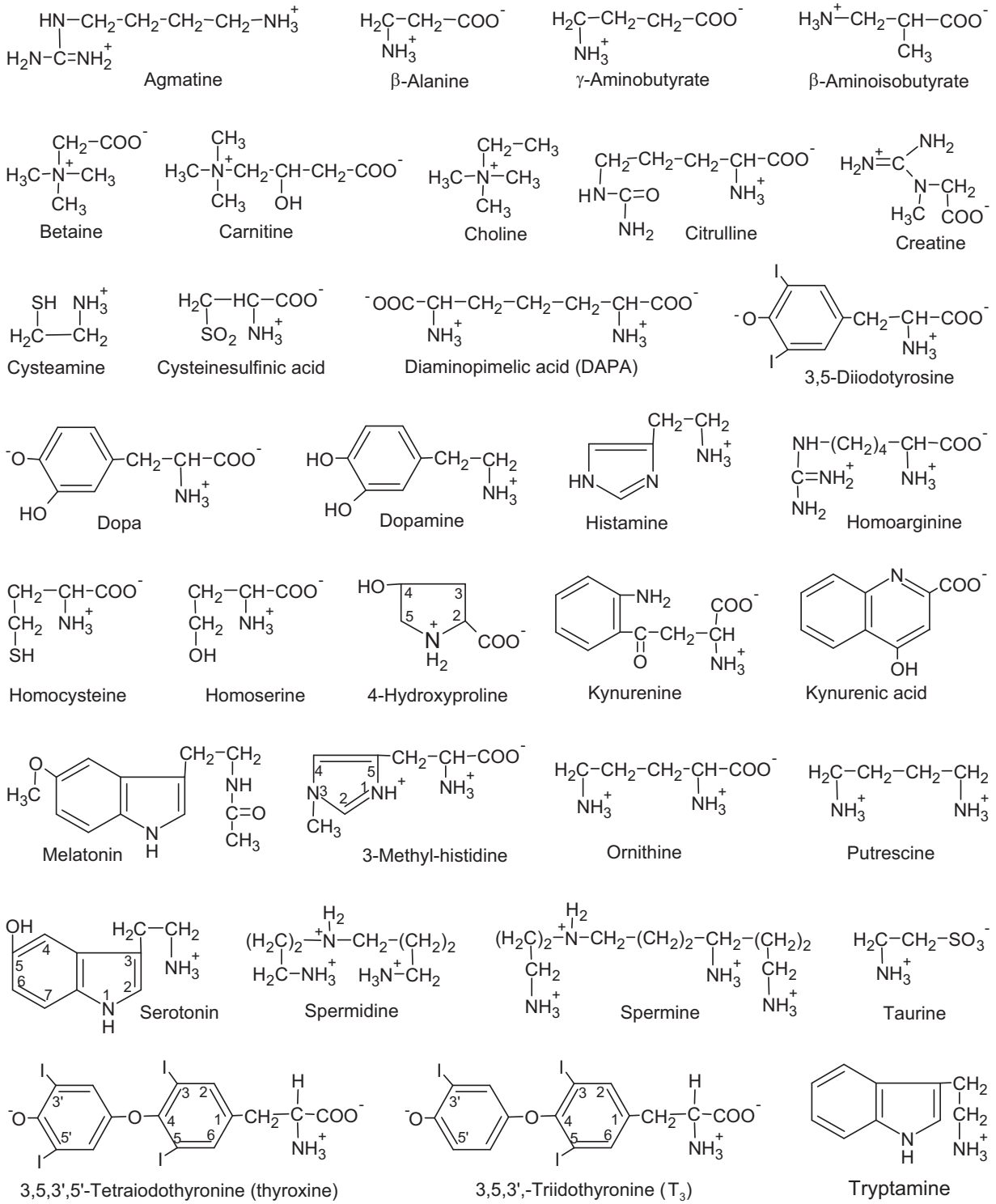
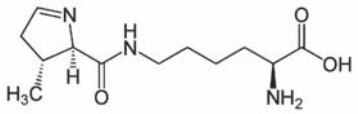


Fig. 2 Chemical structures of non-proteinogenic amino acids (AAs) and their metabolites at neutral pH in humans (including those produced by gastrointestinal bacteria). These substances occur in physiological fluids of animals, and some of them are also present in microorganisms and plants. Taurine is present in humans as a free AA but is absent from bacteria and plants. Reproduced from Wu (2018) with permission.

the French chemist H. Braconnot in 1820. In 1925, threonine was discovered by S.B. Schryver and H.W. Buston as a common proteinogenic AA in oat protein, and R.A. Gortner and W.F. Hoffmann found that this AA was a constituent of teozein. In 1973 and 2002, selenocysteine and pyrrolysine were found to be rare AAs in selenoproteins and the Archaea methylamine methyltransferase, respectively. AAs are present in humans and their foods in both the free and peptide-bound forms. Over 700 AAs occur in nature,

but only 20 of them serve as the precursors of proteins in humans. Because of variations in their side chains (e.g., carbon numbers, structure, and chemical groups), AAs have remarkably different chemical properties and classifications, which are highlighted in this chapter.



L-Pyrrolysine (Pyl, O; $C_{12}H_{21}N_3O_3$; MW = 255.31 Da)

Nomenclature of AAs

Most AAs have an asymmetric carbon atom (chiral carbon). Thus, their aqueous solutions exhibit an optical activity and rotate plane-polarized light (Greenstein and Winitz, 1961). The configuration of an AA (L- or D-isomers) was first arbitrarily defined by Emil Fischer in 1908 according to that of L- and D-glyceraldehyde (**Fig. 3**). Because each asymmetric carbon can have two possible configurations, an AA with n asymmetric carbons has 2^n different possible stereoisomers. Except for glycine, all proteinogenic AAs can have both D- and L-isomers. Although L-AAs are the predominant physiological isomers of AAs in humans, some D-AAs (e.g., D-serine and D-alanine) occur in certain tissues (e.g., the brain) and fulfill physiological functions. A small amount of D-AAs is present in foods (e.g., milk and heat-processed cereal grains). Human proteins consist of only L-AAs and glycine, but certain bacterial peptides contain D-AAs (e.g., D-proline, D-alanine, and D-glutamate). Some non-proteinogenic AAs [e.g., GABA, β-alanine, and taurine (which have important physiological roles in humans), and aminomalonic acid (which is present as a free AA in humans and microbes)] have no D- or L-isomers. The chemical configuration of AAs is not changed in their crystalline form. All L-AAs plus glycine are the most abundant physiological isomers and account for > 99.9% of total AAs in humans.

The nomenclature of an AA with more than one asymmetric carbon includes an allo-form. For example, a synthetic L-threonine (a 4-carbon AA) whose β-carbon has a configuration opposite to that of the α-carbon is called L-allo-threonine. L-Threonine and L-allo-threonine are called diastereomers, and the relationship between the two AAs is diastereomerism. Likewise, a synthetic L-isoleucine (a 5-carbon AA) whose β-carbon has a configuration opposite to that of the α-carbon has a diastereomer of L-allo-isoleucine. This also applies to synthetic D-threonine and D-isoleucine.

To better distinguish some AAs (e.g., 4-hydroxyproline and 3-hydroxyproline) with a pyrrolidine ring structure or more than one chirality center (e.g., threonine and isoleucine), the *R*/*S* nomenclature is also used to name their absolute configurations, where *R* is *rectus* (right) and *S* is *sinister* (left) in Latin (Hu et al., 2021). This nomenclature system does not involve a reference molecule such as glyceraldehyde, but is based on the prioritized spatial arrangement of the four different groups that are attached to the asymmetric carbon. For AAs (e.g., e.g., hydroxyproline) with a pyrrolidine ring structure, *cis*–*trans* isomerism is often used to distinguish their chemical structure, with *cis* and *trans* denoting that the functional groups are on the same and opposite side of a plane, respectively. In the *R*/*S* system, naturally occurring AAs are mostly in the *S* configuration for the first chiral center. *Trans*-(2*S*,4R)-4-hydroxyproline (i.e., *trans*-4-hydroxy-L-proline) and *trans*-(2*S*,3*S*)-3-hydroxyproline (i.e., *trans*-3-hydroxy-L-proline) (**Fig. 4**) play important roles in the structure and stability of collagen and elastin, primarily extracellular proteins in the connective tissue of humans. Human collagen consists of 9.01% *trans*-(2*S*,4R)-4-hydroxyproline and 0.09% *trans*-(2*S*,3*S*)-3-hydroxyproline on the molar basis. *Trans*-(2*S*,4R)-4-hydroxyproline also contribute to the phosphorylation and catalytic activities of some non-collagen protein kinases in mammals.

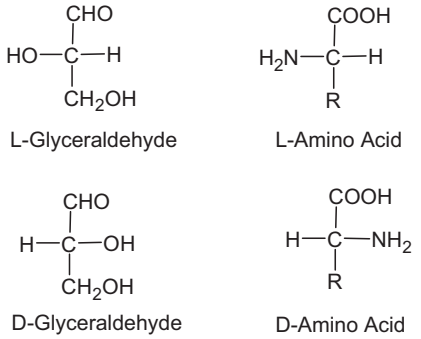


Fig. 3 Fisher projections for the chemical configurations of amino acids relative to L- and D-glyceraldehydes. The general structure of an amino acid in its non-ionized form is shown. Note that glycine, taurine, β-alanine, γ-aminobutyrate, and aminomalonic acid have no asymmetric carbon and, therefore, no L- or D-isoform.

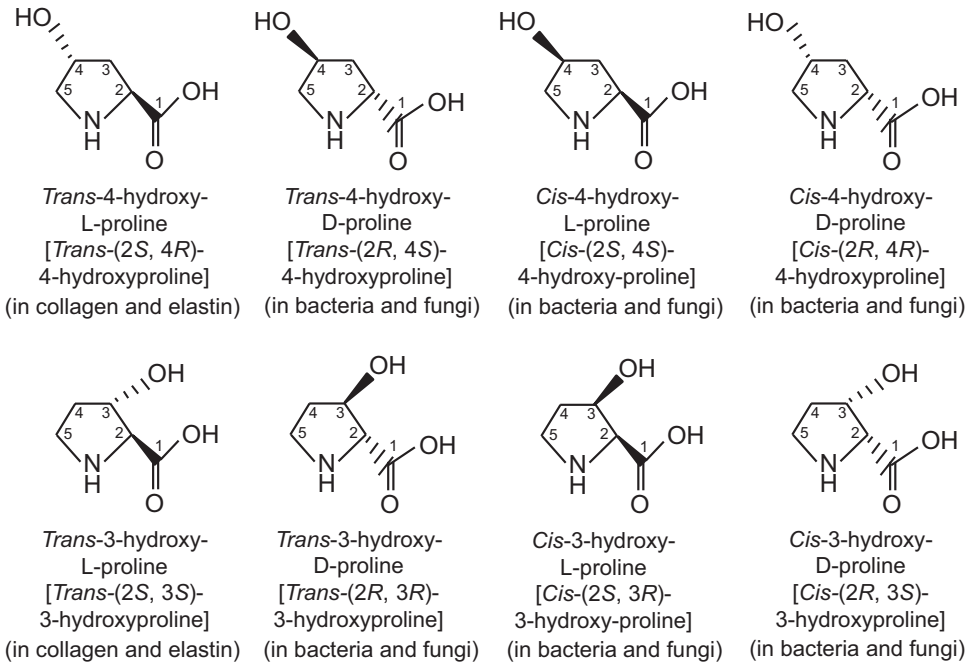


Fig. 4 Chemical structures of 4-hydroxyproline and 3-hydroxyproline in humans, including those produced by gastrointestinal bacteria. Reproduced from Hu et al. (2021) with permission.

Free AAs and Peptide- (or protein)-bound AAs

Free AAs are those AAs that are not covalently bound in dipeptides, oligopeptides, or polypeptides. Aspartate, glutamate, glutamine, and taurine are among the most predominant free AAs in human tissues (Adibi and Mercer, 1973; Ahlman et al., 1993; Barle et al., 1996; Perry et al., 1971). In healthy humans, glutamine is the most abundant free AA in their plasma and skeletal muscle, and is the second most abundant free AA in their brain and liver (**Table 1**). Human skeletal muscle, heart, retina, placenta, brain, and small-intestinal mucosa contain 15–20, 28–40, 20–35, 20–35, 1.16, and 8.74 mM taurine, respectively (Wu, 2020). A healthy 70-kg person has approximately 80 g glutamine and 70 g taurine. In the human brain and small-intestinal mucosa, glutamate is the most abundant and second most abundant free AA, respectively. Interestingly, aspartate is the most abundant free AA in the human liver (**Table 1**). Selenocysteine does not occur as a free AA in humans (Wu, 2022).

Peptide- (or protein)-bound AAs are those AAs that are linked in peptides or proteins via peptide bonds. A healthy 70-kg adult human contains 15.1% protein (or 10.6 kg protein), including 3.72 kg collagen or 35% of the total body protein. Skeletal muscle contains about 20% protein and accounts for ~45% of body mass in non-obese and non-overweight heathy adults. In skeletal muscle proteins, glutamate is the most abundant amino acid, followed by leucine, lysine, and arginine. In the whole body, glycine is the most abundant amino acid, followed by proline, glutamate, and arginine in descending order (**Table 1**). Note that selenocysteine is present in selenoproteins due to the conversion of serine and selenium into selenocysteine during the mRNA translation process.

Total free AAs represent ~ 3% of total AAs (free plus peptide-bound) in humans, but individual free AAs may contribute more (e.g., free glutamine) or less (e.g., free tryptophan) to the whole-body AA pool. In beef, about 97-98% of total AAs are present in proteins and polypeptides (**Table 2**). This is also true for most of plant-sourced foods (e.g., wheat flour, soybean, rice, and corn; **Table 2**). By contrast, free AAs represent 34.4% and 28.5% of total AAs in potatoes and sweet potatoes, respectively (Hou et al., 2019). Interestingly, free asparagine represents 32.3% and 17.5% of total free AAs in potatoes and sweet potatoes, whereas free glutamine constitutes 25.5% of total free AAs in potatoes (Hou et al., 2019). Taurine is abundant in human milk and animal-sourced foods but is absent from plants (Wu, 2022). In addition, meat is an excellent source of β-alanine and 4-hydroxyproline (Wu et al., 2016), but these two AAs are low or negligible in plants (Li and Wu, 2020). Exposure to high temperatures can result in the formation of a relatively small amount of the following AAs: (1) D-AAs from L-AA (Friedman, 1999), (2) lysinoalanine and lanthionine from cystine or cysteine and serine, (3) 3-methyl-lysinoalanine from lysine and threonine, (4) 3-methyl-lanthionine from cysteine and threonine, and (5) a modified AA (e.g., arginine and lysine) with a Schiff base from an AA (e.g., arginine and lysine) in the presence of a carbohydrate (e.g., glucose and starch) (Wu, 2018,; **Fig. 5**).

Not all AAs in proteins are precursors for the synthesis of the polymers. This is because some AA residues in proteins undergo post-translational modifications to exert biological functions, such as hydroxylation, methylation, acetylation, phosphorylation, nitrosylation, sulfation, glycosylation (*O*- and *N*-linkages), deamination, oxidation, deamidation, racemization, γ-carboxylation, biotinylation, and glycation (Wu, 2022). Thus, in humans, some proteins contain modified AA residues, including

Table 1 Concentrations of free amino acids in the tissues and plasma of adult humans, as well as amino acid composition in the whole human body

	Concentrations of free amino acids					*Content of amino acids in*	
	Plasma[a]	*Liver*[b]	*Skeletal muscle*[a]	*Brain*[c]	*Small intestine*[d]	*Skeletal muscle protein*[e]	*The whole human body*[f]
Amino acids	*(mM)*	*(mM)*	*(mM)*	*(mM)*	*(mM)*	*(mg/g protein)*	
Alanine	0.33	1.83	2.34	0.94	3.17	61	72
β-alanine	0.0067	–	1.0	0.014	–	–	–
γ-aminobutyrate	0.0005–0.003[j]	–	–	1.07	–	–	–
Arginine	0.08	0.014	0.51	0.16	0.23[j]	71	77
Asparagine	0.05	0.19	0.47	0.11	0.51[j]	43	–
Aspartate	0.004	10.7	1.82[j]	1.17	3.35	54	49
Citrulline	0.03	–	0.04	0.13	–	–	41
Cysteine[g]	0.11	0.03[h]	0.18	0.36	–	16	–
Glutamate	0.06	2.59	4.38	10.3	6.90	93	81
Glutamine	0.57	2.94	19.5	9.07	0.86[j]	55	49
Glycine	0.21	2.14	1.33	0.76	2.17	38	118
Histidine	0.08	0.44	0.37	0.14	0.11[j]	31	26
4-hydroxyproline	0.02	–	–	–	–	3.3	36
Isoleucine	0.06	0.06	0.11	0.07	0.43	53	35
Leucine	0.12	0.17	0.15	0.16	1.39	83	75
Lysine	0.18	0.14	1.15	0.34	0.31[j]	76	72
Methionine	0.02	0.03	0.11	0.07	0.12	29	20
Ornithine	0.06	0.21	0.30	0.11	0.17[j]	–	–
Phenylalanine	0.05	0.06	0.07	0.07	0.56	51	41
Proline	0.17	0.41[h]	0.83	0.23	–	40	84
Serine	0.12	0.57	0.98	0.57	1.07	44	44
Taurine	0.07	4.86	15.4	1.16	8.74[j]	–	–
Threonine	0.15	0.31	1.03	0.40	0.76	49	41
Tryptophan	0.02	0.014	0.08[i]	0.05	–	14	–
Tyrosine	0.05	0.09	0.10	0.07	0.55	38	29
Valine	0.22	0.19	0.26	0.26	0.92	60	47

[a]Adapted from Bergström et al. (1974) for healthy men and women (20 to 36 years of age). Blood and skeletal muscle (quadriceps femoris muscle) samples were obtained from the study participants after an overnight fast.

[b]Adapted from Barle et al. (1996) for metabolically healthy men and women with no known liver disease. Liver samples were obtained from the study participants after a 12–15 h fast. Concentrations of amino acids in the liver were calculated on the basis of its water content (70%).

[c]Adapted from Perry et al. (1971). Concentrations of amino acids in the brain were calculated on the basis of its water content (70%).

[d]Adapted from Adibi and Mercer (1973) for healthy men (19–24 years of age). The study individuals were fasted overnight and then consumed a meal (containing 50 g of bovine serum albumin). One hour after feeding, jejunal mucosal samples were obtained from the individuals at 1 h for amino acid analysis. The concentration of glutamine plus asparagine was 1.31 mM.

[e]Values were calculated on the basis of the molecular weights of intact amino acids.

[f]Adapted from Wu (2022) and Hu et al. (2021). Values were calculated on the basis of the molecular weights of intact amino acids. The content of asparagine plus aspartate and glutamine plus glutamate in the human body was 90 and 130 mg/g protein, respectively.

[g]Total cysteine (cysteine + ½ cystine). "1/2 cystine" denotes that 1 mol cystine is composed of 2 mol cysteine.

[h]Ryan and Carver (1966).

[i]Borgenvik et al. (2012).

[j]Adapted from Ahlman et al. (1993) for the duodenal mucose of metabolically healthy, well-nourished men and women. Duodenal mucosal samples were obtained from the study participants after an overnight fast.

4-hydroxyproline, 3-hydroxyproline, 5-hydroxylysine, 3-methylhistidine, 1-methylhistidine, methylated lysine, asymmetrical dimethylarginine, symmetrical dimethylarginine, N^G-monomethylarginine, N^ϵ-acetylated lysine, phosphoserine, phosphotyrosine, nitrosylated cysteine, nitrosylated tyrosine, sulfated tyrosine, N-acetylglucosamine-linked serine, N-acetylglucosamine-linked asparagine, citrulline, allysine (in collagen and elastin), hypusine, D-aspartate (in myelin basic protein), γ-carboxylated glutamate (in blood clotting factors), biotin-acylated lysine (in acyl carrier protein), and glucose-glycated hemoglobin (in diabetes). As shown in **Fig. 6**, the cross-linkage of certain modified AA residues (e.g., allysine) results in the formation of condensed molecules (e.g., desmosine and its isomer isodesmosine) required for normal extracellular matrix structure and strength (Schmelzer et al., 2020).

Table 2 Content of free amino acids (FAAs) and peptide-bound amino acids (PAAs) in foods

	Corn grain		*Potato*		*Soybean*		*Sweet potato*		*Wheat flour*		*White rice*		*Beef (loin)*	
AA	*FAA*	*PAA*	*Free*	*PAA*	*FAA*	*PAA*	*FAA*	*PAA*	*FAA*	*PAA*	*FAA*	*PAA*	*FAA*	*PAA*
Ala	224	7.74	243	2.58	146	22.7	1342	3.46	70.4	5.35	33.8	4.55	1258	44.1
Arg	37.8	4.30	2058	3.10	1819	35.4	467	2.61	84.6	7.12	35.5	7.25	170	52.2
Asp	151	4.59	2017	2.04	240	37.1	1803	1.53	256	3.33	21.0	4.25	65	41.0
Asn	114	3.77	10,095	9.15	435	24.1	3686	8.28	237	4.23	93.7	4.51	206	33.2
Cys[a]	32.6	2.14	387	0.75	40.2	7.73	220	1.41	14.7	3.02	5.48	1.62	362	10.7
Glu	410	6.72	112	0.89	601	52.0	2250	3.21	165	2.64	65.7	7.32	801	74.0
Gln	429	11.4	7957	8.36	15.2	43.0	603	1.59	88.2	45.4	10.3	7.95	5069	44.8
Gly	193	4.24	108	2.63	48.5	24.0	401	3.12	24.7	6.28	7.02	3.95	456	33.2
His	40.2	2.64	965	0.74	217	12.5	183	0.99	14.8	3.53	8.20	2.19	338	31.4
Hyp	ND	0.04	ND	0.08	ND	0.78	ND	0.05	ND	0.43	ND	0.04	73	1.70
Ile	119	3.73	731	2.46	32.3	23.9	671	2.64	14.3	4.91	8.28	3.73	236	40.9
Leu	225	12.5	514	4.21	50.7	39.9	813	4.07	52.6	9.95	40.5	7.22	398	66.2
Lys	84.3	2.77	492	4.64	86.2	32.6	297	3.32	28.8	3.62	6.37	2.32	349	71.7
Met	35.7	2.23	483	1.05	34.2	6.39	274	0.95	11.2	2.47	5.85	2.02	180	25.2
Phe	72.5	5.11	824	3.10	56.2	25.6	1319	3.51	15.9	7.51	7.47	4.39	352	33.2
Pro	280	11.0	689	2.33	109	28.3	1246	1.40	78.3	23.2	43.8	5.36	535	32.4
Ser	233	4.82	679	2.88	62.7	30.0	1946	2.60	42.7	6.86	34.2	4.05	365	35.0
Thr	197	3.38	399	2.79	24.3	20.9	1485	2.26	16.8	4.07	10.3	3.01	389	36.5
Trp	20.2	0.73	322	0.69	34.6	6.99	193	0.29	5.73	1.70	4.61	1.12	67	9.95
Tyr	146	4.66	619	1.73	38.7	18.4	632	2.36	20.6	3.84	15.2	2.36	244	29.8
Val	162	4.90	1463	3.58	39.1	24.6	1103	3.36	26.1	6.22	19.8	5.17	278	47.2
β-Ala	9.92	–	46.0	–	69.3	–	12.2	–	3.99	–	2.38	–	712	–
Cit	28.5	–	18.7	–	22.3	–	12.7	–	2.30	–	1.68	–	63	–
Orn	19.2	–	91.8	–	9.08	–	21.7	–	6.79	–	0.83	–	195	–
Tau	ND	–	ND	–	ND	–	ND	–	ND	–	ND	–	2920	–
Total[b]	3266	103.5	31,314	59.8	4231	516.5	20,981	53.0	1282	155.6	482	84.4	16,088	795
Total[c]	–	88.5	–	51.3	–	444.0	–	45.3	–	133.8	–	72.5	–	684

Cit = citrulline; Hyp, 4-hydroxyproline; ND = not detected; Orn = ornithine; PAA = amino acids in proteins and peptides.

[a]Total cysteine (cysteine + ½ cystine). "1/2 cystine" denotes that 1 mol cystine is composed of 2 mol cysteine.

[b]The molecular weights of intact amino acids were used to calculate the amount of peptide-bound amino acids.

[c]The molecular weights of amino acid residues (i.e., the molecular weights of intact amino acids – 18) were used to calculate the amount of peptide-bound amino acids. Total peptide-bound amino acids represent true proteins plus peptides in foods.

Adapted from Hou et al. (2019) and Wu et al. (2016). Values (the means) for free AAs and peptide-bound AAs are expressed as μg/g of dry matter and mg/g of dry matter, respectively.

Chemical Properties of AAs

Physical appearance, fluorescence, and melting points of crystalline AAs

AA crystals are generally white. Phenylalanine, tyrosine, and tryptophan are naturally fluorescent substances, but other AAs are not. This presents a technical challenge for their analysis at a low limit of detection without derivatization with appropriate reagents. AAs generally have a high melting points (Table 2).

Tastes of crystalline AAs

Different AAs confer different tastes (Wu, 2022). Generally speaking, small AAs (e.g., glycine, L-alanine, L-citrulline, L-proline, L-serine and L-threonine) confer sweet taste, whereas some large L-AAs (e.g., L-arginine, L-Isoleucine, L-lysine and L-phenylalanine) have bitter tastes. β-Alanine, L-cystine, L-methionine, and L-tyrosine are flat (lacking taste), whereas L-cysteine, L-histidine, L-leucine, L-ornithine, and L-tryptophan have a flat to bitter taste. L-Glutamate (e.g., monosodium glutamate) and glutamic acid have a "meaty" umami taste. By contrast, D-glutamate is almost tasteless, and D-cystine is flat. Unlike most L-AAs as well as D-glutamate and D-cystine, many D-AAs are sweet substances.

Some AAs have complex tastes. For example, L-aspartate has a sour taste with a slight umami flavor, L-valine has a bitter taste after an initial slightly sweet flavor, and L-histidine-HCl has a slightly bitter taste after an initial sour flavor. The taste of basic AAs is altered by their hydrochloride salts, such that L-arginine-HCl, L-lysine-HCl and L-ornithine-HCl have a mild but characteristic taste. Different individuals may have a different tasting response to the same AA.

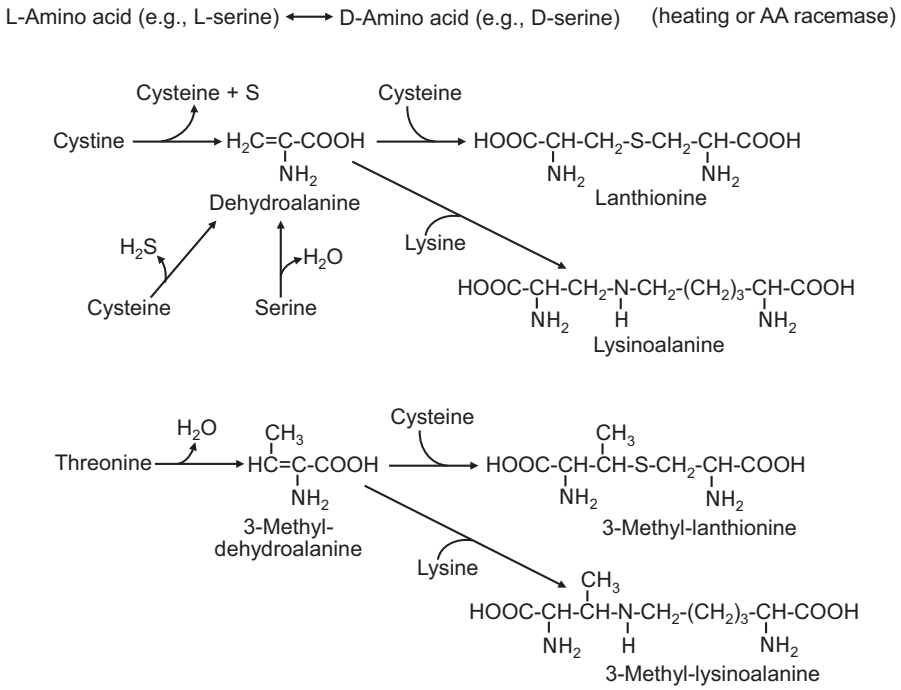


Fig. 5 Formation of D-amino acids (AAs), lanthionine, lysinoalanine, 3-methyl-lanthionine, 3-methyl-lysinoalanine after heating of free AAs. These reactions can also occur in peptide-bound AAs. In addition, AA racemases can catalyze the formation of D-AAs from L-AAs in humans. Reproduced from Wu (2018) with permission.

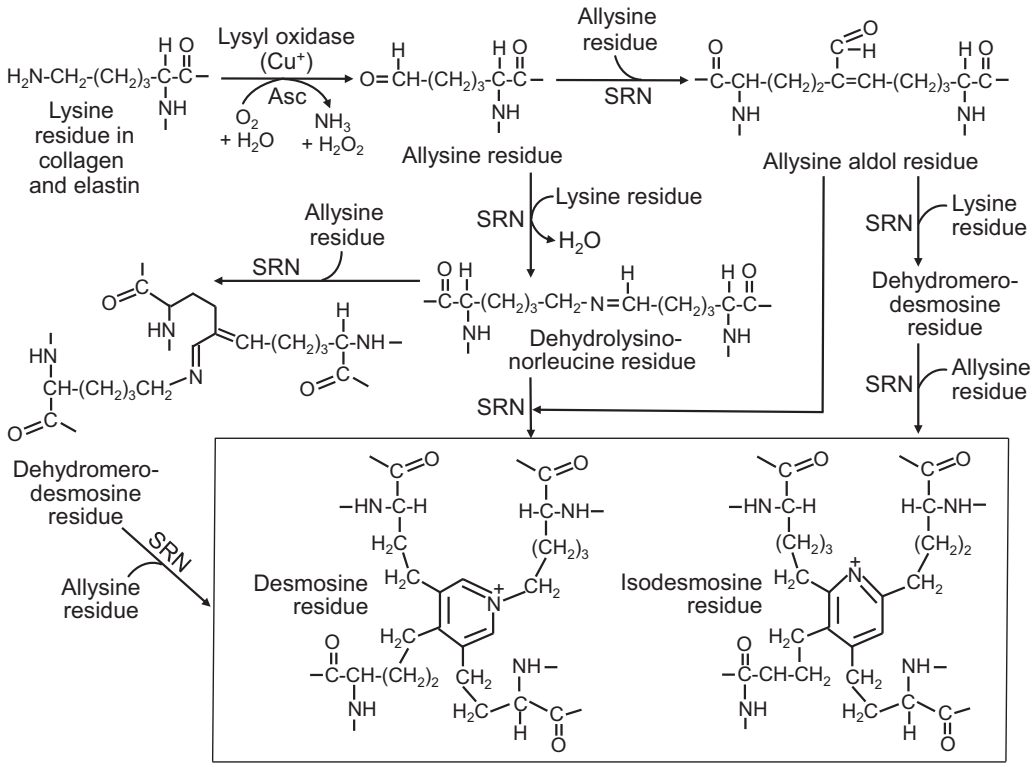


Fig. 6 Formation of desmosine and its isomer isodesmosine from lysine residues in human collagen and elastin for intramolecular and intermolecular cross links. Lysyl oxidase [an ascorbate (Asc, vitamin C) and copper-dependent enzyme] catalyzes the oxidative deamination of lysine residue into allysine residue. Condensations between the indicated molecules occur as spontaneous reactions (SRN) to form desmosine and isodesmosine, which consist of 4 lysine residues to form a pyridinium nucleus. Substrates and products, which are residues in collagen and elastin of all the reactions, are shown. In their intact molecules and non-ionized form, the "–HCO–" group is "–HCOOH" and the "–NH–" group should be "–NH_2". The molecular formula of intact desmosine and isodesmosine is $C_{24}H_{40}N_5O_8$ (MW = 526.603).

Solubility of AAs in water and organic solvents

All AAs are soluble in water. Leucine, isoleucine, valine, phenylalanine, tryptophan, methionine, tyrosine, and cysteine are among the most hydrophobic AAs. Proline and cystine are the most and least soluble, respectively (Table 3). The solubility of tyrosine in water is low. By contrast, *N*-acetyl-cysteine and *N*-acetyl-tyrosine are readily soluble in water (Hou et al., 2015a). The solubility of AAs in water generally increases in acidic or alkaline solutions and with elevated temperatures. With the exception of proline and hydroxyproline, AAs are generally insoluble in organic solvents (e.g., absolute ethanol). The hydrochloride salts of AAs (both neutral and basic) are generally more soluble in water than the corresponding free AAs. Most of the AA hydrochlorides are highly soluble in absolute ethanol.

Table 3 Chemical properties and classification of amino acids

Amino acids	*Side chain*	*Nitrogen content (%)*	*MP (°C)*	*Solub*[a] in H_2O	*pK_1*[b]	*pK_2*[c]	*pK_3*[d]	*pI*
(1) Neutral amino acids								
L-alanine	Aliphatic	15.72	297	16.5	2.35	9.87	–	6.11
β-alanine	Aliphatic	15.72	197	82.8	3.55	10.24	–	6.90
γ-aminobutyrate	Aliphatic	13.58	202	107.3	4.03	10.56	–	7.30
β-aminoisobutyrate	Aliphatic	13.58	176	56.7	4.17	10.32	–	7.25
L-asparagine	Containing an amide group	21.20	236	2.20	2.02	8.80	–	5.41
L-citrulline	Ureido	23.99	222	15.2	2.43	9.41	–	5.92
L-cysteine	Containing sulfur atom	11.56	178	17.4	1.96[e]	8.18[e]	10.28[e]	5.07[e]
L-cystine	Containing sulfur atom	11.66	261	0.011	<1.0[f]	8.02[f]	–	5.06[f]
					2.10[f]	8.71[f]	–	
L-glutamine	Containing an amide group	19.17	185	4.81[e]	2.17	9.13	–	5.65
Glycine	Aliphatic	18.66	290	25.0	2.35	9.78	–	6.07
L-4-hydroxyproline	Containing an imino group	10.68	270	36.1	1.92	9.73	–	5.83
L-isoleucine	Aliphatic, branched-chain	10.68	284	4.12	2.36	9.68	–	6.02
L-leucine	Aliphatic, branched-chain	10.68	337	2.19	2.33	9.75	–	6.04
L-methionine	Containing sulfur atom	9.39	283	5.06	2.28	9.21	–	5.74
L-phenylalanine	With an aromatic ring	8.48	284	2.96	2.20	9.31	–	5.76
L-proline	Containing an imino group	12.17	222	162.3	1.99	10.6	–	6.30
L-serine	Containing an –OH group	13.33	228	41.3	2.21	9.15	–	5.68
Taurine	Containing sulfur atom	11.19	328	10.5	1.50	8.74	–	5.12
L-threonine	Containing an –OH group	11.76	253	9.54	2.15	9.12	–	5.64
L-tryptophan	With an aromatic ring	13.72	282	1.14	2.38	9.39	–	5.89
L-tyrosine	With –OH and an aromatic ring	7.73	344	0.045	2.20	9.11	10.07	5.66
L-valine	Aliphatic, branched-chain	11.96	315	5.82	2.29	9.72	–	6.01
(2) Basic amino acids								
L-arginine	With guanidino group	32.16	238	18.6	2.17	9.04	12.48	10.76
L-histidine	With imidazole group	27.08	277	4.19	1.80	9.33	6.04	7.69
L-homoarginine	With guanidino group	29.77	208	0.158	2.49	9.00	12.30	10.65
L-lysine	With an ϵ-amino group	19.16	224	78.2[g]	2.18	8.95	10.53	9.74
L-ornithine	With an amino group	21.20	231[h]	54.5[h]	1.94	8.65	10.76	9.71
L-pyrrolysine	With an ϵ-amino group	16.46	–	–	–	–	–	–
(3) Acidic amino acids and their salts								
Aminomalonic acid	Containing a carboxyl group	11.76	243	12.6	1.80	8.50	1.80	1.80
L-aspartic acid	Containing a carboxyl group	10.52	270	0.45	1.88	9.60	3.65	2.77
L-Glu-Na (MSG)	Containing a carboxyl group	8.28	232	73.9	–	9.47	4.07	6.77
L-glutamic acid	Containing a carboxyl group	9.52	249	0.86	2.19	9.67	4.25	3.22
L-selenocysteine	Containing a seleno group	8.38	145	32.5	2.50	9.50	5.20	3.90

MP = melting point; MSG = monosodium glutamate (the sodium salt of glutamic acid); MW = molecular weight.

[a]Solubility (Solub); g/100 mL of water at 25°C unless otherwise indicated).

[b]pK_a for α-COOH (SO_3H for taurine), namely pK_1, at 25°C unless otherwise indicated.

[c]pK_a for α-NH_3^+, namely pK_2, at 25°C unless otherwise indicated.

[d]pK_a for the ionized group in the side chain, namely pK_3, at 25°C unless otherwise indicated.

[e]Determined at 30°C.

[f]Determined at 35°C.

[g]L-lysine-H_2O.

[h]L-ornithine-HCl.

Adapted from Wu (2022).

The Zwitterionic form of AAs

In an aqueous solution, AAs are ionized to have both positive and negative electrical charges, which are known as the zwitterion. The ionizable groups of AAs interact with water through ionic interactions, hydrogen bonds, and van der Waal interactions. Both the carboxyl and amino groups of an AA, as well as a side chain with an ionizable group have their respective dissociation constants (pK_a), which is the negative log of the acid dissociation constant (K_a). A lower pK_a value indicates a stronger acid or greater tendency to give up a proton (H^+). When the pH of an aqueous solution is equal to pK_a, the ionizable group is 50% protonated and 50% deprotonated (Wu, 2022). The dissociation constants for the acid, amino, and side-chain groups are termed pK_1 (pK_a for the carboxyl group), pK_2 (pK_a for the amino group), and pK_3 (pK_a for the side-chain functional group), respectively. The pH of an aqueous solution at which an AA has no net electrical charge is called the isoelectric point (pI). When an AA is dissolved in pure water until the solution is saturated, the pH of this solution will approach the pI value of the AA. At pH 7.4 (e.g., in the plasma) and pH 7.0 (e.g., in the cytoplasm), the α-carboxylic acid and α-amino groups of α-AAs are completely ionized to take the zwitterion form. In their ionized states, glutamic acid and aspartic acid are often referred to as glutamate and aspartate, respectively.

Chemical stability of AAs

The crystalline forms of all AAs are stable at 25°C for at least 25 years. However, AAs should be protected from light and a high humidity environment during storage. Except for cysteine and aminomalonic acid, AAs in water are stable at -20°C for 3 months and stable at -80°C for 6 months. Likewise, except for cysteine, glutamine, and aminomalonic acid, AAs are generally stable in aqueous solution at physiological pH and body temperature. Cysteine undergoes rapid oxidation to cystine in an aqueous solution that contains oxygen gas, glutamine slowly undergoes intramolecular cyclization to form pyroglutamate (a neurotoxic substance at high concentrations) at a rate of $< 1\%$/day for 1 mM glutamine at pH 7.0 and 37°C; and aminomalonic acid is decarboxylated to glycine.

Most AAs are stable in a neutral aqueous solution at 105°C for 24 h. Under these conditions, glutamine is completely decomposed, whereas asparagine, glutamate, and methionine are partially lost.

Under high pressure and high temperature conditions in an autoclave, glutamine and asparagine are almost completely destroyed. In a dipeptide form (e.g., L-alanyl-glutamine and glycyl-glutamine, as well as L-leucyl-asparagine and glycyl-asparagine), glutamine and asparagine are stable under these conditions.

All AAs except for aminomalonic acid are stable in 1.5 M perchloric acid (HClO4) at $\leq$ 25°C for at least 15 min and stable in 10% trichloroacetic acid at $\leq$ 25°C for at least 30 min. Under alkaline conditions (e.g., 4 M NaOH or KOH) at $\leq$ 25°C, tryptophan is stable for at least 24 h, aminomalonic acid is relatively stable, and other AAs are not stable.

Under standard conditions of acid hydrolysis (i.e., 6 M HCl, 110°C, and 24 h under nitrogen gas), most AAs are stable, but all glutamine and asparagine are converted to glutamate and aspartate, respectively, and tryptophan is completely destroyed. Under the conditions of alkaline hydrolysis at a high temperature (e.g., 105°C), most AAs (including arginine, asparagine, glutamine, glycine, histidine, serine, and threonine) are almost completely destroyed, and many AAs (e.g., cysteine, cystine, and methionine) are decomposed to a great extent. By contrast, tryptophan is stable in alkaline solutions [e.g., 4.2 M NaOH and 1% thiodiglycol (an antioxidant)] at 105°C for 20 h.

Chemical Reactions of AAs

Chemical reactions of the amino group in α-AAs

The amino group of an α-AA participates in many chemical reactions with a variety of substances. The reactions include: acetylation, benzoylation, carbobenzoxylation, condensation, conjugation, deamination, dinitrophenylation, group protection, methylation, oxymethylation, and transamination, as well as keto-enol tautomerization and amino-imino tautomerization.

Chemical reactions of the carboxyl group in α-AAs

The carboxyl group of α-AAs participates in chemical reactions, such as amidation, decarboxylation, esterification, and reduction. Esterification of an AA can block its α-carboxyl group from binding to undesirable substrates. Furthermore, α-AAs can be chemically converted into the corresponding alcohols through reduction with sodium borohydride and iodine in tetrahydrofuran.

Chemical reactions of the side chain in α-AAs

The amino or carboxyl group of the side chain of α-AAs participates in some chemical reactions. For example, the phenolic ring of tyrosine is iodinated under alkaline conditions. In addition, the ε-amino group of lysine and the guanidino group of arginine contribute to hydrogen bonding, can be methylated, and react with carbohydrates (e.g., glucose and starch). The guanidination of free or protein-bound lysine with methylisourea generates homoarginine. The guanidino group of arginine can also react with diketones, and this chemical property is useful for studying protein structure and function (e.g., the allosteric and active sites of enzymes). Furthermore, the γ-amino group of asparagine in protein form an N-linkage with glucose, and free asparagine can react

with reactive carbonyls at high temperatures to generate acrylamide (a potential carcinogen). Of particular note, the presence of an acid (e.g., 6 M HCl) at an elevated temperature (e.g., 110°C) results in the deamidation of the γ-amino group of asparagine and the ε-amino group of glutamine. Finally, glutamine residues in proteins undergo deamidation by transglutaminase I (in the presence of polyamines) or transglutaminase II (in the presence of protein-bound lysine) to release NH_3.

Chemical reactions involving both the amino and carboxyl groups of α-AAs

Both amino and carboxyl groups simultaneously participate in some unique chemical reactions, such as chelation, cyclization, racemization, formation of N-carboxy anhydride, oxidative deamination (decarboxylation; e.g., reaction with ninhydrin), condensation (e.g., peptide formation), and esterification (e.g., among L-arginine monohydrochloride, ethanol, and lauroyl chloride in the presence of NaOH). The resulting products include AA-chelates (providing highly bioavailable minerals for humans), azlactone, diketopiperazine, *N*-carboxy AA anhydride (NCA), peptide bonds, and aldehyde.

The formation of one peptide bond from two molecules of AAs results in the loss of one molecule of water (Fig. 7). Proteins are large polymers of AAs, with the whole collagen (three cross-linked chains) molecule containing approximately 3,000 AA residues. In the human body, 100 g of protein consist of 100 g of AA residues but not 100 g of intact AAs. The complete hydrolysis of 100 g of protein yields more than 100 g of free, intact AAs. For example, the enzymatic hydrolysis of 100 g of protein in human skeletal muscle or pork meat generates 118 g of intact AAs. This should be kept in mind when analyzing AA composition in proteins (including those in tissues, cells, and foods).

Chemical analyses of AAs

Many methods have been used to analyze free AAs, including gas, ion-exchange, and liquid chromatographies, as well as mass spectrometry. Most chromatographic procedures involve pre- or post-column derivatization with reagents that give rise to fluorescent or colorimetric AA derivatives for appropriate detection. To date, liquid chromatography is more widely used than gas chromatography for AA analysis. Among liquid chromatography methods, high-performance liquid chromatography involving the pre-column derivatization with *o*-phthaldialdehyde (OPA) is most popular for the analysis of free and peptide-bound AAs in foods and biological samples (Dai et al., 2014).

Derivatizing reagents for UV/VIS detection include butylisothiocyanate (BITC), benzylisothiocyanate (BZITC), 4-chloro-3,5-dinitrobenzotrifluoride (CDNBTF), 4-dimethylaminoazobenzene-4-sulfonyl chloride (DABSYL-Cl), diethyl ethoxymethylenemalonate (DEEMM), 2,4-dinitrofluorobenzene (Sanger's reagent; DNFB), 1-fluoro-2,4-dinitrobenzene (FDNB), *N,N*-diethyl-2,4-dinitro-5-fluoroaniline (FDNDEA), 1-fluoro-2-4-dinitrophenyl-5-L-alanine amide (Marfey's reagent; FDNPAA), 2,2-dihydroxyindane-1,3-dione (ninhydrin), 1,2-Naphthoquinone-4-sulfonate (NQS), 4-Nitrophenylisothiocyanate (NPITC), and phenylisothiocyanate (PITC). Derivatizing reagents for fluorescence detection include 3-(4-Carboxybenzoyl)-2-quinolinecarboxaldehyde (CBQCA), fluorescein-5-isothiocyanate (FITC), 9-fluorenylmethyl chloroformate (FMOC-Cl), fluorescamine (FRA), 4-chloro-7-nitrobenzofurazan [CNBF; Also known as 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl)], 4-fluoro-7-nitro-2,1,3-benzoxadiazole (NBD-F), naphthalene-2,3-dialdehyde [NDA; also known as naphthalene-2,3-dicarboxaldehyde (NDB)], OPA, 4,7-phenanthroline-5,6-dione (PAD; also known as phanquinone), and styrene-divinylbenzene (SBD-P). Derivatizing reagents for either UV/VIS or fluorescence detection include 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC), 4-chloro-7-nitrobenzofurazan (CNBF), and 1-dimethylaminonaphthalene-5-sulphonyl chloride (Dansyl-Cl). The use of a derivatizing reagent depends on the availability of laboratory equipment and sample type.

To analyze AAs in proteins and peptides, the polymers must first be completely hydrolyzed by appropriate acid, alkaline, or enzymatic methods. The standard method for the acid hydrolysis of proteins involves 6 M HCl at 110°C for 24 h (Dai et al., 2014). The major disadvantage of this procedure is the complete conversion of glutamine and asparagine into glutamate and aspartate, respectively, and the complete destruction of tryptophan, as noted previously. To overcome this problem, a mixture of proteases and peptidases (e.g., pronase E, prolidase, pyroglutamate aminopeptidase, carboxypeptidase A, and aminopeptidase M) is used to analyze glutamate, glutamine, aspartate and asparagine in proteins (Li et al., 2021), whereas the analysis of peptide-bound

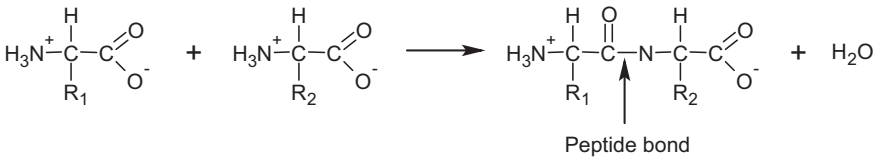


Fig. 7 Synthesis of a dipeptide from two amino acids (AAs). R1 and R2 represent the side chains of the two AAs. With the loss of one H_2O molecule (one "H" atom from the α-amino group of an AA and the "-OH" from the carboxyl group of the other AA), a peptide bond (-CONH-) is formed. This reaction is used to produce peptides through chemical manufacturing, and is catalyzed by enzymes in human cells. Peptides are generally stable in an aqueous solution.

tryptophan involves the hydrolysis of proteins in the presence of an alkaline solution (e.g., 4.2 M NaOH) at 105 °C for 20 h and 1% thiodiglycol (an antioxidant).

Classification of AAs

AAs can be classified as neutral (e.g., alanine and glycine), basic (arginine and lysine), and acidic (aspartate and glutamate) AAs, based on their net charges (e.g., 0, +1, and −1) at neutral pH (**Table 2**). At neutral pH, the side-chain $-NH_2$ group in asparagine, glutamine and citrulline is not ionized and, therefore, has no electric charge. Thus, in contrast to statements in some books and papers, these three AAs are not basic AAs at pH 7.0 or 7.4. Note that the net charges of AAs at acidic pH (e.g., pH 1.5 to 3.5 in the luminal fluid of the stomach of humans) are different from those at neutral pH. The net electrical charges of AAs have important implications for their intestinal absorption, transport by cells, and inter-organ metabolism in the body, as well as dietary supplementation with AAs. For example, the addition of an acidic or basic AA to a solution with a weak buffering capacity will substantially decrease or increase the pH of the solution, respectively. Likewise, arginine-HCl, but not arginine base, is used for intravenous infusion to humans to prevent acid-base imbalance.

Based on the side chain, AAs are classified as aliphatic AAs (e.g., branched-chain AAs); AAs with a hydroxylic (–OH) group (e.g., serine and threonine); AAs containing at least one sulfur atom (e.g., cysteine and methionine); AAs with an acidic group (e.g., aspartate and glutamate); AAs containing amide groups (e.g., asparagine and glutamine); AAs with a basic group (e.g., lysine and ornithine); AAs containing an aromatic group (e.g., phenylalanine and tryptophan); AAs with guainidino groups (e.g., arginine and homoarginine); AAs with imidazole groups (e.g., histidine and 3-methylhistidine); AAs with AAs with a seleno group (e.g., selenocysteine); AAs with an imino group (e.g., proline and 4-hydroxyproline); and phosphorylated AAs (e.g., phosphoserine and phosphotyrosine). Some AAs (e.g., desmosine, isodesmosine, and cystine) contain two or more amino groups and acidic groups. The side-chain of AAs affects their chemical properties and reactions (either spontaneous or enzyme-catalyzed).

Since 1912, AAs have long been classified as nutritionally essential or nonessential for humans (see Hou et al., 2015b for review). AAs that are not synthesized *de novo* by human tissues, which are known as nutritionally essential AAs (EAAs), must be provided in diets for maintenance, survival, and growth (Hou and Wu, 2018). Over the past three decades, a growing number of studies have revealed that endogenous syntheses of AAs are inadequate for the maximum growth or the optimum lactation, reproduction, and health of mammals (Hou et al., 2016). Thus, the term "nutritionally nonessential AAs" (NEAAs) has now been recognized as a misnomer in nutritional sciences (Hou and Wu, 2017). There is no compelling evidence that infants, adolescents and adults can synthesize *de novo* sufficient NEAAs (Hou et al., 2015b; Wu et al., 2013). Rather, results of recent clinical studies have shown that humans are not capable of forming sufficient amounts of all proteinogenic NEAAs and all nonproteinogenic NEAAs (Wu, 2022). For example, infants, particularly preterm neonates, have a limited ability to synthesize taurine, and suffer from severe cardiac and retinal damage when not fed a taurine-containing diet (Wright et al., 1986). In addition, adults fed solely plant-based diets do not produce sufficient taurine to meet optimum physiological needs, particularly under stress or diseased conditions (Wu, 2020). Furthermore, adults consuming little or no arginine exhibit metabolic defects, such as an inadequate synthesis of nitric oxide in the vasculature, reduced blood flow, and impaired fertility (Wu et al., 2021).

Conclusion and outlook

AAs (organic molecules containing both amino and acidic groups) are major molecules in humans and foods. The configuration of AAs is generally defined in reference to L- and D-glyceraldehyde. The absolute configuration of AAs is designated according to the R/S system, and includes *cis* and *trans* for those with a ring structure. AAs exist in both free and peptide-bound forms. All α-AAs, except for glycine and aminomalonic acid, can have both L- and D-isomers. Except for glycine, proteinogenic (protein-creating) AAs in human tissues are L-isomers. All L-AAs plus glycine are the most abundant physiological isomers and account for > 99.9% of total AAs in humans, but some D-AAs have important functions particularly in the neurological and immune systems. AAs in proteins plus peptides and free AAs account for approximately 97% and 3% of the total AAs in animals, respectively. Glycine is the most abundant peptide-bound AA in humans, followed by proline, glutamate, and arginine. Glutamine and alanine are the most and second most abundant free AA in the human plasma, respectively. Human liver, skeletal muscle, brain, and small-intestinal mucosa contain high concentrations of free aspartate plus taurine plus glutamine plus glutamate, free glutamine plus taurine plus glutamate, free glutamate plus glutamine, and free taurine plus glutamate, respectively. In some plant-sourced foods, select AAs (e.g., glutamine and asparagine) are highly abundant in the free form, with free AAs representing 34.4% and 28.5% of the total AAs in potatoes and sweet potatoes, respectively. Animal-sourced foods generally contain high amounts and proper ratios of all proteinogenic AAs. Of note, meat is an abundant source of taurine, β-alanine and 4-hydroxyproline, as well as carnosine (β-alanyl-L-histidine) and anserine (β-alanyl-L-1-methylhistidine), whereas plants lack taurine, carnosine and anserine and contain only a small or negligible amount of β-alanine and 4-hydroxyproline (Wu, 2020). Different AAs have different chemical properties (e.g., solubility, stability, melting points, reactions, taste, and electrical charges). Depending on the availability of laboratory instruments, many methods are available for the analysis of free and peptide-bound AAs, with the precolumn derivatization with OPA being the most widely used. AAs can be classified as neutral, basic, and acidic AAs based on their net charges at neutral pH, or as aliphatic, sulfur-containing, hydroxylated, amidated, carboxylated, phosphorylated, guanidino, aromatic, imidazole, seleno, and

secondary amino acids. Much needs to be done about the content of AAs (particularly glutamate, glutamine, aspartate, asparagine, proline, hydroxyprolines, cysteine, and tryptophan) in foods. Knowledge of AA chemistry and classification is essential for the understanding of human nutrition and health.

Acknowledgments

Work is the author's laboratory is supported by Texas A&M AgriLife Research Hatch Project H-8200 and by Agriculture and Food Research Initiative Competitive Grants from the USDA National Institute of Food and Agriculture.

References

Adibi, S.A., Mercer, D.W., 1973. Protein digestion in human intestine as reflected in luminal, mucosal, and plasma amino acid concentrations after meals. J. Clin. Invest. 52, 1586–1594.
Ahlman, B., Leijonmarck, C.-E., Wernerman, J., 1993. The content of free amino acids in the human duodenal mucosa. Clin. Nutr. 12, 266–271.
Barle, H., Ahlman, B., Nyberg, B., et al., 1996. The concentrations of free amino acids in human liver tissue obtained during laparoscopic surgery. Clin. Physiol. 16, 217–227.
Bergström, J., Furst, P., Norée, L.D., Vinnars, E., 1974. Intracellular free amino acid concentration in human muscle tissue. J. Appl. Physiol. 36, 693–697.
Borgenvik, M., Nordin, M., Mattsson, C.M., et al., 2012. Alterations in amino acid concentrations in the plasma and muscle in human subjects during 24 h of simulated adventure racing. Eur. J. Appl. Physiol. 112, 3679–3688.
Dai, Z.L., Wu, Z.L., Jia, S.C., Wu, G., 2014. Analysis of amino acid composition in proteins of animal tissues and foods as pre-column *o*-phthaldialdehyde derivatives by HPLC with fluorescence detection. J. Chromatogr. B 964, 116–127.
Friedman, M., 1999. Chemistry, biochemistry, nutrition, and microbiology of lysinoalanine, lanthionine, and histidinoalanine in food and other proteins. J. Agric. Food Chem. 47, 1295–1319.
Greenstein, J.P., Winitz, M., 1961. Chemistry of Amino Acids. John Wiley, New York.
He, W.L., Wu, G., 2020. Metabolism of amino acids in the brain and their roles in regulating food intake. Adv. Exp. Med. Biol. 1265, 167–185.
Hou, Y.Q., Wu, G., 2017. Nutritionally nonessential amino acids: a misnomer in nutritional sciences. Adv. Nutr. 8, 137–139.
Hou, Y.Q., Wu, G., 2018. Nutritionally essential amino acids. Adv. Nutr. 9, 849–851.
Hou, Y.Q., Wang, L., Yi, D., Wu, G., 2015a. N-acetylcysteine and intestinal health: a focus on mechanisms of its actions. Front. Biosci. 20, 872–891.
Hou, Y.Q., Yin, Y.L., Wu, G., 2015b. Dietary essentiality of "nutritionally nonessential amino acids" for animals and humans. Exp. Biol. Med. 240, 997–1007.
Hou, Y.Q., Yao, K., Yin, Y.L., Wu, G., 2016. Endogenous synthesis of amino acids limits growth, lactation and reproduction of animals. Adv. Nutr. 7, 331–342.
Hou, Y.Q., He, W.L., Hu, S.D., Wu, G., 2019. Composition of polyamines and amino acids in plant-source foods for human consumption. Amino Acids 51, 1153–1165.
Hu, S.D., He, W.L., Wu, G., 2021. Hydroxyproline in animal metabolism, nutrition, and cell signaling. Amino Acids. https://doi.org/10.1007/s00726-021-03056-x.
Li, P., Wu, G., 2020. Composition of amino acids and related nitrogenous nutrients in feedstuffs for animal diets. Amino Acids 52, 523–542.
Li, P., He, W.L., Wu, G., 2021. Composition of amino acids in foodstuffs for humans and animals. Adv. Exp. Med. Biol. 1332, 189–210.
Meister, A., 1965. Biochemistry of Amino Acids. Academic Press, New York.
Perry, T.L., Hansen, S., Berry, K., Mok, C., Lesk, D., 1971. Free amino acids and related compounds in biopsies of human brain. J. Neurochem. 18, 521–528.
Ryan, W.L., Carver, M.J., 1966. Free amino acids of human foetal and adult liver. Nature 212, 292–293.
Schmelzer, C.E.H., Hedtke, T., Heinz, A., 2020. Unique molecular net works: formation and role of elastin cross-links. IUBMB Life 72, 842–854.
Wright, C.E., Tallan, H.H., Lin, Y.Y., Gaull, G.E., 1986. Taurine: biological update. Annu. Rev. Biochem. 55, 427–453.
Wu, G., Wu, Z.L., Dai, Z.L., et al., 2013. Dietary requirements of "nutritionally nonessential amino acids" by animals and humans. Amino Acids 44, 1107–1113.
Wu, G., 2022. Amino Acids: Biochemistry and Nutrition, second ed. CRC Press, Boca Raton.
Wu, G., Cross, H.R., Gehring, K.B., et al., 2016. Composition of free and peptide-bound amino acids in beef chuck, loin, and round cuts. J. Anim. Sci. 94, 2603–2613.
Wu, G., Meininger, C.J., McNeal, C.J., Bazer, F.W., Rhoads, J.M., 2021. Role of L-arginine in nitric oxide synthesis and health in humans. Adv. Exp. Med. Biol. 1332, 167–187.
Wu, G., 2018. Principles of Animal Nutrition. CRC Press, Boca Raton.
Wu, G., 2020. Important roles of dietary taurine, creatine, carnosine, anserine and hydroxyproline in human nutrition and health. Amino Acids 52, 329–360.

Amino acids: Metabolism

Guoyao Wu, Department of Animal Science and Faculty of Nutrition, Texas A&M University, College Station, TX, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of P.W. Emery, Amino acids: Metabolism, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 72–78, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00010-6.

Introduction 24
Synthesis of AAs in humans 24
Overall view of AA synthesis in humans 24
Cell- and tissue-specific synthesis of AAs in humans 24
Liver 25
Skeletal muscle, heart, brain, lungs, and white adipose tissue 26
The small intestine 27
Kidneys 27
Endothelial cells, smooth muscle cells, macrophages, and lymphocytes 27
Sense organs 27
Placentae, mammary glands, ovaries, testes, and pancreas 27
Degradation of AAs in humans 27
Overall view of AA degradation in humans 27
Cell- and tissue-specific degradation of AAs in humans 28
Liver 29
Skeletal muscle, heart, brain, lungs, and white adipose tissue 29
The small intestine 30
Kidneys 30
Endothelial cells, smooth muscle cells, macrophages, and lymphocytes 30
Sense organs 30
Placentae, mammary glands, ovaries, testes, and pancreas 30
Ammonia detoxification via ureagenesis in the liver and small intestine of humans 30
Production and toxicity of ammonia in humans 30
The urea cycle in the liver and small intestine 31
Protein turnover in humans 32
Metabolites of AAs in humans 32
Summary and outlook 33
Acknowledgments 34
References 34

Key points

- Humans can synthesize *de novo* the following nine proteinogenic AAs (alanine, arginine, aspartate, asparagine, glutamate, glutamine, glycine, proline, and serine), as well as non-proteinogenic AAs (e.g., citrulline, ornithine, and taurine) in a cell- and tissue-specific manner.
- Humans can convert methionine into cysteine in the liver; phenylalanine into tyrosine mainly in the liver and kidneys; and serine plus selenium into selenocysteine as a component of selenoproteins in cells during the mRNA translation.
- Humans can degrade all proteinogenic AAs and physiological non-proteinogenic AAs via compartmentalized metabolic pathways and inter-organ cooperation.
- In the mucosa of the human small intestine, dietary glutamate, glutamine, and aspartate are the major metabolic fuels after feeding, whereas glutamine is the primary energy substrate in the post-absorptive state.
- Microbes in the small and large intestines actively degrade dietary and endogenous AAs.
- Products of AA catabolism and conjugation include CO_2, water, ammonia, urea, nitric oxide (NO), carbon monoxide (CO), hydrogen sulfide (H_2S), γ-aminobutyrate (GABA), heme, serotonin, melanin, melatonin, histamine, polyamines, glutathione, creatine, phospholipids, and nucleic acids.
- Based on their metabolic fates, AAs can be classified as ketogenic, glucogenic, and ketogenic plus glucogenic.
- The mechanistic target of rapamycin (MTOR) cell signaling pathway integrates anabolic factors to promote protein synthesis in cells, whereas ATP and ubiquitin-dependent proteasomes as well as autophagy and lysosomal proteases are responsible for the bulk of intracellular proteolysis.

- Elevated concentrations of ammonia in the plasma are highly toxic to the central nervous system, and this metabolite must be removed as urea primarily via the periportal hepatocytes of the liver and, to a much lesser extent, the enterocytes of the small intestine.
- In adults, adequate intakes of dietary protein and AAs are essential to mitigate muscle loss during aging and under catabolic conditions.

Introduction

In healthy adult humans, concentrations of total amino acids (AAs) in their plasma are increased after the oral consumption of a diet [e.g., soy or milk; providing 0.41 g protein/kg body weight (BW)] to peak values at approximately 1 h post-feeding, maintained at plateau until 3 h post-feeding, decreased thereafter gradually to baseline levels at 6 h post-feeding, and then maintained constant until 8 h post-feeding (Bos et al., 2003). The post-prandial changes in the plasma concentrations of total branched-chain AAs (BCAAs) and total glucogenic AAs (alanine, glutamate, glutamine, glycine, proline, serine, and threonine) followed the same pattern as those for the total AAs. In healthy term neonates (5–7 days of age), concentrations of total AAs and many AAs (including BCAAs, arginine, ornithine, and proline) in their plasma are increased after the oral consumption of human milk and cow's milk formula (providing 0.20 and 0.4–0.8 g protein/kg BW, respectively) to peak values at approximately 0.5 and 1 h post-feeding, respectively, and decreased thereafter gradually until 3 h post-feeding (Tikanoja, 1982; Tikanoja et al., 1982a). Similar phenomena occur in preterm infants (4–12 days of postnatal age; gestational days of 34–37 weeks) fed under the same conditions (Tikanoja et al., 1982b). In both adults and neonates, the patterns of AAs in the plasma differ remarkably from those in diets as a result of tissue-specific AA metabolism (Sarkar et al., 2021; Wu, 2022). The maintenance of AA homeostasis in the body is regulated by many metabolic processes, including gastrointestinal protein digestion and the intestinal absorption of its products (AAs and small peptides), the inter-organ transport of AAs, intracellular protein turnover (protein synthesis and degradation), post-translational protein modifications, and the cell-specific synthesis and catabolism of AAs (Phang, 2019; Tsikas, 2021; Wu, 2022). These aspects of AA metabolism are highlighted in this article.

Synthesis of AAs in humans

Overall view of AA synthesis in humans

Human and cow milk-based formulas contain relatively low concentrations of arginine and glycine and, therefore, do not provide an adequate amount of both AAs to infants (Davis et al., 1994; Wu, 2022). Likewise, individuals consuming plant-based foods (which are all deficient in glycine and proline, Hou et al., 2019; Li and Wu, 2020) do not ingest a sufficient quantity of these two nutrients. Furthermore, because taurine [a physiologically essential AA (Wu, 2020a)] is absent from plants, vegetarians have little or no intake of this AA from their diets. Thus, humans must be able to synthesize certain AAs for survival, growth, and development. In support of this view, the optimal production of collagen (the most abundant protein in the body) requires the endogenous formation of glycine and proline *de novo* from other AAs under both healthy and diseased (e.g., burning and infection) conditions (Hamanaka and Mutlu, 2021; Meléndez-Hevia et al., 2009). Finally, most of dietary aspartate, glutamate, and glutamine are degraded by the small intestine during their first pass into the portal vein, and, therefore, these AAs (which are among the most abundant AAs in the body) must be produced from other AAs in humans (Wu, 1998). AA syntheses require energy and the so-called nutritionally essential AAs (EAAs), and these processes are energetically inefficient and also produce ammonia (a toxic substance at elevated concentrations). Under normal feeding conditions, AAs that are synthesizable *de novo* in animal cells (AASAs) play important roles in mammalian nutrition, metabolism and redox signaling, but the amounts are insufficient for the maximum growth and optimum health of humans (Wu, 2022). AAs that are not synthesized *de novo* in human tissues, which are known as nutritionally essential AAs, must be provided in diets to ensure survival, growth, development, and health (Wu, 2022). The century-old notion of "nonessential AAs" has now been recognized as a misnomer (Hou and Wu, 2017). Likewise, the long-standing nutritional concept of "ideal protein" ignores all AASAs and should no longer be used in research or feeding practices.

Cell- and tissue-specific synthesis of AAs in humans

Humans can synthesize *de novo* the following nine proteinogenic AAs (alanine, arginine, aspartate, asparagine, glutamate, glutamine, glycine, proline, and serine), as well as non-proteinogenic AAs (e.g., citrulline, ornithine, taurine, and GABA) in a cell- and tissue-specific manner (Wu, 2022). In addition, humans can form (a) citrulline from glutamate, glutamine and proline in the small intestine; (b) taurine from cysteine in the liver; (c) tyrosine from phenylalanine in the liver, kidneys, and skin; (d) phosphoethanolamine from serine; and (e) certain L-AAs (e.g., L-serine and L-aspartate) into D-AAs (e.g., D-serine and D-aspartate); and some D-AAs (e.g., D-methionine and D-phenylalanine) into L-AAs (e.g., L-methionine and L-phenylalanine). The overall metabolic pathways are illustrated in **Fig. 1**, and the amounts of AAs synthesized in the whole-body of humans are summarized in

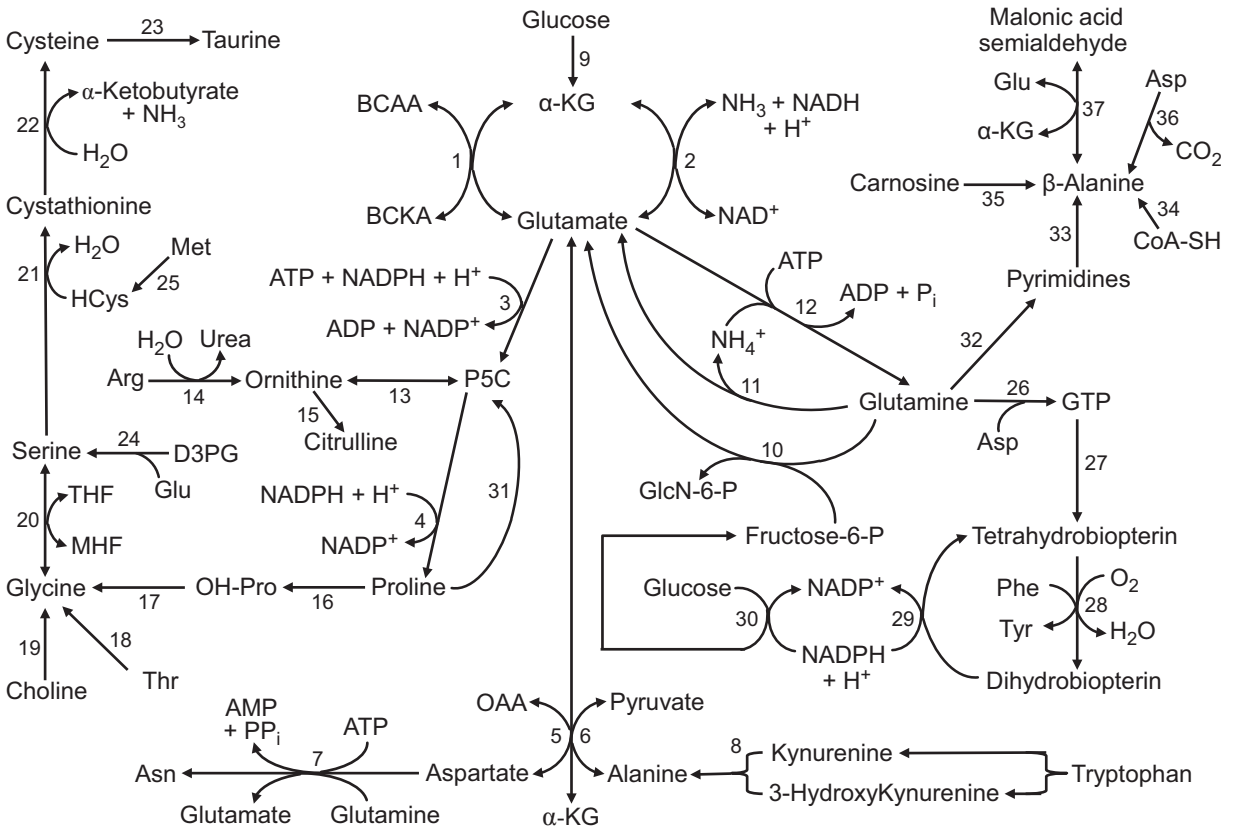


Fig. 1 Synthesis of amino acids in humans in a cell- and tissue-dependent manner. The enzymes catalyzing the indicated reactions are: (1) BCAA transaminase; (2) glutamate dehydrogenase; (3) P5C synthase; (4) P5C reductase; (5) aspartate transaminase; (6) alanine transaminase; (7) asparagine synthetase; (8) enzymes for tryptophan catabolism; (9) enzymes for converting glucose into α-KG; (10) glutamine:fructose-6-phosphate transaminase; (11) phosphate-activated glutaminase; (12) glutamine synthetase; (13) ornithine aminotransferase; (14) arginase; (15) ornithine carbamoyltransferase; (16) enzymes for protein synthesis, hydroxylation of peptide-bound proline, and protein degradation; (17) enzymes for converting 4-hydroxyproline into glycine; (18) enzymes for converting threonine into glycine; (19) enzymes for converting choline into glycine; (20) serine hydroxymethyltransferase; (21) cystathionine β-synthase; (22) cystathionine γ-lyase; (23) enzymes for converting cysteine into taurine; (24) enzymes for converting D-3-phosphoglycerate and glutamate into serine; (25) enzymes for methionine catabolism; (26) enzymes for GTP synthesis; (27) enzymes for tetrahydrobiopterin synthesis; (28) phenylalanine hydroxylase; (29) dihydrobiopterin reductase; (30) enzymes of the pentose cycle; (31) proline oxidase; (32) enzymes for pyrimidine synthesis; (33) enzymes for pyrimidine catabolism; (34) enzymes for coenzyme A (CoA-SH) catabolism; (35) carnosinase; (36) aspartate decarboxylase; (37) β-alanine-α-KG transaminase, with malonic acid semialdehyde being produced from propionyl-CoA and malonyl-CoA semialdehyde. BCAA, branched-chain amino acid; BCKA, branched-chain α-ketoacid; D3PG, D-3-phosphoglycerate; GlcN-6-P, glucosamine-6-phosphate; HCys, homocysteine; MTH, N^5-N^{10}-methylene tetrahydrofolate; OH-Pro, 4-hydroxyproline; OAA, oxaloacetate; P5C, Δ^1-Pyrroline-5-carboxylate; THF, tetrahydrofolate. Reproduced from Wu (2013) with permission.

Table 1. Much evidence shows that endogenous AA production is altered during development and in diseased states (Wu, 2022) (**Fig. 2**).

Microbes in the small intestine have a net consumption of dietary AAs (Dai et al., 2011, 2015). Although microbes in the large intestine can convert free AAs and small peptides into proteins, the microbial proteins are not absorbed into the blood circulation and, therefore, have no nutritive value (Wu, 2013). In the fed and post-absorptive states, only a small or nutritionally insignificant amount of free AAs is absorbed from the lumen of the large intestine into the blood circulation of humans (Bergen and Wu, 2009).

Liver

The liver is an active organ in mammalian AA metabolism (Hou et al., 2020). Under physiological conditions, there is no net synthesis of citrulline or arginine in the human liver. This is because ornithine-derived citrulline is immediately converted into arginine via argininosuccinate synthase and lyase, and the arginine formed via the urea cycle is rapidly hydrolyzed to urea plus ornithine by arginase in hepatocytes. The liver cannot synthesize proline from glutamine or glutamate due to its lack of pyrroline-5-carboxylate (P5C) synthase. By contrast, the human liver can form proline from arginine via arginase, ornithine aminotransferase (OAT), and P5C reductase (Wu and Morris, 1998). The human liver can form (a) aspartate, asparagine, glutamate, glutamine, glycine, and serine from other AAs; (b) cysteine from methionine, taurine from cysteine, and cholyltaurine from taurine plus bile acid; (c) tyrosine from phenylalanine (via phenylalanine hydroxylase); and (d) phenylacetylglutamine from phenylalanine,

Table 1 Rates of the syntheses of amino acids (AAs) in humans.

AA	*Species*	*Nutritional state*	*Flux in arterial plasma (μmol/kg of body weight/h)*	*Endogenous synthesis (μmol/kg of body weight/h)*
Arginine	Human infants	TPN	209	15
	Preterm infants	Fed (enterally)	461	42 (from proline)
	Adult humans	Fed (enterally)	73	5.2
Glutamine	Preterm human infants	Fasted	755	655
	Preterm human infants	Fed (enterally)	644	529
	AGA infants	Fasted	545	472
		Fed (enterally)	494	423
	SGA infants	Fasted	791	679
		Fed (enterally)	677	569
	Adult humans	Post-absorptive	348	245
Glutamate	Human infants	Fed (enterally)	607 (423–791)	482
	Adult humans	Post-absorptive	83	67
Glycine	Adult humans	Fed (enterally)	458	72
Proline	Adult humans	Fed (enterally)	128	39
Serine	Adult humans	Overnight fasted	150	110

AGA = appropriate for-gestational-age; SGA = small-for-gestational-age; TPN = total parenteral nutrition.
Adapted from Wu (2022).

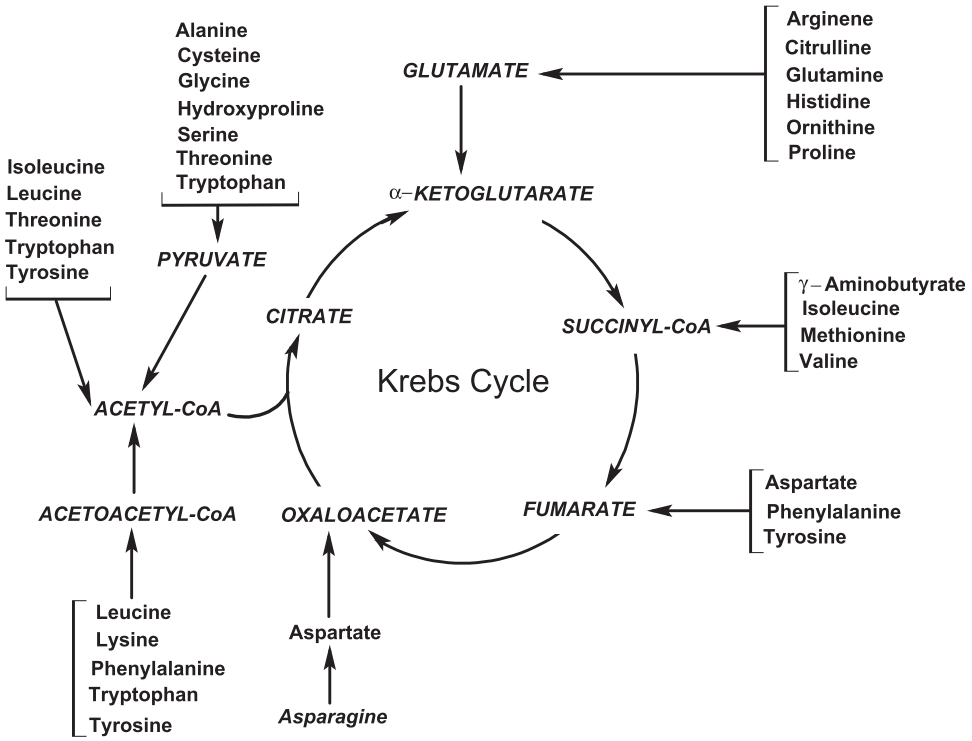


Fig. 2 An overall view of amino acid catabolism in humans. Different metabolic pathways for the catabolism of amino acids converge to common intermediates (α-KG, oxaloacetate, and pyruvate) that feed into the Krebs cycle in tissues. Different amino acids have different metabolic fates in the body. Reproduced from Wu (2013) with permission.

acetyl-CoA, and glutamine. This organ has a limited ability to convert BCAAs into other AAs due to the near absence of BCAA transaminase activity under physiological conditions (Harper et al., 1984). Note that in the human liver, glutamine synthase is expressed primarily in perivenous hepatocytes to convert the ammonia that escapes periportal hepatocytes plus glutamate into glutamine to maximize the removal of ammonia from the blood. In contrast to adults and term infants, preterm humans have a limited ability to synthesize taurine from cysteine in the liver.

Skeletal muscle, heart, brain, lungs, and white adipose tissue

In humans, alanine, aspartate, and glutamate can be synthesized from BCAAs (the donors of the amino group) and glucose-derived carbon skeletons in the skeletal muscle, heart, brain, lungs, and white adipose tissue (Brosnan and Brosnan, 2006). These reactions

require BCAA transaminase, glutamine synthetase, glutamate transaminase, and alanine transaminase. A small amount of aspartate can be further converted into asparagine via asparagine synthetase in these tissues. In fasting individuals, alanine and glutamine account for approximately 50% of total AAs released by their skeletal muscle (Marliss et al., 1971). Because of its large mass (45% and 40% of BW in adults and neonates, respectively), skeletal muscle is the major source of both alanine and glutamine in post-absorptive humans (Ryan et al., 2021). The synthesis of glutamine from NH_3 and glutamate plays a major role in removing ammonia from the blood. Furthermore, the skeletal muscle, heart, brain, lungs, and white adipose tissue of humans can interconvert serine and glycine.

The small intestine

The small intestine is the terminal site for the digestion of dietary proteins and the absorption of its products (small peptides and free AAs). In addition, as for many other mammals, the small intestine of humans can form alanine, arginine, aspartate, asparagine, citrulline, ornithine, and proline from glutamate and glutamine; glutamate from BCAAs plus glucose, glutamine, and proline; glycine from serine; tyrosine from phenylalanine; and methionine from homocysteine (Wu, 2022). The gut releases alanine, arginine, citrulline, ornithine, and proline in the post-absorptive state. The mammalian small intestine has a limited ability to synthesize glutamine due to a low activity of glutamine synthetase. Citrulline bypasses the liver and only about 10% of arginine in the portal vein is taken up by this organ so as to maximize the bioavailability of diet- and gut-derived citrulline and arginine to the extrahepatic tissues. Because of the complex compartmentation of AA metabolism involving both the mitochondrion and cytoplasm, extracellular ornithine is poorly utilized for citrulline or arginine synthesis in enterocytes. These reactions occur in both enterocytes and bacteria. Microorganisms in the lumen of the gastrointestinal tract can synthesize all AAs from ammonia, carbohydrates, and sulfur (Dai et al., 2011).

Kidneys

In humans, the kidneys can synthesize alanine, aspartate, and glutamate from BCAAs, glutamine, and 4-hydroxyproline, while interconverting phenylalanine into tyrosine, serine into glycine, and proline into ornithine (Hu et al., 2022; Li et al., 2020). In contrast to many other mammals (e.g., rats and rabbits), human kidneys lack glutamine synthetase (Wu, 2022). Alanine transaminase activity is present in the kidneys of humans. The kidneys contain argininosuccinate synthase and argininosuccinate lyase for converting citrulline into arginine. These two enzymes are strategically localized within the proximal convoluted tubules which express little arginase activity. In adults, approximately 60% of net arginine synthesis from citrulline occurs in the kidneys.

Endothelial cells, smooth muscle cells, macrophages, and lymphocytes

These cells can synthesize alanine, aspartate, and glutamate from BCAAs and glutamine, and convert citrulline into arginine to conserve arginine for the production of nitric oxide (NO) at the expense of aspartate (Wu, 2022). This is commonly referred to as the arginine–citrulline cycle. By contrast, all these cells do not synthesize: (1) citrulline from glutamine, glutamate and proline; (2) tyrosine from phenylalanine; or (3) taurine from cysteine.

Sense organs

In humans, sense organs (the eyes, ears, nose, tongue, and skin) can synthesize alanine, aspartate, glutamate, and 4-hydroxyproline, while converting citrulline into arginine, phenylalanine into tyrosine, serine into glycine, arginine into ornithine, and ornithine into proline (Wu, 2020b). The eyes and skin also produce asparagine, glutamine and glycine from BCAAs, glutamate, and 4-hydroxyproline, but little is known about these reactions in other sense organs. In addition, retinal cells in the eyes can decarboxylate glutamate to generate GABA. All the sense organs do not convert: (1) methionine into cysteine; or (2) glutamine, glutamate and proline into citrulline.

Placentae, mammary glands, ovaries, testes, and pancreas

In humans, these organs can synthesize glutamate, glutamine, alanine, aspartate, and asparagine from BCAAs, form ornithine and proline from arginine, interconvert serine into glycine, and produce 4-hydroxyproline from proline (via collagen synthesis) (Wu, 2022). During lactation, the mammary glands contributes to a large amount of glutamate, glutamine, aspartate, and proline for milk synthesis. The human placentae, ovaries, and testes lack both phenylalanine hydroxylase activity and P5C synthase, whereas P5C synthase is also absent from the pancreas. By contrast, both the mammary glands and the pancreas can form tyrosine from phenylalanine.

Degradation of AAs in humans

Overall view of AA degradation in humans

Degradation AAs occurs in humans regardless of their nutritional states but at higher rates immediately after feeding than during fasting (Jungas et al., 1992). Because physiological concentrations of free AAs are readily soluble in water but their accumulation in the body can increase the extracellular and intracellular osmolarity, most of the excessive AAs are degraded in a cell- and tissue-specific manner. The major products of AA catabolism are CO_2, water, ammonia, urea, uric acid, pyruvate, and acetyl-CoA, with the metabolic fates of pyruvate (e.g., glucose or fatty acid synthesis) and acetyl-CoA (oxidation to CO_2 or fatty acid synthesis),

depending on nutritional, physiological, and pathological states (Wu, 2022). Reactions for initiating AA degradation include transamination, amidotransferation, cleavage, condensation, deaminated oxidation (FAD-dependent), deamination, decarboxylation, dehydration, dehydrogenation [$NAD(P)^+$-dependent], dioxygenation, hydrolysis (via amidinohydrolysis or deamidation), hydroxylation, one-carbon unit transfer, oxidation (FAD-dependent), oxidative deamination, and reduction. AA metabolism generates bioactive metabolites (e.g., NO, taurine, creatine, polyamines, dopamine, serotonin, and H_2S) with enormous physiological importance. Thus, in humans consuming a normal meal with protein, carbohydrate and lipids, the catabolism of any excess AAs takes precedence over the degradation of carbohydrates and fatty acids. Aspartate, glutamate and glutamine are major metabolic fuels for the human small intestine as noted previously and help to regulate the homeostasis of these AAs in plasma (Bertolo and Burrin, 2008; Rhoads and Wu, 2009). However, the use of AAs for ATP production in the whole body is energetically inefficient in mammals (e.g., 47% for glutamate, 49% for threonine, and 24% for methionine; Wu, 2022).

Cell- and tissue-specific degradation of AAs in humans

In humans, most AAs are degraded in a tissue-, cell-, and concentration-dependent manner, but some pathways (e.g., NO synthesis, the hydrolysis of arginine by arginase, the oxidation of proline to P5C by proline oxidase, and the decarboxylation of ornithine to putrescine) occur in nearly all cell types. There are also diurnal changes in the catabolism of some AAs independent of food intake, such as the conversion of tryptophan to melatonin during the dark period of the light–dark cycle (Armstrong, 1999). Transport into cells via specific and multiple systems is the first step for the utilization of AAs (Bröer and Bröer, 2017; Closs et al., 2000). The complete catabolic pathways for most AAs may involve both the cytosol and mitochondria (Curthoys and Watford, 1995; Morris, 2009). Because of the complex compartmentation of AA metabolism, extracellularly- and intracellularly-derived AAs may have very different metabolic fates. The cooperation of multiple organs is required for the complete oxidation of AAs to CO_2, NH_3 (a gas) and water. In an aqueous solution, free ammonia (NH_3) is at equilibrium with ammonium ion (NH_4^+; pKa = 9.2). At pH 7.4 and 37 °C, approximately 1.6% and 98.4% of ammonia exists as free NH_3 and ammonium ion, respectively (Wu, 2022). In this article, free NH_3 and NH_4^+ are collectively referred to as ammonia. NH_3 is further oxidized primarily to urea in humans and, to a lesser extent, uric acid. Note that AA transaminases and AA decarboxylases require vitamin B_6 as an essential cofactor for their catalytic activities. Other vitamin B_6-dependent enzymes include threonine aldolase, some D-AA racemases, cystathionase, cystathionine synthase, the glycine cleavage system, kynureninase (kynurenine hydrolase), kynurenine transaminase, methionine synthetase, and serine hydroxymethyl transferase. Another unique feature of AA degradation is that tetrahydrobiopterin is essential for the hydroxylation of arginine, phenylalanine, tyrosine, and tryptophan by NO synthase, phenylalanine hydroxylase, tyrosine hydroxylase, and tryptophan hydroxylase, respectively. The physiological half-lives of plasma AAs in healthy adult humans are relatively short; for example, L-arginine, L-citrulline, and L-glutamine: ~1.0 h; L-glutamate: 0.5 h; glycine: 1.05 h; and L-leucine: 0.57–1.0 (Wu, 2022). The rates of whole-body catabolism of some AAs in humans are summarized in **Table 2**, and are altered during development and in diseased states. For example, both acidosis and cancer stimulate the whole-body catabolism of glutamine and BCAAs in humans to regulate pH and energy homeostasis (Curthoys and Watford, 1995). Available evidence shows that infants, children, and adults can tolerate dietary intakes of at least 4.7, 5.1, and 3.5 g protein/kg BW/day (Wu, 2016), as well as high intakes of various AAs (Blachier et al., 2021; McNeal et al., 2018).

Table 2 Rates of the catabolism of amino acids (AAs) in humans.

AA	*Species*	*Nutritional state*	*Endogenous flux (μmol/kg of body weight/h)*	*Degradation (μmol/kg of body weight/h)*
Arginine	Human infants	Total parenteral nutrition	173	34
	Preterm infants	Fed (enterally)	461	29
	Adult humans	Fed (enterally)	58	20
Cysteine	Adult humans	Fasted for 12 h	36–47	5.3
Glutamine	Preterm human infants	Fasted	755	463
	Preterm human infants	Fed (enterally)	619	327
	AGA infants	Fasted	545	253
		Fed (enterally)	488	196
	SGA infants	Fasted	791	499
		Fed (enterally)	671	379
	Adult humans	Post-absorptive	348	306
Glutamate	Human infants	Fed (enterally)	605 (421–789)	155
	Adult humans	Post-absorptive	83	15
Glycine	Adult humans	Fed (enterally)	445	81

Table 2 Rates of the catabolism of amino acids (AAs) in humans.—cont'd

AA	*Species*	*Nutritional state*	*Endogenous flux (μmol/kg of body weight/h)*	*Degradation (μmol/kg of body weight/h)*
Leucine	AGA infants	Fed (enterally)	98	51
	SGA infants	Fed (enterally)	151	67
	Adult humans	Fasted for 12 h	101	15
		Fed (enterally)	96	40
		Post-absorptive (14-h fast)	87	16
Lysine	Adult humans	Fed (enterally)	71	28
		Fed (2–100 mg Lys/kg BW/day)	46	5–31
Methionine	Adult humans	Fasted for 12 h	16	12
		Fasted for 10 h	19	10
		Fed	41	34
Phe	Adult humans	Fasted for 12 h	45	9.3
		Fed (enterally)	44–52	15
Proline	Preterm infants	Fed (enterally)	237	–
	Adult humans	Fed (enterally)	115	52
Serine	Adult humans	Fasted (overnight)	150	115
Threonine	Preterm infants	Fed (enterally)	211	15
	Adult humans	Fed (enterally)	90	16
Tyrosine	Adult humans	Fasted for 12 h	42–48	7.5
		Fed (enterally)	31–40	10
Valine	Adult humans	Post-absorptive (14 h fast)	80	12

BW = body weight; Phe = phenylalanine.
Adapted from Wu (2022).

Liver

In humans, the liver actively degrades most AAs (including lysine and methionine), but neither BCAAs due to the near absence of BCAA transaminases nor extracellular citrulline because of no uptake. Thus, concentrations of aromatic AAs (tyrosine, phenylalanine, and tryptophan) in the plasma of individuals with liver failure are usually increased due to a limited inability of the liver to degrade these AAs. Of particular note, there is metabolic zonation for hepatic AA catabolism. For example, glutamate and aspartate in the arterial and portal venous blood escape periportal hepatocytes and enter perivenous hepatocytes for glutamine synthesis (Brosnan and Brosnan, 2013). AA-derived ammonia is detoxified primarily as urea in hepatocytes via the urea cycle (see the section below). This organ is the only site for the complete catabolism of all α-ketoacids of AAs [including branched-chain α-ketoacids (BCKAs) of BCAAs] released from extrahepatic tissues. The liver produces glucose from alanine and other glucogenic AAs, constituting the glucose-alanine cycle between this organ and other tissues (primarily skeletal muscle). The α-ketoacids of AAs can be converted into glucose, ketone bodies, or fatty acids in the human liver (Table 3), depending on nutritional, physiological and pathological states.

Skeletal muscle, heart, brain, lungs, and white adipose tissue

In humans, these tissues actively transaminate BCAAs and α-KG to form BCKAs and glutamate (Chen et al., 2020; Huston et al., 2005; He and Wu, 2020). Glutamate then undergoes transamination with pyruvate and oxaloacetate to generate alanine and aspartate, respectively, or amidated with NH_4^+ to yield glutamine. All of these tissues contain glutaminase (a mitochondrial enzyme) to hydrolyze glutamine into glutamate and NH_4^+, and, therefore, the intracellular glutamine-glutamate cycle, as originally proposed for rat and avian skeletal muscle (Wu et al., 1991). The balance of this metabolic cycle is the major determinant of glutamine release from the tissues in response to nutritional and physiological needs as well as pathological alterations. Interestingly, concentrations

Table 3 Ketogenic, glucogenic, and ketogenic plus glucogenic amino acids in humans.

Group	*Amino acids (AAs)*
Ketogenic AAs[a]	Leucine, lysine, and taurine
Glucogenic AAs[b]	Alanine, γ-aminobutyrate, arginine, asparagine, aspartate, citrulline, cysteine, glutamate, glutamine, glycine, histidine, 4-hydroxyproline, 3-hydroxyproline, methionine, ornithine, phosphoarginine, proline, serine, and valine
Ketogenic plus glucogenic AAs[c]	Isoleucine, phenylalanine, threonine, tryptophan, and tyrosine

[a]Forming only acetyl-CoA.
[b]Forming pyruvate and 4–5 carbon unit metabolites (e.g., oxaloacetate and α-KG).
[c]Forming acetyl-CoA, pyruvate, and 4–5 carbon metabolites.

of the three branched-chain AAs (leucine, isoleucine, and valine) in the plasma of patients with liver failure are usually decreased possibly due to increased catabolism of these AAs in skeletal muscle and other extrahepatic tissues, as well as reduced food intake (Hou et al., 2020).

The small intestine

Enterocytes actively degrade most of proteinogenic AAs present in the lumen of the small intestine. Intestinal microbes are also capable of degrading dietary AAs (Dai et al., 2011, 2015). Thus, in healthy adult humans, the small intestine actively sequesters many diet-derived AAs during their first pass into the portal circulation (expressed as the % of AAs present in the lumen of the small intestine) mainly via catabolism: arginine, 40%; BCAAs, 20–30%; glutamate, 96%; glutamine, 64%; lysine, 19%; phenylalanine, 27%; and threonine, 18% (Chapman et al., 2013; Haisch et al., 2000; Matthews et al., 1993; Wu, 1998). Likewise, in preterm infants, the small intestine utilizes 73% of dietary glutamine (van der Schoor et al., 2010) and 77% of dietary aspartate (Corpeleijn et al., 2010) primarily for oxidation to CO_2. In addition to the mucosa of the small intestine, its microbes degrade all proteinogenic AAs and some nonproteinogenic AAs (Dai et al., 2015; Davila et al., 2013). Because all dietary AAs but taurine are degraded by the small intestine during their entry into the portal circulation, the pattern of AAs in the plasma differs substantially from that in the diet. Interestingly, among all AAs in the arterial blood, only glutamine is taken up by the small intestine in the post-absorptive state.

Kidneys

In humans, the kidneys plays an important role in the regulation of acid-base balance through hydrolyzing glutamine to glutamate and NH_3 (which takes up H^+ to form NH_4^+), as well as converting glutamate to α-KG and NH_3 (Li et al., 2020). α-KG can be used for renal gluconeogenesis or oxidation to CO_2, depending on nutritional and physiological states. The human kidneys also form tyrosine from phenylalanine, arginine from citrulline, ornithine from arginine, alanine/aspartate from glutamate, glycine from 4-hydroxyproline, and *S*-adenosylmethionine from methionine, while interconverting glycine into serine, proline into ornithine, proline and 4-hydroxyproline (via collagen synthesis and degradation), and serine and glutamate/glucose (via D-3-phosphoglycerate dehydrogenase and phosphatase pathways) (Wu, 2022). Furthermore, in humans, the kidneys convert arginine and glycine into ornithine and guanidinoacetate, with the latter being taken up primarily by the liver for conversion into creatine in the presence of *S*-adenosylmethionine as the methyl group donor.

Endothelial cells, smooth muscle cells, macrophages, and lymphocytes

These cells oxidize arginine to NO and citrulline for the endothelium-dependent vasodilation, innate immunity, and cell signaling, while generating a small amount of H_2S (a vasodilator and an oxidant) from cysteine (Wu, 2022). All these cells can also convert citrulline into arginine in the presence of aspartate; degrade BCAAs for the formation of glutamate; oxidize glutamine, glutamate, and aspartate as metabolic fuels; and converts glutamine to glutamate and glucosamine-6-phosphate via glutamine:fructose-6-phosphate transaminase for the synthesis of glycoproteins. Of particular note, glutamine is a major energy substrate in macrophages and lymphocytes, particularly under the conditions of immunological challenges (Li et al., 2007).

Sense organs

In humans, sense organs (the eyes, ears, nose, tongue, and skin) can degrade alanine, arginine, aspartate, glutamate, and BCAAs (Wu, 2020b). In addition, both the eyes and the skin can degrade asparagine, glutamine, glycine, serine, and tryptophan. There is evidence that the skin catabolize histidine, lysine, methionine, phenylalanine, proline, 4-hydroxyproline, and tyrosine. All these tissues can hydrolyze arginine to ornithine plus urea, while oxidizing arginine to NO and citrulline.

Placentae, mammary glands, ovaries, testes, and pancreas

In humans, these organs can extensively degrade BCAAs to form glutamate, glutamine, alanine, and aspartate, while hydrolyzing arginine to ornithine plus urea (Wu, 2022). All these tissues can oxidize arginine to NO plus citrulline and recycle citrulline to arginine in the presence of aspartate; interconvert ornithine and P5C into proline; and catabolize alanine, aspartate, glutamate, glutamine, glycine, 4-hydroxyproline, methionine, serine, tryptophan, and tyrosine. Furthermore, placentae, ovaries, testes, and pancreas oxidize proline to P5C by proline oxidase, whereas both mammary glands and pancreas can hydrolyze phenylalanine to tyrosine. Finally, the pancreas can convert guanidinoacetate into creatine, while degrading cysteine, GABA, histidine, and threonine (Wu, 2022). By contrast, mammary glands may have no or limited ability to degrade proline, and may also lack phosphate-activated glutaminase for hydrolyzing glutamine to glutamate plus ammonia.

Ammonia detoxification via ureagenesis in the liver and small intestine of humans

Production and toxicity of ammonia in humans

In humans, ammonia is produced from AA catabolism by tissues and cells, as well as the microorganisms of the gastrointestinal tract. The central nervous system is particularly sensitive to adverse effects of ammonia. Because multiple metabolic pathways remove ammonia from the blood, the concentrations of ammonia ($NH_3 + NH_4^+$) in the plasma of healthy children and adults are usually $\leq$35 and $\leq$ 30 μM, respectively (Wu, 2022). Hyperammonemia, which is generally defined as a concentration of ammonia ($NH_3 + NH_4^+$) in plasma being $\geq$50 μM in infants, children and adolescents (Raina et al., 2020) or >30 μM in adults

(Sakusic et al., 2018), frequently occur under various nutritional and pathological conditions. This condition causes multiple organ dysfunctions, vomiting, nausea, seizures, and death in humans, particularly preterm infants maintained on a total parenteral nutrition (TPN) solution. The underlying biochemical mechanisms include reductions in NO bioavailability, blood flow to the brain, and tissue oxygenation; depletion of α-KG in the Krebs cycle; and oxidative stress. Major factors responsible for hyperammonemia include arginine deficiency, liver failure, inherited metabolic disease (a defect in a urea-cycle enzyme or fatty acid oxidation), excessive AA load, AA imbalances, and AA antagonism (Wu, 2022).

The urea cycle in the liver and small intestine

In humans, the urea cycle is the major metabolic pathway for ammonia detoxification (Fig. 3), with the net reaction of $2NH_3 + HCO^- \text{———} > H_2N\text{-}CO\text{-}NH_2$ (urea) and the net consumption of 4 mol ATP/mol urea. This metabolic process occurs primarily in the liver (Meijer et al., 1990) and, to a much lesser extent, the small-intestinal mucosa (Wu, 2022). The urea cycle spans both the mitochondria and the cytoplasm in periportal hepatocytes and enterocytes. NH_3 and HCO_3^- are formed from the

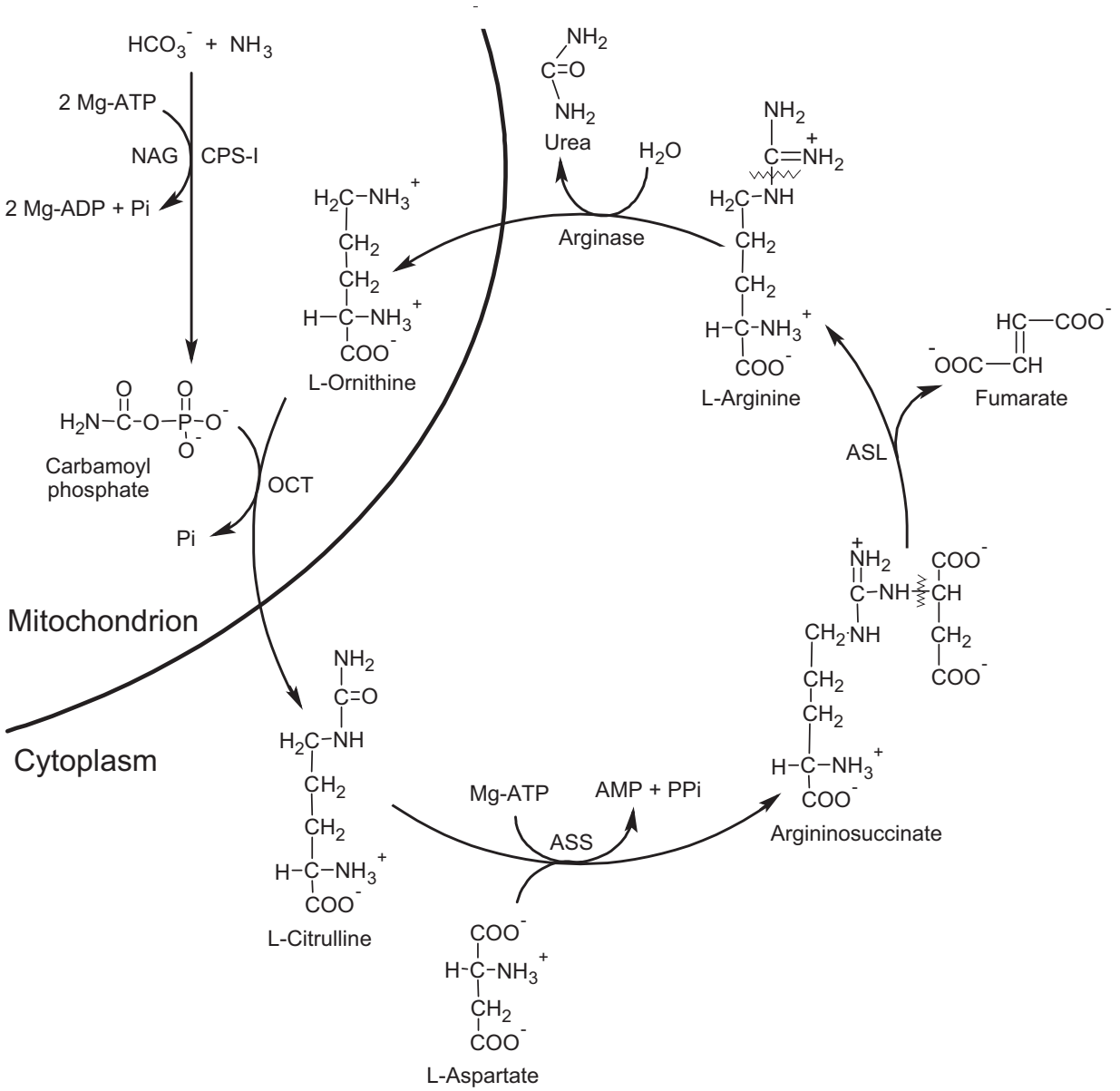


Fig. 3 The urea cycle in mammals. The synthesis of urea from ammonia and bicarbonate involves both the mitochondrion and the cytoplasm. Citrulline exits the mitochondrion into the cytoplasm where it is converted into arginine, which is rapidly hydrolyzed by arginase into urea plus ornithine. Ornithine is then re-used for another turnover of the cycle. ASL, argininosuccinate lyase; ASS, argininosuccinate synthase; CPS-I, carbamoylphosphate synthetase-I; NAG, N-acetylglutamate. Reproduced from Wu (2013) with permission.

catabolism of AAs (including glutamate, glutamine, and glycine) in these cell types, and can also be derived from the blood. Glutamate dehydrogenase and glutaminase are major enzymes for generating ammonia in the liver and intestine, respectively. After carbamoylphosphate synthetase I (CPS-I) converts NH_3 plus HCO_3^- into carbamoyl phosphate, the latter combines with ornithine (an AA from the diet, blood, or cytosolic arginase) to form citrulline by ornithine carbamoyltransferase (OCT). The cytosolic ornithine is transported by ORNT1 (mitochondrial ornithine transporter 1, an antiporter) to the mitochondrial matrix in exchange for the exit of mitochondrial citrulline to the cytosol, where citrulline is rapidly metabolized to arginine in the presence of aspartate and then to urea plus ornithine via a series of three enzymes: argininosuccinate synthase, argininosuccinate lyase, and arginase-I (Meijer et al., 1990). A major source of the cytosolic aspartate is the mitochondrial aspartate, which is exchanged with the cytosolic glutamate through the action of Citrin (a transporter in the mitochondrial membrane). Thus, the urea cycle is characterized by metabolic compartmentation and channeling to facilitate the immediate transfer of intermediates between enzymes and helps maintain a relatively high concentration of substrates in catalytic sites (Watford, 2020). This ensures the rapid and efficient formation of urea from ammonia. In adult humans, about 20%–25% of the urea newly synthesized in the liver is taken up by the small and large intestines, where urea is hydrolyzed into ammonia and CO_2. Some of the ammonia (e.g., 75% in the upper part of the gut) enters the portal vein for ureagenesis in the liver (Bergen and Wu, 2009).

Protein turnover in humans

AAs are constantly used for protein synthesis primarily in the endoplasmic reticulum and, to a much lesser extent, the mitochondria of cells, whereas intracellular proteins undergo continuous degradation to AAs by proteases and peptidases in lysosomes and nonlysosomal compartments of cells. These two processes are collectively referred to as intracellular protein turnover. The nonlysosomal pathway (including proteasomes) is responsible for the degradation of 70–80% of cellular proteins (Wu, 2022). In addition, extracellular proteins (e.g., proteins in the extracellular matrix and plasma) are degraded by proteases in the interstitial space of tissues and in the blood. In growing humans, the rate of protein synthesis is greater than the rate of protein degradation to result in protein accretion. The opposite is true in a catabolic state. In healthy adults gaining no protein, the rate of protein synthesis is equal to the rate of protein degradation, with the value being about 300 g per day in a 70 kg person (Waterlow, 1995). The incorporation of 1 mol AA into protein requires 5 mol ATP, and the hydrolysis of protein to release 1 mol AA consumes 2 mol ATP. In fed adults, 14.5% of dietary energy is utilized for the synthesis of 300 g protein/day, and protein turnover accounts for about 20% of whole-body energy expenditure (Wu, 2022). Over the past decades, there has been active debate regarding quantitative aspects of human requirements for dietary AAs.

Protein turnover is subjected to nutritional and hormonal regulation. Anabolic hormones (e.g., insulin, growth hormone, insulin-like growth factors, and testosterone), AAs (e.g., arginine, glutamine, glycine, and leucine), and muscle contraction stimulate protein synthesis in skeletal muscle. In adults with minimal physical activity, adequate intakes of high-quality dietary protein (e.g., 1.0–1.2 g/kg BW/day) such as an appropriate mixture of highly digestible animal and plant proteins are essential to mitigate muscle loss, and this beneficial effect may be attributed to a group of functional AAs but not all AAs. By contrast, catabolic hormones (e.g., cortisol), starvation, acidosis, heat stress, infections, and diseases reduce muscle protein synthesis. Muscle proteolysis is inhibited by insulin, insulin-like growth factors, testosterone, and AAs, but is promoted under catabolic conditions. The mechanistic target of rapamycin (MTOR) cell signaling pathway integrates anabolic factors to promote protein synthesis in cells (Paudel et al., 2021), whereas ATP and ubiquitin-dependent proteasomes as well as autophagy and lysosomal proteases are responsible for the bulk of intracellular proteolysis (Shen et al., 2021). Provision of dietary AAs in sufficient amounts and proper ratios is crucial for maximum protein synthesis in tissues (Wu, 2022). For example, either a deficiency or an excess of dietary threonine reduces protein synthesis in the jejunum and skeletal muscle of young mammals (Wang et al., 2007). Various metallo-proteases and peptidases catalyze the extracellular degradation of collagens and elastin in tissues to form free AAs (including 4-hydroxyproline), as well as small peptides containing high proportions of glycine, proline, and hydroxyproline residues (Phang et al., 2010). Many functional AAs (e.g., arginine, branched-chain AAs, glutamine, and glycine) inhibit proteolysis in skeletal muscle, small intestine, mammary glands, and many other tissues (Wu, 2022).

Metabolites of AAs in humans

Cell- and tissue-specific catabolism and conjugation of AAs generate hundreds of nitrogen-, sulfur-, and oxygen-containing metabolites in humans, including urea, ammonia, creatine, creatinine, catecholamines (e.g., dopamine, epinephrine, and norepinephrine), amino sugars, glucose, ketone bodies, fatty acids, NO, carbon monoxide (CO), hydrogen sulfide (H_2S), nitrite, nitrate, sulfate, polyamines (putrescine, spermidine, and spermine), homoarginine, serotonin, taurine, GABA, homocysteine, dimethylarginines, formate, and thyroid hormones (**Table 4**). Some of them are end products for excretion in the urine and feces (e.g., creatinine and urea), most of them (e.g., NO and creatine) at physiological concentrations are essential to life, and some of them (e.g., homocysteine and ammonia) are toxic at elevated concentrations (Bazer et al., 2021; Beaumont and Blachier, 2020; Stipanuk, 2004; Tsikas and Wu, 2015). In addition, the enzyme-catalyzed condensation of AAs form di-, tri-, and other oligo-peptides, such as carnosine, glutathione, and oxytocin. Furthermore, some AAs (e.g., taurine and glycine) can form conjugates with lipids (e.g., bile

Table 4 Products of amino acid metabolism in humans.

Amino acid	*Metabolites*
Amino acids	CO_2, water, ammonia, urea, uric acid, nitrite, nitrate, peptides, xenobiotic conjugates, glucose, fatty acids, and ketone bodies
Alanine	Pyruvate, aspartate, and glutamate
β-Alanine	Coenzyme A, pantothenic acid, carnosine (β-alanyl-L-histidine), and carcinine (β-alanyl-histamine)
Arginine	Nitric oxide, agmatine, and ornithine
Asparagine	Aspartate and *N*-acetylglucosamine-linked conjugate (N-linkage)
Aspartate	Purine, pyrimidine, asparagine, arginine, D-aspartate, and N-acetylaspartate
Citrulline	Arginine
Cysteine	Taurine, H_2S, sulfate, and homocysteine
Glutamate	Glutamine, citrulline, arginine, ornithine, proline, γ-aminobutyrate, and *N*-acetylglutamate
Glutamine	Purine, pyrimidine, ornithine, citrulline, arginine, proline, glutamate, aspartate, asparagine, and glucosamine-6-phosphate
Glycine	Bile salts, heme, bilirubin, purines, and serine
Histidine	Dipeptides, histamine, imidazoleacetate, and urocanate
Isoleucine	Glutamate, glutamine, alanine, and aspartate
Leucine	Glutamate, glutamine, alanine, aspartate, and β-hydroxy-β-methylbutyrate
Lysine	Trimethyllysine in calmodulin, desmosine and isodesmosine in collagen and elastin, and 5-hydroxylysine in collagen
Methionine	Homocysteine, betaine, cysteine, *S*-adenosylmethionine, taurine, and phospholipids
Phenylalanine	Tyrosine and phenylacetylglutamine
Proline	H_2O_2, pyrroline-5-carboxylate, ornithine, arginine, and 4-hydroxyproline
Serine	Cysteine, purine, pyrimidine, ceramide, phosphatidylserine, glucose, glycine, D-Serine, N^5-N^{10}-methylene-tetrahydrofolate, and *N*-acetylglucosamine-linked conjugate (O-linkage)
Threonine	Glycine and glucose
Tryptophan	Serotonin, N-acetylserotonin, melatonin, anthranilic acid, niacin, indoles, and kynurenate
Tyrosine	Dopamine, epinephrine, norepinephrine, melanin, triiodothyronine, and thyroxine
Valine	Glutamate, glutamine, alanine, and aspartate
Arg & Met	Polyamines (putrescine, spermidine, and spermine)
Arg, Met & Gly	Creatine
Cys, Glu & Gly	Glutathione
Gln, Asp & Gly	Nucleic acids and uric acid
Lys, Met & Ser	Carnitine
Ser & Met	Choline

Adapted from Wu (2022).

acids) via covalent bonds. Finally, intestinal bacteria produce many methylated, acetylated, hydroxylated, phosphorylated, and conjugated metabolites (Liu et al., 2019).

Summary and outlook

Dietary AAs are metabolized in humans via both synthetic and catabolic pathways in a cell- and tissue-specific manner, as well as intracellular protein turnover (protein synthesis and degradation) and extracellular proteolysis. Humans cannot synthesize *de novo* histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine. By contrast, humans can form alanine, arginine, aspartate, asparagine, glutamate, glutamine, glycine, proline, and serine *de novo*, while converting methionine into cysteine, phenylalanine into tyrosine, and serine plus selenium into selenocysteine. AA syntheses require energy, substrates, and also cofactors /substrates (e.g., NADH, NADPH, FAD, pyruvate, oxaloacetate, and α-KG). Growing evidence shows that endogenous AA syntheses are crucial for human growth, development, and health and that humans have dietary requirements for not only EAAs but also AASAs. As for other mammals, humans can degrade all proteinogenic AAs and physiological non-proteinogenic AAs at various rates via compartmentalized metabolic pathways and inter-organ cooperation, with products including CO_2, water, ammonia, pyruvate, acetyl-CoA, NO, SO_4^{2-}, glutathione, creatine, serotonin, and polyamines. Based on their metabolic fates, AAs can be classified as ketogenic, glucogenic, and ketogenic plus glucogenic. The complete oxidation of most AAs involves the formation of acetyl-CoA, which is oxidized to CO_2, NADH, and $FADH_2$ via the Krebs cycle. NAD(P)H and $FADH_2$ are oxidized into water via the mitochondrial electron transport system for the production of ATP from ADP plus Pi. Microbes in the small and large intestines degrade AAs via multiple pathways to form CO_2, ammonia, H_2S, and polyamines, as well as methylated, acetylated, hydroxylated, phosphorylated, and conjugated metabolites. Degradation of dietary AAs by the small intestine during their first-pass into the portal vein modifies the patterns of AAs in the plasma and reduces the efficiency of the utilization of dietary protein and free AAs for homeostasis and growth in the body. In healthy adults gaining no protein, active protein turnover occurs (i.e., 300 g protein/70 kg person per day) to account for about 20% of whole-body energy expenditure. Because elevated concentrations of ammonia in the plasma are highly

toxic to the brain, this metabolite must be removed as urea via the urea cycle primarily in the liver (periportal hepatocytes) and, to a much lesser extent, the enterocytes of the small intestine. The coordination of AA metabolism via multiple pathways is essential for the growth, development, and health of humans. Much needs to be determined about the optimum requirements of infants, children, and adults for dietary AAs (particularly AASAs). Finally, systematic research is warranted to identify important roles of functional AAs in stimulating muscle protein synthesis and alleviating muscle loss during aging and under catabolic conditions (e.g., heat stress, cancers, and viral infections) in humans.

Acknowledgments

Work is the author's laboratory is supported by Texas A&M AgriLife Research Hatch Project H-8200 and by Agriculture and Food Research Initiative Competitive Grants from the USDA National Institute of Food and Agriculture.

References

Armstrong, S.M., 1999. Melatonin as a chronobiotic for circadian insomnia. Clinical observations and animal models. Adv. Exp. Med. Biol. 460, 283–297.
Bazer, F.W., Seo, H., Johnson, G.A., Wu, G., 2021. One-carbon metabolism and development of the conceptus during pregnancy: lessons from studies with sheep and pigs. Adv. Exp. Med. Biol. 1285, 1–15.
Beaumont, M., Blachier, F., 2020. Amino acids in intestinal physiology and health. Adv. Exp. Med. Biol. 1265, 1–20.
Bergen, W.G., Wu, G., 2009. Intestinal nitrogen recycling and utilization in health and disease. J. Nutr. 139, 821–825.
Bertolo, R.F., Burrin, D.G., 2008. Comparative aspects of tissue glutamine and proline metabolism. J. Nutr. 138, 2032S–2039S.
Blachier, F., Blais, A., Elango, R., et al., 2021. Tolerable amounts of amino acids for human supplementation: summary and lessons from published peer-reviewed studies. Amino Acids 53, 1313–1328.
Bos, C., Metges, C.C., Gaudichon, C., et al., 2003. Postprandial kinetics of dietary amino acids are the main determinant of their metabolism after soy or milk protein ingestion in humans. J. Nutr. 133, 1308–1315.
Bröer, S., Bröer, A., 2017. Amino acid homeostasis and signaling in mammalian cells and organisms. Biochem. J. 474, 1935–1963.
Brosnan, J.T., Brosnan, M.E., 2006. Branched-chain amino acids: enzyme and substrate regulation. J. Nutr. 136, 207S–211S.
Brosnan, J.T., Brosnan, M.E., 2013. Glutamate: a truly functional amino acid. Amino Acids 45, 413–418.
Chapman, K.P., Elango, R., Ball, R.O., Pencharz, P.B., 2013. Splanchnic first pass disappearance of threonine and lysine do not differ in healthy men in the fed state. J. Nutr. 143, 290–294.
Chen, J.Q., Jin, Y.H., Yang, Y., Wu, Z.L., Wu, G., 2020. Epithelial dysfunction in lung diseases: effects of amino acids and potential mechanisms. Adv. Exp. Med. Biol. 1265, 57–70.
Closs, E.I., Scheld, J.S., Sharafi, M., Förstermann, U., 2000. Substrate supply for nitric-oxide synthase in macrophages and endothelial cells: role of cationic amino acid transporters. Mol. Pharmacol. 57, 68–74.
Corpeleijn, W.E., Riedijk, M.A., Zhou, Y., et al., 2010. Almost all enteral aspartate is taken up in first-pass metabolism in enterally fed preterm infants. Clin. Nutr. 29, 341–346.
Curthoys, N.P., Watford, M., 1995. Regulation of glutaminase expression and glutamine metabolism. Annu. Rev. Nutr. 15, 133–159.
Dai, Z.L., Wu, G., Zhu, W.Y., 2011. Amino acid metabolism in intestinal bacteria: links between gut ecology and host health. Front. Biosci. 16, 1768–1786.
Dai, Z.L., Wu, Z.L., Hang, S.Q., Zhu, W.Y., Wu, G., 2015. Amino acid metabolism in intestinal bacteria and its potential implications for mammalian reproduction. Mol. Hum. Reprod. 21, 389–409.
Davila, A.-M., Blachier, F., Gotteland, M., et al., 2013. Intestinal luminal nitrogen metabolism: role of the gut microbiota and consequences for the host. Pharmacol. Res. 68, 95–107.
Davis, T.A., Nguyen, H.V., Garcia-Bravo, R., et al., 1994. Amino acid composition of human milk is not unique. J. Nutr. 124, 1126–1132.
Haisch, M., Fukagawa, N.K., Matthews, D.E., 2000. Oxidation of glutamine by the splanchnic bed in humans. Am. J. Physiol. 278, E593–E602.
Hamanaka, R.B., Mutlu, G.M., 2021. The role of metabolic reprogramming and *de novo* amino acid synthesis in collagen protein production by myofibroblasts: implications for organ fibrosis and cancer. Amino Acids 53, 1851–1862.
Harper, A.E., Miller, R.H., Block, K.P., 1984. Branched-chain amino acid metabolism. Annu. Rev. Nutr. 4, 409–454.
He, W.L., Wu, G., 2020. Metabolism of amino acids in the brain and their roles in regulating food intake. Adv. Exp. Med. Biol. 1265, 167–185.
Hou, Y.Q., Wu, G., 2017. Nutritionally nonessential amino acids: a misnomer in nutritional sciences. Adv. Nutr. 8, 137–139.
Hou, Y.Q., He, W.L., Hu, S.D., Wu, G., 2019. Composition of polyamines and amino acids in plant-source foods for human consumption. Amino Acids 51, 1153–1165.
Hou, Y.Q., Hu, S.D., Li, X.Y., He, W.L., Wu, G., 2020. Amino acid metabolism in the liver: nutritional and physiological significance. Adv. Exp. Med. Biol. 1265, 21–37.
Hu, S.D., He, W.L., Wu, G., 2022. Hydroxyproline in animal metabolism, nutrition, and cell signaling. Amino Acids 54, 513–528.
Hutson, S.M., Sweatt, A.J., LaNoue, K.F., 2005. Branched-chain amino acid metabolism: implications for establishing safe intakes. J. Nutr. 135, 1557S–1564S.
Jungas, R.L., Halperin, M.L., Brosnan, J.T., 1992. Quantitative analysis of amino acid oxidation and related gluconeogenesis in humans. Physiol. Rev. 72, 419–448.
Li, P., Wu, G., 2020. Composition of amino acids and related nitrogenous nutrients in feedstuffs for animal diets. Amino Acids 52, 523–542.
Li, P., Yin, Y.L., Li, D.F., Kim, S.W., Wu, G., 2007. Amino acids and immune function. Br. J. Nutr. 98, 237–252.
Li, X.Y., Zheng, S.X., Wu, G., 2020. Amino acid metabolism in the kidneys: nutritional and physiological significance. Adv. Exp. Med. Biol. 1265, 71–95.
Liu, Y.Y., Tian, X.J., He, B.K., et al., 2019. *Lactobacillus reuteri* DSM 17938 feeding of healthy newborn mice regulates immune responses while modulating gut microbiota and boosting beneficial metabolites. Am. J. Physiol. 317, G824–G838.
Marliss, E.B., Aoki, T.T., Pozefsky, T., Most, A.S., Cahill, G.F., 1971. Muscle and splanchnic glutamine and glutamate metabolism in postabsorptive and starved man. J. Clin. Invest. 50, 814–817.
Matthews, D.E., Marano, M.A., Campbell, R.G., 1993. Splanchnic bed utilization of glutamine and glutamic acid in humans. Am. J. Physiol. 264, E848–E854.
McNeal, C.J., Meininger, C.J., Wilborn, C.D., Tekwe, C.D., Wu, G., 2018. Safety of dietary supplementation with arginine in adult humans. Amino Acids 50, 1215–1229.
Meijer, A.J., Lamers, W.H., Chamuleau, R.A., 1990. Nitrogen metabolism and ornithine cycle function. Physiol. Rev. 70, 701–748.
Meléndez-Hevia, E., De Paz-Lugo, P., Cornish-Bowden, A., Cárdenas, M.L., 2009. A weak link in metabolism: the metabolic capacity for glycine biosynthesis does not satisfy the need for collagen synthesis. J. Biosci. 34, 853–872.
Morris Jr., S.M., 2009. Recent advances in arginine metabolism: roles and regulation of the arginases. Br. J. Pharmacol. 157, 922–930.
Paudel, S., Wu, G., Wang, X.Q., 2021. Amino acids in cell signaling: regulation and function. Adv. Exp. Med. Biol. 1332, 17–33.
Phang, J.M., Liu, W., Zabirnyk, O., 2010. Proline metabolism and microenvironmental stress. Annu. Rev. Nutr. 30, 441–463.
Phang, J.M., 2019. Proline metabolism in cell regulation and cancer biology: recent advances and hypotheses. Antioxid. Redox Signal. 30, 635–649.

Raina, R., Bedoyan, J.K., Lichter-Konecki, U., et al., 2020. Consensus guidelines for management of hyperammonaemia in paediatric patients receiving continuous kidney replacement therapy. Nat. Rev. Nephrol. 16, 471–482.
Rhoads, J.M., Wu, G., 2009. Glutamine, arginine, and leucine signaling in the intestine. Amino Acids 37, 111–122.
Ryan, P., Riechman, S.E., Fluckey, J.D., Wu, G., 2021. Interorgan metabolism of amino acids in health and disease. Adv. Exp. Med. Biol. 1332, 129–149.
Sakusic, A., Sabov, M., McCambridge, A.J., et al., 2018. Features of adult hyperammonemia not due to liver failure in the intensive care unit. Crit. Care Med. 46, e897–903.
Sarkar, T.R., McNeal, C.J., Meininger, C.J., et al., 2021. Dietary intakes of amino acids and other nutrients by adult humans. Adv. Exp. Med. Biol. 1332, 211–227.
Shen, J.Z., Wu, G., Guo, S.D., 2021. Amino acids in autophagy: regulation and function. Adv. Exp. Med. Biol. 1332, 51–66.
Stipanuk, M.H., 2004. Sulfur amino acid metabolism: pathways for production and removal of homocysteine and cysteine. Annu. Rev. Nutr. 24, 539–577.
Tikanoja, T., Simell, O., Viikari, M., Järvenpää, A.L., 1982a. Plasma amino acids in term neonates after a feed of human milk or formula. II. characteristic, changes in individual amino acids. Acta Paediatr. Scand. 71, 391–397.
Tikanoja, T., Simell, O., Järvenpää, A.-L., Räihä, N.C.R., 1982b. Plasma amino acids in preterm infants after a feed of human milk or formula. J. Pediatr. 101, 248–252.
Tikanoja, T., 1982. Plasma amino acids in term neonates after a feed of human milk or formula. Acta Paediatr. Scand. 71, 385–389.
Tsikas, D., Wu, G., 2015. Homoarginine, arginine, and relatives: analysis, metabolism, transport, physiology, and pathology. Amino Acids 47, 1697–1702.
Tsikas, D., 2021. Post-translational modifications (PTM): analytical approaches, signaling, physiology and pathophysiology—part I. Amino Acids 53, 485–487.
van der Schoor, S., Schierbeek, H., Bet, P., et al., 2010. Majority of dietary glutamine is utilized in first pass in preterm infants. Pediatr. Res. 67, 194–199.
Wang, X., Qiao, S.Y., Yin, Y.L., et al., 2007. A deficiency or excess of dietary threonine reduces protein synthesis in jejunum and skeletal muscle of young pigs. J. Nutr. 137, 1442–1446.
Waterlow, J.C., 1995. Whole-body protein turnover in humans—past, present, and future. Annu. Rev. Nutr. 15, 57–92.
Watford, M., 2020. Ornithine cycle. In: The Encyclopedia of Biochemistry, third ed. Elsevier, New York. https://doi.org/10.1016/B978-0-12-819460-7.00062-1.
Wu, G., 2022. Amino Acids: Biochemistry and Nutrition, second ed. CRC Press, Boca Raton.
Wu, G., Morris Jr., S.M., 1998. Arginine metabolism: nitric oxide and beyond. Biochem. J. 336, 1–17.
Wu, G., Thompson, J.R., Baracos, V.E., 1991. Glutamine metabolism in skeletal muscle from the broiler chick (*Gallus domesticus*) and the laboratory rat (*Rattus norvegicus*). Biochem. J. 274, 769–774.
Wu, G., 1998. Intestinal mucosal amino acid catabolism. J. Nutr. 128, 1249–1252.
Wu, G., 2013. Amino Acids: Biochemistry and Nutrition, 1st ed. CRC Press, Boca Raton.
Wu, G., 2016. Dietary protein intake and human health. Food Funct. 7, 1251–1265.
Wu, G., 2020a. Important roles of dietary taurine, creatine, carnosine, anserine and hydroxyproline in human nutrition and health. Amino Acids 52, 329–360.
Wu, G., 2020b. Metabolism and functions of amino acids in sense organs. Adv. Exp. Med. Biol. 1265, 201–217.

Amino acids: Specific functions

Guoyao Wu, Department of Animal Science and Faculty of Nutrition, Texas A&M University, College Station, TX, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of M.C.G. van de Poll, Y.C. Luiking, C.H.C. Dejong, P.B. Soeters, Amino Acids: Specific Functions, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 79–87, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00011-8.

Introduction 37
Physiological functions of AAs 37
Physiological functions of AA metabolites 37
Functions of oligopeptides 37
Functions of NO, CO, and H_2S 40
Functions of polyamines 40
Functions of creatine 41
Functions of glutathione 41
Functions of purine and pyrimidine nucleotides 42
Functions of taurine 42
Functions of histamine 42
Functions of melanin 42
Functions of melatonin 42
Functions of carnosine 43
Functions of 4-hydroxyproline 43
Functions of glucosamine 43
Functions of dopamine 43
Functions of serotonin 43
Functions of quinolinic acid and kynurenic acid 44
Functions of AAs and metabolites as activators of the aryl hydrocarbon receptor 44
Functions of heme 44
Functions of AAs in sensing 44
Functions of D-AAs 44
Summary and outlook 44
Acknowledgments 45
References 45

Key points

- Amino acids (AAs) fulfill structural, metabolic, physiological, and immunological needs in humans.
- Some AAs (e.g., aspartate, glutamate, and glycine) and metabolites (e.g., γ-aminobutyrate, serotonin, nitric oxide, carbon monoxide, and hydrogen sulfide) are neurotransmitters.
- Some AAs (e.g., arginine, glutamine, glycine, leucine, tryptophan, and valine) activate the mechanistic target of rapamycin cell signaling pathway to stimulate protein synthesis and inhibit autophagy-mediated proteolysis.
- Some AAs (e.g., alanine, glutamine, and leucine) inhibit autophagy and lysosomal proteolysis.
- Most physiological metabolites (e.g., creatine, heme, formate, purines, pyrimidines, and polyamines) are essential for biochemical processes (e.g., ATP production, energy storage, oxygen delivery, and cell growth).
- AAs regulate gene expression, cell signaling pathways, food intake, chemical sensing, the digestion and absorption of dietary nutrients, neurological development and behavior, and the metabolism of energy and nutrients.
- AAs are also crucial for immunity, reproduction, hormone secretion, anti-oxidative and anti-inflammatory responses, the detoxification of endogenous and exogenous substances, acid-base and mineral balances, and osmolarity.
- Dietary supplementation with functional AAs (e.g., L-arginine, L-glutamate, and glycine) can enhance immunity, anti-oxidative responses, fertility, wound healing, ammonia detoxification, and lean tissue mass; ameliorate metabolic syndromes (including dyslipidemia, obesity, diabetes, and hypertension); and treat individuals with erectile dysfunction, sickle cell disease, muscular dystrophy, and pre-eclampsia in humans.

Introduction

Amino acids (AAs) are the building blocks of proteins, which fulfill structural, metabolic, and regulatory functions in humans. A healthy adult with a normal body mass index contains 15.1% protein, i.e., 10.6 kg protein/70 kg person (Wang et al., 2003), including 3.72 kg collagen (Meléndez-Hevia et al., 2009). Collagen (the most abundant protein humans) consists of approximately 1/3 glycine and 1/3 proline plus 4-hydroxyproline on the molar basis, and comprises about 30% and 35% of body proteins in newborns and adults, respectively (Wu, 2022). Thus, large amounts of AAs, particularly glycine and proline, are required for protein synthesis and, therefore, the survival, health, growth, and development of humans. In addition, AAs serve as essential substrates for the syntheses of small peptides, low-molecular-weight substances [e.g., nitric oxide (NO), carbon monoxide (CO), hydrogen sulfide (H_2S), glutathione, creatine, polyamines, carnitine, choline, and homoarginine], and non-protein large molecules [e.g., bradykinin (MW 1060.2 Da) and oxytocin (MW 1007.2 Da)] (Bazer et al., 2021; Beaumont and Blachier, 2020; Calder et al., 2002; Closs et al., 2000; Dai et al., 2013; Li et al., 2009; Suryawan and Davis, 2011; Wu, 2022). Furthermore, AAs are signaling molecules to regulate food intake, gene expression, post-translational protein modifications, metabolism, immune and anti-oxidative responses, cell-to-cell communication, and whole-body homeostasis (Brosnan and Brosnan, 2013; Li et al., 2022; Moffett, 2003; Phang, 2019; Rhoads and Wu, 2009; Tsikas, 2021; Wetzel et al., 2019; Wu, 2009). These specific functions of AAs beyond protein synthesis are highlighted in this article.

Physiological functions of AAs

In humans, different AAs have different physiological functions beyond protein synthesis (**Table 1**). The overall functions of AAs include their roles in the regulation of gene expression (including epigenetics), cell signaling pathways, food intake, chemical sensing, digestion and absorption of dietary nutrients, neurological development and behavior, energy and nutrient metabolism, immune responses, female and male reproduction, endocrine status, anti-oxidative and anti-inflammatory responses, the detoxification of endogenous and exogenous substances, acid-base and mineral balances, osmolarity, whole-body homeostasis, and health and well-being (Biefer et al., 2017; Brasse-Lagnel et al., 2010; Calder et al., 2002; Chen et al., 2020; Curthoys and Watford, 1995; Fernstrom, 2013; Field et al., 2002; Gietzen and Aja, 2012; He and Wu, 2020; Li et al., 2022; Paudel et al., 2021; Wu, 2009). For example, some AAs (e.g., aspartate, glutamate and glycine) are neurotransmitters in the brain and spinal cord (**Fig. 1**), whereas some AAs (e.g., BCAAs, glutamine and alanine) participate in the inter-organ transport and metabolism of nitrogen and carbon (Hou et al., 2020; Li et al., 2020; Watford, 2008). In addition, arginine, glutamine, glycine, leucine, tryptophan, and valine activate the mechanistic target of rapamycin (MTOR) cell signaling pathway to stimulate protein synthesis and inhibit autophagy-mediated proteolysis (Bröer and Bröer, 2017; Wu, 2022). Many AAs (e.g., alanine, glutamine, leucine, and tyrosine) directly inhibit autophagy in cells such as hepatocytes (Mortimore and Khurana, 1990; Shen et al., 2021). Thus, dietary supplementation with AAs (e.g., L-arginine, glycine, and branched-chain AAs) can enhance immunity, anti-oxidative responses, fertility, wound healing, ammonia detoxification, lean tissue mass, and lactation; ameliorate metabolic syndromes (including dyslipidemia, obesity, diabetes, and hypertension); and treat individuals with erectile dysfunction, sickle cell disease, muscular dystrophy, and pre-eclampsia in humans (Cynober, 2004; Durante, 2020; Holeček, 2018; Jiang et al., 1993; Le Floc'h et al., 2011; Lei et al., 2012; Li et al., 2007; McNeal et al., 2018; Newsholme et al., 1999; Popovic et al., 2007; Posey et al., 2021; Ryan et al., 2021; Tomé, 2018; Wu et al., 2021; Ziegler et al., 1998). Furthermore, the catabolism, racemization, and conjugation of AAs results in the formation of non-gaseous products (e.g., heme, D-serine, D-aspartate, glutathione, creatine, bile salts, and homoarginine) with enormous versatility and physiological importance (see below). In addition, the human host, dietary AAs may also alter the population, metabolism, and function of its gastrointestinal microbes (Dai et al., 2011, 2015; Davila et al., 2013). Adequate consumption of balanced AAs in the form of protein and supplements is essential for the optimum human nutrition (Wu, 2016). Note that animal-sourced foods generally contain much more proteins than plant-sourced foods (Davis et al., 1994; Hou et al., 2019; Li and Wu, 2020; Li et al., 2021).

Based on recent advances in AA biochemistry and physiology, Wu (2010) proposed a new nutritional concept of functional AAs, which are defined as AAs that regulate key metabolic pathways to improve the growth, development, reproduction, lactation, physical performance, immunity, and health of humans. These AAs include arginine, cysteine, glutamate, glutamine, glycine, leucine, methionine, proline, and tryptophan. The concept of functional AAs is now transforming human protein nutrition research and feeding practice (Hou and Wu, 2017; Posey et al., 2021; Ryan et al., 2021). Humans have dietary requirements for all proteinogenic AAs plus some nonproteinogenic AAs (e.g., taurine and β-alanine) for optimum growth, development, physical strength, health, and well-being (Hou et al., 2015; Wu et al., 2013). Thus, it is counter-intuitive to classify biosynthesizable AAs as nutritionally nonessential (dispensable) in human nutrition.

Physiological functions of AA metabolites

Functions of oligopeptides

AAs are used to synthesize oligopeptides, including antibiotics produced by bacteria and the intestinal mucosa, as well as many other oligopeptides consisting of 9 or 10 AA residues (Wu, 2022). Examples of the physiologically important oligopeptides include:

Table 1 Major physiological functions of amino acids and their metabolites in humans.[a]

Amino acid	*Metabolites or direct action*	*Major functions*
Amino acids	Directly	Osmolarity, inhibition of autophagy, cell signaling, and protein function
	Proteins	Structural components of the body; cell growth, development, and function;
	Peptides	Hormones, antibiotics, metabolic regulation, and antioxidants
	Ammonia	Syntheses of glutamate, glutamine, carbamoyl phosphate, and urea
Alanine	Directly	Inhibition of pyruvate kinase and hepatic autophagy, gluconeogenesis, the glucose-alanine cycle, and therapy of muscular glycogen depletion
	Asp and glu	Major metabolic fuels for the small intestine, the malate shuttle for the transfer of NADH from the cytosol to mitochondria, and the interorgan metabolism and transport of both carbon and nitrogen
β-Alanine	Directly	A component of coenzyme A and pantothenic acid
	Dipeptides	Carnosine (β-alanyl-L-histidine), carcinine (β-alanyl-histamine), (β-alanyl-1-methyl-L-histidine), and balenine (β-alanyl-3-methyl-histidine)
Arginine	Directly	Activation of MTOR and AMPK signaling pathways
	Nitric oxide	Vasodilator, signaling molecule, immune response, regulator of metabolism
	Agmatine	Ligand for α_2-adrenergic and imidazoline receptors, and NOS inhibitor
	Ornithine	Ammonia detoxification; syntheses of proline, glutamate and polyamines
	Proline	Collagen synthesis, wound healing, and many other functions (see below)
	Methylarginines	Competitive inhibition of NOS
Asparagine	Directly	Regulation of gene expression, immunity, immune function, and the survival of leukemic cells
Aspartate	Directly	Syntheses of purine, pyrimidine, and arginine; activation of NMDA receptors
	D-Aspartate	Activation of NMDA receptors in the brain
Citrulline	Directly	Antioxidant; arginine synthesis, osmoregulation, and ammonia detoxification
Cysteine	Directly	Disulfide linkage in protein, and the interorgan transport of sulfur
	Taurine	Antioxidant, and the regulation of cellular redox state
	H_2S	Neurotransmission, vasodilation, killing of pathogens, and cell metabolism
Glutamate	Directly	Syntheses of glutamine and arginine, metabolism via multiple pathways, flavor enhancer for increasing food intake, the activation of NMDA receptors, and a precursor of *N*-acetylglutamate in hepatocytes and enterocytes
	γ-Aminobutyrate	Inhibitory or excitatory neurotransmitter, depending on age, type of receptor, and the region of brain; regulation of neuronal excitability
Glutamine	Directly	Regulation of protein turnover, cell volume, gene expression, and immune responses; syntheses of purine, pyrimidine, citrulline, arginine, and NAD(P)
	Glu and asp	Excitatory neurotransmitters, components of the malate shuttle, cell Metabolism, ammonia detoxification, and major fuels for enterocytes
	Alanine	Interorgan metabolism of nitrogen and carbon; also see above.
	Glucosamine-6-P	Synthesis of aminosugars, anti-inflammation, angiogenesis, and cell growth
	Ammonia	Renal regulation of acid-base balance, glutamate and carbamoyl-P syntheses
Glycine	Directly	Activation of a glycine-gated channel; syntheses of purine, serine, porphyrins, glutathione, and heme; an inhibitory neurotransmitter in the brain; antioxidant; one-carbon metabolism; conjugation with bile acids
	Heme	A component of hemoproteins, carbon monoxide production, iron storage
	Bilirubin	Natural ligand of aryl hydrocarbon receptor in the cytoplasm
Histidine	Directly	Hemoglobin structure, antioxidative dipeptides, and one-carbon metabolism
	Histamine	Allergic reaction, vasodilator, a neurotransmitter, the stimulation of gastrointestinal secretions, and the regulation of immunocyte metabolism
	Imidazoleacetate	Analgesic and narcotic actions
	Urocanate	Immune responses in the skin, and protecting the skin against UV radiation
Isoleucine	Directly	Synthesis of glutamine and alanine
Leucine	Directly	Regulation of protein turnover, activator of glutamate dehydrogenase
	Gln and Ala	Many metabolic functions (see above)
	HMB	Regulation of immune response and skeletal-muscle protein synthesis
Lysine	Directly	Regulation of nitric oxide synthesis; and O-linked glycosylation
	5-Hydroxylysine	Structure and function of collagen
	Desmosine	Structure and function of collagen and elastin
Methionine	Homocysteine	Oxidant; independent risk factor for cardiovascular disease
	Betaine	Methylation of homocysteine to methionine, and one-carbon metabolism
	Cysteine	Cellular metabolism and nutrition
	SAM	Methylation of proteins and DNA; methyl group donor in biosynthesis
	Phospholipids	Synthesis of lecithin and phosphatidylcholine; and cell signaling

Table 1 Major physiological functions of amino acids and their metabolites in humans.[a]—cont'd

Amino acid	*Metabolites or direct action*	*Major functions*
Phenylalanine	Directly	Activation of tetrahydrobiopterin (a cofactor for NOS) synthesis, syntheses of tyrosine and phenylacetylglutamine, and neurological function
Proline	Directly	Collagen structure and function, neurological function, and an antioxidant
	H_2O_2	Killing pathogens, intestinal integrity, a signaling molecule, and immunity
	P5C	Cellular redox state, arginine synthesis, gene expression, and stress response
	4-hydroxyproline	Structure and function of collagen, glycine synthesis, and cell signaling
Serine	Directly	Methionine catabolism; one-carbon metabolism; syntheses of cysteine, purine, pyrimidine, ceramide and phosphatidylserine
	Glycine	Many metabolic and regulatory functions (see above)
	D-serine[b]	Activation of NMDA receptors in brain
Threonine	Directly	Syntheses of mucins and glycine, immunity, O-linked glycosylation, and the therapy of spastic syndromes
Tryptophan	Serotonin	Neurotransmitter; regulation of food intake, mood, sleep, and behavior; stimulation of gut motility
	N-acetylserotonin	Antioxidant, and anti-inflammatory responses
	Melatonin	Antioxidant, anti-inflammatory responses, and circadian rhythms
	Anthranilic acid	Inhibiting production of proinflammatory T-helper-1 cytokines
	Niacin	A component of NAD and NADP, and modifications of proteins
	Indoles[c]	Natural ligands of aryl hydrocarbon receptor in the cytoplasm, and immunity
	Quinolinic acid	An agonist of the NMDA receptor
	Kynurenic acid	A nonselective NMDA receptor antagonist and therefore a blocker of AA- modulated excitation of the brain, neuromodulatory activity, as well as anti-excitotoxic and anti-convulsant effects
Tyrosine	Directly	Protein phosphorylation, nitrosation, and sulfation
	Dopamine	Neurotransmitter (neurological development, brain reward, pleasure, motivation, motor control, emotion, and executive function); and enhancement of immunity
	EPN & NEPN	Neurotransmitters in the brain; the most important neurotransmitters in the sympathetic nervous system; the regulation of cell metabolism; increasing blood pressure and heart rate, and alertness; and decreasing intestinal motility
	Melanin	Antioxidant, sexual activity, and pigmentation of the skin and hair
	T3 and T4	Regulation of energy and protein metabolism, growth, and development
Valine	Directly	Synthesis of glutamine and alanine
A group of AAs[d]	Directly	Activation of the MTOR cell signaling
Arg & Met	Polyamines	DNA and protein synthesis, ion channel function, and MTOR activation
Arg, Met & Gly	Creatine	Antioxidant; energy metabolism in the heart, skeletal muscle and brain
Cys, Glu & Gly	Glutathione	Antioxidant, cell metabolism, cellular redox, immunity, and gene expression
Gln, Asp & Gly	Nucleic acids	Gene expression; cell growth, and syntheses of protein and uric acid
	Uric acid	Antioxidant, and the major metabolite of purines (adenosine and guanosine)
Lys, Met & Ser	Carnitine	Transport of long-chain fatty acids into mitochondria for oxidation
Ser & Met	Choline	A component of acetylcholine (a neurotransmitter) and phosphatidylcholine (a structural lipid); syntheses of betaine, sarcosine and glycine

Note: BCAAs, branched-chain amino acids; SAM, *S*-adenosylmethionine; EPN, epinephrine; HMB, β-hydroxy-β-methylbutyrate; MTOR, mechanistic target of rapamycin; NEPN, norepinephrine; NOS, nitric oxide synthase; T_3, triiodothyronine; T_4, thyroxine.

[a]Adapted from Wu, (2022). Unless indicated, amino acids (except for glycine, taurine, β-alanine, and γ-aminobutyrate) are L-isomers.

[b]Synthesized from L-serine by serine racemase.

[c]Including indole acetic acid, kynurenine, and tryptamine.

[d]Arginine, glutamine, glycine, leucine, tryptophan, and valine.

angiotensin II [10 AAs; Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-His-Leu (NH_2)], bradykinin [(9 AAs); Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg (NH_2)], oxytocin [(9 AAs); Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly (NH_2)], and arginine vasopressin [9 AAs; Cys-Tyr-Phe-Gln-Asn-Cys-Pro-Arg-Gly (NH_2)] (Forbes and Krishnamurthy, 2020). Through actions including vasoconstriction and the stimulation of the sympathetic nervous system, angiotensin II can raise blood pressure and, therefore, is used to treat hypotension resulting from septic shock. Bradykinin causes arterioles to dilate and augment vascular permeability by increasing the release of prostacyclin, NO, and endothelium-derived hyperpolarizing factor from blood vessels. Oxytocin (a peptide hormone and neuropeptide) strengthens the contractions of the uterus during labor and induces the let-down of milk from the lactating mammary glands. Arginine vasopressins (also known as antidiuretic hormone) increases blood pressure in humans by stimulating the renal reabsorption of water into the blood circulation and constricts arterioles and, therefore, is used to treat individuals with vasodilatory shock.

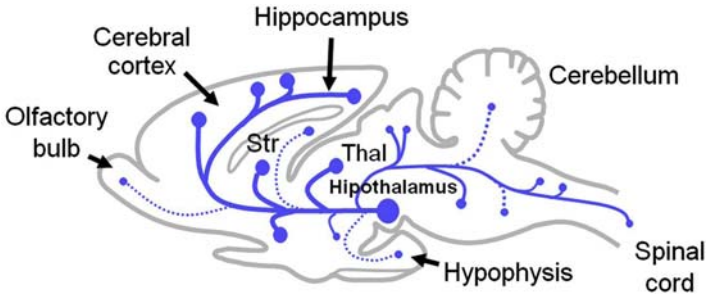


Fig. 1 Structure of the various parts of the brain and the spinal cord for the production and actions of neurotransmitters. GABAergic neurons are located in the hippocampus, thalamus, basal ganglia, hypothalamus, and brainstem; γ-Aminobutyrate (GABA) is a major inhibitory transmitter in the mature brain but is primarily excitatory in the developing brain. Heme oxygenase is localized in many regions of the brain, including the hypothalamus and olfactory bulb. Cholinergic neurons is found in the striatum. Dopaminergic neurons extend from the midbrain (e.g., the substantia nigra pars compacta, the ventral tegmental area, the arcuate nucleus, and the retrorubral area) through the forebrain (e.g., the olfactory bulb). Epinephrine is produced in the medulla oblongata of the brainstem. H_2S is produced by cystathionine β-synthase (CBS), cystathionine γ-lyase (CSE), and 3-mercaptopyruvate sulfurtransferase (3MST); CBS is highly expressed in the astrocytes and microglial cells of the hippocampus and cerebellum; CSE is mainly expressed in the cardiovascular system but is also present in microglial cells, spinal cord, and cerebellar granule neurons; and 3MST occurs in the neurons and astrocytes of the brain. Histaminergic neurons are confined to the tuberomamillary nucleus of the posterior hypothalamus. Neuronal nitric oxide synthase (nNOS) is expressed in neurons and pineal gland of the brain and in the spinal cord; in the juvenile and mature hippocampus and neocortex, nNOS is primarily expressed in subpopulations of GABAergic interneurons. Norepinephrine is produced primarily by neurons of the locus coeruleus and to a lesser extent in the lateral tegmental field of the brain. Serotonergic neurons are predominantly located in the brainstem raphe complex, a midline complex closely associated with the reticular formation within the midbrain, pons, and medulla. Thal, thalamus; Str, striatum. Reproduced from a freely available article by Nieto-Alamilla et al. (2016) at https://molpharm.aspetjournals.org/content/molpharm/90/5/649.full.pdf.

Functions of NO, CO, and H_2S

NO, CO, and H_2S are produced from L-arginine, heme (formed from glycine), and L-cysteine, respectively. NO synthase and heme oxygenase are wide-spread in human tissues. By contrast, H_2S is synthesized by cystathionase in the liver, kidneys, enterocytes, and vascular smooth muscle cells, by cystathionine-β-synthase in the brain, and by 3-mercaptopyruvate sulfurtransferase in cardiac tissues (Kamat et al., 2015). These gases readily penetrate cell membranes to exert their physiological effects (**Fig. 2**). At the molecular level, NO, CO and H_2S activate guanylyl cyclase to convert GTP into cGMP, whereas H_2S stimulates adenylyl cyclase to convert ATP into cAMP and also increases cGMO availability in cells (Beltowski, 2015; Li et al., 2009; Nakahira et al., 2006; Otterbein et al., 2000). As oxidants, NO and H_2S can kill pathogenic microbes. In this regard, it is noteworthy that NO can kill coronaviruses, including severe acute respiratory syndrome coronavirus (SARS-CoV; Åkerström et al., 2005) and SARS-CoV-2 [the virus causing the coronavirus disease-2019 (Covid-19); Akaberi et al., 2020]. Physiological concentrations of NO are essential for the endothelium-dependent relaxation of blood vessels, blood flow, angiogenesis, embryogenesis, mitochondrial biogenesis, neurotransmission, immunity, ovulation in females, spermatogenesis in males, and wound healing (Wu et al., 2021). Physiological concentrations of CO relax blood vessels, act as a retrograde messenger at synapses, stimulate long-term potentiation in the hippocampus, have potent cytoprotective effects in cells (particularly neurons), modulate immune responses, and stimulate the oxidation of glucose and fatty acids in tissues (e.g., white adipose tissue and skeletal muscle; Li et al., 2009). Physiological concentrations of H_2S play an important role in the body through modulating the endothelial arginine-NO pathway, stimulating ATP-sensitive K^+ channels and membrane hyperpolarization, protecting cells from oxidant (e.g., homocysteine)-induced injury, facilitating the induction of hippocampal long-term potentiation, and reducing oxidative metabolism (Li et al., 2009).

Functions of polyamines

Polyamines (putrescine, spermidine, and spermine) are synthesized from arginine- and proline-derived ornithine via ornithine decarboxylase, spermidine synthase, and spermine synthase in humans (Wu, 2022). Methionine is required for the production of spermidine and spermine via the formation of S-adenosylmethionine as the methyl group donor. Polyamines are essential for DNA and protein syntheses, angiogenesis, spermatogenesis, and wound healing (Agostinelli, 2020; Pegg, 2016). Polyamines also participate in the posttranslational modifications of some proteins, such as eIF5A by deoxyhypusine synthase to form the hypusine residue that is critical for eukaryotic translation (Park and Wolff, 2018). Furthermore, polyamines are substrates for some transglutaminases that catalyze the formation of an isopeptide bond (an amide bond) between a primary amine or the γ-NH_2 group of a glutamine residue in a polypeptide chain and the ε-NH_2 of a lysine residue of another polypeptide chain. In humans, mutations in the spermine synthase gene decrease spermine concentrations, resulting in Snyder-Robinson syndrome, an X-linked recessive condition characterized by mental retardation, skeletal defects, hypotonia, and movement disorders (Becerra-Solano et al., 2009).

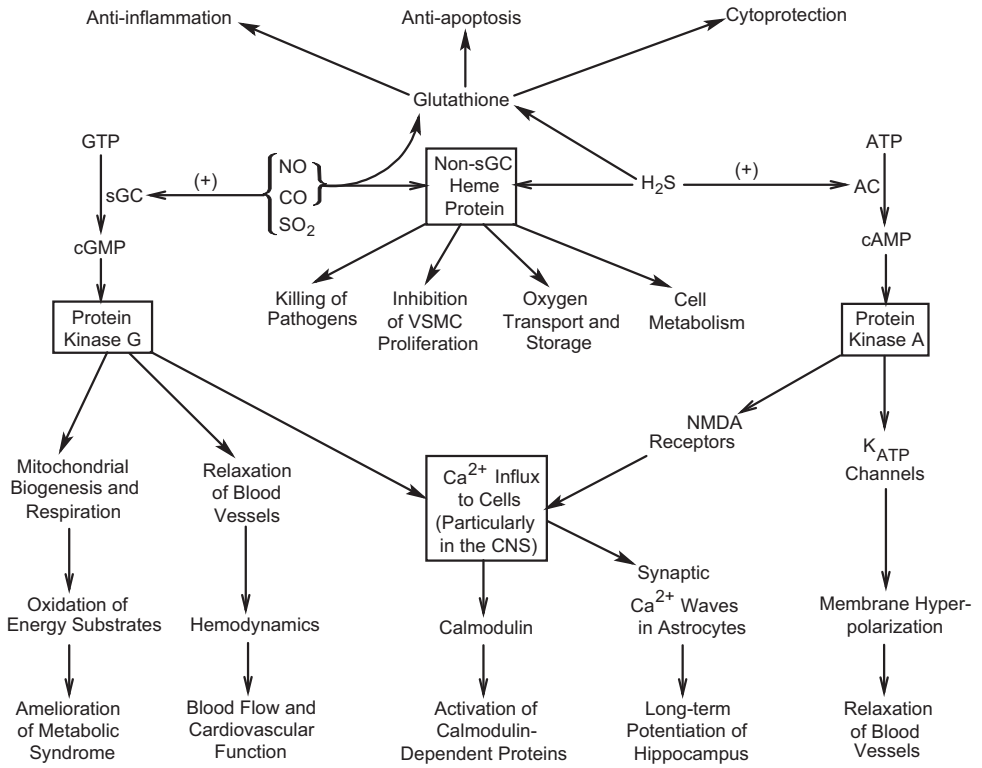


Fig. 2 Gaseous signaling in cells via cGMP and cAMP-dependent and independent pathways in humans. Physiological levels of both NO and CO activate guanylate (guanylyl) cyclase to generate cGMP, which stimulates cGMP-dependent protein kinase. H_2S activates adenylate cyclase activity to yield cAMP, which stimulates cAMP-dependent protein kinase. NO, CO, and H_2S enhance glutathione synthesis to protect cells from oxidative stress. These gases also play important roles in innate and cell-mediated immunities, including cytoprotection for immunocytes, the modulation of inflammatory responses, reductions in the production of proinflammatory cytokines, and increases in the production of anti-inflammatory cytokines. AC, adenylate cyclase; CNS, central nervous system; and sGC, soluble guanylate cyclase. Reproduced from Li et al. (2009) with permission.

Functions of creatine

Creatine is synthesized from arginine, glycine and methionine via inter-organ cooperation (Brosnan and Brosnan, 2007). Creatine kinase converts creatine and ATP into phosphocreatine and ADP as a mechanism for energy storage. The phosphocreatine/creatine kinase system is characterized by cell- and tissue-specific isoforms of creatine kinase, which are differentially localized in the cytoplasm and mitochondria to fulfill metabolic needs (Wallimann et al., 2011). This enzyme is present in many tissues (including skeletal muscle, brain, heart, kidneys, testes, and ovaries), but is absent from the liver (Wyss and Kaddurah-Daou, 2000). Creatine plays an important role in cellular energy metabolism and anti-oxidative reactions, particularly in the nervous, muscular, and reproductive systems and under hypoxic conditions. Therefore, oral administration of creatine can improve the health and exercise performance of humans (Posey et al., 2021). Conversely, defects in creatine synthesis result in neurological and muscular dysfunction and possibly reproductive failure (Philip et al., 2020).

Functions of glutathione

Glutathione (GSH) is synthesized from glycine, cysteine, and glutamate. GSH directly and effectively scavenges free radicals and other reactive oxygen species, while reacting with various electrophiles, physiological metabolites (e.g., estrogen, melanins, prostaglandins and leukotrienes), and xenobiotics (e.g., bromobenzene and acetaminophen) to form mercapturates for excretion (Meister and Anderson, 1983; Sies, 1999). In addition, GSH serves as a cofactor for the conversion of prostaglandin H_2 into prostaglandins D_2 and E_2 by endoperoxide isomerase. Furthermore, GSH contributes to the glutathionylation of proteins (e.g., thioredoxin, ubiquitin-conjugating enzyme, and cytochrome *c* oxidase). Finally, GSH maintains cellular redox state, confers anti-infectious effects (Diotallevi et al., 2017), and reduces the susceptibility of humans to SARS (Wu et al., 2004) and COVID-19 (Polonikov, 2020; Silvagno et al., 2020).

Functions of purine and pyrimidine nucleotides

Purines are synthesized from aspartate, glutamine, glycine, bicarbonate, and formate, whereas and pyrimidines from aspartate, glutamine, and bicarbonate in cells (Wu, 2022). Purine and pyrimidine nucleotides form DNA and RNA. DNA stores genetic information and directs the syntheses of RNA (including mRNA, tRNA, rRNA, sRNA, miRNA, and siRNA) and protein. In addition, ATP and other nucleoside triphosphates (i.e., GTP and CTP) store biological energy and drive endergonic reactions in nutrient transport and metabolic pathways. The nucleoside triphosphates also participate in reactions that generate the phosphorylated metabolites (e.g., glucose-6-phosphate and fructose-6-phosphate) and phosphorylated proteins. Furthermore, cGMP and cAMP are second messengers in cell signaling. Finally, nucleotides are components of FAD, FMN, NAD, NADP, CoA, and SAM, which participate in many reactions, including oxidation, dehydrogenation, and methylation (Wu, 2018).

Functions of taurine

Taurine is synthesized from cysteine in the liver and is transported into extrahepatic cells. A 70 kg adult person has ~70 g taurine (Huxtable, 1992). This AA is a highly abundant in the liver, intestine, eyes, skeletal muscle, heart, brain, and many of other tissues, as well as bile and milk. In humans, taurine, along with glycine, conjugates with bile acids to form bile salts, which facilitate the intestinal digestion and absorption of dietary lipids (including lipid-soluble vitamins) and eliminate cholesterol (Wu, 2020a). In addition, taurine is a major antioxidant, anti-inflammatory, and anti-apoptotic factor; a regulator of intracellular osmoregulation and retinal photoreceptor activity; and a key component of nerve and muscle conduction networks (Wright et al., 1986). Furthermore, taurine stimulates neurological development, and serves as an inhibitory neurotransmitter in the brain. Finally, taurine reacts with HOCl and HOBr in activated granulocytes and neutrophils to produce taurine chloramine (*N*-chlorotaurine) and taurine bromamine (*N*-bromotaurine), which can kill bacteria, fungi, parasites, and viruses (Marcinkiewicz and Kontny, 2014).

$$\text{Taurine} + \text{HOCl} \rightarrow N\text{-Chlorotaurine} + \text{H}_2\text{O}\ [\text{halide-dependent myeloperoxidase system}]$$

$$\text{Taurine} + \text{HOBr} \rightarrow N\text{-Bromotaurine} + \text{H}_2\text{O}\ [\text{halide-dependent myeloperoxidase system}]$$

Functions of histamine

Histamine results from the decarboxylation of histidine in activated mast cells, and act on immunocytes (e.g., monocytes, T cells, macrophages, neutrophils, eosinophils, B cells, and dendritic cells) and other cell types (including smooth muscle cells). This amine mediates allergic reactions in humans (Thangam et al., 2018) by serving as the endogenous ligand of histamine receptors H1R–H4R (a class of G protein-coupled receptors) present in different tissues. For example, H1R is found in many cells (including mast cells, the smooth muscle cells of blood vessels and bronchi, neurons, respiratory epithelium, and endothelial cells); H2R in the gastric mucosa, B cells, T cells, dendritic cells, smooth muscle cells, the brain, and heart; H3R exclusively in neurons; and H4R in immunocytes, intestinal epithelia, lungs, synovial tissue, and brain. In addition, histamine has a vasodilatory effect to increase permeability and lower blood pressure in humans. Furthermore, histamine regulates acetylcholine secretion by the central nervous system, while stimulating gastric acid secretion as well as the contraction of the heart and the smooth muscle cells of the lungs, uterus, and stomach. H1R and H2R antagonists have been used to treat allergy and inhibit gastric acid secretion (e.g., individuals with peptic ulcer disease and acid reflux esophagitis) in humans, respectively.

Functions of melanin

In the skin of humans, melanin (a metabolite of tyrosine) is a color pigment in melanocytes in the basal layer of the epidermis (Wu, 2022). Pigmentation by melanin has important biological, cosmetic and social significance. Eumelanin (black or brown) and pheomelanin (red or yellow in the skin and hair) are the known forms of melanin in the body. There are two types of eumelanin based on their colors: black or brown. Dark-skinned persons have more melanin than light-skinned individuals. Melanin protects the skin from ultraviolet light, and can affect stress response, immunity, sexual activity, and anti-oxidative responses (Lambert et al., 2019).

Functions of melatonin

Melatonin (a metabolite of tryptophan) is produced by the pineal gland and retina (Cipolla-Neto and Amaral, 2018). This hormone activates melatonin receptors (G protein-coupled receptors) on the plasma membrane to induce cell signaling. Concentrations of melatonin in the blood are altered during a daily cycle to regulate the circadian rhythms and sleep-wake pattern of humans. Physiological concentrations of melatonin confer anti-oxidant effects, modulate neurological function (e.g., memory and mood), aging, and immunity in humans (Strasser et al., 2016). Oral administration of melatonin can be used to alleviate sleep disorders due to circadian rhythm abnormalities and travel-related jet lag.

Functions of carnosine

Carnosine (a dipeptide consisting of β-alanine and histidine) is highly abundant in the human skeletal muscle (approximately 25 mM in women and 30 mM in men). This substance quenches reactive oxygen, nitrogen and carbonyl species, and other oxidants (Boldyrev et al., 2013). In addition, carnosine acts on histamine H1 or H3 receptors and on the hypothalamic suprachiasmatic nucleus (a master regulator of the circadian clock) in humans. Furthermore, carnosine activates the signaling cascades involving mitogen-activated protein kinase and cGMP-dependent protein kinase, while inhibiting proapoptotic signaling. Due to its histidine moiety, carnosine also play a role in pH buffering, which is of physiological importance in the contraction of skeletal muscle during intense exercise (Derave et al., 2019).

Functions of 4-hydroxyproline

4-Hydroxyproline is formed post-translationally from some proline residues in collagen, elastin, and certain regulatory proteins. This AA has structural, physiological, and nutritional significance in humans (Hu et al., 2021). For example, 4-hydroxyproline, along with proline, permits the sharp twisting of the collagen helix to establish the rigid structure of the collagen molecule in connective tissues. In addition, the presence of 4-hydroxyproline in the Gly–X–Y collagen peptides inhibits chemotaxis and apoptosis in neutrophils. Furthermore, 4-hydroxy-L-proline is a major substrate for glycine synthesis in human tissues and can scavenge reactive oxygen species. Finally, the formation of 4-hydroxyproline residues in protein kinases A and B, eukaryotic elongation factor 2 activity, and hypoxia-inducible transcription factor plays an important role in regulating their phosphorylation and catalytic activation, as well as metabolic signaling in cells (Hu et al., 2021).

Functions of glucosamine

Glucosamine (formed from glutamine and fructose-6-phosphate) inhibit constitutive NO synthesis in endothelial cells by reducing pentose cycle activity and, therefore, the intracellular levels of NADPH (Wu et al., 2001). This amino sugar also decreases inducible NO synthesis in immunologically activated macrophages and other cell types by suppressing the expression of the iNOS protein (Meininger et al., 2000). Furthermore, N-acetylglucosamine-6-phosphate is used to form all glycoproteins in cells. Thus, oral administration of glucosamine alleviates osteoarthritis in humans (Salazar et al., 2014).

Functions of dopamine

Dopamine (4-(2-aminoethyl)benzene-1,2-diol) is synthesized from tyrosine in dopaminergic neurons of the brain (Kopin, 1985). Tyrosine hydroxylase, a (6*R*)-5,6,7,8-tetrahydro-L-biopterin (BH4)-dependent enzyme, catalyzes the initial reaction of this metabolic pathway. Dopamine is a neurotransmitter in the central nervous system to modulate neurological development, learning, motor control, behavior, reward, emotion, and executive function in humans and to reduce their food intake (Liu et al., 2021). The physiological importance of dopamine is epitomized by the finding that the disruption of dopaminergic neurons in the basal ganglia results in Parkinson's disease in humans. In addition, immune cells, specifically T lymphocytes and dendritic cells, can convert tyrosine into dopamine, thereby modulating immune responses in mammals (Matt and Gaskill, 2020). Mental retardation and neurodegenerative disorders occur in individuals with little or no dopamine in the brain due to a deficiency of BH4 or phenylalanine hydroxylase and in patients with Parkinson's disease. L-3,4-Dihydroxyphenylalanine (L-DOPA), the most immediate precursor of dopamine, has beneficial effects in treating Parkinson's disease.

Functions of serotonin

The catabolism of tryptophan via the hydroxylation and decarboxylation pathway in endocrine cells of the gastrointestinal tract and in neurons of the brain generates serotonin (5-hydroxytryptamine), with 80–90% of the whole-body serotonin being present in the gastrointestinal tract (Sanger, 2008). In the small intestine of humans, this neurotransmitter stimulates gastrointestinal secretion, motility, and food intake. In the brain, serotonin modulates neuropsychological processes and neural activity to improve mood, cognition, behavior, and nocturnal sleep in humans, while reducing aggressive behavior. Thus, drugs that selectively inhibit the reuptake of serotonin from the synaptic cleft (a space between two neurons) into the presynaptic cell to increase its extracellular concentrations are usually used to treat depression, anxiety disorders, and other psychological conditions (Preskorn et al., 2004). In both the gut and the brain, serotonin and its metabolite *N*-acetylserotonin inhibit the production of inflammatory cytokines and superoxide, scavenge free radicals, and enhance host immunity (Schiavone et al., 2013). Furthermore, *N*-acetylserotonin is an inhibitor of sepiapterin reductase, an enzyme for the synthesis of BH4 (an essential cofactor for NO synthase), thereby decreasing NO synthesis by inducible NO synthase under inflammatory conditions. Furthermore, taste cells in the digestive tract may synthesize and release serotonin (Uneyama et al., 2006), which transmits electrical impulses from the vagal nerve to the forebrain to amplify taste signal transduction. Thus, low concentrations of tryptophan in plasma may contribute to mood, behavioral, bowel and neurological disorders. Note that branched-chain AAs (leucine, isoleucine, and valine) compete with tryptophan for uptake by the brain and, therefore, have been used to alleviate tryptophan-induced cognitive impairment and chronic fatigue disorders in humans, particularly those with hepatic encephalopathy (Holeček, 2018).

Functions of quinolinic acid and kynurenic acid

Quinolinic acid and kynurenic acid are two of the metabolites of tryptophan catabolism via the kynurenine pathway in the brain of humans (Wu, 2022). Interestingly, these two substances have opposite effects, as quinolinic acid and kynurenic acid are an agonist and a nonselective antagonist of the NMDA receptor (Schwarcz and Stone, 2017). Because quinolinic acid can activate but kynurenic acid can block AA-modulated excitation of the brain, their imbalance can result in an excitotoxic neuronal cell death. This may be a pathogenesis of some neurological diseases, including Huntington's chorea and epilepsy.

Functions of AAs and metabolites as activators of the aryl hydrocarbon receptor

Tryptophan and its metabolites (e.g., indole acetic acid, kynurenine, and tryptamine), as well as bilirubin (a metabolite of heme), are natural ligands and activators of the aryl hydrocarbon receptor (AhR; Bessede et al., 2014). Upon ligand binding, AhR undergoes a conformational change and translocates from the cytosol to the nucleus, resulting in the expression of target genes (e.g., cytochrome P450s) to catalyze the monooxygenation of various endogenous and exogenous substrates for excretion in the urine and feces. Furthermore, Ahr confers anti-oxidative and anti-inflammatory effects in cells (Rothhammer and Quintana, 2019).

Functions of heme

Heme is synthesized from glycine and succinyl-CoA by virtually all cell types in humans, with approximately 85% and 15% of heme being formed in immature red blood cells of the bone marrow and the liver (mainly hepatocytes), respectively (Ajioka et al., 2006). Heme accounts for 95% of iron in the human body and about two-thirds of dietary iron intake by humans in developed nations. Heme is an essential component of hemoglobin, myoglobin, and other heme-containing proteins (including enzymes), with hemoglobin accounting for 85% of the total heme in humans (Ferreira, 2013). These proteins are essential for oxygen transport in the blood, oxygen storage in tissues (particularly skeletal muscle and heart), proteins in the mitochondrial electron transport system (e.g., cytochromes), and heme enzymes (Wu, 2018). The latter include catalyze, guanylyl cyclase, indoleamine 2,3-dioxygenase, myeloperoxidase, peroxidase, prostaglandin endoperoxide synthase, and tryptophan 2,3-dioxygenase. In addition, intestinal bacteria express heme-containing nitrite reductase to convert nitrite into NO. Anemia, stunting, physical weakness, and metabolic defects occur in individuals with a deficiency of heme synthesis.

Functions of AAs in sensing

AAs are crucial for the functions of sense organs, including the eyes, ears, nose, tongue, and skin that provide senses of sight, hearing, smell, taste, and touch, respectively to aid the survival, development, learning, and adaptation of humans. The chemosensory transduction is the biochemical basis of sensing responses, which help humans to avoid foods with strong bitter tastes, sour tastes, or unusual (e.g., bad or strong) smell, and to protect the organisms against consuming toxic substances or toxic foods (San Gabriel and Uneyama, 2013; Wu, 2020b). For example, taste cells express different taste receptors (transmembrane proteins) for different tastants. The gene *Tas1R* encodes for sweet and umami receptor proteins, which bind sweet AAs and glutamate, respectively. By contrast, the gene *Tas2R* encodes for bitter receptor proteins, which bind bitter AAs. Sour and salty taste receptors are transmembrane H^+ and Na^+ channels, respectively.

Functions of D-AAs

AA racemases can convert certain L-AAs into D-AAs in humans and their intestinal microbes, including D-aspartate, D-serine, D-alanine, and D-glutamate (Friedman and Levin, 2012; Guercio and Panizzutti, 2018; Lee et al., 2020). High concentrations of D-aspartate, D-serine, and D-alanine occur in the endocrine glands and brain to regulate hormone secretion, as well as neurological growth, development and function (Errico et al., 2012; Flynn et al., 2020). For example, D-alanine is an agonist of the glycine site on the NMDA subtype glutamate receptor, thereby possibly affecting memory function and synaptic plasticity (Baccari et al., 2020). In addition, D-aspartate is the precursor for NMDA (Baccari et al., 2020; D'Aniello, 2007) and also binds NMDA receptors to affect long-term potentiation and spatial memory. Likewise, D-serine is an endogenous ligand for the glycine site of the NMDA receptor and functions as a novel neurotransmitter (McKay et al., 2019). Of note, both glycine and D-serine have been proposed as adjuvant therapy to the standard psychopharmacological treatment of individuals with schizophrenia. In the small and large intestines, some peptides that contain D-AAs are potent antimicrobials (Wu, 2018).

Summary and outlook

AAs are substrates for the syntheses of not only proteins and peptides but also other nitrogen-sulfur-, and oxygen-containing substances with enormous versatility and physiological significance. Some AAs (e.g., aspartate, glutamate, and glycine) and metabolites (e.g., GABA, NO, CO, and H_2S) are neurotransmitters. Some AAs (e.g., arginine, glutamine, glycine, leucine, tryptophan, and valine) activate the MTOR cell signaling pathway to stimulate protein synthesis and inhibit autophagy-mediated proteolysis. Some

AAs fulfill structural (e.g., proline, 4-hydroxyproline, and 5-hydroxylysine), metabolic (e.g., alanine, leucine, and glutamate), physiological (e.g., D-aspartate, D-serine, and taurine), and immunological (e.g., arginine, tryptophan, and glycine) needs of humans. Most physiological metabolites (e.g., creatine, heme, formate, purines, pyrimidines, and polyamines) are essential for biochemical processes (e.g., ATP production, energy provision, oxygen delivery, and cell growth). Thus, through direct and indirect actions, specific functions of AAs in humans include their roles in regulating gene expression, cell signaling pathways, food intake, chemical sensing, the digestion and absorption of dietary nutrients, neurological development and behavior, the metabolism of energy and nutrients, immunity, reproduction, hormone secretion, anti-oxidative and anti-inflammatory responses, the detoxification of endogenous and exogenous substances, acid-base and mineral balances, osmolarity, whole-body homeostasis, health and well-being (Bröer and Bröer, 2017; Li and Wu, 2022; Wu, 2021). Therefore, dietary supplementation with functional AAs (e.g., L-arginine, L–glutamate, and glycine) can improve the integrity and health of tissues and cells (e.g., the small intestine, brain, liver, skeletal muscle, heart, blood vessels, testes, ovaries, and immunocytes), enhance immunity, anti-oxidative responses, fertility, wound healing, ammonia detoxification, and lean tissue mass; ameliorate metabolic syndromes (including dyslipidemia, obesity, diabetes, and hypertension); and treat individuals with erectile dysfunction, sickle cell disease, muscular dystrophy, and pre-eclampsia in humans. Like any other nutrients, excessive AA intakes beyond upper safe levels can disturb whole-body homeostasis and must be avoided (Wu, 2022). Further research on dietary requirements (particularly functional AAs) is warranted to optimize human protein nutrition research and feeding practice.

Acknowledgments

Work is the author's laboratory is supported by Texas A&M AgriLife Research Hatch Project H-8200 and by Agriculture and Food Research Initiative Competitive Grants from the USDA National Institute of Food and Agriculture.

References

Agostinelli, E., 2020. Biochemical and pathophysiological properties of polyamines. Amino Acids 52, 111–117.
Ajioka, R.S., Phillips, J.D., Kushner, J.P., 2006. Biosynthesis of heme in mammals. Biochim. Biophys. Acta 1763, 723–736.
Akaberi, D., Krambrich, J., Ling, J., et al., 2020. Mitigation of the replication of SARS-CoV-2 by nitric oxide in vitro. Redox Biol. 37, 101734.
Åkerström, S., Mousavi-Jazi, M., Klingström, J., Leijon, M., Lundkvist, Å., Mirazimi, A., 2005. Nitric oxide inhibits the replication cycle of severe acute respiratory syndrome coronavirus. J. Virol. 79, 1966–1969.
Baccari, G.C., Falvo, S., Santillo, A., Russo, F.D.G., Di Fiore, M.M., 2020. D-amino acids in mammalian endocrine tissues. Amino Acids 52, 1263–1273.
Bazer, F.W., Seo, H., Johnson, G.A., Wu, G., 2021. One-carbon metabolism and development of the conceptus during pregnancy: lessons from studies with sheep and pigs. Adv. Exp. Med. Biol. 1285, 1–15.
Beaumont, M., Blachier, F., 2020. Amino acids in intestinal physiology and health. Adv. Exp. Med. Biol. 1265, 1–20.
Becerra-Solano, L.E., Butler, J., Castañeda-Cisneros, G., et al., 2009. A missense mutation, p.V132G, in the X-linked spermine synthase gene (SMS) causes Snyder-Robinson syndrome. Am. J. Med. Genet. 149, 328–335.
Beltowski, J., 2015. Hydrogen sulfide in pharmacology and medicine—an update. Pharmacol. Rep. 67, 647–658.
Bessede, A., Gargaro, M., Pallotta, M.T., et al., 2014. Aryl hydrocarbon receptor control of a disease tolerance defence pathway. Nature 511, 184–190.
Biefer, H.R.C., Vasudevan, A., Elkhal, A., 2017. Aspects of tryptophan and nicotinamide adenine dinucleotide in immunity: a new twist in an old tale. Int. J. Tryptophan Res. 10, 1–8.
Boldyrev, A.A., Aldini, G., Derave, W., 2013. Physiology and pathophysiology of carnosine. Physiol. Rev. 93, 1803–1845.
Brasse-Lagnel, C.G., Lavoinne, A.M., Husson, A.S., 2010. Amino acid regulation of mammalian gene expression in the intestine. Biochimie 92, 729–735.
Bröer, S., Bröer, A., 2017. Amino acid homeostasis and signaling in mammalian cells and organisms. Biochem. J. 474, 1935–1963.
Brosnan, J.T., Brosnan, M.E., 2007. Creatine: endogenous metabolite, dietary, and therapeutic supplement. Annu. Rev. Nutr. 27, 241–261.
Brosnan, J.T., Brosnan, M.E., 2013. Glutamate: a truly functional amino acid. Amino Acids 45, 413–418.
Calder, P.C., Field, C.J., Gill, H.S. (Eds.), 2002. Nutrition and Immune Function. CABI, Wallingford.
Chen, J.Q., Jin, Y.H., Yang, Y., Wu, Z.L., Wu, G., 2020. Epithelial dysfunction in lung diseases: effects of amino acids and potential mechanisms. Adv. Exp. Med. Biol. 1265, 57–70.
Cipolla-Neto, J., Amaral, F.G.D., 2018. Melatonin as a hormone: new physiological and clinical insights. Endocr. Rev. 39, 990–1028.
Closs, E.I., Scheld, J.S., Sharafi, M., Förstermann, U., 2000. Substrate supply for nitric-oxide synthase in macrophages and endothelial cells: role of cationic amino acid transporters. Mol. Pharmacol. 57, 68–74.
Curthoys, N.P., Watford, M., 1995. Regulation of glutaminase expression and glutamine metabolism. Annu. Rev. Nutr. 15, 133–159.
Cynober, L.A. (Ed.), 2004. Metabolic and Therapeutic Aspects of Amino Acids in Clinical Nutrition, second ed. CRC Press, Boca Raton.
Dai, Z.L., Wu, G., Zhu, W.Y., 2011. Amino acid metabolism in intestinal bacteria: links between gut ecology and host health. Front. Biosci. 16, 1768–1786.
Dai, Z.L., Wu, Z.L., Yang, Y., et al., 2013. Nitric oxide and energy metabolism in mammals. Biofactors 39, 383–391.
Dai, Z.L., Wu, Z.L., Hang, S.Q., Zhu, W.Y., Wu, G., 2015. Amino acid metabolism in intestinal bacteria and its potential implications for mammalian reproduction. Mol. Hum. Reprod. 21, 389–409.
D'Aniello, A., 2007. D-aspartic acid: an endogenous amino acid with an important neuroendocrine role. Brain Res. Rev. 53, 215–234.
Davila, A.-M., Blachier, F., Gotteland, M., et al., 2013. Intestinal luminal nitrogen metabolism: role of the gut microbiota and consequences for the host. Pharmacol. Res. 68, 95–107.
Davis, T.A., Nguyen, H.V., Garcia-Bravo, R., et al., 1994. Amino acid composition of human milk is not unique. J. Nutr. 124, 1126–1132.
Derave, W., De Courten, B., Baba, S.P., 2019. An update on carnosine and anserine research. Amino Acids 51, 1–4.
Diotallevi, M., Checconi, P., Palamara, A.T., et al., 2017. Glutathione fine-tunes the innate immune response toward antiviral pathways in a macrophage cell line independently of its antioxidant properties. Front. Immunol. 8, 1239.
Durante, W., 2020. Amino acids in circulatory function and health. Adv. Exp. Med. Biol. 1265, 39–56.
Errico, F., Napolitano, F., Nisticò, R., Usiello, A., 2012. New insights on the role of free D-aspartate in the mammalian brain. Amino Acids 43, 1861–1871.
Fernstrom, J.D., 2013. Large neutral amino acids: dietary effects on brain neurochemistry and function. Amino Acids 45, 419–430.
Ferreira, G.C., 2013. Heme synthesis. In: Lennarz, W.J., Lane, M.D. (Eds.), Encyclopedia of Biological Chemistry. Academic Press, San Diego, pp. 539–542.

Field, C.J., Johnson, I.R., Schley, P.D., 2002. Nutrients and their role in host resistance to infection. J. Leukoc. Biol. 71, 16–32.
Flynn, N.E., Shaw, M.H., Becker, J.T., 2020. Amino acids in health and endocrine function. Adv. Exp. Med. Biol. 1265, 97–109.
Forbes, J., Krishnamurthy, K., 2020. Biochemistry, Peptide. StatPearls. https://www.ncbi.nlm.nih.gov/books/NBK562260/. (Accessed 18 August 2021).
Friedman, M., Levin, C.E., 2012. Nutritional and medicinal aspects of D-amino acids. Amino Acids 42, 1553–1582.
Gietzen, D.W., Aja, S.M., 2012. The brain's response to an essential amino acid-deficient diet and the circuitous route to a better meal. Mol. Neurobiol. 46, 332–348.
Guercio, G.D., Panizzutti, R., 2018. Potential and challenges for the clinical use of d-serine as a cognitive enhancer. Front. Psychiatr. 9, 14.
He, W.L., Wu, G., 2020. Metabolism of amino acids in the brain and their roles in regulating food intake. Adv. Exp. Med. Biol. 1265, 167–185.
Holeček, M., 2018. Branched-chain amino acids in health and disease: metabolism, alterations in blood plasma, and as supplements. Nutr. Metab. 15, 33.
Hou, Y.Q., Wu, G., 2017. Nutritionally nonessential amino acids: a misnomer in nutritional sciences. Adv. Nutr. 8, 137–139.
Hou, Y.Q., Yin, Y.L., Wu, G., 2015. Dietary essentiality of "nutritionally nonessential amino acids" for animals and humans. Exp. Biol. Med. 240, 997–1007.
Hou, Y.Q., He, W.L., Hu, S.D., Wu, G., 2019. Composition of polyamines and amino acids in plant-source foods for human consumption. Amino Acids 51, 1153–1165.
Hou, Y.Q., Hu, S.D., Li, X.Y., He, W.L., Wu, G., 2020. Amino acid metabolism in the liver: nutritional and physiological significance. Adv. Exp. Med. Biol. 1265, 21–37.
Hu, S.D., He, W.L., Wu, G., 2021. Hydroxyproline in animal metabolism, nutrition, and cell signaling. Amino Acids. https://doi.org/10.1007/s00726-021-03056-x.
Huxtable, R.J., 1992. Physiological actions of taurine. Physiol. Rev. 72, 101–163.
Jiang, Z.M., Wang, L.J., Qi, Y., et al., 1993. Comparison of parenteral nutrition supplemented with L-glutamine or glutamine dipeptides. J. Parenter. Enteral Nutr. 17, 134–141.
Kamat, P.K., Kalani, A., Tyagi, N., 2015. Role of hydrogen sulfide in brain synaptic remodeling. Meth. Enzymol. 555, 207–229.
Kopin, I.J., 1985. Catecholamine metabolism: basic aspects and clinical significance. Pharmacol. Rev. 37, 333–364.
Lambert, M.W., Maddukuri, S., Karanfilian, K.M., Elias, M.L., Lambert, W.C., 2019. The physiology of melanin deposition in health and disease. Clin. Dermatol. 37, 402–417.
Le Floc'h, N., Otten, W., Merlot, E., 2011. Tryptophan metabolism, from nutrition to potential therapeutic applications. Amino Acids 41, 1195–1205.
Lee, C.L., Qiu, T.A., Sweedler, J.V., 2020. D-alanine: distribution, origin, physiological relevance, and implications in disease. Biochim. Biophys. Acta 1868, 140482.
Lei, J., Feng, D.Y., Zhang, Y.L., et al., 2012. Nutritional and regulatory role of branched-chain amino acids in lactation. Front. Biosci. 17, 2725–2739.
Li, X.L., Johnson, G.A., Zhou, H.J., et al., 2022. Microarray analysis reveals an important role for dietary L-arginine in regulating global gene expression in porcine placentae during early gestation. Front. Biosci. (Landmark edition) 27 (1), 033.
Li, P., Wu, G., 2020. Composition of amino acids and related nitrogenous nutrients in feedstuffs for animal diets. Amino Acids 52, 523–542.
Li, P., Wu, G., 2022. Important roles of amino acids in immune response. Br. J. Nutr. 127, 398–402.
Li, P., Yin, Y.L., Li, D.F., Kim, S.W., Wu, G., 2007. Amino acids and immune function. Br. J. Nutr. 98, 237–252.
Li, X.L., Bazer, F.W., Gao, H., et al., 2009. Amino acids and gaseous signaling. Amino Acids 37, 65–78.
Li, X.Y., Zheng, S.X., Wu, G., 2020. Amino acid metabolism in the kidneys: nutritional and physiological significance. Adv. Exp. Med. Biol. 1265, 71–95.
Li, P., He, W.L., Wu, G., 2021. Composition of amino acids in foodstuffs for humans and animals. Adv. Exp. Med. Biol. 1332, 189–210.
Liu, C., Goel, P., Kaeser, P.S., 2021. Spatial and temporal scales of dopamine transmission. Nat. Rev. Neurosci. 22, 345–358.
MacKay, M., Kravtsenyuk, M., Thomas, R., et al., 2019. D-serine: potential therapeutic agent and/or biomarker in schizophrenia and depression? Front. Psychiatr. 10, 25.
Marcinkiewicz, J., Kontny, E., 2014. Taurine and inflammatory diseases. Amino Acids 46, 7–20.
Matt, S.M., Gaskill, P.J., 2020. Where is dopamine and how do immune cells see it? Dopamine-mediated immune cell function in health and disease. J. Neuroimmune Pharmacol. 15, 114–164.
McNeal, C.J., Meininger, C.J., Wilborn, C.D., Tekwe, C.D., Wu, G., 2018. Safety of dietary supplementation with arginine in adult humans. Amino Acids 50, 1215–1229.
Meininger, C.J., Kelly, K.A., Li, H., Haynes, T.E., Wu, G., 2000. Glucosamine inhibits inducible nitric oxide synthesis. Biochem. Biophys. Res. Commun. 279, 234–239.
Meister, A., Anderson, M.E., 1983. Glutathione. Annu. Rev. Biochem. 52, 711–760.
Meléndez-Hevia, E., De Paz-Lugo, P., Cornish-Bowden, A., Cárdenas, M.L., 2009. A weak link in metabolism: the metabolic capacity for glycine biosynthesis does not satisfy the need for collagen synthesis. J. Biosci. 34, 853–872.
Moffett, J.R., 2003. Tryptophan and the immune system. Immunol. Cell Biol. 81, 247–265.
Mortimore, G.E., Khurana, K.K., 1990. Regulation of protein degradation in the liver. Int. J. Biochem. 22, 1075–1080.
Nakahira, K., Kim, H.P., Geng, X.H., et al., 2006. Carbon monoxide differentially inhibits TLR signaling pathways by regulating ROS-induced trafficking of TLRs to lipid rafts. J. Exp. Med. 203, 2377–2389.
Newsholme, P., Curi, R., Curi, T.C.P., et al., 1999. Glutamine metabolism by lymphocytes, macrophages, and neutrophils: its importance in health and disease. JNB 10, 316–324.
Nieto-Alamilla, G., Márquez-Gómez, R., García-Gálvez, A.-M., Morales-Figueroa, G.-E., Arias-Montaño, J.W., 2016. The histamine H_3 receptor: structure, pharmacology, and function. Mol. Pharmacol. 90, 649–673.
Otterbein, L.E., Bach, F.H., Alam, J., et al., 2000. Carbon monoxide has anti-inflammatory effects involving the mitogen-activated protein kinase pathway. Nat. Med. 6, 422–428.
Park, M.H., Wolff, E.C., 2018. Hypusine, a polyamine-derived amino acid critical for eukaryotic translation. J. Biol. Chem. 293, 18710–18718.
Paudel, S., Wu, G., Wang, X.Q., 2021. Amino acids in cell signaling: regulation and function. Adv. Exp. Med. Biol. 1332, 17–33.
Pegg, A.E., 2016. Functions of polyamines in mammals. J. Biol. Chem. 291, 4904–14912.
Phang, J.M., 2019. Proline metabolism in cell regulation and cancer biology: recent advances and hypotheses. Antioxid. Redox Signal. 30, 635–649.
Philip, M., Snow, R.J., Della Gatta, P.A., Bellofiore, N., Ellery, S.J., 2020. Creatine metabolism in the uterus: potential implications for reproductive biology. Amino Acids 52, 1275–1283.
Polonikov, A., 2020. Endogenous deficiency of glutathione as the most likely cause of serious manifestations and death in COVID-19 patients. ACS Infect. Dis. 6, 1558–1562.
Popovic, P.J., Zeh, H.J., Ochoa, J.B., 2007. Arginine and immunity. J. Nutr. 137, 1681S–1686S.
Posey, E.A., Bazer, F.W., Wu, G., 2021. Amino acids and their metabolites for improving human exercising performance. Adv. Exp. Med. Biol. 1332, 151–166.
Preskorn, S.H., Ross, R., Stanga, C.Y., 2004. Selective serotonin reuptake inhibitors. In: Preskorn, S.H., Feighner, H.P., Stanga, C.Y., Ross, R. (Eds.), Antidepressants: Past, Present and Future. Springer, Berlin, pp. 241–262.
Rhoads, J.M., Wu, G., 2009. Glutamine, arginine, and leucine signaling in the intestine. Amino Acids 37, 111–122.
Rothhammer, V., Quintana, F.J., 2019. The aryl hydrocarbon receptor: an environmental sensor integrating immune responses in health and disease. Nat. Rev. Immunol. 19, 184–197.
Ryan, P., Riechman, S.E., Fluckey, J.D., Wu, G., 2021. Interorgan metabolism of amino acids in health and disease. Adv. Exp. Med. Biol. 1332, 129–149.
Salazar, J., Bello, L., Chávez, M., et al., 2014. Glucosamine for osteoarthritis: biological effects, clinical efficacy, and safety on glucose metabolism. Arthritis 2014, 432463.
San Gabriel, A., Uneyama, H., 2013. Amino acid sensing in the gastrointestinal tract. Amino Acids 45, 451–461.
Sanger, G.J., 2008. 5-hydroxytryptamine and the gastrointestinal tract: where next? Trends Pharmacol. Sci. 29, 465–471.
Schiavone, S., Jaquet, V., Trabace, L., Krause, K., 2013. Severe life stress and oxidative stress in the brain: from animal models to human pathology. Antioxid. Redox Signal. 18, 1475–1490.
Schwarcz, R., Stone, T.W., 2017. The kynurenine pathway and the brain: challenges, controversies and promises. Neuropharmacology 112, 237–247.
Shen, J.Z., Wu, G., Guo, S.D., 2021. Amino acids in autophagy: regulation and function. Adv. Exp. Med. Biol. 1332, 51–66.
Sies, H., 1999. Glutathione and its role in cellular functions. Free Radic. Biol. Med. 27, 916–921.
Silvagno, F., Vernone, A., Pescarmona, G.P., 2020. The role of glutathione in protecting against the severe inflammatory response triggered by COVID-19. Antioxidants 9, 624.
Strasser, B., Gostner, J.M., Fuchs, D., 2016. Mood, food, and cognition: role of tryptophan and serotonin. Curr. Opin. Clin. Nutr. Metab. Care 19, 55–61.
Suryawan, A., Davis, T.A., 2011. Regulation of protein synthesis by amino acids in muscle of neonates. Front. Biosci. 16, 1445–1460.

Thangam, E.B., Jemima, E.A., Singh, H., et al., 2018. The role of histamine and histamine receptors in mast cell-mediated allergy and inflammation: the hunt for new therapeutic targets. Front. Immunol. 9, 1873.
Tomé, D., 2018. The roles of dietary glutamate in the intestine. Ann. Nutr. Metabol. 73 (Suppl. 5), 15–20.
Tsikas, D., 2021. Post-translational modifications (PTM): analytical approaches, signaling, physiology and pathophysiology—part I. Amino Acids 53, 485–487.
Uneyama, H., Nijjima, A., San Gabriel, A., Torii, K., 2006. Luminal amino acid sensing in the rat gastric mucosa. Am. J. Physiol. 291, G1163–G1170.
Wallimann, T., Tokarska-Schlattner, M., Schlattner, U., 2011. The creatine kinase system and pleiotropic effects of creatine. Amino Acids 40, 1271–1296.
Wang, Z., Shen, W., Kotler, D.P., et al., 2003. Total body protein: a new cellular level mass and distribution prediction model. Am. J. Clin. Nutr. 78, 979–984.
Watford, M., 2008. Glutamine metabolism and function in relation to proline synthesis and the safety of glutamine and proline supplementation. J. Nutr. 138, 2003S–2007S.
Wetzel, M.D., Gao, T., Venkatachalam, M., Morris Jr., S.M., Awad, A.S., 2019. L-homoarginine supplementation prevents diabetic kidney damage. Physiol. Rep. 7, e14235.
Wright, C.E., Tallan, H.H., Lin, Y.Y., Gaull, G.E., 1986. Taurine: biological update. Annu. Rev. Biochem. 55, 427–453.
Wu, G., Haynes, T.E., Li, H., Yan, W., Meininger, C.J., 2001. Glutamine metabolism to glucosamine is necessary for glutamine inhibition of endothelial nitric oxide synthesis. Biochem. J. 353, 245–252.
Wu, G., Fang, Y.Z., Yang, S., Lupton, J.R., Turner, N.D., 2004. Glutathione metabolism and its implications for health. J. Nutr. 134, 489–492.
Wu, G., Wu, Z.L., Dai, Z.L., et al., 2013. Dietary requirements of "nutritionally nonessential amino acids" by animals and humans. Amino Acids 44, 1107–1113.
Wu, G., Meininger, C.J., McNeal, C.J., Bazer, F.W., Rhoads, J.M., 2021. Role of L-arginine in nitric oxide synthesis and health in humans. Adv. Exp. Med. Biol. 1332, 167–187.
Wu, G., 2009. Amino acids: metabolism, functions, and nutrition. Amino Acids 37, 1–17.
Wu, G., 2010. Functional amino acids in growth, reproduction and health. Adv. Nutr. 1, 31–37.
Wu, G., 2016. Dietary protein intake and human health. Food Funct. 7, 1251–1265.
Wu, G., 2018. Principles of Animal Nutrition. CRC Press, Boca Raton.
Wu, G., 2020a. Important roles of dietary taurine, creatine, carnosine, anserine and hydroxyproline in human nutrition and health. Amino Acids 52, 329–360.
Wu, G., 2020b. Metabolism and functions of amino acids in sense organs. Adv. Exp. Med. Biol. 1265, 201–217.
Wu, G., 2021. Amino acids in nutrition, health, and disease. Front. Biosci. (Landmark Edition) 26, 1386–1392.
Wu, G., 2022. Amino Acids: Biochemistry and Nutrition, second ed. CRC Press, Boca Raton.
Wyss, M., Kaddurah-Daouk, R., 2000. Creatine and creatinine metabolism. Physiol. Rev. 80, 1107–1213.
Ziegler, T.R., Bye, R.L., Persinger, R.L., et al., 1998. Effects of glutamine supplementation on circulating lymphocytes after bone marrow transplantation: a pilot study. Am. J. Med. Sci. 315, 4–10.

Antioxidants: Intervention studies

Rafael A. Casuso[a], **Javier González Gallego**[b], **and Jesús Rodríguez Huertas**[a], [a] Department of Physiology, Institute of Nutrition and Food Technology, Biomedical Research Center, University of Granada, Granada, Spain; and [b] Institute of Biomedicine (IBIOMED), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), University of León, León, Spain

© 2023 Elsevier Ltd. All rights reserved.

Introduction 48
Effects of antioxidant supplementation on aging 50
Effects of antioxidant supplementation on diabetes, metabolic syndrome, obesity and hypertension 50
Inclusion criteria and search strategy for meta-analysis 50
Data extraction and synthesis 52
Meta-analysis 52
Fasting glucose, HbA_{1c} % and body mass index 52
Blood pressure 54
Effects of antioxidant supplementation on cancer development risk 54
Effects of antioxidant supplementation on exercise adaptations and physical performance 56
Conclusions 56
References 58
Further reading 60

Key points

We summarize and discuss findings from intervention trials with antioxidants (vitamin C and vitamin E) supplementation (Fig. 1). Information available for, at least, these two vitamins make possible to shed light on the issue of its usefulness in patients suffering from conditions associated with oxidative stress: diabetes, metabolic syndrome, obesity, hypertension, cancer, other pathologies associated to aging as well as the role of physical performance as a tool to strong the health and prevented these diseases.

- Vitamin C and E supplementation may reduce fasting blood glucose and HbA_{1c} levels in subjects metabolically compromised but this reduction is unlikely to have a clinically meaningful effect.
- Vitamin supplementation concomitant to exercise training could prevent the gain of ~1 mL/kg/min induced by training.
- An analysis of large cohort studies revealed a trend toward higher cancer risk among subjects supplemented with antioxidants.
- In certain situations, loading the cell with high doses of antioxidants leads to a blunting of the positive effects of a balanced diet and/or exercise training and interferes with important ROS-mediated physiological processes.
- We recommend that an adequate intake of antioxidants through a varied and balanced diet remains the best approach to maintain the optimal antioxidant status in individuals.

Introduction

The oxygen molecule was introduced into the earth's atmosphere several billion years ago due to the appearance of photosynthetic oxygen-releasing organisms. Oxygen played a key role in the evolutionary process of erobic organisms, due to its ability to act as a final electron acceptor in the mitochondrial transport chain. Oxidative metabolism has the advantage to generate a large amount of energy through the complete combustion of glucose. However, the process requires a sequence of reactions mediated by enzymatic systems that generate the superoxide anion, which in turn initiates a cascade of production of reactive oxygen species (ROS), such as hydrogen peroxide, the hydroxyl radical or the oxygen singlet. The oxygen molecule can also generate nitric oxide, and this in turn reacts with the superoxide anion to produce peroxynitrite, a highly reactive and deleterious molecule.

ROS play relevant physiological roles; reduced levels of these molecules are required for cell signaling and modulation of processes such as growth and differentiation, immune function, mitochondrial biogenesis, apoptosis, angiogenesis or muscle hypertrophy. From this point of view, molecules such as superoxide anion or nitric oxide are essential for the body. However, when production is excessive and/or the levels of antioxidant defenses are reduced, oxidative/nitrosative stress and cell damage occur, with alterations in macromolecules such as lipid proteins or nucleic acids. Moreover, evidence accumulated in recent years suggests that increased oxidative stress unaccompanied by antioxidant adaptations is the trigger for the pathophysiology of inflammation, through activation of the transcription factor NF-κappaB and other pathways, such as those involving mitogen-activated

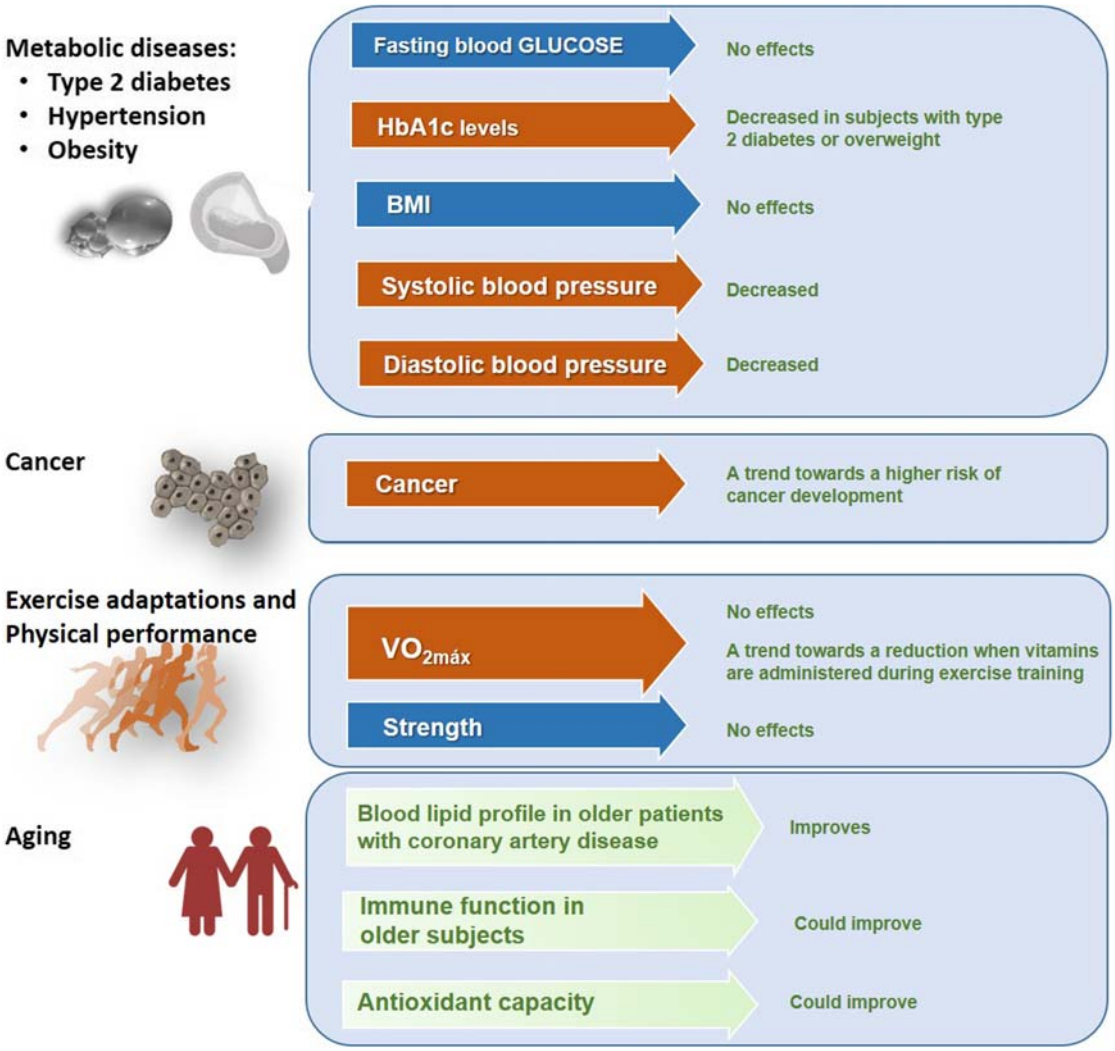


Fig. 1 Graphical abstract of the intervention studies.

protein kinases (MAPK). As a result, a loss of physiological regulatory capacity can occur in various systems, such as cardiovascular, respiratory, or neuromuscular. Furthermore, ROS-mediated damage can contribute to the development of chronic diseases, such as cardiovascular and neurodegenerative diseases, as well as a higher incidence of tumors.

The cell is able to maintain a balance in the redox state when low and moderate levels of these molecules are generated, thanks to the antioxidant systems. The body uses enzymatic and non-enzymatic antioxidant systems to protect membranes and macromolecules from ROS-induced damage. Non-enzymatic antioxidants include dietary vitamins and provitamins (vitamins C, E, and beta-carotene), polyphenols (especially flavonoids), thiols (such as reduced glutathione), and various low-molecular weight protein compounds, such as ubiquinone, uric acid, and others. In contrast to other vertebrates, the human organism is not capable of synthesizing antioxidant vitamins, which means that the plasma and tissue levels of non-enzymatic antioxidants depend on the quality and quantity of the food eaten.

A large number of in vitro and animal research has shown beneficial effects of antioxidants in different conditions associated with oxidative/inflammatory damage, and results from observational studies suggest that lack of dietary antioxidants is casually related to chronic diseases risk. However, data obtained with those approaches do not allow for drawing conclusions regarding causality, and a number of intervention trials have been (and are being) carried out to shed light on this issue. The existing evidence shows potential positive-health effects, but the diversity of findings (including negative effects), the design variation between intervention trials, the small number of patients recruited, or the different characteristics of patients, make difficult to reach valid conclusions for most dietary antioxidants.

In this article we summarize and discuss findings from intervention trials with vitamin C and vitamin E supplementation. Information available for, at least, these two vitamins makes possible to shed light on its usefulness in patients suffering from conditions associated with oxidative stress. First, we performed a systematic review with the aim of summarizing the relationship between

antioxidant intake and a variety of aging markers. We next performed an in-deep search aiming to quantify through meta-analytic procedures the effects of antioxidant supplementation on health markers related to obesity and diabetes as well as hypertension. We also analyzed long-term antioxidant supplementation in healthy older subjects on the risk of cancer development during follow-up (3–10 years). Finally, as exercise exerts powerful protection against these metabolic diseases, we explored the effect of antioxidant supplementation on exercise adaptations.

Effects of antioxidant supplementation on aging

There are numerous and very diverse alterations derived from aging or that have an impact on it. Sufficient homogeneous variables have not been found to perform a meta-analysis about vitamin C and/or vitamin E supplementation and its impact on aging. The physiological parameters related to aging that are measured in clinical trials are very diverse: mortality, antioxidant capacity, various markers indicating genetic damage, blood lipid profile, blood pressure, immune function, eye disease and cognitive impairment. Since the variables were so heterogeneous, it was not possible to perform a meta-analysis, so a systematic review was performed (**Table 1**).

A search was performed in the Web of Science database, using the following keywords: ("aging") and ("vitamin*" or "vitamin C" or "ascorbic acid" or "vitamin E" or "tocopherol"). The filter "Clinical Trial" was used, obtaining 802 results. The following were established as inclusion criteria: trials that only used vitamin C and/or E supplements, without including other vitamins and antioxidant substances; trials that measured physiological parameters that have an impact on or are a cause of aging; and trials published in journals ranked Q1 in JCR.

Clinical trials using cocktails of antioxidant substances, such as multivitamins, fatty acids, zinc, etc., and those published in lower impact journals were excluded. Finally, 21 clinical trials were analyzed, consisting of very diverse methodologies, finding significant variations in the age range of the participating subjects, in the habits or underlying diseases of the subjects, and in the physiological parameters evaluated.

The effects of vitamin E supplementation on mortality depend on the dietary intake of vitamin C and the age of the subjects (Hemilä and Kaprio, 2009, 2011). In supplemented subjects consuming higher dietary amounts of vitamin C, from 66 years of age onwards, mortality decreases (Hemilä and Kaprio, 2009, 2011); whereas mortality increases in supplemented subjects younger than 65 years of age (Hemilä and Kaprio, 2009). Regarding antioxidant capacity, significant positive effects have been seen in subjects supplemented with vitamin C, vitamin E or both (Lara-Padilla et al., 2007; Naeini et al., 2014). However, a higher antioxidant capacity does not influence the cognitive function of patients (Naeini et al., 2014; Cetin et al., 2010). On the other hand, there are contradictory results on the effects of both vitamins on genetic damage, and these effects could be influenced by the age of the subjects (Sacheck et al., 2003; Huang et al., 2000). Immune function seems to be favored by vitamin E supplementation, but there are no significant differences vs. placebo (Meydani et al., 1990; Meydani, 1995; Pallast et al., 1999). Vascular parameters are not affected by vitamin C supplementation (Singh et al., 2002). In older subjects with coronary artery disease, vitamin E supplementation significantly improves the lipid profile vs. placebo (Paolisso et al., 1995). Alzheimer's disease alters telomeric profiles, and vitamin E supplementation shows no effect on these profiles (Guan et al., 2012). It also had no effect on cognitive performance in a trial conducted in patients with Down syndrome (Sano et al., 2016). Age-related eye diseases, such as maculopathy and cataracts, were also unaffected by vitamin C and E supplementation (Christen et al., 1999, 2008, 2010a,b, 2015; Teikari et al., 1997; McNeil et al., 2004).

In conclusion, vitamin E supplementation seems to improve the blood lipid profile in older patients with coronary artery disease, and could improve immune function in older subjects. Moreover, simultaneous supplementation with vitamins C and E could improve antioxidant capacity. However, the evidence currently available reflects more uncertainty than certainty, so further research and new clinical trials, perhaps with a better-defined methodology, are needed.

Effects of antioxidant supplementation on diabetes, metabolic syndrome, obesity and hypertension

Inclusion criteria and search strategy for meta-analysis

We considered any randomized clinical trial comparing vitamin C and/or vitamin E with placebo. Only original articles written in English were included. Exclusion criteria included intravenous vitamin administration, lack of placebo group (i.e. a control group receiving no pill or supplement), pregnant women or performed in children or adolescent. If vitamin supplementation included another component (i.e. other antioxidants, minerals or drugs), the study was only included if there was a group receiving only this component with which it was compared. In addition, we included several large studies assessing the risk of cancer development when vitamin C and/or E are supplemented. Note that the last section assessed the effects of vitamins on aging related parameters but due to the impossibility to normalize the data only a qualitative analysis was provided.

A systematic search of the literature was conducted on Medline, Scopus, CINAHL and Embase, current to October 2021. A combination of terms related to vitamin C, vitamin E, diabetes, metabolic syndrome, obesity, exercise performance, hypertension and cancer was used.

Table 1 Antioxidants and healthy aging.

Author	*Treatment*	*Duration*	*Dose (per day)*	*Age*	*Physiological parameter*	*Effect*	*Observations*
(Hemilä and Kaprio, 2009)	Vit E	5–8 y	50 mg	50–62	Mortality	↑[b]	Male smokers (high vitamin C intake)
				66–69		↓[b]	
(Hemilä and Kaprio, 2011)	Vit E	5–8 y	50 mg	>71	Mortality	↓[b]	Male smokers
(Lara-Padilla et al., 2007)	Vit C	1 wk	500 mg	19–23	ACP	↑[a]	
	Vit E		400 IU				
	Vit C + E		500 mg + 400 IU				
(Naeini et al., 2014)	Vit C + E	1 y	400 mg + 300 mg	60–75	TAC	↑[b]	Mild cognitive impairment patients
					GSH	↑[b]	
					8-OHdG	↓[a]	
(Cetin et al., 2010)	Vit E	6 mo	900 IU	60–85	TAC	=	
					Cognitive performance	=	
(Sacheck et al., 2003)	Vit E	12 wk	1000 IU	26,4 ± 3,3	Ceruloplasmin	↓[a]	
					iPF2α	↑	
					MDA	↓[a]	
					8-OHdG	↓[a]	
				71,1 ± 4,0	Ceruloplasmin	↓[a]	
					iPF2α	↓[a]	
					MDA	↑[a]	
					8-OHdG	↓	
(Huang et al., 2000)	Vit C	2 mo	500 mg	58 ± 14	8-OHdG	=	
	Vit E		400 IU			=	
	Vit C + E		500 mg + 400 IU			=	
(Meydani et al., 1990)	Vit E	1 mo	800 mg	>60	[α-tocopherol] in PBMC	↑[a]	
					Mitogen-induced lymphocite proliferation	↑[a]	
					IL-2 production	↑[a]	
					PG-2 production	↓[a]	
					Plasma lipid peroxides	↓[a]	
					DTH	↑[a]	
(Meydani, 1995)	Vit E	1 mo	800 IU	62–70	[α-tocoferol] in PBMC	↑[a]	
					Mitogen-induced lymphocite proliferation	↑[a]	
					IL-2 production	↑[a]	
					PG-2 production	↓[a]	
					DTH	↑	
(Pallast et al., 1999)	Vit E	6 mo	100 mg	65–80	DTH	↑	
					Positive reactions	=	
					Balance of Th1 to Th2 cytokines	=	
(Singh et al., 2002)	Vit C	6 wk	1 g	67 ± 1	Bradykinin-dependent vasodilatation	↑	
					Blood pressure	= (placebo ↓[a])	
(Paolisso et al., 1995)	Vit E	4 mo	900 mg	73,8 ± 2,1	Oxygen free radicals	↓[b]	Moderately obese patients
					Total cholesterol	↓[b]	
					LDL cholesterol	↓[b]	
					HDL cholesterol	↑[b]	
					Triglycerides	↓[b]	
(Guan et al., 2012)	Vit E	6 mo	400 mg	69,8 ± 5,1	8-iso-PGF2α	↓[b]	Alzheimer's disease patients
					Short telomeres (<4,4 kb)	↓	
(Sano et al., 2016)	Vit E	3 y	2000 IU	>50	Cognitive deterioration	=	Down syndrome patients
(Christen et al., 1999)	Vit C	12,5 y		40–84	Age-related maculopathy	=	Male
	Vit E					↓	
(Christen et al., 2010a)	Vit E	10 y	600 IU (alternate days)	>45	Age-related maculopathy	=	Female
(Christen et al., 2008)	Vit E	9,7 y	600 IU (alternate days)	>45	Age-related cataract	=	Female
(Christen et al., 2010b)	Vit C	8 y	500 mg	>50	Age-related cataract	=	Male
	Vit E		400 IU			=	

(Continued)

Table 1 Antioxidants and healthy aging.—cont'd

Author	*Treatment*	*Duration*	*Dose (per day)*	*Age*	*Physiological parameter*	*Effect*	*Observations*
(Christen et al., 2015)	Vit E	5,6 ± 1,2 y	400 IU	>50	Age-related cataract	=	Male
(Teikari et al., 1997)	Vit E	5-8 y	50 mg	50–69	Age-related cataract	=	Male smokers
(McNeil et al., 2004)	Vit E	4 y	500 IU	55–88	Age-related cataract	=	

ACP = antioxidant capacity of plasma; TAC = total antioxidant capacity; GSH = glutathione; MDA = malondialdehyde; PBMC = peripheral blood mononuclear cells; DTH = delayed-type hypersensitivity.

[a]Significant within the same group.

[b]Significant vs. placebo.

Data extraction and synthesis

The relevant variables extracted included HbA_{1c}, fasting blood glucose, body mass index, blood arterial pressure, maximal oxygen consumption, maximal strength, risk of cancer development as well as study length. The change in mean (ΔMean) and ΔSD was calculated for each treatment (i.e. placebo and vitamin) and outcome. When ΔSD was not reported we calculated it assuming a correlation coefficient of 0.7 as suggested by the Cochrane Handbook for Systematic Reviews of Interventions (Higgins and Green, 2011) as follows:

$$\Delta SD = \sqrt{\left(SD_{pre}{}^2 + SD_{post}{}^2 - 2 \times corr \times SD_{pre} \times SD_{post}\right)}.$$

The assumption of a 0.7 correlation coefficient was also recently validated for a similar paper, showing no substantial differences when varying it from 0.5 to 0.7 (Mason et al., 2021). Crossover trials followed the same treatment as parallel trials. When several doses of vitamin where studied the higher dose was used for calculations. Similarly, when several time-points where analyzed the last one was used.

Meta-analysis

A random-effects meta-analysis with DerSimonian-Laird methods was performed using R software on the change of each outcome (i.e. vitamin vs. placebo). Effects sizes are presented as standardized mean differences (SMD), mean differences (MD) or risk ratio (RR) with means ± SD and 95% CIs. Heterogeneity was assessed by χ^2 and I^2 and significance was set at $p < 0.05$. Additionally, subgroup analysis (i.e. vitamin E vs. vitamin C, or trained vs. untrained) and meta-regression (length of the study) were performed.

Fasting glucose, HbA_{1c} % and body mass index

We analyzed 35 clinical trials including more than 1700 type 2 diabetic and/or overweight to obese subjects (Economides et al., 2005; El-Aal et al., 2018; Paolisso et al., 1993; Ward et al., 2007; Ceriello et al., 1991; Reaven et al., 1995; Fuller et al., 1996; Gazis et al., 1999; Duffy et al., 2001; Botham et al., 2005; Ble-Castillo et al., 2005; Winterbone et al., 2007; Manning et al., 2013; Stonehouse et al., 2016; Rafraf et al., 2016; Farajbakhsh et al., 2019; Mazloom et al., 2011; Bishop et al., 1985; Ragheb et al., 2020; Kunsongkeit et al., 2019; Devanandan et al., 2020; Siavash and Amini, 2014; Sanguanwong et al., 2016; Tousoulis et al., 2007; Gillani et al., 2017; Dakhale et al., 2011; Mason et al., 2016; Paolisso et al., 1995; Gokce et al., 1999; Lu et al., 2005; Chen et al., 2006; Mason et al., 2019; Naylor et al., 1985). When both vitamin C and E are analyzed together there was a trend $p = 0.09$ toward decreasing fasting blood glucose in 1762 subjects from 32 studies (SMD -0.17 [95% CI −0.38%, 0.03%]) (**Fig. 2**). However, the analysis showed a high heterogeneity among studies $I^2 = 74.7\%$, $p < 0.01$. As this could be due to differences in the response to vitamin C and E, we next analyzed and compared both subgroups. We found that neither vitamin C ($n = 1114$, in 17 studies) nor vitamin E ($n = 648$, in 15 studies) decreased fasting blood glucose and no difference was detected between subgroups. It should, however, be noted that heterogeneity was higher among vitamin C studies ($I^2 = 82.1\%$) than in vitamin E studies ($I^2 = 40.6\%$). We also explored whether the length of the interventions could be a source of heterogeneity, however, this was not the case ($Q_m = 2.5$, $df = 1$, $p = 0.109$).

Regarding HbA_{1c}, we analyzed 1404 subjects from 23 studies (**Fig. 3**). The analysis shows that vitamin supplementation decreased HbA_{1c} in subjects with type 2 diabetes or overweight (SMD −0.33 [95% CI −0.58%, −0.09%]). Here, again, we observed a high heterogeneity among studies ($I^2 = 79.9\%$, $p < 0.01$). We therefore tested the hypothesis of subgroup differences, however, no differences were observed between vitamin C and vitamin E treatments ($p = 0.42$). Furthermore, and similar to the analysis of fasting blood glucose, we observed that an important source of heterogeneity comes from vitamin C studies ($I^2 = 80\%$, $p < 0.01$) rather than from vitamin E studies ($I^2 = 7.1\%$, $p = 0.2$). Notably, we further observed that the length of the intervention is also an important source or heterogeneity when addressing HbA_{1c} ($Q_M = 7.65$, $df = 1$, $p = 0.01$). In fact, under our analysis, our data suggest that each week of treatment would add 0.0198 to the estimate in favor to vitamin supplement. Note that this analysis is true when both vitamin E and vitamin C studies are analyzed together (**Fig. 3**).

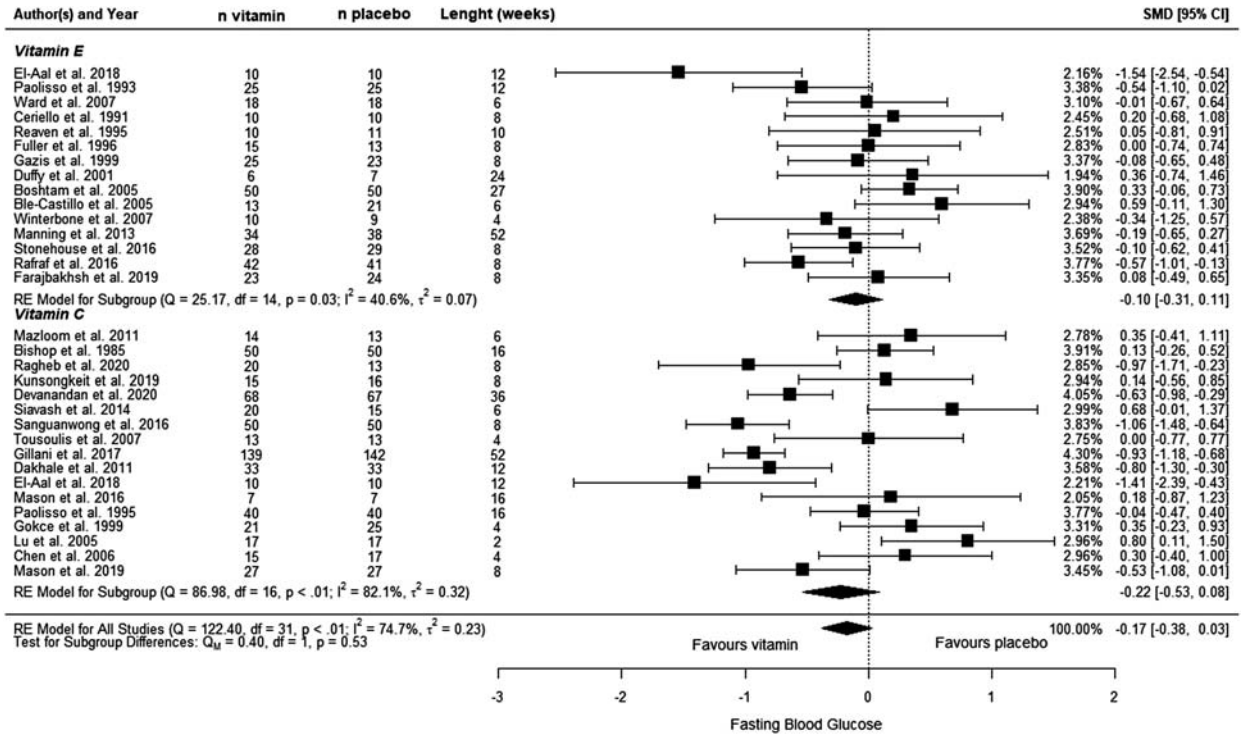


Fig. 2 Forest plot of the results from a random-effects meta-analysis shown as standardized mean difference with 95% CIs on fasting blood glucose. The rhombi represent the weighted vitamin E, vitamin C and total group's standardized mean difference.

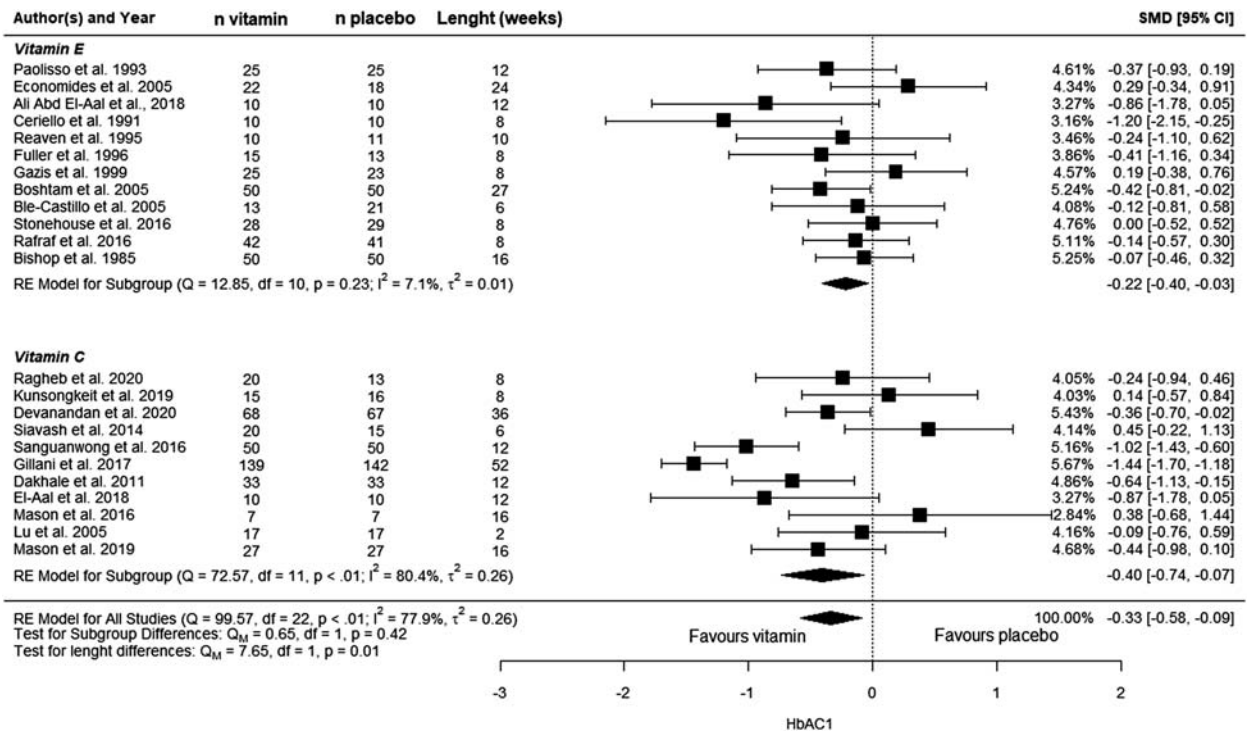


Fig. 3 Forest plot of the results from a random-effects meta-analysis shown as standardized mean difference with 95% CIs on % HbA_{1c}. The rhombi represent the weighted vitamin E, vitamin C and total group's standardized mean difference.

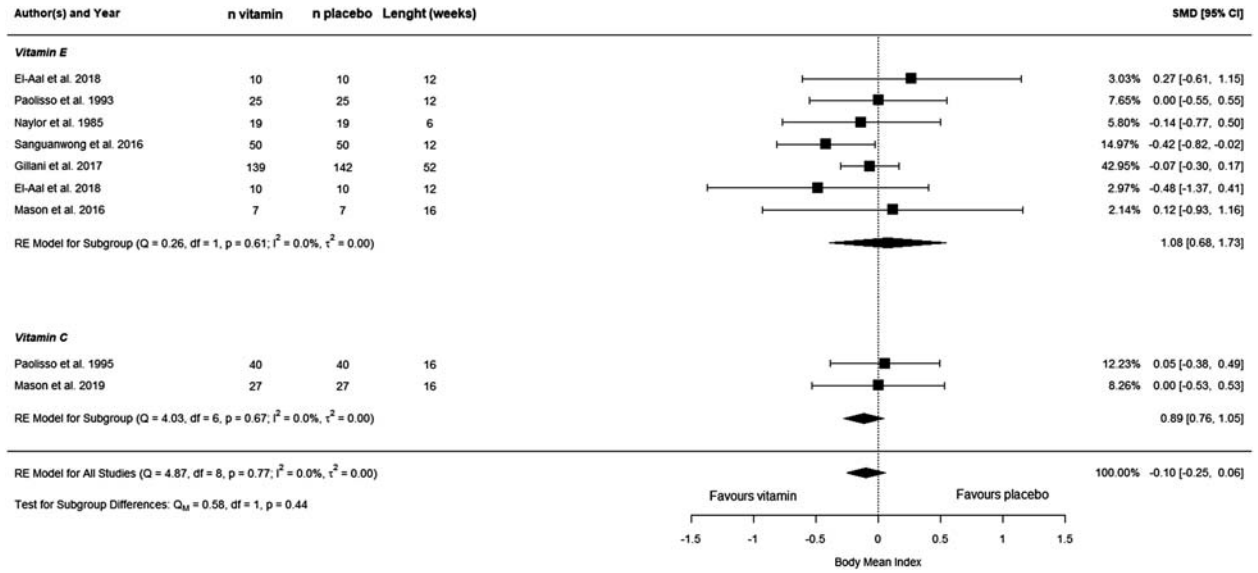


Fig. 4 Forest plot of the results from a random-effects meta-analysis shown as mean difference with 95% CIs on Body Mass Index$_c$. The rhombi represent the weighted vitamin E, vitamin C and total group's mean difference.

Surprisingly, only 9 studied (n = 657) body mass index as an outcome (Fig. 4). Therefore, although no effects of vitamin C and E has been observed in our analysis (SMD −0.10 [95% CI −0.25%, 0.06]) this result may be limited to the low number of studies included.

Blood pressure

We found that 30 clinical trials (n = 1555) studied the effects of vitamin administration on blood pressure (Figs. 5 and 6) (Rodrigo et al., 2014, 2008; Farvid et al., 2004, 2006; Huang et al., 2000; Palumbo et al., 2000; Chappell et al., 1999; Izadi et al., 2019; Farajbakhsh et al., 2019; Aryaeian et al., 2008; Studinger et al., 2004; Jessup et al., 2003; Rasool et al., 2003, 2006, 2008; Boshtam et al., 2002; Paolisso et al., 1995; Mason et al., 2019; Oliveira et al., 2015; Goldstein et al., 2012; Darko et al., 2002; Singh et al., 2002; Duffy et al. 2001; Fotherby et al., 2000; Gokce et al., 1999; Ghosh et al., 1994). Overall, it seems that systolic blood pressure (MD −5.0 [95% CI −8.49, −1.53]; $p < 0.001$) is greatly decreased by vitamin treatment. Which is also observed when vitamin C and E are consumed together ($p = 0.007$) and when vitamin C ($p = 0.006$) and vitamin E ($p = 0.03$) are consumed alone. However, there was a high heterogeneity among the studies analyzed (overall: $I^2 = 93\%$). Which can be mainly attributed to vitamin E trials ($I^2 = 96\%$) as no significant heterogeneity effect was observed in vitamin C trials. In fact, one particular vitamin E trial (Boshtam et al., 2002) which reported a huge hypotensive effect.

Regarding diastolic blood pressure (Fig. 6) we found a similar overall hypotensive effect induced by vitamin treatment (MD −2.06) [95% CI −3.29, −0.83; $p = 0.001$] in 30 clinical trials (n = 1537). However, when analyzing each supplementation on its own, no effects were observed for vitamin C ($p = 0.06$), vitamin E ($p = 0.05$) and vitamin E + C ($p = 0.05$). It is important to note that the length of the supplementation was not a source of heterogeneity neither for systolic nor for diastolic blood pressure as showed by the meta-regression model.

Effects of antioxidant supplementation on cancer development risk

We haven't found any clinical trial analyzing the effects of antioxidant supplementation on cancer patient. Instead, several trials have addressed the long term effect of antioxidant supplementation in healthy women (Lee et al., 2005), subjects at risk of cancer development due to is lifestyle (i.e. smokers) (Albanes et al., 1996; Varis et al., 1998; Rautalahti et al., 1999; Virtamo et al., 2000; Wright et al., 2007), subjects at risk of prostate cancer (high baseline prostate-specific antigen but normal rectal exploration) (Klein et al., 2011) and subjects at risk of cardiovascular disease (Lin et al., 2009). The outcome was cancer development during the follow-up (5–12 years).

We identified 8 trials which studied 116,882 subjects, and cancer development was registered during a period of 5–10 years where there were supplemented with vitamins or placebo (Fig. 7). Only one of these trials was performed with vitamin C, while 7 were performed with vitamin E. Risk ratio analysis showed that there was a trend ($p = 0.09$) toward a higher risk of cancer development in those subjects consuming vitamins (RR −0.07 [95% CI −0.15, 0.01]). Notably, this analysis showed a really low heterogeneity.

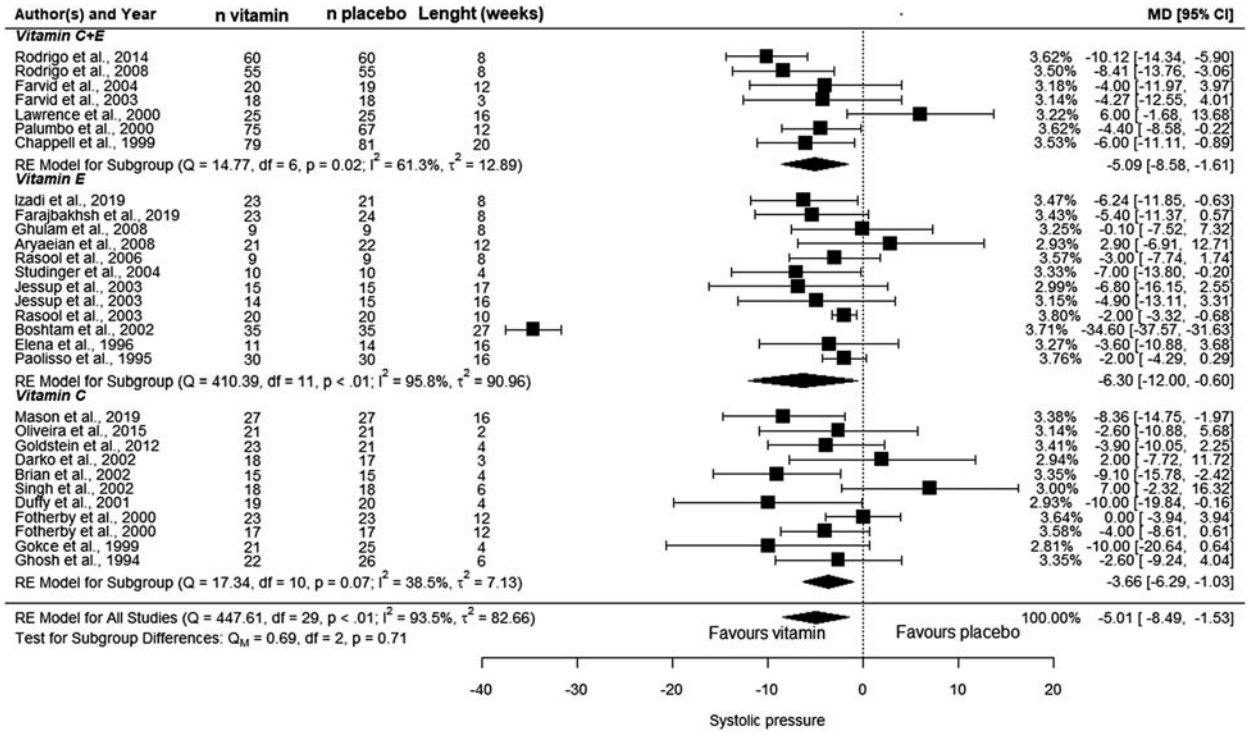


Fig. 5 Forest plot of the results from a random-effects meta-analysis shown as mean difference with 95% CIs on systolic blood pressure. The rhombi represent the weighted vitamin E, vitamin C and total group's mean difference.

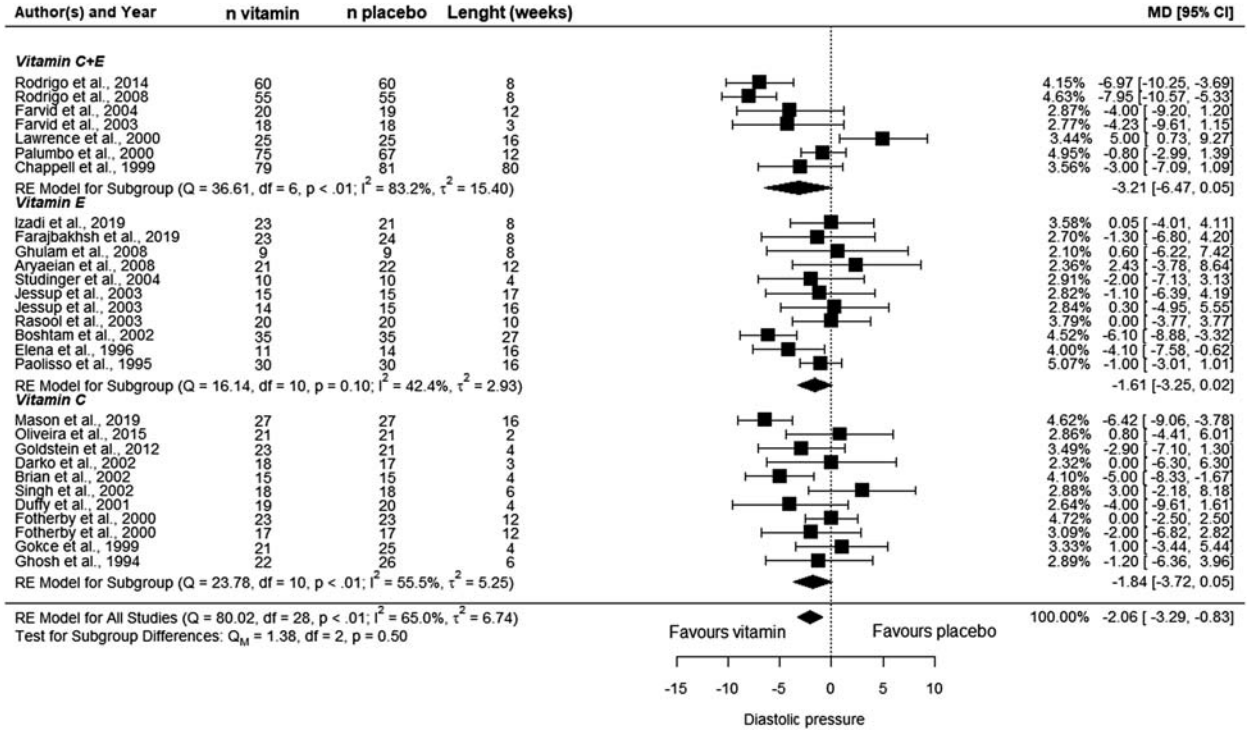


Fig. 6 Forest plot of the results from a random-effects meta-analysis shown as mean difference with 95% CIs on diastolic blood pressure. The rhombi represent the weighted vitamin E, vitamin C and total group's mean difference.

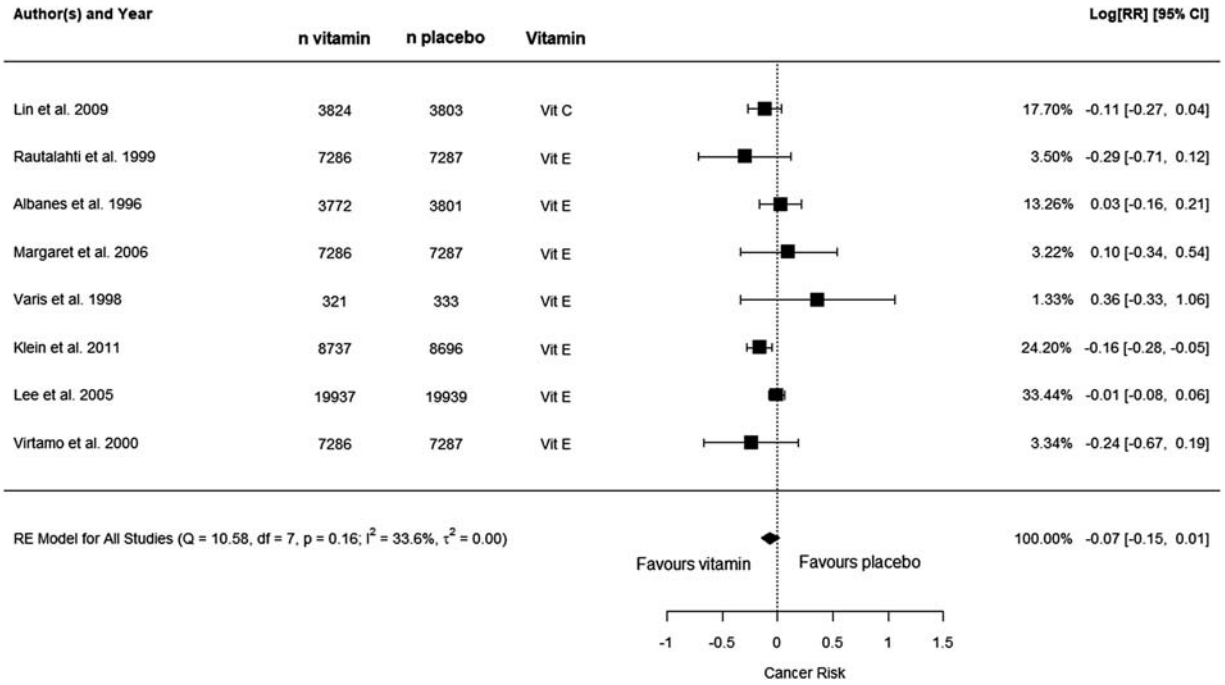


Fig. 7 Forest plot of the results from a random-effects meta-analysis shown as risk ratio with 95% Cis on cancer development. The rhombi represent the weighted vitamin E, vitamin C and total group's risk ratio.

Effects of antioxidant supplementation on exercise adaptations and physical performance

Exercise can improve cardiovascular health, glucose homeostasis as well as lower the risk of cancer development (Handschin and Spiegelman, 2008). Therefore, it is important to know whether antioxidant can prevent exercise-related adaptations.

We next analyzed whether vitamin supplementation could alter physical performance. We found that 14 trials analyzed maximal oxygen consumption during exercise (n = 341 subjects) (Gomez-Cabrera et al., 2008; Zychowska et al., 2021; Itoh et al., 2000; Roberts et al., 2011; Morrison et al., 2015; Jessup et al., 2003; Oberbach et al., 2010; Collins et al., 2003; Margaritelis et al., 2020; Cobley and Marrin, 2012; Paschalis et al., 2016; Gaeini et al., 2006; Patil et al., 2009). Of them, 5 trials were performed with vitamin C, 7 with vitamin E and 2 with both vitamin C and E (**Fig. 8**). Notably, 7 trials included concomitant training while the other 7 trials did not. Therefore, a two-subgroup analysis was performed including those with concomitant training vs. those with no training.

Overall, vitamin supplementation did not alter maximal oxygen consumption; but, similarly to previous analysis, there was a high heterogeneity among studies ($I^2 = 75\%$) (**Fig. 8**). Which was not due to the length of the intervention as suggested by the moderator test. However, the heterogeneity was higher among those studies which did not included concomitant training ($I^2 = 90\%$) than in those who included a trained period ($I^2 = 39\%$). It should be noted that although there was not a difference between subgroups, there was a trend ($p = 0.1$) toward a reduction in VO_{2max} when vitamins are administered during exercise training. In absolute terms, this analysis suggests that vitamin supplementation during exercise would prevent the gain of 0.82 mL/kg/min induced by training [95% CI −0.52, 2.17 mL/kg/min]. Finally, we found 5 studies engaging 6 strength outcomes (n = 142) (Barker et al., 2015; Beaton et al., 2002; Bjørnsen et al., 2016; Close et al., 2006; Dutra et al., 2018). The analysis suggest that vitamin supplementation does not alter strength ($p = 0.3$) (**Fig. 9**).

Conclusions

Here we show that vitamin C and E supplementation may reduce fasting blood glucose and HbA_{1c} levels in subjects metabolically compromised. However, and similar to a previous meta-analysis on the effects of vitamin C (Mason et al., 2021) this reduction is unlikely to have a clinically meaningful effect. In contrast, previous data reported that vitamin c supplementation shows a hypotensive effect (Mason et al., 2021). Here, we show that this effect can be extrapolated to vitamin E and vitamin C + E supplementation. It should however be noted that we found a large heterogeneity among all the studies and most of the variables. Therefore, future work should further investigate the source of this heterogeneity as antioxidant supplementation may not be save for all people in all conditions.

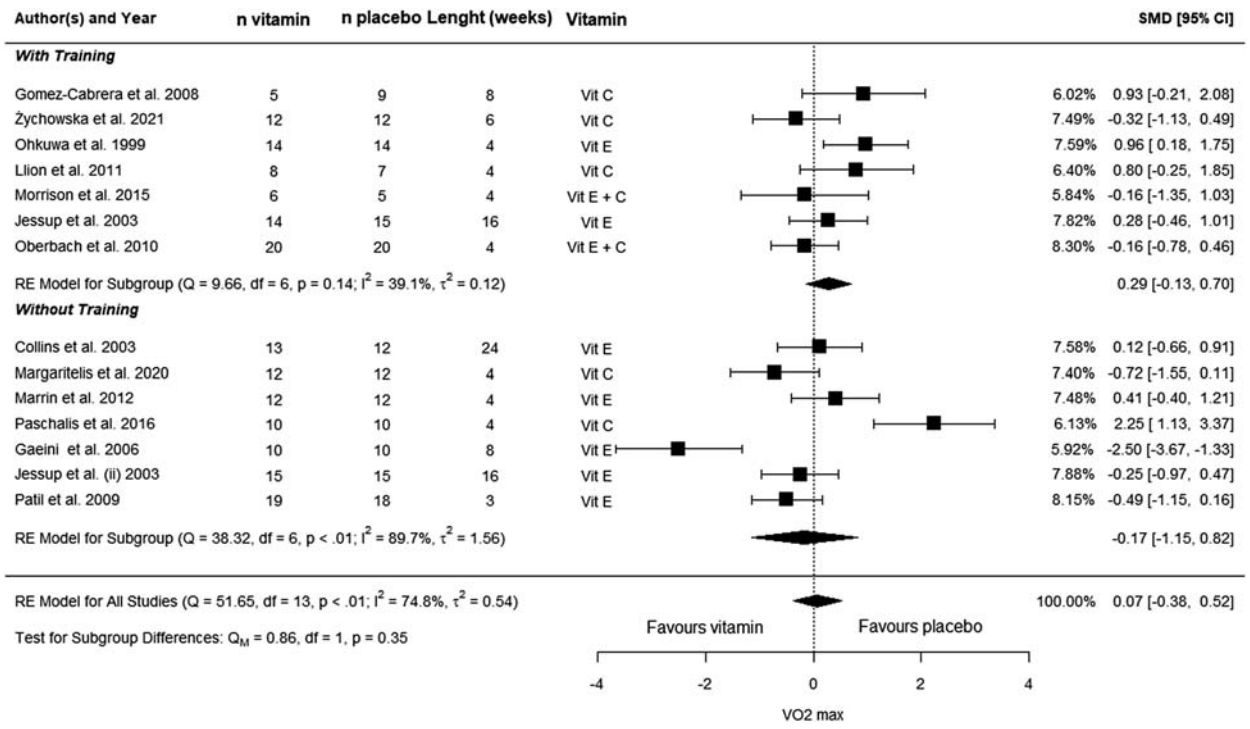


Fig. 8 Forest plot of the results from a random-effects meta-analysis shown as standardized mean difference with 95% CIs on maximal oxygen uptake (VO_{2max}). The rhombi represent the weighted vitamin E, vitamin C and total group's standardized mean difference.

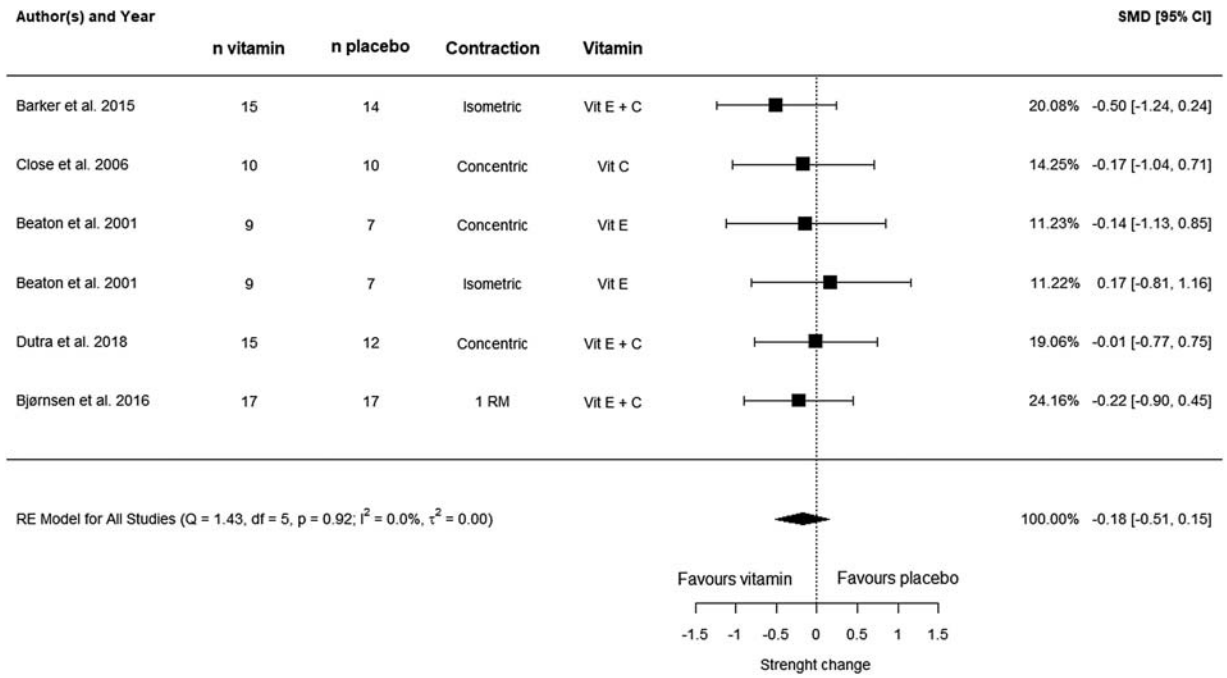


Fig. 9 Forest plot of the results from a random-effects meta-analysis shown as standardized mean difference with 95% CIs on maximal strength. The rhombi represent the weighted vitamin E, vitamin C and total group's standardized mean difference.

For instance, we found that vitamin supplementation concomitant to exercise training could prevent the gain of ~1 mL/kg/min induced by training. More importantly, an analysis of large cohort studies revealed a trend toward higher cancer risk among subjects supplemented with antioxidants. Therefore, antioxidant supplementation must be avoided in subjects at risk to develop cancer.

The main findings of these studies are that, in certain situations, loading the cell with high doses of antioxidants leads to a blunting of the positive effects of a balanced diet and/or exercise training and interferes with important ROS-mediated physiological processes, such as vasodilation and insulin signaling. Moreover, an analysis of large cohort studies revealed a trend toward higher cancer risk among; however, samples from these studies were heterogeneous, including healthy subjects, and people at risk of develop cancer (smokers) and at cardiovascular risk. Therefore, more research is needed to produce evidence-based guidelines regarding the use of antioxidant supplementation. We recommend that an adequate intake of antioxidants through a varied and balanced diet remains the best approach to maintain the optimal antioxidant status in individuals.

References

Albanes, D., Heinonen, O.P., Taylor, P.R., et al., 1996. α-tocopherol and β-carotene supplements and lung cancer incidence in the alpha-tocopherol, beta-carotene cancer prevention study: effects of base- line characteristics and study compliance. J. Natl. Cancer Inst. 88, 1560–1570.

Aryaeian, N., Shahram, F., Djalali, M., et al., 2008. Effect of conjugated linoleic acid, vitamin E and their combination on lipid profiles and blood pressure of Iranian adults with active rheumatoid arthritis. Vasc. Health Risk Manag. 4, 1423–1432.

Barker, T., Henriksen, V.T., Rogers, V.E., Trawick, R.H., 2015. Serum cytokines and muscle strength after anterior cruciate ligament surgery are not modulated by high-doses of vitamins E (α- and γ-tocopherol's) and C. Cytokine 74, 279–286.

Beaton, L.J., Allan, D.A., Tarnopolsky, M.A., et al., 2002. Contraction-induced muscle damage is unaffected by vitamin E supplementation. Med. Sci. Sports Exerc. 34, 798–805.

Bishop, N., Schorah, C.J., Wales, J.K., 1985. The effect of vitamin C supplementation on diabetic hyperlipidaemia: a double blind, crossover study. Diabet. Med. 2 (2), 121–124.

Bjørnsen, T., Salvesen, S., Berntsen, S., et al., 2016. Vitamin C and E supplementation blunts increases in total lean body mass in elderly men after strength training. Scand. J. Med. Sci. Sports 26, 755–763.

Ble-Castillo, J.L., Carmona-Díaz, E., Méndez, J.D., et al., 2005. Effect of α-tocopherol on the metabolic control and oxidative stress in female type 2 diabetics. Biomed. Pharmacother. 59, 290–295.

Boshtam, M., Rafiei, M., Sadeghi, K., Sarraf-Zadegan, N., 2002. Vitamin E can reduce blood pressure in mild hypertensives. Int. J. Vitam. Nutr. Res. 72, 309–314.

Botham, K.M., Bravo, E., Elliott, J., Wheeler-Jones, C.P.D., 2005. Direct interaction of dietary lipids carried in chylomicron remnants with cells of the artery wall: implications for atherosclerosis development. Curr. Pharmaceut. Des. 11, 3681–3695.

Ceriello, A., Giugliano, D., Quatraro, A., et al., 1991. Vitamin E reduction of protein glycosylation in diabetes. New prospect for prevention of diabetic complications? Diabetes Care 14, 68–72.

Cetin, E., Top, E.C., Şahin, G., Özkaya, Y.G., Aydin, H., Toraman, F., 2010. Effect of vitamin E supplementation with exercise on cognitive functions and total antioxidant capacity in older people. J. Nutr. Health Aging 14 (9), 763–769.

Chappell, L.C., Seed, P.T., Briley, A.L., 1999. Effect of antioxidants on the occurrence of pre-eclampsia in women at increased risk: a randomised trial. Nutr. Clin. Pract. 15, 112.

Chen, H., Karne, R.J., Hall, G., et al., 2006. High-dose oral vitamin C partially replenishes vitamin C levels in patients with type 2 diabetes and low vitamin C levels but does not improve endothelial dysfunction or insulin resistance. Am. J. Physiol. Heart Circ. 290, H137–H145.

Christen, W.G., Ajani, U.A., Glynn, R.J., Manson, J.E., Schaumberg, D.A., Chew, E.C., et al., 1999. Prospective cohort study of antioxidant vitamin supplement use and the risk of age-related maculopathy. Am. J. Epidemiol. 149 (5), 476–484.

Christen, W.G., Glynn, R.J., Chew, E.Y., Buring, J.E., 2008. Vitamin E and age-related cataract in a randomized trial of women. Ophthalmology 115 (5), 822–829.

Christen, W.G., Glynn, R.J., Chew, E.Y., Buring, J.E., 2010a. Vitamin E and age-related macular degeneration in a randomized trial of women. Ophthalmology 117 (6), 1163–1168.

Christen, W.G., Glynn, R.J., Sesso, H.D., Kurth, T., MacFadyen, J., Bubes, V., et al., 2010b. Age-related cataract in a randomized trial of vitamins E and C in men. Arch. Ophthalmol. 128 (11), 1397–1405.

Christen, W.G., Glynn, R.J., Gaziano, J.M., Darke, A.K., Crowley, J.J., Goodman, P.J., et al., 2015. Age-related cataract in men in the selenium and vitamin E cancer prevention trial eye endpoints study: a randomized clinical trial. JAMA ophthalmol. 133 (1), 17–24.

Close, G.L., Ashton, T., Cable, T., et al., 2006. Ascorbic acid supplementation does not attenuate post-exercise muscle soreness following muscle-damaging exercise but may delay the recovery process. Br. J. Nutr. 95, 976–981.

Cobley, J.N., Marrin, K., 2012. Vitamin E supplementation does not alter physiological performance at fixed blood lactate concentrations in trained runners. J. Sports Med. Phys. Fit. 52, 63–70.

Collins, E.G., Edwin Langbein, W., Orebaugh, C., et al., 2003. Pole Striding exercise and vitamin E for management of peripheral vascular disease. Med. Sci. Sports Exerc. 35, 384–393.

Dakhale, G.N., Chaudhari, H.V., Shrivastava, M., 2011. Supplementation of vitamin C reduces blood glucose and improves glycosylated haemoglobin in type 2 diabetes mellitus: a randomized, double-blind study. Adv. Pharmacol. Sci. 2011, 195271.

Darko, D., Dornhorst, A., Kelly, F.J., et al., 2002. Lack of effect of oral vitamin C on blood pressure, oxidative stress and endothelial function in Type II diabetes. Clin. Sci. 103, 339–344.

Devanandan, P., Chowdary, R.P., Muthukumar, V.A., 2020. Effects of vitamin C supplementation on the glycemic control and cardiovascular risk in type II diabetes mellitus. J. Res. Pharm. 24, 182–187.

Duffy, S.J., Gokce, N., Holbrook, M., Hunter, L.M., Biegelsen, E.S., Huang, A., Keaney Jr., J.F., Vita, J.A., 2001. Effect of ascorbic acid treatment on conduit vessel endothelial dysfunction in patients with hypertension. Am. J. Physiol. Heart Circ. 280 (2), H528–H534.

Dutra, M.T., Alex, S., Mota, M.R., et al., 2018. Effect of strength training combined with antioxidant supplementation on muscular performance. Appl. Physiol. Nutr. Metab. 43, 775–781.

Economides, P.A., Khaodhiar, L., Caselli, A., et al., 2005. The effect of vitamin E on endothelial function of micro- and macrocirculation and left ventricular function in type 1 and type 2 diabetic patients. Diabetes 54, 204–211.

El-Aal, A.A., El-Ghffar, E.A.A., Ghali, A.A., et al., 2018. The effect of vitamin C and/or E supplementations on type 2 diabetic adult males under metformin treatment: a single-blinded randomized controlled clinical trial. Diabetol. Metab. Syndrome 12, 483–489.

Farajbakhsh, A., Mazloomi, S.M., Mazidi, M., et al., 2019. Sesame oil and vitamin E co-administration may improve cardiometabolic risk factors in patients with metabolic syndrome: a randomized clinical trial. Eur. J. Clin. Nutr. 73, 1403–1411.

Farvid, M.S., Jalali, M., Siassi, F., et al., 2004. The impact of vitamins and/or mineral supplementation on blood pressure in type 2 diabetes. J. Am. Nutr. Assoc. 23, 272–279.

Farvid, M.S., Jalali, M., Siassi, F., Hosseini, M., 2006. Comparison of the effects of vitamins and/or mineral supplementation on glomerular and tubular dysfunction in type 2 diabetes: response to Rossing et al. Diabetes Care 29, 748–749.

Fotherby, M.D., Williams, J.C., Forster, L.A., et al., 2000. Effect of vitamin C on ambulatory blood pressure and plasma lipids in older persons. J. Hypertens. 18, 411–415.
Fuller, C.J., Chandalia, M., Garg, A., Grundy, S.M., Jialal, I., 1996. RRR-alpha-tocopheryl acetate supplementation at pharmacologic doses decreases low-density-lipoprotein oxidative susceptibility but not protein glycation in patients with diabetes mellitus. Am. J. Clin. Nutr. 63 (5), 753–759.
Gaeini, A.A., Rahnama, N., Hamedinia, M.R., 2006. Effects of vitamin E supplementation on oxidative stress at rest and after exercise to exhaustion in athletic students. J. Sports Med. Phys. Fit. 46, 458–461.
Gazis, A., White, D.J., Page, S.R., Cockcroft, J.R., 1999. Effect of oral vitamin E (alpha-tocopherol) supplementation on vascular endothelial function in Type 2 diabetes mellitus. Diabet. Med. 16, 304–311.
Ghosh, S.K., Ekpo, E.B., Shah, I.U., et al., 1994. A double-blind, placebo-controlled parallel trial of vitamin C treatment in elderly patients with hypertension. Gerontology 40, 268–272.
Gillani, S.W., Sulaiman, S.A.S., Abdul, M.I.M., Baig, M.R., 2017. Combined effect of metformin with ascorbic acid versus acetyl salicylic acid on diabetes related cardiovascular complication; a 12-month single blind multicenter randomized control trial. Cardiovasc. Diabetol. 16, 103.
Gokce, N., Keaney, J.F., Frei, B., et al., 1999. Long-term ascorbic acid administration reverses endothelial vasomotor dysfunction in patients with coronary artery disease. Circulation 99, 3234–3240.
Goldstein, B.H., Sandelin, A.M., Golbus, J.R., et al., 2012. Impact of vitamin C on endothelial function and exercise capacity in patients with a Fontan circulation. Congenit. Heart Dis. 7, 226–234.
Gomez-Cabrera, M.C., Domenech, E., Romagnoli, M., et al., 2008. Oral administration of vitamin C decreases muscle mitochondrial biogenesis and hampers training-induced adaptations in endurance performance. Am. J. Clin. Nutr. 87, 142–149.
Guan, J.Z., Guan, W.P., Maeda, T., Makino, N., 2012. Effect of vitamin E administration on the elevated oxygen stress and the telomeric and subtelomeric status in Alzheimer's disease. Gerontology 58 (1), 62–69.
Handschin, C., Spiegelman, B.M., 2008. The role of exercise and PGC1alpha in inflammation and chronic disease. Nature 454 (7203), 463–469.
Hemilä, H., Kaprio, J., 2009. Modification of the effect of vitamin E supplementation on the mortality of male smokers by age and dietary vitamin C. Am. J. Epidemiol. 169 (8), 946–953.
Hemilä, H., Kaprio, J., 2011. Vitamin E may affect the life expectancy of men, depending on dietary vitamin C intake and smoking. Age Ageing 40 (2), 215–220.
Higgins, J.P.T., Green, S., 2011. Cochrane Handbook for Systematic Reviews of Interventions. The Cochrane Collaboration.
Huang, H.Y., Helzlsouer, K.J., Appel, L.J., 2000. The effects of vitamin C and vitamin E on oxidative DNA damage: results from a randomized controlled trial. Cancer Epidemiol. Prev. Biomark. 9 (7), 647–652.
Itoh, H., Ohkuwa, T., Yamazaki, Y., et al., 2000. Vitamin E supplementation attenuates leakage of enzymes following 6 successive days of running training. Int. J. Sports Med. 21, 369–374.
Izadi, A., Shirazi, S., Taghizadeh, S., Gargari, B.P., 2019. Independent and additive effects of coenzyme Q10 and vitamin E on cardiometabolic outcomes and visceral adiposity in women with polycystic ovary syndrome. Arch. Med. Res. 50, 1–10.
Jessup, J.V., Horne, C., Yarandi, H., Quindry, J., 2003. The effects of endurance exercise and vitamin E on oxidative stress in the elderly. Biol. Res. Nurs. 5, 47–55.
Klein, E.A., Thompson, I.M., Tangen, C.M., et al., 2011. Vitamin E and the risk of prostate cancer: the selenium and vitamin E cancer prevention trial (SELECT). J. Am. Med. Assoc. 306, 1549–1556.
Kunsongkeit, P., Okuma, N., Rassameemasmaung, S., Chaivanit, P., 2019. Effect of vitamin C as an adjunct in nonsurgical periodontal therapy in uncontrolled type 2 diabetes mellitus patients. Eur. J. Dent. 13, 444–449.
Lara-Padilla, E., Kormanovski, A., Grave, P.A., Olivares-Corichi, I.M., Santillan, R.M., Hicks, J.J., 2007. Increased antioxidant capacity in healthy volunteers taking a mixture of oral antioxidants versus vitamin C or E supplementation. Adv. Ther. 24 (1), 50–59.
Lee, I.-M., Cook, N.R., Gaziano, J.M., et al., 2005. Vitamin E in the primary prevention of cardiovascular disease and cancer. JAMA 294, 56.
Lin, J., Cook, N.R., Albert, C., et al., 2009. Vitamins C and E and beta carotene supplementation and cancer risk: a randomized controlled trial. J. Natl. Cancer Inst. 101, 14–23.
Lu, Q., Bjorkhem, I., Wretlind, B., Diczfalusy, U., Henriksson, P., Freyschuss, A., 2005. Effect of ascorbic acid on microcirculation in patients with type II diabetes: a randomized placebo-controlled cross-over study. Clin. Sci. 108, 507–513.
Manning, P.J., Sutherland, W.H., Williams, S.M., Walker, R.J., Berry, E.A., De Jong, S.A., Ryalls, A.R., 2013. The effect of lipoic acid and vitamin E therapies in individuals with the metabolic syndrome. Nutr. Metab. Cardiovasc. Dis. 23 (6), 543–549.
Margaritelis, N.V., Paschalis, V., Theodorou, A.A., et al., 2020. Antioxidant supplementation, redox deficiencies and exercise performance: a falsification design. Free Radical Biol. Med. 158, 44–52.
Mason, S.A., Della Gatta, P.A., Snow, R.J., Russell, A.P., Wadley, G.D., 2016. Ascorbic acid supplementation improves skeletal muscle oxidative stress and insulin sensitivity in people with type 2 diabetes: findings of a randomized controlled study. Free Radical Biol. Med. 93, 227–238.
Mason, S.A., Rasmussen, B., van Loon, L.J.C., Salmon, J., Wadley, G.D., 2019. Ascorbic acid supplementation improves postprandial glycaemic control and blood pressure in individuals with type 2 diabetes: findings of a randomized cross-over trial. Diabetes Obes. Metab. 21, 674–682.
Mason, S.A., Keske, M.A., Wadley, G.D., 2021. Effects of vitamin C supplementation on glycemic control and cardiovascular risk factors in people with type 2 diabetes: a GRADE-assessed systematic review and meta-analysis of randomized controlled trials. Diabetes Care 44 (2), 618–630.
Mazloom, Z., Hejazi, N., Dabbaghmanesh, M.H., Tabatabaei, H.R., Ahmadi, A., Ansar, H., 2011. Effect of vitamin C supplementation on postprandial oxidative stress and lipid profile in type 2 diabetic patients. Pakistan J. Biol. Sci. 14, 900–904.
McNeil, J.J., Robman, L., Tikellis, G., Sinclair, M.I., McCarty, C.A., Taylor, H.R., 2004. Vitamin E supplementation and cataract: randomized controlled trial. Ophthalmology 111 (1), 75–84.
Meydani, S.N., 1995. Vitamin E enhancement of T cell-mediated function in healthy elderly: mechanisms of action. Nutr. Rev. 53 (4), S52–S58.
Meydani, S.N., Barklund, M.P., Liu, S., Meydani, M., Miller, R.A., Cannon, J.G., et al., 1990. Vitamin E supplementation enhances cell-mediated immunity in healthy elderly subjects. Am. J. Clin. Nutr. 52 (3), 557–563.
Morrison, D., Hughes, J., Della Gatta, P.A., et al., 2015. Vitamin C and e supplementation prevents some of the cellular adaptations to endurance-training in humans. Free Radical Biol. Med. 89, 852–862.
Naeini, A.A., Elmadfa, I., Djazayery, A., Barekatain, M., Ghazvini, M.A., Djalali, M., Feizi, A., 2014. The effect of antioxidant vitamins E and C on cognitive performance of the elderly with mild cognitive impairment in Isfahan, Iran: a double-blind, randomized, placebo-controlled trial. Eur. J. Nutr. 53 (5), 1255–1262.
Naylor, G.J., Grant, L., Smith, C., 1985. A double blind placebo controlled trial of ascorbic acid in obesity. Nutr. Health 4, 25–28.
Oberbach, A., Kirsch, K., Lehmann, S., et al., 2010. Serum vaspin concentrations are decreased after exercise-induced oxidative stress. Obes. Facts 3, 328–331.
Oliveira, IJL de, de Souza, V.V., Motta, V., Da-Silva, S.L., 2015. Effects of oral vitamin C supplementation on anxiety in students: a double-blind, randomized, placebo-controlled trial. Pakistan J. Biol. Sci. 18, 11–18.
Pallast, E.G., Schouten, E.G., De Waart, F.G., Fonk, H.C., Doekes, G., von Blomberg, B.M., Kok, F.J., 1999. Effect of 50-and 100-mg vitamin E supplements on cellular immune function in noninstitutionalized elderly persons. Am. J. Clin. Nutr. 69 (6), 1273–1281.
Palumbo, G., Avanzini, F., Alli, C., et al., 2000. Effects of vitamin E on clinic and ambulatory blood pressure in treated hypertensive patients. Am. J. Hypertens. 13, 564–567.
Paolisso, G., D'Amore, A., Giugliano, D., Ceriello, A., Varricchio, M., D'Onofrio, F., 1993. Pharmacologic doses of vitamin E improve insulin action in healthy subjects and non-insulin-dependent diabetic patients. Am. J. Clin. Nutr. 57 (5), 650–656.
Paolisso, G., Gambardella, A., Giugliano, D., Galzerano, D., Amato, L., Volpe, C., et al., 1995. Chronic intake of pharmacological doses of vitamin E might be useful in the therapy of elderly patients with coronary heart disease. Am. J. Clin. Nutr. 61 (4), 848–852.

Paschalis, V., Theodorou, A.A., Kyparos, A., et al., 2016. Low vitamin C values are linked with decreased physical performance and increased oxidative stress: reversal by vitamin C supplementation. Eur. J. Nutr. 55, 45–53.
Patil, S.M., Chaudhuri, D., Dhanakshirur, G.B., 2009. Role of alpha-tocopherol in cardiopulmonary fitness in endurance athletes, cyclists. Indian J. Physiol. Pharmacol. 53, 375–379.
Rafraf, M., Bazyun, B., Sarabchian, M.A., et al., 2016. Vitamin E improves serum paraoxonase-1 activity and some metabolic factors in patients with type 2 diabetes: no effects on nitrite/nitrate levels. J. Am. Nutr. Assoc. 35, 521–528.
Ragheb, S.R., El Wakeel, L.M., Nasr, M.S., Sabri, N.A., 2020. Impact of rutin and vitamin C combination on oxidative stress and glycemic control in patients with type 2 diabetes. Clin. Nutr. ESPEN 35, 128–135.
Rasool, A.H.G., Rehman, A., Wan Yusuf, W.N., Rahman, A.R.A., 2003. Vitamin E and its effect on arterial stiffness in postmenopausal women—a randomized controlled trial. Int. J. Clin. Pharmacol. Therapeut. 41, 587–592.
Rasool, A.H.G., Yuen, K.H., Yusoff, K., et al., 2006. Dose dependent elevation of plasma tocotrienol levels and its effect on arterial compliance, plasma total antioxidant status, and lipid profile in healthy humans supplemented with tocotrienol rich vitamin. J. Nutr. Sci. Vitaminol. 52, 473–478.
Rasool, A.H.G., Rahman, A.R.A., Yuen, K.H., Wong, A.R., 2008. Arterial compliance and vitamin E blood levels with a self emulsifying preparation of tocotrienol rich vitamin E. Arch. Pharm. Res. 31, 1212–1217.
Rautalahti, M.T., Virtamo, J.R.K., Taylor, P.R., et al., 1999. The effects of supplementation with α-tocopherol and β-carotene on the incidence and mortality of carcinoma of the pancreas in a randomized, controlled trial. Cancer 86, 37–42.
Reaven, P.D., Herold, D.A., Barnett, J., Edelman, S., 1995. Effects of Vitamin E on susceptibility of low-density lipoprotein and low-density lipoprotein subfractions to oxidation and on protein glycation in NIDDM. Diabetes Care 18, 807–816.
Roberts, L.A., Beattie, K., Close, G.L., Morton, J.P., 2011. Vitamin C consumption does not impair training-induced improvements in exercise performance. Int. J. Sports Physiol. Perform. 6, 58–69.
Rodrigo, R., Prat, H., Passalacqua, W., et al., 2008. Decrease in oxidative stress through supplementation of vitamins C and E is associated with a reduction in blood pressure in patients with essential hypertension. Clin. Sci. 114, 625–634.
Rodrigo, R., Miranda-Merchak, A., Grau, R.V., et al., 2014. Modulation of (Na,K)-ATPase activity by membrane fatty acid composition: therapeutic implications in human hypertension. Clin. Exp. Hypertens. 36, 17–26.
Sacheck, J.M., Milbury, P.E., Cannon, J.G., Roubenoff, R., Blumberg, J.B., 2003. Effect of vitamin E and eccentric exercise on selected biomarkers of oxidative stress in young and elderly men. Free Radic. Biol. Med. 34 (12), 1575–1588.
Sanguanwong, S., Tangvarasittichai, O., Sengsuk, C., Tangvarasittichai, S., 2016. Oral supplementation of vitamin C reduced lipid peroxidation and insulin resistance in patients with type 2 diabetes mellitus. Int. J. Pharmacol. Toxicol. 8, 114–119.
Sano, M., Aisen, P.S., Andrews, H.F., Tsai, W.Y., Lai, F., Dalton, A.J., 2016. Vitamin E in aging persons with down syndrome: a randomized, placebo-controlled clinical trial. Neurology 86 (22), 2071–2076.
Siavash, M., Amini, M., 2014. Vitamin C may have similar beneficial effects to gemfibrozil on serum high-density lipoprotein-cholesterol in type 2 diabetic patients. J. Pharm. Pract. Res. 3, 77–82.
Singh, N., Graves, J., Taylor, P.D., MacAllister, R.J., Singer, D.R., 2002. Effects of a "healthy" diet and of acute and long-term vitamin C on vascular function in healthy older subjects. Cardiovasc. Res. 56 (1), 118–125.
Stonehouse, W., Brinkworth, G.D., Thomson, C.H., Abeywardena, M.Y., 2016. Short term effects of palm-tocotrienol and palm-carotenes on vascular function and cardiovascular disease risk: a randomised controlled trial. Atherosclerosis 254, 205–214.
Studinger, P., Mersich, B., Lénárd, Z., et al., 2004. Effect of vitamin E on carotid artery elasticity and baroreflex gain in young, healthy adults. Auton. Neurosci. 113, 63–70.
Teikari, J.M., Virtamo, J., Rautalahti, M., Palmgren, J., Liesto, K., Heinonen, O.P., 1997. Long-term supplementation with alpha-tocopherol and beta-carotene and age-related cataract. Acta Ophthalmol. Scand. 75 (6), 634–640.
Tousoulis, D., Antoniades, C., Vasiliadou, C., et al., 2007. Effects of atorvastatin and vitamin C on forearm hyperaemic blood flow, asymmetrical dimethylarginine levels and the inflammatory process in patients with type 2 diabetes mellitus. Heart 93, 244–246.
Varis, K., Taylor, P.R., Sipponen, P., et al., 1998. Gastric cancer and premalignant lesions in atrophic gastritis: a controlled trial on the effect of supplementation with alpha-tocopherol and beta-carotene. Scand. J. Gastroenterol. 33, 294–300.
Virtamo, J., Edwards, B.K., Virtanen, M., et al., 2000. Effects of supplemental alpha-tocopherol and beta-carotene on urinary tract cancer: incidence and mortality in a controlled trial (Finland). Cancer Causes Control 11, 933–939.
Ward, N.C., Wu, J.H., Clarke, M.W., Puddey, I.B., Burke, V., Croft, K.D., Hodgson, J.M., 2007. The effect of vitamin E on blood pressure in individuals with type 2 diabetes: a randomized, double-blind, placebo-controlled trial. J. Hypertens. 25 (1), 227–234.
Winterbone, M.S., Sampson, M.J., Saha, S., et al., 2007. Pro-oxidant effect of alpha-tocopherol in patients with type 2 diabetes after an oral glucose tolerance test–a randomised controlled trial. Cardiovasc. Diabetol. 6, 8.
Wright, M.E., Virtamo, J., Hartman, A.M., et al., 2007. Effects of α-tocopherol and β-carotene supplementation on upper aerodigestive tract cancers in a large, randomized controlled trial. Cancer 109, 891–898.
Żychowska, M., Grzybkowska, A., Zasada, M., et al., 2021. Effect of six weeks 1000 mg/day vitamin C supplementation and healthy training in elderly women on genes expression associated with the immune response—a randomized controlled trial. Sports Nutr. Rev. J. 18, 1–10.

Further reading

Mullan, B.A., Young, I.S., Fee, H., McCance, D.R., 2002. Ascorbic acid reduces blood pressure and arterial stiffness in type 2 diabetes. Hypertension 40, 804–809.
Title, L.M., Cummings, P.M., Giddens, K., et al., 2000. Effect of folic acid and antioxidant vitamins on endothelial dysfunction in patients with coronary artery disease. J. Am. Coll. Cardiol. 36, 758–765.

Antioxidants: Physiology and dietary sources

Jesús R. Huertas[a], S. Stanner[b], and E. Weichselbaum[b], [a] Department of Physiology, Institute of Nutrition and Food Technology, Biomedical Research Center, University of Granada, Granada, Spain; and [b] British Nutrition Foundation, London, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of S. Stanner, E. Weichselbaum, Antioxidants, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 88–99, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00013-1.

Introduction 62
Definition of antioxidant 63
Antioxidants sources 63
Normal plasma concentration of the most common antioxidants 63
Antioxidants as signaling 64
New physiological antioxidants 64
The antioxidant hypothesis 66
Cardiovascular disease 67
Primary prevention 67
Secondary prevention 67
Cancer 69
β-carotene 69
Vitamin C 71
Vitamin E 71
Selenium 72
Other diseases associated with oxidative damage 72
Type 2 diabetes 72
Asthma and chronic obstructive pulmonary disease (COPD) 73
Macular degeneration and cataracts 73
Immune function—common cold 74
Possible explanations for the disagreement between the findings of observational studies and clinical trials 74
Conclusion 74
References 75
Further reading 75

Key points

- Free radicals, including reactive oxygen and nitrogen species (ROS/NOS), are formed in the human body as a result of oxidative metabolism.
- Free radicals are capable of modifying important molecules such as DNA, lipids, and proteins, and therefore affecting their ability to function or causing them to function abnormally.
- The body's antioxidant system is complex and prevent formation of ROS and terminate chains of ROS-initiated peroxidation of biological substrates.
- Antioxidants present in the diet include vitamin C, vitamin E, carotenoids, polyphenols, chlorophylls, metals and minerals such as zinc or selenium, etc.
- Diets rich in plant foods (particularly fruit and vegetables) convey health benefits, as do high plasma levels of several antioxidant nutrients found in these foods, a causal link between lack of antioxidants and disease occurrence or between antioxidant administration and disease prevention remains to be established.
- Many antioxidants can also act as pro-oxidants in certain conditions, such as the presence of transition metals or at high concentrations.
- Vitamin E and other nutrients that are classified as antioxidants have also been shown to modulate pathways of cell signaling and gene expression.
- Intervention studies demonstrating higher risk of cancer in smokers with β-carotene supplementation have, for example, attributed this finding to its pro-oxidant effect.
- Further evidence is required regarding the efficacy, safety, and appropriate dosage of anti-oxidants in relation to chronic disease.
- Currently, the most prudent public health advice continues to be to consume a variety of plant foods.

Introduction

Free radicals, including reactive oxygen and nitrogen species (ROS/NOS), are formed in the human body as a result of oxidative metabolism, i.e., as a result of the many chemical reactions and metabolic processes that occur in the body and also have numerous beneficial functions (Fig. 1). Free radicals are capable of modifying important molecules such as DNA, lipids, and proteins, and therefore affecting their ability to function or causing them to function abnormally. These processes are often referred to as oxidative damage. Antioxidants have the ability to scavenge and neutralize free radicals, or are necessary to enable other molecules to perform such a function. The body's antioxidant system is complex and consists of various intracellular and extracellular, endogenous and exogenous, and aqueous and lipid-soluble components that act in concert to prevent formation of ROS and terminate chains of ROS-initiated peroxidation of biological substrates. Antioxidants present in the diet include vitamin C, vitamin E, carotenoids and polyphenols. Metals and minerals that are key components of antioxidant enzymes, such as zinc or selenium, are also referred to as antioxidants. Many antioxidants can also act as pro-oxidants in certain conditions, such as the presence of transition metals or at high concentrations. Intervention studies demonstrating higher risk of cancer in smokers with β-carotene supplementation have, for example, attributed this finding to its pro-oxidant effect (see section on β-Carotene).

Oxidant or oxidative stress is a pro-oxidant shift in the oxidant–antioxidant balance caused by a relative or absolute deficiency of antioxidants. A pro-oxidant shift promotes damaging oxidative changes to important cellular constituents, and this may in turn lead to cellular dysfunction and, ultimately, to aging, disability, and disease. Long-term oxidative stress has therefore been linked with a number of chronic diseases including coronary heart disease (CHD), cancer, cataract, dementia, and stroke (Fig. 2).

Currently, it is known that ROS act as key signaling molecules for the maintenance of the metabolism. Therefore, maintaining ROS production at physiological levels is critical for achieving a correct metabolic function. A deregulation of ROS production, mostly an up-regulation, occurs in most metabolic diseases. Therefore, it could be suggested that the supplementation with antioxidants in these diseases can attenuate the disease. However, the blocking of ROS production through the use of antioxidants could prevent the correct function of molecular signaling and, therefore, it could be counterproductive (Fig. 2).

Other plant components, including polyphenols, have also been shown to possess antioxidant potential. To date, most of the evidence on antioxidant potential of polyphenols comes from *in vitro* studies, many possessing higher antioxidant potential *in vitro* than vitamins and carotenoids. However, although polyphenols show strong antioxidant properties *in vitro*, there has been much discussion about whether they are present in sufficient quantities *in vivo* to influence antioxidant activities. Studies have shown that polyphenols do not appear to be circulating in the blood at high enough concentrations to contribute significantly to the body's total antioxidant capacity. Further, it has been estimated that, after ingestion, around 90–95% of polyphenols undergo molecular changes. These changes to the structure of polyphenols can moderately or even radically change the "biological activities" of polyphenols found in *in vitro* studies. However, it has been suggested that polyphenols may exert an indirect antioxidant effect, by protecting endogenous antioxidant enzymes in the human body.

Recently, a new study provides a first level of evidence on the *in vivo* health benefits of olive oil triterpenes (oleanolic and maslinic acids) in healthy humans, decreasing DNA oxidation and plasma inflammatory biomarkers.

Owing to a lack of concise evidence of the potential antioxidant properties of polyphenols in the human body, these will not be discussed in this article.

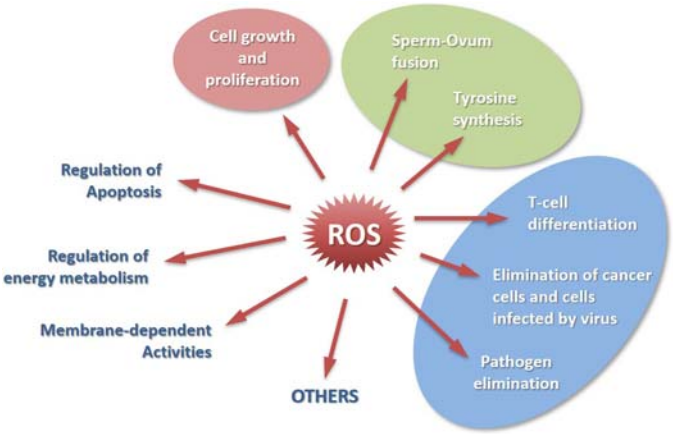


Fig. 1 Relevant actions and physiological functions of ROS. Adapted from Casuso and Huertas (2018).

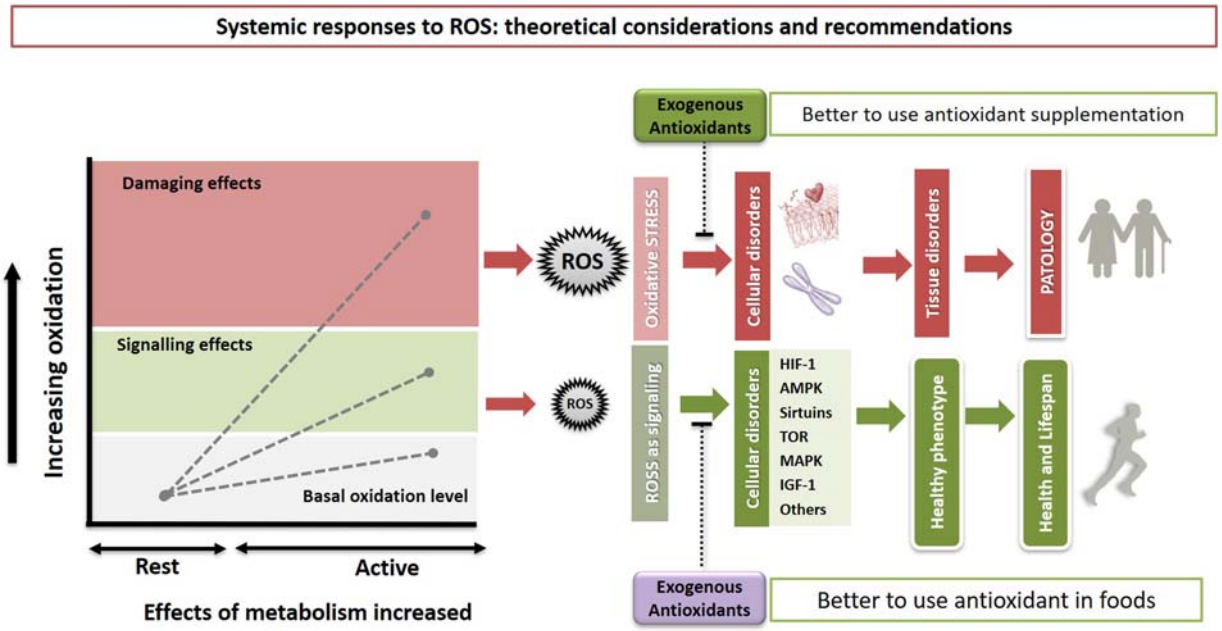


Fig. 2 Hypothetical model depicting body responses to increasing oxidation of redox-sensitive components due to increased ROS exposure. ROS are required in very small amounts as mediators of expression of critical genes for establishing an active and healthy phenotype. Low levels of ROS change the cellular redox state to increase the redox-sensitive signaling pathways and supplementation with antioxidants will be negative. High levels generate damaging effects. IGF-1, insulin-like growth factor-1; HIF-1, hypoxia-inducible factor-1; AMPK, AMP-activated protein kinase; FoxOs, forkhead box O proteins.

Definition of antioxidant

Word such "antioxidant" is difficult to define. Considering its function we could coin numerous definitions, but probably the most complete is the provided by B. Halliwell and Gutteridge in 1992: "*an antioxidant is any substance that when present at low concentrations compared to those of an oxidizable substrate delays or prevents oxidation of that substrate.*" Mechanisms of antioxidant action can include: removal of O2, scavenging reactive oxygen/nitrogen species or their inhibiting ROS/RNS formation, binding metal ions needed for catalysis of ROS generation, repair mechanisms and up regulation of endogenous antioxidant defenses.

Sics in 1996 defined antioxidants as the substances that neutralize free radicals or their actions. Ternay and Sorokin in 1997 defined antioxidant as any substance that hinders a free radical reaction. Azzi and Davies in 2004 defined antioxidant as the substances which counteract free radicals and prevent the damage caused by them.

Antioxidants sources

Antioxidants are molecules widely distributed in foods, especially those of plant origin. The concentration and nature of antioxidants depends on the type of food (plants or fruits), its seasonality and especially on the processing to which these foods have been subjected.

In Tables 1 and 2 we include foods and spices, ordered from the highest to the lowest antioxidant content (mg/100 g) and by their content in typical cooking portions (mg/portion). In addition to the numerical value, we incorporate a color range for quick visual reference.

Normal plasma concentration of the most common antioxidants

At present there are no consensus plasma normal values for antioxidants, probably due to their wide variety and nature. In Table 3 we make a proposal based on our data (Barranco-Ruiz et al., 2017) and on values obtained from numerous investigations and clinical trials in humans, published in the most relevant journals. One of the problems for the elaboration of this proposal is that plasma values usually present a multimodal distribution due to the intake of foods more or less rich in antioxidants or to the consumption or not of supplements. For example, as shown in Fig. 3A, the plasma tocopherol values of 100 healthy subjects show several peaks. Eighty percent of the sample ($n = 75$) presented α-tocopherol values < 80 nmol/mL, which are considered physiologically normal levels from a balanced diet, whereas the remaining 20% ($n = 19$) presented significantly elevated levels (>80 nmol/mL), suggesting vitamin E supplementation (Fig. 3A). In addition, a detailed analysis of the plasmatic α-tocopherol concentrations distribution in the normal levels group showed two peaks (Fig. 3B). The second peak corresponded to only five

Table 1 Antioxidants content of the more frequent spices used in cooking.

Spices	Tochopherol ($\alpha+\beta+\gamma+\Delta$) (mg)	Vitamin C (mg)	β-Carotene (mg)	α-Carotene (mg)	Lycopene (mg)	Lutein + zeaxanthin (mg)	Se (mg)	Zn (mg)	Total (mg/100 g)	Portion (g)	Antioxidant (mg)/portion
Parsley	10.51	125	1.15	0.017	0	2.43	0.0141	5.44	144.56	5	**7.2**
Pepper, red or cayenne	29.8	76.4	21.8	0	0	13.2	0.0088	2.48	143.69	5	**7.2**
Paprika, spice	33.24	0.9	26.2	0.595	0	18.9	0.0063	4.33	84.17	5	**4.2**
Thyme	7.58	50	2.26	0	0	1.9	0.0046	6.18	67.92	5	**3.4**
Dandelion greens	3.44	35	5.854	0.363	0	13.61	0.0005	0.41	58.68	5	**2.9**
Oregano	43.62	2.3	1.01	0.02	0	1.9	0.0045	2.69	51.54	5	**2.6**
Mustard	25.68	7.1	0.018	0	0	0.568	0.208	6.08	39.65	5	**2.0**
Cinnamon	12.98	3.8	0.112	0.001	0.015	0.222	0.0031	1.83	18.96	5	**0.9**
Clove	8.82	0.2	0.045	0	0	0	0.0072	2.32	11.39	5	**0.6**
Basil	0.77	0.8	0.378	0.113	0.393	1.15	0.003	7.1	10.71	5	**0.5**
Tumeric	5.16	0.7	0	0	0	0	0.0062	4.5	10.37	5	**0.5**
Black pepper	7.6	0	0.31	0.012	0.02	0.454	0.0049	1.19	9.59	5	**0.5**
Ginger	3.01	0.7	0.018	0	0	0	0.0558	3.64	7.42	5	**0.4**

participants with values between 46 and 80 nmol/mL, surpassing the average of the group without antioxidant supplementation evidence. These values correspond to subjects with a high antioxidant foods in their diet.

Antioxidants as signaling

In the last years the physiological role of the ROS have evolving and today playing important functions as keys in signaling mechanisms (Casuso and Huertas, 2018). In fact, it is necessary the maintenance of a physiological amount of ROS for maintaining the biologic function (**Fig. 2**). It is harmful having high levels of ROS as those produced in metabolic disease as well as having ROS levels below the physiological minimum. This hormetic response has been tested under different circumstances, among them it should be highlighted the study of supplementation with antioxidants during exercise. As described above, a large amount of ROS is produced during exercise through several independent pathways. Therefore, for years it has been tested if the intake of antioxidants could prevent certain processes related to the oxidative damage such as inflammation and the associated sarcomere damage. It was found that the intake of antioxidants before a strenuous test can, in fact, prevent muscle damage associated with ROS production. These findings affected the society becoming the supplementation with antioxidants extended to athletes. However, the chronic effect of the combination of exercise and supplementation with antioxidants was not considered. It was not until well into the 2000s when a series of researchers used animal and human models to test how the intake of antioxidant supplements interfere in chronic adaptations to training. Among the observed effects there was a performance decrease which was associated with a decrease in mitochondrial function and biogenesis by preventing the expression of PGC-1α. Therefore, the concept of "mitohormesis" was extended, where a minimum of mitochondrial oxidative stress is key for adaptations, whereas the total suppression of ROS as well as a sustained increase of ROS is harmful. There is still much to elucidate and to research with this regard, e.g., the discovery of antioxidants blocking particular pathways of ROS production could help to individualize pharmacological treatments in particular diseases. Indeed, diseases such as type 2 diabetes, cardiovascular diseases, cancer and alzheimer have been associated with an overproduction of ROS. In almost all these diseases the exercise seems to produce a beneficial effect, among other mechanisms by controlling this production of ROS. Nevertheless, the intake of antioxidants seems to prevent these improvements, this is very evident in subjects with risk of developing type 2 diabetes. In these subjects, the supplementation with antioxidants prevented the beneficial effects of exercise on insulin signaling.

New physiological antioxidants

The enzymatic antioxidant systems are synthesized in the body and it has been shown that they can be induced in response to the presence of ROS produced by exercise. This response depends on factors such as exercise duration and intensity. Several studies have shown that training can induce increased levels of antioxidant enzymes in human and laboratory animals muscle tissue, leading to reduced susceptibility of muscle to oxidative stress.

Recently, it has been shown that some of the antioxidant effects of exercise may be explained by an antioxidant mechanism at the mitochondrial level (Huertas et al., 2017). Complex I of the mitochondrial electron transport chain is a site of ROS production during exercise. Data accumulated over the past few years suggest that the distribution of mitochondrial electron transport chain complexes can change in response to cellular stimuli such as oxidative or energetic stress and that they can assemble into

Table 2 Antioxidants content of vegetables food.

Food	*Tochopherol* ($\alpha + \beta + \gamma + \Delta$) *(mg)*	*Vitamin C (mg)*	*Retinol (mg)*	*β-Carotene (mg)*	*Lycopene (mg)*	*Lutein + zeaxanthin (mg)*	*Se (mg)*	*Zn (mg)*	*Total (mg/ 100 g)*	*Portion (g)*	*Antioxidant (mg)/portion*
Guava	0.73	228	6.01	0.374	5.2	0	0.0006	0.23	240.54	150	**360.8**
Black currant	1	181	0	0	0	0	0	0.27	182.27	100	**182.3**
Yellow pepper	0	184	0	0.12	0	0	0.0003	0.17	184.29	70	**129.0**
Broccoli	0.96	89.2	0	0.361	0	1.4	0.0025	0.41	92.33	120	**110.8**
Orange	0	59.1	0	0	0	0	0	0.11	59.21	180	**106.6**
Brussels sprouts	0.88	85	0	0.45	0	1.59	0.0016	0.42	88.34	120	**106.0**
Kale	0.66	93.4	0	2.87	0	6.26	0.0009	0.39	103.58	100	**103.6**
Red pepper	1.78	128	0	1.62	0	0.051	0.0001	0.25	131.70	70	**92.2**
Kiwi	1.3	74.7	0	0.052	0	0.122	0.0002	0.14	76.31	100	**76.3**
Papaya	0.42	60.9	0	0.274	1.83	0.089	0.0006	0.08	63.59	100	**63.6**
Strawberries	0.39	58.8	0	0.007	0	0.026	0.0004	0.14	59.36	100	**59.4**
Cauliflower	0.28	48.2	0	0	0	0.001	0.0006	0.27	48.75	120	**58.5**
Lemon	0.15	53	0	0.003	0	0.011	0.0004	0.06	53.22	100	**53.2**
Grapefruit	0.13	33.3	0	0.014	0	0.01	0.0014	0.07	33.53	150	**50.3**
Spinach	2.21	28.1	0	5.63	0	12.2	0.001	0.53	48.67	100	**48.7**
Oysters	0.85	0	0.013	0	0	0	0.0197	39.3	40.18	100	**40.2**
Sweet potato	1.38	12.1	0	8.968	0	0	0.0002	0.19	22.64	150	**34.0**
Tomato	0	17.8	0	0.276	2.86	0.056	0.0025	0.08	21.07	150	**31.6**
Potato	0.01	19.7	0	0.001	0	0.009	0.0004	0.3	20.02	150	**30.0**
Pumpkin	1.06	9	0	3.1	0	1.5	0.0003	0.32	14.98	150	**22.5**
Watermelon	0.05	8.1	0	0.303	4.53	0.008	0.0004	0.1	13.09	200	**26.2**
Carrots	0.67	5.9	0	8.28	0.001	0	0.0001	0.24	15.09	130	**19.6**
Pork, kidney	0	13.3	0.059	0		0	0.19	2.75	16.30	100	**16.3**
Raspberries	3.39	26.2	0	0.012	0	0.136	0.67	0.42	30.83	50	**15.4**
Blackberries	3.45	21	0	0.128	0	0.118	0.0004	0.53	25.23	60	**15.1**
Wheat germ oil	149.4	0	0	0	0	0	0	0	149.40	10	**14.9**
Lettuce	0.47	4.6	0	5.23	0	2.31	0.0004	0.25	12.86	100	**12.9**
Artichokes	0.19	11.7	0	0.008	0	0.464	0.0002	0.49	12.85	100	**12.9**
Beef, kidney	0.24	9.4	0.419	0	0.02	0	0.141	1.92	12.14	100	**12.1**
Wheat germ	15.99	6	0	0.062	0	0.79	0.065	16.67	39.58	30	**11.9**
Red onion	0	8.1	0	0	0	0	0.0005	0.17	8.27	140	**11.6**
Oat	0.41	21.4	0.99	0	0	0.159	0.0249	13.4	36.38	30	**10.9**
Lentils	4.72	4.5	0	0.023	0	0	0.0001	3.27	12.51	80	**10.0**
Avocado	2.47	10	0	0.062	0	0.271	0.0004	0.64	13.44	70	**9.4**
Sunflower seeds	26.1	1.4	0	5	0	0	0.0793	5.29	37.87	20	**7.6**
Pecan nuts	26.66	1.1	0	0.029	0	0.017	0.0038	4.53	32.34	20	**6.5**
Chickpeas	0.82	4	0	0.04	0	0	0	2.76	7.62	80	**6.1**
Almonds	25.6	0	0	0.001	0	0.001	0.0041	3.12	28.73	20	**5.7**
Blueberries	0.97	9.7	0	0.032	0	0.08	0.0001	0.16	10.94	50	**5.5**
Hazelnuts	15.33	6.3	0	0.011	0	0.092	0.0024	2.45	24.19	20	**4.8**
Hazelnut oil	47.2	0	0	0	0	0	0	0	47.20	10	**4.7**
Sunflower oil	41.08	0	0	0	0	0	0	0	41.08	10	**4.1**
Almond oil	39.2	0	0	0	0	0	0	0	39.20	10	**3.9**
Cottonseed oil	35.3	0	0	0	0	0	0	0	35.30	10	**3.5**
Brazil nuts	5.65	0.7	0	0	0	0	1.917	4.06	12.33	20	**2.5**
Olive oil	14.35	0	0	0	0	0	0	0	14.35	10	**1.4**
Cocoa	0.1	0	0	0	0	0.038	0.0143	6.8	6.95	10	**0.7**

Table 3 Normal plasma concentration of the most common antioxidants.

	Media	*Standard deviations*
Tocopherol (μmol/L)	26.71	4.43
Vitamin C (μmol/L)	113.41	82.5
Retinol (μmol/L)	1.98	0.31
β-Carotene (μmol/L)	0.68	0.43
Ubiquinone (μmol/L)	0.5	0.17

supramolecular structures called mitochondrial supercomplexes. The main function attributed to these structures is an increased efficiency of mitochondrial respiration, which among other effects would prevent an excessive production of ROS. A very recent study has shown that exercise can limit ROS production by increasing the amount of complex I assembled in mitochondrial supercomplexes. The mechanism that leads to lower ROS production by complex I when it is in a supercomplex is not clear, although it has been suggested that these structures stabilize complex I. It has been confirmed that in metabolic diseases the levels of supercomplex assembly are lower, which has been related to oxidative damage. Therefore, research into strategies to improve the assembly of mitochondrial supercomplexes will be important over the next few years. In any case, exercise appears to be a potent stimulus that can decrease oxidative damage and associated metabolic diseases through changes in mitochondrial morphology.

The antioxidant hypothesis

A predominantly plant-based diet, high in fruit and vegetables, reduces the risk of developing several chronic diseases, including cancer and cardiovascular disease (CVD). It is often assumed that antioxidants, including vitamin C, Vitamin E, and carotenoids, contribute to this protection by interfering passively with oxidative damage to DNA, lipids, and proteins. This hypothesis is supported by numerous *in vitro* studies in animals and humans. A large number of descriptive, case-control, and cohort studies have also demonstrated an inverse association between high intakes or plasma levels of anti-oxidants and risk of CVD and cancer at numerous sites, as well as other conditions associated with oxidative damage, such as age-related macular degeneration, cataracts, and chronic obstructive pulmonary disease (COPD).

These findings provided a strong incentive for the initiation of intervention studies to investigate whether a lack of dietary antioxidants is causally related to chronic disease risk and if providing antioxidant supplements confers benefits for the prevention and treatment of these conditions. This article summarizes the findings of the largest primary and secondary trials published to date and considers their implications for future research and current dietary advice.

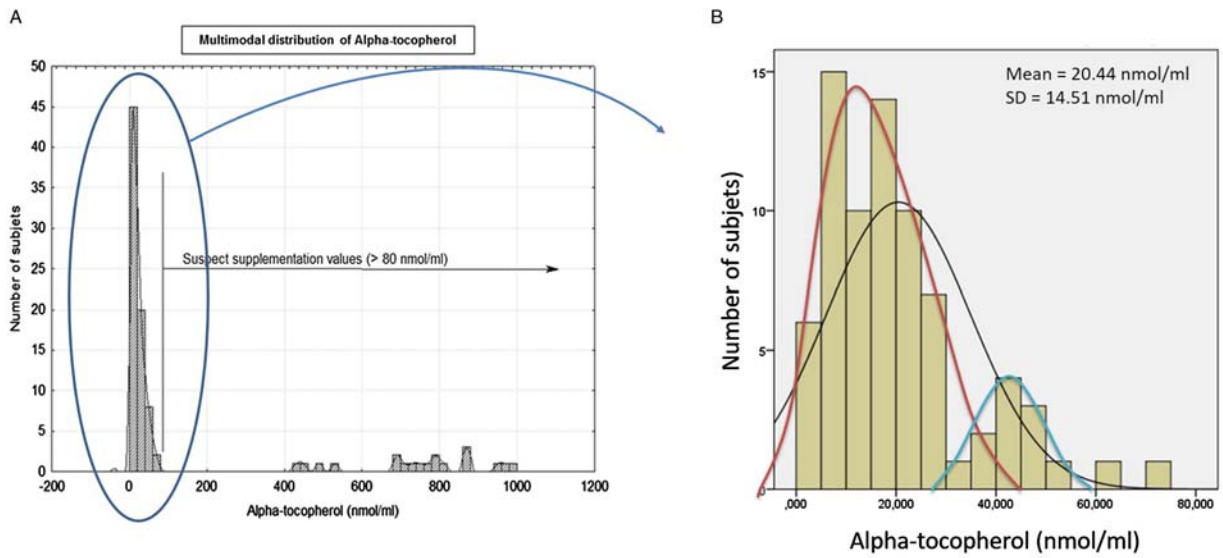


Fig. 3 (A) Multimodal distribution of Alpha-tocopherol concentrations after analysis with high-performance liquid chromatography. (B) Alpha-Tocopherol distribution for normal subjects without supplementations.

Cardiovascular disease

Of all the diseases in which excess oxidative stress has been implicated, CVD has the strongest supporting evidence. Oxidation of low-density lipoprotein (LDL) cholesterol appears to be a key step in the development of atherosclerosis, a known risk factor in the development of CVD. Small studies have demonstrated reductions in LDL oxidation (mostly *in vitro*) following supplementation with dietary antioxidants (particularly vitamin E, which is primarily carried in LDL-cholesterol), suggesting that they may provide protection against the development of heart disease. A number of large intervention trials using disease outcomes (rather than biomarkers such as LDL oxidation) have also been conducted to try to demonstrate a protective effect of vitamin E, β-carotene, and, to a lesser extent, vitamin C supplements on CVD. Most have been carried out in high-risk groups (e.g., smokers) or those with established heart disease (i.e., people with angina or who have already suffered a heart attack).

Primary prevention

The results of most primary prevention trials have not been encouraging (Table 4). For example, in the Finnish Alpha-Tocopherol Beta-Carotene Cancer prevention (ATBC) study, approximately 30,000 male smokers received vitamin E (50 mg d^{-1} of α-tocopherol), β-carotene (20 mg d^{-1}), both, or an inactive substance (placebo) for approximately 6 years. There was no reduction in risk of major coronary events with any of the treatments despite a 50% increase in blood vitamin E concentrations and a 17-fold increase in β-carotene levels. Moreover, with vitamin E supplementation, there was an un-expected increase in risk of death from hemorrhagic stroke and a small but significant increase in mortality from all causes with β-carotene supplementation (relative risk (RR), 1.08; 95% confidence interval (CI), 1.01–1.16). In the Physicians' Health Study (PHS) II, looking at the effect of 400 IU α-tocopherol supplementation on alternate days in approximately 14,000 male physicians, the risk of hemorrhagic stroke almost doubled with vitamin E supplementation during 8 years of follow-up (hazard ratio (HR) 1.99, 95% CI, 1.13–3.52). An increase in CVD deaths was also observed in the Beta-Carotene and Retinol Efficacy Trial (CARET), which tested the effects of combined treatment with β-carotene (30 mg d^{-1}) and retinyl palmitate (25,000 IU d^{-1}) in 18,000 men and women with a history of cigarette smoking or occupational exposure to asbestos compared to the placebo group (RR, 1.26; 95% CI, 0.99–1.61). However, during 6-year follow-up after stopping supplements no effect of supplementation was found. In contrast, in the Women's Health Study (WHS), which looked at the effect of 600 IU α-tocopherol (alternate days) in approximately 40,000 female nurses over a period of 10 years, a significant reduction in cardio-vascular deaths (RR, 0.76; 95% CI, 0.59–0.98), largely attributable to fewer sudden deaths, was observed.

Secondary prevention

The most positive results from secondary prevention trials came from the Cambridge Heart Antioxidant Study (CHAOS), a controlled trial on 2002 heart disease patients with angiographically proven coronary atherosclerosis randomly assigned to receive a high dose of vitamin E (400 or 800 IU d^{-1}) or placebo (Table 5). Those receiving the supplements were 77% less likely to suffer from nonfatal heart disease over the 1^1-year trial period than those who did not receive vitamin E (RR, 0.23; 95% CI, 0.11–0.47), although there was no reduction in CVD deaths. However, other large secondary prevention trials with longer follow-up have been less encouraging. For example, in a further analysis of the ATBC study, the β-carotene supplementation was associated with an increased risk of CHD deaths among men who had a previous heart attack and were thus at high risk of subsequent coronary events. There were significantly more deaths from fatal CHD in the β-carotene group (RR, 1.75; 95% CI, 1.16–2.64) and in the combined β-carotene and vitamin E group (RR, 1.58; 95% CI, 1.05–2.40) compared to the placebo group. The Heart Outcomes Prevention Evaluation (HOPE) study observed no benefit from vitamin E supplementation (400 IU d^{-1}) on CVD or all-cause mortality. The Heart Protection Study (HPS) in the UK examined the effect of 5 years of supplementation with a cocktail of anti-oxidant vitamins (600 mg vitamin E, 250 mg vitamin C, and 20 mg β-carotene) alone or in combination with the lipid-lowering drug Simvastatin or placebo in more than 20,000 adults with CHD, other occlusive arterial disease, or diabetes mellitus. Although blood levels of antioxidant vitamins were substantially increased, no significant reduction in the 5-year mortality from vascular disease or any other major outcome was noted. In the Italian GISSI-Prevenzione Trial dietary fish oils reduced the risk of fatal or nonfatal CVD in men and women who had recently suffered from a heart attack but vitamin E supplementation (300 mg daily for 3^1 years) did not provide any benefit. In these three trials, no significant adverse effects of vitamin E were observed. The PSH II study was designed as a primary prevention study including more than 14,000 male physicians, 754 of which had prevalent CVD. Analysis of this subsample showed that there was a nonsignificant decrease of total CVD (HR 0.82, 95% CI, 0.63–1.09), myocardial infarction (MI) (HR 0.88, 95% CI, 0.50–1.55), stroke (HR 0.74, 95% CI, 0.47–1.16) and cardiovascular mortality risk (HR 0.91, 95% CI, 0.70–1.17) with vitamin E supplementation, and of MI (HR 0.57, 0.32–1.02) with vitamin C supplementation. In the Women's Antioxidant Cardiovascular Study (WACS) in over 8000 post-menopausal women with a history of CVD or at least cardiac risk factors, supplementation with Vitamin E, β-carotene nor vitamin C showed any effect on total cardio-vascular events, MI, stroke, or deaths.

Systematic reviews and meta-analyses of the clinical trials to date have therefore concluded that supplementation with any single antioxidant nutrient or combination of nutrients has not demonstrated any consistent benefit for the treatment or prevention of CVD.

Table 4 Summary of large intervention trials (>1000 subjects) investigating the role of antioxidants and CVD in primary prevention.

Trial	*Characteristics of subjects*	*Sex*	*Length of follow-up (years)*	*Treatment*	*Effect of antioxidant supplementation*
ATBC	29,133 smokers, Finland	Male	6	50 mg α-tocopherol or 20 mg β-carotene	No significant effect on fatal or nonfatal CHD or total strokes with either supplement increase in deaths from hemorrhagic stroke in vitamin E group increase in hemorrhagic stroke (+62%) and total mortality (+8%) in b-carotene group
CARET	14,254 smokers, 4060 asbestos workers, USA	Male and female	3	30 mg β-carotene and 25,000 IU retinol	Increase in deaths from CVD (+26%) (terminated early)
LCPS	29,584 poorly nourished, China	Male and female	4	15 mg β-carotene, 30 mg α-tocopherol, and 50 mg selenium	Small decline in total mortality (−9%)
NPCT	1004 subjects with nonmelanoma skin cancer but without CVD, USA	Male and female	7.6	200 mg selenium daily	No effect on total CVD events, MI, stroke, or CVD mortality reduction in deaths from stroke in men (−55%) but not women
PHS	22,071 physicians, USA	Male	12	50 mg β-carotene or aspirin (alternate days)	No effect on fatal or nonfatal MI or stroke
PHS II	13,887 physicians, USA	Male	8	400 IU α-tocopherol (alternate days) or 500 mg ascorbic acid (daily) or placebo	No effect of vitamin E or vitamin C on CV events, MI, stroke, or CV mortality. Significantly increased risk of hemorrhagic stroke (+99%) with vitamin E
POPADAD	1276 adults with type 1 or type 2 diabetes and asymptomatic peripheral arterial disease, UK	Male and female	6.7	Antioxidant capsule (200 mg α-tocopherol, 100 mg ascorbic acid, 25 mg pyridoxine hydrochloride, 10 mg zinc sulfate, 10 mg nicotinamide, 0.4 mg lecithin, 0.8 mg sodium selenite) or 100 mg aspirin or placebo daily	No effect of antioxidant on CVD deaths or events
PPP	4495 with one or more CVD risk factors, Italy	Male and female	3 $\frac{1}{2}$	Low-dose aspirin or 300 mg α-tocopherol	No effect on CVD deaths or events (but inadequate power due to premature interruption of trial)
SCPS	1720 with recent nonmelanoma skin cancer, Australia	Male and female	8	50 mg β-carotene	No effect on CVD mortality
SUVIMAX	13,017, France	Male and female	7.5	Combination of 120 mg ascorbic acid, 30 mg vitamin E, 6 mg of β-carotene, 100 mg selenium, 20 mg zinc, daily	No effect on incidence of ischemic CVD
VACP II	1204 former asbestos workers, Australia	Male and female	5	30 mg β-carotene or 25,000 IU retinol (no placebo group)	No effect of b-carotene on CHD deaths
WHS	39,876, United States	Female	2 10	50 mg β-carotene (alternate days) 600 IU α-tocopherol (alternate days)	No effect on fatal or nonfatal CVD No effect on total CV events, MI or stroke significant 24% reduction in CV deaths (largely attributable to fewer sudden deaths)

ATBC, Alpha-Tocopherol Beta-Carotene Prevention Study; CARET, Beta Carotene and Retinol Efficacy Trial; LCPS, Linxian Cancer Prevention Study; NPCT, Nutritional Prevention of Cancer Trial; PHS, Physicians' Health Study; POPADAD, Prevention of Progression of Arterial Disease and Diabetes; PPP, Primary Prevention Project; SCPS, Skin Cancer Prevention Study; SUVIMAX, Supplementation en Vitamines et Mineraux Antioxydants (Vitamin and Mineral Antioxidant Supplementation Study); VACP, Vitamin A and Cancer Prevention; WHS, Women's Health Study; CHD, coronary heart disease; MI, myocardial infarction; CV, cardiovascular.

Table 5 Summary of large intervention trials (>1000 subjects) investigating the role of antioxidants and CVD in secondary prevention[a].

Trial	*Characteristics of subjects*	*Sex*	*Length of follow-up (years)*	*Treatment*	*Effect of antioxidant supplementation*
ATBC	1862 smokers with previous MI, Finland 1795 heavy smokers with previous angina, Finland	Male	5 1/2	50 mg α-tocopherol or 20 mg β-carotene	No effect on total coronary events (fatal and nonfatal) increase in deaths from fatal CHD in β-carotene (+75%) and combined b-carotene/vitamin E group (+58%) vs. placebo No effect on symptoms or progression of angina or on total coronary events
CHAOS	2002 patients with coronary atherosclerosis, UK	Male and female	1 1/2	300 or 800 IU α-tocopherol	Reduction in nonfatal MI (−77%) but no effect on CVD mortality
GISSI	11,324 patients with recent MI, Italy			300 mg α-tocopherol or 1 g n-3 PUFA	No benefit from vitamin E
HOPE	9541 known CVD or diabetes, Canada	Male and female	4–6	400 IU α-tocopherol or ACE inhibitor	No effect on MI, stroke, or CVD death
HPS	20,536 with known vascular disease or at high risk, UK	Male and female	≥5	20 mg β-carotene, 600 mg α-tocopherol, and 250 mg vitamin C	No effect on fatal or nonfatal MI or stroke
WACS	8171 postmenopausal women with history of CVD or at least 3 cardiac risk factors, USA	Female	9.4	600 IU α-tocopherol or 50 g β-carotene (both every other day) or 500 mg vitamin C (daily)	No effect of vitamin C, tocopherol or β-carotene on total CV events, MI, stroke or deaths

GISSI, GISSI Prevenzione Trial; HOPE, Heart Outcomes Prevention Evaluation Study; HPS, Heart Protection Study; WACS, Women's Antioxidant Cardiovascular Study; MI, myocardial infarction; PUFA, polyunsaturated fatty acids.

[a]Secondary prevention is defined as including patients with known or documented vascular disease.

Cancer

The oxidative hypothesis of carcinogenesis asserts that carcinogens generate ROS that damage RNA and DNA in cells, predisposing these cells to malignant changes and enhanced cancer risk. Most, but not all, damage is corrected by internal surveillance and repair systems involving dietary antioxidants, as well as endogenous antioxidant mechanisms. Antioxidants are therefore proposed to prevent cell damage by neutralizing free radicals and oxidants, thus preventing subsequent development of cancer.

β-carotene

Many of the randomized controlled trials (RCTs) investigating a protective role for antioxidant nutrients in cancer prevention (Table 6) have focused on β-carotene. A study in Linxian, China, of a rural population with poor nutritional status found that supplementation with a combination of β-carotene, selenium, and vitamin E for 5 years provided a 21% reduction in stomach cancer mortality and a 13% reduction in all cancer deaths. Although interesting, the population studied was likely to have very low intakes of a number of micro-nutrients and this study does not contribute to knowledge about the effects of individual antioxidants or offer any insight into their effects on populations with good nutritional status. The findings of a number of large double-blind RCTs in well-fed subjects using high-dose β-carotene supplements (either alone or in combination with other agents) have generally been unsupportive of any protective effect, although most have only focused on high-risk groups (e.g., smokers, asbestos workers, and older age groups), although newer studies provide data from the general population or subjects with a health issue not related to cancer risk (e.g., established CVD or CVD risk factors). In the ATBC Cancer Prevention Trial, in which 29,000 male smokers were randomly assigned to receive β-carotene, α-tocopherol or placebo each day, β-carotene showed no protective effect on the incidence of any type of cancer after approximately 6 years. In fact, concern was raised following the publication of the findings of this trial because those randomized to receive β-carotene had an 18% higher risk of lung cancer (RR, 1.18; 95% CI, 3–36) as well as an 8% higher total mortality than nonrecipients. Subgroup analyses suggested that the adverse effect of β-carotene on lung cancer risk was restricted to heavy smokers and that the risk appeared to be transient, being lost at follow-up 4–6 years after cessation of supplementation.

The CARET was also terminated early because of similar findings; subjects receiving a combination of supplements (30 mg β-carotene and vitamin A daily) experienced a 28% increased risk of lung cancer incidence compared with the placebo group (RR, 1.28; 95% CI, 1.04–1.57). Subgroup analyses also suggested that the effect was found in current, but not former, smokers. In contrast, in the PHS, supplementation of male physicians with 50 mg β-carotene on alternate days had no effect on cancer incidence (men who were smokers did not experience any benefit or harm). The HPS also demonstrated no effect on 5-year cancer incidence or mortality from supplementation with 20 mg β-carotene in combination with vitamins E and C in individuals at high risk of CVD, despite increases in blood concentrations of these nutrients (plasma β-carotene concentrations increased 4-fold). They did not,

Table 6 Summary of large intervention trials (>1000 subjects) investigating the role of antioxidants and cancer in primary prevention.

Trial	*Characteristics of subjects*	*Sex*	*Length of follow-up (years)*	*Treatment*	*Effect of antioxidant supplementation*
ATBC	29,133 smokers, Finland	Male	5–8	50 mg α-tocopherol or 20 mg β-carotene	18% increase in lung cancer in b-carotene group (no effect in vitamin E group) 34% reduction in incidence of prostate cancer in vitamin E group No effect of either vitamin on colorectal, pancreatic, or urinary tract cancer, or cancer of the oral cavity/pharynx, esophagus and larynx
CARET	14,254 smokers, 4060 asbestos workers, USA	Male and female	4	30 mg β-carotene and 25,000 IU retinol	Lung cancer increased by 28%
HPS	20,536 at high CVD risk, UK	Male and female	>5	20 mg β-carotene, 600 mg α-tocopherol, and 250 mg vitamin C	No effect on cancer incidence or mortality
LCPS	29,584 poorly nourished, China	Male and female	5	15 mg β-carotene, 30 mg α-tocopherol, and 50 mg selenium	Cancer deaths declined by 13% Stomach cancer declined by 21%
NSCPT	1621 (73% without skin cancer at baseline), Australia	Male and female	4 1/2	30 mg β-carotene with or without sunscreen application	No effect on basal cell or squamous cell carcinoma
PHS	22,071 physicians, United States	Male	12	50 mg β-carotene or aspirin (alternate days)	No effect on incidence of malignant neoplasms or nonmelanoma skin cancer
SELECT	35,533, USA, Canada, and Puerto Rico	Male	5.5	200 mg selenium or 400 IU vitamin E or both	No effect of either treatment on prostate cancer incidence No effect on lung, colorectal and all other cancers, and cancer deaths
SUVIMAX	13,017, France	Male and female	7.5	Combination of 120 mg ascorbic acid, 30 mg vitamin E, 6 mg of β-carotene, 100 mg selenium, 20 mg zinc, daily See above	No effect on overall cancer incidence in whole cohort, but significant reduction (−31%) of overall cancer incidence in men No effect on overall skin cancer incidence in total sample; but increased incidence in women taking antioxidants, not in men No effect on nonmelanoma skin cancer No effect on overall prostate cancer incidence Significant reduction (-48%) of prostate cancer in those with baseline PSA levels $<3\ \mu g L^{-1}$
	5141 men, France	Male	8.5–9		
VACP II	1204 former asbestos workers, Australia	Male and female	5	30 mg β-carotene or 25,000 IU retinol (no placebo group)	No effect of β-carotene on cancer mortality
WHS	39,876, USA	Female	2	50 mg β-carotene (alternate days)	No effect on cancer incidence
			10	600 IU α-tocopherol (alternate days)	No effect on overall cancer incidence and death, no effect on site-specific cancers (breast, lung, colon, rectum, and stomach)
WACS	7627 postmenopausal women with history of CVD or at least 3 cardiac risk factors, USA	Female	9.4	600 IU α-tocopherol or 50 g β-carotene (both every other day) or 500 g vitamin C (daily)	No effect of any antioxidant on total cancer incidence and mortality. Higher lung cancer rates in vitamin C group (+84%)

LCPS, Linxian Cancer Prevention Study; NSCPT, Nambour Skin Cancer Prevention Trial; SELECT, Selenium and Vitamin E Cancer Prevention Trial; SUVIMAX, Supple'mentation en Vitamines et Mine'raux Antioxydants (Vitamin and Mineral Antioxidant Supplementation Study); VACP, Vitamin A and Cancer Prevention; PSA, prostate-specific antigen.

however, find any harmful effects from these vitamins. The WACS in approximately 7600 postmenopausal women having established CVD or CVD risk factors showed no effect of supplementation with β-carotene (in combination with vitamin E or vitamin C or placebo) on total cancer incidence and mortality. Interestingly, the Vitamin and Mineral Antioxidant Supplementation Study (SUVIMAX) found a significant 31% reduction of total cancer incidence in men (RR 0.69, 95% CI, 0.53–0.91) taking a low-dose supplement containing β-carotene along with vitamin C, vitamin E, selenium, and zinc, but found no effect in women (RR 1.04, 95% CI, 0.85–1.29). The authors suggested that the difference in men and women may be explained by differences in nutrient

status at baseline, but that lower baseline status alone could not entirely explain the observed differences. In men, supplementation with this mixture of low-dose micronutrients showed no effect on overall prostate cancer incidence, but showed a significant reduction of prostate cancer (HR 0.52, 95% CI, 0.29–0.92) in those with baseline prostate-specific antigen (PSA) levels >3 mg L^{-1} (high PSA levels can be an indicator for higher risk of prostate cancer or existing prostate cancer).

A number of trials have attempted to investigate the effect of β-carotene supplementation on nonmelanoma skin cancer, the most common forms of which are basal cell and squamous cell carcinomas (these types of cells are both found in the top layer of the skin). However, none have shown any significant effect on skin cancer prevention. For example, the PHS found no effect after 12 years of β-carotene supplementation on the development of a first nonmelanoma skin cancer. The Nambour Skin Cancer Prevention Trial (NSCPT) of 1621 men and women followed for nearly 5 years (most of whom had no history of skin cancer at baseline) showed that those supplemented with 30 mg β-carotene did not experience any reduction in risk of basal cell or squamous cell carcinoma or the occurrence of solar keratoses (precancerous skin growths that are a strong determinant of squamous cell carcinoma). The SUVIMAX found no effect of low-dose β-carotene along with vitamin C, vitamin E, selenium, and zinc on risk of skin cancer in the total sample, but found an increased frequency of skin cancer in the female group (1.3% vs. 0.7%, p ¼ 0.02). A 5-year trial of 1805 men and women with recent nonmelanoma skin cancer (the Skin Cancer Prevention Study) also found that supplementation with 50 mg of β-carotene gave no protection against either type of skin cancer, although this may have been because these cancers have a long latency period of approximately 12 years (**Table 7**).

Together, these trials suggest that β-carotene supplements offer no protection against cancer prevention in healthy individuals and, among smokers, may actually increase the risk of lung cancer. Investigators have sought to explain these findings by proposing that components of cigarette smoke may promote oxidation of β-carotene in the lungs, causing it to exert a pro-oxidant (rather than antioxidant) effect and act as a tumor promoter.

Vitamin C

Only one RCT has investigated the effect of vitamin C alone in primary prevention of cancer. In the WACS trial 500 g of vitamin C per day for an average of 9.5 years had no effect on total cancer incidence and mortality in postmenopausal women with history of CVD or cardiac risk factors, but was associated with higher lung cancer rates in vitamin C group (RR 1.84, 95% CI 1.14–2.97). Data from a small number of trials of vitamin C in combination with other nutrients have not provided any support for a role for high-dose vitamin C supplementation in cancer prevention (**Table 6**). The Linxian trial found no significant effect of supplementing Chinese men and women with 120 mg vitamin C and 30 mg molybdenum daily for 5 years on the risk of cancers of the esophagus or stomach. The polyp prevention study, a trial of 864 patients with previous adenoma, found no effect of either β-carotene or a combination of vitamins E and C (1000 mg) on the incidence of subsequent colorectal adenomas. The HPS also found no beneficial effects of supplementation with these three vitamins on cancer mortality. However, trials have generally been carried out on those with diets containing sufficient amounts of vitamin C and there is a need for further studies in people with low intakes. As mentioned above, in the SUVIMAX study a low-dose supplement containing vitamin C along with β-carotene, vitamin E, selenium, and zinc found a significant reduction in total cancer incidence in men (RR 0.69, 95% CI, 0.53–0.91) but not in women.

Vitamin E

The ATBC trial showed no significant effect of α-tocopherol supplementation (50 mg d^{-1}) on risk of lung, pancreatic, colorectal, or urinary tract cancers among heavy smokers (**Table 6**). However, in a post hoc subgroup analysis a 34% reduction in the risk of prostate cancer was seen in men who received this supplement. Although interesting, prostate cancer was not a primary endpoint of this

Table 7 Summary of large intervention trials (>1000 subjects) investigating the role of antioxidants and cancer in secondary prevention[a].

Trial	*Characteristics of subjects*	*Sex*	*Length of follow-up (years)*	*Treatment*	*Effect of antioxidant supplementation*
NPCT	1312 with history of basal or squamous cell carcinoma, United States	Male and female	4 1/2	200 mg selenium	No effect on incidence of skin cancer overall increased risk of squamous cell carcinoma (+25%) and of total nonmelanoma skin cancer (+17%) reduced cancer mortality (−50%), cancer incidence (−37%), prostate cancer (−63%), colorectal cancer (−58%), and lung cancer (−46%)
SCPS	1805 with recent nonmelanoma skin cancer, USA	Male and female	5	50 mg β-carotene	No effect on occurrence of new nonmelanoma skin cancer

NPCT, Nutritional Prevention of Cancer Trial; SCPS: Skin Cancer Prevention Study.

[a]Secondary prevention defined as subjects with documented cancer including nonmelanoma skin cancer (although some of the primary prevention trials did not exclude those with nonmelanoma skin cancer at baseline).

study. In the Selenium and Vitamin E Cancer Prevention Trial (SELECT) study, where prostate cancer was the primary outcome, no effect of vitamin E supplementation over a period of 5.5 years in around 35,000 men was found. In this study, no effect on any other cancer site was observed, including lung, colorectal and all other cancer sites, and there was no effect on cancer death rates. The WHS and WACS trials, both carried out in women, also found no effect of vitamin E supplementation on overall cancer incidence or deaths. The HPS found no effect of vita-min E in combination with vitamin C and β-carotene on cancer incidence or mortality, and the SUVIMAX study found a reduction of total cancer incidence in men but not women with low dose vitamin E along with vitamin C, β-carotene, selenium, and zinc. Two smaller, short-term intervention studies found no effect of α-tocopherol supplementation on mammary dysplasia or benign breast disease. Several trials have also been unable to demonstrate a protective effect of vitamin E supplementation on the risk or recurrence of colorectal adenomatous polyps.

Selenium

A few trials have suggested that selenium supplementation may have a protective effect on liver cancer in high-risk groups living in low-selenium areas. For example, the provision of selenium-fortified salt to a town in Qidong, China, with high rates of primary liver cancer, reduced the incidence of this cancer by 35% compared with towns that did not receive this intervention. Some trials have also demonstrated the incidence of liver cancer to be significantly reduced in subjects with hepatitis B and among members of families with a history of liver cancer receiving a daily supplement of 200 mg of selenium for 4 and 2 years, respectively. The Nutritional Prevention of Cancer Trial (NPCT) in the US also supported a possible protective role of selenium in 1312 patients (mostly men) with a previous history of skin cancer who were supplemented with 200 mg selenium per day for 4^1 years (**Table 7**). Those receiving selenium demonstrated significant reductions in the risk of total cancer incidence (37%) and mortality (50%) compared to those receiving placebo. The selenium-treated group also had substantial reductions in the incidence of lung, colorectal, and prostate cancers of 46, 58, and 63%, respectively. However, recurrent squamous cell carcinoma was increased by 25% and total non-melanoma skin cancer by 17%. Further analysis showed the protective effect on prostate cancer to be confined to those with lower baseline PSA and plasma selenium levels. However, the SELECT study, one of the largest human cancer prevention trials ever undertaken in 45,533 healthy males from the US, Puerto Rico, and Canada, found no effect of 200 mg of selenium on prostate cancer, lung cancer, colorectal cancer and all other cancers, or on cancer deaths. Comparison of this study to other clinical trials involving selenium and to animal studies suggests that the source of the selenium supplement, L-selenomethionine, and the relatively high initial levels of selenium in the enrolled men may have contributed to this outcome.

Other diseases associated with oxidative damage

Type 2 diabetes

Type 2 diabetes is associated with elevated oxidative stress (especially lipid peroxidation) and declines in antioxidant defense. This is thought to be due in part to elevated blood glucose levels (hyperglycemia), but severe oxidative stress may also precede and accelerate the development of type 2 diabetes and then of diabetic complications (CVD and microvascular complications such as retinopathy, neuropathy, and nephropathy).

Some of the trials looking at antioxidants and CVD and cancer have also reported on the association of the examined antioxidants and the risk of diabetes. In the WACS women with a history of CVD or at least 3 CVD risk factors but free of diabetes at baseline (n ¼ 6574) received 500 mg vitamin C daily, or 600 IU α-tocopherol every other day, or 50 mg β-carotene every other day. During a median follow-up of 9.2 years no significant effect of any of the treatments could be found: RR in vitamin C group was 0.89 (95% CI, 0.78–1.02), RR in vitamin E group 1.13 (95% CI, 0.99–1.29) and RR in β-carotene group 0.97 (95% CI, 0.85–1.11). Analysis of data from the ATBC study in 27,379 smokers free of diabetes at baseline showed that neither supplementation with α-tocopherol (50 mg d^{-1}) nor with β-carotene (20 mg d^{-1}) for 5–8 years had an effect on diabetes risk over a median follow-up of 12.5 years. Data from the WHS from almost 39,000 healthy women free of diabetes, cancer, and CVD at baseline and receiving either 600 IU of α-tocopherol or placebo on alternate days showed no effect of study treatment on diabetes risk over a median follow-up period of 10 years. In a smaller study, the NPCT, in 1202 participants with a history of non-melanoma skin cancer but no baseline diabetes treatment with 200 mg d^{-1} selenium was associated with an increased risk of developing diabetes (HR 1.55, 95% CI, 1.03–2.33) after an average follow-up of 7.7 years.

Small-scale human trials have shown administration of high doses of vitamin E to reduce oxidative stress and improve some CVD risk factors, such as blood glycated hemoglobin, insulin, and triglyceride levels, in people with diabetes. Such trials have also indicated benefit from vitamin E in improving endothelial function, retinal blood flow, and renal dysfunction. However, the findings of large clinical trials investigating the role of individual or a combination of antioxidant nutrients in reducing the risk of CVD and microvascular complications in people with diabetes have generally been disappointing. For example, the HOPE trial investigated the effects of vitamin E and the drug Ramipril in patients at high risk for CVD events and included a large number of middle-aged and elderly people with diabetes (more than 3600). An average of 4^1 years of supplementation with 400 IU of vitamin E per day was found to exert no beneficial or harmful effect on CVD outcomes or on nephropathy. The Primary Prevention Project (PPP) trial found no effect of vitamin E (300 mg d^{-1}) supplementation for 3 or 4 years in diabetic subjects, and the HPS, which included a number of people with diabetes, also reported no benefit of a combination of antioxidant vitamins on mortality or incidence of vascular disease. In the ATBC study, no effect of supplementation with α-tocopherol (50 mg d^{-1}) nor

β-carotene (20 mg d^{-1}) was found during the intervention period (median 6.1 years) on the risk of macro-vascular complication or total mortality in 1700 men with type 2 diabetes at baseline. No essential changes were found in these effects when the follow-up was extended up to 19 years. In another study, the Prevention of Progression of Arterial Disease and Diabetes (POPADAD) study, no effect of an antioxidant capsule (containing 200 mg α-tocopherol, 100 mg vitamin C, 25 mg pyridoxine hydrochloride, 10 mg zinc sulfate, 10 mg nicotinamide, 9.4 mg lecithin, and 0.8 mg sodium selenite) was found on CVD risk in 1276 adults with type 1 or type 2 diabetes and an ankle brachial pressure index of 0.99 or less but no symptomatic CVD at baseline over a median follow-period of 6.7 years. However, a meta-analysis of two studies reporting by diabetes subtype, the Haptoglobin (Hp) 2-2 genotype which is characterized by a markedly in-creased risk of CVD compared to other types, found that patients with Hp 2-2 diabetes taking vitamin E supplements reduced the risk of cardiovascular events (combined odds ratio (OR) 0.58, 95% CI 0.40–0.86).

Asthma and chronic obstructive pulmonary disease (COPD)

Asthma is a chronic inflammatory disease resulting in reversible airways bronchoconstriction which affects a large number of children and adults. Epidemiological studies suggest that intake and status of antioxidant nutrients are inversely associated with the risk of asthma and wheezing. A number of studies have also demonstrated a beneficial effect of fruit and vegetable intake on lung function. For example, regular consumption of fresh fruit rich in vitamin C (citrus fruits and kiwi) has been found to have a beneficial effect on reducing wheezing and coughs in children.

In COPD patients, many of the pathophysiological changes associated with the disease are produced through the generation of oxygen free radicals by activated inflammatory cells. Antioxidant nutrients have therefore been suggested to play a role in the prevention and treatment of these conditions. Common examples of COPD are emphysema and chronic bronchitis. COPD mainly affects smokers or people with a smoking history.

Vitamin C is the major antioxidant present in extracellular fluid lining of the lung, and intake in the general population has been inversely correlated with the incidence of asthma, bronchitis, and wheezing and with pulmonary problems. Al- though some trials have shown high-dose supplementation (1–2 g d^{-1}) to improve symptoms of asthma in adults and protect against airway responsiveness to viral infections, allergens, and irritants, this effect has been attributed to the antihistaminic action of the vitamin rather than to any anti-oxidant effect. The results of these trials have also been in-consistent, and a Cochrane review of nine RCTs concluded that there is insufficient evidence to recommend a specific role for vitamin C in the treatment of asthma. However, a need for further trials to address the question of the effectiveness of vitamin C in asthmatic children was highlighted.

Other dietary antioxidants have been positively associated with lung function in cohort studies but the findings of clinical trials have been mixed. In a study of 158 children with moderate to severe asthma, supplementation with vitamin E (50 mg d^{-1}) and vitamin C (250 mg d^{-1}) led to some improvement in lung function following ozone exposure. A study in 72 adults with asthma receiving 500 mg vitamin E daily for 6 weeks did not show any effect on bronchial re-activity to methacholine, lung function, morning peak flow, or any other outcome measures. A Cochrane review looking at selenium and asthma only identified one study meeting their pre-defined criteria. The study in 24 patients suffering from chronic asthma found that 100 m d^{-1} of selenium for 14 days was associated with significant improvements in subjective asthma symptoms but this improvement could not be validated by significant changes in the separate clinical parameters of lung function and airway hyperresponsiveness. A sub-sequent RCT in 197 participants found no effect of 100 mg daily for 24 weeks on asthma-related quality of life, lung function, asthma symptom scores, or any other outcome measures. The ATBC trial found no benefit from supplementation with α-tocopherol (50 mg d^{-1}) and β-carotene (20 mg d^{-1}) on symptoms of COPD, despite the fact that those with high dietary intakes and blood levels of these vitamins at baseline had a lower prevalence of chronic bronchitis and dyspnea.

Macular degeneration and cataracts

The eye is at particular risk of oxidative damage due to high oxygen concentrations, large amounts of oxidizable fatty acids in the retina, and exposure to ultraviolet rays. In Western countries, age-related macular degeneration (AMD) is the leading cause of blindness among older people. Cataracts are also widespread among the elderly and occur when the lens is unable to function properly due to the formation of opacities within the lens. These develop when proteins in the eye are damaged by photooxidation; these damaged proteins build up, clump, and precipitate. It has been proposed that antioxidants may prevent cellular damage in the eye by reacting with free radicals produced during the process of light absorption.

The results of intervention trials in this area have also been mixed. The age-related eye disease study in the US investigating the effects of combined antioxidant vitamins C (500 mg), E (400 IU), and β-carotene (15 mg) with and without 80 mg zinc daily for 6 years showed some protective effect (a reduction in risk of approximately 25%) on the progression of moderately advanced AMD but no benefit on the incidence or progression or early AMD or cataracts. Further analysis of the data showed that an individual's response could be related to a specific genotype. The lutein antioxidant supplementation trial, a 12-month study of 90 patients with AMD, found significant improvements in visual function with 10 mg d^{-1} lutein (one of the major carotenoids found in the pigment of a normal retina) alone or in combination with a number of other antioxidant nutrients. The Roche European Cataract Trial, providing a combined daily supplement of β-carotene, vitamin C, and vitamin E among adults with early signs of age-related cataract, showed a small deceleration in the progression of cataract after 3 years.

However, the Linxian trial found no influence of vitamin supplementation on risk of cataract; the ATBC trial found no reduction in the prevalence of cataracts with vitamin E, β-carotene, or both among male smokers; and the PHS of more than 22,000 men

showed no benefit from 12 years of supplementation with β-carotene (50 mg on alternate days) on age-related maculopathy or cataract incidence. In fact, current smokers at the beginning of this trial who received the supplement experienced an increased risk of cataract (by ap-proximately 25%) compared to the placebo group. The Vitamin E, Cataract, and Age-Related Maculopathy Trial also reported no effect of supplementation with vitamin E for 4 years (500 IU d^{-1}) on the incidence or progression of cataracts or AMD.

Immune function—common cold

Antioxidants, vitamins, and minerals have also been linked with immune function. Deficiency of these nutrients, including vitamin C, vitamin E, vitamin A, selenium, and zinc, has been associated with compromised immune function. A Cochrane systematic review looked at the effects of the use of vitamin C supplements on the incidence of colds in the nor-mal population. Based on 30 trials including almost 12,000 study participants on the use of vitamin C supplements (doses from 200 mg or higher), the reviewers concluded that supplements did not reduce the incidence of colds in the normal population. Vitamin C supplementation taken for prophylaxis reduced the duration of common cold symptoms by 8% in adults and by 13.6% in children, but no effect was found when vitamin C was taken therapeutically (i.e., after the onset of the common cold). Despite the lack of effectiveness in reducing the risk of a common cold infection in the normal population, vitamin C supplementation could reduce the risk of a common cold by half in study participants exposed to short periods of extreme physical or cold stress, or both (including marathon runners and skiers). In the ATBC study long-term vitamin E and β-carotene supplementation had no overall effect on common cold incidence. Among subjects aged 65 years or older, the incidence of colds was slightly lower in the vita-min E group compared with the control (unsupplemented) group (RR ¼ 0.95; 95% CI ¼ 0.90–1.00); this reduction was greatest among older city dwellers who smoked fewer than 15 cigarettes per day (RR ¼ 0.72; 95% CI ¼ 0.62–0.83). A Cochrane review on the use of zinc for common cold treatment based on seven studies including around 800 study participants concluded that, overall, treatment with zinc lozenges did not reduce the duration of cold symptoms.

Possible explanations for the disagreement between the findings of observational studies and clinical trials

Various explanations have been given for the different findings of observational studies and intervention trials. Clearly, non-randomized studies are unable to exclude the possibility that antioxidants are simply acting as a surrogate measure of a healthy diet or lifestyle and that the protective effect of certain dietary patterns, which has been presumed to be associated with dietary antioxidants, may in fact be due to other com-pounds in plant foods, substitution of these foods for others, or a reflection of other health behaviors common to people who have a high fruit and vegetable intake. However, although intervention studies provide a more rigorous source of evidence than observational studies, they are not without weak-nesses from a nutritional perspective. Many of the trials to date have been criticized for their use of high-risk populations and high doses of single supplements, insufficient duration of treatment, and follow-up and lack of consideration of the impact of genetic variability. More recent trials that have at-tempted to answer some of these criticisms. For example, SUVIMAX was designed to assess the efficacy of supplementation among more than 12,000 healthy men and women over a 7.5-year period with a cocktail of antioxidant vitamins and minerals at doses achievable by diet (approximately one to three times the daily recommended dietary allowances) on premature death from CVD and cancer. This low-dose anti-oxidant supplementation had no effect on vascular disease incidence but lowered total cancer incidence in men, but not in women. This suggests that the contradictory results of observational and interventional studies may be due to differences in the effects of antioxidants in relation to supplement doses (nutritional vs. pharmacological), baseline antioxidant status (different between gender or nutritional status), and health status of subjects (healthy vs. cancer high-risk subjects). This study concluded that antioxidant supplementation may have a beneficial effect on cancer incidence only in healthy subjects who are not exposed to cancer risk, and with a particularly low baseline status. In contrast, high doses of anti-oxidant supplements may be deleterious in high-risk subjects without any clinical symptoms in whom the initial phase of cancer development has already started. Studies are also now attempting to look at genetic variation by studying the effects of antioxidants in individuals with specific genotypes. For example, in a meta-analysis of two trials (Hope and ICARE), vitamin E supplementation was associated with reduced risk of CVD in individuals with diabetes mellitus and the haptoglobin 2-2 genotype, both of which are associated with increased risk of the disease.

Conclusion

Although there is a substantial body of evidence that diets rich in plant foods (particularly fruit and vegetables) convey health benefits, as do high plasma levels of several antioxidant nutrients found in these foods, a causal link between lack of antioxidants and disease occurrence or between antioxidant administration and disease prevention remains to be established. There is a lack of understanding of the mechanisms underpinning the apparent protective effect of plant foods and, as yet, no clear picture of which components are effective and hence no way of predicting whether all or just some plant foods are important in this respect.

If future trials do demonstrate a reduction in chronic dis-ease risk with antioxidant supplementation, this cannot be definitively attributed to the antioxidant effect of these nutrients because other biological functions may also play a role. For example, in

addition to retarding LDL oxidation, vitamin E may help to protect against CVD *via* its action on platelet aggregation and adhesion or by inhibition of the proliferation of smooth muscle cells. Vitamin E and other nutrients that are classified as antioxidants have also been shown to modulate pathways of cell signaling and gene expression. Furthermore, although vitamin C, vitamin E, and selenium have been shown to decrease the concentration of some of the biomarkers associated with oxidative stress, the relationship between many of these biomarkers and chronic disease remains to be elucidated.

The intervention studies highlight the lack of information on the safety of sustained intakes of moderate to high doses of micronutrient supplements and long-term harm cannot be ruled out, particularly in smokers. Further evidence is required regarding the efficacy, safety, and appropriate dosage of anti-oxidants in relation to chronic disease. Currently, the most prudent public health advice continues to be to consume a variety of plant foods.

References

Barranco-Ruiz, Y., Aragón-Vela, J., Casals, C., Martínez-Amat, A., Casuso, R.A., Huertas, J.R., 2017. Control of antioxidant supplementation through interview is not appropriate in oxidative-stress sport studies: analytical confirmation should be required. Nutrition 33, 278–284.

Casuso, R.A., Huertas, J.R., 2018. Antioxidant supplements in obesity and metabolic syndrome: angels or demons. In: Obesity: Oxidative Stress and Dietary Antioxidants. Academic Press, London, pp. 263–275.

Huertas, J.R., Al Fazazi, S., Hidalgo-Gutierrez, A., López, L.C., Casuso, R.A., 2017. Antioxidant effect of exercise: exploring the role of the mitochondrial complex I superassembly. Redox Biol. 13, 477–481.

Further reading

Asplund, K., 2002. Antioxidant vitamins in the prevention of cardiovascular disease: a systematic review. J. Intern. Med. 251, 372–392.

Chong, E.W., Wong, T.Y., Kreis, A.J., Simpson, J.A., Guymer, R.H., 2007. Dietary antioxidants and primary prevention of age related macular degeneration: systematic review and meta-analysis. Br. Med. J. 335, 755.

Clarke, R., Armitage, J., 2002. Antioxidant vitamins and risk of cardiovascular disease. Review of large-scale randomised trials. Cardiovasc. Drugs Ther. 16, 411–415.

Evans, J., 2008. Antioxidant supplements to prevent or slow down the progression of AMD: a systematic review and meta-analysis. Eye 22, 751–760.

Galan, P., Brian-con, S., Favier, A., et al., 2005. Antioxidant status and risk of cancer in the SU.VI.MAX study: is the effect of supplementation dependent on baseline levels? Br. J. Nutr. 94, 125–132.

Greenwald, P., Anderson, D., Nelson, S.A., Taylor, P.R., 2007. Clinical trials of vitamin and mineral supplements for cancer prevention. Am. J. Clin. Nutr. 85, 314S–317S.

Hercberg, S., Galan, P., Preziosi, P., et al., 2004. The SU.VI.MAX Study: a randomized, placebo-controlled trial of the health effects of antioxidant vitamins and minerals. Arch. Intern. Med. 164, 2335–2342.

Kaur, B., Rowe, B.H., Arnold, E., 2009. Vitamin C supplementation for asthma. Cochrane Database Syst. Rev. 21 (1), CD000993.

Lawlor, D.A., Davey Smith, G., Kundu, D., et al., 2004. Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? Lancet 363, 1724–1727.

Mares, J.A., 2004. High-dose antioxidant supplementation and cataract risk. Nutr. Rev. 62, 28–32.

Morris, C., Carson, S., 2003. Routine vitamin supplementation to prevent cardiovascular disease: a summary of the evidence for the US Preventive Services Task Force. Ann. Intern. Med. 139, 56–70.

Myung, S.K., Kim, Y., Ju, W., Choi, H.J., Bae, W.K., 2010. Effects of antioxidant supplements on cancer prevention: meta-analysis of randomized controlled trials. Ann. Oncol. 21, 166–179.

Stanner, S.A., Hughes, J., Kelly, C.N.M., et al., 2004. A review of the epidemiological evidence for the antioxidant hypothesis. Publ. Health Nutr. 7, 407–422.

Vivekananthan, D., Penn, M.S., Sapp, S.K., et al., 2003. Use of antioxidant vitamins for the prevention of cardiovascular disease: meta-analysis of randomised trials. Lancet 361, 2017–2023.

Bioavailability

L. Davidsson, International Atomic Energy Agency (IAEA), Vienna International Centre, Vienna, Austria
S.A. Tanumihardjo, University of Wisconsin-Madison, Madison, WI, USA

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of L. Davidsson, S.A. Tanumihardjo, Bioavailability, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 149–155, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00024-6.

Introduction 76
Nonheme Iron Bioavailability 76
Bioavailability as an Important Component in the Development of Food-Based Strategies to Combat Iron Deficiency: Food Diversification 78
Bioavailability as an Important Component in the Development of Food-Based Strategies to Combat Iron Deficiency: Food Fortification 79
Bioavailability of Provitamin A Carotenoids 80
Carotenoids 80
Bioconversion of Carotenoids to Vitamin A 80
Bioavailability of Carotenoids 81
Conversion Factors 82
Effectors of Bioavailability 83
Influence of Vitamin A Status 83
Conclusion 83
Further Reading 83
Relevant Websites 84

Glossary

Bioavailability The fraction of absorbed and utilized micronutrient.
Bioconversion The rate at which absorbed provitamin A is cleaved to vitamin A in the intestine.
Bioefficacy Combines absorption and bioconversion and is the efficiency with which ingested dietary provitamin A carotenoids are absorbed and converted to vitamin A.
Biofortification Agronomically improving stable crops by enhancing micronutrient levels through traditional breeding or genetic modification.
Stable isotope Nonradioactive isotope.

Introduction

Bioavailability refers to the fraction of absorbed and utilized micronutrient. This concept is particularly important for some micronutrients, e.g., nonheme iron, zinc, provitamin A carotenoids, folate, and vitamin B_{12} as bioavailability varies widely depending on a number of factors. These factors include nutritional status, physiological factors such as gastric acid secretion, food matrix, and interactions between nutrients in addition to the presence of enhancers and inhibitors in the diet (**Figure 1**). Stable (nonradioactive) isotope techniques to assess micronutrient bioavailability have been developed over the last 20 years or so and the application of these techniques has contributed significantly to our understanding of the importance of bioavailability in micronutrient nutrition. As stable isotope techniques do not expose the study population, or the investigators, to any potential health hazard related to ionizing radiation, studies in vulnerable population groups at high risk of developing micronutrient deficiencies are feasible. Consequently, over the last 10–15 years, crucial new information, for example, about dietary enhancers and inhibitors of nonheme iron absorption in infants and children has been made available, and important data have been generated on nonheme iron bioavailability from iron compounds used in food fortification programs. In vitamin A nutrition, the development of stable isotope techniques to estimate body pool sizes of vitamin A has provided new important information on the bioefficacy of provitamin A carotenoids and the influence of dietary composition, as well as the nutritional status of the consumer, on provitamin A bioavailability.

Nonheme Iron Bioavailability

Fractional absorption of nonheme iron varies widely in individuals, from less than 1% to 100%, depending on the nutritional status and physiological factors of the individual as well as the composition of the test meal (see **Figure 1**). From a methodological point of view, the rapid incorporation of newly absorbed iron into a target tissue, which can be sampled relatively easily, i.e., erythrocytes,

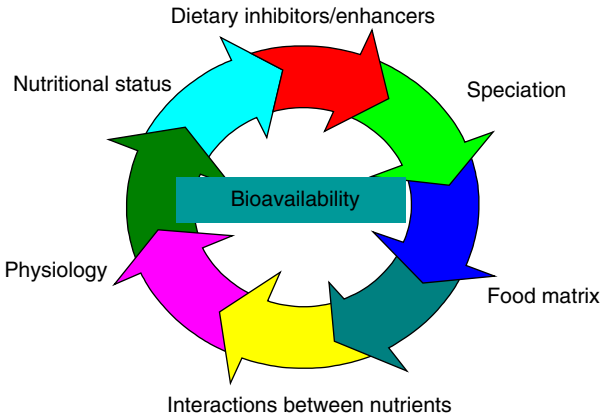


Figure 1 Factors influencing bioavailability of nonheme iron and provitamin A carotenoids.

is a major advantage, and a stable isotope technique to evaluate iron bioavailability has been developed based on the incorporation of stable iron isotopes into red blood cells 14 days after administration of labeled test meals. This methodology was originally developed by investigators using radioactive isotopes of iron, and the development of stable isotope technique has benefited greatly from this original work related to the development of the study protocol (defining the appropriate times for blood sampling, collecting data on incorporation rates etc.). Usually, the incorporation rate of newly absorbed iron into erythrocytes is assumed to be relatively constant, approximately 80–90% (see **Figure 2**). However, when the incorporation rate cannot be assumed to remain stable, for example, during pregnancy or in individuals infected with malaria, incorporation of a stable isotope administered intravenously can be used to correct for changes in incorporation rate.

Considerable interindividual variation in iron bioavailability has been demonstrated, largely due to differences in iron status among individuals, and paired comparisons are therefore essential when evaluating iron bioavailability from different foods or food fortificants. Using a double isotope technique, i.e., administration of two stable isotopes of iron (usually ^{57}Fe and ^{58}Fe) – on consecutive days – information about iron bioavailability from two different test meals can be obtained simultaneously (**Figures 3 and 4**). When evaluating bioavailability from iron compounds used for food fortification, ferrous sulfate is typically used as the reference compound and relative bioavailability is reported, i.e., bioavailability from ferrous sulfate is set at 100% and bioavailability of other compounds are compared to this value (**Table 1**). Blood samples drawn at baseline and 14 days after administration are analyzed for ^{57}Fe- and ^{58}Fe-enrichment by Thermal Ionization Mass Spectrometry (TIMS) or High Resolution Inductively Coupled Plasma Mass Spectrometry. Although the number of suitable mass spectrometers dedicated to nutrition remains limited worldwide, the application of this technique has been used in a wide range of settings, based on close North–South collaboration. The recently installed TIMS in Bangalore, India, will hopefully increase the application of this technique in Asia. Successful implementation of stable isotope studies depend on joint efforts made by a group of people, including analytical chemists, nutritionists, and relevant health professionals such as pediatricians, nurses, and health workers. As many different steps are crucial – both during the preparation for the study, i.e., development of study protocols, preparation of test meals, preparation of stable isotope doses, administration of labeled test meals – as well as in the final analysis of enriched blood samples – a multidisciplinary team is clearly needed to plan and implement this kind of study.

Over the last 10–15 years, stable isotope technique has been used to generate new data on iron bioavailability from iron compounds used in food fortification programs and information about dietary enhancers and inhibitors of iron absorption in

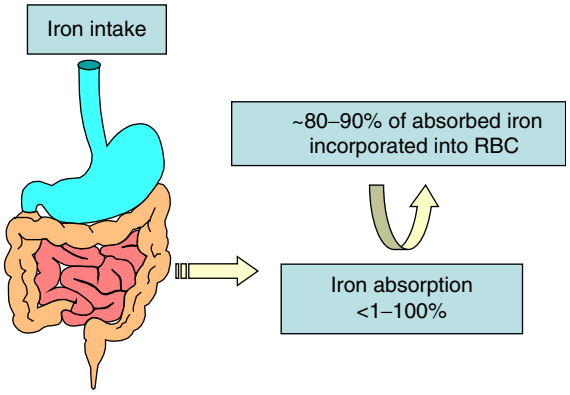


Figure 2 Newly absorbed iron is rapidly incorporated into red blood cells (RBC) at a relatively constant, high rate.

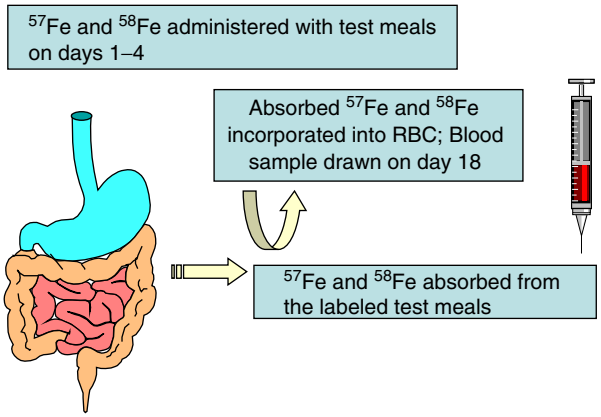


Figure 3 Basic principles of the double stable isotope technique to evaluate iron bioavailability based on incorporation into red blood cells (RBC).

vulnerable population groups. More information about the importance of bioavailability of nonheme iron in the development of food-based strategies to combat iron deficiency, i.e., food fortification and dietary diversification, is provided below.

Bioavailability as an Important Component in the Development of Food-Based Strategies to Combat Iron Deficiency: Food Diversification

A number of dietary factors influencing nonheme iron bioavailability were identified during the earlier studies using radioactive isotopes of iron, as well as more recent data based on stable isotope technique. Inhibitors include phytic acid, polyphenols, and calcium, as well as some proteins, whereas ascorbic acid and muscle tissue (the 'meat factor') enhance nonheme iron absorption. More recently, the importance of host factors such as nutritional deficiencies, infection/inflammation, and genetic disorders has received more attention, and there is clearly a need to further investigate the importance of these factors in different population groups. The identification of hepcidin as a key regulator of iron homeostasis highlights the need for future studies to investigate the role of this peptide and its increased expression during chronic inflammation and obesity on iron absorption. In addition, in spite of numerous studies, including studies to evaluate the potential influence of vitamin A on nonheme iron bioavailability in healthy Western adults and African children with subclinical vitamin A deficiency, the influence of vitamin A and provitamin A carotenoids on nonheme iron absorption still remains largely unknown.

The importance of optimizing nonheme iron bioavailability from the diet can be assumed to be crucial for individuals with high requirements and consuming inhibitory diets, in particular, infants and young children in developing countries. The positive effect of adding meat to a vegetable-based complementary food on nonheme iron bioavailability has been demonstrated in European infants and the inclusion of meat into the diet of young children would thus be beneficial by providing highly bioavailable heme iron as well as by enhancing the bioavailability of nonheme iron. However, meat is often not available or affordable and,

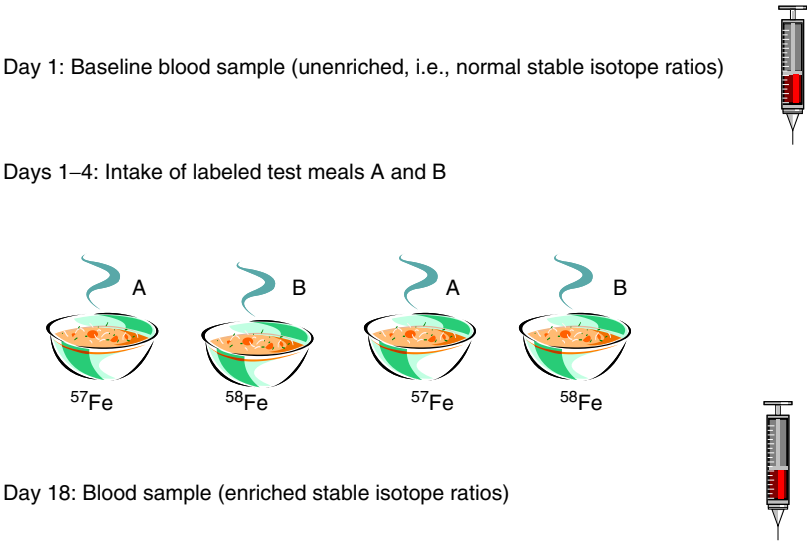


Figure 4 Study design to evaluate iron bioavailability from test meals A and B.

Table 1 Examples of data on relative iron bioavailability in adult humans

Compound	*Relative iron bioavailability (%)*
Ferrous sulfate	100
Sodium iron EDTA	>100
Ferrous fumarate	100
Ferric pyrophosphate	21–74
Elemental iron; electrolytic	75

Source: Adapted from World Health Organization (WHO), Food and Agriculture Organization of the United Nations (FAO) (2006) In: Allen L, De Benoist B, Dary O, and Hurrell RF (eds.) *Guidelines on Food Fortification with Micronutrients*. http://www.who.int/nutrition/publications/guide_food_fortification_micronutrients.pdf

thus, infants and young children living in resource poor communities where monotonous, cereal-based diets are consumed are of special concern.

Stable isotope technique has been used to demonstrate the potent inhibitory effect of phytic acid as well as the strong enhancing effect of ascorbic acid on nonheme iron bioavailability in infants. Dephytinization of cereals and legumes can be made by the addition of exogenous phytase or by the use of whole-grain cereals as a source of phytase, and ascorbic acid can be included into the diet by consumption of fruits, in particular citrus fruits rich in vitamin C. However, inclusion of fruits into the young child's diet can be limited by availability, affordability, and tradition. An alternative source of ascorbic acid in the diet of African infants and young children was identified – human milk – and subsequently tested for its effect on nonheme iron bioavailability from a traditional complementary food in Bangladesh, using stable isotope technique. Although human milk contributed considerable amounts of ascorbic acid, no significant effect on iron bioavailability was observed, indicating that components in human milk modify the influence of ascorbic acid. A more recent study confirmed the potent enhancing effect of ascorbic acid on nonheme iron bioavailability in Pakistani infants and clearly demonstrated that breastfeeding immediately after intake of a complementary food with added ascorbic acid does not blunt the enhancing effect of the vitamin when evaluated by stable isotope technique.

Bioavailability as an Important Component in the Development of Food-Based Strategies to Combat Iron Deficiency: Food Fortification

Food fortification is currently implemented as a public health strategy to combat iron deficiency in many countries, either by fortification of staple foods such as milled cereal flours to reach a large proportion of the population or by targeted approaches based on fortification of products consumed by vulnerable population groups, e.g., fortified commercial infant foods. More recently, other approaches have been developed to reach individuals, in particular, infants and young children, who do not have access to centrally produced foods by providing micronutrients in sachets, crushable tablets, and spreads for 'in-home fortification'.

The concept of bioavailability is central to the development of food fortification strategies as the overall impact of a food fortification program on the consumers' iron status will be influenced by the bioavailability of the iron compound used as well as the presence of inhibitors and enhancers of iron absorption in the diet. Information about bioavailability of iron compounds represents an integral part of the development of food fortification strategies, together with careful selection of the food vehicle and testing of changes to the organoleptic properties of the fortified food (**Figure 5**). These findings include important information on significant

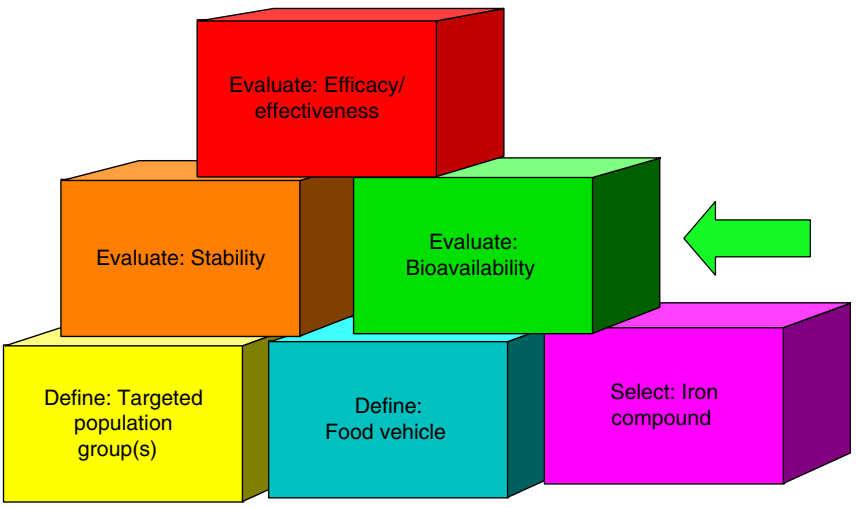


Figure 5 Components of an iron fortification program.

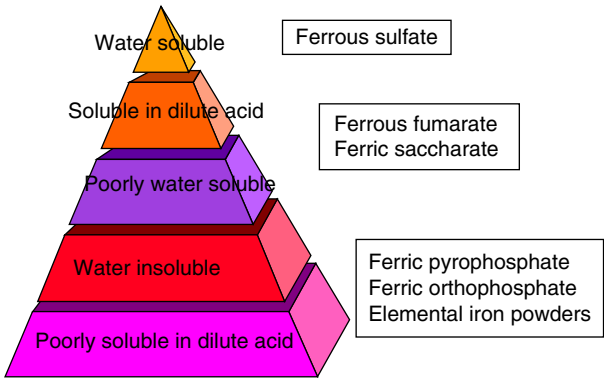


Figure 6 Examples of iron compounds with different solubility, ranging from freely soluble in water to poorly soluble in dilute acid.

differences in iron bioavailability between iron compounds, with different physicochemical properties such as solubility in water and dilute acid (Figure 6). For example, stable isotope technique has been used to demonstrate a three-fold difference in iron bioavailability from iron compounds with different solubility properties used to fortify infant cereals in different parts of the world, i.e., ferrous fumarate and ferric pyrophosphate, in infants. In addition, substantially lower relative bioavailability of ferrous fumarate versus ferrous sulfate has been noted in preschool children in Bangladesh (approximately 30%) as compared to that in Western adult women (100%). These observations clearly highlight the importance of studies in appropriate population groups as opposed to extrapolating data from other groups, e.g., healthy adults in the North to infants and young children in the South, who are less well-nourished and have lower gastric acid output. Furthermore, stable isotope technique has been applied to demonstrate the usefulness of NaFeEDTA for fortification of high extraction rate cereal flours and condiments such as fish sauce and soy sauce. This technique has also been used to evaluate different strategies to optimize iron bioavailability from fortified foods consumed by children, for example, by the addition of enhancers such as ascorbic acid or NaEDTA.

The World Health Organization (WHO) and the Food and Agriculture Organization of the United Nations (FAO) have developed guidelines to assist in the development and evaluation of food fortification programs. The importance of bioavailability data to optimize the impact of food fortification by selecting iron compounds with high relative bioavailability as well as by modifying the composition of the diet to increase iron absorption, is highlighted in these guidelines. In particular, the potent enhancing effect of ascorbic acid and the strong inhibitory effect of phytic acid are emphasized. Briefly, these recommendations provide a list of iron fortificants (based on relative bioavailability data) in the following order of preference: (1) ferrous sulfate, (2) ferrous fumarate, (3) encapsulated ferrous sulfate or fumarate, (4) electrolytic iron, and (5) ferric pyrophosphate. To compensate for low relative bioavailability from electrolytic iron and ferric pyrophosphate, it is recommended to add twice the amount. In addition, it is recommended to add ascorbic acid at a 2:1 molar ratio; for foods high in phytic acid, the molar ratio should be increased to 4:1 to counteract the inhibitory effect of phytic acid. NaFeEDTA is recommended for fortification of cereal flours high in phytic acid and for certain condiments such as fish sauce and soy sauce.

Bioavailability of Provitamin A Carotenoids

Carotenoids

Carotenoids are responsible for the colors of many fruits and vegetables. Bioavailability of carotenoids from food sources is an active area of research because of purported health benefits. For example, a diet high in fruit and vegetables, which are high in carotenoids, is related to a reduced risk of various types of cancer. Many factors are known to impact the bioavailability of these fat-soluble compounds from food. Evidence exists that the structure of the carotenoid impacts its ability to become part of the lipid complex (i.e., micelle) that makes it easily absorbed by the human body. The hydroxyl-group containing carotenoids tend to be more bioavailable than the hydrocarbon carotenoids. Cooking enhances carotenoid bioavailability by disrupting the vegetable matrix, but prolonged heat destroys them or changes the structural configuration. The isomeric form of β-carotene impacts its bioavailability and the forms that are 'bent' or in the *cis* configuration are less bioavailable than the 'straight' or all-*trans* forms.

Provitamin A carotenoids number approximately 50 in nature, but only three of these are common in the human diet, namely, α-carotene, β-carotene, and β-cryptoxanthin (Figure 7). α-Carotene is found predominantly in carrots and some squashes, β-carotene in orange vegetables and green leafy vegetables, and β-cryptoxanthin in citrus fruits and yellow maize.

Bioconversion of Carotenoids to Vitamin A

The process of producing vitamin A from provitamin A carotenoids is termed bioconversion. The nutritional status of the person impacts whether or not the provitamin A carotenoids are converted to vitamin A. For example, if a person is vitamin A deficient they

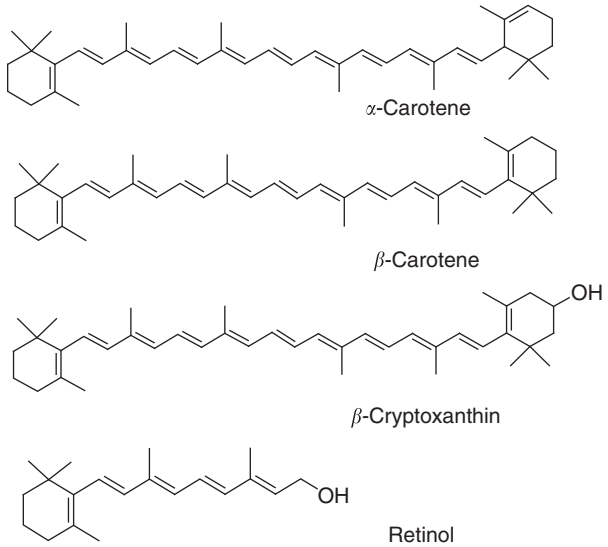
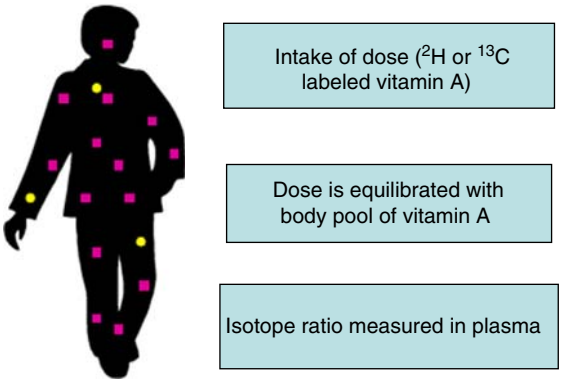


Figure 7 Chemical structures of provitamin A carotenoids in comparison to retinol. β-Carotene is able to supply two molecules of retinol, whereas α-carotene and β-cryptoxanthin supply one.

will convert more of the absorbed provitamin A carotenoids to vitamin A. However, if the person is zinc deficient, less provitamin A will be converted to vitamin A because the enzyme that cleaves the provitamin A carotenoids to vitamin A is zinc dependent.

Bioavailability of Carotenoids

Bioavailability is often measured by looking at differences in serum carotenoid concentrations. However, this only determines relative differences between two comparisons. Stable isotope techniques to evaluate vitamin A status were developed to use in vulnerable populations. The method involves administering a dose of isotopically labeled tracer and waiting for the dose to equilibrate with body pools of vitamin A, usually two to three weeks (**Figure 8**). Basically, the method is an indirect measure of liver reserves of vitamin A, and the test does require blood samples. Typically a blood sample is taken at baseline and then after equilibration to measure the change in enrichment caused by the administered isotope. The application of stable isotope techniques to evaluate changes in vitamin A body pools has contributed significantly to the evaluation of interventions based on provitamin A carotenoids. The principles of the study design used for these applications are described in **Figure 9**. In practice, two stable isotopes have been used including the heavy isotope of hydrogen, deuterium, and the heavy isotope of carbon, carbon 13. Usually, deuterated retinol esterified to acetate with four or eight deuterium atoms is administered, followed by analysis of deuterium enrichment in serum or plasma samples by gas chromatography-mass spectrometry (GCMS), the so-called 'paired-stable isotope dilution technique'. The study design used in 'paired-stable isotope dilution technique' studies is presented in **Figure 10**. In particular, the paired-stable isotope dilution technique has been used to generate new data on provitamin A conversion factors to retinol.



Illustration; Kerstin Gross–Helmert

Figure 8 Stable isotope dilution technique to assess vitamin A body pools.

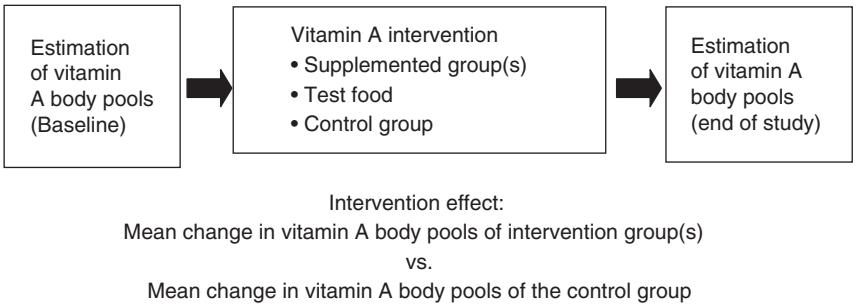


Figure 9 Principle of study design based on vitamin A body pool size.

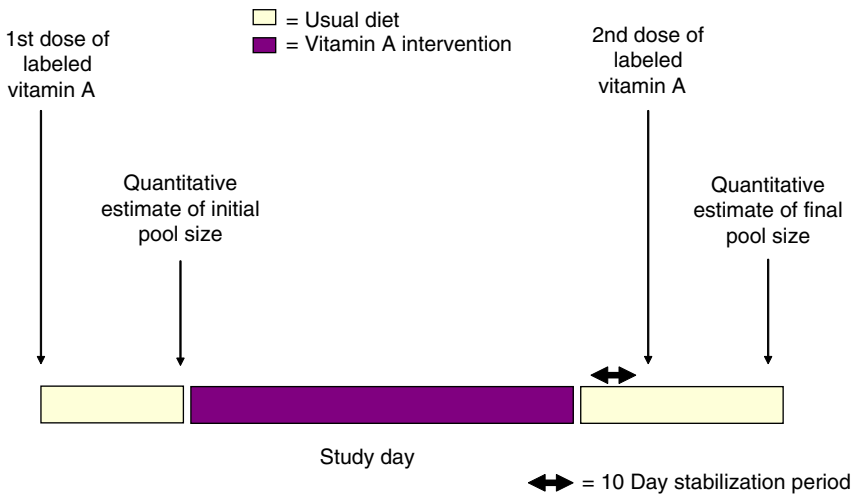


Figure 10 Design for studies based on vitamin A body pool assessment using the paired-isotope dilution technique.

Conversion Factors

Conversion factors are a way to define how much of an ingested provitamin A carotenoid is converted to 1 μg retinol. This is also known as bioefficacy. The Institute of Medicine uses the factors of 12 μg β-carotene:1 μg retinol and 24 μg α-carotene or β-cryptoxanthin:1 μg retinol (Table 2). This equivalency is referred to as retinol activity equivalents (RAE) to distinguish it from the conversion factors still used by the Food and Agriculture Organization, which is 6 μg β-carotene:1 μg retinol and referred to as retinol equivalents or RE. This estimate of vitamin A formation from provitamin A assumes that the person is healthy and eating a mixed diet. In comparison to these generalized conversion factors, a mean conversion factor of 26.7 μg β-carotene:1 μg retinol was reported for green and yellow vegetables consumed by Chinese school-aged children and 13.4 μg β-carotene:1 μg retinol for sweet potato, 9.5:1 for Indian spinach (*Basella alba*), and 6.3:1 for pure β-carotene in oil were obtained in Bangladeshi men. These studies were able to determine conversion factors by applying the paired-stable isotope dilution technique to evaluate the changes in total body vitamin A pools in response to the intervention with the test food.

Table 2 The commonly used bioconversion factors of provitamin A carotenoids to yield 1 μg retinol include those published by the Institute of Medicine and the Food and Agriculture Organization of the United Nations (FAO)

Institute of medicine	*FAO*
2 μg β-Carotene in oil supplement	2 μg β-Carotene in oil supplement
12 μg β-Carotene	6 μg β-Carotene
24 μg β-Cryptoxanthin	12 μg β-Cryptoxanthin
24 μg α-Carotene	12 μg α-Carotene

Other isotopic methods have also been developed to evaluate bioconversion factors. Using a different approach other than the paired-stable isotope dilution technique, intrinsically deuterium-labeled vegetables were prepared by growing carrots and spinach hydroponically in deuterated water. Humans were fed the labeled vegetables and a reference dose of $^{13}C_8$-retinyl acetate was administered for comparison. Intake of vegetables resulted in conversion of 20.9 μg β-carotene:1 μg retinol and 14.8:1 for spinach and carrots, respectively. More recently, the intrinsic labeling technique was used to evaluate bioconversion of provitamin A carotenoids in rice that was genetically modified to contain β-carotene. The process of enhancing micronutrients in staple crops, such as rice, is called biofortification. This rice is called 'Golden Rice' due to its deep yellow hue exclusively due to β-carotene. In comparison to a reference dose, the β-carotene from Golden Rice resulted in a mean conversion factor of 3.8 μg β-carotene to 1 μg of retinol for five adults. These data highlight the potential usefulness of biofortification to impact vitamin A status. In general, animal and human studies that have tested stable crops biofortified with provitamin A have shown very good conversion rates of provitamin A to retinol demonstrating the feasibility of this agronomic technique. At this time, plant sources of provitamin A are often overlooked in infant feeding strategies in developing countries. However, a simulation based on intake of 100 g orange-fleshed sweet potato each day using the Institute of Medicine's 12:1 conversion factor resulted in a significant increase of vitamin A liver stores in infants. Furthermore, feeding sweet potato to children in South Africa and Mozambique improved vitamin A status demonstrating sweet potato β-carotene is bioavailable and converted to vitamin A.

Finally, extrinsic reference methods using stable isotopically labeled β-carotene and retinyl ester have been used to determine bioconversion factors from supplements and foods. These methods have been compared to traditional oral–fecal balance studies and seem to be better suited to supplement studies than mixed diets.

Effectors of Bioavailability

Dietary fat is needed for absorption of carotenoids in order to facilitate the formation of micelles in the small intestine. However, fat does not seem to be a strong limiting factor in human diets when vegetables are consumed as part of a composite meal. For example, studies in Filipino schoolchildren, based on the paired-isotope dilution test, found no difference between three levels of fat (7, 15, or 29 g fat day^{-1}) when fed with 4.2 mg provitamin A carotenoids in the form of carotenoid-rich vegetables. Studies in animals, however, did see enhancement of bioefficacy when high levels of fat were fed with orange-fleshed sweet potato. Some forms of fiber also interfere with the bioavailability of carotenoids, although the addition of dry matter in the form of white sweet potato to orange-fleshed sweet potato did not seem to affect the bioefficacy.

Influence of Vitamin A Status

From a methodological point of view, it is important to recognize that the retinol response to provitamin A carotenoids in humans and animals varies inversely with vitamin A status. Clearly, there is an urgent need for additional, well-designed studies to evaluate the bioavailability of provitamin A carotenoids consumed in settings where vitamin A deficiency remains a public health problem among infants, children, and women of child-bearing age. The WHO continues to advocate high-dose vitamin A supplements, which are oil-based, to preschool children in part due to the lack of consistent high provitamin A sources in the diets of the poor and the questionable bioavailability of the provitamin A carotenoids.

Conclusion

Stable isotope techniques provide excellent tools to assess bioavailability of nonheme iron and provitamin A carotenoids and thus contribute important information to be used in the development of nutrition interventions.

Further Reading

Allen, LH., 2008. Causes of vitamin B_{12} and folate deficiency. Food and Nutrition Bulletin 29 (supplement 2), S20–S34.

Food and Agricultural Organization of the United Nations (FAO) (2001) Chapter 7: Vitamin A. In: *Human Vitamin and Mineral Requirements*. http://ftp://ftp.fao.org/docrep/fao/004/y2809e/y2809e07.pdf

Furr, H.C., Green, M., Haskell, M., et al., 2005. Stable isotope dilution techniques for assessing vitamin A status and bioefficacy of provitamin A carotenoids in humans. Public Health Nutrition 8, 596–607.

Hambidge, K.M., Miller, L.V., Westcott, J.E., Sheng, X., Krebs, N.F., 2010. Zinc bioavailability and homeostasis. American Journal of Clinical Nutrition 91 (supplement), 1478S–1483S.

Hurrell, R.F., Egli, I., 2007. Optimizing the bioavailability of iron compounds for food fortification. In: Kraemer, K., Zimmermann, M.B. (Eds.), Nutritional Anemia. Sight and Life Press, Basel, Switzerland, pp. 77–90.

Hurrell, R., Egli, I., 2010. Iron bioavailability and dietary reference values. American Journal of Clinical Nutrition 91 (supplement), 1461S–1467S.

Institute of Medicine, Food and Nutrition Board, 2001. Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Molybdenum, Nickel, Silicon, Vanadium, and Zinc. National Academy Press, Washington, DC, pp. 65–126.

International Zinc Nutrition Consultative Group (IZiNCG) Technical Document #1, 2004. Techniques for measuring zinc absorption. Appendix 3. Food and Nutrition Bulletin 25 (supplement 2), S198–S199.

Kastenmayer, P., Davidsson, L., Galan, P., Cherouvrier, F., Hercberg, S., Hurrell, R.F., 1994. A double stable isotope technique for measuring iron absorption in infants. British Journal of Nutrition 71, 411–424.

Tanumihardjo, S.A., 2008. Food-based approaches for ensuring adequate vitamin A nutrition. Comprehensive Reviews in Food Science and Food Safety 7, 373–381.

Van Loo-Bouwman, C.A., West, C.E., van Breemen, R.B., et al., 2009. Vitamin A equivalency of beta-carotene in healthy adults: Limitation of the extrinsic dual-isotope dilution technique to measure matrix effect. British Journal of Nutrition 101, 1837–1845.

Vitamin A Tracer Task Force (2004) Appropriate uses of vitamin A tracer (stable isotope) methodology. USAID/HarvestPlus/IAEA/ILSI. http://www-naweb.iaea.org/nahu/nahres/pdf_files/vitatracer.pdf

Vitamin A Tracer Task Force (2005) *Handbook on Vitamin A Tracer Dilution Methods to Assess Status and Evaluate Intervention Programs*. HarvestPlus/IAEA/USAID. http://www-naweb.iaea.org/nahu/nahres/pdf_files/handbook.pdf

Vitamin A Tracer Task Force (2007) Analysis of stable isotope data to estimate vitamin A body stores. IAEA/HarvestPlus/USAID. http://www-pub.iaea.org/MTCD/publications/PubDetails.asp?pubId=7804

World Health Organization (WHO), Food and Agriculture Organization of the United Nations (FAO) (2006) In: Allen L, De Benoist B, Dary O, and Hurrell RF (eds.) *Guidelines on Food Fortification with Micronutrients*. http://www.who.int/nutrition/publications/guide_food_fortification_micronutrients.pdf

Relevant Websites

http://www.iaea.org/NewsCenter/Focus/Humanhealth/index.html. – International Atomic Energy Agency (IAEA).

Biotin

Donald M. Mock[a] **and Luis Fontana**[b,c,d], [a] Departments of Biochemistry and Molecular Biology and Pediatrics, Biomedical Research Building I, University of Arkansas for Medical Sciences, Little Rock, AR, United States; [b] Department of Biochemistry and Molecular Biology 2, School of Pharmacy, Campus de Cartuja s/n, Granada, Spain; [c] Institute of Nutrition and Food Technology "José Mataix", Biomedical Research Center, Parque Tecnológico Ciencias de la Salud, Avda. del Conocimiento s/n, Granada, Spain; and [d] Instituto de Investigación Biosanitaria ibs.GRANADA, Complejo Hospitalario Universitario de Granada, Granada, Spain

© 2023 Elsevier Ltd. All rights reserved.

This is an update of D.M. Mock, Biotin: Physiology, Dietary Sources, and Requirements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 182–190, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00026-X.

Introduction 85
Biotin is a vitamin 86
Digestion of protein-bound biotin 86
Intestinal absorption 86
The contribution of microbial biotin to absorbed biotin 86
Transport from the intestine to peripheral tissues 87
Biotin uptake by liver and most peripheral tissues 87
Biotin transport 87
Transport of biotin into the central nervous system 87
Placental transport of biotin 87
Transport of biotin into human milk 87
Metabolism and urinary excretion of biotin and metabolites 87
Biliary excretion of biotin and metabolites 89
Metabolic functions 89
Biotin deficiency diseases 90
HCS deficiency 90
Biotinidase deficiency 91
Role of biotin in gene expression 91
Assessment of biotin status 92
Measurement of biotin 92
Laboratory findings of biotin deficiency 92
Requirements 92
Dietary sources, deficiency, and high intakes 93
Dietary sources 93
Circumstances leading to deficiency 93
Clinical findings of frank deficiency 94
Pharmacology 94
Conclusions 94
References 94

Key points

- To obtain an overview of the role of biotin in intermediary metabolism.
- To understand the relationship between the chemical structure of this vitamin and its biological function.
- To know the processes of absorption, transport, excretion and metabolism of this vitamin.
- To have information about the dietary intakes of this vitamin.
- To know the metabolic disorders related to biotin.
- To know the most important dietary sources and the requirements of this vitamin.

Introduction

Biotin is known as vitamin B_7 (Combs and McClung, 2016; Rodwell et al., 2018). As other vitamins of the B complex, it is water-soluble and widely distributed in food, is not particularly stored in the organism and does not usually produce toxicity due to

overdosage. Together with thiamine, riboflavin, niacin, pantothenic acid and pyridoxine, biotin shares the participation as coenzymes in the metabolism of macronutrients (Stipanuk and Caudill, 2018).

This article will consider the metabolic functions, the main food sources, the requirements, the deficiencies, the processes of absorption, transport, excretion and metabolism, and the assessment of the nutritional status of biotin.

Biotin is a vitamin

Biotin (also known as Vitamin B_7, Vitamin H, and Coenzyme R) is a water-soluble vitamin that is generally classified in the B complex group. Mammals (including humans) cannot synthesize biotin and depend on biotin synthesized by microbes and plants (Combs and McClung, 2016; Rodwell et al., 2018). Biotin was discovered in nutritional experiments that demonstrated a factor in many foodstuffs capable of curing the scaly dermatitis, hair loss, and neurologic signs induced in rats fed dried egg white.

Avidin, a glycoprotein found in egg white, binds biotin very specifically and tightly. From an evolutionary standpoint, avidin probably serves as a bacteriostat in egg white; consistent with this hypothesis is the observation that avidin is resistant to a broad range of bacterial proteases in both the free and biotin-bound form. Because avidin is also resistant to pancreatic proteases, ingested avidin binds to dietary biotin (and probably any biotin from intestinal microbes) and prevents absorption, carrying the biotin on through the GI tract. Cooking denatures avidin, rendering avidin susceptible to digestion and unable to interfere with absorption of biotin (Bistas et al., 2021).

Digestion of protein-bound biotin

The content of free biotin and protein-bound biotin in foods is variable, but the majority of biotin in meats and cereals appears to be protein-bound via an amide bond between biotin and lysine. The determinants of biotin bioavailability have been not clearly delineated. Biotin release from covalent protein binding is likely mediated by a specific biotin-amide hydrolase, biotinidase (EC 3.5.1.12). Biotinidase mRNA is present in pancreas and intestinal mucosa. Biotinidase also similarly plays a critical role in intracellular recycling of biotin by releasing biotin from intracellular proteins during protein turnover.

Intestinal absorption

At physiologic pH, the carboxylate group of biotin is negatively charged. Thus, biotin is at least modestly water-soluble and requires a transporter to cross cell membranes. Biotin transport must occur across two structurally and functionally different membrane domains of human intestinal epithelial cells: the brush border (apical) membrane that faces the intestinal lumen and the basolateral membrane that faces the interstitium in contact with blood that perfuses the intestine.

A biotin transporter is present in each of the membrane domains. In the brush border membrane, transport occurs via a Na^+-dependent, electroneutral, carrier-mediated mechanism that saturates at the micromolar range, accounting for the overall limitation in nondiffusion transport. In the presence of a Na^+ gradient, biotin transport occurs against a concentration gradient. This biotin transporter can also transport pantothenic acid and lipoic acid and hence has been named Sodium-dependent MultiVitamin Transporter (SMVT). Human SMVT is the product of the *SLC5A6* gene, which is located on chromosome 2p23 and consists of 17 exons. SMVT is exclusively targeted to the brush border membrane (Sabui et al., 2016).

Biotin transport across the basolateral membrane is also a carrier-mediated mechanism. However, this carrier is Na^+-independent, electrogenic, and cannot accumulate biotin against a concentration gradient.

The intestinal biotin transport upregulates in response to biotin deficiency. Upregulation likely is accomplished primarily by induction of SMVT mRNA synthesis and, hence, an increased number of SMVT transporters per cell. The increase in SMVT is likely mediated by an induction in the activity of P1, which is one of the two promoter regions upstream from the SMVT gene.

Based on a study in which biotin was administered orally in pharmacologic amounts, bioavailability of biotin is approximately 100%. Thus, both physiologic intakes and the pharmacologic doses of biotin given to treat biotin-dependent inborn errors of metabolism are likely to be well absorbed.

The contribution of microbial biotin to absorbed biotin

The contribution of microbial biotin to absorbed biotin, if any, remains unknown. Biotin is synthesized by many intestinal microbes. Based on rat studies, carrier-mediated transport of biotin is most active in the proximal small bowel, where microbes are the least numerous. However, biotin absorption from the proximal colon where microbes are the most numerous is still significant, supporting the potential nutritional significance of biotin synthesized and released by enteric flora.

Transport from the intestine to peripheral tissues

Biotin concentrations in plasma are small relative to other water-soluble vitamins. Most biotin in plasma is free and dissolved in the aqueous phase of plasma. However, approximately 7% is reversibly bound to plasma protein, and approximately 12% is covalently bound to plasma protein. Binding to human serum albumin likely accounts for reversible binding. Biotinidase has been proposed as a biotin-binding protein or biotin-carrier protein for transport into cells. A biotin-binding plasma glycoprotein has been observed in pregnant rats. Although the importance of protein binding in the transport of biotin from the intestine to the peripheral tissues is not yet clear, the immunoneutralization of this protein led to decreased transport of biotin to a fetus and early death of the embryo.

Biotin uptake by liver and most peripheral tissues

SMVT is widely expressed in human tissues. Studies by Said and coworkers provide strong evidence that biotin uptake by liver (and likely many other somatic tissues) occurs via SMVT. Metabolic trapping, (e.g., biotin bound covalently to intracellular proteins) is also important.

Studies by Zempleni and coworkers provide evidence in favor of monocarboxylate transporter 1 (MCT1) as the lymphocyte biotin transporter. MCT1 may also be responsible for biotin transport in keratinocytes.

A child with biotin-responsive neurologic problems and a pattern of organic aciduria consistent with multiple carboxylase deficiency has been reported. An autosomal recessive defect in lymphocyte biotin transport was identified. SMVT gene sequence was normal. These investigators speculated that lymphocyte biotin transporter is expressed in other tissues and mediates some critical aspect of tissue biotin homeostasis.

Biotin transport

Transport of biotin into the central nervous system

Biotin is transported across the blood–brain barrier. The transporter is saturable and structurally specific for the free carboxylate group on the valeric acid side chain. Transport into the neuron also appears to involve a specific transport system with subsequent trapping of biotin by covalent binding to brain proteins, presumably carboxylases.

Ozand and collaborators have described several patients in Saudi Arabia with biotin-responsive basal ganglia disease. Symptoms include confusion, lethargy, vomiting, seizures, dystonia, dysarthria, dysphagia, seventh nerve paralysis, quadriparesis, ataxia, hypertension, chorea, and coma. Gusella and coworkers provided evidence of a genetic defect in SLC19A3 and speculated that this might be a biotin transporter. However, Said and coworkers conclusively proved that SLC19A3 codes for THTR2, a thiamine transporter located in the apical membrane of intestinal, renal tubule, and hepatic cells. THTR2 does not transport biotin, leaving the biotin responsiveness of these patients unexplained.

Placental transport of biotin

Biotin concentrations are 3–17-fold greater in plasma from human fetuses compared to their mothers in the second trimester, consistent with active placental transport. SMVT is likely responsible for placental biotin transport based on the observations that SMVT is expressed in normal human placenta and that the microvillus membrane of the placenta contains a biotin transport system that is saturable, Na^+-dependent and actively accumulates biotin. However, in the isolated, perfused single cotyledon from placenta, transport of biotin across the placenta is relatively weak, potentially allowing greater fetal deficiency than maternal deficiency. Indeed, in mice, the degree of fetal biotin deficiency is substantially greater than maternal biotin deficiency.

Transport of biotin into human milk

More than 95% of the biotin in human milk is free in the skim fraction. The concentration of biotin varies substantially in some women and exceeds the concentration in serum by 1–2 orders of magnitude, suggesting that there is a system for transport into milk. Metabolites account for more than half of the total biotin plus metabolites in early and transitional human milk. With postpartum maturation, the biotin concentration increases. No soluble biotin-binding protein has been detected in human milk.

Metabolism and urinary excretion of biotin and metabolites

Biotin is a bicyclic compound (Fig. 1). One of the rings contains an ureido group (–N–CO–N–). The tetrahydrothiophene ring contains sulfur and has a valeric acid side chain. A significant proportion of biotin undergoes catabolism before excretion (Fig. 1). Two principal pathways of biotin catabolism have been identified in mammals. In the first pathway, the valeric acid side chain of biotin is degraded by β-oxidation. β-Oxidation of biotin leads to the formation of bisnorbiotin, tetranorbiotin, and related intermediates that are known to result from β-oxidation of fatty acids. The cellular site of this β-oxidation of biotin

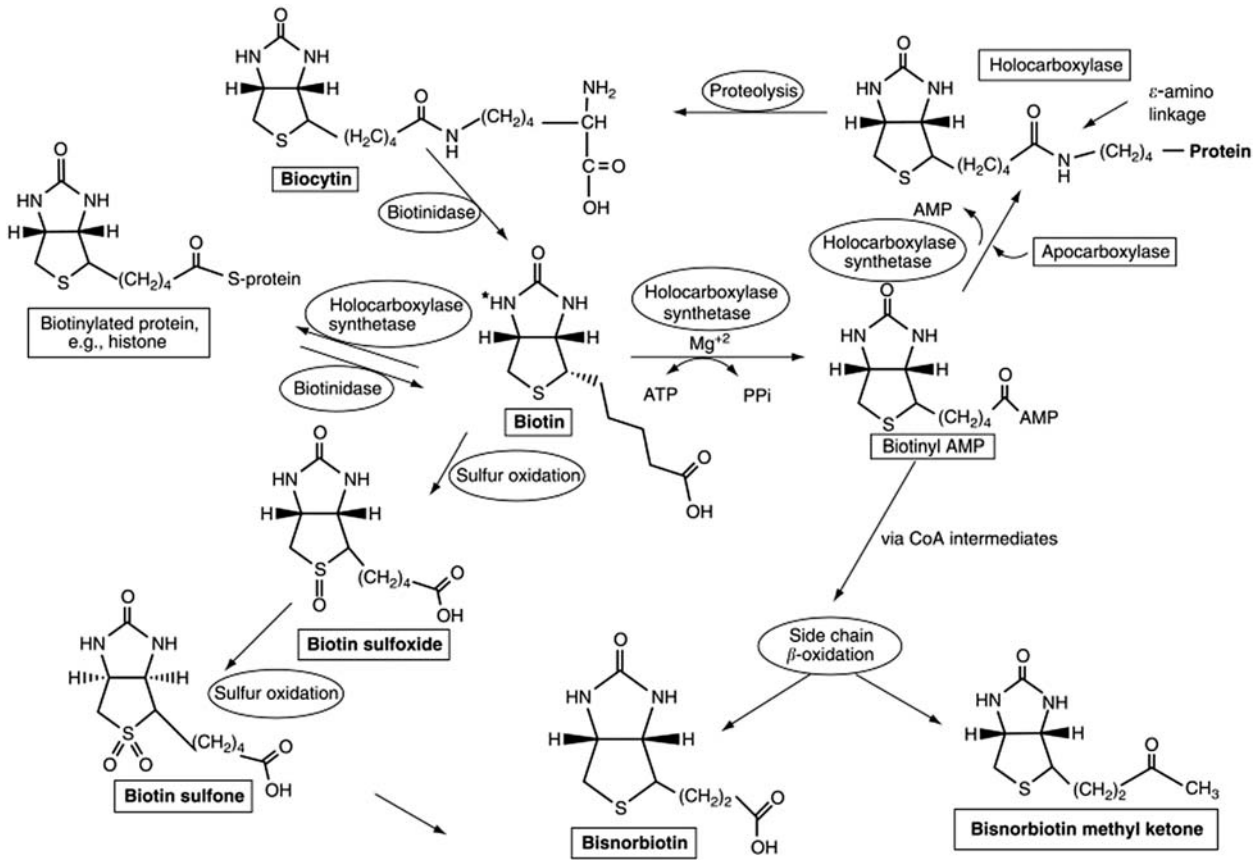


Fig. 1 Biotin metabolism and degradation. Ovals denote enzymes or enzyme systems; rectangles denote biotin, intermediates and metabolites. AMP, adenosine monophosphate; ATP, adenosine triphosphate; CoA, coenzyme A; PPi, pyrophosphate; and *, site of attachment of carboxyl moiety.

is uncertain. Spontaneous (nonenzymatic) decarboxylation of the unstable β-ketoacids (β-ketobiotin and β-ketobisnorbiotin) leads to formation of bisnorbiotin methylketone and tetranorbiotin methylketone; these catabolites appear in urine.

In the second pathway, the sulfur in the thiophane ring of biotin is oxidized leading to the formation of biotin l-sulfoxide, biotin d-sulfoxide, and biotin sulfone. Sulfur oxidation may be catalyzed by an NADPH-dependent process in the smooth endoplasmic reticulum. Combined oxidation of the ring sulfur and β-oxidation of the side chain lead to metabolites such as bisnorbiotin sulfone. In mammals, degradation of the biotin ring to release carbon dioxide and urea is quantitatively minor.

On a molar basis, biotin accounts for approximately half of the total avidin-binding substances in human serum and urine (Table 1). Biocytin, bisnorbiotin, bisnorbiotin methylketone, biotin-d,l-sulfoxide, and biotin sulfone account for most of the balance. Biotin catabolism to these inactive forms is accelerated by anticonvulsant therapy, by cigarette smoking, and during pregnancy, thereby increasing the ratio of biotin metabolites to biotin in urine and potentially contributing to an increased requirement.

Because biotin and its metabolites are small molecules ($\leq$244 Da) and mainly free in plasma, most of the biotin will pass into the glomerular filtrate. Thus, specific systems for the reabsorption of biotin from the glomerular filtrate are important to avoid substantial losses in urine. Based on work by Said and coworkers, SMVT is the principal transporter responsible for biotin reabsorption. Biotin uptake by SMVT is adaptively regulated by biotin deficiency, consistent with previous studies demonstrating reduced biotin excretion in marginal biotin deficiency induced experimentally in human subjects.

Subsequent egress of biotin from the tubular cells occurs via a basolateral membrane transport system that is not dependent on Na^+. Studies in patients with biotinidase deficiency suggest that there may be a role for biotinidase in the renal handling of biotin.

Table 1 Normal range of urinary excretion of biotin and major metabolites.

Biotin	*Bisnorbiotin*	*Biotin sulfoxide*
18–77	11–39	8–19

Results are expressed in nmol/24 h.

Biliary excretion of biotin and metabolites

Biliary excretion of biotin and metabolites is quantitatively negligible compared to urine based on a study in rats. Although the concentrations of biotin, bisnorbiotin, and biotin-d,l-sulfoxide were approximately 10-fold greater in bile than serum of pigs, the bile to serum ratio for biotin in still more than 10-fold less than the bile to serum ratio for bilirubin, which is actively excreted in bile.

Metabolic functions

In mammals, biotin serves as an essential cofactor for five carboxylases, each of which catalyzes a critical step in intermediary metabolism (Mock, 2014). All five of the mammalian carboxylases catalyze the incorporation of bicarbonate as a carboxyl group into a substrate. All five employ a similar catalytic mechanism.

Biotin is attached to the apocarboxylase by a condensation reaction catalyzed by holocarboxylase synthetase (HCS) (Fig. 1). An amide bond is formed between the carboxyl group of the valeric acid side chain of biotin and the ε-amino group of a specific lysyl residue in the apocarboxylase; these regions contain sequences of amino acids that are highly conserved for the individual carboxylases both within and between species.

In the carboxylase reaction, the carboxyl moiety is first attached to biotin at the ureido nitrogen opposite the side chain; then the carboxyl group is transferred to the substrate. Because the valeric acid side chain of biotin is coupled to the side chain of lysine in each holocarboxylase, this CO_2 is at the end of a long, flexible chain, allowing the biotinyl coenzyme to be carboxylated at one site and used as a CO_2 donor at a second site. The reaction is driven by the hydrolysis of ATP to ADP and inorganic phosphate. Subsequent reactions in the pathways of the mammalian carboxylases release carbon dioxide from the product of the carboxylase reaction. Thus, these reaction sequences rearrange the substrates into more useful intermediates but do not violate the classic observation that mammalian metabolism does not result in the net fixation of carbon dioxide.

The five biotin-dependent mammalian carboxylases (Fig. 2) are acetyl-CoA carboxylase (EC 6.4.1.2) isoforms ACC-1 and ACC-2 (formerly known as α ACC and β ACC), pyruvate carboxylase (EC 6.4.1.1, PC), methylcrotonyl-CoA carboxylase (EC 6.4.1.4, MCC), and propionyl-CoA carboxylase (EC 6.4.1.3, PCC).

Both ACC-1 and ACC-2 catalyze the incorporation of bicarbonate into acetyl CoA to form malonyl CoA (Fig. 2), but ACC-1 and ACC-2 are thought to have two very different roles in cellular metabolism; one controls fatty acid synthesis, and the other controls fatty acid oxidation. ACC-1 is located in the cytosol and produces malonyl CoA, because availability of malonyl CoA is rate limiting, activity of ACC-1 controls fatty acid synthesis (elongation) and is tightly regulated in a sophisticated fashion. Cytosolic ACC-1 exists as a very large polymer with a molecular mass in the millions of Daltons and is inactivated by dissociation into its protomer units. Citrate activates ACC-1 by increasing polymerization. CoA itself activates ACC-1 by lowering the K*m* for acetyl CoA. ACC-1 is inhibited by the products of fatty acid synthesis, the long-chain acyl CoAs, which also act to depolymerize the enzyme. In addition,

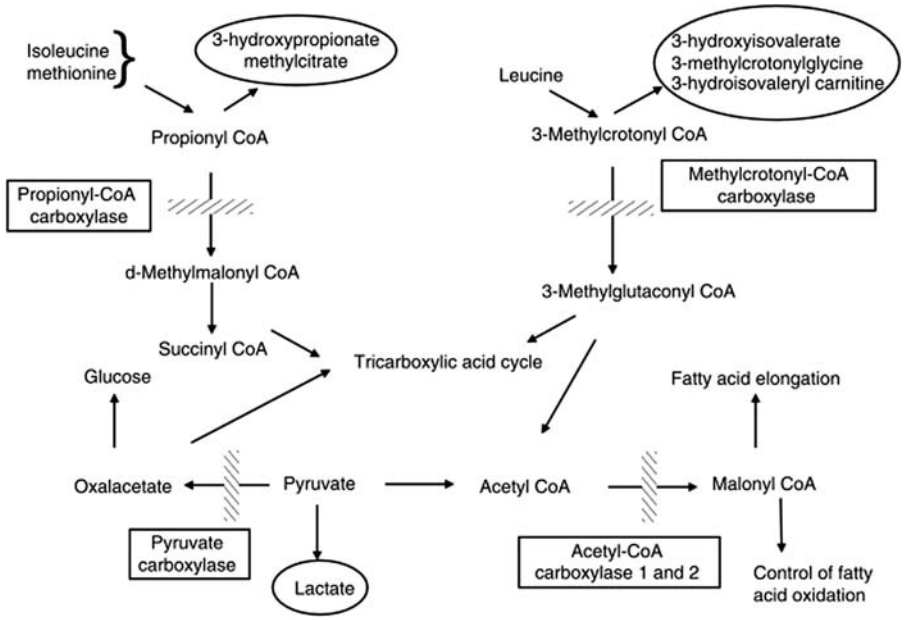


Fig. 2 Interrelationship of pathways catalyzed by biotin-dependent enzymes (shown in boxes). Organic acids and odd-chain fatty acids accumulate because biotin deficiency causes reduced activity of biotin-dependent enzymes. Hatched bars denote metabolic blocks at deficient carboxylases; ovals denote accumulation of products from alternative pathways.

ACC-1 activity is regulated by covalent modification (phosphorylation) in response to the hormones insulin and glucagon. A high insulin-to-glucagon ratio typical of the immediate postprandial state with increased blood glucose level favors dephosphorylation of ACC-1 to an active form, whereas a low insulin-to-glucagon ratio (typical of fasting) favors phosphorylation to the inactive form. The amount of ACC-1 protein also responds to changes in dietary and hormonal conditions.

ACC-2 is located on the outer mitochondrial membrane and regulates the availability of fatty acids for oxidation through the inhibition of carnitine palmitoyltransferase I by malonyl CoA. Carnitine palmitoyltransferase I catalyzes the rate-limiting transfer of the fatty acid from CoA to carnitine; the fatty acyl carnitine is transported into the mitochondrial matrix converted back to the CoA derivative and oxidized.

The three remaining carboxylases are mitochondrial. PC catalyzes the incorporation of bicarbonate into pyruvate to form oxaloacetate, an intermediate in the Krebs tricarboxylic acid cycle (Fig. 2). Thus, PC catalyzes an anaplerotic reaction. In gluconeogenic tissues (that is, liver and kidney), the oxaloacetate can be converted to glucose. Deficiency of PC is probably the cause of the lactic acidemia, central nervous system lactic acidosis and abnormalities in glucose regulation observed in biotin deficiency and in genetic biotinidase deficiency and HCS deficiency. PC also plays a role in the formation of the protective myelin sheath that surrounds certain nerve cells and the production of neurotransmitters, both of which may contribute to the phenotype of inherited PC deficiency.

In the "A" form of inherited PC deficiency, affected infants present in the first few months of life with mild or moderate lactic acidemia and psychomotor retardation. Lactic acidemia can lead to vomiting, abdominal pain, fatigue, muscle weakness, and difficulty breathing. Most children die within the first few years, and survivors have severe mental retardation.

In the more severe "B" form of PC deficiency, the initial presentation usually occurs shortly after birth with hypotonia, seizures, coma, severe lactic acidemia, and signs of liver failure such as elevated blood ammonia concentrations. Death usually occurs before 3 months of age.

MCC catalyzes an essential step in the degradation of the branched-chain amino acid leucine (Fig. 2). MCC is composed of two non identical subunits: a biotinylated α subunit that is encoded by the gene MCCC1 and a nonbiotinylated β subunit, which is encoded by the gene MCCC2. MCC is not regulated by small molecules or by dietary or hormonal factors. Deficient activity of MCC leads to metabolism of 3-methylcrotonyl CoA to 3-hydroxyisovaleric acid, 3-hydroxyisovaleryl carnitine and 3-methylcrotonylglycine by an alternate pathway. Thus, increased urinary excretion of these abnormal metabolites reflects deficient activity of MCC. The inherited deficiency of MCC characteristically presents with recurrent episodes of vomiting, diarrhea, lethargy, hypotonia, severe metabolic acidosis, hypoglycemia, and carnitine depletion. Moderate restriction of dietary protein to limit leucine intake and carnitine supplementation to correct or prevent carnitine deficiency generally result in normal development. Some cases respond to biotin supplementation. Newborn screening of acyl carnitines has identified a much higher incidence of asymptomatic MCC deficiency than expected from the number of patients ascertained by clinical symptoms; this observation suggests that many patients may have a benign clinical course.

PCC catalyzes the incorporation of bicarbonate into propionyl CoA to form methylmalonyl CoA; methylmalonyl CoA undergoes an isomerization to succinyl CoA catalyzed by the B_{12}-dependent enzyme methylmalonyl-CoA mutase and enters the tricarboxylic acid cycle (Fig. 2). In a fashion analogous to MCC deficiency, deficiency of PCC leads to increased urinary excretion of 3-hydroxypropionic acid and 3-methylcitric acid. Sources of propionyl CoA include catabolism of the amino acids valine, isoleucine, threonine, and methionine; β-oxidation of odd-numbered or branched-chain fatty acids; byproducts of bile acid synthesis from cholesterol; and intestinal microflora. PCC is not rate limiting in the metabolism of propionyl CoA, and the enzyme activity is not sensitively regulated by allosteric effectors or by dietary or hormonal changes. PCC is composed of two nonidentical subunits: a biotinylated α subunit that is encoded by the gene PCCA and a nonbiotinylated β subunit, which is encoded by the gene PCCB.

Propionic acidemia is the disease caused by an inherited deficiency of PCC. Affected individuals have repeated, life-threatening episodes of severe ketosis and metabolic acidosis that often begin in infancy. Findings include vomiting, dehydration, and lethargy, which can progress to coma and death if not treated. Frequent neurological complications include developmental delay, seizures, and cerebral atrophy. The concentration of propionyl CoA increases proximal to the metabolic block caused by decreased PCC activity and results in increased urinary excretion of a constellation of propionate metabolites that are diagnostic of propionic acidemia. These include 3-hydroxypropionate, propionylglycine, propionylcarnitine, and methylcitrate.

The most important treatment for propionic acidemia is restriction of dietary protein, thereby limiting the amino acid precursors of propionate. Use of special formulas that have very low levels of isoleucine, valine, methionine, and threonine has the same goal. The minimum requirement for these essential amino acids is met by the addition of other proteins after calculation of the content of each amino acid. Loss of propionylcarnitine can lead to a secondary deficiency of carnitine. Therapy includes dietary carnitine to prevent carnitine deficiency; this promotes the conversion of propionyl CoA to propionylcarnitine, which helps restore free CoA concentrations and facilitates excretion of propionate.

Biotin deficiency diseases

HCS deficiency

Genetic deficiencies of HCS and biotinidase cause the two types of multiple carboxylase deficiency that were previously designated the neonatal and juvenile forms. The inherited deficiency of HCS activity results in decreased activities of all five of the biotin-dependent carboxylases. In turn, these multiple carboxylase deficiencies result in clinical findings arising from the roles of all

five carboxylases in metabolism. In patients with a severe HCS deficiency, illness often occurs in the neonatal period and includes severe ketoacidosis, seizures, and lethargy, if not recognized and treated, coma and death can ensue. In patients who have a milder form of HCS deficiency, hair loss (alopecia), and an erythematous skin rash typical of biotin deficiency can appear at several months of age. Elevated urinary excretions of the metabolites characteristic of deficiency of several of the biotin-dependent carboxylases deficiencies are seen (Saleem and Soos, 2021).

Treatment with large oral doses of biotin (e.g., 10–60 mg/day), usually gives dramatic improvement of the biochemical abnormalities, skin rash, alopecia, and neurological findings, provided irreversible neurological damage has not occurred. A variety of mutations of the HCS gene have been reported. When studied, the concentration of biotin needed to attain half of the maximal reaction rate (the K*m*) generally was found to be increased far above biotin levels found in normal cells. On the contrary, the maximal enzyme activity (V*max*) usually was importantly greater than zero and approached normal. These observations explain why treatment with very large doses of biotin that increased tissue levels of biotin far above normal can result in enough active of HCS to convert enough apocarboxylases to active holocarboxylases, correcting the multiple carboxylase deficiencies. A complete absence of HCS activity, which would mean no activity of any of the five carboxylases, would probably be fatal *in utero*.

Biotinidase deficiency

In the normal turnover of cellular proteins, holocarboxylases are degraded to biocytin or biotin linked to an oligopeptide containing at most a few amino acid residues (**Fig. 1**). Biotinidase similarly plays a critical role by releasing biotin for recycling from intracellular proteins such as carboxylases and histones during protein turnover. Consistent with this global role, biotinidase is found in many tissues including heart, brain, liver, lung, skeletal muscle, kidney, plasma, and placenta in addition to pancreas and intestine. The liver is thought to be the source of serum biotinidase.

Individuals with less than 10% of normal activity in serum exhibit seizures, hypotonia, skin rash, and alopecia, usually presenting in infancy. Many children have ataxia, developmental delay, conjunctivitis, hearing loss, and visual problems, including optic atrophy as well as a characteristic organic aciduria. If untreated, some progress to coma or death. Some only manifest one or two features, or present later in life with motor limb weakness, spastic paresis, and eye problems, such as loss of visual acuity and scotomata. Once hearing loss, optic atrophy, and moderate or severe developmental delay appear, they are often irreversible despite treatment with biotin. If treatment is begun before onset of clinical findings, signs and symptoms appear to be preventable (Saleem and Simpson, 2021).

Thus, the clinical findings and biochemical abnormalities of biotinidase deficiency resemble those of biotin deficiency (dermatitis, alopecia, conjunctivitis, ataxia, and developmental delay) suggesting that they are caused by biotin deficiency. However, the signs and symptoms of biotin deficiency and biotinidase deficiency are not identical. Seizures, irreversible neurosensory hearing loss, and optic atrophy have been observed in biotinidase deficiency, but not in biotin deficiency. A knockout mouse model has recently been reported that recapitulates many of these findings.

Role of biotin in gene expression

Hymes and Wolf suggested that the posttranslational biotinylation of histones might play a role as a covalent modifier in the epigenetic code that regulates DNA transcription. However, subsequent work showed that <0.001% of human histones (primarily H3 and H4) are biotinylated, suggesting that the abundance is too low to elicit biological effects in vivo. However, HCS is located prominently in the nucleus, and the knockout of HCS in drosophila as well as in human and other mammalian cells in culture produces distinct phenotypes, including the derepression of long terminal repeats and chromosomal instability; aspects of this HCS knockout phenotype have been attributed to the effects on gene expression rather than reduced activities of the biotin-dependent carboxylases.

Zempleni and colleagues have proposed that the biological effects of biotin on gene expression are caused by a multiprotein complex, including proteins involved in DNA methylation, histone methylation, and histone deacetylation. They proposed that the docking of HCS in chromatin causes an occasional biotinylation of histones ("marks") near the various HCS binding sites. Their studies provide evidence that HLCS enters the nuclear compartment and is recruited to chromatin through physical interactions with DNA methyltransferase 1 and methyl CpG-binding protein 2. Chromatin-bound HCS has been shown to recruit the eukaryotic histone H3 methyltransferase euchromatic histone lysine N-methyltransferase 1, which creates abundant Lys9-methylated histone H3 gene repression marks. In addition, HLCS interacts with nuclear receptor corepressor, a protein known to facilitate the binding of histone deacetylases (HDACs) in chromatin. HDACs remove histone acetylation marks and thus play a critical role in gene repression. Overall, emerging data suggest histone biotinylation marks are a side effect of HCS being in close physical proximity to histones and play no direct role in gene repression, despite contributing toward chromatin condensation.

Biotin status affects gene expression (León-Del-Río, 2018). The participation of biotin in gene expression was derived from two different observations. First, biotin deprivation in different experimental models was shown to affect the transcription or enzymatic activity of hepatic enzymes including glucokinase, pyruvate kinase, 6-phosphofructokinase, and ornithine transcarbamylase. Second, in cells in culture and rat tissues biotin was shown to have a nitric oxide (NO)-like function capable of increasing the intracellular concentration of the second messenger cGMP through the activation of the soluble form of the enzyme guanylate cyclase

(sGC). In human hepatocytes, biotin was found to regulate the expression of the asialoglycoprotein receptor (ASGR) through the activation of sGC and the cGMP-dependent protein kinase (PKG).

The sGC-PKG signal transduction pathway was found to regulate the transcription of genes encoding PC, PCC, ACC1, SMVT, and HCS in human hepatocytes. Interestingly, fibroblasts from a patient carrying a homozygous mutation in HCS, R508W, which increases the Km of the enzyme for biotin 370-fold, required a 100-fold increase in biotin concentration, compared to normal fibroblasts, to increase HCS mRNA levels to within normal values.

These results demonstrated that biotinyl-AMP, the product of HCS, is the transcriptionally active form of biotin and indicates that the HCS-sGC-PKG pathway regulates transcription of the biotin-cycle enzymes.

The study of biotin-deficient rats showed that the mRNA levels of enzymes of the biotin cycle are down-regulated during biotin deficiency in the liver and kidney, but remain constitutively expressed in the brain.

Assessment of biotin status

Measurement of biotin

For measuring biotin at physiological concentrations (that is, from 100 pmol L^{-1} to 100 nmol L^{-1}), most recent studies have used an avidin-binding assay to evaluate biotin status. Avidin-binding assays generally detect all avidin-binding substances after chromatographic separation of biotin analogs. This method appears to be both sensitive and chemically specific.

Laboratory findings of biotin deficiency

In humans, laboratory indicators of biotin deficiency have been validated in studies in which progressive, but asymptomatic biotin deficiency was induced experimentally by feeding diets high in egg white. The urinary excretion of biotin declines dramatically, reaching frankly abnormal values in approximately 90% of subjects after 3 weeks. Urinary excretion of 3-hydroxyisovaleric acid and plasma and urinary 3-hydroxyisovalerylcarnitine increase to greater than the normal range in approximately 90% of subjects after 2 weeks of egg-white feeding, providing evidence that biotin depletion decreases the activity of MCC. The most sensitive indicator of biotin status appears to be activity of PCC in lymphocytes isolated from venous blood samples. Unfortunately, this assay is technically demanding, and the blood samples require special handling and storage.

Serum concentrations of free biotin decrease to abnormal values in less than half of the subjects. This observation is consistent with the impression of many investigators in this field that blood biotin concentration is not an early or sensitive indicator of impaired biotin status.

Requirements

Data providing an accurate estimate of the dietary and parenteral biotin requirements for infants, children, and adults are lacking. However, recommendations for biotin intakes have been formulated by both the European Food Safety Authority (EFSA) and the Institute of Medicine (USA) for preterm infants, term infants, children, and adults (**Tables 2** and **3**).

The EFSA considers that the available data are insufficient to derive Adequate Requirements (ARs) and Population Reference Intakes (PRIs) for biotin, and therefore proposes to set an Adequate Intake (AI) for all population groups (European Food Safety Authority, 2017). The setting of an AI for biotin is based on observed biotin intakes with a mixed diet and the apparent absence of signs of deficiency in the EU, suggesting that current intake levels are adequate. There is no indication that the AI should differ according to sex. The AI for adults also applies to pregnant women. For lactating women, an increment in the adult AI is proposed, in order to compensate for biotin losses through secretion of breast milk. An AI is also proposed for infants aged 7–11 months based on extrapolation from the estimated intake of infants aged 0–6 months using allometric scaling, and for children and adolescents based on observed intakes in the EU (**Table 2**).

Table 2 Adequate Intakes (AIs) for biotin.

Age	*AI (μg/day)*
7–11 months	6
1–3 years	20
4–10 years	25
11–17 years	35
≥18 years[a]	40
Lactation	45

[a]Including pregnancy.

Taken from European Food Safety Authority (EFSA) (2017).

Table 3 Adequate Intakes (AIs) for biotin by life stage group.

Age	*AI (μg/day)*
Life stage group[a]	
0–6 months	5
7–12 months	6
1–3 years	8
4–8 years	12
9–13 years	20
14–18 years	25
19–30 years	30
31–50 years	30
51–70 years	30
>70 years	30
Pregnancy	
≤18 years	30
19–50 years	30
Lactation	
≤18 years	35
19–50 years	35

[a]All groups except pregnancy and lactation represent males and females.
Taken from Institute of Medicine (2006).

The Institute of Medicine also provides AIs, in this case based on extrapolation from the amount of biotin in human milk (**Table 3**) (Institute of Medicine, 2006).

Dietary sources, deficiency, and high intakes

Dietary sources

There is no published evidence that biotin can be synthesized by mammals; thus, the higher animals must derive biotin from other sources. The ultimate source of biotin appears to be *de novo* synthesis by bacteria, primitive eukaryotic organisms such as yeast, molds, and algae, and some plant species.

The great majority of measurements of biotin content of foods have used bioassays. Recent publications provide evidence that the values are likely to contain substantial errors. However, some worthwhile generalizations can be made. Biotin is widely distributed in natural foodstuffs so requirements are easily met with a balanced diet, but the absolute content of even the richest sources is low when compared to the content of most other water-soluble vitamins.

Foods that contain the most biotin include organ meats, eggs, fish, meat, seeds, nuts, and certain vegetables (such as sweet potatoes). The biotin content of food can vary; for example, plant variety and season can affect the biotin content of cereal grains, and certain processing techniques (e.g., canning) can reduce the biotin content of foods.

Circumstances leading to deficiency

That normal humans have a requirement for biotin has been clearly documented in three situations: prolonged consumption of raw egg white, parenteral nutrition without biotin supplementation in patients with short bowel syndrome, and infant feeding with an elemental formula devoid of biotin. Because biotin could not legally be added as a supplement to infant formulas in Japan until 2003, all reports related to infant formula have come from Japan. Often feeding of an elemental formula was required to treat intractable, chronic diarrhea. The infants typically developed both the classic cutaneous manifestations of biotin deficiency and the characteristic pattern of organic aciduria.

Based on lymphocyte carboxylase activities and plasma biotin levels, some children with severe protein-energy malnutrition are biotin deficient. Investigators have speculated that the effects of biotin deficiency may be responsible for part of the clinical syndrome of protein-energy malnutrition.

Long-term anticonvulsant therapy in adults can lead to biotin depletion severe enough to interfere with leucine metabolism and cause increased urinary excretion of 3-hydroxyisovaleric acid. The mechanism of biotin depletion during anticonvulsant therapy is not known, but may involve accelerated biotin catabolism, impaired biotin absorption, impaired biotin transport in plasma, impaired renal reclamation biotin, or a combination of these.

Recent studies of biotin status during pregnancy and of biotin supplementation during pregnancy provide evidence that a marginal degree of biotin deficiency develops in at least one-third of women during normal pregnancy. Although the degree of biotin deficiency is not severe enough to produce overt manifestations of biotin deficiency, the deficiency is severe enough to

produce metabolic derangements. A similar marginal degree of biotin deficiency causes high rates of fetal malformations in some mammals. Moreover, data from a multivitamin supplementation study provide significant, albeit indirect, evidence that the marginal degree of biotin deficiency that occurs spontaneously in normal human gestation is teratogenic.

Biotin deficiency has also been reported or inferred in several other clinical circumstances including Leiner's disease, sudden infant death syndrome, renal dialysis, gastrointestinal diseases, and alcoholism.

Clinical findings of frank deficiency

The clinical findings of frank biotin deficiency in adults, older children and infants are similar. Typically, the findings appear gradually after weeks to several years of egg-white feeding or parenteral nutrition. Thinning of hair and progression to loss of all hair including eyebrows and lashes has been reported. A scaly (seborrheic), red (eczematous) skin rash was present in the majority; in several, the rash was distributed around the eyes, nose, mouth, and perineal orifices. These cutaneous manifestations, in conjunction with an unusual distribution of facial fat, have been dubbed "biotin deficiency facies." Depression, lethargy, hallucinations, and paresthesia of the extremities were prominent neurologic symptoms in the majority of adults. The most striking neurologic findings in infants were hypotonia, lethargy, and developmental delay.

The clinical response to administration of biotin has been dramatic in all well-documented cases of biotin deficiency. Healing of the rash was striking within a few weeks, and growth of healthy hair was generally present by 1–2 months. Hypotonia, lethargy, and depression generally resolved within 1–2 weeks, followed by accelerated mental and motor development in infants. Pharmacological doses of biotin (e.g., 1–10 mg) have been used to treat most patients.

Pharmacology

Mounting reports of biotin deficiency in commercial animals and humans have led to several studies of plasma levels, pharmacokinetics, and bioavailability after acute or chronic oral, intramuscular, or intravenous administration of biotin in cattle, swine, fish, and humans (Mock, 2017). High doses (e.g., 1200 mg) result in high biotin concentrations in blood and the urinary excretion of a large proportion as the unchanged biotin. Increased blood concentrations of bisnorbiotin and biotin sulfoxide and increased urinary excretion of bisnorbiotin and biotin sulfoxide are also reported, providing evidence that the biotin metabolites originate from human tissues rather than enteric bacteria.

Daily doses up to 200 mg orally and up to 20 mg intravenously have been given to treat biotin-responsive inborn errors of metabolism and acquired biotin deficiency. Toxicity has not been reported.

In recent years, dietary intervention using vitamins has been proposed as an alternative therapeutic approach for progressive multiple sclerosis (MS), including biotin. However, in 2017 the use of biotin to treat MS patients received a major setback, when the Committee for Medicinal Products for Human Use of the European Medicines Agency, after reviewing the available evidence, considered that the data were not robust enough and concluded that the observed benefits of high-dose biotin (HDB) did not outweigh its potential risks. A very recent systematic review and meta-analysis of randomized controlled trials evaluated the administration of HDB (at least 300 mg/day administered orally and for at least 3 months) to multiple sclerosis patients. According to this study, a moderate certainty of evidence suggests a potential benefit in favor of HDB administered for 12–15 months in terms of 25-foot walk time (ITW25) in patients with progressive multiple sclerosis. However, an important trade-off of this benefit is the high certainty of evidence suggesting an increased incidence of laboratory test interference when HDB is taken (Espiritu and Remalante-Rayco, 2021).

Conclusions

Among all the vitamins with coenzymatic functions in metabolism, biotin is the only one that behaves as a coenzyme without the need for structural modifications. Bound by covalent bonds to enzymatic proteins, it is always involved in carboxylation reactions that affect the functioning of the Krebs cycle, lipogenesis and the degradation of some amino acids. Nutritional deficiencies in biotin are very rare, because it is found in most foods and is synthesized by the intestinal microbiota.

References

Bistas, K.G., Tadi, P., 2021. Biotin. In: StatPearls. StatPearls Publishing, Treasure Island (FL). Last update 2021 Feb 12.

Combs Jr., G.F., McClung, J.P., 2016. The Vitamins. Fundamental Aspects in Nutrition and Health, fifth ed. Academic Press, London.

Espiritu, A.I., Remalante-Rayco, P.P.M., 2021. High-dose biotin for multiple sclerosis: a systematic review and meta-analyses of randomized controlled trials. Mult. Scler. Relat. Disord. 55, 103159. https://doi.org/10.1016/j.msard.2021.103159.

European Food Safety Authority (EFSA), 2017. Dietary Reference Values for Nutrients. Summary report. https://doi.org/10.2903/sp.efsa.2017.e15121.

Institute of Medicine, 2006. Dietary Reference Intakes: The Essential Guide to Nutrient Requirements. The National Academies Press, Washington, DC. https://doi.org/10.17226/11537.

León-Del-Río, A., 2018. Biotin in metabolism, gene expression, and human disease. J. Inherit. Metab. Dis. 42, 647–654. https://doi.org/10.1002/jimd.12073.

Mock, D.M., 2014. Biotin. In: Ross, A.C., Caballero, B., Cousins, R.J., Tucker, K.L., Ziegler, T.R. (Eds.), Modern Nutrition in Health and Disease, eleventh ed. Wolters Kluwer Health, Lippincott Williams & Wilkins, Philadelphia.
Mock, D.M., 2017. Biotin: from nutrition to therapeutics. J. Nutr. 147, 1487–1492. https://doi.org/10.3945/jn.116.238956.
Rodwell, V.W., Bender, D.A., Botham, K.M., Kennelly, P.J., Weil, P.A., 2018. Harper's Illustrated Biochemistry, 31st ed. McGraw-Hill Education, New York.
Sabui, S., Bohl, J.A., Kapadia, R., Cogburn, K., Ghosal, A., Lambrecht, N.W., Said, H.M., 2016. Role of the sodium-dependent multivitamin transporter (SMVT) in the maintenance of intestinal mucosal integrity. Am. J. Physiol. Gastrointest. Liver Physiol. 311, G561–G570. https://doi.org/10.1152/ajpgi.00240.2016.
Saleem, H., Simpson, B., 2021. Biotinidase deficiency. In: StatPearls. StatPearls Publishing, Treasure Island (FL). Last update 2021 Aug 14.
Saleem, F., Soos, M.P., 2021. Biotin deficiency. In: StatPearls. StatPearls Publishing, Treasure Island (FL). Last update 2021 Apr 14.
Stipanuk, M.H., Caudill, M.A., 2018. Biochemical, Physiological, and Molecular Aspects of Human Nutrition, fourth ed. Saunders, Philadelphia.

Caffeine

Rafael Urrialde, Complutense University of Madrid, San Pablo CEU University of Madrid, Iberoamerican Nutrition Foundation, Granada, Spain

© 2023 Elsevier Ltd. All rights reserved.

This is an update of D.P. Evatt, R.R. Griffiths, Caffeine, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 221–227, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00033-7.

Introduction 96
General concepts 96
Caffeine pharmacokinetics 97
Mechanisms of action 97
Physiological and health effects of caffeine 98
Caffeine and cognitive performance 98
Caffeine and physical exercise 99
Caffeine genetics 99
Caffeine subjective effects 99
Caffeine reinforcement 100
Caffeine withdrawal 100
Caffeine tolerance 101
Caffeine intoxication 102
Caffeine-induced sleep disorder 102
Caffeine-induced anxiety disorder 102
Caffeine dependence 102
Caffeine and food 103
Conclusion 103
Acknowledgments 104
References 104
Further reading 104

Introduction

Caffeine (1, 3, 7-trimethylxanthine) in food and beverages, an adenosine antagonist, is the most widely used mood-altering drug in the world. Caffeine is rapidly absorbed and distributed throughout the body with peak plasma concentrations typically reached 30–45 min after ingestion. The average half-life of caffeine is 4–6 h. Genetics account for some of the variability in responses to caffeine and individual differences in caffeine pharmacokinetics.

Many people around the world consume it every day to increase wakefulness, alleviate fatigue including to have a good physical activity, and improve concentration and focus for having an optimal cognitive performance.

It is also important to say caffeine is a drug, and with any drug, the withdrawals are hard. Food Drug Administration approve a statement in 2018 warning companies to stop selling dangerous and illegal pure and highly concentrated caffeine products.

General concepts

Caffeine is the most widely used psychoactive drug in the world. In the US, an estimated 87% of the population regularly consume beverages containing caffeine. Mean caffeine consumption among all users is 193 mg per day, with the highest intake among men aged 35–54 who consume an average of 336 mg of caffeine per day.

Caffeine is a natural constituent of more than 60 species of plants, including coffee, tea, cola nut, cacao, yerba maté, and guarana. Caffeine (1,3,7-trimethylxanthine) is a member of the methylxanthine class of alkaloids that include theobromine and theophylline. In its free base form, caffeine is a bitter white powder that is moderately soluble in water (21.7 mg mL^{-1}). Worldwide, caffeine is most commonly consumed as coffee and tea. Consumption of beverages with added caffeine (i.e., soft drinks) has markedly increased over the past half century, with consumption volume of soft drinks now being approximately twice that for coffee in the US. A notable recent trend has been the increasing popularity of "energy drinks" which vary considerably in the amount of caffeine (from 30 mg to over 75 mg of caffeine per 100 mL). Hundreds of such products are now marketed in the US (Table 1).

Table 1 Update caffeine content of some beverages.

Substance	*Typical serving size (volume)*	*Typical caffeine content (mg)*	*Caffeine content range (mg)*
Coffee			
Brewed/drip	12 oz[a]	200	108–420
Starbucks hot brewed coffee	16 oz	330	
Espresso	1 oz	70	60–95
Starbucks espresso (solo)	1 oz	75	
Instant	6 oz	70	20–130
Decaffeinated	12 oz	8	0–10
Tea			
Brewed	6 oz	40	30–90
Instant	6 oz	30	10–35
Can or bottle (typical)	12 oz	20	8–32
Arizona brand iced black tea	20 oz	37.5	
Arizona brand iced green tea	20 oz	18.75	
Soft drinks (typical)			
Typical caffeinated soft drink	12 oz	40	22–69
Coke classic	12 oz	35	
Diet coke	12 oz	47	
Coke zero sugar	12 oz	34	
Pepsi–cola	12 oz	38	
Diet pepsi	12 oz	36	
Mountain dew	12 oz	55	
Barq's root beer	12 oz	23	
Energy drinks (typical)			
Typical energy drink	Varies	Varies	50–375[b]
Red bull	8.3 oz	80	
Red bull free sugar	8.3 oz	80	
Rockstar	16 oz	160	
Monster	16 oz	151	
Monster zero sugar	16 oz	142	
Wired X344	16 oz	344	

[a]Historically, a "typical" serving size of coffee has been considered to be 6 oz. However, standard servings of coffee tend to be much larger than this. The smallest serving size offered at many fast food restaurants and coffee houses (e.g., Starbucks) is 12 oz.
[b]Specified range does not include energy shot products, which can contain as much as 500 mg caffeine per 1 oz serving.

Caffeine pharmacokinetics

Caffeine is rapidly and completely absorbed from the gastrointestinal tract after oral administration. Caffeine is readily distributed throughout the body and is found in all body fluids. Peak plasma concentrations are typically reached 30–45 min following oral ingestion. The fraction of caffeine bound to plasma protein is 10–35%.

More than 25 caffeine metabolites have been identified in humans. The primary metabolic pathways involve the cytochrome P-450 liver enzyme system which produces three demethylated active metabolites: paraxanthine, theobromine, and theophylline, accounting for 80%, 10%, and 4% of caffeine metabolism, respectively. The average half-life of caffeine is 4–6 h, with elimination rates varying by more than 10-fold across individuals. Caffeine half-life is prolonged in individuals with liver disease and during the end of pregnancy. Caffeine metabolism is inhibited by numerous compounds including oral contraceptive steroids, cimetidine, some quinoline antibiotics, fluvoxamine, mexiletine, and high doses of caffeine itself. Neonates have a markedly increased caffeine half-life (80–100 h) due to immature liver enzyme systems which are fully developed at approximately 6 months of age. Tobacco smoking increases caffeine metabolism by stimulating the cytochrome P-450 1A2 enzyme (CYP1A2), with smokers metabolizing caffeine about twice as fast as nonsmokers. Genetic variations in CYP1A2 activity are a significant determinant of caffeine metabolism. For more information on individual differences in genetics, see the Section on **Caffeine genetics**.

Mechanisms of action

The primary cellular site of action of caffeine is the adenosine receptor. Among the adenosine receptor subtypes that have been identified, A_1 and A_{2A} receptors are the preferential targets of caffeine. A_1 receptors are widely expressed throughout the brain, whereas A_{2A} receptors are concentrated in dopaminergic-rich areas, such as the striatum. Adenosine is an endogenous purine nucleoside that generally exerts inhibitory effects throughout the central and peripheral nervous system (e.g., excitatory neurotransmitter inhibition,

suppression of motor activity, and inhibition of gastric secretion). Caffeine is a nonselective competitive A_1 and A_{2A} receptor antagonist. Thus, caffeine produces a variety of central and peripheral effects that are opposite to the effects of adenosine.

Of most relevance to caffeine's central nervous system (CNS) stimulating effects, caffeine enhances dopamine activity indirectly by competitive antagonism of adenosine receptors that are co-localized and functionally interact with dopamine. Adenosine receptors can form functional receptor heteromers with dopamine receptors (i.e., A_1-D_1 and A_{2A}-D_2). There is some evidence that the motor stimulant and reinforcing effects of caffeine are mediated by dopamine. Preclinical studies demonstrate that caffeine produces behavioral effects similar to the dopamine-mediated effects of classic stimulants, such as cocaine and amphetamine. Moreover, dopamine depletion or blockade of dopamine receptors significantly impairs the motor stimulant and discriminative stimulus effects of caffeine.

Although caffeine can inhibit phosphodiesterase and increase intracellular calcium concentration, typical dietary doses of caffeine are believed to be too low to be significantly influenced by these nonadenosine mechanisms. Thus, caffeine's effects appear to be primarily mediated by direct antagonism of adenosine and indirect enhancement of brain dopamine activity. For more information on ergogenic mechanisms of action, see the Section on Caffeine and Exercise.

Physiological and health effects of caffeine

Caffeine modestly increases blood pressure but appears to have no effects on or to reduce heart rate. Hypertensive and hypertensive-prone caffeine users appear to be particularly sensitive to the pressor effects of caffeine. Caffeine produces increases in gastric acid secretion, colonic stimulation, diuresis (30% or more increased volume), respiratory stimulation, and bronchodilation. Caffeine also increases plasma epinephrine, norepinephrine, adrenocorticotropic hormone, cortisol, renin, and free fatty acids. Acute caffeine administration produces increased cerebral blood flow velocity and electroencephalography (EEG) beta power activity. In addition, caffeine has prominent sleep-disrupting effects. For more information on the sleep disrupting effects of caffeine see the Section on Caffeine-Induced Sleep Disorder.

Although studies of the association between caffeine consumption and coronary heart disease have yielded inconsistent findings, one recent investigation demonstrated an association between slow caffeine metabolism and the incidence of coronary heart disease in moderate and heavy caffeine consumers. For more information on caffeine genetics, see the Section on Caffeine Genetics. Several studies have found that moderate coffee intake is associated with decreased risk for coronary heart disease, possibly because of protective effects of antioxidant and other protective compounds in coffee.

Over the last few decades, studies have yielded inconsistent findings regarding the effects of caffeine on reproductive and perinatal outcomes. Although some investigations have not found evidence of a significant association between caffeine exposure and adverse birth outcomes, several recent studies did show a relationship. Although equivocal findings preclude definitive conclusions regarding the effects of caffeine on pregnancy, some governmental health agencies have taken a prudent stance and issued health warnings to limit the use of caffeine during pregnancy. Health Canada recommends that women of reproductive age consume no more than 300 mg of caffeine per day. The Food Standards Agency of the UK advises that pregnant women keep their daily intake of caffeine below 200 mg. European Food Safety Authority approved on May 2015 a scientific opinion on the safety of caffeine, saying: "single doses of caffeine up to 200 mg (about 3 mg/kg bw for a 70 kg adult) do not give rise to safety concerns. The same amount does not give rise to safety concerns when consumed <2 h prior to intense physical exercise under normal. Habitual caffeine consumption up to 200 mg per day by pregnant women does not give rise to safety concerns for the fetus. Single doses of caffeine and habitual caffeine intakes up to 200 mg consumed by lactating women do not give rise to safety concerns for breastfed infants. For children and adolescents, the information available is insufficient to derive a safe caffeine intake. The Panel considers that caffeine intakes of no concern derived for acute caffeine consumption by adults (3 mg/kg bw per day) may serve as a basis to derive single doses of caffeine and daily caffeine intakes of no concern for these population subgroups" (EFSA, 2015).

Caffeine and cognitive performance

Numerous investigations have examined the effects of caffeine on human cognitive performance. The most consistent finding to emerge is that caffeine restores cognitive performance that has been degraded by sleep deprivation, fatigue, or prolonged vigilance. At normal dietary doses, caffeine may improve tapping speed, reaction time, and sustained attention (vigilance), although results have been variable, and sizes of the effects are often modest. A large number of experimental studies have examined the effects of caffeine on memory, but evidence is insufficient to conclude that caffeine produces acute improvements in memory.

A significant limitation of the majority of studies that have found cognitive performance-enhancing effects of caffeine is that subjects in these studies have been regular caffeine users who were required to abstain from caffeine before testing (e.g., overnight abstinence). Thus, the observed cognitive performance-enhancing effects of caffeine in these studies may reflect restoration of deficits that are produced by caffeine withdrawal. For more information on the caffeine withdrawal syndrome, see the Section on Caffeine Withdrawal. It is important to note, however, that a few studies have found cognitive performance-enhancing effects of caffeine in nondependent caffeine users and nonusers. Some studies conclude that low doses of caffeine (32 mg), equivalent to those found in commonly consumed beverages and foods, can significantly improve performance on auditory and visual tasks, without having negative effects on mood profiles such as anxiety (Peeling and Dawson, 2007). In support of these outcomes, Smith

et al. in 1999 showed that a number of common drink choices (water, tea, coffee, or cola) supplemented with an additional low dose (40 mg) of caffeine could significantly enhance performance of a choice reaction time test, a semantic memory task, and a delayed recognition memory task compared with a decaffeinated placebo drink.

A few studies have also demonstrated cognitive performance increases in caffeine users who were not required to abstain from usual caffeine use, suggesting that complete tolerance to the cognitive performance-enhancing effects of caffeine does not occur at usual dietary doses. However, in high-dose caffeine consumers, cognitive performance-enhancement beyond withdrawal reversal is likely to be modest.

Caffeine and physical exercise

A large body of research has examined the effects of caffeine on exercise performance. Numerous well-controlled studies have found that relative to placebo, caffeine enhances performance during endurance physical exercise. Studies have also generally found that caffeine reduces ratings of perceived exhaustion or effort during physical exercise. Ergogenic effects of caffeine are typically demonstrated at doses of 3–6 mg kg^{-1}; higher doses of caffeine (e.g., 9 mg kg^{-1}) appear to exert little or no additional benefit on endurance physical exercise. There is some evidence that caffeine produces greater endurance physical exercise benefits in caffeine nonusers and in athletes who abstained from caffeine for several days before dosing. Findings from studies examining the effects of caffeine on short-term, high-intensity physical exercise performance have generally been equivocal, however a recent review suggested that caffeine can improve performance in team-sports physical exercise and power-based sports, with this effect more common in elite athletes who do not regularly consume caffeine. Supplementation with caffeine has been shown to acutely enhance many aspects of exercise, including prolonged aerobic-type activities and brief duration, high intensity exercise. Caffeine is ergogenic when consumed in doses of 3–6 mg/kg body mass. The most commonly used timing of caffeine supplementation is 60 min pre-exercise (Guest et al., 2021). Although not rigorously studied, findings are suggestive that tolerance occurs to the ergogenic effects of caffeine. Several non-independent mechanisms have been proposed for caffeine's effects on exercise physical performance, including increased fatty acid oxidation, increased availability of muscle glycogen, mobilization of intracellular calcium, increased muscle contractile force, and direct CNS effects via adenosine antagonism. Caffeine may improve cognitive and physical performance in some individuals under conditions of sleep deprivation. Caffeine at the recommended doses does not appear significantly influence hydration, and the use of caffeine in conjunction with exercise in the heat and at altitude is also well supported (Guest et al., 2021).

Caffeine genetics

Much of the variability in caffeine consumption and individual differences in response to caffeine can be accounted for by genetic factors. Findings from twin studies indicate that there may be common genetic factors underlying the use of caffeine, nicotine, and alcohol. Moreover, twins studies indicate that genetic factors may influence total caffeine consumption, heavy caffeine consumption, caffeine tolerance, caffeine withdrawal, caffeine intoxication, and caffeine-related sleep disturbances.

The CYP1A2 gene codes for the isoenzyme P-450 1A2, which is responsible for the demethylation of caffeine to paraxanthine, theobromine, and theophylline. For more information on caffeine metabolism, see the Section on Caffeine Pharmacokinetics. More than 150 CYP1A2 single-nucleotide polymorphisms have been identified. Individual variability in the pharmacokinetics of caffeine can be in large part accounted for by variations in CYP1A2 activity. Recent evidence suggests that individuals homozygous for the allele associated with slow metabolism (CYP1A2*1F) are at increased risk for nonfatal myocardial infarction associated with caffeinated coffee intake. Thus, caffeine consumption may increase risk for myocardial infarction in individuals with slow caffeine metabolism.

Genetic differences in adenosine A2A receptors have been implicated in individual differences in human caffeine responses. Variations in A2A receptor polymorphisms have been associated with caffeine sensitivity, caffeine-induced anxiety, caffeine-related sleep impairment, and caffeine consumption. One study reported evidence that a polymorphism in dopamine DRD2 receptors is associated with caffeine-induced anxiety.

Caffeine subjective effects

The qualitative subjective effects of caffeine depend on caffeine dose, individual differences in sensitivity, and degree of tolerance to caffeine. Low to moderate doses of caffeine typically produce positive subjective effects, including increased well-being, arousal, energy, alertness, concentration, motivation to work, and sociability, and decreased feelings of sleepiness or tiredness. Positive subjective effects are more likely to be reported in individuals who have undergone overnight caffeine abstinence.

At higher acute doses of caffeine (e.g., 400–800 mg), negative subjective effects of caffeine typically emerge. Negative subjective effects include increased anxiety, nervousness, jitteriness, tense negative mood, and upset stomach. Anxiogenic subjective effects are more likely to be reported in individuals with panic disorder or generalized anxiety disorder, and in nonclinical populations who

endorse high levels of anxiety sensitivity (i.e., fear of anxiety). For more information about high dose caffeine effects, see the Section on **Caffeine intoxication**.

Caffeine reinforcement

The efficacy of a substance in establishing or maintaining self-administration behavior reflects the reinforcing effects of the substance. The circumstantial evidence indicating that caffeine functions as a reinforcer is compelling. Caffeine is the most widely used mood-altering drug in the world. Regular daily consumption of pharmacologically active doses occurs in widely varying cultural and social contexts. Historically, efforts to restrict or eliminate consumption of caffeine-containing foods and beverages have been unsuccessful. Caffeine consumption occurs in a wide variety of vehicles (e.g., coffee, tea, maté soft drinks, energy drinks; chewing kola nuts). Finally, caffeine-containing beverages tend to be more popular than their caffeine-free counterparts. As an example, in 2009, in the US, the top six selling carbonated soft drink brands, and eight of the top 10 selling brands, contained added caffeine.

Numerous well-controlled experimental studies have demonstrated caffeine reinforcement in various subject populations. Across studies, approximately 40% of normal caffeine users demonstrate caffeine reinforcement. Higher rates of reinforcement have been observed among individuals with high levels of caffeine consumption or a history of drug or alcohol abuse. Caffeine can function as a reinforcer at very low doses (i.e., 25 mg per cup of coffee), but may produce avoidance at higher doses (e.g., 400 or 600 mg).

In habitual caffeine consumers, avoidance of caffeine withdrawal plays an important role in the reinforcing effects of caffeine. This relationship has been shown in retrospective questionnaire studies and in experimental studies that have used direct behavioral measures of reinforcement. For example, in one experimental study, moderate caffeine consumers who reported withdrawal symptoms (i.e., headaches and drowsiness) were more than twice as likely to show caffeine reinforcement. Other studies that have prospectively manipulated caffeine physical dependence have demonstrated that subjects were more than twice as likely to exhibit caffeine reinforcement when they were caffeine physically dependent (and thus prone to experiencing withdrawal symptoms when they abstain).

Studies using a conditioned flavor preference paradigm have provided indirect evidence of caffeine reinforcement. In these studies, caffeine abstinent subjects develop a liking and preference for caffeine-paired flavored beverages, relative to beverages paired with placebo. Further studies showed that, in subjects who were repeatedly exposed to a caffeine-paired flavored beverage, the development of liking and preference for the beverage was determined by alleviation of withdrawal symptoms. These studies suggest that conditioned flavor preferences (driven at least in part by alleviation of unpleasant withdrawal symptoms) likely play an important role in consumer preferences for caffeine-containing beverages.

Caffeine withdrawal

Cessation or reduction of daily caffeine consumption results in withdrawal symptoms in many caffeine users. Caffeine withdrawal has been well characterized in numerous rigorous experimental studies and in survey studies. Caffeine withdrawal headache, which is the hallmark feature of the caffeine withdrawal syndrome, has been the most frequently assessed withdrawal symptom. Approximately half of regular caffeine users report headache when abstaining from caffeine. The caffeine withdrawal headache, which develops gradually, is described as diffuse, throbbing, severe, and phenomenologically distinct from migraine headache. In addition to headache, other withdrawal symptoms that have been reliably observed across experimental and survey studies include: fatigue, decreased energy/activeness, decreased alertness, drowsiness, decreased contentedness, depressed mood, difficulty concentrating, irritability, foggy/not clearheaded, nausea/vomiting, and muscle stiffness/pain. Based on these observed symptoms, five primary clusters of withdrawal symptoms have been proposed: (1) headache, (2) fatigue and drowsiness, (3) dysphoric mood, depressed mood, or irritability, (4) difficulty concentrating, and (5) flu-like somatic symptoms, nausea, vomiting, or muscle pain/stiffness.

Onset of withdrawal typically occurs 12–24 h after abrupt caffeine cessation, although onset has been observed as early as 6 h and as late as 43 h after abstinence in some individuals. Typically, peak intensity of symptoms occurs 1–2 days after abstinence. The duration of caffeine withdrawal symptoms is generally 2–9 days. Re-administration of caffeine (usually within 30–60 min of onset) rapidly and often completely reverses caffeine withdrawal. **Fig. 1** shows the time-course of caffeine withdrawal from an illustrative double-blind experimental study.

The severity of caffeine withdrawal symptoms can range from mild to extreme and depends on several factors. Some caffeine users report clinically significant functional impairment associated with caffeine withdrawal (e.g., interference with work or child care activities). Studies show that clinically significant distress occurs in approximately 13% of caffeine users. A much higher rate (73%) of withdrawal-related clinically significant distress occurs in individuals meeting criteria for caffeine dependence. Severity of caffeine withdrawal is positively associated with increases in caffeine maintenance dose, such that greater withdrawal is experienced following cessation of higher maintenance doses. Caffeine withdrawal can occur after daily doses of caffeine as low as 100 mg day^{-1}. Caffeine withdrawal may also occur when lower doses of caffeine are substituted for the maintained caffeine dose. As the substituted dose of caffeine decreases, withdrawal severity increases. Nevertheless, even a small amount of substituted caffeine (e.g., 25 mg) can mitigate severity of caffeine withdrawal symptoms.

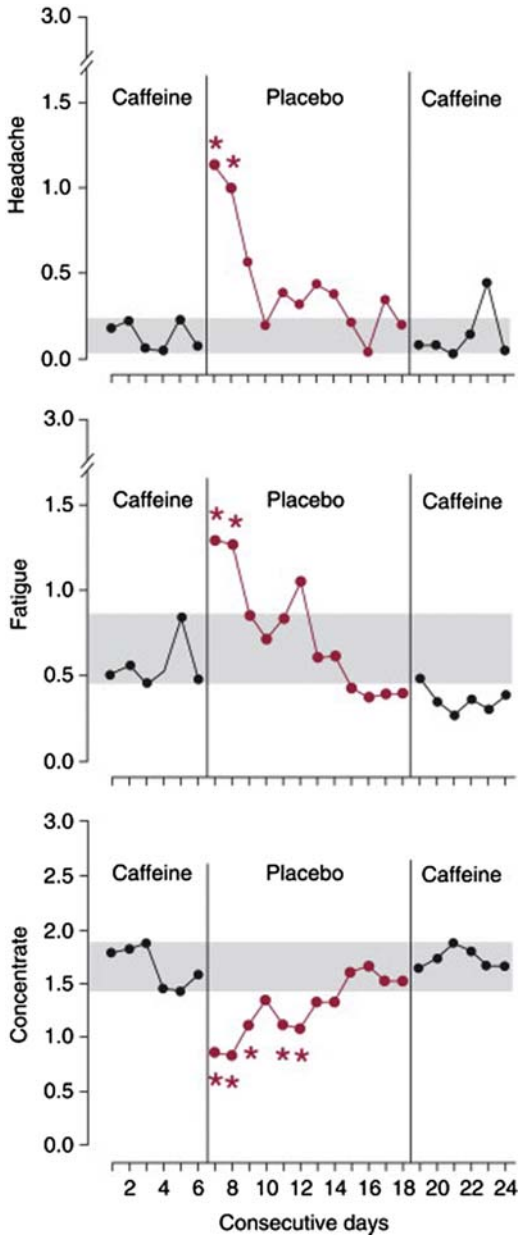


Fig. 1 Time course of caffeine withdrawal symptoms in four volunteers. Under double-blind conditions volunteers received either 100 mg day^{-1} of caffeine or placebo. Assessments included subjective ratings of headache (top), feelings of lethargy/fatigue/tired/sluggish (middle), and ability to concentrate (bottom). Ratings ranged from 0 (not at all) to 3 (very much). Shaded areas indicate the range of means from the initial 6-day caffeine period. Mean ratings are presented in black for caffeine days and red for placebo days. Asterisks indicate placebo days that are significantly different from the initial caffeine period ($p \leq 0.05$). Reproduced from Griffiths, R.R., Evans, S.M., Heishman, S.J., Preston, K.L., Sannerud, C.A., Wolf, B., Woodson, P.P., 1990. Low-dose caffeine physical dependence in humans. J. Pharmacol. Exp. Therapeut. 255: 1123–1132, with permission from ASPET.

Caffeine tolerance

Caffeine tolerance may occur in response to daily caffeine consumption. Tolerance can occur to the subjective, sleep disrupting, and physiological effects of caffeine. The degree of caffeine tolerance depends on several factors including the challenge and maintenance doses, frequency of administration and individual differences in elimination rate. The prevalence of self-reported tolerance among current caffeine users varies from 8% to 50%. Rates as high as 92% have been reported among caffeine-dependent individuals.

Regular caffeine users may acquire complete tolerance (i.e., no difference between placebo and caffeine after prolonged exposure to caffeine) to some, but not all, of the subjective effects of caffeine. For example, experimental studies showed that volunteers who

received moderate to high doses of caffeine (400 mg–900 mg day^{-1}) for at least two weeks developed complete tolerance to ratings of subjective stimulant effects. Other studies indicate that complete tolerance to caffeine subjective effects does not occur at lower caffeine doses and over shorter periods of dosing.

Tolerance development may differ across different outcome measures. One study showed that tolerance or complete tolerance developed to subjective ratings but not to measures of cerebral blood flow or EEG in volunteers receiving 400 mg day^{-1}.

With regard to other physiological effects, tolerance may develop to the effects of caffeine on diuresis, parotid gland salivation, metabolic rate, plasma norepinephrine and epinephrine levels, and plasma renin activity. Findings suggest that regular caffeine users develop partial, but not complete tolerance to the effects of caffeine on cerebral blood flow and EEG measures. Two studies, which tested a small number of volunteers demonstrated the development of complete tolerance to the pressor effects of high doses of caffeine (e.g., 600–850 mg day^{-1}). However, several more recent studies examining a larger number of volunteers showed that tolerance to the pressor effects of caffeine is variable across individuals, with some subjects showing complete tolerance, whereas others show only incomplete tolerance.

Caffeine intoxication

Caffeine intoxication, which is a DSM-IV-TR recognized disorder, is defined by the development of symptoms and clinical features in response to acute caffeine consumption that cause clinically significant distress or impairment. Although the DSM-IV-TR definition specifies that the diagnosis depends on recent consumption of at least 250 mg of caffeine, symptoms typically emerge at doses greater than 500 mg. Symptoms of caffeine intoxication include restlessness, nervousness, insomnia, flushed face, diuresis, gastrointestinal disturbance, muscle twitching, rambling flow of thought and speech, tachycardia or cardiac arrhythmia, periods of inexhaustibility, and psychomotor agitation. Although children and caffeine-intolerant individuals may be particularly sensitive to the acute adverse effects of caffeine, habitual caffeine users may also experience episodes of caffeine intoxication. Several case reports and experimental studies suggest that caffeine consumption may produce hallucinations in some individuals, particularly under conditions of stress. Intoxications with caffeine are rare, however can be fatal. In general, caffeine use seems not harmful within typical doses of intake. Gahr concluded in 2020 that caffeine features several, however, not all characteristics of potentially addictive drugs; withdrawal after termination of a longer period of use and tolerance are known. In the DSM-5 "caffeine use disorder" is categorized as a possible future disorder that currently needs further study. The pattern of caffeine use of patients should be considered in the medical practice.

Caffeine-induced sleep disorder

Caffeine reduces total sleep time and limits latency to sleep onset, most probably by blocking the sleep promoting effects of adenosine. In addition, caffeine decreases stage 3–4 sleep and suppresses EEG slow wave activity during sleep. The sleep-disrupting effects of caffeine are well documented even at low doses (e.g., one cup of coffee). Surveys have found associations between daily dietary caffeine intake and sleep problems in both adults and adolescents. Eid et al. concluded in 2021 that the risk factors for objective sleep impairment included parenting young children and watching television at night, whereas better sleep outcomes were associated with greater engagement with physical activity.

Some caffeine users may develop caffeine-induced sleep disorder, which is a DSM-IV-TR recognized disorder typically characterized by insomnia. Some caffeine users may present with caffeine-induced hypersomnia with daytime sleepiness due to withdrawal symptoms. Sleep disturbances secondary to caffeine may increase in severity as caffeine dose increases and proximity to caffeine administration at bedtime decreases. Individuals who are not regular caffeine users and are not tolerant, or have only partial tolerance to the sleep-disrupting effects of caffeine are more likely to experience caffeine-related sleep disruption. Lunsford-Avery et al. concluded 2021 that for experimental factors, past studies often lacked a baseline control for diet and sleep and none discussed the possible reversal of withdrawal effect due to pre-experimental fasting.

Caffeine-induced anxiety disorder

In addition to the symptom of anxiety that can be a component of caffeine intoxication, caffeine can also produce caffeine-induced anxiety disorder, a DSM-IV-TR disorder. Presentation of a caffeine-induced anxiety disorder may include symptoms of generalized anxiety, panic attacks, obsessive–compulsive disorder, or phobic disorder. Individuals who have an existing anxiety disorder, or who endorse symptoms of anxiety sensitivity, are at increased risk of experiencing anxiety symptoms in response to caffeine.

Caffeine dependence

Substance dependence is characterized by a cluster of cognitive, behavioral, and physiological symptoms indicating that an individual is continuing to use a substance despite experiencing clinically significant substance-related problems. Caffeine dependence

is recognized as a diagnosis in ICD-10, the official diagnostic system of the World Health Organization. In contrast, the DSM-IV-TR currently excludes caffeine from a diagnosis of substance dependence despite using very similar diagnostic criteria to ICD-10. A growing literature from experimental studies, clinical interviews, and survey studies indicates that some caffeine users manifest a pattern of symptoms consistent with a DSM-IV-TR diagnosis of substance dependence as applied to caffeine.

One population-based survey study of 162 randomly-selected caffeine users found that 9% of the sample endorsed three or more of four DSM-IV-TR substance dependence criteria that are thought to be most relevant of a meaningful diagnosis of caffeine dependence. The criteria and past-year incidence were: (1) Persistent desire or unsuccessful efforts to cut down or control use (56%); (2) Characteristic withdrawal syndrome or substance taken or relieve or avoid withdrawal (18%); (3) Use is continued despite a physical or psychological problem likely caused or exacerbated by the substance (14%); and (4) Tolerance defined by either a need for markedly increased amounts to achieve desired effect or markedly diminished effect with continued use of the same amount (8%).

Individuals meeting criteria for caffeine dependence vary considerably in the amount of caffeine consumed per day and in the types of caffeine-containing products that they regularly consume (e.g., coffee, soft drinks, tea). Importantly, the problems associated with caffeine dependence are not trivial. These include, but are not limited to anxiety, insomnia, stomach problems, and cardiovascular problems. One survey found that 13% of caffeine users had been advised by a physician or counselor to reduce or cut down caffeine in the last year. Fifteen percent of caffeine consumers were particularly resistant to modifying their use, indicating that they would not change when or how much caffeine they used, no matter what they were doing or where they were.

Caffeine and food

In addition to being consumed in its natural plant forms (e.g., coffee and tea), caffeine is also frequently consumed as an added ingredient with or without added sugars for both kind of beverages in many popular sugar-sweetened soft drinks and energy drinks. The bitter taste profile of caffeine is often masked or obscured by the addition of sugar, fat, and other flavors in caffeine-containing foods and beverages. Beverage manufacturers have made the claim that caffeine is added to beverages in order to enhance flavor, but most individuals are unable to detect flavor differences between sugar-sweetened soft drinks with and without caffeine at the caffeine concentration found in most soft drink beverages. It is likely that caffeine-containing beverages are widely consumed because caffeine can function as a reinforcer, increase flavor preferences for caffeine-containing beverages, and produce physical dependence, which results in a substance dependence syndrome.

As described above, caffeine-dependent users may be unable to cut down or control caffeine use despite a persistent desire to do so and may also continue to use caffeine despite medical problems associated with caffeine consumption. Thus, caffeine dependence may exacerbate adverse health outcomes associated with the consumption of caffeine-containing sugar-sweetened beverages. Sugary drinks have been associated with weight gain, obesity, and type-2 diabetes even after controlling for other factors. Of particular concern is that sugar-sweetened beverages together other food containing added sugars consumption may be associated with weight gain and obesity in children and may displace milk and other important nutrients in the diets of children and adolescents.

Caffeine is an added ingredient in many over-the-counter weight loss medications. Experimental studies have generally found that acute caffeine consumption is associated with increased energy expenditure, decreased food intake, and reduced ratings of hunger. There is also some evidence that caffeine increases fat oxidation. It is not clear whether any or all of these effects are due to acute caffeine effects *per se* versus reversal of caffeine withdrawal (e.g., the observed increased energy expenditure may be due to the reversal of suppressed energy expenditure in the caffeine-deprived comparison condition). Prospective longitudinal studies have shown that caffeine consumption is negatively associated with weight gain, but few well-controlled studies have examined the long-term efficacy of caffeine alone as an intervention for weight loss and weight loss maintenance.

Experimental studies have shown that the co-ingestion of caffeine and catechins enhances energy expenditure and fat oxidation more than an equivalent amount of caffeine without added catechins. Numerous investigations have examined the combined effects of caffeine and catechins, most commonly co-ingested in green tea, on weight loss outcomes. A recent meta-analysis of studies comparing a caffeine/catechin condition with either a placebo condition or a low-dose caffeine/catechin condition found that the caffeine-catechin combination had small positive effects on weight loss and maintenance of weight loss after a period of negative energy balance. There is some evidence that effects are attenuated in habitual caffeine consumers with high daily caffeine intake (e.g., >300 mg day^{-1}). This effect may be mediated through caffeine tolerance. Several studies have found that caffeine in combination with ephedrine is efficacious for weight loss. However, ephedrine alkaloid supplements have been associated with adverse events and have been banned by the Food and Drug Administration (FDA) in the US.

Conclusion

Caffeine can function as a reinforcer and produces a wide range of physiological, behavioral, and subjective effects, including anxiety, insomnia, diuresis, alertness, and enhancement in some types of performance. Regular caffeine consumption can produce tolerance, withdrawal, and a substance dependence syndrome.

Acknowledgments

Preparation of the original chapter was supported, in part, by the United States Public Health Service grant R01 DA03890 from the National Institute on Drug Abuse.

References

EFSA Panel on Dietetic Products, Nutrition and Allergies (NDA), 2015. Scientific opinion on the safety of caffeine. EFSA J. 13 (5), 4102. https://doi.org/10.2903/j.efsa.2015.4102.

Eid, S.W., Brown, R.F., Birmingham, C.L., Maloney, S.K., 2021. A comparison of physical activity levels, sleep disrupting behavior, and stress/affective distress as predictors of sleep as indexed by actigraphy. J. Phys. Activ. Health 13, 1–12. https://doi.org/10.1123/jpah.2020-0042.

Food and Drug Administration (USA), 2018. Highly Concentrated Caffeine in Dietary Supplements: Guidance for Industry. U.S. Department of Health and Human Services, Center for Food Safety and Applied Nutrition. https://www.fda.gov/media/112363/download.

Gahr, M., 2020. Caffeine, the most frequently consumed psychostimulant: a narrative review article. Fortschr. Neurol. Psychiatr. 88 (5), 318–330. https://doi.org/10.1055/a-0985-4236.

Guest, N.S., VanDusseldorp, T.A., Nelson, M.T., Grgic, J., Schoenfeld, B.J., Jenkins, N.D.M., Arent, S.M., Antonio, J., Stout, J.A.J.R., Trexler, E.T., Smith-Ryan, A.E., Goldstein, E.R., Kalman, D.S., Campbell, B.I., 2021. International society of sports nutrition position stand: caffeine and exercise performance. Sports Nutr. Rev. J. 18, 1. https://doi.org/10.1186/s12970-020-00383-4.

Lunsford-Avery, J.R., Kollins, S.H., Kansagra, S., Wang, K.W., Engelhard, M.M., 2021. Impact of daily caffeine intake and timing on electroencephalogram-measured sleep in adolescents. J. Clin. Sleep Med. https://doi.org/10.5664/jcsm.9736.

Peeling, P., Dawson, B., 2007. School of human movement and exercise influence of caffeine ingestion on perceived mood states, concentration, and arousal levels during a 75-min university lecture. Adv. Physiol. Educ. 31, 332–335. https://doi.org/10.1152/advan.00003.2007.

Smith, A., Sturgess, W., Gallagher, J., 1999. Effects of a low dose of caffeine given in different drinks on mood and performance. Hum. Psychopharmacol. 14, 473–482. https://doi.org/10.1002/(SICI)1099-1077(199910)14:7<473::AID-HUP129>3.0.CO;2-D.

Further reading

Astorino, T.A., Roberson, D.W., 2010. Efficacy of acute caffeine ingestion for short term high-intensity exercise performance: a systematic review. J. Strength Condit. Res. 24 (1), 257–265. https://doi.org/10.1519/JSC.0b013e3181c1f88a.

Cornelis, M.C., El-Sohemy, A., 2007. Coffee, caffeine, and coronary heart disease. Curr. Opin. Clin. Nutr. Metab. Care 10, 745–751. https://doi.org/10.1097/MOL.0b013e3280127b04.

Ferré, S., 2008. An update of the mechanisms of the psychostimulant effects of caffeine. J. Neurochem. 105, 1067–1079. https://doi.org/10.1111/j.1471-4159.2007.05196.x.

Ganio, M.S., Klau, J.F., Casa, D.J., Armstrong, L.E., Maresh, C.M., 2009. Effect of caffeine on sport-specific endurance performance: a systematic review. J. Strength Condit. Res. 23 (1), 315–324. https://doi.org/10.1519/JSC.0b013e31818b979a.

Juliano, L.M., Griffiths, R.R., 2004. A critical review of caffeine withdrawal: empirical validation of symptoms and signs, incidence, severity, and associated features. Psychopharmacology 176 (1), 1–29. https://doi.org/10.1007/s00213-004-2000-x.

Yang, A., Palmer, A.A., de Wit, H., 2010. Genetics of caffeine consumption and responses to caffeine. Psychopharmacology 211 (3), 245–257. https://doi.org/10.1007/s00213-010-1900-1.

Calcium

V. Loria-Kohen, A.M. López-Sobaler, A. Aparicio, and R.M. Ortega, Department of Nutrition and Food Science, Complutense University of Madrid. UCM Research Group VALORNUT-920030, Madrid. Spain

© 2023 Elsevier Ltd. All rights reserved.

This is an update of K.J. Schulze, Calcium, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 228–234, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00034-9.

Introduction **106**
Functions **106**
Dietary calcium intake **106**
Sources 106
Metabolism **107**
Balance 107
Absorption **107**
Mechanisms 107
Bioavailability 108
Nutrient interactions 109
Calcium losses 109
Hormonal control 110
Influence of life stages 111
Race and genetics 112
Health consequences of calcium deficiency **112**
Skeletal 112
Nonskeletal 112
Health consequences of calcium excess **112**
Conclusion **113**
References **113**

Key points

The main objectives for the article are:

- To identify the essential functions that calcium exerts in our body.
- To describe calcium sources and the variability in their bioavailability.
- To review the current guidelines for calcium and vitamin D intake according to The Institute of Medicine of the National Academy of Sciences.
- To describe the mechanisms implicated in calcium absorption and metabolism.
- To outline the health consequences of calcium deficiency or excess according to the new scientific evidence.

Glossary

Calbindin D Transcellular calcium transport protein involved in calcium absorption (9 kD form) and renal calcium reabsorption (28 kD form)

Calmodulin Calcium binding protein that undergoes conformational changes to affect cell signaling

Osteopenia Defect in bone mass less severe but potentially leading to osteoporosis

Osteoporosis Defect in bone mass and microarchitecture resulting in porous and brittle bones with normal mineral to collagen ratio

Parathyroid hormone Eighty-four amino acid polypeptide hormone released from the parathyroid gland in response to low circulating calcium concentrations

Rickets Disorder of bone characterized by poor mineralization

TRPV5, TRPV6 Calcium channels for transcellular calcium transport, members of the Transient Receptor Protein superfamily

Vitamin D Family of fat-soluble secosteroids, comprised of cholecalciferol (synthesized from dermal conversion of 7-dehyrocholesterol) and ergocalciferol (from irradiated plant sources); the form of vitamin D that reflects nutritional status is 25(OH) vitamin D, derived from the hydroxylation of cholecalciferol or ergocalciferol in the liver; another hydroxylation step in the kidney results in $1,25(OH)_2$ vitamin D, or calcitriol

Introduction

Calcium is a divalent cation with a molecular weight of 40 g mol^{-1}. It is the fifth-most abundant element in the human body, making up approximately 1000–1200 g of total body weight in an adult. Over 99% of calcium is present in bones and teeth of the body, whereas less than 1% is distributed among cells and extracellular fluid. Calcium circulates bound to albumin (40%), sulfates, phosphate and citrates (10%), or as ionized calcium (50%) in concentrations that are tightly maintained at ~10 mg dL^{-1} (2.5 mmol L^{-1}) by sensitive homeostatic processes (Song, 2017). A steep gradient between extracellular or cellular organelle and cytosolic calcium concentrations allows for a rapid flux of calcium into the cell to activate cell-signaling (Weaver and Heaney, 2014). Calcium is lost from the body daily in urine and through the gastrointestinal tract and skin, so calcium intakes must be sufficient to balance these obligate losses. The calcium economy is regulated through parathyroid hormone (PTH) and the conversion of 25(OH) vitamin D to $1,25(OH)_2$ vitamin D that it promotes, resulting in increased renal reabsorption of calcium, resorption of calcium from bone, and more efficient absorption of dietary calcium when circulating calcium concentrations are low (Song, 2017). Adequate calcium intake is critical for the achievement of peak bone mass in the first several decades of life, the retention of bone during middle adulthood, and the minimization of bone loss during the last several decades of life. Both skeletal and nonskeletal roles for calcium in human health continue to be explored.

Functions

Over 99% of the body's calcium occurs in bone where it is deposited on a cartilage matrix as hydroxyapatite ($Ca_{10}(PO_4)_6(OH)_2$); over 30% of bone mineral is calcium. Bone provides structural integrity and also serves as a reservoir of calcium. The remaining <1% of the body's calcium is present in blood, extracellular space, muscle, and other tissues. Maintenance of constant serum calcium concentrations at approximately 10 mg dL^{-1} (2.5 mmol L) is critical for a number of cellular functions (Song, 2017). Disorders of calcium metabolism are pretty frequently in routine clinical practice. Hypocalcemia is not as frequently seen as hypercalcemia is, but it can be potentially life-threatening if not appropriately recognized and swiftly treated (Goyal et al., 2021). Intracellular calcium is bound within organelles such as the nucleus, endoplasmic reticulum, and vesicles, with cytosolic calcium approximately 104 times less concentrated than that of the extracellular fluid and organelles. Therefore, rapid changes in cytosolic calcium concentrations occur with the influx of extracellular or sequestered calcium. In the best known cell signaling pathway, intracellular calcium binds to calmodulin, resulting in conformational changes in the protein to trigger cellular events. Resumption of normal cell concentrations requires extrusion of calcium via pumps or sequestration via calcium binding proteins. Muscle contraction, nerve conduction, cell movement and differentiation, cell division, cell-to-cell communication, and secretion of hormones such as insulin are all dependent on intracellular calcium signaling. These functions are largely protected from dietary fluctuations in calcium intake until deficiency becomes extreme. Nonetheless, marginal levels of intake may not only affect skeletal health but predispose to subtle changes in cellular conditions that contribute to alterations in cellular metabolism (Weaver and Heaney, 2014).

Dietary calcium intake

Sources

There are striking inequities in calcium intake between rich and poor populations (Cormick y Belizan, 2019). Calcium intakes in some regions of the world average less than 400 mg (10 mmol) per day. Across the 74 countries with data, average national dietary calcium intake ranges from 175 to 1233 mg day^{-1}. Many countries in Asia have average dietary calcium intake less than 500 mg day^{-1}. Countries in Africa and South America mostly have low calcium intake between about 400 and 700 mg day^{-1}. Only Northern European countries have national calcium intake greater than 1000 mg day^{-1}. Average calcium intake is generally lower in women than men (Balk et al., 2017).

Dietary intake of calcium in the US is typically 900–1200 mg (22.5–30 mmol) per day unless supplements are consumed. Approximately 65–75% of dietary calcium is consumed in dairy products, 8–10% in fruits and vegetables, 5% in grains, and the rest from all other sources. Whereas in China, only around 7% of total calcium intake comes from dairy products, while most comes from vegetables (30%) and legumes (17%) (Cormick and Belizan, 2019; Weaver and Heaney, 2014). One serving of dairy products (i.e., 250 mL milk or yogurt or 40 g cheese) contains approximately 300 mg (7.5 mmol) of calcium. Grains consumed in substantial amounts in bread or maize products can be important sources of calcium, although this is a less bioavailable source than dairy products. Other foods high in calcium are tofu set with a calcium salt, kale, broccoli, and calcium-fortified juices and cereals. Also, calcium present in mineral drinking waters is an important quantitative source of calcium intake. Supplement use is reported in over 40% of the US population overall, and in nearly 70% of older women (Cormick and Belizan, 2019).

In countries or regions where calcium intake is low, milk-based products tend not to be a major component of the diet, and thus a higher percent of total calcium intake is supplied by plant products. Other sources of calcium are experimentally used by the food industry, such as eggshells (Palacios et al., 2021). To some extent, calcium requirements may depend on other nutrients that enhance or inhibit calcium absorption or utilization. Recognizing the presence of low calcium intake is necessary to develop national strategies to optimize intake (Balk et al., 2017).

Table 1 Current intake recommendations for calcium.

Group	*Age range*	*EAR (mg)*	*RDA/AI (mg)*	*UL (mg)*
Infants	0–6 months	–	200	1000
	6–12 months	–	260	1500
Children	1–3 years	500	700	2500
	4–8 years	800	1000	2500
Males	9–18 years	1100	1300	3000
	19–50 years	800	1000	2500
	51–70 years	800	1000	2000
	>70 years	1000	1200	2000
Females	9–18 years	1100	1300	3000
	19–50 years	800	1000	2500
	51–70 years	1000	1200	2000
	>70 years	1000	1200	2000
Pregnant Lactating	No changes in age-appropriate intakes advocated			

Abbreviations: EAR, estimated average requirement; RDA, recommended dietary allowance; AI, adequate intake recommendation; UL, tolerable upper intake level. RDAs are established for all age groups except infants.
Source: Institute of Medicine (2011).

The Institute of Medicine of the National Academy of Sciences (IOM, 2011), currently called National Academy of Medicine, reviewed guidelines for calcium and vitamin D intake in 2011 and updated recommendations for US and Canadian populations (Table 1). These recommendations are based on the most current data available regarding the amount of dietary calcium i.e., required to optimize bone calcium deposition or maintenance, accounting for estimates of fractional absorption and usual losses. In contrast to previous reports suggesting that a considerable number of people in the US were at risk of consuming inadequate amounts of calcium, more recent national data suggest that most population groups, other than adolescent girls, are generally consuming adequate amounts of calcium with reference to the new Estimated Average Requirements.

Metabolism

Balance

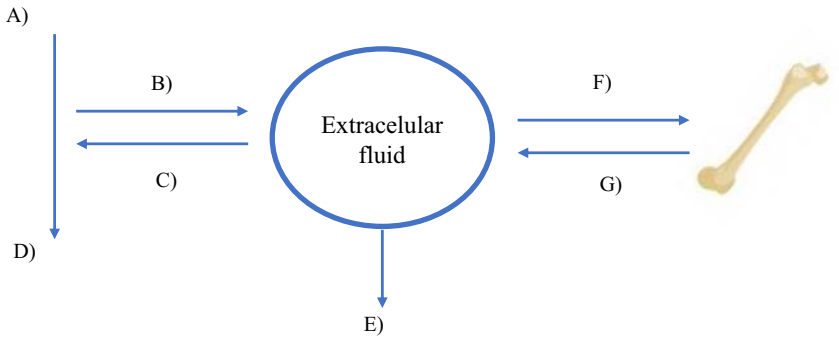
Calcium balance can be calculated as the difference between calcium entering the body through absorption of dietary calcium and obligate losses of calcium through urine, gastrointestinal tract, and skin. When uptake exceeds losses, an individual is in positive calcium balance, and net bone calcium accrual occurs, although bone continually remodels such that bone calcium turns over entirely every 5 or 6 years. Negative balance occurs when dietary intake is insufficient to cover losses, despite mechanisms to conserve body calcium (Song, 2017). Typically, approximately 25% of dietary calcium is absorbed in adults and delivered to the exchangeable calcium pool, which turns over 20–30 times per day. A remarkably large amount of calcium is filtered through the kidneys, approximately 10,000 mg (250 mmol) per day, of which approximately 98% is reabsorbed, so that urinary excretion of the mineral is only 100–200 mg (2.5–5 mmol) per day. Calcium balance benefits from increased dietary intake up to a "threshold," above which excess calcium is excreted rather than contributing to bone mass. These thresholds are highest during periods of growth. Differences in calcium economy also exist between racial/ethnic groups, such that the lower urinary calcium and better calcium conservation in African–Americans relative to Caucasians probably contributes to their higher bone mineral density. Aspects of calcium balance for a typical adult, pubertal and elderly female are shown in Fig. 1.

Absorption

Mechanisms

Dietary calcium is complexed to food constituents such as proteins, phosphate, and oxalate, from which it needs to be released before absorption. Conditions that promote the solubility of calcium may enhance its absorption, and achlorhydria may therefore impair calcium absorption, as has been shown with some calcium supplements when consumed in a fasted state (Weaver and Heaney, 2014).

Calcium crosses the intestinal mucosa by both active and passive transport. The active process is saturable, transcellular, and occurs primarily in the duodenum. This pathway is upregulated during calcium deficiency and predominates when calcium intakes are low. Luminal calcium enters the enterocyte from the microvillus border of the apical plasma membrane via a calcium channel



		Adult female	Puberal female	Elderly (75 years)
A)	Dietary calcium intake (mg/d)	1000	1000	1000
B)	Total calcium absorption (mg/d)	250	350	200
	Fractional calcium absorption (%) = (100*B)/A	25%	35%	20%
C)	Net endogenous losses (mg/d)	110	70	120
D)	Fecal calcium (mg/d)= (A-B) + C	860	720	920
E)	Urinary calcium (mg/d)	150	100	150
F)	Bone calcium deposition (mg/d)	500	1800	500
G)	Bone calcium resorption (mg/d)	500	1620	570
	Calcium balance (mg/d)= B - (C+E) = F - G	0	+180	-70

Fig. 1 Hypothetical example of calcium balance on 1000 mg day^{-1} dietary calcium intakes in an adult woman, pubertal girl and elderly woman. Calcium uptake by the body is indicated by the red fill, losses by the blue fill. The difference between calcium intake and losses must equal the difference between bone deposition and resorption, as the extracellular fluid compartment calcium concentration remains constant, although bone turnover rates are higher during growth. Calcium absorption efficiency is higher in the girl, and endogenous losses are lower due to smaller body size. Note that the puberal female and the elderly is not consuming the RDA for calcium, and a higher intake would favor a greater calcium balance.

and is translocated to the basolateral membrane where it is released through another calcium channel. Calbindin D_{9k}, a calcium binding protein regulated by the hormonal form of vitamin D, $1,25(OH)_2D$ vitamin D, transports calcium across the enterocyte following its uptake through an apical membrane channel, TRPV6, and calcium is extruded via an ATPase, PMCA1. Details of this process are still under review, and other transcellular processes may exist (Weaver and Heaney, 2014).

The passive transport pathway is no saturable and paracellular. It occurs throughout the small intestine and is unaffected by PTH and relatively independent of $1,25(OH)_2$ vitamin D, although this metabolite has been found by some investigators to increase the permeability of the paracellular pathway. A substantial amount of calcium is absorbed by passive transport in the ileum due to the relatively slow passage of food through this section of the intestine. The amount of calcium absorbed by passive transport is proportional to the intake and the bioavailability of calcium consumed (Weaver and Heaney, 2014).

Although fractional calcium absorption increases in response to low calcium intake, this compensatory mechanism is incomplete, and total calcium absorption increases in relation to calcium intake. The efficiency of calcium absorption also varies throughout life. It is highest during infancy, time when the absorption fraction can range from somewhat above 60% with lower intakes to about 30% with higher intakes. As the infant transitions into childhood, fractional calcium absorption declines, only to rise again in early puberty, a time when modeling of the skeleton is maximal. A calcium absorption rate of 28% before puberty, 34% during early puberty (the age of the growth spurt), and 25% 2 years after early puberty has been reported. Fractional absorption remains about 25% in young adults (IOM, 2011), and then declines with age (by approximately 2% per decade after menopause). During pregnancy, calcium absorption doubles, although inter-individual variation is high. Thus, the active transport pathway is most efficient when calcium requirements are high: during infancy, adolescence, and pregnancy. Increased absorption is also observed during primary hyperparathyroidism, sarcoidosis, and estrogen and growth hormone administration. Physiological/pathological factors, which decrease intestinal calcium absorption include low serum $1,25(OH)_2$ vitamin D, chronic renal insufficiency, and hypoparathyroidism. Metabolic status also influences calcium absorption such that severe obesity is associated with higher calcium absorption and dieting reduces the fractional calcium absorption by 5% (IOM, 2011).

Bioavailability

Dietary factors that may inhibit calcium absorption include phosphate, oxalate, phytate, and fiber. Fermentation of bread during leavening reduces phytate content, making calcium more bioavailable. Fiber in fruits and vegetables can also inhibit calcium absorption; the uronic acids in hemicellulose are strong calcium binders, as is the oxalic acid present in high concentrations in foods

such as spinach. Calcium bioavailability from beans is approximately half and that from spinach approximately one-tenth of the bioavailability from milk. In contrast, calcium absorption from low-oxalate vegetables, such as kale, broccoli, and collard greens, is comparable to that of milk (Weaver and Heaney, 2014). Therefore, veganism is associated with a low calcium intake, which has less bioavailability, and people who are following a vegan diet should be aware of the risk of potential dietary deficiencies (Bakaloudi et al., 2021).

Most calcium salts used in food fortification for human consumption have a bioavailability ranging from 20% to 40%. Their solubility, source, interaction with the food and whole meal, and bioavailability vary widely. Calcium citrate malate has been shown to have greater absorption compared with other salts (Palacios et al., 2021), and it may be recommended instead of the less costly calcium carbonate in individuals with achlorhydria or if supplements cannot be taken with a meal.

Dietary fat does not affect calcium absorption except in individuals with diseases that impair fat absorption (e.g., short bowel syndrome, celiac disease, and pancreatitis), where calcium may form an insoluble "soap" with the unabsorbed fat in the alkaline lumen of the small intestine. Neither dietary phosphorus nor a wide range of phosphorus-to-calcium ratios affect intestinal calcium absorption.

Factors that increase calcium absorption include protein (or specific amino acids, lysine, and arginine). Lactose improves calcium absorption in young infants, in whom absorption of calcium is predominantly by passive transport, but in adults, lactose has little effect on the efficiency of calcium absorption (Weaver and Heaney, 2014). New research also suggests that certain oligosaccharides, such an inulin, may enhance absorption of calcium. It is possible that these prebiotics function through a variety of mechanisms to enhance the solubility of luminal calcium and improve the ability of the enterocyte to bind and take up calcium. In addition to these mechanisms, current evidence suggests that prebiotic fibers may offer an alternative approach to enhance calcium absorption through altering gut microbiota to improve bone health ultimately (Cao et al., 2020). According to a recent study, an inulin intake of at least 8–10 g day^{-1} supports calcium absorption and total body bone mineral content/density in adolescents through its known mechanisms of action (Bakirhan and Karabudak, 2021).

Nutrient interactions

Calcium can inhibit iron and zinc absorption, although this interaction may not be an issue at typical levels of calcium intake. The mechanism by which this occurs remains controversial, but the inhibition probably occurs within the mucosal cells rather than in the intestinal lumen. This interaction is of concern because calcium supplements are taken by many women who may have difficulty maintaining adequate iron stores. The inhibitory effect on iron absorption is relatively unimportant when iron stores are adequate (ferritin 50–60 μg L^{-1}), but consideration should be given to monitoring the iron status of menstruating women with low iron stores who take calcium supplements. There is no inhibitory effect when calcium and iron supplements are consumed together in the absence of food, and inhibition may be less with calcium citrate. Moreover, evidence shows no effect on iron status of prolonged calcium supplementation taken at the same time or separate of meals (Cormick and Belizan, 2019; Weaver and Heaney, 2014).

Earlier, it was common to restrict dietary calcium in patients with a history of calcium oxalate stones. More recent studies have suggested that dietary calcium restriction is not recommended for stone formers with nephrolithiasis; on the contrary, diets with more than 1 g of calcium per day could be protective against stone formation. Kidney stones are mainly composed of calcium combined with oxalate or phosphate. The calcium remaining in the intestine would impede the absorption of products associated with the risk of renal stones, such as oxalates, and the intake of calcium supplements during meals would decrease the absorption of oxalates and thus the formation of stones (Cormick and Belizan, 2019).

Calcium losses

Very large amounts of calcium are filtered by the kidney each day, but nearly all is reabsorbed throughout the nephron, such that typically, 100–200 mg (2.5–5 mmol) of calcium is excreted in urine daily. Most calcium is reabsorbed passively in the proximal tubule, whereas active absorption takes place in the distal convoluted tubule via a calcium channel, TRPV5, and transcellular movement via calbindin D_{28k}, a calcium transport protein. Dietary calcium has a relatively small impact on urinary calcium in adults (e.g., only 6–8% of an increased dietary calcium intake is excreted in the urine); in children, increased intake is utilized for bone accretion rather than excreted. The major food components that affect urinary calcium are protein, phosphorus, caffeine, and sodium. For each 50 g increment in dietary protein, approximately 60 mg (1.5 mmol) of additional calcium is lost in urine (Weaver and Heaney, 2014). The higher amounts of phosphorus consumed concurrently with a high-protein diet can blunt, but not eliminate, this phenomenon. Dietary phosphorus (as well as intravenously administered phosphorus) increases PTH synthesis and subsequently stimulates renal calcium reabsorption, reducing the urinary excretion of calcium. Caffeine causes a transient loss of urinary calcium through diuresis. It has been shown repeatedly in animals and humans that dietary sodium, in the form of salt (NaCl), increases urinary calcium excretion. On average, for every 2300 mg (100 mmol) of sodium excreted in urine, there is an approximately 24–40 mg (0.6–1 mmol) loss of calcium in free-living healthy populations of various ages. Thus, when urinary calcium excretion is excessive it is often recommended to limit salt intake. Diets that produce more alkaline metabolic conditions, such as those with potassium-rich fruits and vegetables, may reduce calcium losses through the buffering effects of potassium bicarbonate on metabolic pH (Weaver and Heaney, 2014).

Endogenous losses of calcium occur in the gastrointestinal tract as cells are sloughed and calcium-containing pancreatic and bile secretions are released during digestion. Endogenous losses are proportional to body size and average approximately 2 mg kg^{-1}

day^{-1} in adults and 1.5 mg kg^{-1} day^{-1} in children. Greater losses may occur under some conditions that affect gut integrity, such as protein-losing enteropathies. Because of the difficulties in measuring endogenous losses, requiring balance studies that utilize calcium stable isotope tracers, these losses are typically estimated rather than directly measured. Dermal losses of calcium occur on the order of approximately 25 mg day^{-1}, but may increase with sweating, and are as well typically estimated rather than measured (Weaver and Heaney, 2014).

Hormonal control

The principal regulators of calcium homeostasis in humans and most terrestrial vertebrates are PTH and 1,25(OH)$_2$ vitamin D, the active form of the vitamin (Fig. 2). PTH is a single-chain polypeptide i.e., released from the parathyroid when a decrease in the circulating calcium concentration is detected by parathyroid gland calcium-sensing receptors. It restores extracellular calcium concentrations by increasing the renal reabsorption of calcium and decreasing phosphate reabsorption, and by enhancing the renal conversion of 25(OH) vitamin D to the active, hormonal form of the vitamin, 1,25(OH)$_2$ vitamin D, and by stimulating the resorption of bone to release calcium. In turn, 1,25(OH)$_2$ vitamin D enhances calcium absorption through the active pathway. PTH release is inhibited when serum calcium and 1,25(OH)$_2$ vitamin D increase or when serum phosphate is decreased. The highly regulated interactions among PTH, calcium, 1,25(OH)$_2$ vitamin D, and phosphate maintain blood calcium levels at remarkably constant levels despite significant changes in calcium intake or absorption, bone metabolism, or renal function. PTH regulation of circulating calcium concentration occurs on a minute-by-minute basis, and acute PTH administration leads to release of the rapidly turning over pool of calcium near the bone surface. Chronic administration of PTH increases osteoclast cell number and activity. Interestingly and paradoxically, intermittent PTH administration is anabolic, increasing formation of trabecular bone (Weaver and Heaney, 2014).

There are two sources of vitamin D: the diet (where it is found as vitamin D$_2$ (ergocalciferol) or vitamin D$_3$ (cholecalciferol)) or synthesis in skin during exposure to ultraviolet radiation (sunlight). The vitamin enters the circulation and is transported on a vitamin D binding protein to the liver, where it is hydroxylated to 25(OH) vitamin D—the major circulating form of the vitamin. It is hydroxylated again in the kidney to 1,25(OH)$_2$ vitamin D, or calcitriol, the most active metabolite of the vitamin. Vitamin D acts through nuclear receptors to enhance production of proteins involved in calcium absorption, as well as renal calcium

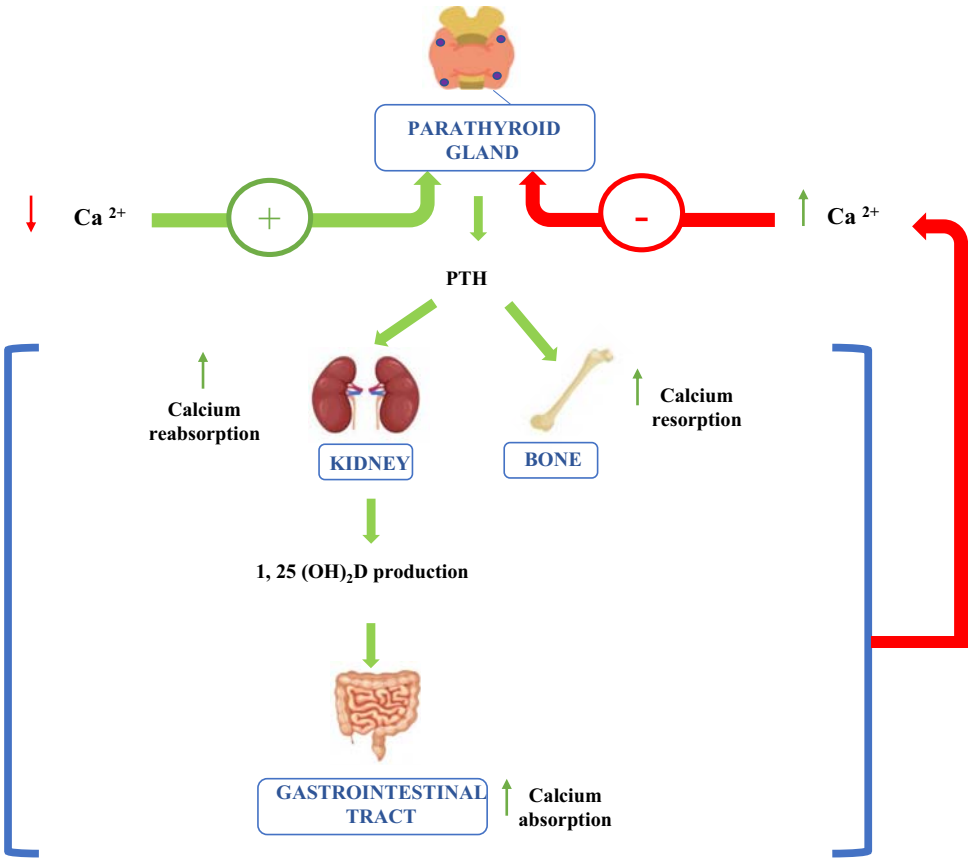


Fig. 2 Regulatory system for maintaining calcium homeostasis via the PTH-vitamin D axis. 1,25(OH)$_2$ vitamin D works with PTH to enhance calcium availability from bone and kidney.

reabsorption and bone calcium resorption in concert with PTH. Because dietary sources of vitamin D are relatively uncommon (e.g., fatty fish) and endogenous production through cutaneous exposure of 7-dehydrocholesterol to UVB radiation may be limited by latitude, season, skin tone, or availability of cutaneous precursors (i.e., in the elderly), inadequate vitamin D status has become an area of concern globally. Although debate remains regarding optimal 25(OH) vitamin D concentrations, sustained elevated PTH levels occur at low 25(OH)D concentrations, contributing to the bone loss that occurs in vitamin D deficiency. This secondary hyperparathyroidism causes increased osteoclastic activity and calcium loss from bone, and contributes to osteoporosis (Weaver and Heaney, 2014).

Other hormones also affect calcium metabolism, although whether they function directly or via the PTH-vitamin D pathway is not always delineated. Notably, estrogens are necessary for the maintenance of balance between bone resorption and accretion. The decrease in serum estrogen concentrations at menopause is the primary factor contributing to the elevated rate of bone resorption that occurs at this stage of life, contributing to osteoporosis. Testosterone also inhibits bone resorption, and lack of this hormone with aging can contribute to osteoporosis in men. It may work through the growth hormone/insulin-like growth factor-1 (IGF-1) axis, which stimulates cartilage formation, the formation of $1,25(OH)_2D$ vitamin D, and intestinal calcium absorption. IGF-1 stimulates osteocalcin production, which is required for bone mineralization. Thyroid hormones stimulate bone resorption, and calcium metabolism abnormalities occur in both hyper- and hypothyroidism. Insulin stimulates collagen production by osteoblasts and impairs the renal reabsorption of calcium. Glucocorticoids, sometimes used to treat conditions such as osteoid arthritis, inflammatory bowel disease, and asthma, inhibit both osteoclastic and osteoblastic activity, impair collagen and cartilage synthesis, and reduce calcium absorption. Excessive bone loss often occurs with glucocorticoid treatment or when excessive amounts of the hormone are secreted, such as in Cushing's disease (Weaver and Heaney, 2014).

Influence of life stages

The total body calcium content of the newborn infant is approximately 30 g (0.75 mol), with most skeletal calcium accrued during the third trimester of pregnancy. The efficiency of calcium absorption is highest during infancy (approximately 60%), and the amount absorbed from breast milk does not appear to be affected by calcium consumed in solid foods.

Childhood remains a time of bone mineral accrual, culminating in the pubertal growth spurt, when calcium absorption increases and bone calcium retention peaks at 200–300 mg (5–7.5 mmol) per day under the orchestration of growth hormone, IGF-1, and sex steroids. Rates of bone calcium deposition and net calcium accrual peak in girls just before menarche and decline thereafter; peak rates of calcium accrual occur approximately 1.5 years later in boys and bone mineralization persists later in boys as well. Forty percent of adult bone mass is acquired during pubertal growth, and while bone mass may continue to accumulate until approximately 30 years of age, relatively little is gained after 18 years of age. Thus, it is important to optimize bone mineralization during growth to ensure adequate bone mineral stores are present to defer the risk of osteoporosis later in life (Weaver and Heaney, 2014).

During pregnancy, a relatively small amount of calcium, approximately 625–750 mmol, is transported to the fetus. Most of this calcium is thought to be obtained through greater efficiency of maternal intestinal calcium absorption, possibly induced by increases in $1,25(OH)_2D$ vitamin D production. For this reason, a greater calcium intake during pregnancy is probably not required, although urinary calcium excretion also increases as plasma volume expansion increases the filtered load of the kidneys. Most studies have reported that there is no increase in intestinal calcium absorption during lactation even when dietary intake of the mineral is relatively low. Changes in biochemical markers and kinetic studies using isotopes indicate that the source of much of the calcium secreted in breast milk is the maternal skeleton, as well as more efficient renal reabsorption and subsequently lower urinary excretion of the mineral. Bone calcium is restored at the end of lactation as the infant is weaned, when ovarian function returns and menstruation resumes. At this time, intestinal calcium absorption increases, urinary calcium remains low, and bone turnover rates decline to normal levels. There is no strong evidence that lactation per se or maternal calcium intake during lactation affect later risk of osteoporosis in women. Thus, there is no strong rationale for increasing maternal calcium intake during lactation. Breast milk calcium concentration is relatively unaffected by maternal intake, and it remains stable throughout lactation (Weaver and Heaney, 2014).

Menopause begins a period of bone loss that extends until the end of life. It is the major contributor to higher rates of osteoporotic fractures in older women. The decrease in estrogen at menopause is associated with accelerated bone loss, particularly of the spine; in the first five postmenopausal years, approximately 15% of skeletal calcium may be lost. Calcium absorption becomes less efficient, and urinary calcium excretion increases, resulting in a declining calcium balance of approximately 30 mg day^{-1}. Hormone replacement therapy was an effective means of improving calcium balance and bone health, but is generally no longer recommended in this age group due to other potential health risks (e.g., increased risk of cardiovascular outcomes such as coronary heart disease and stroke). Bone turnover markers (such as serum N-terminal propeptide of type I procollagen and C-terminal collagen cross-link) are biological molecules liberated during bone formation or bone resorption, can be detected in blood or urine and have been suggested as an additional tool to monitor the response of treatment of osteoporosis. Nevertheless, these new markers remain emerging and are primarily used in research studies (Song, 2017). Calcium supplements alone have only a moderate impact in preventing postmenopausal bone loss, although calcium and vitamin D supplements are recommended to complement other therapies to treat osteoporosis.

Race and genetics

Dietary calcium intake and fecal calcium excretion explained almost 85% of the heterogeneity of calcium retention. Chinese adults could maintain a positive calcium balance with plant-based diets at calcium intakes as low as 300 mg/d through increasing fractional calcium absorption and decreasing calcium excretion in urine and feces (Fang et al., 2016).

Health consequences of calcium deficiency

Skeletal

Between 60% and 80% of the variance in peak bone mass is explained by genetics, including polymorphisms in the vitamin D-receptor gene and in genes responsible for IGF-1 and collagen production. Moreover, dietary and lifestyle factors besides calcium contribute to bone health, and studies that have provided calcium supplements have often concurrently provided vitamin D, making it difficult to distinguish calcium-specific effects. Nonetheless, it is accepted that calcium intake is critical for assuring optimal bone mass and its preservation (Balk et al., 2017; Cormick and Belizan, 2019; Li et al., 2018). Studies of calcium supplementation in children have shown modest improvements in skeletal mineralization, and although these gains may be transient, with unsupplemented children acquiring similar levels of bone mineral content later, they may help prevent fractures. Studies of calcium or calcium and vitamin D supplementation during menopause have also demonstrated modest gains in bone mineral density, particularly among those whose calcium intakes are lowest (Weaver and Heaney, 2014).

Other skeletal effects of low calcium diets include rickets observed in equatorial regions of the world, where calcium intakes are extremely low and vitamin D status is typically adequate. Calcium deficiency occurs in children who are somewhat older than those in whom vitamin D-deficient rickets might appear (i.e., 1–3 years of age versus infancy), and is resolved with provision of calcium rather than vitamin D. It is possible that susceptibility of some children to rickets is due to efficiency of calcium utilization that might be explained by different vitamin D receptor types.

Nonskeletal

Calcium has been investigated as a protective agent in a variety of chronic diseases, including cancer, cardiovascular disease, hypertension, gastrointestinal diseases, metabolic syndrome and other important diseases (Das and Choudhuri, 2021; Li et al., 2018). Appropriate calcium intake has shown many health benefits, such as reduction of hypertensive disorders of pregnancy, lower blood pressure particularly among young people, prevention of osteoporosis and colorectal adenomas, lower cholesterol values, and lower blood pressure in the progeny of mothers taking sufficient calcium during pregnancy (Cormick and Belizan, 2019). Most data come from observational studies linking dietary calcium intake to these outcomes, and associations may therefore be explained by overall healthful dietary patterns that are associated with increased calcium intake. Data from intervention trials have been less consistent in demonstrating benefits of calcium on these health outcomes, but data from high quality trials with chronic disease incidence as primary outcomes are lacking. The best evidence to date is for a protective effect of calcium on colorectal cancer. There is debate about the role of calcium on cardiovascular outcomes, with some metanalysis showing an increased risk with increased calcium intake, whereas other data demonstrate no effect or protective effects. Calcium supplementation reduces risk of preeclampsia among pregnant women on low calcium intakes (Weaver and Heaney, 2014). Diabetic patients also show characteristic potassium, magnesium, phosphate, and calcium depletion and have been found a homeostatic mechanism that links calcium and diabetes mellitus.

Calcium-rich diets may play a role in energy regulation and reduce the risk of obesity by maintaining low intracellular calcium concentrations in adipocytes, reducing lipogenesis. Although this metabolic pathway has been elucidated, application of calcium-rich diets has not always favored weight loss in controlled trials (Das and Choudhuri, 2021).

In the immune system, calcium signals play a central role in various cellular functions such as proliferation, differentiation, apoptosis, and numerous gene transcriptions (Park et al., 2020). Different microRNAs (small non-coding RNA which regulate the gene expression post-transcriptionally) have been studied recently to elucidate the mechanisms implicated in the protective effect of calcium on chronic diseases. Deregulated microRNAs may have a central role in severe diseases, including cancer, heart disease and a wide range of immune disorders (Diener et al., 2018; Park and Kho, 2021).

Health consequences of calcium excess

Excessive calcium intakes can rarely be achieved through diet, but with increasing utilization of supplements some groups may be particularly susceptible to consequences of calcium excess. Hypercalcemia, circulating calcium above 12 mg dL^{-1} (3 mmol L^{-1}), is the acute effect of calcium excess and is associated with weight loss, fatigue, heart arrhythmias, and soft tissue calcification. In addition, it may result in hypercalciuria and affect renal function, resulting in the so-called milk-alkali syndrome. Kidney stones may result from the precipitation of calcium oxalate in renal tissue, and are more likely to be associated with supplemental rather than dietary calcium, which may be protective. An association of prostate cancer with excess calcium intake is being explored. A recent meta-analysis of clinical trials showed that the use of calcium supplements was significantly associated with the increased

risk of cardiovascular and coronary heart disease by 15%, specifically in postmenopausal women (Li et al., 2018; Myung et al., 2021). These data suggest that calcium supplementation should be prescribed and taken cautiously, accounting for individual patients' risks and benefits. It is desirable to reach the recommended contributions through food, whenever possible (Li et al., 2018).

Conclusion

Calcium is the fifth most abundant element in the human body, and it has numerous structural and regulatory functions. Marginal intake levels may affect skeletal health and lead to changes in cellular conditions that contribute to alterations in cellular metabolism.

Dairy products are the best source of calcium in quantity and quality due to their bioavailability. Nowadays, there is a large gap between countries in calcium intake, and it is mandatory to develop national strategies to improve and optimize intakes. The absorption of this mineral changes throughout life, being relevant during childhood and puberty to achieve peak bone mass.

Calcium has a crucial role in skeletal disease such as osteoporosis, and it is considered a protective agent in a variety of chronic diseases, including cancer, cardiovascular disease, hypertension, gastrointestinal diseases and metabolic syndrome.

References

Bakaloudi, D.R., Halloran, A., Rippin, H.L., et al., 2021. Intake and adequacy of the vegan diet. A systematic review of the evidence. Clin. Nutr. 40 (5), 3503–3521.
Bakirhan, H., Karabudak, E., 2021. Effects of inulin on calcium metabolism and bone health. Int. J. Vitam. Nutr. Res. 1–12.
Balk, E.M., Adam, G.P., Langberg, V.N., et al., 2017. Global dietary calcium intake among adults: a systematic review. Osteoporos. Int. 28 (12), 3315–3324.
Cao, S., Cladis, D.P., Weaver, C.M., 2020. Use of calcium isotopic tracers to determine factors that perturb calcium metabolism. J. Agric. Food Chem. 68 (46), 12886–12892.
Cormick, G., Belizán, J.M., 2019. Calcium intake and health. Nutrients 11 (7), 1606.
Das, S., Choudhuri, D., 2021. Role of dietary calcium and its possible mechanism against metabolic disorders: a concise review. J. Food Biochem. 45 (4), e13697.
Diener, C., Hart, M., Alansary, D., et al., 2018. Modulation of intracellular calcium signaling by microRNA-34a-5p. Cell Death Dis. 9 (10), 1008.
Fang, A.P., Li, K.J., Shi, H.Y., He, J.J., Li, H., 2016. Habitual dietary calcium intakes and calcium metabolism in healthy adults Chinese: a systematic review and meta-analysis. Asia Pac. J. Clin. Nutr. 25 (4), 776–784.
Goyal, A., Anastasopoulou, C., Ngu, M., Singh, S., 2021. Hypocalcemia. In: StatPearls [Internet]. StatPearls Publishing, Treasure Island (FL). Available from. https://www.ncbi.nlm.nih.gov/books/NBK430912/.
Institute of Medicine, 2011. Dietary Reference Intakes for Calcium and Vitamin D. The National Academies Press, Washington, DC. https://doi.org/10.17226/13050.
Li, K., Wang, X.F., Li, D.Y., Chen, Y.C., Zhao, L.J., et al., 2018. The good, the bad, and the ugly of calcium supplementation: a review of calcium intake on human health. Clin. Interv. Aging 13, 2443–2452.
Myung, S.K., Kim, H.B., Lee, Y.J., Choi, Y.J., Oh, S.W., 2021. Calcium supplements and risk of cardiovascular disease: a meta-analysis of clinical trials. Nutrients 13 (2), 368.
Palacios, C., Cormick, G., Hofmeyr, G.J., et al., 2021. Calcium-fortified foods in public health programs: considerations for implementation. Ann. N. Y. Acad. Sci. 1485 (1), 3–21.
Park, J.H., Kho, C., 2021. MicroRNAs and calcium signaling in heart disease. Int. J. Mol. Sci. 22 (19), 10582.
Park, Y.J., Yoo, S.A., Kim, M., Kim, W.U., 2020. The role of calcium-calcineurin-NFAT signaling pathway in health and autoimmune diseases. Front. Immunol. 11, 195.
Song, L., 2017. Calcium and bone metabolism indices. Adv. Clin. Chem. 82, 1–46.
Weaver, C.M., Heaney, R.P., 2014. Calcium. In: Ross, A.C., Caballero, B., Cousins, R.J., Tucker, K.L., Ziegler, T.R. (Eds.), Modern Nutrition in Health and Disease. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, pp. 133–149.

Carbohydrates: Chemistry and classification

Chrysanthi Stylianopoulou, Public Health Nutritionist, Veria, Greece

© 2023 Elsevier Ltd. All rights reserved.

This is an update of C. Stylianopoulos, Carbohydrates: Chemistry and Classification, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 265–271, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00041-6.

Introduction **115**
Functions and types of carbohydrates **115**
Monosaccharides **115**
Classification 115
Stereoisomerism and nomenclature 116
Physical properties 117
Chemical reactions 117
Oxidation to produce sugar acids 117
Reduction to produce sugar alcohols 117
Reactions in alkaline solutions (Maillard) 117
Ester formation 118
Ether formation 118
Disaccharides **118**
Sucrose 118
Maltose 118
Lactose 118
Sugar alcohols 118
Mannitol 119
Xylitol 120
Sorbitol 120
Erythritol 121
Oligosaccharides **121**
Physical properties 121
Chemical reactions 121
Polysaccharides **122**
Starch 122
Glycogen 123
Cellulose 123
Dietary fiber 123
Physical properties 124
Chemical reactions 124
Conclusion **124**
References **124**
Relevant websites **125**

Key points

- Define and describe the types and main functions of carbohydrates
- Distinguish carbohydrates, according to their chemical structure and polymerization
- Define and classify monosaccharides
- Describe stereoisomerism, physical properties and chemical reactions exhibited by monosaccharides
- Briefly describe sugar alcohols and their importance
- Outline the most important oligosaccharides and polysaccharides and explain their physical properties and chemical reactions

Introduction

Carbohydrates are the most abundant constituents of cereals, fruits, vegetables, and legumes and a dietary staple in most countries. Carbohydrate oxidation is the main energy-yielding pathway in the majority of human cells and the most important energy source in human nutrition. They contribute to the texture and flavor of foods and diet variability and palatability.

Carbohydrates are polyhydroxy aldehyde or ketone molecules and their derivatives. They comprise a group of substances with different structures and varying physical, chemical and physiological properties. Dietary carbohydrates are important in maintaining glycemic homeostasis, and gastrointestinal health.

Functions and types of carbohydrates

Carbohydrates are chemically stable organic molecules and comprise of carbon, hydrogen and oxygen atoms (Nelson and Cox, 2017). They are produced in green plants containing the green pigment chlorophyll, by the conversion of carbon dioxide and water, utilizing solar energy, a process known as photosynthesis.

Carbohydrates, also known as saccharides or glycans, are polyhydroxy aldehydes or ketones, or substances hydrolyzed to produce such compounds (Nelson and Cox, 2017). Some carbohydrate molecules have the empirical formula $(CH_2O)_n$, but others may also contain nitrogen, phosphorus, or sulfur.

Carbohydrates are compounds of great biological importance and have an extensive role in all forms of life. They provide energy through oxidation and comprise the greatest part of caloric intake in most human diets (Lewis, 2000). They serve as a short-term energy reserve for bodily functions, form metabolic intermediates and provide the carbon backbone for synthesis of biochemical substances and cell components (Nelson and Cox, 2017). They form part of the structural framework of the nucleotide genome sequence. Furthermore, they are part of structural constituents of cells and tissues. Carbohydrates are linked to proteins and lipids and have important functions in cell interactions.

Carbohydrates are classified depending on their chemical structure and their degree of polymerization into the following categories: sugars (monosaccharides and disaccharides), polyols (sugar alcohols), oligosaccharides and polysaccharides (World Health Organization, 1998). **Table 1** displays carbohydrates based on their chemical structure. "Saccharide" is derived from the Greek word sakcharon, meaning sugar. Monosaccharides and disaccharides are jointly called sugars. Polyols (sugar alcohols) are monosaccharide and disaccharide derivatives that have nutritional importance.

Monosaccharides

Classification

Monosaccharides are the simplest form of carbohydrate and cannot be further hydrolyzed to smaller subunits (Nelson and Cox, 2017). They are comprised of a single polyhydroxy aldehyde (the aldoses) or a ketone unit (the ketoses). Monosaccharides with three, four, five, six, and seven carbon atoms in their backbones are called, according to their chain length, trioses, tetroses, pentoses, hexoses, and heptoses respectively. The two simplest monosaccharides are glyceraldehyde, an aldotriose, and dihydroxyacetone, a ketotriose. The most nutritionally important monosaccharides are the pentoses (5-carbon atom skeleton), e.g., ribose, and the hexoses (6-carbon atom skeleton), e.g., glucose. Other examples of monosaccharides include fructose, galactose, and xylose.

The most nutritionally important and abundant monosaccharide is glucose, which is used as the major cell fuel in the human body and can be found unbound in body tissues and fluids (Lewis, 2000). Glucose is the building block of several polysaccharides. Galactose is the monosaccharide produced from the hydrolysis of lactose, which is present in milk (Ball et al., 2014). It can also be synthesized from glucose in the human body. Galactose can be used as cell fuel, but it is also a component of glycolipids, found in nerve cells. Fructose (levulose) is present in honey and fruits, and can also be used as cell fuel. Fructose is the sweetest of all monosaccharides.

Table 1 Chemical classification of carbohydrates.

Class	*Subgroup*	*Components*
Sugars	Monosaccharides	Glucose, fructose, galactose
	Disaccharides	Sucrose, lactose, maltose
Polyols (sugar alcohols)	Monosaccharide derivatives	Erythritol, xylitol, mannitol, sorbitol
	Disaccharide derivatives	Lactitol, isomalt, maltitol
Oligosaccharides	Malto-oligosaccharides	Maltodextrins
	Non-digestible oligosaccharides	Raffinose, stachyose, fructo-oligosaccharides, verbacose
Polysaccharides	Starch	Amylose, amylopectin, modified starches
	Non-starch polysaccharides	Cellulose, hemicellulose, pectins, hydrocolloids (gums)

Adapted from World Health Organization (1998). Carbohydrates in Human Nutrition. Report of a Joint FAO/WHO Expert Consultation. FAO Food and Nutrition Paper 66, 1–140.

Table 2 Some nutritionally important monosaccharides and monosaccharide derivatives.

Class	*Species*	*Significance*
Hexoses	D-glucose	Major cell fuel, unbound in body fluids and tissues, building block of several polysaccharides
	D-fructose	Cell fuel, constituent of sucrose
	D-galactose	Cell fuel, constituent of lactose
	D-mannose	Constituent of plant cell wall polysaccharides and gums
Pentoses	L-arabinose, D-xylose	Constituent of plant cell wall polysaccharides
	D-ribulose, D-xylulose	Metabolite in pentose pathway
	D-ribose	RNA constituent
Uronic acids	D-glucuronic, D-galacturonic	Constituent of plant cell wall polysaccharides
	D-mannuronic, D-guluronic	Constituent of algal polysaccharides
Sugar alcohols	D-glucitol, D-xylitol	Food ingredient, sweetener
	D-galactitol	Metabolite of galactose
Deoxysugars	D-deoxyribose	DNA constituent
	D-deoxygalactose	Constituent of algal polysaccharides
	L-fucose	Constituent of bacterial polysaccharides
	L-rhamnose	Constituent of pectic plant polysaccharides
Aminosugars	D-glucosamine, D-galactosamine	Constituent of aminosaminoglycans, cartilage

Aldose or ketose derivatives of monosaccharides can be produced by substitution of a hydroxyl group or by oxidation of a carbon atom to a carboxyl group to supply esters, ethers or deoxy and amino compounds (Lewis, 2000). In addition, monosaccharides can form sugar alcohols and sugar acids (uronic, aldonic, and aldaric acids). When they are attached to proteins and lipids, important compounds are synthesized, such as glycoproteins, proteoglycans and glycolipids. These compounds have important structural and functional roles. The most important monosaccharides and monosaccharide derivatives and their significance are outlined in **Table 2**.

Stereoisomerism and nomenclature

Carbohydrates demonstrate stereoisomerism. In general, stereoisomers have the same chemical and structural formulas, but different spatial arrangement of their atoms (Nelson and Cox, 2017).

Monosaccharides have optical isomers, which is a type of stereoisomer (Nelson and Cox, 2017). If a molecule's mirror image is non-superimposable on itself, then it has an optical isomer. These non-superimposable mirror images are named enantiomers of each other. Asymmetric or chiral carbons are carbons with different functional groups attached to them. The presence of chiral carbons in monosaccharides gives rise to optical activity. Optical activity is the phenomenon occurring when polarized light is passed through a solution of monosaccharide compounds, during which the plane of light will be rotated either to the left (Levorotatory or L form) or to the right (Dextrorotatory or D form). Consequently, similar structures of the same monosaccharide are produced, namely of the D or the L configuration. Monosaccharides of the D form are biologically significant and nutritionally important, because most naturally occurring monosaccharides are D stereoisomers, and metabolic and digestive enzymes are specific for them. Monosaccharides that have a different configuration about a single carbon atom are called epimers.

For simplicity, three-dimensional monosaccharide structures are drawn as straight chain molecules using Fischer projection formulas (**Fig. 1**). However, in aqueous solutions, monosaccharides with five or more carbon atoms exist mainly as cyclic or

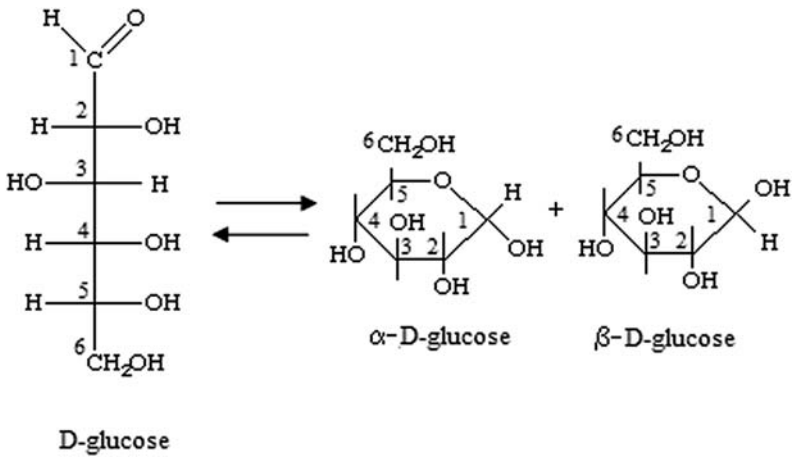


Fig. 1 D-glucose molecule shown as open chain and as a cyclic pyranose ring in the α and β configuration.

ring structures (Nelson and Cox, 2017). This is another type of stereoisomerism exhibited by monosaccharides, in which the carbonyl group has formed a covalent bond with the oxygen of a hydroxyl group along the chain. This stereoisomerism results from a general reaction between alcohols and aldehydes or ketones, to form hemiacetals or hemiketals.

Cyclization can produce two stereoisomers of the α and β configuration, binding either in the "front" or in the "back" of the carbon atom (Nelson and Cox, 2017). Monosaccharide isomers, which have different configuration about the hemiactal or hemiketal carbon atom, are named anomers and the carbonyl carbon atom is named anomeric carbon. Generally, an equilibrium mixture of the straight and the cyclic form exists in monosaccharide solutions. Carbohydrates can change spontaneously between the two α and β configurations by forming an intermediate straight chain, a process called mutorotation. The pentoses form furanose rings (5-carbon ring) and the hexoses form pyranose rings (6-carbon ring). For example, D-glucose occurs in solution as a hemiacetal where the unbound hydroxyl group at C-5 has attached to the C-1 of the aldehyde, resulting in an asymmetric carbon and two potential anomers, α-D-glucose and β-D-glucose. **Fig. 1** illustrates D-glucose in its pyranose form in the α and β configuration. The isomerization produces compounds with different properties and has major metabolic importance, because of enzyme specificity for particular stereoisomers.

The nomenclature for monosaccharides specifies the configuration at each anomeric carbon (for example, α or β), the optical rotation (for example, D or L) and the common name of the monosaccharide (for example, glucose).

Physical properties

Monosaccharides are usually colorless and crystalline solids at room temperature (Ball et al., 2014). They exhibit increased solubility in water. Monosaccharides are much more water soluble than other molecules of similar molecular weight. For example, glucose can dissolve in very small amounts of water (1 g glucose to 1 mL of water) to form a syrup. On the contrary, they are insoluble in nonpolar organic solvents. Furthermore, they exhibit limited solubility in pure alcohols, but are very soluble in aqueous alcohol solutions, so these solutions are widely used for extraction and analysis. The majority of monosaccharides are characterized by their sweet taste.

Chemical reactions

The chemical properties of monosaccharides are established by the functional groups they contain (Ball et al., 2014). Hydroxyl groups accessible for reactions are present in all carbohydrate molecules (Shendurse and Khedkar, 2016). Most monosaccharides exist in cyclic hemiacetals. However, when they are in solution, they exist in equilibrium with the aldehydes or ketones in open chain conformation. Consequently, they undergo chemical reactions of aldehydes, ketones, alcohols and hemiacetals.

Oxidation to produce sugar acids

Since the cyclic hemiacetal forms of monosaccharides exist in equilibrium with the open chain aldehydes, they undergo oxidation to form carboxylic acids. (Stoker, 2013). Consequently, monosaccharides are reducing sugars, meaning that they are capable of getting reduced without being hydrolyzed. A reducing sugar will contain an aldehyde group (e.g., glyceraldehyde), a hydroxyketone group (e.g., fructose) or a cyclic hemiacetal group (e.g., glucose).

Oxidation results in three types of acids, depending on the type of oxidizing agent used:

- Mild reagents oxidize the aldehyde end of the monosaccharide resulting in aldonic acids: Tollen's reagent ($Ag+$ in aqueous ammonia), Fehling reagent ($Cu2+$ complex with tartrate ion), and Benedict's reagent ($Cu2+$ complex with citrate ion). The Fehling and Benedict's reagents produce a brick red copper (I) oxide product and the Tollen's reagent results in silver mirror formation, and therefore confirm the presence of monosaccharides in solution (Stoker, 2013).
- Strong oxidizing agents oxidize both ends of the monosaccharide simultaneously to produce aldaric acids (Stoker, 2013).
- Enzymes catalyze the oxidation of the primary alcohol end of the monosaccharides, without oxidizing the aldehyde end and produce alduronic acids (Stoker, 2013). These enzymes are called oxidases and are substrate specific. For example, the enzyme glucose oxidase catalyzes only the conversion of β-D-glucopyranose (not the α-D-glucopyranose) to D-glucono-δ-lactone.

Reduction to produce sugar alcohols

Monosaccharides are powerful reducing agents toward a range of metal salts in alkaline solution, due to the presence of aldo- and keto-groups and formation of hemiacetal or acetal groups (Lewis, 2000). The extent of reduction differs among different monosaccharides. The carbonyl group in a monosaccharide (either an aldose or a ketose) goes through reduction, using a hydrogen catalyst. Important compounds formed in this way are sugar alcohols, which typically have a sweet taste. For example, sorbitol is produced by reduction of the aldo group of glucose. Sugar alcohols are utilized widely in the industry, as sweeteners, since they do not promote tooth decay.

Reactions in alkaline solutions (Maillard)

Monosaccharides when present in weak alkaline solutions result in the isomerization of the aldose-keto group, referred as enolization (Lewis, 2000). In stronger alkaline solutions, they produce a series of degradation compounds, saccharinic acids, which condense repeatedly to generate a series of highly colored products, in the presence of ammonia, amino acids and proteins. This

reaction, known as Maillard reaction, results in the browning of food products, and it is utilized in the food industry for the production of caramel colors.

Ester formation

Monosaccharides contain hydroxyl groups and react with acids to form a variety of esters (Nelson and Cox, 2017). The phosphate esters play a major role in carbohydrate metabolism. For example, the first step of glycolysis involves the production of the glucose 6-phosphate ester in a reaction catalyzed by the enzyme glucokinase in the presence of adenosine triphosphate (ATP). Uronic acids react with alcohols to form esters. The methyl esters of uronic acids are the most important in determining the physical properties of uronans.

Ether formation

Monosaccharides undergo substitution reactions with methyl iodide to produce methyl ether derivatives (Lewis, 2000). These compounds have been used to identify the structure of polymers, because the site of nonmethyl substituted groups is indicative of the site of the branch points after hydrolysis. Monosaccharides undergo acetylation, which occurs on the free or the reduced molecule, to produce acetylated alditols. These volatile compounds have been used to identify sugar mixtures by Gas Liquid Chromatography (GLC).

Disaccharides

Disaccharides are composed of two monosaccharide units joined together by acetal formation (Nelson and Cox, 2017). One monosaccharide plays the role of a hemiacetal and the other one plays the role of an alcohol; the resulting ether bond is a glycolytic linkage (Stoker, 2013). Glycosidic bond is the bond formed by the condensation of the hydroxyl group of one monosaccharide hemiacetal with another monosaccharide hydroxyl group. Disaccharides are linked together with glycosidic bonds in the α or β orientation. The common nomenclature for disaccharides specifies the order of monosaccharide units, the configuration at each anomeric carbon, and the carbon atoms involved in the glycosidic bond(s).

Sucrose

Sucrose is a non reducing disaccharide and it is the most abundant disaccharide found in plants (Stoker, 2013). Sucrose consists of a molecule of an α-anomeric carbon 1 of glucose linked together with a β-anomeric carbon 2 of fructose, resulting in an α-1,2-glycosidic bond (**Fig. 2**A). Sucrose is a commercial product of sugar cane and sugar beet.

Maltose

Maltose is a reducing disaccharide and is encountered in corn syrup, malt and germinating seeds (Stoker, 2013). Maltose is mainly produced by partial hydrolysis of starch and consists of two glucose units linked by an α-1,4 glycosidic bond (**Fig. 2**C).

Lactose

Lactose is a reducing disaccharide found in milk and dairy products (Stoker, 2013). Human milk contains 7–8% lactose, whereas cow's milk contains 4–5% lactose. Lactose consists of galactose and glucose linked by a β-1,4 glycosidic bond (**Fig. 2**B).

A number of important disaccharide derivatives are synthesized by substitution of a hydroxyl group or by oxidation of a carbon atom to a carboxyl group, such as glycoproteins and glycolipids (Lewis, 2000). These compounds have important structural and functional roles. Some nutritionally important disaccharides and disaccharide derivatives and their significance are outlined in **Table 3**.

Disaccharides have similar physical and chemical properties to their corresponding monosaccharide constituents (Nelson and Cox, 2017). Disaccharides are very soluble in water and insoluble in organic polar solvents (Lewis, 2000). Sucrose is extremely soluble in water, while lactose is soluble to a lesser extent. Disaccharides have similar reducing properties with monosaccharides, due to the presence of an aldo- or a keto-group and formation of hemiacetal or acetal groups. However, sucrose is a non-reducing sugar, since it has no free hemiacetal group. Disaccharides in mild acidic conditions are hydrolyzed to their constituent monosaccharides. They are also hydrolyzed in specific enzymatic solutions; therefore, enzymatic hydrolysis is a useful method for the analysis of sugar mixtures. Disaccharides undergo esterification reactions, as do monosaccharides. Sucrose, for example, reacts with fatty acids to produce nondigestible esters, which have similar properties to triacylglycerols.

Sugar alcohols

Sugar alcohols (polyols, polyhydric alcohols, polyalcohols, alditols or glycitols) are monosaccharide and disaccharide derivatives that have the aldehyde or ketone moiety reduced to an alcohol (Yebra-Biurrun, 2005). They occur naturally in many fruits and vegetables or can be produced industrially from sugars, primarily by enzymatic hydrolysis or fermentation (Wang, 2003). They have

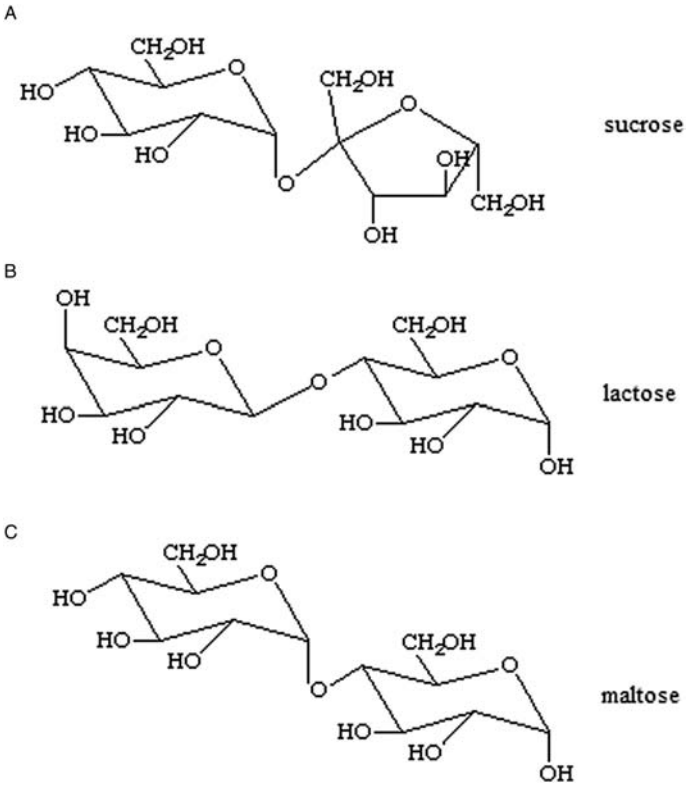


Fig. 2 The molecular structures of (A) sucrose (B) lactose and (C) maltose.

Table 3 Some nutritionally important disaccharides and disaccharide derivatives.

Class	*Species*	*Significance*
Disaccharide	Sucrose	Constituent of fruits, vegetables and sweetener
	Lactose	Milk and dairy products
	Maltose, isomaltose	Constituent of starch
	Trehalose	Food additive and constituent of mushrooms
	Lactulose	Lactose derivative, laxative
Disaccharide alcohols	Maltitol	Constituent of starch, sweetener
	Lactitol	Constituent of lactose, sweetener

a sweet taste and are used extensively in the industry as sugar substitutes and thickeners (Awuchi, 2017). Most commonly used are the monosaccharide derivatives sorbitol, mannitol, xylitol and erythritol, the disaccharide derivatives isomalt, maltitol and lactitol, as well as the sugar alcohol mixture hydrogenated starch hydrolyzates (HSH).

Sugar alcohols have received increased attention, because of their desirable properties: relative sweetness, cooling sensation in the mouth, limited digestion and absorption (Awuchi, 2017). They provide low caloric content, blunted glycemic response, and reduced risk for dental caries. Every sugar alcohol can have the health claim, according the United States Food and Drug Administration (FDA) and European Commission, that is "does not promote tooth decay" (Food and Drug Administration, 2005; European Food Safety Agency, 2011). Sugar alcohols have been approved by the FDA as Generally Recognized as Safe (GRAS) or as food additives (American Dietetic Association, 2004). The Joint Expert Committee on Food Additives (JECFA) of the World Health Organization (WHO) has not specified an acceptable daily intake (ADI) for sugar alcohols, placing no limits for their use (World Health Organization, 2021).

Nonetheless, their excessive consumption may cause gastrointestinal distress, such as diarrhea, flatulence and bloating, with the exception of erythritol (Awuchi, 2017). The majority of people develop tolerance to sugar alcohols and increased consumption diminishes the gastrointestinal side effects. **Table 4** outlines the relative sweetness and food energy of the most commonly encountered sugar alcohols.

Mannitol

Mannitol is utilized as a sweetener, bulking and texturing agent. It is naturally occurring in figs, olives, larches, edible fungi, yeast, seaweed and exudates of some tress (Grembecka, 2015).

Table 4 Sugar alcohols' relative sweetness.

Name	*Sweetness (relative to sucrose = 1.0)*	*Food energy (kcal gram^{-1}) (sucrose = 4.0)*	*Sweetness per food energy (relative to sucrose = 1.0)*	*Food energy for equal sweetness (relative to sucrose = 100%)*
Erythritol	0.8	0.21	15	6.7%
Hydrogenated starch hydrolyzates (HSH)	0.4–0.9	3.0	0.52–1.2	83–190%
Isomalt	0.5	2.0	1.0	100%
Lactitol	0.4	2.0	0.8	125%
Maltitol	0.9	2.1	1.7	59%
Mannitol	0.5	1.6	1.2	83%
Sorbitol	0.6	2.6	0.92	108%
Xylitol	1.0	2.4	1.6	62%

Adapted from Awuchi, C. (2017). Sugar Alcohols: Chemistry, Production, Health Concerns and Nutritional Importance of Mannitol, Sorbitol, Xylitol, and Erythritol. International Journal of Advanced Academic Research 3(2), 32–66.

Mannitol is approximately half as sweet as sucrose and non-cariogenic. It has a pleasant taste and a low caloric value. It is only partially taken up by the small intestine (Awuchi, 2017). It has a minimal effect on blood glucose, much smaller than the effect of sucrose (Grembecka, 2015). Mannitol inhibits moisture absorption from the air and sugar crystallization, and is chemically inert (Awuchi, 2017). Mannitol has low solubility in water, limited solubility in organic solvents (for example ethanol), and is almost completely insoluble in ether, ketones and hydrocarbons (Awuchi, 2017). Mannitol is an isomer of sorbitol and has a different orientation of the hydroxyl group at carbon 2. It has a cooling effect in the mouth, because it has a negative heat of solution. It is stable at all temperatures and has a high melting point (Grembecka, 2015).

Mannitol is obtained from the reduction of mannose. Industrial methods of mannitol production include industrial synthesis, biosynthesis, chemical process, fermentation etc. The most common industrial method is fructose hydrogenation, and it is derived from starch or sucrose (Awuchi, 2017).

Mannitol is extensively utilized in the pharmaceutical and nutraceutical industry, to make chewable tablets, granulated powders and to mask the bad taste of certain drugs (Awuchi, 2017). Furthermore, it is used in the food industry to lengthen the shelf life of packaged products, as a bulking agent and in chewing gum (Awuchi, 2017; Grembecka, 2015). Mannitol is utilized in the medical field as a diuretic medication to reduce ocular and intracranial pressures (Awuchi, 2017). Furthermore, it is used for the treatment of high blood pressure. Mannitol may facilitate drug crossing across the blood-brain barrier and may prevent kidney failure and cerebral edema in various surgeries. It is also known as an antioxidant and a scavenger of hydroxyl radicals (Grembecka, 2015).

Xylitol

Xylitol is mainly utilized as a sweetener. It is naturally occurring in fruits and vegetables and can be obtained from berries, oats, mushrooms, corn husks and sugar cane (Awuchi, 2017).

Xylitol has about a third fewer calories than sucrose, with the same sweetness and no unpleasant aftertaste (Awuchi, 2017). About half of xylitol is taken up by the small intestine and 50–75% is metabolized in the liver or by the intestinal flora (Grembecka, 2015). Xylitol has a low glycemic index and has a minimal effect on blood glucose (Awuchi, 2017). Its metabolism does not need the presence of insulin and it has a cooling effect, since it has a negative heat of solution (Grembecka, 2015). Xylitol is non cariogenic and may cause tooth remineralization and decrease plaque formation (Grembecka, 2015).

Xylitol may be produced by xylan, found in wood material and subsequently hydrolyzed to xylose (Grembecka, 2015). Industrial production of xylitol occurs by catalytic hydrogenation of xylose or by microbial, fungal or yeast fermentation from xylose or hemicellulosic hydrolyzate (Awuchi, 2017).

Xylitol is widely used in the pharmaceutical, nutraceutical and food industries, in products as pastilles, chewing gums and candies, since it has few gastrointestinal side effects and desirable properties (Grembecka, 2015; Awuchi, 2017).

Sorbitol

Sorbitol (also known as glucitol) has a sweet, cool and pleasant taste and it is widely utilized as a sweetener, humectant, softener, texturizing and anti-crystallizing agent (Grembecka, 2015). Sorbitol is naturally occurring in apples, pears, peaches, nectarines, apricots, prunes, dates, raisins and in some vegetables (Grembecka, 2015).

Sorbitol is 60% sweeter than sucrose, with fewer calories and non-cariogenic. As previously stated, sorbitol is an isomer of mannitol. Sorbitol is different than mannitol in physical characteristics, food sources and uses (Awuchi, 2017). It is 20 times more soluble in water than mannitol and it is encountered as a liquid and crystalline solid (Grembecka, 2015). It has a cooling effect in the mouth, since is has a negative heat of solution, similar to xylitol. Sorbitol is chemically inert and stable at all temperatures, and does not take part in Maillard browning reaction (Grembecka, 2015). It can be readily compressed, but it is also highly hygroscopic.

Sorbitol is produced by glucose reduction from corn syrup (Awuchi, 2017). Industrially, it is produced from concurrent catalytic hydrolysis and hydrogenation of maize starch under pressure (Awuchi, 2017).

Sorbitol is widely used in the pharmaceutical, nutraceutical, food and cosmetics industries. It is used in several different products, including mints, cough syrups, chewing gum, confectionery, baked goods and chocolate (Grembecka, 2015). Sorbitol is also used in the production of Vitamin C and may have a laxative effect if consumed excessively.

Erythritol

Erythitol is used as a sweetener for industrial applications. It is naturally occurring in vegetables, fruits (melon, peaches), mushrooms and fermented foods (Grembecka, 2015).

Erythritol has a clean sweet taste, with no aftertaste, and is 60–70% sweeter than sucrose (Awuchi, 2017; Grembecka, 2015). It has almost no calories, it is non-cariogenic. The majority of erythritol (60–90%) is taken up in the small intestine, but it is excreted in the urine almost intact and it has no effect on blood glucose (Awuchi, 2017; Grembecka, 2015). Since minute amounts undergo intestinal fermentation, laxative effects are not usually observed unless excessive amounts are consumed (Grembecka, 2015).

Erythritol demonstrates a strong cooling effect, very similar to xylitol, and among the strongest of all sugar alcohols (Awuchi, 2017). Its taste resembles to sucrose, when blended with artificial sweeteners, other sugar alcohols or stevia (Grembecka, 2015). Its sweetness can by greatly enhanced (by about 30%) by the addition of very small amounts of artificial sweeteners. Erythritol, like sorbitol, does not take part in Maillard browning reaction and is stable in changes of temperature and pH (Grembecka, 2015).

In contrast to other sugar alcohols, erythritol is not produced directly by catalytic hydrogenation of erythrose, due to the elevated cost of erythrose (Grembecka, 2015). Industrial production of erythritol occurs by enzymatic hydrolysis of maize starch and subsequent fermentation of glucose (Awuchi, 2017).

Erythritol is used in the pharmaceutical, nutraceutical and food industries, in products such as chewing gums, candy, ice cream and beverages. It is also used to mask aftertaste of other sweeteners and improve micrological stability, texture and mouthfeel (Grembecka, 2015).

Oligosaccharides

Oligosaccharides consist of short chains of monosaccharide units (usually 3–9 monosaccharides), covalently linked to form large units (Groff and Gropper, 2000). They are named trioses, tetroses, etc., denoting the number of carbons in their molecule. Oligosaccharides are distributed widely in plants, and when digested yield their constituent monosaccharides. The major oligosaccharides consist of the raffinose series, formed by the linkage of galactose, sucrose and glucose units, and the maltose series, formed by the linkage of glucose units. Some nutritionally important oligosaccharides and their significance are outlined in **Table 5**. The common nomenclature for oligosaccharides specifies the order of monosaccharide units, the configuration at each anomeric carbon, and the carbon atoms involved in the glycosidic linkage(s).

Physical properties

Oligosaccharides generally exhibit properties similar to mono- and disaccharides with similar functional groups, but some oligosaccharides with nine monosaccharide units may exhibit similar properties to polysaccharides (Lewis, 2000). Oligosaccharides are less soluble in aqueous alcohol solutions than monosaccharides, and their solubility decreases as the number of monosaccharide units increases.

Chemical reactions

Oligosaccharides in mild acidic conditions are hydrolyzed to their constituent monosaccharides (Lewis, 2000). The fructofuranosyl linkages of the fructooligosaccharides are quite susceptible to acid hydrolysis. Oligosaccharides are also susceptible to enzymatic hydrolysis. The maltooligosaccharides can be rapidly hydrolyzed by glucosidase enzymes. Furthermore, they undergo esterification reactions.

Table 5 Some nutritionally important oligosaccharides.

Class	*Species*	*Significance*
Maltoses	Maltotriose, maltotetraose	Constituent of starch
Raffinoses	Raffinose, stachyose, verbascose	Constituent of vegetables and legumes
Fructoses	Fructotriose	Constituent of cereals, tubers
Lactoses	Fucosyl lactoses	Constituent of human breast milk

Polysaccharides

Polysaccharides consist of long chains (>9) of monosaccharide residues linked by glycosidic bonds (Nelson and Cox, 2017). These compounds consist of several hundred or even thousands of monosaccharide units. The properties of polysaccharides are determined by the species of monosaccharides in the polymer backbone, the type of linkages between residues, and the extent and type of chain branching.

Glucans are polymers of glucose and the major polysaccharides in the diet (Groff and Gropper, 2000). The most important glucans are starch, glycogen, and cellulose. Glycogen is the short-term storage form of glucose in animal tissues. Starch is the most common digestible storage polysaccharide in plants and cellulose is a major structural component of plant cell walls (**Fig. 3**). Some nutritionally important polysaccharides and their significance are outlined in **Tables 6** and **7**.

Polysaccharide molecules can be linear or branched (Lewis, 2000). Branches can be formed through any unlinked hydroxyl group, and vary from alternating and consecutive single unit branches to multiple unit branches (ramified structure). Polysaccharides with α linkages have a helical shape, e.g., the amylose starch molecule, while those with β linkages generally have a linear or flat ribbon-like molecule, e.g., cellulose (**Fig. 3**).

Starch

The most important, abundant and digestible polysaccharide in human nutrition is starch. Starch comprises of large chains of α linked D-glucose residues (Nelson and Cox, 2017). There are two forms of starch: amylose and amylopectin. Both forms of starch can be found in cereals, potatoes, legumes and other vegetables, with amylopectin comprising 80–85% and amylose 15–20% of total starch.

Amylose is a linear, unbranched form of starch, which consists of α-1,4 linked glucose units (**Fig. 3**A). Amylose forms a flexible chain and its molecules are shaped into a helix. This tight conformation of the amylose molecules result in a decreased digestion rate compared to other starches. Amylose helices bind to iodine molecules, resulting in a blue purple color, which is often used to confirm the presence of starch.

Amylopectin is a branched-chain polymer, which consists primarily of α-1,4-linked glucose units with occasional α-1,6 linked glucose units (Ball et al., 2014). The α-1,6-glycosidic bond in amylopectin is responsible for the branching. An amylopectin molecule may include thousands of glucose molecules. Branching occurs every 25–30 glucose units and disrupts the helical structure of amylopectin. As a result, when amylopectin binds to iodine molecules, a less intense reddish color is produced compared to amylose.

Starches are usually insoluble in water because they have a high molecular weight (Stoker, 2013). Starches, similarly to other polysaccharides, form thick colloidal dispersions when heated in water solutions, since they contain large number of hydroxyl groups. Consequently, they are used as thickening agents.

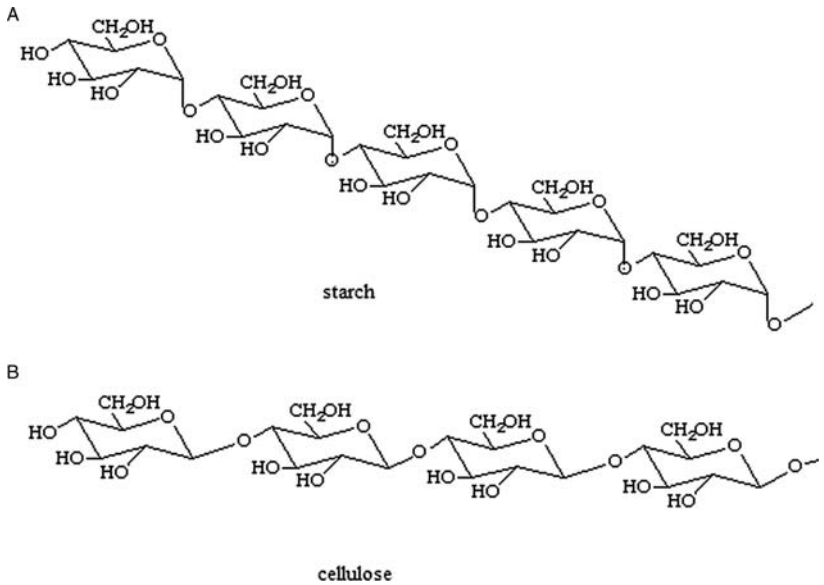


Fig. 3 (A) Five units of an α-1,4-D-glucopyranose chain form a starch molecule (amylose) (B) Four units of a β-1,4-D-glucopyranose chain form a cellulose molecule.

Table 6 Some nutritionally important polysaccharides.

Class	*Species*	*Significance*
Glucans	Starch	Storage polysaccharide in plants
	Glycogen	Short-term storage form of glucose in animal tissues
	Cellulose	Major structural component of plant cell walls
Galactans		Major constituents of noncellulosic matrix of plant cell wall
Xylans		Constituents of mature plant tissues
Mannans		Storage forms in several plants
Uronans	Galacturonans	Major components of water-soluble pectic fraction of plants
	Mannuronans	Components of algal polysaccharides
	Guluronans	Components of algal polysaccharides

Table 7 Some nutritionally important starch and nonstarch polysaccharides.

Class	*Species*	*Significance*
Starch	Amylose, amylopectin	Most common digestible plant polysaccharides
Nonstarch	Cellulose	Major component of plant cell wall
	Pectin	Constituent of plant cell wall, food additive
	Hemicellulose	Constituent of plant cell wall
	Gums, mucilages	Plant hydrocolloids, food additives
	Algal polysaccharides	Constituents of algae and seaweed, food additives

Glycogen

Glycogen is a branched polysaccharide, similar to amylopectin, consisting of both α-1,6, and α-1,4 glycosidic bonds (Ball et al., 2014). Glycogen is more highly branched than amylopectin, and branching occurs every 8–12 glucose units. Glycogen branches are shorter than amylopectin branches. Glycogen is degraded into D-glucose to produce energy.

All mammalian cells comprise of glycogen, but it is found more abundantly in liver (4–8% by tissue weight) and muscle (0.5–1.0% by tissue weight) (Ball et al., 2014). During fasting periods, the organism initially utilizes glycogen to achieve metabolic balance. The majority of glycogen is stored in skeletal muscle, since the body is comprised of a significantly greater muscle mass compared to liver.

Cellulose

Cellulose is a long, unbranched polysaccharide, similar to amylose, consisting of β-1,4-glycosidic bonds (Ball et al., 2014). Cellulose consists of 300–3000 glucose units and it has a very rigid and extended structure. Cellulose is an important structural polysaccharide and the most plentiful organic compound on the planet. Furthermore, it forms part of plant cell walls.

The majority of animals do not have the enzymes to break down cellulose, but it promotes intestinal contraction and digestion (Ball et al., 2014). Ruminant animals, such as cows, sheep and goats, break down cellulose, with the assistance of bacterial colonies in their digestive track and the use of multiple stomachs to provide additional time needed for cellulose digestion. Other animals (for example horses) have longer intestinal tracts, while others (for example rabbits) have the ability to re-digest food so that cellulose can be processed. Cellulose is an important compound in various industrial processes, since it is part of wood, paper, cotton, cellophane, rayon, linen, nitrocellulose, etc.

Dietary fiber

Dietary Fiber generally refers to complex carbohydrates which compose the cells walls and structural components of plants (World Health Organization, 1998). Good sources of fiber include whole cereal grains, fruits and vegetables. Dietary fiber may be soluble in water, for example pectin. Soluble fiber binds to other carbohydrates and delays digestion and absorption, therefore assisting in maintaining stable blood glucose levels. Insoluble fiber, such as cellulose, bulks up stool and aids in intestinal waste elimination.

Dietary fiber consumption has potential important health benefits, including general gastrointestinal health and prevention of several non-communicable diseases, through blood cholesterol reduction and regulation of blood sugar levels (World Health Organization, 1998).

Physical properties

Polysaccharides are insoluble in water, and have no sweet taste (Nelson and Cox, 2017). In general, they form colloidal solutions in water, while some other polymers are extremely insoluble in water and require prior treatment with acid, alkali or organic solvents to get them to dissolve (Lewis, 2000). For example, β-1,4 mannans and glucans (e.g., cellulose) are very insoluble due to hydrogen bonding between parallel chains. On the other hand, arabinoxylans are readily soluble in water, because the arabinosyl chains inhibit hydrogen bonding. Galactomannans are also readily soluble in water, producing viscous solutions, and are used as food additive gums. The α-linked glucans (e.g., amylose and amylopectin) have completely different solubilities. The α-1,4 amylose is very soluble in warm water and forms colloidal solutions. When the amylose chains cool down, they form an amylose gel, which subsequently forms an insoluble crystalline material. Amylopectins are also very soluble in hot water, but do not form an insoluble crystalline material to the same degree as amylose.

Chemical reactions

Polysaccharides show slower reactions rates than other saccharides due to steric effects. Polysaccharides usually contain one reducing group at the terminal end of the polymer chain and, as a result, have low reducing properties.

Polysaccharides are also hydrolyzed to their constituent monosaccharides by acid hydrolysis, but the conditions necessary for complete hydrolysis depend on the solubility of the polymers (Lewis, 2000). The majority of polysaccharides (e.g., starch) are completely hydrolyzed under weak acidic conditions. On the contrary, cellulose requires treatment with strong acid for several hours prior to hydrolysis, and subsequent heating under weak acidic conditions for the completion of the reaction. The uronans are very resistant to complete acid hydrolysis and generally produce disaccharides of aldobiuronic acids. Acid hydrolysis of polysaccharides results in extensive losses of their monosaccharide constituents.

Polysaccharides are more efficiently hydrolyzed to their monosaccharide constituents using specific enzymes (Stoker, 2013). Fungal enzymes act specifically for the hydrolysis of different polysaccharides. The α-1,4 glycosidic linkages in starch can be hydrolyzed by various α-amylases (e.g., salivary and pancreatic), producing maltose and isomaltose. The β-1,6 glycosidic linkages in amylopectin are not as easily hydrolyzed and require the presence of the fungal enzyme pullulanase to complete the hydrolysis.

Polysaccharides can also produce ester compounds (Lewis, 2000). For example, galacturonans, which are composed of an α-1,4 galacturonic acid chain with integrated rhamnose units, form salts with cations and may be esterified with methoxyl groups.

Conclusion

Carbohydrates are compounds of great biological importance and have an extensive role in all forms of life. Carbohydrates vary enormously from short and compact sugar molecules to long and highly branched polysaccharide chains. They are encountered in every living cell and provide energy, structural support, and metabolic efficiency. Carbohydrates are nutritionally important as a dietary staple and have extensive industrial applications.

References

American Dietetic Association, 2004. Position of the American Dietetic Association: nutritive and nonnutritive sweeteners. J. Am. Diet Assoc. 104 (2), 255–275.

Awuchi, C., 2017. Sugar alcohols: chemistry, production, health concerns and nutritional importance of mannitol, sorbitol, xylitol, and erythritol. Int. J. Adv.Acad. Res. 3 (2), 32–66.

Ball, D.W., Hill, J.W., Scott, R.J., 2014. The Basics of General, Organic, and Biological Chemistry. Flat World Knowledge.

European Food Safety Agency, 2011. Scientific opinion on the substantiation of health claims related to the sugar replacers xylitol, D-tagatose, xylitol, sorbitol, mannitol, maltitol, lactitol, isomalt,erythritol, D-tagatose, isomaltulose, sucralose and polydextrose and maintenance of tooth mineralisation by decreasing tooth demineralisation. EFSA J. 9 (4), 2076.

Food and Drug Administration, 2005. Food labeling; health claims; dietary noncariogenic carbohydrate sweeteners and dental caries. Fed. Regist. 70 (92), 25496–25502.

Grembecka, M., 2015. Sugar alcohols—their role in the modern world of sweeteners: a review. Eur. Food Res. Technol. 241, 1–14.

Groff, J.L., Gropper, S.S., 2000. Advanced Nutrition and Human Metabolism, third ed. Wadswoth, Belmont.

Lewis, B.A., 2000. Structure and properties of carbohydrates. In: Stipanuk, M. (Ed.), Biochemical and Physiological Aspects of Human Nutrition, first ed. WB Saunders Company, Philadelphia, pp. 3–22.

Nelson, D.L., Cox, M.M., 2017. Lehninger Principles of Biochemistry, seventh ed. WH Freeman, New York.

Shendurse, A.M., Khedkar, C.D., 2016. Glucose: properties and analysis. In: Caballero, B., Finglas, P., Toldrá, F. (Eds.), The Encyclopedia of Food and Health, vol. 3. Academic Press, Oxford, pp. 239–247.

Stoker, H.S., 2013. General, Organic, and Biological Chemistry, sixth ed. Cengage Learning, Boston.

Wang, Y.-J., 2003. Saccharides: modifications and applications. In: Tomasik, P. (Ed.), Chemical and Functional Properties of Food Saccharides, first ed. CRC Press, Boca Raton, pp. 35–421212.

World Health Organization, 1998. Carbohydrates in Human Nutrition. Report of a Joint FAO/WHO Expert Consultation, vol. 66. FAO Food and Nutrition Paper, pp. 1–140.

World Health Organization, 2021. Joint Expert Committee on Food Additives Database. Available at. https://www.who.int/foodsafety/areas_work/chemical-risks/jecfa/en/. (Accessed 22 March 2021).
Yebra-Biurrun, M.C., 2005. Sweeteners. In: Worsfold, P., Poole, C., Townsend, A., Miro, M. (Eds.), Encyclopedia of Analytical Science, second ed. Elsevier, London, pp. 562–572.

Relevant websites

http://www.dh.gov.uk/.
http://www.efsa.europa.eu/en.
http://www.fda.gov/.
http://www.iom.edu/.
http://www.nutrition.org.uk/.
http://www.who.int.

Carbohydrates: Regulation of metabolism

Chrysanthi Stylianopoulou, Public Health Nutritionist, Veria, Greece

© 2023 Elsevier Ltd. All rights reserved.

This is an update of C. Stylianopoulos, Carbohydrates: Regulation of Metabolism, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 272–277, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00042-8.

Introduction **127**
Starch digestion 127
Digestion of disaccharides 127
Undigested glycosidic bonds 128
Absorption 128
Transport 128
Factors affecting carbohydrate assimilation 128
Carbohydrates and energy metabolism **128**
Glucose 128
Fructose and galactose 129
Glucose production by the liver and kidneys (glycogen) **130**
Gluconeogenesis 130
Glycogenolysis 130
Control of carbohydrate metabolism **130**
Hormonal regulation 130
Insulin 131
Glucagon 131
Catecholamines 131
Glucocorticoids 132
Allosteric enzyme regulation 132
Directional shifts 132
Regulation of gene expression 132
Glycogen synthesis and breakdown 133
Diseases of carbohydrate metabolism **133**
Carbohydrate malabsorption 133
Fructose intolerance and essential fructosuria 133
Glucose and galactose malabsorption 133
Lactose deficiency or lactose intolerance 133
Glycogen storage diseases 133
Diabetes mellitus 134
Type 1 diabetes 134
Type 2 diabetes 134
Gestational diabetes 134
Monogenic diabetes 135
Summary **135**
References **135**

Key points

- Outline and describe the process of starch and disaccharide digestion, absorption and transport
- Explain factors that influence carbohydrate assimilation
- Discuss the process of breakdown of glucose, fructose and galactose for energy production
- Explain the process of glucose production by the liver and kidneys
- Describe the control mechanisms that regulate carbohydrate metabolism
- Discuss the principal diseases from defects in carbohydrate metabolism

Introduction

Carbohydrates are an inexpensive and versatile staple of developed and developing countries, accounting for 40–80% of consumption. Simple and complex carbohydrates are important human dietary components. Digestible carbohydrates provide energy and nutrients for the human body. They include monosaccharides (glucose, fructose and galactose), disaccharides (sucrose, lactose and maltose), oligosaccharides (components of polysaccharide breakdown), and polysaccharides (mainly starch), as well as sugar alcohols (sorbitol, mannitol, xylitol and erythritol). Nondigestible carbohydrates, which are also described as dietary fiber, are present in carbohydrate rich foods and are not assimilated by the human body.

All carbohydrates, in order to be utilized by the human body, need to be broken down to monosaccharides through enzymatic cleavage of the glycosidic bond, which is the bond that unites different monosaccharides to form large and complex carbohydrate molecules. In human metabolism, all simple sugars (monosaccharides and disaccharides) are converted into glucose, which is the circulating form of carbohydrates in the bloodstream. Appropriate regulation of glucose metabolism is necessary for cell function and essential for health and survival.

Starch digestion

Carbohydrate digestion occurs with the secretion of the enzyme α-amylase by the salivary and pancreatic exocrine glands in the mouth and small intestine respectively, and it is completed by the secretion of oligosaccharidases in the apical membrane of the intestinal mucosal cells (Leturque and Brot-Laroche, 2013; Gray, 2000).

The digestion of starch starts in the mouth with mastication and enzymatic conversion of starch to dextrins and maltose, catalyzed by salivary α-amylase (Gray, 2000). When the masticated bolus of food mixes with the acidic stomach secretions, salivary α-amylase becomes inactive, since it requires neutral pH for activity (Leturque and Brot-Laroche, 2013). However, salivary α-amylase seems to be protected by the starch content of the meal and may pass into the duodenum, where it will continue to break down starches, with the help of pancreatic α-amylase (Rosenblum et al., 1988). Then, the majority of starch comes in contact with large amounts of pancreatic amylase inside the duodenal lumen, which subsequently cleave glycosidic bonds and complete the digestion process in the upper part of the small intestine (Leturque and Brot-Laroche, 2013).

The two major starches are amylose, consisting of linear segments with α-1,4 glycosidic bonds, and amylopectin, which is a complex form of starch, consisting of both linear and branched segments with α-1,6 glycosidic bonds (Gray, 2000). The active enzyme amylase binds to the interior glycosidic bonds of amylose and specifically to five consecutive glucose subunits with high affinity, leading to the production of pentasaccharides (oligosaccharides with five glucose units linked together) (Leturque and Brot-Laroche, 2013). The resulting pentasaccharide is hydrolyzed to produce the trisaccharide maltotriose and the disaccharide maltose, which are the final products of amylose digestion and cannot be broken down by the amylase due to low affinity for the enzyme's active site (Gray, 2000).

Amylopectin is also digested by amylase, but results in different end products than those of amylose. While amylase binds and cleaves the α-1,4 glycosidic bonds of amylose, it cannot bind and cleave the α-1,6 glycosidic bonds of amylopectin nor can it bind to adjacent α-1,4 bonds, due to structural angulation inhibition of binding (Leturque and Brot-Laroche, 2013). Consequently, the digestion of amylopectin produces branched oligosaccharides called α-dextrins, as well as maltose and maltotriose (Gray, 2000).

Salivary and pancreatic amylases break down amylose to produce glucose oligosaccharides and disaccharides with α-1,4 glycosidic bonds, and amylopectin to produce glucose oligosaccharides and disaccharides with α-1,4 and α-1,6 glycosidic bonds (Leturque and Brot-Laroche, 2013). In both cases, free glucose molecules account for only a small amount (about 4%) of the digestion products released (Yook and Robyt, 2002).

The brush border of the small intestine contains an ample amount of oligosaccharidases, which consist primarily of α glucosidases, and catalyze the final reactions of starch digestion by cleaving the α-1,4 glycosidic bonds between glucose molecules one at a time (Semenza and Auricchio, 1991). Alpha glucosidases (i.e., sucrase, α-dextrinase, glucoamylase, and trehalase) have the ability to break down α-1,4 glycosidic bonds, but they bind with high affinity to the substrate indicated by their nomenclature (Gray, 2000). All these enzymes work in a complementary manner and generate unbound glucose to be transported into enterocytes (Gray, 2000). In the case of maltose, digestion is catalyzed by two α glycosidase enzymes, maltase-glucoamylase and sucrase-isomaltase, which work synergistically to completely digest starches into glucose molecules (Leturque and Brot-Laroche, 2013).

Digestion of disaccharides

In addition to maltose digestion, an end product of starch breakdown, humans have the capacity to hydrolyze unique linkages present in sucrose, lactose, and trehalose. The hydrolysis of these disaccharides occurs with the action of disaccharidases found on the luminal surface of enterocytes (Leturque and Brot-Laroche, 2013). The hydrolysis of sucrose (table sugar) is carried out by the sucrase–isomaltase enzyme, and produces glucose and fructose (Leturque and Brot-Laroche, 2013). The hydrolysis of lactose (milk sugar) is carried out by the lactase–phlorizin hydrolase enzyme, and produces glucose and galactose (Gray, 2000). The hydrolysis of trehalose (sugar found in yeast and fungi) occurs by trehalase and produces two molecules of glucose (Leturque and Brot-Laroche, 2013).

Undigested glycosidic bonds

Complex carbohydrates molecules, known as dietary fiber, include oligosaccharides with α-galactosidic linkages (i.e., raffinose, stachyose, verbascose, and ajucose), polysaccharides with glucose or fructose molecules joined together by β-glycosidic bonds (cellulose, inulin), heteropolysaccharides (hemicelluloses, gums and mucilages) with various sugar molecules and types of bonds (Leturque and Brot-Laroche, 2013). These complex carbohydrate molecules cannot be digested from pancreatic and intestinal enzymes, since the human body cannot break down β-glucosyl or fructosyl and α-galactosyl bonds (Gray, 2000). Instead, dietary fiber molecules remain intact and pass through, from the ileum to the colon, where they are metabolized by intestinal bacteria (Leturque and Brot-Laroche, 2013).

Absorption

Starch and disaccharide digestion produce the monosaccharides glucose, galactose and fructose (Gray, 2000). Glucose consists of the largest quantity of absorbed carbohydrate (~80%) and galactose and fructose only account for a small amount (~20%). The body quickly absorbs and transports the simple sugars, which enter the portal circulation via the capillaries of the intestinal villi and are transported to the liver. In the liver, fructose and galactose are converted to glucose, which is either used immediately for energy or it is stored in the form of glycogen. Liver can store approximately 4–8% of its mass in the form of glycogen, which can be readily converted to glucose for the production of energy (Ball et al., 2014).

Transport

Glucose, fructose and galactose are transported through the enterocytes by facilitated diffusion via a family of transporters known as glucose transporters (GLUT). Glucose and galactose are isomers and structurally related, therefore, they utilize the same glucose transporter, called sodium glucose transporter or SGLT1 (Gray, 2000). The passage of glucose and galactose through the apical membrane of the intestinal villi requires the presence of 2 Na^+ ions for each molecule of glucose or galactose respectively (Leturque and Brot-Laroche, 2013). On the contrary, fructose is transported into the enterocyte utilizing the glucose transporter GLUT5, a brush border protein which is not dependent on Na^+ (Gray, 2000). All three monosaccharides exit the enterocyte towards the blood stream through another sugar transporter GLUT2 (Leturque and Brot-Laroche, 2013).

Factors affecting carbohydrate assimilation

Initially, the factors influencing dietary carbohydrate assimilation are degree of mastication and length of time in the oral cavity, which determine the digestion by salivary amylase (Gray, 2000). Then, the carbohydrates consumed are transported to the stomach and become exposed to an acidic pH, which stops the action of salivary amylase (Leturque and Brot-Laroche, 2013). Gastric emptying of the ingested food varies based on its nutrient composition. For example, presence of nondigestible carbohydrate (dietary fiber) and high osmolarity (sucrose sweetened foods) will slow down gastric emptying (Gray, 2000). The stomach contents are then transported to the small intestine, a process that is highly efficient yet variable, depending on the composition of ingested food (Leturque and Brot-Laroche, 2013). When healthy intestinal and pancreatic function is present, there is no significant disturbance of assimilation of nutrients due to intestinal mobility (Gray, 2000).

Certain physicochemical properties of carbohydrate foods, for example crystallization, moisture, physical barriers, may influence the degree of starch digestion (Leturque and Brot-Laroche, 2013). Furthermore, dietary carbohydrates naturally occur in foods in combination with hydrophobic proteins, which may hinder the luminal interaction of glucose moieties within the starch molecule with polar amylases. Processing of dietary carbohydrate, including cracking, milling and cooking, will facilitate the binding of amylase to starch. Consequently, based on factors affecting starch digestion, a small amount (1–10%) of the ingested starch will pass undigested and will be hydrolyzed by colonic bacteria (Leturque and Brot-Laroche, 2013).

Carbohydrates and energy metabolism

Glucose

The breakdown of glucose can be divided into two major parts: the anaerobic conversion of glucose to pyruvate, known as glycolysis and the aerobic breakdown of pyruvate to carbon dioxide (CO_2) and water (H_2O), which comprises the tricarboxylic acid cycle and the electron transport chain (Nelson and Cox, 2017).

Glycolysis is the series of enzymatic steps, leading to the breakdown of one molecule of glucose for the production of two molecules of pyruvate (**Fig. 1**). Glycolysis occurs in the cytosol of different cells, and all human cells are capable of carrying out the process of glycolysis (Nelson and Cox, 2017). However, the largest part of glycolysis occurs in the liver, muscle and adipose tissue (McGrane, 2013).

The fate of pyruvate is determined by the cell type and the availability of oxygen (Nelson and Cox, 2017). In the absence of oxygen, pyruvate is reduced to lactate in the cytosol (Wong, 2021). This would occur in the muscle during strenuous exercise, when the demands for energy are high (Nelson and Cox, 2017). In cells that do not contain mitochondria, such as the erythrocytes, the glycolysis pathway is the only mechanism of energy production.

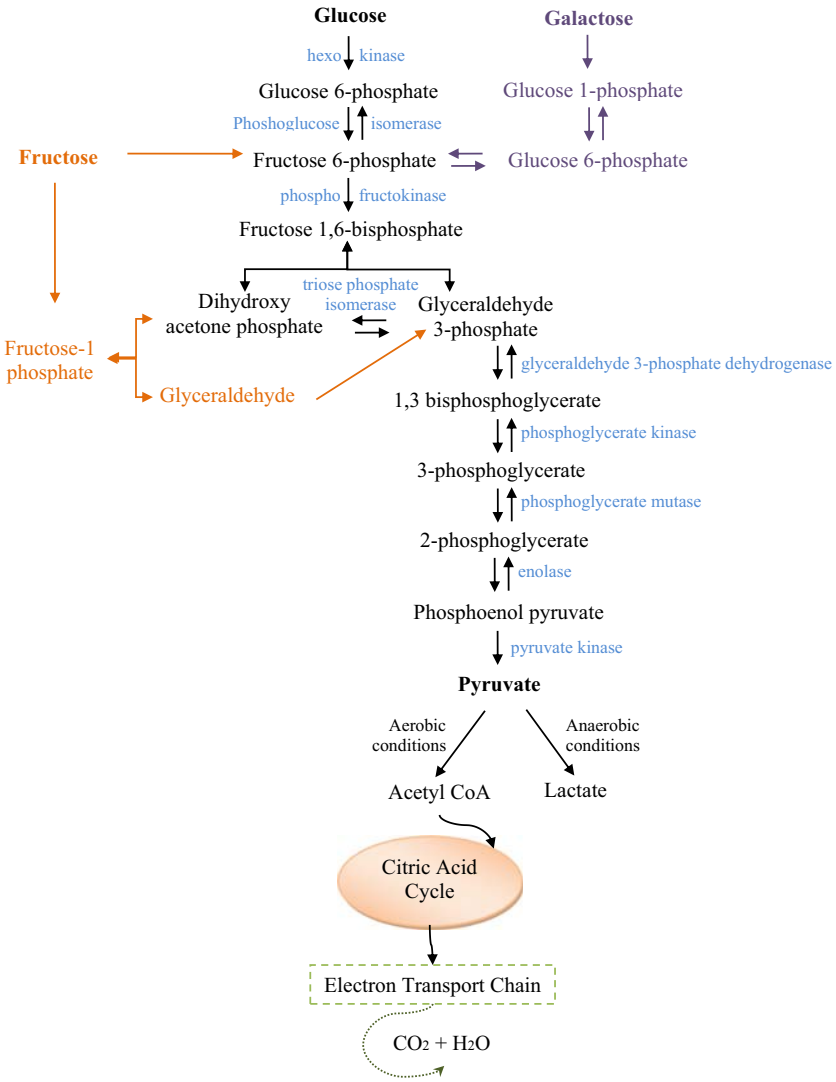


Fig. 1 Outline of glucose metabolism, with entry points for fructose and galactose.

In the presence of oxygen, pyruvate is converted to acetyl coenzyme A (acetyl CoA) in the mitochondria, enters the tricarboxylic acid cycle and subsequently the electron transport chain (Wong, 2021). As a result, pyruvate is fully oxidized to CO_2 and H_2O and large amounts of energy are produced.

Fructose and galactose

Fructose and galactose enter the glycolytic pathway through their conversion to intermediate compounds (**Fig. 1**). This occurs primarily in the liver and, as a result, these two monosaccharides are not generally available for uptake by other tissues (Nelson and Cox, 2017). The end products of the catalysis of these monosaccharides are similar to glucose; however, when they are absorbed, they do not elicit the same hormonal response as glucose.

In the liver, fructose breakdown, known as fructolysis, is initiated by the conversion of fructose to fructose 1-phoshate catalyzed by fructokinase, and subsequent hydrolysis to glyceraldehyde and dihydroxyacetone phosphate catalyzed by fructose 1-phosphate aldolase (Wong, 2021). These products of hydrolysis can be used for further glycolytic conversion. Fructolysis in the liver bypasses the highly regulated step of phosphofructokinase and can produce a large amount of glycolytic metabolites (Nelson and Cox, 2017). In the muscle and kidney cells, fructose can enter the glycolysis pathway through its conversion to fructose 6-phosphate, prior to the highly regulated phosphofructokinase step (McGrane, 2013).

In the liver, galactose enters the glycolytic pathway through its phosphorylation to galactose 1-phosphate and subsequent epimerization to glucose 1-phosphate (Nelson and Cox, 2017). This metabolic intermediate can either enter glycolysis by its conversion to glucose 6-phosphate or can be used in glycogen synthesis, depending on the nutritional state of the organism (McGrane, 2013).

Glucose production by the liver and kidneys (glycogen)

Gluconeogenesis

The biosynthesis of glucose from pyruvate, lactate, glycerol or other precursors is known as gluconeogenesis (Nelson and Cox, 2017). It is not a direct reversal of glycolysis, since several steps of glycolysis are irreversible (Wong, 2021). Gluconeogenesis mainly occurs in the liver and less in the kidney (Nelson and Cox, 2017). These tissues contain all the necessary enzymes for gluconeogenesis and the enzymatic activity of glycerol kinase, which allows glycerol to enter the gluconeogenic pathway at the level of glyceraldehyde 3-phosphate (**Fig. 2**).

In humans, liver glycogen stores can sustain the organism for 12–24 h without the ingestion of dietary carbohydrates (Nelson and Cox, 2017). After this period, the liver must produce glucose and transport it to other organs. Liver is the main gluconeogenic contributor (~90%), while the kidney produces glucose gluconeogenically to a lesser extent (~10%).

Glycogenolysis

Glycogen is a branched polymer of glucose, which may contain over 100,000 glucose units (Wong, 2021). The breakdown of glycogen for the production of glucose units is known as glycogenolysis. Glycogen breakdown is initiated at the non-reducing ends of its branches. It consists of phosphorolysis of single glucose units by the cooperating enzymatic action of glycogen phosphorylase and the debranching enzyme (Nelson and Cox, 2017). The product of phosphorolysis, glucose 1-phosphate, needs the additional action of phosphoglucomutase to be converted to glucose 6-phosphate. The liver contains the enzyme glucose 6-phosphatase for hydrolysis of glucose 6-phosphate to free glucose and for export from the organ to target tissues. However, the muscle and the brain do not contain such an enzyme and the produced glucose 6-phosphate enters the glycolytic pathway for energy production. Glycogen is a very efficient storage from of glucose, having an overall storage efficiency of approximately 97%.

Control of carbohydrate metabolism

Hormonal regulation

Hormones regulate (activate or inhibit) specific enzymes which catalyze the reactions of metabolic pathways (Nelson and Cox, 2017). This is achieved mainly by covalent regulation or conversion of the enzymes into their active or inactive form 7(Groff and Gropper, 2000). Furthermore, hormones can regulate enzymes by induction or regulation of their transcription. Regulation

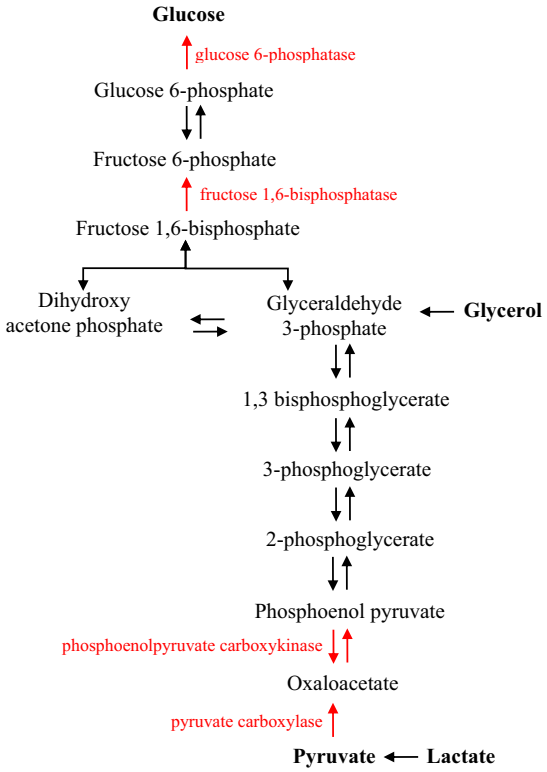


Fig. 2 Outline of gluconeogenesis, with entry points for lactate and glycerol. The reactions shown in red are distinctive for gluconeogenesis, while the other reactions are common with glycolysis.

of the expression of specific genes controls the concentration of enzymes and transport proteins necessary for carbohydrate metabolism (Nelson and Cox, 2017).

Insulin

When a meal is ingested, glucose is liberated from hydrolysis of dietary carbohydrate in the small intestine, and then it is absorbed into the blood. Increased glucose concentrations stimulate the production and secretion of insulin by the β cells of the pancreas (Nelson and Cox, 2017). Insulin promotes the transfer of glucose into the target cells (i.e., skeletal muscle, liver and adipose tissue) for utilization as energy, and for storage in the form of glycogen, primarily in the liver.

Insulin also stimulates glycolysis, by increasing the activity of glycogen synthase and the transcription of glycolytic enzymes (Stipanuk, 2013). Insulin inhibits gluconeogenesis, by decreasing the transcription of several gluconeogenic enzymes, and by moderating peripheral release of gluconeogenic precursors (**Figs. 3** and **4**). Insulin also increases the concentration of the glucose transporter GLUT4, resulting in increased glucose uptake by the skeletal muscle and adipose tissue (Nelson and Cox, 2017).

Fasting results in a decrease in insulin concentration and reduction of glucose uptake by the muscle and adipose tissue, which may use alternate forms of energy, such as free fatty acids (Stipanuk, 2013). Glucose then becomes available for uptake by the brain, red blood cells and renal medulla, which are strongly dependent on glucose for energy (Stipanuk, 2013; Nelson and Cox, 2017).

Glucagon

Glucagon is a hormone secreted in the bloodstream by the α cells of the pancreas in response to low glucose levels (Nelson and Cox, 2017). Glucagon counteracts the action of insulin and its main role is to stimulate hepatic glucose output and to maintain glucose homeostasis (Stipanuk, 2013). Glucagon stimulates glycogenolysis by activating glycogen phosphorylase, and inhibits glycogen synthesis by inactivating glycogen synthase (**Fig. 4**). Furthermore, glucagon stimulates gluconeogenesis by increasing the gene expression of gluconeogenic enzymes and blocking glycolysis (Groff and Gropper, 2000). In the liver, glucagon enhances the rate of gluconeogenesis by lipolysis, resulting in increased concentrations of free fatty acids and glycerol (Nelson and Cox, 2017).

Catecholamines

Epinephrine and norepinephrine are catecholamines that have a regulatory effect on carbohydrate metabolism, which is mainly dependent on the type of adrenergic receptor present on each cell surface (McGrane, 2013). Catecholamine receptors are divided into two types: two α (α1, α2) and three β (β1, β2, β3) receptors (Nelson and Cox, 2017). The α-1 and β receptors stimulate catabolic reactions, while the α-2 receptor inhibits them (McGrane, 2013). The presence of catecholamine receptors on different cell

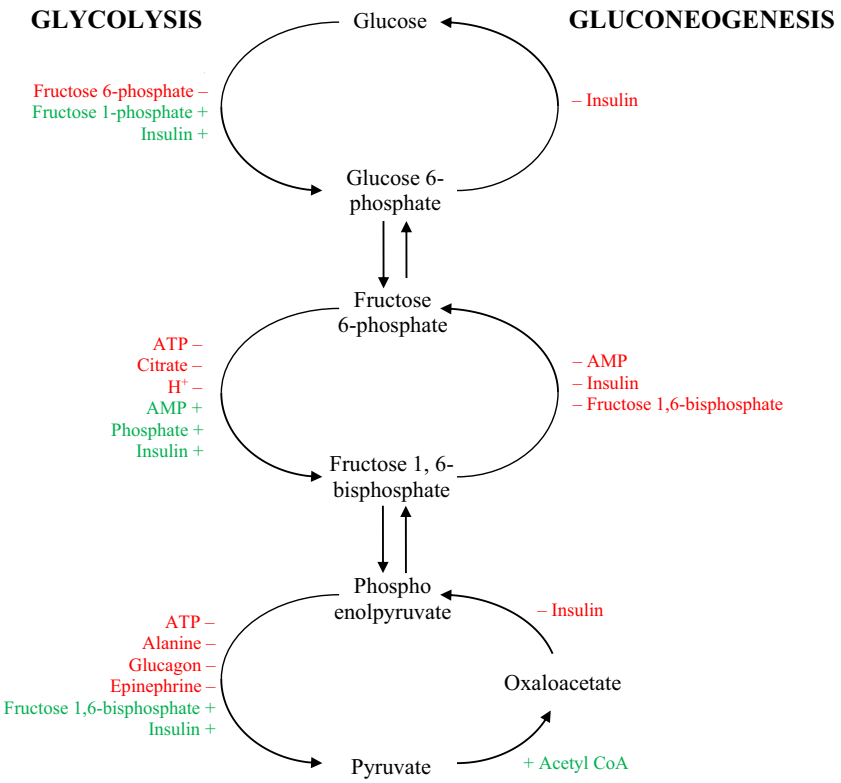


Fig. 3 Regulation of glycolysis and gluconeogenesis in the liver.

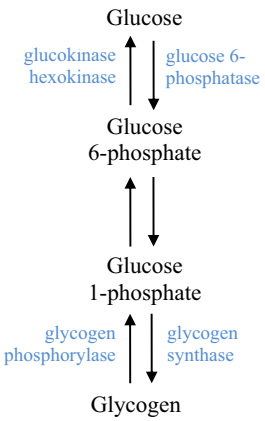


Fig. 4 Points of regulation of glycogen synthesis and breakdown.

types explains the selective breakdown of stores from certain tissues. Specifically, when epinephrine stimulates β-receptors, there is an increase in the rate of glycogenolysis and gluconeogenesis in the liver and lipolysis in adipose tissue and muscle (McGrane, 2013). Furthermore, when epinephrine stimulates α1-receptors, there is an increase in the rate of glycogenolysis, and when it stimulates α2-receptors, there is a decrease in lipolysis (Nelson and Cox, 2017).

During fasting, catecholamine regulation has not been well elucidated, but hypoglycemia is known to cause a rise in catecholamine levels in adipose tissue and liver, and/or catecholamine release by the adrenal glands (Stipanuk, 2013; McGrane, 2013).

Glucocorticoids

Glucocorticoids are steroid hormones, which are produced in reaction to mental or physical stress (McGrane, 2013) Glucocorticoids are responsible for suppression of insulin signaling. Cortisol, the principal glucocorticoid, stimulates hepatic glucose output and the expression of genes encoding for gluconeogenic enzymes, thus stimulating gluconeogenesis (Groff and Gropper, 2000). Cortisol is essential for the action of several hormones and has a much slower effect on hepatic glucose production in comparison to glucagon and the catecholamines.

Allosteric enzyme regulation

Allosteric enzymes are activated or inhibited by substances produced in the pathway in which the enzymes function (Stipanuk, 2013). These substances are called modulators and can alter the activity of allosteric enzymes by changing their conformation (Groff and Gropper, 2000). Adenosine monophosphate (AMP), adenosine diphosphate (ADP), and adenosine triphosphate (ATP) are important modulators of allosteric enzymes in carbohydrate metabolism. The effect of ATP is opposed by AMP and ADP. When energy supply is adequate, ATP accumulates and negatively modulates enzymes which catalyze energy-producing or catabolic pathways, e.g., glycolysis (Stipanuk, 2013). When energy is depleted and ATP concentration is decreased, AMP and ADP accumulate. As a result, there is a need for energy production and a positive modulation of allosteric enzymes in catabolic pathways (Groff and Gropper, 2000). An increase in ATP inhibits further energy production and blocks glycolytic enzymes, while an increase in AMP or ADP stimulates glycolytic enzymes for energy production (**Fig. 3**).

Directional shifts

The majority of enzymes catalyze reversible reactions and their action is highly dependent on the concentration of the reactants involved (Groff and Gropper, 2000). An increase in the concentration of one reactant will drive the reaction in the direction that results in the breakdown of the reactant and the achievement of homeostasis. An example of a directional shift is the interconversion of glucose 1-phosphate and glucose 6-phosphate. During glucogenolysis, the concentration of glucose 1-phosphate increases and the reaction is driven toward the production of glucose 6-phosphate (McGrane, 2013). During glycogen synthesis and gluconeogenesis, the concentration of glucose 6-phosphate increases and the reaction is driven toward the production of glucose 1-phosphate and subsequently toward the formation of glycogen (Groff and Gropper, 2000).

Regulation of gene expression

Gene expression regulation enables the human body to respond to changes in nutrient concentration. During increased availability of a specific nutrient, there is no need for expression of the genes encoding for enzymes involved in the metabolism of that nutrient.

Gene expression is highly regulated by hormones, which respond to the concentration of nutrients in the blood (Nelson and Cox, 2017). Selective expression of specific genes plays a major role in regulation of carbohydrate metabolism.

Hormonal and nutrient concentrations affect several regulatory domains of genes, which encode for enzymes involved in anabolic and catabolic pathways (Nelson and Cox, 2017). Insulin and glucose concentrations increase messenger ribonucleic acid (mRNA) levels and transcription rates of the glycolytic enzymes, and decrease those of the gluconeogenic enzymes. On the contrary, glucagon has the opposite effect of insulin.

Glycogen synthesis and breakdown

The regulatory mechanism of glycogen synthesis and breakdown involves two counteracting enzymes: glycogen synthase and glycogen phosphorylase (**Fig. 4**). Insulin activates glycogen synthase and, therefore, increases glycogen synthesis in the liver and muscle (McGrane, 2013). When blood glucose levels decrease, glucagon inhibits glycogen synthase, and activates glycogen phosphorylase for the breakdown of glycogen in the liver (Nelson and Cox, 2017).

Diseases of carbohydrate metabolism

Carbohydrate malabsorption

Carbohydrate malabsorption is manifested by the inability to metabolize fructose, glucose, or galactose (Leturque and Brot-Laroche, 2013). Carbohydrate malabsorption causes undigested disaccharides to reach the colon, which are converted to short-chain fatty acids, carbon dioxide and hydrogen gases by colonic bacteria (Topping and Clifton, 2001). Part of the short-chain fatty acids produced are absorbed, hydrogen gases can be observed in the breath, and there is a rise in osmolarity and amount of liquid and gas in the lower intestinal area and colon (Leturque and Brot-Laroche, 2013). As a result, mild symptoms of abdominal pain, discomfort and diarrhea occur in adults, while more severe symptoms of vomiting occur in infants. If carbohydrate malabsorption is not treated, it causes deficit of energy and micronutrients, which may result in unwanted weight loss and developmental delays.

Fructose intolerance and essential fructosuria

Fructose intolerance and essential fructosuria are the two genetic defects of fructose metabolism (van den Berghe, 1995). Fructose intolerance is an autosomal recessive disease, caused by a genetic defect of fructose 1-phosphate aldolase (aldolase B) in the liver (McGrane, 2013). The symptoms of aldolase B deficiency start when the infant is exposed to fructose. Aldolase B deficiency results in phosphate depletion and fructose 1-phosphate accumulation in the liver (van den Berghe, 1995). Consequently, gluconeogenesis and glycogenolysis are blocked, resulting in inhibition of protein synthesis and subsequent liver failure (McGrane, 2013). Essential fructosuria is triggered by a defect in the gene coding for fructokinase (van den Berghe, 1995). This disease is not harmful and does not have any symptoms. It causes fructose to be excreted in the urine and to be converted to fructose 6-phosphate in the muscle and adipose tissue (McGrane, 2013).

Glucose and galactose malabsorption

Glucose and galactose malabsorption is caused by a gene mutation of the glucose and galactose transporter SGLT1 which impairs its transport function (Leturque and Brot-Laroche, 2013). It is diagnosed at an early age and can be lethal if not treated by elimination of glucose and galactose from the diet (Meeuwisse and Dahlqvist, 1968). Patients with this disease can digest fructose, but not sucrose, since sucrose is converted to fructose and glucose. It is possible for patients to improve tolerance to glucose and galactose as they get older (Leturque and Brot-Laroche, 2013).

Lactose deficiency or lactose intolerance

All infants are born with some degree of lactase activity, which enables them to digest the lactose present in breast milk or formula. Lactose deficiency or intolerance usually occurs after weaning due to a reduction in lactase expression, resulting in complete loss or decline of lactase activity in intestinal cells (Nelson and Cox, 2017). As a result, lactose passes partly undigested into the large intestine, where colonic bacteria convert it to toxic compounds, which cause abdominal pain and diarrhea. Furthermore, undigested lactose causes intestinal water retention by enhancing intestinal osmolarity (Nelson and Cox, 2017). Lactose intolerance in adults is very common, and it is resolved by avoiding or decreasing consumption of dairy products. Furthermore, another approach to treating lactose intolerance is using enzyme replacement in commercial milk products, including treatment with lactase (Leturque and Brot-Laroche, 2013).

Glycogen storage diseases

Several types of glycogen storage disorders are caused by gene mutations encoding for enzymes, which are implicated in glycogen metabolism. Von Gierke disease (type I glycogen storage disease) is a rare autosomal recessive trait caused by a deficiency in glucose 6-phosphatase in the liver, kidney and intestine (McGrane, 2013) It is presented in early infancy, and it is characterized by fasting hypoglycemia, hepatomegaly, and recurrent acidosis (Dunger and Holton, 1994). Genetic defects in glucose 6-phosphatase, glucose

6-phosphatase translocase, or the pyrophosphate transporter result in a metabolic imbalance and inability of the liver to maintain glucose homeostasis by either glycogenolysis or gluconeogenesis (McGrane, 2013).

Hers' disease (type IV glycogen storage disease) is another rare autosomal recessive disorder and it is caused by a defect in glycogen phosphorylase function (McGrane, 2013). It is presented in childhood, and it is characterized by hypoglycemia, hepatomegaly and growth delay (Fernandes and Chen, 1995).

McArdle's disease (type V glycogen storage disease) is caused by deficiency of the phosphorylase kinase gene in the liver and muscle (McGrane, 2013). It is presented in adulthood and it is characterized by progressive muscle weakness and glycogen accumulation. This disease causes exercise intolerance, due to reduced ability to release glucose from glycogen to meet the increased energy requirements. However, hypoglycemia and hepatomegaly are not present.

Phosphorylase kinase deficiency is an X-linked recessive trait and it is caused by absence of phosphorylation and activation of glycogen phosphorylase (McGrane, 2013). It occurs more frequently than glycogen phosphorylase deficiency and it has similar symptoms.

Cori's disease (type III glycogen storage disease) is caused by a mutation in the debranching enzyme (McGrane, 2013). It is presented in infancy and it is characterized by glycogen accumulation in liver and muscle, fasting hypoglycemic convulsions, hepatomegaly, and myopathy (Dunger and Holton, 1994). The symptoms of hypoglycemia and hepatomegaly often decrease at the age of puberty (McGrane, 2013).

Diabetes mellitus

Diabetes mellitus is a chronic disease characterized by high levels of blood glucose (impaired glucose tolerance), and results from defects in insulin secretion, insulin action or both. Diabetes mellitus is a major cause of premature mortality, stroke, cardiovascular disease, peripheral vascular disease, congenital malformations, perinatal mortality, and long- and short-term disability. Diabetes mellitus is diagnosed by increased fasting plasma glucose (hyperglycemia), by glucose presence in the urine (glycosuria), or by elevated glucose values during an oral glucose tolerance test. There are four principal types of diabetes mellitus: type 1, type 2, gestational diabetes, and monogenic diabetes.

Type 1 diabetes

Type 1 or juvenile or insulin-dependent diabetes is caused by autoimmune pancreatic β cell exhaustion and loss of insulin secretion. Onset of the disease occurs when most of the pancreatic β cells have been destroyed by the immune system (Stipanuk, 2013). This form of diabetes is generally diagnosed in children and young adults, and it accounts for approximately 5–10% of all cases of diabetes mellitus (Centers for Disease Control and Prevention, 2017).

Type 1 diabetes usually develops in children, teenagers or young adults, but it can be diagnosed at any age (Centers for Disease Control and Prevention, 2017). In patients with type 1 diabetes, when pancreatic β cells are damaged, additional metabolic problems with glucagon secretion by the pancreatic α cells develop (Stipanuk, 2013). Hormonal dysfunctions result in dysregulation of macronutrient metabolism observed in diabetic patients. Postprandial hyperglycemia causes decreased suppression of glucagon secretion by the pancreatic α cells, dysregulation of hepatic glucose output, and dysfunctional rate of gastric emptying following a meal (Stipanuk, 2013). Since type 1 diabetics cannot produce insulin, the treatment for this type of diabetes involves insulin injections.

Type 2 diabetes

Type 2 or non-insulin-dependent diabetes is a complex heterogenous disorder caused by interactions of various genetic and environmental factors. It is characterized by insulin resistance, obesity, a sedentary lifestyle, and occasionally by decreased insulin secretion. The majority of the cases of diabetes mellitus are type 2 (Centers for Disease Control and Prevention, 2017).

Type 2 diabetes usually develops after 40 years of age, but it has become increasingly prevalent in younger children and adolescents, as a result of increasing rates of obesity and physical inactivity (Stipanuk, 2013). Type 2 diabetics have a wide range of plasma insulin values, but are insulin resistant. Insulin resistance is the inability of the muscle, liver or fat cells to respond to insulin and remove plasma glucose from the circulation. Initially, this condition is compensated by increased insulin secretion by the pancreatic β cells and maintenance of healthy blood glucose levels, but eventually the pancreas is not able to secrete enough insulin and hyperglycemia develops. Dietary changes, weight loss and increased physical activity, as well as oral diabetic medication may be used to treat and control type 2 diabetes, enhancing insulin action. However, insulin injections may also be used as treatment.

Gestational diabetes

Gestational diabetes is a condition characterized by high blood glucose levels in pregnant women without previously diagnosed diabetes (Centers for Disease Control and Prevention, 2019). Gestational diabetes is usually diagnosed in the second half of pregnancy by abnormal values in an oral glucose tolerance test. Gestational diabetes increases the mother's risk of elevated blood pressure during pregnancy and of a cesarean delivery (Center for Disease Control and Prevention, 2019). It also increases the risk of having a premature baby with health problems, including hypoglycemia and being very large for gestational age (Center for Disease Control and Prevention, 2019). Gestational diabetes is treated by dietary modification, increase in physical activity, and potentially insulin injections. Blood glucose should be closely monitored during pregnancy and after delivery. Gestational diabetes is usually ameliorated after childbirth, but it increases the future risk of developing type 2 diabetes for the mother and the baby.

Monogenic diabetes

Monogenic diabetes mellitus is a rare case of the disease caused by mutations in a single gene and is diagnosed by genetic testing (National Institute of Diabetes and Digestive and Kidney Diseases, 2017). In the United States, these cases account for 1–4% of all cases of diabetes (National Institute of Diabetes and Digestive and Kidney Diseases, 2017). Most cases of monogenic diabetes are treated with insulin or oral medications. The two main types of monogenic diabetes are neonatal diabetes mellitus (NDM) and maturity-onset diabetes of the young (MODY).

NDM occurs up to 1 in 400,000 infants in the United States, usually in newborns and young infants (National Institute of Diabetes and Digestive and Kidney Diseases, 2017). These infants do not produce enough insulin, resulting in hyperglycemia. NDM is often misdiagnosed as type 1 diabetes, however, type 1 diabetes is very seldom seen in infants younger than 6 months old. Half of infants with NDM develop permanent neonatal diabetes mellitus, while the other half of infants develop a temporary condition which disappears in infancy and may appear at a later time. NDM is characterized by frequent urination, rapid breathing, and dehydration and is diagnosed by increased glucose levels in blood or urine. Most fetuses with NDM suffer from intrauterine growth retardation, but appropriate treatment may ameliorate growth and development.

MODY occurs usually in adolescence or early adulthood and it accounts for nearly 2% of all diabetes cases in the United States in young adults and children below 20 years old. MODY cases are characterized by mild hyperglycemia. MODY is often misdiagnosed as type 1 or type 2 diabetes.

Summary

Carbohydrates are an important source of energy in the human diet. All carbohydrates, in order to be utilized by the human body, need to be broken down to simple sugars. Carbohydrate digestion begins in the oral cavity and it is completed in the apical membrane of the intestinal mucosal cells. The monosaccharides produced (glucose, fructose and galactose) are transported through the enterocytes to the blood stream, principally in the form of glucose. Glucose is then aerobically or anaerobically broken down to produce energy, and is subject to hormonal regulation, allosteric enzyme control, and regulation of gene expression. Defective regulation of carbohydrate metabolism is the cause for a variety of diseases, which have mild to serious implications for human health.

References

Ball, D.W., Hill, J.W., Scott, R.J., 2014. The Basics of General, Organic, and Biological Chemistry. Flat World Knowledge.

Centers for Disease Control and Prevention, 2017. National Diabetes Statistics Report, 2017. Available at. www.cdc.gov/diabetes/. (Accessed 20 May 2021).

Centers for Disease Control and Prevention, 2019. Gestational Diabetes. Available at. www.cdc.gov/diabetes/. (Accessed 20 May 2021).

Dunger, D.B., Holton, J.B., 1994. Disorders of carbohydrate metabolism. In: Holton, J.B. (Ed.), The Inherited Metabolic Diseases, second ed. Churchill Livingstone, London, pp. 21–65.

Fernandes, J., Chen, Y.-T., 1995. Carbohydrate metabolism: glycogen storage diseases. In: Fernandes, J., Saudubray, J.-M., van den Berghe, G. (Eds.), Inborn Metabolic Diseases: Diagnosis and Treatment, first ed. Springer-Verlag, Berlin, pp. 71–131.

Gray, G.M., 2000. Digestion and absorption of carbohydrate. In: Stipanuk, M. (Ed.), Biochemical and Physiological Aspects of Human Nutrition, first ed. WB Saunders Company, Philadelphia, pp. 91–106.

Groff, J.L., Gropper, S.S., 2000. Advanced Nutrition and Human Metabolism, third ed. Wadswoth, Belmont.

Leturque, A., Brot-Laroche, E., 2013. Digestion and absorption of carbohydrate. In: Stipanuk, M., Caudill, M. (Eds.), Biochemical, Physiological, and Molecular Aspects of Human Nutrition, third ed. WB Saunders Company, Philadelphia, pp. 142–161.

McGrane, M.M., 2013. Carbohydrate metabolism – synthesis and oxidation. In: Stipanuk, M., Caudill, M. (Eds.), Biochemical, Physiological, and Molecular Aspects of Human Nutrition, third ed. WB Saunders Company, Philadelphia, pp. 209–255.

Meeuwisse, G.W., Dahlqvist, A., 1968. Glucose-galactose malabsorption. A study with biopsy of the small intestinal mucosa. Acta Paediatr. Scand. 57, 273–280.

National Institute of Diabetes and Digestive and Kidney Diseases, 2017. Diabetes Health Information. Available at. https://www.niddk.nih.gov/health-information/diabetes/. (Accessed 20 March 2021).

Nelson, D.L., Cox, M.M., 2017. Lehninger Principles of Biochemistry, seventh ed. WH Freeman, New York.

Rosenblum, J.L., Irwin, C.L., Alpers, D.H., 1988. Starch and glucose oligosaccharides protect salivary-type amylase activity at acid pH. Am. J. Physiol. 254, G775–G780.

Semenza, G., Auricchio, S., 1991. The lactase history: from physiopathology to biochemistry, molecular and cell biology—and back?. In: Gracey, M., Kretchmer, N., Rossi, E. (Eds.), Sugars in Nutrition, vol. 25. Nestec Ltd., Vevey/Raven Press Ltd, New York, pp. 93–102.

Stipanuk, M.H., 2013. Regulation of fuel utilization in response to food intake. In: Stipanuk, M., Caudill, M. (Eds.), Biochemical, Physiological, and Molecular Aspects of Human Nutrition, third ed. WB Saunders Company, Philadelphia, pp. 434–460.

Topping, D.L., Clifton, P.M., 2001. Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. Physiol. Rev. 81, 1031–1064.

van den Berghe, G., 1995. Disorders of fructose metabolism. In: Fernandes, J., Saudubray, J.-M., van den Berghe, G. (Eds.), Inborn Metabolic Diseases: Diagnosis and Treatment, first ed. Springer-Verlag, Berlin, pp. 95–99.

Wong, E.V., 2021. Cells-Molecules and Mechanisms. Axolotl Academica Publishing. Available at. https://bio.libretexts.org/. (Accessed 18 May 2021).

Yook, C., Robyt, J.F., 2002. Reactions of alpha amylases with starch granules in aqueous suspension giving products in solution and in a minimum amount of water giving products inside the granule. Carbohydr. Res. 337, 1113–1117.

Carbohydrates: Requirements and dietary importance

Chrysanthi Stylianopoulou, Public Health Nutritionist, Veria, Greece

© 2023 Elsevier Ltd. All rights reserved.

This is an update of C. Stylianopoulos, Carbohydrates: Requirements and Dietary Importance, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 278–282, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00043-X.

Introduction 136
Dietary sources 136
Sugars 137
Natural vs. added or free sugars 137
Polysaccharides 137
Starch 137
Non-starch 137
Dietary fiber 138
Health effects of carbohydrates 138
Carbohydrates and nutrient density 139
Carbohydrate and obesity 139
Carbohydrate and cardiovascular disease 139
Carbohydrates and type 2 diabetes 140
Carbohydrates and dental caries 140
Carbohydrate and cancer 140
Carbohydrates and gastrointestinal health 140
Low carbohydrate or ketogenic diets 141
Requirements and recommendations 141
Conclusion 142
References 142
Relevant websites 143

Key points

- Provide a brief overview of the dietary sources of carbohydrates
- Examine the main sources of dietary intake of carbohydrates
- Relate health issues associated with carbohydrate intake
- Summarize studies and evidence on the most important health issues associated with carbohydrate intake
- Discuss the method of carbohydrate restriction for weight management
- Outline the recommendation of various authorities on total carbohydrate, added/free sugar and fiber intake

Introduction

Carbohydrates are an important energy source in the human diet. They generally supply about 40% of our energy requirement in developed countries and up to 80% in developing countries. Carbohydrates have been considered a fundamental source of nourishment and an inexpensive and versatile staple of the diet. The current global emphasis for healthy eating focuses on increasing carbohydrate consumption in the form of vegetables, fruits, legumes and whole grains.

Dietary sources

The major sources of carbohydrates are cereals, consisting of over 50% of carbohydrate consumed in both developed and developing countries, followed by sweeteners, root crops, pulses, vegetables, fruit and milk products (World Health Organization, 1998). Carbohydrate and nutrient intake in general can be estimated using data from food production and balance sheets, household surveys and individual assessments (World Health Organization, 1998).

In developing countries, the majority of carbohydrate comes from a single food source, for example rice, cassava, or corn (World Health Organization, 1998). Carbohydrates are a significant medium for micronutrients (vitamins and minerals), and other nutritionally important elements, like phytochemicals. When a single carbohydrate food source is consumed in developing countries,

diets can suffer from nutritional deficiencies. Therefore, variety in carbohydrate intake is very important, in order to provide to maximize nutrient sufficiency and health benefits.

The type and composition of dietary carbohydrates varies greatly among different food products. Dietary carbohydrates can be predominantly found in the form of sugar (monosaccharides and disaccharides) and starch or non-starch polysaccharides. Furthermore, in the food industry, they can be used in the form of hydrolyzed cornstarch, high-fructose corn syrups, modified starches, gums, mucilages and sugar alcohols.

Sugars

The term "sugar" includes monosaccharides and disaccharides (Brody, 1999). The most common monosaccharides are glucose (or dextrose), fructose and galactose. Glucose is found in fruits, honey, maple syrup and vegetables. Glucose is also formed from sucrose hydrolysis in honey, maple syrup and invert sugar, and it is produced from starch hydrolysis in corn syrups. The properties of glucose are important for improving food texture, flavor and palatability. Glucose is the major cell fuel, and the principal energy source for the brain (Nelson and Cox, 2017). Fructose is found in honey, maple sugar, fruits and vegetables. Fructose is also formed from sucrose hydrolysis in honey, maple syrup and invert sugar. It is commonly used as a sweetener in soft drinks, bakery products and candy, in the form of high-fructose corn syrups. Galactose is found primarily in milk and dairy products.

The most common disaccharides are sucrose, lactose and maltose (Brody, 1999). Sucrose is mostly found in sugar cane and beet, and in lesser amounts in honey, maple syrup, fruits and vegetables. The properties of sucrose are important in improving viscosity, sweetness, and flavor of baked foods, ice cream, and desserts. Maltose is formed from starch digestion and it is also produced from the germination of grain for malt liquors. Lactose is found in milk and dairy products, and is not as sweet as glucose or sucrose.

Natural vs. added or free sugars

Sugars occur naturally in fruits, vegetables and milk. These sugars are different from "added" or "free" sugars which are sugars that have been incorporated as part of food processing and preparation. These sugars can include sucrose, syrups and honey, and sugars from concentrated fruit or vegetable juices (US Food and Drug Administration, https://ask.usda.gov/s/article/).

The European Food Safety Authority (EFSA) defines "added sugars" as sucrose, fructose, glucose, starch hydrolyzates and other isolated sugar preparations utilized in the processes of food preparation and manufacturing (European Food Safety Authority, 2010).

The World Health Organization uses the term "free sugars" and defines them as sugars "added to foods and beverages by the manufacturer, cook or consumer, and sugars naturally present in honey, syrups, fruit juices and fruit juice concentrates" (World Health Organization, 2015). The Scientific Advisory Committee on Nutrition (SACN) of the United Kingdom defines "free sugars" as "all monosaccharides and disaccharides added to foods by the manufacturer, cook or consumer, plus sugars naturally present in honey, syrups and unsweetened fruit juices" (Scientific Advisory Committee on Nutrition, 2015). This definition excludes naturally occurring sugars and captures the consumption of high fructose corn syrup.

Global sugar consumption has increased over the last decade from 130 to 178 million tones (World Cancer Research Fund International, 2015). This increase has been driven by an increase in the sugar consumption in developing countries (Siervo et al., 2014). Even though there is a decline of sugar consumption in the developed countries, sugar intake is still high, consisting of approximately 20% of total dietary energy intake. This is due, in particular, to increased consumption of added or free sugars, as a result of their greater use in beverages and foods.

Polysaccharides

Starch

Starch is the most important and abundant food polysaccharide. Starch is predominantly derived from plant seed, such as wheat, maize, rice, oats and rye, and from plant roots, such as potatoes (World Health Organization, 1998). Legumes and vegetables also contribute to the starch content of the diet. Bread and pasta are popular forms of starch, while tropical starchy foods, such as plantains, cassava, sweet potatoes and yams are increasingly contributing to carbohydrate intake. Starch accounts for 20–50% of total energy intake, depending on the total carbohydrate consumption (World Health Organization, 1998).

Non-starch

Non-starch polysaccharides (NSP), formerly referred to as "dietary fiber," can either be soluble or insoluble and are mainly derived from cereals, especially wholegrain (World Health Organization, 1998). Wheat, rice and maize are good sources of insoluble NSP, while oats, rye and barley are good sources of soluble NSP. Vegetables are also good sources of NSP, both soluble and insoluble. Intakes of NSP range from about 19 g/day in Europe and North American countries to 30 g/day in rural Africa (World Health Organization, 1998).

Dietary fiber

Dietary fiber generally refers to complex carbohydrates, which compose the cells walls and structural components of plants (World Health Organization, 1998). Good sources of fiber consist of whole cereal grains, fruits, and vegetables. Dietary fiber may be soluble in water, for example pectin. Soluble fiber binds to other carbohydrates and delays digestion and absorption, therefore assisting in maintaining stable blood glucose levels. Insoluble fiber, such as cellulose, bulks up stool and aids in intestinal waste elimination. Dietary fiber consumption has potential important health benefits, including general gastrointestinal health and prevention of several non-communicable diseases, through blood cholesterol reduction and regulation of blood sugar levels (World Health Organization, 1998).

The definition of dietary fiber has been inconsistent for many decades. Various definitions have emerged and different methods of quantifying dietary fiber have been utilized during the past decades. Several authorities have provided official definitions of dietary fiber, in order to facilitate analysis of research studies, and provide comparable intake recommendations and food labeling guidelines.

In the United States, the Institute of Medicine (IOM) of the National Academies of Science has defined Dietary Fiber as "nondigestible carbohydrates and lignin that are intrinsic and intact in plants" (Institute of Medicine, 2005). IOM defined Functional Fiber as "isolated nondigestible carbohydrates that have beneficial physiological effects in humans". Furthermore, Total Fiber has been defined as the sum of Functional and Dietary Fiber (Institute of Medicine, 2005).

EFSA has defined dietary fiber as non-digestible carbohydrates plus lignin (European Food Safety Authority, 2010). This definition includes NSP (cellulose, hemicelluloses, pectins, hydrocolloids), resistant oligosaccharides, fructo and galactooligosaccharides, other resistant oligosaccharides, resistant starch (physically enclosed starch, some types of raw starch granules, retrograded amylose, chemically and/or physically modified starches), and lignin associated with the dietary fiber polysaccharides (European Food Safety Authority, 2010).

According to the WHO, dietary fiber is defined as a heterogeneous mixture of polysaccharides and lignin that cannot be degraded by the endogenous enzymes of vertebrate animals (World Health Organization, 2003). WHO makes a distinction of water-soluble and water-insoluble fiber. Examples of water-soluble fibers are pectins, gums, mucilages and some hemicelluloses, while water-insoluble fibers are cellulose and some hemicelluloses.

A formal definition for fiber has been recently incorporated in Codex Alimentarius, which is a collection of standards, codes of practice, guidelines and recommendations (Codex Alimentarius, 2017). At the 2008 meeting of the Codex Committee on Nutrition and Foods for Special Dietary Uses, dietary fiber was defined as carbohydrates with 10 or more subunits, which cannot be hydrolyzed by endogenous enzymes of the human small intestine and can belong to the following categories:

1. Naturally occurring edible carbohydrates
2. Carbohydrates obtained from food raw material by physical, chemical or enzymatic methods, and scientifically shown to have a physiological effect or health benefit
3. Synthetic carbohydrates scientifically shown to have a physiological effect or health benefit

This comprehensive definition has generated the need for the development of an integrated analytical method for the determination of total dietary fiber. The American Association of Cereal Chemists International and the Association of Official Analytical Chemists International have previously developed methods for measurement of dietary fiber and dietary fiber components. Under the supervision of these two international organizations, a method has been developed for determination of total dietary fiber. This method was successfully evaluated by interlaboratory testing (McCleary et al., 2010, 2012). Modification of this method has been made to improve several of its aspects and, as a result, the Rapid Integrated Total Dietary Fiber method has been developed and evaluated by interlaboratory testing (McCleary and Cox, 2017).

Health effects of carbohydrates

Carbohydrates are the major energy source in most human diets. Since the majority of amino acids and lipids can be converted to glucose via gluconeogenesis, there is no absolute requirement for carbohydrates (Institute of Medicine, 2005). Furthermore, in the absence of dietary carbohydrates, the human body can utilize the products of lipolysis of stored triacylglycerol or ketone bodies, for the production of energy. This process known as ketogenesis may supply energy to the brain, heart and skeletal muscle (Nelson and Cox, 2017).

Carbohydrates are stored in the human body as glycogen, mainly in the liver and muscle. The human body has a limited storage capacity for carbohydrates compared to fat. The total amount of carbohydrates stored in tissues as glycogen and circulating in the blood as glucose is less than a day's energy supply (Nelson and Cox, 2017). Diets high in carbohydrate ensure adequate glycogen storage available for immediate energy utilization. Carbohydrates are the preferred energy source for the human brain, and have an important role in reducing protein breakdown when energy intake is inadequate.

Dietary carbohydrates are absorbed in their hexose form (glucose, fructose, galactose) and provide 15.6 kJ per gram (3.75 kcal per gram) of energy. Although sugars and polysaccharides provide similar amounts of energy, they differ in their physiological and metabolic properties. The effects of carbohydrate-containing foods on blood glucose levels during digestion and absorption are variable, depending on the type of dietary carbohydrate. Post-prandial glucose response is reduced when glucose absorption is slow.

Several studies have demonstrated that foods which are absorbed slowly and those which have not been greatly processed may have greater health benefits, compared to rapidly absorbed and highly processed foods (Institute of Medicine, 2005). In order to quantify the absorption and digestion rate of carbohydrate foods, glycemic index (GI) and glycemic load (GL) have been utilized. GI and GL are measures of the blood glucose response after carbohydrate consumption. GI is the area under the curve of the blood glucose increase 2 h after carbohydrate ingestion of a set amount of a particular food (for example, 50 g) as compared to the blood glucose increase 2 h after the ingestion of the same amount of a reference food (white bread or glucose). GI is influenced significantly by the carbohydrate type and physical determinants of digestion rate (intact vs. ground grains, cooked vs. uncooked food and soluble fiber content). GL is the product of GI and the content of available carbohydrate, and considers both the quality and the amount of carbohydrate present in a particular food (Brouns et al., 2005). GI (and also GL) is determined mostly by the type of carbohydrates in a particular food, and to a small degree by the food matrix of the meal, that is the amount of protein and fat the meal contains. Beneficial effects of low GI foods have been linked to reduced glucose response after a meal and several positive health outcomes.

The role of carbohydrates in health is a growing area of research and has received a great amount of interest in the past decade.

Carbohydrates and nutrient density

Increased sugar consumption has generated a concern in the recent years, because of the potential to displace the micronutrient content of the diet by increasing "empty calories" and energy intake. There is some evidence that essential nutrient intake decreases with increasing total sugar intake (Institute of Medicine, 2005). However, sugar intake has not been shown to accurately predict micronutrient ingestion. Moderate intakes of sugar coincide with sufficient nutrient intake. The risk of low micronutrient status is increased for individuals with a diet high in sugars and low in total energy intake, as in the case of children or people on restrictive diets (Institute of Medicine, 2005). Data analysis on food intake of preschool children suggests that the intake of some micronutrients (calcium, zinc, thiamin, riboflavin and niacin) is inversely related to sugar intake. However, the dilutional effects of sugars may be somewhat distorted by the fact that some rich sources of added sugars are also fortified with micronutrients, as in the case of breakfast cereals. The IOM of the National Academies of Science, using national food intake data, reported that a clear dilutional effect on micronutrient intake starts when sugar intake approaches 25% of total calories.

Several human studies have demonstrated that diets rich in NSP may reduce the bioavailability of minerals, such as iron, calcium, and zinc (Institute of Medicine, 2005). Nevertheless, this effect may be attributed to the presence of phytate, which inhibits the absorption of those minerals, rather than the NSP content of the diet (Institute of Medicine, 2005).

Carbohydrate and obesity

Several studies have been conducted to establish an association between sugar ingestion and total energy intake (Institute of Medicine, 2005). There have been consistent reports of a negative association between sugar intake and body mass index in adults and children (Scientific Advisory Committee on Nutrition, 2015). Some ad lib dietary studies have shown that diets low-in-sugar are associated with weight loss, maybe as a result of reduced calorie intake (Institute of Medicine, 2005). Foods high in sugars or GI are highly palatable and can create a potential risk for energy overconsumption and weight gain (Institute of Medicine, 2005). However, there is insufficient evidence to support this claim or confirm the role of GI on body weight regulation (Scientific Advisory Committee on Nutrition, 2015). Foods high in sugar have high energy density, and thus decreasing their consumption can assist in weight reduction. On the contrary, foods rich in NSP are bulky and, as a result, induce greater satiety when ingested. Consequently, diets rich in NSP may be useful for obesity prevention, since they prevent energy overconsumption and assist in weight control (European Food Safety Authority, 2010).

The consumption of sugar-sweetened drinks may contribute to weight gain because of the low satiety of liquid foods (Institute of Medicine, 2005). Short-term human studies have shown that sugar-sweetened drink consumption results in an increase of total energy intake (Scientific Advisory Committee on Nutrition, 2015). Consumption of these drinks has been associated with childhood obesity (Scientific Advisory Committee on Nutrition, 2015).

Carbohydrate and cardiovascular disease

Dietary carbohydrates influence risk factors for the development of cardiovascular disease. A diet rich in carbohydrates, in the form of whole grain cereals, fruits and vegetables, may assist in the increase of the antioxidant and phytonutrient content of the diet, therefore reducing the risk of heart disease (World Health Organization, 2003). Certain NSP have been shown to reduce plasma total and low-density lipoprotein cholesterol and, therefore, result in lowering the risk of coronary heart disease in the short-term (Scientific Advisory Committee on Nutrition, 2015). Nevertheless, this effect has not been demonstrated with NSP-containing supplements or in the long-term (Institute of Medicine, 2005). Dietary fiber has been shown to lower blood pressure in dietary interventions, which is an additional benefit for protection against heart disease (World Health Organization, 2003). Furthermore, dietary fiber has been associated with lower incidence of cardiovascular disease in prospective cohort studies (Scientific Advisory Committee on Nutrition, 2015).

On the contrary, a high intake of carbohydrates (>65% of total calories), especially in the form of refined sugars and starch, may increase serum triacylglycerol levels and adversely affect plasma lipoprotein profile (Institute of Medicine, 2005). Short-term studies

show a consistent relationship between sugar consumption and elevation of triacylglycerol levels, as well as a decrease in plasma high-density lipoprotein levels, which could result in increased atherosclerosis and heart disease risk (Scientific Advisory Committee on Nutrition, 2015). However, longitudinal cohort studies have failed to show a consistent association of sugar consumption and cardiovascular disease, mainly because of the confounding factors associated with increased heart disease risk (Institute of Medicine, 2005).

High GI diets have been shown to slightly increase hemoglobin A1c, total serum cholesterol and triacylglycerols, and decrease high-density lipoprotein levels and urinary C-peptide in diabetic and hyperlipidemic individuals (Scientific Advisory Committee on Nutrition, 2015). On the contrary, low GI diets have been shown to decrease cholesterol and triacylglycerol levels in dyslipidemic individuals (Scientific Advisory Committee on Nutrition, 2015). Nonetheless, there are insufficient studies performed on healthy individuals, and further research on the role of GI on lipid profile and cardiovascular risk factors is warranted.

Carbohydrates and type 2 diabetes

There is little evidence from prospective studies to support a positive association between total dietary carbohydrate consumption and type 2 diabetes risk (Institute of Medicine, 2005). Some evidence suggests that rapidly digested refined sugars, which have a high GI, may increase the risk of type 2 diabetes (Scientific Advisory Committee on Nutrition, 2015). Short-term studies have shown that decreasing the GI of a meal can improve glucose tolerance and insulin sensitivity in healthy people (Scientific Advisory Committee on Nutrition, 2015). Furthermore, the substitution of high GI with low GI carbohydrates can decrease post-prandial glucose and insulin levels (Institute of Medicine, 2005).

Some studies have demonstrated a protective effect of cereal fiber and whole grain consumption against type 2 diabetes (Institute of Medicine, 2005; Scientific Advisory Committee on Nutrition, 2015).

Carbohydrates and dental caries

Prospective cohort studies show that the quantity and frequency of dietary sugar intake play a significant role in the development of dental caries (World Health Organization, 2015; Scientific Advisory Committee on Nutrition, 2015). Dental caries is a multifaceted disease, affected not only by the frequency and type of sugar consumed, but also by oral hygiene, fluoride supplementation and use (Scientific Advisory Committee on Nutrition, 2015). Despite the increase in sugar consumption, the incidence of dental caries has decreased worldwide, because of the increased use of fluoride and improvement of oral hygiene (World Health Organization, 2015).

Dental caries results from the loss of minerals from the tooth enamel in the presence of acid (Scientific Advisory Committee on Nutrition, 2015). Fermentation and hydrolysis of dietary sugars occur in the oral cavity, and reduce the pH of the mouth (Scientific Advisory Committee on Nutrition, 2015). Saliva has the ability to neutralize and repair the loss of minerals from this process (Scientific Advisory Committee on Nutrition, 2015). When this loss of minerals exceeds the capacity of saliva to remineralize the tissues, tooth decay occurs (Scientific Advisory Committee on Nutrition, 2015).

Data indicates that dental caries are reduced in countries where sugar intake is below 15–20 kg per person per year (World Health Organization, 2003). This figure translates to a daily intake of 40–55 g per person and to a total energy intake of 6–10% (World Health Organization, 2003). Worldwide ecological studies and randomized controlled trials have consistently shown a relationship between high sugar consumption, mainly sucrose, and tooth demineralization and tooth decay (Scientific Advisory Committee on Nutrition, 2015; World Health Organization, 2003). When annual sugar consumption increases over 15–20 kg per person per year, dental caries also increase (World Health Organization, 2003).

Carbohydrate and cancer

Case-control studies have shown that colorectal cancer risk increases with high intakes of sugar-rich foods, while other studies have failed to prove such a relationship (Institute of Medicine, 2005). Therefore, there is insufficient evidence to support the role of sugar in the risk for colorectal cancer.

On the contrary, increased dietary fiber consumption in the form of fruits, vegetables and cereals has been shown to be protective against colorectal cancer (Institute of Medicine, 2005; Scientific Advisory Committee on Nutrition, 2015).

Carbohydrate foods are a good source of phytoestrogens, which may protect against breast cancer (Institute of Medicine, 2005). However, due to the fact that studies related to carbohydrate intake and breast cancer have been inconsistent, it is not possible to establish an association (Institute of Medicine, 2005).

Prospective cohort study data indicate a negative association of prostate cancer risk with high fructose intakes; however, evidence on the role of sugar consumption on prostate cancer risk is scarce (Institute of Medicine, 2005). Some data suggest that increased fiber intakes are related to decreased prostate cancer risk (Institute of Medicine, 2005).

Carbohydrates and gastrointestinal health

High intakes of NSP and dietary fiber have been shown to contribute to the prevention and treatment of constipation, to the decrease of intestinal transit times, and to the increase of fecal mass (Institute of Medicine, 2005; Scientific Advisory Committee

on Nutrition, 2015). Population studies have linked the prevalence of hemorrhoids, diverticular disease and appendicitis to NSP intakes, although there are several dietary and lifestyle confounding factors that could directly affect these relationships (Institute of Medicine, 2005). High carbohydrate diets may be related to bacterial growth in the gut, and subsequent reduction of acute infective gastrointestinal disease risk (Institute of Medicine, 2005).

Low carbohydrate or ketogenic diets

Carbohydrate restriction has been investigated for many decades, as part of diabetes management, especially before insulin therapy was available (Katz and Meller, 2014). Carbohydrate restriction was also used as a treatment tool for childhood epilepsy (Swink et al., 1997). Recently, carbohydrate restriction has gained popularity in the arena of obesity management. The Atkins diet was the first to gain popularity, followed by the South Beach diet, the Zone diet, and the ketogenic diet. These diets restrict carbohydrates below a given level, either below 100 g per day or even below 50 g per day. By restricting carbohydrates, these diets shift the body's metabolism from glucose utilization to fat oxidation. Fat and protein composition of these diets is variable, but the main determinant of a low carbohydrate or ketogenic diet is the reduction of dietary carbohydrates.

The low carbohydrate diet is based on the Carbohydrate-Insulin Model of obesity (Ludwig and Ebbeling, 2018). This model proposes that a diet high in carbohydrates results in post-prandial increased insulin secretion, causing excess energy storage in adipose tissue, and subsequent weight gain. Carbohydrate has the strongest effect on insulin secretion, and this effect depends on the type and amount of carbohydrate consumed. Protein also has an effect on insulin secretion, but this effect is counterbalanced by glucagon secretion. Dietary fat does not have an appreciable effect on insulin secretion. When carbohydrates are restricted, two metabolic pathways are turned on: gluconeogenesis and ketogenesis (Brouns, 2018). Gluconeogenesis occurs due to the reduced flux of glucose to the tissues. However, the process of gluconeogenesis ceases when substrates become unavailable, and ketogenesis occurs to provide ketone bodies as an alternate energy source. Subsequently, reduced insulin levels promote lipolysis and increase the supply of fatty acids. These fatty acids are used for fuel or are further converted to ketone bodies, which can also be utilized as energy, even by the brain and nervous system through metabolic adaptation.

Low carbohydrate diets decrease appetite due to higher satiety of increased protein intake, regulation of appetite control hormones, and potential appetite suppression by ketone bodies (Paoli et al., 2013). They result in increased lipolysis, reduced lipogenesis, and increased metabolic efficiency (Paoli et al., 2013; Barber et al., 2021).

Short to medium duration intervention studies have demonstrated that low carbohydrate diets are effective for weight loss and have metabolic health benefits (Katz and Meller, 2014; Boden et al., 2005; Volek et al., 2008). Low carbohydrate diets have been shown to reduce hunger and promote fat loss, which may contribute to adherence to a weight loss diet. Furthermore, carbohydrate restriction seems to improve insulin sensitivity and glycemic control in diabetic and pre-diabetic subjects (Barber et al., 2021; Goldenberg et al., 2020). When longer term trials are considered, the metabolic benefits of low carbohydrate diets diminish (Goldenberg et al., 2020).

Potential risks associated with low carbohydrate diets are essentially attributed to increased protein intake and entail possible kidney damage, because of increased levels of nitrogen turnover. However, animal studies, meta-analyses and human trials do not show any kidney damage associated with high protein intake (Paoli et al., 2013). Another health issue with very low carbohydrate diets (<50 g per day) may be acidosis, but that does not occur in individuals with normal insulin functions (Paoli et al., 2013). Long term safety of low carbohydrate diets remains to be evaluated.

Despite the metabolic health benefits of carbohydrate restriction, low carbohydrate diets remain controversial (Barber et al., 2021). Nonetheless, they have been shown to be an effective weight loss tool in short to medium-term studies (Katz and Meller, 2014). Studies are lacking to demonstrate a health benefit and safety in the long term.

Requirements and recommendations

The recommendation for total carbohydrate intake from various authorities is starting from 45% up to 75% of total energy intake. It is a broad range of intake, and it is meant to represent population consumption, and provide an adequate macronutrient balance, increased dietary fiber intake, and sufficient energy requirements. High variability exists in added/free sugar intake recommendation, but most authorities agree that there should be a reduction in consumption of added/free sugars. Increased dietary fiber intake is recommended, at a level over 25 g per day, due to the positive health benefits associated with high levels of consumption. Dietary recommendations for total carbohydrate, added/free sugars, and dietary fiber from various authorities are summarized in **Table 1**.

More specifically, the IOM of the United States recommended in 2005 a range of carbohydrate intake of 45–65% of total energy intake and provided a Recommended Dietary Allowance (RDA) of 130 g of total carbohydrates per day for adults, which was based on the average minimum amount of glucose needed daily by the brain. The expert panel of IOM suggested a maximal intake of less than 25% of energy from added sugars, based on the dilutional effect of added sugars on certain micronutrients. IOM proposed that total sugar intake can be decreased by limiting foods high in added sugars and consuming naturally occurring sugar products, like milk, dairy products, and fruit. The IOM did not specify dietary requirements or recommendations for NSP consumption, but provided recommended intakes for fiber, which includes NSP. Adequate intake of total fiber, according to the IOM, should be

Table 1 Dietary recommendations for total carbohydrate, added or free sugars and dietary fiber.

Authority - year adopted	*Carbohydrates (% of total energy)*	*Free/added sugars (% of total energy)*	*Dietary fiber (daily intake)*
Institute of Medicine (IOM) - 2005	45–65%	< 25%	38 g for men 25 g for women
Scientific Advisory Committee on Nutrition (SANS) - 2015	50%	<5%	30 g
European Food Safety Authority (EFSA) - 2010	45–60%	No limit set	25 g
World Health Organization (WHO) - 2003	50–75%	<5–10%	>25 g

Data sources: Institutes of Medicine (2005), SANS (2015), EFSA (2010), WHO (2003) and Mann et al. (2007).

approximately 38 g per day for men and 25 g per day for women, based on an adequate intake of 14 g per 1000 kcal per day of total fiber.

Similarly, the SACN of the United Kingdom proposed in 2015 the consumption of total carbohydrates to comprise half of the total dietary intake (50% of total energy intake). This estimate was proposed to achieve an adequate macronutrient intake (adequate fat and protein balance). SACN estimated that the dietary recommendation for free sugars should not be higher that 5% of total energy intake, in order to provide a 5% reduction in energy, and lead to a moderate weight loss for most individuals (Scientific Advisory Committee on Nutrition, 2015). SACN recommended that the average adult population intake for dietary fiber should be 30 g per day, from a wide variety of sources (Scientific Advisory Committee on Nutrition, 2015). This recommendation by SACN was based on the positive health benefits associated with high dietary fiber intake, especially in colorectal function and cardiovascular risk factors (Scientific Advisory Committee on Nutrition, 2015).

EFSA has also published dietary reference values for the intake of carbohydrates and dietary fiber (European Food Safety Authority, 2010). EFSA recognized that an intake of 130 g per day will sufficiently cover the glucose requirements for the brain. EFSA acknowledged that this level of intake is not adequate to cover energy needs in terms of acceptable intakes of fat and protein. Consequently, EFSA proposed a range of intake of total carbohydrates from starchy foods and from simple sugars of 45–60% of total energy intake, partly based on current levels of intake and dietary patterns. The EFSA panel found insufficient evidence to pose an upper limit for sugars or added sugars and proposed such recommendations to be considered at a national level. However, the EFSA panel set a minimum daily intake of 25 g for dietary fiber, based on health benefits associated with high fiber intakes. These benefits, according to EFSA, included decreased risk of coronary heart disease and type 2 diabetes, as well as improved weight control.

The Joint Committee of Food and Agriculture Organization (FAO) and the WHO published in 2003 a report regarding the role of carbohydrates in human nutrition, and recommend the consumption of 55–75% of total energy in the form of carbohydrates from a variety of sources (World Health Organization, 2003). This lower limit has been recently revised to 50–75% to align with other authorities (Mann et al., 2007). The Joint FAO/WHO committee suggested that free sugars should be restricted to less than 10% of total energy. A recent WHO publication proposed that the consumption of free sugars should be further restricted to less than 5% of total energy intake, in order to further decrease the risk of dental caries (World Health Organization, 2015). WHO reports recognized that there is no direct causal link between sugar consumption and chronic disease, but acknowledged that sugars increase the energy density of the human diet significantly and that high sugar drinks are associated with childhood obesity. Furthermore, WHO proposed that people should consume more than 400 g per day of fruits and vegetables, and that the majority of carbohydrates should originate from NSP, principally from cereals, vegetables, legumes, and fruits. This recommendation, according to WHO, should provide adequate amounts of NSP (>20 g per day) and dietary fiber (>25 g per day).

Conclusion

Carbohydrates are an important energy source in the human diet. Research studies confirm the health promoting role of dietary fiber and non-starch polysaccharides in improving cardiovascular disease risk and promoting colorectal health. Furthermore, they demonstrate a link of increased sugar consumption to obesity and dental caries. Therefore, public health nutrition guidelines propose an increase in consumption of carbohydrates in the form of fruits, vegetables, legumes and whole grains, and concurrent reduction of added/free sugars in order to promote health and well-being.

References

Barber, T.M., Hanson, P., Kabisch, S., Pfeiffer, A.F.H., Weickert, M.O., 2021. The low-carbohydrate diet: short-term metabolic efficacy versus longer-term limitations. Nutrients 13, 1187.

Boden, G., Sargrad, K., Homko, C., Mozzoli, M., Stein, T.P., 2005. Effect of a low-carbohydrate diet on appetite, blood glucose levels, and iinsulin resistance in obese patients with type 2 diabetes. Ann. Intern. Med. 142, 403–411.

Brody, T., 1999. Nutritional Biochemistry, second ed. Academic Press, San Diego.

Brouns, F., 2018. Overweight and diabetes prevention: is a low-carbohydrate-high-fat diet recommendable? Eur. J. Nutr. 57, 1301–1312.

Brouns, F., Bjorck, I., Frayn, K.N., Gibbs, A.L., Lang, V., et al., 2005. Glycaemic index methodology. Nutr. Res. Rev. 18, 145–171.
Codex Alimentarius, 2017. Guidelines on Nutrition Labelling CAC/GL 2-1985 as Last Amended 2017. Joint FAO/WHO Food Standards Programme, Secretariat of the Codex Alimentarius Commission, Rome.
European Food Safety Agency, 2010. Scientific opinion on dietary reference values for carbohydrates and dietary fibre. EFSA J. 8 (3), 1462.
Goldenberg, J.Z., Day, A., Brinkworth, G.D., Sato, J., Yamada, S., et al., 2020. Efficacy and safety of low and very low carbohydrate diets for type 2 diabetes remission: systematic review and meta-analysis of published and unpublished randomized trial data. Br. Med. J. 372, 4743–4755.
Institute of Medicine, 2005. Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids. National Academies Press, Washington, D.C.
Katz, D.L., Meller, S., 2014. Can we say what diet is best for health? Annu. Rev. Publ. Health 35, 83–103.
Ludwig, D.S., Ebbeling, C.B., 2018. The carbohydrate-insulin model of obesity: beyond "calories in, calories out". JAMA Intern. Med. 178, 1098–1103.
Mann, J., Cummings, J.H., Englyst, H.N., Key, T., Liu, S., et al., 2007. FAO/WHO scientific update on carbohydrates in human nutrition: conclusions. Eur. J. Clin. Nutr. 61 (Suppl. 1), S132–S137.
McCleary, B.V., Cox, J., 2017. Evolution of a definition for dietary fiber and methodology to service this definition. Luminacoids Res. 21 (2), 9–21.
McCleary, B.V., DeVries, J.W., Rader, J.I., Cohen, G., Prosky, L., et al., 2010. Determination of total dietary fiber (CODEX definition) by enzymatic gravimetric method and liquid chromatography: collaborative Study. J. AOAC Int. 93, 221–233.
McCleary, B.V., DeVries, J.W., Rader, J.I., Cohen, G., Prosky, L., et al., 2012. Determination of insoluble, soluble, and total dietary fiber (CODEX definition) by enzymatic-gravimetric method and liquid chromatography: collaborative study. J. AOAC Int. 95, 824–844.
Nelson, D.L., Cox, M.M., 2017. Lehninger Principles of Biochemistry, seventh ed. WH Freeman, New York.
Paoli, A., Rubini, A., Volek, J.S., Grimaldi, K.A., 2013. Beyond weight loss: a review of the therapeutic uses of very-low-carbohydrate (ketogenic) diets. Eur. J. Clin. Nutr. 67, 789–796.
Scientific Advisory Committee on Nutrition, 2015. SACN Carbohydrates and Health Report. The Scientific Advisory Committee on Nutrition Recommendations on Carbohydrates, Including Sugars and Fibre. Public Health England.
Siervo, M., Montagnese, C., Mathers, J., Soroka, K., Stephan, B., Wells, J., 2014. Sugar consumption and global prevalence of obesity and hypertension: an ecological analysis. Public Health Nutr. 17 (3), 587–596.
Swink, T.D., Vining, E.P., Freeman, J.M., 1997. The ketogenic diet. Adv. Pediatr. 44, 297–329.
Volek, J.S., Fernandez, M.L., Feinman, R.D., Phinney, S.D., 2008. Dietary carbohydrate restriction induces a unique metabolic state positively affecting atherogenic dyslipidemia, fatty acid partitioning, and metabolic syndrome. Prog. Lipid Res. 47, 307–318.
World Cancer Research Fund International, 2015. Curbing Global Sugar Consumption Effective Food Policy Actions to Help Promote Healthy Diets & Tackle Obesity. London.
World Health Organization, 1998. Carbohydrates in Human Nutrition. Report of a Joint FAO/WHO Expert Consultation. FAO Food and Nutrition Paper 66, pp. 1–140.
World Health Organization, 2003. Diet, Nutrition and the Prevention of Chronic Diseases: Report of a Joint WHO/FAO Expert Consultation. WHO Technical Report Series No. 916. Geneva.
World Health Organization, 2015. Guideline: Sugars Intake for Adults and Children. Geneva.

Relevant websites

http://www.dh.gov.uk/.
http://www.efsa.europa.eu/.
http://www.fao.org/.
http://www.gov.uk/.
http://www.iom.edu/.
http://www.usda.gov/.
http://www.who.int/.

Carotenoids: Chemistry, sources and physiology

E. Rodríguez-Rodríguez[a], L.M. Bermejo[b], and R.M. Ortega[b], [a] Department of Chemistry in Pharmaceutical Sciences, Analytical Chemistry, F. Pharmacy, Complutense University of Madrid, Madrid, Spain; and [b] Department of Nutrition and Foods Science, F. Pharmacy, Complutense University of Madrid, Madrid, Spain

© 2023 Elsevier Ltd. All rights reserved.

This is an update of B.J. Burri, Carotenoids: Chemistry, Sources and Physiology, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 283–291, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00044-1.

Introduction 145
Structure and classification 145
Stereochemistry 145
Biosynthesis 146
Properties 147
Chemical and physical properties 147
Electrochemical properties 148
Reaction with reactive oxygen species 148
Carotenoids as pro-oxidants 148
Effects of storage and processing 149
Dietary sources 149
Physiology 149
Digestion 149
Absorption 150
Transport 151
Tissue distribution 152
Effects of carotenoids on health 152
Conclusions 153
References 153
Further reading 153

Key points

The main objectives for the chapter are:

- To describe the structure and classification of carotenoids.
- To summarize the main chemical, physical and electrochemical properties of carotenoids.
- To describe the effect of processing and storage conditions in carotenoid levels and bioavailability.
- To enumerate the main carotenoids sources.
- To describe the physiology of carotenoids.
- To outline the main effects of carotenoids on health.

Glossary

Carotene A carotenoid formed only from carbon and hydrogen, such as lycopene, β-carotene, and α-carotene

Carotenoid A family of pigments with a characteristic conjugated double bond system, widely distributed in nature

Isoprenoid Compounds formed from the basic five-carbon building block, isoprene

Polyene An organic compound containing many double bonds, especially one having double bonds in a long aliphatic hydrocarbon chain

Provitamin A carotenoids The subfamily of carotenoids that have appropriate chemical structures for vitamin A formation. The most important of these carotenoids are β-carotene, α-carotene, and β-cryptoxanthin

Xanthophyll Oxygenated carotenoids, such as β-cryptoxanthin, lutein, and zeaxanthin

Introduction

Carotenoids are fat-soluble pigments which can be found in a wide variety of plants but also in animals, algae, fungi and bacteria. Majority of carotenoids are tetraterpenoids lipid compounds made of eight isoprenoid units with a 40-carbon skeleton and possess a C22 central unit consisting of nine conjugated double bonds and an end group at both ends of the polyene chain. Carotenoids can be classified into two groups regarding the presence or absence of oxygen in their molecules: Carotenes (non-oxygenated) and Xanthophylls (contains oxygen atoms). The novo synthesis of carotenoids does not occur in humans, and they are primarily obtained from fruit and vegetables but also from several animal-derived foods such as salmon and eggs (Maoka, 2020). More than 1204 natural carotenoids have been identified in the nature in 722 source organisms, and approximately 40 are present in the human diet. Nevertheless, only 6 carotenoids (β-carotene, α-carotene, lycopene and the oxy-carotenoids lutein, zeaxanthin and β-cryptoxanthin) represent the 90–95% of the total carotenoids detected in human plasma (Olmedilla et al., 2001).

Carotenoids can show both antioxidant and pro-oxidant characteristics depending on conditions and so many of the known actions and functions of carotenoids can be directly attributed to their physico-chemical properties. From a nutritional point of view, the importance of carotenoids lies in their provitamin A activity, in their antioxidant capacity and their beneficial effect on the prevention of various diseases such as cancer, eye and vascular disorders, among others (Böhm et al., 2021).

Structure and classification

Carotenoids are lipid-soluble pigments widely distributed in nature and are present in photosynthetic bacteria, some species of archaea and fungi, algae, plants, and animals. Chemically the majority of naturally occurring carotenoids are tetraterpenoids made of eight isoprenoid units with a 40-carbon skeleton and possess a C22 central unit consisting of nine conjugated double bonds and an end group at both ends of the polyene chain (Maoka, 2020) (**Fig. 1**).

There are seven different terminal groups: ψ, β, γ, ε, φ, χ, and κ. In general terms, the terminal rings β, γ, and ε rings are formed from ψ ends, whereas φ, χ, and κ rings are formed from β end groups (**Fig. 2**) (Fernandes et al., 2018; Maoka, 2020). Depending on the presence or absence of rings in their molecules, they can be classified into two main groups: cyclic or acyclic carotenoids, respectively (Meléndez-Martínez et al., 2019).

Carotenoids can also be classified into two groups regarding the presence or absence of oxygen in their molecules: Carotenes and Xantophylls (**Fig. 3**). Carotenes (as β-carotene, α-carotene, γ-carotene, lycopene, phytoene and phytofluene) are non-oxygenated carotenoids that may be linear or possess cyclic hydrocarbons at one (monocyclic) or both (bicyclic) ends of the molecule. Xanthophylls are carotenoids containing oxygen atoms that are usually hydroxy, epoxy, carboxy, or carbonyl groups. Most of the oxygen substituents appear on β-ionyl ring positions (as β-cryptoxanthin, lutein, zeaxanthin, la meso-zeaxanthin, astaxanthin, canthaxanthin and fucoxanthin), although 5,8-epoxides (such as flavoxanthin) and apo-carbonyls (as β-citraurin) are known (Maoka, 2020).

Apart from these two general classifications of carotenoids as cyclic or acyclic or carotenes and xanthophylls, other subgroups of carotenoids can be distinguished by their structure. Some carotenoids have a 45- or 50-carbon skeleton, which are called higher carotenoids. On the other hand, carotenoids composed of carbon skeletons with fewer than 40 carbons are called apocarotenoids. Apocarotenoids correspond to products that are generated by oxidative cleavage of carotenoids, catalyzed by a family of enzymes called carotenoid excision dioxygenases (Fernandes et al., 2018).

Finally, some of the carotenoids also act as precursors of vitamin A, allowing their classification into provitamin A group (β-carotene, α-carotene and β-cryptoxanthin) and non-provitamin A carotenoid group.

Stereochemistry

Due to the presence of double bonds in carotenoid molecules, geometrical isomers may exist considering the relative position of substituents around a planar carbon–carbon double bond, thus existing all-trans (all-E) or cis (Z) (cis/trans o Z/E) isomers, which markedly differ in shape. For instance, α- and β-carotene that differ only in the position of a double bond in the cyclic end-group can both show further cis/trans-isomerism along the terpene chain. In general, the (all-E) isomers of carotenoids are the most stable and therefore the most abundant. Although the presence of cis isomers in foods and other matrices can be attributed to isomerization caused by factors such as heat or light, it has been established that some can also occur naturally, as in the case of acyclic carotenoids such as phytoene or phytofluene (Meléndez et al., 2019).

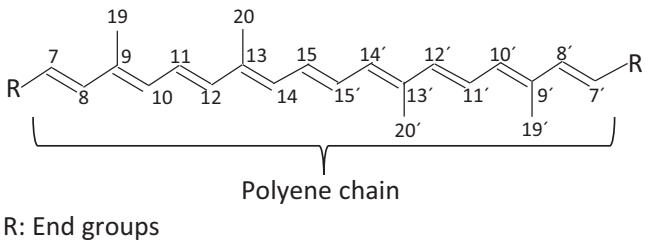


Fig. 1 Basic structure of carotenoids.

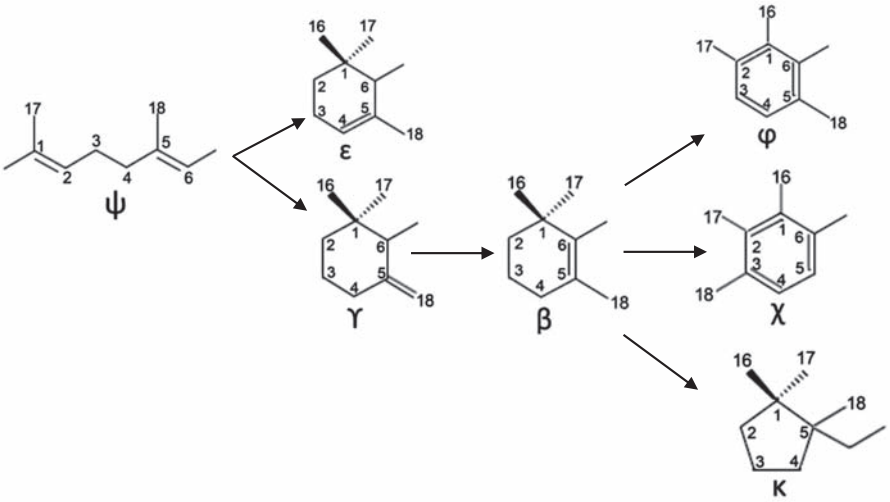


Fig. 2 Terminal end groups of carotenoids. Adapted from Fernandes et al. (2018).

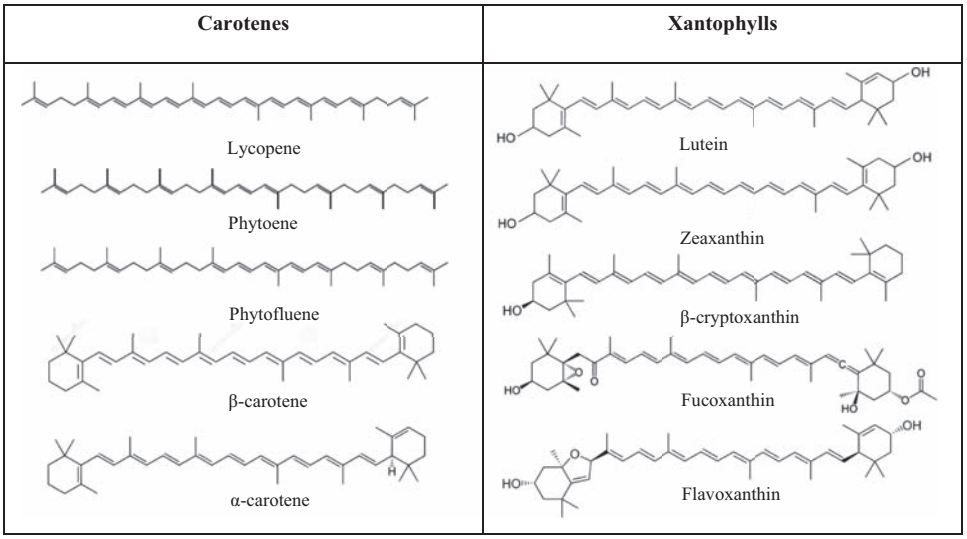


Fig. 3 Examples of carotenes and xantophylls.

Furthermore, many carotenoids have chiral centers in their molecules (carbon atoms to which four different substituents are attached), so that different optical isomers can exist (Meléndez et al., 2019). Classic examples of carotenoids with a chiral center are zeaxanthin and astaxanthin. Two optical isomers (3R-3′R)-zeaxanthin and (3R-3′S)-zeaxanthin, commonly referred to as meso-zeaxanthin and (3S, 3′S)-zeaxanthin are found in the macula lutea of the human retina. Conversely, optical isomers different from astaxanthin, (3R, 3′R), (3S, 3′S) and (3R, 3′S or meso), are found in marine organisms in varying proportions (Fernandes et al., 2018) (**Fig. 4**).

Biosynthesis

All isoprenoids are derived from two common precursors of five carbon atoms (C5), the isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), which can be generated from two different metabolic pathways, the mevalonate pathway (MVA) and the Methyl Erythritol 4-phosphate pathway (MEP). DMPA condenses with three IPP molecules and geranyl-geranyl diphosphate (GGDP) is produced, from which two molecules are condensed, in a reaction catalyzed by phytoene synthase (PSY), producing phytoene (the first carotene, which is colorless). The phytoene goes through a series of desaturation reactions to finally obtain the lycopene molecule (first carotenoid with color), which can take two routes. In the first of them, by the

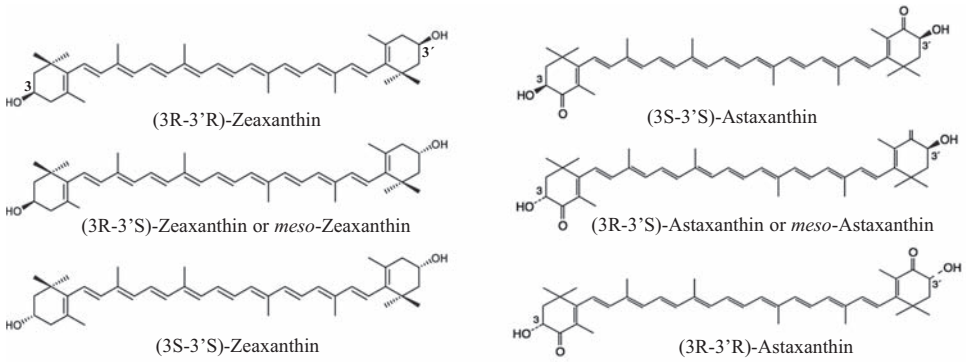


Fig. 4 Optical isomers of zeaxanthin and astaxanthin.

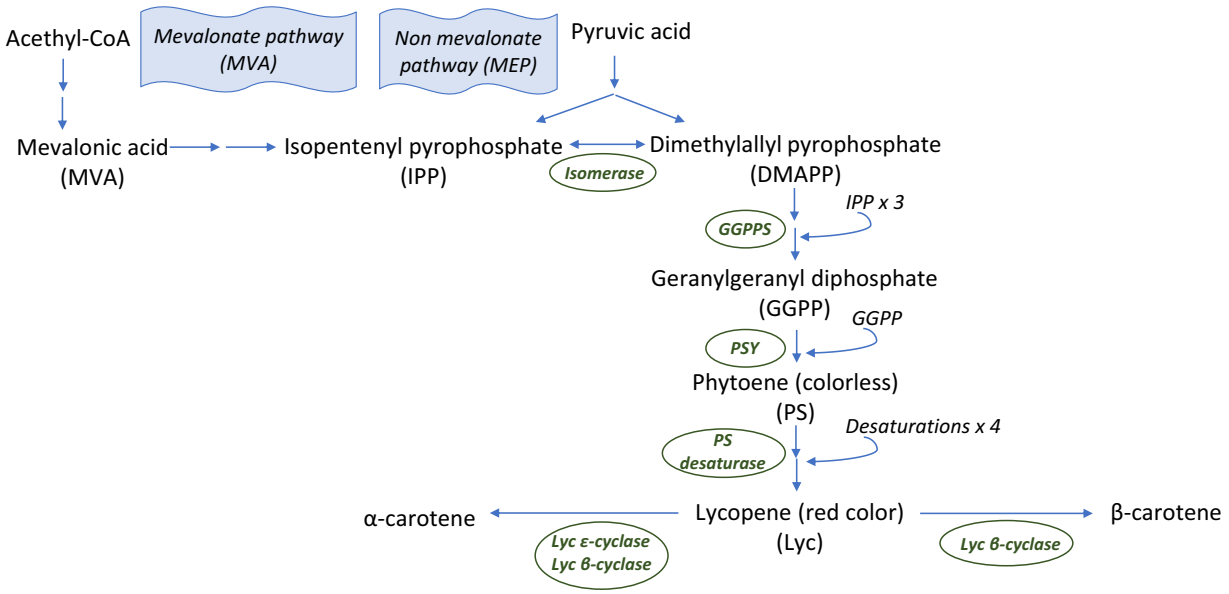
action of the ε-lycopene cyclase (ε-LCY) one of the ends of the lycopene is cycled and with the action of the β-lycopene cyclase (β-LCY) the other end is cycled, obtaining the α-carotene. By the second way, the two ends are cycled with the help of the β-LCY, obtaining the β-carotene. From these two compounds, through oxidation reactions catalyzed by ε- and/or β-carotene hydroxylases and zeaxanthin epoxidase, xanthophylls are formed (Maoka, 2020) (**Fig. 5**).

Properties

Chemical and physical properties

In general, carotenoids are hydrophobic molecules and are soluble only in organic solvents such as acetone, methanol, diethyl ether, hexane, chloroform and pyridine, among many others. Due to its hydrophobic character, they are associated with lipophilic sites in cells, such as bilayer membranes. Polar substituents such as hydroxyl groups decrease their hydrophobicity and alter their orientation with respect to membranes. Lycopene and β-carotene are aligned parallel to membrane surfaces to maintain a hydrophobic environment, whereas the more polar xanthophylls such as lutein become oriented perpendicular to membrane surfaces to keep their hydroxyl groups in a more hydrophilic environment (Gruszecki and Strzałka, 2005). These differences can affect the physical nature of a membrane as well as its function.

However, it must be taken into account that they are soluble in aqueous environments if they are integrated into liposomes or cyclic oligosaccharides such as cyclodextrins. In addition, their association with proteins or glycosylation reactions also allow them to be present in aqueous media. Carotenoproteins have been found mainly in plants and invertebrates, but intracellular



MVA: Mevalonate pathway; MEP: Methyl Erythritol 4-phosphate pathway; GGPPS: Geranylgeranyl diphosphate synthase; PSY: Phytoene sintasa

Fig. 5 Biosynthetic pathways of carotenoids.

β-carotene-binding proteins have been found in bovine liver and intestine and in livers of the rat and ferret. In addition, a xanthophyll-binding protein has been found in human retina and macula. Carotenoids are also present in nature as crystalline aggregates (lycopene in chromoplasts) or as fine dispersions in aqueous media (β-carotene in oranges). Typically carotenoid molecules are very sensitive to elevated temperatures and to the presence of oxygen, acid, and light when in solution, and are subject to oxidative degradation (Meléndez et al., 2019).

Electrochemical properties

What sets carotenoids apart from other molecules and gives them their electrochemical properties is their conjugated double bond (c.d.b.) system. In this alternating double and single bond system, π-electrons are delocalized over the entire polyene chain. This polyene chain imparts characteristic electronic spectra and photophysical and photochemical properties to this group of molecules. The highly delocalized π-electrons require little energy to reach an excited state so that light energy can cause a transition. The length of the conjugated polyene or chromophore affects the amount of energy needed to excite the π-electrons. The longer the conjugated system, the easier it is to excite, so longer wavelengths of light can be absorbed, but for there to be a perceptible coloration, at least seven c.d.b are necessary. Thus phytoene, having three conjugated double bonds, and phytofluene, having five, are colorless, zeta-carotene having seven conjugated double bonds, absorbs light at ~400 nm and appears yellow, neurosporene having nine, absorbs light at ~451 nm, and appears orange, and lycopene having eleven conjugated double bonds, absorbs at ~472 nm, and appears red.

The UV–Vis absorption spectrum of carotenoids is of interest to clarify their structure. Usually there are three wavelengths of maximum absorption (λ_{max}) whose value depends on the number of c.d.b. and the solvent used for the measurement. In addition, the λ_{max} increases with the length of the chromophore, so that the unconjugated double bonds do not significantly affect the spectrum. However, when there are c.d.b. in a ring, λ_{max} appears at shorter wavelengths compared to non-cyclic carotenoids with the same number of c.d.b. Carbonyl groups conjugated with the polyenic chain also increase the length of the chromophore, while the hydroxyl and methoxyl groups do not affect the chromophore. The spectrum of cyclic carotenoids in which the chromophore extends to the rings presents a hypsochromic and hypochromic displacement, as occurs in zeaxanthin, while the spectrum of carotenoids with carbonyl groups conjugated with the polyenic chain presents a bathochromic displacement, as it occurs in astaxanthin. When comparing the spectrum of Z isomers with respect to those of all-E isomers, it is observed that the maximum absorption is located at values between 2 and 6 nm lower and that a new absorption band appears in the ultraviolet region (Meléndez-Martínez et al., 2019).

Reaction with reactive oxygen species

Carotenoids are synthesized in photoautotrophic and in heterotrophic organisms as antioxidants to protect especially against reactive Oxygen Species (ROS) that are mainly produced as a result of aerobic metabolism or in response to biotic and abiotic stresses. The major members of the ROS family include free radicals like superoxide anion radicals $O_2\cdot^-$, hydroxyl radicals HO·, peroxyl radicals (ROO·) and non-radicals like H_2O_2 and $^1O_2$. Among the various defense strategies, carotenoids are very efficient quenchers of singlet oxygen and scavengers of other reactive oxygen species (Stahl and Sies, 2003).

Carotenoids belong to the most efficient physical quenchers of $^1O_2$, both *in vitro* and *in vivo* and protect organisms from photosensitized formation and accumulation of $^1O_2$ because they can absorb the radiation energy from the photosensitizer preventing the transfer of excitation energy to ground state oxygen and because the triplet energy level of carotenoids is close or below that of $^1O_2$ and they can also efficiently drain the excitation energy from $^1O_2$. Thus, a prominent function of carotenoids is the protection of the photosynthesis apparatus from damage under high light conditions. This includes quenching of photosensitized triplet chlorophyll as the first line of defense and of single oxygen once formed by reaction with triplet chlorophyll (Sandmann, 2019).

The reaction of carotenoids as scavengers of both short-lived and long-lived oxidizing radicals has been investigated. In homogeneous solutions reaction pathways have been found to depend on both the nature of the reacting free radical and the structure of carotenoid. There are four generally accepted major types of reactions of free radical scavenging by carotenoids: (1) addition reactions, (2) oxidation, (3) reduction, and (4) hydrogen atom abstraction (Ribeiro et al., 2018).

(1) Adduct formation R-Car· by radical addition (ROO·, HO·, …) to the carotenoid polyene chain. This adduct can react with another radical species and form a non-radicalar product (R-Car-R) with low concentration of oxygen.

(2) Removal of an electron from the conjugated system of the carotenoid by oxidizing radical species ($R\cdot^+$) possessing high redox potential, leading to the formation of a carotenoids radical cation $Car\cdot^+$ with a delocalized unpaired electron.

(3) Reduction of carotenoids, resulting in the formation of a carotenoid radical anion ($Car\cdot^-$).

(4) Hydrogen atom abstraction from the carotenoid results in the formation of a resonance-stabilized radical Car·.

Carotenoids as pro-oxidants

Although at low oxygen pressure, carotenoids act as a chain breaking antioxidants, at high oxygen pressure they are readily autoxidized and exhibit pro-oxidant activity. With higher concentrations of oxygen, the carotenoid radical (Car·) may react with dioxygen

(O_2) forming a carotenoid peroxyl radical (Car-OO·) (auto-oxidation). These radicals act as pro-oxidants as they cause lipid (LH) peroxidation and promote oxidative damage in other biomolecules (e.g., DNA, proteins). Hydrophobic lipid (LOO·) may react with carotenoids to produce oxidation products that will be further metabolized and eliminated from the body (Ribeiro et al., 2018).

Effects of storage and processing

Carotenoids are susceptible to oxidative degradation and isomerization resulting from storage and processing conditions (Dias et al., 2021). These reactions result in both loss of color and biological activity and formation of unpleasant volatile compounds. Degradation occurs on exposure to oxygen, light, heat, and conditions that destroy cell walls and ultrastructural integrity. Degradation is accelerated by the presence of metals, enzymes, unsaturated lipids, and prooxidants. Heating can also promote isomerization of the naturally occurring all-trans to various cis isomers, affecting bioavailability, the color saturation and can change the antioxidant activity of the compounds. Food processing (e.g., blanching, home cooking, pasteurization, and drying) affects bioavailability by macerating tissues, destroying or weakening cell ultrastructure, denaturing or weakening complexes with proteins, and cleaving ester linkages, thereby releasing carotenoids from the food matrix. Processing can also result in carotenoid degradation and, to alleviate such problems, some advanced processing methods (e.g., non-thermal and vacuum processes) have been proposed (Böhm et al., 2021; Dias et al., 2021). Processed foods are frequently fortified with carotenoids for example, annatto, an extract from the seeds of the *Bixa orella* tree containing the carotenoids bixin and norbixin, is added to butter, margarine, and processed cheese to give a yellow-orange color to these products. Similarly, tomato oleoresin is added to processed tomato products, increasing lycopene content while enhancing their attractive red color and increase nutritive value or enhance attractiveness.

Dietary sources

Carotenoids cannot be synthesized by humans; therefore, they must be obtained from dietary sources.

Fruit and vegetables are considered the most important sources for carotenoids in the human diet. However, the contribution of some animal food must not be overlooked: egg yolk, dairy products (milk, butter, etc.) and seafood may provide a significant amount of certain carotenoids.

The richest sources are highly pigmented in red, orange, and yellow colors, depending on the type of main carotenoid contained. For example, β-carotene, α-carotene, β-cryptoxanthin, lutein, and zeaxanthin are yellow to orange, and they responsible for giving foods as carrot, kale, apricot, mango, egg yolk, butter, orange cheeses … their colors. The carotenoid lycopene is red, and it is responsible for giving tomato, watermelon, papaya, pink grapefruit and others their colors. Xanthophylls, astaxanthin and canthaxanthin, are responsible to provide the pink color of salmon, which they obtain from eating small crustaceans and krill. By other hand, it is important to highlight that carotenoids are also present in green leafy vegetables, but they are masked by chlorophyll color. All these foods provide approximately 40 carotenoids in the human diet. **Table 1** lists the contain of the most common carotenoids found in some foods from USDA database reference (USDA, 2021).

Moreover, an European Database of carotenoids levels in food has been published in 2021 (Dias et al., 2021).

Finally, it is important to take into account that the variable carotenoid content in food products is affected by diverse nature factors: for example genotype, seasons and climatic conditions of the production area (temperature, amount of sunlight, degrees of stress from extremes in climate such as drought, heat, and cold), agronomic factors (reduced irrigation, high salinity or electrical conductivity, high leaf to fruit ratio, nitrogen fertilization …), cooking, and storage and processing conditions as mentioned above.

Physiology

Digestion

Digestion of food in the stomach increases accessibility of carotenoids for absorption by maceration in acid and digestive enzymes. The acidic environment of the stomach helps to disrupt cell walls and other cellular ultrastructure of raw fruits and vegetables and causes further breakdown of cooked foods to release carotenoids from food matrices in which they are contained or bound. Carotenoids in green leafy vegetables are found in chloroplasts; those in fruit are located in chromoplasts. Absorption studies comparing plasma levels of β-carotene and retinol after consuming fruit versus green leafy vegetables show that β-carotene is more efficiently absorbed from fruit, indicating that chloroplasts (or the bonds linking chloroplast proteins and carotenoids) are more resistant to disruption in the digestive tract than chromoplasts. Thus, the location of a carotenoid in the cell affects its accessibility.

Moreover, carotenoid isomerization also affect the intestinal absorption of carotenoids, being *cis* isomers more bioavailable. The isomerization can occur in the acidic gastric milieu. Lycopene present in fruits and vegetables occurs almost exclusively as the all-*trans* isomer, but is converted in the stomach to *cis* isomers. Although almost 100% of lycopene in red tomatoes is in the all-*trans* form, plasma and tissue profiles show that *cis* isomers make up more than 50% of the total lycopene present. However, studies show that no *trans*/*cis* isomerization of β-carotene occurs in the stomach. Evidence has been found for transfer of a significant portion of both β- and α-carotene to the fat phase of the meal in the stomach, which would increase bioavailability of these carotenoids for absorption.

Table 1 Carotenoid content (mg per 100 g fresh weight) in foods.[b]

Carotenoid	*Concentration (mg per 100 g fresh weight)*	*Food*[a]
Lycopene	27.0–776.0	Gac aril
	1.14–3.42	Tomato (*Tomatoes, red, ripe, raw, year round average*)
	3.04–5.59	Watermelon (*Watermelon, raw*)
	4.7–5.5	Guava (*Guavas common, raw*)
	0.68–3.67	Papaya (*Papayas, raw*)
	0.16–3.36	Grapefruit, pink (*Grapefruit raw, pink and red, all areas*)
β-carotene	20.0–82.7	Gac aril
	1.99–21.0	Carrot (*Carrot, raw*)
	1.38–4.71	Cantaloupe (*Melons, cantaloupe, raw*)
	2.16–3.83	Kale (*Kale, raw*)
	4.03–16.0	Sweet potato (*Sweet potato, raw, unprepared*)
	3.97–8.9	Spinach (*Spinach, raw*)
	6.95	Turnip greens (*Turnip greens, raw*)
	0.62–1.77	Apricot (*Apricots, raw*)
	0.18–0.57	Tomato (*Tomatoes, red, ripe, raw, year round average*)
	1.36–8.38	Squash, butternut (*Squash, winter, butternut, raw*)
	2.72–4.57	Swiss chard (*Chard, swiss, raw*)
	0.19–1.68	Mango (*Mangos, raw*)
	1.0–5.4	Collards (*Collards, raw*)
	0.25–2.34	Grapefruit, pink (*Grapefruit raw, pink and red, all areas*)
Lutein + zeaxanthin	4.46–8.56	Kale (*Kale, raw*)
	0.67–10.2	Parsley (*Parsley, fresh*)
	9.5–15.9	Spinach (*Spinach, raw*)
	3.13–5.99	Collards (*Collards, raw*)
	0.43–2.06	Broccoli (*Broccoli, raw*)
	5.77	Watercress (*Watercress, raw*)
	1.09	Egg yolk, chicken (*Egg, yolk, raw, fresh*)
	0.8–1.34	Cilantro (*Coriander (cilantro) leaves, raw*)
	0.03–0.07	Pepper, sweet red (*Peppers, sweet, red, raw*)
	13.2	Pepper, chilli (*Spices, pepper, red or cayenne*)
β-cryptoxanthin	0.04–2.2	Pepper, sweet red (*Peppers, sweet, red, raw*)
	1.23–1.66	Japanese persimmon (*Persimmons, Japanese, raw*)
	6.25	Pepper, chilli (*Spices, pepper, red or cayenne*)
	0.04–0.49	Tangerine (*Tangerines, (mandarin oranges), raw*)
	0–0.4	Cilantro (*Coriander (cilantro) leaves, raw*)
	0.01–1.26	Papaya (*Papayas, raw*)
	0–0.31	Watermelon (*Watermelon, raw*)
α-carotene	0.91–9.71	Carrot (*Carrot, raw*)
	0.83	Squash, butternut (*Squash, winter, butternut, raw*)
	0–0.03	Collards (*Collards, raw*)
	0–0.11	Tomato (*Tomatoes, red, ripe, raw, year round average*)

[a]The reference name of the food identified in the USDA database is shown in brackets.
[b]All data have been obtained from USDA database, except Gac aril Wimalasiri et al. (2017).

In the intestinal lumen (Fig. 6) where carotenoids are released from the food matrix, cleavage of carotenoproteins and fatty acid esters by carboxylic ester hydrolase (CEH), which is secreted by the pancreas, can occur. Carotenoids are then solubilized into lipid micelles (Reboul, 2019).

Absorption

Carotenoid absorption is saturable and somewhat specific (e.g., all-*trans*-β-carotene is preferentially absorbed to *cis*-β-carotenes and α-carotene). The presence of other carotenoids can affect the absorption of carotenoids into intestinal mucosal cells, because one carotenoid can compete for or facilitate the absorption of another. Human studies show that β-carotene decreases lutein absorption, whereas lutein has either no effect or a lowering effect on β-carotene absorption. The inhibitory effect of lutein on β-carotene absorption might be partly attributed to the inhibition of the β-carotene cleavage enzyme by lutein, shown in rats. β-carotene also seemed to lower absorption of canthaxanthin, whereas canthaxanthin did not inhibit β-carotene absorption. Studies showed that β-carotene increased lycopene absorption, although lycopene had no effect on β-carotene. α-Carotene and β-cryptoxanthin show high serum responses to dietary intake compared to lutein. In addition, *cis* isomers of lycopene seem to be more bioavailable than the all-*trans*, and selective intestinal absorption of all-*trans* β-carotene occurs, as well as conversion of the 9-*cis* isomer to all-

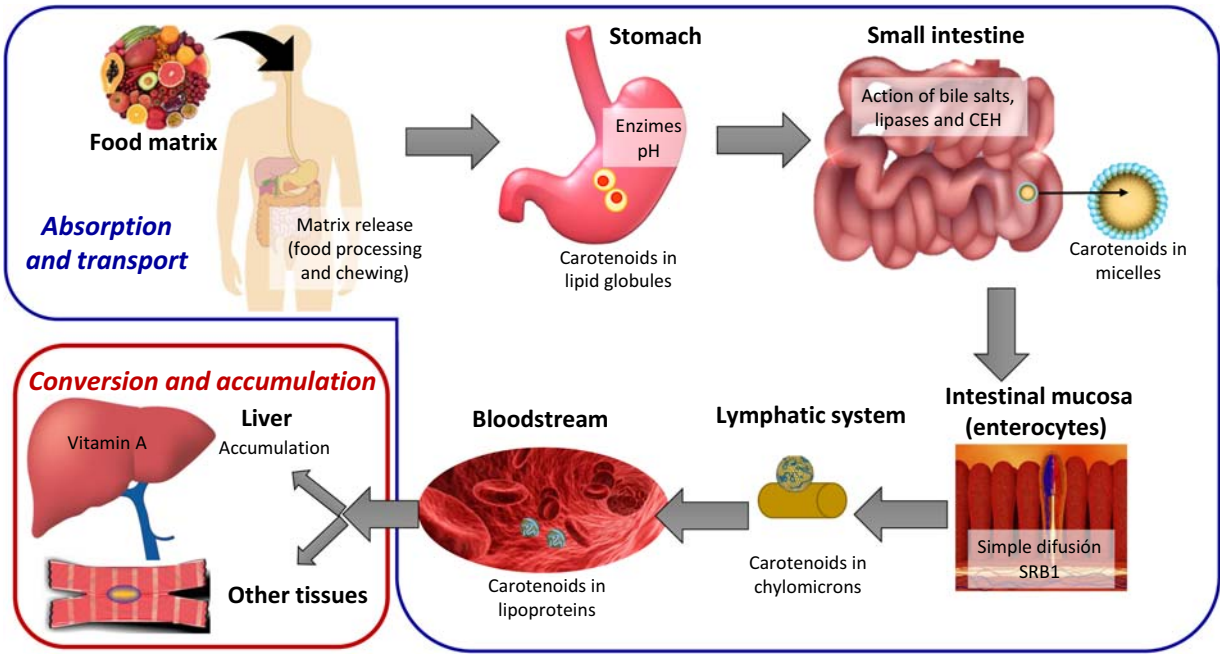


Fig. 6 Bioavailability of carotenoids. CEH: carboxylic ester hydrolase (secreted by the pancreas). Adapted from Chacón-Ordóñez and Esquivel (2013).

trans β-carotene. It is clear, then, that selective absorption of carotenoids takes place into the intestinal mucosal cell. These results, as well as cell culture and isotope data suggest that much of carotenoid absorption is controlled by an active transport mechanism (Fig. 5). A major receptor facilitating this absorption is scavenger receptor B1 (SRB1), which also facilitates the absorption of cholesterol and vitamin E. However, abolishing SRB1 activity does not completely abolish carotenoid absorption, which can also occur through passive diffusion. SRB1 may also have a role in the differential accumulation of carotenoids in some tissues, such as increased carotenoid uptake in the retina.

Carotenoids are more efficiently absorbed when accompanied by at least a small amount of fat. The amount of fat for optimal carotenoid absorption seems to differ among carotenoids. For example, lutein esters require more fat for optimal absorption than β-carotene. These differences have not been quantified for each carotenoid. In addition, the presence of a nonabsorbable, fat-soluble component decreases carotenoid absorption. Sucrose polyester, a nonabsorbable fat replacer decreased carotenoid levels in plasma by 10–60%. The extent of this inhibition depends on the amount of nonabsorbable compound ingested, as well as the particular carotenoid under consideration. The mechanism for this inhibition is apparently similar to the action of fiber, i.e., sequestration. The type of fat that is ingested along with carotenoids will also affect carotenoid absorption. As macerated food passes into the intestinal lumen, carotenoids freed from the food matrix then become incorporated into micelles, consisting of free fatty acids, monoglycerides, phospholipids, and bile acids. Many other factors can affect intestinal absorption such as micelle size, phospholipid composition, and solubilization of carotenoids into mixed micelles, and concentration of available bile salts, among others.

Another complicating factor in the intestinal mucosal cell is the partial conversion of provitamin A carotenoids (β- and α-carotenes and cryptoxanthin) to vitamin A (primarily to retinyl esters). Therefore, in absorption studies these metabolic reactions must be accounted for in measuring intestinal transport. Non-provitamin A carotenoids such as lycopene, lutein, and zeaxanthin are incorporated intact, although some cleavage can occur (Reboul, 2019).

Transport

In the intestinal mucosa, both carotenoids and retinyl esters are incorporated into chylomicrons and secreted into the lymph for transport to blood. In blood, lipoprotein lipase rapidly degrades chylomicrons, and the liver sequesters the resulting carotenoid-containing fragments. The liver then secretes carotenoids back into the bloodstream in association with hepatic very low-density lipoproteins (VLDL). Most carotenoids in fasting plasma are carried by low-density lipoproteins (LDL) and high-density lipoproteins (HDL). Seventy-five percent of the hydrocarbon carotenoids, e.g., lycopene and β-carotene, are associated with LDL, the rest is associated with HDL and, in smaller amounts, with VLDL. More polar carotenoids such as lutein and zeaxanthin are found equally distributed between HDL and LDL. After ingestion, carotenoids first appear in the bloodstream in chylomicrons, resulting from excretion from intestinal mucosal cells (4–8 h). HDL carotenoid levels peak in the circulation between 16 and 28 h; LDL carotenoid levels peak between 24 and 48 h. The bloodstream then transports carotenoids to different tissues (e.g., liver, prostate gland, fat, ocular macula) where they are sequestered by various mechanisms (Reboul, 2019).

Tissue distribution

In general, carotenoid concentrations in serum and tissues reflect concentrations contained in the food that is ingested. Carotenoids have been found in various human organs and tissues. These include human liver, lung, breast, cervix, skin, adipose, and ocular tissues. Only about 20 of them have been detected in the human body, and only 6 of these represent the 90–95% of the total carotenoids detected in human plasma: α-Carotene, β-carotene, β-cryptoxanthin, lutein, zeaxanthin, and lycopene (Olmedilla et al., 2001).

The major storage organs are adipose tissue (probably because of its volume) and the liver. Tissues containing large amounts of LDL receptors seem to accumulate high levels of carotenoids, probably as a result of nonspecific uptake by lipoprotein carriers. Preferential uptake, however, is indicated in some cases. For example, unusually high concentrations of phytoene in the lung, ζ-carotene and phytofluene in breast tissue, lycopene in the prostate and colon, lycopene, β-carotene, and phytofluene in cervical tissue, and lutein and zeaxanthin in ocular tissues have been found.

Effects of carotenoids on health

Numerous studies indicate that carotenoids and their metabolites play a role in combating oxidative damage and other degradative reactions that are harmful to human health. Most of these functions seem to be related to their antioxidant nature and ability to dissipate energy from light and free radical-generating reactions. Some carotenoids (α-carotene, β-carotene and β-cryptoxanthin) are also precursors of vitamin A, an essential nutrient. Much research is still required to shed light onto mechanisms involved in these functions. Other fascinating roles in nature are also being discovered, for example, the signaling of apparent good health and consequently good potential parenting in birds by the red coloration of beaks, which seems to serve as an attractant to prospective mates. All this mechanism of actions they play significant roles in promoting human health and reducing the risk of chronic diseases Different studies have provide that all these mechanisms of action may provide desirable health benefits to reduce the risk of the development of chronic diseases such as eye and skin disease, type II diabetes, coronary artery disease, cancer, etc. (**Fig. 7**) (Böhm et al., 2021).

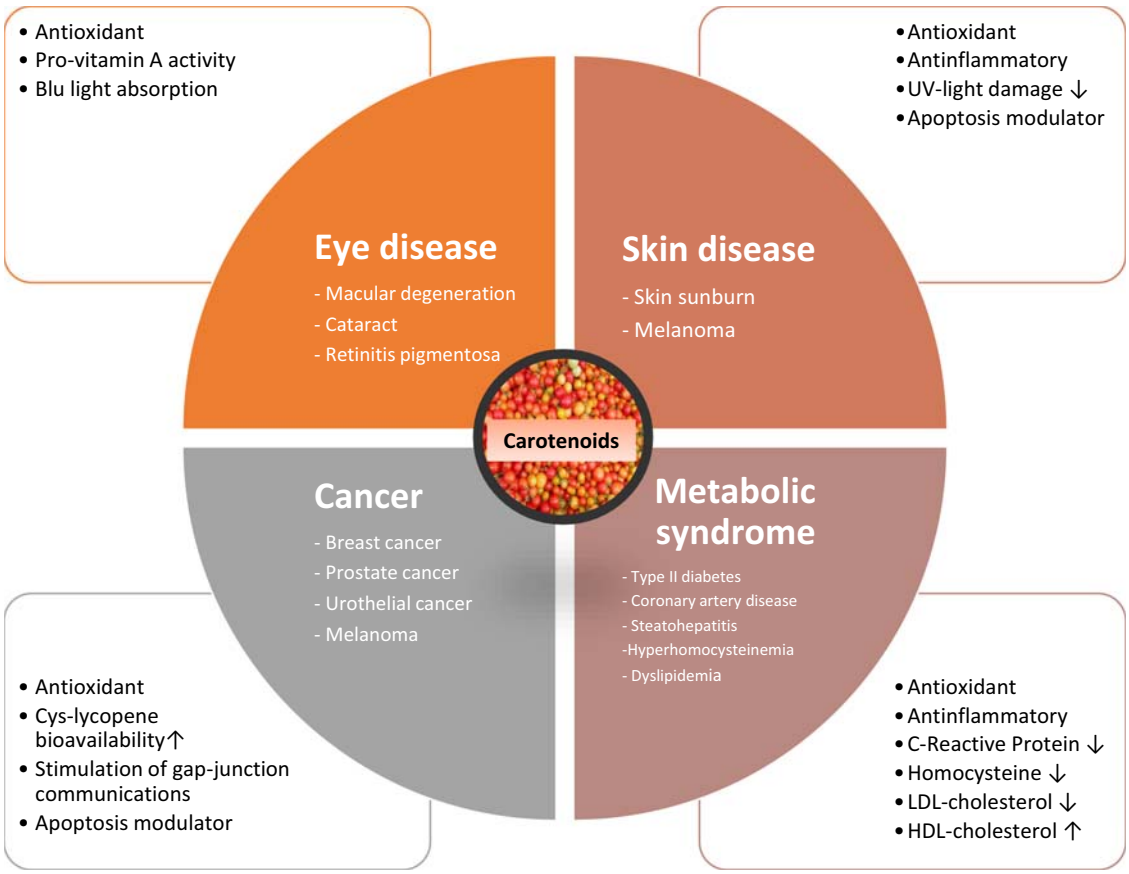


Fig. 7 Mechanism of action of carotenoids against chronic diseases. Adapted from Cicero and Colletti (2017).

Conclusions

Carotenoids are fat-soluble pigments that cannot be synthesized by humans, so they must be obtained from dietary sources, being present in the human diet around 40 of the 1204 natural carotenoids exiting in the nature. The richest sources are those highly pigmented in red, orange, and yellow colors. Fruit and vegetables are considered the most important sources for carotenoids in the human diet. However, the contribution of some animal food such as egg yolk, dairy products and seafood is also important. During storage and processing, carotenoids are susceptible to oxidative degradation and isomerization, resulting from exposure to oxygen, light, heat. This situation affects their bioavailability and antioxidant activity and must be kept in mind. Only about 20 carotenoids have been detected in the human body (liver, lung, breast, cervix, skin, adipose, and ocular tissues), and only 6 of them (α-carotene, β-carotene, β-cryptoxanthin, lutein, zeaxanthin, and lycopene) are mainly detected in human plasma. Some carotenoids are precursors of vitamin A, an essential nutrient. Other carotenoids and their metabolites provide desirable health benefits to reduce the risk of the development of chronic diseases such as eye and skin disease, type II diabetes, coronary artery disease, cancer, etc., linked to their antioxidant nature and their ability to dissipate energy from light and free radical-generating reactions.

References

Böhm, V., Lietz, G., Olmedilla-Alonso, B., Phelan, D., Reboul, E., Bánati, D., Borel, P., Corte-Real, J., De Lera, A.R., Desmarchelier, C., Dulinska-Litewka, J., 2021. From carotenoid intake to carotenoid blood and tissue concentrations–implications for dietary intake recommendations. Nutr. Rev. 79 (5), 544–573.

Cicero, A., Colletti, A., 2017. Effects of carotenoids on health: are all the same? results from clinical trials. Curr. Pharm. Des. 23 (17), 2422–2427.

Chacón-Ordóñez, T., Esquivel, P., 2013. Frutos tropicales como fuente de carotenoides: biosíntesis, composición, biodisponibilidad y efectos del procesamiento. Revista Venezolana de Ciencia y Tecnología de Alimentos 4, 001–023.

Dias, M.G., Borge, G., Kljak, K., et al., 2021. European database of carotenoid levels in foods. Factors affecting carotenoid content. Foods 10 (5), 912.

Fernandes, A.S., Nascimento, T.C., Jacob-Lopes, E., Rosso, V.V., Zepka, L.Q., 2018. Introductory chapter: carotenoids—a brief overview on its structure, biosynthesis, synthesis, and applications. In: Zepka, L., Jacob-Lopes, E. (Eds.), Progress in Carotenoid Research. IntechOpen, Rijeka.

Gruszecki, W.I., Strzałka, K., 2005. Carotenoids as modulators of lipid membrane physical properties. Biochim. Biophys. Acta 1740 (2), 108–115.

Maoka, T., 2020. Carotenoids as natural functional pigments. J. Nat. Med. 74 (1), 1–16.

Meléndez-Martínez, A.J., Mapelli-Brahm, P., Hornero-Méndez, D., Vicario, I.M., 2019. Chapter 1: structures, nomenclature and general chemistry of carotenoids and their esters. In: Mercadante, A.Z. (Ed.), Food Chemistry, Function and Analysis. Royal Society of Chemistry, London, pp. 1–50.

Olmedilla, B., Granado, F., Southon, S., et al., 2001. Serum concentrations of carotenoids and vitamins A, E, and C in control subjects from five European countries. Br. J. Nutr. 85 (2), 227–238.

Reboul, E., 2019. Mechanisms of carotenoid intestinal absorption: where do we stand? Nutrients 11 (4), 838.

Ribeiro, D., Freitas, M., Silva, A., et al., 2018. Antioxidant and pro-oxidant activities of carotenoids and their oxidation products. Food Chem. Toxicol. 120, 681–699.

Sandmann, G., 2019. Antioxidant protection from UV- and light-stress related to carotenoid structures. Antioxidants 8 (7), 219.

Stahl, W., Sies, H., 2003. Antioxidant activity of carotenoids. Mol. Aspect. Med. 24 (6), 345–351.

United States Department of Agriculture, 2021. United States Department of Agriculture (2021) USDA National Nutrient Database for Standard Composition of Foods Raw, Processed, Prepared. Available from. https://fdc.nal.usda.gov/.

Wimalasiri, D., Brkljača, R., Piva, T.J., Urban, S., Huynh, T., 2017. Comparative analysis of carotenoid content in Momordica cochinchinensis (Cucurbitaceae) collected from Australia, Thailand and Vietnam. J. Food Sci. Technol. 54 (9), 2814–2824.

Further reading

Britton, G., 1995. Structure and properties of carotenoids in relation to function. FASEB J. 9 (15), 1551–1558.

Jomova, K., Valko, M., 2013. Health protective effects of carotenoids and their interactions with other biological antioxidants. Eur. J. Med. Chem. 70, 102–110.

Kiokias, S., Proestos, C., Varzakas, T., 2016. A review of the structure, biosynthesis, absorption of carotenoids-analysis and properties of their common natural extracts. Curr. Res. Nutr. Food Sci. J. 4, 25–37.

Meléndez-Martínez, A.J., Vicario, I.M., Heredia, F.J., 2007. Pigmentos carotenoides: consideraciones estructurales y fisicoquímicas [Carotenoid pigments: structural and physicochemical considerations]. Archivos latinoamericanos de nutricion 57 (2), 109–117.

Mordi, R.C., Ademosun, O.T., Ajanaku, C.O., Olanrewaju, I.O., Walton, J.C., 2020. Free radical mediated oxidative degradation of carotenes and xanthophylls. Molecules 25 (5), 1038.

Namitha, K.K., Negi, P.S., 2010. Chemistry and biotechnology of carotenoids. Crit. Rev. Food Sci. Nutr. 50 (8), 728–760.

Rodriguez-Amaya, D.B., 1999. A Guide to Carotenoid Analysis in Foods. ILSI Press, Washington, D.C.

Rodriguez-Concepcion, M., Avalos, J., Bonet, M.L., et al., 2018. A global perspective on carotenoids: metabolism, biotechnology, and benefits for nutrition and health. Prog. Lipid Res. 70, 62–93.

Yabuzaki, J., 2020. Carotenoids Database: Structures, Chemical Fingerprints and Distribution Among Organisms. Database, Oxford. Available from. http://carotenoiddb.jp/.

Carotenoids: Health effects

Sherry A. Tanumihardjo, Department of Nutritional Sciences, University of Wisconsin-Madison, Madison, WI, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of S.A. Tanumihardjo, Carotenoids: Health Effects, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 292–297, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00045-3.

Introduction 154
Hydrocarbon carotenoid: β-Carotene 157
Hydrocarbon carotenoid: α-Carotene 158
Xanthophyll: β-Cryptoxanthin 158
Hydrocarbon carotenoid: lycopene 158
Xanthophylls: lutein and zeaxanthin 159
Further steps in carotenoids and health research 160
References 160

Key points

- Carotenoids are a group of phytochemicals that are not currently considered essential nutrients
- Bioavailability of carotenoids from foods is very important, especially when consuming plant-based diets
- Carotenoids have been implicated in mitigation of various diseases such as age-related macular degeneration (AMD) and cataract, certain cancers and cardiovascular disease
- Lutein and zeaxanthin, xanthophyll carotenoids, are found in a specific location in the human retina

Glossary

Antioxidants Compounds that neutralize free radicals in the body formed from normal biological processes
Bioaccessibility The amount of a nutrient or compound of interest that is released from food and available for absorption
Bioavailability How much of a compound of interest is absorbed from food and available for physiological function or storage
Bioconversion The amount of active compound that is formed from an absorbed precursor
Bioefficacy The amount of active compound that is made by the body from how much was theoretically contained in food or a supplement
Carotenoids A group of compounds found in fruits and vegetables, which generally consist of hydrocarbon isoprenoid groups
Hydrocarbon carotenoids Carotenoids that do not contain oxygen but are composed of hydrogen and carbon chains
Xanthophylls Oxygen-containing carotenoids

Introduction

The colors of many fruits and vegetables are due to carotenoids (Fig. 1). Over 700 carotenoids have been identified in nature. Humans absorb carotenoids from the foods that we consume whereas many other animals do not circulate carotenoids. Carotenoids are considered an important class of phytochemicals, which are compounds derived from plants that may or may not have nutritional value. Many carotenoids circulate in humans and the most commonly studied include β-carotene, α-carotene, β-cryptoxanthin, lycopene, lutein, and zeaxanthin. Carotenoids are nutritionally significant because about 50 of them are cleaved by the body to make vitamin A. These are known as provitamin A carotenoids and the three most abundant in fruit and vegetables are β-carotene, α-carotene, and β-cryptoxanthin (Tanumihardjo et al., 2010). Provitamin A carotenoids, especially β-carotene, provide less than one-half of the vitamin A supply in North America, but provide more than one-half in Africa and Asia (Haskell, 2013).

Dietary recommendations for the intake of specific carotenoids have not been established due to lack of evaluation of scientific evidence. To date, carotenoids are not considered essential nutrients (Institute of Medicine, Food and Nutrition Board, 2000). Current dietary recommendations for vitamin A are 900 retinol activity equivalents (RAE) for men and 700 RAE for women (Institute of Medicine, Food and Nutrition Board, 2001). An RAE is equivalent to 1 μg of retinol, the active form circulating in humans and the storage form of vitamin A when esterified to fatty acids as retinyl esters. The vitamin A recommendations for infants and children are less than adults and range from 300 to 600 RAE dependent upon age (Institute of Medicine, Food and Nutrition Board, 2001). If dietary preformed vitamin A is not sufficient, consumers need to consume sufficient amounts of carotenoid-rich fruits and

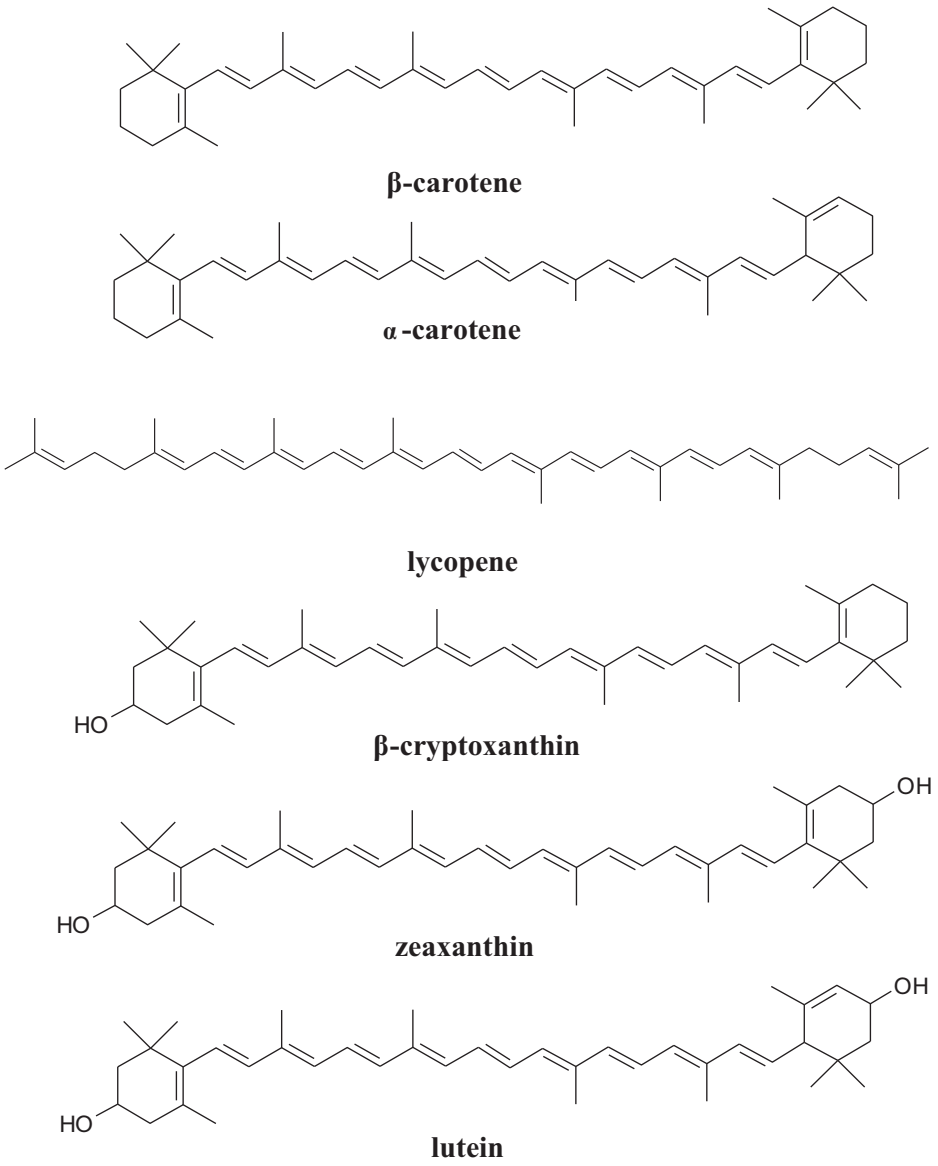


Fig. 1 The structures of the most common carotenoids found in the human body. Three of them, β-carotene, α-carotene and β-cryptoxanthin, can be used by the body for vitamin A. All carotenoids are antioxidants found in fruit and vegetables.

vegetables to meet their daily vitamin A requirement, achieve optimal dietary carotenoid intake, and lower the risk of certain chronic diseases. In 2001, the Institute of Medicine established the amount of carotenoids needed to provide vitamin A from foods as 12 μg β-carotene or 24 μg other provitamin A carotenoids to yield 1 RAE (Institute of Medicine, Food and Nutrition Board, 2001). Currently, high-dose pharmacological supplementation with carotenoids is not advised. Despite this, a tolerable upper intake level, the maximum daily amount of a nutrient that appears to be safe, has not been established for any individual carotenoid; although, supplemental β-carotene at 20 mg/day or more is likely contraindicated for use in current, heavy smokers based on the studies described below.

Many factors affect the release and absorption of carotenoids from foods (**Fig. 2** from Tanumihardjo et al., 2010). When most sources of dietary vitamin A are from provitamin A carotenoids, bioavailability from the food matrix becomes important. Bioavailability of preformed vitamin A, i.e., retinol and retinyl esters, is not a major concern because 80–95% is absorbed when adequate fat is consumed at the same time. Foods that are high in preformed retinol (e.g., liver, eggs, and fortified milk), however, are not consumed by everybody due in part to the cost of animal-source foods and availability (Haskell, 2013). When discussing carotenoids from fruit and vegetables, there are four terms that need to be defined: bioaccessibility, bioavailability, bioconversion, and bioefficacy (Tanumihardjo et al., 2010) (see **Table 1**).

- Bioaccessibility refers to how much carotenoid is released from the food and available for absorption.
- Bioavailability is how much carotenoid is absorbed from that released from the food and available for physiological function.

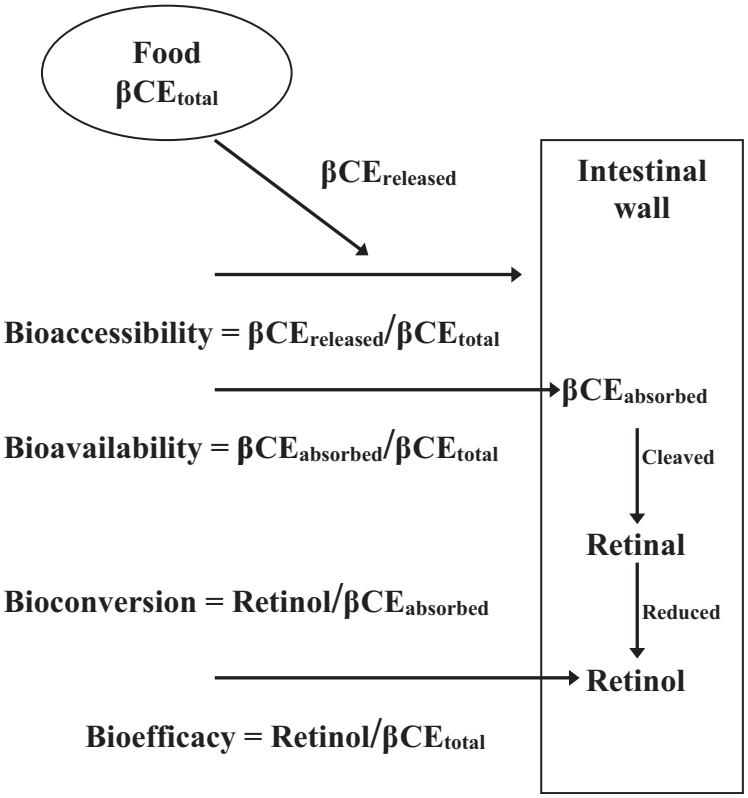


Fig. 2 A schematic outlining the path of provitamin A carotenoids as β-carotene equivalents (βCE) as they move out from the food into the intestinal wall. The definition of terms associated with understanding provitamin A release, absorption and conversion to retinol are illustrated: bioaccessibility, bioavailability, bioconversion, and bioefficacy. βCE refers to β-carotene equivalents and is usually equal to β-carotene plus ½ α-carotene and ½ β-cryptoxanthin. Reproduced with permission from the Tanumihardjo et al. (2010).

Table 1 Terms which are associated with the provitamin A carotenoid value of foods based on β-carotene and subsequent utilization as retinol.

Term	*Definition*	*100%*
Bioaccessibility	β-Carotene released	1 μmol released
	β-Carotene in food	1 μmol in food
Bioavailability	β-Carotene absorbed	1 μmol absorbed
	β-Carotene in food	1 μmol in food
Bioconversion	Retinol formed	2 μmol formed
	β-Carotene absorbed	1 μmol absorbed
Bioefficacy	Retinol formed	2 μmol formed
	β-Carotene in food	1 μmol in food

- Bioconversion relates to the provitamin A carotenoids and is defined as the amount of retinol that is formed from absorbed provitamin A carotenoids.
- Bioefficacy encompasses all of the biological processing of provitamin A carotenoids and is the amount of retinol formed from the amount of carotenoid contained in the food or supplement.

The study of provitamin A carotenoid bioefficacy from foods is important in international health because the most frequently consumed sources of vitamin A are fruit and vegetables. A 100% bioefficacy means that 1 μmol of dietary β-carotene provides 2 μmol of retinol in the body; however, 100% bioefficacy does not actually occur in the process of digestion and provitamin A carotenoid uptake by the body. Determining bioefficacy is an important research endeavor due to the introduction of provitamin

A-enhanced staple crops into countries at risk for vitamin A deficiency (Tanumihardjo et al., 2010). Staple crops targeted for provitamin A-enhancement, a process known as biofortification, include sweet potato, maize, cassava, and rice.

Once in the body, carotenoids can act as potent antioxidants, which are substances that neutralize free radicals formed from the natural metabolic processes of cells. Free radicals damage tissues and cells through oxidative processes. While free radical formation is a natural process in the body, environmental factors such as smoking and pollution can increase free radical load and thus increase disease risk. Carotenoids may counter these influences by functioning as antioxidants and quenching oxygen-containing free radicals (Landrum, 2013). At the whole-body level, some population studies have indicated that certain carotenoids from either dietary intake or blood concentration data are associated with better immune response, lower rates of age-related macular degeneration (AMD) and cataract, as well as lower risk for certain cancers, cardiovascular disease, and osteoporosis. These diseases and degenerative processes are linked to reactive oxygen species cascades (Landrum, 2013). The associations between specific carotenoids and decreased risk of various diseases are summarized in **Table 2**. Some of these purported health effects may be due to their function as antioxidants.

Blood levels of specific carotenoids are often used as biomarkers for fruit and vegetable intake to strengthen dietary intake data or show compliance to an intervention (Howe et al., 2009). A wide variation in analytical methods exists and standardization between laboratories does not routinely occur. Nonetheless, higher blood concentrations have been favorably correlated with certain disease states. In a review, total carotenoid, α-carotene, β-carotene, lycopene, and lutein concentrations were significantly inversely associated in women with breast cancer compared with controls who had not developed breast cancer in multiple studies (Kim et al., 2021). Vitamin A concentrations in contrast were usually not influenced by breast cancer risk. Nonetheless, vitamin A concentrations are homeostatically controlled and affected by inflammation, which are known confounders in their interpretation. Carotenoids may be protective against or utilized during breast cancer. Furthermore, the Nurses' Health Study, which included a cohort of over 83,000 women, also showed a significant inverse association between dietary β-carotene intake and breast cancer risk (Zhang et al., 1999). This was especially strong for premenopausal women with a family history of breast cancer or high alcohol consumption. Other prospective studies, however, have had mixed results (Kim et al., 2021).

Hydrocarbon carotenoid: β-Carotene

β-Carotene is one of the most widely studied carotenoids, both for its vitamin A activity and its abundance in fruits and vegetables. Epidemiological studies have often pointed to the abundance of dietary carotenoids as being protective against many diseases. Diets rich in fruits and vegetables are recommended to reduce the risk of disease and promote optimal health. Current dietary guidelines for a 2000-calorie intake level recommend 4.5 cups of fruit and vegetables each day (U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2020).

When removed from the plant matrix and administered as supplements, however, these benefits sometimes disappear. For example, lung cancer is a leading cause of death in many countries and the β-Carotene and Retinol Efficacy Trial (CARET) in the 1990s set out to test whether β-carotene conferred cancer protection. CARET was an intervention study based on a number of observations that showed high levels of β-carotene from food sources were protective against lung cancer; however, the trial demonstrated an increased risk for lung cancer in the treatment group over the control (Christen et al., 2000). A similar outcome was observed in the α-Tocopherol β-Carotene (ATBC) Study Group (1994). Although evidence clearly exists showing an association between β-carotene and enhanced lung function, as in the CARET study, the ATBC trial also found an increase in lung cancer rates among smokers. It is plausible that the lung cancer had already been initiated in the smokers and supplementation with β-carotene could not prevent development. Subsequent studies in ferrets showed that the amounts of β-carotene commonly consumed from fruit and vegetables were protective against lung damage but higher amounts, equivalent to those in CARET, increased the formation of abnormal tissue in the lung (Liu et al., 2003). The ATBC study also showed an increased incidence of angina pectoris, a mild warning sign of heart disease characterized by chest pain, among heavy smokers (Rapola et al., 1996).

In both the CARET and ATBC interventions, much higher doses of β-carotene were used than could be obtained from the typical diet, and the blood levels attained were two to six times higher than the 95th percentile of β-carotene in a survey of a representative sample of the United States population. Thus, it remains unclear whether β-carotene is a pro-carcinogen or an anti-carcinogen. The

Table 2 A summary of epidemiologic and/or clinical studies where carotenoids and a significant association to a specific disease risk has been shown in at least one study.[a]

Carotenoid	*Cardiovascular disease*	*Cataract*	*Macular degeneration*	*Lung cancer*	*Prostate cancer*	*Bone health*
ß-Carotene	Yes	–	–	Yes[b]	–	–
α-Carotene	Yes	–	–	Yes	Yes	–
ß-Cryptoxanthin	Yes	–	–	Yes	–	Yes
Lycopene	Yes	–	–	Yes	Yes	Yes
Lutein/zeaxanthin	Yes	Yes	Yes	Yes	–	–

[a]For a more complete discussion of the association of specific carotenoids to cardiovascular disease please refer to Ciccone et al. (2013); and for cancer please refer to Rowles and Erdman (2020).

[b]The opposite finding has been observed in clinical trials with smokers.

associations for lower disease risk observed in epidemiologic studies may reflect other protective dietary agents or an interaction between dietary components. Furthermore, people with higher intake of fruits and vegetables may have healthier lifestyles that contribute to their lower risk of chronic diseases. The higher disease risk observed in clinical trials may be correlated to high dose β-carotene with yet unidentified mechanisms, the limited treatment duration, and/or the timing of the interventions in regard to cancer development due to a history of heavy smoking. More research on β-carotene's biological actions is needed to explore mechanisms. Current consensus is that the beneficial effects of β-carotene are associated with dietary consumption, whereas the harmful effects in some subpopulations are with supplements at pharmacological levels.

Another explanation for a lack of beneficial outcome with β-carotene supplementation may be that not all people respond to β-carotene treatment. A number of polymorphisms in the β-carotene cleavage enzymes have been identified (Ross and Moran, 2020). Furthermore, individuals who do not respond to β-carotene supplementation may be better at converting it to vitamin A, which could be driven by vitamin A status. Blood response to β-carotene supplementation may also be inversely related to body mass index (BMI) due to increased sequestration of lipophilic β-carotene by fat stores present in people with larger BMI. However, some individuals with larger BMI's do not necessarily have a high body fat percentage, but rather increased lean muscle mass.

Excellent food sources of β-carotene include carrots, winter squash, red-orange sweetpotato, and various types of dark green leafy vegetables. No deficiency or toxicity has been observed from dietary β-carotene intake, although high intakes can be associated with yellow pigmentation of the skin because carotenoids are stored in adipose tissue, a condition known as hypercarotenodermia (Tanumihardjo et al., 2015). Supplements containing β-carotene are common. In a large observational/intervention study in postmenopausal women, the Women's Health Initiative, approximately 50% reported using a supplement containing β-carotene. The Women's Health Initiative included both a clinical trial and observational study with more than 160,000 women and outcomes have been reviewed (Cauley and Crandall, 2020). The Physicians' Health Study II also included β-carotene as an intervention to determine the balance of risks and benefits of this carotenoid with cancer, cardiovascular disease, and eye disease (Christen et al., 2000).

Hydrocarbon carotenoid: α-Carotene

α-Carotene, another carotenoid frequently quantified in food, has provitamin A activity. Based on its structure, it is converted to one molecule of biologically active retinol after central cleavage and twice the molar amount is equivalent to the retinol from β-carotene (Tanumihardjo and Howe, 2005). Like other carotenoids, it has antioxidant and possibly anti-carcinogenic properties, and may enhance immune function as well. Some, but not all, epidemiological studies observed that higher α-carotene intake was associated with lower risk of cardiovascular disease and cancer, whereas others did not (Rowles and Erdman, 2020). High dietary intake of α-carotene was associated with decreased prostate cancer risk (Rowles and Erdman, 2020). Clinical trials to test isolated α-carotene's influences in humans have not been conducted to date. This is probably because α-carotene is usually associated with ample amounts of β-carotene when found in fruits and vegetables and singling out dietary α-carotene is difficult.

α-Carotene is especially high in orange carrots and some pumpkins; high serum concentrations are associated with carrot and pumpkin intake. Serum α-carotene concentrations were elevated in children who developed hypercarotenodermia during mango season (Tanumihardjo et al., 2015). α-Carotene is found in some combined carotenoid supplements.

Xanthophyll: β-Cryptoxanthin

β-Cryptoxanthin has provitamin A activity and may be more bioefficacious than α-carotene even though theoretically both carotenoids supply one vitamin A molecule based on structure (Tanumihardjo and Howe, 2005). This is likely due to the bipolar nature of β-cryptoxanthin. Several epidemiological studies suggest that dietary β-cryptoxanthin is associated with lower rates of lung cancer and improved lung function in humans (Burri et al., 2016). A pooled analysis of dietary carotenoid intake and cancer risk, which included seven cohort studies, identified β-cryptoxanthin intake as protective against lung cancer even after evaluating smoking status (Mannisto et al., 2004). The beneficial effects for β-cryptoxanthin suggested by these results, however, could be merely an indicator for other antioxidants and/or a measure of a healthy lifestyle that are more common in people with high dietary intakes of β-cryptoxanthin. However, β-carotene was not associated with lung cancer risk in the same analysis (Mannisto et al., 2004).

In tissue culture, β-cryptoxanthin has a direct stimulatory effect on bone formation and an inhibitory effect on bone resorption (Yamaguchi, 2012). In postmenopausal women from Italy and the United States, β-cryptoxanthin was lower in women with osteoporosis than those without (Tanumihardjo and Binkley, 2013), perhaps due to increased utilization. On the other hand serum retinol concentrations were lower (Tanumihardjo and Binkley, 2013). Further investigations are needed to elucidate the action of β-cryptoxanthin in bone health.

No deficiency or toxicity has been observed from dietary β-cryptoxanthin intake. The best food sources for β-cryptoxanthin are oranges, papaya, peaches, tangerines, and yellow and orange maize. Tropical fruit intake is directly proportional to β-cryptoxanthin blood concentrations.

Hydrocarbon carotenoid: lycopene

Lycopene, while having no provitamin A activity, is a potent antioxidant with twice the activity of β-carotene for quenching singlet oxygen and 10 times the antioxidant activity of α-tocopherol in some model systems. The antioxidant potential of food chemicals

varies widely according to location in the body and the presence of other body chemicals. Epidemiological evidence shows an inverse association between lycopene consumption and the incidence and development of certain cancers (Rowles and Erdman, 2020). This association is especially strong for prostate cancer, which is the most common cancer among men in western countries and the second leading cause of cancer death in American men. Prostate cancer rates in Asian countries are lower, but appear to be increasing rapidly due to change in lifestyle. The current consensus is that high consumption of tomatoes or high circulating concentrations of lycopene are associated with risk reduction for prostate cancer (Rowles et al., 2017). Animal and cell culture studies were reviewed to determine mechanisms of action (Applegate et al., 2019). Some studies have suggested that tomato products are more protective against prostate cancer than isolated lycopene. Epidemiologic studies have also observed lower rates of bladder, cervical, and breast cancers as well as cancers of the gastrointestinal tract among people with high intake of lycopene (Rowles and Erdman, 2020). The discovery of significant concentrations of lycopene in specific tissues in the body, in addition to prostate, suggests that lycopene may play a role in these tissues, i.e., plasma, testes, adrenal glands, liver, and kidney.

Lycopene, as an antioxidant, may be protective against heart disease by slowing down the oxidation of polyunsaturated fats in the low-density lipoprotein particles in the blood. Epidemiological and clinical studies show that higher blood lycopene concentrations are associated with lower risk and incidence of cardiovascular disease (Ciccone et al., 2013). The evidence for protective cardiovascular effects is compelling, because studies have shown an improvement in cardiovascular parameters with higher blood concentrations of lycopene but it is usually linked to increased intake of fruit and vegetables (Ciccone et al., 2013).

In addition to links with cancer and cardiovascular effects, lycopene intake was associated with lower risk for hip and nonvertebral fractures in the Framingham study (Sahni et al., 2009). Serum concentrations of lycopene were also noted to be lower in midwestern postmenopausal women with osteoporosis perhaps indicating higher utilization because dietary intake did not differ (Tanumihardjo and Binkley, 2013). Finally, higher intake of fruits and vegetables is associated with better lung function, and specifically, high tomato intake is associated with higher timed expiratory volume.

While the body of evidence seems strong, several studies have found either no or weak associations between lycopene consumption and disease. Some of this may be explained by the fact that blood lycopene concentrations were much lower in these studies than in those that showed a beneficial effect. Dietary-based studies should include blood sampling to further define the range of blood lycopene concentrations in the population and ideally with method standardization so that studies can be directly compared. The prostate cancer association is usually stronger for cooked tomato products rather than raw tomatoes or total lycopene intake (Rowles and Erdman, 2020). This supports the idea that whole foods with a broad array of nutrients and non-nutritive bioactive components are important for overall health rather than isolated compounds. The beneficial effects of tomatoes may be increased by processed, concentrated products that enhance the nutrient bioavailability.

The major food source of lycopene globally is tomatoes and tomato products. In the United States, more than 80% of dietary lycopene comes from tomatoes. Other sources include watermelon, pink grapefruit, and red carrots. Red carrots are especially prevalent in India and China.

Xanthophylls: lutein and zeaxanthin

Lutein and zeaxanthin, structural isomers, are non-provitamin A carotenoids that are measurable in human blood and tissues. Lutein and zeaxanthin are the xanthophylls that constitute the macular pigment of the human and non-human primate retina (Mares, 2016) and they may be important in cognitive function (Johnson, 2014). Biological functions include light absorption, oxidative stress protection, and inflammation mitigation. The relative concentration of lutein to zeaxanthin in the macula is distinctive. Zeaxanthin is more centralized and lutein predominates toward the outer area of the macula (Mares, 2016). A putative xanthophyll binding protein has been described that may explain the high variability of people to accumulate these carotenoids in eye tissues. Increased lutein intake from both food sources and supplements is positively correlated with increased macular pigment density, which is theorized to lower risk for macular degeneration. AMD is the leading cause of irreversible blindness in the elderly in developed countries. AMD adversely affects the central field of vision and the ability to see fine detail. Some, but not all, population studies suggest lower rates of AMD among people with higher levels of lutein and zeaxanthin in the diet or blood (Mares, 2016). Possible mechanisms of action for these carotenoids include antioxidant protection of the retinal tissue and the macular pigment filtering of damaging blue light.

Free radical damage is also linked to the development of cataracts, which causes opaqueness in the lens of the eyes. Cataract remains the leading cause of visual disability in the United States and about one-half of the 30–50 million cases of blindness throughout the world. Cataract is treatable, but blindness occurs because individuals have either chosen not to correct the disease or do not have access to the appropriate medical treatment. Several epidemiological studies have shown inverse associations between the risk of cataracts and carotenoid intake (Mares, 2016). Lutein and zeaxanthin are found in the lens and are thought to protect cells in the eye against oxidative damage, which may prevent cataract.

Human data on the consumption of lutein and zeaxanthin is important to understand disease prevention. Lutein may protect against some forms of cancer and enhance immune function, but the evidence is mixed (Rowles and Erdman, 2020). Lutein may work in concert with other carotenoids such as β-carotene to lower cancer risk due to anti-mutagenic and anti-tumor properties.

Major dietary sources of lutein and zeaxanthin include corn, green leafy vegetables, and eggs. Lutein tends to be the predominant isomer in foods although some varieties of maize contain significant amounts of zeaxanthin. One complicating factor that requires better understanding is the bioavailability of lutein from food sources and supplements. The food matrix is an important factor influencing lutein bioavailability and the amount and type of food processing generally influences the bioavailability of all

carotenoids. For example, the processing of spinach does not affect bioavailability of lutein, but it does enhance that of β-carotene. Bioavailability studies have been conducted with lutein supplements and/or foods containing lutein (Mares, 2016). In humans, lutein from vegetables seems to be more bioavailable than that of β-carotene; however, this may be partially explained by bioconversion of β-carotene to vitamin A. Competition between carotenoids, such as lutein and β-carotene, for incorporation into chylomicra has been noted in humans consuming vegetables and supplements. The amount of fat consumed with the lutein source also affects bioavailability, as higher fat increases the bioavailability of lipid-soluble carotenoids. Lutein from egg yolk and oil-based supplements is very bioavailable due to the fat matrix. Due to potential health benefits, lutein and zeaxanthin supplements are sold commercially and incorporated into some multivitamins. Levels of these xanthophylls in single supplements vary widely, and neither benefit nor safety have been adequately studied (Mares, 2016). Lutein supplements are often derived from marigold flowers.

Further steps in carotenoids and health research

Fruit and vegetable consumption, good sources of many antioxidants including carotenoids, is a preventative measure for many diseases. Most of the epidemiological evidence suggests that carotenoids are a very important class of phytochemicals. While some of the effects may be attributable to a diet high in fruits and vegetables, and an overall healthy lifestyle, the presence of specific carotenoids localized in different areas of the human body lend evidence to their overall importance in optimal human health. Non-invasive methods have been developed to assess carotenoid levels in the skin and eye. Large-scale studies that determine carotenoid levels in blood, skin, and the eye may lead to a better understanding of their importance in human health and disease prevention. Additional epidemiologic studies to further strengthen the associations that have been observed in populations are needed.

Study design and statistical analyses vary across published work and no one study can give conclusive evidence. An integrated multidisciplinary approach to study the functions and actions of carotenoids in the body is necessary to fully understand the role of carotenoids in health and disease prevention. This includes comparisons of carotenoids in whole fruits and vegetables and their effect on human health and well-being. High fruit and vegetable intake is associated with a decreased risk of cancer, cardiovascular disease, diabetes, AMD and osteoporosis. Removing any one class of phytochemicals from the intricate matrix of the whole plant may not give the same beneficial outcome in terms of human health. Considering that the average intake of fruit and vegetables is still less than recommendations, programs that promote their consumption may be more effective at preventing disease in the long-term than using individual pharmacological carotenoid supplements.

A question that remains is whether or not carotenoids can be considered nutrients. A variety of phytochemicals contained in fruits and vegetables, including carotenoids, are assumed to be needed for optimal health and reduction of chronic disease risk, but have not been classified as nutrients. Daily Reference Intakes are still not available for any carotenoid (Ross and Moran, 2020). Factors have been defined that categorize substances as nutrients: (1) they must be obtained from the diet because the body cannot synthesize the bioactive form, and be used for growth, maintenance, or tissue repair; (2) studies must be done to determine the essentiality of the substance and its specific function in the body and these are typically randomized, controlled interventions in humans (Courtney Gaine et al., 2013); (3) the concentration in specific tissues needs to be defined, and consumption and/or supplementation must result in tissue concentration increases and improved tissue function, such as has been done for lutein and zeaxanthin (Mares, 2016); and (4) accurate intake assessment needs to be defined with biomarkers identified to assess status (Courtney Gaine et al., 2013).

A large body of observational studies suggests that high blood concentrations of carotenoids obtained from food are associated with chronic disease risk reduction. However, there is little evidence of their specific role in the body. Lutein and zeaxanthin are the only carotenoids found in a specific tissue, the macular region of the retina, which seem to have specific functions. Providing lutein in the diet increases macular pigment in humans. Animal studies show that a diet low in lutein can deplete macular pigment, but the influence on the health of the eye is not yet well-understood. To further our understanding, large randomized prospective intervention trials need to be conducted to explore the essentiality of lutein supplementation for reducing ocular disease risk in humans. Although the evidence is mounting for lutein, to date, no specific carotenoid has been classified as an essential nutrient.

References

Applegate, C.C., Rowles III, J.L., Erdman Jr., J.W., 2019. Can lycopene impact the androgen axis in prostate cancer?: a systematic review of cell culture and animal studies. Nutrients 11, 633.

Burri, B.J., La Frano, M.R., Zhu, C., 2016. Absorption, metabolism, and functions of β-cryptoxanthin. Nutr. Rev. 74 (2), 69–82.

Cauley, J.A., Crandall, C., 2020. The Women's Health Initiative: a landmark resource for skeletal research since 1992. J. Bone Miner. Res. 35, 845–860.

Christen, W.G., Gaziano, J.M., Hennekens, C.H., 2000. Design of Physicians' Health Study II–a randomized trial of beta-carotene, vitamins E and C, and multivitamins, in prevention of cancer, cardiovascular disease, and eye disease, and review of results of completed trials. Ann. Epidemiol. 10, 125–134.

Ciccone, M.M., Cortese, F., Gesualdo, M., Carbonara, S., Zito, A., Ricci, G., De Pascalis, F., Scicchitano, P., Riccinoni, G., 2013. Dietary intake of carotenoids and their antioxidant and anti-inflammatory effects in cardiovascular care. Mediat. Inflamm. 2013, 782137.

Courtney Gaine, P., Balentine, D.A., Erdman Jr., J.W., Dwyer, J.T., Ellwood, K.C., Hu, F.B., Russell, R.M., 2013. Are dietary bioactives ready for recommended intakes? Adv. Nutr. 4 (5), 539–541.

Haskell, M.J., 2013. Provitamin A carotenoids as a dietary source of vitamin A. In: Tanumihardjo, S.A. (Ed.), Carotenoids and Human Health. Springer Science, New York, pp. 249–260.

Howe, J.A., Valentine, A.R., Hull, A.K., Tanumihardjo, S.A., 2009. ^{13}C natural abundance in serum retinol acts as a biomarker for increases in dietary provitamin A. Exp. Biol. Med. 234, 140–147.
Institute of Medicine, Food and Nutrition Board, 2000. Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids. National Academy Press, Washington, DC.
Institute of Medicine Food and Nutrition Board, 2001. Vitamin A. in: Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc. National Academies Press, Washington, DC, pp. 82–161.
Johnson, E.J., 2014. Role of lutein and zeaxanthin in visual and cognitive function throughout the lifespan. Nutr. Rev. 72 (9), 605–612.
Kim, J.A., Jang, J.-H., Lee, S.-Y., 2021. An updated comprehensive review on vitamin A and carotenoids in breast cancer: mechanisms, genetics, assessment, current evidence, and future clinical implications. Nutrients 13 (9), 3162.
Landrum, J.T., 2013. Reactive oxygen and nitrogen species in biological systems: reactions and regulation by carotenoids. In: Tanumihardjo, S.A. (Ed.), Carotenoids and Human Health. Springer Science, New York, pp. 57–101.
Liu, C., Russell, R.M., Wang, X.D., 2003. Exposing ferrets to cigarette smoke and a pharmacological dose of beta-carotene supplementation enhance in vitro retinoic acid catabolism in lungs via induction of cytochrome P450 enzymes. J. Nutr. 133, 173–179.
Männistö, S., Smith-Warner, S.A., Spiegelman, D., Albanes, D., Anderson, K., van den Brandt, P.A., Cerhan, J.R., Colditz, G., Feskanich, D., Freudenheim, J.L., Giovannucci, E., Goldbohm, R.A., Graham, S., Miller, A.B., Rohan, T.E., Virtamo, J., Willett, W.C., Hunter, D.J., 2004. Dietary carotenoids and risk of lung cancer in a pooled analysis of seven cohort studies. Cancer Epidemiol. Biomark. Prev. 13 (1), 40–48.
Mares, J., 2016. Lutein and zeaxanthin isomers in eye health and disease. Annu. Rev. Nutr. 36, 571–602.
Rapola, J.M., Virtamo, J., Haukka, J.K., Heinonen, O.P., Albanes, D., Taylor, P.R., Huttunen, J.K., 1996. Effect of vitamin E and beta-carotene on the incidence of angina pectoris. A randomized, double-blind, controlled trial. J. Am. Med. Assoc. 275, 693–698.
Ross, A.C., Moran, N.E., 2020. Our current dietary reference intakes for vitamin A—now 20 years old. Curr. Dev. Nutr. 4 (10), nzaa096.
Rowles III, J.L., Erdman Jr., J.W., 2020. Carotenoids and their role in cancer prevention. BBA Mol. Cell Biol. Lipids 1865, 158613.
Rowles III, Ranard, K.M., Smith, J.W., An, R., Erdman Jr., J.W., 2017. Increased dietary and circulating lycopene are associated with reduced prostate cancer risk: a systematic review and meta-analysis. Prostate Cancer Prostatic Dis. 20, 361–377.
Sahni, S., Hannan, M.T., Blumberg, J., Cupples, L.A., Kiel, D.P., Tucker, K.L., 2009. Protective effect of total carotenoid and lycopene intake on the risk of hip fracture: a 17-year follow-up from the Framingham Osteoporosis Study. J. Bone Miner. Res. 24, 1086–1094.
Tanumihardjo, S.A., Binkley, N., 2013. Carotenoids and bone health. In: Tanumihardjo, S.A. (Ed.), Carotenoids and Human Health. Springer Science, New York, pp. 237–245.
Tanumihardjo, S.A., Howe, J.A., 2005. Twice the amount of α-carotene isolated from carrots is as effective as β-carotene in maintaining the vitamin A status of Mongolian gerbils. J. Nutr. 135, 2622–2626.
Tanumihardjo, S.A., Palacios, N., Pixley, K.V., 2010. Provitamin A carotenoid bioavailability: what really matters. Int. J. Vitam. Nutr. Res. 80 (4–5), 336–350.
Tanumihardjo, S.A., Gannon, B.M., Kaliwile, C., Chileshe, J., 2015. Hypercarotenodermia in Zambia: which children turned orange during mango season? Eur. J. Clin. Nutr. 69, 1346–1349.
The Alpha-Tocopherol Beta Carotene Cancer Prevention Study Group, 1994. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. N. Engl. J. Med. 330, 1029–1035.
U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2020. Dietary Guidelines for Americans, 2020–2025, ninth ed. DietaryGuidelines.gov. Available at.
Yamaguchi, M., 2012. Role of carotenoid β-cryptoxanthin in bone homeostasis. J. Biomed. Sci. 19 (1), 36. https://doi.org/10.1186/1423-0127-19-36.
Zhang, S., Hunter, D.J., Forman, M.R., Rosner, B.A., Speizer, F.E., Colditz, G.A., Manson, J.E., Hankinson, S.E., Willett, W.C., 1999. Dietary carotenoids and vitamins A, C, and E and risk of breast cancer. J. Natl. Cancer Inst. 91 (6), 547–556.

Choline and phosphatidylcholine

Kevin C. Klatt[a,b], [a] Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX, United States; and [b] Center for Precision Environmental Health, Baylor College of Medicine, Houston, TX, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of M.D. Niculescu, Choline and Phosphatidylcholine, In Encyclopedia of Human Nutrition (Third Edition), edited by Benjamin Caballero, Academic Press, 2013, https://doi.org/10.1016/B978-0-12-375083-9.00053-2.

Introduction 162
Physiology and metabolism 164
Intestinal absorption and handling 164
Transport and metabolism 164
Excretion 165
Dietary recommendations and choline adequacy 165
Dietary reference intakes 165
General population 166
Clinical populations 166
Assessment of choline intake and status 167
Emerging areas of investigation 168
Pregnancy 168
TMAO and cardiovascular disease 169
Neurocognitive health and aging 170
Nutritional genomics of choline 170
Nutrigenetics 170
Nutrigenomics 170
Conclusion 171
References 171

Key points

- Understand the basis for establishing nutrient reference values for choline and its classification as an essential nutrient.
- Identify the key metabolic fates of choline and the physiological functions of these metabolites.
- Appreciate global differences in nutrient reference values for choline and relevant food sources of choline.
- Comprehend challenges in understanding choline's relationship to health outside of choline restriction, including its biochemical heterogeneity in the diet, heterogeneity in the endogenous capacity to synthesize choline, and reliance on self-reported dietary intakes in prospective cohorts.
- Distinguish different aspects of nutritional genomics and choline, including genetic factors that influence choline requirements and choline's ability to influence gene expression.

Introduction

Choline (IUPAC name 2-hydroxyethyl(trimethyl)azanium hydroxide) is a water-soluble, quaternary saturated amine found in foods of animal and plant origin (Fig. 1). Choline can be found in the diet as a composite of 6 metabolites that contain a choline moiety and contribute to total dietary choline intakes: free choline, phosphocholine (phoCho), glycerophosphocholine (GPC), phosphatidylcholine (PC) and sphingomyelin (SM). Choline natively found in foods, as well as added in the form of various lecithin emulsifiers (commonly derived from egg and soy), is predominantly PC; one exception to this

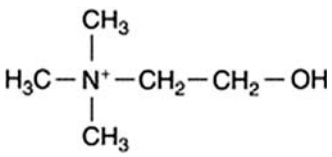


Fig. 1 Secondary chemical structure of choline molecule.

observation is during infancy, whereby breastmilk and some infant formulas provide substantial amounts of water-soluble choline-containing compounds (Holmes-McNary et al., 1996; Davenport et al., 2015; Shunova et al., 2020). Dietary supplements containing choline are typically in the form of PC, as water-soluble choline salts (e.g., choline-chloride, bitartrate), or as CDP-choline (citicoline).

The requirement for a source of choline in the diet was established for numerous animal species throughout the 20th century, with the earliest evidence indicating the nutritional importance of choline in preventing fat accumulation in the liver (Best and Huntsman, 1935). The buildup of triacylglycerols in the liver is a common phenotype across animal species when subjected to choline deficient diets (National Research Council (US) Subcommittee on Laboratory Animal Nutrition, 1995), including humans (Institute of Medicine, National Academy of Sciences (1998)), and is linked to the requirement for choline to produce PC and facilitate hepatic lipoprotein export. A series of dietary depletion-repletion studies in men and women across the reproductive lifecycle (i.e., pre- and post-menopausal) have demonstrated that choline restriction (<50 mg/day), in most participants, manifests in elevated circulating markers of liver (alanine amino transferase, ALT; aspartate aminotransferase, AST) and muscle (creatine phosphokinase, CPK) dysfunction/damage, an effect that is reversible by choline repletion (Zeisel et al., 1991; Niculescu et al., 2006; Fischer et al., 2007, 2010a,b). The requirement for dietary choline is inherently influenced by the hepatic capacity to synthesize a choline moiety de novo in the form of phosphatidylcholine, generated by the triple methylation of phosphatidylethanolamine catalyzed by the enzyme, phosphatidylcholine *N*-methyltransferase (PEMT) (Vance, 2013). The PEMT pathway relies on methyl donors, contributed by choline itself, or other dietary compounds, including betaine, methionine, and folates/vitamin B12. PEMT activity, discussed further below, is also regulated by estrogen availability (Resseguie et al., 2007), with men and postmenopausal women who are not utilizing estrogen replacement therapy being most sensitive to increased markers of organ dysfunction when placed on choline deficient diets. Thus, factors influencing PEMT activity, as well as the availability of other dietary one carbon metabolism nutrients, influence the de novo capacity to synthesize choline and subsequent dietary choline needs.

The requirement for choline relates to its incorporation into 3 primary metabolite pools (**Fig. 2**): **(1) phosphatidylcholines** (Kennedy, 1954; Bremer and Greenberg, 1960), that play a critical role in membrane structure and integrity, provide a source of the potent secondary messenger, diacylglycerol, facilitate lipoprotein assembly and export, serve as critical pools of polyunsaturated fatty acids and have recently been described as nuclear receptor ligands; **(2) acetylcholine** (Dale, 1914; Cohen and Wurtman, 1976), a neurotransmitter capable of binding nicotinic and muscarinic receptors in both the neuronal and non-neuronal cholinergic system, linking choline supply to cholinergic regulation of learning, memory, mood, skeletal and cardiac muscle function, and respiration; **(3) betaine** (Mann and Quastel, 1937), a major methyl donor capable of regenerating methionine from homocysteine and contributing to the universal methyl donor pool of S-adenosyl methionine (SAM); betaine is also an important osmolyte in the kidney and is notably present in the diet in high (mg) quantities. In addition to these classical metabolite pools, it has long been known that choline also contributes to trimethylamine (TMA), and the trimethylamine N-oxide (TMAO) (Simenhoff et al., 1976), which has received considerable attention for its relationship to a wide variety of chronic diseases and speculation about choline's influence on circulating TMAO concentrations and disease (discussed further below). It is evident that choline's entry into such diverse metabolites and their associated metabolic pathways provide underlying biological plausibility through which choline intakes may exert its influence on health and disease across the lifestyle.

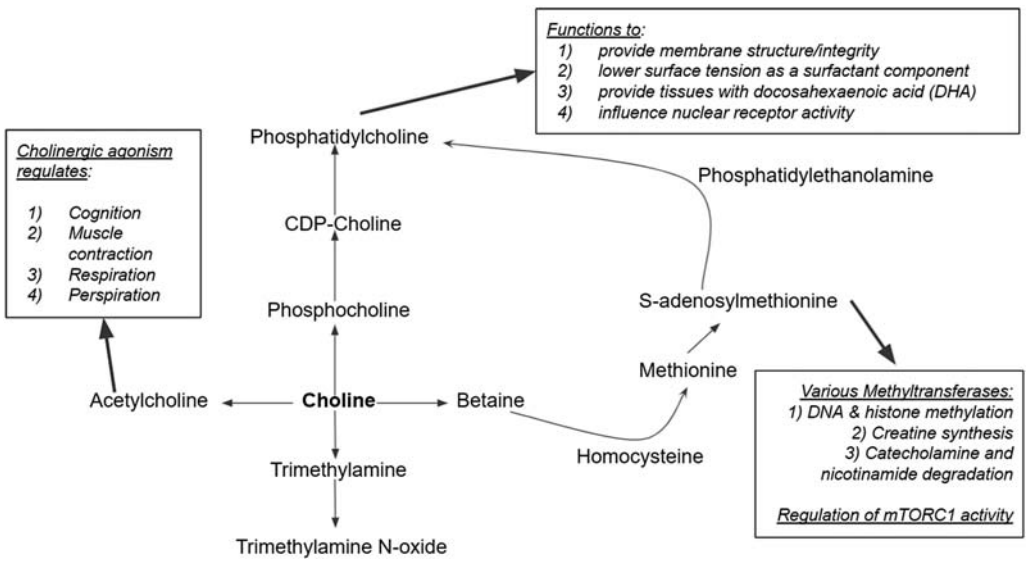


Fig. 2 Simplified schematic of choline metabolism and functions of its metabolites.

Physiology and metabolism

Intestinal absorption and handling

Few if any gastric factors have been found that influence choline absorption and choline appears relatively stable in the stomach. Upon entry into the small intestine, lipid and water-soluble choline are subjected to differential metabolism dependent upon their chemical structure. PCs must be hydrolyzed by the actions of phospholipase A2 to a free fatty acid and lysophosphatidylcholine (LPC) prior to uptake by the enterocyte (Fox et al., 1979) or further metabolized to glycerophosphocholine and an additional free fatty acid. LPC can be reacylated in the enterocyte to regenerate PC and/or incorporated into chylomicrons (Hui, 2016), or further hydrolyzed and degraded into a free fatty acid, glycerophosphate and free choline, with the latter capable of entering the portal vein as a source of choline for use by the liver or other peripheral tissues.

Water-soluble forms of choline are absorbed and/or degraded to free choline in the intestinal environment and taken up by the enterocyte. As a cation, choline cannot readily cross the cell membrane lipid bilayer via passive diffusion and requires specific transporters at the plasma membrane. In the intestine, evidence indicates that choline is efficiently absorbed in the jejunum and ileum via a substrate-specific and sodium-independent carrier-mediated transport system (Saitoh et al., 1992; Kamath et al., 2003); this transporter is likely a member of the 5-member, plasma and mitochondrial membrane-associated SLC44 family, of which SLC44A2 and A4 have been detected in the small intestine (Traiffort et al., 2013), while SLC44A1 appears localized to the large intestine (Wang et al., 2020).

Few investigations employing isotopic techniques have been undertaken to determine the relative bioavailability of different choline forms in animals or humans. Studies in rodents utilizing radioactive CDP-choline demonstrate near 100% bioavailability after being hydrolyzed to free choline in the intestine (Galletti et al., 1985). Studies employing chick growth assays comparing choline chloride relative to PC natively in soybean, corn, or peanut meal suggest significant effects of the plant food matrix on relative PC bioavailability, with soy and peanut exhibiting >70% bioavailability and canola exhibiting lower estimates (<35%), despite canola meal containing significantly higher total choline (Emmert and Baker, 1997). Heat treatment did not significantly influence growth and inclusion of meals alongside choline chloride treatment did not influence growth promotion. Unknown factors in the food matrix or potentially food particle microstructure might influence PC bioavailability.

The availability of choline is also limited by microbial leaching in the oral cavity and throughout the gastrointestinal tract (Chao and Zeisel, 1990). A bacterial gene cluster (choline-utilization cluster or *Cut*) encoding a glycyl radical enzyme with choline trimethylamine-lyase activity (*CutC*) and a glycyl radical-activating protein (*CutD*) are essential for microbiota metabolism of choline to TMA (Craciun and Balskus, 2012), and ultimately, leaching of luminal choline from the host. Colonization of gnotobiotic mice with the TMA-producing species *C. sporogenes* resulted in significant depletions of fecal choline, reductions in serum choline, elevated cecal TMA and increased serum TMAO, indicative of significant utilization of luminal choline (a composite of diet-derived choline compounds and biliary PC) (Romano et al., 2015). The relative loss of a choline dose to microbial metabolism likely depends on its form, as water-soluble choline compounds result in greater TMA formation than PC (Zeisel et al., 1983), consistent with PC digestion resulting in limited free choline availability to the microbiota. The high bioavailability of water-soluble choline compounds in the small intestine, coupled to their strong TMAO increasing capacity, has led to preliminary investigation indicating that the small intestinal microbiota, as well as the large intestinal microbiota, contributes to TMAO (Stremmel et al., 2017). This poses a challenge for identifying how gut microbial composition influences TMAO formation and quantifying its contribution to net choline losses, given that the small bowel microbiome is markedly different from that in stool (Leite et al., 2020), the primary sample matrix available for microbiome assessment in human population studies.

Transport and metabolism

Following absorption and intracellular metabolism, free choline enters the portal vein and undergoes primary hepatic metabolism. Alternatively, lipid-soluble choline containing moieties (PC, LPC) facilitate chylomicron formation and export into the lymphatic system. Ultimately, free choline enters the aqueous phase of plasma, whereas phosphorylated choline compounds are found as components of lipoproteins. Investigations into the tissue-level metabolic fate of different forms of dietary choline in postnatal rat pups employing radioactively labeled (^{14}C) choline chloride, phosphocholine, glycerophosphocholine and PC have reveal significantly different distribution and kinetic profiles (Cheng et al., 1996). The 3 water-soluble choline metabolites appears rapidly and reaches peak levels between 1 and 5 h in the blood and liver whereas PC takes 5–8 h to appear in blood and remains elevated in blood for at least 24 h.

Tissues uptake phosphorylated choline forms via lipoprotein mediated mechanisms; additionally, LPC can transport bound to albumin, for which a transporter of the major facilitator superfamily, Mfsd2a, has recently been characterized as a sodium-dependent transporter (Nguyen et al., 2014), with high expression levels in tissues with barrier functions (e.g., blood-brain barrier, placenta, mammary gland). Tissue uptake of free choline is facilitated by several transporters (Lockman and Allen, 2002), with a major role for the aforementioned, SLC44 family, of which SLC44A1 has received been most well-characterized. Cholinergic neurons additionally express a high affinity, sodium-dependent choline transporter, CHT1 (encoded by *SLC5A7*) at the presynaptic nerve terminal; the activity of this transport is typically the rate-limiting step for acetylcholine synthesis. A third, low-affinity, sodium-independent transport system is nearly ubiquitously expressed across tissues, facilitated by organic ion transporters (encoded by *SLC22*), and contributes to choline supply for phospholipid synthesis (Michel et al., 2006).

Following uptake, all nucleated cells possess the capacity to produce PC from choline via the CDP-choline pathway (Fagone and Jackowski, 2013). Free choline is rapidly phosphorylated within the cytoplasm of the cell, a reaction catalyzed by choline kinase and then activated to CDP-choline via CTP:phosphocholine cytidylyltransferase, the second and rate-limiting step of the CDP-choline pathway typically localized to the nucleus. CDP-choline subsequently undergoes condensation with diacylglyercol via the enzyme choline phosphotransferase to form a PC species (Henneberry et al., 2002). Synthesized PC has multiple metabolic fates depending on the tissue; it can be exported as PC, degraded by phospholipases to form LPC and/or free choline, exported into bile, used to produce surfactant (lung) or donate its phosphocholine group to ceramide to facilitate sphingomyelin (SM) production.

In the hepatocyte and kidney, choline uptake also results in its oxidation to betaine via the activity of choline dehydrogenase (encoded by *CHDH*), first producing betaine aldehyde, and subsequently betaine via the activity of betaine aldehyde dehydrogenase (Zhang et al., 1992). The hepatocyte and kidney are also major sites of expression for betaine homocysteine methyltransferase (encoded by *BHMT*), an enzyme that catalyzes the transfer of a methyl group from betaine to homocysteine, producing methionine and dimethylglycine (DMG) (Obeid, 2013). This pathway serves as an alternate pathway for methionine regeneration from homocysteine via methionine synthase activity, an enzyme that utilizes 5-methyltetrahydrofolate and vitamin B12 as cofactors and links choline intakes and metabolism to folate and vitamin B12. Methionine can subsequently be activated to S-adenosylmethionine (SAM) and used by methyltransferases for a wide-variety of reactions, including placement of DNA and histone methylation (epigenetic marks), and synthesis of neurotransmitters, hormones creatine and phospholipids. In addition to serving as a source of methyl donors, SAM was recently shown to bind a previously uncharacterized protein, now referred to as SAMTOR; SAM binding to SAMTOR disrupts SAMTOR's interactions with GATOR1, a major direct inhibitor of the mammalian target of rapamycin (mTOR) complex 1, thus linking SAM availability to the regulation of cell growth and nutrient sensing (Gu et al., 2017). The other product of BHMT activity, DMG, can be further metabolized to sarcosine and glycine, and its methyl groups utilized in folate-mediated one carbon metabolism.

The hepatocyte expresses, nearly exclusively, the phosphatidylethanolamine N-methyltransferase (PEMT) pathway of PC synthesis (Vance, 2013), a prominent consumer of methyl groups from SAM. Thus, choline can produce PC through the CDP-choline pathway directly or produce PC indirectly through its oxidation to betaine and donation of methyl groups to SAM, ultimately used by PEMT. This redundant pathway of PC synthesis in hepatocytes provides not only a mechanism to produce a choline moiety de novo during times of dietary choline inadequacy, but also produces a PC species that is enriched with the omega 3 fatty acid, docosahexaenoic acid (DHA), (DeLong et al., 1999; Watkins et al., 2003). In addition to regulation by SAM supply and SAH concentrations, PEMT is also regulated transcriptionally by estrogen (Resseguie et al., 2007, 2011), a major determinant of its activity, and contributor to reduced sensitivity of premenopausal women and postmenopausal women on hormone replacement therapy to organ dysfunction when placed on low choline diets.

In cholinergic neurons, as well as some non-neuronal tissues, a major metabolic fate of choline is the synthesis of acetylcholine (Beckmann and Lips, 2013). In neurons, choline acetyltransferase catalyzes acetylcholine synthesis from choline and acetyl-CoA. In the extracellular space, acetylcholine can bind to both the ligand-gated cation channel nicotinic receptors, and the G-protein coupled muscarinic receptors. In non-neuronal tissues, including cell types of the integumentary, respiratory, digestive, immune and reproductive systems, acetylcholine can be synthesized by choline acetyltransferase, as well as carnitine acetyltransferase activities (White and Scates, 1990). Acetylcholine degradation to yield free choline and acetate is predominantly achieved via acetylcholinesterase, whereby choline can be recycled for metabolic use.

Choline that is metabolized by the enteric microbiota to TMA is passively absorbed into portal circulation and metabolized by flavin monooxygenase 3 (FMO3) in the hepatocyte, yielding TMAO (Al-Waiz et al., 1987a,b; Lang et al., 1998). TMAO serves general roles as an osmolyte and protein stabilizer in physiological systems, though its functional role in humans remains less clear.

Excretion

Excretion of choline occurs as either choline or one of its primary metabolites. Dietary choline forms can be found in the feces, as well as TMA and TMAO. Feces also contains choline metabolites derived from enterocyte sloughing and PC-derived from bile that is not reabsorbed during enterohepatic recycling; in mice, the latter is a major route for PC loss in the liver, and genetic loss of the multiple drug-resistant protein 2 (MDR2 in mice; MDR3 in humans), responsible for exporting PC into bile, protects *PEMT* knockout mice fed choline deficient diets from rapid liver failure (Igolnikov and Green, 2006). Small amounts of choline and its metabolites are lost in the urine, through skin and hair loss, and from exhalation of formate.

Dietary recommendations and choline adequacy

Dietary reference intakes

In 1998, the United States' Food and Nutrition Board of the Institute of Medicine (now National Academies of Medicine (NAM)) established "Adequate Intake" (AI) values for choline (DRI Dietary Reference Intakes choline - Books - NCBI, no date). AI values are typically determined when dose-response relationships are unclear and an established average requirement (EAR) cannot be determined; in the case of choline, the use of depletion-repletion studies, repleting with a single dose of 7 mg choline/kg body weight, led to limited confidence in the intake of choline needed to prevent markers of hepatic organ dysfunction and reliance on this value as

Table 1 Nutrient reference values for dietary choline intakes.

EFSA adequate intakes (2016)		*NAM adequate intakes (1998)*	
Age/life-stage	*mg choline/day*	*Age/life-stage*	*mg choline/day*
7–11 months	160	7–12 months	150
1–3 years	140	1–3 years	200
4–6 years	170	4–8 years	250
7–10 years	250	9–13 years	375
11–14 years	340	14–18 years	M: 550 F: 400
15–17 years	400		
Adults	400	Adults	M: 550 F: 425
Pregnancy	480	Pregnancy	450
Lactation	520	Lactation	550

the AI. For the adult, the choline AI is 550 mg/d for men, and 425 mg/day for non-pregnant women; for pregnant and lactating women, a factorial approach was used to estimate the amount of choline deposited in fetal tissues and secreted in human milk to adjust the AI to 450 mg/d and 550 mg/d, respectively. All other age group DRIs were derived from body weight adjustment methods and/or consideration of the amount of choline in human milk. The upper limit (UL) for choline for adults is 3500 mg/day, a value linked to reduced likelihood of fishy body odor and hypotension, after consideration of an additional safety factor. In 2016, the European Food Safety Authority (EFSA) updated the Dietary Reference Values (DRVs) for choline following a request from the European Commission (European Food Safety Authority, 2016). The EFSA panel relied on similar data, with some additional feeding and supplementation trials, to derive the DRVs. For all adults, the AI was established for all adults as 400 mg/day, based on the amount of choline needed to replete 70% of individuals who demonstrated organ dysfunction in depletion-repletion studies. For pregnancy and lactation, the AI values were determined to be 480 mg/day and 520 mg/day respectively. A summary of EFSA and NAM nutrient reference values can be found in **Table 1**. Choline does not yet have a DRI value for the newly established Chronic Disease Risk Reduction (CDRR) value from the NAM. Dose-response data and assessment of hard endpoints continue to limit establishment of stronger DRIs with outcomes beyond preventing organ dysfunction as seen in depletion-repletion studies (**Table 1**).

General population

Despite its name, the establishment of an Adequate Intake value over an EAR for choline precludes assessment of the proportion of the population at risk of inadequacy. Analysis of the National Health and Nutrition Examination Survey in the United States indicates that few individuals >9 years of age consume the respective AI for their age grouping and life-stage (Wallace and Fulgoni, 2016, 2017). Generally, males consume more choline than females, though the predominant determinant of high choline intakes and likelihood of achieving the AI relates to egg consumption; eggs contain around 235 mg choline per 100 g (143 kcal) serving, making them a rich source of choline (predominantly PC) in the diet. Other protein foods (meat, poultry, and seafood) also contribute meaningfully (~100 mg/d) to choline intakes. Foods less commonly consumed in the general population, such as beef liver, shrimp, oysters, salmon, cauliflower and asparagus, are relatively rich sources of choline on a per 100 g and per kilocalorie basis. Common food sources of choline and betaine and their nutrient content are shown in **Fig. 3**. No subpopulation groups achieve the UL for choline, a value that is extremely challenging to achieve without significant supplement utilization. In infants, choline contents of formulas vary substantially, with few formulas exhibiting choline compositions similar to human breastmilk (utilizing PC or free choline, as opposed to glycerophosphocholine or phosphocholine) (Shunova et al., 2020).

Clinical populations

Choline is not known to exhibit significant drug-nutrient interactions; needs may be higher in individuals with long-term methotrexate use but evidence from non-rodent models is limited (Varela-Moreiras et al., 1995; Hardwick et al., 2014). No dietary reference values have been clearly established by authoritative bodies focused on guideline setting in the clinical population. However, some clinical populations may exhibit altered choline intakes, exhibit low intakes and/or increased losses of choline. Patients with genetic mutations in FMO3 exhibit trimethylaminuria, resulting from a buildup of trimethylamine and inability to oxidize TMA to TMAO; patients limit excessive dietary choline intakes to limit TMA buildup and the associated fishy body odor (Schmidt and Leroux, 2020).

Choline losses are likely greater in clinical populations characterized by impaired intestinal homeostasis, including intestinal failure, exocrine pancreas insufficiency, and small intestinal bacterial overgrowth, though rigorous investigation is generally lacking. Studies of patients with cystic fibrosis and exocrine pancreas insufficiency demonstrate significant choline losses (as PC and

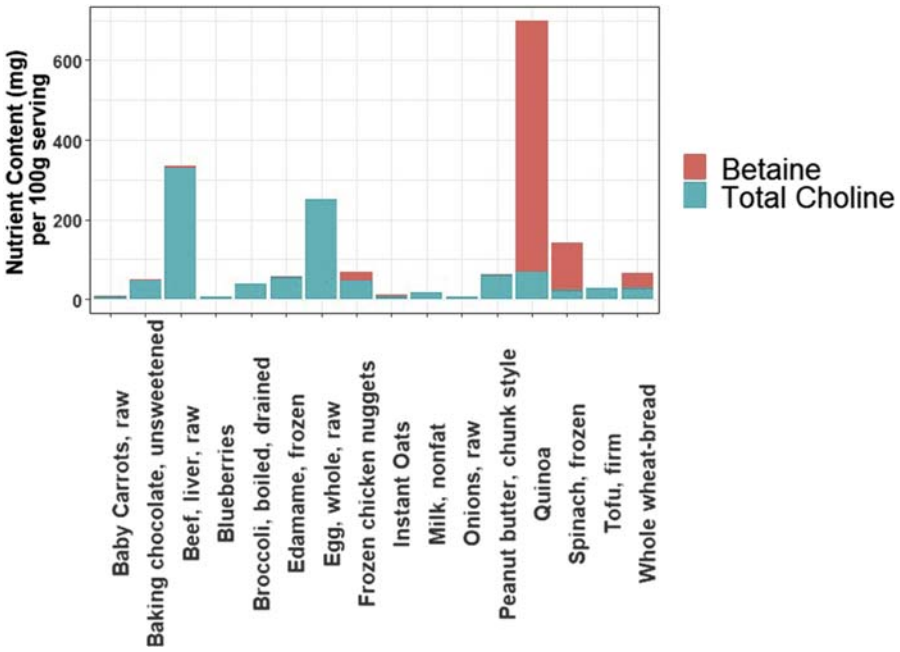


Fig. 3 Total choline and betaine content in selected foods, raw or cooked. The source for the data was the USDA Database for the Choline Content of Common Foods (Release Two) available at http://www.ars.usda.gov.

LPC, primarily) in the stool, and supplementation improves low plasma choline/PC, and improve elevated homocysteine levels (Chen et al., 2005; Innis et al., 2007; Bernhard, 2021). Patients with impaired intestinal function regularly receive parenteral nutrition (PN); parenteral choline is not readily available, limited by its stability in solution long-term. Patients on long-term PN often exhibit liver disease associated with PN or underlying causes of PN (i.e., intestinal failure) that may result from choline deficiency, though other dietary and metabolic factors also likely contribute to the observed broad liver dysfunction (Burrin et al., 2014). Choline supplementation to cohorts of adult patients receiving PN appears to improve hepatic steatosis, as well as memory impairments (Buchman, 2009); particular concern over choline nutriture in the enterally and PN-fed preterm infant has emerged, given the importance of choline for brain development, reduced exposure to high *in utero* choline concentrations and recently observed low PEMT activity and thus, capacity to generate choline de novo (Goss et al., 2020). Calls to action to develop a readily available source of choline for adult and pediatric PN formulations have been increasingly made (Vanek et al., 2012, 2015). A commercial choline chloride product has recently completed phase 2 studies in intestinal failure-associated liver disease and received Fast Track designation from the United States' Food and Drug Administration.

Plasma choline is altered in several other patient populations, such as burns, sepsis and elective abdominal surgery, as well as being present in high quantities in renal dialyzate; however, few cohort studies have assessed functional consequences of these alterations and/or controlled trials testing the impacts of choline supplementation.

Assessment of choline intake and status

Objective biomarkers of choline intake and status would be beneficial in assessing dietary adequacy of both populations and individuals, as well as serving as an adequate exposure variable in the assessment of choline's relationship to health outcomes in prospective cohort studies. Unfortunately, such biomarkers are broadly lacking.

Hepatic choline levels have been suggested to be the best marker of choline status, consistent with low choline availability in the hepatocyte resulting in hepatic triacylglycerol accumulation; whether other tissue choline content, or a related metabolite, may serve as a better marker of choline status, remains less well-characterized, though advances in proton magnetic resonance spectroscopy (MRS) allow for the in vivo quantification of free choline, glycerophosphocholine and phosphocholine in tissues (Mazzetti et al., 2013). A biomarker of choline status should clearly track with dose-response changes in choline intake and be strongly associated with health and disease. Plasma choline and its metabolites are frequently measured as markers of choline status, typically by liquid chromatography/electrospray ionization, stable isotope dilution (tandem) mass spectrometry, though other methods to measure choline, such as colorimetric kits, are available. Fasting plasma choline levels do not always track with moderate changes in choline intake (Veenema et al., 2008; Abratte et al., 2009), and rarely drop below 4.5 μM in individuals consuming deficient diets; nevertheless, fasting plasma choline levels, particularly in repleted individuals previously consuming low intakes and in those consuming high supplemental intakes, can be responsive to choline intakes. Plasma choline levels should be considered as not only an

indicator of dietary intakes but a composite of its relative degree of absorption, endogenous synthesis, release and metabolism of choline stores, relative partitioning toward other metabolites (e.g., betaine), and losses in urine.

Recent attempts to relate varying dietary choline intakes (~25, 50 and 100% of the AI) to hepatic choline contents, as measured by MRS, have not demonstrated a clear linear correlation, though technical challenges may have influenced this finding (Horita et al., 2021). Plasma betaine levels showed clear correlation with dietary choline intakes in this study; however, variation in choline intakes were achieved through bread rolls fortified with choline chloride, a readily oxidizable form of choline, and may not be a clear indicator of dietary choline adequacy in the general population, which consumes primarily PC. While discovery work examining gene expression and metabolomic signatures of choline deficiency work has yielded promise (Niculescu et al., 2007; Sha et al., 2010), stronger evidence is needed to find readily measurable correlates of varying choline intakes, considering both water- and lipid-soluble intakes forms, that ideally track with tissue choline contents.

Given the lack of readily measurable indicators of choline status and intake adequacy, assessment at the population and individual level relies heavily on self-reported dietary intakes. Underreporting of dietary intakes likely results in choline underreporting; comparisons between 3-day dietary records and known choline intakes in a metabolic ward suggest that absolute choline intakes are significantly underreported by self-reported dietary measures, but that choline intakes per kilocalorie are not (Fischer et al., 2005). While not supported by this study, additional concern exists that, in free living individuals consuming self-selected diets with significant proportions of processed foods, choline intakes will be significantly influenced by the lecithin content of foods, a factor not captured by most recall methodologies or in nutrient composition databases for many commercially available foods. Concerns about inadequate choline intakes may be complemented with additional functional tests, including liver functional tests and CPK, to assess the potential for dietary inadequacy; notably, however, dietary choline depletion-repletion studies suggest that choline's effect on many readily available liver function tests is subtle and may not necessarily fall outside of "normal" clinical ranges.

Emerging areas of investigation

While current nutrient reference values (NRVs) continue to rely on choline depletion-repletion studies to assess dietary requirements for choline, growing bodies of evidence link choline to health outcomes beyond preservation of normal liver and muscle dysfunction.

Pregnancy

NRVs from NAM and EFSA for pregnancy are ultimately derived from studies conducted in non-pregnant participants with adjustments based on assumptions about the amount of choline needed to support fetal growth. To date, no NRVs are based on functional outcomes related to alterations in fetal development and offspring health and/or reduced pregnancy complications. Pregnancy is a physiological state that stresses choline metabolism, with evidence from animal models demonstrating depletion of maternal hepatic choline contents, despite consuming dietary intakes adequate to prevent abject organ dysfunction; such a depletion is pronounced as gestation progresses as the placenta increasingly transfers choline to the fetal compartment (Sweiry et al., 1986; Ilcol et al., 2002); fetal cord plasma exhibits a choline concentration 5–10X higher than maternal plasma (Zeisel et al., 1980; Yan et al., 2012). Feeding of choline deficient diets to dams further stresses choline needs and has broad consequences for fetal neurodevelopment, impairing neurogenesis and angiogenesis in the hippocampus, decreasing cell proliferation and increasing apoptosis, ultimately impairing several domains of cognition in adult animals (Meck et al., 1988; Albright et al., 1999; Cheng et al., 2008; Blusztajn et al., 2017).

A large body of literature in rodents demonstrate that maternal choline supplementation throughout gestation, beyond usual chow intakes, alters the development of fetal septohippocampal circuitry and improves postnatal spatial learning and memory, as well as attention, phenotypes in the offspring (Strupp et al., 2016; Blusztajn et al., 2017). Doses used in the rodent literature are typically 4X the standard chow diet and achieved via supplementation with water-soluble choline salts. In humans, a small body of literature during health pregnancy has investigated choline supplementation's impact on infant cognitive development (Cheatham et al., 2012; Ross et al., 2013; Caudill et al., 2018). Of the 3 studies, 2 utilized phosphatidylcholine (750 mg and 900 mg/d) in free-living pregnant women and one controlled feeding trial (Caudill et al., 2018) compared 930 vs 480 mg choline/d, with the differential achieved via choline chloride. One PC trial and the one choline salt trial demonstrated cognitive improvements, whereas the largest trial, utilizing PC, demonstrated null effects. Trials are diverse with regard to design and methods used to assessment infant cognitive development, and future trials, with dose-response (>2 arm) designs and consideration of dietary form are required to better understand choline's link to cognition and refine NRVs. Evidence for prenatal choline supplementation in alcohol-exposed pregnancies in both preclinical models and human trials has also demonstrated promise to buffer against alcohol's toxic impacts (Zeisel, 2011; Coles et al., 2015; Jacobson et al., 2018; Sawant et al., 2019).

Choline intakes in rodent models and in some epidemiological investigations are also linked to the risk of birth defects (Shaw et al., 2004, 2009; Chan et al., 2010); however, no trials have been designed to assess these relationships in humans. To date, significant amounts of choline are still absent from commercial prenatal vitamins, and insufficient physical space in existing prenatal vitamin supplements precludes fortification with significant quantities of choline without increasing supplement size.

Limited investigation has linked choline intakes to reduced risk of maternal pregnancy complications. A recent post-hoc analysis of the aforementioned Caudill et al. feeding trial observed that the higher choline intake arm (930 mg/d), relative to the lower intake arm (480 mg/d) resulted in lower placental and circulating levels of the vascular endothelial growth factor (VEGF) decoy receptor, soluble fms-like tyrosine kinase 1 (sFLT1), a finding identified by microarray analysis of placental tissue. sFLT1 is an anti-angiogenic factor secreted by the placenta and has been causally linked to pre-eclampsia (PE) (Ives et al., 2020). While many dietary factors have been investigated for their link to PE risk, future observational and intervention investigation are needed to assess the potential for dietary choline supplementation to reduce PE risk.

TMAO and cardiovascular disease

Choline was originally suggested to be potentially cardioprotective, due to its role as a major dietary methyl donor and maintaining homocysteine levels in a normal range (da Costa et al., 2005; Olthof et al., 2005). However, choline's role as a dietary TMA(O) precursor has led to concern that choline intakes may promote cardiovascular disease. In prospective cohorts, TMAO exhibits a dose-response relationship, per 1 μmol/L increment, with major adverse cardiovascular events, making it an attractive candidate for therapeutic targeting (Heianza et al., 2017). Whether TMAO represents a causal factor in disease development has been questioned, as elevated TMAO may be a marker of other underlying factors such as a pathogenic gut microbiota, insulin resistance, an inducer of FMO3 activity (Miao et al., 2015), and/or altered kidney function (TMAO levels are much higher in advanced renal disease (Pelletier et al., 2019)); experimental evidence, however, suggests that TMAO is likely a causal mediator of disease risk. In mechanistic investigations, TMAO has been shown to impair reverse cholesterol transport, promote macrophage inflammation, reduced bile acid pool sizes, and increased platelet aggregation, ultimately resulting in increased foam cell formation, atherosclerotic plaque volume, and risk of thrombosis (Zeisel and Warrier, 2017). Cellular mediators of the response to TMAO are an active area of investigation; recent advancements have indicated that TMAO's impact on metabolic disease is mediated via its binding and activation of protein kinase R (PKR)—like endoplasmic reticulum kinase (PERK), a branch of the unfold protein response in the endoplasmic reticulum (Chen et al., 2019). Efforts are ongoing to test the potential of CutC/CutD inhibitors to reduce TMA production and assess its impact on cardiovascular events (Roberts et al., 2018).

Randomized controlled trials manipulating dietary choline and observing its effect on cardiovascular disease endpoint are unlikely, due to their impracticality, large financial investment, and inherent design challenges (e.g., reducing egg consumption, a major source of choline, also reduces dietary cholesterol, and may be replaced by foods with additional factors that influence CVD risk). Thus, investigations have been limited to rodent models of cardiovascular disease, and observational population studies with self-reported dietary choline intakes. In rodent models, findings are discordant for an effect of choline on atherosclerotic plaque volume. Indeed, the most common model, the ApoE knockout mouse, has been employed by several investigators feeding high (∼1% choline) vs low choline (∼0.1%) diets, with several investigators observed increased atherosclerosis, and others finding no effect (Tang et al., 2015; Zhu et al., 2016; Geng et al., 2018; Aldana-Hernández et al., 2020, 2021). No effect of choline has also been reported in the LDL receptor knockout mouse. While not all investigators have reported plasma TMAO levels, those that have demonstrate striking variation in achieved levels (1.5–200 μM), suggestive of different microbiota composition and capacity to generate TMA from TMAO underlying variability in response to choline supplementation; variation in microbiota composition across vivariums has been reported, and affects other research domains dependent upon microbiota composition (e.g., inflammatory bowel disease) (Reinoso Webb et al., 2018). Null findings have generally achieved lower TMAO levels.

The translatability of these animal models to free-living humans consuming habitual choline intakes is uncertain. In addition to species differences in atherosclerosis, it should be noted that dietary manipulations in mice have utilized high doses of water-soluble choline salts (more readily metabolized to TMA), typically achieving very high TMAO levels, challenging the translation of these findings to human populations consuming predominantly PC with generally lower TMAO levels. To complement such animal models, investigations into dietary choline intakes in human cohort studies have been undertaken, but are challenging to interpret, due to their correlation with the associated food matrix; major sources of choline in the human diet are eggs, red meat, and dairy products, each with their own nutrient matrices that may contribute to cardiovascular risk increase or reduction. Additionally, individuals who consume low choline foods/diets may consume other foods that modify cardiovascular risk, obscuring or amplifying the risks associated with choline. Systematic reviews of dietary choline and CVD do not support a clear risk (Meyer and Shea, 2017), though individual cohorts have supported an association. The lack of an effect in the general population consuming predominantly PC may not be entirely surprising, given the complexity of TMAO generation; while diet is a precursor to TMA, persistent and substantial elevations in TMAO levels might require a pathogenic gut microbiota composition, alterations in FMO3 activity (notably regulated by insulin), and/or impaired capacity of the kidney to clear TMAO to achieve persistently elevated TMAO concentrations associated with clearly pathogenic levels. To date, estimates of the variance in plasma TMAO levels contributed by diet remain relatively small, with dietary factors such as fish and red meat, explaining the most variation (Krüger et al., 2017). Future studies in higher risk populations are needed to determine the degree to which diet influences plasma TMAO levels and whether this may result in likely risk reduction while still achieving intakes adequate to prevent nutrient deficiency phenotypes. Ultimately, efforts to make dietary recommendations based on modifying TMAO levels must take into account that fish, including fatty fish, are major sources of dietary TMA(O) (Cho et al., 2017), and that other dietary factors, such as a yet to be identified factor in red meat, appears to influence renal handling of TMAO (Wang et al., 2019).

Neurocognitive health and aging

The relationship of dietary choline to neurocognitive health and disease has received some attention in the literature, though the evidence base is limited. Rationale for such investigations relate not only to acetylcholine synthesis and cholinergic agonism but to choline's role in SM production, a component of the myelin sheath, its role as a methyl donor and maintaining low homocysteine levels, and potential epigenetic effects (DNA & histone methylation). One investigation in the Framingham Offspring study has linked self-reported choline intakes to cognitive function and reduced white matter hyperintensity volume, following adjustment for several covariates (Poly et al., 2011). While evidence from long-term PN use results in neurocognitive abnormalities (as discussed above), evidence to support choline supplementation or increased dietary choline intakes on cognitive function in adults or in those with dementia is of low quality and lacking clear evidence (Higgins and Flicker, 2003; Leermakers et al., 2015), though individual acute studies demonstrate proof of principle neuromodulation by high dose choline supplementation (Naber et al., 2015). Dietary choline's relationship to age-related dementia, including Alzheimer's Disease, requires further investigation.

Nutritional genomics of choline

Efforts to advance our understanding of choline requirement and links between choline intake and broader health outcomes should consider nutrient-genome interactions. Indeed, a growing body of evidence indicates that common genetic variants influence the risk of deficiency when fed choline-deficient diets, while altering choline metabolism. Choline, through several direct and indirect mechanisms, also influences gene expression. Consideration of the nutrigenetics and nutrigenomics of choline may refine and inform investigations aiming to bring clarity to choline's links to health and disease.

Nutrigenetics

Common genetic variants throughout one carbon metabolism-related genes have been the subject of several post-hoc analyses from choline intervention studies. In the choline depletion-repletion literature, single nucleotide polymorphisms in choline and folate metabolizing enzymes predict the odds of developing organ dysfunction when placed on choline deficient (folate, b12 and methionine replete) diets (Kohlmeier et al., 2005; da Costa et al., 2006); odds ratio were substantial for several variants, including a 25-fold increased odds for carriers of the PEMT variant, rs12325817, a polymorphism in the promoter region linked to impaired upregulation by estrogen. Several variants, such as those in SLC44A1, were also more prevalent in individuals who experienced elevated markers of muscle damage relative to liver damage (da Costa et al., 2014). Several of the described variants, such as those in the *PEMT* locus, are in linkage disequilibrium, and the prevalence of SNPs across typical racial/ethnic groupings varies considerably, with individuals of African descent having lowest frequency of *PEMT* risk alleles and choline kinase (*CHKA*) protective alleles. The impact of these variants on choline requirements outside the context of choline deficiency studies remains underexplored, with limited investigation into their impacts on choline partitioning and flux and some evidence indicating SNPs in MTHFR and PEMT influence maternal diet-breastmilk choline metabolite concentrations (Fischer et al., 2010a,b; Ganz et al., 2016, 2017). While proof of concept for the field of nutrigenetics and the potential for individualized nutrition, studies examining genetic variants in a prospective manner and their interactions with more usual choline intakes are required to realize the full potential of incorporating genotype subgroups into dietary guidance. Consideration of effect modification by genetic variants prospectively in all study designs assessing dietary choline intakes and its relationship to health and disease will be critical for explaining heterogeneity in response in a statistically rigorous manner.

Nutrigenomics

Epigenetics refers to heritable marks on both DNA and on the histone proteins that DNA wraps around to form the structure of chromatin; methylation is a common mark on both DNA and histones, and thus, choline's role as a methyl donor has led to significant investigation into the epigenetic mechanisms by which it influences gene expression and in which tissues. However, it is notably challenging to attribute gene expression and broader phenotypic changes to its role as a methyl donor relative to changes downstream of its other metabolic roles (e.g., cholinergic signaling; PC synthesis), and few multi-omic, systems biology approaches have been employed in the context of feeding different levels of choline. Early investigations employed methyl deficient diets, of which choline depletion was a component, with an emphasis on identifying hypomethylated genes associated with the development of hepatocellular carcinomas, including aberrant regulation of *KRAS* and *c-myc* (Bhave et al., 1988; Tsujiuchi et al., 1999). Choline was also a component of the methyl donor cocktail fed to agouti mice, resulting in hypermethylation at the Avy locus and both coat color and body composition changes (Waterland and Jirtle, 2003). Additional studies in rodents selectively altering choline both through depletion and supplementation have demonstrated global and gene-specific alterations in DNA methylation and histone methylation in the liver and frontal cortex at embryonic day 17, corresponding to the end of the period of gestation (E11-17) that is critical for improved cognition in offspring (Davison et al., 2009). A recent investigation in humans who consumed 930 mg vs 480 mg choline/d throughout the third trimester of pregnancy demonstrated elevated promoter methylation in 2 cortisol regulating genes (*CRH*; *NR3C1*), reduced placental expression of CRH, and lower cord plasma cortisol in response to higher choline supplementation (Jiang et al., 2012). Dimethylated histone H3 at lysine residue 9 was also significantly elevated. The causal

relationships between choline intake, altered epigenetic marks and changes in gene expression require further interrogation by emerging systems biology approaches, including single-cell sequencing technologies.

Nutrients, such as Vitamins A and D, classically modify gene expression via binding of nuclear receptors (i.e., serving as ligands); nuclear receptors are a family of protein transcription factors that bind consensus sequences of DNA and influence transcription. There is emerging interest in choline's effects on gene expression, secondary to the finding that specific PC species are ligands for nuclear receptors of the 5A family (Li et al., 2005; Ortlund et al., 2005; Musille et al., 2016; Petruzzelli et al., 2016). One member of this family, Liver Receptor Homolog-1 (LRH-1) has been shown to exhibit transcriptional and phenotypic responses to specific PC supplementation, though conflicting evidence exists to support the CDP-choline or PEMT pathway as the primary producer of LRH-1 ligands and both may play roles in producing specific PC species that bind LRH-1 (Lee et al., 2011; Wagner et al., 2016; Choi et al., 2020). Further investigation is needed to determine the effects of varying choline intakes on nuclear receptor ligand production and function.

Choline influences the transcriptome through various, more indirect stress responses when choline intakes are limited, secondary to altered mitochondrial membrane integrity and reactive oxygen species leakage, increased consumption of folates and associated uracil misincorporation and DNA strand breaks, and stimulation of apoptotic pathways (Zeisel, 2012). Such effects likely influence cell signaling networks and underlie the increased risk of cancer from inadequate choline intakes seen in animal models, though observational data in human populations supporting such links remains of low quality (Sun et al., 2016).

Conclusion

Despite more than 3 decades since the seminal observation in humans that choline restriction compromises hepatic function in humans, many questions remain regarding dietary choline intakes relationship to health and disease. There is a great need to undertake dose-response feeding trials to establish an Estimated Average Requirement and Recommended Dietary Allowance for Choline, moving beyond the uncertain AI values. Evidence for increased choline requirements during pregnancy, significantly beyond current recommendations, to optimize neurocognitive health of the offspring continues to emerge, but more evidence is likely needed to see changes to DRIs and current obstetric practice. Choline serves as an important case study for all human nutrition as the field attempts to tackle "Precision Nutrition", given the complexity of measuring choline in the diet and assessing choline status, lack of clear surrogate intakes linking choline to health and disease outcomes outside of frank deficiency states, and the evidence pointing to significant nutrient-nutrient, nutrient-microbiota, and nutrient-host genome interactions.

References

Abratte, C.M., et al., 2009. Choline status is not a reliable indicator of moderate changes in dietary choline consumption in premenopausal women. J. Nutr. Biochem. 20 (1), 62–69. https://doi.org/10.1016/j.jnutbio.2007.12.002.

Albright, C.D., et al., 1999. Choline availability alters embryonic development of the hippocampus and septum in the rat. Dev. Brain Res. 113 (1), 13–20. https://doi.org/10.1016/S0165-3806(98)00183-7.

Aldana-Hernández, P., et al., 2020. Dietary choline or trimethylamine N-oxide supplementation does not influence atherosclerosis development in Ldlr−/− and Apoe−/− male mice. J. Nutr. 150 (2), 249–255. https://doi.org/10.1093/jn/nxz214.

Aldana-Hernández, P., et al., 2021. Dietary phosphatidylcholine supplementation reduces atherosclerosis in Ldlr−/− male mice2. J. Nutr. Biochem. 92, 108617. https://doi.org/10.1016/j.jnutbio.2021.108617.

Al-Waiz, M., et al., 1987a. Disclosure of the metabolic retroversion of trimethylamine N-oxide in humans: a pharmacogenetic approach. Clin. Pharmacol. Therapeut. 42 (6), 608–612. https://doi.org/10.1038/clpt.1987.207.

Al-Waiz, M., et al., 1987b. The metabolism of 14C-labelled trimethylamine and its N-oxide in man. Xenobiotica 17 (5), 551–558. https://doi.org/10.3109/00498258709043962.

Beckmann, J., Lips, K.S., 2013. The non-neuronal cholinergic system in health and disease. Pharmacology 92 (5–6), 286–302. https://doi.org/10.1159/000355835.

Bernhard, W., 2021. Choline in cystic fibrosis: relations to pancreas insufficiency, enterohepatic cycle, PEMT and intestinal microbiota. Eur. J. Nutr. 60 (4), 1737–1759. https://doi.org/10.1007/s00394-020-02358-2.

Best, C.H., Huntsman, M.E., 1935. The effect of choline on the liver fat of rats in various states of nutrition. J. Physiol. 83 (3), 255–274. https://doi.org/10.1113/jphysiol.1935.sp003227.

Bhave, M.R., Wilson, M.J., Poirier, L.A., 1988. c-H-ras and c-K-ras gene hypomethylation in the livers and hepatomas of rats fed methyl-deficient, amino acid-defined diets. Carcinogenesis 9 (3), 343–348. https://doi.org/10.1093/carcin/9.3.343.

Blusztajn, J.K., Slack, B.E., Mellott, T.J., 2017. Neuroprotective actions of dietary choline. Nutrients 9 (8), E815. https://doi.org/10.3390/nu9080815.

Bremer, J., Greenberg, D.M., 1960. Biosynthesis of choline in vitro. Biochim. Biophys. Acta 37, 173–175. https://doi.org/10.1016/0006-3002(60)90104-9.

Buchman, A.L., 2009. The addition of choline to parenteral nutrition. Gastroenterology 137 (5 Suppl. l), S119–S128. https://doi.org/10.1053/j.gastro.2009.08.010.

Burrin, D.G., et al., 2014. Impact of new-generation lipid emulsions on cellular mechanisms of parenteral nutrition-associated liver disease. Adv. Nutr. 5 (1), 82–91. https://doi.org/10.3945/an.113.004796.

Caudill, M.A., et al., 2018. Maternal choline supplementation during the third trimester of pregnancy improves infant information processing speed: a randomized, double-blind, controlled feeding study. Faseb. J. 32 (4), 2172–2180. https://doi.org/10.1096/fj.201700692RR.

Chan, J., et al., 2010. Low dietary choline and low dietary riboflavin during pregnancy influence reproductive outcomes and heart development in mice. Am. J. Clin. Nutr. 91 (4), 1035–1043. https://doi.org/10.3945/ajcn.2009.28754.

Chao, C.K., Zeisel, S.H., 1990. Formation of trimethylamine from dietary choline by *Streptococcus sanguis* I, which colonizes the mouth. J. Nutr. Biochem. 1 (2), 89–97. https://doi.org/10.1016/0955-2863(90)90055-p.

Cheatham, C.L., et al., 2012. Phosphatidylcholine supplementation in pregnant women consuming moderate-choline diets does not enhance infant cognitive function: a randomized, double-blind, placebo-controlled trial. Am. J. Clin. Nutr. 96 (6), 1465–1472. https://doi.org/10.3945/ajcn.112.037184.

Chen, A.H., et al., 2005. Phosphatidylcholine and lysophosphatidylcholine excretion is increased in children with cystic fibrosis and is associated with plasma homocysteine, S-adenosylhomocysteine, and S-adenosylmethionine. Am. J. Clin. Nutr. 81 (3), 686–691. https://doi.org/10.1093/ajcn/81.3.686.
Chen, S., et al., 2019. Trimethylamine N-Oxide binds and activates PERK to promote metabolic dysfunction. Cell. Metab. 30 (6), 1141–1151.e5. https://doi.org/10.1016/j.cmet.2019.08.021.
Cheng, W.-L., et al., 1996. Bioavailability of choline and choline esters from milk in rat pups. J. Nutr. Biochem. 7 (8), 457–464. https://doi.org/10.1016/0955-2863(96)00079-4.
Cheng, R.-K., et al., 2008. Prenatal choline supplementation alters the timing, emotion, and memory performance (TEMP) of adult male and female rats as indexed by differential reinforcement of low-rate schedule behavior. Learn. Mem. 15 (3), 153–162. https://doi.org/10.1101/lm.729408.
Cho, C.E., et al., 2017. Trimethylamine-N-oxide (TMAO) response to animal source foods varies among healthy young men and is influenced by their gut microbiota composition: a randomized controlled trial. Mol. Nutr. Food Res. 61 (1). https://doi.org/10.1002/mnfr.201600324.
Choi, S., et al., 2020. Methyl-sensing nuclear receptor liver receptor homolog-1 regulates mitochondrial function in mouse hepatocytes. Hepatology 71 (3), 1055–1069. https://doi.org/10.1002/hep.30884.
Cohen, E.L., Wurtman, R.J., 1976. Brain acetylcholine: control by dietary choline. Science 191 (4227), 561–562. https://doi.org/10.1126/science.1251187.
Coles, C.D., et al., 2015. Dose and timing of prenatal alcohol exposure and maternal nutritional supplements: developmental effects on 6-month-old infants. Matern. Child Health J. 19 (12), 2605–2614. https://doi.org/10.1007/s10995-015-1779-x.
Craciun, S., Balskus, E.P., 2012. Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. Proc. Natl. Acad. Sci. U. S. A. 109 (52), 21307–21312. https://doi.org/10.1073/pnas.1215689109.
da Costa, K.-A., et al., 2005. Choline deficiency in mice and humans is associated with increased plasma homocysteine concentration after a methionine load. Am. J. Clin. Nutr. 81 (2), 440–444. https://doi.org/10.1093/ajcn.81.2.440.
da Costa, K.-A., et al., 2006. Common genetic polymorphisms affect the human requirement for the nutrient choline. Faseb. J. 20 (9), 1336–1344. https://doi.org/10.1096/fj.06-5734com.
da Costa, K.-A., et al., 2014. Identification of new genetic polymorphisms that alter the dietary requirement for choline and vary in their distribution across ethnic and racial groups. Faseb. J. 28 (7), 2970–2978. https://doi.org/10.1096/fj.14-249557.
Dale, H.H., 1914. The action of certain esters and ethers of choline, and their relation to muscarine. J. Pharmacol. Exp. Therapeut. 6 (2), 147–190.
Davenport, C., et al., 2015. Choline intakes exceeding recommendations during human lactation improve breast milk choline content by increasing PEMT pathway metabolites. J. Nutr. Biochem. 26 (9), 903–911. https://doi.org/10.1016/j.jnutbio.2015.03.004.
Davison, J.M., et al., 2009. Gestational choline supply regulates methylation of histone H3, expression of histone methyltransferases G9a (Kmt1c) and Suv39h1 (Kmt1a), and DNA methylation of their genes in rat fetal liver and brain. J. Biol. Chem. 284 (4), 1982–1989. https://doi.org/10.1074/jbc.M807651200.
DeLong, C.J., et al., 1999. Molecular distinction of phosphatidylcholine synthesis between the CDP-choline pathway and phosphatidylethanolamine methylation pathway. J. Biol. Chem. 274 (42), 29683–29688. https://doi.org/10.1074/jbc.274.42.29683.
Dietary Reference Values for Choline | European Food Safety Authority, 2016. Available at. https://www.efsa.europa.eu/en/efsajournal/pub/4484.
DRI Dietary Reference Intakes Choline—Books—NCBI, (no date). Available at https://www.ncbi.nlm.nih.gov/books/NBK114310/?term=DRI%20Dietary%20Reference%20Intakes%20choline.
Emmert, J.L., Baker, D.H., 1997. A chick bioassay approach for determining the bioavailable choline concentration in normal and overheated soybean meal, canola meal and peanut meal. J. Nutr. 127 (5), 745–752. https://doi.org/10.1093/jn/127.5.745.
Fagone, P., Jackowski, S., 2013. Phosphatidylcholine and the CDP-choline cycle. Biochim. Biophys. Acta 1831 (3), 523–532. https://doi.org/10.1016/j.bbalip.2012.09.009.
Fischer, L.M., et al., 2005. Ad libitum choline intake in healthy individuals meets or exceeds the proposed adequate intake level. J. Nutr. 135 (4), 826–829. https://doi.org/10.1093/jn/135.4.826.
Fischer, L.M., et al., 2007. Sex and menopausal status influence human dietary requirements for the nutrient choline. Am. J. Clin. Nutr. 85 (5), 1275–1285. https://doi.org/10.1093/ajcn/85.5.1275.
Fischer, L.M., da Costa, K.A., et al., 2010a. Choline intake and genetic polymorphisms influence choline metabolite concentrations in human breast milk and plasma. Am. J. Clin. Nutr. 92 (2), 336–346. https://doi.org/10.3945/ajcn.2010.29459.
Fischer, L.M., da Costa, K.-A., et al., 2010b. Dietary choline requirements of women: effects of estrogen and genetic variation. Am. J. Clin. Nutr. 92 (5), 1113–1119. https://doi.org/10.3945/ajcn.2010.30064.
Fox, J., Betzing, H., Lekim, D., 1979. Pharmacokinetics of Orally Ingested Phosphatidylcholine.
Galletti, P., et al., 1985. Transport and metabolism of double-labelled CDPcholine in mammalian tissues. Biochem. Pharmacol. 34 (23), 4121–4130. https://doi.org/10.1016/0006-2952(85)90204-7.
Ganz, A.B., et al., 2016. Genetic impairments in folate enzymes increase dependence on dietary choline for phosphatidylcholine production at the expense of betaine synthesis. Faseb. J. 30 (10), 3321–3333. https://doi.org/10.1096/fj.201500138RR.
Ganz, A.B., et al., 2017. Genetic variation in choline-metabolizing enzymes alters choline metabolism in young women consuming choline intakes meeting current recommendations. Int. J. Mol. Sci. 18 (2), E252. https://doi.org/10.3390/ijms18020252.
Geng, J., et al., 2018. Trimethylamine N-oxide promotes atherosclerosis via CD36-dependent MAPK/JNK pathway. Biomed. Pharmacother. 97, 941–947. https://doi.org/10.1016/j.biopha.2017.11.016.
Goss, K.C.W., et al., 2020. Postnatal adaptations of phosphatidylcholine metabolism in extremely preterm infants: implications for choline and PUFA metabolism. Am. J. Clin. Nutr. 112 (6), 1438–1447. https://doi.org/10.1093/ajcn/nqaa207.
Gu, X., et al., 2017. SAMTOR is an S-adenosylmethionine sensor for the mTORC1 pathway. Science 358 (6364), 813–818. https://doi.org/10.1126/science.aao3265.
Hardwick, R.N., et al., 2014. Increased susceptibility to methotrexate-induced toxicity in nonalcoholic steatohepatitis. Toxicol. Sci. 142 (1), 45–55. https://doi.org/10.1093/toxsci/kfu156.
Heianza, Y., et al., 2017. Gut microbiota metabolites and risk of major adverse cardiovascular disease events and death: a systematic review and meta-analysis of prospective studies. J. Am. Heart Assoc. 6 (7), e004947. https://doi.org/10.1161/JAHA.116.004947.
Henneberry, A.L., Wright, M.M., McMaster, C.R., 2002. The major sites of cellular phospholipid synthesis and molecular determinants of fatty acid and lipid head group specificity. Mol. Biol. Cell 13 (9), 3148–3161. https://doi.org/10.1091/mbc.01-11-0540.
Higgins, J.P.T., Flicker, L., 2003. Lecithin for dementia and cognitive impairment. Cochrane Database Syst. Rev. (3), CD001015. https://doi.org/10.1002/14651858.CD001015.
Holmes-McNary, M.Q., et al., 1996. Choline and choline esters in human and rat milk and in infant formulas. Am. J. Clin. Nutr. 64 (4), 572–576. https://doi.org/10.1093/ajcn/64.4.572.
Horita, D.A., et al., 2021. Two methods for assessment of choline status in a randomized crossover study with varying dietary choline intake in people: isotope dilution MS of plasma and in vivo single-voxel magnetic resonance spectroscopy of liver. Am. J. Clin. Nutr. 113 (6), 1670–1678. https://doi.org/10.1093/ajcn/nqaa439.
Hui, D.Y., 2016. Intestinal phospholipid and lysophospholipid metabolism in cardiometabolic disease. Curr. Opin. Lipidol. 27 (5), 507–512. https://doi.org/10.1097/MOL.0000000000000334.
Igolnikov, A.C., Green, R.M., 2006. Mice heterozygous for the Mdr2 gene demonstrate decreased PEMT activity and diminished steatohepatitis on the MCD diet. J. Hepatol. 44 (3), 586–592. https://doi.org/10.1016/j.jhep.2005.08.026.
Ilcol, Y.O., et al., 2002. Changes of plasma free choline and choline-containing compounds' concentrations and choline loss during hemodialysis in ESRD patients. Clin. Biochem. 35 (3), 233–239. https://doi.org/10.1016/s0009-9120(02)00298-9.

Innis, S.M., et al., 2007. Choline-related supplements improve abnormal plasma methionine-homocysteine metabolites and glutathione status in children with cystic fibrosis. Am. J. Clin. Nutr. 85 (3), 702–708. https://doi.org/10.1093/ajcn/85.3.702.
Ives, C.W., et al., 2020. Preeclampsia-pathophysiology and clinical presentations: JACC state-of-the-art review. J. Am. Coll. Cardiol. 76 (14), 1690–1702. https://doi.org/10.1016/j.jacc.2020.08.014.
Jacobson, S.W., et al., 2018. Efficacy of maternal choline supplementation during pregnancy in mitigating adverse effects of prenatal alcohol exposure on growth and cognitive function: a randomized, double-blind, placebo-controlled clinical trial. Alcohol Clin. Exp. Res. 42 (7), 1327–1341. https://doi.org/10.1111/acer.13769.
Jiang, X., et al., 2012. Maternal choline intake alters the epigenetic state of fetal cortisol-regulating genes in humans. Faseb. J. 26 (8), 3563–3574. https://doi.org/10.1096/fj.12-207894.
Kamath, A.V., Darling, I.M., Morris, M.E., 2003. Choline uptake in human intestinal caco-2 cells is carrier-mediated. J. Nutr. 133 (8), 2607–2611. https://doi.org/10.1093/jn/133.8.2607.
Kennedy, E.P., 1954. Synthesis of phosphatides in isolated mitochondria. II. Incorporation of choline into lecithin. J. Biol. Chem. 209 (2), 525–535.
Kohlmeier, M., et al., 2005. Genetic variation of folate-mediated one-carbon transfer pathway predicts susceptibility to choline deficiency in humans. Proc. Natl. Acad. Sci. U. S. A. 102 (44), 16025–16030. https://doi.org/10.1073/pnas.0504285102.
Krüger, R., et al., 2017. Associations of current diet with plasma and urine TMAO in the KarMeN study: direct and indirect contributions. Mol. Nutr. Food Res. 61 (11). https://doi.org/10.1002/mnfr.201700363.
Lang, D.H., et al., 1998. Isoform specificity of trimethylamine N-oxygenation by human flavin-containing monooxygenase (FMO) and P450 enzymes: selective catalysis by FMO3. Biochem. Pharmacol. 56 (8), 1005–1012. https://doi.org/10.1016/s0006-2952(98)00218-4.
Lee, J.M., et al., 2011. A nuclear-receptor-dependent phosphatidylcholine pathway with antidiabetic effects. Nature 474 (7352), 506–510. https://doi.org/10.1038/nature10111.
Leermakers, E.T.M., et al., 2015. Effects of choline on health across the life course: a systematic review. Nutr. Rev. 73 (8), 500–522. https://doi.org/10.1093/nutrit/nuv010.
Leite, G.G.S., et al., 2020. Mapping the segmental microbiomes in the human small bowel in comparison with stool: a reimagine study. Dig. Dis. Sci. 65 (9), 2595–2604. https://doi.org/10.1007/s10620-020-06173-x.
Li, Y., et al., 2005. Crystallographic identification and functional characterization of phospholipids as ligands for the orphan nuclear receptor steroidogenic factor-1. Mol. Cell 17 (4), 491–502. https://doi.org/10.1016/j.molcel.2005.02.002.
Lockman, P.R., Allen, D.D., 2002. The transport of choline. Drug Dev. Ind. Pharm. 28 (7), 749–771. https://doi.org/10.1081/ddc-120005622.
Mann, P.J., Quastel, J.H., 1937. The oxidation of choline by rat liver. Biochem. J. 31 (6), 869–878. https://doi.org/10.1042/bj0310869.
Mazzetti, S., et al., 2013. Choline-containing compounds quantification by 1H NMR spectroscopy using external reference and noise measurements. Phys. Med. 29 (6), 677–683. https://doi.org/10.1016/j.ejmp.2012.07.002.
Meck, W.H., Smith, R.A., Williams, C.L., 1988. Pre- and postnatal choline supplementation produces long-term facilitation of spatial memory. Dev. Psychobiol. 21 (4), 339–353. https://doi.org/10.1002/dev.420210405.
Meyer, K.A., Shea, J.W., 2017. Dietary choline and betaine and risk of CVD: a systematic review and meta-analysis of prospective studies. Nutrients 9 (7), E711. https://doi.org/10.3390/nu9070711.
Miao, J., et al., 2015. Flavin-containing monooxygenase 3 as a potential player in diabetes-associated atherosclerosis. Nat. Commun. 6, 6498. https://doi.org/10.1038/ncomms7498.
Michel, V., et al., 2006. Choline transport for phospholipid synthesis. Exp. Biol. Med. 231 (5), 490–504. https://doi.org/10.1177/153537020623100503.
Musille, P.M., et al., 2016. Unexpected allosteric network contributes to LRH-1 co-regulator selectivity. J. Biol. Chem. 291 (3), 1411–1426. https://doi.org/10.1074/jbc.M115.662874.
Naber, M., Hommel, B., Colzato, L.S., 2015. Improved human visuomotor performance and pupil constriction after choline supplementation in a placebo-controlled double-blind study. Sci. Rep. 5 (1), 13188. https://doi.org/10.1038/srep13188.
National Research Council (US) Subcommittee on Laboratory Animal Nutrition, 1995. Nutrient Requirements of Laboratory Animals: Fourth Revised Edition. National Academies Press (US), Washington (DC). Available at. http://www.ncbi.nlm.nih.gov/books/NBK231927/. (Accessed 2 August 2021).
Nguyen, L.N., et al., 2014. Mfsd2a is a transporter for the essential omega-3 fatty acid docosahexaenoic acid. Nature 509 (7501), 503–506. https://doi.org/10.1038/nature13241.
Niculescu, M.D., Craciunescu, C.N., Zeisel, S.H., 2006. Dietary choline deficiency alters global and gene-specific DNA methylation in the developing hippocampus of mouse fetal brains. Faseb. J. 20 (1), 43–49. https://doi.org/10.1096/fj.05-4707com.
Niculescu, M.D., et al., 2007. Lymphocyte gene expression in subjects fed a low-choline diet differs between those who develop organ dysfunction and those who do not. Am. J. Clin. Nutr. 86 (1), 230–239. https://doi.org/10.1093/ajcn/86.1.230.
Obeid, R., 2013. The metabolic burden of methyl donor deficiency with focus on the betaine homocysteine methyltransferase pathway. Nutrients 5 (9), 3481–3495. https://doi.org/10.3390/nu5093481.
Olthof, M.R., et al., 2005. Choline supplemented as phosphatidylcholine decreases fasting and postmethionine-loading plasma homocysteine concentrations in healthy men. Am. J. Clin. Nutr. 82 (1), 111–117. https://doi.org/10.1093/ajcn.82.1.111.
Ortlund, E.A., et al., 2005. Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP. Nat. Struct. Mol. Biol. 12 (4), 357–363. https://doi.org/10.1038/nsmb910.
Pelletier, C.C., et al., 2019. Elevation of trimethylamine-N-oxide in chronic kidney disease: contribution of decreased glomerular filtration rate. Toxins 11 (11), 635. https://doi.org/10.3390/toxins11110635.
Petruzzelli, M., et al., 2016. Biliary phospholipids sustain enterocyte proliferation and intestinal tumor progression via nuclear receptor Lrh1 in mice. Sci. Rep. 6, 39278. https://doi.org/10.1038/srep39278.
Poly, C., et al., 2011. The relation of dietary choline to cognitive performance and white-matter hyperintensity in the Framingham Offspring Cohort. Am. J. Clin. Nutr. 94 (6), 1584–1591. https://doi.org/10.3945/ajcn.110.008938.
Reinoso Webb, C., et al., 2018. Differential susceptibility to T cell-induced colitis in mice: role of the intestinal microbiota. Inflamm. Bowel Dis. 24 (2), 361–379. https://doi.org/10.1093/ibd/izx014.
Resseguie, M., et al., 2007. Phosphatidylethanolamine N-methyltransferase (PEMT) gene expression is induced by estrogen in human and mouse primary hepatocytes. Faseb. J. 21 (10), 2622–2632. https://doi.org/10.1096/fj.07-8227com.
Resseguie, M.E., et al., 2011. Aberrant estrogen regulation of PEMT results in choline deficiency-associated liver dysfunction. J. Biol. Chem. 286 (2), 1649–1658. https://doi.org/10.1074/jbc.M110.106922.
Roberts, A.B., et al., 2018. Development of a gut microbe-targeted nonlethal therapeutic to inhibit thrombosis potential. Nat. Med. 24 (9), 1407–1417. https://doi.org/10.1038/s41591-018-0128-1.
Romano, K.A., et al., 2015. Intestinal microbiota composition modulates choline bioavailability from diet and accumulation of the proatherogenic metabolite trimethylamine-N-oxide. mBio 6 (2), e02481. https://doi.org/10.1128/mBio.02481-14.
Ross, R.G., et al., 2013. Perinatal choline effects on neonatal pathophysiology related to later schizophrenia risk. Am. J. Psychiatr. 170 (3), 290–298. https://doi.org/10.1176/appi.ajp.2012.12070940.
Saitoh, H., et al., 1992. Carrier-mediated transport system for choline and its related quaternary ammonium compounds on rat intestinal brush-border membrane. Biochim. Biophys. Acta 1112 (1), 153–160. https://doi.org/10.1016/0005-2736(92)90265-n.
Sawant, O.B., et al., 2019. Maternal choline supplementation mitigates alcohol-induced fetal cranio-facial abnormalities detected using an ultrasonographic examination in a sheep model. Alcohol 81, 31–38. https://doi.org/10.1016/j.alcohol.2019.05.001.

Schmidt, A.C., Leroux, J.-C., 2020. Treatments of trimethylaminuria: where we are and where we might be heading. Drug Discov. Today 25 (9), 1710–1717. https://doi.org/10.1016/j.drudis.2020.06.026.
Sha, W., et al., 2010. Metabolomic profiling can predict which humans will develop liver dysfunction when deprived of dietary choline. Faseb. J. 24 (8), 2962–2975. https://doi.org/10.1096/fj.09-154054.
Shaw, G.M., et al., 2004. Periconceptional dietary intake of choline and betaine and neural tube defects in offspring. Am. J. Epidemiol. 160 (2), 102–109. https://doi.org/10.1093/aje/kwh187.
Shaw, G.M., et al., 2009. Choline and risk of neural tube defects in a folate-fortified population. Epidemiology 20 (5), 714–719. https://doi.org/10.1097/EDE.0b013e3181ac9fe7.
Shunova, A., et al., 2020. Choline content of term and preterm infant formulae compared to expressed breast milk-how do we justify the discrepancies? Nutrients 12 (12), E3815. https://doi.org/10.3390/nu12123815.
Simenhoff, M.L., et al., 1976. Amine metabolism and the small bowel in uraemia. Lancet 2 (7990), 818–821. https://doi.org/10.1016/s0140-6736(76)91207-1.
Stremmel, W., et al., 2017. Blood trimethylamine-N-oxide originates from microbiota mediated breakdown of phosphatidylcholine and absorption from small intestine. PLoS One 12 (1), e0170742. https://doi.org/10.1371/journal.pone.0170742.
Strupp, B.J., et al., 2016. Maternal choline supplementation: a potential prenatal treatment for down syndrome and Alzheimer's disease. Curr. Alzheimer Res. 13 (1), 97–106.
Sun, S., et al., 2016. Choline and betaine consumption lowers cancer risk: a meta-analysis of epidemiologic studies. Sci. Rep. 6, 35547. https://doi.org/10.1038/srep35547.
Sweiry, J.H., et al., 1986. Evidence of saturable uptake mechanisms at maternal and fetal sides of the perfused human placenta by rapid paired-tracer dilution: studies with calcium and choline. J. Dev. Physiol. 8 (6), 435–445.
Tang, W.H.W., et al., 2015. Gut microbiota-dependent trimethylamine N-oxide (TMAO) pathway contributes to both development of renal insufficiency and mortality risk in chronic kidney disease. Circ. Res. 116 (3), 448–455. https://doi.org/10.1161/ClRCRESAHA.116.305360.
Traiffort, E., O'Regan, S., Ruat, M., 2013. The choline transporter-like family SLC44: properties and roles in human diseases. Mol. Aspect. Med. 34 (2–3), 646–654. https://doi.org/10.1016/j.mam.2012.10.011.
Tsujiuchi, T., et al., 1999. Hypomethylation of CpG sites and c-myc gene overexpression in hepatocellular carcinomas, but not hyperplastic nodules, induced by a choline-deficient L-amino acid-defined diet in rats. Jpn. J. Cancer Res. 90 (9), 909–913. https://doi.org/10.1111/j.1349-7006.1999.tb00834.x.
Vance, D.E., 2013. Physiological roles of phosphatidylethanolamine N-methyltransferase. Biochim. Biophys. Acta 1831 (3), 626–632. https://doi.org/10.1016/j.bbalip.2012.07.017.
Vanek, V.W., et al., 2012. A.S.P.E.N. position paper: recommendations for changes in commercially available parenteral multivitamin and multi-trace element products. Nutr. Clin. Pract. 27 (4), 440–491. https://doi.org/10.1177/0884533612446706.
Vanek, V.W., et al., 2015. A call to action to bring safer parenteral micronutrient products to the U.S. Market. Nutr. Clin. Pract. 30 (4), 559–569. https://doi.org/10.1177/0884533615589992.
Varela-Moreiras, G., Ragel, C., Pérez de Miguelsanz, J., 1995. Choline deficiency and methotrexate treatment induces marked but reversible changes in hepatic folate concentrations, serum homocysteine and DNA methylation rates in rats. J. Am. Coll. Nutr. 14 (5), 480–485. https://doi.org/10.1080/07315724.1995.10718539.
Veenema, K., et al., 2008. Adequate Intake levels of choline are sufficient for preventing elevations in serum markers of liver dysfunction in Mexican American men but are not optimal for minimizing plasma total homocysteine increases after a methionine load. Am. J. Clin. Nutr. 88 (3), 685–692. https://doi.org/10.1093/ajcn/88.3.685.
Wagner, M., et al., 2016. LRH-1 is a critical determinant of methyl-pool metabolism. Hepatology 63 (1), 95–106. https://doi.org/10.1002/hep.28124.
Wallace, T.C., Fulgoni, V.L., 2016. Assessment of total choline intakes in the United States. J. Am. Coll. Nutr. 35 (2), 108–112. https://doi.org/10.1080/07315724.2015.1080127.
Wallace, T.C., Fulgoni, V.L., 2017. Usual choline intakes are associated with egg and protein food consumption in the United States. Nutrients 9 (8), E839. https://doi.org/10.3390/nu9080839.
Wang, Z., et al., 2019. Impact of chronic dietary red meat, white meat, or non-meat protein on trimethylamine N-oxide metabolism and renal excretion in healthy men and women. Eur. Heart J. 40 (7), 583–594. https://doi.org/10.1093/eurheartj/ehy799.
Wang, Y., et al., 2020. Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. J. Exp. Med. 217 (2), e20191130. https://doi.org/10.1084/jem.20191130.
Waterland, R.A., Jirtle, R.L., 2003. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. Mol. Cell Biol. 23 (15), 5293–5300. https://doi.org/10.1128/MCB.23.15.5293-5300.2003.
Watkins, S.M., Zhu, X., Zeisel, S.H., 2003. Phosphatidylethanolamine-N-methyltransferase activity and dietary choline regulate liver-plasma lipid flux and essential fatty acid metabolism in mice. J. Nutr. 133 (11), 3386–3391. https://doi.org/10.1093/jn/133.11.3386.
White, H.L., Scates, P.W., 1990. Acetyl-l-carnitine as a precursor of acetylcholine. Neurochem. Res. 15 (6), 597–601. https://doi.org/10.1007/BF00973749.
Yan, J., et al., 2012. Maternal choline intake modulates maternal and fetal biomarkers of choline metabolism in humans. Am. J. Clin. Nutr. 95 (5), 1060–1071. https://doi.org/10.3945/ajcn.111.022772.
Zeisel, S.H., Warrier, M., 2017. Trimethylamine N-oxide, the microbiome, and heart and kidney disease. Annu. Rev. Nutr. 37, 157–181. https://doi.org/10.1146/annurev-nutr-071816-064732.
Zeisel, S.H., Epstein, M.F., Wurtman, R.J., 1980. Elevated choline concentration in neonatal plasma. Life Sci. 26 (21), 1827–1831. https://doi.org/10.1016/0024-3205(80)90585-8.
Zeisel, S.H., Wishnok, J.S., Blusztajn, J.K., 1983. Formation of methylamines from ingested choline and lecithin. J. Pharmacol. Exp. Therapeut. 225 (2), 320–324.
Zeisel, S.H., et al., 1991. Choline, an essential nutrient for humans. Faseb. J. 5 (7), 2093–2098.
Zeisel, S.H., 2011. What choline metabolism can tell us about the underlying mechanisms of fetal alcohol spectrum disorders. Mol. Neurobiol. 44 (2), 185–191. https://doi.org/10.1007/s12035-011-8165-5.
Zeisel, S.H., 2012. Dietary choline deficiency causes DNA strand breaks and alters epigenetic marks on DNA and histones. Mutat. Res. 733 (1–2), 34–38. https://doi.org/10.1016/j.mrfmmm.2011.10.008.
Zhang, J., Blusztajn, J.K., Zeisel, S.H., 1992. Measurement of the formation of betaine aldehyde and betaine in rat liver mitochondria by a high pressure liquid chromatography-radioenzymatic assay. Biochim. Biophys. Acta 1117 (3), 333–339. https://doi.org/10.1016/0304-4165(92)90033-q.
Zhu, W., et al., 2016. Gut microbial metabolite TMAO enhances platelet hyperreactivity and thrombosis risk. Cell 165 (1), 111–124. https://doi.org/10.1016/j.cell.2016.02.011.

Cholesterol: Sources, absorption, function, and metabolism

D.J. McNamara, Eggs for Health Consulting, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of D.J. McNamara, Cholesterol: Sources, Absorption, Function, and Metabolism, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 341–345, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00052-0.

Absorption, Transport, and Storage **175**
Cholesterol absorption 175
Exogenous Cholesterol Transport 176
Tissue Uptake and Storage 176
Biosynthesis **177**
Tissue Cholesterol Synthesis 177
Regulation of Synthesis 177
Metabolism and Excretion 177
Bile Acid Synthesis 178
Biliary Cholesterol Secretion 178
Fecal Excretion 178
Metabolic Function **178**
Steroid Hormones 178
Bile Acid Synthesis 178
Very Low-Density Lipoprotein Synthesis 179
Dietary Cholesterol and Plasma Cholesterol 179
Dietary Sources **179**
Dietary Cholesterol Intake Patterns 179
Major Dietary Sources 179
Further Reading **179**

Absorption, Transport, and Storage

Cholesterol absorption

Cholesterol in the intestinal lumen typically consists of one-third dietary cholesterol and two-thirds biliary cholesterol. The average daily diet contains 300–500 mg of cholesterol obtained from animal products. The bile provides an additional 800–1200 mg of cholesterol throughout each day as gallbladder contractions provide a flow of bile acids, cholesterol, and phospholipids to facilitate lipid digestion and absorption. Dietary cholesterol is a mixture of free and esterified cholesterol whereas biliary cholesterol is nonesterified and is introduced into the small intestine as a cholesterol–bile salt–phospholipid water-soluble complex. The only other source of intraluminal cholesterol is mucosal cell cholesterol, derived from either sloughed mucosal cells or cholesterol secreted by the mucosal cells into the intestinal lumen. Measurements of exogenous and endogenous cholesterol absorption in humans indicate that there is probably very little direct secretion of newly synthesized cholesterol from mucosal cells into the intraluminal contents.

Cholesterol absorption occurs primarily in the duodenum and proximal jejunum of the small intestine and is dependent on the presence of bile salts. In the absence of bile secretion, or in the presence of bile acid-binding resins, there is virtually no intestinal absorption of cholesterol. On average, humans absorb 50–60% of the intestinal contents of cholesterol, but there is a large interindividual variance in absorption, with values ranging from as low as 20% to as high as 80%. Intestinal transit time is related to cholesterol absorption with slower transit times resulting in higher fractional absorption rates. Dietary factors that affect the relative percent absorption of cholesterol include the total mass of dietary cholesterol, the concentration of plant sterols in the diet, and the type and amount of dietary fiber. Studies suggest that the ratio of polyunsaturated to saturated fat (P:S) in the diet has little effect on cholesterol absorption rates in humans, nor does the amount of dietary fat.

Two interesting, and as yet undefined, aspects of cholesterol absorption are that it decreases as the mass of cholesterol increases above an intake of 1500 mg per day, and that the fractional absorption below this level is relatively constant for an individual. For example, at a daily cholesterol intake of 800 mg a subject might absorb 60% or 480 mg a day, whereas at a daily intake of 400 mg the absorption remains at 60%, equaling 240 mg a day absorbed. The quandary is, if the system can accommodate absorption of 480 mg at the high cholesterol intake, then why is the amount absorbed 240 mg at the low intake? Clearly the upper value of cholesterol absorption is achievable, yet at the lower intake level the absorption rate stays at a fixed fractional value. The mechanisms controlling this aspect of cholesterol absorption have not been defined.

Experimental evidence indicates that biliary cholesterol and dietary cholesterol are absorbed equally; however, the pattern of exogenous and endogenous cholesterol absorption differs along the length of the intestinal lumen. Dietary cholesterol enters

the small intestine solubilized in the oil phase of the stomach digest, whereas the binary cholesterol enters in the micelle phase of the bile. This differential distribution results in a greater absorption of biliary cholesterol in the upper portion of the small intestine with dietary cholesterol absorption increasing as the oil phase of the intestinal contents are hydrolyzed. As the oil phase is reduced, dietary cholesterol moves from the oil phase to the aqueous micelle phase and becomes available for absorption. In the case of cholesteryl esters in the diet, it is necessary that the esters are hydrolyzed by pancreatic cholesterol esterase (CEase) before the cholesterol is available for absorption. Pancreatic CEase requires the presence of bile salts for activity and may play a key role in the actual absorption process.

The process, and selectivity, of sterol absorption involves a complex interplay of regulated transporters, transporting sterols into and out of the enterocyte, and the assembly and secretion of chylomicrons into the lymph. The enterocyte takes up both cholesterol and phytosterols from the intestinal lumen by what appears to be a common sterol transporter or permease in the brush border membrane. Preliminary studies suggest that the Neiman–Pick C1 Like 1 (NPC1L1) protein is involved in this process. Once the sterols enter the enterocyte, the ATP-binding cassette (ABC) hemitransporters ABCG5 and ABCG8 function in the apical excretion of sterols back into the intestinal lumen. The selectivity of this process accounts for the higher absorption rates of cholesterol (50–60%) compared to the phytosterols, which are very poorly absorbed. Loss of ABCG5/G8 function results in excessive absorption of both cholesterol and phytosterols. Studies in mice have shown that ABCG5/G8 are expressed primarily in the liver and intestine, are coordinately up-regulated at the transcriptional level by dietary cholesterol intake, and require the liver X receptor α (LXRα), a nuclear receptor that regulates the expression of a number of key genes involved in lipid metabolism.

Evidence is accumulating that the fractional cholesterol absorption rates are regulated by one or more genetic determinants. The apolipoprotein (apo) E phenotype has a significant effect on fractional cholesterol absorption and appears to play a major role in determining the plasma lipoprotein response to changes in dietary cholesterol intake. Men with the apoE4 allele have a high cholesterol absorption rate whereas those with the apoE2 allele have a low cholesterol absorption efficiency. The absorption values for the more common apoE3/3 fall between the apoE2 and apoE4 patterns. Polymorphisms of the apolipoprotein A-IV and of the low-density lipoprotein (LDL) receptor gene have also been related to differences in fractional cholesterol absorption. These genetic variants affecting cholesterol absorption no doubt play a significant role in determining an individual's fractional absorption of cholesterol as well as accounting for much of the heterogeneity of plasma lipid responses to changes in dietary cholesterol intakes (see the Section on Dietary Cholesterol and Plasma Cholesterol below).

Exogenous Cholesterol Transport

Cholesterol is absorbed in the unesterified state, whereas the cholesterol secreted into the lymph is 70–80% esterified. This esterification process generates a concentration gradient of free cholesterol within the mucosal cell, which could facilitate absorption rates. Cholesterol is esterified in intestinal mucosal cells by acyl-coenzyme A: cholesterol acyltransferase-2 (ACAT-2) to form cholesteryl esters, which are secreted from the basolateral surface of the enterocyte as part of the chylomicrons. At this stage it is assumed that cholesterol molecules from exogenous and endogenous sources are indistinguishable, and have similar effects on endogenous cholesterol and lipoprotein metabolism. Chylomicrons are large particles (>70 nm in diameter) composed mainly of triacylglycerols (95% by weight) and containing 3–7% cholesterol by weight, the esterified cholesterol localized in the hydrophobic core and the free cholesterol primarily in the hydrophilic outer layer. The data indicate that the amount of dietary cholesterol consumed has little effect on the cholesterol content of chylomicrons. The chylomicrons are released from the intestinal cells, enter the lymphatic system and are transported *via* the lymphatics (thoracic duct) to the bloodstream. Because chylomicrons are too large to pass through the capillaries, this is the only mechanism by which they can enter the bloodstream.

In the plasma compartment the chylomicrons pick up a number of apolipoproteins, which are required for intravascular metabolism of the particles. The initial metabolism of chylomicrons involves hydrolysis of the associated triacylglycerols by endothelial cell lipoprotein lipase (LPL) located in adipose, muscle, and heart tissues which results in production of chylomicron remnants. The chylomicron remnants, depleted of triacylglycerol and enriched with cholesteryl ester, are taken up by the liver *via* the LDL receptor-related protein (LRP). The ligand for hepatic uptake of the chylomicron remnant appears from various transgenic mouse studies to be the apo-E moiety of the particle. The clearance of chylomicrons from the bloodstream is rapid, with particles having a half-life of less than an hour. The liver cannot take up native chylomicrons but rather takes up the chylomicron remnant, which has lost approximately 90% of its triacylglycerol content and become relatively enriched in free and esterified cholesterol through the actions of the plasma cholesteryl ester transfer protein (CETP), which transfers cholesteryl ester from HDL to the apo-B-containing lipoproteins.

The chylomicron remnants taken up by the liver are subjected to lysosomal hydrolysis resulting in the release of the absorbed dietary and biliary cholesterol into the hepatocyte as free cholesterol. The influx of cholesterol contained in the chylomicron remnant has the ability to affect a number of regulatory sites of hepatic cholesterol metabolism, which function to maintain cholesterol homeostasis in the liver. The liver has four primary fates for the newly delivered cholesterol: catabolism to bile acids; secretion as biliary cholesterol; storage in lipid droplets as cholesteryl ester; or incorporation into very low-density lipoprotein (VLDL) for secretion from the liver.

Tissue Uptake and Storage

The body pool of cholesterol is approximately 145 g with one-third of this mass localized in the central nervous system. The remainder of the metabolically active cholesterol pool exists in the plasma compartment (7.5–9 g) and as constituents of body

tissues. In humans, tissue cholesterol levels are relatively low, averaging 2–3 mg gm^{-1} wet weight. Little information exists regarding changes in hepatic and extrahepatic tissue cholesterol concentrations with changes in dietary cholesterol intake. Animal studies, which are usually carried out using very high levels of dietary cholesterol, have shown that hepatic cholesterol can increase from 2-fold up to 10-fold, depending on the species and other dietary constituents, when dietary cholesterol is increased.

Biosynthesis

Tissue Cholesterol Synthesis

Cholesterol biosynthesis occurs in every nucleated cell in the body. Although it is often thought that the majority of cholesterol synthesis occurs in the liver, studies have shown that the bulk tissues of the body account for the overwhelming majority of endogenous cholesterol production. Hepatic cholesterol synthesis in humans is thought to contribute 10–20% of the total daily synthesis rate. Because the majority of cholesterol synthesis in the body occurs in extrahepatic tissues, and the only quantitatively significant site for excretion and catabolism of cholesterol is the liver, some 600–800 mg of cholesterol each day must be transported from peripheral tissues through the plasma compartment to the liver to account for daily cholesterol catabolism and binary secretion. Approximately 9 mg cholesterol per kilogram body weight is synthesized by peripheral tissues every day and must be moved to the liver for catabolism *via* a process termed 'reverse cholesterol transport' (RCT).

RCT describes the metabolism, and important antiatherogenic function, of the HDL-mediated efflux of cholesterol from nonhepatic cells and its subsequent delivery to the liver and steroidogenic organs for use in the synthesis of lipoproteins, bile acids, vitamin D, and steroid hormones. A cellular ABC transporter (ABCA1) mediates the first step of RCT involving the transfer of cellular cholesterol and phospholipids to lipid-poor apolipoproteins. Lecithin:cholesterol acyltransferase (LCAT) mediated esterification of cholesterol generates spherical particles that continue to expand with ongoing cholesterol esterification and phospholipid transfer protein (PLTP) mediated particle fusion and surface remnant transfer. Larger HDL_2 particles are converted into smaller HDL_3 particles when CETP facilitates the transfer of cholesteryl esters from HDL onto apo-B-containing lipoproteins. The scavenger receptor B1 (SR-BI) promotes selective uptake of cholesteryl esters into liver and steroidogenic organs whereas hepatic lipase (HL) and LPL mediated hydrolysis of phospholipids and triglycerides. SR-BI mediates the selective uptake of cholesteryl esters from HDL and also LDL into hepatocytes and steroid hormone-producing cells without internalizing HDL proteins, which can recycle through the RCT sequence moving cholesterol from peripheral tissues to the liver.

Regulation of Synthesis

The rate-limiting enzyme in cholesterol biosynthesis is 3-hydroxy-3-methylglutaryl coenzyme A (HMG- CoA) reductase, a microsomal enzyme which converts HMG-COA to mevalonic acid in the polyisoprenoid synthetic pathway. Peripheral tissue cholesterol synthesis is much less responsive to regulatory factors compared with the liver, which is controlled by a variety of dietary, hormonal, and physiological variables. Studies indicate that endogenous cholesterol synthesis is significantly increased in obesity and in patients with the metabolic syndrome. Obesity, insulin resistance, and diabetes have pronounced effects on both cholesterol absorption and synthesis. Findings in type I diabetes appear related to low expression of ABCG5/G8 genes resulting in high absorption and low synthesis of cholesterol. Cholesterol absorption efficiency is lower and cholesterol synthesis higher in obese subjects with type II diabetes compared to obese subjects without diabetes, suggesting that diabetes modulates cholesterol metabolism to a greater extent than obesity alone. In a similar manner, low cholesterol absorption and high synthesis appears to be part of the insulin resistance (metabolic) syndrome.

Research shows that in most individuals, dietary cholesterol alters endogenous cholesterol synthesis and that this feedback regulation can effectively compensate for increased cholesterol input from dietary sources. The precision of these regulatory responses depends on a number of genetic factors, and data suggest that multiple genetic loci are involved. For example, family studies have shown that in siblings of low cholesterol absorption families, cholesterol absorption percentages are significantly lower, and cholesterol and bile acid synthesis, cholesterol turnover, and fecal steroids significantly higher than in siblings of high absorption families.

Metabolism and Excretion

The body's metabolic processes cannot break the sterol rings of cholesterol and therefore must either catabolize cholesterol to other products, which can be excreted in the urine or feces, or directly excrete cholesterol in the bile, with a fraction of the biliary cholesterol lost daily as fecal neutral sterols. In humans, the major route of excretion is as biliary cholesterol (two-thirds of the total lost each day), with catabolism to bile acids and bile acid excretion being the second most important route, accounting for approximately one-third of the daily turnover.

For all practical purposes, the body must excrete daily an amount of neutral and acidic sterols equivalent to the combined inputs of total dietary and newly synthesized cholesterol. Given an average fecal excretion of 1020 mg a day with 250 mg as acidic sterols, it can be calculated that the 770 mg per day excreted as neutral steroids comes from unabsorbed biliary (650 mg) and unabsorbed dietary (120 mg) cholesterol (**Table 1**). It is easy to see that even small changes in the daily balance between a cholesterol input and output value of 1020 mg per day could, over years, result in significant tissue cholesterol accumulation.

Bile Acid Synthesis

The results from numerous sterol balance studies carried out in subjects fed diets low and high in cholesterol indicate that in humans dietary cholesterol has little effect on fecal bile acid excretion rates. This finding is in striking contrast to results from studies in some rodent models, which show that intake of pharmacological doses of dietary cholesterol can result in several-fold increases in bile acid synthesis and excretion. In contrast, some rodent species and nonhuman primates have little if any increase in bile acid excretion with increased intakes of cholesterol. Although there have been a few reports of enhanced bile acid excretion on a high-cholesterol diet in some patients, this does not appear to be a major regulatory response in humans.

Biliary Cholesterol Secretion

The majority of cholesterol entering the intestinal tract is biliary cholesterol. Biliary cholesterol secretion averages 1000 mg per day as part of the bile system and enters as free cholesterol already solubilized with bile acids and phospholipids. Both cholesterol absorption by enterocyte and biliary cholesterol secretion by hepatic cells are regulated by expression of the half-transporters ABCG5 and ABCG8. Studies in animals have shown that treatment with a LXR agonist decreases cholesterol absorption, increases biliary cholesterol secretion, and increases fecal neutral sterol excretion. Studies in transgenic mouse models demonstrate that increased expression of ABCG5 and ABCG8 increases biliary neutral sterol secretion and reduces intestinal cholesterol absorption, leading to increased neutral sterol excretion and cholesterol synthesis.

Fecal Excretion

The only route of significant cholesterol excretion is through fecal excretion of neutral sterols. The combination of unabsorbed biliary and dietary cholesterol accounts for the total neutral sterol output, and under most conditions equals 750–850 mg a day. Dietary patterns or drugs that interfere with intestinal cholesterol absorption result in increased fecal neutral steroid excretion. In the colon, intestinal bacteria are able to metabolize cholesterol to a variety of neutral steroids as well as to nonsteroid end products. There have been some studies suggesting that the intestinal metabolism of cholesterol by bacteria, which can be altered by diet and drugs, can influence endogenous cholesterol metabolism as well as plasma cholesterol levels. What these relationships might be and the mechanisms involved have not been defined.

Metabolic Function

Steroid Hormones

Daily production of steroid hormones is quantitatively a very small fraction of the daily turnover of dietary and newly synthesized cholesterol in the body. For men, the average daily excretion of steroid hormones is approximately 50 mg per day, whereas in women the value can be substantially higher depending on the menstrual phase.

Bile Acid Synthesis

The enterohepatic circulation of bile acids is essential for fat and cholesterol digestion and absorption. Each day the bile acid pool (approximately 3–5 g) cycles through the intestine 6–10 times. The absorption of bile acids by the ileum is very effective and 98–99% of bile acids secreted in the bile are returned to the liver *via* the portal vein. The small amount of bile acids lost each day as fecal acidic steroids are replaced through the conversion of hepatic cholesterol to the primary bile acids, cholic acid, and chenodeoxycholic acid. This catabolism of cholesterol can be as little as 250 mg per day up to 500 mg per day depending on the diet. The bile acids represent the only major catabolic product of cholesterol metabolism and in humans account for some 25–30% of the daily loss of cholesterol from the body.

Table 1 Average cholesterol metabolism values for a 70 kg adult

Cholesterol pools and flux	*Mass*
Cholesterol pool (70 kg adult)	160 g
Plasma cholesterol pool	8 g
Dietary cholesterol intake	300 mg day^{-1}
Absorption (average 60%)	180 mg day^{-1}
Synthesis (12 mg kg-day^{-1})	840 mg day^{-1}
Total cholesterol input	1020 mg day^{-1}
Bile acid synthesis (=fecal excretion)	250 mg day^{-1}
Neutral steroid excretion	770 mg day^{-1}

Very Low-Density Lipoprotein Synthesis

The endogenous pathway for cholesterol transport focuses on the liver with the synthesis and secretion of VLDL particles. Cholesterol in these triacylglycerol-rich particles comes from multiple sources: endogenous synthesis, diet, and plasma lipoproteins. Catabolism of VLDL by LPL leads to formation of intermediate-density lipoproteins (IDL), which can either be taken up by the liver or undergo further metabolism to form LDL. Low-density lipoproteins contain apo-B_{100} and account for 60–80% of the plasma cholesterol in most individuals. During lipolysis of VLDL triacylglycerol, the lipoproteins containing apo-B becomes enriched with cholesteryl ester through the plasma CETP-catalyzed net transfer of cholesteryl ester from HDL. This process, termed reverse cholesterol transport, moves cholesterol from extrahepatic tissues to HDL to VLDL-IDL-LDL and eventual uptake by the liver. Some 70% of the LDL degraded each day is degraded by the hepatic LDL receptor pathway.

Dietary Cholesterol and Plasma Cholesterol

The effect of dietary cholesterol on plasma cholesterol levels has been an area of considerable debate. In 1972 the American Heart Association recommended that dietary cholesterol intake should average less than 300 mg per day as part of a 'heart-healthy', plasma cholesterol-lowering diet. Since that initial recommendation, a number of other public health dietary recommendations in the USA have endorsed the 300 mg daily limit. Interestingly, few dietary recommendations from other countries contain a dietary cholesterol limitation. The evidence for a relationship between dietary cholesterol and plasma cholesterol indicates that the effect is relatively small, and that on average a change of 100 mg per day in dietary cholesterol intake results in a 0.057 mmol l^{-1} (2.2 mg dl^{-1}) change in plasma cholesterol concentrations. Studies have also shown that the majority of individuals are resistant to the plasma cholesterol-raising effects of dietary cholesterol (nonresponders) and have less than the predicted response. In contrast, a segment of the population (estimated to be between 10% and 20%) are sensitive to dietary cholesterol (responders) and exhibit a greater than expected plasma cholesterol response to a change in dietary cholesterol intake. To date there are no defined physiological or clinical characteristics to differentiate responders from nonresponders, but studies suggest that the apo-E phenotype plays a role, as does the clinical condition of combined hyperlipidemia. Data also suggest that sensitivity to dietary cholesterol is associated with sensitivity to dietary fat, and that overall adiposity may also play a role. Although on a population basis the plasma cholesterol response to dietary cholesterol is relatively small, and in most epidemiological analyses not related to hypercholesterolemia, some individuals are sensitive to dietary cholesterol changes and, if hypercholesterolemic, would experience plasma cholesterol reduction with dietary cholesterol restrictions. For the majority, however, dietary cholesterol restrictions have little effect on plasma cholesterol levels.

Dietary Sources

Dietary Cholesterol Intake Patterns

Dietary cholesterol intakes in the USA have been declining, from an average of 500 mg a day in men and 320 mg a day in women in 1972 to levels in 1990 of 360 mg a day in men and 240 mg a day in women. This decline arises in part from dietary recommendations to the American public to reduce total and saturated fat intake and to reduce dietary cholesterol daily intake to less than 300 mg, and in part from the increased availability of products with reduced fat and cholesterol content. Major efforts in the early 1970s by public health agencies and advertising emphasized reducing dietary cholesterol as a means to lower plasma cholesterol levels, leading to a high degree of consumer concern regarding cholesterol-containing foods and demand for low-cholesterol products. Today, practically all foods sold in the USA are labeled for their cholesterol content and their percentage contribution to the daily value of 300 mg for cholesterol.

Major Dietary Sources

The major sources of cholesterol in the diet are eggs, meat, and dairy products. A large egg contains approximately 185 mg of cholesterol and contributes some 30–35% of the total dietary cholesterol intake in the USA. Meat, poultry, and fish contribute 45–50%, dairy products 12–15% and fats and oils 4–6%. In the USA, the range of dietary cholesterol intakes is 300–400 mg per day for men and 200–250 mg per day for women; thus for much of the population the national goal of a dietary cholesterol intake of less than 300 mg a day has already been met.

Further Reading

Dietschy, JM, Turley, SD, Spady, DK, 1993. Role of liver in the maintenance of cholesterol and low density lipoprotein homeostasis in different animal species, including humans. Journal of Lipid Research 34, 1637–1659.

Gylling, H, Miettinen, TA, 2002. Inheritance of cholesterol metabolism of probands with high or low cholesterol absorption. Journal of Lipid Research 43, 1472–1476.

Jia, L, Betters, JL, Yu, L, 2011. Niemann-Pick C1-Like 1 (NPC1L1) protein in intestinal and hepatic cholesterol transport. Annual Review of Physiology 73, 239–259.

McNamara, DJ, 1987. Effects of fat-modified diets on cholesterol and lipoprotein metabolism. Annual Review of Nutrition 7, 273–290.

McNamara, DJ, 1990. Relationship between blood and dietary cholesterol. Advances in Meat Research 6, 63–87.

McNamara, DJ, 2000. Dietary cholesterol and atherosclerosis. Biochim Biophys Acta 1529, 310–320.
Millatt, LJ, Bocher, V, Fruchart, JC, Staels, B, 2003. Liver X receptors and the control of cholesterol homeostasis: Potential therapeutic targets for the treatment of atherosclerosis. Biochimica Biophysica Acta 1631, 107–118.
Oram, JF, 2003. HDL apolipoproteins and ABCA1 - Partners in the removal of excess cellular cholesterol. Arteriosclerosis Thrombosis and Vascular Biology 23, 720–727.
Ordovas, JM, Tai, ES, 2002. The babel of the ABCs: Novel transporters involved in the regulation of sterol absorption and excretion. Nutrition Reviews 60, 30–33.
Sehayek, E, 2003. Genetic regulation of cholesterol absorption and plasma plant sterol levels: Commonalities and differences. Journal of Lipid Research 44, 2030–2038.
Thompson, GR, Naumova, RP, Watts, GF, 1996. Role of cholesterol in regulation of apolipoprotein B secretion by the liver. Journal of Lipid Research 37, 439–447.
Wilson, MD, Rudel, LL, 1994. Review of cholesterol absorption with emphasis on dietary and biliary cholesterol. Journal of Lipid Research 35, 943–955.
Yu, LQ, Li-Hawkins, J, Hammer, RE, et al., 2002. Overexpression of ABCG5 and ABCG8 promotes biliary cholesterol secretion and reduces fractional absorption of dietary cholesterol. Journal of Clinical Investigation 110, 671–680.

Chromium

John B. Vincent, Department of Chemistry and Biochemistry, The University of Alabama, Tuscaloosa, AL, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of R.A. Anderson, Chromium, In Encyclopedia of Human Nutrition (Third Edition), edited by Benjamin Caballero, Academic Press, 2013, https://doi.org/10.1016/B978-0-12-375083-9.00054-4.

Introduction 181
Intake guidelines 182
Lack of symptoms of deficiency 182
Body mass 182
Dietary sources 183
Digestion 183
Transport 183
Toxicity 184
Chromium supplementation and human health 185
Pharmacological effects in rodents 185
Molecular mechanism(s) of action 186
Conclusion 187
Conflict of interest 187
References 187

Key Points

- Chromium cannot currently be considered an essential element.
- No definitive symptoms of chromium deficiency have been established.
- Chromium supplementation has no effects on body mass or composition.
- Chromium supplementation has no clinically significant effects on the symptoms of insulin resistance or type 2 diabetes.
- Chromium supplementation appears safe at doses of 1 mg/day or less.

Glossary

Adequate Intake (AI) The recommended daily intake of a nutrient estimated to meet or exceed the amount needed to maintain adequate nutrition for most people in a particular life stage and gender group. An AI is established when not enough information is available from scientific research to determine a Recommended Dietary Allowance.

Estimated Average Requirement (EAR) A nutrient intake value estimated to meet the requirement of half the healthy individuals in a group.

Recommended Daily Allowance (RDA) Average daily level of intake sufficient to meet the nutrient requirements of nearly all (97%–98%) healthy people. RDAs vary by age, gender and whether a woman is pregnant or breastfeeding.

Total Parenteral Nutrition (TPN) Nutrition administered into the body through the veins, bypassing the process of eating and digestion, used when patients cannot eat or absorb enough food through tube feeding formula or by mouth to maintain good nutrition status.

Upper Limit (UL) Also called the Tolerable Upper Intake level. The largest daily intake of a nutrient considered safe for most people. Intake of more than the UL may be harmful.

Introduction

In the presence of air and water, trivalent chromium, Cr^{3+}, is the stable form of chromium. This is the oxidation state relevant to human nutrition. Hexavalent chromium, Cr^{6+}, particularly in the form of chromate or dichromate, is also present in the environment, usually as a result of industrial activity. Chromate and dichromate are kinetically stable but thermodynamically unstable, which is responsible for their toxic, mutagenic, and carcinogenic properties when inhaled. Over the last thirty years, the status of chromium (as Cr^{3+}) has gone from being an essential element for which supplementation was suggested to being an essential trace element whose need would be met by a well-balanced diet to not being considered an essential element for lack of convincing evidence. Despite its current status and the lack of clear demonstration of any beneficial effects from chromium supplementation,

chromium-containing supplements annually generate hundreds of million dollars in sales as these supplements are available over the counter in numerous forms including pills, chewing gums, sports drinks, and nutrition bars.

Intake guidelines

The Institute of Medicine of the National Research Council of the Food and Nutrition Board of the National Academies of Science (US) established the first Estimated Safe and Adequate Dietary Intake (ESADDI) for chromium in 1980 at 50–200 μg/day for adults. Insufficient data existed for to set an Estimated Average Requirement (EAR) and consequently to develop a Recommended Daily Allowance (RDA). The ESADDI was left unchanged by the Council in 1989. However, after 1980 several studies found levels of chromium consumption for adults below 50 μg per day although equilibrium (positive net Cr balance) was achieved at these levels (Vincent and Brown, 2018).

Finally, in 2001, the Institute of Medicine reaffirmed insufficient evidence existed to set an EAR or consequently a RDA for chromium but found that the previous recommendations required modification. Consequently, AI of 35 μg per day and 25 μg per day for young men and women, respectively, were set based on estimated mean intakes (Institute of Medicine, 2001). (The AI is a replacement for the ESADDI). As an AI is more conservative than a RDA, the AI means that more than greater 98% of adult Americans would be chromium sufficient at ~30 μg chromium per day; hence, essentially all Americans are not chromium deficient. Thirty micrograms chromium per day is approximately the amount of chromium found in both nutritionist-designed and self-selected American diets, a function of the ubiquitous presence of chromium in foods at very low concentrations. More recently in 2014, the Panel on Dietetic Products, Nutrition and Allergies of the European Food Safety Authority determined "there is no evidence of beneficial effects associated with chromium intake in healthy subjects" and "the setting of an Adequate Intake for chromium is also not appropriate" (European Food Safety Authority, 2014).

The United States and Canada need to reexamine the status of chromium as the 2001 Institute of Medicine recommendations are outdated.

Lack of symptoms of deficiency

The reason that the chromium requirements and status of chromium have been modified with time is no symptoms of chromium deficiency have been definitively established (European Food Safety Authority, 2014). The most commonly cited support for chromium potentially being an essential element is case studies of patients who developed insulin resistance and received total parenteral nutrition (TPN) supplemented with chromium. However, the evidence is rather limited, being less than 10 case studies reporting beneficial effects (Vincent, 2017). In these case studies, subjects on TPN received ~5–16 μg of chromium per day before their TPN was supplemented with chromium. Given the average Adequate Intake of 30 μg chromium per day and assuming 1% of chromium is absorbed, an individual on a typical American diet would have ~0.3 μg of chromium enter the bloodstream daily. Thus, subjects (before the TPN was supplemented with chromium) were receiving several fold more chromium than should be necessary based on the AI and cannot be considered to have been chromium deficient. Subjects in the studies treated with TPN supplemented with chromium received an additional 12–250 μg of Cr per day. The supplemented diet clearly provided supranutritional, rather than nutritionally relevant, quantities of chromium. Additionally, care must be taken in interpreting these case studies as the conditions of the subjects before entering the studies varied considerably as did their time on TPN, the amount of chromium received, and several other variables.

Thus, no relationship can be established between the results of these case studies and chromium deficiency, although the studies could hint at a pharmacological effect in humans with altered glucose or carbohydrate metabolism (Vincent, 2017). In analyzing the results of studies of chromium supplementation of TPN patients, the European Food Safety Authority determined that "it is unclear on the basis of these case reports whether deficiency of chromium could be considered the only cause of glucose intolerance in these patients, whether deficiency of chromium has occurred in these patients, and whether chromium deficiency occurs in healthy populations" (European Food Safety Authority, 2014).

Chromium holds a similar status in Europe in terms of farm animal nutrition. The Panel on Additives and Products or Substances Used in Animal Feed (FEEDAP) in 2009 determined that "No symptoms of Cr deficiency in animals have been demonstrated in experimental conditions or observed in the field. The FEEDAP Panel considers that no evidence exists of essentiality of Cr(III) as trace element in animal nutrition and consequently no Cr^{3+} requirements could be established (EFSA Panel on Additives and Products or Substances Used in Animal Feed, 2009)."

Body mass

During the late 1980s and early 1990s, advertisements touting chromium, particularly as chromium picolinate, became commonplace in popular media, rocketing sales of chromium dietary supplements. However, because of the lack of "competent and reliable scientific evidence", the Federal Trade Commission (FTC) of the United States in 1997 ordered entities associated with the nutritional supplement chromium picolinate to stop making several representations (Federal Trade Commission, 1997). These representations include that chromium picolinate reduces body fat; causes weight loss; chromium picolinate causes weight loss without dieting or exercise; chromium picolinate causes long-term or permanent weight loss; chromium picolinate increases lean body

mass or builds muscle; chromium picolinate increases human metabolism; chromium picolinate controls appetite or craving for sugar; chromium picolinate reduces serum cholesterol; chromium picolinate lowers elevated blood sugar levels; chromium picolinate is effective in the treatment or prevention of diabetes; or 90% of US adults do not consume diets with sufficient chromium to support normal insulin function, resulting in increased risk of obesity, heart disease, elevated blood fat, high blood pressure, diabetes, or some other adverse effect on health.

A recent review of the effects of chromium on body mass and body composition concluded "the preponderance of experimental findings emphasizes the conclusion that Cr(III) supplementation, regardless of dosage (~200 to ~1000 μg/d), is neither a significant promoter of body mass loss nor a facilitator of body fat loss or body mass accretion of adults. Only studies with questionable experimental quality report beneficial changes in body composition after Cr(III) supplementation" (Lukaski, 2018).

Dietary sources

Chromium is present in essentially all foods but at very low concentrations. Dependable tables of chromium contents of foods are lacking as insufficient data exists to determine whether the chromium content of foodstuffs measured to date are truly reflective or dependent on locality and other variables. During the processing of food, particularly in stainless steel (normally ~10–40% chromium) equipment, the concentration of chromium increases; thus, the majority chromium in some foods may come from processing. Foods reported to rich in chromium (i.e., >100 ppb) include broccoli, black pepper, and certain beers; however, the reported contents for vegetables must be considered carefully as the amount of chromium coming from soil contamination can vary. The low concentrations of chromium in food, the ease of contamination, and the low suggested dietary requirement for chromium make preparation of a low-chromium diet difficult. Given the amount of chromium in the diet that comes from modern processing, the chromium intake of early humans should have been limited; thus, humans possibly could have evolved on a diet containing significantly less chromium than the current diets of people of developed nations (upon which the AI's are based).

Digestion

Understanding any potential biological effects from dietary chromium requires establishing how chromium is transported and distributed in the body. Approximately only ~1% of chromium ingested orally is absorbed, with the rest lost through the feces. Chromium is absorbed from the gastrointestinal tract by passive diffusion. Chromium supplements are believed to breakdown under the acidic conditions in the stomach, explaining the similar degree of absorption of chromium from organic complexes of chromium (e.g., chromium picolinate and chromium nicotinate) to that of chromium from inorganic chromium salts such as $CrCl_3$ (Vincent and Edwards, 2018). This also explains why supplements with varying solubilities ($CrCl_3$ is soluble in water, while chromium picolinate is only slightly soluble in water and chromium nicotinate is insoluble) have similar degrees of absorption. However, no studies have followed the forms of chromium present during digestion, although double radiolabeling experiments (^{1}H and ^{51}Cr) have shown that chromium and picolinate from chromium picolinate are rapidly divorced and have different fates in the gastrointestinal tract. Intestinal perfusate studies suggest while chromium is passively absorbed it could be actively transported out of the intestinal cells, as ~94% of the chromium entering the cells was rapidly cleared from the cells (leaving only ~6% behind to be stored) (Dowling et al., 1989). However, no membrane transport protein is known for chromium; thus, the mechanism for this potential transport is not known.

Transport

The transport of chromium after absorption has been an active area of research in recent years. Cr^{3+} is generally believed to be transported from the bloodstream to tissues by the iron transport protein transferrin. Cr^{3+} is proposed to be transported from the tissues back to the bloodstream for elimination via the urine by the peptide low-molecular-weight chromium-binding substance (LMWCr). Notably, questions have arisen about the proposals, and not all data appears to be consistent with these proposals. The movement of Cr^{3+} by transferrin followed by LMWCr has been proposed to function as a chromium detoxification pathway and to be potentially be associated with the pharmacological action of chromium at high doses (Vincent and Edwards, 2018).

When provided orally to rats to rats, over 80% of the chromium in the bloodstream is bound to transferrin. Under physiological conditions, transferrin selectively binds Fe^{3+}, as the metal-binding sites have evolved to bind metal ions with large charge-to-size ratios. Transferrin possesses two lobes, the N-terminal and C-terminal lobes, sharing a significant degree of sequence homology; each lobe contains a ferric ion-binding site. In humans, transferrin has the potential to bind metal ions other than Fe^{3+} as the protein is normally about 30% saturated with Fe^{3+}. The binding of manganese and chromium, as their trivalent ions, may be physiologically relevant (Vincent and Love, 2012).

Cr-transferrin movement has been followed using $^{51}Cr^{3+}$-labeled transferrin administered intravenously to rats. Introduction of the labeled transferrin initiates a rapid movement of chromium from the bloodstream into the tissues (Clodfelder and Vincent, 2005). More than 50% of the chromium is transported to the tissues within 30 min resulting in the amount of labeled chromium in the tissues being maximal ~30 min after injection. The movement from the bloodstream to the tissues is insulin-sensitive; plasma membrane recycling of transferrin receptors is sensitive to insulin as increases in insulin result in a stimulation of the movement of transferrin receptors from vesicles to the plasma membrane. The subsequent decreases in tissue chromium over time were

mirrored by increases in urine chromium. Consequently, transferrin, in an insulin-dependent fashion, can deliver chromium to tissues from which chromium is ultimately excreted in the urine. The removal of ^{51}Cr from the blood is faster than the appearance of ^{51}Cr in the urine; the lag in time indicates that the Cr-transferrin in the blood and chromium in the urine are not in direct equilibrium and that intermediates in the transport of chromium must be involved. One intermediate, LMWCr, has been identified, and recent studies of the binding and release of chromium from transferrin have suggested some of the intermediate steps.

Yet, the *in vivo* transport of chromium by transferrin has been questioned. The largest questions arise around the rate of chromium binding and release. Cr^{3+} has a d^3 electron configuration in its normal octahedral or pseudo-octahedral coordination environment. This makes exchange of ligands at the metal center slow compared to those of most other first row transition metal ions. Transferrin binds chromium concomitantly with an equivalent of (bi)carbonate, making the binding of chromium first order in the concentration of bicarbonate; in the presence of approximately 25 mM bicarbonate (as in the bloodstream), Cr^{3+} binds tightly to transferrin in minutes. In the absence of added bicarbonate, loading of transferrin with Cr^{3+} requires up to two weeks. Whether Cr^{3+} loss from transferrin is fast enough to be physiologically relevant as been a question. Transferrin delivers metal ions to cells via endocytosis. While release of Cr^{3+} from transferrin is slow, Cr^{3+} is rapidly released at the pH of endosomes (pH $\sim$ 5.5) from the transferrin/transferrin receptor complex (Edwards et al., 2020). How Cr^{3+} moves from the endosome to the rest of the cell is unknown; however, a recent study has shown at pH 5.5 that LMWCr binds Cr^{3+} $\sim$250-fold faster in the presence transferrin/transferrin receptor complex, making the peptide a candidate for the carrier of Cr^{3+} across the endosomal membrane.

LMWCr is an oligopeptide 10 or 11 amino acids in length composed of only four different amino acids (glutamate, aspartate, glycine, and cysteine) with the majority of amino acids having carboxylate side chains. The first seven residues have been cleaved from the rest, and the resulting peptide has been sequenced yielding the sequence EEEEGDD. How the remining glycine and cysteine residues are attached is unknown. LMWCr tightly and cooperatively binds four Cr^{3+} ions. Spectroscopic studies suggest that the chromic ions comprise an anion-bridged multinuclear assembly supported by carboxylates from the oligopeptide (Vincent and Edwards, 2018).

The intravenous injection of rats with $^{51}Cr(III)_2$-transferrin results in the appearance of LMWCr loaded with ^{51}Cr. LMWCr appears in the bloodstream where it the only form of Cr other than transferrin and in the urine, where it is the only Cr-containing species. Approximately 50% of the ^{51}Cr appeared in the urine within 6 h of injection of Cr-transferrin into the tail vein of rats in the absence of added insulin; insulin treatment concurrent with injection of ^{51}Cr-labeled transferrin results in approximately 80% of the label appearing in the urine within 180 min, an appreciable enhancement in the movement of Cr from the blood to the tissues to the urine (Clodfelder and Vincent, 2005). LMWCr appears to the mechanism by which Cr is bound in the tissues for elimination via the bloodstream and the urine.

Toxicity

No Upper Limit (UL) could be established for trivalent chromium ion in 2001 by the Institute of Medicine (Institute of Medicine, 2001); no adverse effects could be convincingly associated with excess intake of chromium from food or supplements. In turn in 2003, the Scientific Committee on Food (SCF) found insignificant evidence to set an UL. Although the number of studies available from which to draw data was limited, no evidence of adverse effects was found to be associated with supplemental chromium intake up to a dose of 1 mg per day; this recommendation was specifically for sources of Cr^{3+} other than Cr picolinate (Scientific Committee on Food, 2003). Also in 2003, the Food Standard Agency (United Kingdom) has determined that doses up to 10 mg of chromium daily should be safe for humans (Expert Group on Vitamins and Minerals, 2003).

In 2005, the US FDA found that use of chromium supplements was safe up to 1 mg chromium per day, the highest amount used in clinical trials (Food and Drug Administration, 2005). A Tolerable Daily Intake (TDI) of 300 μg chromium per kilogram body mass daily was derived from the lowest No Observed Adverse Effect Level (NOAEL) identified in a chronic oral toxicity study in rats by the EFSA Panel on Contaminants in the Food Chain in 2014 (CONTAM Panel) (EFSA Panel on Contaminants in the Food Chain, 2014).

The use of chromium supplementation is not without its concerns. For example, the use of supplemental chromium in TPN solutions has been questioned as subjects on TPN solutions providing 10 or more micrograms chromium per day accumulate chromium. As a result, the American Society for Parenteral and Enteral Nutrition (ASPEN) has made examining chromium requirements in parenteral nutrition their number one urgent priority (Vanek et al., 2012). To put the situation into perspective, TPN solutions currently provide 10–15 μg chromium per day in the United States. Using the average AI for men and women of 30 μg chromium per day and an $\sim$1% absorption efficiency for chromium, humans would have $\sim$0.3 μg chromium entering their bloodstream daily; as a result, TPN solutions would provide to the bloodstream $\sim$33–50 times more chromium than a standard diet. ASPEN has recommended that TPN without added chromium should be made available although no reported cases of chromium toxicity in patients on long term TPN are known.

While no significant health concerns have been clearly identified for commercial chromium supplements at current doses, questions about the use of chromium, in particular chromium picolinate, have been raised. The redox potential of Cr(III) shifts when bound to ligands binding through imine nitrogen atoms. Compounds with 4 or more imine nitrogens bound to Cr(III) are susceptible to undergoing oxidation to deleterious Cr(IV) and Cr(V) species, making the compounds clastogenic and toxic. Chromium picolinate with three coordinated imine nitrogens has a redox potential near the borderline for being susceptible to oxidation. Cell culture study and *in vivo* studies using mammals where the compound was administered intravenously or using fruit flies

suggest it is toxic and clastogenic, mutagenic, and possibly carcinogenic (Vincent, 2013). Yet, a study commissioned by the National Toxicology Program (National Institutes of Health) has found that providing chromium picolinate up to 5% of the diet of male and female rats and mice for 2 years had no conclusive deleterious health effects (Stout et al., 2009). This discrepancy between the studies probably arises from chromium picolinate dissociating in the gastrointestinal tract when provided orally. In the cell culture study and *in vivo* studies, the chromium picolinate molecule potentially reached cells intact (Vincent, 2013). Consequently, the supplement should not be used intravenously. The potential for deleterious effects probably arises from its redox potential compared to other current forms of supplemental chromium, which could possibly allow the chromic center of chromium picolinate to undergo redox chemistry *in vivo*.

Chromium supplementation and human health

Currently chromium is not recommended for any clinical uses. Simply, no consistent dose-response relationships between chromium and a beneficial health outcome in humans have been established. One issue is that clinical studies have focused primarily on type 2 diabetic subjects, tend to have small subject pools, and tend not be well designed. Agencies of the United States government have commissioned two meta-analyses on whether chromium affects symptoms of type 2 diabetes; both generated inconclusive results. In 2002 with funding from the Office of Dietary Supplements of the National Institutes of Health, only four quality studies were identified for analysis (Althuis et al., 2002). The second, published in 2007 and funded by the Department of Health and Human Services, identified 18 studies. The authors concluded Cr supplementation "may have a modest effect" on glucose metabolism in type 2 diabetics but that "large heterogeneity and the overall poor quality limit the strength of our conclusions" (Balk et al., 2007). Unfortunately, the positive effects of greatest magnitude used in the analysis came from the 12 studies ranked lowest in quality. Additionally, a trend was observed that commercial industry-sponsored studies were more likely to observe beneficial effects. A thorough review of the effects of chromium in human disease, focusing on recent meta-analyses, determined "the data from meta-analyses and [randomized clinical trials] utilizing Cr interventions continue to be mixed compared to more effective results with certain other therapies for the same diseases" as "the quality of the [randomized clinical trials] in the existing meta-analyses left much to be desired" (Costello et al., 2018).

The US Food and Drug Administration (FDA) received a petition in December 2003 from the nutraceutical company Nutrition 21 for eight qualified health claims. The claims were chromium picolinate may reduce the risk of insulin resistance; chromium picolinate may reduce the risk of cardiovascular disease when caused by insulin resistance; chromium picolinate may reduce abnormally elevated blood sugar levels; chromium picolinate may reduce the risk of cardiovascular disease when caused by abnormally elevated blood sugar levels; chromium picolinate may reduce the risk of type 2 diabetes; chromium picolinate may reduce the risk of cardiovascular disease when caused by type 2 diabetes; chromium picolinate may reduce the risk of retinopathy when caused by abnormally high blood sugar levels; and chromium picolinate may reduce the risk of kidney disease when caused by abnormally high blood sugar levels. At the conclusion of its review, the FDA issued a letter of enforcement discretion allowing only one qualified health claim for the labeling of dietary supplements stating "One small study suggests that chromium picolinate may reduce the risk of type 2 diabetes. FDA concludes that the existence of such a relationship between chromium picolinate and either insulin resistance or type 2 diabetes is highly uncertain" (Food and Drug Administration, 2005).

In recent years, the American Diabetes Association has been consistent in its position: "there is insufficient evidence to support the routine use of herbal supplements and micronutrients, such as … chromium, to improve glycemia in people with diabetes" and "There is no clear evidence that dietary supplementation with vitamins, minerals (such as chromium and vitamin D), herbs, or spices … can improve outcomes in people with diabetes who do not have underlying deficiencies, and they are not generally recommended for glycemic control" (American Diabetes Association, 2021).

Most recently, case studies examining intravenous infusions of chromium (generally 3 μg/h) as a treatment for glucose intolerance found chromium administration reduced insulin requirements for subjects with hyperglycemia. These studies may be suggestive of a pharmacological effect of chromium at supra-nutritional levels. However, drawing any firm conclusions is not possible based on the small number of subjects to date (Vincent, 2017). These case studies along with observation beneficial effects of chromium supplementation in rodent studies require supra-nutritional doses of chromium could suggest beneficial effects from chromium supplementation in humans are possible at higher doses than have been used to date in clinical trials. However, clinical trials using >10 mg chromium per day will be required before any definitive conclusions regarding potential beneficial effects of pharmacological doses of chromium in humans could be reached other than that no clinical effects are reliably and reproducibly reported at doses of 1 mg chromium per day.

Pharmacological effects in rodents

Rodent models of diabetes and insulin resistance have been utilized to examine the effects of chromium supplementation (Vincent, 2014; Krejpcio, 2018). Studies suffer from the use of varying chemical forms of chromium supplements, the range of dosages utilized, the range of duration of treatment, and the numbers of different animal models examined.

While still preliminary, dividing these studies by the type of model utilized appears to be useful. Three models examined develop peripheral insulin resistance arising from related mutations; the JCR:LA-cp, Zucker obese, and Zucker diabetic fatty rats all have

mutations of the leptin receptor generating peripheral insulin resistance. When chromium is administered at a young age to these models, it generally has no effect on body mass and food intake. Chromium administration has little, if any, effect on fasting blood glucose levels but lowers glycated hemoglobin levels. Chromium administration also appears generally to be beneficial to lipid metabolism, lowering total cholesterol levels; however, effects on other lipid variables are inconsistent. In contrast, in Goto-Kakizaki rats, possessing hepatic insulin resistance opposed to peripheral insulin resistance, chromium supplementation does not lead to any significant effects. These conflicting results between the two types of genetic models might suggest the effects of chromium supplementation are primarily derived from the skeletal muscle (Vincent, 2014).

The effects of chromium supplementation of chemically induced insulin resistance models, most notably rats treated with streptozotocin or alloxan, have also studied. The results of chromium supplementation are too inconsistent to allow firm conclusions to be drawn. Finally, dietary models have also been examined. Chromium administration to rats on high-sugar diets results in increased insulin sensitivity. Similarly rats on high-fat diets generally have increased insulin sensitivity, although for both diets conflicting results have been observed. Several studies have looked at the effects of chromium supplementation on rats on high-fat diet that were also treated with streptozotocin. While these studies are similar in observing reduced cholesterol and triglycerides levels, other results, most notably fasting glucose and insulin levels and body mass changes, are not consistent (Vincent, 2014; Krejpcio, 2018).

These rodent studies have ultimately utilized pharmacologically, rather than nutritionally, relevant doses of chromium. This was demonstrated in a seminal study of "low-chromium" diets (Di Bona et al., 2011). Rats were provided the AIN-93G purified diet with no added chromium in the mineral mix (elemental analysis revealed less than 20 μg of chromium per kilogram of diet, the lowest quantity of chromium in a rat diet examined to date) and were kept in cages with no access to metal for six months. (The chromium content of this diet is similar to that of a human consuming the AI of 30 μg chromium daily). No differences in body mass, insulin sensitivity, or response to a glucose challenge were observed compared to rats on the complete AIN-93G diet (containing 1000 μg chromium per kilogram of diet). Adding additional Cr^{3+} to the diet (200 or 1000 μg/kg) also had no effects on body mass but resulted in increased insulin sensitivity as measured by reduced areas under the curve for insulin concentrations in glucose tolerance tests. Consequently, insulin sensitivity was shown to increase as a function of added chromium but required supra-nutritional doses. In fact, this study suggests rats on the AIN-93G are already obtaining enough chromium to have beneficial effects on insulin sensitivity from supplementation. Thus, the basal chromium levels in rodent diets must be considered when designing experiments, and effects from supplemental Cr could be difficult to detect on top of beneficial effects from high chromium content in the basal diets. The Cr content of the AIN-93 diets needs to be reexamined, and Cr probably should be removed from the mineral mix.

Molecular mechanism(s) of action

Several pathways have been proposed by which a chromium-containing biomolecule could manifest effects on insulin sensitivity. Research results both in *in vitro* and *in vivo* systems are contradictory, such that the state of the field is far from clear (Vincent, 2013). Attention has been focused on two mechanisms: the insulin signaling cascade, most specifically insulin receptor and Akt (protein kinase B), and on cholesterol synthesis and metabolism. Recently, mechanisms involving oxidative stress and cytokine generation have also been suggested (Nair, 2018).

One major concern with many *in vitro* studies is the forms of chromium utilized. Most commercial chromium supplements are believed to degrade in the gastrointestinal tract, so that the cells of the body never encounter the chromium-containing molecule comprising the supplement. For example, the best-selling supplement, chromium picolinate, is apparently toxic to cells *in vitro* but harmless at nutritional doses when given orally. Thus, the effects of adding the commercial supplements directly to cultured cells could have drastically different effects on insulin sensitivity, and the mechanisms thereof, than those of chromium *in vivo*, as presumably cells would normally be exposed to chromium as a complex with transferrin. The use of transferrin, which appears to be the form in which Cr^{3+} is primarily transported into cells, would appear optimal, but no studies using Cr-containing transferrin to look for effects on insulin signaling in cultured cells have been reported. Additionally, control experiments using the free ligands, e.g., picolinic acid, are not often performed, leaving in doubt whether any observed effects arise from chromium, the chromium complex, or the free ligand.

In vivo studies have suggested potential targets for the origins of the effects of chromium at a molecular level; yet, results are not consistent, perhaps in part, because different chromium compounds in different doses in different animals and models have been utilized. One of the most notable studies because of its design suggested a role of chromium directly in the insulin signaling pathway. Cefalu and coworkers (Wang et al., 2006) observed increased phosphorylation of insulin receptor substrate-1 (IRS-1) tyrosine and kinase activity of IRS-1–associated phosphoinositol-3 (PI-3) in skeletal muscle after insulin stimulation in obese JCR:LA-cp rats administered chromium as chromium picolinate compared with obese controls. The enhanced signaling occurred without changes in the concentration of IRS-1, Akt, or PI-3 kinase; however, levels of phosphotyrosine phosphatase 1B, (PTP 1B) that dephosphorylates insulin receptor, were reduced. Reproduction of this study, potentially with another diabetes model, could be most significant to the field. Sreejayan and coworkers (Dong et al., 2008) found supplementation of sucrose-fed mice with chromium tris(phenylalanine) significantly enhanced insulin-stimulated Akt phosphorylation and membrane-associated glucose transporter-4 (Glut-4) in skeletal muscles. Cefalu and coworkers (Cefalu et al., 2002) also have found increased insulin-stimulated membrane-associated Glut-4 in JCR:LA-cp rats administered chromium as chromium picolinate. Yet, other studies

are not consistent with either of these sets of results, including observing changes in protein phosphorylation without insulin stimulation, observing changes in protein concentration, and failing to observe changes in PTP 1B levels or activity; however, a consensus does seem to appear that chromium supplementation augments insulin-dependent Glut-4 translocation under insulin resistant conditions (Hua et al., 2012). Increases in insulin-dependent Glut-4 translocation have also been observed in several *in vitro* studies as well and appears to be the most consistently observed effect from chromium administration.

Conclusion

Currently, evidence does not support chromium being an essential trace element so that setting dietary guidelines is not appropriate. However, dietary chromium does not appear to have any toxic effects. Despite the lack of toxicity, chromium supplementation is not recommended for healthy individuals or individuals with any medical condition. Randomized clinical trials using higher doses of chromium than in current nutritional supplements potentially are required to determine whether chromium might have beneficial effects at supra-nutritional doses.

Conflict of interest

The author has no known conflict of interest to acknowledge.

References

Althuis, M.D., Jordan, N.E., Ludington, E.A., Wittes, J.T., 2002. Glucose and insulin responses to dietary chromium supplements: a meta-analysis. Am. J. Clin. Nutr. 76, 148–155.

American Diabetes Association, 2021. 5. Facilitating behavior change and well-being to improve health outcomes: Standards of Medical care in Diabetes—2021. Diabetes Care 44 (Suppl. 1), S53–S72.

Balk, E.M., Tatsioni, A., Lichenstein, A.H., Pittas, A.G., 2007. Effect of chromium supplementation on glucose metabolism and lipids: a systematic review of randomized clinical trials. Diabetes Care 30, 2154–2163.

Cefalu, W.T., Wang, Z.Q., Zhang, X.H., Baldor, L.C., Russell, J.C., 2002. Oral chromium picolinate improves carbohydrate and lipid metabolism and enhances skeletal muscle Glut-4 translocation in obese, hyperinsulinemic (JCR-LA corpulent) rats. J. Nutr. 132, 1107–1114.

Clodfelder, B.J., Vincent, J.B., 2005. The time dependent transport of chromium in adult rats from the bloodstream to the urine. J. Biol. Inorg. Chem. 6, 383–393.

Costello, R.B., Dwyer, J.T., Merkel, J.M., 2018. Chromium supplements in health and disease. In: Vincent, J.B. (Ed.), The Nutritional Biochemistry of Chromium(III), second ed. Elsevier, Amsterdam, pp. 219–249.

Di Bona, K.R., Love, S., Rhodes, N.R., McAdory, D., Halder Sinha, S., Kern, N., Kent, J., Strickland, J., Wilson, A., Beaird, J., Ramage, J., Rasco, J., Vincent, J.B., 2011. Chromium is not an essential trace element for mammals: effects of a "low-chromium" diet. J. Biol. Inorg. Chem. 16, 381–390.

Dong, F., Kendadi, M.R., Ren, J., Sreejayan, N., 2008. Chromium (D-phenylalanine)$_3$ supplementation alters glucose disposal, insulin signaling, and glucose transporter-4 membrane translocation in insulin-resistant mice. J. Nutr. 138, 1846–1851.

Dowling, H.J., Offenbacher, E.G., Pi-Sunyer, F.X., 1989. Absorption of inorganic, trivalent chromium from the vascularly perfused rat small intestine. J. Nutr. 119, 1138–1145.

Edwards, K.C., Kim, H., Vincent, J.B., 2020. Release of trivalent chromium from serum transferrin is sufficiently rapid to be physiologically relevant. J. Inorg. Biochem. 202, 110901.

EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP), 2009. Safety and efficacy of Availa® Cr (chromium chelate of DL-methionine) as a feed additive for all species. EFSA J. 1043, 1–69.

EFSA Panel on Contaminants in the Food Chain, 2014. Scientific opinion on the risks to public health related to the presence of chromium in food and drinking water. EFSA J. 12, 3595.

European Food Safety Authority, 2014. Scientific opinion on dietary reference values for chromium. EFSA J. 12, 3845.

Expert Group on Vitamins and Minerals, 2003. Safe Upper Levels for Vitamins and Minerals. Food Standards Agency, London.

Federal Trade Commission, 1997. Docket No. C-3758 Decision and Order, 1997. https://www.ftc.gov/sites/default/files/documents/cases/1997/07/nutrit2.htm. (Accessed 25 February 2021).

Food and Drug Administration, 2005. Qualified Health Claims: Letter of Enforcement Discretion—Chromium Picolinate and Insulin Resistance (Docket No. 2004Q0144). http://wayback.archive-it.org/7993/20171114183739/https://www.fda.gov/Food/IngredientsPackagingLabeling/LabelingNutrition/ucm073017.htm. (Accessed 11 March 2021).

Hua, Y., Clark, S., Ren, J., Sreejayan, N., 2012. Molecular mechanisms of chromium in alleviating insulin resistance. J. Nutr. Biochem. 23, 313–319.

Institute of Medicine, 2001. Dietary Reference Intakes for Vitamin A, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc. National Academies Press, Washington, DC.

Krejpcio, Z., 2018. Effects of chromium(III) supplementation on rodent models of insulin resistance and diabetes. In: Vincent, J.B. (Ed.), The Nutritional Biochemistry of Chromium(III), second ed. Elsevier, Amsterdam, pp. 195–218.

Lukaski, H.C., 2018. Effects of chromium(III) as a nutritional supplement. In: Vincent, J.B. (Ed.), The Nutritional Biochemistry of Chromium(III), second ed. Elsevier, Amsterdam, pp. 61–77.

Nair, S., 2018. Metabolic effects of chromium—potential molecular mechanisms. In: Vincent, J.B. (Ed.), The Nutritional Biochemistry of Chromium(III), second ed. Elsevier, Amsterdam, pp. 175–191.

Scientific Committee on Food, 2003. Opinion on the Scientific Committee on Food on the Tolerable Upper Intake Level of Trivalent Chromium. SCF/CS/NUT/UPPLEV/67 Final. https://ec.europa.eu/food/sites/food/files/safety/docs/sci-com_scf_out197_en.pdf. (Accessed 11 March 2021).

Stout, M.D., Nyska, A., Collins, B.J., Witt, K.L., Kissling, G.E., Malarkey, D.E., Hooth, M.J., 2009. Chromic toxicity and carcinogenicity studies of chromium picolinate monohydrate administered in feed to F344/N rats and B6C3F1 mice for 2 years. Food Chem. Toxicol. 47, 729–733.

Vanek, V.W., Borum, P., Buchman, A., Fessler, T.A., Howard, L., Jeejeebhoy, K., Kochevar, M., Shenkin, A., Valentine, C.J., 2012. A.S.P.E.N. position paper: recommendation for changes in commercially available multivitamin and multi-trace element products. Nutr. Clin. Pract. 27, 440–491.

Vincent, J.B., Brown, S., 2018. Introduction: a history of chromium studies (1955–2007). In: Vincent, J.B. (Ed.), The Nutritional Biochemistry of Chromium(III), second ed. Elsevier, Amsterdam, pp. 1–58.

Vincent, J.B., Edwards, K.C., 2018. The absorption and transport of chromium in the body. In: Vincent, J.B. (Ed.), The Nutritional Biochemistry of Chromium(III), second ed. Elsevier, Amsterdam, pp. 129–174.

Vincent, J.B., Love, S., 2012. The binding and transport of alternative metals by transferrin. Biochim. Biophys. Acta 1820, 362–378.
Vincent, J.B., 2013. The Bioinorganic Chemistry of Chromium. John Wiley & Sons, Chichester.
Vincent, J.B., 2014. Is chromium pharmacologically relevant? J. Trace Elem. Med. Biol. 28, 397–405.
Vincent, J.B., 2017. New evidence against chromium as an essential trace element. J. Nutr. 12, 2212–2219.
Wang, Z.Q., Zhang, X.H., Russell, J.C., Hulver, M., Cefalu, W.T., 2006. Chromium picolinate enhances skeletal muscle cellular insulin signaling in vivo in obese, insulin-resistant JCR:LA-cp rats. J. Nutr. 136, 415–420.

Cofactors: Inorganic

E.D. Harris, Texas A&M University, College Station, TX, USA

© 2013 Elsevier Ltd. All rights reserved.
This is an reproduction of E.D. Harris, Cofactors: Inorganic, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 357–365, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00057-X.

Introduction 189
History 190
Macro- Versus Microminerals 190
Metal-activated Versus Metalloenzymes 190
Individual Metal Cofactors 192
Macrominerals 192
Sodium 192
Potassium 192
Magnesium 192
Calcium 192
Micro (Trace) Minerals 192
Iron 192
Zinc 194
Copper 195
Manganese 196
Cobalt 197
Vanadium 197
Molybdenum 197
Nickel 197
Other 198
Nonmetal Mineral Cofactors 198
Selenium 198
Reactivity 198
Silicon 198
Reactivity 199
Boron 199
Conclusions 199
Further reading 199

Glossary

Carbonium ion A carbon ion with a positive charge.
Chloroplasts Light energy sequestering organelles in photosynthetic plants.
Cytosol The internal milieu of a cell.
Endoplasmic reticulum An internal cell compartment enclosed by membranes.
Hemoglobin The major iron protein in the blood.
Noncarbonaceous Not having the element carbon in the structure.
Peroxisomes An internal compartment in a cell.
Ruminants Cattle and such where a four-chambered rumen replaces the stomach.
Surrogate metal A metal that may act as a substitute.
Transferrin A protein that transfers iron to the tissues.
Transition series Belonging to a certain division of the Periodic Table of Elements.

Introduction

The word 'inorganic' implies a noncarbon component. Such is true of inorganic cofactors, which consist mainly of metal ions and noncarbonaceous components. More importantly, one-third of all the enzymes require an inorganic cofactor for function. Their role

is by no means minor. Enzymes that exchange electrons between substrates or quench dangerous free radicals or reactive oxygen species use metal ion cofactors to lessen the risk of irreversibly modifying the structure of the enzyme. Some cofactors perform mainly structural roles; for example, stabilizing the overall shape of the enzyme so it is poised to act on a substrate. Its more common, however, for the cofactor to engage the substrate directly and assist the catalysis of the ensuing reaction. More recently attention has turned to metal ions as modulators of genetic expression. These understandings have promoted the view that cofactors may be more than just passive components in biological systems. Here, we define a cofactor as any nonenzyme component that promotes the catalytic prowess of an enzyme. The definition emphasizes function rather than structure. This article will expand the discussion on enzyme cofactors by focusing on the properties of their 'inorganic' counterparts.

History

The nutritional history of the mineral elements, unlike the vitamins, had an early focus on domestic livestock foraging on mineral-poor soils. Typical symptoms were the crimping of wool in sheep, aortic rupture in pigs and cattle, and decrease in myelin in brains of newborn lambs. Symptoms were lessened sharply by supplementing the feed with salts of metal ions such as $CuSO_4$, $Fe(NH_4)_2(SO_4)_2$, $ZnCl_2$. Reversing symptoms and reestablishing optimal growth to livestock provided the first solid evidence for the essentiality of metals in nutrition. Thus, no longer were metal ions in the blood and tissues of animals and humans considered a sign of toxic exposure. With this view instead came the realization that metals can be indispensable elements to the animal's health. An early study by Hart *et al*. showed that rebuilding hemoglobin to normal levels in an anemic rat required both iron and copper. Others showed a similar copper–iron interaction in humans. Coupled with the advent of semipurified diets in that same era, the science of nutrition made major steps forward in defining the importance of metal ions in biological systems and many of the studies tended to focus on their role with enzymes. The past decades have witnessed a tremendous interest in metal ions as cofactors and as regulators of enzyme activity in tissues and organs. These studies have focused not only what they do but also on how the system is able to adroitly handle potentially dangerous nutrients present in macro and micro quantities and ensure that their functionality and the organism's safety are realized.

Macro- Versus Microminerals

Of the 27 nutritionally essential elements in the Periodic Table, nearly half are minerals, more specifically metal ions. Because they vary in quantity both in the diet and within the organism, minerals are divided into two major classes, the macro- and microminerals. A list of some of the more prominent ones in each class is shown in **Table 1**. Macrominerals, which include sodium, potassium, magnesium, and calcium tend to be present in larger quantities both in the diet and the tissues. As such they perform functions that are attuned to their bulk; functions such as regulating osmotic balance, or forming the matrix of the skeleton, or providing high energy gradients to drive membrane transport. In contrast, the microminerals as their name implies are present in very small amounts in the tissues and have a much lower requirement for adequacy. Despite their scarcity, as a class the microminerals make up the bulk of inorganic cofactors. The reason for this will be made clear below (*see* Metal-activated Versus Metalloenzymes). This class, which is also referred to as trace elements or trace metals, includes iron, zinc, copper, manganese, etc. (**Table 1**). All of those mentioned in the table are known to perform cofactor functions with enzymes.

In determining their role with enzymes, it is important to realize that most of the microminerals make up the 3*d* transition series of elements in the Periodic Table of Elements. As transition metals, they can exist in multiple valence states and are capable of forming specific geometric complexes with proteins. More importantly, perhaps, the multivalency common with many of the micro metal ions permits them to exchange electrons between substrate and cofactor. Nonmetal microminerals such as selenium also has this property. In contrast nearly all metal ions in the macromineral category (K^+, Ca^{2+}, Mg^{2+}) are monovalent and are not capable of donating or accepting electrons (**Table 1**).

Another important property of transition metals is the ability to bind firmly to the structure of the enzyme. Binding is generally through coordinate covalent bonds and exhibits a specific geometric pattern (**Figure 1**). The precision of the angles and distances in the binding pocket are a basis for pairing a particular metal with the enzyme. Macrominerals have more of a tendencey to engage an enzyme by ionic attraction and thus their binding to the enzyme surface is much weaker. The differences in these two properties carries over to classifying inorganic cofactors on the basis of activation by the metal or having the metal be an integral part of the enzyme's structure.

Metal-activated Versus Metalloenzymes

Freedom of movement carries over to the strength of binding when the mineral (as a biometal) engages the structure of an enzyme. Because biometals in the macromineral class tend to form bonds that are easily broken, the biometal exists in a state of equilibrium with the enzyme. Enzymes that use weakly bound cofactors are referred to as metal-activated; a term that signifies the enzyme can be primed to greater activity by adding its metal ion cofactor. Because the metal cannot be bound in a more permanent way, metal-activated enzymes typically lose activity during purification. An example is pyruvate kinase, which has a specific requirement for K^+

Table 1 Examples of inorganic cofactors

	Stable biological form or valence
Macrominerals	
Sodium	Na^{+}
Potassium	K^{+}
Calcium	Ca^{2+}
Magnesium	Mg^{2+}
Chloride	Cl^{-}
Phosphate	HPO_4^{2-}
Microminerals	
Iron	Fe^{2+}, Fe^{3+}
Zinc	Zn^{2+}
Copper	Cu^{+}, Cu^{2+}
Manganese	Mn^{2+}, Mn^{4+}, Mn^{5+}
Cobalt	Co^{+}, Co^{2+}, Co^{3+}
Molybdenum	MoO_2^{2+}, MoO_4^{2-}
Nickel	Ni^{+}, Ni^{2+}

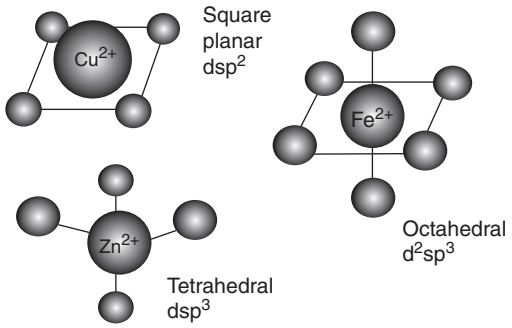


Figure 1 Some common geometries of metal complexes.

and is inactivated by dialysis (diffusion through a semiporous membrane). Another example is the class of enzymes referred to as kinases. These enzyme transfer the terminal phosphate group from ATP to the substrate, a reaction that requires Mg^{2+} to activate the substrate (Mg^{2+}–ATP).

In contrast, enzymes that have the metal ion bound firmly are referred to as metalloenzymes. Such enzymes have the metal ion bound mainly through coordinate covalent bonds. Because of the tight binding the metal ion is basically a firm fixture of the protein's structure. With few exceptions, metals in the micromineral class fit into the picture as cofactors for metalloenzymes. Tight binding precludes loss of the metal ion by dialysis or loss to weakly dissociating agents. Metal ion chelators with a strong affinity for the metal, however, can out-compete the enzyme for the metal ion and render the enzyme inactive. As prosthetic groups, metals in metalloenzymes have a stoichiometric relationship (metal ion–enzyme protein ratio expressed as a whole integer) with the enzyme and are seldom primed to greater activity by adding more metal ion to the enzyme. Spatial geometry is also a concern. Examples of the more common geometrical arrangements are shown in **Figure 1**. For metals in the first transition series one takes note of the 3*d* orbitals. Examples of both classes of enzymes are shown in **Table 2**.

An exception to note in the table is selenium. Because selenium has properties similar to sulfur, selenium can replace sulfur in the structure of amino acids that make up the enzyme's structure. Thus, when functioning as a cofactor, selenium is present as selenocysteine and selenomethionine and not as elemental selenium coordinated to the protein structure. Another exception to note is in the enzyme *Cu_2, Zn_2 superoxide dismutase*. On rare occasions an enzyme may require more than one metal ion to perform catalysis. Other than those with Zn, enzymes and proteins with first transition series metal ions tend to be highly colorful. Consider, for example the beautiful red color of hemoglobin (iron), the green of chlorophyll or the blue color of ceruloplasmin (whose name means heavenly blue) associated with copper. **Table 2** gives some examples of metalloenzymes and the specific metal each has bound to the structure.

Table 2 Metal-activated enzymes and metalloenzymes

Metal or metal cofactor	*Enzyme*	*Function*
Metal-activated enzymes		
K^+	Pyruvate kinase	Synthesize pyruvate
Mg^{2+}	Hexokinase	Phosphorylate glucose
–	DNAase	Cleave DNA
–	RNAase	Cleave RNA
–	ATPase	Cleave ATP
Metalloenzymes		
Cu^{2+}, Zn^{2+}	Superoxide dismutase	Destroy superoxide anion
Fe	Catalase	Destroy H_2O_2
Zn	Alcohol dehydrogenase	Metabolize alcohol
–	DNA polymerase	Synthesize DNA
Mn	Pyruvate carboxylase	Synthesize oxaloacetate
–	Arginase	Synthesize urea
Ca	Alpha amylase	Cleave glycogen, starch

Individual Metal Cofactors

Macrominerals

Although their presence in the diet and within the system far exceeds that of the microminerals, the macrominerals as a whole are not the category of abundance when considering enzyme cofactors. Examples of some of the more familiar are:

Sodium

As a monovalent ion, Na^+ is generally not considered a cofactor because one has yet to demonstrate an enzyme whose catalysis depends strictly on sodium ions. Sodium-activated enzymes often respond to surrogate metal cofactors such as Li^+ or even divalent cations. Sodium ions, however, form a major class as cotransporters for a series of transport proteins referred to as solute-linked carriers. Working with transporters for amino acids and sugars, sodium ion gradients across the membrane provide the driving energy for movement of amino acids, monosaccharides, etc. into cells.

Potassium

The potassium ion (K^+) makes a rare appearance as a specific cofactor for *pyruvate kinase* in the glycolysis pathway. Both potassium and magnesium form no permanent bonds with their respective enzymes and hence act more as activators.

Magnesium

Magnesium ions (Mg^{2+}) are required by a large number of enzymes referred to as *kinases*. Kinases transfer the terminal phosphate group of ATP to substrates. They figure prominently in many biochemical pathways such as glycolysis (*hexokinase*, *fructose-6-phosphate kinase*, *pyruvate kinase)*, hormone responses mediated by cyclic AMP, cell signaling, and regulation of cell division. Mg^{2+} also modulates muscle contraction by competing with Ca^{2+} on proteins that trigger muscle contraction.

Calcium

As a group IIa metal ion, Ca^{2+} is limited to a +2 valence state and serves primarily as a divalent cation in its interactions with enzymes. The role of Ca^{2+} is limited mainly to structure stability although it is a cofactor for a limited number of important enzymes apart from the more familiar actin–myosin complex in muscle. *Alpha amylase* and *thermolysin* are two of the most familiar. As a free ion or working through calmodulin, calcium is better understood as an activator of enzymes in hormone-dependent cell signaling pathways. Enzymes, which have been referred to as Ca-ATPases and H^+/Ca-ATPases are not to be mistaken as calcium-dependent. This is a misnomer in that the Ca^{2+} is the object of the enzyme's action rather than the cofactor for activity. The ATPases comprise a large group of membrane-bound enzymes that either pump Ca^{2+} from the cytosol into the endoplasmic reticulum or expel Ca from the cell through membrane channels.

Micro (Trace) Minerals

Iron

Most iron enzymes engage iron either as heme or as a special arrangement of iron with sulfur groups referred to as iron–sulfur centers (Fe_nS_n) (Table 3). Iron in heme bears a striking resemblance to magnesium ion in chlorophyll (Figure 2). Heme, basically a porphyrin ring system with iron positioned in the center, is the most common form of iron in biological proteins. In cytochrome *c*, a common heme protein in the mitochondria, the axial ligands to the iron are occupied by histidine and methionine from the protein. Heme enzymes include *calalase* and *peroxidase*. As components of iron–sulfur centers, iron enters into multiple cluster

Table 3 Important iron enzymes

Enzyme	*Source*	*Function*	*Form of Fe*
Cytochrome *c* oxidase	Mitochondria	Electron Transport	Heme
Aconitase	Mitochondria	Krebs Cycle	Fe_4S_4
Succinate dehydrogenase	Mitochondria	Krebs Cycle	Fe_4S_4
Catalase	Peroxisomes	H_2O_2 destruction	Heme
Peroxidase	Peroxisomes	Peroxide destruction	Heme
Prolyl hydroxylase	Cytosol	Collagen Synthesis	Fe^{2+}
Ribonucleotide reductase	Cytosol	DNA Synthesis	Fe–O–Fe
Cytochrome P450	Microsomes	Sterol Synthesis	Heme

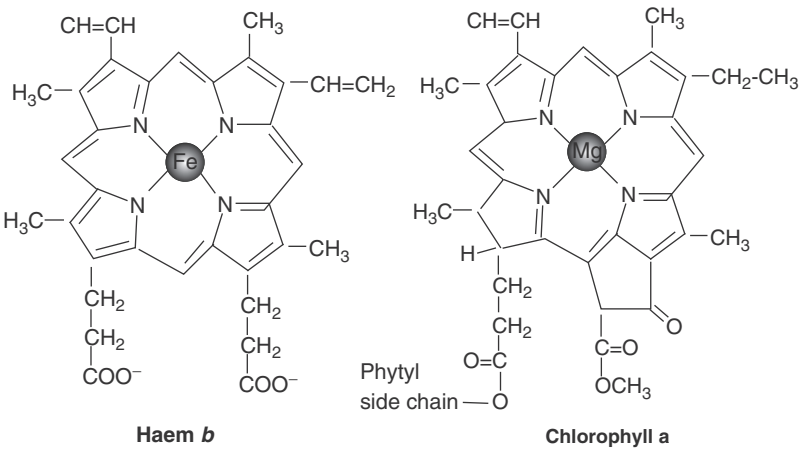


Figure 2 Heme iron in hemoglobin. Heme is a porphyrin ring with iron in the center. Four heme b groups are present in hemoglobin, the iron protein in erythrocytes. A similar structural arrangement is seen with magnesium in chlorophyll *a* from plants.

arrangements with cysteine residues on enzymes that offer a more direct contact with the protein. These centers differ in their complexity from the simple 2Fe–2S to the more elaborate 4Fe–4S (Figure 3). Iron in these centers binds substrates as well as transfer electrons and takes part in reactions involving dehydrations and rearrangements. Enzymes with iron–sulfur centers include *xanthine oxidase, succinate dehydrogenase, aconitase, and nitrogenase*. A third class, represented by *ribonucleotide reductase* has an FeO_2 cluster with a dioxygen as a peroxide-anion O_2^{2-} straddled between two iron centers (Figure 4). This arrangement allows the enzyme to remove a hydrogen atom from a very stable C–H bond. No metal can replace iron in these complexes. Enzymes with a heme group generally are reddish-brown in color (depending on the oxidation state of the iron). The color led to early interest in these proteins and was the motivating factor behind naming heme proteins in the mitochondria 'cytochromes'. Although only a relatively few soluble enzymes have iron as a cofactor, iron is especially prominent in membrane-bound proteins that comprise electron transport pathways. Examples of the latter include the cytochromes in the mitochondria, endoplasmic reticulum, and photosystems I and II in chloroplasts. Perhaps the most unusual iron protein is ferritin, a huge multisubunit iron storage protein that has the capacity to bind more than 2500–5000 iron atoms in its structure.

Reactivity

The redox property of iron carries over to much of its chemistry as a cofactor. Iron is nearly always involved with the transfer of electrons and many times donates the electrons to a molecule of oxygen. Two important properties that fit that role are: (1) an iron atom can readily undergo reversible valence changes from Fe^{2+} to Fe^{3+}, which allows facile exchange of electrons, and (2) the ferrous–ferric ion pair has a relatively low electrochemical potential (−.1 V), which allows iron to be on the high (reducing) end of a electron transport chain. In cytochrome P450 a single oxygen atom is transferred to the substrate after O_2 binds to Fe(II). In the mechanism the Fe(II)–O_2 complex is converted into FeO, which features an Fe(V) species that attacks the substrate and incorporates the single oxygen atom into its structure. Although higher valence states such as Fe(IV) and Fe(VI) are formed by the loss of additional 3*d* electrons, only rarely are these higher valences of Fe seen in biological systems. As noted above, *catalase* and *peroxidase*, two heme enzymes, use iron to engage dangerous oxidants. Both enzymes are located in the cytosol and in peroxisomes where harmful oxidation reactions occur during the course of normal metabolic events. Perhaps the most familiar iron-

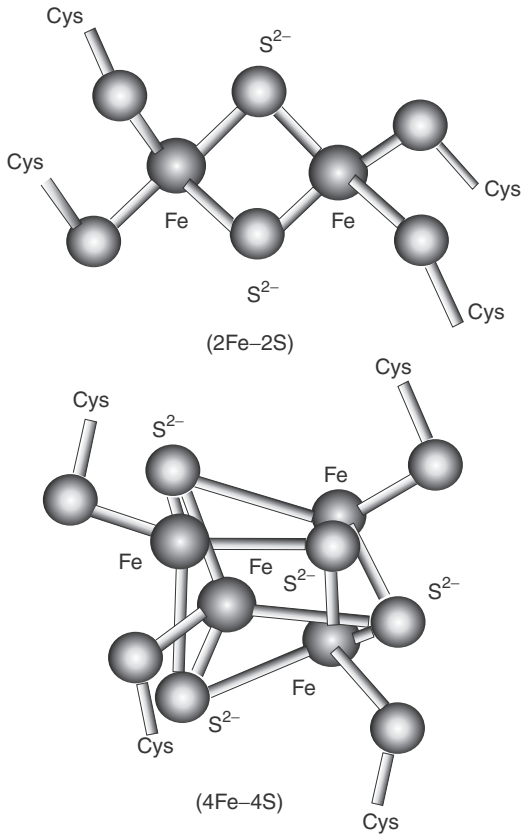


Figure 3 Iron–sulfur clusters. Both the Fe_2S_2 and Fe_4S_4 clusters are bound to the protein via cysteine residues. The iron in these complexes either engages a substrate or holds and passes electrons.

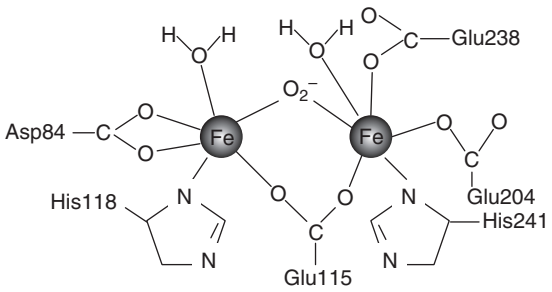


Figure 4 Fe–O–Fe center in ribonucleotide reductase. The two iron atoms are in close juxtaposition to bind dioxygen as the peroxide-anion O_2^{2-}. Side chains of aspartic and glutamic acid residues as well as two histidine residues assist in linking the center to the protein. The center assists in the formation of a free radical that forms on a neighboring tyrosine residue (after Fraústo da Silva and Williams).

containing enzyme is *cytochrome c oxidase*, the terminal electron acceptor in the mitochondrial electron transport chain and the enzyme capable of splitting a molecule of oxygen to form water.

$$4 \text{ cytochrome } c\left(Fe^{2+}\right) + O_2 + 4H^+ \rightarrow$$
$$4 \text{ cytochrome } c\left(Fe^{3+}\right) + 2H_2O$$

Zinc

Zinc is perhaps the most ubiquitous and versatile of all metal cofactors. Sequencing the human genome exposed more than 900 proteins that have zinc-binding domains in their structure. Not all of these proteins function as enzymes, however. Hence, it is wrong to say zinc is a cofactor for 900 enzymes. Some of the better characterized zinc enzymes are shown in **Table 4**.

Table 4 Important zinc enzymes

Enzyme	*Source*	*Function*	*Zn/protein*
Alcohol dehydrogenase	Liver	Alcohol metabolism	4
Alkaline phosphatase	Placenta	Unknown	4
Carbonic anhydrase	Erythrocyte	CO_2 hydration	1
Carboxypeptidase	Pancreas	Protein catabolism	1
Glutamate dehydrogenase	Liver	Glutamate synthesis	2–6
Leucine aminopeptidase	Intestine	Peptide catabolism	4–6

Zinc-binding proteins that engage DNA, the so-called zinc finger proteins, are examples of nonenzyme protein but nonetheless proteins whose function in a noncatlaytic way is dependent on zinc. Approximately 3% of the genome of mammals codes for zinc finger protein. As a cofactor, zinc can perform both structural and catalytic functions. In *carbonic anhydrase*, for example, Zn enters into a coordinate bond with the CO_2 substrate (**Figure 5**). In *carboxypeptidase*, zinc takes an active part in the cleavage of the peptide bond (**Figure 6**). Multisubunit enzymes such as *aspartate transcarbamylase* use Zn to coordinate the positions of the catalytic and regulatory subunits, a structural role. *Cu_2, Zn_2 superoxide dismutase* requires zinc to position the copper atom in the channel accessed by the substrate $HO_2^{\cdot}$, another structural role. In zinc finger proteins, Zn^{2+} contributes to the stability of the loop structure that contacts the major and minor grooves of DNA. These examples illustrate why zinc is an important companion to enzymes and proteins.

One of the lesser known, and perhaps less appreciated functions of zinc is that of a neurotransmitter that modulates neural activity in the brain. Although not to be considered a cofactor role because no enzyme appears to be involved in the action, zinc ions amass in synaptic vesicles and are released into the synaptic junction of glutamanergic neurons. Releasing zinc tends to modulate the activity of the neurons and thus control the impact of the glutamate.

Reactivity

Zinc is considered a bland metal because it behaves as a divalent cation with weak geometric preference. It is perhaps this blandness that allows zinc to adapt to so many different enzyme environments. Zn exists in one valence state, Zn^{2+}, and hence cannot donate or accept electrons. Zn^{2+}ion is configured as a $3d^{10}$, which denotes a filled 3*d* orbital. For that reason, zinc complexes lack color and zinc itself behaves mostly as a cation. The Zn^{2+} ion is capable of recognizing electron pairs (typical of a Lewis acid) and thus can enter into a coordinate bonding arrangement that polarizes groups to which it binds. This property allows zinc to increase the susceptibility of a chemical bond to attack. For example, Zn^{2+} polarizes water, which makes the water behave more like hydroxide ion and be more effective in attacking the CO_2 to form HCO_3^-:

$$Zn^{2+} + H_2O \Leftrightarrow Zn^{2+}(OH^-) + H^+$$

in the reaction catalyzed by *carbonic anhydrase*. Another example is the use of zinc to polarize the ester or amide bonds thus promoting nucleophilic attack of water on the bond as in reactions catalyzed by *carboxypeptidase* and *aminopeptidase*.

Copper

Copper, like iron, can donate and accept electrons. Thus like iron, Cu exists in multiple valence states; Cu^+ and Cu^{2+}(cuprous and cupric) are the most stable. Copper enzymes, although not nearly as numerous as zinc, fill a variety of important biological

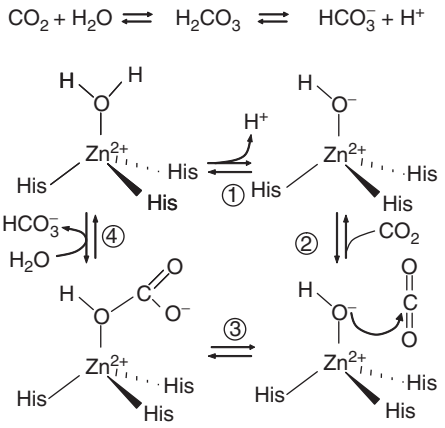


Figure 5 Zinc in carbonic anhydrase. Zinc in the enzyme 'activates' a water molecule (1) creating a better nucleophile to attack the CO_2 (2). Once formed (3) the hydrated CO_2 as HCO_3^- is displaced from the enzyme via a second water molecule (4) regenerating the active enzyme.

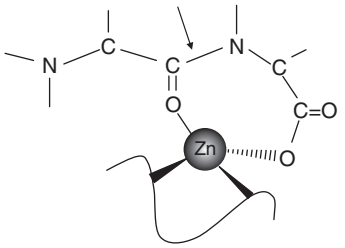


Figure 6 Zinc in carboxypeptidase. In carboxypeptidase, zinc atom forms a binary complex with groups on the C-terminal end of the protein. Arrow shows that bond that will be cleaved with water. Only the C-terminus residue is released from the protein.

functions mostly with membrane-bound enzyme (**Table 5**). A common denominator to these, however, is the donor–acceptor electron fit the category of oxidoreductases, or more specifically 'oxidases', meaning they catalyze reactions in which electrons from the substrate are transferred to O_2. Copper enzymes can be simple or complex, depending on the number of Cu atoms in the enzyme. Simple enzymes generally contain one Cu per subunit. The more complex include the multicopper oxidases, which may have as few as four, e.g., *laccase*, or as many as eight copper atoms per enzyme e.g., *dopamine-β-monooxygenase*. Copper in these enzymes exists in three different chemical environments referred to as type 1, type 2, and type 3 copper sites. *Ceruloplasmin*, for example, contains 6–7 Cu atoms in three distinct sites. The type 1 copper site gives a blue color to ceruloplasmin and other blue Cu proteins. The copper-binding sites in a multicopper oxidase form a triad consisting of one type 2 and two type 3 coppers arranged as an isosceles triangle. Oxygen binds to the two type 3 coppers at the base of the triangle. Examples of copper enzymes include *cytochrome c oxidase*, *lysyl oxidase*, and *ascorbate oxidase*.

Reactivity

Because it is prone to accept electrons, Cu is a powerful oxidant in biological systems. The Cu sites in ceruloplasmin have the capacity to oxidize Fe^{2+} to Fe^{3+}, which prepares ferric ions to bind to transferrin and delivers iron to the organs and tissues. This reaction links iron with copper metabolism and could explain how an absence of copper in the diet impairs the transport of iron and causes anemia in humans. In *Cu_2, Zn_2 superoxide dismutase*, the Cu^{2+} at the active site removes the single nonbonding electron from one superoxide anion (O_2^-) and transfers it to another:

$$O_2^- + O_2^- + 2H^+ \Leftrightarrow H_2O_2 + O_2$$

Seldom is copper destined to perform only a structural role and many enzymes that possess copper as a cofactor use the metal at the active site. More recent studies have linked Cu ions with the formation of blood vessels, or angiogenesis. One of the more exciting discoveries yet to be fully understood is that depriving an animal or human of Cu delays or even inhibits the growth of cancerous tumors. From a nutritional perspective, this could mean that Cu is essential for the development of the microvascular system.

Manganese

Whereas zinc may be the most common transition metal in enzymes, manganese is perhaps the least common. Part of the reason is that complexes of manganese with proteins tend to be weakly stable with a tendency to dissociate. Notable manganese metalloenzymes include *pyruvate carboxylase* and *manganese superoxide dismutase* in the mitochondria and *arginase* in the urea cycle. Manganese can also function as a metal-activating cofactor for many enzymes that require magnesium.

Reactivity

Although manganese is not considered a redox metal based on reactivity, it nonetheless can exist in six oxidation states (Mn^{2+} to Mn^{7+}) three of which (Mn^{5+} to Mn^{7+}) are not seen in biological systems. The most common form of manganese is Mn^{2+}. The highest number of multiple valences of manganese occur in the *water splitting enzyme* that is found in chloroplasts of plants as part of photosystem II.

Table 5 Important copper enzymes

Enzyme	*Source*	*Function*	*Cu/protein*
Ascorbate oxidase	Squash	Ascorbate catabolism	8
Ceruloplasmin	Plasma	Iron oxidation	6–7
Cytochrome *c* oxidase	Mitochondria	Electron transport	2
Dopamine-β-monooxygenase	Adrenal	Noradrenaline synthesis	8
Lysyl oxidase	Aorta	Collagen, elastin synthesis	1
Superoxide dismutase	Erythrocyte	Superoxide radical destruction	2

Cobalt

The role of cobalt as a cofactor is limited to its presence in vitamin B_{12}. Cobalt can exist in three valence states, Co^{+}, Co^{2+} and Co^{3+} with Co^{2+} being the most common in 5′-deoxyadenosylcobalamin, the familiar form of vitamin B_{12} coenzyme. Cobalt is bound by a planar ring system analogous to heme but with very special features. Cobalt (and nickel) are ions that may have figured more prominently in primitive systems when the atmosphere contained H_2 and CH_4 as common environmental gases. The argument has been made that as biological system gradually adapted to O_2 the necessity for these two metals became less.

Reactivity

Cobalt in the structure of vitamin B_{12} resembles iron in heme by being bound in a square planar arrangement to a ring (corrin). Unlike heme, however, cobalt has two axial ligands that are free from the protein, which allows nonprotein groups to access the central metal from above and below the plane. In the octahedral complex, one axial position (the fifth coordinate) is normally occupied by a benzimidazole and the other by a methyl group (as in methyl cobalamin). The arrangement is unique and allows cobalt to form carbon–metal bonds with the potential for two different reactivities. The methyl group, for example, may be removed as a carbonium ion retaining both electrons on the cobalt, which then reverts to a less stable Co(I). This is typical of the reaction in which B_{12} acts as a methyl donor. In positional rearrangements, cobalt retains only one electron and forms a stable Co(II) or d^7 ion with the release of a free radical. Free radicals are highly reactive and overcome energy barriers that would stymie other reactants. Thus, cobalt's chemical properties transfer groups as carbonium ions or highly reactive carbon-centered radicals. Both products are possible and hence explains the necessity for Co as a cofactor for a reaction that precede via a free radical mechanism. An example of the latter is the intramolecular rearrangement of methylmalonyl–CoA to succinyl–CoA as catalyzed by *methylmalonyl–CoA mutase*.

Vanadium

A well defined biochemical function for vanadium in higher animals and humans is yet to be described. Recent reports of vanadium in bacteria and algae have provided clues as to the functional necessity of this metal in enzyme catalysis. Approximately 10 years ago, vanadium was found to be essential for the activity of *bromoperoxidase*, an enzyme found in brown and red algae. Shortly thereafter, a vanadium-dependent *iodoperoxidase* was characterized. Vanadium was also found in high concentrations in mushrooms and was shown to accumulate in large quantities in ascidians, specifically the blood cells (vanocytes) of these organisms. Speculation as to function of vanadium in microorganisms range from antimicrobial action to electron transfer and the trapping of oxygen. In higher animals, however, vanadium has been shown to have insulin-mimetic properties and to stimulate cell proliferation and differentiation. It is also believed to regulate phosphorylation and dephosphorylation reactions through control of ATPases, phosphatases, and adenylcyclase that have wide spread effects on cell functions. These are the most plausible understandings to date. However, it should be emphasized that vanadium has not been shown to be a specific activator (or inhibitor) of any enzyme in humans.

Reactivity

Vanadium is like molybdenum in being able to form both oxyanions and oxycations, $VO_4^{2-}(MoO_4^{2-})$, $VO_2^{+}(MoO^{2+}$, MoO_2^{+}, and $MoO_2^{2+})$ and sulfur centers as well, e.g., VS_4^{3-} (MoS_4^{2-}). Vanadate differs from molybdate in being a rather strong oxidizing agent, E~+.5 V at pH 7, which may relate to its electron transfer function in lower life forms but has questionable significance in humans.

Molybdenum

Molybdenum is widely distributed in plants and animals. The metal exists in three valence states, Mo^{4+}, Mo^{5+}, and Mo^{6+}. A limited number of redox reactions exploit the multivalence states. Molybdenum-dependent enzymes are found in pathways that metabolize purines, pyrimidines, pterins, aldehydes, and sulfites. A cofactor structure for molybdenum has been proposed (**Figure 7**) and referred to as molybdopterrin. Enzymes that use the cofactor include *xanthine oxidase, sulfite oxidase,* and *aldehyde oxidase.* In microorganisms, molybdenum is a key metal for the fixation of nitrogen. *Xanthine oxidase* is the enzyme with importance relevance to a mammalian system.

Reactivity

A major nutritional concern of molybdenum is its ability to antagonize copper. Indiscriminant spraying of soils with molybdenum has been shown to affect the growth and productivity of ruminants. The effect relates to the formation of thiomolybdates in the rumen. The thiomolybdates interact and bind copper preventing its absorption from the rumen. Thiomolybdates have a very high affinity for copper almost to the exclusion of other metal ions. Lately, thiomolybdates have been used to control copper toxicity in Wilson's disease, a genetic disease of copper poisoning in humans.

Nickel

As a cofactor, nickel occurs infrequently. About the only known occurrence of nickel is in microbial and plant enzymes such as *urease* from jack bean, soybean, rice, and tomatoes. There are roughly two gram-atoms of nickel per mole of the 96 000 dalton subunits of the enzyme. Other metalloenzymes containing nickel include Factor F430 found in the membrane of methanogenic bacteria, *carbon*

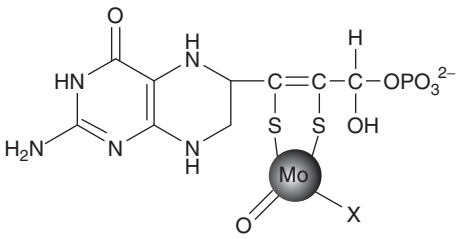


Figure 7 Proposed structure for the molybdenum cofactor in nitrogenase. This center consists of a special pterin cofactor, a relative of tetrahydrofolate. The molybdenum engages two sulfur atoms as a dithiolate complex.

monoxide dehydrogenase and *hydrogenases I and II*. Nickel has drawn the attention of nutritionists because of the observation that nickel concentrations in serum of women rise sharply immediately after parturition.

Reactivity

Some consider nickel the 'metal that was'. As biosystems evolved and moved from an atmosphere of no oxygen to one rich in oxygen, where methane and H_2 have tended to be minimized as energy substrates, metals that formed a major cofactor in the anaerobic environment and used by the more primitive organisms such as archaebacteria have been replaced in favor of a metal or cofactor more suitable to the present day environment. Thus, nickel, like cobalt may have had its greatest era in enzymes that catabolized CH_4 or H_2.

Other

Although metals such as Cr, Sn, As, and Sr are known to be essential for optimum growth and health of organisms as well as having a major influence on biological systems, cofactor functions for these metal ions have not been assigned because specific enzymes which may require them for activity have not been found.

Nonmetal Mineral Cofactors

Selenium

Selenium belongs to the category of redox nonmetals. Selenium is included in the same class with sulfur (sometimes referred to as metalloids), which implies that selenium should be able to substitute for sulfur in biological complexes. As a congener of sulfur, selenium becomes part of a protein's structure as selenocysteine and selenomethionine, not as a selenium atom ligated directly to the protein as a prosthetic group. The former are the active cofactors in selenium enzymes.

Reactivity

Although a selenium ion is clearly capable of redox reactions, there is still little information available as to how selenium functions as a cofactor. Enzymes such as *glutathione peroxidase* are soluble enzymes that transfer electrons to and from substrates. Replacing the selenium with sulfur in the enzyme negates the activity. With only a few selenoenzymes available, there is little information as to the precise catalytic role of selenium. *Glutathione peroxidase* in the reduced (resting) form is believed to contain an ionized selenol that can react with either organic peroxides or H_2O_2 according to the reaction (a) shown below: a selenol enzyme is

$$\text{(a)} \quad \text{Enz–Se}^- + \text{ROOH} \rightarrow \text{Enz–SeOH} + \text{ROH}$$

$$\text{(b)} \quad \text{Enz} + \text{Se}^- + \text{RI}_4 \rightarrow \text{Enz–SeI} + \text{RI}_3$$

also believed to be an intermediate in the reaction (b) catalyzed by *5′-deiodinase*, the enzyme that catalyzes the removal of a single iodine atom from thyroxin, the major thyroid hormone. This T4 to T3 transition gives rise to the more active form of the hormone. Because the *5′-deiodinase* activity is suppressed in an iodine deficiency, there have reports of goiter-like conditions being manifest in people with low levels of selenium in the diet.

Silicon

There is still some question as to whether silicon is a cofactor. It is included here because of the importance of silicon in a number of biochemical reactions leading to the synthesis of glycoproteins and polysaccharides in the extracellular matrix of connective tissue ground substance. Silicon as $Si(OH)_4$ is very abundant in soils and minerals and is as common in human tissues as magnesium. In plants, especially grasses, silicon is a major component of a mineral skeleton and has a metabolic turnover nearly on a par with carbon. In humans, the highest concentrations of silicon occur in connective tissues such as aorta, trachea, tendon, bone, and skin. Lesser amounts are found in liver, heart, and muscle. The epidermis and hair are significantly high in silicon.

Reactivity

Silicon, as silicic acid, has been shown to be required for maximal activity of *prolyl hydroxylase*, the enzyme that converts proline residues to hydroxyproline in collagen. High levels (0.2–2.0 mM) are needed to stimulate the enzyme, which catalyzes a rate-determining factor in collagen biosynthesis.

Boron

Manipulating the boron content of a diet leads to a wide number of metabolic responses, which is testament to the potential importance of boron in human nutrition. Early studies reported increased levels of steroid hormones, testosterone, and estradiol in animals supplemented with boron. Further studies suggest that boron has a regulatory role in the metabolism of other minerals such as calcium and may affect bone metabolism. In a comparative way the role of boron is well established in vascular plants, diatoms, and marine algal flagellates. Zebra fish deprived of boron tend to suffer developmental defects. These data have impelled investigations into the biological functions of boron in higher vertebrates. To date, however, few studies have supported boron's essential role in vertebrates. In a comparison to Zebra fish, pregnant rats fed one-fiftieth the level of boron as control rats exhibited no impairment in fetal growth or development. Fewer two-cell embryos from the deficient rats, however, reached the blastocyst stage when cultured *in vitro*, suggesting boron deprivation did have an impact at a very early stage of development. Perhaps the strongest holdup to accepting boron as essential is the failure to define and link a specific organoboron compound with a physiological function. A report of boron associated with a naturally occurring antibiotic is an exception. The data, however, tend to support the notion that boron complexes with biological components are too unstable to be isolated and studied. This clearly has put a damper on the forward thrust of knowing boron's precise metabolic function.

Conclusions

The mineral cofactors described here may be thought of as representing a special subset of the biominerals. Rather than contributing to skeletal mass and fluid homeostasis, however, mineral cofactors are more subtle and are devoted specifically to enzymes. The words ‘mineral’ and ‘cofactor’ combine to designate an inorganic component required by an enzyme in order to achieve optimum catalytic efficiency. In seeking a reason for mineral cofactors, one must consider that to meet its functional obligations, an enzyme faces many challenges. The protein surface can easily be modified chemically through interaction with substrates and the enzyme protein can readily lose its biological form through denaturation. Electrons and groups that are transferred to and from substrates have the potential to permanently modify the enzyme. This happens frequently and instead of undergoing repair, old enzymes are replaced by new ones. The mineral cofactors fit into the daily wear and grind by making the enzyme better able to stand up to the harsh environment of their existence. They also have been shown to be effective binders of substrate and to interact with oxidants and reductants in a facile manner. Some trace metals such as Zn can accept electron pairs in forming a covalent attachment that polarizes and facilitates rupture of the chemical bonds in the substrate. Other metals such as copper and iron can accept electrons from the substrate and pass them to oxygen. Catalysis and structure stability are the two primary functions of metals in enzymes. Many organic factors serve as electron-capturing and group-transferring agents. This suggests that metalloenzymes may backup enzymes with organic cofactors. This view is rather narrow and oversimplified because there are many enzyme-catalyzed reactions where only a metal will suffice, such as in the metalloenzymes that catalyze the destruction of oxygen radicals. In biology seldom does one factor become indispensable. What nutritionists refer to as essential metals are on the same level as vitamins in that they are needed in very small quantities to maintain a status quo system and, like vitamins, available strictly through the diet. One must, therefore, conclude that essential minerals and vitamins have common ground in the enzymes, which they literally permit to function.

Further reading

Berthon, G, 1995. Handbook of Metal-ligand Interactions in Biological Fluids. In: Bioinorganic Medicine, vols. I–II. Marcel Dekker, New York.
Eichhorn, GL (Ed.), 1973. Inorganic Biochemistry, vols. 1–2. Elsevier Scientific Publishing, Amsterdam.
Fraústo da Silva, JJR, Williams, RJP, 1991. The biological chemistry of the elements. The Inorganic Chemistry of Life. Oxford University Press, Oxford.
Harris, ED, 2003. Basic and clinical aspects of copper. Critical Reviews of Clinical and Laboratory Sciences 40, 547–586.
King, TE, Mason, HS, Morrison, M, 1988. Oxidases and Related Redox Systems. In: Progress in Clinical and Biological Research, vol. 274. Alan R. Liss, New York.
Lanoue, L, Taubeneck, MW, Muniz, J, et al., 1998. Assessing the effects of low boron diets on embryonic and fetal development in rodents using in vitro and in vivo model systems. Biological Trace Element Research 66, 271–298.
Mertz, W, 1987. Trace Elements in Human and Animal Nutrition, 5th edn.,, vols. I–II. Academic Press, Inc, New York.
Prasad, AS, 1993. Essential and Toxic Trace Elements in Human Health and Disease: An Update. Wiley-Liss, New York.
Stadtman, TC, 1996. Selenocyteine. Annual Reviews of Biochemistry 65, 83–100.
Stryer, L, 1995. Biochemistry, 4th edn. WH Freeman & Co, New York.

Copper

Peter J. Aggett, University of Central Lancashire, Preston, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of P.J. Aggett, Copper, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 397–403, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00062-3.

Introduction 200
Copper function 201
Dietary sources and reference values 201
Copper trafficking and kinetics 203
Copper absorption (intestinal uptake of copper and its transfer to the body) 203
Copper distribution 204
Copper excretion and homeostasis 205
Inborn errors of metabolism 205
Measuring copper status 206
Copper deficiency 206
Copper excess 207
Conclusion 207
References 208

Key points

- The switch between cuprous and cupric states underpins the metal's essentiality in systemic substrate metabolism, energy production, tissue synthesis, neurotransmission and peptide endocrine activity.
- Elements of its roles and homeostasis are evolutionarily conserved thus enabling a systems biology approach to studying its essentialty.
- Reference intakes of copper for adults approximate 1.0-1.5 mg daily. It is widely distributed in the diet particularly in cereals and grains.
- The liver is the main regulator of systemic metabolism and homeostasis of copper by means of its seqestration with cysteine rich isoproteins, collectively called metallthionein.
- Two inborn errors of copper metabolism affecting its systemic distribution from the liver (Menkes Disease) and its biliary excretion (Wilson Disease) have highlighted the affects of low and high body burdens of copper.
- Copper excess may arise from environmental contamination, and may be predisposed by ecogenetic factors.
- Copper deficiency occurs in malnutrition states and with synthetic feeds. The is interest that foods rich in fructose or glucose may cause adverse cardiovascular effects in adults on low copper intakes.

Introduction

For more background information and specific information covering points in the text the following reviews are recommended Crichton (2019), Davis and Mertz (1987), EFSA NDA Panel (2015), EFSA Scientific Committee (2022) *to be released shortly*. Moller and Aaseth (2022), Online Mendelian Inheritance in Man, www.omim.org, and Stern et al. (2007).

Copper (atomic number 29 and atomic weight 63.55) is an essential trace element. It was one of the first metals to have been isolated and used by man, this was about 10,000 years ago somewhere between present day Greece, and India. Certainly, the metal had been isolated before 3500 BC, which is when it was realized that adding tin (one part to nine parts copper) produced a harder more versatile material, namely bronze. After iron and aluminum, copper is the most widely used metal. Its abundance in the Earth's crust is 68 ppm, and although small amounts of metallic copper can be found naturally, most exist as an ore. Fifty percent of crustal copper occurs with iron as sulfide ores, chalcopyrite, bornite, chalcocite and covellite. Carbonate ores include malachite and azurite, and an oxide ore is cuprite. In its biological role and activity, copper is bound in proteins by N and S ligands provided by the amino acids histidine and cysteine. Copper has two oxidation states that are of physiological importance these are the cuprous (Cu^{+}) which is easily oxidized, and cupric (Cu^{2+}) which is the more abundant and stable. The oxidation of cuprous copper is controlled when the copper is bound in a complex as a cuproprotein. The two states have different coordination chemistries; cupric copper favors a planar configuration with 4N ligands, whereas cuprous ions form tetrahedral complexes involving 4S ligands. More complex structures involving two or more copper atoms and others with Cu^{3+} and Cu^{4+} states are known but, as yet their functions have not been fully characterized. One exception to this generalization is hemocyanin which is found in invertebrates

such as molluscs and arthropods. Hemocyanin contains two Cu^{+} atoms that interact to bind molecular oxygen reversibly at the low oxygen tensions and temperatures of the environment at which these creatures live; it is not found in mammals in whom hemoglobin has the same role. However, many cuproproteins found in humans and other mammals are ubiquitous and highly conserved thereby encouraging speculation that, like iron, copper may have been functionally important before, and indeed contributed to, the development of an aerobic environment on earth (Crichton, 2019).

The oxidation of cuprous to cupric copper generates free radicals which, in turn, can cause extensive functional and architectural tissue damage by oxidizing organic molecules such as lipids, proteins, and nucleic acids. Thus, to acquire, distribute, use, and excrete copper, organisms have had to evolve a system of specific protein carriers which forestalls the production of free copper ions and thereby minimizes the risk of oxidative toxicity. Similar systems of chaperoned traffic exist for other essential transition elements such as iron, zinc, and manganese. Since these elements have similar physicochemical properties, interactions may occur between them at separate stages of their respective trafficking systems, however, at normal levels of exposure each system has an accumulative specificity and selectivity which discriminates effectively for their delivery to their respective depots and functional sites. However, there is evidence that dietary iron, in some circumstances and high molar ratios, does interfere with the metabolism and utilization of copper.

Copper function

Copper-dependent enzyme activities (Table 1) are crucial to energy metabolism and muscle efficiency, connective tissue and bone formation, neurotransmission and catecholamine synthesis, the turnover of peptide hormones, free radical elimination, iron trafficking and hemoglobin formation. The features of copper deficiency are described later. Genomic analyses indicate the existence of many potential copper binding motifs for which the proteins and functions have yet to be identified.

Dietary sources and reference values

The copper content of some common foods is shown in Table 2. Good dietary sources of copper are seeds, grains, nuts (cashews and pecans), beans, wheat bran, crustacea, shellfish, cocoa products, and a relatively small component from liver. Although copper is naturally abundant in the liver, the amount in the livers of farmstock may have been enhanced either from the animals having been reared in areas which had been surface treated with slurry containing copper, or in the case of pigs from copper given to them as a microbicidal to alter their gut flora to prevent enteric infections, and improve their growth. These exposures and practices are subject to regulation and this should be considered in the appropriate perspective when preparing dietary advice for individuals and populations. The entry of copper into the food chain via vegetation is limited by the innate homeostatic regulation of copper uptake by plants and by factors such as the redox potential, alkalinity, and sulfur and molybdenum content of the soil which limit the availability of copper from soil to vegetation, and also to grazing livestock which ingest soil to the extent that in some areas, copper deficiency may be endemic in grazing livestock despite copper being abundant in the soil (Davis and Mertz, 1987).

The copper content of water delivered by copper piping can contribute significantly to intakes. Soft acidic water can leach copper from copper piping. If the copper content of water is 0.1 mg L^{-1} water contributes less than 10% of daily intake of copper, whereas if

Table 1 Copper-dependent enzyme activities.

Cytochrome *c* oxidase	Mitochondrial oxidative phosphorylation transfer of electron to O_2
Dopamine-b-hydroxylase	Norepinephrine and epinephrine synthesis
Tyrosinase	Tyrosine-dopa-dopaquinone-melanin production peptidylglycine a-amidating mono–oxygenase
Neuropeptide synthesis including melanocyte stimulating hormone Tryptophan 2,3-dioxygenase	Tryptophan to *N*-formyl-L-kynurenine
Indole 2.3-dioxygenase	Indole to 2-formylaminobenzaldehyde
Monoamine oxidases	Degradation of amines, e.g., serotonin, catecholamines, dopamine, and tyramine
Diamine oxidases	Degrades histamine and polyamines
(Cu–Zn) superoxide dismutase	Cytosolic antioxidant; $2HO_2 > O_2 + H_2O_2$ superoxide conversion
Lysyl (and hydroxylysine) oxidase	Collagen and elastin cross linking; lung, bone matrix, cardiovascular integrity
Uricase	Hepatic and renal metabolism of uric acid
Hephaestin	Ferroxidase for cellular export of iron
Ceruloplasmin	Plasma metallo/ferroxidase
Thiol oxidase	Formation of disulfide linkages

Table 2 Illustrative copper contents of foods.

Food	*Copper content (mg/100 g wet weight)*	*Size of typical serving (g)*	*Copper in a typical serving (wet weight) (mg)*
Fish	61	120	0.07
Turkey	71	120	0.09
Chicken	34	120	0.04
Hamburger	95	120	0.11
Roast beef	82	120	0.10
Steak	120	120	0.14
Sheep liver	15,700	120	18.9
Pork liver	14,100	120	17.0
Egg	80	40	0.03
Hard cheese	43	120	0.05
Whole wheat	107	30	0.03
Scallops	608	120	0.03
Clams	739	120	0.73
Crab	175	120	0.89
Shrimp	175	120	0.21
Oysters	289	120	0.35
Mussels	475	120	0.57
Lobster	3600	120	4.40
Chocolate	118	15	0.02
Milk	33	120	0.04
Peas	238	120	0.29
Soy beans	109	120	0.13
Avocado	168	120	0.20
Raisins	168	30	0.05
Peanut butter	853	30	0.26

it is >1–2 mg L^{-1} upto 50% (i.e., 0.5 mg copper) of daily intake may derive from water. Whether or not this constitutes a potential benefit or hazard needs to be determined by a case-by-case risk assessment; this applies also to assessing the potential impact of using utensils made of copper alloys (e.g., brass) or of copper for cooking or for storing water, milks, and drinks.

Dietary reference values for copper for North America are given in **Table 3**. Most advisory bodies give similar values either as a single value, or as a range (e.g., 1.0–1.2 mg day^{-1} in Australia and New Zealand). Some bodies, including that in the UK, do not think that additional intakes are necessary during pregnancy or lactation, considering that maternal adaptation such as increased absorption or mobilization of endogenous depots would meet the needs of pregnancy and lactation. The uncertainty of the information on copper nutrition in humans has meant that rather than estimating a range of lower, average, and population intake reference values, most agencies provide figures that are seen as adequate intakes e.g., the values from the European Union (**Table 4**). Failure to meet these intakes by individuals or groups of people is not diagnostic of deficiency but can be regarded as an indicator of a risk of deficiency. These uncertainties are the source of variability among the values produced by different advisory bodies and because such variations, no matter how small, could be barriers to trade there are international initiatives to harmonize approaches for the setting of reference values. The dietary intake of copper is approximately 1–2 mg day^{-1}, but individual intakes of copper vary widely over 2–3 weeks, according to the pattern of intakes of rich sources of copper. Few dietary intake surveys are long enough to capture this variability. A tolerable upper intake limit, based on the risk of liver damage, of 10 mg for daily intakes of

Table 3 Dietary reference intakes (RDAs) (mg day^{-1}) for copper in North America.

Age	*RDA*
0–6 months	0.2
6–12 months	0.3
1–3 years	0.34
4–8 years	0.44
9–13 years	0.7
14–18 year	0.89
19– years	0.9
Pregnancy	1.0
Lactation	1.3

Reproduced from IOM recommendations and those of EFSA NDA Panel (2015).

Table 4 Adequate Intakes (AIs) for copper in the European Union (EFSA NDA Panel, 2015).

Adequate intakes for copper (mg/day)		
Age	*Males*	*Females*
7–11 months	0.4	0.4
1–<3 year	0.7	0.7
3–<10 years	1.0	1.0
10–<18 years	1.3	1.1
18 + years	1.6	1.3
Pregnancy		1.5
Lactation		1.5

copper over a lifetime has been advised in North America. In the European Union, an upper level of 5 mg has been recommended (EFSA NDA Panel, 2015; EFSA, 2022).

Copper trafficking and kinetics

Knowledge of the acquisition, distribution, metabolism and excretion (ADME) i.e., the metallomics of copper has been deduced from studies in cell line systems *in vitro*, bacteria, yeasts, drosophila, xenopus, zebra fish, mice, sheep, and humans. Knockout models and inborn errors of metabolism in mammals have contributed appreciably, as have molecular biological techniques which have enabled the integration of information from these sources to demonstrate an appreciable evolutionary conservation of mechanisms for the acquisition and systemic utilization of copper in life forms (Nevitt et al., 2012). Studies of these processes in humans are hampered because there are few tracer forms of copper: the abundance of two natural stable isotopes of copper ^{63}Cu and ^{65}Cu of 69% and 31% make them difficult to use in tracer studies, as do the short half-lives of its radio isotopes (^{67}Cu –61.9 h and ^{64}Cu – 12.9 h).

The principal proteins in the trafficking pathways for copper are a high-affinity copper transporter, (CTR1) which functions as a trimer which forms s a channel to enable an energy independent influx of copper into cells; two carriers involved with the energy-dependent efflux of cupric copper from cells, ATP7A ATPase and ATP7B ATPase; a depot protein, metallothionein, or more specifically isometallothioneins (MTN); and a number of target-specific chaperone carrier proteins that take copper to its functional sites. Additionally, there is a copper transport protein, transcuprein, in the circulation. These proteins are homologous to a varying extent with proteins with similar functions in bacteria and other species, in some of which proteins such as CTR1 homologs are essential for embryonic morphogenesis and cell differentiation, thereby raising interest in the copper dependence of this role, and speculation that CTR may have similar functions in human embryogenesis.

Copper absorption (intestinal uptake of copper and its transfer to the body)

The intestine absorbs 5–90% or more of ingested copper. The higher figure relates to intakes of copper in aqueous solutions among which copper oxides are less well absorbed than acetate, sulfate, gluconate, and citrate salts. Otherwise, absorption varies considerably, approximately 10–60%, depending on the amount in the diet, the character of the diet and most importantly the host's need for copper. Copper is released from the dietary matrix by gastric acidity and proteolytic activity. Gastric mucosal uptake and transfer of copper have been modeled in animal models and plasma appearance curves after ingesting copper salts suggest that gastric absorption occurs in humans, but this is probably not a major route in normal physiology. The predominant site for uptake of copper in adults is the duodenum and proximal jejunum; however, in young animals, at least, the uptake of copper extends throughout the small intestine. The intraluminal solubility of copper and thus its availability for intestinal uptake is enhanced by an acid milieu, anions such as sulfate and nitrate, and low molecular weight ligands including sulfur amino acids, histidine, lactose, glucose, and starch, presumably after its hydrolysis to glucose. The uptake of copper from foodstuffs is impaired by Maillard reaction products produced during food preparation, and by interactions involving phosphate compounds (such as phytate), amino acids, magnesium, and calcium which precipitate copper in the gut lumen. Similarly, vitamin C, by oxidizing cuprous copper to cupric, impairs copper uptake in animal models, but this does not necessarily happen in humans.

The first stage of copper uptake is its aggregation by the mucus and glycocalyx on the enterocytic microvilli, whence it reaches membrane-associated reductases (STEAP2 and Dcytb) which convert any residual cupric to cuprous copper for translocation into the enterocyte. There are at least two mechanisms for enterocytic uptake of copper; one which is dependent on CTR1, which accounts for 70% or more of uptake, and one which probably uses the divalent metal transporter (DMT1). The former mechanism involves copper uptake into subapical vesicles in the enterocyte, and because this process occurs in the absence of CTR1, and has not been associated with DMT1-mediated uptake of other metals, it is thought that copper may be first taken up by endocytosis, it is not clear whether the metal is reduced before endocytosis or within the endosome. Then CTR1 transfers the cuprous metal out of the

vesicle. This endocytic route would be consistent with early literature reports of copper uptake. In the enterocyte, copper is either bound to metallothionein or to a chaperone, Atox1, which translocate the copper to ATP7A for vesicular transfer of the metal across the basolateral membrane into the portal circulation.

This association is not predictable, and neither is the effect of coincidental intakes of zinc and iron on copper uptake and transfer. When given in solution with copper both iron and zinc might reduce the systemic absorption and use of copper. Reduced red cell superoxide dismutase activity has been seen in infants given iron supplements, and it is possible that iron supplements impair the utilization of copper when given during the management of pan-malnutrition. However, when given in the complex milieux of diets and infant formulas the interactions are not always observed. In another context mentioned later persistent intakes of zinc can block intestinal transfer of copper by inducing enterocytic MTN.

A phenomenon that has not been fully explained is that high fructose and sucrose intakes apparently increase the risk of marginal intakes of copper, whereas isocaloric intakes of complex carbohydrates do not. This was first observed experimentally and may not be relevant to humans. However, in adults, experimental diets providing 0.7–1.0 mg of copper daily have induced cardiac dysrhythmias, conduction defects, bradycardia, and elevated low-density lipoprotein (LDL) cholesterol levels and reduced high-density lipoprotein (HDL) cholesterol levels in the circulation. These changes might have arisen from the character of the diet, i.e., the use of corn syrup, and the phenomena have not been reproduced or observed using more usual diets. Nonetheless, given the character of "Western Diets" there are concerns that copper inadequacy might be widespread in the population and partly responsible for common defects in lipid metabolism with adverse sequelae on cardiac health attributed to such diets (Harder et al., 2020).

Copper has several carriers in the portal circulation. These include transcuprein, albumin, and complexes with amino acids usually histidine, threonine, and glutamine. Copper is transported in the portal circulation to the liver and thence to the systemic circulation. Copper complexed with phosphatidic acid and fatty acids has been found in mesenteric lymph; this implies that some copper may bypass the liver and reach the systemic circulation via the thoracic duct. Usually, within 2 h of ingestion nearly all absorbed copper is taken up by the liver. This is mediated by cell membrane bound reductases and CTR1. The liver is the primary depot for copper, and the principal mediator of its systemic homeostasis, whereas fundamental control of homeostasis may depend on mitochondria (Baker et al., 2017).

In the hepatocyte, the copper joins a high turnover labile pool, probably centered on copper–glutathione complexes from which the metal is distributed to at least four targets: (1) a metallothionein pool; (2) the copper chaperone for Cu, Zn – SOD(CCS) to SOD1; (3) a series of chaperones involving COX 17 (with COX 19 and COX23) to COX 11 and SCO1 which is the chaperone; taking copper to mitochondrial inner membrane cytochrome *c*-oxidase or (4) to Atox1, for transport to ATP7B in the TGN which incorporates copper into apoproteins to form caeruloplasmin, and other cupric enzymes. It is noteworthy that knockout models of ATOX1 accumulate intracellular copper and develop a copper-deficient phenotype. Caeruloplasmin is then secreted into the circulation or excreted directly in the bile.

ATP7A and ATP7B transfer copper to apoenzymes, or into vesicles for export from the cell. The loss of either role is responsible for a specific disturbance of copper turnover; the absence of a normally functioning ATP7A is associated with Menkes disease and that of ATP7B with Wilson disease. These are discussed later. The difference between these two diseases and their variants can be appreciated from the different distributions of the two transporters. Although both are ubiquitous, ATP7A predominates in the kidneys, lungs, blood–brain barrier, gastrointestinal tract, and muscle, whereas ATP7B predominates in the hepatocytes where it is responsible for the synthesis of caeruloplasmin, and for the excretion of copper into the bile. There is little ATP7A in the liver. Within cells both ATP7A and ATP7B are usually distributed around the nucleus where they donate copper to apoenzymes in the Trans Golgi network (TGN). It is noteworthy that cisplatin and related anticancer agents are trafficked via CTR1 which mediates their uptake; additionally, cisplatins are excreted by ATP7B into the bile and therapeutic resistance to these compounds is associated with an upregulation of ATP7B.

MTN are ubiquitous intracellular monomeric polypeptides with a relative molecular mass of 6500 that contain 60 amino acids 30% of which are cysteine. MTN binds 6–10 atoms of copper, and it is induced among other things by endotoxemia, infections, calorie restriction, glucocorticoids, exercise, oestrogens, and hypothermia, as well as by zinc and, high exposure to copper. As has been said, MTN may have a role as a sequester of excess copper or a mobilisable depot of the metal.

Copper distribution

Copper is probably distributed by the systemic circulation to peripheral tissues by the same complexes that are formed in the portal circulation and peripheral tissues are thought to take up and utilize copper in the same way as does the liver.

Caeruloplasmin contains 60–70% of copper in the circulation. It binds six atoms of copper and was thought to be a copper transport protein. However, evidence that caeruloplasmin has cuprous oxidase and ferroxidase activity, and a role in facilitating the binding of manganese to transferrin has prompted the revised concept that caeruloplasmin, and the related hephaestin are metallo-oxidases rather than just ferroxidases. The inherited deficiency of caeruloplasmin has little effect on copper trafficking and function but has appreciable effects on iron utilization.

The copper content of some tissues in adults and infants, and in the functional absence of ATP7A (Menkes' disease) is shown in Table 5. The adult distribution, and metabolism of copper develops during infancy and early childhood. Overall, an adult human contains 50–120 mg of copper: 40% of this is in muscle; 15% in hepatocytes; 10% in brain, and 6% in blood. Approximately 60% of the copper in red blood cells is in superoxide dismutase. In plasma, 60–70% of copper is found in caeruloplasmin, 10–30% is associated with a transport protein transcuprein, and 15–20% is bound to albumin and amino acids.

Table 5 Tissue copper content (mg g^{-1} wet weight) in adults, infants, and infants with ATP7A dysfunction (Menkes Syndrome).

Organ	*Adults*	*Infants*	*Menkes syndrome*
Placenta		4.1–7.5	8.3–14.5
Liver	4.2–16.9	30–80	3–12
Brain	3.6–7.5	0.3–1.2	0.2–1.04
Intestine	1.2–3.4	4.1–7.5	6.4–12.4
Muscle	0.6–1.4	0.25–1.02	1.7–2.6
Spleen	0.9–1.7	0.6–1.9	6.4–15.4
Kidney	2.1–3.7	0.5–1.9	5.9–36.8
Lung	1.02–2.0	0.4–1.0	1.8–4.6

Reproduced with permission from Aggett (1998).

Copper excretion and homeostasis

The major route for systemic homeostasis of copper at customary intakes is hepato-biliary excretion; this accounts for 98% of excretion, and the rest is lost via urine. Hepato-biliary excretion involves ATP7B, and in a minor capacity ATP7A, secreting copper as a variety of complexes into vesicles which merge with the hepatocytic apical plasma membrane and transfer the copper and complexes into the bile for elimination. The excreted copper is not reabsorbed, the reason for this is unclear, and is lost in the feces.

The excreted copper pool may be supplemented by copper acquired from senescent intestinal cells and degraded caeruloplasmin.

At customary intakes, the amount of copper in the body is regulated by the changes in the amount of copper excreted by the liver. Usually, this entails the body retaining approximately 15% of absorbed copper and excreting the rest. The amount retained compensates loss of copper from shed epithelia (skin, intestinal, and other mucosae) and hair, menstruation, and adventitious blood loss. At times of increased copper need from potential deficiency, arising from inadequate intakes and depleted depots or from new tissue synthesis during growth or nutritional rehabilitation, hepatic excretion of copper is reduced, and intestinal uptake is increased; whereas with excessive copper exposure, hepatic excretion is increased, intestinal uptake and transfer are reduced, and at really high exposures increased metallothionein levels in the gut mucosa and liver, respectively, reduce intestinal transfer of the element, and increase its deposition in the liver. All these phenomena occur over a wide range of intakes and are deduced from a variety of studies in different models, but the adaptive phenomena at copper intakes between 0.8 and 7.5 mg daily at which absorption was 56% and 12%, respectively, have been well documented in humans. Note though that although the % absorption declined, the absolute amount of copper increased at the higher intake.

Studies at the cellular level show that when there is an increased need for copper both ATP7A and ATP7B traffic to and become tightly associated with the TGN; CTR1 increases in several tissues, and aggregates near the plasma membrane, while, in the enterocytes, CTR1 aggregates at the apical membrane.

High copper states in hepatocytes induce a migration of ATP7B from the TGN to cytosolic vesicular compartments that support biliary elimination of copper. In enterocytes, and other cells, increased copper leads to the endocytosis and degradation of CTR1, as well as the relocalization of ATP7A and ATP7B to cytosolic vesicles at the basolateral plasma membrane for transfer of copper into the blood. It is not known how these processes are regulated. A mechanism for the posttranslational regulation of ATP7B has been demonstrated. This involves a protein, COMMD1, which stabilizes ATP7B. However, the stability of COMMD1 itself depends on the protection from proteomic degradation which is mediated by X-linked Inhibitor of Apoptosis (XIAP). XIAP appears to be a copper sensor in that when it binds to copper it loses its inhibitory action on the proteomic degradation of COMMD1 and in turn ATP7B is degraded.

Inborn errors of metabolism

Collectively inborn errors of copper trafficking have contributed considerably to the understanding of the systemic use of copper. A source of detailed information on these is the Website *Online Mendelian Inheritance in Man*. Two conditions, Menkes and Wilson diseases with respective incidences of 1:300,000 and 1:30,000 have captured interest.

The role of copper in the neurodegenerative condition, Menkes (MIM 309400) disease was suspected when its phenotypic similarity to copper deficiency affecting grazing sheep was realized. The defect is in the X chromosome gene (Xq21.1) for ATP7A and because this is an X linked recessive defect Menkes' Disease usually affects boys. The synthesis of a dysfunctional ATP7A results in the affected boy not being able to transfer copper out of the intestine effectively, and the copper which is transferred is neither distributed appropriately around the body or within organs and cells. The latter may accumulate copper but are unable to transfer it to apoenzymes: an illustration of this is that patients with Menkes have low circulating levels of copper and functioning caeruloplasmin, but the circulating level of immunoreactive apocaeruloplasmin may be normal. Since ATP7A is responsible for the transfer of copper to the brain, this organ is affected by a severe copper deficiency. There is an accumulation of copper in the placenta, but the

baby has little or no features of copper deficiency when born. The boy presents after 3 months of age with developmental regression, apneic episodes, failure to thrive and a propensity to infection. This is a heterogeneous condition with a number of defects affecting the allele, and the functional implications of these are being explored in a number of mouse models of ATP7A dysfunction.

Defects in ATP7B synthesis result in Wilson disease (MIM277900), in which there is an inability to transfer copper to apoproteins, including caeruloplasmin, in the TGN and to excrete copper into the bile. There are more than 60 causative genetic mutations of the allele (13q14.3) and the genotype can to a certain extent predict both the age of presentation of the disease, which has been as early as 4 years, and its character. Usually, copper slowly accumulates in the liver and leaks into the circulation where it is deposited in a variety of tissues predominant among which, possibly because ATP7A is functioning normally, is the central nervous system (CNS). The onset of the disease is insidious; early biochemical evidence of liver damage is not detected because the patients have no symptoms or signs to prompt any investigations. However, liver damage has been found at 1 year of age in the sibling of a known case. Other than liver disease resembling hepatitis, or in later presentations, hepatic cirrhosis, the first evidence of disease is often deterioration in schoolwork, and neuromotor and psychiatric defects. Overall, in cases that have not been diagnosed until adulthood most tissues in the body are affected by copper deposition and associated oxidative damage. On a positive note, if the condition is detected early, patients can be "decoppered" with clinical improvement, by using a chelating agent such as D-penicillamine or trientine. Additionally, large oral doses of zinc which induces enterocytic metallothionein thereby blocking the transfer of copper taken up from the gut, or by tetrathiomolybdate with a view to mimicking the cause of copper deficiency in livestock by complexing dietary copper and preventing its uptake by the gut; it does work but its mechanism may involve inhibiting the function of ATOX1 and other metallochaperones.

Hypoceruloplasminemia (MIM 604290) is manifest by low circulating levels of caeruloplasmin and consequently low plasma copper levels, but with no other abnormalities. However, aceruloplasminemia (MIM 604290) in which there is a complete loss of ferroxidase (metallo-oxidase) activity has marked effects on iron metabolism with deposition of iron in tissues, including the CNS and basal ganglia, causing oxidative damage to tissues and resulting in endocrine and severe neurological disturbances. There have been case descriptions of humans with defects of the cytochrome *c* oxidase copper chaperones; a defect in SCO1 caused neonatal liver failure and encephalopathy, and defective SCO2 was linked to heart, brain, and muscle damage in an infant. Not all of these effects can be accounted for by loss of cytochrome *c* oxidase activity, so perhaps the SCOs have other roles.

Measuring copper status

It is difficult to determine by a single measurement if an individual's copper status is deficient, adequate, or excessive. Easily accessible markers such as serum copper (reference range 10-22 μmol l^{-1}) or caeruloplasmin (reference range 0.9-2.65 μmol l^{-1}) concentrations represent a small proportion of the body pools of copper and are susceptible to many confounders. Since caeruloplasmin can contain up to 95% of the serum copper, measuring serum and plasma copper or caeruloplasmin is essentially measuring the same thing. However, caeruloplasmin levels, and copper, are increased as a component of the acute phase 1 reaction to infection, pregnancy, inflammation, and by estrogen, and hypoxia among other factors. Levels are low in copper deficiency but in human conditions the deficiency is invariably accompanied by conditions that increase caeruloplasmin levels, thus obscuring the deficiency marker. Only with severe copper deficiency are serum caeruloplasmin and copper levels reduced. Other approaches have been to measure copper-dependent enzyme activities in tissues with a rapid turnover such as erythrocyte superoxide dismutase, or platelet cytochrome *c* oxidase. Strategies based on measuring the non-caeruloplasmin component of circulating copper, multiple indices, or genomic and postgenomic markers of homeostatic adaptation are being explored. Currently, suspicion and alertness to the possibility of copper deficiency coupled with monitoring any response to copper supplements are the best clinical approaches to diagnosing copper deficiency, but this approach is not useful for large populations or for the subtle deficiency associated with multiple micronutrient deficiencies (McArdle 2021).

Copper deficiency

Copper deficiency is seldom an isolated phenomenon, although it might occur as such in infants, children, and adults on synthetic diets, and parenteral or enteral nutrition. The appearance of copper deficiency in both term and ex-preterm infants develops when their hepatic deposits of copper are exhausted. Copper-deficient infants have presented at 4 weeks and 8 months of age. At 26 weeks of gestation the fetal liver has accrued 3 mg of copper, and at term this depot approximates 10–12 mg so early presentation of copper deficiency may involve defective accumulation of copper *in utero*. In recent clinical practice, copper deficiency is emerging as a complication of gastric surgery in the management of obesity.

The most common cause of copper deficiency is a complex of malabsorption and increased losses from the body because of hemolytic anemias, and gut infections and parasitism, causing protein losing enteropathies. These circumstances are widespread affecting perhaps 25% of the global population, in whom the effects on copper nutriture would be compounded by an absolute or relatively inadequate intake of copper.

The features of copper deficiency observed in infants and children with copper deficiency are listed in **Table 6**. They are the features of marked copper deficiency; the features in classic Menkes' disease are more marked with varicose vasculature, hernias, as well as neurodegeneration secondary to defective synaptogenesis and axon formation.

Table 6 Features of copper deficiency in infants and children.

(Hypocupraemia, and hypoceruloplasminaemia)[a]
Hypochromic microcytic iron resistant anemia, Arrested maturation of erythroid and myeloid bone marrow
Neutropenia and propensity to infections
Hypotonia and muscle weakness, poor feeding failure to thrive
Pallor
Hypothermia, apneic episodes,
Skeletal change
Scurvy like bone changes: metaphyseal irregularities, epiphyseal flaring and cupping, bony spurs, and chip fractures
Epiphyseal porosis and separation
Periosteal reaction and subperiosteal new bone formation
Wormian bones, delayed bone maturation
Osteoporosis, fractures
Abnormal elastic and connective tissues, hernias tortuous vasculature, varices and aneurysms
Fishy odor: trimethylaminemia[a] hypoproteinemia with edema, neurodegeneration, developmental regression

[a]Not always evident.

The propensity to infections is attributable to neutropenia and reduced microbicidal activity of the cells secondary to loss of the respiratory burst and reduced superoxide dismutase activity.

Copper excess

Exposure to excess copper can occur acutely, usually by ingestion of copper salts as free solutions, or by chronic exposure over an extended period. The upper levels mentioned earlier are set to help manage the latter risk. However, they do not prevent large systemic burdens of copper arising from failure to excrete copper as may occur with hepato-biliary obstruction, and Wilson disease. Additionally, patients receiving renal dialysis via a copper-based dialysis membrane may accumulate copper.

Ingested copper may come from water supplied via copper pipes. The metallic and salty bitter taste of copper in the water can be detected at concentrations of 2.5–3.5 mg copper per l. This is just below the threshold (4–5 mg L^{-1}) at which nausea, retching, vomiting, and abdominal pain develop. The current regulatory limit advised by the World Health Organization for the copper content of water is 2 mg L^{-1}, but a lower limit may be introduced. A high copper content in water might cause hair to turn green. Self-poisoning with copper salts, usually in suicide attempts, involves doses of 20–70 g of copper. In these circumstances, the above clinical features rapidly progress to hematemesis, diarrhea, and hypovolemic shock all of which reflect intestinal oxidative damage; following this systemic damage resulting in hemolysis, renal and liver failure develop. The stools and vomit may be green.

Chronic copper exposure from copper contamination during food storage and preparation was thought to contribute to the etiology and pathophysiology of the accumulation of copper seen in Indian Childhood Cirrhosis (ICC) (Idiopathic Copper Toxicosis), but it is now uncertain if the accumulated copper is a primary feature, or a secondary and variable sequel of exposure to another hepatotoxin. Thus, the precise role of the increased copper deposition as a toxicant is uncertain. ICC may be an ecogenetic condition in which patients have genetic predisposition to be abnormally sensitive to an environmental exposure. This is probably the case in Bavarian or Tyrolean liver disease that seems to be an autosomal recessive trait which only emerges with exposure to elevated copper intakes from the use of copper cooking utensils (Moller and Aaseth, 2022).

In contrast to the earlier comments on a possible endemic copper deficiency in Western populations, there are also concerns that the public is exposed to too much dietary copper, for example, via water supplies and that this predisposes the population to copper overload and an increased exposure to oxidative damage, which in the CNS might contribute to the development of Alzheimer's disease.

Conclusion

Copper is an essential micronutrient with extensive roles in enzyme activities that are responsible for the integrity of connective tissue and bone matrix, the development of the nervous system, and the functioning of neurotransmitters, and, not least for effective energy metabolism. Two inborn errors of metabolism affecting copper have provided extensive information on severe copper

deficiency and copper overload. These are Menkes' disease and Wilson's disease, respectively, and their respective fundamental defects are the dysfunction of ATP7A and ATP7B which are energy-dependent copper transport proteins. The features of these conditions have facilitated a better understanding of how the body distributes and uses copper, and also of the subtler features of copper deficiency in particular. This has increased awareness that copper deficiency is probably a treatable component of most if not all malnutrition states. In public health nutrition there is probably a need for a through risk assessment of the risk of copper deficiencies and excesses in all populations and of the possible interactions with dietary intakes of fructose and glucose. These will be difficult to achieve until sound markers of adequacy, deficiency, and excess are available. It would be hoped that such markers could be forthcoming from increased information on the homeostatic control of copper turnover at cellular, organ and systemic levels.

Insight into copper's overall metabolism has been derived from observations of its metabolism in many life forms, genetic defects and conditions. The synthesis of this evidence demonstrates the potential role of metabolomics and systems biology and genomics in advancing biological and nutritional science.

References

Aggett, P., 1998. Neonatal trace elementt metabolism. In: Cowett, R.M. (Ed.), Principles of Perinatal-Neonatal Metabolism, 2nd. Springer Verlag, New York, pp. 909–941.

Baker, Z.N., Cobine, P.A., Leary, S.C., 2017. The mitochondrion: a central architect of copper homeostasis. Metallomics 9 (11), 1501–1512. https://doi.org/10.1039/c7mt00221a.

Crichton, R., 2019. Biological Inorganic Chemistry, third ed. Academic Press, San Diego and London.

Davis, G.K., Mertz, W., 1987. Copper. In: Mertz, W. (Ed.), Trace Elements in Human and Animal Nutrition, vol. 1. Academic Press, San Diego and London, pp. 301–364.

EFSA NDA Panel, 2015. Scientific opinion on dietary reference values for copper. EFSA J. 13 (10), 4253. https://doi.org/10.2903/j.efsa.2015,4253.

EFSA Scientific Committee, 2022. Re-evaluation of the existing health based guidance values for copper and exposure assessment from all sources. EFSA J. to be released shortly.

Harder, N.H.O., Hieronimus, B., Stanhope, K.L., et al., 2020. Effects of dietary glucose and fructose on copper, and zinc metabolism parameters in humans. Nutrients 12, 2581. https://doi.org/10.3390/nu12092581.

McArdle, H.J., 2021. Principles of Nutritional Assessment. Copper. https://nutritionalassessment.org/copper/.

Moller, L.B., Aaseth, J., 2022. Copper. In: Nordberg, G.F., Costa, M. (Eds.), Handbook on the Toxicology of Metals, fifth ed., vol. II. Elsevier, San Diego and London, pp. 243–266. https://doi.org/10.1016/B978-0-12-822946-0.00009-X. Chap 10.

Nevitt, T., Ohrvik, H., Thiele, D.J., 2012. Charting the travels of copper in eukaryotes from yeast to mammals. Biochim. Biophys. Acta 1823, 1580–1593. https://doi.org/10.1016/j.bbamcr.2012.02.011.

Online Mendelian Inheritance in Man. www.omim.org.

Stern, B.R., Solioz, M., Krewski, D., et al., 2007. Copper and human health: biochemistry, genetics, and strategies for modeling dose—response relationships. J. Toxicol. Environ. Health Part B 10, 157–222.

Dietary fiber: Classification and physiological role

Madeline Timm and Joanne Slavin, Department of Food Science and Nutrition, St. Paul, MN, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of I.T. Johnson, Fiber: Physiological and Functional Effects, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 240–245, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00105-7.

Introduction 209
What is dietary fiber? 209
Classification of dietary fiber 210
What is resistant starch? 212
Fermentation of dietary fiber as a biomarker 212
What is a prebiotic? 213
Health benefits of fiber, prebiotics, and probiotics? 214
What are postbiotics? 214
Links between dietary fiber intake and health outcomes 214
Fiber intake and recommendations—mind the gap 214
Conclusion 214
References 215

Key points

- Dietary fiber is an essential nutrient in the human diet that is crucial for human health
- Not all fiber is alike in both chemistry and physiological effect
- Fiber intake is about half of recommended intakes and often fiber supplements or fiber fortification is needed to obtain the recommended amounts of dietary fiber
- Differences in fermentation of fiber affect the physiological response to dietary fibers supporting that adaptation to increased fiber intake is essential in fiber advice
- Although dietary fiber has a known role in gut health, its effects extend beyond the gut with benefits for glycemic response, brain health, and cardiovascular health

Introduction

Dietary fiber is widely recognized as beneficial for human health, and increased intake of certain fibers improves serum lipid concentrations, promotes regularity of bowel movements, improves blood glucose control, aids in weight maintenance, and improves immune function (Klosterbuer et al., 2011). Lack of agreement regarding what dietary fiber is and how it should be measured continues to confound the scientific literature on dietary fiber (Slavin, 1987). In 2001, the Food and Nutrition Board appointed a panel to develop a proposed definition of dietary fiber for North America (Dietary reference intakes: proposed definition of dietary fiber, 2001). Dietary fiber consists of non-digestible carbohydrates and lignin that are intrinsic to and intact in plants. Added or functional fiber consists of isolated, non-digestible carbohydrates that have beneficial physiologic effects in humans. Total fiber is the sum of dietary fiber and functional fiber (Slavin and Greenberg, 2003). The intent of these proposed definitions was to recognize the physiologic actions of fiber and its demonstrable health effects and to reduce the emphasis on dietary fiber as a constituent of food requiring quantification. The emphasis on the physiological benefits of isolated fibers will require that isolated fibers have clinical data that support their effectiveness in human clinical trials. Thus, many of the recent clinical trials of dietary fiber examine commercial fiber supplements and their link to a physiological endpoint.

What is dietary fiber?

Recent definitions of dietary fiber have been offered by various groups and they all agree that dietary fiber is a group of carbohydrate polymers and oligomers (and lignin) that escape digestion in the small intestine and pass into the large bowel, where they are partially, or completely, fermented by the gut microbiota (Jones, 2013). Differences among definitions for fiber are found in three main areas: (1) Degree of polymerization; (2) Relationship to food and food processing; (3) Physiological effects (Fuller et al., 2016). Disagreements on these issues have led to confusion in fiber definition and labeling. For example, in the United States, fibers include oligosaccharides with a degree of polymerization of 3–9, whereas other bodies such as the Codex Alimentarius Commission (CAC) support that the national authorities should decide if carbohydrate polymers with a DP 3–9 should be included. This creates

challenges for measuring dietary fiber and databases of dietary fiber as total dietary fiber values will depend on whether the lower chain DP fibers are included.

The North American definition includes all resistant carbohydrates >3 DP but requires that isolated fibers must show a physiological benefit to be included as dietary fiber sources on the Nutrition Facts label. Physiological benefits long accepted include improved laxation, serum lipid lowering, and glucose control (Slavin, 1987; Korczak and Slavin, 2020). Newer benefits such as improved mineral absorption or blood pressure control can also be accepted as physiological benefits of dietary fiber.

Definitions recognize that dietary fibers can be extracted from edible material (intrinsic) or modified and added back into a food (extrinsic). Because recommendations for dietary fiber are based on protective effects of intrinsic fiber on risk of cardiovascular disease, regulatory bodies have discussed the need to divide fiber into dietary fiber (intrinsic) and functional or added fiber (extrinsic). Earlier discussions linking fiber to physiological effects divided fiber into soluble and insoluble fractions (Slavin et al., 2009). Soluble fibers dissolve in water and may form a gel. They have been linked to blood lipid lowering and glucose control (Slavin, 1987). Soluble fibers survive transit through the small intestine and may be fermented in the gut to form short-chain fatty acids (SCFAs) such as butyrate, acetate, and propionate. Insoluble fibers do not dissolve in water and are the fibers we traditionally think of as roughage, fibers like cellulose and hemicellulose. These fibers are more closely linked to laxation. Because of these early associations between the solubility of fibers and physiological effects, many regulatory bodies allow foods to list total dietary fiber, soluble fiber, and insoluble fiber on labels.

Classification of dietary fiber

Classifying fiber has challenged analytical chemists through the ages. Animal feeds originally measured crude fiber as the nondigestible carbohydrate that might be utilized by ruminants, but it was not well tolerated in monogastric animals. Crude fiber was referred to as the residue of plant-based food left after extraction with solvent, dilute acid, and dilute alkali (Dai and Chau, 2017). In 1953 Hipsley defined dietary fiber as a sum of indigestible constituents that made up the plant cell wall, encompassing the "unavailable carbohydrate." Dietary fiber consists of remnants of edible plant cells, polysaccharides, lignin, and associated substances resistant to digestion by the alimentary enzymes of humans. Dietary fiber includes cellulose, hemicellulose, lignin, gums, mucilage, oligosaccharides, pectin, and other associated substances, including waxes, cutin, and suberin.

Dietary fiber chemistry is critical to classifying dietary fiber. Dietary fiber can be divided into noncellulosic polysaccharides, cellulose, and lignin. Human foodstuffs contain mainly noncellulosic polysaccharides, some cellulose, and little lignin (Slavin, 1987). The average content of noncellulosic polysaccharides, cellulose, and lignin for cereals is 75%, 17%, and 7%, respectively; for raw vegetables, 66%, 31%, and 3%, respectively; and for fruits, 63%, 20%, and 17%, respectively. Generally, fruits and vegetables tend to be higher in cellulose than cereals. Lignin is highest in fruits with edible seeds or in mature vegetables such as carrots or other root vegetables. Dietary fiber content of food depends on plant species, maturity, and components (i.e., leaf, root, or stem). Dietary fiber also includes cutin, waxes, and small amounts of proteins and lipids that are closely associated with the plant cell wall.

The solubility of dietary fiber was thought to help explain physiological effects, but more recent work suggests that solubility does not always predict physiological effects. Viscous, soluble fibers such as oat bran or psyllium are known to lower serum cholesterol and have health claims for their ability to lower blood lipids. Most other soluble fibers, for example, fructo-oligosaccharides, have no effect on blood lipids. Generally, only viscous, soluble fibers play a role in blood glucose control. Even in the 2001 Dietary Reference Intakes, the proposed definition of dietary fiber, it was concluded that separating soluble and insoluble fiber to predict physiological effects was problematic. Yet nutrition labels had already included this designator and consumers were told that insoluble fibers would aid in laxation and soluble fibers would lower blood lipids and improve glucose control.

Similar problems are seen with the belief that insoluble fibers aid laxation and soluble fibers do not. The increase in stool weight is caused by the presence of fiber in the colon, by the water that the fiber holds, and by partial fermentation of the fiber, which increases the amount of bacteria in stool (Slavin and Greenberg, 2003). Fiber in mixed diets including insoluble cellulose and hemicellulose are particularly effective promoters of normal laxation, as are cereal brans, psyllium seed husk, and methylcellulose in the form of supplements If the fiber is rapidly fermented in the large bowel, as are many oligosaccharides, there is little increase in stool weight. But this rapid fermentation of easily fermented fiber can cause intestinal gas and discomfort in some individuals.

Studying dietary fiber in human subjects is difficult, as balance studies require collection of all food consumed and feces produced and measuring the fiber consumed and the fiber left over in feces. Balance studies must be done on controlled food intake and 5-day fecal composites collected as indicators of fiber surviving gut transit. Few contemporary studies exist to include in this review. Stephen et al. (2017) summarized the current state of knowledge on fiber definitions, sources, recommendation, intake and relationships to health in a comprehensive review paper of European fiber work. Disagreements on definitions of dietary fiber and methods of analysis continue, but there is more agreement that standardized definitions are first needed before there is agreement on analytical methods. Nutrient databases for dietary fiber, soluble fiber, and insoluble fiber continue to be lacking.

Cumming et al. (1978) fed 20 g/day of concentrated dietary fiber from carrot, cabbage, apple, bran, and guar gum to the controlled basal diet of nineteen healthy volunteers. Fiber from four commonly eaten food sources and guar gum produced very different response in colonic function. Fecal weight increased by 127% on bran and 20% on guar gum with carrot, cabbage, and apple producing intermediate changes in fecal weight. Differences were related to the intake of pentose-containing polysaccharides in the fiber, generally hemicelluloses. Individual variation in response to a given fiber was significant with stool weights ranging from 65 to 194 g/day on a controlled diet with tightly controlled fiber intake.

Mechanism of action of dietary fiber in the human colon by a novel method modified from nutrition for isolating bacteria found that the main component of human feces is bacteria (Stephen and Cummings, 1980). Cabbage fiber which is extensively broken down in the gut provides a readily useable substrate for stimulation of microbial growth, whereas wheat fiber remains largely undigested and retains water in the gut lumen. Additional work by this research group examined water-holding capacity and its relationship to fecal output in human subjects. Pectin had the greatest water holding capacity but produced the smallest change in fecal weight while bran had the lowest water holding capacity and produced the largest increase in fecal weight (Stephen and Cummings, 1979).

Early work with consumption of purified cellulose in humans, a poorly fermentable fiber, found that when purified cellulose, 16 g/day, was ingested with a semi-purified liquid diet only 8 % of the cellulose was broken down (Slavin et al., 1981). In contrast, more than half of the fiber in a habitual low fiber diet containing fruits, vegetables, and refined grains was degraded. This early work supports that the breakdown of fiber in the large intestine varies greatly depending on the chemical composition of the dietary fiber.

Fiber breakdown on enteral diets with and without fiber supplementation were examined and attempts made to relate fiber breakdown to breath gas production (McNamara et al., 1986). Subjects consumed enteral diets with 0, 30 and 60 g per day of soy polysaccharide, an isolated fiber source. Feces were collected, homogenized, dried, and analyzed for neutral detergent fiber (NDF). NDF was also determined in the diets to measure fiber balance. Fiber breakdown was extensive on all doses of fiber, about 90%. No relationship was found between fiber fermentability and production of breath gases, although there was large individual variability in the breath gas production levels. NDF from the soy polysaccharide fiber in the enteral formula was extensively fermented.

Our laboratory group studied differences in fiber fermentation when human subjects were fed controlled enteral diets with quick bread with 0 g, 10 g, and 30 g of dietary fiber as wheat bran and mixed vegetable fiber (Lampe et al., 1993). Each treatment was consumed for 23 days. Women and men seemed to respond differently to wheat bran and vegetable fiber regarding NDF excretion and breakdown. Fecal fiber excretion was greater in men than women; correspondingly, women tended to digest more fiber than men. This study again found large interindividual responses to dietary fiber and identified factors beyond fiber, such as gender, that alter the gut response to dietary fiber consumption.

Few metabolic studies on rigidly controlled diets have examined the effect of different fiber sources on fiber breakdown. Tucker et al. (1981) published a study on dietary fiber and personality factors as determinants of stool output as they had collected many fecal samples while fiber intakes were controlled at the USDA laboratory in Grand Forks, ND. Even when food and fiber intake was controlled, they found substantial individual differences in stool output. Personality measures were used to predict stool weight and fecal frequency and accounted for as much variance in stool output as did dietary fiber intake. They conclude that personality factors predispose humans to low stool output.

Grabitske and Slavin (2009) reviewed the gastrointestinal effects of low-digestible carbohydrates (LDCs) including definitions, classifications, and mechanisms of LDCs. LDCs are carbohydrates that are incompletely or not absorbed in the small intestine but are at least partly fermented by bacteria in the large intestine. Fiber, resistant starch, and sugar alcohols are types of LDCs. LDCs are increasingly added to processed foods because of potential health benefits including a reduced caloric content, reduced or no effect on blood glucose levels, and non-cariogenic effect. The benefits of LDCs are related to the inability of human digestive enzymes to break down completely the carbohydrates into absorbable saccharides and the subsequent fermentation of unabsorbed carbohydrates in the colon.

Thus, LDCs may affect laxation and cause gastrointestinal effects, including abdominal discomfort, flatus, and diarrhea, especially at higher or excessive intakes. Current recommendations for fiber intake do not consider total LDC consumption nor recommend an upper limit for LDC intake based on potential gastrointestinal effects. In fact, the Institute of Medicine (IOM) did not set an Upper Limit (UL) for dietary fiber intake when fiber recommendations were finalized as it is accepted that humans are omnivores and can adapt to high intakes of dietary fiber. A vegan individual may consume 75 g of fiber per day or more. Of course this fiber comes from foods and is not isolated and potentially fed as a large bolus.

The term "laxation" refers to the elimination of feces (Grabitske and Slavin, 2008). Normal laxation varies widely among individuals. It is clinically defined as having a range of 3 bowel movements per week to 3 bowel movements per day. Stools are generally about 75% water, although this is highly variable and water content increases greatly with diarrhea. Of the remaining stool, about one-third is dead bacteria, one-third is undigested carbohydrate, and one-third is protein, fat mucus, dead cells, and inorganic material.

The frequency, consistency, composition, and weight of bowel movements are all indicators of bowel function and digestive health, but unlike blood, stool samples are generally not collected in large cohorts of healthy subjects. Stool weights are significantly related to fiber intake and studies find a wide range of stool weights, from 72 g/day to 470 g/day (Cummings et al., 1992). Different fibers have different effects on stool weight. Average increase in fecal weight per gram of fiber fed ranged from 5.4 g for wheat bran to 1.2 g for pectin (Cummings, 1993). This information supports why fiber sources high in insoluble fiber (wheat bran) were linked to increased stool weight and fiber sources high in soluble fiber (pectin) were linked to low stool weight.

Factors that affect laxation and digestive health are many. Lifestyle factors, diet, drugs, physical activity, and stress are all known to alter digestive function (Grabitske and Slavin, 2008). Personal factors including age, gender, genetics, the microflora, personality, food intolerances, and gastrointestinal pathology, such as celiac disease or lactose intolerance, will also alter digestive health. Bowel habits are also affected by changes in routine, such as travel or new work schedules. Typically, regular bowel habits depend on eating breakfast, a morning cup of coffee, or other habits. Drugs to alter bowel function and motility range from fiber supplements to irritants or other mechanisms. Antibiotics that decrease the gut microbiota will alter fiber fermentation in the gut and must be considered in any study of dietary fiber fermentation.

All the limitations to human studies of fiber fermentation and metabolism suggest more support for animal studies. Pigs are considered the monogastric animal with the closest digestive tract to the human, so findings from pig studies may provide insight to human physiology. The classification of potentially fermentable carbohydrates into soluble and insoluble is no longer enough for the information required to elucidate mechanisms by which dietary fiber has beneficial effects on monogastric health (Williams et al., 2019). Characteristics including fermentability (both kinetics of fermentation and end-products) are needed to help us understand how dietary fiber affects health in monogastric animals. Unfortunately, these in-vivo studies are expensive so results for in-vitro studies must inform our understanding of dietary fiber fermentation.

What is resistant starch?

Of course, describing and measuring such a diverse mixture of non-digestible carbohydrates and associated substances is a major undertaking. Should non-digestible starch or resistant starch be included in dietary fiber? According to Codex, if resistant starch is naturally present in food, it could be classified as dietary fiber. However, if it is derived from an artificial synthesis, such as physical, enzymatic, or chemical synthesis, it should provide desirable physiological benefits to be considered as dietary fiber.

Foods naturally high in resistant starch include potatoes, cereals, whole grains, beans, legumes, rice, and bananas (Patterson et al., 2020). Barley, potatoes, and rice with higher amylose concentrations have more resistant starch than lower amylose varieties and foods cooked and then chilled have higher resistant starch than foods cooked without chilling. Resistant starch can also be produced from isolated starches and then added back to foods and beverages as dietary fiber, if allowed by local regulations on fiber definitions and labeling.

Starch has been divided into three fractions based on its digestive rate, including rapidly digesting starch, slowly digesting starch, and resistant starch. Resistant starch is a broad and diverse term and was originally divided into four types, i.e., physical inaccessible starch (RS1), ungelatinized starch granules (RS2), retrograded starch (RS3), and chemically modified starch (RS4). More recently, RS5 has been introduced as the amylose-lipid complex, which has a similar resistant mechanism toward starch digestive enzymes as RS3 (Li and Hu, 2022).

In the 1980s, RS was proposed and categorized as a kind of insoluble fiber as it could not be digested in the small intestine. Interest in RS grew as clinical data supported that resistant starch increased bacterial mass in feces, was a good substrate for growth of colonic microbiota, but preferentially increased production of the short chain fatty acid butyrate (Bird et al., 2010). Butyrate is the most important energy source for the colonocyte cell, increasing interest in resistant starch fermentation beyond its dietary fiber content (Carlson and Slavin, 2016).

Resistant starch intervention studies have been conducted on aging, insulin resistance, metabolic syndrome, kidney disease, and schizophrenia and it is thought to be particularly relevant for diseases characterized by dysregulated epithelial integrity and immune function, like inflammatory bowel disease (Dobranowski and Stintzi, 2021). The variable effects of RS on the gut microbiome are striking, with one subject showing increased butyrate production and another showing less butyrate production. Appreciate that measuring SCFAs production in human subjects is challenging as SCFAs are quickly absorbed in the colon and blood levels are not useful biomarkers of SCFA production.

Fermentation of dietary fiber as a biomarker

Dietary fiber is classified based on chemical analysis and solubility, but also fermentability. As described, in vivo studies to measure fiber fermentation in human subjects are difficult to conduct and few recent studies have been published to add to the body of evidence on this topic. Thus, in vitro studies give us our best idea of differences in fermentation of dietary fibers of interest to human nutrition.

Although databases exist for the total dietary fiber, soluble fiber, and insoluble fiber content of many foods and fiber supplements, no accepted databases exist for fermentability of isolated fibers. Although fermentability is just one attribute of dietary fiber, the fermentation of fiber to yield "good bacteria" or associated metabolites (SCFAs) are critical data points for studies of prebiotics and postbiotics, discussed later in this review. The physiologic importance of SCFAs from nondigestible carbohydrate fermentation has been described and it is proposed that the degree of fermentability of a NDC, rather than the endpoints of clinical trials, may be appropriate for classifying it as a dietary fiber (Alexander et al., 2019). This need relates back to 2016 and the USA FDA releasing its first official definition of dietary fiber. Dietary fiber is defined as nondigestible soluble and insoluble carbohydrates with >3 monomeric units and lignin that are either intrinsic and intact in plants or isolated and synthetic and demonstrate a physiologic health benefit in humans. FDA reviewed the scientific literature and determined that 15 isolated and synthetic nondigestible carbohydrates had >1 beneficial physiologic effect. Included were cellulose, pectin, guar gum, locust bean gum, hydroxypropylmethylcellulose, beta-glucan, psyllium husk, mixed plant cell waters, arabinoxylan, alginate, inulin, and inulin-type fructans (fructooligosaccharides), high amylose starch/soluble corn fiber (resistant starch 2), galactooligosaccharides, polydextrose, and resistant maltodextrin/dextrin. Other isolated and synthetic NDCs are being evaluated for accepted physiologic effects. It should be noted the FDA has accepted the following physiological benefits: (1) Attenuation of blood glucose or insulin; (2) Lower fasting cholesterol concentration; (3) Improved laxation; (4) Increased gastrointestinal (GI) mineral absorption; and (5) Reduced energy intake 9 increased satiety. Production of fermentative end-products and change in the GI microbiota are not accepted physiological endpoints, although they are the endpoints of interest in prebiotic research.

Thus, in vitro studies that compare the fermentation aspects of isolated fibers are of mechanistic interest, they are not accepted as a measurement of prebiotic potential or fiber fermentation dynamics. Standard in vitro methods for fiber fermentability do not exist, but generally human donor fecal samples are used from healthy volunteers and then fiber samples incubated with the fecal slurries. Endpoints measured include fiber fermentation, gas production, pH changes, SCFA production, and changes in the microbiota. Recent work from our laboratory with resistant starches find differences in gas production with different resistant starches, but all were fermentable and promoted formation of beneficial SCFAs (Erickson et al., 2018). Examining the in vitro fermentation of commercial fiber supplements promoted as prebiotics found difference in production of *Bifidobacteria* and *Collinsella* (Carlson et al., 2017). All fibers were extensively fermented and produced beneficial SCFAs.

Earlier work found that shorter chain fructo-oligosaccharides exhibited more rapid fermentation than long-chain inulin in an in vitro fermentation (Stewart et al., 2008). Particle size and fraction of wheat bran also influenced SCFAs production in vitro (Stewart and Slavin, 2009). Five samples were compared: large-particle bran, small-particle bran, aleurone, coarse by-product, and fine-by product. Small/fine particle size increased SCFA production compared with large/coarse particle size. The molar percentage of butyrate at 24 h was significantly greater for large-particle bran than the other sample. Bran characteristics and composition should be considered when manufacturing foods due to the diversity of physiologic effects.

Fiber blends of fibers with different fermentation rates were compared in vitro (Koecher et al., 2014). FOS and inulin were fermented most readily, but when included in blends, fermentation was slower and less gas was produced. This study suggested that fiber blends may be more useful in enteral products to combine the desirable solubility profile and fermentation profile for optimum health.

What is a prebiotic?

The health effects of dietary fiber have been extensively reviewed and are accepted worldwide (Carlson et al., 2018). Prebiotics were first defined in 1995 and their definition has continued to evolve over time. Prebiotic dietary fibers are specific, microbiota-shaping compounds that function as a carbon source for growth of beneficial taxa, thus delivering a specific or selective change that confers the host health related to its metabolism of the prebiotic dietary fibers.

Most of the original prebiotics were also dietary fibers, including inulin, oligofructose, lactulose, and resistant starch. As there is no official method of becoming a prebiotic, other carbohydrates including galactooligosaccharides, trangalactooligosaccharides, polydextrose, wheat dextrin, acacia gum, psyllium, and partially hydrolyzed guar gum (PHGG), to name a few, have published studies that show that consumption of these fibers increases fecal levels of bifidobacterial or lactobacillus (Slavin, 2013). Often production of SCFAs in vitro have been used as a biomarker of prebiotic effect, since SCFAs are the byproduct of fiber fermentation.

Recent reviews have attempted to classify prebiotics based on in vitro studies and short-chain fatty acid production (Ashaolu et al., 2020). Isolated dietary fibers including resistant starch, pectin, hemicellulose, beta-glucan, and fructans were reviewed and their support for being a prebiotic offered. Regulatory acceptance of the importance of prebiotics beyond their ability to list their fiber content on the label is lacking.

The effect of fructans, prebiotics and fibres on the gut microbiome was reviewed (Swanson et al., 2020). Molecular studies confirmed the selective bifidogenic effect of fructans and galactooligosaccharides (GOS) in human population. Resistant starches, polydextrose and beta-glucan showed broader effects with more and different types of gut microbial species being enhanced, often including phylotypes of *Ruminococcaceae*. There was significant variation in magnitude of response and in individual responses to a specific fiber that may be due to numerous factors, including presence and relative abundance of microbial types, diet, genetics of the host, and intervention parameters, including intervention duration and fiber dose. The authors conclude that the field could benefit from a more systematic approach that will support defining the impact of prebiotics and fibers on the gut microbiome, identify biomarkers that link gut microbes to health, and address the personalized responses of an individual's microbiota to prebiotics and dietary fibers. Because the definition of prebiotics requires a health benefit, it is currently much easier to stay in the dietary fiber classification as definitions and regulatory approval are already realized.

Another recent review described dietary fibers as prebiotic, prebiotic candidates, or fibers without prebiotic potential (Rezende et al., 2021). Resistant oligosaccharides, fructans (FOS, oligofructose and inulin) and galactans are recognized in the literature as prebiotics according to the authors. Galactosides and some non-starch polysaccharides and resistant starches may act as prebiotic fibers or have prebiotic potential. They suggest that associated substances (non-carbohydrates) are considered dietary fibers but apparently have no prebiotic effect on humans. Intrinsic dietary fiber in plants with associated substances have never been studied for prebiotic potential, so this conclusion can be questioned. Proteins and fats have been described as prebiotics, so the lack of specificity for the term will continue to limit the labeling of products as "prebiotics" (Sanders et al., 2019).

So et al. (2018) published a systematic review and meta-analysis of dietary fiber intervention on gut microbiota composition in healthy adults. A total of 64 studies involving 2099 participants were included. Dietary fiber intervention resulted in higher abundance of *Bifidobacterium* spp. and *Lactobacillus* spp. as well as fecal butyrate concentration compared with placebo/low-fiber comparators. Subgroup analysis found that fructans and galacto-oligosaccharides led to significantly greater abundance of both *Bifidobacterium* spp. and *Lactobacillus* spp. compared with comparators. No differences in effect were found between fiber intervention and comparators for alpha-diversity, abundances of other prespecified bacteria, or other SCFA concentrations. The authors conclude that dietary fiber interventions, particularly for fructans and GOS lead to high fecal abundance of Bifidobacterium and Lactobacillus, but do not affect alpha-diversity. Most of the data has been generated because of the increased interest in prebiotics starting with their definition in 1995.

Health benefits of fiber, prebiotics, and probiotics?

Gastrointestinal health and the gut microbiome is a rapidly emerging field and fiber, prebiotics, and probiotics are all dietary components that play a role in maintaining a healthy gut microflora (Carlson and Slavin, 2016; Liu et al, 2022). Generally, we consider prebiotics as fiber-like compounds that generate bacteria in the colon while probiotics are live bacteria that you consume. Synbiotics are often combining prebiotics and probiotics. Unfortunately, the marketplace is awash in these products and little regulatory oversight is available for these products. Prebiotics operate in the safe space of dietary fiber, an accepted nutrient in the human diet that can be measured by accepted analytical tools and regulatory approval throughout the world. Enter postbiotics.

What are postbiotics?

A consensus statement by the International Scientific Association of Probiotics and Prebiotics (ISAPP) on the definition and scope of postbiotics defined postbiotic as a "preparation of inanimate microorganisms and/or their components that confers a health benefit on the host." (Salminen et al., 2021). The review paper lists past proposed definitions of the term postbiotic which perhaps explains the confusion in the literature and with consumers on this concept. Wegh et al. (2019) define postbiotics as functional bioactive compounds, generated in a matrix during fermentation which may be used to promote health. The term postbiotics can be regarded as an umbrella term for all synonyms and related terms of these microbial fermentation compounds. Therefore, postbiotics can include many different constituents including metabolites, SCFAs, microbial cell fractions, functional proteins, extracellular polysaccharides (EPS), cell lysates, teichoic acid, peptidoglycan-derived muropeptides and pili-type structures. The concept suggests that the microbes produced by fermentation of fiber are not the important mechanisms. The byproducts of fiber fermentation, for example, SCFAs are more important.

The need for established clinical markers and microbial endpoints are still essential to determine the causal role of the GI microbiota in protecting health (Deehan et al., 2017). Few studies collect fecal samples and even when fecal samples are collected, there is little information provided on the dietary exposure and the kinds and amounts of dietary fiber that were consumed. Collection of fecal samples to measure the gut microbiota with the suggestion that we can predict health from such information is premature and potentially misleading.

Links between dietary fiber intake and health outcomes

Intake of intrinsic fiber is linked to less risk of cardiovascular disease. Additionally, dietary fiber is accepted as an important piece in gut health and a first step in treating constipation and diarrhea. Fibers differ in their chemistry and there are some relationships between fiber chemistry and physiological effect. Other aspects of fiber, including fermentability, will also predict physiological effects.

Additional attributes of fiber, for example, viscosity, will determine the possibility that fiber will lower blood lipids or modulate blood glucose. Because of the inconsistency of fiber's effect on physiological parameters, the FDA now requires isolated fiber to show at least 1 physiological benefit before the fiber can list its content on the Nutrition Facts label.

Fiber intake and recommendations—mind the gap

We have long appreciated the importance of dietary fiber in the diet and have promoted the need to consume adequate dietary fiber across the lifecycle. Additionally, the importance of dietary fiber in treating diseases such as constipation and high blood lipids is well accepted. Yet intakes of dietary fiber are inadequate across the globe and despite public health guidance to eat more plant foods and fiber, usual intake is about half of recommended levels. In the past, public health measures to add deficient nutrients to the diet in foods has been promoted to solve dietary deficiencies. As more fibers gain acceptance by the FDA as having a physiological benefit, it might be considered to add fiber to popular foods or beverages to increase consumption of dietary fiber to recommended levels. The large individual variability in response to dietary fiber and concerns with intestinal gas and bloating must be considered as those are often reported as the reasons that consumers are not willing to consume more dietary fiber.

Conclusion

Even in 300 BC, Hippocrates said "wholemeal bread makes larger feces than refined bread." The importance of dietary fiber in laxation and bowel health has long been appreciated. More recently prospective, cohort studies find that higher intakes of dietary fiber are linked to decreased risk of developing cardiovascular disease. Recommendations for intake of dietary fiber in North America are based on these epidemiologic studies that find that 14 g dietary fiber/1000 kcalories is associated with less cardiovascular risk. Thus, the Nutrition Facts panel recommends 28 g of dietary fiber for the average consumer of 2000 kcalories. Usual intake of dietary fiber is approximately half that amount which makes dietary fiber a nutrient of concern. Although dietary guidance supports consumption of foods concentrated in dietary fiber, whole grains, legumes, vegetables, nuts, and fruit, many consumers turn to fiber supplements to obtain their needed fiber intake.

Current research interest in the effect of dietary fiber on the gut microbiota has expanded the interest of fiber in disease prevention into fields such as immunology, brain health, and diabetes control. Fiber is active throughout the gastrointestinal tract, but much of the current interest is in the fermentation of dietary fiber in the gut. Although earlier thinking supported that certain microbiota were "healthy," the prebiotic concept, more recent thinking is that the products of fiber's fermentation are important players in fiber's protective role in disease prevention. SCFAs from fiber fermentation help maintain the integrity of the colonic cells and are thought to trigger a cascade of additional health benefits.

We know that intake of dietary fiber is protective, but we also know that fibers cover a range of physiological benefits, solubility, fermentability, and viscosity, for example. Thus, public health guidance supports that consumers choose a variety of fiber-rich foods to obtain the Adequate Intake (AI) daily. Ideally, fiber should be provided by whole foods, but since dietary fiber intake continues to lag significantly behind dietary recommendations, isolated fibers with health benefits supported by research trials are also part of the solution to increasing dietary fiber intakes to recommended levels.

References

Alexander, C., Swanson, K.S., Fahey, G.C., Garleb, K.A., 2019. Perspective: physiologic importance of short-chain fatty acids from nondigestible carbohydrate fermentation. Adv. Nutr. 10, 576–589.

Ashaolu, T.J., Ashaolu, J.O., Adeyeye, S.A.O., 2020. Fermentation of prebiotics by human colonic microbiota in vitro and short-chain fatty acids production: a critical review. J. Appl. Microbiol. 130, 677–687.

Bird, A.R., Conlon, M.A., Christophersen, C.T., Topping, D.L., 2010. Resistant starch, large bowel fermentation and a broader perspective of prebiotics and probiotics. Benef. Microbes 1, 423–431.

Carlson, J.L., Slavin, J.L., 2016. Health benefits of fibre, prebiotics and probiotics: a review of intestinal health and related health claims. Qual. Assur. Saf. Crop Foods 8, 539–554.

Carlson, J.L., Ericonson, J.M., Hess, J.M., Gould, T.J., Slavin, J.L., 2017. Prebiotic dietary fiber and gut health: comparing the in vitro fermentations of beta-glucan, inulin and xylooligosaccharide. Nutrients 9, 1361.

Carlson, J.L., Erickson, J.M., Lloyd, B.B., Slavin, J.L., 2018. Health effects and sources of prebiotic dietary fiber. Curr. Dev. Nutr. 2, NZY005.

Cummings, J.H., Southgate, D.A.T., Branch, W., et al., 1978. Colonic response to dietary fibre from carrot, cabbage, apple, bran, and guar gum. Lancet 1 (8054). January 7, 1978.

Cummings, J.H., Bingham, S.A., Heaton, K.W., Eastwood, M.A., 1992. Fecal weight, colon cancer risk, and dietary intake of nonstarch polysaccharides (dietary fiber). Gastroenterology 103, 1783–1789.

Cummings, J.H., 1993. The effect of dietary fiber on fecal weight and composition. In: Spiller, G.A. (Ed.), CRC Handbook of Dietary Fiber in Human Nutrition. CRC Press, Boca Raton, FL, pp. 263–333.

Dai, F., Chau, C., 2017. Classification and regulatory perspectives of dietary fiber. J. Food Drug Anal. 25, 37–42.

Deehan, E.C., Duar, R.M., Armet, A.M., et al., 2017. Modulation of the gastrointestinal microbiome with nondigestible fermentable carbohydrates to improve human Health. Microbiol. Spectr. 5 (5). BAD-0019.

Dobranowski, P.A., Stintzi, A., 2021. Resistant starch, microbiome, and precision modulation. Gut Microb. 13, e1926842.

Erickson, J.M., Carlson, J.L., Stewart, M.L., Slavin, J.L., 2018. Fermentability of novel type-4 resistant starches in in vitro system. Foods 7, 18.

Fuller, S., Beck, E., Salman, H., Tapsell, L., 2016. New horizons for the study of dietary fiber and health: a review. Plant Foods Hum. Nutr. 71, 1–12.

Grabitske, H.A., Slavin, J.L., 2008. Laxation and the like. Nutr. Today 43, 193–198.

Grabitske, H.A., Slavin, J.L., 2009. Gastrointestinal effects of low-digestible carbohydrates. Crit. Rev. Food Sci. Nutr. 49, 327–360.

Hipsley, E.H., 1953. Dietary "fibre" and pregnancy toxaemia. Br. Med. J. 2, 420–422.

Jones, J.M., 2013. Dietary fiber future directions: integrating new definitions and findings to inform nutrition research and communication. Adv. Nutr. 4, 8–15.

Klosterbuer, A., Roughhead, Z.F., Slavin, J., 2011. Benefits of dietary fiber in clinical nutrition. Nutr. Clin. Pract. 26, 625–635.

Koecher, K.J., Noack, J.A., Timm, D.A., et al., 2014. Estimation and interpretation of fermentation in the gut: coupling results from a 24 h batch in vitro system with fecal measurements from a human intervention feeding study using fructo-oligosaccharides, inulin, gum acacia, and pea fiber. J. Agric. Food Chem. 62, 1332–1337.

Korczak, R., Slavin, J.L., 2020. Definitions, regulation, and new frontiers for dietary fiber and whole grains. Nutr. Rev. 78, 6–12.

Lampe, J.W., Fredstrom, S.B., Slavin, J.L., Potter, J.D., 1993. Sex differences in colonic function: a randomized trial. Gut 34, 531–536.

Li, C., Hu, Y., 2022. New definition of resistant starch types from the gut microbiota perspectives—a review. Crit. Rev. Food Sci. Nutr. https://doi.org/10.1080/10408398.2022.2031101.

Liu, Y., Wang, J., Wu, C., 2022. Modulation of gut microbiota and immune system by probiotics, pre-biotics, and post-biotics. Front. Nutr. 8, 634897.

McNamara, E.A., Levitt, M.D., Slavin, J.L., 1986. Breath hydrogen and methane: poor indicators of apparent digestion of soy fiber. Am. J. Clin. Nutr. 43, 898–902.

Patterson, M.A., Maiya, M., Stewart, M.L., 2020. Resistant starch content in foods commonly consumed in the United States: a narrative review. J. Acad. Nutr. Diet. 120, 230–244.

Rezende, E.S.V., Lima, G.C., Naves, M.M.V., 2021. Dietary fibers as beneficial microbiota modulators: a proposed classification by prebiotic categories. Nutrition 89, 111217.

Salminen, S., Collado, M.C., Endo, A., et al., 2021. The International Scientific Association of Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of postbiotics. Nat. Rev. Gastroenterol. Hepatol. 18, 649–667.

Sanders, M.E., Merenstein, D.J., Reid, G., et al., 2019. Probiotics and prebiotics in intestinal health and disease: from biology to the clinic. Nat. Rev. Gastroenterol. Hepatol. 16, 605–616.

Slavin, J.L., Greenberg, N.A., 2003. Partially hydrolyzed guar gum: clinical nutrition uses. Nutrition 19, 549–552.

Slavin, J.L., Brauer, P.M., Marlett, J.A., 1981. Neutral detergent fiber, hemicellulose and cellulose digestibility in human subjects. J. Nutr. 111, 287–297.

Slavin, J.L., Savarino, V., Paredes-Diaz, A., Fotopoulos, G., 2009. A review of the role of soluble fiber in health with specific reference to wheat dextrin. J. Int. Med. Res. 37, 1–17.

Slavin, J.L., 1987. Dietary fiber: classification, chemical analyses, and food sources. J. Am. Diet Assoc. 87, 1164–1171.

Slavin, J.L., 2013. Fiber and prebiotics: mechanisms and health benefits. Nutrients 5, 1417–1435.

So, D., Whelan, K., Rossi, M., et al., 2018. Dietary fiber intervention on gut microbiota composition in healthy adults: a systematic review and meta-analysis. Am. J. Clin. Nutr. 107, 965–983.

Stephen, A.M., Cummings, J.H., 1979. Water-holding by dietary fibre in vitro and its relationship to faecal output in man. Gut 20, 722–729.

Stephen, A.M., Cummings, J.H., 1980. Mechanism of action of dietary fibre in the human colon. Nature 284, 283–284.

Stephen, A.M., Champ, M.M., Cloran, S.J., et al., 2017. Dietary fibre in Europe: current state of knowledge on definitions, sources, recommendations, intakes and relationships to health. Nutr. Res. Rev. 30, 149–190.

Stewart, M.L., Slavin, J.L., 2009. Particle size and fraction of wheat bran influence short-chain fatty acid production in vitro. Br. J. Nutr. 102, 1404–1407.

Stewart, M.L., Timm, D.A., Slavin, J.L., 2008. Fructooligosaccharides exhibit more rapid fermentation than long-chain inulin in an in vitro fermentation system. Nutr. Res. 28, 329–334.

Swanson, K.S., de Vos, W., Martens, E.C., et al., 2020. Effect of fructans, prebiotics and fibres on the human gut microbiome assessed by 16W rRNA-based approaches: a review. Benef. Microbes 11, 101–129.

Tucker, D.M., Sandstead, H.H., Logan, G.M., et al., 1981. Dietary fiber and personality factors as determinants of stool output. Gastroenterology 81, 879–883.

Wegh, C.A.M., Geerlings, S.Y., Knol, J., et al., 2019. Postbiotics and their potential applications in early life nutrition and beyond. Int. J. Mol. Sci. 20, 4673.

Williams, B.A., Mikkelsen, D., Flanagan, B.M., Gidley, M.J., 2019. "Dietary fibre": moving beyond the "soluble/insoluble" classification for monogastric nutrition, with an emphasis on humans and pigs. J. Anim. Sci. Biotechnol. 10, 45.

Energy requirements

Susan B. Roberts and Rachel E. Silver, Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, Medford, MA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of W.P.T. James, Energy Requirements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 186–192, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00090-8.

Introduction 217
Estimated energy requirements determined using the doubly labeled water method 218
The components of energy requirements 218
Basal metabolic rate 218
Postprandial thermogenesis 220
Physical activity 221
Extra energy costs of growth, pregnancy and lactation 221
Conclusion 222
References 222

Key points

- The doubly labeled water method has become the gold standard for determining energy requirements because it allows for measurement of energy requirements in individuals leading their usual daily lives.
- The major components of energy requirements are: basal metabolic rate, thermic effect of feeding, energy expenditure for physical activity, and (in children) energy needs for growth.
- The major factors influencing energy requirements are body size, level of physical activity, and age and sex; equations are presented that allow calculation of typical requirements based on these parameters.

Glossary

BMR, RMR, and REE Basal Metabolic Rate is the rate of energy expenditure in the thermo-neutral state 12–18 h after last eating. This term is often used interchangeably with Resting Metabolic Rate and Resting Energy Expenditure, although the extended period of fasting required for a BMR measurement is not necessary for RMR and REE, which typically have overnight fasting that is not necessarily >12 h

MET The Metabolic Equivalent is a unit of energy expenditure expressed relative to resting energy expenditure that provides a simple index of the relative strenuousness of different activities

PAL Physical activity level: The ratio of the total energy expenditure to BMR or REE over an extended period, typically assessed as

Doubly Labeled Water Method The DLW method is a stable isotope method that estimates total energy expenditure over periods of 1–3 weeks by measurement of the differential rates of disappearance of two stable isotopes of water

Introduction

The term energy requirement is the amount of metabolizable[1] energy from food needed to match the body's energy expenditure needed to maintaining all its functions and engage in all forms of physical activity, plus the additional energy needed for growth in children, for pregnancy and lactation, when recovering from some illnesses, and when exposed to cold. The amount of food energy needed by an individual depends on his/her size and several other factors. For example if the individual has overweight or obesity, the food needs to maintain the extra weight will be greater than if they are underweight for the same level of physical activity.

[1]Metabolizable energy is the amount of energy available to the body after obligatory losses in stool and urine. Typically, approximately 90% of the total energy in food is available as metabolizable energy.

Estimated energy requirements determined using the doubly labeled water method

Measures of total energy expenditure estimated over 2–3 weeks can be obtained by the use of the double-labeled water technique (Speakman and Hambly, 2016), which relies on the difference in labeling of body water with two stable isotopes of water, deuterium and ^{18}O. The differential dilution of deuterium and $^{18}0$ in urinary water over time is monitored in urine during a 1–3-week period following a single oral dose of $D_2{}^{18}O$. The ^{18}O concentration in urine decreases more rapidly than the deuterium because the oxygen in water exchanges rapidly with the body's bicarbonate pool, which is turning over rapidly as carbon dioxide is produced by tissue metabolism and excreted in expired air. Thus, the difference in rate constants for disappearance of ^{18}O and deuterium provides a measure of the rate of carbon dioxide production. Carbon dioxide production can then be converted to a rate of energy expenditure with knowledge of the balance of macronutrients being oxidized, which can be estimated from food records. The technique is an important advance over previous methods for determining energy requirements, because it allows scientists to measure energy expenditure in the free living state while participants lead their usual daily lives.

The doubly labeled water method has been adopted as the gold standard for assessment of energy requirements both in the US and Europe (e.g. Institute of Medicine, Food and Nutrition Board, 2002). For individuals where weight stability is optimal, energy requirements are equal to total energy expenditure in the weight stable state. For children and pregnant and lactating mothers, additional allowance needs to be made for the cost of tissue deposition and milk excretion, but in humans this amount is usually a small fraction of total energy requirements. Thus, using doubly labeled water as the basis for energy requirements has allowed for more accurate determination of energy requirements than in the past, when methods such as activity diaries combined with minute-to-minute estimates of the energy costs of different physical activities were used. A limitation of using the doubly labeled water method to estimate energy requirements is that most of the available data was not obtained for the purpose of measuring energy requirements, but rather for other scientific purposes relating to energy expenditure, and thus the population group means may not be normative.

The US Institute of Medicine complied all available doubly labeled water for a comprehensive reevaluation of energy requirements in 2002 (Institute of Medicine, Food and Nutrition Board, 2002), and are currently updating this information. The available equations from 2002 were generated from all available data at the time, with empiracle approaches used to add energy needed for growth when relevant for different population groups. A summary of the equations obtained are summarized in **Table 1**. Ages from birth to old age are covered, and quartiles of energy expenditure were derived for individuals aged 3 years and above, allowing for estimated typical energy requirements to be derived for individuals with different lifestyles.

The components of energy requirements

As noted above, total energy expenditure is the major component of energy requirements and has three main components: the BMR, postprandial thermogenesis, and energy expenditure for physical activity. BMR usually amounts to 50–70% of an individual's total energy expenditure. Postprandial thermogenesis amounts to approximately 10% of total energy expenditure, with this amount used for the metabolic cost of processing, i.e. eating, absorbing, transporting, and storing food. The remaining energy is used for physical activity and non-exercise activity thermogenesis. In addition to total energy expenditure, estimated energy requirements allow for energy deposition in tissue and milk production. These separate components of estimated energy requirements are summarized below.

Basal metabolic rate

The rate at which the body burns its own stored fuels in the fasted, resting, and relaxed state, i.e. in the basal state in a thermoneutral environment, is called the basal metabolic rate (BMR). The process of oxidation of stored fuels involves a series of enzymatically controlled biochemical reactions leading eventually to the combination of oxygen with the carbon and hydrogen components of the body's fuels thereby yielding carbon dioxide and metabolically derived water. The incompletely oxidized nitrogen is excreted in urine primarily as urea, after synthesis by the liver. The intermediate steps in the metabolism of the body's fuels are linked biochemically to drive the generation of phosphate-containing organic molecules, such as adenosine triphosphate (ATP), which in turn serve as the direct energy sources for all the body's cell activities, including the synthesis of complex molecules, the maintenance of tightly controlled ionic gradients in the cell, and the excretion of ions and molecules outside the cell. Thus, the oxygen being taken up by the lungs reflects the tissue metabolism of the fuels needed to regenerate the ATP used up in either biochemical "internal" work or mechanical external work undertaken by the body's muscles.

BMR varies with the age of the individual, at least in part because of the varying sizes of metabolically very different organs at different ages. Thus, a child has a relatively large brain, liver, and intestine with a higher metabolic rate per kilogram of body weight than a more muscular adult. Body fat cells are metabolically active but contain a substantial amount of inert fat, so that the larger fat mass of a woman results in a lower BMR per unit body weight than a man. If energy expenditure is adjusted for the amount of fat-free mass and fat mass, then the metabolic rate of men and women is similar. As men and women age, they tend to lose fat free mass and store extra fat, and some organs including the brain can shrink, so the BMR on a weight basis falls with age.

Table 1 Equations to estimate energy requirements by age-group and gender from US Institute of Medicine Dietary Reference Intakes, which were derived from doubly labeled water measures of energy expenditure (Institute of Medicine, Food and Nutrition Board, 2002).

EER for boys and girls aged 0–3 years:
0–3 months = (89 × weight − 100) + 175
4–6 months = (89 × weight − 100) + 56
7–12 months = (89 × weight − 100) + 22
13–36 months = (89 × weight − 100) + 20
EER for boys 3–18 years:
EER age 3–8 = 88.5 − (61.9 × age) + PA × (26.7 × weight) + (903 × height) + 20
EER age 9–18 = 88.5 − (61.9 × age) + PA × (26.7 × weight) + (903 × height) + 25
Where ages is in years, weight in kg, height in m and PA is physical activity coefficient:
PA = 1.00 if PA is 1.0 − <1.4 (sedentary)
PA = 1.13 if PA is >1.4 − <1.6 (low active)
PA = 1.26 if PA is >1.6 − <1.9 (active)
PA = 1.42 if PA is >1.9 − <2.5 (very active)
EER for girls 3–8 years:
EER age 3–8 = 135.2 − (30.8 × age) + PA × (10.0 × weight) + (934 × height) + 20
EER age 9–18 = 135.2 − (30.8 × age) + PA × (10.0 × weight) + (934 × height) + 25
Where ages is in years, weight in kg, height in m and PA is physical activity coefficient:
PA = 1.00 if PA is 1.0 − <1.4 (sedentary)
PA = 1.16 if PA is >1.4 − <1.6 (low active)
PA = 1.31 if PA is >1.6 − <1.9 (active)
PA = 1.56 if PA is >1.9 − <2.5 (very active)
EER for adult men 19 years and older (includes normal weight and individuals with overweight and obesity):
EER = 864 − (9.72 × age) + PA × (14.2 × weight) + 503 × height
Where ages is in years, weight in kg, height in m and PA is physical activity coefficient:
PA = 1.00 if PA is 1.0 − <1.4 (sedentary)
PA = 1.11 if PA is >1.4 − <1.6 (low active)
PA = 1.25 if PA is >1.6 − <1.9 (active)
PA = 1.48 if PA is >1.9 − <2.5 (very active)
EER for adult women 19 years and older (includes normal weight and individuals with overweight and obesity):
EER = 387 − (7.31 × age) + PA × (10.9 × weight) + 660.7 × height
Where ages is in years, weight in kg, height in m and PA is physical activity coefficient:
PA = 1.00 if PA is 1.0 − <1.4 (sedentary)
PA = 1.12 if PA is >1.4 − <1.6 (low active)
PA = 1.27 if PA is >1.6 − <1.9 (active)
PA = 1.45 if PA is >1.9 − <2.5 (very active)
EER for pregnancy aged 19–50 years
1st trimester: adult EER + 0 energy expended in pregnancy + 0 energy deposition
2nd trimester: adult EER + 160 kcal/day energy expended in pregnancy + 180 kcal/day energy deposition
3rd trimester: adult EER + 272 kcal/day energy expended in pregnancy + 180 kcal/day energy deposition
EER for lactation aged 19–50 years
0–6 months: adult EER + 500 kcal/day milk energy − 170 kcal/day body fat loss
6–12 months: adult EER + 400 kcal/day milk energy − 0 kcal/day body fat loss

If oxygen consumption and carbon dioxide production are measured using an "indirect" calorimeter system (so called because it does not measure energy expenditure directly), classic equations originally developed by J.S. de Weir can be used to calculate energy expenditure (de Weir, 1949). The conditions of the measurement are lying quietly and awake, fasting 12–18 h in a thermoneutral environment, BMR can be accurately calculated:

$$\mathrm{RMR(kcal\,/\,day)} = 1440(3.941\ VO2\ +\ 1.106\ VCO2)$$

where VO_2 and VCO2 are measured in liters per minute.

Equations have also been developed to predict BMR from readily measured anthropometric variables, and **Table 2** shows equations that can be used to calculate based on biological sex, weight, height and age (Henry, 2005). There is a range of BMR amounting to ±20% of the mean value at each weight. Thus, in a young 25-year-old woman of 55 kg, the anticipated mean BMR is 1305 kcal/day but could vary under normal conditions from 1063 to 1547 kcal/day. The differences in BMR of different individuals of the same weight in part reflects differences in their proportion of lean to fat tissues. Approximately 40% of the BMR variation between sexes and the age of individuals may be explained by differences in the size of the body organs, for example, liver, intestine, and

Table 2 BMR equations for different age groups and genders from Henry (2005).

Gender	*Age (years)*	*BMR (kcal/day)*		
		Coefficient weight (kg)	*Coefficient height(m)*	*Constant*
Males	<3	28.2	859	−371
	3–10	15.1	313	306
	10–18	15.6	266	299
	18–30	14.4	313	133
	30–60	11.4	541	−137
	>60	11.4	541	−256
Females	<3	30.4	703	−287
	3–10	15.9	210	349
	10–18	9.40	249	462
	18–30	10.4	615	−282
	30–60	8.18	502	−11.6
	>60	8.52	421	10.7

Reproduced from Department of Health, 1991. Dietary reference values for food energy and nutrients for the United Kingdom. Report on Health and Social Subjects, vol. 41. HMSO, London.

muscle, but there is a residual difference between individuals who seem to be explicable only in terms of differences in the rate at which different tissues metabolize fuel. This is controlled in part by the circulating concentration of thyroid hormones. Adults with an above-average level of circulating thyroid hormones tend to have a higher but still normal BMR. Smokers' BMRs are approximately 5% above normal, perhaps because of activation of the sympathetic nervous system.

Young women show a swing in BMR that is at its lowest in the late follicular phase of the menstrual cycle, just before ovulation. On ovulation, the basal body temperature rises rapidly by approximately 0.5 °C and the BMR immediately increases but rises further to a peak in the later luteal phase. This metabolic swing of ±5% is independent of changes in food intake, but the recognized 5–10% fall in intake during the follicular phase with a similar rise in the luteal phase may accentuate the hormonally dependent swing in metabolism. The effects of contraceptives that inhibit ovulation and the subsequent rise in basal temperature are not well documented. The previous day's food intake may somewhat affect the BMR, especially if there has been substantial overeating or undereating. In addition, the mixture of fuels combusted during fasting is influenced by the proportion of the previous 3–4 days' intakes derived from carbohydrate; much of the glucose from glycogen is metabolized in the fasting state if carbohydrate intake was previously high. When glycogen stores in the liver near exhaustion, the body switches to using body fat with a small fall in carbon dioxide output; a carbohydrate-rich diet induces a slightly higher fasting metabolic rate probably because of a slight induction of thyroid metabolism by dietary carbohydrates.

The BMR falls by 2–5% when individuals transfer to live in a tropical warm environment; in uninsulated houses, seasonal BMR can increase 5–10% in winter are seen as in Japan before World War II. The BMR formulas shown in **Table 2** ignore any temperature effects. The observed lower BMR of some people living in the tropics may also reflect the effects of malnutrition. Poor nutrition can have both immediate and longer-term effects in lowering the BMR. After 4 days of semistarvation the BMR falls, and after 2 weeks the BMR is approximately 15% lower as the body's organs become more efficient. More prolonged or severe semistarvation induces a progressive loss of the body lean tissues as well as fat, and the BMR therefore continues to decline in proportion to the loss of lean tissues. Body weight can eventually stabilize at a new low level and, if the physical activity is also reduced, semistarved volunteers can come back into energy balance on as little as 50% of their initial intake. However, this requires a large loss of weight and marked lethargy if energy balance is to be preserved on such a low intake.

Postprandial thermogenesis

The increase in oxygen uptake and carbon dioxide production after a meal, known as postprandial thermogenesis, has also been described as the specific dynamic action of food, dietary-induced thermogenesis, and the thermic effect of feeding (de Jonge and Bray, 1997). The maximum postprandial thermogenesis occurs after ingesting protein, and both glucose and fat have only a small effect, with a summary of postprandial thermogenesis studies indicating that 13.2% of protein metabolizable energy is lost to postprandial thermogenesis compared to 5.1% for fat and 5.4% for carbohydrate, while alcohol induces a variable 0–8% effect (Eisenstein et al., 2002). Nevertheless, the overall effect of protein on energy expenditure is relatively modest in diets of typical composition, in which protein is likely to be in the range 10–25% of energy. Some additional dietary components increase metabolism: caffeine equivalent to two cups of tea induces a 1–3% increase and spices, as in Indian curry, by up to 25% in some reports.

Differences in postprandial energy expenditure have been sought as an explanation for the propensity of some individuals to obesity but there is no strong evidence to support this thesis. A proportion of obese subjects have a reduced metabolic response to a meal; this effect may prove to depend on the degree of abdominal insulation because the response is reduced if volunteers are swathed in insulation to reduce the abdominal heat loss, thereby increasing the temperature of the blood entering and leaving the liver. Insulin resistance may also explain the difference. Lactating mothers (and pregnant women) also have a lower postprandial thermogenesis that returns to normal after they have stopped breast-feeding. Smoking and postprandial thermogenesis interact synergistically so the thermic output after a meal is enhanced. The small postprandial response during lactation is consistent with that observed in many species of animal in which brown adipose tissue is used as the organ for modulating heat production as a mechanism to maintain body temperature. However, this organ is normally not very active in humans although recent analyses with new scanning techniques have shown that its activity does continue in adults and small seasonal changes in BMR may reflect changes in brown fat metabolism. So postprandial thermogenesis as well as the response to cold in adults may involve brown adipose activity.

Prolonged overfeeding may produce an increase in postprandial response providing the intensity of overfeeding (especially with carbohydrate) is high. Thus, progressive overconsumption of 1500 kcal/day has been reported to lead to up to a 33% increase in BMR combined with postprandial thermogenesis. Nevertheless, this apparent mechanism for dissipating excess energy is limited because energy is stored, at least two-thirds as fat and the rest as lean tissue. The majority of the increased metabolism may be accounted for the cost of fat synthesis from carbohydrate, although the human capacity to transform carbohydrate into fat is relatively limited unless carbohydrate overfeeding is extreme, preference being given to the selective storage of the fat component of the ingested energy.

Physical activity

The energy cost of physical activity can be measured with calorimetry systems that collect expired gasses. Based on these methods, typical costs of different physical activities have been generated. Weight-bearing movement and antigravitational moves, for example, walking up a hill with a load, are particularly energetic. The simplest way of estimating individual costs for specific tasks involves the use of extensive tables listing the energy cost of different movements expressed as metabolic equivalents or METs, as summarized in **Table 3A** (Ainsworth et al., 2011). For simplicity, METs are energy expenditure values expressed relative to energy expenditure in the resting state. METs for typical activities range from 1 to more than 10 depending on the amount of work being performed. Whole body calorimetry performed over 24 h can also measure energy expenditure for physical activity, as well as non-exercise activity thermogenesis, by correlating fluctuations in energy expenditure with motion sensors embedded in the chamber. There are also internationally recognized questionnaires for assessing individuals' physical activity patterns and sophisticated accelerometers and wrist-worn activity sensors worn by the individual, with a standardization of their energy equivalence from measures of oxygen uptake. Accelerometers, and commercial variants such as smart watches, are increasingly used to measure the pattern and energy equivalence of daily activities in population studies.

The ratio of the total daily energy expenditure to the BMR is designated the physical activity level (PAL). Physical activity is important to general health so it is desirable that the overall PAL of individuals should be high. However, very high minute-to-minute MET values cannot be sustained, with the consequence that the usual average daily PAL level is in the 1.0–2.0 range. Even very active individuals are unlikely to have a daily PAL value over 2.5. **Table 3**B also shows a common definition for activity level in relation to PAL level.

Extra energy costs of growth, pregnancy and lactation

The cost of growth amounts to 2.5–6.0 kcal/g of new tissue deposited; the value being higher if fat with little lean tissue is laid down. A newborn has a high energy requirement of approximately 110 kcal/kg with a cost of weight gain amounting to 6 kcal/g, but by 1

Table 3A Typical energy expenditures and MET ratios (activity energy expenditure/resting energy expenditure) for some different activities from Ainsworth et al. (2011).

Activity	*MET value*
Inactivity, lying quietly; or sleeping	1.0
Inactivity, sitting or standing quietly	1.3
Walking, <2.0 mph	2.0
Walking, 2.8–3.2 mph, level firm ground	3.5
Walking, carrying 1–15 lb load	5.0
Running, 4 mph	6.0
Swimming, leisurely pace	6.0
Running, 10 mph (6 min mile)	14.5

Table 3B Classification system for level of physical activity in adults in relation to quartiles of physical activity level as defined in doubly labeled water studies (Institute of Medicine, Food and Nutrition Board, 2002).

Activity level classification	*PAL range*
Sedentary	1.0 – <1.4
Low active	>1.4 – <1.6
Active	>1.6 – <1.9
Very active	>1.9 – <2.5

year of age, the total daily requirement has fallen to approximately 80 kcal/kg as growth slows with growth now costing 2.5 kcal/g. Without sufficient energy, a child will fail to grow, but the causes of growth failure usually also correlate with deficiencies of protein and or vitamins and minerals or to infection, rather than only to a lack of dietary energy. Adolescents, particularly boys, who are physically very active, may have a high demand for energy. However, the actual cost of even rapid growth rates at this age is modest relative to total energy expenditure.

Traditionally, pregnancy is considered, incorrectly, a time of great demand for food. Good nutrition is extremely important and a weight gain in pregnancy of approximately 11–16 kg for a woman with a Body Mass Index in the healthy range (18.5–24.9 kg/m^2), and 5–9 kg for mothers who have obesity when they become pregnant, is considered appropriate for reducing the risk of maternal and fetal complications and preterm and low birth weight babies (Institute of Medicine, 2009). With a weight gain of 12 kg for women with a healthy BMI, the net increase in energy requirements taking into account changes in BMR and tissue deposition is 0, 320 and 452 kcal/day, respectively, for first, second and third trimesters, in addition to energy requirements prior to pregnancy. Women who enter pregnancy with obesity are recommended to gain less weight, with a corresponding reduction in additional energy requirements to support weight gain and fetal growth.

In lactation the net extra energy needed to cover the energy in the milk and the cost of producing, taking into account fat mobilization from body stores of an average 170 kcal/day on average, is estimated at 330 kcal/day in the first 6 months of lactation. During later lactation, milk production typically falls and maternal body weight stabilizes, with an estimated increment in energy requirements for lactation of 400 kcal/day.

Conclusion

With the widespread availability of the doubly labeled water method, nutrition science has an objective way to create estimated energy requirement equations for different population groups including adults of all ages, children, and pregnant and lactating mothers. This advance will increasingly have important consequences for next-generation dietary reference intakes for other nutrients, because energy requirements define the total amount of food energy that can be consumed, which in turn has implications for the types of foods that will ensure adequacy of macronutrients and micronutrients. Improving the scientific basis of nutritional requirements in this way will support public health initiatives to improve human health through healthy nutrition.

References

Ainsworth, B.E., Haskell, W.L., Herrmann, S.D., Meckes, N., Bassett Jr., D.R., Tudor-Locke, C., Greer, J.L., Vezina, J., Shitt-Glover, M.C., Leon, A.S., 2011. Compendium of physical activities. A second update of codes and MET values. Med. Sci. Sports Exerc. 43, 1575–1581.

de Jonge, L., Bray, G.A., 1997. The thermic effect of food and obesity: a critical review. Obes. Res. 5, 622–631.

Eisenstein, J., Roberts, S.B., Dallal, G., Saltzman, E., 2002. High-protein weight loss diets: are they safe and do the work? A review of the experimental and epidemiological data. Nutr. Rev. 60, 189–200.

Henry, C.J., 2005. Basal metabolic rate studies in humans: measurement and development of new equations. Publ. Health Nutr. 8, 1133–1152.

Institute of Medicine, 2009. Weight Gain During Pregnancy: Re-examining the Guidelines. National Academies of Sciences, Washington, DC. www.iom.edu/Reports/../Weight-Gain-During-Pregnancy-Reexamining-the-Guidelines.aspx.

Institute of Medicine, Food and Nutrition Board, 2002/2005. Dietary Reference Intakes for Energy, Carbohydrate, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids. The National Academies Press, Washington, D.C.

Speakman, J.E., Hambly, C., 2016. Using doubly-labelled water to measure free-living energy expenditure: some old things to remember and some new things to consider. Comp. Biochem. Physiol. Mol. Integr. Physiol. 202, 3–9.

Weir, J.B., 1949. New methods for calculating metabolic rate with special reference to protein metabolism. J. Physiol. 109 (1–2), 1–9.

Fatty acids: Omega-3 polyunsaturated

A.P. Simopoulos, The Center for Genetics, Nutrition and Health, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of A.P. Simopoulos, Omega-3 polyunsaturated fatty acids: Nutrigenetic and Nutrigenomic Aspects in the Determination of Dietary Requirements, Development, and Chronic Diseases, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 405–412, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00099-4.

Introduction 224
Eicosanoid Metabolism and Biological Effects of Omega-6 and Omega-3 Fatty Acids 224
Genetic Variation: Nutrigenetics 226
Genetic Variants, FADS1 and FADS2 in Estimating Nutritional Requirements of Omega-3 and Omega-6 Fatty Acids 226
Genetic Variants of the FADS1 and FADS2 Gene Cluster Influence Omega-6 and Omega-3 Fatty Acids Composition in Both Plasma and Red Cell Membrane Phospholipids During Pregnancy and Lactation 226
Genetic Variants in Omega-6 and Omega-3 Fatty Acids Metabolism and IQ 227
Genetic Variants in FADS1 and FADS2 and CHD Risk 228
Genetic Variants in the 5-LO, the Role of Omega-6 and Omega-3 Fatty Acids in CHD 229
Genetic Variants in the 5-LO, Omega-6 Fatty Acids, and Breast Cancer 229
Genetic Variants of Cyclooxygenase-2 (COX-2) and the Protective Effect of Long Chain Omega-3 Fatty Acids in Cancer of the Prostate 230
Nutrigenomics: The Role of Omega-6 and Omega-3 Fatty Acids in Gene Expression 230
Anti-Inflammatory Aspects: Omega-3 Fatty Acids Downregulate the Expression of Genes Involved in Inflammation and Obesity 230
Summary and Conclusions 231
Further Reading 231

Glossary

Allele Alternative form of a genetic locus; a single allele for each locus is inherited from each parent (e.g., at a locus for eye color the allele might result in blue or brown eyes).

Gene The fundamental physical and functional unit of heredity. A gene is an ordered sequence of nucleotides located in a particular position on a particular chromosome that encodes a specific functional product (i.e., a protein or RNA molecule).

Gene expression The process by which a gene's coded information is converted into the structures present and operating in the cell. Expressed genes include those that are transcribed into mRNA and then translated into protein and those that are transcribed into RNA but not translated into protein (e.g., transfer and ribosomal RNAs).

Genetic polymorphism Difference in DNA sequence among individuals, groups, or populations (e.g., genes for blue eyes versus brown eyes).

Genotype The genetic constitution of an organism, as distinguished from its physical appearance (its phenotype).

Nutrigenetics Refers to an individual's specific response to diet due to genetic variants or polymorphisms (i.e., individuals responding differently to the same diet by having different levels of, for example, serum cholesterol and blood pressure because of genetic variation).

Nutrigenomics Refers to the role of nutrients in gene expression (i.e., polyunsaturated fatty acids suppress fatty acid synthase (mRNA) gene expression).

Polymorphism Difference in DNA sequence among individuals that may underlie differences in health. Genetic variations occurring in more than 1% of a population would be considered useful polymorphisms for genetic linkage analysis.

Abbreviations

AA Arachidonic acid
AI Adequate intake
ALA α-Linolenic acid
CAD Coronary artery disease
CHD Coronary heart disease
COX-2 Cyclooxygenase-2

CRP C-reactive protein
CTSS Cysteine protease cathepsin S
DHA Docosahexaenoic acid
EFA Essential fatty acid
EPA Eicosapentaenoic acid
EPG Ethanolamine phosphoglyceride
FADS Fatty acid desaturase
HDL High-density lipoprotein
IL Interleukin
LA Linoleic acid
LC-PUFA Very-long-chain polyunsaturated fatty acid
5-LO 5-Lipoxygenase
NF-κB Nuclear factor-Kappa-B
PAI-1 Plasminogen activator inhibitor factor-1
PG Prostaglandin
PLAVR Plasminogen activator urokinase receptor
PPARα Peroxisome proliferator-activated receptor alpha
PUFA Polyunsaturated fatty acid
TNF Tumor necrosis factor
VLDL Very-low-density lipoprotein

Introduction

Approximately 80 years ago (1929–1930) Burr and Burr were the first to discover the importance of linoleic acid (LA) 18:2ω-6 and alpha-linolenic acid (ALA) 18:3ω-3 in restoring the effects caused by the fat-free diet in deprived animals. They coined the term 'essential fatty acids' (EFA). Although healthy skin and successful growth, reproduction, and lactation were obtained in mammals fed with LA as the only source of EFA, ALA was found to permit growth, but was unable to prevent the skin lesions of EFA deficiency, or support reproduction. Omega-6 fatty acids are the predominant polyunsaturated fatty acids (PUFAs) in all diets, especially the US and other Western diets (Table 1). The major omega-6 fatty acid in Western diets is LA, representing approximately 90% of all the PUFA in North American Diets. Today we know that LA and ALA and their long fatty acid derivatives, arachidonic acid (AA), eicosapentaenoic acid (EPA), and docosahexaenoic acid (DHA) respectively are essential for normal growth and development of human beings, and in the prevention and management of chronic diseases.

Over the past five years major advances have taken place in the understanding of mechanisms by which omega-6 and omega-3 fatty acids carry out their metabolic effects at the molecular level. Genetic variants in the LA and ALA metabolic pathways indicate the need to consider their genetic variants in the determination of estimating dietary requirements and in influencing gene expression. Genetic variants most likely account for the conflicting results of epidemiologic studies relative to the effects of omega-3 fatty acids in cardiovascular disease and cancer. Therefore this article focuses on Nutrigenetics – how genetic variation influences dietary response and on Nutrigenomics – how omega-6 and omega-3 fatty acids influence gene expression.

Eicosanoid Metabolism and Biological Effects of Omega-6 and Omega-3 Fatty Acids

The two families of omega-6 (LA) and omega-3 (ALA) fatty acids are physiologically and metabolically distinct, they cannot be synthesized in the human body, and they must be obtained from the diet. Figure 1 shows the metabolism of LA and ALA into

Table 1 *n*-6/*n*-3 ratio in various populations

Population	*n-6/n-3*
Paleolithic	0.79
Greece before 1960	1.00–2.00
Current Japan	4.00
Current India, rural	5–6.1
Current UK and northern Europe	15.00
Current US	16.74
Current India, urban	38–50

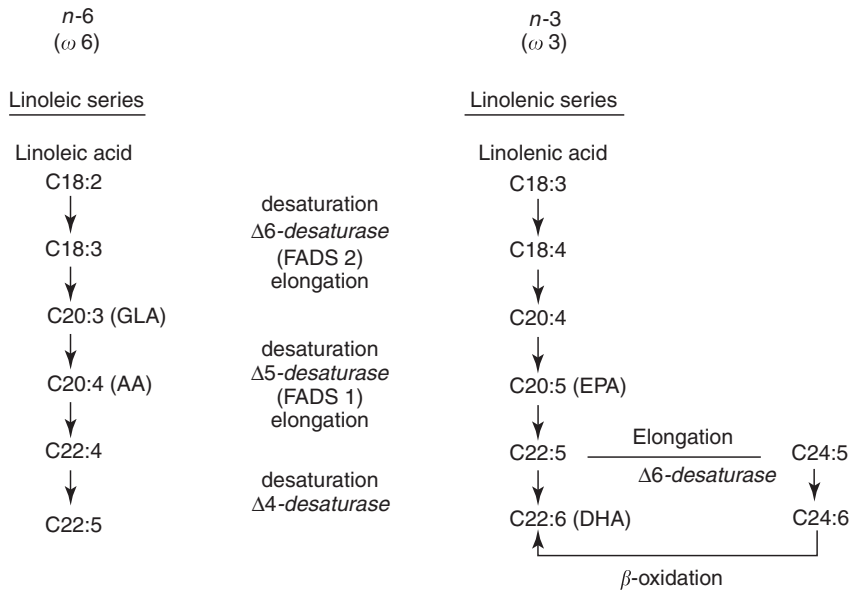


Figure 1 Desaturation and elongation of ω3 and ω6 fatty acids.

very-long-chain polyunsaturated fatty acids (LC-PUFA) through a series of desaturases and elongases. Both LA and ALA use the same enzymes (desaturases and elongases) and compete with each other for enzyme availability. During evolution there was a balance in the intake of LA and ALA with a ratio of ω-6/ω-3=1, whereas today in Western societies the ratio is approximately 16/1 ω-6/ω-3 due to the high intake of vegetable oils-corn oil, sunflower, safflower, soybean, and linseed oil, which are high in LA (Figure 2). LA is found in high amounts in grains with the exception of flaxseed, perilla, rapeseed, and walnuts that are rich in ALA. The green leaves of plants, particularly wild plants are higher in ALA than LA. A low LA/ALA ratio, that is, an increase of ALA leads to higher EPA levels, which competes with AA, and increases the production of anti-inflammatory prostaglandins and leukotrienes.

The PUFA composition of phospholipids has been shown to be associated with normal growth and development, as well as in the outcome of chronic diseases such as coronary heart disease (CHD), hypertension, cancer, diabetes, mental health, neurodegenerative diseases, arthritis, allergies, and other autoimmune diseases, because both omega-6 and omega-3 PUFAs are processed to powerful promoters of eicosanoids such as prostaglandins and leukotrienes, influence gene expression, and telomere length. Figure 1 shows the metabolic pathways of omega-6 and omega-3 fatty acids. The key enzymes in this pathway are the delta-5 and delta-6 desaturases, which are encoded by fatty acid desaturase (FADS) 1 and (FADS) 2, respectively. They are the rate limiting enzymes in the synthesis of the long chain PUFA, AA, EPA, and DHA from their dietary precursors LA and ALA.

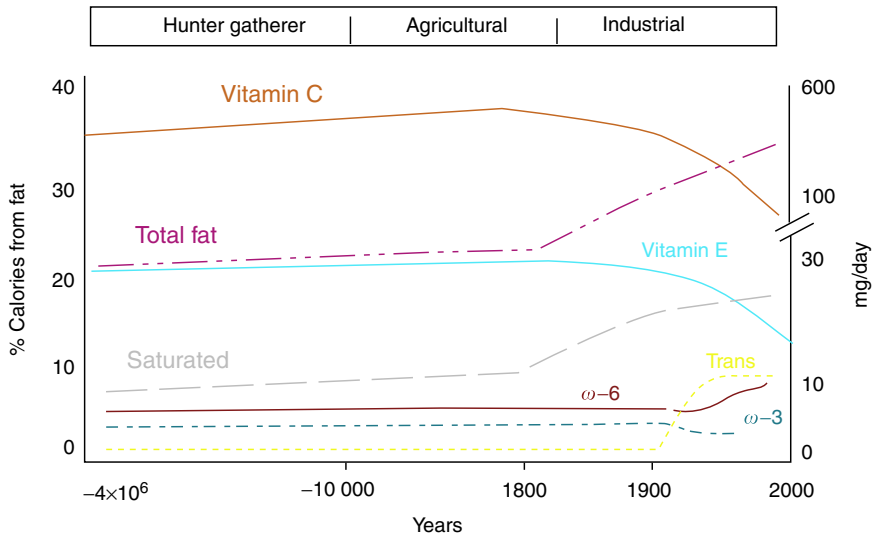


Figure 2 Hypothetical scheme of fat, fatty acid (ω6 and ω3, trans and total) intake (as percent of calories from fat) and intake of vitamins E and C (mg day^{-1}).

Table 2 Effects of ingestion of EPA and DHA from fish or fish oil

Decreased production of prostaglandin E_2 (PGE_2) metabolites
A decrease in thromboxane A_2, a potent platelet aggregator and vasoconstrictor
A decrease in leukotriene B_4 formation, an inducer of inflammation, and a powerful inducer of leukocyte chemotaxis and adherence
An increase in thromboxane A_3, a week platelet aggregator and weak vasoconstrictor
An increase in prostacyclin PGI_3, leading to an overall increase in total prostacyclin by increasing PGI_3 without a decrease in PGI_2, both PGI_2 and PGI_3 are active vasodilators and inhibitors of platelet aggregation
An increase in leukotriene B_5, a weak inducer of inflammation and a weak chemotactic agent

Competition between the omega-6 and omega-3 fatty acids occur in prostaglandin formation that are metabolically and physiologically distinct and have suppressing properties. EPA competes with AA, for prostaglandin and leukotriene synthesis at the cyclooxygenase and lipoxygenase level (**Figure 1**). **Table 2** shows the changes that occur when humans ingest fish or fish oil.

Omega-3 fatty acids modulate prostaglandin metabolism, decrease triglycerides, raise high-density lipoprotein (HDL), and in high doses lower cholesterol, and have antithrombotic and anti-inflammatory properties that account for decreasing the risk for CHD and other chronic conditions, i.e., cognition in the elderly (**Table 3**). Recent studies show that EPA and DHA attenuate the rate of decrease in telomere length, which may account for their beneficial effects on aging and CHD, whereas LA is the most potent nutrient in decreasing telomere length leading to cell apoptosis and aging, and increasing the risk for CHD. In the early phase of inflammation, excessive amounts of interleukins and lipid mediators are released and these play a crucial role. Proinflammatory eicosanoids of AA metabolism are released from membrane phospholipids in the course of inflammatory activation. EPA is released to compete with AA for enzymatic metabolism inducing the production of less inflammatory and chemotactic derivatives (**Table 2**). During the resolution phase of inflammation EPA produces Resolvins E1 and E2 and DHA produces Resolvins D1-D2 and neuroprotectin D1.

Genetic Variation: Nutrigenetics

Genetic Variants, FADS1 and FADS2 in Estimating Nutritional Requirements of Omega-3 and Omega-6 Fatty Acids

The FADS1 and FADS2 gene cluster involved in the metabolic pathway of LA and ALA as well as the enzymes involved in the production of eicosanoids, 5-LO, and cyclooxygenase from the AA and EPA, are polymorphic. Recent studies on their polymorphisms indicate that the minor alleles of the genetic variants in FADS1 and FADS2 are associated with higher LA and lower AA levels in red cell membrane and plasma phospholipids, which may influence the estimation of dietary requirements particularly during pregnancy and lactation as well as the infant's IQ whereas an increase in the activity of the desaturases increases the AA to LA ratio and the risk for CHD. Furthermore genetic variants in the 5-LO and COX-2 genes have been associated with increased risk for CHD and cancer.

The levels of LC-PUFA in plasma serum or red cell membrane phospholipids depend on dietary intake and endogenous metabolism (**Figure 1**). There have been many indications for considerable inter-individual variation in the capacity for endogenous formation of LC-PUFAs. For example, more than 20 years ago there was a rather close correlation of omega-6 and omega-3 fatty acids content in mature milk in human beings even though the main dietary sources were different. Thus it appears that some women have a higher ability to synthesize and secrete milk LC-PUFAs of both the omega-6 and omega-3 series, than others. In addition there was a tracking of plasma LC-PUFA levels in the absence of tracking of dietary intake patterns, suggesting that there is inter-individual variation in the ability to endogenously synthesize LC-PUFAs among children, which persists over time and could most likely be due to genetically determined differences in metabolic turnover. Changes in PUFA conversion have been shown with stable isotope studies.

Genetic Variants of the FADS1 and FADS2 Gene Cluster Influence Omega-6 and Omega-3 Fatty Acids Composition in Both Plasma and Red Cell Membrane Phospholipids During Pregnancy and Lactation

AA, EPA, and DHA play central roles in infant growth, neural development, and immune function. The maternal status of AA, EPA, and DHA during gestation influences maternal to infant transfer via the placenta and breast milk provides fatty acids to infants after birth. FADS1 and FADS2 single nucleotide polymorphisms influence plasma phospholipid and erythrocyte ethanolamine phosphoglyceride (EPG) omega-6 and omega-3 fatty acids during pregnancy and their breast milk during lactation. Minor allele

Table 3 Effects of *n*-3 fatty acids on factors involved in the pathophysiology of atherosclerosis, inflammation, and aging

Factor	*Function*	*Effect of n-3 fatty acid*
Arachidonic acid	Eicosanoid precursor; aggregates platelets; stimulates white blood cells	↓
Thromboxane A_2	Platelet aggregation; vasoconstriction; increase of intracellular Ca^{2+}	↓
Prostacyclin ($PGI_{2/3}$)	Prevent platelet aggregation; vasodilation; increase cAMP	↑
Leukotriene (LTB_4)	Neutrophil chemoattractant; increase of intracellular Ca^{2+}	↓
Fibrinogen	A member of the acute phase response; and a blood clotting factor	↓
Tissue plasminogen activator	Increase endogenous fibrinolysis	↑
Platelet activating factor (PAF)	Activates platelets and white blood cells	↓
Platelet-derived growth factor (PDGF)	Chemoattractant and mitogen for smooth muscles and macrophages	↓
Oxygen free radicals	Cellular damage; enhance LDL uptake via scavenger pathway; stimulate arachidonic acid metabolism	↓
Lipid hydroperoxides	Stimulate eicosanoid formation	↓
Interleukin 1 and tumor necrosis factor	Stimulate neutrophil O_2 free radical formation; stimulate lymphocyte proliferation; stimulate PAF; express intercellular adhesion molecule-1 on endothelial cells; inhibit plasminogen activator, thus, procoagulants	↓
Interleukin-6	Stimulates the synthesis of all acute phase proteins involved in the inflammatory response: C-reactive protein (CRP); serum amyloid A; fibrinogen; α_1-chymotrypsin; and haptoglobin	↓
CRP	An acute phase reactant and an independent risk factor for cardiovascular disease	↓
Endothelial-derived relaxation factor	Reduces arterial vasoconstrictor response	↑
Insulin sensitivity		↑
VLDL		↓
HDL	Decreases the risk for CHD	↑
Lp(a)	Lipoprotein(a) is a genetically determined protein that has atherogenic and thrombogenic properties	↓
Triglycerides and chylomicrons	Contribute to postprandial lipemia	↓
Telomeres	Have anti-aging effects whereas LA promotes shortening of telomeres and aging	↑
Resolvin E1–E2 (EPA)	Anti-inflammatory important in the resolution of inflammation	↑
Resolvin D1–D2 (DHA)	Anti-inflammatory important in the resolution of inflammation	↑
Neuroprotectin (DHA)	Protects brain; important in patients with strokes or trauma	↑
PPAR	Upregulates the expression of genes involved in lipid metabolism and downregulates the expression of genes involved in inflammation and suppresses NFκB	↑

VLDL: very-low-density lipoprotein.

homozygotes of rs 174553 (GG), rs 99780 (TT), and rs 174583 (TT) had lower AA but higher LA in plasma phospholipids and erythrocyte EPG and decreased omega-6 and omega-3 fatty acids product to precursor ratio at 16 and 36 weeks gestation $P<0.001$. Breast milk fatty acids were influenced by genotype, with significantly lower 14.0, AA and EPA, but higher 20:2ω-6 in the minor allele homozygotes of rs 174553 (GG), rs 99780 (TT), and rs 174583 (TT) and lower AA, EPA, DPA 22:5ω-3, and DHA in the minor allele homozygotes GG of rs 174575. The results indicate a robust association between minor alleles of the 4 SNPs and lower AA and other omega-6 fatty acids relative to precursor LA. Similar results were found for the omega-3 series. Genetic variation in the FADS1 and FADS2 gene cluster is important for the composition of fatty acids provided to breastfed infants in mother's milk.

Genetic Variants in Omega-6 and Omega-3 Fatty Acids Metabolism and IQ

Children's intellectual development is influenced by both genetic and environmental experiences. Breastfed children attain higher IQ scores than children not fed breast milk. Breastfeeding is thought to influence brain development through nutritional processes involving fatty acids. The predominant LC-PUFA present in human milk but not in cow's milk, are DHA and AA. Substantial

amounts of DHA and AA accumulate in the human brain during the first postnatal months and infants who are breastfed have higher concentrations of DHA and AA than infants fed unsupplemented formulas. Randomized controlled clinical trials comparing the neurodevelopment of infants fed DHA supplemented versus unsupplemented formula have produced inconsistent results. A search in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database for genes involved in LC-PUFA metabolism that might moderate the effect of breastfeeding on IQ led to FADS2. FADS2 gene expression is regulated through end product inhibition and dietary LC-PUFA such as those available in breast milk. Two SNPs (rs 174575 and rs 1535) were selected as candidate biomarkers, to test the hypothesis that the cognitive advantage associated with breastfeeding in human beings is related to genetic differences in LC-PUFA metabolism, which was replicated in two birth cohorts. It was found that the difference in IQ test scores between breastfed and not breastfed to be 5.6 IQ points in one birth cohort and 6.3 IQ points in the other cohort. Analysis revealed that the rs 174575 interacted with breastfeeding to influence IQ in both cohorts. There was a dominant effect of the C allele in response to breastfeeding, with those carrying the C allele having a statistically significant 6.4-IQ-point advantage relative to children not breastfed. In contrast, GG homozygous neither gained an advantage from breastfeeding nor suffered a disadvantage from not being breastfed. The interaction between children's 174575 genotype and breastfeeding suggests that C carrying children benefit from breast milk more than GG homozygotes. There were no significant IQ differences among children fed breast milk as a function of maternal genotypes. Therefore these results suggest that the rs 174575 influence of breastfeeding effects on IQ involves genetic differences in children's LC-PUFA metabolism rather than rs 174575 differences among lactating women in their milk composition. Among GG homozygote children the IQ advantage associated with breastfeeding was nil. But children who were C-carriers the difference in IQ was 6.8 IQ points with C-carriers having the advantage. This advantage corresponds to a moderate effect size that is associated with many important life outcomes. This very important finding needs to be replicated in much larger cohorts and populations. The molecular mechanism by which rs 174575 may influence cognitive development is not known **(Table 4)**. The rs 174575 C allele is linked with the major alleles of FADS1 and FADS2 SNPs, which are associated with more efficient fatty acid processing, possibly due to increased transcriptomal activity or a more active protein **(Table 5)**.

Genetic Variants in FADS1 and FADS2 and CHD Risk

Desaturase activity is assayed in vitro or in animals by measurement of the rate of conversion of radiolabeled precursor fatty acids to their respective products, but ethical and practical reasons prevent this possibility in humans. Instead a product to precursor ratio (e.g., AA/LA or EPA/ALA) as a surrogate measure to estimate desaturase activity is well established. In an ongoing case–control study with or without angiographic evidence of coronary artery disease (CAD), both AA/LA and the ratio of EPA to ALA were higher in

Table 4 Adequate intake (AI) for adults

Fatty acid	*Grams per day (2000 kcal diet)*	*% Energy*
LA	4.44	2.0
(upper limit)[a]	6.67	3.0
ALA	2.22	1.0
DHA+EPA	0.65	0.3
DHA to be at least[b]	0.22	0.1
EPA to be at least	0.22	0.1
TRANS-FA		
(upper limit)[c]	2.00	1.0
SAT		
(upper limit)[d]	–	<8.0
MONOs[e]	–	–

ALA, α-linolenic acid; DHA, docosahexaeonic acid; EPA, eicosapentaenoic acid; LA, linoleic acid; MONOs, monounsaturated fatty acids; SAT, saturated fatty acids; *TRANS*-FA, *trans*-fatty acids.

Note: If sufficient scientific evidence is not available to calculate an estimated average requirement, a reference intake called an adequate intake is used instead of a recommended dietary allowance. The AI is a value based on experimentally derived intake levels or approximations of observed mean nutrient intakes by a group (or groups) of healthy people. The AI for children and adults is expected to meet or exceed the amount needed to maintain a defined nutritional state or criterion of adequacy in essentially all members of a specific healthy population.

[a]Although the recommendation is for AI, the Working Group felt that there is enough scientific evidence to also state an upper limit (UL) for LA of 6.67 g day^{-1} based on a 2000 kcal diet or of 3.0% of energy.

[b]For pregnant and lactating women, ensure 300 mg day^{-1} of DHA.

[c]Except for dairy products, other foods under natural conditions do not contain *trans*-FA. Therefore, the Working Group does not recommend *trans*-FA to be in the food supply as a result of hydrogenation of unsaturated fatty acids or high-temperature cooking (reused frying oils).

[d]Saturated fats should not comprise more than 8% of energy.

[e]The Working Group recommended that the majority of fatty acids are obtained from monounsaturates. The total amount of fat in the diet is determined by the culture and dietary habits of people around the world (total fat ranges from 15% to 40% of energy) but with special attention to the importance of weight control and reduction of obesity.

Table 5 Adequate intake (AI) for infant formula/diet

Fatty acid	*Percent of fatty acids*
LA[a]	10.00
ALA	1.50
AA[b]	0.50
DHA	0.35
EPA[c] (upper limit)	<0.10

AA, arachidonic acid; ALA, α-linolenic acid; DHA, docosahexaenoic acid; EPA, eicosapentaenoic acid; LA, linoleic acid; MONOs, monounsaturated fatty acids; SAT, saturated fatty acids; *TRANS-FA*, *trans*-fatty acids.

Note: If sufficient scientific evidence is not available to calculate an estimated average requirement, a reference intake called an adequate intake is used instead of a recommended dietary allowance. The AI is a value based on experimentally derived intake levels or approximations of observed mean nutrient intakes by a group (or groups) of healthy people. The AI for children and adults is expected to meet or exceed the amount needed to maintain a defined nutritional state or criterion of adequacy in essentially all members of a specific healthy population.

[a]The Working Group recognizes that in countries like Japan the breast milk content of LA is 6–10% of fatty acids and the DHA is higher, approximately 0.6%. The formula/diet composition described here is patterned on infant formula studies in Western countries.

[b]The Working Group endorsed the addition of the principal long chain polyunsaturates, AA and DHA, to all infant formulas.

[c]EPA is a natural constituent of breast milk, but in amounts more than 0.1% in infant formula may antagonize AA and interfere with infant growth.

participants with CAD than in those without CAD, but in a multiple logistic regression model only a higher AA/LA resulted as an independent risk factor for CAD. Concentrations of high sensitivity C-reactive protein (hs-CRP) increased progressively across tertiles of AA/LA. Graded increases in hs CRC concentrations and CAD risk were related to the carriership of FADS haplotypes, including the alleles associated with a higher ratio of AA/LA.

A number of studies have shown that higher amounts of AA in adipose tissue are associated with higher risk of acute myocardial infarction. In populations eating a Western diet rich in omega-6 PUFA a high desaturase activity due to FADS1 and FADS2 polymorphism may promote an increased bioavailability of AA with prevailing synthesis of AA-derived proinflammatory eicosanoids leading to atherosclerosis and vascular damage. On the other hand, high desaturase activity in subjects on a diet rich in omega-3 fatty acids or receiving EPA and DHA supplementation could result in an opposite situation with a preferential synthesis of anti-inflammatory eicosanoids.

Genetic Variants in the 5-LO, the Role of Omega-6 and Omega-3 Fatty Acids in CHD

Because atherosclerosis involves arterial inflammation, investigators determined 5-lipoxygenase (5-LO) genotypes, carotid artery intima–media thickness, markers of inflammation, CRP, interleukin-6 (IL-6), dietary AA, EPA, DHA, LA, and ALA with the use of six 24-hour recalls of food intake. The results showed that 5-LO variant genotypes were found in 6.0% of the cohort. Mean intima–media thickness adjusted for age, sex, height, and racial or ethnic group was increased among the carriers of two variant alleles as compared with the carrier of the common (wild-type) allele. In multivariate analysis, the increase in intima–media thickness among carriers of two variant alleles was similar in this cohort to that associated with diabetes, the strongest common cardiovascular risk factor. Increased dietary AA significantly enhanced the apparent atherogenic effect of genotype, whereas increased dietary intake of omega-3 fatty acids EPA and DHA blunted this effect. Furthermore, the plasma level of CRP of two variant alleles was increased by a factor of 2, as compared with that among carriers of the common allele. Thus, genetic variation of 5-LO identifies a subpopulation with increased risk for atherosclerosis. The diet–gene interaction further suggests that dietary omega-6 fatty acids LA and AA promote, whereas marine omega-3 fatty acids EPA and DHA inhibit leukotriene-mediated inflammation that leads to atherosclerosis in this subpopulation (**Figure 1**). The study constitutes evidence that genetic variation in an inflammatory pathway – in this case the leukotriene pathway – can trigger atherogenesis in humans. These findings could lead to new dietary and targeted molecular approaches for the prevention and treatment of cardiovascular disease according to genotype.

Genetic Variants in the 5-LO, Omega-6 Fatty Acids, and Breast Cancer

A number of epidemiologic studies and animal experiments suggest that omega-6 fatty acids increase the risk of cancer and omega-3s decrease. However not all studies have produced consistent results. The 5-LO pathway has been implicated in carcinogenesis and tumor progression in many types of cancer; lung, colon, prostate, kidney, bladder. Earlier epidemiologic studies on dietary fat intake

and breast cancer did not find positive association between omega-6 and breast cancer risk. Those studies however did not take into account genetic predisposition related to omega-6 fatty acid metabolism. A recent study on genetic variants in the 5-LO gene (ALOX5) and 5-lipogenase-activating protein gene (ALOX5AP) in combination with dietary LA intake in a population-based multiethnic case–control study on breast cancer in Latin, African-American and White women in the San Francisco area, did not find significant main effects of ALOX5 and ALOX5AP genotypes on breast cancer risk that were consistent across race or ethnicity. There was a significant interaction between the ALOX5AP −4900 A> G polymorphisms and dietary LA intake among women consuming a diet high in LA (top quartile of intake >17.4 g day^{-1}), carrying the AA genotype. The AA genotype was associated with higher breast cancer risk, compared to genotype AG or GG. Among women consuming ≤17.4 g day^{-1} of LA ALOX5AP −4900 genotype was not associated with breast cancer risk. These findings indicate that studies on dietary fat intake and cancer should take into consideration both type of fat and genetic variants. Furthermore. in the U.S. 17.4 g day^{-1} is the intake that a significant portion of the population ingests.

Genetic Variants of Cyclooxygenase-2 (COX-2) and the Protective Effect of Long Chain Omega-3 Fatty Acids in Cancer of the Prostate

Prostate cancer is one of the most common types of cancer in men. Increasing evidence points to chronic inflammation as one of the factors leading to cancer. Inflammation may result from bacterial or viral infections, intra-prostatic urine reflux, or diet. Dietary components that are potent anti-inflammatory agents are the omega-3 PUFAs. Studies have shown that genetic variants at the COX-2 gene modify prostate inflammation through the COX-2 enzymatic pathway. COX-2 is a key enzyme in fatty acid metabolism and inflammation. In a case–control study of 466 men diagnosed with aggressive prostate cancer and 478 age- and ethnicity-matched controls nine COX-2 tag SNPs were genotyped. Dietary history was assessed with a semiquantitatively food frequency questionnaire. Increasing omega-3 intake was associated with a decreased risk of aggressive prostate cancer, and this inverse association was even stronger among men with genetic variants rs 4648310 (+8897 A/G) flanking the 3 region of COX-2. The patient with the lowest intake of omega-3s and the genetic variant had the most aggressive tumor whereas the omega-3 PUFAs were protective and this effect was modified by the genetic variant. This gene by diet (omega-3s) interaction clearly shows that the main dietary effect was modified by the genetic variant whereas men with the variant genotype AG or GG and low intake of omega-3s had much higher risk than men with the variant genotype and high intake of omega-3s. A study in Swedish men found similar inverse association between consumption of fatty fish and prostate cancer risk, an effect also modified by rs 5275 (+6364 A>G) SNP in COX-2.

Nutrigenomics: The Role of Omega-6 and Omega-3 Fatty Acids in Gene Expression

Anti-Inflammatory Aspects: Omega-3 Fatty Acids Downregulate the Expression of Genes Involved in Inflammation and Obesity

Animal experiments and human studies have shown that EPA and DHA have the ability to upregulate and downregulate genes in various tissues including adipose tissue and peripheral blood mononuclear cells (PBMCs) in humans. Clinical studies indicate that inflammation is at the base of many diseases including cardiovascular disease, aging, mental health, obesity, diabetes, and even cancer. EPA and DHA have been shown to have beneficial effects in these conditions but the exact mechanisms by which EPA and DHA suppress inflammation are still under investigation. Previous studies have focused on the ability of EPA and DHA to suppress interleukin-1beta (IL1b) and IL-6 cytokines and to play an important role in the resolution of inflammation as well as through the production of resolvins E1 and E2 from EPA, and D1 and D2 from DHA, and neuroprotectins D1 from DHA. EPA and DHA activate peroxisome proliferator-activated receptor alpha (PPARα), which upregulates the expression of genes involved in lipid metabolism and downregulates the expression of genes involved in inflammation by suppressing Nuclear factor-Kappa-B (NF-kB).

In studying the effect of fish oil on adipose tissue it was noted that 1.8 g of EPA and DHA decreased the total omega-6/omega-3 ratio in the plasma phospholipids from 12.9±1.1 to 5.6±0.7. Total fat mass and subcutaneous adipocyte diameter were lower in the group receiving 1.8 g of EPA and DHA than in the placebo group. In addition, with EPA and DHA supplementation, significant correlations were found between the adipocyte markers (adipocyte diameter and whole fat mass) and the main adipokines-plasma leptin and adiponectin as well as plasma atherogenic factors (plasminogen activator inhibitor factor-1 (PAI-1), insulin, and triacylglycerol). There was no correlation between adipocyte diameter and plasma tumor necrosis factor-alpha (TNF-α) or plasma IL-6. As expected adipocyte diameter and fat mass percentage were correlated with atherogenic (cysteine protease cathepsin S or CTSS) and inflammation related genes (the chemoattractant gene plasminogen activator urokinase receptor, or PLAVR), the macrophage surface marker CD11b, and the macrophage phagocytic activity marker CD68. There was no change in weight but there was significant loss in body fat mostly in the trunk and subcutaneous tissue, but not in visceral tissue. After fish oil treatment, PAI-1 was lower whereas leptin, IL-6, TNF-α, and serum amyloid A did not change significantly after 2 months with 1.8 g EPA and DHA supplementation. Epidemiologic studies have shown a correlation between adipocyte size and the omega-6 and omega-3 fatty acids content in subcutaneous abdominal adipose tissue, in a group of overweight patients who had undergone abdominal surgery, and metabolic studies indicate a beneficial role of EPA and DHA in lowering adiposity in humans.

Studies on the relationship between plasma omega-3 PUFA composition and weight status found that higher omega-3 PUFA intake was associated with a healthier BMI, waist circumference, and hip circumference. These findings suggest that omega-3 PUFA may play a role in weight status and abdominal adiposity in human beings. In earlier studies EPA and DHA supplementation

reduced body fat mass and stimulated lipid oxidation in healthy adults. Others subsequently concluded that omega-3 fatty acid intake by itself or along with exercise increases weight loss. Caloric restriction is recommended for weight loss. Of interest is the fact that caloric restriction affects gene expression in a manner similar to EPA and DHA supplementation.

Summary and Conclusions

Fatty acid composition in red cell membranes and serum phospholipids plays an important role in cellular processes, and has been shown to be associated with the etiology of several complex diseases in humans. The metabolism of EFAs, LA, and ALA and their metabolic derivatives are controlled by enzymes encoded by polymorphic genes. Therefore the availability of PUFAs to various tissues is of major importance to health and depends on both dietary intake and endogenous production or metabolic turnover.

The genetic variants FADS1 and FADS 2 lead to differences in the conversion of omega-6 and omega-3 fatty acids catalyzed by the delta-5 and delta-6 desaturases, which suggests that individuals may require different amounts of dietary PUFAs or LC-PUFAs to achieve comparable biological effects. Furthermore further studies addressing the biological effects of PUFAs and LC-PUFAs should include genotyping for FADS1 and FADS2 polymorphisms.

Variants in the human genes of delta-5 and delta-6 desaturase FADS1 and FADS2 that influence both serum and red cell membrane phospholipid levels of PUFA, have a frequency of 26% in the population. The minor alleles are associated with lower AA and higher LA and account for 28% of the variation in serum phospholipid AA and up to 12% of its precursor fatty acids. Smaller percentage values were found for omega-3 fatty acids. These findings suggest that individuals may require different amounts of dietary LA, ALA, or AA, EPA, and DHA, for both normal development and in the prevention and management of chronic diseases.

The interaction between dietary factors and genetic variants could explain the differences noted in association studies. Considering that a low omega-3 intake in the presence of certain genetic variants leads to a more aggressive disease, an increase in omega-3 intake and a decrease in omega-6 leading to a balanced omega-6/omega-3 ratio may be a sensible recommendation to reduce disease risk in the general population. Nutrigenetics/nutrigenomics will continue to provide data on mechanisms of nutrients and gene interactions in both health and disease.

Further Reading

Caspi, A, Williams, B, Kim-Cohen, J, et al., 2007. Moderation of breastfeeding effects on the IQ by genetic variation in fatty acid metabolism. Proceedings of the National Academy of Sciences 104 (47), 18860–18865.

Dwyer, JH, Allayee, H, Dwyer, KM, et al., 2004. Arachidonate 5-lipoxygenase promoter genotype, dietary arachidonic acid, and atherosclerosis. New England Journal of Medicine 350, 29–37.

Farzaneh-Far, R, Lin, J, Epel, ES, Harris, WS, Blackburn, EH, Whooley, MA, 2010. Association of marine Omega-3 fatty acid levels with telomeric aging in patients with coronary heart disease. Journal of the American Medical Association 303 (3), 250–257.

Fontani, G, Corradeschi, F, Felici, A, Alfatti, F, Migliorini, S, Lodi, L, 2005. Cognitive and physiological effects of Omega-3 polyunsaturated fatty acid supplementation in healthy subjects. European Journal of Clinical Nutrition 35 (11), 691–699.

Serhan, CN, Chiang, N, Van Dyke, TE, 2008. Resolving inflammation: Dual anti-inflammatory and pro-resolution lipid mediators. Nature Reviews Immunology 8 (5), 349–361.

Simopoulos AP (2005) Fatty acids: Omega-3 polyunsaturated. *Encyclopedia of Human Nutrition*, 2nd edn., pp. 205–219. Elsevier: Oxford.

Simopoulos, AP, 2008. The importance of the omega-6/omega-3 fatty acid ratio in cardiovascular disease and other chronic diseases. Federation of American Societies for Experimental Biology 233 (6), 674–688.

Simopoulos, AP, 2010. Nutrigenetics/Nutrigenomics. Annual Reviews of Public Health 31, 53–68. Annual Reviews.

Simopoulos, AP, Bazan, NG (Eds.), 2009. World Review of Nutrition and Dietetics, vol. 99. Karger, Basel.

Simopoulos, AP, Leaf, A, Salem Jr, N, 1999. Essentiality of and recommended dietary intakes for omega-6 and omega-3 fatty acids. Annals of Nutrition and Metabolism 43, 127–130.

Simopoulos, AP, Milner, JA (Eds.), 2010. World Review of Nutrition and Dietetics, vol. 101. Karger, Basel.

Simopoulos, AP, Robinson, J, 1999. In the Omega Diet. The Lifesaving Nutritional Program Based on the Diet of the Island of Crete. Harper Collins, New York.

Wang, J, John, EM, Ingles, SA, 2008. 5-lipoxygenase and 5-lipoxygenase-activating protein gene polymorphisms, dietary linoleic acid, and risk for breast cancer. Cancer Epidemiology, Biomarkers & Prevention 17 (10), 2748–2754.

Xie, L, Innis, SM, 2008. Genetic variants of the FADS1 FADS2 gene cluster are associated with altered (n-6) and (n-3) essential fatty acids in plasma and erythrocyte phospholipids in women during pregnancy and in breast milk during lactation. Journal of Nutrition 138 (11), 2222–2228.

Yurko-Mauro, K, McCarthy, D, Rom, D, et al., 2010. Beneficial effects of docosahexaenoic acid on cognition in age-related cognitive decline. Alzheimer's & Dementia 1–9.

Fatty acids: Omega-6 polyunsaturated

T.A. Mori and J.M. Hodgson, School of Medicine and Pharmacology, University of Western Australia, Perth, WA, Australia

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of T.A. Mori, J.M. Hodgson, Fatty acids: Health Effects of Omega-6 Polyunsaturated Fatty Acids, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 209–214, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00100-8.

Introduction 233
Structure, Function, and Nutritional Requirements 233
Endothelial Function, Atherosclerosis, and Cardiovascular Disease 234
Eicosanoids: Relevance to Endothelial Function, Thrombosis, Inflammation, and Atherosclerosis 234
Prostaglandins and Leukotrienes 235
Fatty Acid Intake and Eicosanoids 235
***n*-6 Fatty Acids and Risk of Cardiovascular Disease** 235
Cardiovascular Disease: Population Studies 235
Thrombosis 236
Inflammation 236
Cholesterol and Lipoproteins 236
Oxidative Stress 237
Blood Pressure 237
Conclusions 237
Further Reading 237

Glossary

Arachidoinic acid (20:4*n*-6) A 20 carbon omega-6 fatty acid which is a precursor of eicosanoids.

Eicosanoids A family of signaling molecules made by oxidation of twenty-carbon essential fatty acids; they include the prostaglandins, thromboxanes, leukotrienes and epoxyeicosatrienoic acids.

Essential fatty acids Fatty acids that the body cannot manufacture and that may cause nutritional defiency if not supplied through diet. The two essential fatty acids for humans are α-linolenic acid and linoleic acid.

Linoleic acid (18:2*n*-6) The precursor fatty acid of the *n*-6 fatty acids, and the primary dietary *n*-6 fatty acid comprising greater than 97% of total *n*-6 fatty acid intake.

Omega-3 (*n*-3) polyunsaturated fatty acid A class of polyunsaturated fatty acids characterized by the presence of two or more *cis*-double bonds, with the position of the first double bond three carbon atoms from the methyl end of the molecule.

Omega-6 (*n*-6) polyunsaturated fatty acid A class of polyunsaturated fatty acids characterized by the presence of two or more *cis*-double bonds, with the position of the first double bond six carbon atoms from the methyl end of the molecule.

Polyunsaturated fatty acids Fatty acids in which more than one double bond exists within the molecule.

α-Linolenic acid (18:3*n*-3) The precursor fatty acid of the *n*-3 fatty acids.

Abbreviations

AA	Arachidonic acid – 20:4*n*-6
ALA	α-Linolenic acid – 18:3*n*-3
COX-2	Cyclooxygenase-2
DGLA	Dihomo-γ-linolenic acid – 20:3*n*-6
DHA	Docosahexaenoic acid – 22:6*n*-3
DiHETE	Dihydroxyeicosatetraenoic acid
DPA	Docosapentaenoic acid – 22:5*n*-3
EET	Epoxyeicosatrienoic acids
EFA	Essential fatty acid
EPA	Eicosapentaenoic acid – 20:5*n*-3

GLA	γ-Linolenic acid – 18:3*n*-6
HDL	High-density lipoprotein
HETE	Hydroxyeicosatetraenoic acid
LA	Linoleic acid – 18:2*n*-6
LDL	Low-density lipoprotein
LT	Leukotriene
LXA_4	Lipoxin A_4
LXB_4	Lipoxin B_4
***n*-6**	Omega-6
***n*-3**	Omega-3
PG	Prostaglandin
PUFA	Polyunsaturated fatty acid
VLDL	Very low density lipoprotein

Introduction

The omega-6 (*n*-6) fatty acids have the potential to influence a number of chronic diseases and disorders. This article focuses on the effects of *n*-6 fatty acids in relation to cardiovascular disease and atherosclerosis.

Structure, Function, and Nutritional Requirements

The *n*-6 fatty acids are a class of polyunsaturated fatty acids (PUFAs) characterized by the presence of two or more *cis*-double bonds, with the position of the first double bond six carbon atoms from the methyl end of the molecule. The general formula of *n*-6 fatty acids is $CH_3(CH_2)_4(CH{=}CHCH_2)_x(CH_2)_yCOOH$. Linoleic acid [*cis*-9, *cis*-12-octadecadienoic acid, 18:2*n*-6, LA] and α-linolenic acid [*cis*-9, *cis*-12, *cis*-15-octadecatrienoic acid, 18:3*n*-3, ALA] are the precursor fatty acids of the *n*-6 and omega-3 (*n*-3) fatty acids, respectively.

LA is the primary dietary *n*-6 fatty acid comprising greater than 97% of total *n*-6 fatty acid intake. The average daily intake of LA for adults older than 19 years is approximately 11 g or 4.4% of energy in Australia, and 14.8 g and 6.7% of energy in the United States. In contrast, the average intake of ALA is approximately 1.1 g day^{-1} or 0.6% of energy. LA and ALA cannot be made by mammals and are therefore termed essential fatty acids (EFAs). In addition, mammals are unable to interconvert LA and ALA, or any of the *n*-6 and *n*-3 fatty acids, because mammalian tissues do not contain the necessary desaturase enzyme. Plant tissues and plant oils are rich sources of LA. ALA is also present in plant sources such as green vegetables, flaxseed, canola, and some nuts. Once consumed in the diet, LA can be converted via chain elongation and desaturation to γ-linolenic acid (GLA, 18:3*n*-6), dihomo-γ-linolenic acid (DGLA, 20:3*n*-6), and arachidonic acid (AA, 20:4*n*-6) (Figure 1). The same enzymes involved in elongation and desaturation of the *n*-6 fatty acids are common to the omega-3 (*n*-3) series of fatty acids (Figure 1). Thus, ALA can be converted to eicosapentaenoic acid (EPA, 20:5*n*-3) and docosahexaenoic acid (DHA, 22:6*n*-3). In humans, however, the conversion is relatively inefficient and gender dependent. The major source of dietary EPA and DHA is fish, seafood, and marine oils, although small quantities of *n*-3 fatty acids, particularly of docosapentaenoic acid (DPA, 22:5*n*-3), are derived from meat and poultry.

The *n*-6 and *n*-3 fatty acids are metabolically and functionally distinct and often have important opposing physiological functions. Indeed, the balance of EFA is important for good health and normal development. Historically, human beings evolved on a diet in which the ratio of *n*-6 to *n*-3 fatty acids was approximately 1:1. In contrast, Western diets have a ratio of approximately 15:1. Evidence for this change in diet through history comes from studies on the evolutionary aspects of diet, modern-day hunter–gatherers, and traditional diets. Modern agriculture has led to a substantial increase in *n*-6 fatty acids at the expense of *n*-3 fatty acids, which has resulted in excessive consumption of *n*-6 fatty acids by humans.

The *n*-6 EFAs have two main functions. First, they act as structural components of membranes forming the basis of the phospholipid component of the lipid bilayer of plasma membranes in every cell in the body, thus providing a membrane impermeable to most water-soluble molecules. The length and degree of saturation of the fatty acids determine how the phospholipid molecules pack together and consequently affect membrane fluidity, signal transduction, and the expression of cellular receptors. The second role of *n*-6 fatty acids is to act as precursors to the eicosanoids (Figure 1). The eicosanoids are a family of 'hormone-like' compounds including prostaglandins (PGs), leukotrienes (LTs), and hydroxy- (HETEs), dihydroxy- (DiHETEs), and epoxy- (EETs) fatty acids. Eicosanoids, however, are distinct from most hormones in that they act locally, near their sites of synthesis, and they are catabolized extremely rapidly. Thus, they are considered to be locally acting hormones. The eicosanoids modulate renal and pulmonary functions, vascular tone, and inflammatory responses. The enzymes involved in AA metabolism include the cyclooxygenases and lipoxygenases, which yield the two-series PGs and four-series LTs, respectively. Lipoxygenases also utilize AA for the formation of the hydroxyeicosatetraenoic acids (HETEs). A third pathway for the utilization of AA involves the cytochrome P-450 enzymes found in the liver, kidney, lung, intestines, heart, small blood vessels, and white blood cells. AA metabolized via cytochrome P-450 yields epoxyeicosatrienoic acids (EETs), dihydroxyeicosatetraenoic acids (DiHETEs), as well as HETEs. The cytochrome P-450 metabolites

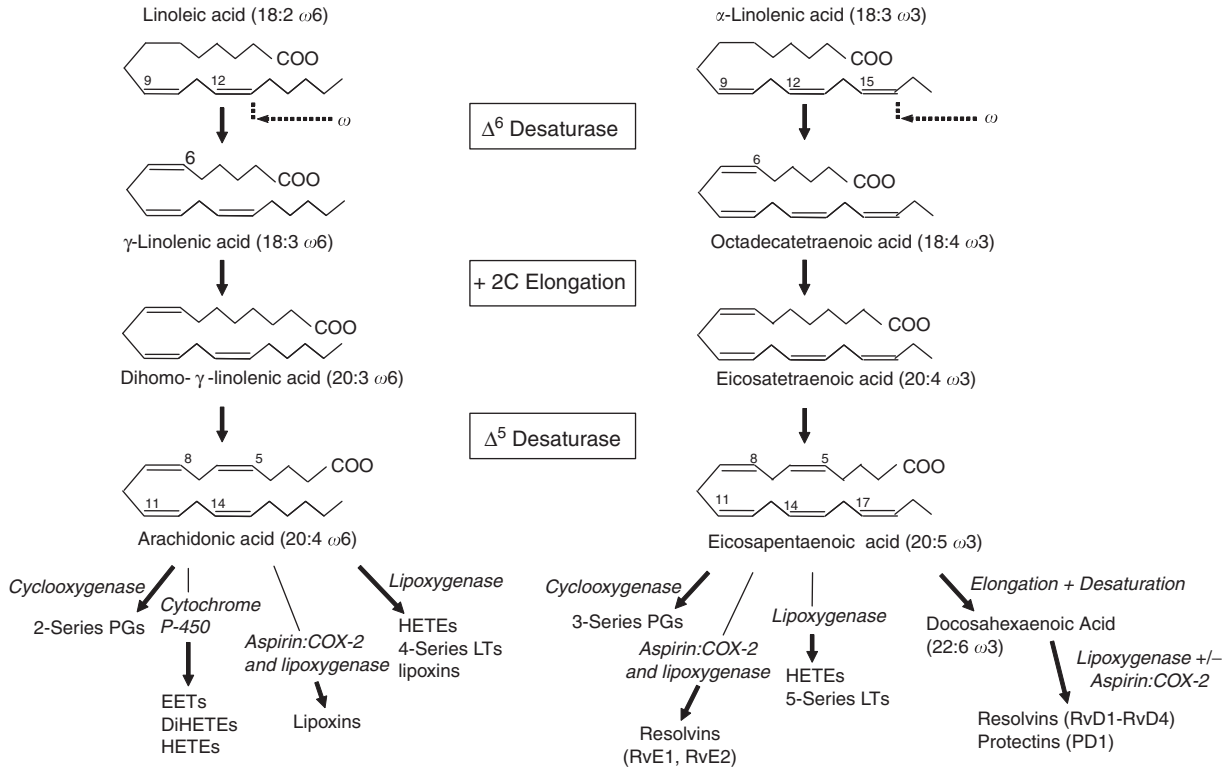


Figure 1 Essential fatty acid metabolism.

play an important role as paracrine factors and second messengers in the regulation of pulmonary, cardiac, renal, and vascular function and modulate inflammatory and growth responses. Finally, AA can be utilized by a pathway comprising the interaction of aspirin with cyclooxygenase-2 (COX-2) and lipoxygenase to generate a family of compounds called the lipoxins. COX-2 is an inducible form of the cyclooxygenase enzyme activated during inflammation. The lipoxins (LXA_4 and LXB_4) are potent anti-inflammatory mediators. They can also be derived from the action of lipoxygenase in the absence of aspirin–COX-2 (Figure 1).

Endothelial Function, Atherosclerosis, and Cardiovascular Disease

The vascular endothelium is the most important organ controlling vascular function and consists of a single layer of epithelial cells lining blood vessels. Its primary function is to regulate vascular tone, but it plays a critical role in modulating coagulation and fibrinolysis, inflammation, smooth muscle cell proliferation, and macrophage function. Many of these functions are regulated through the release of various mediators including eicosanoids. There is multiple and close interaction of the endothelial cells with circulating cells, smooth muscle cells, and macrophages. There is also evidence that endothelial dysfunction precedes clinically apparent atherosclerosis.

Atherosclerosis is an inflammatory disease involving multiple cellular and molecular responses that lead to an alteration in vascular function and structure, and the development and progression of cardiovascular disease. Atherosclerosis is characterized by degenerative changes, deposition of cholesterol, proliferation of smooth muscle cells, involvement of a range of circulating proinflammatory cells, and fibrosis. Resulting atheromatous plaques cause narrowing of arteries and increase the likelihood of thrombosis and occlusion. When this occurs in the coronary arteries, the outcome is myocardial infarction with possible death.

Eicosanoids: Relevance to Endothelial Function, Thrombosis, Inflammation, and Atherosclerosis

In general, the eicosanoids derived from AA have potent prothrombotic and proinflammatory activity. In contrast, the eicosanoids derived from EPA have reduced biological activity and are less prothrombotic and proinflammatory. Eicosanoid production is generally tightly controlled through homeostatic mechanisms. However, eicosanoid production can be significantly altered in situations in which endothelial dysfunction, atherosclerosis, and plaque rupture, or various thrombotic or inflammatory conditions are present.

Prostaglandins and Leukotrienes

Prostaglandins have a central role in the regulation of platelet aggregation and vascular tone. In this regard, two of the major prostaglandins derived from AA are thromboxane A_2, produced in platelets, and prostacyclin I_2, produced in endothelial cells. Thromboxane A_2 promotes platelet aggregation and blood vessel constriction, whereas prostacyclin I_2 has opposite effects. An increase in availability of EPA can decrease platelet thromboxane A_2 and increase thromboxane A_3, the latter having considerably less physiological activity. EPA supplementation also stimulates formation of prostacyclin I_3, whereas prostacyclin I_2 is unaffected. Prostacyclin I_3 and prostacyclin I_2 are equipotent in their biological activity. The net result following intake of *n*-3 fatty acids is a shift in the thromboxane/prostacyclin balance toward a reduced prothrombotic state.

Leukotriene B_4 is a potent inflammatory mediator produced by neutrophils from 20:4*n*-6 at the site of injury. Leukotriene B_4 is also a powerful chemotactic factor responsible for attracting neutrophils to the site of injury. Leukotriene B_5, which is produced from EPA, has significantly lower biological activity. Therefore, an increased availability of EPA has the potential to reduce inflammation.

Fatty Acid Intake and Eicosanoids

The concentration of the eicosanoid precursor fatty acids both circulating and in tissues depends on dietary intake. DGLA and AA can be obtained from animal meat and fat, and by desaturation and chain elongation of LA. The major dietary source of EPA and DHA is fish and seafood. EPA can also be obtained indirectly from ALA, although desaturation and chain elongation of ALA appears to be a less important pathway in humans.

Only the free form of the fatty acid precursors of eicosanoids can be utilized by the enzymes for conversion to the biologically active metabolites. However, the amount of precursor free fatty acid in the cytoplasm and circulating is usually low and so too is basal eicosanoid formation. Furthermore, basal eicosanoid formation may depend on dietary and adipose tissue fatty acid composition. The amount of eicosanoid precursor free fatty acids is controlled to a large extent by incorporation and release from cellular phospholipids. Which eicosanoids are produced during stimulated synthesis may depend on membrane fatty acid composition as well as the cell type involved. Dietary fatty acid composition, therefore, has the potential to affect basal and stimulated synthesis of eicosanoids and influence endothelial function, and thrombotic and inflammatory responses.

n-6 Fatty Acids and Risk of Cardiovascular Disease

Evidence that differences in *n*-6 fatty acid intake can influence cardiovascular disease risk derives from several sources. Population studies may provide useful data for establishing optimal intakes of *n*-6 fatty acids. However, valuable information on the potential mechanisms and effects of these fatty acids is derived from randomized controlled studies focusing on their impact on thrombosis, inflammation, endothelial function, and other cardiovascular risk factors.

Cardiovascular Disease: Population Studies

The incidence of cardiovascular disease within populations with either very high or very low intakes of *n*-6 fatty acids may provide some indication for optimal intakes of *n*-6 fatty acids. Within populations with low *n*-6 fatty acid intakes (<~3%), there would appear to be a benefit of having a higher *n*-6 fatty acid intake on cardiovascular disease risk reduction. These observations suggest that very low *n*-6 fatty acid intakes increase the risk for cardiovascular disease. The presence of EFA deficiency in a significant proportion of such populations may explain the increased risk. Several populations, including the Israelis, Taiwanese, and !Kung bushmen in the African Kalahari desert, have high to very high intakes of *n*-6 fatty acids. The contribution of *n*-6 fatty acids to total energy intake is approximately 10% in the Israelis and Taiwanese and approximately 30% in the !Kung bushmen. Rates of cardiovascular disease are low in the Taiwanese, where dietary *n*-6 fatty acids are obtained mainly from soybean oil, and estimated to be very low in the !Kung bushmen, where dietary *n*-6 fatty acids were obtained mainly from the monongo fruit and nut. In the Taiwanese, the soybean oil is refined but is accompanied by a diet rich in antioxidant polyphenols, notably from tea, fruits, and vegetables. In the !Kung bushmen, the oil is unrefined and is therefore likely to contain a range of phytochemicals. There is, however, a high prevalence of cardiovascular disease in the Israeli population, where *n*-6 PUFAs are obtained largely from refined sources. These observations suggest that a high *n*-6 fatty acid intake can be compatible with low risk of cardiovascular disease, but the dietary context may be very important. Given that *n*-6 fatty acids are susceptible to lipid peroxidation, high *n*-6 fatty acid intake may increase risk for cardiovascular disease when consumed against a background diet low in antioxidants. The potential impact on eicosanoid metabolism remains uncertain.

Several factors may need to be considered in the interpretation of the results of population studies. (1) The effect of LA on atherosclerosis and cardiovascular disease may depend on the background intake in the population being studied. (2) Any relationships observed may be confounded by intake of other foods from which LA derives. (3) LA may have differential effects on aspects of the etiology of cardiovascular disease, including endothelial function, thrombosis, arrhythmia, and atherosclerosis.

Thrombosis

Dietary fatty acids influence thrombosis by altering the activity and function of endothelial cells, platelets, and other circulating cells: effects that can be mediated, in part, by alterations in eicosanoid metabolism. Replacement of dietary saturated fatty acids with unsaturated fatty acids, including *n*-6 fatty acids, generally lowers the risk of thrombosis and cardiovascular disease. Furthermore, studies have shown that an increase in *n*-3 fatty acid intake can increase vasodilation, attenuate platelet aggregation, and alter circulating concentrations of factors involved in coagulation and fibrinolysis. The net effect of increasing *n*-3 fatty acid intake is a tendency toward reduced risk for thrombosis. These findings are supported by population studies demonstrating that *n*-3 fatty acids may reduce the risk of thrombosis. It remains uncertain whether the major factor influencing these functions is the absolute increase in *n*-3 fatty acids, or the relative proportions of *n*-6 and *n*-3 fatty acids in the diet and cell membranes. There is evidence, however, that increased *n*-3 fatty acid intake may be more beneficial in populations consuming relatively small quantities of fish, which includes many Western populations.

Much of the evidence for a potential impact of *n*-6 fatty acids on thrombosis derives from research on platelet function. The role of platelets in thrombosis is established and the influence of fatty acid intake on platelet function has been assessed in many studies. Platelets play a part in thrombosis by adhering to, and aggregating at, the site of injury. Platelet reactivity and increased platelet activation may increase the risk of thrombosis. *In vitro* and *in vivo* studies assessing effects of *n*-6 fatty acids on platelet aggregation are inconsistent. To date there is little evidence that a high *n*-6 fatty acid diet in humans decreases platelet aggregation and some studies are suggestive of increased aggregation with high *n*-6 fatty acid diets, primarily in the form of LA. The effects of AA on platelet aggregation are also not clear. One of the main difficulties in interpreting these studies is the unresolved issue as to how the *in vitro* aggregation test reflects platelet function *in vivo*.

Inflammation

Inflammation is involved in many human diseases including cardiovascular conditions such as thrombosis, stroke, and atherosclerosis. Conditions associated with increased inflammation, such as inflammatory arthritis, dermatological conditions such as psoriasis and atopic dermatitis, chronic inflammatory bowel disease, autoimmune diseases, and bronchial asthma, appear to be beneficially influenced to a greater extent by *n*-3 fatty acids than by *n*-6 fatty acids.

Whether increased intake of *n*-6 fatty acids can exacerbate inflammation *via* increased production of proinflammatory eicosanoids remains uncertain. Results of *in vitro* studies and intervention studies in humans are generally consistent with this theoretical potential of *n*-6 fatty acids to enhance inflammation, at least in comparison to *n*-3 fatty acids and probably omega-9 (*n*-9) monounsaturated fatty acids. The importance of absolute and relative intakes of *n*-6 fatty acids to inflammatory processes also remains unclear. The effects of changes in *n*-6 fatty acid intake on inflammatory processes may depend on the background dietary fatty acid intake, as well as proportional and absolute intake of *n*-3 fatty acids.

Resolution of inflammation involves the reduction or removal of inflammatory cells and debris from inflamed sites, enabling the return to homeostasis. This process, which was initially considered to be a passive process, is now known to be rapidly initiated after acute challenges by cellular pathways that lead to the synthesis of lipid mediators such as lipoxins, resolvins, and protectins. The lipoxins (LXA_4 and LXB_4) are derived from AA and have anti-inflammatory and proresolution properties (**Figure 1**). They are biosynthesized by the sequential action of lipoxygenases or via interaction of aspirin with COX-2 and lipoxygenase. Recently, a novel family of lipid mediators, the resolvins and protectins, which are derived from EPA and DHA, has also been described and implicated in the resolution of inflammation (**Figure 1**). Resolvins and protectins are local mediators that are generated during spontaneous resolution of inflammation.

Cholesterol and Lipoproteins

The major classes of circulating lipoproteins in human plasma are chylomicrons, very low density lipoproteins (VLDL), low-density lipoproteins (LDL), and high-density lipoproteins (HDL). High fasting plasma concentrations of LDL cholesterol and triglycerides – predominantly circulating as part of VLDL – and low plasma concentrations of HDL cholesterol, are associated with increased risk of cardiovascular disease. Dietary fatty acids can influence lipoprotein metabolism and therefore have the potential to influence atherosclerosis and cardiovascular disease risk. Most studies examining the effects of *n*-6 PUFAs on cholesterol metabolism have focused on LA.

It is now established that LDL cholesterol lowering reduces the risk of cardiovascular disease. In the fasting state, LDL is the major cholesterol-carrying lipoprotein in human plasma. The mechanisms through which raised plasma LDL cholesterol concentrations increase cardiovascular disease risk are not entirely understood, but oxidative modification of LDL is thought to be involved. An increase in LA intake results in a lowering of plasma LDL cholesterol concentrations and therefore has the potential to reduce cardiovascular disease risk. These effects may not be linear over the entire range of LA intake and most of the benefits appear to be gained by moving from lower (<2% of energy) to moderate (~4–5% of energy) intakes. In addition, the effects of dietary *n*-6 PUFAs are less than half that of lowering dietary saturated fatty acids. Therefore, if total fat intake is maintained, the LDL cholesterol lowering effects of increasing *n*-6 PUFA intake are greatly enhanced if saturated fatty acid intake is also decreased.

HDL cholesterol is inversely associated with cardiovascular disease risk. The mechanism by which HDL reduces cardiovascular disease risk may involve reverse cholesterol transport and reductions in cholesterol accumulation in the arterial wall. Intakes of LA

within the normal ranges of intakes in most populations do not appear to alter HDL cholesterol concentrations. However, very high intakes – more than 12% of energy – can lower HDL cholesterol concentrations.

Oxidative Stress

Several lines of evidence suggest that oxidatively modified LDL plays an important role in the development of atherosclerosis. Oxidative modification of LDL involves peroxidation of PUFAs. LDL particles enriched in PUFAs have been shown to be more susceptive to oxidative modification compared with LDL particles rich in monounsaturated fatty acids. Others have also suggested that a diet high in PUFAs may overwhelm the antioxidant defenses of cells. In particular, studies have shown that LA-enriched LDL is more prone to *in vitro* oxidation than oleic acid-enriched LDL. Concern also remains with respect to the potential for increased lipid peroxidation following *n*-3 fatty acids. However, much of the early literature relating to PUFAs and lipid peroxidation is based on indirect and nonspecific assays, including measurement of LDL oxidative susceptibility, which relies on the isolation of LDL from plasma. In this regard, the recent discovery of F_2-isoprostanes, which are nonenzymatic prostaglandin-like products of free radical peroxidation of AA, has allowed for the direct assessment of *in vivo* lipid peroxidation. There is now good evidence that quantitation of F_2-isoprostanes provides a reliable measure of *in vivo* oxidative stress. Using measurement of F_2-isoprostanes, recent data have demonstrated that *n*-3 fatty acids decrease oxidative stress. It has also been suggested that the concentration of PUFAs may be a more important factor affecting lipid peroxidation than the degree of unsaturation. Further research using better markers of lipid peroxidation is required before definitive statements can be made relating to the effect of *n*-6 fatty acids and indeed PUFAs in general, on oxidative stress.

Blood Pressure

The possible effects of dietary fatty acids on blood pressure have been explored in population studies and dietary intervention trials. Although a hypotensive influence of *n*-6 fatty acids was suggested in early clinical studies, this has not been confirmed in subsequent randomized controlled trials. In normotensive individuals, randomized controlled trials have shown no consistent effects on blood pressure with dietary modifications to change the intake of *n*-6 fatty acids. LA supplements have also produced either no change or reduction in blood pressure. Additionally, studies in hypertensives have shown no consistent effects of *n*-6 fatty acids on blood pressure. Current data suggest that *n*-6 fatty acids, when substituted for saturated fatty acids, may have some blood pressure lowering effect if part of complex dietary changes that include increases in fruit, nuts, and vegetable consumption and low-fat dairy products. These effects may also be enhanced by moderation of salt intake.

Conclusions

Diets low in *n*-6 fatty acids, principally LA, appear to be associated with an increased risk of cardiovascular disease. The results of studies examining the effects of LA on risk factors for atherosclerosis and cardiovascular disease are consistent with this observation. An increase in *n*-6 PUFA intake from a low to a moderate intake level, in conjunction with decreases in total and saturated fat intake, may beneficially influence lipoprotein metabolism, lower blood pressure, and reduce cardiovascular disease risk. Observations in populations with high *n*-6 PUFA intake indicate that high intakes of *n*-6 fatty acids (>10%) can occur together with low rates of cardiovascular disease and possibly also cancer. However, where antioxidant composition of the diet is low, there is the potential for increased risk of cardiovascular disease. An increased susceptibility of PUFAs to oxidative damage, particularly in the presence of low concentrations of protective antioxidants, may be an important factor. The source of *n*-6 PUFAs in the diet, refined versus unrefined, and the composition of the background diet, may therefore be important determinants of whether high *n*-6 fatty acid intake increases or decreases the risk of cardiovascular disease. In addition, the quantity of *n*-6 to *n*-3 fatty acids in the diet may also play an important role in determining cardiovascular risk.

Available data suggest that the relative proportion of all the classes of dietary fatty acids may be more important and relevant to cardiovascular risk reduction than any single class of fatty acids. The importance of *n*-6 fatty acids in relation to cardiovascular risk was recently summarized in an American Heart Association statement, which concluded that consumption of at least 5–10% of energy from *n*-6 fatty acids in the context of other appropriate lifestyle and dietary behavior may reduce the risk of coronary heart disease relative to low intake.

Further Reading

Grundy, SM., 1996. Dietary fat. In: Ziegler, EE, Filer Jr, LJ (Eds.), Present Knowledge in Nutrition, 7th edn. ILSI Press, Washington, DC, pp. 44–57.

Harris, WS, Mozaffarian, D, Rimm, E, et al., 2009. Omega-6 fatty acids and risk for cardiovascular disease: A science advisory from the American Heart Association Nutrition Subcommittee of the Council on Nutrition, Physical Activity, and Metabolism; Council on Cardiovascular Nursing; and Council on Epidemiology and Prevention. Circulation 119, 902–907.

Hodgson, JM, Wahlqvist, ML, Boxall, JA, Balazs, NDH, 1993. Can linoleic acid contribute to coronary artery disease? American Journal of Clinical Nutrition 58, 228–234.

Hodgson, JM, Wahlqvist, ML, Hsu-Hage, B., 1995. Diet, hyperlipidaemia and cardiovascular disease. Asia Pacific Journal of Clinical Nutrition 4, 304–313.

Hornsrtra, G, Barth, CA, Galli, C, et al., 1998. Functional food science and the cardiovascular system. British Journal of Nutrition 80 (supplement 1), S113–S146.

Horrobin, DF (Ed.), 1990. Omega-6 Essential Fatty Acids: Pathophysiology and Roles in Clinical Medicine. Wiley-Liss, New York.
Howe, P, Meyer, B, Record, S, Baghurst, K., 2006. Dietary intake of long-chain omega-3 polyunsaturated fatty acids: Contribution of meat sources. Nutrition 22, 47–53.
Jones, GP., 1997. Fats. In: Wahlqvist, ML (Ed.), Food and Nutrition, Australia, Asia and the Pacific. Allen & Unwin, Sydney, pp. 205–214.
Jones, PJH, Kubow, S, 1999. Lipids, sterols and their metabolism. In: Shils, ME, Olson, JA, Shike, M, Ross, AC (Eds.), Modern Nutrition in Health and Disease, 9th edn. Williams and Wilkins, Baltimore, pp. 67–94.
Kitson, AP, Stroud, CK, Stark, KD., 2010. Elevated production of docosahexaenoic acid in females: Potential molecular mechanisms. Lipids 45, 209–224.
Knapp, HR., 1997. Dietary fatty acids in human thrombosis and hemostasis. American Journal of Clinical Nutrition 6 (supplement 5), 1687S–1698S.
Lyu, LC, Shieh, MJ, Posner, BM, et al., 1994. Relationship between dietary intake, lipoproteins and apolipoproteins in Taipei and Framingham. American Journal of Clinical Nutrition 60, 765–774.
Mensink, R, Connor, W (Eds.), 1996. Nutrition. Current Opinion in Lipidology, 7, pp. 1–53.
Mori, TA., 2004. Effect of fish and fish oil-derived omega-3 fatty acids on lipid oxidation. Redox Report 9 (4), 193–197.
Mori, TA, Burke, V, Beilin, LJ., 2007. Dietary Fats and Blood Pressure. In: Lip, GYH, Hall, JE (Eds.), Comprehensive Hypertension. Section I: Epidemiology. Elsevier, Philadelphia, pp. 77–88.
National Health and Medical Research Council (1992) The role of polyunsaturated fats in the Australian diet: Report of the NHMRC Working Party, pp.1–81. Canberra: Australian Government Publishing Service. ISBN: 0644 143126.
Salem N, Simopoulos AP, Galli C, Lagarde M, and Knapp HR (eds.) (1996) Fatty acids and lipids from cell biology to human disease: Proceedings of the 2nd International Congress of the International Society for the Study of Fatty Acids and Lipids. *Lipids* 31(supplement): S1–S325.
Serhan, CN, Yacoubian, S, Yang, R., 2008. Anti-inflammatory and proresolving lipid mediators. Annual Review of Pathology-Mechanisms of Disease 3, 279–312.
Truswell, AS, 1977. Diet and nutrition of hunter gatherers. Ciba Foundation Symposium 49, 213–221.
Yam, D, Eliraz, A, Berry, EM., 1996. Diet and disease – The Israeli paradox: Possible dangers of a high omega-6 polyunsaturated fatty acid diet. Israeli Journal of Medical Sciences 32, 1134–1143.

Fiber intake and resulting health benefits

Rebekah Schulz and Joanne Slavin, Department of Food Science and Nutrition, University of Minnesota, St. Paul, MN. United States

© 2023 Elsevier Ltd. All rights reserved.

Introduction 240
Definition of dietary fiber 240
Current fiber dietary guidelines (US, Europe, UK) 240
Food labeling 241
Fiber: food sources and functional fiber 242
Health claims 242
Fiber intake: amount and sources 242
Health impacts of fiber 243
Cardiovascular disease 243
Metabolic health 243
Gut microbiome 243
Gut health 244
Barriers and solutions to meeting recommendations 244
Conclusion/summary/outlook 245
References 245

Key points

- Fiber is an essential component of a healthy diet
- Due to the various ways to quantify and define dietary fiber, reaching a consensus on recommended daily intake is challenging for both regulatory bodies and practitioners
- Recommended daily dietary fiber intakes across age groups and countries
- Overall, most age groups are unable to meet general fiber intake recommendations
- Fiber is associated with short term and long-term health benefits, both directly related to gut health as well as other physiological benefits
- Plant based diets will benefit not only planetary health but also human health through overall increased fiber intake

List of abbreviations

AI Adequate Intake
CVD Cardiovascular Disease
DRI Dietary Reference Intakes
DP Degree of Polymerization
DRV Daily Reference Value
EAR Estimated Average Requirement
EFSA European Food Safety Authority
FAO Food and Agriculture Organization
FDA Food and Drug Administration
MU Monomeric Units
NAM National Academy of Medicine
NDNS National Diet and Nutrition Survey
NHANES National Health and Nutrition Examination Survey
NSP Non-Starch Polysaccharides
IOM Institute of Medicine
MU Monosaccharide Unit
RACC Reference Amount Customarily Consumed
RCT Randomized Control Trial
RDA Recommended Daily Allowance

RO Resistant Oligosaccharides
RS Resistant Starches
SACN Scientific Advisory Committee on Nutrition
SCFA Short Chain Fatty Acids
TDF Total Dietary Fiber
WHO World Health Organization

Introduction

Fiber is an essential component of a healthy diet. As knowledge surrounding fiber has evolved, so too has understanding regarding its resulting physiological impacts. However, despite the consensus on its health benefits, the quantification methods and definitions vary greatly, making a universally agreed upon intake recommendation challenging across age groups. Consequently, international recommendations on total dietary fiber (TDF) intake vary by regulatory bodies, with most agreeing on an intake of greater than 25 g/day for the average adult and more diversity existing among recommendations for children. Working toward further collaboration on a fiber definition and recommendation will aid in international harmonization as it will be beneficial in being able to accurately measure intake as well as establish internationally accepted food labeling, conduct nutrition research, and accurately synthesize food composition tables.

Despite the consensus on fiber's beneficial health effects, fiber intake across the globe generally only meets half of the recommended amount. This is due to a number of barriers including the consumption of processed food due to a standard Western diet, and the unattainability of recommendations, especially for children. Potential solutions to this include the fortification of foods with dietary fiber, keeping the calorie content at an advisable level while helping individuals meet the recommendations, as well as the general move toward plant-based foods to mitigate climate change, thereby increasing overall fiber intake.

As fiber intake is linked to beneficial health outcomes, individuals consuming higher fiber diets will experience the beneficial and preventative role that appropriate fiber consumption can play in human health.

Definition of dietary fiber

Fiber is classified in a variety of ways including source, such as cereal and legume fiber, chemistry, such as oligosaccharides (carbohydrates with degree of polymerization (DP) 3–9) and polysaccharides (carbohydrates with DP > 9), physiochemical properties, such as solubility and viscosity, and food matrix such as pre-processing like grinding and cooking. These numerous classifications translate into a variety of definitions from regulatory bodies. Prebiotics are an additional area related to dietary fiber with many qualifying as dietary fiber. They are described as substances selectively utilized by the host to benefit specific growth or activity of bacteria in the colon and therefore improving host health.

The Institute of Medicine (IOM), now titled the National Academy of Medicine (NAM) classifies fiber as plant derived, nonstarch polysaccharides and lignin that are not digested or absorbed by the small intestine (Turner and Lupton, 2011). The UK based Scientific Advisory Committee on Nutrition (SCAN) extends this definition to including certain oligosaccharides as falling under the fiber definition (SACN Carbohydrates, 2015). The Codex Alimentarius Commission details perhaps the most comprehensive definition, describing dietary fiber as carbohydrates and lignin that are neither digested nor absorbed with three or more monomeric units (MU), including intact and intrinsic non-starch polysaccharides (NSP) (MU > 10), resistant oligosaccharides (RO) (MU 3–9), and resistant starch (RS) (MU > 10) (WHO/FAO, 2009). Countries who adhere to Codex guidelines can determine how much of this definition they align with and thus adhere to. Codex updated their definition in 2014 to include any substance as dietary fiber regardless of its source as long as it provides the stated physiological effect (Jones, 2014). This was with the goal of creating international harmonization around nutrition labeling, food composition tables, nutrition research and dietary recommendations for healthy people. However, there is still discrepancy between different regulatory bodies on the definition of fiber and therefore the recommendations, measured intake, and study exposure guidelines.

Current fiber dietary guidelines (US, Europe, UK)

Historically, anecdotal evidence indicated fiber as beneficial due to an observed link between carbohydrate and fiber consumption in non-Western countries and stool size and transit time related to overall health. As scientific methods developed to analyze fiber and measure its intake against epidemiological observances, regulatory bodies in Western countries have developed recommended intakes of dietary fiber for their populations since the 1970s.

Table 1 Fiber recommendations for children.

Country/region	*Source*	*Year of publication*	*Daily fiber recommendation*
USA	National Academy of Medicine	2017 (Quagliani and Felt-Gunderson, 2016)	Age 1–3 years: 19 g/day Age 4–8 years: 25 g/day Age 9–13 years: 31 g/day (M) 26 g/day (W) Age 14–19: 38 g/day (M) 26 g/day women)
UK	UK Scientific Advisory Committee on Nutrition	2015 (SACN Carbohydrates, 2015)	Age 2–5 years: 15 g/day Age 5–11 years: 20 g/day Age 11–16 years: 25 g/day Age 16–18: 30 g/day
Europe	European Food Safety Authority	2010 (Efsa Panel on Dietetic Products, 2010)	Age 1–3 years: 10 g/day Age 4–6 years: 14 g/day Age 7–10 years: 16 g/day Age 11–14 years: 19 g/day Age 15–17 years: 21 g/day

For different nutrients, dietary reference intakes (DRIs) are established as reference ranges to set dietary guidelines and assess intakes. The DRIs for fiber include an adequate intake (AI), the nutrient amount required to meet the needs of an observed or experimental group of healthy people when an Estimated Average Requirement (EAR—the nutrient amount needed to meet the needs of half of the healthy individuals in a group) or the Recommended Dietary Allowance (RDA—a nutrient amount needed to meet the daily requirements of nearly all (97–98%) of healthy individuals in a group) cannot be set. Overall, countries make AI recommendations for adults, older adults, and children, aged 1–18 years old. There are no current recommendations for healthy infants age 0–6 months who are milk-fed as breast milk does not contain dietary fiber. Similarly, no recommendations exist for infants 7–12 months due to lack of data as well as low consensus of whole food intake overall.

The AI for TDF in US children and adults was informed by epidemiological studies examining cardiovascular disease (CVD) (**Tables 1** and **2**). European countries have a wide range of recommended intakes for TDF in children and adults, also informed by epidemiological studies examining its benefits for bowel function, blood lipids, glucose response, and protection from noncommunicable diseases (**Tables 1** and **2**). Similarly, SCAN established recommended intakes for TDF in children and adults based on epidemiological findings.

Overall, fiber recommendations are for TDF and do not specify the types of fiber. This is largely due to lack of consensus on how to quantify fiber. However, as different foods have different fiber types and these types vary in their physiological benefit, future guidelines could potentially begin to set intakes for specific fiber types to best serve their population health. For example, vegetables contain primarily non-starch polysaccharides (NSP) such as cellulose, pectic polysaccharides, and hemicelluloses, while legumes additionally contain resistant starch (RS) and resistant oligosaccharides (RO) such as alpha-Galactosides, beta-Fructo-oligosaccharides, resistant dextrins, and polydextrins.

Food labeling

In the US, TDF is required to be listed on the nutrition facts label. Differentiating soluble and insoluble fiber was also historically required on the label but recently removed from the requirements due to inconsistent scientific findings on their previously understood physiological effects (Medicine et al., 2001). In 2016, the United States Food and Drug Administration (FDA) changed the dietary fiber Daily Value (DV) recommendation from 25 g/day to 28 g/day based on a 2000 calorie diet, requiring food manufacturers to update their labels and to reassess whether they would need to reformulate their product to keep in alignment with previously stated health claims.

Table 2 Fiber recommendations for adults.

Country/region	*Source*	*Year of publication*	*Daily fiber recommendation*
USA	National Academy of Medicine	2017 (Quagliani and Felt-Gunderson, 2016)	Age 19–50: 38 g/day (M) 25 g/day (W) Age 50+: 30 g/day (M) 21 g/day (W)
UK	UK Scientific Advisory Committee on Nutrition	2015 (SACN Carbohydrates, 2015)	Adults: 30 g/day
Europe	European Food Safety Authority	2010 (Efsa Panel on Dietetic Products, 2010)	Adults: >25 g/day

Fiber: food sources and functional fiber

Most fiber intake comes from plant foods, including fruit, vegetables, nuts, pulses, and whole grains. As fruit and vegetables have a high-water content, most only contain 1–3 g of fiber per serving (Slavin, 2008). As a result, whole grains are the primary source of fiber intake overall as they typically contain more grams of fiber per serving. Additionally, fortified high-fiber cereals are a good choice when it comes to obtaining a large amount of fiber in a food product serving. As these sources are whole food sources, they also contain vitamins, minerals, and other micronutrients that can provide additional benefit to the consumer.

Due to even most plant foods containing low amounts of fiber per serving, enriching foods with added fiber to increase the total amount or to aid in functional properties is a common approach in the food industry. Added or functional fiber is approved differently depending on the regulatory body. In the US, the NAM defines added or functional fiber as isolated non-digestible carbohydrates that confer a beneficial physiological response to the consumer. This isolated fiber can either be synthetically made or derived from nondigestible carbohydrates and is allowed on the nutrition facts panel if it analytically and physiologically behaves like fiber. As a result, while some functional fibers can count toward the TDF, others cannot if they have not yet been proven to demonstrate a beneficial physiological effect to the consumer. Physiological effects include lowering blood glucose and cholesterol levels, lowering blood pressure, improving laxation, increasing mineral absorption, and reducing energy intake. Trials that further examine various fibers' physiological effects will aid in developing and labeling products that better help consumers meet their daily requirements.

Health claims

Claims related to fiber differ across countries. For a European health claim to state that a food is "high in" a particular fiber, the food must contain at least 6 g per 100 g or 3 g per 100 kcal of the fiber as defined by the EU commission (Stephen et al., 2017; Commission, 2006). Similarly, for a food to be a "source of fiber," it must contain at least 3 g of fiber per 100 g or at least 1.5 g of fiber per 100 kcal. Specific to Europe, health claims indicate specific fibers and their demonstrated health benefits, such as pectin improving glycemic response or wheat bran fiber related to fecal bulk. In contrast, the US does not specify type of fiber, only the amount and the food source. For a food to be "high" in fiber in the US, the United States Food and Drug Administration (FDA) outlines that the food must contain greater than 20% of the DRV per reference amount customarily consumed (RACC), and between 10 and 19% to be considered a "good source of fiber" (Regulations, 2012a,b) While these claims have proven benefits in an experimental setting, the intake amount required to meet the health claim is often difficult for the average consumer to achieve outside of an experimental diet.

Fiber intake: amount and sources

National food surveys and 24 h recalls are the preferred method for collecting population dietary intake. In the US, the National Health and Nutrition Examination Survey (NHANES) measures intakes via 24 h recalls every 2 years from age categories: 1–4, 4–12, 13–18, 18+, 18–39, 40–59, and 60 years old and over. The UK utilizes the National Diet and Nutrition Survey (NDNS) which assesses the diet, nutrient intake, and nutritional status of the UK population. Europe utilizers a variety of surveys based on the country of interest.

A review by Stephen et al. assessed 29 studies on fiber intake in Europe, the UK, and the US, and found that overall, fiber intake in adult males and females ranged from 15 to 25 g/day and 14–21 g/day respectively (Stephen et al., 2017). The authors noted that overall intake was higher in Europe as compared to North America. Intakes in children varied more widely and thus were less generalizable. **Table 3** outlines the most recent data from NHANES and the NDNS for various age groups (Stephen et al., 2017).

Table 3 Total dietary fiber intake for specific age groups as assessed by the National Health and Nutrition Examination Survey (NHANES) and the National Diet and Nutrition Survey (NDNS).

Country	*Survey*	*Year of survey*	*Age*	*Total fiber intake (g/d)*
USA	NHANES	2011–2012 (Agricultre, 2012)	20+	M: 19.1; F: 15.5
			60+	M: 19.6; F: 15.9
			13–18	M: 18.1; F: 12.5
			4–12	M: 15.4; F: 13.9
			0–4	12.1
UK	NDNS	2009–2012 (Agency, 2014)	0–4	8.2
			4–10	M: 11.5; F: 10.7
			11–18	M: 12.8; F: 10.7
			19–64	M: 14.7; F: 12.8
			65+	M: 14.9; F: 13.9

Stephen et al. also noted that grain products were the largest source of fiber in all countries for adults, while there was greater variance in fiber sources in children across and within countries.

Health impacts of fiber

There are both short- and long-term benefits from consuming dietary fiber. These benefits are both directly related to gastrointestinal health as well as extend to other organs and systems. Short-term benefits include immune system support, increasing mineral absorption in the colon, and preventing or ameliorating autoimmune diseases. Long-term benefits include reducing the risk of non-communicable diseases and reducing bowel conditions such as constipation.

Cardiovascular disease

Dietary fiber has long shown its ability to protect against cardiovascular disease (CVD) and this relationship is the main epidemiological finding by which the US sets their AI. A systematic review by Threapleton et al. assessing 22 prospective cohort studies found an inverse association between total dietary fiber intake and risk of CVD and coronary heart disease (Threapleton et al., 2013a,b). However, despite this well-established relationship, the mechanisms of action (MoA) are not well known. Some proposed MoA include fiber's ability to mitigate CVD risk factors such as soluble fiber's ability to decrease low density lipoprotein (LDL) cholesterol, thereby decreasing overall serum cholesterol, and attenuating CVD risk (Buttriss and Stokes, 2008). Additional studies have shown higher fiber diets' ability to decrease C-reactive protein, a marker of systemic inflammation, modulate postprandial glycemic response, decrease body fat through overall caloric intake modulation, and decrease hypertension, thereby protecting against CVD.

In the US, the FDA has allowed specific authorized health claims linking beta-glucan consumption from oats, barley, and psyllium husk to reduced risk of CVD. EFSA also allows similar function health claims, extending the food sources to oat, oat bran, barley, and barley bran.

Metabolic health

Various trials have examined the impacts of high fiber diet on overall metabolic health, analyzing markers such as insulin sensitivity, HbA1C, lipid profiles, body weight, and C-reactive protein. An interventional trial examining 111 overweight adults with features of metabolic syndrome found that the adults supplemented with the high cereal fiber (HCF) diet had improved insulin sensitivity as compared to control, high protein (HP), and HCF/HP groups (Weickert et al., 2011a,b). An observational trial examining 190 Japanese adults without type 2 diabetes found that an increase of dietary fiber to carbohydrate ratio in the diet resulted in decreased HbA1C (Morimoto et al., 2018). A systematic review and meta-analysis also found that a high-fiber diet resulted in improved insulin sensitivity, HbA1C, lipid profile, body weight, and C-reactive protein (Reynolds et al., 2020).

Regarding body weight and abdominal adiposity, current evidence-based trials provide conflicting results, with some indicating that increased fiber diets result in lower body weights and decreased overall caloric intake and others indicating that changes in weight and intake were minor or nonsignificant. Most of these studies indicate the ability of dietary fiber intake to decrease overall caloric intake due to appetite suppression related to increase satiation. Future studies should further examine these links as well as design longer-term trials to better understand any potential linkages. Additionally, trials examining dietary fiber intake's effect on body weight maintenance would be of interest.

Two systematic reviews examining prospective cohort studies found that fiber intake (TDF) was associated with a decrease in the long-term risk of the development of type 2 diabetes mellitus (Threapleton et al., 2013a,b; Yao et al., 2014). Cereal fiber was the primary fiber responsible for reduced diabetes risk. Additional prospective cohort studies have also indicated that insoluble cereal dietary fiber and whole grains are those associated with decreased risk for type 2 diabetes mellitus (Schulze et al., 2007; de Munter et al., 2007). While the MoA for this association is not fully known, proposed mechanisms include fiber's ability to increase gastric distention and delay emptying, thereby decreasing overall caloric intake, soluble fiber's ability to modulate the absorption of glucose, and fiber's ability to increase the absorption of nutrients, and stimulate beneficial colonic bacteria which may impact the gut-brain axis and not yet known type 2 diabetes linkages.

Gut microbiome

The gut microbiome and its interconnected and complex role in overall health is currently of great scientific and consumer interest. As of yet, most studies examining fiber's effect on the gut microbiome come from animal models, specifically rodent-based studies. One such study demonstrated the role of dietary fiber's role in protecting gnotobiotic rodents' colonic mucosal barrier from erosion and therefore overall gastrointestinal health (Desai et al., 2016). While these studies have not been conducted in humans, they point toward potentially similar benefits as well as trial designs to emulate.

Dietary fiber is known to improve overall microbiome health and metabolite production. However, a randomized control trial (RCT) ($n = 69$) examining MoA for fiber's ability to improve insulin sensitivity found that among overweight adults, those who were supplemented with the high cereal fiber isocaloric diet had no alterations in short-chain fatty acid (SCFAs) production or gut microbiota ecology despite improved insulin sensitivity in this group(Weickert et al., 2011a,b). As cereal fibers tend to be both primarily insoluble and nonfermentable, perhaps their role in gut health is more related to their ability to increase stool bulk through microbial mass growth.

As a subset of fiber, prebiotics are also of great consumer interest as well as scientific exploration. While various studies exist pointing to their beneficial effects on various colonic bacterial strains, there are currently no prospective, cohort studies which link changes in the microbiota as a result of prebiotic exposure to health outcomes (Brownawell et al., 2012). As a result, for prebiotics to be listed as a dietary fiber, studies need to be designed to indicate a linkage between their consumption and a resulting health benefit. Standardization of health outcome quantification methods as well as prebiotic protocols in human studies are key to establishing this linkage.

Gut health

Dietary fiber's beneficial effects on gastrointestinal health are well established. Fiber impacts the gut overall by increasing chewing and saliva production, increasing gastric distention and delaying emptying (potentially decreasing caloric intake), delaying nutrient absorption (thereby modulating glucose and insulin response), altering gut hormones such as GLP-1, PYY, and ghrelin (thereby modulating hunger cues), and aiding in colonic fermentation and the production of short chain fatty acids (SCFAs) (Dayib et al., 2020).

Due to differing physiochemical effects, different fibers will have differing effects on stool weight. Some fibers, such as pectin and inulin are extensively fermented (over 90%), while others are poorly fermented, such as purified cellulose (less than 10%). Overall, however, even fibers that are extensively fermented are able to increase stool weight since the fermentation process increases the bacterial mass, resulting in a greater water binding capacity.

Regarding gut motility, if transit time is already normal (2 days, 48 h), increasing fiber intake will not alter or improve this transit time (Marlett et al., 2000). However, if intake is low, fiber intake will normalize transit time to 2–4 days (Harvey et al., 1973). A review conducted in 2007 found that of 150 studies examining the effects of fiber intake on stool bulk, specific types of fiber such as raw wheat and fruit and vegetable fiber were more effective in increasing stool bulk as compared to soy, legume, and pectin fibers (Elia and Cummings, 2007). An additional meta-analysis found that dietary fiber intake as compared to placebo in constipated patients resulted in increased stool frequency (Yang et al., 2012). However, they noted that associations between fiber intake and improved stool consistency, treatment success, laxative use and painful defecation would need to be further explored.

Barriers and solutions to meeting recommendations

As fiber is labeled a nutrient of concern in the US due to low intake, also reflected worldwide, it is important to assess the barriers to meeting dietary fiber intake recommendations and propose solutions.

Processed food has been touted both as a barrier to fiber intake as well as a way to increase fiber consumption. A typical Western diet often contains high amounts of ultra-processed foods which could either contain large amounts of refined grains and therefore lack fiber, or alternatively, could include ultra-processed foods that have been enriched or fortified to include fiber in the form of whole grains or functional fibers. As a result, a solution could be the development of standards around fiber benchmarks in ultra-processed foods to ensure that populations consuming these foods are able to meet the fiber intake recommendations and therefore confer the beneficial health effects.

Clean food labeling is proposed as a barrier in fiber intake as consumers are looking for labels that have five or less ingredients, specifically ingredients that they can pronounce and understand. As a result, if an enriched or fortified food contains a chemically based and described fiber ingredient, this subgroup of consumers may avoid the product, therefore decreasing their consumption of fiber. Additionally, as described above, these consumers also avoid ultra-processed foods due to the high number of ingredients, therefore also potentially missing food products that have a high source of fiber. Education and awareness around nutrition facts labels, ingredient descriptions, and the health benefits of ultra-processed foods could aid in increasing literacy and therefore consumption of high fiber food products.

Many sources note the unattainability of most fiber intake recommendations for both children and adults due to fiber intake also being linked to calorie intake. Enriching and fortifying foods products could aid in providing consumers with appropriate fiber amounts while keeping calorie intake at desirable level (Jones, 2014). Additionally, as a greater consensus is reached on the categorization of dietary fiber and the specific health benefits of each fiber type, perhaps more realistic and defined dietary fiber intake recommendations will be developed which will be both more realistic for the consumer and more precise for overall health.

Recent diet trends have tended to favor less fiber in the diet such as low carbohydrate diets, low gluten diets, the ketogenic diet, the paleo diet, and the low FODMAP (fermentable oligosaccharides, disaccharides, monosaccharides, and polyols) diet. Individuals participate in these diets for a range of reasons, including weight loss, specific health outcomes related to communicable and

noncommunicable diseases, and popular culture trends. While each of these diets have various trials indicating health outcomes, most are short-term in nature and fail to examine any negative consequences due to low fiber consumption. However, the recent plant-based movement could counter some of these effects. To mitigate climate change related to food production and thus meet sustainability targets established by organizations such as the Food Agriculture Organization (FAO), EAT-Lancet, and the World Health Organization (WHO) for planetary health, plant-based diets are encouraged across cultural and socioeconomic groups. As a result, fiber intake could being to increase overall as more individuals adopt plant-based diets and move away from the majority of calories coming from fiber-poor food sources.

Conclusion/summary/outlook

Dietary fiber is an essential nutrient with many known and yet to be discovered health benefits. These benefits include both short and long-term health impacts such as immune system support, increasing mineral absorption in the colon, ameliorating autoimmune disease, and supporting colonic health. While countries recommend different dietary fiber amounts across populations and age groups, dietary intakes often fall short, averaging at less than half of recommended amounts. This is due to various barriers such as the consumption of processed food, low fiber diets, lofty intake recommendations, overall caloric intake considerations, and clean-label preferences. Fiber can also be added to the diet as a dietary supplement, bulk laxative, or be added to popular foods and drinks with functional fiber. Thus, processed foods can also be a solution to adding fiber into the food supply. As education surrounding the importance of fiber and its resulting health benefits continues, as well as overall adoption of plant-based diets to meet climate goals increase, overall fiber intake should better match recommended amounts with consequent health benefits for populations worldwide.

References

Brownawell, A.M., Caers, W., Gibson, G.R., Kendall, C.W., Lewis, K.D., Ringel, Y., Slavin, J.L., 2012. Prebiotics and the health benefits of fiber: current regulatory status, future research, and goals. J. Nutr. 142 (5), 962–974.

Buttriss, J.L., Stokes, C.S., 2008. Dietary fibre and health: an overview. Nutr. Bull. 33 (3), 186–200.

Code of Federal Regulations, 2012a. Health Claims: Fiber-Containing Grain Products, Fruits, and Vegetables and Cancer.

Code of Federal Regulations, 2012b. Health Claims: Fruits, Vegetables, and Grain Products that Contain Fiber, Particularly Soluble Fiber, and Risk of Coronary Heart Disease.

Dayib, M., Larson, J., Slavin, J., 2020. Dietary fibers reduce obesity-related disorders: mechanisms of action. Curr. Opin. Clin. Nutr. Metab. Care 23 (6), 445–450.

de Munter, J.S.L., Hu, F.B., Spiegelman, D., Franz, M., van Dam, R.M., 2007. Whole grain, bran, and germ intake and risk of type 2 diabetes: a prospective cohort study and systematic review. PLoS Med. 4 (8), 1385–1395.

Desai, M.S., Seekatz, A.M., Koropatkin, N.M., Kamada, N., Hickey, C.A., Wolter, M., Pudlo, N.A., Kitamoto, S., Terrapon, N., Muller, A., et al., 2016. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. Cell 167 (5), 1339–1353.e21.

EFSA Panel on Dietetic Products Nutrition and Allergies, 2010. Scientific opinion on dietary reference values for carbohydrates and dietary fibre. EFSA J. 8 (3).

Elia, M., Cummings, J.H., 2007. Physiological aspects of energy metabolism and gastrointestinal effects of carbohydrates. Eur. J. Clin. Nutr. 61 (S1), S40–S74.

European Commission, 2006. Regulation (EC) No 1924/2006 of the European Parliament and of the Council of 20 December 2006 on Nutrition and Health Claims Made on Foods. Official Journal of the European Union.

Harvey, R.F., Pomare, E.W., Heaton, K.W., 1973. Effects of increased dietary fibre on intestinal transit. Lancet 1 (7815), 1278–1280.

Institute of Medicine, National Academy of Sciences, 2001. Dietary Reference Intakes: Proposed Definition of Dietary Fiber. National Academies Press, Washington, DC.

Jones, J.M., 2014. CODEX-aligned dietary fiber definitions help to bridge the 'fiber gap'. Nutr. J. 13, 34.

Marlett, J.A., Kajs, T.M., Fischer, M.H., 2000. An unfermented gel component of psyllium seed husk promotes laxation as a lubricant in humans. Am. J. Clin. Nutr. 72 (3), 784–789.

Morimoto, N., Kasuga, C., Tanaka, A., Kamachi, K., Ai, M., Urayama, K.Y., Tanaka, A., 2018. Association between dietary fibre:carbohydrate intake ratio and insulin resistance in Japanese adults without type 2 diabetes. Br. J. Nutr. 119 (6), 620–628.

Public Health England and the Food Standards Agency, 2014. Results of the National Diet and Nutrition Survey (NDNS) Rolling Programme for 2008 and 2009 to 2011 and 2012.

Quagliani, D., Felt-Gunderson, P., 2016. Closing America's fiber intake gap: communication strategies from a food and fiber summit. Am. J. Lifestyle Med. 11 (1), 80–85.

Reynolds, A.N., Akerman, A.P., Mann, J., 2020. Dietary fibre and whole grains in diabetes management: systematic review and meta-analyses. PLoS Med. 17 (3), e1003053.

SACN Carbohydrates and Health Report, 2015. Public Health England, London, UK.

Schulze, M.B., Schulz, M., Heidemann, C., Schienkiewitz, A., Hoffmann, K., Boeing, H., 2007. Fiber and magnesium intake and incidence of type 2 diabetes: a prospective study and meta-analysis. Arch. Intern. Med. 167 (9), 956–965.

Slavin, J.L., 2008. Position of the American Dietetic Association: health implications of dietary fiber. J. Am. Diet Assoc. 108 (10), 1716–1731.

Stephen, A.M., Champ, M.M.J., Cloran, S.J., Fleith, M., van Lieshout, L., Mejborn, H., Burley, V.J., 2017. Dietary fibre in Europe: current state of knowledge on definitions, sources, recommendations, intakes and relationships to health. Nutr. Res. Rev. 30 (2), 149–190.

Threapleton, D.E., Greenwood, D.C., Evans, C.E.L., Cleghorn, C.L., Nykjaer, C., Woodhead, C., Cade, J.E., Gale, C.P., Burley, V.J., 2013a. Dietary fibre intake and risk of cardiovascular disease: systematic review and meta-analysis. BMJ 347, f6879.

Threapleton, D.E., Greenwood, D.C., Evans, C., Cleghorn, C.L., Nykjaer, C., Woodhead, C., Burley, V.J., 2013b. Dietary fibre intake and diabetes risk: a systematic review and meta-analysis of prospective studies. Proc. Nutr. Soc. 72 (OCE4).

Turner, N.D., Lupton, J.R., 2011. Dietary fiber. Adv. Nutr. 2 (2), 151–152.

U.S. Department of Agriculture, 2012. What We Eat in America. Data Tables.

Weickert, M.O., Roden, M., Isken, F., Hoffmann, D., Nowotny, P., Osterhoff, M., Blaut, M., Alpert, C., Gögebakan, O., Bumke-Vogt, C., et al., 2011a. Effects of supplemented isoenergetic diets differing in cereal fiber and protein content on insulin sensitivity in overweight humans. Am. J. Clin. Nutr. 94 (2), 459–471.

Weickert, M.O., Arafat, A.M., Blaut, M., Alpert, C., Becker, N., Leupelt, V., Rudovich, N., Möhlig, M., Pfeiffer, A.F., 2011b. Changes in dominant groups of the gut microbiota do not explain cereal-fiber induced improvement of whole-body insulin sensitivity. Nutr. Metabol. 8 (1), 90.

WHO/FAO, 2009. Report of the 30th Session of the Codex Committe on Nutrition and Food for Special Dietary Uses. ALINORM.

Yang, J., Wang, H.-P., Zhou, L., Xu, C.-F., 2012. Effect of dietary fiber on constipation: a meta analysis. World J. Gastroenterol. 18 (48), 7378–7383.

Yao, B., Fang, H., Xu, W., Yan, Y., Xu, H., Liu, Y., Mo, M., Zhang, H., Zhao, Y., 2014. Dietary fiber intake and risk of type 2 diabetes: a dose-response analysis of prospective studies. Eur. J. Epidemiol. 29 (2), 79–88.

Folate/folic acid

Joshua W. Miller, Department of Nutritional Sciences, School of Environmental and Biological Sciences, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of J.W. Miller, Folic Acid, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 262–269, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00111-2.

Introduction 247
Chemistry and biochemical functions 248
Chemical forms 248
Overview of biochemical functions 248
Source of one-carbon units 250
Removal of one-carbon units 251
Regulation of folate metabolism 251
Enzyme polymorphisms 251
Nutritional aspects 252
Dietary sources 252
Bioavailability 252
Dietary reference intakes 252
Absorption, transport, and excretion 252
Absorption 252
Transport and cellular uptake 253
Excretion 253
Deficiency and excess 253
Megaloblastic anemia 253
Hyperhomocysteinemia 254
Neurological and cognitive dysfunction 254
Cancer 255
Neural tube and other birth defects 255
Folic acid fortification beyond NTDs 255
Biomarkers and status assessment 256
Summary 256
References 257
Further reading 258

Key points

- Folates are a class of B vitamins consisting of multiple interconvertible chemical forms that are essential for one-carbon metabolism-related biochemistry, including DNA and RNA synthesis, amino acid metabolism, methylation reactions, and formate metabolism.
- Consequences of folate deficiency include megaloblastic anemia, hyperhomocysteinemia, cancer initiation, neurological and cognitive impairment, depression, and neural tube defects and other negative birth outcomes.
- Folic acid is the synthetic, oxidized form of folate that is used as a supplement and as a fortificant in foods.
- Folic acid fortification of cereal and grain food products has significantly reduced the incidence of neural tube defects.
- Excess folic acid exposure due to supplementation and fortification may have additional effects on cancer rates, vascular disease, development, and the severity of B12 deficiency, but these effects are controversial and require further investigation.

Introduction

In the 1930s, first Lucy Wills (Wills, 1931) and then Robert Stokstad (Stokstad and Manning, 1938) isolated a natural component of yeast (termed "Wills factor" and "factor U", respectively) that prevented megaloblastic anemia of pregnancy and promoted growth in chickens. In the next decade, Herschell Mitchell, Esmond Snell, and Robert Williams (Mitchell et al., 1941) isolated a factor from spinach that could support the growth of lactic acid bacteria, *Streptococcus faecalis* and *Lactobacillus casei*. They named this factor "folic acid" based on *folium*, the Latin word for leaf. Shortly thereafter, Stokstad (Stokstad, 1943) isolated the pure crystalline form of the

vitamin, and it was recognized that the Wills factor, factor U, and folic acid were the same substance. Beginning in the 1950s and 1960s, the components and details of folic acid metabolism, and the metabolism of its many interconvertible reduced forms collectively known as "folates", were elucidated. These included metabolic interrelationships with vitamin B_{12} and methionine/homocysteine metabolism, roles in pyrimidine and purine synthesis, and the determination of the molecular basis for deficiency diseases. In subsequent decades, key discoveries included identification of polymorphic variants of folate metabolizing enzymes, interrelationships with formate metabolism, and characterization of subcellular compartmentalization of folate metabolism within the cytosol, the mitochondria and, most recently, the nucleus. In the late 1980s and 1990s, clinical interest in folates and folic acid heightened with the identification of elevated plasma homocysteine (hyperhomocysteinemia) as a risk factor for vascular disease, cognitive dysfunction, and dementia. During this time, it was also determined that periconceptional supplements of the vitamin were effective in reducing the risk of neural tube defects (NTDs). This led to wide-spread fortification of cereal and grain products with folic acid in the US and Canada beginning in the mid-to late-1990s. Today, more than 80 countries and territories have instituted folic acid fortification programs, which have been highly effective in reducing NTDs. There are, however, lingering questions regarding the safety of excess folic acid consumption.

Chemistry and biochemical functions

Chemical forms

Though used generically, the term "folic acid" refers specifically to the synthetic form of the vitamin, which is used in supplements and as a fortificant in foods. The term "folate" refers generally to all forms of the vitamin. As shown in **Fig. 1**, folic acid consists of a pterin moiety linked via a methylene group to a para-aminobenzoyl-glutamate moiety. Its metabolic activity requires reduction to the tetrahydrofolate (THF) derivative, addition of a chain of glutamate residues in γ-peptide linkage, and acquisition of one-carbon units. One-carbon units at various levels of oxidation are generated metabolically and are reactive only as moieties attached to the N5 or N10 positions of the folate molecule (**Table 1**). The range of oxidation states for folate one-carbon units extends from methanol to formate as methyl, methylene, methenyl, formyl, or formimino moieties. When one-carbon units are incorporated into folate derivatives, they may be converted from one oxidation state to another by the gain or loss of electrons.

Overview of biochemical functions

The biochemical function of folate substrates is to transfer and use one-carbon units in a variety of essential reactions (**Fig. 2**), including *de novo* purine biosynthesis (formylation of glycinamide ribonucleotide and 5-amino-4-imidazole carboxamide ribonucleotide), pyrimidine nucleotide biosynthesis (methylation of deoxyuridylic acid to thymidylic acid), amino acid interconversions (interconversion of serine to glycine, catabolism of histidine to glutamic acid, and vitamin B_{12}-dependent conversion of homocysteine to methionine), and the generation and use of formate.

Many of the enzymes involved in these reactions are multifunctional and are capable of channeling substrates and one-carbon units from reaction to reaction within a protein matrix. Another feature of intracellular folate metabolism is the compartmentalization of folate coenzymes among the cytosol, the mitochondria, and the nucleus (**Fig. 3**). For instance, 5-methylTHF is associated with the cytosolic fraction of the cell, whereas most of 10-formylTHF is located in the mitochondria. Similarly, some

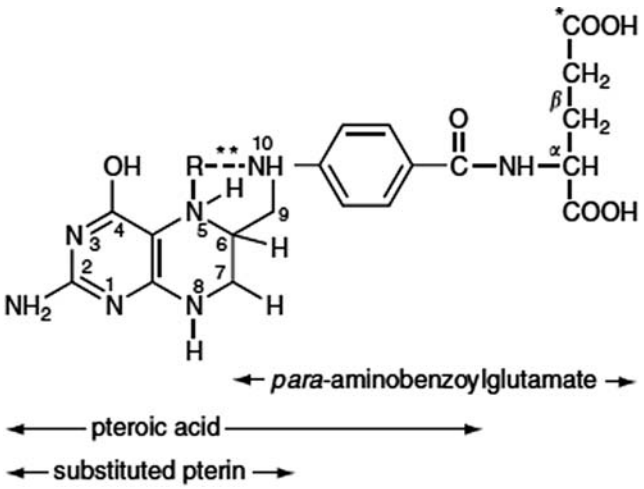


Fig. 1 Structural formula of tetrahydrofolate (THF) compounds. In tetrahydrofolate R = H; other substituents are listed in **Table 1**. The asterisk indicates the site of attachment of extra glutamate residues; the hatched line and double asterisk indicates the N5 and/or N10 site of attachment of one-carbon units.

Table 1 Structure and nomenclature of folate compounds.

Compound	*R-group*	*Oxidation state*
5-formylTHF	—CHO	Formate
10-formylTHF	—CHO	Formate
5-formiminoTHF	—CH=NH	Formate
5,10-methenylTHF	—CH=	Formate
5,10-methyleneTHF	—CH_2—	Formaldehyde
5-methylTHF	—CH_3	Methanol

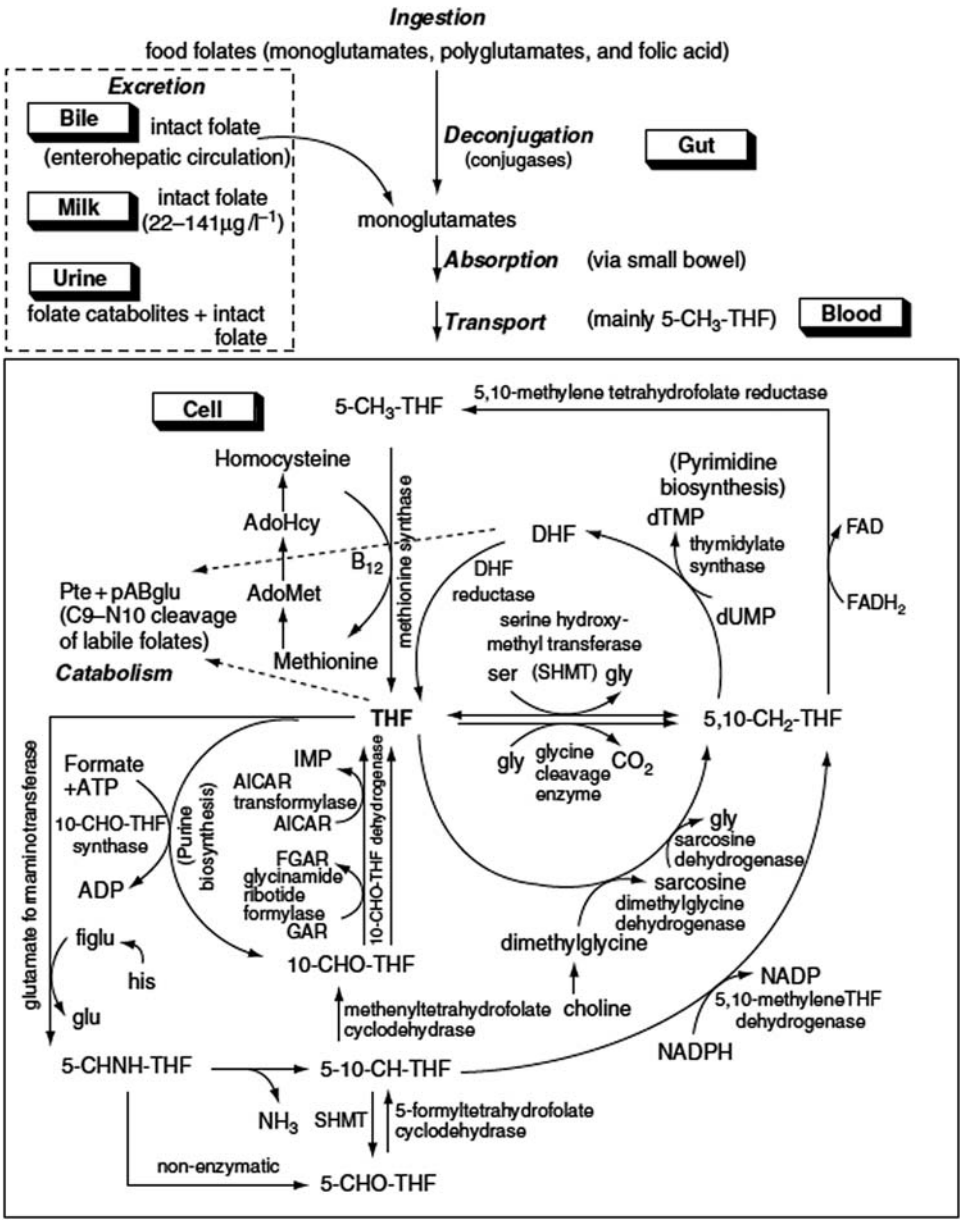


Fig. 2 Physiology and metabolism of folate. GAR, glycinamide ribonucleotide; FGAR, formylglycinamide ribonucleotide; AICAR, aminoimidazolecarboxamide ribonucleotide; FIGLU, formiminoglutamic acid; IMP, inosine monophosphate.

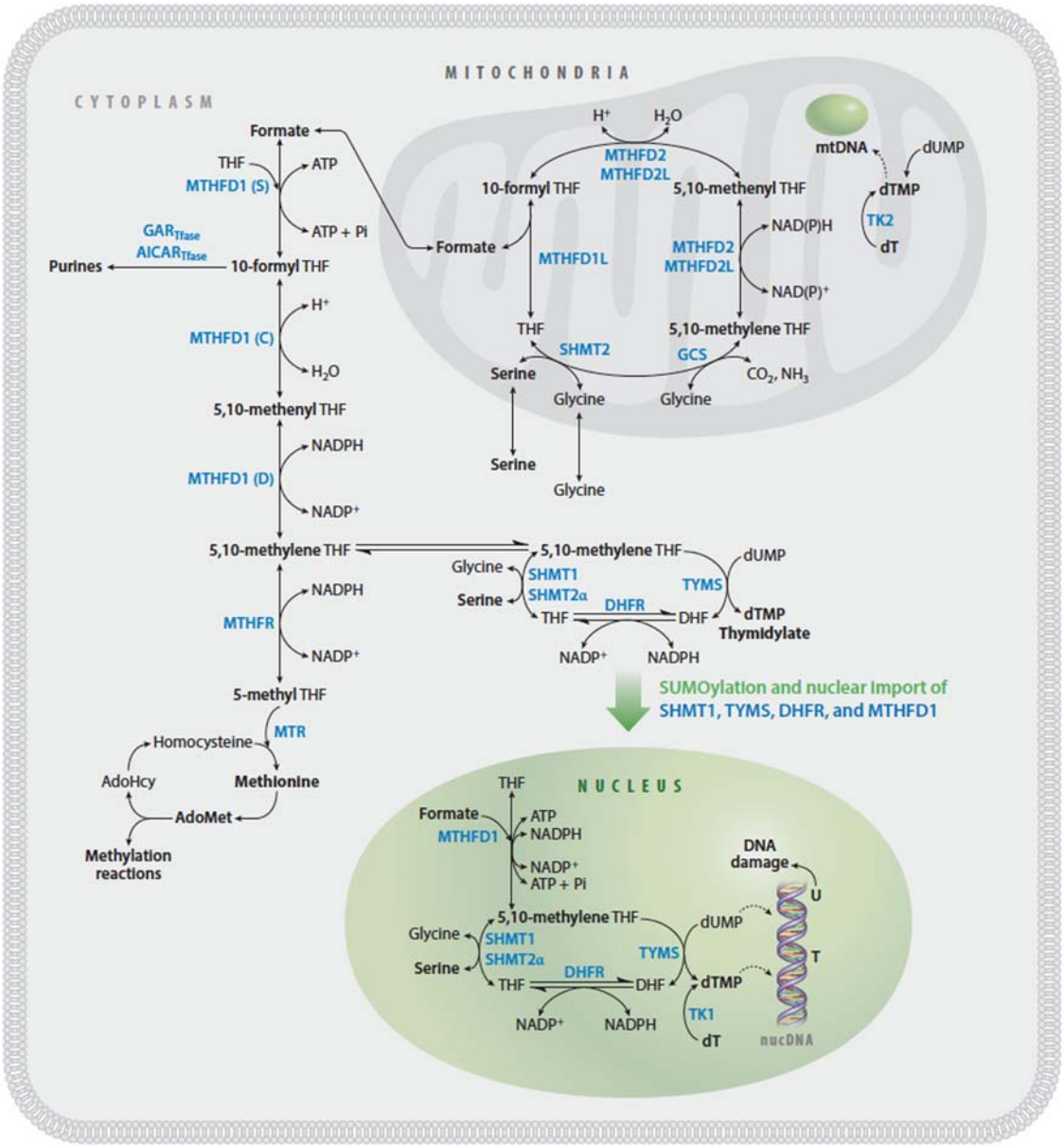


Fig. 3 Subcellular compartmentalization of folate metabolism. AdoHcy, S-adenosylhomocysteine; AdoMet, S-adenosylmethionine; AICARTfase, aminoimidazolecarboxamide ribonucleotide transformylase; DHFR, dihydrofolate reductase; dT, deoxythymidine; dTMP, deoxythymidine monophosphate; dUMP, deoxyuridine monophosphate; GARTfase, glycinamide ribonucleotide tranformylase; MTHFD1, methylenetetrahydrofolate dehydrogenase 1; MTHFR, methylenetetrahydrofolate reductase; MTR, methionine synthase; mtDNA, mitochondrial DNA; SHMT, serine hydroxymethyltransferase; SUMO, small ubiquitin-like modifier; THF, tetrahydrofolate; TK, thymidine kinase; TYMS, thymidylate synthase. Reprinted with permission from Field et al. (2018).

folate-dependent enzymes are associated with one or other compartment, though some are found in both. Metabolic products of folate-dependent reactions, such as serine and glycine, are readily transported between the two locations, but the folate coenzymes are not. In recent years, the existence and importance of nuclear folate metabolism, first suggested by reports from the 1960s and early 1980s, has come to be appreciated. This stems from the discovery that key cytosolic enzymes undergo small-ubiquitin-like modifier-dependent modification (a process called SUMOylation), which facilitates the nuclear import of these enzymes from the cytosol into the nucleus (**Fig. 3**) (Field et al., 2018).

Source of one-carbon units

One-carbon units at the oxidation level of formate can enter directly into the folate pool as formic acid in a reaction catalyzed by 10-formylTHF synthase (EC 6.3.4.3) (**Fig. 2**). Entry at the formate level of oxidation can also take place via a catabolic product of

histidine, formiminoglutamic acid. The third mode of entry at the formate level of oxidation involves the formation of 5-formylTHF from 5,10-methenylTHF by the enzyme serine hydroxymethyltransferase (SHMT) (EC 2.1.2.1). The 5-formylTHF may be rapidly converted to other forms of folate.

The enzyme SHMT is also involved in the entry of one-carbon units at the formaldehyde level of oxidation by catalyzing the transfer of the ß-carbon of serine to form glycine and 5,10-methyleneTHF. Other sources of one-carbon entry at this level of oxidation include the glycine cleavage system and the choline-dependent pathway; both enzyme systems generate 5,10-methyleneTHF in the mitochondria of the cell.

Removal of one-carbon units

One-carbon units are removed from folate by a number of reactions. The enzyme 10-formylTHF dehydrogenase (EC 1.5.1.6) provides a mechanism for disposing of excess one-carbon units as carbon dioxide. (Folate administration to animals enhances the conversion of ingested methanol and formate to carbon dioxide, diminishing methanol toxicity.) Additionally, single-carbon units from 10-formylTHF are used for the biosynthesis of purines (**Fig. 2**).

The one-carbon unit of 5,10-methyleneTHF is transferred in two ways. Reversal of the SHMT reaction produces serine from glycine, but as serine is also produced from glycolysis via phosphoglycerate, this reaction is unlikely to be important. However, one-carbon transfer from 5,10-methyleneTHF to deoxyuridylate to form thymidylic acid, a precursor of deoxyribonucleic acid (DNA), is of crucial importance to the cell. Although the source of the one-carbon unit, namely 5,10-methyleneTHF, is at the formaldehyde level of oxidation, the one-carbon unit transferred to form thymidylic acid appears at the methanol level of oxidation. Electrons for this reduction come from THF itself to generate dihydrofolate as a product. The dihydrofolate must in turn be reduced back to THF in order to accept further one-carbon units.

A solitary transfer of one-carbon units takes place at the methanol level of oxidation. It involves the transfer of the methyl group from 5-methylTHF to homocysteine to form methionine and THF. This reaction is catalyzed by the enzyme methionine synthase (EC 2.1.1.13) and requires vitamin B_{12} as a cofactor. The substance 5-methylTHF is the dominant folate in the body, and it remains metabolically inactive until it is demethylated to THF, whereupon polyglutamylation takes place to allow subsequent folate-dependent reactions to proceed efficiently. Methionine is activated by reaction with ATP to form S-adenosylmethionine (SAM). SAM serves as the universal methyl donor for a variety of essential methylation reactions, including those involving DNA, ribonucleic acid (RNA), histones, proteins, phospholipids, neurotransmitters, and creatine synthesis.

Regulation of folate metabolism

The biochemical functions of folate can be divided into two general categories: (1) the synthesis of bases for incorporation into DNA and RNA, and (2) the metabolism of homocysteine to form methionine and subsequently SAM. Whether folate is utilized for the former or the latter purpose is controlled by biochemical feedback mechanisms that specifically affect the conversion of 5,10-methyleneTHF to 5-methylTHF by the enzyme, 5,10-methyleneTHF reductase (MTHFR) (EC 1.5.1.20). MTHFR is subject to allosteric inhibition by SAM. When intracellular SAM concentrations are low, MTHFR actively promotes the synthesis of 5-methylTHF for methionine and SAM synthesis. When SAM concentrations are high, inhibition of MTHFR by SAM reduces the synthesis of 5-methylTHF, methionine, and SAM, and promotes utilization of 5,10-methyleneTHF for thymidylate and purine synthesis (Selhub and Miller, 1992). Another important inhibitor of MTHFR is dihydrofolate. Accumulation of DHF as a product of thymidylate synthesis will inhibit MTHFR, thus conserving 5,10-methyleneTHF for further thymidylate synthesis. In this way, DHF may serve as a sensor of active DNA synthesis and cellular proliferation (Matthews and Daubner, 1982).

Enzyme polymorphisms

Common polymorphisms have been identified in key enzymes involved in folate metabolism. The most well studied is the C677T single nucleotide polymorphism in the MTHFR gene, which encodes for either alanine (wild type) or valine (variant) at residue 222 within the amino acid sequence (Frosst et al., 1995). The variant form of the enzyme is thermolabile *in vitro* and has approximately 70% lower activity than the wild-type enzyme due to reduced affinity for its substrate (5,10-methyleneTHF) and its cofactor, flavin adenine dinucleotide (FAD) (Kang et al., 1988). Homozygosity for the variant form is found in 10–15% of Caucasians, with lower prevalence in blacks and higher in Hispanics. The lower activity of the variant MTHFR is associated with a variety of conditions, including hyperhomocysteinemia, greater risk of NTDs, and lower risk of some cancers. Hyperhomocysteinemia and NTD risk can be attenuated by increased dietary folate intake or folic acid supplements. Prevalent polymorphisms in several other folate metabolizing enzymes have been identified, including those within thymidylate synthase (EC 2.1.1.45), methionine synthase, methionine synthase reductase (EC 1.16.1.8), reduced folate carrier, and folylpoly-γ-glutamate carboxypeptidase (EC 3.4.17.21), among others. The effects of these polymorphisms on enzyme activity and associated pathophysiological conditions range from none to weak or moderate, depending on the nature of the specific polymorphism.

Nutritional aspects

Dietary sources

Folate is synthesized by microorganisms and higher plants, but not by mammals, for which it is an essential vitamin. The most concentrated food folate sources include liver, yeast extract, green leafy vegetables, legumes, certain fruits (e.g., oranges and strawberries), and fortified cereal and grain products. Folate content is likely to depend on the maturity and variety of particular sources. Prolonged exposure to heat, air, or ultraviolet light is known to inactivate the vitamin; thus, food preparation and cooking can make a difference to the amount of folate ingested. Boiling, in particular, results in substantial food losses. The major source of folate loss from vegetables during boiling may be leaching as opposed to folate degradation. Broccoli and spinach are particularly susceptible to loss through leaching during boiling, compared with potatoes, because of their larger surface areas. The retention of folate during cooking depends on the food in question as well as the method of cooking. Folates of animal origin are stable during cooking by frying or grilling. Steaming in preference to boiling is likely to double the amount of folate consumed from green vegetables. In countries with voluntary or mandatory fortification policies, folic acid from enriched cereal grain products, including bread, cereals, pasta, flour, and rice, may constitute the largest proportion of dietary intake of the vitamin.

Bioavailability

The amount of folate that is absorbed and utilized physiologically varies among different food sources and among different chemical forms of the vitamin. Folic acid consumed as a supplement separate from food is the most highly bioavailable. The bioavailability of folic acid taken with a meal or as a food fortificant is approximately 85%, compared with folic acid consumed while fasting. Natural food folates are less bioavailable, at approximately 50% of the value for folic acid alone. Based on these differences in bioavailability, dietary folate equivalents (DFEs) have been defined as: 1 DFE = 1 μg food folate = 0.6 μg folic acid added to food = 0.5 μg folic acid taken without food (Institute of Medicine, 1998; Suitor and Bailey, 2000).

Dietary reference intakes

The recommended dietary allowance (RDA) for males and nonpregnant and nonlactating females aged 15 years or older is 400 μg DFE day^{-1} (**Table 2**) (Institute of Medicine, 1998). The RDA ranges from 65 to 300 μg DFE day^{-1} for ages 0–14 years. The RDAs for pregnant and lactating women are, respectively, 600 and 500 μg DFE day^{-1}, which accounts for the increased demands for folate of the growing fetus and breast-feeding infant. There is no upper tolerable intake level (UL) established for food folates. However, a UL for folic acid has been set at 1000 μg day^{-1}. This is based not specifically on direct toxic effects of folic acid, but rather the possible masking of vitamin B_{12} deficiency by high dose folic acid, which can correct hematological abnormalities but not the neuropathological manifestations of B_{12} deficiency. In recent years, concerns have been raised that folic acid consumption exceeding the current UL may have other negative consequences related to cancer promotion, exacerbation (over and beyond masking) of the metabolic and clinical features of B_{12} deficiency, impairment of immune function, and altered developmental programming. However, to date these concerns have not been considered in the establishment of the UL for folic acid (see the Section on **Folic acid fortification beyond NTDs**).

Absorption, transport, and excretion

Absorption

Food folates mainly consist of reduced polyglutamates, which are hydrolyzed to monoglutamates in the gut prior to absorption across the intestinal mucosa. The conjugase enzyme that hydrolyzes dietary folates, folylpoly-γ-glutamate carboxypeptidase, is found on the luminal brush border membrane in the human jejunum and has equal affinity for folate polyglutamates of various chain lengths. Uptake of folate monoglutamate into the intestinal cell is facilitated by two saturable carrier-mediated processes.

Table 2 Recommended dietary allowances (RDA) for folate (US and Canada).

Category	*Age*	*RDA (μg/day)*
Infants	0–6 months	65
	6–12 months	80
	1–3 years	150
Children	4–6 years	200
	7–14 years	300
Adults	15+ years	400
Pregnancy		600
Lactation		500

These include the reduced folate carrier (RFC), a transporter protein that mediates the uptake of food folates, but has low affinity for folic acid, and folate binding protein (FBP), which mediates the uptake of both reduced and oxidized folates by receptor-mediated endocytosis. The luminal pH optimum for intestinal folate uptake is ~5.0–6.0, and changes in luminal pH, as well as the presence of conjugase inhibitors, folate binders, or other food components can adversely affect the rate of hydrolysis and intestinal absorption. Such factors account for the wide variation in the bioavailability of the vitamin from foods of plant and animal origins. Some metabolism of the resultant monoglutamate, mainly to 5-methylTHF, appears to occur during the absorption process, though this may not be necessary for transport across the basolateral membrane of the intestinal mucosa into the portal circulation. The degree of metabolic conversion of supplemental folic acid depends on the dose; significant portions of pharmacological amounts are transported unaltered into the circulation.

Transport and cellular uptake

Folate circulates in the blood predominantly as 5-methylTHF. A variable proportion circulates freely or is bound either to low-affinity protein binders such as albumin, which account for about 50% of bound folate, or to a high-affinity folate binder in serum, which carries less than 5% of circulating folate. The physiological importance of serum binders is unclear, but they may control folate distribution and excretion during deficiency.

Though most folate is initially taken up by the liver following absorption, it is delivered to a wide variety of tissues in which many types of folate transporters have been described. Because these transporters have affinities for folate in the micromolar range, they would not be saturated by normal ambient concentrations of folate. Therefore, folate uptake into tissues should be responsive to any increases in serum folate concentration arising from folate supplementation. An important determinant of folate uptake into cells is their mitotic activity, as would be expected given the dependence of DNA biosynthesis on folate coenzyme function. Folate accumulation is more rapid in actively dividing cells than in quiescent cells, a factor that is probably related to the induction and activity of folylpoly-γ-glutamate synthase (EC 6.3.2.17). This enzyme catalyzes the addition of glutamate by γ-peptide linkage to the initial glutamate moiety of the folate molecule. Although polyglutamate derivatization may be considered a storage strategy, this elongation is the most efficient coenzyme form for normal one-carbon metabolism. The activity of folylpoly-γ-glutamate synthase is highest in the liver, the folate stores of which account for half of the estimated 15–30 mg adult total body complement. Retention within the cell is facilitated by the high proportion of folate associated with proteins, and this is likely to be increased in folate deficiency.

The mobilization of liver and other stores of folate in the body is not well understood, particularly in deficiency states, though some accounts describe poor turnover rates in folate-depleted rats. Transport across cell membranes during redistribution requires deconjugation of the large negatively charged polyglutamates. Mammalian γ-glutamylhydrolases that hydrolyze glutamate moieties and transpeptidases that hydrolyze folylpolyglutamates directly to mono- or di-glutamate forms of the vitamin have been described for several tissues. Thus, mammalian cells possess two types of enzymes that can play a key role in folate homeostasis and regulation of one-carbon metabolism: folylpolyglutamate synthetase that catalyzes the synthesis of retentive and active folate, and a number of deconjugating enzymes that promote the release of folate from the cell. Polyglutamate forms released into the circulation either through cell death or by a possible exocytotic mechanism would be hydrolyzed rapidly by plasma γ-glutamyl-hydrolase to the monoglutamate form.

Excretion

Folate is concentrated in bile and enterohepatic recirculation from the intestine accounts for considerable reabsorption and reuse of folate (about 100 μg day^{-1}). Fecal folates arise mainly through biosynthesis of the vitamin by the gut microflora, with only a small contribution from unabsorbed dietary folate. Urinary excretion of intact folates accounts for only a small fraction (1–2%) of ingested folate under normal physiological conditions. Free folates (i.e., non-protein bound) are filtered in the glomerulus and reabsorbed in the proximal tubules, which contain a high concentration of FBP. The greater amount of excretion in urine is accounted for by products that arise from cleavage of the folate molecule at the C9-N10 bond, consisting of one or more pteridines and *p*-acetamido-benzoylglutamate. The rate of scission of the folate molecule increases during rapid-mitotic conditions such as pregnancy and growth. Scission of folate is perhaps the major mechanism of folate turnover in the body.

Deficiency and excess

Megaloblastic anemia

Megaloblastic anemia (a form of macrocytic anemia) is characterized by larger than normal circulating red blood cells and hypersegmented neutrophils. Deficiencies of either folate or vitamin B_{12} induce anemias that are clinically indistinguishable (Chanarin, 1979). The hematological effect in both cases is the result of intracellular concentration of 5,10-methyleneTHF that is inadequate to sustain thymidylate synthesis. In the case of folate deficiency, this is directly the result of a lack of dietary folate. In vitamin B_{12} deficiency, this is due to what is known as the "methylfolate trap" (Scott and Weir, 1981). This occurs because the conversion of 5,10-methyleneTHF to 5-methylTHF by MTHFR is an irreversible reaction. When vitamin B_{12} is deficient, the utilization of 5-methylTHF for methionine synthesis is inhibited. Consequently, the metabolism of 5-methylTHF cannot proceed forward or backward, and

thus the 5-methylTHF becomes metabolically trapped and a secondary folate deficiency occurs. Regardless of the cause of the folate deficiency, the resulting inhibition of DNA synthesis affects rapidly proliferating cells, in particular the blood cell precursors in the bone marrow. Inhibition of DNA synthesis leads to a block in cell replication, but not in cytoplasmic protein synthesis, which results ultimately in the release of megaloblastic or macrocytic red cells into the circulation. Other rapidly proliferating cells that are similarly affected by folate deficiency are the mucosal cells of the intestine. This leads to blunted intestinal villi and consequent reduced nutrient absorptive capacity.

It is important for the treatment of megaloblastic anemia to differentiate between folate and B_{12} deficiencies as the cause in any given patient. High dose folic acid supplements correct hematological abnormalities, but not the neuropathological manifestations of B_{12} deficiency. This occurs because folic acid is taken up into cells and is metabolized first to dihydrofolate and then THF by the enzyme, dihydrofolate reductase (EC 1.5.1.3) (**Fig. 2**). The THF can then be converted to 5,10-methyleneTHF and used for thymidylate synthesis. In this way, the folic acid bypasses the methylfolate trap, reverses the megaloblastic anemia, and essentially "masks" the B_{12} deficiency. The patient nonetheless remains B_{12} deficient and is susceptible to potentially irreversible neuronal damage if the B_{12} deficiency continues untreated. Notably, there are clinical observations that have been interpreted to be evidence that folic acid supplementation in this context may actually exacerbate or accelerate the clinical manifestations of B_{12} deficiency, particularly the neuropathology and its consequences, but this is controversial (Reynolds, 2002).

Hyperhomocysteinemia

An important consequence of folate deficiency is the inability to remethylate homocysteine to form methionine (**Fig. 2**). Indeed, there is an inverse correlation between the concentration of folate and that of homocysteine in the blood (Jacques et al., 1999). Many clinical studies, beginning with the observations of children with homocystinuria presenting with vascular abnormalities and thromboembolism (McCully, 1969), have demonstrated an association between hyperhomocysteinemia and increased risk of premature atherosclerosis in the coronary, carotid, and peripheral vasculatures (Refsum et al., 1998). Even mild hyperhomocysteinemia is recognized to be an independent risk factor for cardiovascular disease. The risk of heart disease increases proportionately in most, but not all, studies, throughout the full of range of blood homocysteine concentration. An increase in plasma homocysteine of 5 $\mu mol\ L^{-1}$ has been associated with a combined odds ratio of 1.3 for cardiovascular disease (Boushey et al., 1995).

Metabolically, homocysteine may be disposed of by the methionine synthase reaction (dependent on folate and vitamin B_{12}), the transsulfuration pathway (dependent on vitamin B_6), or the choline degradation pathway (independent of folate, B_{12}, and B_6) (Selhub and Miller, 1992). Deficiencies of each of these three vitamins and choline are associated with hyperhomocysteinemia. Of the three vitamins, folate has been shown to be most effective in lowering homocysteine concentration in the blood. Evidence for the potential role of folate intake in the prevention of vascular disease has come from an observed significant inverse relationship between serum folate concentration and fatal coronary heart disease. In addition to homocysteine-lowering, proposed mechanisms by which folate may affect vascular disease risk include antioxidant actions and interactions with the enzyme, endothelial nitric oxide synthase (EC 1.14.14.47) (Psara et al., 2020). It is important to note that folic acid supplements, with or without concomitant vitamin B_{12} and B_6 supplements, though effective in lowering plasma homocysteine, have not been proven to be particularly effective in reducing overall vascular disease risk (Clarke et al., 2010). There is some evidence, however, that homocysteine-lowering is effective in preventing stroke (Huo et al., 2015).

Neurological and cognitive dysfunction

Folate deficiency is associated with a variety of neurological consequences, including neuropsychiatric manifestations (insomnia, irritability, fatigue, and forgetfulness), peripheral neuropathies (decreased or absent reflexes, diminished vibration sense, and restless leg syndrome), optic neuropathy, and progressive vision loss. Folate deficiency has also been associated with cognitive impairment and dementia (Clarke et al., 1998). Both cross-sectional and longitudinal studies have verified that low folate intake or status is a risk factor for poor scores on objective measures of cognitive function and Alzheimer's disease. In addition, some, but not all studies have demonstrated a beneficial effect of folic acid supplements in preventing cognitive decline in older adults (Clarke et al., 2014). The benefit of folic acid supplements may be limited to the period prior to the development or the early stages of cognitive impairment (before significant structural damage and dementia have occurred). Folate deficiency may also contribute to altered mood and depression, and may limit the efficacy of anti-depressant medications (Alpert and Fava, 1997).

Several potential mechanisms have been postulated to explain the effects of folate deficiency on neurological function (Smith and Refsum, 2016). Most attention has focused on hyperhomocysteinemia which, in addition to cardiovascular disease, is associated with atherosclerotic and thrombotic damage that acutely (i.e., stroke and infarction) or chronically limits blood flow to the brain. Homocysteine or products of its oxidative metabolism, including homocysteine sulfinic acid and homocysteic acid, may also induce excitotoxicity or oxidative stress within the brain, with resultant neurodegeneration. Alternatively, hyperhomocysteinemia may be a marker of altered intracellular concentrations of SAM and S-adenosylhomocysteine (SAH). Reduced SAM and increased SAH induced by folate deficiency may inhibit key methylation reactions in the central nervous system involving proteins, neurotransmitters, phospholipids, and DNA. Inhibition of these processes may in turn cause metabolic impairments and structural damage. In addition, depressive symptoms associated with folate deficiency may result from altered SAM concentration in the brain. Oral SAM supplements have been shown to have anti-depressant properties and folate deficiency may induce depressive symptoms

by reducing SAM in the brain. This may also be relevant to cognition because depressive symptoms are a strong determinant of cognitive impairment in older adults.

Cancer

Folate has a complex relationship with cancer. Folate deficiency limits the synthesis of thymidylic acid and causes the accumulation of uridylic acid. This leads to misincorporation of uracil into DNA. Cellular repair mechanisms efficiently excise the misincorporated uracil. However, because thymidylic acid availability is reduced due to the folate deficiency, DNA strand breaks occur, which leaves DNA vulnerable to mutation. In this way, folate deficiency is a risk factor for the initiation of cancer, while folate sufficiency is protective. In contrast, proliferating cancer cells, like all mammalian cells, require folate for replication and proliferation. When folate is deficient, replication and proliferation are inhibited. Thus, after cancer is initiated, folate deficiency will actually retard its progression. This phenomenon, first explored by Sidney Farber in the treatment of pediatric acute lymphoblastic leukemia (Farber and Diamond, 1948), is the basis for several *anti*-folate cancer chemotherapeutic drugs, including methotrexate (an inhibitor of dihydrofolate reductase) and 5-fluorouracil (an inhibitor of thymidylate synthase), among others.

Neural tube and other birth defects

In early vertebral development, the neural tube forms from the invagination of neural crest cells, which subsequently differentiate into the brain and spinal column. NTDs are malformations in which there is failure of the neural tube to close properly during the fourth week of embryonic life. Incomplete closure of the spinal cord results in spina bifida, while incomplete closure of the cranium results in anencephaly. Observations in both humans and animals in the 1960s and 1970s suggested that poor nutritional status and specifically low folate status was a cause of NTDs. In the 1980s and early 1990s, randomized controlled trials demonstrated that folic acid supplements prevent the recurrence and occurrence of NTDs, which led to government-mandated fortification of cereal and grain products with folic acid in the US and Canada in 1998 (Miller, 2011; Centers for Disease Control and Prevention, 1992; Food and Drug Administration, 1996). Today, more than 80 countries and territories have instituted mandatory folic acid fortification programs. Assessments of the efficacy of folic acid fortification in the US, Canada, Chile, Costa Rica, and South Africa, among others, show that NTD incidence has decreased between 19 and 78%, depending on the rate of NTDs and folate status within a population prior to the institution of fortification (Heseker et al., 2009). These data indicate that folic acid fortification, for its intended purpose, has been one of the most highly successful public health interventions ever devised.

International agencies have published folic acid recommendations for the prevention of NTDs. To prevent recurrence, 4 mg folic acid day^{-1} is recommended, while 400 μg day^{-1} is recommended to prevent occurrence. Because women do not usually become aware that they are pregnant until after neural tube closure has occurred in the developing fetus, it is essential that folic acid supplements be commenced prior to conception. For informed women and their doctors, this is typically achieved through folic acid supplements. However, for the general population, in which a high proportion of pregnancies are unplanned, fortification is the more effective strategy for prevention. The level of fortification mandated in the US is 140 μg of folic acid per 100 g of flour, calculated to increase individual consumption of folic acid by 100 μg day^{-1} and to bring overall folate intake to the recommended 400 μg day^{-1}. Measurement of the actual folic acid in fortified foods indicates that manufacturers have added an amount closer to 200 μg folic acid per 100 g flour, presumably to provide a margin of error in achieving the mandated amount (Choumenkovitch et al., 2002).

Folate deficiency and hyperhomocysteinemia have also been associated with other abnormal pregnancy outcomes. These include premature delivery, low birth weight, fetal growth retardation, *abruptio placentae* (placental infarction), preeclampsia, and congenital heart defects. Randomized controlled clinical trials are needed to determine if folic acid supplements can reduce the incidence of these abnormal outcomes.

Folic acid fortification beyond NTDs

Because of the essential role of folate in one-carbon metabolism, other effects of folic acid fortification have been considered, including the possibility that excess intake may have negative consequences. With respect to cancer, it is postulated that high folic acid intake at the population level could have both preventive and promoting effects. Folic acid supplements increase the grade and multiplicity of colorectal adenomas in patients who previously had colonic polyps removed (Cole et al., 2007). In addition, ecological studies suggest that temporary increases in the incidence of colorectal cancer occurred in the US and Chile concurrent with the institution of folic acid fortification (Mason et al., 2007; Hirsch et al., 2009). In contrast, a reduction in the rate of pediatric neuroblastoma was observed after the start of fortification (French et al., 2003). With respect to breast cancer, conflicting studies have found that risk may be decreased or increased with folic acid supplements (Stolzenberg-Solomon et al., 2006; Maruti et al., 2009). It is possible that these contradictory findings are explained by the timing of folic acid exposure. If exposure occurs before neoplastic initiation, folic acid may reduce the risk of cancer. However, if exposure occurs after initiation, then folic acid may promote proliferation and accelerate progression to clinical disease. Notably, a meta-analysis of folic acid intervention trials including more than 49,000 participants found that the relative risk (95% confidence interval) for cancer of any kind was 1.06 (0.99, 1.13) ($p = 0.10$) indicating that there was a nonsignificant estimated 6% increase in cancer risk associated with folic acid

supplementation ($\geq$0.5 mg day^{-1}; mean follow up time of 5.2 years) (Vollset et al., 2013). This suggests that, at the population level, the risk associated with folic acid supplementation is relatively small.

One of the positive consequences of folic acid fortification has been a significant reduction in the prevalence of hyperhomocysteinemia in the general population. It is unclear if this has translated into reduced risk of cardiovascular disease. Meta-analysis of homocysteine-reduction intervention trials suggests that there is limited or no benefit on overall cardiovascular disease risk (Clarke et al., 2010). However, subgroup analyses in some trials found significant reductions in stroke risk associated with homocysteine-lowering interventions, particularly for individuals with specific characteristics, such as age less than 69 years, those with elevated homocysteine or cholesterol, those not taking anti-platelet or lipid-lowering drugs, and those not exposed to folic acid fortification (Smith and Refsum, 2021). In addition, reduced rates of stroke were observed in the US and Canada around the time fortification was initiated, but not in England and Wales, where fortification had not been initiated (Yang et al., 2006). These latter studies were ecological in design, however, and the strength of this type of evidence is limited. Taken together, whether homocysteine-lowering with folic acid supplementation or fortification reduces cardiovascular disease risk may depend on the characteristics of the individual and may be limited to stroke and not other manifestations of cardiovascular disease (Smith and Refsum, 2021).

One of the primary concerns with folic acid fortification was the possible masking of vitamin B_{12} deficiency, as described above. Further, observational studies suggest that excess folic acid intake, particularly in folic acid supplement users also exposed to folic acid fortification, may actually exacerbate both the neurological and hematological consequences of vitamin B_{12} deficiency (Morris et al., 2007). Older adults with a combination of low serum vitamin B_{12} and high serum folate concentrations were found to be at higher risk of cognitive impairment and anemia than adults of similar age who had both low vitamin B_{12} and low folate. The combination of low vitamin B_{12} and high folate was also associated with higher circulating homocysteine and methylmalonic acid and lower holotranscobalamin concentrations (metabolic indicators of functional vitamin B_{12} deficiency) than the combination of low vitamin B_{12} and low folate (Selhub et al., 2007; Miller et al., 2009). These findings are consistent with clinical observations from several decades ago, where patients with pernicious anemia (severe vitamin B_{12} deficiency caused by autoimmune-induced vitamin B_{12} malabsorption) exhibited further reduction in circulating vitamin B_{12} concentration after being treated with high-dose folic acid (Lear and Castle, 1956; Bok et al., 1958). Recently, a hypothesis to explain this putative effect of folic acid on serum biomarkers of vitamin B_{12} status has been put forward. The hypothesis states that in individuals with already low vitamin B12 status, excess folic acid causes increased mobilization of circulating holotranscobalamin to the bone marrow to support reticulocyte formation and/or increased excretion of holotranscobalamin in the urine. This reduces the availability of B12 to other organs and, thus, exacerbates manifestations of vitamin B12 deficiency, including elevations of homocysteine and methylmalonic acid in the blood and neurological outcomes (Selhub et al., 2022). This hypothesis has not yet been systematically tested, and the conclusion that excess folic acid exacerbates vitamin B_{12} deficiency remains controversial.

Biomarkers and status assessment

The main clinical indicator of folate deficiency is megaloblastic anemia, as described above. The primary screening measurements for assessing folate status are serum and red blood cell folate concentrations. The gold-standard for folate measurement is a microbiological assay utilizing *L. casei*. Radioactive competitive binding and automated chemiluminescence assays are commonly used for research and clinical purposes, respectively. Mass spectrometry methods have also been devised. Though there is usually a high correlation between serum and red cell folate concentrations, serum folate typically reflects recent folate intake and short-term status, while red cell folate reflects long-term status, over the 120 days lifespan of the red cell. Common cut-off values used to define folate deficiency are $<$3 ng mL^{-1} ($<$7 nmol L^{-1}) in serum and $<$140 ng mL^{-1} ($<$305 nmol L^{-1}) in red cells. Elevated plasma homocysteine ($>$10 μmol L^{-1}) is another indicator of folate deficiency. However, hyperhomocysteinemia has several potential causes, including vitamin B_{12} and vitamin B_6 deficiencies, renal disease, and hypothyroidism; thus homocysteine is a sensitive, but not specific, indicator of folate status. Another biomarker sensitive to low folate status is global DNA methylation, as a consequence of the role of folate in the synthesis of SAM used for DNA methylation reactions.

Summary

Folates, through their participation in one-carbon metabolism, play key biochemical roles in the synthesis of nucleotides for DNA and RNA synthesis, amino acid interconversions (i.e., serine to glycine and homocysteine to methionine), production of the universal methyl donor SAM, and formate metabolism. Predictably, folate deficiency leads to impairment of one-carbon metabolism, which underlies a variety of pathophysiological outcomes, including: megaloblastic anemia due to impaired nucleotide synthesis; increased risk of vascular disease related to hyperhomocysteinemia (particularly stroke); increased risk of cancer resulting from uracil misincorporation, DNA strand breaks, and increased susceptibility to mutation; increased risk of neurological/cognitive deficits and depression possibly related to hyperhomocysteinemia, oxidative stress, or impaired methylation reactions in the brain; and neural tube and other birth defects. Folic acid fortification in more than 80 countries around the world has significantly reduced the incidence of neural tube defects and represents one of the most successful public health interventions for its intended purpose ever implemented. In addition, to preventing neural tube defects, folic acid fortification may reduce risk of cardiovascular disease

and cancer. However, there is lingering concern that excess consumption of folic acid may exacerbate biochemical and pathophysiological manifestations of B12 deficiency. Additional basic and translational research is needed to address this issue.

References

Alpert, J.E., Fava, M., 1997. Nutrition and depression: the role of folate. Nutr. Rev. 55, 145–149.

Bok, J., Nieweg, H.O., Kroese, W.F., Faber, J.G., De Vries, A., 1958. Vitamin B12 content of blood during folic acid therapy of untreated patients with pernicious anemia. Ned. Tijdschr. Geneeskd. 102, 795–798.

Boushey, C.J., Beresford, S.A.A., Omenn, G.S., Motulsky, A.G., 1995. A quantitative assessment of plasma homocysteine as a risk factor for vascular disease: probable benefits of increasing folic acid intakes. JAMA 274, 1049–1057.

Centers for Disease Control and Prevention, 1992. Recommendations for the use of folic acid to reduce the number of cases of spina bifida and other neural tube defects. Morb. Mortal. Wkly. Rep. 41 (No. RR-14), 1–7.

Chanarin, I., 1979. The Megaloblastic Anaemias, second ed. Blackwell Scientific Publications, Oxford.

Choumenkovitch, S.F., Selhub, J., Wilson, P.W., Rader, J.I., Rosenberg, I.H., Jacques, P.F., 2002. Folic acid intake from fortification in United States exceeds predictions. J. Nutr. 132, 2792–2798.

Clarke, R., Bennett, D., Parish, S., et al., 2014. Effects of homocysteine lowering with B vitamins on cognitive aging: meta-analysis of 11 trials with cognitive data on 22,000 individuals. Am. J. Clin. Nutr. 100, 657–666.

Clarke, R., Halsey, J., Lewington, S., et al., 2010. Effects of lowering homocysteine levels with B vitamins on cardiovascular disease, cancer, and cause-specific mortality: meta-analysis of 8 randomized trials involving 37,485 individuals. Arch. Intern. Med. 170, 1622–1631.

Clarke, R., Smith, A.D., Jobst, K.A., et al., 1998. Folate, vitamin B12, and serum total homocysteine levels in confirmed Alzheimer disease. Arch. Neurol. 55, 1449–1455.

Cole, B.F., Baron, J.A., Sandler, R.S., et al., 2007. Folic acid for the prevention of colorectal adenomas: a randomized clinical trial. JAMA 297, 2351–2359.

Farber, S., Diamond, L.K., 1948. Temporary remissions in acute leukemia in children produced by folic acid antagonist, 4-aminopteroyl-glutamic acid. N. Engl. J. Med. 238, 787–793.

Field, M.S., Kamynina, E., Chon, J., Stover, P.J., 2018. Nuclear folate metabolism. Annu. Rev. Nutr. 38, 219–243.

Food and Drug Administration, 1996. Food labeling: health claims and label statements; folate and neural tube defects. Fed. Regist. 61, 8752–8781.

French, A.E., Grant, R., Weitzman, S., Ray, J.G., Vermeulen, M.J., Sung, L., Greenberg, M., Koren, G., 2003. Folic acid food fortification is associated with a decline in neuroblastoma. Clin. Pharmacol. Therapeut. 74, 288–294.

Frosst, P., Blom, H.J., Milos, R., et al., 1995. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat. Genet. 10, 111–113.

Heseker, H.B., Mason, J.B., Selhub, J., Rosenberg, I.H., Jacques, P.F., 2009. Not all cases of neural tube defects can be prevented by increasing the intake of folic acid. Br. J. Nutr. 102, 173–180.

Hirsch, S., Sanchez, H., Albala, C., de la Maza, M.P., Barrera, G., Leiva, L., Bunout, D., 2009. Colon cancer in Chile before and after the start of the flour fortification program with folic acid. Eur. J. Gastroenterol. Hepatol. 21, 436–439.

Huo, Y., Li, J., Qin, X., et al., 2015. Efficacy of folic acid therapy in primary prevention of stroke among adults with hypertension in China: the CSPPT randomized clinical trial. JAMA 313, 1325–1335.

Institute of Medicine, 1998. Folate. In: Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline. National Academies Press, Washington, DC, pp. 196–305.

Jacques, P.F., Selhub, J., Bostom, A.G., Wilson, P.W., Rosenberg, I.H., 1999. The effect of folic acid fortification on plasma folate and total homocysteine concentrations. N. Engl. J. Med. 340, 1449–1454.

Kang, S.S., Wong, P.W., Zhou, J.M., Sora, J., Lessick, M., Ruggie, N., Grcevich, G., 1988. Thermolabile methylenetetrahydrofolate reductase in patients with coronary artery disease. Metabolism 37, 611–613.

Lear, A.A., Castle, W.B., 1956. Supplemental folic acid therapy in pernicious anemia: the effect on erythropoiesis and serum vitamin B12 concentrations in selected cases. J. Lab. Clin. Med. 47, 88–97.

Mason, J.B., Dickstein, A., Jacques, P.F., Haggarty, P., Selhub, J., Dallal, G., Rosenberg, I.H., 2007. A temporal association between folic acid fortification and an increase in colorectal cancer rates may be illuminating important biological principles: a hypothesis. Cancer Epidemiol. Biomarkers Prev. 16, 1325–1329.

Matthews, R.G., Daubner, S.C., 1982. Modulation of methylenetetrahydrofolate reductase activity by S-adenosylmethionine and by dihydrofolate and its polyglutamate analogues. Adv. Enzym. Regul. 20, 123–131.

McCully, K., 1969. Vascular pathology of homocysteinemia: implications for the pathogenesis of arteriosclerosis. Am. J. Pathol. 56, 111–128.

Maruti, S.S., Ulrich, C.M., White, E., 2009. Folate and one-carbon metabolism nutrients from supplements and diet in relation to breast cancer risk. Am. J. Clin. Nutr. 89, 624–633.

Miller, J.W., 2011. Folic acid fortification. In: Herrmann, W., Obeid, R. (Eds.), Vitamins in the Prevention of Human Diseases. De Gruyter, Berlin, Germany, pp. 273–293.

Miller, J.W., Garrod, M.G., Allen, L.H., Haan, M.N., Green, R., 2009. Metabolic evidence of vitamin B12 deficiency, including high homocysteine and methylmalonic acid and low holotranscobalamin, is more pronounced in older adults with elevated plasma folate. Am. J. Clin. Nutr. 90, 1586–1592.

Mitchell, H.K., Snell, E.E., Williams, R.J., 1941. The concentration of "folic acid". J. Am. Chem. Soc. 63, 2284.

Morris, M.S., Jacques, P.F., Rosenberg, I.H., Selhub, J., 2007. Folate and vitamin B-12 status in relation to anemia, macrocytosis, and cognitive impairment in older Americans in the age of folic acid fortification. Am. J. Clin. Nutr. 85, 193–200.

Psara, E., Pentieva, K., Ward, M., McNulty, H., 2020. Critical review of nutrition, blood pressure and risk of hypertension through the lifecycle: do B vitamins play a role? Biochimie 173, 76–90.

Refsum, H., Ueland, P.M., Nygård, O., Vollset, S.E., 1998. Homocysteine and cardiovascular disease. Annu. Rev. Med. 49, 31–62.

Reynolds, E., 2002. Benefits and risks of folic acid to the nervous system. J. Neurol. Neurosurg. Psychiatr. 72, 567–571.

Scott, J.M., Weir, D.G., 1981. The methyl folate trap: a physiological response in man to prevent methyl group deficiency in kwashiorkor (methionine deficiency) and an explanation for folic-acid-induced exacerbation of subacute combined degeneration in pernicious anaemia. Lancet 318, 337–340.

Selhub, J., Miller, J.W., 1992. The pathogenesis of homocysteinemia: interruption of the coordinate regulation by S-adenosylmethionine of the remethylation and transsulfuration of homocysteine. Am. J. Clin. Nutr. 55, 131–138.

Selhub, J., Miller, J.W., Troen, A.M., Mason, J.B., Jacques, P.F., 2022. The high folate/low vitamin B-12 interaction is a novel cause of vitamin B-12 depletion with a specific etiology: a hypothesis. Adv. Nutr. 13, 16–33.

Selhub, J., Morris, M.S., Jacques, P.F., 2007. In vitamin B12 deficiency, higher serum folate is associated with increased total homocysteine and methylmalonic acid concentrations. Proc. Natl. Acad. Sci. U. S. A. 104, 19995–20000.

Smith, A.D., Refsum, H., 2016. Homocysteine, B vitamins, and cognitive impairment. Annu. Rev. Nutr. 36, 211–239.

Smith, A.D., Refsum, H., 2021. Homocysteine – from disease marker to disease prevention. J. Intern. Med. https://doi.org/10.1111/joim.13279.

Stokstad, E.L.R., 1943. Some properties of a growth factor for *Lactobacillus casei*. J. Biol. Chem. 149, 573–574.

Stokstad, E.L.R., Manning, P.D.V., 1938. Evidence of a new growth factor required by chicks. J. Biol. Chem. 125, 687–696.
Stolzenberg-Solomon, R.Z., Chang, S.C., Leitzmann, M.F., Johnson, K.A., Johnson, C., Buys, S.S., Hoover, R.N., Ziegler, R.G., 2006. Folate intake, alcohol use, and post-menopausal breast cancer risk in the prostate, lung, colorectal, and ovarian cancer screening trial. Am. J. Clin. Nutr. 83, 895–904.
Suitor, C.W., Bailey, L.B., 2000. Dietary folate equivalents: interpretation and application. J. Am. Diet Assoc. 100, 88–94.
Vollset, S.E., Clarke, R., Lewington, S., et al., 2013. Effects of folic acid supplementation on overall and site-specific cancer incidence during the randomised trials: meta-analyses of data on 50,000 individuals. Lancet 381, 1029–1036.
Wills, L., 1931. Treatment of "pernicious anaemia of pregnancy" and "tropical anaemia" with special reference to yeast extract as a curative agent. Br. Med. J. 1, 1059–1064.
Yang, Q., Botto, L.D., Erickson, J.D., Berry, R.J., Sambell, C., Johansen, H., Friedman, J.M., 2006. Improvement in stroke mortality in Canada and the United States, 1990 to 2002. Circulation 13, 1335–1343.

Further reading

Bailey, L.B., 2007. Folic acid. In: Zempleni, J., Rucker, R.B., McCormick, D.B., Suttie, J.W. (Eds.), Handbook of Vitamins, fourth ed. CRC Press, Boca Raton, Florida, pp. 385–412.
Bailey, L.B. (Ed.), 2010. Folate in Health and Disease, second ed. CRC Press, Boca Raton, Florida.
Duthie, S.J., 1999. Folic acid deficiency and cancer: mechanisms of DNA instability. Br. Med. Bull. 55, 578–592.
Homocysteine Lowering Trialists' Collaboration, 1998. Lowering blood homocysteine with folic acid based supplements: meta-analysis of randomised trials. BMJ 316, 894–898.
Lamarre, S.G., Morrow, G., Macmillan, L., Brosnan, M.E., Brosnan, J.T., 2013. Formate: an essential metabolite, a biomarker, or more? Clin. Chem. Lab. Med. 51, 571–578.
Reynolds, E.H., 2002. Folic acid, ageing, depression, and dementia. BMJ 324, 1512–1515.
Selhub, J., Jacques, P.F., Wilson, P.W.F., Rush, D., Rosenberg, I.H., 1993. Vitamin status and intake as primary determinants of homocysteinemia in an elderly population. JAMA 270, 2693–2698.
Seshadri, S., Beiser, A., Selhub, J., Jacques, P.F., Rosenberg, I.H., D'Agostino, R.B., Wilson, P.W., Wolf, P.A., 2002. Plasma homocysteine as a risk factor for dementia and Alzheimer's disease. N. Engl. J. Med. 346, 476–483.
Smith, A.D., Kim, Y.I., Refsum, H., 2008. Is folic acid good for everyone? Am. J. Clin. Nutr. 87, 517–533.

Homocysteine

Joshua W. Miller, Department of Nutritional Sciences, School of Environmental and Biological Sciences, Rutgers, The State University of New Jersey New Brunswick, NJ, USA

© 2023 Elsevier Ltd. All rights reserved.

This is an update of J.W. Miller, Homocysteine, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 424–430, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00144-6.

Introduction 259
Structure and forms 259
Biosynthesis and metabolism 260
Regulation of metabolism 261
Hyperhomocysteinemia 261
Homocysteine and clinical outcomes 263
Measurement of blood concentrations 265
Summary 266
References 266
Further Reading 267

Key Points

- Homocysteine is a sulfur amino acid and product of methionine metabolism.
- Homocysteine metabolism requires several B-vitamins, including folate, B12, B6, and riboflavin.
- Hyperhomocysteinemia is caused by B vitamin deficiencies, genetic defects, and other pathophysiological conditions, including renal disease and hypothyroidism.
- Hyperhomocysteinemia is a risk factor for vascular disease, and neurodegenerative disease.
- B-vitamin supplementation lowers blood homocysteine concentrations.
- It is unclear if B-vitamin supplementation reduces the risk of vascular disease and neurodegenerative disease.

Introduction

Homocysteine is a sulfur amino acid and an intermediate in the biochemical conversion of methionine to cysteine, a process called transsulfuration. The biochemistry of homocysteine was elucidated by Nobel Laureate Vincent Du Vigneaud and colleagues from the 1930s to the 1950s (Finkelstein, 2000). In the early 1960s, the description and characterization of the inborn error of metabolism, homocystinuria (Mudd et al., 1995), initiated a 60-year (and continuing) period of investigation that has revealed elevated blood homocysteine (hyperhomocysteinemia) as an independent risk factor for over 100 conditions, with most attention focused on vascular and neurodegenerative diseases (Smith and Refsum, 2021). Randomized controlled trials of B vitamins (folic acid with or without vitamin B12 and vitamin B6) effectively lower blood homocysteine concentration, but have been mostly ineffective in demonstrating significant effects on cardiovascular disease incidence (Clarke et al., 2010) or risk of age-related cognitive decline and dementia/Alzheimer's disease (Clarke et al., 2014). This has led many to the conclusion that hyperhomocysteinemia is a risk marker, but not a causative factor, in the pathogenesis of cardiovascular and neurodegenerative diseases. However, there have been several clinical trials with compelling evidence that homocysteine-lowering with B vitamins may be specifically effective in preventing stroke and slowing brain atrophy and cognitive decline in older adults with mild cognitive impairment (Smith and Refsum, 2021).

Structure and forms

The most prominent features of homocysteine and cysteine are the free sulfhydryl groups located at the end of the side chains of both amino acids (**Table 1**). These sulfhydryl groups are highly susceptible to oxidation and the formation of disulfide linkages with other sulfhydryl compounds. The primary forms of homocysteine found in the blood (**Table 1**) consist of homocysteine in disulfide linkage with (1) cysteine residues within the primary sequences of albumin and other plasma proteins (protein-bound form), (2) free cysteines or cysteine containing peptides such as glutathione (mixed disulfides), and (3) other homocysteine molecules (homocystine). Only approximately 1% of homocysteine in the blood is in the free-reduced form. Methionine, in contrast, does not have a free sulfhydryl group, and therefore, does not form disulfide compounds.

Biosynthesis and metabolism

The ultimate source of homocysteine is dietary methionine (Fig. 1) (Selhub and Miller, 1992). Methionine is first activated by addition of an adenosyl group (from adenosine triphosphate) to form S-adenosylmethionine (SAM). SAM is an important intermediate known as the universal methyl donor for its role as the methylating agent in a variety of essential reactions, including those involving DNA, RNA, proteins, membrane phospholipids, neurotransmitters, and the synthesis of creatine. A product of all SAM-dependent methylation reactions is S-adenosylhomocysteine (SAH) which, in turn, is metabolized to form adenosine and homocysteine. Homocysteine is then at a metabolic intersection: It can be remethylated to form methionine or catabolized through cystathionine synthesis.

In remethylation, homocysteine reacquires a methyl group in a reaction catalyzed by the zinc-dependent enzyme, methionine synthase (5-methyltetrahydrofolate-homocysteine methyltransferase, EC 2.1.1.13), with methyltetrahydrofolate serving as the methyl donor and vitamin B12 serving as a cofactor. This reaction occurs in all mammalian cells. Alternatively, homocysteine can be remethylated in a folate- and vitamin B12-independent reaction using betaine as the methyl donor and catalyzed by betaine-homocysteine methyltransferase (EC 2.1.1.5). This reaction occurs primarily in the liver, to a lesser extent in the kidney, and possibly the brain.

Table 1 Structures and forms of homocysteine and related amino acids.

Structures		*Forms of homocysteine in blood*	
Homocysteine	NH_3^+ H-C-CH_2-CH_2-S-H COOH	Protein-bound	HCY-S-S-CYS-albumin
Cysteine	NH_3^+ H-C-CH_2-SH COOH	Mixed disulfide Homocystine	HCY-S-S-CYS HCY-S-S-HCY
Methionine	NH_3^+ H-C-CH_2-CH_2-S-CH_3 COOH	Free reduced	HCY-SH

Abbreviations: HCY, homocysteine; CYS, cysteine.

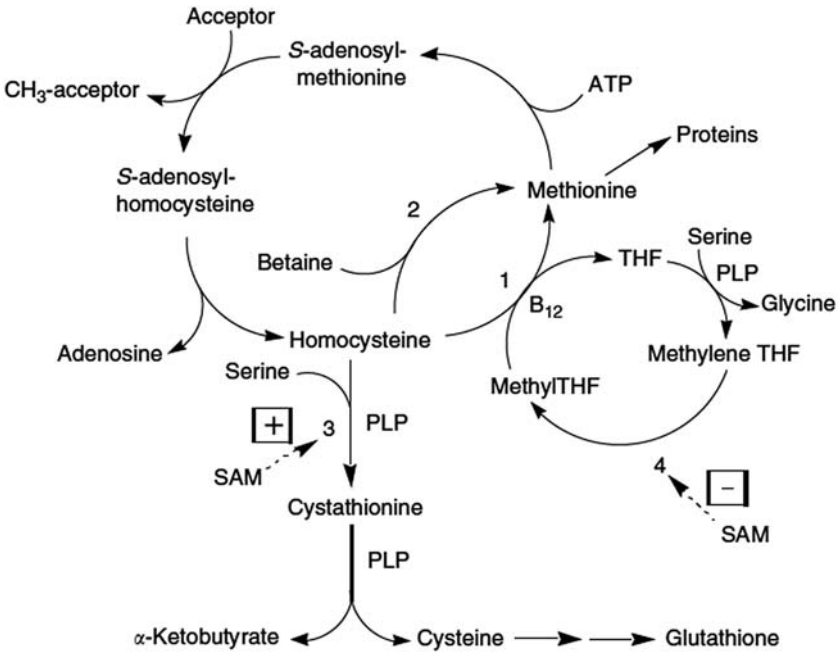


Fig. 1 The biosynthesis and metabolism of homocysteine. Reactions that are regulated by S-adenosylmethionine (SAM) are indicated by positive and negative signs. Key enzymes: (1) methyltetrahydrofolate-homocysteine methyltransferase or methionine synthase; (2) betaine-homocysteine methyltransferase; (3) cystathionine ß-synthase; (4) methylenetetrahydrofolate reductase. Abbreviations: THF, tetrahydrofolate; PLP, pyridoxal-5′-phosphate (vitamin B6).

Homocysteine catabolism occurs through cystathionine synthesis in a condensation reaction with serine. This reaction is catalyzed by cystathionine ß-synthase (EC 4.2.1.22), which requires vitamin B6 in the form of pyridoxal-5′-phosphate (PLP) as a cofactor. Cystathionine is then cleaved to form α-ketobutyrate and cysteine in a second PLP-dependent reaction catalyzed by cystathionine γ-lyase (EC 4.4.1.1). Further metabolism of cysteine leads to the formation of glutathione or inorganic sulfate.

Regulation of metabolism

An important aspect of homocysteine metabolism is that it is subject to allosteric control (Selhub and Miller, 1992). In addition to serving as the universal methyl donor, SAM is also an activator of cystathionine ß-synthase and an inhibitor of methylenetetrahydrofolate reductase (MTHFR) (EC 1.7.99.5), the enzyme responsible for the synthesis of methyltetrahydrofolate (**Fig. 1**). These allosteric functions serve to control whether homocysteine is recycled to form methionine or catabolized to form cystathionine. When dietary supply of methionine is high, such as after a protein meal, intracellular SAM concentration increases. High SAM activates cystathionine ß-synthase and inhibits MTHFR, thereby promoting homocysteine catabolism and limiting homocysteine remethylation. This serves to reduce the recycling of homocysteine when there is an adequate dietary supply of methionine. Conversely, under fasting conditions when there is no dietary influx of methionine, intracellular SAM concentration decreases. Cystathionine ß-synthase is then not activated and the inhibition of MTHFR is relieved, thus promoting homocysteine remethylation over catabolism. Consequently, this maintains intracellular methionine concentration during times of limited dietary supply.

An additional level of control on homocysteine metabolism is exerted by oxidative stress, which reduces methionine synthase activity. This may occur by oxidative inactivation of the vitamin B12 cofactor or by oxidation of cysteine residues that are important for zinc binding. By inhibiting methionine synthase, oxidative stress tends to divert homocysteine toward cystathionine synthesis away from methionine synthesis. This serves to increase synthesis of glutathione, an important intracellular antioxidant.

As discussed in the following Section (Hyperhomocysteinemia – Other Causes of Hyperhomocysteinemia), alterations in homocysteine metabolism also occur after menopause, in diabetes, and in hypothyroidism. These observations suggest that hormones, including estrogen, insulin, thyroxine, and thyroid-stimulating hormone, may directly or indirectly affect homocysteine metabolism. The mechanisms by which these hormones affect homocysteine metabolism are poorly understood.

Hyperhomocysteinemia

Under conditions of maximal metabolic efficiency, plasma homocysteine ranges from 4 to 10 μmol L^{-1} (Refsum et al., 2004). Metabolic blocks in homocysteine metabolism lead to accumulation of intracellular homocysteine with subsequent export into the blood. Depending on the magnitude of the metabolic impairment, plasma homocysteine concentration can rise to varying degrees, as defined in **Table 2**.

Genetic defects. Severe elevations in plasma homocysteine (concentrations as high as several hundred micromoles per liter) are observed in individuals with homozygous genetic defects affecting cystathionine ß-synthase, MTHFR, or any of several enzymes responsible for the conversion of vitamin B12 to its methionine synthase-associated cofactor form (Blom et al., 1998). These autosomal recessive genetic disorders, collectively termed homocystinuria because homocysteine accumulates in the urine as well as the blood, are associated with severe premature vascular disease, including thrombosis and atherosclerosis, mental retardation, dislocation of the eye lens (*ectopia lentis*), and skeletal malformations. Premature death (often in childhood) usually results from a major thrombotic or embolic event. Notably, one of the genetic defects that afflicts a significant proportion of homocystinuria patients reduces the affinity of cystathionine ß-synthase for its vitamin B6 cofactor, PLP. For these patients, the metabolic defect can be overcome, to some extent, with high dose vitamin B6 supplements, which significantly lower plasma homocysteine, reduce morbidity, and increase life expectancy. Interestingly, for other genetic defects involving cystathionine ß-synthase that cause homocystinuria independent of the affinity of the enzyme for PLP, high dose vitamin B6 supplements also have a therapeutic effect, despite having little or no influence on plasma homocysteine concentration.

B-vitamin deficiencies. Hyperhomocysteinemia is also caused by B-vitamin deficiencies (Selhub et al., 1993). Deficiencies of folate and vitamin B12 lead to impaired remethylation of homocysteine causing mild, moderate, or severe elevations in plasma homocysteine, depending on the severity of the deficiency, as well as coexistence of genetic or other factors that interfere with homocysteine metabolism (see below, in this Section). Because riboflavin is required for synthesis of flavin adenine dinucleotide (FAD), and because FAD serves as a cofactor for MTHFR, riboflavin deficiency can also affect homocysteine remethylation, and, therefore,

Table 2 Degrees of hyperhomocysteinemia.

Total plasma homocysteine	Designation
4–10 μmol L^{-1}	Normal
11–25 μmol L^{-1}	Mild to moderate
26–50 μmol L^{-1}	Intermediate
>50 μmol L^{-1}	Severe

contribute to elevation in plasma homocysteine. Vitamin B6 deficiency leads to impairment of homocysteine catabolism and, thus, also causes hyperhomocysteinemia. However, the nature of hyperhomocysteinemia caused by vitamin B6 deficiency differs from that caused by folate and vitamin B12 deficiencies: In vitamin B6 deficiency, fasting blood homocysteine is usually not elevated or is only slightly elevated. Only after a protein meal, or after consumption of an oral methionine load (see the Section on Measurement of Blood Concentrations), does plasma homocysteine become abnormally elevated in vitamin B6-deficient patients. In contrast, plasma homocysteine tends to be elevated regardless of prandial state in patients with folate or vitamin B12 deficiency. The basis for these different manifestations is likely due to differential effects of the vitamin deficiencies on intracellular SAM and consequent disruption of the allosteric control of homocysteine metabolism (Selhub and Miller, 1992).

Over the last two decades, there has been growing interest in nutritional genomics. This refers to genetic variability among individuals and its effect on nutritional requirements. A prime example of this concept is a common polymorphism in MTHFR (677C→T) in which an alanine is replaced by valine at codon 222 in the primary sequence of the enzyme (Frosst et al., 1995). Individuals with the homozygous variant (677 TT) of this gene (10–15% of the general population, lower in blacks, higher in Latinos and in some parts of Europe, e.g. Southern Italy) have an enzyme that is thermolabile, with reduced affinity for its substrate (methylenetetrahydrofolate) and its cofactor (FAD). Consequently, 677 TT individuals require a higher intake of folate and riboflavin to maintain optimal enzyme activity than those with the wild-type isoform of the enzyme (677 CC). This is reflected by the fact that blood homocysteine is higher in people with the 677 TT isoform than in those with the 677 CC isoform, but only when overall folate and riboflavin status is low. When overall folate and riboflavin status is high, no difference in homocysteine is observed between the isoforms.

The clinical and public health importance of the MTHFR polymorphism is reflected by two observations: First, women with the 677 TT isoform are at increased risk of having a child with a neural tube defect (e.g., *spina bifida* and anenencephaly) (Reilly et al., 2014). This risk can be reduced by folic acid supplements, an observation that underlies the decision by the US and Canadian governments to mandate folic acid fortification of grain products in 1998 (Miller, 2011). This program has been highly successful, having reduced the prevalence of folate deficiency from over 20% to approximately 1%, the prevalence of hyperhomocysteinemia by approximately 50%, and the incidence of neural tube defects by at least 20%. The success of the folic acid fortification program in the USA and Canada has spawned similar programs in more than 80 countries and territories throughout the world. Notable exceptions are the countries of the European Union, which have been slow to adopt this intervention strategy. This is due to concern about the feasibility of fortification, hesitancy to impose mandatory fortification on the population, and lingering concern about masking B12 deficiency and the possibility of other unrecognized health consequences associated with excess folic acid intake.

Individuals homozygous for the 677 TT isoform of MTHFR have been found to have elevated blood pressure, compared with wild-type (677 CC) and heterozygous (677 CT) individuals (Psara et al., 2020). Notably, supplementation of 677 TT individuals with riboflavin, the vitamin precursor of the MTHFR cofactor, FAD, effectively lowers blood homocysteine and both diastolic and systolic blood pressure to levels similar to wild-type and heterozygous individuals. The influence of these effects of riboflavin supplements on homocysteine and blood pressure on risk of stroke and other vascular disease outcomes remains to be determined.

Other polymorphisms in MTHFR and other enzymes involved in homocysteine metabolism [e.g. methionine synthase, methionine synthase reductase (EC 1.16.1.8), cystathionine ß-synthase] have been identified and their overall influence on homocysteine metabolism, B vitamin requirements, and disease risk have been, and continue to be, evaluated.

Other causes of hyperhomocysteinemia. Other pathophysiological causes of hyperhomocysteinemia include renal dysfunction and hypothyroidism. The kidney is a major site of homocysteine metabolism and renal disease leads to a significant reduction in the body's overall capacity to metabolize this amino acid (Bostom and Lathrop, 1997). The resulting moderate to severe hyperhomocysteinemia can be attenuated, in part, by high dose B-vitamin supplements, which putatively maximize the residual renal metabolism, as well as the metabolic capacities of the extrarenal organs. Mild elevation in homocysteine occurs in patients with hypothyroidism, which resolves to normal with thyroid replacement therapy (Hussein et al., 1999). This observation implies that thyroxine and thyroid-stimulating hormone influence homocysteine metabolism directly, perhaps through up- or downregulation of key homocysteine-metabolizing enzymes. Alternatively, homocysteine may become elevated in hypothyroid patients secondary to mild impairment of renal function that may accompany the disorder.

Patients with diabetes (both insulin-dependent and insulin-independent) tend to have mild hyperhomocysteinemia (Wijekoon et al., 2007). However, this seems to be confined to those patients whose diabetes condition has progressed to involve renal insufficiency. Notably, in the absence of renal involvement, homocysteine concentration in diabetes patients tends to be lower than normal. Insulin has been shown to inhibit homocysteine catabolism through cystathionine synthesis. Therefore, reduced insulin in diabetes patients may actually promote homocysteine catabolism, thus leading to lower plasma concentration.

Premenopausal women tend to have lower plasma homocysteine than men of similar age, and homocysteine tends to rise in women after menopause. Hormone replacement therapy reduces homocysteine back to premenopausal concentration. Moreover, homocysteine decreases in male to female transsexuals, and increases in female to male transsexuals, effects primarily related to the estrogen and androgen regimens that such individuals follow. Taken together, these observations suggest an influence of sex hormones on homocysteine metabolism, though the mechanisms are not well understood.

Drugs can also affect homocysteine metabolism and lead to elevations of homocysteine in the blood. Certain anti-cancer drugs, such as methotrexate, and anti-epilepsy medications, such as valproate and carbamazepine, are inhibitors of folate metabolism. Resulting functional folate deficiency leads to hyperhomocysteinemia. The anti-Parkinsonian drug, levodopa or L-dopa, causes elevation in blood homocysteine by a different mechanism: A significant proportion of an oral dose of L-dopa is methylated by SAM, leading to increased intracellular synthesis of SAH and homocysteine. The excess synthesis of homocysteine can overwhelm

the capacities of the homocysteine metabolic pathways, particularly when B vitamin status is suboptimal, leading to hyperhomocysteinemia (Miller et al., 2003).

Homocysteine and clinical outcomes

Homocysteine and vascular disease. The continuing interest in homocysteine is primarily related to its recognized status as an independent risk factor for cardiovascular, cerebrovascular, and peripheral vascular disease. This homocysteine theory of vascular disease is directly descendant from a seminal observation made by Kilmer McCully (McCully, 1969). In the early to mid-1960s, it was recognized that a prominent characteristic of patients with homocystinuria caused by defects in cystathionine ß-synthase were very high elevations of both homocysteine and methionine in the blood. Therefore, it was not clear whether the vascular complications of this disorder were the consequence of hyperhomocysteinemia or hypermethioninemia. McCully observed that a patient with homocystinuria caused by a defect in a B12-metabolizing enzyme had hyperhomocysteinemia, but not hypermethioninemia. Nonetheless, this patient had similar (although not identical) vascular pathology to that observed in patients with homocystinuria caused by cystathionine ß-synthase deficiency. From this, McCully concluded that the vascular culprit was homocysteine, and not methionine.

McCully's hypothesis was not immediately embraced. The prevailing theory of atherosclerosis at the time centered on cholesterol, and it proved difficult for McCully to convince his peers and national funding agencies of the potential importance of this new and competing hypothesis. Contributing to this was a lack of reproducible animal models of homocysteine-induced vascular disease, or a sensitive method to measure homocysteine in the blood. Consequently, McCully's hypothesis went into temporary obscurity.

In the mid-1970s, David and Bridget Wilcken reinvigorated McCully's hypothesis with their observation that a subset of patients with premature coronary artery disease had reduced ability to metabolize homocysteine (Wilcken and Wilcken, 1976). Notably, this association was observed in individuals who did not have any of the severe genetic defects that underlie homocystinuria, suggesting that less severe or modest impairment of homocysteine metabolism may contribute to vascular disease risk. Subsequently, the advent of reliable, high-throughput assays for total plasma or serum homocysteine in the 1980s (see the Section on Measurement of Blood Levels) allowed for large-scale epidemiological assessment of associations between homocysteine and vascular (and other) diseases, both cross-sectionally and longitudinally. Through the 1990s, an explosion of population and case-control studies established that hyperhomocysteinemia is, indeed, a risk factor for heart attack, stroke, thrombosis, and peripheral atherosclerotic disease. Moreover, the risk associated with hyperhomocysteinemia is independent of other prominent risk factors, such as hypertension, hypercholesterolemia, hyperlipidemia, smoking, male gender, and others. Further indication of the importance of homocysteine with respect to vascular disease is the estimate that the relative risk of coronary artery disease associated with hyperhomocysteinemia is about equivalent to that associated with hypercholesterolemia (Boushey et al., 1995). As the evidence mounted, McCully was vindicated and his contribution became widely recognized.

Homocysteine, Cognitive Function, and Dementia. As the relationship between homocysteine and vascular disease became more and more apparent, researchers also addressed the hypothesis that hyperhomocysteinemia may affect cognitive function and the risk of dementia in older adults. This was based primarily not only on the recognized association between homocysteine and cerebrovascular disease, but also the observation that homocysteine and its metabolite, homocysteic acid, can induce excitotoxicity in neurons. Throughout the 1990s and into the new century, many cohort studies revealed significant inverse correlations between plasma homocysteine concentration and performance on a variety of cognitive function tests. Moreover, individuals with Alzheimer's disease were found to have higher plasma homocysteine than age and sex-matched controls (Clarke et al., 1998), whereas baseline homocysteine predicted risk of incident dementia (Seshadri et al., 2002). Baseline homocysteine also predicted the rate of brain atrophy in older adults with mild cognitive impairment, and lowering of homocysteine with B vitamin supplements was shown to slow the rate of atrophy (Smith et al., 2010).

Homocysteine and Pregnancy Outcomes. Hyperhomocysteinemia has also been suspected as a risk factor for pregnancy complications and birth defects. Elevated plasma homocysteine has been associated with placental vasculopathy, pre-eclampsia, and placental infarction, as well as recurrent premature delivery, low birth weight, and spontaneous abortion. Birth defects associated with hyperhomocysteinemia in the mother include neural tube defects, orofacial clefts, clubfoot, and Down's syndrome. The protective effect of folic acid supplementation and fortification against neural tube defects, and perhaps the other abnormal birth outcomes cited, may be related to reduced homocysteine concentration.

Mechanisms. In parallel with epidemiological studies, a significant amount of basic research has focused on the mechanism(s) by which homocysteine may induce atherosclerosis and thrombosis. A definitive answer has proven elusive. Potential mechanisms with significant experimental support include, but are not limited to, the following: (1) modification of the endothelial cell surface, (2) *S*- or *N*-homocysteinylation of plasma and cellular proteins, (3) activation of platelets, (4) modification of monocyte functions, (5) increased expression or activity of vascular adhesion molecules, and (6) oxidative damage induced by peroxides formed during disulfide bond formation.

A sixth potential mechanism relates to a known quirk of homocysteine synthesis and metabolism. The equilibrium of the interconversion between SAH and homocysteine (catalyzed by SAH hydrolase, EC 3.3.1.1) actually favors SAH synthesis (**Fig. 1**). *In vivo*, this reaction proceeds toward homocysteine synthesis because of product removal, i.e., the efficient metabolism of homocysteine back to methionine or through cystathionine synthesis. However, when there is a block in homocysteine metabolism, as occurs in the genetic defects, B-vitamin deficiencies, and other causes delineated previously (see Section on Hyperhomocysteinemia above),

homocysteine accumulates intracellularly. Consequently, SAH also accumulates within cells. The significance of this phenomenon is that SAH is a feedback inhibitor of all SAM-dependent methylation reactions. Therefore, hyperhomocysteinemia may cause or contribute to vascular disease through SAH-mediated inhibition of methylation.

Another area that has received attention is the relationship between homocysteine, nitric oxide, and endothelial function. One of the roles of nitric oxide is as a vasodilator. Homocysteine has been shown to be an inhibitor of nitric oxide synthesis, and thus can inhibit vasodilatation. This has led to the hypothesis that hyperhomocysteinemia, by inhibiting nitric oxide synthesis, impairs the ability of the vascular endothelium to maintain homeostasis of vascular tone. This in turn may directly or indirectly increase susceptibility to vascular insults, thus promoting atherosclerosis and thrombosis. This may underlie the aforementioned association between the MTHFR 677C→T polymorphism and blood pressure (Psara et al., 2020). Alternatively, elevated homocysteine may reflect reduced intracellular production of the product of the MTHFR reaction, 5-methyltetrahydrofolate, which is both a required substrate for the folate- and B12-dependent conversion of homocysteine to methionine and a putative antioxidant that protects nitric oxide synthesis from uncoupling by superoxide radicals.

An array of possible mechanisms by which homocysteine contributes to neurodegenerative disease also have been postulated. Those that have compelling experimental support include: (1) promotion of cerebrovascular disease by one or more of the mechanisms described above in this Section, (2) hyperphosphorylation of tau through activation of tau kinases leading to the formation of neurofibrillary tangles, and (3) inhibition of SAM-dependent methylation reactions leading to deposition of ß-amyloid in the brain, reduced activity of protein phosphatase-2A (responsible for dephosphorylating tau), and impaired formation of phosphatidylcholine enriched in omega-3 fatty acids (Smith and Refsum, 2016).

The search for the definitive pathogenetic mechanism implicating homocysteine as a cause of vascular and neurodegenerative diseases continues, and it is recognized that several mechanisms may contribute synergistically. However, some have questioned whether homocysteine is a cause of vascular disease, or simply a consequence.

Cause or effect? Although there is considerable evidence, both epidemiological and experimental, that homocysteine is a causative factor in vascular disease, there are data that contradict this conclusion (Brattstrom and Wilcken, 2000). First, although cross-sectional and case-control studies fairly consistently demonstrate that hyperhomocysteinemia is associated with vascular disease, some prospective studies have found no relationship between baseline homocysteine and risk of incident vascular events. Second, several studies have found no relationship between the MTHFR 667C→T polymorphism and venous thrombosis, despite the association of this polymorphism with elevated plasma homocysteine.

In the late 1990s, several large-scale intervention trials were initiated to determine if folic acid supplements with or without vitamin B12 and vitamin B6, which effectively lower blood homocysteine, reduce vascular outcomes and mortality in individuals at increased risk of cardiovascular disease (**Table 3**). A meta-analysis evaluated the results of these studies. Overall, folic acid with or without vitamin B12 or vitamin B6, was effective in lowering homocysteine (an average of 25% across all the studies). However, over a median follow-up of 5 years, no significant reductions in major vascular events, major coronary events, stroke, or vascular mortality were observed. These findings suggest that lowering homocysteine with B vitamins is not an effective strategy for preventing major vascular outcomes (Clarke et al., 2010). A similar conclusion was made regarding homocysteine-lowering via B-vitamin supplements for the prevention of age-associated cognitive decline and risk of Alzheimer's disease/dementia. In the 2000s and early 2010s, a series of B-vitamin supplementation trials were conducted in older adults, and the results pooled and analyzed by meta-analysis. While, again, supplementation effectively lowered homocysteine (approximately 26–28%), no significant effects on global cognitive function or specific cognitive domains were found (Clarke et al., 2014).

Based on these analyses, some have concluded that there is no benefit of B-vitamin supplementation and homocysteine-lowering with respect to vascular and neurodegenerative disease risk. However, these conclusions have been challenged. *Post-hoc* reviews of the conditions and protocols of intervention trials, the characteristics of study participants, and sub-group analyses reveal that homocysteine-lowering may be beneficial for specific conditions under specific circumstances (Smith and Refsum, 2021). These include the following:

- Homocysteine-lowering may be more effective in preventing stroke than myocardial infarction and other major cardiovascular events.
- The effect of homocysteine-lowering on the risk of stroke may be modified by age of at-risk individuals, exposure to folic acid fortification, baseline homocysteine and cholesterol concentrations, and treatment with anti-platelet or lipid-lowering drugs.
- Homocysteine-lowering may be more effective in primary stroke prevention rather than recurrence.
- The effect of homocysteine-lowering on age-associated cognitive decline and risk of neurodegenerative disease may be modified by baseline homocysteine, as well as treatment with aspirin and omega-3 fatty acid status.
- The effect of homocysteine-lowering on age-associated cognitive decline may be most likely to be observed in older adults who are already experiencing cognitive decline, but who have not yet progressed to clinical dementia, that is, those with mild cognitive impairment.

Consensus on these issues has not yet been achieved. The uncertain relationship between hyperhomocysteinemia, B vitamins, and vascular disease/neurodegenerative disease is summarized in **Fig. 2**. If homocysteine is not a direct causative factor, it may still serve as a biomarker of both disease and the efficacy of B vitamin supplementation.

Table 3 Intervention trials to determine the effect of B vitamin supplements on homocysteine and the risk of vascular events and mortality.

Study	*Location*	*Trial Period*	*Intervention*	*Outcomes*
Cambridge Heart Antioxidant Study 2 (CHAOS-2)	UK	1998–2002	FA (5.0 mg)	HCY reduced 11%; No effect on major Vascular events
Heart Outcomes Prevention Evaluation 2 (HOPE-2)	Canada	2000–2006	FA (2.5 mg) B12 (1.0 mg) B6 (50 mg)	HCY reduced 24%; No effect on death from cardiovascular events, MI, or stroke
Homocysteinemia in Kidney and End Stage Renal Disease (HOST)	USA	2000–2007	FA (40 mg) B12 (2 mg) B6 (100 mg)	HCY reduced 25%; No effect on all cause mortality, MI, or stroke
Norwegian Multi-Center B-Vitamin Intervention Study (NORVIT)	Norway	1999–2006	FA (0.8 mg) B12 (0.4 mg) B6 (40 mg)	HCY reduced 28%; No effect on MI, stroke, or sudden death due to coronary artery disease
Study of Effectiveness of Additional Reductions in Cholesterol and Homocysteine (SEARCH)	UK	1999–2008	FA (2.0 mg) B12 (1.0 mg)	HCY reduced 27%; No effect on major coronary events, stroke, non-coronary revascularization, or death due to vascular disease
Vitamins in Stroke Prevention (VISP)	USA	1996–2003	FA (2.5 mg) B12 (0.4 mg) B6 (25 mg)	HCY reduced 17%; No effect on stroke
Western Norway B-Vitamin Intervention Trial (WENBIT)	Norway	1999–2006	FA (0.8 mg) B12 (0.4 mg) B6 (40 mg)	HCY reduced 26%; No effect on MI, storke, or hospitalization for angina pectoris
Women's Antioxidant and Folic Acid Cardiovascular Study (WAFACS)	USA	1998–2006	FA (2.5 mg) B12 (1.0 mg) B6 (50 mg)	HCY reduced 18%; No effect on MI, stroke, coronary revascularization, or cardiovascular mortality
B-Vitamin Treatment Trialists' Collaboration - Meta-Analysis	Multiple	–	–	HCY reduced 25% (mean of all studies); No overall effect on major vascular events, including cardiovascular events, stroke, and revascularization

Abbreviations: FA, folic acid; B12, vitamin B12; B6, vitamin B6; HCY, homocysteine; MI, myocardial infarction.

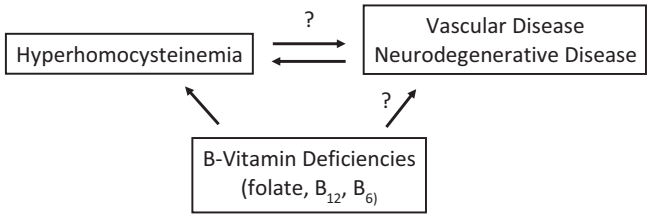


Fig. 2 Hyperhomocysteinemia, B vitamins, and vascular disease/neurodegenerative disease. There is still some question whether elevated plasma homocysteine is a cause or consequence of vascular disease and neurodegenerative disease, and whether there are influences of B vitamins on disease risk that are independent of homocysteine.

Measurement of blood concentrations

A variety of assays have been developed to quantify blood homocysteine, with those employing high-pressure liquid chromatography the most common (Refsum et al., 2004). These assays have proven to be relatively accurate and precise (coefficients of variation $< 10\%$) and are relatively simple and rapid to perform. The development of such assays in the 1980s was the technological breakthrough that spurred the exponential increase in homocysteine-related research over the last 40 years, leading to the establishment of hyperhomocysteinemia as an independent risk factor for vascular disease, neurodegenerative disease, and other conditions.

As described above (in the Section Structure and Forms), homocysteine exists in several forms in the blood, including the protein-bound form, mixed disulfides, homocystine, and the free-reduced form. Assays for homocysteine usually measure the sum total of all these forms, i.e., total homocysteine. To accomplish this, the first procedure in homocysteine assays is a reduction step to break all disulfide bonds, thus converting all homocysteine to the free-reduced form. The free-reduced form is then quantified by one of various methods.

Blood sample collection and processing are critical factors in the determination of homocysteine concentration. Typically, blood samples for homocysteine analysis are collected in tubes containing an anticoagulant (e.g., ethylenediaminetetraacetic acid, heparin). Prompt separation of plasma from the blood cells after centrifugation is required to avoid excess release of intracellular homocysteine into the plasma or removal of homocysteine from the plasma by metabolically active leukocytes after blood withdrawal. Keeping the blood sample cold until centrifugation and separation (ideally within 4 h of blood withdrawal) minimizes

this problem. Serum homocysteine concentration typically exceeds plasma concentration by 20%. This is likely due to the fact that blood collected to isolate serum (i.e., without an anticoagulant) must clot at room temperature for 30–60 min before centrifugation and separation. Therefore, plasma is preferred for measurement of homocysteine. Once separated from blood cells, the concentration of homocysteine in plasma or serum remains stable for years when stored frozen.

Another important issue in the measurement of homocysteine is the prandial state of the individual. For individuals with adequate B-vitamin status, no genetic abnormalities, and no pathophysiological conditions that affect homocysteine metabolism, plasma homocysteine concentration after an overnight fast is similar to that after a meal (even high protein meals containing methionine). However, for individuals with low B6 status or heterozygous genetic defects in cystathionine ß-synthase, post-prandial homocysteine can be significantly higher than fasting concentration. Because of the nutritional or genetic block in the conversion of homocysteine to cystathionine, there is decreased capacity to metabolize the influx of homocysteine synthesized from dietary methionine. This is the basis for the methionine load test for detection of impaired cystathionine ß-synthase activity. In this test, baseline blood is drawn after an overnight fast, and then again 4 h after consumption of a large dose of methionine dissolved in orange juice (100 mg methionine per kilogram body weight). Plasma homocysteine increases to a greater extent in individuals with low vitamin B6 status or heterozygous genetic defects in cystathionine ß-synthase than in individuals without these problems. Importantly, individuals with elevated fasting homocysteine and those with normal fasting, but elevated post-methionine load concentrations, are both at increased risk of vascular disease.

Summary

Homocysteine is an important intermediate in the interconversion of methionine to cysteine (transsulfuration) and in folate/one-carbon metabolism. Elevation of homocysteine in the blood (hyperhomocysteinemia) primarily results from impaired metabolism due to B-vitamin deficiencies, genetic defects, and pathophysiological conditions such as renal disease and hypothyroidism. Hyperhomocysteinemia has been clearly established as a risk marker for many diseases and conditions, most prominently vascular disease, age-related cognitive decline and Alzheimer's disease/dementia. Whether lowering blood homocysteine concentration with B-vitamin supplements reduces the risk of vascular disease and neurodegenerative disease remains an open question. *Post hoc* review of clinical trials suggests that B-vitamin supplementation and homocysteine-lowering may be particularly beneficial for prevention of stroke and the slowing of cognitive decline and brain atrophy in older adults with mild cognitive impairment. Additional randomized control trials are needed to provide more definitive answers.

References

Blom, H., Fowler, B., Jakobs, C., Koch, H.-G., 1998. Disorders of homocysteine metabolism. Eur. J. Pediatr. 157 (Suppl. 2), S39–S142.

Bostom, A., Lathrop, L., 1997. Hyperhomocysteinemia in end-stage renal disease: prevalence, etiology, and potential relationship to arteriosclerotic outcomes. Kidney Int. 52, 10–20.

Boushey, C.J., Beresford, S.A.A., Omenn, G.S., Motulsky, A.G., 1995. A quantitative assessment of plasma homocysteine as a risk factor for vascular disease: probable benefits of increasing folic acid intakes. JAMA 274, 1049–1057.

Brattström, L., Wilcken, D.E.L., 2000. Homocysteine and cardiovascular disease: cause or effect? Am. J. Clin. Nutr. 72, 315–323.

Clarke, R., Bennett, D., Parish, S., et al., 2014. Effects of homocysteine lowering with B vitamins on cognitive aging: meta-analysis of 11 trials with cognitive data on 22,000 individuals. Am. J. Clin. Nutr. 100, 657–666.

Clarke, R., Halsey, J., Lewington, S., et al., 2010. Effects of lowering homocysteine levels with B vitamins on cardiovascular disease, cancer, and cause-specific mortality: meta-analysis of 8 randomized trials involving 37 485 individuals. Arch. Intern. Med. 170, 1622–1631.

Clarke, R., Smith, A.D., Jobst, K.A., et al., 1998. Folate, vitamin B12, and serum total homocysteine levels in confirmed Alzheimer disease. Arch. Neurol. 55, 1449–1455.

Finkelstein, J.D., 2000. Homocysteine: a history in progress. Nutr. Rev. 58, 193–204.

Frosst, P., Blom, H.J., Milos, R., et al., 1995. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat. Genet. 10, 111–113.

Hussein, W.I., Green, R., Jacobsen, D.W., Faiman, C., 1999. Normalization of hyperhomocysteinemia with L-thyroxine in hypothyroidism. Ann. Intern. Med. 131, 348–351.

McCully, K., 1969. Vascular pathology of homocysteinemia: implications for the pathogenesis of arteriosclerosis. Am. J. Pathol. 56, 111–128.

Miller, J.W., 2011. Folic acid fortification. In: Herrmann, W., Obeid, R. (Eds.), Vitamins in the Prevention of Human Diseases. De Gruyter, Berlin, pp. 273–293.

Miller, J.W., Selhub, J., Nadeau, M., et al., 2003. Effect of L-Dopa on plasma homocysteine in PD patients: relationship to B vitamin status. Neurology 60, 1125–1129.

Mudd, S.H., Levy, H.L., Skovby, F., 1995. Disorders of transsulfuration. In: Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D. (Eds.), The Metabolic and Molecular Bases of Inherited Disorders, seventh ed. McGraw Hill, Inc, New York, pp. 1279–1327.

Psara, E., Pentieva, K., Ward, M., McNulty, H., 2020. Critical review of nutrition, blood pressure and risk of hypertension through the lifecycle: do B vitamins play a role? Biochimie 173, 76–90.

Refsum, H., Smith, A.D., Ueland, P.M., et al., 2004. Facts and recommendations about total homocysteine determinations: an expert opinion. Clin. Chem. 50, 3–32.

Reilly, R., McNulty, H., Pentieva, K., Strain, J.J., Ward, M., 2014. MTHFR 677TT genotype and disease risk: is there a modulating role for B-vitamins? Proc. Nutr. Soc. 73, 47–56.

Selhub, J., Miller, J.W., 1992. The pathogenesis of homocysteinemia: interruption of the coordinate regulation by S-adenosylmethionine of the remethylation and transsulfuration of homocysteine. Am. J. Clin. Nutr. 55, 131–138.

Selhub, J., Jacques, P.F., Wilson, P.W.F., Rush, D., Rosenberg, I.H., 1993. Vitamin status and intake as primary determinants of homocysteinemia in an elderly population. JAMA 270, 2693–2698.

Seshadri, S., Beiser, A., Selhub, J., et al., 2002. Plasma homocysteine as a risk factor for dementia and Alzheimer's disease. N. Engl. J. Med. 346, 476–483.

Smith, A.D., Refsum, H., 2016. Homocysteine, B vitamins, and cognitive impairment. Annu. Rev. Nutr. 36, 211–239.

Smith, A.D., Refsum, H., 2021. Homocysteine – from disease marker to disease prevention. J. Intern. Med. 290 (4), 826–854.

Smith, A.D., Smith, S.M., de Jager, C.A., et al., 2010. Homocysteine-lowering by B vitamins slows the rate of accelerated brain atrophy in mild cognitive impairment: a randomized controlled trial. PLoS One 5, e12244.
Wijekoon, E.P., Brosnan, M.E., Brosnan, J.T., 2007. Homocysteine metabolism in diabetes. Biochem. Soc. Trans. 35 (Pt 5), 1175–1179.
Wilcken, D.E., Wilcken, B., 1976. The pathogenesis of coronary artery disease. A possible role for methionine metabolism. J. Clin. Invest. 57, 1079–1082.

Further Reading

Carmel, R., Jacobsen, D.W., 2001. Homocysteine in Health and Disease. Cambridge University Press, Cambridge, United Kingdom.
Graham, I., Refsum, H., Rosenberg, I.H., Ueland, P.M., 1997. Homocysteine Metabolism: From Basic Science to Clinical Medicine. Kluwar Academic Publishers, Norwell, MA.
Herrmann, W., 2005. Homocysteine research: where do we stand and where are we going? Clin. Chem. Lab. Med. 43 (10), 977–979.
Jacques, P.F., Selhub, J., Bostom, A.G., Wilson, P.W., Rosenberg, I.H., 1999. The effect of folic acid fortification on plasma folate and total homocysteine concentrations. N. Engl. J. Med. 340, 1449–1454.
McCaddon, A., Miller, J.W., 2015. Assessing the association between homocysteine and cognition: reflections on Bradford Hill, meta-analyses and causality. Nutr. Rev. 73, 723–735.
Robinson, K. (Ed.), 2000. Homocysteine and Vascular Disease. Kluwar Academic Publishers, Norwell, Massachusetts.

Iodine: Deficiency disorders and prevention programs

M.B. Zimmermann, Laboratory of Human Nutrition, Institute of Food, Nutrition and Health, ETH Zurich, Zurich, Switzerland

© 2023 Elsevier Ltd. All rights reserved.

This is an update of M.B. Zimmermann, Iodine: Deficiency Disorders and Prevention Programs, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 28–32, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00163-X.

Introduction 268
Dietary sources, absorption, and metabolism 268
Iodine deficiency disorders 269
Iodine requirements 270
Assessment of iodine status 270
Prophylaxis and treatment of iodine deficiency 271
Iodine excess and toxicity 272
Conclusions 272
References 272

Key points

- Iodine deficiency has multiple adverse effects on growth and development in humans; these are collectively termed the iodine deficiency disorders
- In nearly all regions affected by iodine deficiency, the most effective way to control iodine deficiency is through salt iodization

Glossary

Cretinism A congenital condition caused by a deficiency of thyroid hormone during prenatal development and characterized in childhood by dwarfed stature and mental retardation

Goiter A noncancerous enlargement of the thyroid gland, visible as a swelling at the front of the neck, that is often associated with iodine deficiency

Hypothyroidism A physiological state characterized by insufficient production and/or action of thyroid hormone, which can result in a decreased basal metabolic rate, causing weight gain and fatigue

Iodine deficiency disorders The collective term for the multiple adverse effects on growth and development in animals and humans caused by iodine deficiency and the resulting inadequate thyroid hormone production

Thyroglobulin A thyroid protein that is the precursor to iodine-containing hormones and is typically present in the colloid of thyroid gland follicles

Introduction

Bernard Courtois, a French chemist, discovered iodine (atomic weight 126.9) in 1811. In 1819, Jean-François Coindet, in Geneva, reported treating goiter with tincture of iodine. The connections between goiter, hypothyroidism and iodine were established in 1896, when Baumann and Roos discovered iodine in the thyroid. In the early 20th century, ground-breaking studies by Swiss and American scientists showed the efficacy of iodine prophylaxis in the prevention of goiter and cretinism.

Dietary sources, absorption, and metabolism

Iodine (as iodide) is widely but unevenly distributed in the Earth's environment. In many regions, leaching from glaciation, flooding, and erosion have depleted surface soils of iodide, and most iodide is found in the oceans. Iodide ions in seawater are oxidized to elemental iodine, which volatilizes into the atmosphere and is returned to the soil by rain, completing the cycle. However, iodine cycles in many regions are slow and incomplete, leaving soils and drinking water iodine depleted. Crops grown in these soils will be low in iodine, and humans and animals consuming food grown in these soils become iodine deficient. Iodine-deficient soils are common in mountainous areas (e.g., the Alps, Andes, Atlas, and Himalaya ranges) and areas of frequent flooding, especially in

South and Southeast Asia (for example, the Ganges River plain of northeastern India). Many inland areas, including central Asia and Africa, and central and eastern Europe are iodine deficient. Iodine deficiency in populations residing in these areas will persist until iodine enters the food chain through addition of iodine to foods (e.g., iodization of salt) or dietary diversification introduces foods produced outside the iodine-deficient area.

The native iodine content of most foods and beverages is low. In general, commonly consumed foods provide 3–80 μg per serving. Foods of marine origin have higher iodine content because marine plants and animals concentrate iodine from seawater. Major dietary sources of iodine in the US are bread and milk. In Switzerland, based on direct food analysis, mean intake of dietary iodine is ~140 μg day^{-1}, mainly from bread and dairy products. Much of the iodine content of dairy products is added during processing, for example, through the use of iodine-containing disinfectants. In many countries, use of iodized salt in households for cooking and at the table provides additional iodine. Dietary supplements often contain iodine. Based on data from the US Third National Health and Nutrition Examination Survey, 12% of men and 15% of nonpregnant women took a supplement that contained iodine, and the median intake of iodine from supplements was ~140 μg day^{-1} for adults.

Iodine is ingested in several chemical forms. Iodide is rapidly and nearly completely absorbed in the stomach and duodenum. Iodate, widely used in salt iodization, is reduced in the gut and absorbed as iodide. In healthy adults, the absorption of iodide is >90%. Iodine is cleared from the circulation mainly by the thyroid and kidney, and although renal iodine clearance is fairly constant, thyroid clearance varies with iodine intake. In conditions of adequate iodine supply, ≤10% of absorbed iodine is taken up by the thyroid. In chronic iodine deficiency, this fraction can exceed 80%. During lactation, the mammary gland concentrates iodine and secretes it into breast milk to provide for the newborn.

The body of a healthy adult contains up to 20 mg of iodine, of which 70–80% is in the thyroid. In chronic iodine deficiency, the iodine content of the thyroid may fall to <20 μg. In iodine-sufficient areas, the adult thyroid traps about 60 μg of iodine/day, either from dietary iodine or from iodine released during thyroid hormone turnover, to balance losses and maintain thyroid hormone synthesis. Thyroglobulin (Tg), a large glycoprotein (molecular weight 660,000), is the carrier of iodine in the follicles of the thyroid. Thyrocytes produce and secrete the two thyroid hormones from Tg, thyroxine (T4) (the major form) and triiodothyronine (T3) (Yen, 2001). In the circulation, thyroid hormones are bound noncovalently to carrier proteins, mainly thyroxine-binding globulin. In target tissues, including liver, kidney, heart, muscle, pituitary, and the developing brain, T4 is converted to T3. T3 is the main physiologically active form of thyroid hormone and binds to nuclear receptors.

Thyroid hormone regulates a variety of physiologic processes, including reproductive function, growth and development, as well as the basal metabolic rate. During pregnancy, thyroid hormone crosses the placenta to the fetus early in the first trimester, before the fetal thyroid is functioning. In the developing brain, it influences cell growth and migration. It also promotes growth and maturation of peripheral tissues and the skeleton. Thyroid hormone increases energy metabolism in most tissues. It also increases the basal metabolic rate.

Both T4 and T3 are degraded through a complex series of pathways, and their turnover is relatively slow: the half-life of T4 is ~5 days and for T3, 1.5–3 days. The released iodine enters the plasma iodine pool and can be taken up again by the thyroid or excreted by the kidney. More than 90% of ingested iodine is ultimately excreted in the urine, with only a small amount appearing in the feces.

Iodine deficiency disorders

Iodine deficiency has multiple adverse effects on growth and development in animals and humans (Zimmermann, 2009). These are collectively termed the iodine deficiency disorders (IDD) (**Table 1**), and are one of the most important and common human diseases. They result from inadequate thyroid hormone production due to lack of sufficient iodine.

Table 1 Iodine deficiency disorders, by age group.

Age groups	*Health consequences of iodine deficiency*
All ages	Goiter
	Increased susceptibility of the thyroid gland to nuclear radiation
Fetus	Abortion
	Stillbirth
	Congenital anomalies
	Perinatal mortality
Neonate	Infant mortality
	Endemic cretinism
Child and adolescent	Impaired mental function
	Delayed physical development
Adults	Reduced work productivity
	Toxic nodular goiter; iodine-induced hyperthyroidism
	Hypothyroidism in moderate-to-severe iodine deficiency

Thyroid enlargement (goiter) is the classic sign of iodine deficiency. It is a physiologic adaptation to chronic iodine deficiency. As iodine intake falls, the ratio of T4 to T3 produced by the gland decreases, secretion of TSH increases in an effort to maximize uptake of available iodine, and thyroid-stimulating hormone (TSH) stimulates thyroid hypertrophy and hyperplasia. Large goiters may be cosmetically unattractive, can obstruct the trachea and esophagus, and may damage the recurrent laryngeal nerves and cause hoarseness.

Although goiter is the most visible effect of iodine deficiency, the most serious adverse effect is damage to reproduction and fetal development (Gowachirapant, 2017). Severe iodine deficiency during pregnancy is associated with a greater incidence of stillbirths, abortions, and congenital abnormalities. The fetal brain is particularly vulnerable to iodine deficiency. Normal levels of thyroid hormones are required for neuronal migration and myelination of the central nervous system (Morreale de Escobar, 2004). The most severe form of neurological damage from fetal hypothyroidism is termed cretinism. It is characterized by gross mental retardation along with varying degrees of short stature, deaf mutism, and spasticity. Up to 10% of populations with severe iodine deficiency may be cretinous. Iodine prophylaxis has completely eliminated the appearance of new cases of cretinism in previously iodine-deficient Switzerland and other countries.

Although new cases of cretinism are now rare, iodine deficiency still affects up to 30% of the global population (see below), and can impair cognitive development. A meta-analysis of 18 studies concluded that moderate-to-severe iodine deficiency reduces mean IQ scores by 13.5 points. Iodine deficiency is thus considered one of the most common causes of preventable mental retardation worldwide. Even in areas of mild-to-moderate iodine deficiency, cognitive impairment in school-aged children is at least partially reversible by administration of iodine. Overall, iodine deficiency produces widespread adverse effects in a population, including decreased educability and productivity, resulting in impaired social and economic development.

Only a few countries, including Switzerland, the Scandinavian countries, Australia, the US, and Canada, were completely iodine sufficient before 1990, due to iodized salt programs and adventitious iodine added during processing of foods. Since then, widespread introduction of iodized salt has produced dramatic reductions in iodine deficiency. The Global Scorecard of Iodine Nutrition, published by the Iodine Global Network, is based on the most recent available median urinary iodine concentration data from 194 World Health Organization Member States. In 2020, only 21 countries have insufficient iodine intake, while 118 countries have adequate iodine intake and 13 have excessive iodine intake (Zimmermann and Andersson, 2021).

Iodine requirements

The US Food and Nutrition Board of the National Academy of Sciences has set an Adequate Intake (AI) for iodine in infancy and a Recommended Dietary Allowance (RDA) for children, adults, and pregnant and lactating women (Institute of Medicine, Academy of Sciences, 2001) (**Table 2**). The WHO has established recommended nutrient intakes for iodine (World Health Organization et al., 2007) (**Table 2**).

Assessment of iodine status

Several methods are available for assessment of iodine nutrition. The most commonly used are measurement of thyroid size, concentration of urinary iodine (UI), and serum or dried blood spot thyroglobulin. As discussed below, UI is a sensitive indicator of recent iodine intake (days) and serum Tg shows an intermediate response (weeks to months), whereas changes in the goiter rate reflect long-term iodine nutrition (months to years) (Zimmermann, 2009).

Two methods are available for measuring goiter: neck inspection and palpation, and thyroid ultrasonography. Goiter surveys are usually done in school-aged children. By palpation, a thyroid is considered goitrous when each lateral lobe has a volume greater than the terminal phalanx of the thumbs of the subject being examined. In areas of mild-to-moderate iodine deficiency, where goiters are small, measurement of thyroid size by ultrasonography is a more objective and precise method, and is preferable to

Table 2 Recommendations for iodine intake ($\mu g\ day^{-1}$) by age or population group.

	US Institute of Medicine			*World Health Organization*
Age or population group	*EAR*	*AI or RDA*	*Age or population group*	*RNI*
Infants 0–12 months	–	110–130	Children 0–5 years	90
Children 1–8 years	65	90	Children 6–12 years	120
Children 9–13 years	73	120		
Adults ≥14 years	95	150	Adults >12 years	150
Pregnancy	160	220	Pregnancy	250
Lactation	209	290	Lactation	250

Abbreviations: AI, adequate intake; EAR, estimated average requirement; RDA, recommended daily allowance; RNI, recommended nutrient intake.

Table 3 Epidemiological criteria from the World Health Organization for assessment of iodine nutrition in a population based on median urinary iodine concentrations (World Health Organization et al., 2007).

Urinary iodine concentrations ($\mu g\ L^{-1}$)	*Iodine intake*	*Iodine nutrition*
School-aged children		
<20	Insufficient	Severe iodine deficiency
20–49	Insufficient	Moderate iodine deficiency
50–99	Insufficient	Mild iodine deficiency
100–199	Adequate	Optimum
200–299	More than adequate	Risk of iodine-induced hyperthyroidism in susceptible groups
>300	Excessive	Risk of adverse health consequences (iodine-induced hyperthyroidism, autoimmune thyroid disease)
Pregnant women		
<150	Insufficient	
150–249	Adequate	
250–499	More than adequate	
≥500[a]	Excessive	
Lactating women[b]		
<100	Insufficient	
≥100	Adequate	
Children less than 2 years of age		
<100	Insufficient	
≥100	Adequate	

palpation. Portable ultrasound equipment can be used in the field, and goiter classified according to international reference criteria for iodine-sufficient children by age, gender, and body surface area. The total goiter rate is used to define severity using the following criteria: <5%, iodine sufficiency; 5.0–19.9%, mild deficiency; 20.0–29.9%, moderate deficiency; and >30%, severe deficiency (World Health Organization et al., 2007).

Because >90% of ingested iodine is excreted in the urine, UI is an excellent indicator of recent iodine intake. Most methods of measuring UI are based on the Sandell–Kolthoff reaction, in which iodide catalyzes the reduction of yellow ceric ammonium sulfate to the colorless cerous form, in the presence of arsenious acid. For populations, because it is impractical to collect 24 h samples in field studies, UI can be measured in spot urine specimens from a representative sample of the target group, and expressed as the median, in $\mu g\ L^{-1}$ (**Table 3**). Spot UI measurements from populations are often misinterpreted; it is a common mistake to assume that all subjects with a spot UI $< 100\ \mu g\ L^{-1}$ are iodine deficient, but even in iodine sufficient regions, individual spot UI concentrations are highly variable from day-to-day.

Tg is synthesized only in the thyroid, and is the most abundant intrathyroidal protein. In iodine sufficiency, small amounts of Tg are secreted into the circulation, and serum Tg is normally $<10\ \mu g\ L^{-1}$. In areas of endemic goiter, serum Tg increases due to greater thyroid cell mass and TSH stimulation. Serum Tg is well correlated with the severity of iodine deficiency as measured by UI, and is a sensitive indicator of iodine repletion. Tg can also be assayed on dried blood spots taken by a finger prick, simplifying collection and transport. A reference range for Tg on dried blood spots in school-aged children has recently been published, and Tg is now recommended to assess iodine status.

Prophylaxis and treatment of iodine deficiency

There are two methods commonly used to correct iodine deficiency in a population: iodized oil and iodized salt (Zimmermann, 2009). In nearly all regions affected by iodine deficiency, the most effective way to control iodine deficiency is through salt iodization. All salt for human consumption, including salt used in the food industry, should be continuously iodized. Iodine can be added to salt in the form of potassium iodide (KI) or potassium iodate (KIO_3). Because KIO_3 has higher stability in the presence of salt impurities, humidity, and porous packaging, it is the recommended form. Iodine is usually added at a level of 20–40 mg iodine/kg salt, depending on local salt intake.

However, in industrialized countries, because about 90% of salt consumption is from purchased processed foods, if only household salt is iodized it will not supply adequate iodine. Thus, it is critical that the food industry uses iodized salt. The current push to reduce salt consumption to prevent chronic diseases and the policy of salt iodization to eliminate iodine deficiency do not conflict: iodization methods can fortify salt to provide recommended iodine intakes even if per capita salt intakes are reduced to $<5\ g\ day^{-1}$.

In some regions, iodization of salt may not be practical for control of iodine deficiency, at least in the short term. This may occur in remote areas where communications are poor or where there are numerous small-scale salt producers. In these areas, other options for correction of iodine deficiency should be considered, such as iodized oil (Zimmermann, 2009). Iodized oil is prepared from unsaturated fatty acids in seed or vegetable oils, by addition of iodine to the double bonds. It can be given orally or by

intramuscular injection. The intramuscular route has a longer duration of action (up to 2 years), but oral administration is more common because it is simpler.

Iodized oil is recommended for populations with moderate-to-severe iodine deficiency that do not have access to iodized salt, and may be targeted toward women of child-bearing age, pregnant women, and children. The recommended dose is 400 mg of iodine per year for women and 200 mg of iodine per year for children 7–24 months of age. Iodine can also be given as potassium iodide or iodate as drops or tablets, and in drinking or irrigation water. Iodine supplements ($\sim$150 μg day^{-1}) are recommended for pregnant and lactating women residing in areas of mild-to-moderate iodine deficiency.

Iodine excess and toxicity

Ingestion of excess amount of iodine leads to acute iodine poisoning, which causes gastrointestinal irritation, abdominal pain, nausea, vomiting, and diarrhea, as well as cardiovascular symptoms, coma, and cyanosis. Most people are remarkably tolerant to high dietary intakes of iodine (Farebrother et al., 2019). The US Food and Nutrition Board of the National Academy of Sciences has set a Tolerable Upper Intake Level (UL) for iodine. The UL is the highest level of daily intake that is likely to pose no risk of adverse health effects in almost all individuals. The UL is 200 μg day^{-1} for ages 1–3 years, 300 μg day^{-1} for ages 4–8 years, 600 μg day^{-1} for ages 9–13 years, 900 μg day^{-1} for ages 14–18 years, and 1100 μg day^{-1} thereafter (Institute of Medicine, Academy of Sciences, 2001). Individuals with autoimmune thyroid disease or chronic iodine deficiency may respond adversely to intakes lower than these.

A rapid increase in iodine intake in populations with chronic iodine deficiency may precipitate iodine-induced hyperthyroidism (IIH). This is more likely to occur if the iodine is given in excess, for example, if the iodine content of iodized salt is too high, or when iodine-containing medication is given. IIH occurs mainly in older people with nodular goiter. Thyrocytes in nodules often become insensitive to TSH control, and if iodine supply is suddenly increased, these autonomous nodules may overproduce thyroid hormone (Zimmermann and Boelaert, 2015). Symptoms of IIH include weight loss, tachycardia, and muscle weakness. It is dangerous when superimposed on underlying heart disease, and may be lethal. To reduce risk for IIH, the iodine level in salt should be monitored and reduced if too high.

Conclusions

Because the native iodine content of most foods and beverages is low, the populations of many countries are at risk of iodine deficiency. It is critical to ensure adequate iodine intake in humans, because iodine deficiency has multiple adverse effects on growth and development. In most countries, the most effective and sustainable way to control iodine deficiency is through salt iodization. But salt iodization programs need to be carefully monitored to avoid both iodine deficiency and iodine excess.

References

Farebrother, J., Zimmermann, M.B., Andersson, M., 2019. Excess iodine intake: sources, assessment, and effects on thyroid function. Ann. N. Y. Acad. Sci. 1446, 44–65.

Gowachirapant, S., Jaiswal, N., Melse-Boonstra, A., et al., 2017. Effect of iodine supplementation in pregnant women on child neurodevelopment: a randomized, double-blind, placebo-controlled trial. Lancet Diabetes Endocrinol. 5, 853–863.

Institute of Medicine, Academy of Sciences, 2001. Iodine. Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc. National Academy Press, Washington, DC, pp. 258–289.

Morreale de Escobar, G., Obregon, M.J., Escobar del Rey, F., 2004. Role of thyroid hormone during early brain development. Eur. J. Endocrinol. 151 (Suppl. 3), U25–U37.

World Health Organization, United Nations Children's Fund, International Council for Control of Iodine Deficiency Disorders, 2007. Assessment of Iodine Deficiency Disorders and Monitoring Their Elimination: A Guide for Program Managers, third ed. World Health Organization, Geneva.

Yen, P.M., 2001. Physiological and molecular basis of thyroid hormone action. Physiol. Rev. 81 (3), 1097–1142.

Zimmermann, M.B., Andersson, M., 2021. Global perspectives in endocrinology: coverage of iodized salt programs and iodine status in 2020. Eur. J. Endocrinol. 185, R13–R21.

Zimmermann, M.B., Boelaert, K., 2015. Iodine deficiency and thyroid disorders. Lancet Diabetes Endocrinol. 3, 286–295.

Zimmermann, M.B., 2009. Iodine deficiency. Endocr. Rev. 30 (4), 376–408.

Iodine: Physiology, dietary sources, and requirements

Sonja Y. Hess[a] **and Elizabeth N. Pearce**[b], [a] Department of Nutrition, University of California Davis, Davis, CA, United States; and [b] Section of Endocrinology, Diabetes, and Nutrition, Boston University School of Medicine, Boston, MA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of S.Y. Hess, Iodine: Physiology, Dietary Sources, and Requirements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 33–38, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00162-8.

Introduction 274
Ecology of iodine 274
Absorption and metabolism 274
Adaptations of thyroid metabolism to iodine deficiency 275
Adaptations of thyroid metabolism to iodine excess 276
Impact of other micronutrients on thyroid and iodine metabolism 276
Requirements and dietary sources of iodine 276
Dietary and environmental factors that affect iodine requirements 277
Assessment of iodine status 278
Thyroid volume 278
Urinary iodine concentration 278
Thyroid hormone and thyroglobulin concentrations 278
Conclusions 279
References 279
Relevant websites 281

Key points

- Iodine is essential for the synthesis of thyroid hormone, which regulates growth, development, and metabolism.
- Multiple homeostatic mechanisms help to maintain stable thyroid hormone synthesis, but prolonged exposure to iodine deficiency or excess can lead to thyroid dysfunction.
- Pregnant and lactating women and young children are the groups most vulnerable to the effects of iodine deficiency.
- In addition to iodized salt, the primary means for prevention of iodine deficiency disorders, important dietary iodine sources include seafood and dairy products.
- Iodine deficiency may be exacerbated by concomitant micronutrient deficiencies and by exposure to environmental goitrogens.
- The iodine status of populations can be assessed based on median urinary iodine concentrations, blood thyroglobulin levels, and on prevalence of goiter.

Glossary

Goiter Abnormally enlarged thyroid gland; often resulting from underproduction or overproduction of thyroid hormones or from iodine deficiency

Goitrogens Dietary substances that interfere with thyroid metabolism and aggravate the effects of iodine deficiency

Iodine An essential element in the diet needed by the thyroid gland for the synthesis of thyroid hormones

Thyroglobulin A protein that is the precursor to iodine-containing hormones and is typically present in the colloid of the thyroid gland. Abbreviated as Tg

Thyroid stimulating hormone Anterior pituitary hormone that stimulates the thyroid gland to produce thyroid hormones. Abbreviated as TSH

Thyroxine Hormone containing four atoms of iodine produced by the thyroid gland. Abbreviated as T4

Triiodothyronine Bioactive thyroid hormone with one less iodine atom per molecule than thyroxine. Abbreviated as T3

Introduction

Iodine is an essential component of the hormones produced by the thyroid gland. Inadequate intake of iodine impairs thyroid function and results in a spectrum of disorders, including goiter, adverse obstetric outcomes, impaired cognitive development, and congenital abnormalities, collectively referred to as iodine deficiency disorders (IDDs). One billion people worldwide are at risk of iodine deficiency, with the highest risks in South Asia and sub-Saharan Africa (Zimmermann and Andersson, 2021). In nearly all countries the best strategy to control iodine deficiency is iodization of salt (Gorstein et al., 2020).

Ecology of iodine

Although iodine is widely present in the environment, it is distributed unevenly. The median concentration of iodine in soils worldwide is 5 $\mu g\ g^{-1}$ but can range from 1 to 150 $\mu g\ g^{-1}$ (Steinnes, 2009). The highest amounts are found in soils rich in organic components and located near the coast. Iodine was present during the primordial development of the Earth, but large amounts were leached from the surface soil by glaciation, snow, or rain and were carried by rivers and floods to the ocean. Thus, most of the world's iodine resides in the ocean. The major mechanism of iodine transfer from the ocean to land is based on volatilization of seawater iodine into the atmosphere (Carpenter et al., 2021). Another major source seems to be the release of volatile methyl iodide by marine organisms. However, the return of iodine to soil is insufficient compared to the original loss, leaving soils and drinking water iodine depleted. Iodine-deficient soils are common in mountainous areas (e.g., the Alps, Andes, Atlas, and Himalayan ranges) and areas of frequent flooding, especially in South and Southeast Asia (Zimmermann, 2009). Many inland regions in central Asia and Africa and central and eastern Europe are also affected by iodine-deficient soils. Populations in these areas that depend on locally grown food, consequently, become iodine deficient unless additional iodine is provided through imported iodine-rich foods, iodine fortification, or supplementation.

Absorption and metabolism

Iodine is present in food in different chemical forms. Most ingested iodine is reduced to iodide in the gut and is absorbed almost completely in the duodenum (Alexander et al., 1967). Iodide is cleared from the circulation mainly by the thyroid and kidney. The thyroid adjusts the amount of iodide uptake to amounts required for adequate thyroid hormone synthesis. In conditions of adequate iodine supply, no more than 10% of the absorbed iodine is taken up by the thyroid. In chronic iodine deficiency, this fraction can exceed 80% (DeGroot, 1966). Similarly, during lactation, the mammary gland regulates iodine uptake for secretion in breast milk (Andersson and Braegger, 2021). Several other tissues can also concentrate iodine, including the salivary glands, choroid plexus, and gastric mucosa, but these are minor pathways of uncertain significance. Once the thyroidal iodine requirement has been met, excess iodine is excreted by the kidney.

The body of a healthy adult contains 15–20 mg of iodine, of which 70–80% is in the thyroid (Fisher and Oddie, 1969). In chronic iodine deficiency, the iodine content of the thyroid can fall below 20 μg. In iodine-sufficient areas, the adult thyroid traps approximately 60 μg of iodine per day to balance losses and maintain thyroid hormone synthesis. A transmembrane protein, the sodium/iodide symporter (NIS), mediates the first and key step in the process of supplying iodide to the gland. This protein transfers iodide into the cytoplasm of the follicular cells at a concentration gradient 20–50 times that of serum (Eskandari et al., 1997). Iodine in the thyroid gland then participates in a complex series of reactions to produce thyroid hormones. Iodide must first be oxidized to a higher oxidation state before it can act as an effective iodinating agent. Only H_2O_2 is sufficiently potent to oxidize iodide. At the apical membrane, thyroid peroxidase (TPO) catalyzes the iodination of tyrosyl residues of thyroglobulin (Tg) producing either monoiodotyrosine (MIT) or diiodotyrosine (DIT), the precursors of thyroid hormones (Dunn, 1995). TPO then catalyzes the coupling of the phenyl groups of the iodotyrosines through a diether bridge to form the thyroid hormones. Two residues of DIT couple within Tg to form thyroxine (T4), or one DIT and one MIT to form triiodothyronine (T3). In the thyroid, mature Tg, containing 0.1–1.0% of its weight as iodine, is stored extracellularly in the luminal colloid of the thyroid follicle. Approximately one-third of Tg's iodine is in T4 and T3, the remainder being stored in the inactive precursors DIT and MIT. The iodine in the inactive precursors is not released into circulation, but instead is removed from the tyrosine moiety by a specific deiodinase and then recycled within the thyroid gland. This process is particularly important for iodine conservation in situations of iodine deprivation (Fig. 1).

Before secretion from the thyroid, T4 and T3 must be released from peptide linkage within Tg. Tg retrieved by micropinocytosis passes first through the endosomal system, where proteolysis and hormone release is initiated, then into lysosomes, where the process is completed and T4 and T3 are released into the circulation. T4 is secreted in higher quantities from the thyroid compared with T3. Once in the circulation, T4 and T3 rapidly attach to several binding proteins synthesized in the liver, including thyroxine-binding globulin, transthyretin, and albumin. The bound hormone then migrates to target tissues where T4 is deiodinated to T3, the metabolically active form.

T3 is essential for the development and differentiation of all cells of the human body. It acts mostly through nuclear receptors regulating gene expression, although other mechanisms have also been described (Mendoza and Hollenberg, 2017). T3 plays a key role in normal skeletal development, linear growth, and the acquisition and maintenance of bone mass. T3 also has a critical role in

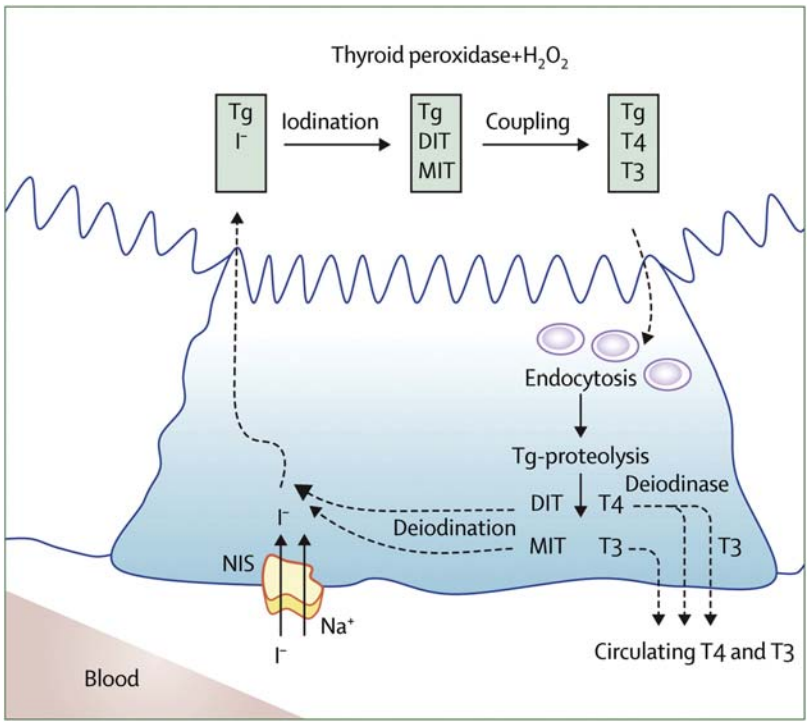


Fig. 1 Iodine pathway in the thyroid cell. Iodide (I^-) is transported into the thyrocyte by the sodium/iodide symporter (NIS) at the basal membrane and migrates to the apical membrane. I^- is oxidized by the enzymes thyroid peroxidase (TPO) and hydrogen peroxidase (H_2O_2) and attached to tyrosyl residues in thyroglobulin (Tg) to produce the hormone precursors monoiodotyrosine (MIT) and diiodotyrosine (DIT). Residues then couple to form thyroxine (T4) and triiodothyronine (T3) within the Tg molecule in the follicular lumen. Tg enters the cell by endocytosis and is digested. T4 and T3 are released into the circulation, and iodine on MIT and DIT is recycled within the thyrocyte. Reprinted with permission from Zimmermann, M.B., Jooste, P.L., Pandav, C.S., 2008. Iodine-deficiency disorders. Lancet 372: 1251–1262.

the development and function of the human central nervous system. Moreover, thyroid hormones are the major endocrine regulators of basal metabolic rate. Hypothyroidism results from insufficient production of thyroid hormones, whereas hyperthyroidism is due to excessive synthesis and secretion of thyroid hormones.

Adaptations of thyroid metabolism to iodine deficiency

When dietary iodine intake is low, the thyroid may still achieve adequate secretion of thyroid hormones by modifications of metabolism. Thyroid stimulating hormone (TSH) is the primary factor that regulates the function of thyroid follicular cells and, ultimately, thyroid hormone secretion. In a classic negative feedback system, thyroid hormone inhibits the synthesis of TSH directly at the pituitary level and indirectly at the hypothalamic level by reducing the secretion of thyrotropin-releasing hormone (TRH). TSH stimulates the trapping of iodine into the thyroid, the breakdown of Tg, and an increase in the ratio of T3 to T4 synthesized and released into the blood (Abrams and Larsen, 1973). As a greater fraction of circulating iodine enters the thyroid, renal iodide excretion is reduced. As long as the iodine intake remains above approximately 50 μg day^{-1}, the absolute uptake of iodide by the thyroid remains normal and the iodine content of the thyroid will remain stable at about 15–20 mg, despite the decrease in serum iodine concentration (Zimmermann, 2009). Below this threshold, there is an increased risk of iodine depletion in the thyroid, and individuals may develop goiter.

The basic process in the transformation of the normal thyroid to a goiter is the generation of new thyrocytes and follicles (hyperplasia) in addition to increasing cell volume (hypertrophy). The optimal thyroid response to iodine deficiency would be an increase in thyroid blood flow, in iodide trapping capacity, and in iodination rate, and a low Tg content in a much reduced colloid space. However, the goitrous thyroid is often large and filled with colloid. The low iodine and high Tg concentration lead to a lesser iodination of Tg, thus reducing the efficiency of thyroid hormone synthesis (Dumont et al., 1995). Iodine deficiency-induced goiter typically is initially diffuse, but over time somatic mutations may lead to the development of nodularity and then to autonomously-functioning nodules (Krohn et al., 2005). In areas of moderate-to-severe iodine deficiency the characteristic pattern of circulating thyroid hormones is a variably elevated TSH, a low serum T4, and a normal or mildly increased T3 concentration. Overt hypothyroidism (elevated serum TSH and low T4) and cretinism usually develop only in regions of chronic, severe iodine deficiency. The effects of iodine deficiency on the development of goiter and thyroid function are extremely variable among

populations and individuals, even in endemic areas. The dietary, environmental, and genetic factors that account for this variability remain largely unknown.

Adaptations of thyroid metabolism to iodine excess

Intakes up to 600 μg day^{-1} in the European Union and 1100 μg day^{-1} in the United States are defined as safe for adults (European Food Safety Authority Scientific Committee on Food, 2002; US Institute of Medicine, 2001). Nevertheless, most individuals tolerate higher intakes, whereas a few may have untoward effects at lower intakes. The NIS is key in maintaining normal thyroid hormone concentration as it reduces the transport of iodide into the thyroid cells under conditions of excessive iodine intake. Even before NIS reacts, a sudden iodine overload paradoxically blocks the second step of hormone synthesis, the organification of iodide. This so-called acute Wolff–Chaikoff effect requires a high ($\geq 10^{-3}$ M) intracellular concentration of iodide and provides a short-term block against further thyroidal iodine uptake (Wolff and Chaikoff, 1948). After a few days of continued exposure to high iodine levels there is an escape from the acute Wolff-Chaikoff effect mediated by downregulation of NIS which lowers intrathyroidal iodine content and allows normal thyroid hormone synthesis to resume (Eng et al., 1999). Another response to iodine excess occurs in the first step of thyroid hormone synthesis (Taurog and Nakashima, 1978). When iodine is abundant, DIT predominates over MIT in favor of T4 biosynthesis, which is less active than T3. Thus, a euthyroid state is maintained despite an increased amount of iodine taken up by the gland.

There are no clinical symptoms specific to these adaptations of excessive intake. Although most individuals suffer no disturbance from iodine excess, some persons develop thyroid dysfunction despite the multiple control systems (Leung and Braverman, 2014). Iodine excess may cause hyperthyroidism (from a failure of the acute Wolff-Chaikoff effect), hypothyroidism (from a failure of the escape from the acute Wolff-Chaikoff effect), euthyroid goiter, or thyroid autoimmunity. Factors responsible for this variety of responses are mostly unknown.

Impact of other micronutrients on thyroid and iodine metabolism

Besides iodine, other highly prevalent micronutrient deficiencies, such as deficiencies of iron, selenium, and vitamin A, adversely affect thyroid function. Although there is little information on the prevalence of concomitant iodine deficiency with iron, selenium, or vitamin A deficiency, in view of the high prevalence of these individual micronutrient deficiencies in low- and middle-income countries, it is highly likely that a substantial number of individuals may be affected by multiple micronutrient deficiencies.

Numerous studies in animals have shown that iron deficiency anemia impairs thyroid metabolism (Zimmermann, 2006). Iron deficiency anemia may influence thyroid metabolism by altering the central nervous system control and reducing TPO activity. Iron deficiency anemia could also impair thyroid metabolism through lowered oxygen transport. It is likely that these mechanisms jointly contribute to the impairment of thyroid function in iron deficiency. There is strong evidence for such an interaction between iron and iodine and thyroid metabolism from randomized, controlled intervention trials in humans, which have repeatedly shown that providing iron along with iodine, either as an iron supplement or as dual-fortified salt, results in significantly greater improvements of thyroid metabolism (Zimmermann et al., 2002; Eftekhari et al., 2006).

Selenium is an integral component of two important enzymes—iodothyronine deiodinase and glutathione peroxidase (GPX)—that are present in many tissues, including the thyroid gland. Briefly, there are three types of deiodinases. Two 5′-deiodinases (DIO1 and DIO2) catalyze the activation of the prohormone T4 to the active thyroid hormone T3. DIO1 is also involved in the degradation to the inactive hormone reverse T3 (Bianco and da Conceição, 2018). The third selenocysteine-containing deiodinase (DIO3) inactivates thyroid hormones, both the prohormone T4 and its active metabolites such as T3 and 3,5-T_2. GPX and thioredoxin reductase are expressed in thyroid tissue and protect the thyroid gland from hydrogen peroxide produced during the synthesis of thyroid hormone, thereby protecting against oxidative damage (Duntas, 2010). In conditions of inadequate supply of both iodide and selenium, complex rearrangements of thyroid hormone metabolism enable adaptation by increasing retention of selenium in the brain, endocrine tissues, and especially in the thyroid gland. However, to date, most controlled intervention trials in humans have failed to confirm an effect of selenium supplementation on thyroid metabolism (Winther et al., 2020).

Vitamin A deficiency has multiple effects on the pituitary–thyroid axis. Vitamin A modulates thyroid hormone metabolism in the thyroid gland and the periphery, and the production of TSH by the pituitary. Similarly to iron, but with slightly less evidence from randomized clinical trials in children, vitamin A supplementation may provide a beneficial impact on thyroid metabolism, either when given alone or in combination with iodized salt (Hess, 2010).

Requirements and dietary sources of iodine

Recommendations for iodine intake by age and population group, as defined by the US Institute of Medicine (IOM) and the World Health Organization (WHO), are shown in **Table 1** (US Institute of Medicine, 2001; World Health Organization, United Nation Children's Fund and Intrnational Council for the Control of Iodine Deficiency, 2007). Infants are born with minimal intrathyroidal

Table 1 Recommendations for iodine intake (μg day^{-1}) by age and population group.

Age or population group	*U.S. Institute of Medicine*[a] *(μg day^{-1})*	*Age or population group*	*World Health Organization*[b] *(μg day^{-1})*
Infants 0–12 months	110–130	Children 0–5 years	90
Children 1–8 years	90	Children 6–12 years	120
Children 9–13 years	120		
Adults ≥14 years	150	Adults ≥12 years	150
Pregnant women	220	Pregnant women	250
Lactating women	290	Lactating women	250

[a]Adequate intake for infants ≤12 months; recommended daily allowance for individuals >1 year.
[b]Recommended nutrient intake.
Data Sources: U.S. Institute of Medicine Panel on Micronutrients, 2001. Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc. National Academies Press, Washington, DC; World Health Organization, United Nation Children's Fund, International Council for the Control of Iodine Deficiency Disorders, 2007. Assessment of Iodine Deficiency Disorders and Monitoring Their Elimination. A Guide for Program Managers. Geneva: World Health Organization.

iodine stores and require daily iodine intakes of about 80 μg to sustain a high rate of thyroid hormone synthesis (Dold et al., 2016). Compared to nonpregnant adults, pregnant women need increased iodine intakes because of a physiologic 50% increase in thyroid hormone synthesis starting in early gestation, transfer of some iodine to the fetus in the second half of gestation, when the fetal thyroid gland is functional, and because of increased renal iodine losses (Rodriguez-Diaz and Pearce, 2020). During lactation, as described above, iodine is actively secreted into breast milk, leading to increased daily iodine requirements to support the nutritional needs of both the mother and the breastfed infant. The most vulnerable population groups for iodine deficiency are young children and pregnant/lactating women due to their increased physiological needs and due to the importance of adequate iodine for neurodevelopment *in utero* and in early life.

The native iodine content of most foods and beverages is low. Most foods consumed contain 3–80 μg per serving (Zimmermann, 2009). Although plant-based foods are generally low in iodine, iodine content is affected by the iodine in soil and irrigation water, and by the use of fertilizers (Ershow et al., 2018). Seafood and seaweed are generally rich sources of iodine. However, the iodine content of fish varies greatly depending on the water they inhabit. In countries such as Japan and Iceland, where there is a high consumption of seafood, some population groups have been found to consume excessive amounts of iodine (Laurberg et al., 1991; Fuse et al., 2021). In some regions of China excessive iodine levels occur naturally in drinking water (Zhao et al., 1998). Due to iodine in cattle feed and the use of iodophor cleansers by the dairy industry, cows' milk may be high in iodine and in many parts of the world dairy products are an important dietary source (van der Reijden et al., 2017).

The most important source of dietary iodine in most countries is iodized salt. The addition of iodine to salt is the most effective way to control iodine deficiency, and WHO, the United Nations Children's Fund (UNICEF), and the Iodine Global Network (IGN) recommend that salt in regions of iodine deficiency should be fortified at a concentration of 20–40 mg iodine per kg salt, depending on local salt intake (World Health Organization, United Nation Children's Fund and Intrnational Council for the Control of Iodine Deficiency, 2007). Since 1990, the proportion of households worldwide using iodized salt has risen from less than 20%–88%, thereby markedly reducing the problem of iodine deficiency (Zimmermann and Andersson, 2021).

The iodine content of breast milk varies depending on maternal iodine status and dietary intake (Andersson and Braegger, 2021). Colostrum contains the greatest amount of iodine, with concentrations as high as 200–400 μg L^{-1}. In mature breast milk, median iodine concentrations of 100–200 μg L^{-1} are consistent with adequate maternal iodine status.

Non-food sources of iodine include topical antiseptics (povidone-iodine), iodinated contrast media used for computed tomography scans and other imaging, and the antiarrhythmic medication amiodarone (Leung and Braverman, 2014). Annual dosing with 400 μg iodized oil supplements may occasionally be employed for vulnerable populations in regions of severe iodine deficiency (World Health Organization, United Nation Children's Fund and Intrnational Council for the Control of Iodine Deficiency, 2007). Lower-dose daily oral iodine supplements are also available in many regions (Rodriguez-Diaz and Pearce, 2020).

Dietary and environmental factors that affect iodine requirements

Goitrogens are dietary substances that interfere with thyroid metabolism and can aggravate the effect of iodine deficiency. Most goitrogens do not have a major clinical effect unless they are consumed at high levels and iodine deficiency is present. Vegetables of the Brassica family (i.e., broccoli, cabbage, cauliflower, kale, turnips, rapeseed) contain glucosinolates, which are potent goitrogenic substances (Felker et al., 2016). The metabolites of glucosinolates compete with iodine for thyroidal uptake. More important, however, are the naturally occurring goitrogens cyanoglucosides in several staple foods, such as cassava, maize, bamboo shoots, sweet potatoes, and lima beans (Ermans et al., 1972). Cyanoglucosides are metabolized to thiocyanates which are anions that compete with iodine in thyroid hormone synthesis. Flavonoids in millet and soy may impair TPO activity, which has raised

concerns about potential adverse effects of soy-based infant formulas on thyroid function of young children. However, evidence from clinical trials on soy consumption remains inconclusive (Otun et al., 2019).

Some substances that are commonly found in the environment may also affect thyroid function (Pearce and Braverman, 2009). The anions perchlorate, thiocyanate, and nitrate are competitive inhibitors of NIS at pharmacological doses. When present in high amounts, these substances can decrease the active transport of iodine into the thyroid and thereby reduce thyroid hormone synthesis. The most vulnerable population groups are the developing fetus and the newborn, as sufficient iodine is essential for their normal thyroid function at this crucial time of neurodevelopment. A low level of perchlorate exposure appears to be ubiquitous. However, the potential effects of environmental perchlorate exposure on thyroid function remain controversial, and more research is needed. Cigarette smoke contains cyanide that is metabolized to thiocyanate. Smoking has been shown to adversely affect thyroid hormone status and iodine concentration in breast milk (Laurberg et al., 2004). Nitrates occur naturally in soil, groundwater, and plants, and sodium nitrite is used as a preservative in cured meats and other foods. Studies in areas of very high nitrate contamination of water have found an increased risk of goiter or hypothyroidism (García Torres et al., 2020). Multiple other environmental substances are known to have adverse effects on thyroid hormone synthesis, metabolism, and action and may worsen the effects of iodine deficiency (Köhrle and Frädrich, 2021). Continued monitoring of chemical exposures is important to detect potential thyroidal inhibitors as industrial practices and governmental regulations change over time.

Assessment of iodine status

The assessment of population iodine status is important to instigate salt iodization programs, guide program modifications, and contribute to greater safety of these programs (Brown et al., 2021). Four indicators are generally recommended for assessment of iodine status: thyroid volume, urinary iodine, serum TSH, and serum Tg concentrations. Thyroid volume reflects long-term iodine nutrition. Urinary iodine concentration is an indicator of recent and Tg concentration of medium-term iodine intake. More details are provided in the following sections.

Thyroid volume

As described in the section on Adaptations of Thyroid Metabolism to Iodine Deficiency above, increased thyroid volume is a common consequence of long-term iodine deficiency. Thyroid volume can be determined by neck inspection and palpation or by ultrasonography. Goiter surveys are usually done in school-aged children. By palpation, a thyroid is considered goitrous when each lateral lobe has a volume greater than the terminal phalanx of the thumbs of the individual being examined. Because palpation has poor sensitivity and specificity in areas of mild iodine deficiency, measurements of thyroid volume by ultrasound is preferable. Thyroid ultrasound is non-invasive, quick, and feasible even in remote areas using portable equipment. Thyroid volume can be classified as goiter according to international reference criteria (Zimmermann et al., 2014). Although thyroid size decreases in children in response to iodine repletion, thyroid size may not return to normal values for months or years, if at all, after correction of iodine deficiency. Therefore, because of the lack of sensitivity to acute changes in iodine intake, this method is of limited usefulness in assessing the impact of salt iodization programs (United Nations Children's Fund, 2018).

Urinary iodine concentration

Because the absorption of dietary iodine is high and approximately 90% of iodine consumed is excreted in urine, the urinary iodine concentration serves as a good reflection of iodine nutrition (Rohner et al., 2014). Urinary iodine concentration can be measured in spot urine specimens from a representative sample of the target group, and expressed as the median, in $\mu g\ L^{-1}$. Because there is substantial day-to-day and hour-to-hour variability in urinary iodine excretion, urinary iodine concentrations cannot be used as a biomarker for the iodine status of individuals (König et al., 2011). For program evaluation, urinary iodine concentration is the most useful indicator because it reflects the current dietary iodine intake, and WHO and UNICEF have proposed cut-off points for classifying population iodine nutrition into different degrees of public health significance (**Table 2**) (World Health Organization, United Nation Children's Fund and Intrnational Council for the Control of Iodine Deficiency, 2007; United Nations Children's Fund, 2018). Surveys to assess iodine deficiency have often relied on school-aged children because of their vulnerability and easy access in schools. However, children's iodine status does not always reflect iodine status in pregnant women and thus may result in underestimating the risk of iodine deficiency in the most vulnerable populations (Wong et al., 2011). In lactating women, although the WHO has provided thresholds for median urinary iodine, iodine status cannot be understood without also assessing breast milk iodine content (Dold et al., 2017).

Thyroid hormone and thyroglobulin concentrations

Thyroid hormone concentrations are poor indicators of iodine status due to the strong regulatory adaptations of thyroid metabolism. In iodine-deficient populations, serum T3 increases or remains unchanged, and serum T4 usually decreases. However, these values are often within the normal range and overlap with iodine-sufficient populations (Rohner et al., 2014). The use of blood TSH concentration is not recommended as a biomarker for iodine status in school-age children or adults (Pearce and Caldwell, 2016).

Table 2 Epidemiological criteria for assessing population iodine nutrition based on median urinary iodine concentrations in different population groups.

Median urinary iodine ($\mu g\ L^{-1}$)	*Iodine intake*	*Iodine status*
School-aged children		
<20	Insufficient	Severe iodine deficiency
20–49	Insufficient	Moderate iodine deficiency
50–99	Insufficient	Mild iodine deficiency
100–299	Adequate	Adequate iodine nutrition
≥300	Excessive	Risk of adverse health consequences
Pregnant women		
<150	Insufficient	
150–249	Adequate	
250–499	Above requirements	
≥500	Excessive	
Lactating women[a]		
<100	Insufficient	
≥100	Adequate	

[a]In lactating women, the cut-offs for median urinary iodine concentrations are lower than the iodine requirement because of the iodine excretion in breastmilk. Breast milk as well as urinary iodine concentration measurements may be needed to assess the iodine status of lactating women.
Data sources: World Health Organization, United Nation Children's Fund, International Council for the Control of Iodine Deficiency Disorders, 2007. Assessment of Iodine Deficiency Disorders and Monitoring Their Elimination. A Guide for Program Managers. Geneva: World Health Organization; United Nations Children's Fund, 2018. Guidance on the Monitoring of Salt Iodization Programs and Determination of Population Iodine Status. https://www.unicef.org/nutrition/files/Monitoring-of-Salt-Iodization.pdf.

Although serum TSH increases slightly in iodine deficiency, the concentrations often stay within the normal range. In contrast, neonatal TSH concentration reflects iodine nutrition more accurately. TSH increases when the supply of thyroid hormone and iodine from the maternal circulation to the fetus has been compromised. WHO suggests that when a sensitive assay is used on samples collected during the first few days of life, a <3% frequency of TSH concentrations >5 mIU L^{-1} indicates iodine sufficiency in a population (World Health Organization, United Nation Children's Fund and Intrnational Council for the Control of Iodine Deficiency, 2007).

When iodine intake is adequate, small amounts of Tg are secreted into the circulation. In iodine deficiency, thyroid hyperplasia and goiter increase serum Tg levels. Serum Tg levels are also increased in the setting of excessive iodine intake. Tg has been found to be a good indicator of iodine status reflecting iodine nutrition over months or years. Tg can be assessed in serum or whole blood stored on filter paper (Stinca et al., 2015; Zimmermann et al., 2013). Median thyroglobulin values <13 μg L^{-1} with <3% of thyroglobulin values >40 μg L^{-1} indicate optimal iodine nutrition in school-aged children.

Conclusions

Adequate iodine intake is essential for the synthesis of thyroid hormone. Although multiple homeostatic mechanisms help to maintain stable thyroid hormone synthesis, prolonged exposure to either iodine deficiency or excess can lead to thyroid dysfunction. The most vulnerable population groups for iodine deficiency are young children and pregnant and lactating women due to their increased physiological needs and the importance of iodine for normal brain development. Iodized salt is the primary strategy to prevent iodine deficiency disorders, and seafood and dairy products are important dietary sources of iodine. Iodine deficiency may be exacerbated by concomitant micronutrient deficiencies and by exposure to environmental goitrogens. There is currently no validated biomarker for the iodine status of individual patients. The iodine status of populations is most frequently assessed based on median urinary iodine concentrations and blood thyroglobulin levels.

References

Abrams, G.M., Larsen, P., 1973. Triiodothyronine and thyroxine in the serum and thyroid glands of iodine-deficient rats. J. Clin. Invest. 52, 2522–2531.
Alexander, W.D., Harden, R.M., Harrison, M.T., Shimmins, J., 1967. Some aspects of the absorption and concentration of iodide by the alimentary tract in man. Proc. Nutr. Soc. 126, 62–66.
Andersson, M., Braegger, C.P., 2021. The role of iodine for thyroid function in lactating women and infants. Endocr. Rev. https://doi.org/10.1210/endrev/bnab029. PMID: 34791141.
Bianco, A.C., da Conceição, R.R., 2018. The deiodinase trio and thyroid hormone signaling. Methods Mol. Biol. 180, 67–83.

Brown, K.H., Moore, S.E., Hess, S.Y., et al., 2021. Increasing the availability and utilization of reliable data on population micronutrient (MN) status globally: the MN Data Generation Initiative. Am. J. Clin. Nutr. 114, 862–870.
Carpenter, L.J., Chance, R.J., Sherwen, T., et al., 2021. Marine iodine emissions in a changing world. Proc. Roy. Soc. A 477, 20200824.
DeGroot, L.J., 1966. Kinetic analysis of iodine metabolism. J. Clin. Endocrinol. Metabol. 26, 149–173.
Dold, S., Zimmermann, M.B., Baumgartner, J., et al., 2016. A dose-response crossover iodine balance study to determine iodine requirements in early infancy. Am. J. Clin. Nutr. 104, 620–628.
Dold, S., Zimmermann, M.B., Aboussad, A., et al., 2017. Breast milk iodine concentration is a more accurate biomarker of iodine status than urinary iodine concentration in exclusively breastfeeding women. J. Nutr. 147, 528–537.
Dumont, J.E., Ermans, A.M., Maenhaut, C., Coppée, F., Stanbury, J.B., 1995. Large goitre as a maladaptation to iodine deficiency. Clin. Endocrinol. 43, 1–10.
Dunn, J.T., 1995. Thyroglobulin, hormone synthesis and thyroid disease. Eur. J. Endocrinol. 132, 603–604.
Duntas, L.H., 2010. Selenium and the thyroid: a close-knit connection. J. Clin. Endocrinol. Metabol. 95, 5180–5188.
Eftekhari, M.H., Simondon, K.B., Jalali, M., et al., 2006. Effects of administration of iron, iodine and simultaneous iron-plus-iodine on the thyroid hormone profile in iron-deficient adolescent Iranian girls. Eur. J. Clin. Nutr. 60, 545–552.
Eng, P.H., Cardona, G.R., Fang, S.L., et al., 1999. Escape from the acute Wolff-Chaikoff effect is associated with a decrease in thyroid sodium/iodide symporter messenger ribonucleic acid and protein. Endocrinology 140, 3404–3410.
Ermans, A.M., Delange, F., Van der Velden, M., Kinthaert, J., 1972. Possible role of cyanide and thiocyanate in the etiology of endemic cretinism. Adv. Exp. Med. Biol. 30, 455–486.
Ershow, A.G., Skeaff, S.A., Merkel, J.M., Pehrsson, P.R., 2018. Development of databases on iodine in foods and dietary supplements. Nutrients 10, 100.
Eskandari, S., Loo, D.D., Dai, G., et al., 1997. Thyroid Na+/I- symporter. Mechanism, stoichiometry, and specificity. J. Biol. Chem. 272, 27230–27238.
European Food Safety Authority (EFSA) Scientific Committee on Food, 2002. Opinion of the Scientific Committee on Food on the Tolerable Upper Intake Level of Iodine. https://www.efsa.europa.eu/sites/default/files/efsa_rep/blobserver_assets/ndatolerableuil.pdf.
Felker, P., Bunch, R., Leung, A.M., 2016. Concentrations of thiocyanate and goitrin in human plasma, their precursor concentrations in brassica vegetables, and associated potential risk for hypothyroidism. Nutr. Rev. 74, 248–258.
Fisher, D.A., Oddie, T.H., 1969. Thyroid iodine content and turnover in euthyroid subjects: validity of estimation of thyroid iodine accumulation from short-term clearance studies. J. Clin. Endocrinol. Metabol. 29, 721–727.
Fuse, Y., Ito, Y., Shishiba, Y., Irie, M., 2021. Current iodine status in Japan: a cross-sectional nationwide survey of schoolchildren, 2014–2019. J. Clin. Endocrinol. Metabol. https://doi.org/10.1210/clinem/dgab919. PMID: 34935958.
García Torres, E., Pérez Morales, R., González Zamora, A., et al., 2020. Consumption of water contaminated by nitrate and its deleterious effects on the human thyroid gland: a review and update. Int. J. Environ. Health Res. 31, 1–18.
Gorstein, J.L., Bagriansky, J., Pearce, E.N., Kupka, R., Zimmermann, M.B., 2020. Estimating the health and economic benefits of universal salt iodization programs to correct iodine deficiency disorders. Thyroid 30, 1802–1809.
Hess, S.Y., 2010. The impact of common micronutrient deficiencies on iodine and thyroid metabolism: the evidence from human studies. Best Pract. Res. Clin. Endocrinol. Metabol. 24, 117–132.
Köhrle, J., Frädrich, C., 2021. Thyroid hormone system disrupting chemicals. Best Pract. Res. Clin. Endocrinol. Metabol. 35, 101562.
König, F., Andersson, M., Hotz, K., Aeberli, I., Zimmermann, M.B., 2011. Ten repeat collections for urinary iodine from spot samples or 24-hour samples are needed to reliably estimate individual iodine status in women. J. Nutr. 141, 2049–2054.
Krohn, K., Führer, D., Bayer, Y., et al., 2005. Molecular pathogenesis of euthyroid and toxic multinodular goiter. Endocr. Rev. 26, 504–524.
Laurberg, P., Pedersen, K.M., Vestergaard, H., Sigurdsson, G., 1991. High incidence of multinodular toxic goiter in the elderly population in a low iodine intake area vs. high incidence of Graves' disease in the young in a high iodine intake area: comparative surveys of thyrotoxicosis epidemiology in East-Jutland Denmark and Iceland. J. Intern. Med. 229, 415–420.
Laurberg, P., Nøhr, S.B., Pedersen, K.M., Fuglsang, E., 2004. Iodine nutrition in breast-fed infants is impaired by maternal smoking. J. Clin. Endocrinol. Metabol. 89, 181–187.
Leung, A.M., Braverman, L.E., 2014. Consequences of excess iodine. Nat. Rev. Endocrinol. 10, 136–142.
Mendoza, A., Hollenberg, A.N., 2017. New insights into thyroid hormone action. Pharmacol. Therapeut. 173, 135–145.
Otun, J., Sahebkar, A., Östlundh, L., Atkin, S.L., Sathyapalan, T., 2019. Systematic review and meta-analysis on the effect of soy on thyroid function. Sci. Rep. 9, 3964.
Pearce, E.N., Braverman, L.E., 2009. Environmental pollutants and the thyroid. Best Pract. Res. Clin. Endocrinol. Metabol. 23, 801–813.
Pearce, E.N., Caldwell, K.L., 2016. Urinary iodine, thyroid function, and thyroglobulin as biomarkers of iodine status. Am. J. Clin. Nutr. 104 (Suppl. 3), 898S–901S.
Rodriguez-Diaz, E., Pearce, E.N., 2020. Iodine status and supplementation before, during, and after pregnancy. Best Pract. Res. Clin. Endocrinol. Metabol. 34, 101430.
Rohner, F., Zimmermann, M., Jooste, P., et al., 2014. Biomarkers of nutrition for development–iodine review. J. Nutr. 144, 1322S–1342S.
Steinnes, E., 2009. Soils and geomedicine. Environ. Geochem. Health 31, 523–535.
Stinca, S., Andersson, M., Erhardt, J., Zimmermann, M.B., 2015. Development and validation of a new low-cost enzyme-linked immunoassay for serum and dried blood spot thyroglobulin. Thyroid 25, 1297–1305.
Taurog, A., Nakashima, T., 1978. Dissociation between degree of iodination and iodoamino acid distribution in thyroglobulin. Endocrinology 103, 633–640.
U.S. Institute of Medicine Panel on Micronutrients, 2001. Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc. National Academies Press, Washington, DC.
United Nations Children's Fund, 2018. Guidance on the Monitoring of Salt Iodization Programs and Determination of Population Iodine Status. https://www.unicef.org/nutrition/files/Monitoring-of-Salt-Iodization.pdf.
van der Reijden, O.L., Zimmermann, M.B., Galetti, V., 2017. Iodine in dairy milk: sources, concentrations and importance to human health. Best Pract. Res. Clin. Endocrinol. Metabol. 31, 385–395.
Winther, K.H., Rayman, M.P., Bonnema, S.J., Hegedüs, L., 2020. Selenium in thyroid disorders - essential knowledge for clinicians. Nat. Rev. Endocrinol. 6, 165–176.
Wolff, J., Chaikoff, I.L., 1948. The inhibitory action of excessive iodide upon the synthesis of diiodotyrosine and of thyroxine in the thyroid gland of the normal rat. Endocrinology 43, 174–179.
Wong, E.M., Sullivan, K.M., Perrine, C.G., Rogers, L.M., Peña-Rosas, J.P., 2011. Comparison of median urinary iodine concentration as an indicator of iodine status among pregnant women, school-age children, and nonpregnant women. Food Nutr. Bull. 32, 206–212.
World Health Organization, United Nation Children's Fund, International Council for the Control of Iodine Deficiency Disorders, 2007. Assessment of Iodine Deficiency Disorders and Monitoring Their Elimination. A Guide for Program Managers. World Health Organization, Geneva.
Zhao, J., Chen, Z., Maberly, G., 1998. Iodine-rich drinking water of natural origin in China. Lancet 352, 2024.
Zimmermann, M.B., Andersson, M., 2021. Global endocrinology: global perspectives in endocrinology: coverage of iodized salt programs and iodine status in 2020. Eur. J. Endocrinol. 185, R13–R21.
Zimmermann, M.B., Zeder, C., Chaouki, N., et al., 2002. Addition of microencapsulated iron to iodized salt improves the efficacy of iodine in goitrous, iron-deficient children: a randomized, double-blind, controlled trial. Eur. J. Endocrinol. 147, 747–753.
Zimmermann, M.B., Aeberli, I., Andersson, M., et al., 2013. Thyroglobulin is a sensitive measure of both deficient and excess iodine intakes in children and indicates no adverse effects on thyroid function in the UIC range of 100–299 μg/L: a UNICEF/ICCIDD study group report. J. Clin. Endocrinol. Metabol. 98, 1271–1280.

Zimmermann, M.B., Hess, S.Y., Molinari, L., et al., 2014. New reference values for thyroid volume by ultrasound in iodine-sufficient schoolchildren: a World Health Organization/Nutrition for Health and Development Iodine Deficiency Study Group Report. Am. J. Clin. Nutr. 79, 231–237.
Zimmermann, M.B., 2006. The influence of iron status on iodine utilization and thyroid function. Annu. Rev. Nutr. 26, 367–389.
Zimmermann, M.B., 2009. Iodine deficiency. Endocr. Rev. 30, 376–408.

Relevant websites

Global Fortification Data Exchange, https://fortificationdata.org/.
Iodine Global Network, www.ign.org.

Iron: Physiology, dietary sources, and requirements

B. Lönnerdal, University of California, Davis, CA, USA
O. Hernell, Umeå University, Umeå, Sweden

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of B. Lönnerdal, O. Hernell, Iron: Physiology, Dietary Sources, and Requirements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 39–46, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00164-1.

Iron Chemistry and Physiology **282**
Body Content, Forms, and Function 282
Absorption, Excretion, Transport, and Storage 283
Absorption 283
Transport 285
Storage 285
Excretion 285
Body Iron Balance 286
Clinical Assessment of Iron Status 287
Iron Nutrition **287**
Iron Deficiency 287
Consequences of Iron Deficiency 287
Physical Work Capacity 287
Cognitive Development 287
Reproduction 288
Other 288
Recommended Dietary Intakes 288
Dietary Iron **288**
Food Sources 288
Bioavailability 289
Heme Iron 289
Nonheme Iron 289
Supplementation and Fortification 289
Excessive Intakes 290
Further Reading **290**

Iron Chemistry and Physiology

Body Content, Forms, and Function

Iron, the 26th element of the periodic table, has an atomic weight of 55.85. Two common aqueous oxidation states, ferrous (Fe^{2+}) and ferric (Fe^{3+}), enable iron to participate in oxidation/reduction reactions that are essential to energy metabolism by accepting or donating electrons. However, this property also enables free iron to catalyze oxidative reactions, resulting in reactive and damaging free radicals. Accordingly, body iron must be chemically bound to facilitate appropriate physiological function, transport, and storage, with minimal opportunity for free ionic iron to catalyze harmful oxidative reactions.

Most of the body's iron functions in heme protein complexes that transport oxygen as hemoglobin and myoglobin. Approximately two-thirds of the body iron is in hemoglobin, a 68 000 MW structure containing four subunits of heme, a protoporphyrin ring with iron in the center (**Figure 1**), and four polypeptide chains (two chains each of α- and β-globin). For transport by hemoglobin, oxygen bonds directly to the iron atom, stabilized in an Fe^{2+} oxidation state surrounded by the protoporphyrin ring and histidine residues. Hemoglobin iron easily binds and releases oxygen, circulating in blood erythrocytes. Myoglobin, consisting of a single heme molecule and globin, enables oxygen transfer from erythrocytes to the mitochondria in muscle cells.

Smaller quantities of iron in the heme form function in mitrochondrial cytochromes involved in electron transfer, oxygen utilization, and the production of ATP. A small fraction of body iron functions in heme-containing hydrogen peroxidases such as catalase that protect against excessive hydrogen peroxide accumulation by catalyzing its conversion to hydrogen and oxygen.

Iron also functions in nonheme proteins that contain an iron–sulfur complex, a cubical arrangement of four iron and four sulfur atoms. This is the principal form of iron in the mitochondria, functioning in enzymes of energy metabolism such as aconitase, NADH dehydrogenase, and succinate dehydrogenase. In both mitochondria and cytosol, aconitase is sensitive to iron concentrations. When iron is abundant, the aconitase enzyme assumes the full iron–sulfur cubic structure that is associated with carbohydrate metabolism. However, when intracellular iron concentrations are reduced, the protein loses aconitase activity and functions as an

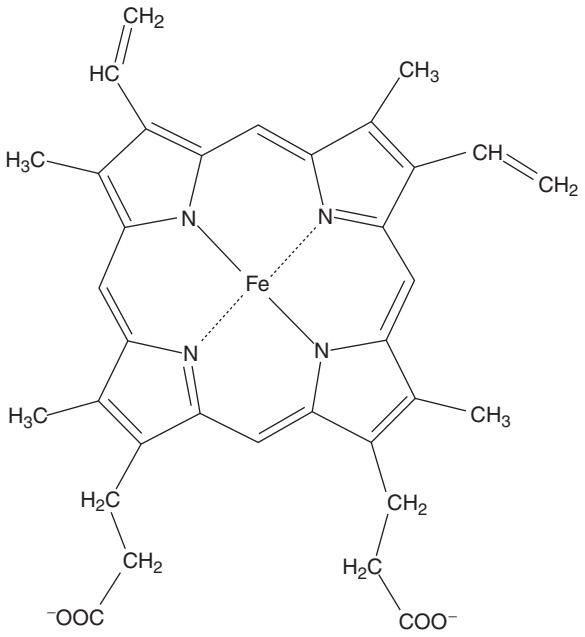


Figure 1 Heme (ferroprotoporphyrin).

iron-binding protein (IRP). IRPs interact with iron response elements (IREs) located in the mRNA to regulate the synthesis of proteins involved in iron transport, storage, and use, in response to changes in cellular iron concentrations.

Absorption, Excretion, Transport, and Storage

Absorption

Both heme and nonheme (inorganic) iron are absorbed in an inverse proportion to body iron stores (indicated by serum ferritin; Figure 2). Heme iron is absorbed more efficiently than the nonheme form. Nonheme iron absorption can vary from 0.1% to >35% and that of heme iron from 20% to 50%, depending on body iron status (stores, erythropoiesis, and hypoxia) and bioavailability

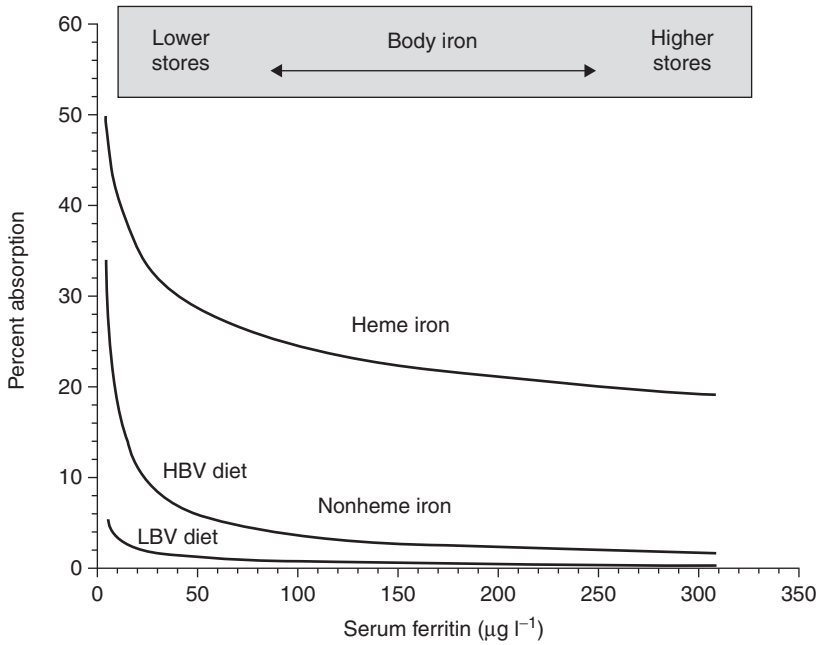


Figure 2 Heme and nonheme iron absorption as influenced by body iron stores and dietary bioavailability. HBV and LBV indicate high and low dietary bioavailability, respectively.

from the diet. These ranges indicate greater control of nonheme compared with heme iron absorption. When iron stores are high, absorption of nonheme iron can be minimized more completely, and when iron stores are low, nonheme iron is absorbed nearly as efficiently as heme iron. Because there is considerably more nonheme iron in the diet (usually ~85–100%), this form of iron contributes most to the physiological control of iron absorption in relation to iron needs.

The upper portion of the duodenum, with its low pH luminal conditions, is the primary site for both heme and nonheme iron absorption (Figure 3). Nonheme iron absorption is better understood than heme iron absorption, but a heme-carrying protein that may be responsible for mucosal uptake of heme iron has been identified. The globin proteins of hemoglobin are proteolytically digested in the intestinal lumen, producing peptide remnants that may enhance the absorption of the heme molecule by preventing heme polymerization. The heme molecule is absorbed as an intact porphyrin structure, most likely *via* a heme-carrying protein (HCP1), possibly involving endocytosis. In the mucosal cell, heme iron is split into ferrous iron and bilirubin by heme oxygenase, adding to a common pool of cellular iron for transport into plasma or intracellular storage and exfoliation.

Nonheme iron is best absorbed if presented to the intestinal villi as soluble ions (preferably reduced, ferrous ions) or as low-affinity, low-molecular-weight iron ligands. Gastric acid facilitates these conditions. Ascorbic acid concurrently ingested with iron helps to maintain the iron in a soluble, reduced, low-molecular ligand form in the intestinal lumen.

Proteins involved in mucosal uptake and transfer of nonheme iron and possible regulatory molecules have been identified (Figures 3 and 4). These include duodenal cytochrome *b* (Dcytb), which converts ferric to ferrous iron at the apical mucosal surface. A divalent metal transporter (DMT-1) transfers ferrous iron into the mucosal cell. Mutations in DMT-1 impair iron absorption and produce microcytic anemia in rodents. Ferrous iron has the highest affinity for DMT-1, but it can also transport other divalent ions, such as manganese, lead, cadmium, zinc, and copper. This may contribute to competitive inhibition observed in the absorption of these metals. Iron transported into the enterocyte may be further transported to the body at the basolateral membrane, completing absorption, or may be held and returned to the intestinal lumen with cellular desquamation. Ferroportin, or Ireg-1, is involved in efflux of iron from the mucosal cell at the basolateral membrane. A mutation in ferroportin results in an uncommon form of hemochromatosis, an iron storage disorder. The mRNA for both DMT-1 and ferroportin contains an IRE, enabling the regulation of mRNA translation by intracellular iron concentrations. Dcytb, DMT-1, and ferroportin are all upregulated in iron deficiency. Intestinal transfer of iron to the circulation also involves hephaestin, an intestinal ferroxidase with a protein sequence similar to that of ceruloplasmin (a copper-containing ferroxidase in serum). A defective hephaestin gene in mice results in anemia and accumulation of iron in intestinal cells.

Iron absorption is responsive to recent iron intake, iron stores, erythropoiesis, hypoxia, pregnancy, and inflammation. A newly identified peptide, hepcidin, is related to several of these stimuli of regulatory control. Hepcidin is an antimicrobial peptide found in human blood and urine that serves as a signal for limiting iron absorption. Some control of absorption also likely involves the HFE protein located in the basolateral membrane of intestinal crypt cells. A specific point mutation in the HFE gene is associated

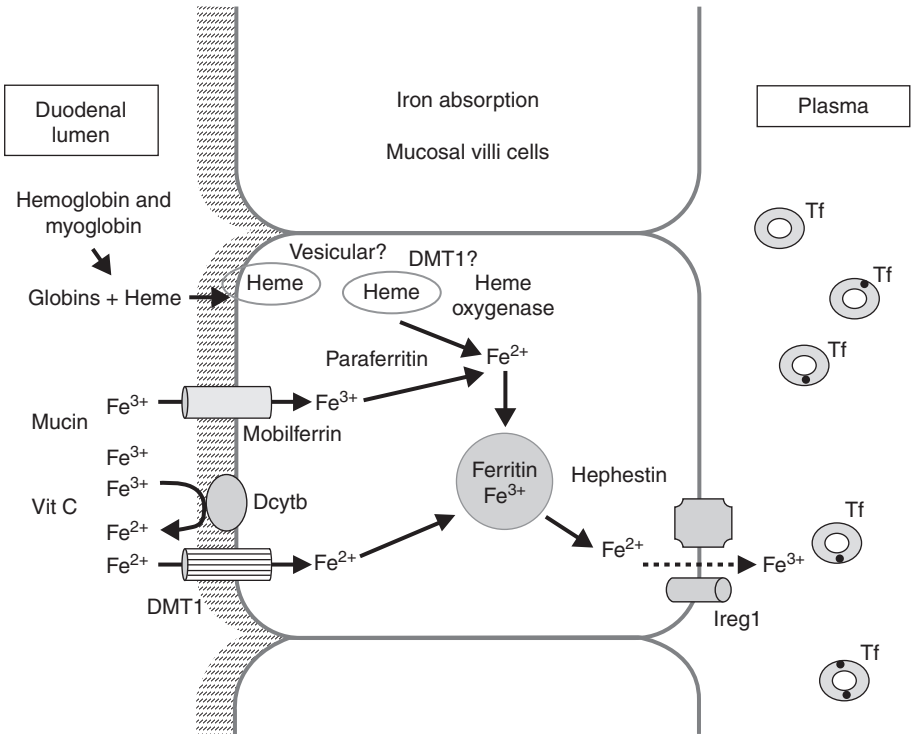


Figure 3 Absorption of iron in the intestinal mucosa.

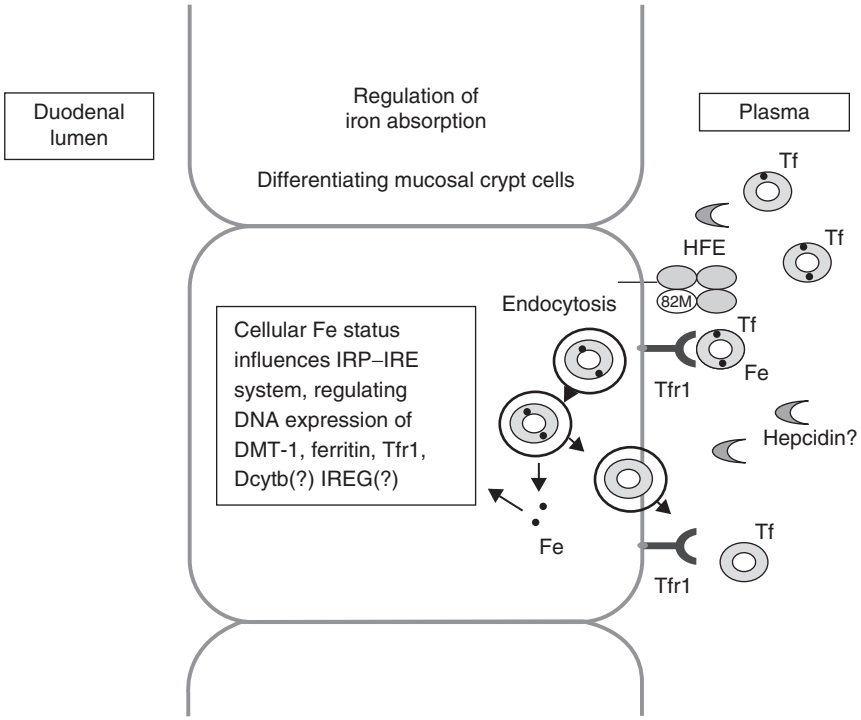


Figure 4 Regulation of iron absorption in the mucosal crypt cells before differentiation and development into actively absorbing intestinal villi cells.

with the most common form of hemochromatosis, a disorder involving excessive iron absorption and accumulation. The HFE protein interacts with β_2-microglobin (B2M) and transferrin receptor, apparently influencing iron uptake from serum transferrin, the primary protein involved in serum iron transport (Figure 4). Knowledge of the control of iron absorption is growing rapidly.

Transport

Transferrin transports essentially all of the 3 or 4 mg of iron in serum, including dietary iron absorbed from the duodenum and iron from macrophages after the degradation of hemoglobin. Each transferrin molecule binds two ferric ions; the transferrin in serum is normally approximately one-third saturated with iron. The amount of iron that can be bound by transferrin is measured as the total iron-binding capacity (TIBC). In iron deficiency, serum iron is reduced, and TIBC is elevated; expressing serum iron as a fraction of the TIBC defines the transferrin saturation, which is reduced in iron deficiency. As iron deficiency develops, these measures of iron transport signal iron deficiency before the functional pool of circulating hemoglobin is reduced (Figure 5).

Membrane transferrin receptors enable the cellular uptake of iron. Transferrin receptors complex with transferrin, the complex is internalized by endocytosis, and the iron is released from transferrin inside the cell on vesicular acidification (Figure 4). Transferrin receptors are abundant in erythrocyte precursors, placenta, and liver, and the number of receptors changes inversely with cellular iron status. Serum transferrin receptors are soluble, truncated forms of the cellular receptors, present in proportion to the cellular receptors, which serve as a clinical indicator of cellular iron status that is useful in distinguishing between iron deficiency and other causes of anemia.

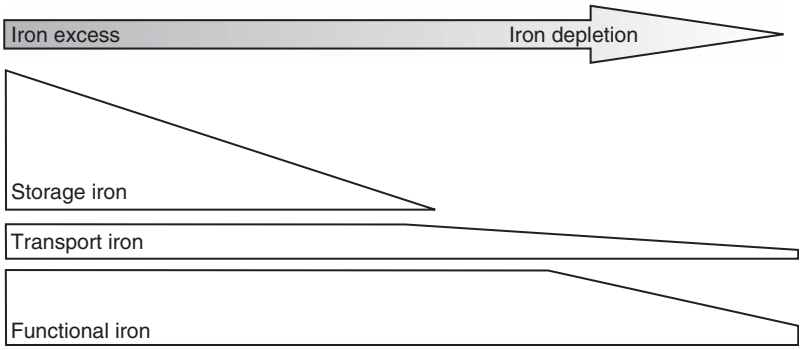
Other proteins involved in iron transport include lactoferrin, which is the major IRP in human milk. Haptoglobin and hemopexin proteins clear hemoglobin and heme, respectively, from circulation as they are released from senescent red blood cells or by disease.

Storage

Iron is primarily stored in the liver, spleen, and bone marrow in the form of ferritin or hemosiderin. Ferritin is a water-soluble protein complex of 24 polypeptide subunits in a spherical cluster with a hollow center that contains up to 25% iron by weight or 4000 atoms of iron per molecule. Hemosiderin is a water-insoluble complex, immunologically similar to ferritin, containing up to 35% iron. Ferritin and hemosiderin each account for approximately half of the storage iron in the liver.

Excretion

The approximately 1 mg of iron lost daily by men and postmenopausal women represents mainly obligatory fecal losses from exfoliated mucosal cells, bile, and extravasated red cells, with minor additional amounts in desquamated skin cells and sweat. Urine contains minimal amounts of iron.



	Excessive body iron	Adequate iron	Low iron stores	Iron deficiency	Iron deficiency anemia
Serum ferritin	↑	---	↓	↓↓	↓↓
Transferrin receptor (sTfR)	---	---	---	↑	↑↑
Index: Log (sTfR/ferritin)	↑	---	↓	↓↓	↓↓↓
Transferrin saturation	↑↑	---	---	↓	↓
Serum iron	↑↑	---	---	↓	↓
Total iron-binding capacity	---	---	---	↑	↑
Erythrocyte protoporphyrin	---	---	---	↑	↑↑
Hemoglobin, hematocrit	---	---	---	---	↓

Figure 5 Clinical indicators of body iron status.

Adolescent girls and premenopausal women lose considerable amounts of iron through menstruation. Menstrual losses of individual women vary considerably; half of women lose less than 14 mg of iron per menstrual period, but the distribution is skewed, and 5% lose 50 mg or more. Iron deficiency among women in affluent countries is commonly attributable to these high iron losses rather that to differences in dietary intakes.

Body Iron Balance

The adult human has 2–4 g of total iron, or approximately 50 mg kg^{-1} in men and 40 mg kg^{-1} in women. Red blood cells contain approximately two-thirds of body iron and have an average life span of 120 days; consequently, approximately 20 mg of iron daily is efficiently recycled from senescent to newly formed erythrocytes through the reticuloendothelial system.

In contrast to other nutrients, controlled through both absorption and excretion, body iron balance is controlled almost exclusively by absorption. Adults consume approximately 10–20 mg iron daily from food. The average absorption and excretion of iron for adult men or postmenopausal women is approximately 1 mg daily. Menstruation can more than double iron losses in women of child-bearing age, increasing their requirement for absorbed iron. Pregnancy also increases the iron requirements considerably. Total body iron in fetuses and newborns is approximately 75 mg kg^{-1} translating into an iron accretion rate of 1–1.5 mg kg^{-1} per day, which, however, does not apply to newborn infants because the normal decline in hemoglobin concentration after birth causes significantly increased iron stores. Therefore, a healthy, term infant is initially independent of external iron and can double its birth weight before iron stores are depleted. Breast milk is low in iron (0.2–0.4 mg l^{-1}), and even though this iron is well utilized, infants breastfed for longer than 4–6 months without receiving iron supplements or iron-fortified complementary foods are at a risk of developing iron deficiency anemia. After that age and through the first years of life , when the growth rate continues to be high, the iron requirements are higher than during any period later in life. Therefore, the introduction of iron-rich complementary food is essential to prevent iron deficiency. Compared to term infants, preterm infants have lower body iron and hemoglobin contents at birth, as well as serum and storage iron. Iron stores may be depleted already during the first months of life, coinciding with the onset of erythropoiesis and catch-up growth. In very low birth weight infants, iron losses due to phlebotomy can amount to 6 mg kg^{-1} per week. Therefore, in contrast to term infants, in whom iron deficiency typically develops after the first half of infancy, preterm infants are at a risk of iron deficiency already during the first half of infancy. Preterm infants of short gestational age or of lower birth weight

are at a particular risk of developing iron deficiency as are preterm infants in low-income countries and those exclusively breastfed without iron supplementation. Maternal iron deficiency does not appear to compromise the iron endowment of their infants, but severe iron deficiency, i.e., iron deficiency anemia, does have an adverse effect on iron status of the newborn. Infants of moderately and severely anemic mothers have lower iron stores and a threefold increased risk of low birth weight, placing them at a higher risk of iron deficiency at an early age. The timing of umbilical cord clamping also affects the iron endowment of the newborn. Early cord clamping decreases iron transfer to the infant, whereas delayed cord clamping increases the red cell volume in the infants and, in turn, increases the iron endowment.

Clinical Assessment of Iron Status

With adequate iron status, there is sufficient iron to meet all of iron's functional roles and a small reserve of storage iron that can be mobilized when needed (**Figure 5**). Excessive body iron, stored in the liver and bone marrow, is marked by elevated serum ferritin and also serum iron and transferrin saturation. Ferritin in plasma corresponds well with body iron stores, but its use as an indicator is limited under inflammatory conditions, which increase circulating ferritin concentrations. Iron deficiency occurs when iron stores are depleted and the iron transported for physiological function is reduced. Iron deficiency without anemia is commonly detected by abnormal values for two out of three blood indices, usually serum ferritin, transferrin saturation, and free erythrocyte protoporphyrin. The ratio of serum transferrin receptor to serum ferritin is considered a single, sensitive indicator of iron status across the full range of body iron status, except under conditions of inflammatory stress and during infancy (**Figure 5**). As iron deficiency becomes more severe, iron deficiency anemia results, with small, pale erythrocytes and reduced blood hemoglobin and hematocrit. The cut-off values used for different age groups when defining anemia and iron deficiency, respectively, may need some further consideration. A study on Swedish adolescents has pointed out how the cut-off values chosen to define anemia (hemoglobin) and iron deficiency (serum ferritin) affect the prevalence of these conditions. This dilemma is even more pronounced in infants and young children for whom appropriate reference values largely are missing and few attempts have been made to establish the appropriate cut-off values. Attempts have recently been made to establish age-appropriate cut-off values for infants with the conclusion that the prevalences of iron deficiency and iron deficiency anemia may have been overestimated.

Iron Nutrition

Iron Deficiency

Iron deficiency is the most common nutrient deficiency, affecting as many as two-thirds of all children and women of child-bearing age worldwide. It has been estimated that iron deficiency severe enough to cause anemia affects 20–25% of infants and as many as 40% of women and 25% of men. Iron deficiency occurs in industrially developed and developing countries. In the United States, 9–11% of toddlers, adolescent girls, and women of child-bearing age have iron deficiency, and 2–4% have iron deficiency anemia. The prevalence of iron deficiency is approximately doubled in US black and Hispanic women. The magnitude of this problem, however, is related not only to the socio-economic status of the population studied, but also how iron deficiency and anemia are defined (see Section on Clinical Assessment of Iron Status). UNICEF estimated that 40–50% of children below 5 years of age in low-income countries suffer from iron deficiency. Some estimates indicate that even in affluent societies it can be as high as 30% or as low as 5%, or less.

Consequences of Iron Deficiency

Iron deficiency adversely affects pregnancy, impairs early childhood development, and cognitive function, and reduces the ability to do physical work. These serious problems are almost exclusively associated with iron deficiency severe enough to cause anemia; however, small reductions in exercise capacity, detectable in a laboratory setting, are also detectable in women with low iron stores and no anemia.

Physical Work Capacity

Iron deficiency anemia adversely affects physical work capacity, reflecting the element's key role in oxygen and energy utilization. Maximal oxygen consumption during exercise is reduced, in association with decreased muscle myoglobin and other iron-containing enzymes. Iron supplementation has improved productivity among Guatemalan sugar and coffee plantation workers, Indian tea pickers, and Indonesian road construction workers and rubber tappers. Iron supplementation programs are clearly cost-effective in addition to providing a positive impact on human health and well-being.

Cognitive Development

In infants, iron deficiency anemia has been associated with reduced mental and motor test scores and behavioral changes. This impaired mental and motor functioning appears to persist after treatment with iron, emphasizing the need for early detection and treatment and preferably prevention of iron deficiency during early development. However, it should be noted that iron deficiency in children is associated with a large number of psychosocial and economic disadvantages, which could account for some or all of the children's functional deficits. There are relatively few double-blind randomized-controlled trials of iron supplementation

with adequate power, but there is considerable evidence showing that children with iron deficiency anemia usually have poor cognitive and motor development and that iron supplementation usually has beneficial effects on motor development in children with iron deficiency anemia under 3 years of age and on cognition in iron-deficient anemic school-aged children.

Reproduction

Iron deficiency anemia has been associated with premature birth and low birth weight. Iron supplementation during pregnancy is not always completely effective in preventing maternal anemia in women with low iron stores early in gestation. leading to suggestions for promoting adequate iron stores in all women of child-bearing age before conception.

Other

Iron deficiency increases the susceptibility to lead poisoning. It may also impair resistance to infection and regulation of body temperature. Iron deficiency has been associated with the eating of nonfood material (pica) or ice (pagophagia). Clinical signs may include spoon-shaped fingernails and abnormalities of the mucosa of the mouth and gastrointestinal tract.

Recommended Dietary Intakes

The US and Canadian recommended iron intakes are intended to meet the requirements of 97.5% of the healthy population, replacing excreted iron and maintaining essential iron functions with a minimal supply of body iron stores. They also assume a relatively high bioavailability of the dietary iron. The recommended 8 mg daily for adult men and postmenopausal women can easily be met with varied Western-style diets. More careful food choices are needed to obtain the 18 mg recommended to meet requirements for 97.5% of adult menstruating women. This higher recommendation reflects the high menstrual iron losses of some women; the median iron requirement is 8.1 mg for menstruating women.

During pregnancy, dietary iron recommendations are increased to 27 mg daily, based on the iron content of the fetus and placenta (approximately 320 mg) and the expanded blood volume associated with a healthy pregnancy. Meeting this recommendation generally requires iron supplementation. Supplementation with 30–60 mg daily is commonly recommended. Lactation has minimal impact on maternal iron balance and recommendations, largely due to lactational amenorrhea.

The high iron requirements of early growth put infants and toddlers at a risk of iron deficiency. Breast-feeding is recommended for the first year of life. Although iron in breast milk is relatively low (0.35 mg l^{-1}, or 0.27 mg daily), it is well absorbed, possibly mediated by lactoferrin. Breast milk alone is assumed to be adequate for the first 6 months of infancy, with the addition of iron-rich foods in the next 6 months. When infant formula is used, iron fortification of the formula is recommended. Most infant formulas contain 4–12 mg of iron per liter, which is at least 10–30 times higher than the level of iron in breast milk. It may be questioned whether infant formula used during the first 6 months of life should contain a vast excess of iron, which provides no benefit in order to cover perceived increased iron requirements during 6–12 months of age. In areas where the same type of infant formula is used during the first 12 months of age, increasing the level of iron fortification in complementary foods may be an alternative possibility, whereas in areas where different types of infant formulas are used between 0–6 and 6–12 months of age, the follow-on formula may have a higher level of iron fortification. Reflecting the high but decreasing growth rate, the recommended daily intake of iron is 11 mg for infants 7–12 months, 7 mg for children between 1 and 3 years, and 10 mg for children between 4 and 8 years of age. There is no recommended intake for infants 0–6 months old, only an adequate intake (AI), which is 0.27 mg day^{-1} and based on the mean iron intake of exclusively breastfed infants. Iron stores are low in preterm infants as they are built during the last trimester of pregnancy. Iron supplementation and blood transfusion are therefore routinely used to prevent iron deficiency anemia in this population. The proper level and timing of iron supplementation is still controversial, but the ESPGHAN Committee on Nutrition recommends an intake of 2–3 mg kg^{-1} per day, corresponding to 1.8–2.7 mg per 100 kcal, and that prophylactic enteral iron should be started at 2–6 weeks of age (2–4 weeks in extremely low birth weight infants).

Western dietary recommendations have been based on mixed diets with a relatively high bioavailability of iron and may need to be increased twofold or more for low-meat, plant-based diets with greater phytic acid content (see section Bioavailability). Other factors that may increase dietary requirements include achlorhydria, which decreases iron absorption, hookworm, or other parasites that increase gastrointestinal blood loss, or intrauterine contraceptive devices that may increase menstrual losses by 30–50%. In contrast, hormonal contraceptives reduce iron requirements by reducing menstrual losses by approximately 50%.

Dietary Iron

Food Sources

Typical Western diets contain approximately 6 mg iron per 1000 kcal. Men and women consume approximately 16–18 and 12–14 mg daily, respectively. In the United States, 24% of dietary iron is supplied by breads, pasta, and bakery products. An additional 21% comes from (mostly fortified) cereal products. Other abundant dietary sources are red meats (9% from beef), poultry, legumes, and lentils. In countries such as the United States, fortification practices increase the influence of grain and cereal products as sources of iron. In countries without fortification to at least replace the iron lost during milling, the refinement of grain products considerably reduces dietary iron content. The populations of developing countries that eat little meat and do not include legumes or lentils as a dietary staple are at an increased risk of inadequate iron intake.

Bioavailability

In underdeveloped countries, diets may be inadequate in both iron content and bioavailability (the amount that is absorbed and utilized by the body). However, the bioavailability of iron can be more important than the iron content in determining the amount of iron absorbed from food. Diets with similar total iron contents can differ 8–10-fold in the amount of absorbable iron. Dietary iron bioavailability is high from refined Western diets containing meat, poultry, and fish and abundant sources of ascorbic acid with low consumption of phytic acid from whole grains and legumes and limited drinking of coffee and tea with meals. On average, men absorb 1 mg daily from such diets, and women, with their lower iron stores, absorb approximately 2 mg. Individuals may absorb considerably more or less, depending on their body iron stores (**Figure 2**). It has been shown in adults that iron status is the strongest regulator of iron absorption. However, iron homeostasis in infants may not be fully developed.

It has been shown that iron absorption was similar in iron-supplemented and unsupplemented 6-month-old healthy, exclusively breast-fed infants born at term, whereas at 9 months of age, unsupplemented infants had considerably higher iron absorption. This strongly suggests that homeostatic regulation of iron absorption is absent in young infants but matures and is present at 9 months of age.

Heme Iron

Approximately 10%, or 1–2 mg, of the iron in a mixed, Western diet is in the well-absorbed heme form. Heme iron accounts for approximately 40% of the iron in meat, poultry, or fish flesh. There is little to no heme iron in dairy products, or foods of plant origin. Heme iron is absorbed as an intact porphyrin structure. Heme iron absorption is enhanced by meat, poultry, or fish but it is not influenced by the other enhancers and inhibitors of nonheme iron absorption. Meat can be a significant source of heme iron for older infants and children and has a positive effect on iron status.

Nonheme Iron

Nonheme iron usually accounts for 85–100% of dietary iron. In contrast to heme iron, the absorption of nonheme iron is substantially influenced by dietary enhancers and inhibitors consumed concurrently. These factors appear to affect the solubility of a single exchangeable pool of nonheme iron absorbed from the intestinal digesta.

Absorption of nonheme iron is enhanced by ascorbic acid, which reduces ferric iron to ferrous iron, resulting in a soluble iron–ascorbic acid complex. Enhanced absorption has been demonstrated with synthetic and several food sources of ascorbic acid. The enhancement increases logarithmically with the dose, approximately doubling absorption with as little as 25 mg of ascorbic acid and increasing absorption by nearly 10-fold with 1000 mg of ascorbic acid.

Nonheme iron absorption is also enhanced by concurrently consuming meat, poultry, or fish. Despite intensive study, the factor responsible for this enhancement by animal flesh has not been identified and may involve the general matrix of low-molecular-weight peptides released during digestion.

Nonheme iron absorption is reduced by phytic acid (inositol hexaphosphate), present in legumes, rice, and grains, that binds iron and makes it insoluble. Both phytate and iron are concentrated in the aleurone layer and germ of grains, and they are reduced with milling, which increases the bioavailability of the remaining iron. An additional unidentified factor in soy beans, independent of the phytic acid, also impairs iron absorption. Polyphenols in grains, fruits, and vegetables, and including the tannins in tea and coffee, also inhibit nonheme iron absorption. Ascorbic acid consumed concurrently can partially reduce the inhibition of nonheme iron absorption by both phytic acid and polyphenols. Calcium in supplemental quantities has been shown to inhibit both heme and nonheme iron absorption from foods in short term studies but shows no effect on long-term iron status.

Supplementation and Fortification

The serious international problem of iron deficiency has been met with limited success by supplementation and fortification efforts. Both approaches suffer from difficulties in delivery and acceptance. Supplements that readily ionize into the ferrous form, such as ferrous sulfate, ferrous fumarate, or ferrous gluconate, are highly bioavailable but may cause gastrointestinal discomfort. Iron injections are poorly tolerated and can result in serious infections. Because daily supplementation reduces the physiological efficiency of iron absorption, routine weekly iron supplementation with 60 mg iron has been suggested in developing countries for women of child-bearing age, beginning in adolescence. Menstruating women in more affluent countries are advised to undergo an assessment from a health professional before taking iron supplements in excess of 20 mg daily.

Fortification of staple foods with 3–10 mg iron daily, depending on the needs of the population, is a long-term preventative strategy. In the United States, bread and cereal products are routinely fortified with 20 mg iron per pound (460 g) of flour, and additional fortification at the option of food suppliers is common. However, fortification is difficult when food processing is decentralized, as is common in low-income countries. Food fortification carries the additional challenge that the chemical forms of iron most bioavailable also tend to be the most reactive with the food fortified, resulting in adverse changes in flavor, color, and shelf-life. Promising approaches include the fortification of food sauces with iron chemically bound with amino acids or with EDTA (sodium iron ethylenediaminetetraacetic acid), from which it is well absorbed even in the presence of phytic acid. Elemental iron powders, commonly referred to as carbonyl, electrolytic, and reduced forms of iron, are relatively inert in foods and inexpensive, but their bioavailability may be 30–80% less than iron from ferrous sulfate, depending on the dissolution in the gastrointestinal tract. Ferric orthophosphate and ferric pyrophosphate do not adversely affect foods but are poorly bioavailable; however, efforts are under way

to enhance their bioavailability by reducing the particle size and encapsulating the particles with various lipids or carbohydrates to prevent agglomeration.

The form of iron given to infants may affect indicators of iron status differently. It has been shown that infants provided iron-fortified cereals between 6 and 9 months of age had significantly higher hemoglobin concentrations than infants given the same amount of iron daily in the form of iron drops. Similarly, it has been found that the addition of meat at 8 months of age had a positive effect on hemoglobin but not on serum ferritin. In contrast, infants given iron drops had significantly higher serum ferritin concentrations than those fed iron-fortified cereals. This suggests that dietary iron and iron from supplements are metabolized differently, with the latter preferentially being deposited in stores, whereas iron in meat and fortified foods is incorporated into erythrocytes. Further studies are needed to elucidate the mechanisms behind these observations.

Excessive Intakes

An extensive biological control system limits the occurrence of free ionic iron that can readily participate in toxic, free radical-producing reactions. Large quantities of ingested iron are acutely toxic, and accidental ingestion of medicinal iron preparations is a cause of poisoning deaths in young children. Iron supplementation is also associated with gastrointestinal irritation. Iron supplements also adversely affect absorption of zinc. Iron absorption is well controlled, but iron overload can result from excessive parenteral iron administration or blood transfusions. Dietary iron overload, possibly exacerbated by genetic factors, occurs in sub-Saharan tribes that consume a high-iron traditional beer prepared and stored in iron containers. Genetic factors can substantially influence body iron retention, as indicated by hemochromatosis, a relatively frequent iron storage disorder of northern European descendants characterized by excessive iron absorption and leading to life-threatening iron damage of organs in adulthood. The possible association of high iron stores with increased risk of diseases related to oxidative stress, including cardiovascular disease, diabetes, and cancer, is an area of epidemiological investigation.

Iron overload as such has not been recognized in term human infants, and is only implicated in premature infants with a known, or feared, consequence of increased iron-associated oxidative damage. Indications of excessive iron intakes by infants have recently been observed in some studies. An adverse effect of iron supplements on linear growth has been shown in iron-replete infants in both affluent and developing countries. Other studies have shown negative effects of iron supplements on weight gain. However, the nutritional status of the infants in those studies was compromised overall, which is known to decrease linear growth and cause stunting. Thus, when linear growth is compromised, it is possible that the adverse effect of excess iron is manifested differently and instead impairs weight gain. The mechanism(s) behind the adverse effect of excess iron is still not known, but may involve the pro-oxidative effects of excess iron or, possibly, an interaction between iron and nutrients involved in growth, such as zinc.

Further Reading

Anderson, GJ, Frazer, DM, McLaren, GD, 2009. Iron absorption and metabolism. Current Opinion in Gastroenterology 25, 129–135.
Centers for Disease Control and Prevention, 1998. Recommendations to prevent and control iron deficiency in the United States. MMWR Recommendations and Reports 47, 1–29.
Collard, KJ, 2009. Iron homeostasis in the neonate. Pediatrics 123, 1208–1216.
Collins, JF, Wesslingresnick-Resnick, M, Knutson, MD, 2008. Hepcidin regulation of iron transport. Journal of Nutrition 138, 2284–2288.
Institute of Medicine, Food and Nutrition Board, 2001. Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc. National Academy Press, Washington, DC.
Kaplan, J, Ward, DM, De Domenico, I, 2011. The molecular basis of iron overload disorders and iron-linked anemias. International Journal of Hematology 93, 14–20.
Lynch, S, Stoltzfus, R, Rawat, R, 2007. Critical review of strategies to prevent and control iron deficiency in children. Food and Nutrition Bulletin 28 (4 supplement), S610–S620.
Wallander, ML, Leibold, EA, Eisenstein, RS, 2006. Molecular control of vertebrate iron homeostasis by iron regulatory proteins. Biochimica et Biophysica Acta 1763, 668–689.

Lipids (fats and oils)

A.H. Lichtenstein, Tufts University, Boston, MA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A.H. Lichtenstein, Fats and Oils, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 201–208, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00097-0.

Dietary fats and oils—the good, bad, and ugly 292
Introduction 292
Lipids—in food and in the body 292
Fatty acids 292
Triacylglycerol (triglyceride) 294
Phospholipid 295
Cholesterol 295
Other sterols 296
Dietary fats and oils 297
Fatty acid profile of common dietary fats 297
Major contributors of dietary saturated, monounsaturated and polyunsaturated fatty acids, and cholesterol 297
Dietary fat and cardiovascular disease prevention 298
Quantity of dietary fat 298
Quality of dietary fat 298
Composition of dietary fats 299
Dietary guidance 299
Summary 299
References 300
Further reading 300

Key points

- Discuss the types of lipids found in foods and the human body.
- Summarize the fatty acid profile in different types of common dietary fats.
- Discuss the major contributors of different types of lipids to the diet.
- Review the relation between dietary lipids and cardiovascular disease; amount and types
- Summarize current dietary guidance with respect to dietary fat for the prevention and treatment of cardiovascular disease.

Glossary

ω-3 Fatty acid Polyunsaturated fatty acids in which the first double bond in the acyl chain is three carbons from the methyl end. Of the common dietary ω-3 fatty acids, carbon chain lengths range from 18 to 22 carbons. Soybean and canola oils contain α-linolenic acid (ALA, 18:3); fish, seafood and algae contain eicosapentaenoic acid (EPA, 20:5) and docosahexaenoic acid (DHA, 22:6)

Atherosclerotic plaque Accumulation of cholesterol in blood vessel walls forming atherosclerotic plaques. Atherosclerotic plaques can narrow the vessel lumen and impede blood flow, and/or trigger formation of blood clots that can subsequently occlude blood flow, resulting in angina, heart attack, stroke or phlebitis

Dietary cholesterol Fat-soluble compound found in animal fats, and is particularly high in eggs. It is a critical component of cell membranes, maintaining fluidity, and lipoproteins, maintaining structure. Cholesterol serves as a substrate for steroid hormones, bile acids, and vitamin D synthesis. In cholesterol-sensitive individuals, high intake of dietary cholesterol is associated with elevated plasma low density lipoprotein (LDL) cholesterol concentrations and cardiovascular disease risk

Dietary fat Macronutrient which provides 9 kcal g^{-1}. Hydrophobic component of food. Primarily in the form of triglyceride. Also includes cholesterol, fat soluble vitamins and phospholipids

Lipoproteins Spherical particles with a hydrophilic surface composed primarily of phospholipid, apoproteins and unesterified cholesterol, and hydrophobic core composed primarily of triacylglycerol (triglyceride) and cholesteryl ester. Particles facilitate the transport of lipid in the aqueous milieu of the blood stream to peripheral tissue for subsequent metabolism or storage

Monounsaturated fatty acid Fatty acid containing a single double bond in the acyl chain. Oils rich in monounsaturated fatty acids are usually liquid at room temperature and semi-solid at refrigerator temperature. When fats rich in monounsaturated

fatty acids replace fats rich in saturated fatty acids in the diet, cardiovascular disease risk is lowered. Common dietary sources are canola and olive oils

Polyunsaturated fatty acid Fatty acid containing two or more double bonds in the acyl chain. Oils rich in polyunsaturated fatty acids are liquid at room and refrigerator temperatures. When fats rich in polyunsaturated fatty acids replace fats rich in saturated fatty acids in the diet, cardiovascular disease risk is lowered. Common dietary sources include soybean, corn, and unmodified safflower and sunflower oils

Saturated fatty acid Fatty acid containing no double bonds in the acyl chain. Fats rich in saturated fatty acids are usually solid at room and refrigerator temperatures. Diets high in saturated fatty acids and low in monounsaturated and polyunsaturated fatty acids are associated with elevated cardiovascular disease risk. Common dietary sources include fatty cuts of meat and full fat dairy products

***Trans* fatty acid** Monounsaturated or polyunsaturated fatty acid in which at least one double bond is in the *trans*, rather than *cis*, configuration. There are two main sources of dietary *trans* fatty acids: animal fat and partially-hydrogenated fat. Recently, due to regulatory changes in Nutrient Facts labels and the FDA's Generally Recognized as Safe (GRAS) list, the latter has been a dramatic decrease in the food supply. Dietary *trans* fatty acids are associated with elevated cardiovascular disease risk

Triacylglycerol (triglyceride) Triacylglycerol is the major form of dietary lipid in fats and oils, whether derived from plants or animals, and the major form of lipid in the body. Triacylglycerol is composed of three fatty acids esterified to a glycerol molecule

Dietary fat is a macronutrient that has historically engendered considerable controversy and continues to do so. Contentious areas include optimal amount and type for cardiovascular disease risk reduction, and role in body weight regulation.

Dietary fats and oils—the good, bad, and ugly

Introduction

Dietary fats and oils are unique in modern times because they encompass good, bad, and ugly characteristics. The aspects of dietary fat that are classified as good include serving as a carrier of fat soluble vitamins (vitamins A, D, E, and K), enhancing the bioavailability of fat-soluble bioactive substances (e.g., absorption of fat-soluble micronutrients), providing essential substrate for the synthesis of metabolically active compounds (e.g., eicosanoid, vitamin D and bile acids), providing critical structural components (e.g., cell membranes and lipoprotein particles), preventing carbohydrate-induced hypertriglyceridemia, insulating the body against extreme temperature changes, cushioning vital organs, and serving an energy-dense form of reserve metabolic fuel (triacylglycerol [triglyceride]). The aspect of dietary fat that are classified as bad include serving as a reservoir for fat-soluble toxic compounds (adipose tissue in our bodies and fats we eat [e.g., fish]). The aspects of dietary fat that are classified as ugly include providing a concentrated form of metabolic fuel in times of excess and contributing saturated and *trans* fatty acids that promote atherosclerotic plaque formation, the underlying cause of heart disease, stroke, and phlebitis.

Lipids—in food and in the body

Fatty acids

Fatty acids are the basic components of larger lipid compounds or serve as substrates for bioactive molecules. They are composed of an acyl (hydrocarbon) chain with a methyl and a carboxyl group at either end. The majority of fatty acids have an even number of carbons. The range of chain lengths for common fatty acids is broad, 12–22 carbons, although shorter- and longer-chain fatty acids occur naturally. The predominant fatty acids in the human body and food are depicted in **Table 1**. In addition to chain length, fatty acids differ from each other in number, type, and location of double bonds. Fatty acids with no double bonds are referred to as saturated, with one double bond as monounsaturated, and with two or more double bonds as polyunsaturated fatty acids (**Fig. 1**).

The double bonds within unsaturated fatty acids, either monounsaturated or polyunsaturated, can be in either the *cis* (hydrogen atoms on same side of the acyl chain) or *trans* (hydrogen atoms on opposite sides of the acyl chain) conformation (**Fig. 2**). The presence of a *cis* relative to a *trans* double bond results in a greater bend or kink in the hydrocarbon chain. This kink impedes the fatty acids from aligning (packing together). In a cell membrane, this results in a higher level of fluidity; in food, it results in oils that are liquid or fats that are soft at room temperature. The vast majority of fatty acid double bonds occur in the *cis* conformation. Fatty acids with an identical number of carbons atoms and double bonds, but differ structurally are referred to as isomers. If the difference is in the conformation of the double bond, they are referred to as geometric isomers (e.g., oleic acid (18:1*cis*) and elaidic acid (18:1*trans*)) (**Fig. 2**).

If the difference is in the location of a double bond in the acyl chain, they are referred to as positional isomers (e.g., α-linolenic acid [18:3*n*-3] and γ-linolenic acid [18:3*n*-6]) (**Fig. 2**). Fatty acids in which the first double bond from the methyl end of the acyl chain occurs at the sixth carbon are termed an ω-6 (*n*-6) fatty acids (e.g., linoleic acid [18:2*n*-6] and γ-linolenic acid [18:3*n*-6]). Fatty acids in which the first double bond from the methyl end of the acyl chain occurs at the third carbon are termed an ω-3 (*n*-3) fatty

Table 1 Common fatty acids.

Code	*Common name*
Saturated	
12:0	Lauric acid
14:0	Myristic acid
16:0	Palmitic acid
18:0	Stearic acid
Monounsaturated	
16:1*n*-7 *cis*	Palmitoleic acid
18:1*n*-9 *cis*	Oleic acid
18:1*n*-9 *trans*	Elaidic acid
Polyunsaturated	
18:2*n*-6,9 all *cis*	Linoleic acid
18:3*n*-3,6,9 all *cis*	α-linolenic acid
18:3*n*-6,9,12 all *cis*	γ-linolenic acid
18:4n-3-6,9,12,15 all *cis*	Stearidonic acids
20:4*n*-6,9,12,15 all *cis*	Arachidonic acid
20:5*n*-3,6,9,12,15 all *cis*	Eicosapentaenoic acid
22:5*n*-3,6,9,12,15 all *cis*	Docosapentaenoic acid
22:6*n*-3,6,9,12,15,18 all *cis*	Docosahexaenoic acid

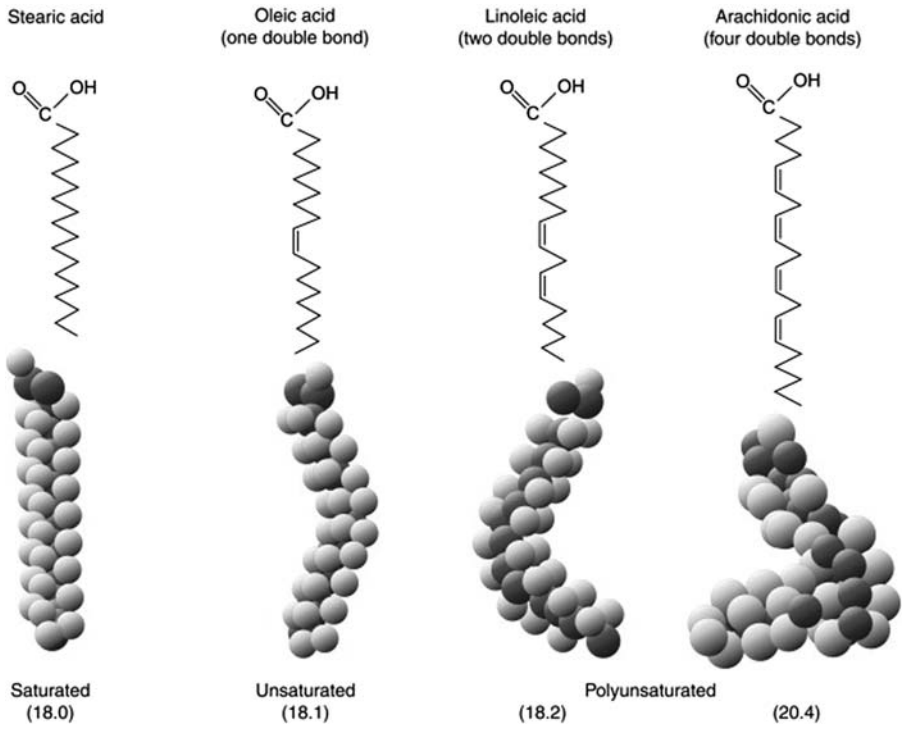


Fig. 1 Structures of saturated, monounsaturated, and polyunsaturated fatty acids.

acids (e.g., eicosapentaenoic acid [20:5*n*-3] and α-linolenic acid [18:3*n*-3]). Most double bonds occur in a nonconjugated sequence, that is, a carbon atom with only single carbon–carbon bonds separates carbon atoms that are part of double bonds. Some double bonds occur in the conjugated form, without an intervening carbon separating carbons as part of double bonds. Conjugated double bonds tend to be more reactive chemically (e.g., more likely to become oxidized). The vast majority of enzymes that metabolize fatty acids distinguish not only on the basis of chain length and number of double bonds, but among isomers. The metabolic products of different fatty acid isomers, especially positional isomers, have distinct and at times opposing biological effects.

Some fatty acids are classified as an essential nutrient. An essential nutrient is a nutrient that the body cannot synthesize or cannot synthesize in amounts adequate to meet requirements. Linoleic acid (18:2*n*-6) and linolenic acid (18:3*n*-3) and their derived fatty acids are essential because unlike plants, humans cannot introduce a double bond after the ninth carbon in a fatty acyl chain (from the carboxyl end).

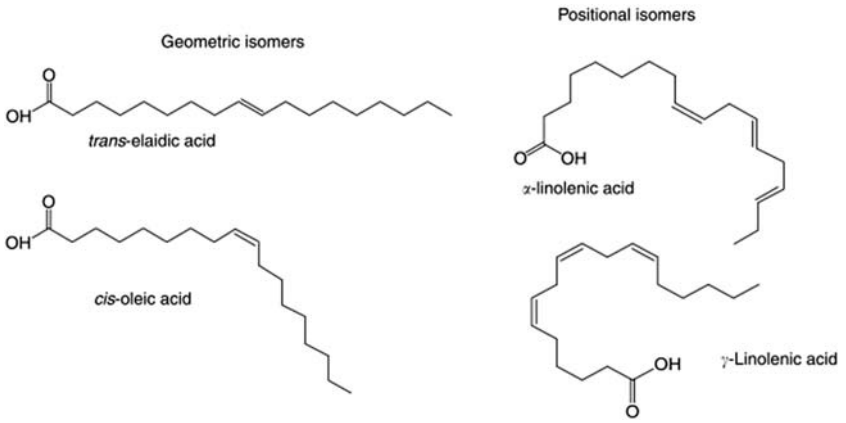


Fig. 2 Examples of fatty acids geometric isomers (*trans*-elaidic acid and *cis*-oleic acid) and positional isomers (α-linolenic acid and γ-linolenic acid).

Triacylglycerol (triglyceride)

Triacylglycerol is the predominant form of dietary lipid in fats and oils, whether derived from plants or animals, and the major form of lipid in the body. Triacylglycerol is composed of three fatty acids esterified to a glycerol molecule (Fig. 3). The physical properties of a triacylglycerol molecule are determined by the specific fatty acids esterified to the glycerol moiety and the actual position the fatty acids occupy. Each of the three carbons comprising the glycerol molecule allows for a stereochemically distinct fatty acid bond position: *sn*-1, *sn*-2, and *sn*-3. A triacylglycerol with three identical fatty acids is termed a simple triacylglycerol. These are exceedingly rare in nature. A triacylglycerol with two or three different fatty acids is termed a mixed triacylglycerol and make up the bulk of the triacylglycerol that occur naturally. The melting point of a triacylglycerol is determined by the physical characteristics and position of the fatty acids esterified to glycerol,—their chain length; number, position, and conformation of the double bonds; and the stereochemical position.

Approximately 90% of the molecular weight of triacylglycerol is accounted for by the three fatty acids. The fatty acid profile of the diet is reflected, in part, in the fatty acid profile of the adipose tissue triacylglycerol. Such data have been used to approximate long-term food intake patterns of humans.

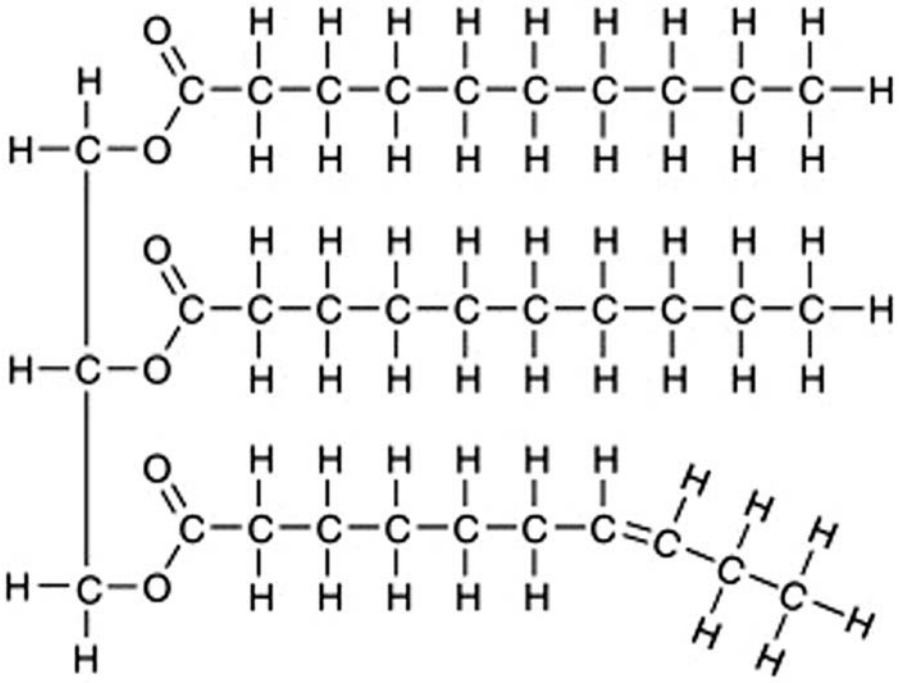


Fig. 3 Triacylglycerol molecule.

Mono- and diglycerides have one and two fatty acids, respectively, esterified to glycerol. In nature, they occur only in trace amounts. They are primarily intermediate products of triacylglycerol digestion, clearance from the bloodstream, or intracellular metabolism. They are used as emulsifiers in processed food.

Once consumed, triacylglycerol is hydrolyzed into non-esterified (free) fatty acids and monoglycerides by lipase that are synthesized in the pancreas and stored in the gallbladder until lipid in the small intestine signals for release. During absorption, these breakdown compounds enter the intestinal cells (enterocyte) and are used to resynthesize triacylglycerol. This triacylglycerol, along with intestinally derived cholesterol and fat soluble compounds such as fat soluble vitamins, are incorporated into nascent triglyceride-rich lipoprotein particles, termed chylomicrons. Chylomicrons are first secreted into the thoracic duct (lymph) which drains into the subclavian vein (peripheral circulation). Once in circulation, fatty acids are hydrolyzed from the triacylglycerol which enables them to cross the plasma membrane of peripheral cells. The primary enzyme that hydrolyzes triacylglycerol in plasma is lipoprotein lipase. The products of lipoprotein lipase are non-esterified fatty acids and 2-monoacylglycerol. The enzyme is attached to the luminal surface of capillary endothelial cells via a highly charged membrane-bound chain of heparin sulfate-proteoglycans. The ability of lipoprotein lipase to bind both the chylomicron particle and the cell surface ensures the cellular uptake of the liberated non-esterified fatty acids. Once inside the cell, the non-esterified fatty acids can be oxidized to provide energy, metabolized to biologically active compounds, or resynthesized into triacylglycerol for storage and subsequent use.

Phospholipid

There are only trace amounts of phospholipid in dietary fats and oils. However, because the fatty acids in fats and oils provide substrate for the endogenous synthesis of phospholipid for the subsequent synthesis of cell membranes and lipoprotein particles, this subtype of fat is important to discuss. Phospholipid is composed of two fatty acids esterified to the *sn*-1 and *sn*-2 positions and a moiety frequently referred to as a polar head group is esterified to the *sn*-3 position of glycerol, the latter group via a phosphate bond (Fig. 4). Phospholipid molecules are amphipathic, that is, there are both hydrophobic and hydrophilic domains in the molecule. The two fatty acids confer hydrophobic properties and the polar head group confers hydrophilic properties. The specific fatty acids esterified to the glycerol backbone of phospholipid tend to be unsaturated. The most common polar head groups include phosphorylcholine, phosphorylserine, phosphorylinositol, and phosphorylethanolamine. In cells, because of their amphipathic nature, phospholipid form bilayers. These bilayers are the structural components of both plasma and intracellular membranes, and serve as a platform for proteins and cholesterol that sit either within or on the surface of the bilayers. The more unsaturated the fatty acids in phospholipid, the more fluid the membrane. The degree of fluidity can affect the activity of proteins located in the membrane (e.g., enzyme activity, receptor ligands). Cell membrane phospholipid also serves as a reservoir for metabolically active fatty acids. In the small intestine, phospholipids, because of their amphipathic properties, plays a critical role facilitating the emulsification of intestinal fat, subsequent formation of micelles and absorption of lipid compounds. As a monolayer on the surface of lipoprotein particles, phospholipids provide a critical component in the packaging and transport of lipid in circulation.

Cholesterol

Dietary sources of cholesterol are limited to foods of animal origin. Cholesterol is an amphipathic molecule that is composed of a steroid nucleus and a branched hydrocarbon tail (Fig. 5). Cholesterol occurs naturally in two forms: non-esterified (free cholesterol) and esterified (cholesteryl ester). If esterified, the fatty acid is linked to cholesterol at the number 3 carbon of the sterol ring.

Non-esterified cholesterol is a component of cell membranes and along with degree of phospholipid fatty acid saturation, determines membrane fluidity. The more unsaturated, the more fluid. Cholesterol intercalates into the phospholipid bilayer restricting motility of the fatty acyl chains and decreases fluidity. In synaptic membranes, non-esterified cholesterol is critical for normal nerve

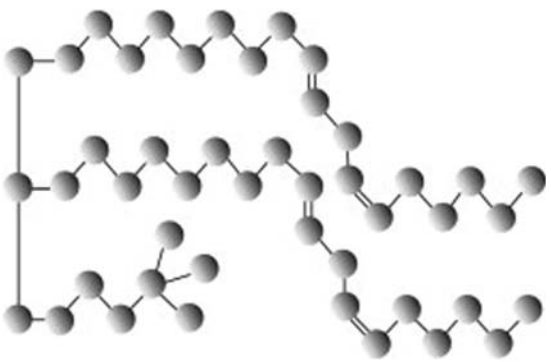


Fig. 4 Phospholipid molecule (phosphatidylcholine).

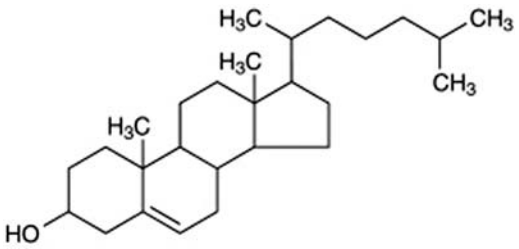


Fig. 5 Cholesterol molecule.

transmission. It makes up approximately 10% (dry weight) of total brain lipids. Cholesterol is a precursor of steroid hormones (e.g., estrogen, testosterone), vitamin D, adrenal steroids (e.g., hydrocortisone, aldosterone), and bile acids. This latter property is exploited in certain approaches to decrease elevated plasma LDL cholesterol concentrations by preventing the resorption of bile acids (recycling), hence, forcing the liver to use additional cholesterol for bile acid synthesis. In so doing, a potential route for net cholesterol excretion is expanded.

The receptor-mediated cellular uptake of cholesterol laden plasma lipoprotein particles is critical in maintaining intracellular and whole-body cholesterol homeostasis. Once internalized, the cholesterol component of lipoprotein particles is hydrolyzed to non-esterified cholesterol. This internalized cholesterol has three major metabolic effects. One, it inhibits the activity of 3-hydroxy 3-methylglutaryl CoA (HMGCoA) reductase, the rate-limiting enzyme in endogenous cholesterol biosynthesis. This property serves to decrease the rate of intracellular cholesterol biosynthesis commensurate with the uptake of cholesterol from extracellular sources (plasma lipoproteins), thereby minimizing intracellular accumulation. Two, intracellular non-esterified cholesterol inhibits the synthesis of receptors that take up lipoprotein particles containing apoproteins B100 or E from the plasma, thereby limiting cholesterol accumulation in the cell and avoiding excess accumulation. Lastly, intracellular non-esterified cholesterol increases the activity of acyl CoA cholesterol acyltransferase (ACAT), the intracellular enzyme that converts non-esterified cholesterol to cholesteryl ester. A high level of intracellular non-esterified cholesterol is cytotoxic, whereas cholesteryl ester is a highly nonpolar molecule that coalesces to form intracellular lipid droplets, averting high concentrations of non-esterified cholesterol interacting with intracellular components and having subsequent detrimental effects. ACAT uses primarily oleoyl CoA as substrate, resulting in the synthesis of cholesteryl oleate.

Non-esterified cholesterol can be not only esterified intracellularly but also in plasma. The reaction is catalyzed by the enzyme lecithin cholesterol acyltransferase (LCAT). The source of fatty acid is phosphatidylcholine. The products of LCAT are cholesteryl linoleate and lysolecithin. The resulting apolar cholesteryl ester migrates to the core of the lipoprotein particle.

Cholesterol in circulation is carried on all lipoprotein particle species: both intestinally derived chylomicrons and hepatically derived very LDL, intermediate-density lipoprotein, and high-density lipoprotein (HDL) particles. Non-esterified cholesterol is incorporated into the phospholipid monolayer surface of lipoprotein particles, whereas cholesteryl ester is incorporated into the core of lipoprotein particles. The majority of the cholesterol in circulation is carried on LDL. Cholesteryl ester is the major component of atherosclerotic plaque. In the arterial wall, cholesteryl ester is either derived from the infiltration of lipoprotein-associated cholesteryl ester resulting from LCAT activity or synthesized *in situ* as a result of ACAT activity. The fatty acid profile of the cholesteryl ester in arterial plaque can provide an indication of its origin.

Historically, dietary cholesterol has been associated with elevated LDL cholesterol concentrations and cardiovascular disease risk. However, within the range currently consumed in Western countries and on the basis of data indicating that dietary fat type has a greater effect on cardiovascular disease risk factors than dietary cholesterol, the emphasis has shifted (Carson et al., 2020). Hence, for the general public, target intakes have not been included in updated dietary guidance, however, recommendations not to increase intake have been included.

Other sterols

Oils derived from plants contain a wide range of phytosterols, compounds structurally similar to cholesterol. The difference between phytosterols and cholesterol is related to their side-chain configuration and steroid ring double bonds. The most common dietary phytosterols found in edible oils are β-sitosterol, campesterol, and stigmasterol (**Fig. 6**). In contrast to cholesterol, phytosterols are absorbed in trace amounts. Because they compete with cholesterol for incorporation into intestinal micelles, they decrease the bioavailability of intestinal cholesterol, hence, its absorption efficiency. For this reason, plant sterols have been used therapeutically to reduce elevated plasma LDL cholesterol concentrations (Katan et al., 2003). They have been shown to have an additive effect to that of other LDL cholesterol lowering therapies, such as statins (Han, 2016). The concentration of plant sterols in plant oils is relative low. For use to treat hypercholesterolemia, doses only achievable with supplements, either from capsules or function foods, are efficacious.

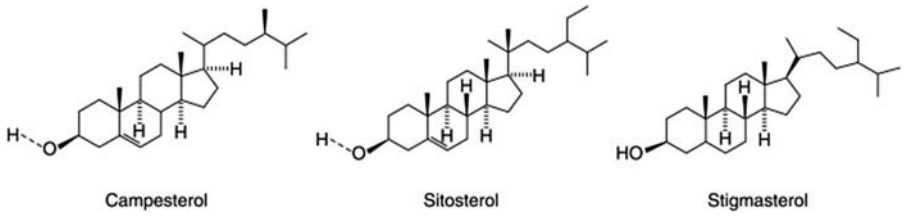


Fig. 6 Plant sterols and stanols.

Dietary fats and oils

Fatty acid profile of common dietary fats

Dietary fats and oils come from both animal and plant sources, primarily in the form of triacylglycerol. The fatty acid profile of commonly consumed dietary fats varies considerably (Fig. 7). In general, fats of animal origin tend to be relatively high in saturated fatty acids, contain cholesterol, and are solid at room temperature. Oils of plant origin tend to be relatively high in unsaturated fatty acids (monounsaturated and polyunsaturated) and are liquid at room temperature. Notable exceptions include tropical oils (e.g., palm, palm kernel, coconut oils) and partially-hydrogenated fat. Tropical oils are high in saturated fatty acids but remain liquid at room temperature because they contain a high proportion of very short-chain fatty acids. Partially-hydrogenated fat has a high proportion of fatty acids containing *trans* double bonds. The larger bond angle of *trans* double bonds compared to *cis* double bonds confers a straighter acyl chain conformation, more similar to saturated than unsaturated fatty acids. This results in a consistency comparable to solid animal fats than liquid plant oils.

Major contributors of dietary saturated, monounsaturated and polyunsaturated fatty acids, and cholesterol

The major types of dietary fats and oils are generally broken down on the basis of animal and plant sources. The relative balance of animal to plant foods is an important determinant of the fatty acid profile of the diet. However, with the increasing prominence of processed, ultraprocessed, reformulated, and genetically modified foods, it is becoming difficult to predict the fatty acid profile of the diet on the basis of the animal verses plant distinction.

According to the National Health and Nutrition Examination Survey (NHANES) recall data, the 10 major dietary sources of saturated fatty acids in the US diet are regular cheese, whole milk, regular ice cream, 2% low-fat milk, pizza with meat, French fries, Mexican dishes with meat, regular processed meat, chocolate candy, and mixed dishes with beef (Table 2). Hence, regular dairy products contribute the majority (~16%) of saturated fatty acids in the diet, and the top 10 sources contribute ~30% of the saturated fatty acids consumed. The increased prevalence of fat-free and low-fat dairy products provides a viable option with which to encourage a population-wide decrease in saturated fat intake. To put the value of decreasing population-wide intakes of saturated fat into perspective, it has been estimated that the isocaloric replacement of 5% of energy from saturated fatty acids with complex carbohydrate, on average, would reduce total cholesterol concentrations by 10 mg dL^{-1} (0.26 mmol L^{-1}) and LDL cholesterol by 7 mg dL^{-1} (0.18 mmol L^{-1}). For a person at moderately high risk of developing cardiovascular disease with total cholesterol concentration of 220 mg dL^{-1} (5.69 mmol L^{-1}) and LDL cholesterol concentration of 140 mg dL^{-1} (3.62 mmol L^{-1}), such a dietary modification would decrease total and LDL cholesterol concentrations by 4.5% and 5%, respectively. Each 1% decrease in total cholesterol concentrations has been associated with a 2% reduction in the incidence of coronary heart disease. Using this example, such a difference would theoretically translate into a 9% decrease in cardiovascular disease risk. However, it is important to note that decreasing the saturated fatty acid content of the diet should not necessarily be done by replacing the fat with carbohydrate (Jakobsen et al., 2009). As will be discussed in the next section, the quantity of dietary fat, relative to carbohydrate and protein, also

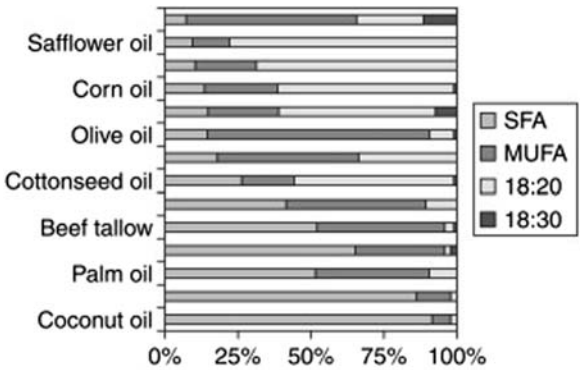


Fig. 7 Relative composition of common dietary fats.

Table 2 Ten major sources of saturated, monounsaturated, and polyunsaturated fatty acids, and cholesterol in the US diet.[a]

Saturated fatty acids	*Monounsaturated fatty acids*	*Polyunsaturated fatty acids*	*Cholesterol*
Cheese, regular	Fried potatoes	Salad dressing, regular	Eggs, fried
Whole milk	Processed meat, regular	White bread, regular	Eggs, regular, including scrambled
Ice cream, regular	Cookies, regular	Mayonnaize, regular	Eggs, mixed dishes
2% milk	Snacks, regular	Fried potatoes	Beef, mixed dishes
Pizza with meat	Pizza with meat	Cakes, regular	Whole milk
Fried potatoes	Salad dressing, regular	Cookies, regular	Cheese, regular
Burritos/tacos with meat	Cheese, regular	Chicken/Turkey, mixed dishes	Fish, fried
Processed meat, regular	Burritos/tacos with meat	Snacks, regular	Chicken/Turkey, mixed dishes
Chocolate candy	Sausage	Potato chips, regular	Beef, lean/trimmed
Beef, mixed dishes	Beef, mixed dishes	Fish, fried	Processed meat, regular

[a]Ranks of food for adults older than 19 years from the NHANES recall data 1999–2000.

impacts on blood lipid concentrations and lipoprotein profiles. Current data suggest that replacing saturated fatty acids with polyunsaturated fatty acids would result in the greatest decrease in cardiovascular disease risk (Sacks et al., 2017).

The 10 major dietary sources of monounsaturated fatty acids in the US diet are French fries, regular processed meat, regular cookies, regular miscellaneous snacks, pizza with meat, regular salad dressing, regular cheese, Mexican dishes with meat, sausage, and mixed dishes with beef (Table 2).

The 10 major dietary sources of *n*-6 polyunsaturated fatty acids in the US diet are regular salad dressing, regular white bread, regular mayonnaize, French fries, regular cake, regular cookies, mixed dishes with chicken and turkey, regular miscellaneous snacks, regular potato chips, and fried fish (Table 2). The distribution of polyunsaturated fatty acids among commonly consumed foods is wide.

The 10 major dietary sources of cholesterol in the US diet are fried eggs, regular eggs including scrambled eggs, mixed dishes with eggs, mixed dishes with beef, whole milk, regular cheese, fried fish, mixed dishes with chicken and turkey, lean cut meat, and regular processed meat (Table 2). Eggs or foods high in eggs contribute ~30% of dietary cholesterol.

Dietary fat and cardiovascular disease prevention

Quantity of dietary fat

When considering the percentage of energy contributed by dietary fats and oils (amount of fat) and cardiovascular disease prevention and management, there are two major factors to consider, impact on body weight and plasma lipoprotein profiles. The potential relation with body weight is important because overweight and obesity are strongly associated with elevated dyslipidemia, hypertension, and type 2 diabetes. These factors are all associated with elevated cardiovascular disease risk. Plasma lipoprotein concentrations that are responsive to dietary modification include LDL and HDL cholesterol, the lipid triglyceride, and the ratio of total cholesterol to HDL cholesterol. When body weight is maintained at a constant level, decreasing the total fat content of the diet, expressed as a percentage of total energy, and replacing it with carbohydrate frequently results in an increase in triglyceride concentrations, decrease in HDL cholesterol concentrations, and higher (less favorable) total cholesterol/HDL cholesterol ratio. Low HDL cholesterol concentrations are an independent risk factor for cardiovascular disease. Very low fat diets are of particular concern to individuals with glucose intolerance and excess body weight who have a predisposition to low HDL cholesterol and high triglyceride concentrations or those individuals with metabolic syndrome (having three or more of the following: abdominal obesity, elevated triacylglycerol concentrations, low HDL concentrations, elevated blood pressure, elevated fasting glucose concentrations). Because of these findings, dietary guidance for cardiovascular disease risk reduction has shifted from low to moderate fat diets, with emphasis shifted to replacing saturated fat with unsaturated fat, particularly polyunsaturated fat. These modifications have been echoed in more recent updates of these guidelines.

With respect to the quantity of dietary fats and oils and body weight, comprehensive reviews of the long-term data have concluded that energy balance, not level or ratio of macronutrients, is the strongest determinant of body weight. Two relatively recent intervention study has concluded that there is no advantage, with respect to weight loss or cardiovascular disease risk indicators, of diets with different proportions of fat, carbohydrate, and protein (Gardner et al., 2018; Del Gobbo et al., 2018). The major determinant of successful weight loss was adherence to the protocol.

Quality of dietary fat

Early evidence demonstrated that diets relatively high in saturated fatty acids increase plasma total cholesterol concentrations. Subsequent work demonstrated that this elevation in total cholesterol concentrations is contributed to by increases in both LDL and HDL cholesterol concentrations, the former more so than the latter. Replacing saturated fatty acids with unsaturated fatty acids, monounsaturated or polyunsaturated fatty acids, is associated with lower LDL cholesterol concentrations and cardiovascular risk,

with polyunsaturated having a somewhat greater effect than monounsaturated (Hu and Willett, 2002; Sacks et al., 2017). Replacing saturated fatty acids with carbohydrate, particularly refined carbohydrate, elevates triglyceride and lowers HDL cholesterol concentrations. Observational data indicate that dietary patterns low in saturated fatty acids and high in polyunsaturated fatty acids are associated with the lowest cardiovascular disease risk (Blondin et al., 2019).

Quantitatively, α-linolenic acid (ALA, 18:3*n*-3) is the most abundant *n*-3 fatty acid in the diet. Two other *n*-3 polyunsaturated fatty acids, sometimes referred to as the very long chain *n*-3 fatty acids, eicosapentaenoic acid (EPA, 20:5*n*-3) and docosahexaenoic acid (DHA, 22:6*n*-3), are present in smaller amounts. The major source of dietary ALA is canola and soybean oils. The major sources of dietary EPA and DHA is fish and seafood, specifically dark flesh fish such as salmon and mackerel. Dietary patterns higher in fish are associated with lower cardiovascular disease risk. Likewise, dietary patterns rich in oils high in ALA, such as soybean and canola oils (**Fig. 7**), are associated with lower cardiovascular disease risk. Although humans have the ability to elongate and desaturate ALA to form EPA and subsequently DHA, the capacity is low. Data on the benefits of supplementation with EPA and DHA have been inconsistent and more data is likely to emerge in the near future. Whether the benefit of diets rich in n-3 fatty acids on cardiovascular disease risk is due solely to direct effects of ALA, EPA and DHA or also to the additional benefits of dietary patterns rich in these fatty acids has yet to be determined. Current recommendations are to consume at least two fish meals per week, particularly those rich in EPA and DHA (salmon, mackerel) and replace fats rich in saturated fat with unsaturated fat, including those rich in ALA.

Trans fatty acids occur naturally in meat and dairy products as a result of anaerobic bacterial fermentation in ruminant animals (Mensink and Katan, 1990). *Trans* fatty acids are also formed as a result of the partial hydrogenation of plant oils. Oils are partially hydrogenated to increase viscosity (change a liquid oil into a semiliquid or solid) and extend shelf life (decrease susceptibility to oxidation). *Trans* fatty acid intake, regardless of the source, is associated with elevated LDL cholesterol concentrations and cardiovascular disease risk (Lichtenstein et al., 1999). Major contributors of dietary *trans* fatty acids had been commercially baked products, animal fats, traditional margarines and shortenings, and commercially fried foods. Mandatory inclusion of *trans* fatty acid content on Nutrient Facts Panels and removal of partially-hydrogenated fat from the FDA Generally Recognized As Safe list (GRAS) has resulted in secular decreased intake (Dahmubed et al., 2013). Replacing animal fat with plant oils resulted in a decrease in *trans* fatty acid intake from ruminant fat.

Composition of dietary fats

Types of fat relatively high in saturated fatty acids include butterfat (62%), beef fat (50%), tropical oils (coconut 87%, palm kernel 81%, palm oil 49%), and lard (pork) fat (39%) (**Fig. 7**). The content of cholesterol in these fats is 33, 14, 0, and 12 mg tablespoon^{-1}, respectively. Types of fat relatively high in monounsaturated fatty acids include canola oil (56%), olive oil (73%), and peanut oil (46%). Types of fat relatively high in polyunsaturated fatty acids include soybean oil (51%), corn oil (58%), and unmodified safflower oil (74%), and sunflower oil (66%). Plant oils do not contain cholesterol.

Dietary guidance

The US government in the form of Dietary Guidelines for Americans and health advocacy organizations such as the American Heart Association and American Diabetes Association issue and update dietary guidance, including that for fats and oils (Lichtenstein et al., 2006; Evert et al., 2019). In general, current recommendations are to consume a diet moderate in total fat and rich in fruits and vegetables, whole-grain products, fish, low-fat and nonfat dairy products, legumes and nuts, and lean meat and poultry. Liquid plant oils are recommended in place of other types of fats (animal fat, partially-hydrogenated fat). Much dietary guidance now emphasizes the benefits, from a nutrition and environmental perspective, of plant-based foods. Important for any type of dietary guidance, especially when it is intended promote shifts in dietary patterns, is to take availability, affordability and personal preference, including regional, cultural and ethnic, into consideration. Also important is regardless of specific foods choices, that total energy intakes achieve and maintain a healthy body weight.

Summary

Dietary fats and oils have both positive and negative attributes with respect to health outcomes. Fats and oils are made up primarily of triacylglycerol. The fatty acid profile of the triacylglycerol dictates the physical properties of the fat. During fatty acid biosynthesis, humans are unable to insert a double bond above the ninth carbon of the acyl chain. For this reason, linoleic acid and fatty acids derived from linoleic acid are classified as essential; hence, these must be consumed preformed. Animal fats are the major contributors of dietary saturated fatty acids. Plant oils, such as canola and olive, are the major contributors of dietary monounsaturated fatty acids. Plant oils, such as soybean and corn oils, are the major contributors of dietary polyunsaturated fatty acids. The plant oils canola and olive, and foods of marine origin (fish and seafood) are major contributors of *n*-3 fatty acids ALA, and EPA and DHA, respectively. Partially-hydrogenated fat has been major contributors of *trans* fatty acids; changes to labeling and regulatory policy has resulted in a dramatic divine in the food supply. Dietary patterns high in polyunsaturated and low in saturated fatty acids have been associated with optimal health outcomes. Very long chain *n*-3 fatty acids have been independently associated with

reduced cardiovascular disease risk. *Trans* fatty acids have been associated with elevated cardiovascular disease risk (Ascherio et al., 1999). Dietary fatty acid intakes are determined by the sum of individual foods. Current dietary guidance from the US government and major health advocacy organization recommend moderate-fat diets rich in fruits and vegetables, whole-grain products, fish, low-fat and nonfat dairy products, legumes and nuts, and unprocessed lean meat and poultry. There are multiple dietary patterns consistent with these recommendations that can accommodating personal preferences.

References

Ascherio, A., Katan, M.B., Zock, P.L., Stampfer, M.J., Willett, W.C., 1999. Trans fatty acids and coronary heart disease. N. Engl. J. Med. 340 (25), 1994–1998.

Blondin, S.A., Campbell, W.W., Emlen, E., Guasch-Ferré, M., Hu, F.B., Janiszewski, M., O'Connor, L.E., Satija, A., Stampfer, M.J., Willett, W.C., 2019. Meta-analysis of randomized controlled trials of red meat consumption in comparison with various comparison diets on cardiovascular risk factors. Circulation 139 (15), 1828–1845.

Carson, J.A.S., Lichtenstein, A.H., Anderson, C.A.M., Appel, L.J., Kris-Etherton, P.M., Meyer, K.A., Petersen, K., Polonsky, T., Van Horn, L., American Heart Association Nutrition Committee of the Council on Lifestyle and Cardiometabolic Health, Council on Arteriosclerosis, Thrombosis and Vascular Biology, Council on Cardiovascular and Stroke Nursing, Council on Clinical Cardiology, Council on Peripheral Vascular Disease, Stroke Council, 2020. Dietary cholesterol and cardiovascular risk: a science advisory from the American Heart Association. Circulation 141 (3), e39–e53.

Dahmubed, A., Jacobson, M.F., Mozaffarian, D., Otite, F.O., 2013. Trends in *trans* fatty acids reformulations of US supermarket and brand-name foods from 2007 through 2011. Prev. Chronic Dis. 10, E85.

Del Gobbo, L.C., Desai, M., Gardner, C.D., Hauser, M.E., Ioannidis, J.P.A., King, A.C., Rigdon, J., Trepanowski, J.F., 2018. Effect of low-fat vs low-carbohydrate diet on 12-month weight loss in overweight adults and the association with genotype pattern or insulin secretion: the DIETFITS randomized clinical trial. J. Am. Med. Assoc. 319, 667–679.

Evert, A.B., Dennison, M., Gardner, C.D., Garvey, W.T., Lau, K.H.K., MacLeod, J., Mitri, J., Pereira, R.F., Rawlings, K., Robinson, S., Saslow, L., Uelmen, S., Urbanski, P.B., Yancy Jr., W.S., 2019. Nutrition therapy for adults with diabetes or prediabetes: a Consensus Report. Diabetes Care 42 (5), 731–754.

Gardner, C.D., Trepanowski, J.F., Del Gobbo, L.C., Hauser, M.E., Rigdon, J., Ioannidis, J.P.A., Desai, M., King, A.C., 2018. Effect of low-fat vs low-carbohydrate diet on 12-month weight loss in overweight adults and the association with genotype pattern or insulin secretion: the DIETFITS randomized clinical trial. J. Am. Med. Assoc. 319 (7), 667–679.

Han, S., Jiao, J., Xu, J., Zimmermann, D., Actis-Goretta, L., Guan, L., Zhao, Y., Qin, L., 2016. Effects of plant stanol or sterol-enriched diets on lipid profiles in patients treated with statins: systematic review and meta-analysis. Sci. Rep. 6, 31337.

Hu, F.B., Willett, W.C., 2002. Optimal diets for prevention of coronary heart disease. J. Am. Med. Assoc. 288, 2569–2578.

Jakobsen, M.U., O'Reilly, E.J., Heitmann, B.L., Pereira, M.A., Balter, K., Fraser, G.E., Goldbourt, U., Hallmans, G., Knekt, P., Liu, S., Pietinen, P., Spiegelman, D., Stevens, J., Virtamo, J., Willett, W.C., Ascherio, A., 2009. Major types of dietary fat and risk of coronary heart disease: a pooled analysis of 11 cohort studies. Am. J. Clin. Nutr. 89 (5), 1425–1432.

Katan, M.B., Grundy, S.M., Jones, P., Law, M., Miettinen, T., Paoletti, R., Participants Stresa Workshop, 2003. Efficacy and safety of plant stanols and sterols in the management of blood cholesterol levels. Mayo Clin. Proc. 78 (8), 965–978.

Lichtenstein, A.H., Ausman, L.A., Jalbert, S.M., Schaefer, E.J., 1999. Comparison of different forms of hydrogenated fats on serum lipid levels in moderately hypercholesterolemic female and male subjects. N. Engl. J. Med. 340, 1933–1940.

Lichtenstein, A.H., Appel, L.J., Brands, M., Carnethon, M., Daniels, S., Franch, H.A., Franklin, B., Kris-Etherton, P., Harris, W.S., Howard, B., Karanja, N., Lefevre, M., Rudel, L., Sacks, F., Van Horn, L., Winston, M., Wylie-Rosett, J., 2006. Diet and Lifestyle Recommendations Revision 2006: a scientific statement from the American Heart Association Nutrition Committee. Circulation 114 (1), 82–96.

Mensink, R.P., Katan, M.B., 1990. Effect of dietary *trans* fatty acids on high-density and low-density lipoprotein cholesterol levels in healthy subjects. N. Engl. J. Med. 323 (7), 439–445.

Sacks, F.M., Lichtenstein, A.H., Wu, J.H.Y., Appel, L.J., Creager, M.A., Kris-Etherton, P.M., Miller, M., Rimm, E.B., Rudel, L.L., Robinson, J.G., Stone, N.J., Van Horn, L.V., American Heart Association, 2017. Dietary fats and cardiovascular disease: a presidential advisory from the American Heart Association. Circulation 136, e1–e23.

Further reading

Bhupathiraju, S.N., Chiuve, S.E., Hu, F.B., Manson, J.E., Rexrode, K.M., Rimm, E.B., Satija, A., Spiegelman, D., Willett, W., 2017. Healthful and unhealthful plant-based diets and the risk of coronary heart disease in U.S. Adults. J. Am. Coll. Cardiol. 70 (4), 411–422.

Manson, J.E., Cook, N.R., Lee, I.M., Christen, W., Bassuk, S.S., Mora, S., Gibson, H., Albert, C.M., Gordon, D., Copeland, T., D'Agostino, D., Friedenberg, G., Ridge, C., Bubes, V., Giovannucci, E.L., Willett, W.C., Buring, J.E., VITAL Research Group, 2019. Marine n-3 fatty acids and prevention of cardiovascular disease and cancer. N. Engl. J. Med. 380 (1), 23–32.

Manson, J.E., Cook, N.R., Lee, I.M., Christen, W., Bassuk, S.S., Mora, S., Gibson, H., Gordon, D., Copeland, T., D'Agostino, D., Friedenberg, G., Ridge, C., Bubes, V., Giovannucci, E.L., Willett, W.C., Buring, J.E., VITAL Research Group, 2019. Vitamin D supplements and prevention of cancer and cardiovascular disease. N. Engl. J. Med. 380 (1), 33–44. https://doi.org/10.1056/NEJMoa1809944.

Sacks, F.M., Bray, G.A., Carey, V.J., Smith, S.R., Ryan, D.H., Anton, S.D., McManus, K., Champagne, C.M., Bishop, L.M., Laranjo, N., Leboff, M.S., Rood, J.C., de Jonge, L., Greenway, F.L., Loria, C.M., Obarzanek, E., Williamson, D.A., 2009. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. N. Engl. J. Med. 360 (9), 859–873.

Virani, S.S., Alonso, A., Aparicio, H.J., Benjamin, E.J., Bittencourt, M.S., Callaway, C.W., Carson, A.P., Chamberlain, A.M., Cheng, S., Delling, F.N., Elkind, M.S.V., Evenson, K.R., Ferguson, J.F., Gupta, D.K., Khan, S.S., Kissela, B.M., Knutson, K.L., Lee, C.D., Lewis, T.T., Liu, J., Loop, M.S., Lutsey, P.L., Ma, J., Mackey, J., Martin, S.S., Matchar, D.B., Mussolino, M.E., Navaneethan, S.D., Perak, A.M., Roth, G.A., Samad, Z., Satou, G.M., Schroeder, E.B., Shah, S.H., Shay, C.M., Stokes, A., VanWagner, L.B., Wang, N.Y., Tsao, C.W., American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee, 2021. Heart disease and stroke statistics-2021 update: a Report From the American Heart Association. Circulation 143 (8), e254–e743.

Lipoproteins

J.M. Ordovas, Tufts University, Boston, MA, USA

© 2013 Elsevier Ltd. All rights reserved.

TThis is a reproduction of J.M. Ordovas, Lipoproteins, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 80–86, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00171-9.

Classification of Lipoproteins **301**
Classification of Serum Lipoproteins According to Their Electrophoretic Mobilities 301
Classification of Serum Lipoproteins According to Their Ultracentrifugal Characteristics 302
Classification of Serum Lipoproteins According to Their Apolipoprotein Composition 302
Synthesis and Catabolism of Lipoproteins **302**
Metabolism of Lipoproteins Carrying Exogenous Lipids 302
Transport of Endogenous Lipids 303
Reverse Cholesterol Transport 304
Effects of Dietary Fats and Cholesterol on Lipoprotein Metabolism **304**
Effects of Diet on Chylomicron Metabolism 304
Effects of Diet on VLDL Metabolism 304
Effects of Diet on LDL Metabolism 305
Effects of Diet on HDL Metabolism 305
Disorders of Lipoprotein Metabolism **305**
Type I or Familial Chylomicronemia 305
Type II or Familial Hypercholesterolemia 305
Familial Defective apo B-100 306
Type III or Familial Dysbetalipoproteinemia 306
Familial Type IV and Type V Hypertriglyceridemias 306
Familial Dyslipidemia 306
Familial Combined Hyperlipidemia 306
Familial Hyperapobetalipoproteinemia 307
Familial Hypoalphalipoproteinemia 307
Familial Lipoprotein (a) Excess 307
General Guidelines for the Treatment of Lipoprotein Abnormalities for CHD Prevention **307**
Further Reading **307**

Classification of Lipoproteins

Classification of Serum Lipoproteins According to Their Electrophoretic Mobilities

With the development of techniques to separate proteins according to their electrophoretic behavior, it could be demonstrated that most of the lipid present in serum was associated with proteins migrating with α_1- and β-globulin mobilities. This resulted in the first classification of lipoproteins as α_1- and β-lipoproteins. The ratio of lipid to protein on the α_1-lipoproteins was approximately 1:1, whereas the β-lipoproteins had a greater relative content of lipids. Application of more advanced electrophoretic techniques resulted in further discrimination among the lipoprotein classes and for many years lipoproteins were classified as β-, pre-β-, and α-lipoproteins. Careful observation of the electrophoretic lipoprotein profiles in normals and subjects with familial lipoprotein

Table 1 Classification of plasma lipoproteins

Lipoprotein	*Diameter (nm)*	*Density (g ml^{-1})*	*Electrophoretic mobility*	*Major lipids*	*Major apolipoproteins*
Chylomicrons	80–500	<0.95	Origin	Dietary triacylglycerols, cholesteryl esters	A-I, A-II, A-IV, B-48, C-I, C-II, C-III, E
Remnants	>30	<1.006	Origin	Dietary cholesteryl esters	B-48, E
VLDL	30–80	<1.006	pre-β	Endogenous triacylglycerols	B-100, C-I, C-II, C-III, E
IDL	25–35	1.006–1.019	pre-β and β	Cholesteryl esters, triacylglycerols	B-100, E
LDL	18–28	1.019–1.063	β	Cholesteryl esters	B-100
HDL_2	9–12	1.063–1.125	α	Cholesteryl esters, phospholipids	A-I, A-II
HDL_3	5–9	1.125–1.210	α	Cholesteryl esters, phospholipids	A-I, A-II

disorders gave rise to the first classification of lipoprotein disorders by Fredrickson and colleagues. The equivalence between electrophoretic and ultracentrifugal separation is presented in Table 1.

Several electrophoretic supports have been used to separate plasma lipoproteins. These include paper, cellulose acetate, agarose, and polyacrylamide. Agarose gel electrophoresis remains the most commonly used for easy and rapid assessment of lipoprotein patterns in the clinical laboratory. This technique is especially useful for identifying the presence of a broad β band in the diagnosis of type III hyperlipidemia. Gradient agarose–polyacrylamide gel electrophoresis under nondenaturing conditions has been an essential tool to analyze low-density lipoprotein (LDL) and high-density lipoprotein (HDL) subclasses, providing a greater resolution than ultracentrifugation. LDL subfractions have been resolved by nondenaturing polyacrylamide gradient gel electrophoresis (2–16%) in up to seven LDL subclasses with densities ranging from 1.020 to 1.063 g ml^{-1} and diameters ranging from 22.0 to 28.5 nm. Usually a major subpopulation and several (one–four) minor LDL subpopulations are found in most subjects examined. A predominance of smaller, more dense LDL, versus larger, more buoyant LDL particles in plasma has been associated with increased coronary heart disease (CHD) risk. There is evidence supporting the genetic origin of the distribution of LDL subfractions; however, age, gender, and environmental factors strongly influence the penetrance. HDL subfractions have been resolved using a similar technique, with a polyacrylamide gradient ranging from 4% to 30%, into five subclasses (HDL_{3c}, HDL_{3b}, HDL_{3a}, HDL_{2a}, and HDL_{2b}). More recently 11–14 subclasses have been described, including β-migrating particles, using an improved electrophoresis technique. The clinical importance of these subfractions is still under investigation.

Classification of Serum Lipoproteins According to Their Ultracentrifugal Characteristics

The presence of lipids within the lipoprotein particles confers these macromolecular complexes with a lower density compared with other serum proteins. With the arrival of the analytical ultracentrifugation in the 1940s, this characteristic allowed its initial separation as a discrete peak using this technique. During the following years, it was demonstrated that this fraction was made up of a wide spectrum of particle sizes and densities (d) ranging from 0.92 to 1.21 g ml^{-1}.

Lipoproteins were classically separated into four major classes designated as chylomicrons (exogenous triacylglycerol-rich particles of d<0.94 g ml^{-1}), very low-density lipoproteins (VLDL, endogenous triacylglycerol-rich particles of d=0.94–1.006 g ml^{-1}), LDL (cholesteryl ester-rich particles of d=1.006–1.063 g ml^{-1}), and HDL (particles containing approximately 50% protein of d=1.063–1.21 g ml^{-1}). With subsequent improvements to the ultracentrifugation techniques, further heterogeneity was detected within each of those major lipoprotein classes; this resulted in the need for further subdivision into several density subclasses such as HDL_{2a} (d=1.10–1.125 g ml^{-1}), HDL_{2b} (d=1.063–1.10 g ml^{-1}), and HDL_3 (d=1.125–1.21 g ml^{-1}).

There is no doubt that the separation of lipoproteins by ultracentrifugation has been essential for the advances in this field; however, this technique is very labor intensive and the isolated lipoproteins are usually modified due to the high g force and salt concentrations used in this process. The development of new vertical and near vertical rotors has shortened considerably the runs and thus diminished some of these negative effects.

Classification of Serum Lipoproteins According to Their Apolipoprotein Composition

Recent interest in the study of lipoprotein subfractions has resulted in an increased use of methods of separation based on affinity chromatography, especially those using immunoaffinity. Using columns containing antibodies against specific apolipoproteins (Table 2), a large number of HDL subpopulations have been resolved. Similarly, this technique allows the separation of several triacylglycerol-rich lipoproteins subfractions.

Lipoproteins containing apo A-I can be separated into two major species: those containing both apo A-I and apo A-II, known as LpAI:AII, and those containing apo A-I but not apo A-II (LpAI). Small numbers of particles containing apo A-II, but not apo A-I, have been detected in normal subjects; however, these particles could become predominant in the presence of rare genetic disorders associated with HDL deficiency. Another HDL species containing apo A-I and apo E is important in reverse cholesterol transport by transporting cholesterol from the cell membranes to the liver for elimination from the body.

Lipoproteins containing apo B consist of four lipoprotein families. Lipoproteins containing apo B only (Lp(B)) are cholesteryl ester-rich and are found primarily within the LDL density range, but they have also been detected within the VLDL range. Particles containing both apo B and apo C (LpB:C), apo B and apo E (LpB:E), and all three apolipoprotein groups (LpB:E:C) are triacylglycerol-rich and are found within the VLDL and intermediate-density lipoprotein (IDL) density range. The apo C and apo E contents decrease as density increases.

More recently, the affinity for lectins of Lp(a), a lipoprotein containing apo B-100 as well as an antigenically unique apolipoprotein (apo(a)), has been used to develop a new technique to measure the levels of this lipoprotein in plasma.

Synthesis and Catabolism of Lipoproteins

Metabolism of Lipoproteins Carrying Exogenous Lipids

Dietary fats absorbed in the intestine are packaged into large, triacylglycerol-rich chylomicrons for delivery through the bloodstream to sites of lipid metabolism or storage. These lipoproteins interact with lipoprotein lipase (LPL) and undergo lipolysis, forming chylomicron remnants. The major sites of LPL activity are adipose tissue, skeletal muscle, the mammary gland, and the

Table 2 Classification and properties of apolipoproteins

Apolipoprotein	*Amino acids*	*Tissue expression*	*Chromosomal localization*	*Functions*
apo A-I	243	Liver Intestine	11	Major structural component of HDL Ligand for HDL binding Activator of LCAT Reverse cholesterol transport
apo A-II	77	Liver	1	Structural component of HDL Activator of hepatic lipase
apo A-IV	377	Intestine Liver	11	Regulator of LPL activity Activator of LCAT Intestinal lipid absorption
apo B-48	2152	Intestine	2	Structural component of TRL Secretion of chylomicrons
apo B-100	4536	Liver	2	Structural
apo C-I	57	Liver Intestine	19	Activator of LCAT Inhibitor of the LRP
apo C-II	79	Liver Intestine	19	Activator of LPL
apo C-III	79	Liver Intestine	11	Inhibits LPL
apo D	169	Most tissues	3	Radical scavenger Reverse cholesterol transport Binding of heme-related compounds
apo E	299	Liver Macrophage	19	Ligand for the LDL receptor Ligand for the LRP Reverse cholesterol transport
apo(a)	Variable	Liver	6	?

myocardium. In these sites, the fatty acids from the triacylglycerols are used for storage, oxidation, or secretion back to the circulation. The triacylglycerol-depleted particles resulting from the lipolysis, known as chylomicron remnants, pick up apo E and cholesteryl ester from HDL and are rapidly taken up by the liver via a process mediated by the apo E receptor. This is a fast process and chylomicron particles are not usually present in the blood after a prolonged fasting period. The occurrence of chylomicronemia can be easily detected by the presence of a creamy supernatant floating on top of the plasma or serum kept for several hours at 4 °C.

Transport of Endogenous Lipids

The liver cell secretes triacylglycerol-rich VLDL, which can be converted first to IDL and then to LDL through lipolysis by a mechanism similar to that described for chylomicrons. The excess surface components are usually transferred to HDL, and the triacylglycerol-depleted VLDL becomes an IDL. Some of these particles may be taken up by the liver via an apo E receptor, whereas others are further depleted of triacylglycerols, becoming cholesteryl ester-enriched particles known as LDL, which contain apo B as their only apolipoprotein. Consumption of fat-rich meals or glucose enhances VLDL production.

Some primary causes of elevated VLDL or IDL levels are familial endogenous hypertriglyceridemia (type IV according to Fredrickson's classification) and familial dysbetalipoproteinemia (type III hyperlipidemia). Genetic mutations at the apo E gene locus are responsible for the type III phenotype. Some secondary causes for elevated VLDL levels are obesity, diabetes mellitus, and alcohol consumption, as well as the use of high doses of certain drugs (e.g., thiazide diuretics and estrogens). The presence of elevated levels of IDL has been associated with an increased atherosclerotic risk.

LDL particles are major carriers of cholesteryl ester in the blood. An LDL receptor that recognizes apo B-100 and apo E, but not apo B-48, allows the liver and other tissues to catabolize LDL. High-fat and high-cholesterol diets can decrease the activity of the LDL receptor, leading to increased levels of circulating LDL. These particles supply cholesterol to cells in the periphery for synthesis of cell membranes and steroid hormones. Modified or oxidized LDL can also be taken up by the scavenger receptor on macrophages in various tissues, including the arterial wall. This process is a potential initiator of foam cell formation and atherosclerosis.

Several LDL subclasses have been identified using gradient gel electrophoresis. Large, less dense LDL particles are commonly found in premenopausal women and men at low risk for CHD, whereas the small, more dense particles have been associated with a significant increased risk for myocardial infarction. The distribution of these particles appears to have a significant genetic component modulated by age and environmental factors.

Reverse Cholesterol Transport

HDL is synthesized by both the liver and the intestine. Its precursor form is discoidal in shape and matures in circulation as it picks up unesterified cholesterol from cell membranes and other lipids (phospholipid and triacylglycerol) and proteins (A-I, E, and C apolipoproteins) from triacylglycerol-rich lipoproteins (chylomicron and VLDL) as these particles undergo lipolysis. The cholesterol is esterified by the action of the lecithin–cholesterol acyltransferase (LCAT) and the small HDL_3 particle becomes a larger HDL_2 particle. The esterified cholesterol is either delivered to the liver or transferred by the action of cholesteryl ester transfer protein (CETP) to other lipoproteins (such as chylomicron, VLDL remnants, or LDL) in exchange for triacylglycerols. This cholesterol may then be taken up by the liver via receptors specific for these lipoproteins, or it can be delivered again to the peripheral tissues. The triacylglycerol received by HDL_2 is hydrolyzed by hepatic lipase and the particle is converted back to HDL_3, completing the HDL cycle in plasma. In the liver, cholesterol can be excreted directly into bile, converted to bile acids, or reutilized in lipoprotein production.

Several genetic disorders have been identified associated with low levels or total deficiency of HDL.

Effects of Dietary Fats and Cholesterol on Lipoprotein Metabolism

The cholesterolemic effects of dietary fatty acids have been extensively studied. The saturated fatty acids $C_{12:0}$, $C_{14:0}$, and $C_{16:0}$ have a hypercholesterolemic effect, whereas $C_{18:0}$ has been shown to have a neutral effect. Monounsaturated and polyunsaturated fatty acids in their most common *cis* configuration are hypocholesterolemic in comparison with saturated fatty acids. The effects of *trans* fatty acids on lipid levels are under active investigation. Our current knowledge shows that their effect is intermediate between those of saturated and unsaturated fats. The effect of dietary cholesterol on lipoprotein levels is highly controversial. This may be due in part to the dramatic interindividual variation in response to this dietary component. Specific effects of dietary fats and cholesterol on each lipoprotein fraction are the focus of other articles and they are only briefly summarized below and in **Table 3**.

Effects of Diet on Chylomicron Metabolism

Diets very high in saturated fat have been associated with increased postprandial chylomicrons and chylomicron remnants compared with diets rich in n-6 polyunsaturated fats; however, human experiments carried out using moderate to high fat intake have not shown significant effects of different types of dietary fat or dietary cholesterol on postprandial lipoproteins.

The effects of dietary carbohydrates on postprandial lipoproteins have also been studied. Most protocols have used diets very high in simple carbohydrates. In general, high carbohydrate intake has been associated with increased levels of fasting triacylglycerols and increased postprandial levels of chylomicrons and chylomicron remnants.

Effects of Diet on VLDL Metabolism

It is well-known that diets high in simple carbohydrate increase hepatic secretion of VLDL. This carbohydrate induction of hypertriglyceridemia is the source of the current controversy regarding the optimal diet for subjects at high risk for cardiovascular disease. Some authors have demonstrated that the increased hepatic triacylglycerol secretion induced by high-carbohydrate diets was not accompanied by parallel increases in apo B-100 secretion. In other words, the consumption of low-fat, high-carbohydrate diets did not affect the number of particles but resulted in larger, more triacylglycerol-enriched VLDL particles.

Intake of saturated fat results in an increased secretion of the number of VLDL particles by the liver, whereas the opposite effect is observed with polyunsaturated fat. Of special note are the dramatic effects on VLDL production found following high intakes of n-3 fatty acids. These diets are associated with marked decreases in triacylglycerol secretion by mechanisms not fully understood. It has been speculated that n-3 fatty acids may stimulate intracellular degradation of apo B in hepatocytes. Dietary cholesterol, within the physiological range, appears to play a minor role in hepatic VLDL production.

Table 3 Effects induced on the major lipoprotein fractions by different dietary components following isoenergetic replacement of saturated fatty acids

	MUFA	*PUFA n-6*	*PUFA n-3*	trans *FA*	*Simple carbohydrate*	*Carbohydrate plus fiber*
VLDL-C	≈	≈/↓	↓	↑	↑	≈
LDL-C	↓	↓	≈/↓	↑	↓	↓
HDL-C	≈/↑	≈/↓	↓	↓	↓	≈/↓

≈ Equivalent effect; ↓ concentration reduced; ↑ concentration increased.

Effects of Diet on LDL Metabolism

The effects of dietary fat and cholesterol on LDL metabolism have been extensively studied. However, the effects of dietary cholesterol are still highly controversial. Whereas some studies have demonstrated increased LDL production and decreased catabolism associated with high cholesterol intakes, others have failed to find such associations.

Replacement of saturated by polyunsaturated fats has been associated with decreased LDL apo B production in some studies, whereas in other studies, increased ratios of polyunsaturated to saturated fats resulted in increased LDL apo B catabolism. Unlike the effects described for VLDL metabolism, intake of n-3 fatty acids appears to play a minor role on LDL metabolism.

Effects of Diet on HDL Metabolism

Diets high in simple carbohydrates reduce HDL cholesterol levels. This effect appears to be mediated by increases in the catabolism of apo A-I; however, one study has also demonstrated an additional decrease in apo A-I production.

Disorders of Lipoprotein Metabolism

For historical reasons the classification of disorders of lipoprotein metabolism will be presented according to the classical Fredrickson's classification (Table 4).

Type I or Familial Chylomicronemia

This disorder is characterized by greatly elevated levels of exogenous triacylglycerols and it is the result of impaired lipolysis of chylomicrons due to a deficiency of LPL or its activator, the apo C-II. Several genetic mutations at the structural genes for both LPL and apo C-II have been reported. These are autosomal recessive traits. In the heterozygous state, subjects have normal to slightly elevated plasma triglycerides, whereas homozygotes have triacylglycerol levels that may exceed 1000 mg dl^{-1} in the fasting state. The diagnosis of the homozygous state takes place during the first years of life from the presence of recurrent abdominal pain and pancreatitis. Eruptive xanthomas and lipemia retinalis may also occur.

The recommended treatment includes a diet low in simple carbohydrates and with a fat content below 20% of total energy. The use of medium-chain triglycerides (MCT) has also been reported to be efficacious. Body weight should be maintained within the normal limits and alcohol consumption should be avoided.

Other secondary causes leading to the presence of chylomicrons in the fasting state include uncontrolled diabetes mellitus, alcoholism, estrogen use, and hypothyroidism.

Fasting chylomicronemia has not been clearly associated with increased risk for atherosclerosis; however, there is considerable evidence supporting the atherogenic properties of chylomicron remnants.

Type II or Familial Hypercholesterolemia

Familial hypercholesterolemia (FH) is an autosomal dominant disorder characterized by elevation of plasma LDL cholesterol levels. Mutations at the LDL receptor gene locus on chromosome 19 are responsible for this disorder. Multiple different mutations have been described at this locus resulting in the FH phenotype. In the heterozygous state, subjects develop tendinous xanthomas, corneal arcus, and CHD. Elevations of LDL can result from well-characterized genetic disorders such as FH or familial defective apo B-100.

Table 4 Classification of hyperlipidemias according to Fredrickson

Type	*Plasma cholesterol*	*Plasma triacylglycerol*	*Lipoprotein fraction(s) affected*	*Atherosclerosis risk*	*Genetic disorder*
I	Normal to elevated	Very elevated	Chylomicrons	No	Familial LPL deficiency Apo C-II deficiency
IIa	Elevated	Normal	LDL	High	Familial hypercholesterolemia Familial combined hyperlipidemia Polygenic hypercholesterolemia
IIb	Elevated	Elevated	LDL and VLDL	High	Familial hypercholesterolemia Familial combined hyperlipidemia
III	Elevated	Very elevated	IDL	High	Familial dysbetalipoproteinemia
IV	Normal or elevated	Elevated	VLDL	Moderate	Familial hypertriglyceridemia Familial combined hyperlipidemia
V	Normal or elevated	Very elevated	VLDL and chylomicrons	Moderate	Familial hypertriglyceridemia

The ranges of LDL cholesterol levels in plasma of FH subjects are 200–400 mg dl^{-1} in heterozygotes and above 450 mg dl^{-1} in homozygotes. The frequency of defects at the LDL receptor locus is approximately 1 in 500 for the heterozygous state and 1 in a million in the homozygous state.

Inhibitors of 3-hydroxy-3-methylglutaryl (HMG) coenzyme A are useful in the treatment of hypercholesterolemia. Most pharmacological therapies are ineffective in the homozygous state. FH homozygotes may be treated with LDL apheresis, liver transplantation, and portacaval shunt. More recently, encouraging results have been obtained using *ex vivo* gene therapy.

The genetic defect(s) associated with a common form of hypercholesterolemia present in most subjects with cholesterol levels between 250 and 300 mg dl^{-1} has (have) not been elucidated. This disorder may be due to a combination of minor gene defects (i.e., presence of apo E-4 allele) that in combination with the environment (i.e., diet, lack of exercise) predispose individuals to moderately elevated LDL cholesterol levels. This disorder has been also named polygenic hypercholesterolemia.

Familial Defective apo B-100

Familial defective apo B-100 is an autosomal dominant genetic disorder that presents with a phenotype similar to FH. The frequency of this disorder may be similar to that of FH; however, it varies considerably depending on the ethnicity of the population studied. The specific mutation responsible for this disorder is a point mutation at amino acid 3500 of the mature apo B. The diagnosis of this disorder requires molecular biology techniques.

Type III or Familial Dysbetalipoproteinemia

In this disorder both plasma triacylglycerol and cholesterol are increased. Several mutations within the apo E gene locus have been found to be responsible for this disease; however, in most patients the complete expression of the clinical genotype needs additional interactions such as age, obesity, and diabetes. In addition to the accumulation in plasma of VLDL remnants and chylomicrons, other characteristics of this disorder are tuboeruptive xanthomas and in some cases also planar xanthomas. Therapies include diet and hypolipidemic agents such as fibrates, statin, or nicotinic acid. In most cases, diagnosis can be carried out first by agarose gel electrophoresis, followed by molecular biology techniques to detect the presence of the apo E-2 allele.

Familial Type IV and Type V Hypertriglyceridemias

These two disorders may have overlapping phenotypes. In type IV or familial endogenous hypertriglyceridemia, triacylglycerol levels are increased and HDL is usually decreased. This disorder appears to be autosomal dominant and relatively frequent in populations consuming high-fat diets. The precise molecular defect has not been defined; however, the increase in triacylglycerol is associated with overproduction of triacylglycerol by the liver and often with consequent reduced clearance. Diet should be the first step in therapy, followed if necessary by pharmacotherapy using fibrates or nicotinic acid. Premature CHD has been seen in some but not all cases presenting with this phenotype.

Type V hyperlipidemia is a much more rare disorder. Usually the first signs of this abnormality are abdominal pain or pancreatitis. VLDL levels are high and chylomicrons are present in the fasting state. This abnormality has not been linked to any specific molecular defect. Besides the primary genetic defect, other secondary causes of type V hyperlipidemia are poorly controlled diabetes mellitus, nephrotic syndrome, hypothyroidism, glycogen storage disease, and pregnancy. Recent data indicate increased susceptibility to atherosclerosis.

Familial Dyslipidemia

Familial dyslipidemia may be a variant of the familial hypertriglyceridemias described previously. It is characterized by hypertriglyceridemia in combination with low HDL cholesterol. Patients are generally overweight, with male pattern obesity, insulin resistance, diabetes, and hypertension. These subjects have both increased hepatic triacylglycerol secretion and increased HDL apo A-I catabolism.

Familial Combined Hyperlipidemia

Familial combined hyperlipidemia (FCH) was initially described as the combination of hypercholesterolemia and hypertriglyceridemia within the same kindred, and with kindred members having one of these abnormalities or both. Moreover, most subjects with FCH have HDL cholesterol levels below the 10th percentile. Affected subjects have elevation in VLDL, LDL, or both. This disorder has a frequency of approximately 10% in survivors of premature myocardial infarction (less than 60 years of age) and approximately 14% in kindred with CHD.

It has been reported that affected subjects have overproduction of apo B-100. The precise molecular defect has not been elucidated, although there are already several candidate gene loci, including the LPL. The expression of this disorder may be triggered by other factors, such as overweight, hypertension, diabetes, and gout. The treatment should include diet and exercise and, if necessary, niacin, HMC CoA reductase inhibitors, or fibrates, depending on the major lipid present in excess.

Familial Hyperapobetalipoproteinemia

Familial hyperapobetalipoproteinemia is characterized by apo B values above the 90th percentile in the absence of other lipid abnormalities; it has been suggested to be a variant of FCH. This disorder is relatively common (5%) in kindreds with premature CHD. The molecular defect is not known, but metabolic studies suggest overproduction of apo B-100.

Familial Hypoalphalipoproteinemia

Severe HDL deficiency, characterized by HDL cholesterol levels <10 mg dl^{-1} is rare and may be due to Tangier disease, apo A-I deficiencies, LCAT deficiency, or fish-eye disease. The apo A-1 deficiency states are due to rare deletions, rearrangements, or point mutations within the apo A-I/C-III/A-IV gene complex. Familial hypoalphalipoproteinemia is relatively common and is characterized by HDL cholesterol levels below the 10th percentile of normal. These subjects have been reported to have either decreased HDL production or increased HDL apo A-I catabolism. This phenotype is present in approximately 4% of kindred with premature CHD.

The genetic defect or defects are not known; however, it has been suggested that FCH, familial hyperapobetalipoproteinemia, familial dysbetalipoproteinemia, and familial hypoalphalipoproteinemia may be variants of a single disorder. This disorder is characterized by a genetic predisposition in subjects consuming high-fat, high-cholesterol diets to an increased secretion of apo B-containing lipoproteins and an increased catabolism of apo A-I-containing lipoproteins. The expression of the phenotype is usually enhanced by the presence of male pattern obesity.

Familial Lipoprotein (a) Excess

Lipoprotein (a) (Lp(a)) is an LDL particle with one molecule of apolipoprotein (a) attached to it. Elevated levels of Lp(a) (>35–40 mg dl^{-1} or 90th percentile) have been associated with premature CHD. This increased risk appears to result from two different mechanisms: cholesterol deposition in the arterial wall and inhibition of fibrinolysis.

Lp(a) concentrations are highly variable among individuals; however, they tend to remain constant during a person's lifetime. Between 80% and 90% of the variability appears to be of genetic origin, owing, for the most part, to variations at the structural apo(a) gene locus. Lp(a) concentrations are inversely associated with a size polymorphism of apo(a). This polymorphism is due to differences in the number of a multiple repeat of a protein domain highly homologous to the kringle 4 domain of plasminogen. Diets and medications used to lower LDL cholesterol levels do not appear to have a significant effect on Lp(a) concentrations; however, niacin has been reported to decrease Lp(a) levels. There have been reports suggesting that diets high in *trans* fatty acids have some raising effects on Lp(a) levels, whereas estrogen replacement lowers Lp(a) in postmenopausal women.

General Guidelines for the Treatment of Lipoprotein Abnormalities for CHD Prevention

There is a clear benefit from lowering LDL cholesterol with diet or drug therapy in patients with hyperlipidemia or CHD or both. Dietary therapy includes using diets that are restricted in total fat (<30% of calories), saturated fat (<7% of calories), and cholesterol (<200 mg day^{-1}). Pharmacological therapies include anion exchange resins, niacin, and HMG CoA reductase inhibitors. The latter agents have been demonstrated to also lower CHD mortality. It should be noted that dramatic interindividual variations have been demonstrated in response to diet and drug therapies. Consequently the efficacy of hypolipidemic therapies will vary from individual to individual. More information is needed about the benefits of HDL cholesterol raising in patients with low HDL cholesterol levels as well as the benefits of lowering triacylglycerol plasma concentrations, and more specifically the triacylglycerol carried in lipoprotein remnants. This is also true regarding the benefits of Lp(a) lowering using niacin in patients with elevated Lp(a) levels.

Further Reading

Alaupovic, P., 1996. Significance of apolipoproteins for structure, function, and classification of plasma lipoproteins. In: Bradley, W.A., Gianturco, S.H., Segrest, J.P. (Eds.), Methods in Enzymology, Plasma Lipoproteins, vol. 263. Academic Press, San Diego, pp. 32–60 (part C, Abstract).

Alwaili, K., Awan, Z., Alshahrani, A., Genest, J., 2010. High-density lipoproteins and cardiovascular disease: 2010 update. Expert Review of Cardiovascular Therapy 8 (3), 413–423.

Austin, M.A., Breslow, J.L., Hennekens, C.H., Buring, J.E., Willett, W.C., Krauss, R.M., 1988. Low density lipoprotein subclass patterns and risk of myocardial infarction. Journal of the American Medical Association 260, 1917–1921.

Calabresi, L., Franceschini, G., 2010. Lecithin: Cholesterol acyltransferase, high-density lipoproteins, and atheroprotection in humans. Trends in Cardiovascular Medicine 20 (2), 50–53.

Dallinga-Thie, G.M., Franssen, R., Mooij, H.L., et al., 2010. The metabolism of triglyceride-rich lipoproteins revisited: New players, new insight. Atherosclerosis 211 (1), 1–8.

Hegele, R.A., 2009. Plasma lipoproteins: Genetic influences and clinical implications. Nature Reviews Genetics 10 (2), 109–121.

Li, Z., McNamara, J.R., Ordovas, J.M., Schaefer, E.J., 1994. Analysis of high density lipoproteins by a modified gradient gel electrophoresis method. Journal of Lipid Research 35, 1698–1711.

National Cholesterol Education Program, 1994. Second Report of the Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel II). Circulation 89, 1329–1445.

Ordovas, J.M., 1991. Molecular biological approaches to the understanding of lipoprotein metabolism. In: Witiak, D.T., Newman, H.A.I., Feller, D.R. (Eds.), Medical, Chemical and Biochemical Aspects of Antilipidemic Drugs. Elsevier, Amsterdam, pp. 97–121.

Ordovas, J.M., 1993. Metabolism of triglyceride-rich lipoproteins: Genetic mutations associated with its pathology. Cardiovascular Risk Factors 3, 1–8.
Ordovas, J.M., 1994. Genetic and environmental factors: Effects on plasma lipoproteins. In: Serrano, Rios (Ed.), Dairy Products in Human Health and Nutrition. Balkema, Rotterdam, pp. 303–307.
Ordovas, J.M., Civeira, F., Genest, J., Schaefer, E.J., 1990. Genetic high density lipoprotein deficiency states. In: Lenfant, C., Albertini, A., Paoletti, R., Catapano, A. (Eds.), Atherosclerosis Reviews. Biotechnology of Dyslipoproteinemias. Application in Diagnosis and Control, vol. 20. Raven Press, New York, pp. 261–274.
Ordovas, J.M., Lopez-Miranda, J., Mata, P., Perez-Jimenez, F., Lichetenstein, A.H., Schaefer, E.J., 1995. Gene-diet interaction in determining plasma lipid response to dietary intervention. Atherosclerosis 118, S11–S27.
Schaefer, E.J., Genest Jr, J.J., Ordovas, J.M., Salem, D.N., Wilson, P.W.F., 1993. Familial lipoprotein disorders and premature coronary artery disease. Current Opinion in Lipidology 4, 288–298.
Schaefer, E.J., Lichtenstein, A.H., Lamon-Fava, S., McNamara, J.R., Ordovas, J.M., 1995. Lipoproteins, nutrition, aging, and atherosclerosis. American Journal of Clinical Nutrition 61 (supplement), 726S–740S.
Schaefer, E.J., Ordovas, J.M., 1992. Diagnosis and management of HDL deficiency states. In: Miller, N.E., Tall, A.R. (Eds.), High Density Lipoproteins and Atherosclerosis, vol. III. Elsevier, Amsterdam, pp. 235–251.
Virella, G., Lopes-Virella, M.F., 2008. Atherogenesis and the humoral immune response to modified lipoproteins. Atherosclerosis 200 (2), 239–246.
Zannis, V.I., Kardassis, D., Zanni, E.E., 1993. Genetic mutations affecting human lipoproteins, their receptors, and their enzymes. Advances in Human Genetics 21, 145–319.

Magnesium

L.H. Allen, University of California, Davis, CA, USA

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of L.H. Allen, Magnesium, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 131–135, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00177-X.

Magnesium Metabolism 309
Distribution of Mg within the Body 309
Intestinal Absorption 310
Urinary Excretion 310
Dietary Sources of Magnesium 311
Requirements 311
Assessment of Mg Status 311
Magnesium Deficit 312
Dietary Mg Deficiency 312
Secondary Mg Deficiency 312
Causes of Mg Deficit 312
Consequences of Mg Deficit and Implications in Various Metabolic Diseases 312
Magnesium Excess 313
Recommended Dietary Allowances 313
Conclusion 314
Further Reading 314

Glossary

Hypermagnesemia High levels of magnesium in blood or serum.

Hypermagnesuria High levels of magnesium in urine.

Magnesium An alkaline earth element with symbol Mg, atomic number 12, the seventh most abundant element in the earth's crust and the 11th most abundant in the human body.

Reabsorption Movement of filtered substances (such as magnesium) and water from the kidney tubules back into the plasma.

Magnesium (Mg), the second intracellular cation after sodium, is an essential mineral. It is a critical cofactor in more than 300 enzymatic reactions. It may be required for substrate formation (Mg-ATP) and enzyme activation. It is critical for a great number of cellular functions, including oxidative phosphorylation, glycolysis, DNA transcription, and protein synthesis. It is involved in ion currents and membrane stabilization. Mg deficiency may be implicated in various metabolic disorders, including cardiovascular diseases, immune dysfunction, and free radical damage.

Magnesium Metabolism

Distribution of Mg within the Body

The normal adult body contains approximately 25 g of Mg, with more than 60% in bone tissue (Table 1). Only a fraction of bone Mg (at the surface of the bone crystal) is exchangeable with extracellular Mg. The muscle contains 25% of total body Mg, and extracellular Mg accounts for only 1%. Plasma Mg is approximately 0.8 mmol l^{-1}, half of which is ionized and active in physiological reactions half bound to proteins or complexed to anions. In cells, Mg is associated with various structures, such as the nucleus and intracellular organelles, and free Mg accounts for 1–5% of total cellular Mg. Intracellular free Mg is maintained at a relatively constant level, even if extracellular Mg level varies. This phenomenon is due to the limited permeability of the plasma membrane to Mg and the existence of specific Mg transport systems that regulate the rates at which Mg is taken up by cells or extruded from cells. Mechanisms by which Mg is taken up by cells have not been completely elucidated, and Mg efflux particularly requires the antiport Na^+/Mg^{2+}. Various hormonal and pharmacological factors influence Mg transport, and it can be assumed that recent developments in molecular genetics will lead to the identification of proteins implicated in Mg transport.

Table 1 Magnesium in human tissues

	% distribution	*Concentration*
Bone	60–65	0.5% of bone ash
Muscle	27	6–10 mmol per kg wet weight
Other cells	6–7	6–10 mmol per kg wet weight
Extracellular	<1	
Erythrocytes		2.5 mmol l^{-1}
Serum		0.7–1.1 mol l^{-1}
Free	55	
Complexed	13	
Bound	32	
Mononuclear blood cells		2.3–3.5 fmol per cell
Cerebrospinal fluid		1.25 mmol l^{-1}
Free	55	
Complexed	45	
Sweat		0.3 mmol l^{-1} (in hot environment)
Secretions		0.3–0.7 mmol l^{-1}

Source: Reproduced from Vormann J (2003) Magnesium: Nutrition and metabolism. *Molecular Aspects of Medicine* 24: 27–37.

Intestinal Absorption

Net Mg absorption results from dietary Mg absorption and Mg secretion into the intestinal tract via bile and gastric and pancreatic juice. In healthy adults, 30–50% of dietary Mg is absorbed. The secreted Mg is efficiently reabsorbed and endogenous fecal losses are only 20–50 mg day^{-1}. Mg absorption occurs along the entire intestinal tract, but the distal small intestine (jejunum and ileum) is the primary site. It is essentially a passive intercellular process by electrochemical gradient and solvent drag. The active transport occurs only for extremely low dietary Mg intake and its regulation is unknown. Mg uptake in the brush border may be mediated by an Mg/anion complex, and Mg efflux across the basolateral membrane may involve Na^+/Mg^{2+} antiport systems. A gene implicated in Mg deficit in humans has been identified. It is expressed in intestine and kidney and appears to encode for a protein that combines Ca- and Mg-permeable channel properties with protein kinase activity. This gene may be implicated in Mg absorption. Because of the importance of the passive process, the quantity of Mg in the digestive tract is the major factor controlling the amount of Mg absorbed.

The possibility of an adaptive increase in the fraction of Mg absorbed as Mg intake is lowered is controversial. In fact, experimental studies indicate that fractional intestinal absorption of Mg is directly proportional to dietary Mg intake. Because only soluble Mg is absorbed, all the factors increasing Mg solubility increase its absorption while formation of insoluble complexes in the intestine may decrease Mg absorption. Most well-controlled studies indicate that high calcium intake does not affect intestinal Mg absorption in humans. In contrast, dietary phytate in excess impairs Mg absorption by formation of insoluble complexes in the intestinal tract. Negative effects of a high intake of dietary fiber have often been reported, but these actions have certainly been overestimated. In fact, only the impact of purified fiber was considered, but fiber-rich diets are a major source of Mg and roles of the intestinal fermentation and the large bowel in mineral absorption were neglected. It was demonstrated in animal models that fermentable carbohydrates (oligosaccharides and resistant starch) enhance Mg absorption in the large bowel and that a similar effect exists in humans. Other nutrients may influence Mg absorption but these effects are important only at low dietary Mg intake.

Urinary Excretion

Magnesium homeostasis is essentially regulated by a process of filtration–reabsorption in the kidney. Urinary Mg excretion increases when Mg intake is in excess, whereas the kidney conserves Mg in the case of Mg deprivation. Usually, 1000 mmol per 24 h of Mg is filtered and only 3 mmol per 24 h is excreted in urine.

A total of 10–15% of the filtered Mg is reabsorbed in the proximal tubule by a passive process. The majority of filtered Mg (65%) is reabsorbed in the thick ascending loop of Henle. The reabsorption in this segment is mediated by a paracellular mechanism involving paracellin-1. It is also related to sodium transport by a dependence on the transepithelial potential generated by NaCl absorption. Thus, factors that impair NaCl reabsorption in the thick ascending loop of Henle, such as osmotic diuretics, loop diuretics, and extracellular fluid volume expansion, increase Mg excretion. At least 10–15% of the filtered Mg is reabsorbed in the distal tubule. The reabsorption occurs via an active transcellular mechanism and is under the control of special divalent cation-sensing receptors. Thus, elevated plasma Mg concentrations inhibit reabsorption of Mg from the distal tubule, leading to an increased magnesuria. Other active transport may also exist because some hormones (parathyroid hormone (PTH), glucagon, calcitonin, and insulin) may increase Mg reabsorption. Other factors may also influence Mg reabsorption, such as hypercalciuria

or hypophosphatemia, which inhibit the tubular reabsorption of Mg. Metabolic alkalosis leads to renal Mg conservation, whereas metabolic acidosis is associated with urinary Mg wasting. Thus, the chronic low-grade metabolic acidosis in humans eating Western diets may contribute to decreased Mg status.

Dietary Sources of Magnesium

Mg is present in all foods, but the Mg content varies substantially (Table 2). Cereals and nuts have high Mg content. Vegetables are moderately rich in Mg, and meat, eggs, and milk are poor in Mg. A substantial amount of Mg may be lost during food processing, and refined foods generally have a low Mg content. In addition to Mg content, it is important to consider the Mg density of food (i.e., the quantity of Mg per unit of energy). Vegetables, legumes, and cereals thus contribute efficiently to daily Mg intake, whereas fat- and sugar-rich products have a minor contribution. Some water can also be a substantial source of Mg, but it depends on the area from which the water derives.

Requirements

Assessment of Mg Status

Several potential markers for estimating daily Mg requirement have been suggested. Plasma Mg concentration is the most commonly used marker to assess Mg status. In healthy populations, the plasma Mg value is 0.86 mmol l^{-1} and the reference value is 0.75–0.96 mmol l^{-1}. A low plasma Mg value reflects Mg depletion, but a normal plasma Mg level may coexist with low intracellular Mg. Thus, despite its interest, plasma Mg is not a good marker of Mg status.

Ion-specific electrodes have become available for determining ionized Mg in plasma, and this measurement may be a better marker of Mg status than total plasma Mg. However, further investigation is necessary to achieve a standardized procedure and to validate its use as an appropriate marker of Mg status.

Erythrocyte Mg level is also commonly used to assess Mg status, and the normal value is 2.06–2.54 mmol l^{-1}. However, erythrocyte Mg level is under genetic control, and numerous studies have shown no correlation between erythrocyte Mg and other tissue Mg.

The total Mg content of white blood cells has been proposed as an indicator of Mg status. However, lymphocytes, polymorphonuclear blood cells, and platelets may have protective mechanisms against intracellular Mg deficiency, and the determination of total Mg content in leukocytes and platelets to assess Mg status is of questionable usefulness.

Mg excretion determination is helpful for the diagnosis of Mg deficit when there is a hypomagnesemia. In healthy populations, the urinary Mg value is 4.32 mmol day^{-1} and the reference value is 1.3–8.2 mmol day^{-1}. In the presence of hypomagnesemia, normal or high urinary Mg excretion is suggestive of renal wasting. On the contrary, Mg urinary excretion lower than normal values is a convincing evidence of Mg deficiency.

The parenteral loading test is probably the best available marker for the diagnosis of Mg deficiency. The Mg retention after parenteral administration of Mg seems to reflect the general intracellular Mg content, and an Mg retention more than 20% of the administered Mg suggests Mg deficiency. However, this test is not valid in the case of abnormal urinary Mg excretion and is contraindicated in renal failure.

Table 2 Mg density of foods

Food	*Magnesium density (mg MJ^{-1})*
Vegetables (lettuce, broccoli)	211
Legumes (bean)	113
Whole cereal (wheat)	104
Nuts (almond)	105
Fruits (apple)	30
Fish (cod)	75
Meat (roast beef)	40
Whole milk	38
Cheese (camembert)	15
Eggs	18
Dessert	
Biscuit	10
Chocolate	52

Source: Reproduced from Répertoire Général des Aliments (1996) *Table de Composition Minérale*. Lavoisier, Paris: Tec & Doc.

Determination of exchangeable Mg pools using Mg stable isotopes is an interesting approach to evaluate Mg status. In fact, Mg exchangeable pool sizes vary with dietary Mg in animals. However, more studies are necessary to better appreciate the relationship between Mg status and exchangeable Mg pool size in humans.

Magnesium Deficit

Two types of Mg deficit must be differentiated. Dietary Mg deficiency results from an insufficient intake of Mg. Secondary Mg deficiency is related to dysregulation of the control mechanisms of Mg metabolism.

Dietary Mg Deficiency

Severe Mg deficiency is very rare, whereas marginal Mg deficiency is common in industrialized countries. Low dietary Mg intake may result from a low energy intake (reduction of energy output necessary for physical activity and thermoregulation, and thus of energy input) or from low Mg density of the diet (i.e., refined or processed foods). Moreover, in industrialized countries, diets are rich in animal source foods and low in vegetable foods. This leads to a dietary net acid load and thus a negative effect on Mg balance. In fact, animal source foods provide predominantly acid precursors (sulfur-containing amino acids), whereas fruits and vegetables have substantial amounts of base precursor (organic acids plus potassium salts). Acidosis increases Mg urinary excretion by decreasing Mg reabsorption in the loop of Henle and the distal tubule, and potassium depletion impairs Mg reabsorption. Mg deficiency treatment simply requires oral nutritional physiological Mg supplementation.

Secondary Mg Deficiency

Failure of the mechanisms that ensure Mg homeostasis, or endogenous or iatrogenic perturbing factors of Mg status, leads to secondary Mg deficit. Secondary Mg deficiency requires a more or less specific correction of its causal dysregulation.

Intestinal Mg absorption decreases in the case of malabsorption syndromes, such as chronic diarrhea, inflammatory enteropathy, intestinal resection, and biliary and intestinal fistulas.

Hypermagnesuria is encountered in the case of metabolic and iatrogenic disorders, such as primary and secondary hyperaldosteronism (extracellular volume expansion), hypercalcemia (competition Ca/Mg at the thick ascending loop of Henle), hyperparathyroidism, and phosphate or potassium depletion. Hypermagnesuria may also result from tubulopathy, as the selective defect of the Mg tubular reabsorption (chromosome 11q23), Bartter's syndrome (thick ascending loop of Henle), or Gitelman's syndrome (distal convoluted tubule).

Administration of medications can be a causal factor in the development of secondary Mg deficiency. Administration of diuretics is the main cause of iatrogenic deficit because it decreases NaCl reabsorption in the thick ascending loop of Henle and thus increases the fractional excretion of Mg.

Causes of Mg Deficit

Complex relations exist between Mg and carbohydrate metabolism. Diabetes is frequently associated with Mg deficit and insulin may play an important role in the regulation of intracellular Mg content by stimulating cellular Mg uptake. Hypomagnesemia is the most common ionic abnormality in alcoholism because of poor nutritional status and Mg malabsorption, alcoholic ketoacidosis, hypophosphatemia, and hyperaldosteronism secondary to liver disease.

Stress can contribute to Mg deficit by stimulating the production of hormones and thus increasing urinary Mg excretion and by impairing neurohormonal mechanisms that spare Mg.

Consequences of Mg Deficit and Implications in Various Metabolic Diseases

Mg deficit causes neuromuscular manifestations, including positive Chvostek and Trousseau signs, muscular fasciculations, tremor, tetany, nausea, and vomiting. The pathogenesis of the neuromuscular irritability is complex, and it implicates the central and peripheral nervous system, the neuromuscular junction, and muscle cells.

Mg deficit perturbs Ca homeostasis and hypocalcemia is a common manifestation of severe Mg deficit. Impaired release of PTH and skeletal end organ resistance to PTH appear to be the major factors implicated, probably by a decrease in adenylcyclase activity.

Perturbations in the action and metabolism of vitamin D may also occur in Mg deficit. Because Mg plays a key role in skeletal metabolism, Mg deficit may be a possible risk factor for osteoporosis. However, epidemiologic studies relating Mg intake to bone mass or rate of bone loss have been conflicting, and further investigation is necessary to clarify the role of Mg in bone metabolism and osteoporosis.

Hypokalemia is frequently encountered in Mg deficit. This is due to an inhibition of Na,K-ATPase activity that impairs K and Na transport in and out of the cell and to stimulation of renin and aldosterone secretion that increases K urinary excretion.

There is increasing evidence that Mg deficiency may be involved in the development of various pathologies. Mg deficit is frequent in diabetes and can be a factor in insulin resistance. It can modify insulin sensitivity, probably by influencing intracellular signaling

and processing. Mg deficit has also been implicated in the development or progression of micro- and macroangiopathy and neuropathy.

Mg deficit appears to act as a cardiovascular risk factor. Experimental, clinical, and epidemiological evidence points to an important role of Mg in blood pressure regulation. Mg deficit can lead to cardiac arrhythmias and to increased sensitivity to cardiac glucosides. Mg deficit may also play a role in the development of atherosclerosis. In experimental animal models, dietary Mg deficiency results in dyslipidemia, increased sensitivity to oxidative stress, and a marked proinflammatory effect, thus accelerating atherogenosis. Macrophages and polynuclear neutrophils are activated and synthesize a variety of biological substances, some of which are powerful inducers of inflammatory events (cytokines, free radicals, and eicosanoids). The effect of Mg depletion or Mg supplementation may result in the ability of Mg to modulate intracellular calcium. Pharmacological doses of Mg may reduce morbidity and mortality in the period following infarction. The beneficial effect of Mg may result from calcium-antagonist action, decreased platelet aggregation, and decreased free radical damage.

Magnesium Excess

Magnesium overload can occur in individuals with impaired renal function or during massive intravenous administration of Mg. It is most often iatrogenic. Clinical symptoms such as drowsiness and hyporeflexia develop when plasma Mg is two- or three-fold higher than the normal value.

Recommended Dietary Allowances

The estimated average requirement (EAR) is the nutrient intake value that is estimated to meet the requirement of 50% of individuals in a life stage and a gender group. Balance studies and data on stable isotopes suggest an EAR of 5 mg kg^{-1} day^{-1} for males and females. This value is greater during growth in adolescents and is estimated to be 5.3 mg kg^{-1} day^{-1}. The Mg requirement is also higher during pregnancy because of Mg transfer to the fetus in the last 3 months; therefore, an additional 35 mg day^{-1} is recommended.

In infants, the determination of the adequate intake (AI) is based on the Mg content of mother's milk and the progressive consumption of solid food. Thus, the AI is 30 mg day^{-1} during the first 6 months of life and 75 mg day^{-1} during the second 6 months of life.

The Recommended Dietary Allowance (RDA) is the average daily dietary intake that is sufficient to meet the nutrient requirement of 97.5% of individuals and is set at 20% above the EAR+2 CVs where the CV is 10%. During recent years, dietary reference intakes for the US and Canada have been revised by the Institute of Medicine. The recommended intakes of Mg are given in **Table 3**. It is not known whether decreased urinary Mg and increased maternal bone resorption provide sufficient amounts of Mg to meet increased needs during lactation. Thus, the French Society for Nutrition suggests adding 30 mg day^{-1} to intake for lactation, whereas no increase is recommended during lactation for the US and Canada.

The intake of Mg has been determined in various populations. Evidence suggests that the occidental diet is relatively low in Mg compared to recommended intakes, whereas the vegetarian diet is rich in Mg. For instance, the mean Mg intake of the subjects in the French Supplementation with Antioxidant Vitamins and Minerals Study was estimated to be 369 mg day^{-1} in men and

Table 3 Recommended dietary allowances of Mg

Age	*RDA (mg day^{-1})*		*AI (mg day^{-1})*	
	Male	Female	Male	Female
0–6 months			30	30
6–12 months			75	75
1–3 years	80	80		
4–8 years	130	130		
9–13 years	240	240		
14–18 years	410	360		
19–30 years	400	310		
31–50 years	420	320		
51–70 years	420	320		
<70 years	420	320		
Pregnancy		+40		
Lactation		+0		

Source: Reproduced from Institute of Medicine (1997) *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D and Fluoride*. Washington, DC: National Academy Press.

280 mg day^{-1} in women. However, it is possible that the recommended intakes are set somewhat high, as clinical problems are uncommon when such intakes are not caused or accompanied by metabolic diseases such as diabetes and alcoholism.

Conclusion

Based on evidence of low Mg intake in industrialized countries, intervention studies are needed to test whether improving Mg status would improve health outcomes.

Further Reading

Arnaud, M.J., 2008. Update on the assessment of magnesium status. British Journal of Nutrition 99, S24–S36.
Coudray, C., Demigné, C., Rayssiguier, Y., 2003. Effects of dietary fibers on magnesium absorption in animals and humans. Journal of Nutrition 133, 1–4.
Durlach, J., 1988. Magnesium in Clinical Practice. John Libbey, London.
Elin, R.J., 1989. Assessment of magnesium status. In: Itokawa, Y., Durlach, I. (Eds.), Magnesium in Health and Disease. John Libbey, London, pp. 137–146.
Feillet-Coudray, C., Coudray, C., Gueux, E., Mazur, A., Rayssiguier, Y., 2002. A new approach to evaluate magnesium status: Determination of exchangeable Mg pool masses using Mg stable isotope. Magnesium Research 15, 191–198.
Galan, P., Preziosi, P., Durlach, V., et al., 1997. Dietary magnesium intake in a French adult population. Magnesium Research 10, 321–328.
Institute of Medicine, 1997. Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D and Fluoride. National Academy Press, Washington, DC.
Rayssiguier, Y., Mazur, A., Durlach, J., 2001. Advances in Magnesium Research, Nutrition and Health. John Libbey, London.
Répertoire général des aliments, 1996. Table de Composition Minérale. Tec & Doc, Lavoisier, Paris.
Rude, R.K., 1998. Magnesium deficiency: A cause of heterogeneous disease in humans. Journal of Bone Mineral Research 13 (4), 749–758.
Shils, M.E., 1994. Magnesium. In: Shils, M.E., Olson, J.A., Shike, M. (Eds.), Modern Nutrition in Health and Disease, 8th edn. Lea & Febiger, Philadelphia, PA, pp. 164–184.
Vormann, J., 2003. Magnesium: Nutrition and metabolism. Molecular Aspects of Medicine 24, 27–37.
Wilkinson, S.R., Welch, R.M., Mayland, H.F., Grunes, D.L., 1990. Magnesium in plants: Uptake, distribution, function and utilization by man and animals. In: Sigel, H., Sigel, A. (Eds.), Compendium of Magnesium and Its Role in Biology, Nutrition and Physiology. Marcel Dekker, New York, pp. 33–56.
Yu, A.S.L., 2007. Disorders of magnesium and phosphorus. In: Goldman, L., Ausiello, D. (Eds.), Cecil Medicine, 23rd edn. Saunders Elsevier, Philadelphia, PA. ch. 120.

Manganese

C.L. Keen, J.L. Ensunsa, B. Lönnerdal, and S. Zidenberg-Cherr, University of California, Davis, CA, USA

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of C.L. Keen, J.L. Ensunsa, B. Lönnerdal, S. Zidenberg-Cherr, Manganese, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 148–154, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00182-3.

Chemical and Physical Properties 315
Dietary Sources 315
Analysis 316
Physiological Role 316
Tissue Concentrations 316
Absorption, Transport, and Storage 316
Metabolic Function and Essentiality 317
Manganese Deficiency 318
Manganese Toxicity 319
Assessment of Manganese Status 320
Further Reading 321

Chemical and Physical Properties

Manganese is the twelfth most abundant element in the Earth's crust and constitutes approximately 0.1% of it. Chemical forms of manganese in their natural deposits include oxides, sulfides, carbonates, and silicates. Anthropogenic sources of manganese are predominantly from the manufacturing of steel, alloys, and iron products. Manganese is widely used as an oxidizing agent, as a component of fertilizers and fungicides, and in dry cell batteries. Methylcyclopentadienyl manganese tricarbonyl (MMT) improves combustion in boilers and motors and can substitute for lead in gasoline as an antiknock agent. Concentrations of manganese in groundwater normally range between 1 and 100 $\mu g\, l^{-1}$, with most values being below 10 $\mu g\, l^{-1}$. Typical airborne levels of manganese (in the absence of excessive pollution) range from 10 to 70 $ng\, m^{-3}$.

Manganese is a transition element located in group VIIA of the periodic table. It occurs in 11 oxidation states ranging from −3 to +7, with the physiologically most important valences being +2 and +3. The +2 valence is the predominant form in biological systems and is the form that is thought to be maximally absorbed. The +3 valence is the form in which manganese is primarily transported in biological systems.

The solution chemistry of manganese is relatively simple. The aquo-ion is resistant to oxidation in acidic or neutral solutions. It does not begin to hydrolyze until pH 10, and therefore free Mn^{2+} can be present in neutral solutions at relatively high concentrations. Divalent manganese is a $3d^5$ ion and typically forms high-spin complexes lacking crystal field stabilization energies. The previous properties, as well as a large ionic radius and small charge-to-radius ratio, result in manganese tending to form weak complexes compared with other first-row divalent ions, such as Ni^{2+} and Cu^{2+}. Free Mn^{2+} has a strong isotropic electron paramagnetic resonance (EPR) signal that can be used to determine its concentration in the low micromolar range. Mn^{3+} is also critical in biological systems. For example, Mn^{3+} is the oxidative state of manganese in superoxide dismutase, is the form in which transferrin binds manganese, and is probably the form of manganese that interacts with Fe^{3+}. Given its smaller ionic radius, the chelation of Mn^{3+} in biological systems would be predicted to be more avid than that of Mn^{2+}. Cycling between Mn^{3+} and Mn^{2+} has been suggested to be deleterious to biological systems because it can generate free radicals. However, at low concentrations, Mn^{2+} can provide protection against free radicals, and it appears to be associated with their clearance rather than their production.

Dietary Sources

Manganese concentrations in typical food products range from 0.4 $\mu g\, g^{-1}$ (meat, poultry, and fish) to 20 $\mu g\, g^{-1}$ (nuts, cereals, and dried fruit). Breast milk is exceptionally low in manganese, containing only 0.004 $\mu g\, g^{-1}$, whereas infant formula can contain up to 0.4 $\mu g\, g^{-1}$. Teas can be particularly rich in manganese, containing up to 900 $\mu g\, g^{-1}$ of the element. An important consideration with respect to food sources of manganese is the extent to which the manganese is available for absorption. For example, although tea contains high amounts of the element, the tannin in tea can bind a significant amount of manganese, reducing its absorption from the gastrointestinal tract. Similarly, the high content of phytates and fiber constituents in cereal grains may limit the absorption of manganese. Conversely, although meat products contain low concentrations of manganese, absorption and retention of manganese from them is relatively high. Based on studies utilizing whole body retention curves after dosing with ^{54}Mn, the estimated percentage absorption of 1 mg of manganese from a test meal was 1.35%, whereas that from green leafy vegetables (lettuce and spinach) was closer to 5%. Absorption from wheat and sunflower seed kernels was somewhat lower than that from the leafy greens

at 1 or 2%, presumably due to a higher fiber content or to higher amounts of phytates and similar compounds in the wheat and sunflower seeds. The dephytinization of soy formula increased manganese absorption 2.3-fold from 0.7 to 1.6%.

Analysis

Although manganese is widely distributed in the biosphere, it occurs in only trace amounts in animal tissues. Serum concentrations can be as low as 20 nM and typical tissue concentrations are less than 4 μmol g^{-1} wet weight; tissue concentrations of 4–8 μmol g^{-1} wet weight are considered high. Because of the high environmental levels of manganese relative to its concentration in animal tissues, considerable effort must be made to minimize contamination of samples during their collection and handling.

The most common analytical methods that can sensitively measure manganese include neutron activation analysis, X-ray fluorescence, proton-induced X-ray emission, inductively coupled plasma emission, EPR, and flameless atomic absorption spectrophotometry (AAS). Currently, the most common method employed is flameless AAS. All of these methods, with the exception of EPR, measure the total concentration of manganese in the samples. EPR allows selective measurement of bound versus free manganese.

Physiological Role

Tissue Concentrations

The average human body contains between 200 and 400 μmol of manganese, which is fairly uniform in distribution throughout the body. There is relatively little variation among species with regard to tissue manganese concentrations. Manganese tends to be highest in tissues rich in mitochondria; its concentration in mitochondria is higher than in cytoplasm or other cell organelles. Hair can accumulate high concentrations of manganese, and it has been suggested that hair manganese concentrations may reflect manganese status. High concentrations of manganese are normally found in pigmented structures, such as retina, dark skin, and melanin granules. Bone, liver, pancreas, and kidney tend to have higher concentrations of manganese (20–50 nmol g^{-1}) than do other tissues. Concentrations of manganese in brain, heart, lung, and muscle are typically <20 nmol g^{-1}; blood and serum concentrations are approximately 200 and 20 nmol l^{-1}, respectively. Typical concentrations in cow milk are on the order of 800 nmol l^{-1}, whereas human milk contains 80 nmol l^{-1}. Bone can account for up to 25% of total body manganese because of its mass. Bone manganese concentrations can be raised or lowered by substantially varying dietary manganese intake over long periods of time, but bone manganese is not thought to be a readily mobilizable pool. The fetus does not accumulate liver manganese before birth, and fetal concentrations are significantly lesser than adult concentrations. This lack of fetal storage can be attributed to the apparent lack of storage proteins and the low prenatal expression of most manganese enzymes.

Absorption, Transport, and Storage

Absorption of manganese is thought to occur throughout the small intestine. Manganese absorption is not thought to be under homeostatic control. For adult humans, manganese absorption has been reported to range from 2 to 15% when ^{54}Mn-labeled test meals are used and to be 25% when balance studies are conducted; given the technical problems associated with balance studies, the ^{54}Mn data are probably more reflective of true absorption values. Data from balance studies indicate that manganese retention is very high during infancy, suggesting that neonates may be particularly susceptible to manganese toxicosis.

The higher retention of manganese in young animals relative to adults in part reflects an immaturity of manganese excretory pathways, particularly that of bile secretion, which is very limited in early life. The avid retention of the small amount of manganese from milk and the postnatal changes in its excretory pattern underscore the considerable changes in manganese metabolism that occur during the neonatal period.

In experimental animals, high amounts of dietary calcium, phosphorus, fiber, and phytate increase the requirements for manganese; such interactions presumably occur via the formation of insoluble manganese complexes in the intestinal tract with a concomitant decrease in the soluble fraction available for absorption. The significance of these dietary factors with regard to human manganese requirements remains to be clarified. Studies in avian species have demonstrated that high dietary phosphorus intakes decrease manganese deposition in bone by approximately 50%. Given that the diet of many individuals may be marginal in manganese (≤2 mg day^{-1} intake) whereas high in phosphorus (≥2000 mg day^{-1} intake), this antagonism may have important implications for human health. For example, the low fractional absorption of manganese from soy formula has been related to its relatively high phytate content. The mechanism underlying this effect of soy protein on manganese absorption/retention has not been fully delineated. However, dephytinization of soy formula with microbial phytase can markedly enhance manganese absorption.

An interaction between iron and manganese has been demonstrated in experimental animals and humans. Manganese absorption increases under conditions of iron deficiency, whereas high amounts of dietary iron can accelerate the development of manganese deficiency. The chronic consumption of high levels of iron supplements (>60 mg Fe day^{-1}) can have a negative effect on manganese balance in adult women. The mechanisms underlying the interactions between iron and manganese have not been fully elucidated; however, they likely involve competition for either a transport site or a ligand. Both iron and manganese can utilize divalent metal transporter 1 (DMT1); however, the expression of DMT1 is regulated by iron status via the IRE/IRP system. Thus,

during iron deficiency, DMT1 is upregulated causing an increase in manganese absorption. Rats fed iron-deficient diets accumulate manganese in several brain regions compared with rats fed control diets; the involvement of DMT1 in this accumulation of manganese is an area of active study. It should be noted that the interaction between manganese and iron can also affect the functions of some enzymes. For example, manganese can replace iron in the iron–sulfur center of cytosolic aconitase (IRP-1), resulting in an inhibition of the enzyme and an increase in iron regulatory protein (IRP) binding activity. Given the central role of IRPs in cellular iron metabolism, elevated cellular manganese concentration could in theory disrupt numerous translational events dependent on IRPs. That this in fact occurs is illustrated by the observation that following the addition of manganese to cells in culture, there can be sharp reductions in ferritin protein abundance, whereas there are increases in transferrin receptor abundance. This results in changes in intracellular iron metabolism, as reflected by decreases in mitochondrial aconitase (m-aconitase) abundance.

Manganese entering the portal blood from the gastrointestinal tract may remain free or become associated with α_2-macroglobulin, which is subsequently taken up by the liver. A small fraction enters the systemic circulation, where it may become oxidized to Mn^{3+} and bind to transferrin. Studies *in vivo* suggest that the Mn^{3+} complex forms very quickly in blood, in contrast to the slow oxidation of the Mn^{2+}–transferrin complex *in vitro*. Manganese uptake by the liver has been reported to occur by a unidirectional, saturable process with the properties of passive mediated transport. After entering the liver, manganese enters one of at least five metabolic pools. One pool represents manganese taken up by the lysosomes, from which it is transferred subsequently to the bile canaliculus. The regulation of manganese is maintained in part through biliary excretion of the element; up to 50% of manganese injected intravenously can be recovered in the feces within 24 h. A second pool of manganese is associated with the mitochondria. Mitochondria have a large capacity for manganese uptake, and the mitochondrial uptake and release of manganese and calcium are thought to be related. A third pool of manganese is found in the nuclear fraction of the cell; the roles of nuclear manganese have not been fully delineated, but one function may be to contribute to the stability of nucleosome structure. A fourth manganese pool is incorporated into newly synthesized manganese proteins; biological half-lives for these proteins have not been agreed upon. The fifth identified intracellular pool of manganese is free Mn^{2+}. Fluctuations in the free manganese pool may be an important regulator of cellular metabolic control in a manner analogous to those for free Ca^{2+} and Mg^{2+}. Consistent with this concept, in pancreatic islets manganese blocks glucose-induced insulin release by altering cellular calcium fluxes, and manganese directly augments contractions in smooth muscle by a mechanism comparable to that of calcium.

The mechanisms by which manganese is transported to, and taken up by, extrahepatic tissues have not been identified. Transferrin is the major manganese binding protein in plasma; however, it is not known to what extent transferrin facilitates the uptake of manganese by extrahepatic tissue. The concentration of manganese citrate in blood can be fairly high, and this complex may be important for manganese movement across the blood–brain barrier. DMT1 may be involved in manganese transport because it is expressed in discrete areas of the brain. Manganese uptake by extrahepatic tissue does not seem to be increased under conditions of manganese deficiency, suggesting that manganese, in marked contrast to iron, does not play a role in the induction (or suppression) of manganese transport proteins.

There is limited information concerning the hormonal regulation of manganese metabolism. Fluxes in the concentrations of adrenal, pancreatic, and pituitary–gonadal axis hormones affect tissue manganese concentrations; however, it is not clear to what extent hormone-induced changes in tissue manganese concentrations are due to alterations in cellular uptake of manganese-activated enzymes or metalloenzymes.

Metabolic Function and Essentiality

Manganese functions as a constituent of metalloenzymes and as an enzyme activator. Manganese-containing enzymes include arginase (EC 3.5.3.1), pyruvate carboxylase (EC 6.4.1.1), and manganese–superoxide dismutase (MnSOD) (EC 1.15.1.1). Arginase, the cytosolic enzyme responsible for urea formation, contains 4 mol Mn^{2+} per mole of enzyme. Reductions in arginase activity resulting from manganese deficiency result in elevated plasma concentrations of ammonia and lowered plasma concentrations of urea. Reductions in arginase activity due to manganese deficiency may affect flux of arginine through the nitric oxide synthase (NOS) pathway, resulting in alterations in NO production. It has been suggested that arginase plays a regulatory role in NO production by competing with NOS for the same substrate, arginine. Rats fed manganese-deficient diets have shown effects indicative of increased NO production, such as increases in plasma and urinary nitrates plus nitrites and decreased blood pressure; however, neither NOS activity nor NO production have been measured directly. In addition, manganese binding by arginase is critical for the pH-sensing function of this enzyme in the ornithine cycle, suggesting that manganese plays a role in the regulation of body pH. With experimental diabetes, liver and kidney manganese concentrations and arginase activity can be markedly elevated. This manganese effect on arginase has been suggested to be due to an effect of Mn^{2+} on the conformational properties of the enzyme with a resultant modification of arginase activity. Whether this finding implies an increased manganese requirement for people with diabetes has not been determined.

Pyruvate carboxylase, the enzyme that catalyzes the first step of carbohydrate synthesis from pyruvate, also contains 4 mol Mn^{2+} per mole enzyme. Although the activity of this enzyme can be lower in manganese-deficient animals than in controls, gluconeogenesis has not been shown to be markedly inhibited in manganese-deficient animals.

MnSOD catalyzes the disproportionation of $^{\bullet}O_2^-$ to H_2O_2 and $^{\bullet}O_2$. The essential role of MnSOD in the normal biological function of tissues has been clearly demonstrated by the homozygous inactivation of the *SOD2* gene for MnSOD in mice. Mice with this phenotype die within the first 10 days of life with a dilated cardiomyopathy, accumulation of lipid in liver and skeletal muscle, and

metabolic acidosis. The activity of MnSOD in tissues of manganese-deficient rats can be significantly lower than in controls due to downregulation of MnSOD at the (pre)transcriptional level. That this reduction is functionally significant is suggested by the observation of higher than normal levels of hepatic mitochondrial lipid peroxidation in manganese-deficient rats. Tissue MnSOD activity can be increased by several diverse stressors, including alcohol, ozone, irradiation, interleukin-1, and tumor necrosis factor-α, presumably as a consequence of stressor-associated increases in cellular free radical (or oxidized target(s)) concentrations. Stressor-induced increases in MnSOD activity can be attenuated in manganese-deficient animals, potentially increasing their sensitivity to these insults. Transgenic mice have also been produced that overexpress MnSOD; a decreased severity of reperfusion injury has been noted in these animals, further supporting its physiological significance.

Considerable research is focused on the introduction of the human MnSOD gene into research animals utilizing viral vectors or plasmid/liposome delivery. This gene therapy has been shown to decrease radiation-induced injury, extend pancreatic islet transplant function, and slow the growth of malignant tumors in animal models via overexpression of the MnSOD protein. Another field of research that is rapidly advancing utilizes MnSOD mimetics for treatment of a variety of diseases in which the native SOD enzyme has been found to be effective. These mimetics are small manganese-containing synthetic molecules that have catalytic activity equivalent or superior to the native enzyme. They possess the additional beneficial properties of being nonimmunogenic because they are nonpeptides, able to penetrate cells, selective for superoxide (they do not interact with biologically important molecules), stable *in vivo*, and not deactivated by the destructive free radical peroxynitrite, which is capable of deactivating native MnSOD via nitration of tyrosine. These mimetic compounds have been found to be protective in animal models of acute and chronic inflammation, reperfusion injury, shock, and radiation-induced injury. Both of these therapies, MnSOD gene delivery and MnSOD mimetics, hold promise for future treatments in human chronic and acute conditions.

Finally, further evidence for the biological and research relevance of MnSOD is that experiments have been undertaken on the International Space Station to improve three-dimensional growth of MnSOD crystals in order to develop a better understanding of the role of structure in the reaction mechanism of this enzyme.

In contrast to the relatively few manganese metalloenzymes, there are a large number of manganese-activated enzymes, including hydrolases, kinases, decarboxylases, and transferases. Manganese activation of these enzymes can occur as a direct consequence of the metal binding to the protein, causing a subsequent conformation change, or by binding to the substrate, such as adenosine triphosphate (ATP). Many of these metal activations are nonspecific in that other metal ions, particularly Mg^{2+}, can replace Mn^{2+}. An exception is the manganese-specific activation of glycosyltransferases. Several manganese deficiency-induced pathologies have been attributed to a low activity of this enzyme class. A second example of an enzyme that may be specifically activated by manganese is phosphoenolpyruvate carboxykinase (PEPCK; EC 4.1.1.49), the enzyme that catalyzes the conversion of oxaloacetate to phosphoenolpyruvate, GDP, and CO_2. Although low activities of PEPCK can occur in manganese-deficient animals, the functional significance of this reduction is not clear.

A third example of a manganese-activated enzyme is glutamine synthetase (EC 6.3.1.2). This enzyme, found in high concentrations in the brain, catalyzes the reaction NH_3+glutamate+ATP→glutamine+ADP+P_i. Brain glutamine synthetase activity can be normal even in severely manganese-deficient animals, suggesting that the enzyme either has a high priority for this element or magnesium can act as a substitute when manganese is lacking. It should be noted that this enzyme can be inactivated by oxygen radicals; therefore, a manganese deficiency-induced reduction in MnSOD activity theoretically could act to depress further the activity of glutamine synthetase.

Manganese Deficiency

Manganese deficiency has been demonstrated in several species, including rats, mice, pigs, and cattle. Signs of manganese deficiency include impaired growth, skeletal abnormalities, impaired reproductive performance, ataxia, and defects in lipid and carbohydrate metabolism.

The effects of manganese deficiency on bone development have been studied extensively. In most species, manganese deficiency can result in shortened and thickened limbs, curvature of the spine, and swollen and enlarged joints. The basic biochemical defect underlying the development of these bone defects is a reduction in the activities of glycosyltransferases; these enzymes are necessary for the synthesis of the chondroitin sulfate side chains of proteoglycan molecules. In addition, manganese deficiency in adult rats can result in an inhibition of both osteoblast and osteoclast activity. This observation is particularly noteworthy, given the reports that women with osteoporosis tend to have low blood manganese concentrations and that the provision of manganese supplements might be associated with an improvement in bone health in postmenopausal women.

One of the most striking effects of manganese deficiency occurs during pregnancy. When pregnant animals (rats, mice, guinea pigs, and mink) are deficient in manganese, their offspring exhibit a congenital, irreversible ataxia characterized by incoordination, lack of equilibrium, and retraction of the head. This condition is the result of impaired development of the otoliths, the calcified structures in the inner ear responsible for normal body-righting reflexes. The block in otolith development is secondary to depressed proteoglycan synthesis due to low activity of manganese-requiring glycosyltransferases.

Defects in carbohydrate metabolism, in addition to those described previously, have been shown in manganese-deficient rats and guinea pigs. In the guinea pig, perinatal manganese deficiency results in pancreatic pathology, with animals exhibiting aplasia or marked hypoplasia of all cellular components. Manganese-deficient guinea pigs and rats given a glucose challenge often respond with a diabetic-type glucose tolerance curve. In addition to its effect on pancreatic tissue integrity, manganese deficiency can directly

impair pancreatic insulin synthesis and secretion as well as enhance intracellular insulin degradation. The mechanism(s) underlying the effects of manganese deficiency on pancreatic insulin metabolism have not been fully delineated, but they are thought to be multifactorial. For example, the flux of islet cell manganese from the cell surface to an intracellular pool may be a critical signal for insulin release. It is also known that insulin messenger ribonucleic acid levels are reduced in manganese-deficient animals, which is consistent with their depressed insulin synthesis. In addition, insulin sensitivity of adipose tissue is reduced in manganese-deficient rats, a phenomenon that may be related to fewer insulin receptors per adipose cell. Manganese deficiency may also affect glucose metabolism by means of a reduction in the number of glucose transporters in adipose tissue by an unidentified mechanism. Finally, the effect of manganese deficiency on insulin production may also be due to the destruction of pancreatic J3 cells. It is worth noting that constitutive pancreatic MnSOD activity is lower than in most tissues; this, coupled with the observation that most diabetogenic agents function via the production of free radicals with subsequent tissue damage, suggests that an additional mechanism underlying pancreatic dysfunction in manganese-deficient animals may be free radical mediated.

In addition to its effect on endocrine function, manganese deficiency can affect pancreatic exocrine function. For example, manganese-deficient rats can be characterized by an increase in pancreatic amylase content. The mechanism underlying this effect of manganese deficiency has not been delineated; however, it is thought to involve a shift in amylase synthesis or degradation because secretagogue-stimulated acinar secretion is comparable in control and manganese-deficient rats.

Although a majority of studies concerning the influence of manganese deficiency on carbohydrate metabolism have been conducted with experimental animals, there is one report in the literature of an insulin-resistant diabetic patient who responded to oral doses of manganese (doses ranged from 5 to 10 mg) with decreasing blood glucose concentrations. Although this is an intriguing case report, others have reported a lack of an effect of oral manganese supplements (up to 30 mg) in diabetic subjects, and low blood manganese concentrations have not been found to be a characteristic of diabetics.

Abnormal lipid metabolism is also characteristic of manganese deficiency: Specifically, a lipotrophic effect of manganese has been suggested in the literature. Severely manganese-deficient animals can be characterized by high liver fat, hypocholesterolemia, and low high-density lipoprotein (HDL) concentrations. Deficient animals can also be characterized by a shift to smaller plasma HDL particles, lower HDL apolipoprotein (apoE) concentrations, and higher apoC concentrations. As stated previously, tissue lipid peroxidation rates can be increased in manganese-deficient animals, possibly as a result of low tissue MnSOD activity.

There is considerable debate as to the extent to which manganese deficiency affects humans under free-living conditions. Manganese deficiency can be induced in humans under highly controlled experimental conditions. In one study, manganese deficiency was induced in adult male subjects by feeding a manganese-deficient diet (0.1 mg Mn day^{-1}) for 39 days. The subjects developed temporary dermatitis, as well as increased serum calcium and phosphorus concentrations and increased alkaline phosphatase activity, suggestive of bone resorption. Since the late 1980s, several diseases have been reported to be characterized, in part, by low blood manganese concentrations. These diseases include epilepsy, Mseleni disease, maple sirup urine disease and phenylketonuria, Down's syndrome, osteoporosis, and Perthes' disease. The finding of low blood manganese levels in subsets of individuals with the previously mentioned diseases is significant because blood manganese levels can reflect soft tissue manganese concentrations. The reports of low blood manganese concentrations in individuals with epilepsy are particularly intriguing, given the observations that manganese-deficient animals can show an increased susceptibility to drug and electroshock-induced seizures, and a genetic model for epilepsy in rats (the GEPR rat) is characterized by low blood manganese concentrations. It is evident that a deficiency of manganese may contribute to the pathology of epilepsy at multiple points, given that Mn^{2+} is implicated in activation of glutamine synthetase, a Mn^{2+}-specific brain ATPase; production of cyclic adenosine monophosphate (AMP); altered synaptosomal uptake of noradrenalin and serotonin; glutamate, GABA, and choline metabolism; and biosynthesis of acetylcholine receptors.

Evidence of widespread manganese deficiency in human populations is lacking. Typically, manganese intakes approximate the 2001 US Institute of Medicine's suggested adequate intakes as follows: 3 μg day^{-1} for infants 0–6 months old, 0.6 mg day^{-1} for infants 7–12 months old, 1.2–1.9 mg day^{-1} for children 1–13 years old, 1.6–2.2 mg day^{-1} for older children, and 1.8–2.6 mg day^{-1} for adults. The Tolerable Upper Intake Level (UL) is the highest level of a daily nutrient intake that is likely to pose no risk of adverse health effects in almost all individuals. The Institute of Medicine's recommended intakes for manganese set ULs at 2, 3, and 6 mg day Upper Intake Level (UL) for children 1–3, 4–8, and 9–13 years old, respectively. Values were set at 9 mg day^{-1} for adolescents 14–18 years old and at 11 mg day^{-1} for adults.

Manganese Toxicity

In domestic animals, the major reported lesion associated with chronic manganese toxicity is iron deficiency, resulting from an inhibitory effect of manganese on iron absorption. Additional signs of manganese toxicity in domestic animals include depressed growth, depressed appetite, and altered brain function.

In humans, manganese toxicity represents a serious health hazard, resulting in severe pathologies of the central nervous system. In its most severe from, the toxicosis is manifested by a permanent crippling neurological disorder of the extrapyramidal system, which is similar to Parkinson's disease. In its milder form, the toxicity is expressed by hyperirritability, violent acts, hallucinations, disturbances of libido, and incoordination. The previous symptoms, once established, can persist even after the manganese body burden returns to normal. Although a majority of reported cases of manganese toxicity occur in individuals exposed to high concentrations of airborne manganese (>5 mg m^{-3}), subtle signs of manganese toxicity, including delayed reaction time, impaired motor coordination, and impaired memory, have been observed in workers exposed to airborne manganese concentrations less than 1 mg m^{-3}. Therefore, an

inhalation reference concentration range for manganese has been established by the US Environmental Protection Agency to be between 0.09 and 0.2 μg m^{-3}. Manganese toxicity has been reported in individuals who have consumed water containing high levels (≥10 mg Mn) of manganese for long periods of time. Recently, there has been concern that the risk for manganese toxicity may be increasing in some areas because of the use of MMT in gasoline as an antiknock agent, although there is little evidence that air, water, or food manganese concentrations have increased where this fuel is used.

In addition to neural damage, reproductive and immune system dysfunction, nephritis, testicular damage, pancreatitis, lung disease, and hepatic damage can occur with manganese toxicity, but the frequency of these disorders is unknown. Although there is a limited body of epidemiological data that suggests that high levels of manganese can result in an increased risk for colorectal and digestive tract cancers, most investigators do not consider manganese to be a carcinogen. In contrast, both divalent ($MnCl_2$) and heptavalent forms ($KMnO_4$) of manganese are recognized to be strong clastogens both *in vitro* and *in vivo*; exposure to high concentrations of either form results in chromosomal breaks, fragments, and exchanges. High concentrations of manganese can also induce forward and point mutations in mammalian cells. High levels of dietary manganese have not been reported to be teratogenic in the absence of overt signs of maternal toxicity. However, there are reports that exposure to high levels of manganese during prenatal development can result in behavioral abnormalities. High levels of brain manganese have been reported in subjects with amyotrophic lateral sclerosis, and it has been suggested that this increase may contribute to the progression of the disease. Similar to the cases in humans, chronic manganese toxicity in rhesus monkeys is characterized by muscular weakness, rigidity of the lower limbs, and neuron damage in the substantia nigra. Findings from a recent study suggest that iron and aluminum, which accumulate in the globus pallidus and the substantia nigra of these animals, induce tissue oxidation that may contribute to the damage associated with manganese toxicity. Neural toxicity is a consistent finding in rats exposed to chronic manganese toxicity. Significant manganese accumulation was accompanied by an increase in cholesterol content in the hippocampal region of manganese-treated rats, which was associated with impaired learning; this impairment was corrected by an inhibitor of cholesterol synthesis. The development of manganese toxicity in individuals with compromised liver function, or compromised biliary pathways, is well documented. Significantly, these individuals can have abnormal magnetic resonance imaging (MRI) patterns, which improve following the alleviation of the manganese toxicity. For example, in some cases improvements in brain function have been achieved after liver transplant. The mechanisms underlying the toxicity of manganese have not been agreed on but may involve multiple etiologies, including endocrinological dysfunction, excessive tissue oxidative damage, manganese-mediated disruptions in intracellular calcium and iron metabolism, and mitochondrial dysfunction caused by manganese inhibition of some pathways of the mitochondrial respiratory chain.

Severe cases of manganese toxicity in humans have been reported for adults, as well as isolated cases in other groups of individuals who are vulnerable, including children on long-term parenteral nutrition and parenteral nutrition patients who have cholestasis or other hepatic disease. In many cases, the previously mentioned groups of individuals have been reported to be characterized by high brain manganese concentrations based on MRI. Although no known cases have been reported, infants may be at a high risk for manganese toxicity due to a high absorptive capacity for the element or an immature excretory pathway for it. If manganese is taken up by extrahepatic tissues via the manganese–transferrin complex, the developing brain may be particularly sensitive to manganese toxicity due to the high number of transferrin receptors elaborated by neuronal cells during development, coupled with the putative need by neural cells for transferrin for their differentiation and proliferation. Newborn rats given daily doses of dietary manganese at a level equivalent to that of soy formula exhibited significant neurodevelopmental delays as assessed by several behavioral tests. It should be noted that the concentration of manganese in soy formula is relatively modest but approximately 60–100 times higher than that of breast milk. Brain manganese concentration was increased and striatal dopamine concentrations were significantly decreased even 45 days after the supplementation ended, suggesting that the impact of manganese on the brain and behavior was irreversible. Thus, dietary exposure to high levels of manganese during infancy can be neurotoxic to rat pups and result in developmental deficits. Further studies on human infants fed diets with different levels of manganese are needed to assess whether there are any long-term consequences of early manganese exposure of newborns.

Another group of neuropathological conditions that has been associated with elevated levels of brain manganese is transmissible spongiform encephalopathies. These diseases found in animals and humans are also referred to as prion diseases. There is strong evidence that in their native state, prions are normal brain glycoproteins that bind copper and have an antioxidant function. However, it has been suggested that in the disease process an abnormal isoform of the protein is generated in which manganese is substituted for copper. This isoform is proteinase resistant, no longer has antioxidant activity, and may play a role in the etiology of these diseases. Indeed, elevated levels of brain manganese, along with lower than normal levels of brain copper, have been measured in patients with the prion disease, Creutzfeld–Jakob disease. Whether the elevated levels of brain manganese observed in these patients as well as in animal models of these diseases play an important role in their pathogenesis or are secondary to other factors remains to be determined.

Assessment of Manganese Status

Reliable biomarkers for the assessment of manganese status have not been identified. Whole blood manganese concentrations are reflective of soft tissue manganese levels in rats; however, it is not known whether a similar relationship holds for humans. Plasma manganese concentrations decrease in individuals fed manganese-deficient diets and are slightly higher than normal in individuals consuming manganese supplements. Lymphocyte MnSOD activity and blood arginase activity are increased in individuals who

consume manganese supplements; however, their value as biomarkers for manganese status may be complicated due to the number of cytokines and disease states that may also increase their expression. Urinary manganese excretion has not been found to be sensitive to dietary manganese intake. With respect to the diagnosis of manganese toxicosis, the use of MRI appears to be promising because the images associated with manganese toxicity are relatively specific. Whole blood manganese concentrations can be correlated with MRI intensity and Ti values in the globus pallidus even in the absence of symptoms of neurological damage. Thus, although it is relatively expensive, MRI may be particularly useful as a means of identifying susceptible individuals in, or around, manganese-emitting factories. In addition, the method may be useful in the evaluation of patients with liver failure.

Further Reading

Brown, D.R., 2002. Metal toxicity and therapeutic intervention. Biochemical Society 30, 742–745.

Crossgrove, J.S., Allen, D.D., Bukaveckas, B.L., Rhineheimer, S.S., Yokel, R.A., 2003. Manganese distribution across the blood–brain barrier. I. Evidence for carrier-mediated influx of managanese citrate as well as manganese and manganese transferring. Neurotoxicology 24, 3–13.

Davey, C.A., Richmond, T.J., 2002. DNA-dependent divalent cation binding in the nucleosome core particle. Proceedings of the National Academy of Sciences of the United States of America 99, 11169–11174.

Garrick, M.D., Dolan, K.G., Horbinski, C., et al., 2003. DMT1: A mammalian transporter for multiple metals. BioMetals 16, 41–54.

Gerber, G.B., Léonard, A., Hantson, P.H., 2002. Carcinogenicity, mutagenicity and teratogenicity of manganese compounds. Critical Reviews in Oncology/Hematology 42, 25–34.

Guo, H., Seixas-Silva, J.A., Epperly, M.W., et al., 2003. Prevention of radiation-induced oral cavity mucositis by plasmid/liposome delivery of the human manganese superoxide dismutase (SOD2) transgene. Radiation Research 159, 361–370.

Kwik-Uribe, C.L., Reaney, S., Zhu, Z., Smith, D., 2003. Alterations in cellular IRP-dependent iron regulation by in vitro manganese exposure in undifferentiated PC12 cells. Brain Research 973, 1–15.

Normandin, L., Hazell, A.S., 2002. Manganese neurotoxicity: An update of pathophysiologic mechanisms. Metabolic Brain Disease 17, 375–387.

Sabbatini, M., Pisani, A., Uccello, F., et al., 2003. Arginase inhibition slows the progression of renal failure in rats with renal ablation. American Journal of Renal Physiology 284, F680–F687.

Salvemini, D., Muscoli, C., Riley, D.P., Cuzzocrea, S., 2002. Superoxide dismutase mimetics. Pulmonary Pharmacology and Therapeutics 15, 439–447.

Takagi, Y., Okada, A., Sando, K., et al., 2002. Evaluation of indexes of in vivo manganese status and the optimal intravenous dose for adult patients undergoing home parenteral nutrition. American Journal of Clinical Nutrition 75, 112–118.

Takeda, A., 2003. Manganese action in brain function. Brain Research Reviews 41, 79–87.

Tran, T., Chowanadisai, W., Crinella, F.M., Chicz-DeMet, A., Lönnerdal, B., 2002. Effect of high dietary manganese intake of neonatal rats on tissue mineral accumulation, striatal dopamine levels, and neurodevelopmental status. Neurotoxicology 158, 1–9.

Vahedi-Faridi, A., Porta, J., Borgstahl, G.E.O., 2002. Improved three-dimensional growth of manganese superoxide dismutase crystals on the International Space Station. Biological Crystallography 59, 385–388.

Yokel, R.A., Crossgrove, J.S., Bukaveckas, B.L., 2003. Manganese distribution across the blood–brain barrier. II. Manganese efflux from the brain does not appear to be carrier mediated. Neurotoxicology 24, 15–22.

Niacin and pellagra

C.J. Bates, MRC Human Nutrition Research, Cambridge, UK

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of C.J. Bates, Niacin and Pellagra, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 182–188, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00190-2.

History, Signs and Symptoms of Deficiency, and Antimetabolites 322
Absorption, Transport, and Storage 326
Metabolism and Excretion 326
Metabolic Function and Essentiality 327
Assessment of Niacin Status and Requirements 328
Dietary Sources and High Intakes 328
Further Reading 329

Glossary

B-vitamins There are seven B-group vitamins that are all essential components of human diets. All are either enzyme cofactors *per se* or are converted in the body to enzyme cofactors (which are small molecules that form part of the active site of specific enzymes and participate in their catalytic reactions). B-vitamins are efficiently absorbed in the gastrointestinal tract and are resistant to urinary excretion, unless intakes and blood concentrations exceed tissue requirements.

Clinical deficiency Often results from severe prolonged tissue (nutrient) deficiency, which may in turn result from inadequate (nutrient) intake, but may also arise from impaired absorption, increased turnover, and excretion, increased tissue demand, etc. Deficiency signs are directly observable; symptoms can be elicited by testing, and the response of these to interventions (e.g., nutrient supplementation) can help to confirm a diagnosis of deficiency.

Estimated Average Requirement (EAR) Similar to RDA and RNI, except that this is the mean (i.e., average) nutrient requirement of the individuals in a defined population group (USA and UK).

Nutrient (e.g., vitamin) status Is commonly assessed by measuring the concentration of the nutrient or a derivative in an accessible body fluid such as serum or urine or else the functionality of an enzyme or a biochemical pathway (functional status). Published 'normal ranges' enable the result to be classified as, for example, deficient, low, normal, or high.

Recommended Dietary Allowance (RDA) The amount of a nutrient (per day) that covers the needs of the majority (usually approximately 97.5%) of the individuals in a defined population group (e.g., adult males) in the USA. The term 'Reference Nutrient Intake' (RNI) is used for a similar concept in the UK.

History, Signs and Symptoms of Deficiency, and Antimetabolites

Pellagra (meaning 'rough' or 'raw' skin) was common in western Europe (e.g., France, Italy) and especially in the southern half of the USA, up to the early twentieth century: it caused approximately 10 000 deaths in the USA in 1929 alone. The typical signs and symptoms of human pellagra are listed in Table 1 and the history of its recognition and causation in Table 2.

By 1810, a European source concluded that the disease was neither contagious nor hereditary but probably caused by poor diets, especially those in which grains such as corn (i.e., maize) were the principal staple; a good hospital diet was curative. By 1860, it was known that poor Mexican peasants, whose diet was mainly corn-based but roasted the corn with lime did not exhibit pellagra. However, for approximately the next 50 years, the 'toxin' theory of pellagra-causation held sway. In Europe, the prevalence of pellagra declined markedly, just as it was beginning to emerge as a major scourge in the southern USA. Poor sanitation, infection, insect-borne disease, and toxins from bacteria or molds were variously blamed – until Joseph Goldberger, from 1914 until his death in 1929, carried out controlled feeding studies in human convicts and an animal model ('blacktongue', a corn diet-induced condition in dogs (Table 1)). The important outcome was that pellagra was now seen not as an infectious disease, but as primarily diet related.

Funk's newly formulated hypothesis of essential dietary vitamins and Gowland Hopkins' concept of 'accessory food factors' then initiated a search for curative organic substance(s). Goldberger classified foods according to their 'pellagra-preventative'(PP) properties, and found that both dried yeast and a water-soluble extract from yeast were curative in small quantities. During the 1930s, and following recognition of the role of pyridine nucleotide coenzymes in food energy release, the central roles of nicotinic acid and nicotinamide were elucidated and equated with the 'PP' factor. The term 'niacin' was then coined, because nicotinic acid was associated with tobacco and hence considered unsuitable for an essential dietary factor.

Table 1 Signs and symptoms of niacin deficiency in humans and animals

1. *Human deficiency*
   - Loss of appetite and weight
   - Dermatosis (hyperpigmentation, hyperkeratosis, desquamation of the epidermis, especially where frequently exposed to strong sunlight)
   - Anorexia
   - Achlorhydria
   - Angular stomatitis, cheilosis, magenta tongue
   - Diarrhea
   - Anemia
   - Neuropathy (headache, dizziness, tremor, neurosis, apathy)
   - Death in severe and prolonged cases
2. *Blacktongue in dogs and cats*
   - Pustules in mouth and excessive salivation, darkening, and necrosis of the tongue
   - Diarrhea
3. *Pigs*
   - Neurological lesions affecting ganglion cells; histopathology of nerves
   - Anemia
   - Degeneration of intestinal mucosa and diarrhea
4. *Rats*
   - Reduced growth rate
   - Alopecia (roughened skin)
   - Damage to peripheral nerves (cells and axons)
5. *Birds (e.g., chickens, ducks)*
   - Inflammation of the upper gastrointestinal tract
   - Dermatitis
   - Diarrhea
   - Poor growth of feathers; bowed and weakened legs

Although pellagra in dogs and humans responds well to supplements of pure niacin, there were further strands to the story. For instance, the niacin content of foods, as measured by chemical analysis, was not always a good guide to their pellagra-producing or -preventing properties (as in **Table 3**). During the 1940s, it was shown that in rats (which respond to pellagragenic diets by a reduced growth rate but not by skin lesions), high dietary tryptophan levels reduced the requirement for niacin. Tryptophan was then shown to be effective in humans in reducing the pellagragenic properties of diets, and the metabolic pathway linking tryptophan to niacin and to the pyridine nucleotide coenzymes was elucidated (**Figure 1**). Between 34 and 86 mg tryptophan in human diets is now considered to be equivalent to 1 mg niacin, with a mean conversion ratio of 60 mg tryptophan per milligram of niacin, now the basis of niacin equivalents (NEs) value of a diet. This is a much better index of antipellagra potency than niacin content alone (**Table 3**).

The causes of pellagra in human populations, and indeed in Goldberger's experimental studies, are a complex mixture of B-vitamin deficiencies, of which niacin and tryptophan are the dominant effectors, but riboflavin is also important, followed by thiamin, vitamin B_6, and other nutrients. Several bacterial, fungal, and other toxins can deplete the cellular levels of nicotinamide adenine dinucleotide phosphate (NAD(P)), and niacin in corn and other grains is often chemically bound into a macromolecular complex, niacytin, from which niacin cannot be released by digestive enzymes in the gastrointestinal tract, but that requires heat and alkali treatment (as in the preparation of Mexican tortillas) to make it bioavailable.

In India, the millet staple 'jowar' is frequently associated with pellagra signs and symptoms, even though it is apparently a reasonably good source of available niacin and tryptophan. Local studies suggested an association with jowar's high leucine content, which may impair the conversion of tryptophan to niacin coenzymes, but this remains controversial: balance studies in humans have failed to show a consistent effect of leucine on excretion of tryptophan metabolites, a sensitive test for impairment of tryptophan-conversion pathways. In parts of South Africa, iron overload has been reported to complicate the metabolic effects of low niacin intakes.

Estrogenic hormones affect the conversion of tryptophan to niacin coenzymes, hence women (except during pregnancy) are more susceptible to pellagra than men. Several drugs have antiniacin (iatrogenic) effects. Thus isoniazid, commonly used in the treatment of tuberculosis, inhibits kynureninase (an enzyme in the tryptophan conversion pathway) activity by inactivating the enzyme cofactor, pyridoxal phosphate. Any interference with vitamin B_6 metabolism is also likely to affect niacin economy as well. Because *ca*. 60% of Indians are genetically slow acetylators (i.e., deactivators) of isoniazid, this drug is especially likely to cause pellagra. Anti-Parkinsonism drugs Carbidopa and Benseride can cause pellagra symptoms and reduce the excretion of *N*-methylnicotinamide. Some antineoplastic drugs, for example 6-dimethylaminonicotinamide and 6-aminonicotinamide, inhibit key

Table 2 History of pellagra, and recognition of its causes, in human populations

1. A poorly understood disease (dermatitis, gastrointestinal, and mental signs/symptoms) appeared in Europe in the eighteenth and nineteenth centuries, and then in southern USA in the first decade of the twentieth century and was named 'pellagra' (meaning raw/rough skin).
2. Favored causal hypotheses included infection, moldy grain, and insects.
3. 1914–1916, Joseph Goldberger disproved the infection hypothesis: by self-experimentation and then producing pellagra in prisoners fed mainly corn diets.
4. 1920s, Joseph Goldberger developed the 'blacktongue' model of pellagra in dogs, with corn diets.
5. 1930s, nicotinic acid was isolated as a pure water-soluble compound (vitamin) of known structure, from yeast and liver extracts, able to cure pellagra and blacktongue.
6. Mid-twentieth century, unavailable niacin is bound in the complex niacytin in corn. Heating in an alkaline environment (as in Mexican tortillas) can release it.
7. Niacin can be produced from tryptophan in the body. Pellagragenic diets are low in tryptophan as well as in niacin. The concept of 'niacin equivalents' (milligram niacin plus one-sixtieth of milligram tryptophan) in food developed. Human requirements estimated.
8. The Indian cereal 'jowar' is shown to be pellagragenic. Some, but not all, studies have implicated its high leucine content.
9. Niacin and riboflavin deficiencies often coexist; the signs and symptoms of pellagra-like disease are often attributable to multiple nutrient deficiencies.
10. Certain inborn errors of metabolism (genetic defects), or iatrogenic effects of drugs, can mimic the signs, symptoms, and metabolic defects of pellagra.

enzymes whose substrates are nicotinamide adenine dinucleotide (NAD) or NADP by being converted *in vivo* to analogs of the coenzymes. 3-Acetyl-pyridine, which also forms an analog of NAD, has either antagonistic or niacin-replacing properties, depending on the dose used. Metronidazole is a niacin antagonist, as is 2-amino,1,3,4-triazole.

Some inborn errors of metabolism can result in pellagra-like symptoms in humans. Thus in Hartnup disease, an autosomal recessive condition in which the cellular transport of tryptophan (and other neutral amino acids) is impaired, tryptophan is lost in the urine through a failure of renal tubular reabsorption. Supplementation with niacin or with tryptophan peptides (but not free tryptophan) can be palliative. Another genetic disease that may respond to niacin supplements is Fredrikson type I familial hypercholesterolemia; nicotinic acid lowers the raised blood cholesterol levels that are associated with it.

Other inborn errors of tryptophan economy include xanthurenic aciduria, hydroxykynurenuria, tryptophanuria (i.e., tryptophan dioxygenase deficiency), and excessive picolinate carboxylase activity. Tumors of the enterochromaffin cells, which synthesize excessive amounts of 5-hydroxytryptophan and 5-hydroxytryptamine, can also result in pellagra, because hyperactivity of this pathway results in the diversion of tryptophan.

Chronic alcoholics represent a high-risk for the development of pellagra in western society, having poor diets and liver damage due to alcohol abuse. Certain psychoses, including depression and schizophrenia, are associated with abnormalities of tryptophan

Table 3 Niacin equivalents in selected foods

	Niacin equivalents from preformed niacin (mg per 100 g)	*Niacin equivalents from tryptophan (mg per 100 g)*	*Total Niacin equivalents (mg per 100 g)/(mg per MJ)*
Milk	0.2	0.6	0.8/2.9
Raw beef	5.0	4.7	9.7/18
Raw white fish	2.4	3.4	5.8/17
Raw eggs	0.1	3.7	3.8/6.1
Raw potatoes	0.6	0.5	1.1/3.4
Raw peas	2.5	1.1	3.6/10.5
Raw peanuts	13.8	2.9	19.3/8.2
White bread	1.6	2.0	3.6/2.5
Rice, uncooked	4.2	1.6	5.8/3.6
Maize (sweetcorn)	2.0	0.5	2.5/5.3
Cornflakes (fortified)	15.0	0.9	15.9/10.0
Coffee	24.8	2.9	27.7/87

Note: Coffee is a good source of niacin, because it is released from the bound form, trigonelline, by roasting. Much of the niacin in maize and part of that in other cereals, including unfortified bread, is unavailable for use by the body, being bound in large molecular complexes.
On average, 60 mg tryptophan yields 1 mg NE.

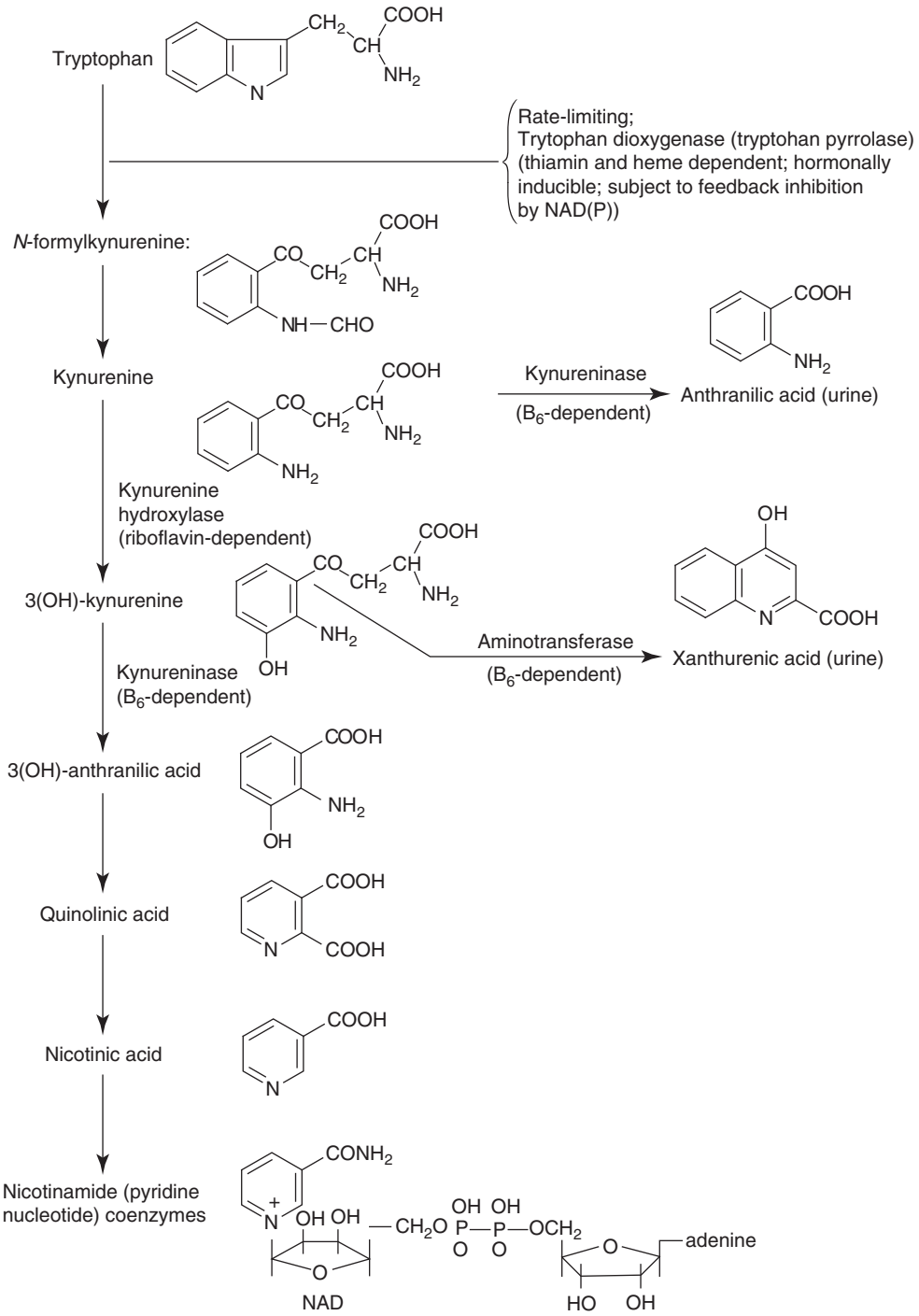


Figure 1 *In vivo* conversion of tryptophan to nicotinic acid and NAD.

metabolism pathways, including those involved in the formation of 5-hydroxytryptamine (serotonin) and 5-hydroxytryptophan in the central nervous system. Some patients may benefit from the modulation of these pathways by drugs and supplements. People with AIDS exhibit impairment of NAD production, which can respond to niacin supplements, and high-dose nicotinic acid is being used as part of the treatment of cardiovascular diseases (see the Section on Dietary Sources and High Intakes below). Diet-associated pellagra is still prevalent in some African countries, e.g., refugees in Malawi, and in parts of India and China.

Table 4 Recommended and reference intakes of niacin equivalents[a]

Age group	*UK*[b]		*USA*[c]
	LRNI	*RNI*	*RDA (mg Neiacin equivalents per day^{-1})*
	(mg Niacin equivalents per day)		
0–6 months	2	3	2
6–12 months	3–4	4–5	4
12 months–13 years	5–10	8–15	6–12
Adult	8–12	13–18	14–16
Lactation	10	15	17

This table provides a simplified summary of the published values.

[a]One niacin equivalent (mg NE) is equivalent to 1 mg niacin or one-sixtieth of the milligram tryptophan consumed.

[b]UK values are calculated on the basis of Lower Reference Nutrient Intake (LRNI) of 4.4 mg NE per 1000 kcal food energy, and Reference Nutrient Intake (RNI) of 6.6 mg NE per 1000 kcal, both of which are constant for all the population groups. The LRNI is intended to cover the needs of the lower 2.5% of a healthy population, whereas the RNI is intended to cover 97.5% of a healthy population.

[c]The US Recommended Dietary Allowances (RDAs) are intended to cover the needs of 97.5% of a healthy population.

Absorption, Transport, and Storage

Preformed niacin (formerly vitamin B_3) occurs in foods either as nicotinamide (niacinamide) or as the pyridine (i.e., nicotinamide) nucleotide coenzymes derived from it, or as nicotinic acid, which lacks the amide nitrogen, and is the form known as niacin in North America. Both nicotinamide and nicotinic acid are equally effective in vitamin potency, but in large doses, they exert markedly different pharmacological effects; thus, it is important, at least in that context, to make the distinction.

The most important sources of preformed niacin in most foods, particularly animal foods, are the pyridine (i.e., nicotinamide) nucleotides, NAD(H) and NADP(H). Hydrolases and pyrophosphatases in biological tissues convert these coenzymes to tissue sources of the vitamin. NAD glycohydrolase and pyrophosphatase enzymes are present in the gut mucosa, and absorption of nicotinamide or nicotinic acid by the mammalian intestine consists of a saturable transport system, dominant at low intakes and dependent on sodium, energy, and pH, plus a nonsaturable component that is dominant at high doses or intakes. Absorption is efficient even at high doses of 3 g or more: as much as 85% then being excreted in the urine. Absorption of test niacin doses introduced directly into the human upper ileum is rapid, peak levels appearing in blood plasma within 5–10 min. Absorption also occurs in the large intestine, making niacin from gut bacteria available to the intestinal cells.

Transport of niacin between the liver and the intestine can occur *in vivo*, as indicated by radioactive probes in animals, and the liver is a major site of conversion of niacin to its functional products: the nicotinamide–nucleotide coenzymes. Nicotinamide can pass readily between the cerebrospinal fluid and the plasma, thus ensuring a sufficient supply to the brain and the spinal cord. The liver contains greater niacin coenzyme concentrations than most other tissues, but all metabolically active tissues contain these essential coenzymes. As in the gut, both facilitated diffusion (which is sodium- and energy-dependent and saturable) and passive diffusion (which is nonsaturable) contribute to tissue uptake from the bloodstream. Except for muscle, brain, and testis, nicotinic acid is a better precursor of coenzymes than nicotinamide. The liver is the most important site of conversion of tryptophan to the nicotinamide coenzymes.

NAD is present mainly in an oxidized form in the tissues, whereas NADP is principally present in the reduced form, NADPH. There are important homeostatic regulation mechanisms that maintain appropriate ratios of the oxidized and reduced forms in healthy tissues. Once converted to coenzymes within the cells, the niacin is effectively trapped and can only diffuse out again after degradation into smaller molecules. This implies, of course, that the synthesis of coenzyme nucleotides must occur within each tissue and cell type, each of which must possess the enzymatic apparatus for their synthesis from niacin. Loss of nicotinamide and nicotinic acid into the urine is minimized (except when the intake exceeds requirements) by efficient reabsorption from glomerular filtrate.

Metabolism and Excretion

The conversion pathway of tryptophan to nicotinic acid *in vivo* is shown in **Figure 1**. The rate of conversion of tryptophan to niacin and the pyridine nucleotides is controlled by the activities of tryptophan dioxygenase (alternatively known as tryptophan pyrrolase), kynurenine hydroxylase, and kynureninase enzymes. These depend on other B-vitamins, glucagon, glucocorticoid hormones, and estrogen metabolites; moreover, competing pathways can affect the rate of conversion. The conversion may be increased threefold in pregnant women and in women taking oral contraceptives. Thus, a variety of nutrient deficiencies, toxins, genetic and metabolic abnormalities, etc. all influence the niacin status and requirements.

The two pyridine (i.e., nicotinamide) nucleotide coenzymes, formerly known as 'coenzymes I and II', then for a period as 'DPN and TPN,' but now as 'NAD' and 'NADP' are involved in hundreds of enzyme-catalyzed redox reactions *in vivo*. Although a few enzymes can use either cofactor, most are highly specific for one or the other.

Catabolism of the pyridine nucleotide coenzymes *in vivo* is achieved by four enzymes: NAD glycohydrolase, ADP ribosyl transferase, and poly (ADP ribose) synthetase (acting in sequence to liberate nicotinamide), and NAD pyrophosphatase (which releases nicotinamide mononucleotide for hydrolysis to nicotinamide). Turnover of nicotinamide then results in the formation of 1-methylnicotinamide (N^1-methyl nicotinamide or NMN), which is excreted into urine by the kidney, together with 1-methyl-2-pyridone-5-carboxamide and 1-methyl-4-pyridone-3-carboxamide (referred to as 2-pyridone and 4-pyridone, respectively). The concentrations of these excretory products can be used as indicators of whole-body niacin status (see the Section on Assessment of Niacin Status and Requirements below). At high intakes of niacin, as much as 85% may be excreted unchanged; however, the excretion of nicotinamide always predominates over that of nicotinic acid.

Other urinary excretion products of niacin include nicotinuric acid (nicotinoyl glycine), nicotinamide *N*-oxide, and trigonelline (N^1-methyl nicotinic acid); the latter may arise from bacterial action in the gut or from its absorption from foods. The pattern of metabolites varies between species and between diets (depending partly on the ratio of nicotinamide to nicotinic acid) and with niacin status, indicating complex regulatory mechanisms.

Hydrolysis of hepatic NAD to nicotinamide allows the release of niacin for other tissues. Protection of the coenzyme content of key enzymes such as glyceraldehyde 3-phosphate dehydrogenase confers protection on key metabolic pathways. In contrast, enzymes that catalyze pyridine nucleotide turnover may be hyperactivated in cells damaged by carcinogens, such as mycotoxins, thus starving them of essential cofactors and causing their death. This may help explain why moldy grain in the diet can increase the risk of pellagra when niacin and tryptophan intakes are marginal. In normal healthy cells, the compartmentalization of hydrolytic enzymes limits coenzyme turnover, but this is impaired in damaged cells.

Metabolic Function and Essentiality

The key functions of niacin involve its coenzymes, NAD and NADP, in the hydrogen/electron-transfer redox reactions in living cells. Like most B-vitamins, niacin is not extensively stored. Inadequate dietary intake leads to tissue depletion within 1–2 months and then successively to biochemical abnormalities, clinical signs of deficiency, and eventually death. As with other B-vitamins, rates of turnover and hence rates of excretion of breakdown products decline progressively as dietary deficiency becomes more severe. Thus the tissue coenzymes are relatively spared.

NAD is responsible for the release of energy during the oxidation of energy-producing fuels. NADP, however, functions mainly in the reductive reactions of lipid biosynthesis, and the reduced form of this coenzyme is generated via the pentose phosphate cycle. NAD is essential for the synthesis and repair of DNA and for supplying adenosine diphosphate (ADP) ribose ligands to lysine, arginine, and asparagine residues in proteins such as histones, DNA lyase II, and DNA-dependent RNA polymerase, and to polypeptides such as bacterial (e.g., diphtheria and cholera) toxins. In the nucleus, poly (ADP ribose) synthetase is activated by binding to DNA breakage points and is involved in DNA repair. It is also concerned with condensation and expansion of chromatin during the cell cycle and in DNA replication. It regulates the fidelity of DNA transcription, and some inflammatory and immune responses. Three different enzyme classes catalyze ADP-ribose transfer: (1) poly-ADP-ribose polymerases (PARP), (2) mono-ADP-ribosyl transferases (which modify G-proteins that regulate cell signalling and glucose-regulated proteins such as GRP78, which is a molecular chaperone, regulating protein-folding in the endoplasmic reticulum), and (3) enzymes that form cyclic ADP-ribose, which mobilizes calcium from intracellular stores. Nicotinic acid-ADP (NAADP), formed by desamidation of NADP, is also a calcium transport regulator. NAD is also required for the activity of silent information regulators (SIR or sirtuin enzymes), which are protein deacetylases for transcription regulation, genome stability, neuronal protection, and longevity. One of their substrates is p53, a protein of genome stability, DNA repair, and apoptosis. Niacin status affects the level of ADP ribosylation of these and other proteins, and may affect (1) chronic degenerative diseases like cancer, diabetes, and dementia, and (2) acute inflammatory conditions such as septic shock or stroke or myocardial infarction, with complex outcomes that are difficult to predict precisely. Nicotinic acid is part of the chromium-containing glucose-tolerance factor, whose functions are still being studied. A high level of poly (ADP ribose) synthetase activity, which is found in some tumors, can result in lower levels of NAD.

Because the electron-transport functions of NAD frequently involve flavin coenzymes, and because both flavin coenzymes and vitamin B_6 coenzymes are involved in the conversion of tryptophan to niacin *in vivo*, there are important metabolic interactions between these B-vitamins. Their clinical deficiency signs also converge, sometimes making it difficult to distinguish between deficiencies of different B-vitamins.

The body's need for niacin can be met completely by dietary tryptophan; thus, it is not, strictly speaking, an essential vitamin. In this respect, it resembles carnitine, which can be synthesized entirely from lysine, but for which, under some circumstances, a dietary requirement arises. Traditionally, however, niacin is classified as an essential vitamin, because some human diets lack both niacin and its precursor, tryptophan. Some animals such as sheep and cattle can synthesize sufficient niacin from tryptophan and thus do not need preformed niacin.

Assessment of Niacin Status and Requirements

Although the measurement of B-vitamin status is usually performed in blood samples, blood-based tests for niacin status are poorly developed. Some studies have suggested that erythrocyte concentrations of NAD or a reduction in the ratio of NAD to NADP may provide evidence of deficiency. The 95% range for healthy US adults for the NAD/NADP ratio was found to be 127–223, with a wide variation between different populations. The niacin coenzymes can be quantified either by enzyme-linked reactions or by their natural fluorescence in an alkaline solution.

At present, however, niacin status is most commonly assessed by assaying its breakdown products in urine. Of these, N^1-methyl nicotinamide (NMN) is the easiest to measure, by conversion *in vitro* to a fluorescent product. However, more reliable information can be obtained by measurement of urinary NMN plus the urinary pyridones (N^1-methyl-2-pyridone-5-carboxamide and N^1-methyl-4-pyridone-3-carboxamide), all of which can be detected and quantified by UV absorption following high-pressure liquid chromatography. The Interdepartmental Committee on Nutrition for National Defense (USA) has recommended an NMN excretion rate criterion of less than 5.8 μmol (0.8 mg) NMN per day in 24-h urine samples as evidence of biochemical deficiency.

The requirement for niacin to prevent or reverse the human deficiency signs is not known very precisely and depends on ancillary dietary deficiencies or other insults and pathologies. To estimate niacin requirements for dietary reference values, restoration of urinary excretion of NMN during controlled human depletion–repletion studies has been used, and on this basis, the average adult requirement has been estimated as 5.5 mg (45 μmol) of NEs per 1000 kcal (4200 kJ). After adding a 20% allowance for individual variation, this becomes 6.6 mg (54 μmol) per 1000 kcal (4200 kJ), which is the current reference nutrient intake in the UK (Table 4). Niacin requirements used to be expressed as a ratio to energy expenditure.

The average content of niacin in human breast milk is 8 mg (65.6 μmol) per 1000 kcal (4200 kJ), and this content forms the basis for intake recommendations (and dietary reference values) for infants up to 6 months of age. In the UK, the Reference Nutrient Intake niacin increment during pregnancy is nil, but during lactation, it is 2 mg day^{-1}.

Dietary Sources and High Intakes

From Table 3, different foods differ considerably, not only in their total contribution to niacin equivalents but also in the ratio of the contribution from preformed niacin and from tryptophan. In a typical Western diet, preformed niacin provides approximately 50% of the niacin supply in the diet. As for other B-vitamins, meat, poultry, and fish are excellent sources of niacin equivalents, followed by dairy and grain products, but certain grains such as maize and highly polished rice are very poor sources. Both nicotinamide and nicotinic acid have potentially useful pharmacological properties at high intakes.

Nicotinic acid, which has marked antihyperlipidemic properties at daily doses of 2–6 g, is of greatest interest in pharmacological terms. Large doses of nicotinic acid reduce the mobilization of fatty acids from adipose tissue by inhibiting the breakdown of triacylglycerols through lipolysis and by inhibiting hepatic triacylglycerol synthesis (by inhibiting hepatic diacylglycerol acyltransferase 2), thus limiting the assembly and secretion of very low-density lipoproteins from the liver and reducing serum cholesterol. It inhibits high-density lipoprotein (HDL) catabolism by decreasing the surface expression of the hepatic ATP synthase beta chain, thereby increasing HDL levels, and it may increase the ratio of HDL_2 to HDL_3, together with a reduced rate of synthesis of apolipoprotein A-II and a transfer of some apolipoprotein A-I from HDL_3 to HDL_2. Such blood lipid effects appear to be due to binding of nicotinic acid to a receptor called a 'niacin receptor' or HM74A, linked to an inhibitory G-protein, and leading to reduced cyclic-AMP levels and inhibition of a hormone-sensitive lipase. Nicotinic acid and niacin can also increase the redox potential in arterial cells. The unwanted side-effect of skin flushing is linked to prostaglandin D2 and E2 formation by macrophages; this can now be reduced by a prostaglandin (DP1) receptor antagonist such as laropiprant. These effects of high-dose nicotinic acid may be beneficial in managing dyslipidemias and in reducing the risk of cardiovascular disease, and it has recently been successfully combined with statin therapy plus laropiprant. If given intravenously, large doses of nicotinic acid can, however, produce undesirable side-effects such as vasodilatation, hypotension, nausea, vomiting, diarrhea, and gastrointestinal disturbance, headache, fatigue, difficulty in focusing, skin discoloration, dry hair, sore throat, jaundice, changes in liver function tests, changes in carbohydrate tolerance, and changes in uric acid metabolism including hyperuricemia. Hyperuricemia may result from effects on intestinal bacteria and enzymes, or on renal tubular function, and is especially severe if sustained-release preparations of nicotinic acid are used. Except under medical supervision, the upper tolerable limit for niacin intake has been set at 35 mg NE per day for adults and 10–30 mg day^{-1} for children and adolescents (1–18 years), and is linked to the skin flushing reaction to nicotinic acid supplements.

A large trial for the secondary prevention of myocardial infarction by high-dose nicotinic acid, with a 15-year follow-up, has produced convincing evidence for moderate but significant protection against mortality, which was attributed to the cholesterol-lowering effect or reduction of reinfarction.

Nicotinamide does not share these effects of nicotinic acid on lipid metabolism or its associated toxicity. However, it has been shown to act as an inhibitor of poly (ADP ribose) synthetase in pancreatic β-cells in animal studies, thereby protecting them from chemically-induced diabetes, albeit sometimes complicated by beta-cell tumor formation. Studies on human diabetes have so far yielded inconsistent results.

Because niacin deficiency or dependency can exacerbate some types of mental illness such as depression or dementia, and because the correction of depressed brain levels of serotonin would be advantageous, there have been attempts to treat depression

with tryptophan or niacin; however, these have so far had only limited success. Schizophrenic patients have been treated with nicotinic acid, because the synthesis of NAD is impaired in some parts of their brains. Recent research has indicated that nicotinamide riboside uniquely supports neuronal NAD synthesis, thereby deserving future therapeutic studies.

In conclusion, niacin and some of its related compounds have many wide-ranging metabolic effects, and their medical significance has attracted renewed interest in recent years.

Further Reading

Bender, DA, 1992. Niacin. Nutritional Biochemistry of the Vitamins. Cambridge University Press, Cambridge ch. 8, pp. 184–222.
Bourgeois, C, Cervantes-Laurean, D, Moss, J, 2006. Niacin. In: Shils, ME, Shike, M, Ross, AC, Caballero, B, Cousins, RJ (Eds.), Modern Nutrition in Health and Disease. Lippincott Williams and Wilkins, Baltimore, pp. 442–451.
Carpenter, KJ (Ed.), 1981. Pellagra: Benchmark Papers in the History of Biochemistry/II. Lewden, Hutchinson and Ross Publishing Co, Pennsylvania.
Carpenter, KJ, Lewin, WJ, 1985. A reexamination of the composition of diets associated with pellagra. American Journal of Clinical Nutrition 115, 543–552.
Combs Jr, GF, 1998. Niacin. The Vitamins. Fundamental Aspects in Nutrition and Health, 2nd edn. Academic Press, New York. ch. 8, pp. 312–331.
Di Palma, JR, Thayer, WS, 1991. Use of niacin as a drug. Annual Review of Nutrition 2, 169–187. View Record in Scopus | Cited By in Scopus (31).
Fu, CS, Swendseid, ME, Jacob, RA, McKee, RW, 1989. Biochemical markers for assessment of niacin status in young men: Levels of erythrocyte niacin coenzymes and plasma tryptophan. Journal of Nutrition 119, 1949–1955.
Goldberg, AC, 2004. A meta-analysis of randomized controlled studies on the effects of extended-release niacin in women. American Journal of Cardiology 94, 121–124.
Horwitt, MK, Harvey, CC, Rothwell, WS, Cutler, JL, Haffron, D, 1956. Tryptophan–niacin relationship in man. Journal of Nutrition 60 (supplement 1), 1–43.
Jacob, RA, 2006. Niacin. In: Bowman, BA, Russell, RM (Eds.), Present Knowledge in Nutrition, 9th edn. ILSI Press, Washington DC, pp. 260–268. ch. 20.
Jacob, RA, Swendseid, ME, McKee, RW, Fu, CS, Clemens, RA, 1989. Biochemical markers for assessment of niacin status in young men: Urinary and blood levels of niacin metabolites. Journal of Nutrition 119, 591–598.
Kamanna, VS, Ganji, SH, Kashyap, ML, 2009. Niacin: An old drug rejuvenated. Current Atherosclerosis Reports 11, 45–51.
Kamanna, VS, Kashyap, ML, 2008. Mechanism of action of niacin. American Journal of Cardiology 101, 20B–26B.
Kirkland, JB, 2007. Niacin. In: Zempleni, J, Rucker, RB, McCormick, DB, Suttie, JW (Eds.), Handbook of Vitamins, 4th edn. CRC Press, Boca Raton, USA, pp. 191–232. ch. 6.
Sauberlich, HE, 1999. Niacin (nicotinic acid, nicotinamide). Laboratory Tests for the Assessment of Nutritional Status, 2nd edn. CDC Press, Boca Raton, USA. pp. 161–174.

Nucleic acids, purine, and pyrimidine nucleotides and nucleosides: Physiology, toxicology, and dietary sources

E.A. Carrey, UCL Institute of Child Health, London, UK
D. Perrett, Queen Mary University of London, London, UK
H.A. Simmonds†

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of E.A. Carrey, D. Perrett, H.A. Simmonds, Nucleic Acids, Purine, and Pyrimidine Nucleotides and Nucleosides: Physiology, Toxicology, and Dietary Sources, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 189–196, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00191-4.

Introduction 330
Physiology 330
Structure 330
Nucleic Acid Biosynthesis in Humans 331
Metabolic Roles of Nucleotides 331
Synthesis of Nucleic Acids 332
Metabolism of Endogenous Nucleic Acids and Excretion of Metabolic End Products 332
Metabolism of Dietary Nucleic Acids in Humans 333
Nucleic Acid Content of Foods 334
Effect of Cooking on Nucleotide Content of the Diet 334
Nucleic Acid and Related Compounds in Beverages 334
Nucleotides in Human Breast Milk and Infant Formula Milks 337
Beneficial Effects of Dietary Nucleosides and Nucleotides 337
Purine Ribonucleotides as Flavor-Enhancing Additives 337
Purine Ribonucleosides and Bases as Markers of Food Quality 338
Toxicology 338
Pharmacological Uses for Nucleosides and Nucleotides 338
Toxicity of Exogenous Nucleic Acids to Humans 338
Urolithiasis and Other Kidney Stones 339
Further Reading 339

Introduction

Nucleic acids, vital components of all living cells, were isolated in 1869 from the nuclei of pus cells and the spermatozoa of Rhine salmon. Later, it was shown that the major constituents of nucleic acids are sugars, phosphate groups, and the characteristic purine and pyrimidine bases. The chemical structures of the purine bases, including uric acid – the end (waste) product of purine metabolism in humans – were established at the end of the 19th century. The role of the nucleic acids in storing and translating the genetic information in the cells was elucidated in the 20th century. Although their calorific contribution to the diet is trivial, nucleotides, nucleosides, and bases have essential roles in metabolism and signaling within the cell and organism.

This article outlines the structure of nucleic acids and gives a brief overview of the physiological functions of nucleosides, nucleotides, and nucleic acids. It describes the toxicity that may arise from degradation of both endogenous and dietary (exogenous) nucleic acids in humans and contains a summary of the nucleic acid content of foods.

Physiology

Structure

The hereditary material in the nucleus of human cells is packed into 46 chromosomes and additional DNA is found in the mitochondria. The capacity of DNA to be copied into two complementary strands arises from the well-known double-helical structure and underlies the transfer of genetic information in all living organisms. Interactions between the DNA and transcription factors determine the time and place in the body where genes are transcribed, controlling development and metabolism.

†Deceased.

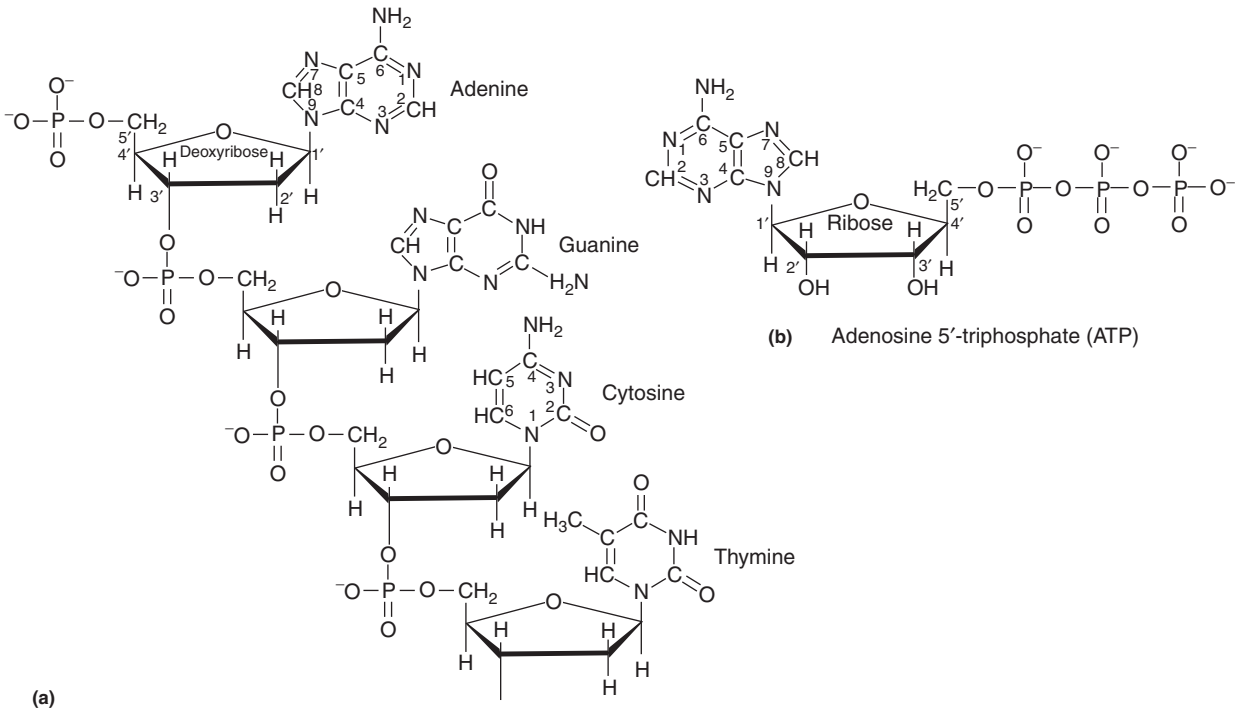


Figure 1 (a) Schematic representation of part of a DNA strand showing the structural formulae of the four constituent bases, adenine, guanine, cytosine, and thymine, linked *via* the 3′-OH group of the deoxyribose moiety to the 5′-phosphate group of the next nucleotide. Also shown is the numbering of the atoms in the deoxyribose, as well as the pyrimidine and purine rings. The bases are adenine (A), guanine (G), cytosine (C), and thymine (T). Two strands are wound in opposing chemical directions to allow the well-known double-helix structure, with hydrogen-bonding between complementary nucleotides (i.e., A–T and G–C), to form. The deoxyribose and phosphate groups form the outer sides of the 'ladder'. The RNA molecule is single-stranded, but double-helical regions arise when stretches of complementary sequences allow hairpin loops to form. In addition, the base uracil (U) is found instead of thymine, and the pentose is ribose. (b) Structural formula of ATP indicating that the ribose, as distinct from deoxyribose, has an OH group at the 2' position on the pentose ring. When a nucleoside triphosphate (NTP) is linked through the 5′ phosphate groups to the 3′ position of the previous residue on the growing chain, the chemical energy for the polymerization is provided by the removal of the second and third phosphate groups.

RNA molecules are synthesized initially on a DNA template by a DNA-dependent RNA polymerase in a process called transcription, where ribonucleotides complementary to the bases of one strand of DNA are joined by 3′–5′ phosphodiester bonds (Figure 1(a)).

Nucleic Acid Biosynthesis in Humans

The first step in nucleic acid synthesis involves the formation of the purine and pyrimidine ribonucleotides. There are two endogenous routes: either the energetically expensive *de novo* route from small molecules such as carbon dioxide, amino acids, and ribose sugars, or the energetically less expensive 'salvage' pathway. Purine bases and pyrimidine nucleosides from the breakdown of nucleic acids and nucleotide cofactors are salvaged within the cells, generating nucleotides that can be incorporated into nucleic acids. In most cells, salvage processes are more important, and the ribonucleotides recycled in this way exert feedback control on the *de novo* routes.

Metabolic Roles of Nucleotides

The most abundant ribonucleotide in the body is adenosine 5′–triphosphate (ATP), which is the universal energy carrier in living organisms (Figure 1(b)). In addition, nucleotides are precursors of several coenzymes, used in many reactions including the conversion of food into energy. Within cells adenosine and guanosine nucleotides also have roles in the transduction of external signals into cellular responses, and in the translation and synthesis of proteins. Pyrimidine nucleotides are present at much lower concentrations in cells but also fulfill diverse functions. Uridine diphosphate (UDP)-glucose and Cytidine diphosphate (CDP)-lipids are active intermediates in the synthesis of glycogen and membranes, respectively, and sugars linked to UDP or GDP are used in the glycosylation of proteins. UDP-glucuronic acid is an essential component of the pathways that convert exogenous molecules and endogenous steroids into soluble forms for disposal from the body.

The free deoxyribonucleotides are very scarce in the normal cell because they are used exclusively for the synthesis of DNA.

Synthesis of Nucleic Acids

Synthesis of both DNA and RNA is prominent in cells and tissues with high turnover or metabolism (e.g., liver, gut epithelium, skin, dividing lymphocytes, bone marrow, and hair follicles). Different complements of enzymes are expressed in each cell type, and therefore tissues have characteristic profiles of internal metabolites, including nucleotides and nucleosides. For example, in cells that do not continuously divide, such as heart and muscle, nucleotide profiles are relatively simple, relating to the major requirement to sustain high levels of ATP and cofactors. Contrastingly, rapidly dividing cells in liver and intestine show a complex nucleotide pattern, identifying these organs as major sites of nucleic acid metabolism. The gut is particularly important in this respect. The rate of cell turnover in the luminal villi is high, and it has been calculated that in rats approximately 30 mg of endogenous nucleic acid derived from dead cells enters the gut lumen daily. This means that nucleic acid synthesis in liver and intestine is much higher than in tissues such as muscle.

Metabolism of Endogenous Nucleic Acids and Excretion of Metabolic End Products

There is a considerable daily turnover of endogenous nucleic acids and ribonucleotides during muscle work, wound healing, erythrocyte senescence, mounting an immune response, etc. However, only a small fraction of these vital endogenous compounds is degraded and lost from the body. Because *de novo* purine and pyrimidine synthesis is energetically expensive the contents of dead cells are normally used by other cells, and degraded RNA or cofactors are recycled within living cells using active 'salvage' routes.

Breakdown products of DNA and RNA enter the salvage pathways in the form of the purine bases hypoxanthine (Hx) and guanine or the pyrimidine bases uracil and thymine (Figure 2).

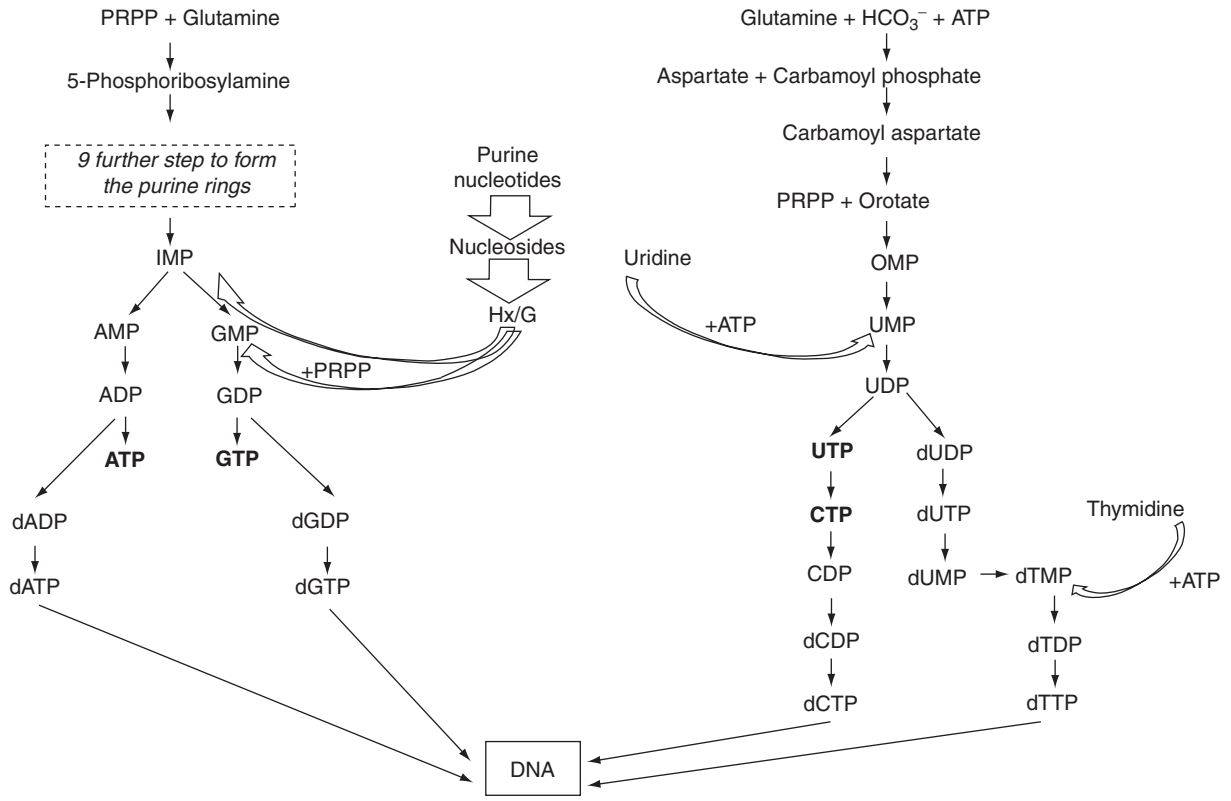


Figure 2 *De novo* synthesis of ribonucleotides uses small molecules and amino acids. The purine base of IMP is built up in several steps on a ribose phosphate molecule, and separate pathways generate ATP and GTP. In pyrimidine biosynthesis, the six-membered ring is formed before addition of the ribose phosphate to orotate, and decarboxylation to form UMP. Ribonucleotides (ATP, GTP, UTP, and CTP, shown in bold) are used in the synthesis of RNA, whereas DNA is synthesized after conversion of the ribose to deoxyribose, and of dUMP to dTMP. Salvage pathways (shown with open arrows) all follow two patterns: purine nucleotides and nucleosides, entering the cells or derived from hydrolysis of cofactors or nucleic acids, are converted to the free bases hypoxanthine or guanine and then converted by a specific phosphoribosyltransferase enzyme to the nucleoside; in contrast pyrimidines are salvaged from the nucleosides uridine or thymidine using specific kinase enzymes. It should be noted that PRPP (phosphoribosyl pyrophosphate) is essential for both *de novo* pathways and the purine salvage pathways.

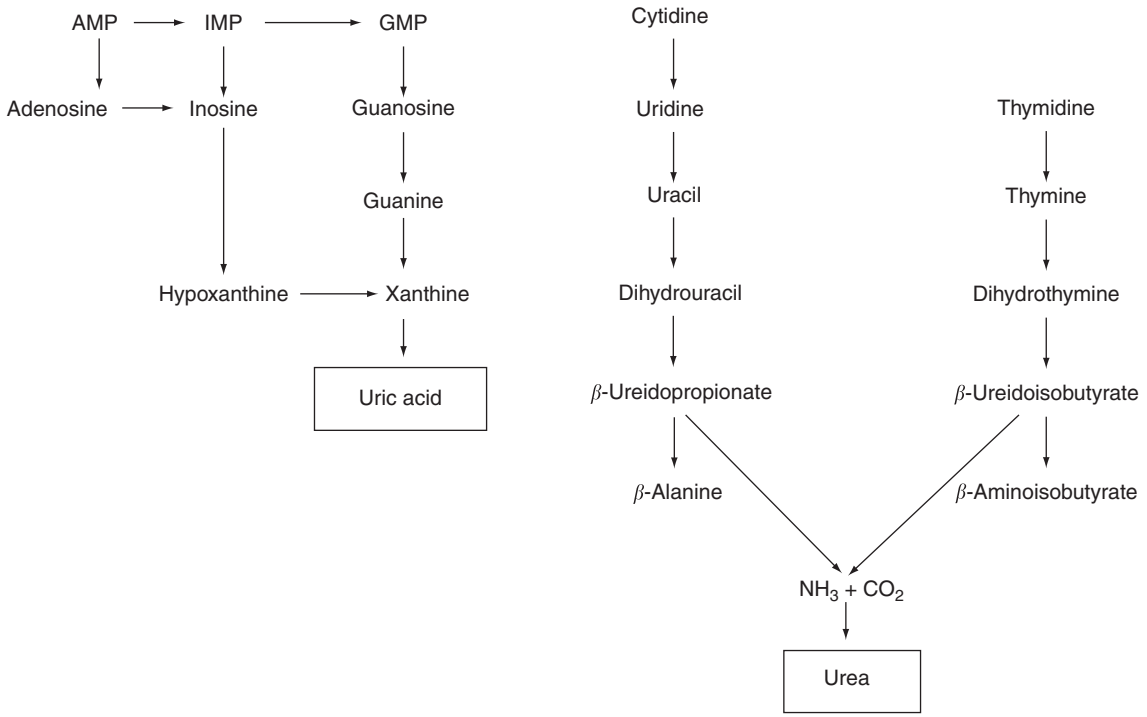


Figure 3 Breakdown of DNA and RNA begins when the molecules are degraded by nuclease enzymes to liberate nucleotides. The next step, degradation by specific 5′–nucleotidases (removing the phosphate groups) to nucleosides or deoxynucleosides, is essentially irreversible. Nucleoside phophorylases generate (a) the purine bases hypoxanthine and xanthine, which are converted to uric acid or (b) the pyrimidine bases uracil and thymine, which are converted to β-amino acids, ammonia, and CO_2, and thus to urea, a solute in urine.

Any pyrimidine bases that are not salvaged are then further catabolized in a series of steps to β-amino acids, which are soluble and readily excreted. There is thus normally no measurable toxic pyrimidine end product from endogenous or dietary nucleic acids, except in the case of a small number of very rare metabolic disorders (**Figure 3**).

The purine base Hx is converted in the liver to the insoluble metabolites xanthine and then uric acid by the enzyme xanthine oxidase/xanthine dehydrogenase (XDH) in man. Urate can normally be disposed of in the urine, but high concentrations can crystallize and form kidney stones or deposits in the joints and under the skin. Some rare genetic disorders can remove feedback regulation of purine biosynthesis, or excessive breakdown of cells may overload the salvage system, each resulting in very high endogenous levels of uric acid. However most other animal species (with the notable exception of some dog varieties) possess an additional catabolic enzyme, uricase, which cleaves the purine ring of uric acid forming the readily soluble allantoin. This compound in turn may be degraded to ammonia in water-dwelling species such as fish.

Metabolism of Dietary Nucleic Acids in Humans

The human diet is naturally rich in nucleic acids because food is derived from once-living organisms. Because, as already described, nucleotides and nucleosides can be synthesized *de novo* and nucleobases liberated during catabolism can be salvaged, they are often considered to be dispensable nutrients in food. The metabolism of these exogenous nucleic acids follows a similar pattern to the intracellular process described above, but the bacterial flora of the intestine are the first point of degradation. This digestion is rapid. Studies in both pigs and humans demonstrated that up to 50% of dietary purine was degraded to carbon dioxide within 30 min, 43% was recovered in the urine and 5% excreted in the feces. Less than 2% of dietary purines is incorporated into nucleic acids.

Humans thus have no apparent essential requirement for purines from the diet and the intestinal mucosa provides an effective barrier to their uptake through a battery of enzymes that can rapidly degrade purine nucleotides, nucleosides, and bases especially unusual purines found in plant materials. Because of this enzyme activity, and the rapid turnover of intestinal mucosa, approximately 200 mg of urate is excreted daily in the feces. This phenomenon ensures that levels of ATP do not fluctuate in concert with the dietary intake of purines, or may represent an important evolutionary development to protect the integrity of cellular DNA.

On the other hand, pyrimidine ribonucleoside monophosphates (NMPs) and ribonucleosides are absorbed readily from the intestine and utilized for nucleic acid synthesis. This has been demonstrated by studies of humans with hereditary oroticaciduria, a rare defect in conversion of orotic acid to uridine monophosphate (UMP) in *de novo* pyrimidine synthesis. Such patients have severe megaloblastic anemia. They have been sustained on oral uridine, indicating that the dietary pyrimidine nucleoside can

Table 1 A quick reference guide to the purine (nucleic acid) content of foods

Foods and Beverages Rich in Nucleic Acids/Purines
Offal: sweetbreads, liver, kidney, heart, and paté
Wild or farmed game meats (venison, pheasant, rabbit, hare)
Seafoods: sardines, sprats, herring, bloaters, anchovies, fish roe, caviar, taramasalata, trout or salmon; lobster, crab, prawns
Vegetables: asparagus, avocado pears, peas, spinach, mushrooms, broad beans, cauliflower
Pulses and grains: legumes, pulses and soya products such as bean curd, tofu, Quorn
Cereals: all bran, oat, rye or wheat cereals and products; wholemeal, rye and brown breads
Other: beer and yeast extracts/tablets (Barmene™, Tastex™). Meat or vegetable extracts (Marmite™, Vegemite™, Bovril™, Oxo™) Blue-green alga extracts (Spirulina)
Foods that are Moderate or Low Sources of Purines
Beef, lamb, pork (steak or chops), bacon, ham, sausages, some poultry, tongue (all should be eaten in moderation)
Carrots, parsnip, other root vegetables, potatoes, lettuce, leeks, cabbage, sprouts, marrow, squash, courgettes
Peanuts, cashew nuts
Breakfast cereals
Some fish (see Table 2)
Foods and Beverages that are Essentially Purine-Free
Milk, cheese, eggs, butter, margarine, cream, ice cream
White bread or flour, cakes, scones, biscuits
Sugar, jam, marmalade, honey, and sweets
Cucumber, tomato, onions, pumpkin
Fresh, cooked or tinned fruits, nuts
Puddings, custards, yogurt
Fruit juices, soft drinks

compensate totally for lack of *de novo* synthesis in humans. Studies using radiolabelled purines and pyrimidines in mice provided further evidence for the incorporation of dietary pyrimidine nucleosides, but not purine nucleosides, into hepatic RNA.

Nucleic Acid Content of Foods

The nucleic acid content of foodstuffs is expressed generally in terms of purine equivalents or 'total potentially available nucleosides (TPAN)' released from food by hydrolysis with sodium hydroxide or hydrochloric acid and enzymes. The data for purines are thus derived by analysis of the resultant constituent bases. Analysis by Robert McCance, Elsie Widdowson, and colleagues from the 1930s onward forms the basis of tables listing the composition of foodstuffs although, with a few exceptions, this demanding work has not been repeated using more modern analytical methods such as liquid chromatography–mass spectrometry (LC–MS).

Foods can be classified into three groups; high, low, or essentially free of purines (and hence of pyrimidines too) (**Tables 1** and **2**). As a general rule growing organisms such as yeast, or rapidly metabolizing tissues such as liver, will be rich in both nucleic acids. Seeds and grain are good sources of the genetic material, DNA, as well as free nucleotides, which are stored for use in germination. Muscle tissue is an excellent source of the nucleotide ATP and the nucleic acids in mitochondria. Offal is also metabolically very active so is usually high in free nucleotides as well as nucleic acids. Fish and shellfish that are eaten 'whole' and fish eggs and roe are also high in nucleic acids. Extracts of meat and yeast e.g., Bovril, Marmite, Vegemite, have very high purine contents, as do supplements such as Spirulina for sale in 'Health Food' shops, but are usually eaten in small quantities.

Fats, white flour, sugar, and fruit juices have been separated from the 'living' part of the food and so they are poor sources of nucleic acids.

Effect of Cooking on Nucleotide Content of the Diet

Nucleic acids are relatively resistant to hydrolysis at the moderate temperatures and short periods of time associated with cooking in water or frying in oil. On the other hand nucleoside triphosphates (NTPs) and nucleoside diphosphates (NDPs) break down readily during boiling in water forming first their related NMP and then their base. The rate of hydrolysis is significantly increased in acidic solutions. The rate of degradation is enhanced if any enzyme activity is still present.

The levels of nucleic acids and NTPs are well maintained during prolonged storage at −20 °C or below.

Nucleic Acid and Related Compounds in Beverages

Tea, coffee, and cola drinks contain a number of unusual nucleobases based on xanthine (**Figure 4**). Caffeine (1,3,7-trimethylxanthine) and theobromine are the best known. Caffeine is found in various quantities in the beans, leaves, and fruit of many plants. It is mainly consumed by humans in infusions extracted from the coffee bean and leaves of the tea bush. A cup of coffee can contain up to 175 mg of caffeine whereas a cup of tea contains approximately 40 mg. Many popular soft drinks,

Table 2 Concentrations of purines in some common foods; results are recorded relative to 100 g of food for purine and for protein, although serving size for each ingredient may be larger or smaller than 100 g based on Diem and Lentner (1970).

Food	*Purine (mg per 100 g)*	*Protein (g per 100 g)*
Meat		
Beef liver	333	19.7
Beef kidney	285	15.4
Beef heart	285	16.8
Beef tongue	167	16.4
Beef steak	151	19.5
Calf liver	348	19
Sweetbreads	1212	19.6
Veal cutlet	152	19.2
Sheep kidney	312	16.8
Lamb chop	196	14.9
Pork liver	289	22
Pork cutlet	164	16.4
Bacon	85	9.1
Ham	136	19.5
Sausage (beef)	79	13.8
Sausage (pork)	66	11.5
Rabbit	118	20.4
Venison	156	20
Vegetables		
Asparagus	32	2.1
Cauliflower	32	2.1
Celery	20	1.1
Kohlrabi	44	2.1
Mushrooms	72	3.5
Peas	72	6.7
Spinach	96	2.2
Dried Legumes		
Split peas	195	21
Red bean	162	20
Lentils	222	28
Haricot beans	230	22
Lima bean	149	21
Other		
Bovril™	340	18
Marmite™	356	2
Oxo™ cubes	236	10
Yeast extracts	2257	46
Poultry		
Chicken flesh	181	20.6
Chicken liver	372	22.1
Chicken heart	223	18
Duck	181	16
Goose	177	16.4
Turkey	239	20.1
Fish, Seafoods		
Anchovies	411	20
Bass	73	19.5
Bloaters	133	22.6
Bream	72	19.7
Cod	62	18
Crab	61	19.2
Clams	136	17
Eel	108	18.6
Fish cakes	36	12.1
Herring	378	17
Kippers	91	21.2
Lobster	100	20

(Continued)

Table 2 Concentrations of purines in some common foods; results are recorded relative to 100 g of food for purine and for protein, although serving size for each ingredient may be larger or smaller than 100 g based on Diem and Lentner (1970).—cont'd

Food	*Purine (mg per 100 g)*	*Protein (g per 100 g)*
Lemon sole	54	19.9
Mackerel	246	29
Plaice	53	18.1
Salmon	250	23
Sardines	345	23
Scallops	117	22.3
Sprats	250	25.1
Squid	135	15
Trout	92	19.2
Canned Seafoods		
Anchovies	321	30
Herring	378	17
Mackerel	246	26
Oysters	116	6
Salmon	88	26
Sardines	399	24
Shrimp	231	22
Tuna	142	29

e.g., cola, may contain up to 300 mg per can. Theobromine (dimethylxanthine) occurs naturally in cocoa beans (20 mg g^{-1} of cocoa powder) and is therefore present in chocolate. Theophylline (1,3-dimethylxanthine) occurs at trace levels in tea leaves and is now chemically synthesized for use as a drug treatment for asthma.

Caffeine is rapidly absorbed by the stomach and small intestine and is distributed throughout total body water. It is rapidly metabolized in the liver to paraxanthine (84%), theobromine (12%), and theophylline (4%).

In man, caffeine acts as a central nervous system stimulant, temporarily warding off fatigue and restoring alertness: it readily enters the brain and acts as a nonselective antagonist of purinergic adenosine receptors. Caffeine is the most widely consumed psychoactive substance; in North America, 90% of adults consume caffeine beverages daily. 'Energy drinks' contain very high levels of caffeine and some also contain alcohol, so over-use can lead to a 'wide-awake but drunken' state. The US Food and Drug Administration (FDA) lists caffeine as a 'multiple purpose drug generally recognized as safe food substance'.

Because beers and related drinks are produced by fermentation of grains with yeasts, the process is associated with vast increases in cell numbers and this leads to drinks with a considerable nucleic acid and nucleotide content, even if the yeasts are removed by filtration. The economic importance of brewing, as well as the clinical relevance of beer and wine in gout, means that there is considerable literature on the composition of beers and wines including their purine levels. A traditional British beer (ale or bitter) can contain up to 25 mg purine per litre (250 µmol l^{-1}) but lager beers have up to 20 mg l^{-1}. Ciders contain <1 mg l^{-1} of purine. Wine also contains significant amounts of purines. Some low alcohol beers may contain three times these levels of purines. Spirits contain very little in the way of purines because these compounds are removed in the distillation step.

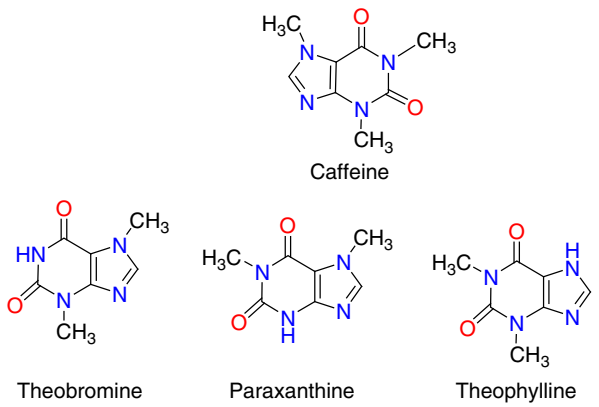


Figure 4 Chemical structures of unusual nucleobases derived from xanthine, and found in plant tissues such as cocoa beans, coffee beans, and tea leaves.

Nucleotides in Human Breast Milk and Infant Formula Milks

Human breast milk contains nucleic acids, nucleotides, and nucleosides, particularly cytidine and uridine, with a profile that reflects the diet of the mother. Average TPAN concentrations in human milk are 172–222 μmol l^{-1} (59–76 mg l^{-1}) at all stages of lactation. The content derived from cells is approximately 18% of TPAN during the first few days of lactation but drops to less than 10% later. Proportions of nucleosides derived from RNA (43–48%); free nucleotides (36–40%); free nucleosides (6.6–8%); and nucleotide adducts (9–10%) are similar in milk from women of several races. It is not known if all of these compounds in human milk are used by the breast-fed infant. However there is also a movement to supplement formula milks (based on cow's milk) with ribonucleotides derived from hydrolyzed RNA. Cow's milk contains lower levels of nucleotides, with a different profile from the human, but significant levels of the *de novo* pyrimidine intermediate orotic acid, which is low in humans (raised in milk from mothers who smoke).

In the late 1990s the FDA agreed to the nucleotide supplementation of infant formula based on cow's milk, but at lower concentrations than in human milk. The EU Food Committee recommended in 2003 (updated advice in 2007) that the content of nucleotides, if added to infant formulae and in follow-on formulae, should not exceed 5 mg per 100 kcal. If added the maximum nucleotide contents should be: cytidine monophosphate (CMP) 2.5 mg per 100 kcal, uridine monophosphate (UMP) 1.75 mg per 100 kcal, adenosine monophosoate (AMP) 1.5 mg per 100 kcal, guanosine monophosphate (GMP) 0.5 mg per 100 kcal, inosine monophosphate (IMP) 1 mg per 100 kcal.

Several trials have evaluated the effects of nucleotide addition to formula milk in infants but only two of the trials have studied formulae with 'human' nucleotide levels of 72 mg l^{-1}. Thus, there is no adequate scientific basis at present to conclude that the addition of nucleotides in higher concentrations than presently permitted for infant formula would provide additional benefits. Formula milks based on soy protein have a high natural content of nucleotides and are therefore not supplemented.

Beneficial Effects of Dietary Nucleosides and Nucleotides

In healthy adults, a normal varied diet is a good source of nucleic acids, nucleotides, and nucleosides, and supplementation is thought to be unnecessary.

There is substantial evidence (principally from research in animal models) that the presence of nucleotides or nucleosides in the diet helps cellular proliferation in the gut, in postoperative trauma, and in the development of the immune response in infants. A medical food supplement containing arginine, glutamine, nucleotides, and omega-3 fatty acids, demonstrates a better clinical outcome for (adult) surgical patients, and a 30% reduction in risk of infection. It should be noted that glutamine is a precursor for *de novo* synthesis of nucleosides as well as a source of energy for proliferating cells.

Dietary nucleotides have been shown to promote the incorporation of essential fatty acids into membrane lipids in healthy newborn babies, and to enhance the integrity and maturation of the intestine and of the immune system, and thus may contribute to the improved immunity seen in breast-fed infants. Studies in lower socioeconomic groups have found that supplementation of formula with 14.2 mg free nucleotides per 100 g milk powder resulted in a significant reduction of first episodes of diarrhea. This may be linked to an alteration in the bowel flora, leading to a predominance of lactobacilli as seen in breast-fed babies.

An extract from sugar cane (trade name NucleomaxX), containing 17% nucleosides, is used in the HIV-positive community to counteract the unpleasant side-effects of HIV drugs that inhibit the formation of mitochondrial DNA and hence energy-producing processes. The use of oral uridine in metabolic disorders is described later.

Nucleotides based on both adenosine and uridine can activate the purinergic receptors on a wide range of cell types. Nucleotides influence the transcription of several genes in intestinal cells, and have been shown to improve growth and maturation of the gut in weanling rats.

In lymphocytes and other cells, synthesis of nucleotides *de novo* is expanded dramatically when a signal for proliferation is received; the rate of pyrimidine biosynthesis increases more than purine biosynthesis. Thus nucleotides are now considered to be 'conditionally essential' because their provision in the diet may provide help through the salvage system where cells are dividing rapidly or where other nutrients, used as precursors, are scarce.

Purine Ribonucleotides as Flavor-Enhancing Additives

The purine 5′-nucleoside monophosphates IMP and GMP, derived from degradation of RNA, have received much attention as the taste-active components in a variety of seafoods and meat. These purine 5′-nucleotides, but not the pyrimidine nucleotides CMP and UMP, enhance the savory flavor generated by monosodium glutamate (MSG), by interaction with receptors on the specific *umami* taste buds in the mouth. Because ATP is the major free nucleotide in muscle cells, its breakdown into the flavor-enhancing IMP provides a scientific rationale for the improved palatability of meat or game birds that have been hung for several days after slaughter. Similarly, the distinctive flavors of several cheeses are related to the metabolism, by bacteria, of the characteristic range of nucleotides present in the original milk.

Purine Ribonucleosides and Bases as Markers of Food Quality

Related to the above topic is the role of hypoxanthine (Hx) in the determination of food quality. As described earlier, when an animal is killed the tissues become ischemic and the intracellular ATP starts to degrade, forming first AMP and then Hx. At room temperature the majority of ATP will have degraded within 24 h. The Hx level will be maximal at approximately 2 weeks in meat stored at 4 °C. The change in Hx content of the food alters the sensory perception of the food, with higher Hx levels causing a bitterness in the taste of meat. This aspect of purine catabolism has been particularly well documented in seafood, which is perhaps the most perishable of foods. Hx in fish and fish products such as fish fingers increases linearly with storage time, and measurement of the Hx levels has been recommended as a marker of fish spoilage.

Toxicology

Pharmacological Uses for Nucleosides and Nucleotides

Rare genetic disorders highlight the sensitivity of lymphocytes to the efficient removal of waste from DNA catabolism, and in fact have provided the basis for development of novel immunosuppressant drugs which inhibit enzyme activities crucial to removal of purine nucleotides, and lead to mis-incorporation of nucleotides in DNA synthesis. On the other hand, the use of certain nucleotide analogs as drugs depends on their incorporation into DNA – for example, analogs used in HIV therapy are incorporated by the reverse transcriptase of the virus, and bring the reaction to a halt. Toxicity associated with several analogs is now known to arise from erroneous incorporation into the patient's mitochondrial DNA, because of less stringent proof-reading by the mitochondrial DNA polymerase enzyme. Azidothymidine (AZT) remains one of the most effective and least toxic drugs for AIDS, albeit now usually taken in triple therapy.

Nucleotide analogs have been used to inhibit the *de novo* pathways for the synthesis of the precursor nucleosides and nucleotides, leading to depletion of metabolites and imbalance of dNTPs, and hence to mis-incorporation of nucleotides in RNA or DNA, respectively. Malaria and other parasites rely exclusively on *de novo* pyrimidine biosynthesis, thus they may be susceptible at drug doses that do not affect the host, because the human body can obtain nucleotides from the salvage pathway. Similarly, because of the increased requirement for nucleotides in rapidly proliferating cells, almost all the enzyme reactions (**Figure 2**) have been investigated as potential targets for treatment of cancer, inflammation, or to prevent rejection of transplanted organs. Once again, combinations of drugs with different modes of action have often proved most effective.

Oral uridine, as described earlier, can be used where *de novo* biosynthesis of pyrimidines is defective, and it may be useful in reversing some effects of mitochondrial dysfunction, and to minimize the toxic effect of the antitumor drug 5-fluorouracil. Oral administration of compounds such as benzylacyclouridine, or 2′3′5′ tri-O-acetyl uridine (PN401), inhibits the degradative processes in the liver and delivers more uridine into the circulation than oral uridine alone. Uridine is also a precursor for UDP-glucose, essential for the deposition of glycogen in the liver, and UDP-glucose has been proposed as a dietary supplement.

Oral CDP-choline is rapidly converted to its components, CDP (which can be recycled to uridine) and choline, an essential component of lipid membranes. Each molecule can then cross the blood–brain barrier where CDP-choline is used in regeneration of membranes within and around nerve cells, and its pharmacological effects may extend to protection against dementia, memory loss, visual degeneration, and to recovery from ischemic strokes.

Toxicity of Exogenous Nucleic Acids to Humans

The potential toxicity of dietary nucleic acids to humans usually arises from their metabolic end products (principally uric acid). Many investigators have shown that when normal subjects are fed RNA, the increase in the urate excretion is dramatic, but there is only a modest rise in plasma urate concentrations.

The body pool of urate, and hence the plasma urate concentration, is the result of a balance between production, ingestion, and excretion. If kidney function is normal, the chief causes of high plasma uric acid concentrations are either a high intake of exogenous nucleic acid in the diet or overproduction of endogenous purine. A low-purine (low-nucleic acid) diet containing less meat, seafood, and other purine-rich foods and beverages (**Tables 1** and **2**) leads to a lower risk of gout symptoms. In contrast, subjects with genetic defects that remove the usual controls on purine biosynthesis may have overwhelmingly high endogenous levels of the waste product, uric acid.

The contribution of the two sources can be assessed by placing the subject on a purine-free diet for a week, and measuring the total urinary uric acid. In this way fewer than 5% of patients with gout are found to excrete abnormally large amounts of urate (>3 millimoles per day) derived from endogenous purines. In these cases, overproduction of purine nucleotides leading to excess uric acid can be traced to a genetic defect. Two such sex-linked disorders are hypoxanthine-guanine phosphoribosyltransferase (HPRT) deficiency and phosphoribosyltransferase superactivity (PRPS). Boys presenting in infancy usually have severe and eventually fatal neurological deficits. In those presenting as adolescents, neurological problems are milder or absent, and only gout may be evident.

Urolithiasis and Other Kidney Stones

Although modest overindulgence in purine-rich food does not precipitate gout in normal subjects, it can predispose to uric acid lithiasis. Uric acid stones are relatively common in countries where the consumption of nucleic acid-rich beverages and food is high, and in hot climates if insufficient fluids are consumed.

A number of compounds, such as vitamin C, increase uric acid clearance and thus can precipitate urolithiasis. Perhaps not so well recognized is the uricosuric effect of a high-protein diet and the fact that purine-rich foods also predispose to renal calcium stones. This may be because many purine-rich foods such as spinach are equally rich in calcium oxalate. Some vegetables may provoke gout attacks by virtue of their oxalic acid content rather than of purines, but legumes, fast-growing parts of brassicas and asparagus tips may also have significant nucleic acid content. Pulses and grains have a particularly high nucleic acid content. Approximately 25% of vitamin C intake is also excreted as oxalate, which can compound the problem.

The solubility of uric acid is very sensitive to the pH of the urine, which in turn may be made more acidic by a high-protein diet. The solubility of uric acid in urine at pH 5.0 is low (approximately 1 mmol l^{-1}), but it can be increased 12-fold by alkalinizing regimens such as sodium bicarbonate or potassium citrate, which raise the pH to 8.0.

Excess uric acid from dietary purines can also precipitate symptoms that may draw attention to endogenous uric acid accumulating in adults with milder forms of HPRT deficiency or PRPS superactivity, or to a defect leading to raised levels of a uric acid analog related to adenine. The ideal diet for subjects at risk of gout or of uric acid lithiasis is no more than one meat meal per day, using only the low-purine meat and vegetables indicated, and treatment with allopurinol.

The most common and effective treatment for gout is the drug allopurinol, which prevents conversion of xanthine to uric acid by inhibiting the enzyme xanthine oxidase. Although the uricase gene appears to be present in human cells, the promoter is not activated, so no enzyme activity is detected in the liver. Biochemical drugs using recombinant uricase are effective in refractory gout.

Further Reading

Becker, MA, 2001. Purines and pyrimidines. In: Scriver, CR, Beaudet, AL, Sly, WS, Valle, D (Eds.), The Metabolic and Molecular Basis of Inherited Disease, 8th edn. McGraw-Hill, New York, pp. 2513–2537. Chapter 106.

Carver, JD, 2003. Advances in nutritional modifications of infant formulas. American Journal of Clinical Nutrition 77, 1550S–1554S.

Christopherson, RI, Lyons, SD, Wilson, PK, 2002. Inhibitors of de novo nucleotide biosynthesis as drugs. Accounts of Chemical Research 35, 961–971.

Diem, K, Lentner, C (Eds.), 1970. Scientific Tables - Chemical Composition of Foodstuffs, 7th edn. Basle, Geigy. pp. 230–243.

Fuke, S, Konosu, S, 1991. Taste-active components in some foods: A review of Japanese research. Physiology and Behaviour 49, 863–868.

Grahame, R, Simmonds, HA, Carrey, EA, 2003. Gout: The 'At Your Fingertips' Guide. Class Publishing, London.

Gutierrez-Castrellon, P, Mora-Magana, I, Diaz-Garcia, L, Jimenez-Gutierrez, C, Ramirez-Mayans, J, Solomon-Santibanez, A, 2007. Immune response to nucleotide-supplemented infant formulae: Systematic review and meta-analysis. British Journal of Nutrition 98, S64–S67.

Lee, H, Hanes, J, Johnson, KA, 2003. Toxicity of nucleoside analogs used to treat AIDS and the selectivity of the mitochondrial DNA polymerase. Biochemistry 42, 14711–14719.

Richette, P, Brière, C, Hoenen-Clavert, V, Loeuille, D, Bardin, T, 2007. Rasburicase for tophaceous gout not treatable with allopurinol: An exploratory study. The Journal of Rheumatology 34, 2093–2098.

Rolls, ET, 2000. The representation of umami taste in the taste cortex. Journal of Nutrition 130, 960S–965S.

Secades, JJ, Frontera, G, 1995. CDP-choline: Pharmacological and clinical review. Methods and Findings in Experimental and Clinical Pharmacology 17 (supplement B), 1–54.

Tressler, RL, Ramstack, MB, White, NR, et al., 2003. Determination of total potentially available nucleosides in human milk from Asian women. Nutrition 19, 16–20.

Yu, VYH, 2002. Scientific rationale and benefits of nucleotide supplementation of infant formula. Journal of Paediatrics and Child Health 38, 543–559.

Pantothenic acid

Christopher J. Bates[a,†] **and Luis Fontana**[b,c,d,*], [a] MRC Human Nutrition Research, Cambridge, United Kingdom; [b] Department of Biochemistry and Molecular Biology 2, School of Pharmacy, Granada, Spain; [c] Institute of Nutrition and Food Technology "José Mataix", Biomedical Research Center, Parque Tecnológico Ciencias de la Salud, Granada, Spain; and [d] Instituto de Investigación Biosanitaria ibs.GRANADA, Complejo Hospitalario Universitario de Granada, Granada, Spain

© 2023 Elsevier Ltd. All rights reserved.

This is an update of C.J. Bates, Pantothenic Acid, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 1–5, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00222-1.

Introduction 340
Absorption, transport, storage and status measurement 340
Metabolism and turnover 341
Metabolic function and essentiality 343
Requirements 344
Dietary sources and high intakes 344
Conclusions 345
References 345

Key points

- To obtain an overview of the role of pantothenic acid in intermediary metabolism.
- To understand the relationship between the chemical structure of this vitamin and its biological function.
- To know the processes of absorption, transport, storage and metabolism of this vitamin.
- To have information about the dietary intakes of this vitamin.
- To know the metabolic disorders related to pantothenic acid.
- To know the most important dietary sources and the requirements of this vitamin.

Introduction

Vitamins constitute a group of chemically heterogeneous substances whose functions are also very diverse. Many of them share, however, a common mechanism of action: their participation as coenzymes in the metabolism of macronutrients (Combs and McClung, 2016). Within this group of vitamins, two subgroups can be considered. The first of these consists of thiamine, riboflavin, niacin, pantothenic acid, pyridoxine and biotin. These vitamins exert their physiological functions as coenzymes, which act in a very general way in the metabolism. Vitamin B_{12} and folic acid belong to the other subgroup. These vitamins are also characterized by their metabolic action as coenzymes, but in this case their coenzyme functions are especially and directly involved in cell proliferation (Stipanuk and Caudill, 2018).

The vitamins of the subgroup with general coenzyme functions belong to the so-called "B complex", are all water-soluble, are widely distributed in food, are not particularly stored in the organism and do not usually produce toxicity due to overdosage (Rodwell et al., 2018). Deficiencies in some of these vitamins have had historical relevance and are still important in some underdeveloped countries. In industrialized countries, however, these vitamin deficiencies have less clinical repercussions (Ross et al., 2014).

This article will consider the metabolic functions, the main food sources, the requirements, the processes of absorption, transport, storage and metabolism, and the assessment of the nutritional status of pantothenic acid.

Absorption, transport, storage and status measurement

Much of the pantothenic acid (formerly vitamin B_5; discovered by Williams, Elvehjem, and Jukes in the 1930s, see **Fig. 1**) that is present in food eaten by animals or humans exists as derivatives such as coenzyme A (CoA) and acyl carrier protein (ACP) (Kelly, 2011; Sanvictores and Chauhan, 2021). It is released as free pantothenic acid or pantetheine by pancreatic enzymes, and absorbed along the entire length of the small intestine by a combination of active transport and passive diffusion, of which the active

*Corresponding author
†Deceased

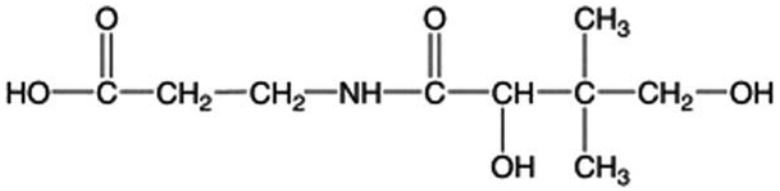


Fig. 1 Structure of pantothenic acid.

transport process predominates at physiological intakes. This active transport process is dependent on sodium, energy, and pH, and is saturable: the Km is approximately 17 μM and V_{max} is approximately 1000 pmol cm^{-2} h^{-1}, with minor variations among species. The transport pathway is shared by biotin in colonic epithelial cells and is regulated by an intracellular protein kinase C-mediated pathway. Calmodulin is also implicated in cellular pantothenic acid transport pathways.

In mice, there is no evidence for adaptation of absorption to low or high intakes. However, studies on rats suggest that a limited secretion of enzymes that degrade CoA in the gut lumen may limit the availability of pantothenic acid from dietary CoA.

In humans, studies of urinary excretion of pantothenic acid after oral doses of either free pantothenic acid or the pantothenic acid in food have indicated an availability of approximately 50% from the food-derived vitamin. Urinary excretion of pantothenate was approximately 0.8 mg day^{-1} when a pantothenate-deficient diet was eaten, rising to 40–60 mg day^{-1} at an intake of l00 mg day^{-1}. At intermediate intakes, in the range 2.8–12.8 mg day^{-1}, the urinary excretion rate varied between 4 and 6 mg day^{-1}. Excretion levels of less than 1 mg day^{-1} are considered low. Urinary excretion rates reflect recent intakes more closely than most other biochemical indices.

The contribution of the gut microbiota to absorbed pantothenate in humans is unknown, but there is evidence that bacterial synthesis of the vitamin may be important in animals, especially ruminants, because severe deficiency can only be achieved here by using antibiotics or antagonists, and recent evidence suggests that absorption of pantothenate synthesized by the flora of the human large intestine may make an important contribution to the pantothenate supply of this tissue. Clinical conditions such as ulcers or colitis can adversely affect pantothenate status and excretion rates, and dietary fiber may affect its absorption.

After a dose of ^{14}C-labeled pantothenate, approximately 40% of the dose appears in muscle tissue and approximately 10% in the liver, with smaller amounts elsewhere. The differential affinities of tissues determine their individual contents of the coenzyme derivatives, CoA and ACP, because there is no major surplus store of the vitamin anywhere in the body. Most organs, including placenta, exhibit evidence of a unidirectional active transport process for the intracellular accumulation of pantothenate, which again is dependent on sodium, energy, and pH. In placenta (and probably elsewhere), this transport process is also shared, and competed for, by biotin and some of its analogs.

The only tissues that have been shown to differ with respect to the transport mechanisms are red cells and the central nervous system. The uptake and efflux of pantothenate in red blood cells are unaffected by sodium, energy, or pH. They contain pantothenate, 4-phosphopantothenate, and pantetheine, but do not contain mitochondria, or carry out CoA-dependent processes. The functions of the pantothenate derivatives in red cells are unknown, but their formation results in higher concentrations of total pantothenate in red cells than in plasma, and red cell (or whole blood) total pantothenate is considered a better status index, and more predictably related to intake, than serum or plasma pantothenate. A concentration less than 1 μmol L^{-1} of pantothenate in whole blood is considered low; the normal range being 1.6–2.7 μmol L^{-1}. Pantothenate in serum appears to be a very short-term marker and it is not well correlated with changes in intake or status.

Concentrations in body fluids are traditionally measured by microbiological assay using *Lactobacillus plantarum*. If CoA is present, enzymatic hydrolysis is needed to liberate free pantothenic acid for the microbiological assay. Other assay methods include gas chromatography (after conversion to a volatile derivative), radioimmunoassay (RIA), or enzyme-linked immunoabsorbent assay (ELISA).

Unlike several other B-vitamin precursors of enzyme cofactors, pantothenate is not entirely converted to coenzyme forms inside the cell, and metabolic "trapping" is therefore less dominant than it is for some other B vitamins. Free pantothenate in tissues seems more closely related to dietary pantothenate than the coenzyme forms are; the latter are protected during periods of dietary deficiency. Uptake of pantothenate from plasma into most tissues is proportional to the plasma concentration because the active transport process is not saturated at typical plasma concentrations of approximately 1 μM (or 1.6–2.7 μM in whole blood).

Pantothenate in acetyl CoA is required for the hepatic acetylation of drugs, and pantothenate deficiency can impair it; moreover, 20–60% of human populations are slow acetylators, varying with ethnicity.

Metabolism and turnover

The primary role of pantothenic acid is in acyl group activation for lipid metabolism, involving thiol acylation of CoA or of ACP, both of which contain 4-phosphopantotheine, the active group of which is β-mercaptoethylamine. CoA is essential for oxidation of fatty acids, pyruvate and α-ketoglutaric, for metabolism of sterols, and for acetylation of other molecules, so as to modulate their

transport. Acyl carrier protein (ACP), which is synthesized from apo-ACP and coenzyme A, is involved specifically in fatty acid synthesis. Its role is to activate acetyl, malonyl, and intermediate-chain fatty acyl groups during their anabolism by the biotin-dependent fatty acid synthase complex (i.e., acyl-CoA: malonyl-CoA-acyl transferase (decarboxylating, oxoacyl- and enoyl-reducing, and thioester-hydrolyzing), EC 2.3.1.85) (Chan and Vogel, 2010).

The liver has the highest concentration of pantothenate, followed by adrenal cortex, because of the requirement for steroid hormone metabolism in these tissues. Ninety-five percent of the CoA within each tissue is found in the mitochondria. However, the initial stages of activation of pantothenate and conversion to CoA occur in the cytosol, and CoA is then transported across the mitochondrial membrane. β-Oxidation within the peroxisomes is also CoA dependent and is reduced by pantothenate deficiency.

The pathways of conversion of pantothenic acid to CoA and to ACP are summarized in Fig. 2. There are three ATP-requiring reactions and one CTP-requiring reaction in the synthesis of CoA. The rate of CoA synthesis is under close metabolic control by energy-yielding substrates, such as glucose and free fatty acids (via CoA and acyl CoA), acting at the initial activation step, which is catalyzed by pantothenate kinase (ATP: pantothenate 4-phosphotransferase, EC 2.7.1.33). There are also hormonal effects of insulin, corticosteroids, and glucagon, which result in important changes in tissue distribution, uptake, etc. (e.g., in diabetics). The mechanisms are complex and not yet fully understood; however, insulin represses and glucagon induces this enzyme. Although

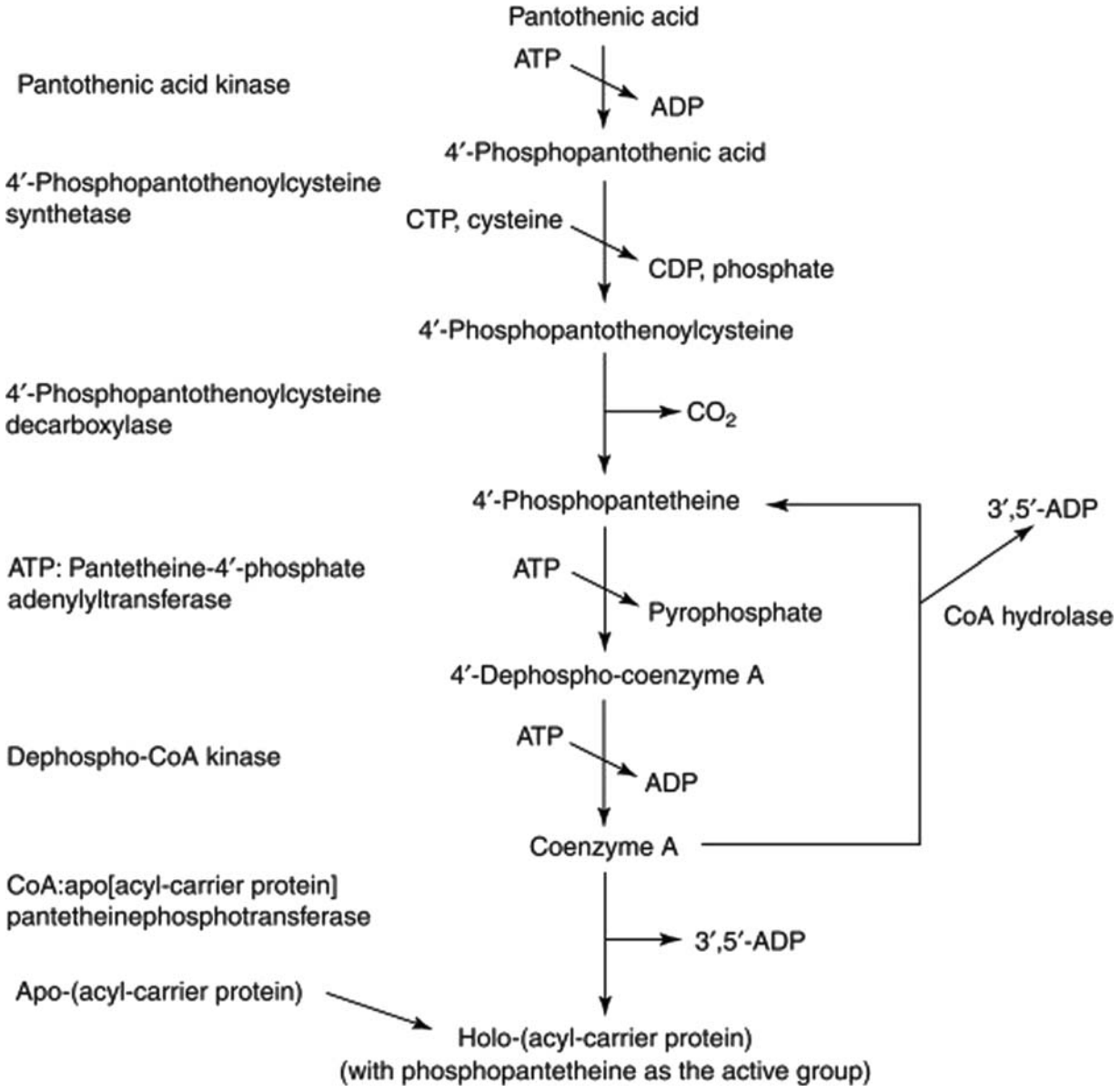


Fig. 2 Synthetic pathway between pantothenic acid, coenzyme A, and acyl carrier protein.

prokaryotes and eukaryotes carry out the same biochemical reactions in the synthesis of CoA, genome studies have revealed important protein-sequence differences, and there is new interest in this pathway for the design of antibacterial drugs.

The discovery of a human neurodegenerative disorder also stimulated research in this area. This rare genetic disease, formerly called Hallervorden–Spatz syndrome, results from mutations and hence a functional deficiency of pantothenate kinase, and is now known as pantothenate kinase-associated neurodegeneration (PKAN). Dystonia, involuntary movements, spasticity, optic atrophy, and iron deposits in the basal ganglia and globus pallidus occur, and although there is no cure, some palliative replacement is possible (Di Meo et al., 2019).

In humans, fasting results in a reduction of fatty acid synthase activity with loss of the coenzyme of ACP, which thus results in a shift away from fatty acid synthesis and toward breakdown. This interconversion of apo-ACP and holo-ACP is thus a very important process for the short-term regulation of fatty acid synthesis.

A deficiency of sulfur amino acids or copper overload (by interfering with sulfur amino acid function) can result in reduced CoA synthesis.

Free pantothenate in urine is the primary excretion route in humans; however, in other mammals its glucuronide or glucoside may be excreted, and pantothenate is very efficiently conserved in animals. Some bacteria can cleave it to pantoic acid and β-alanine. One breakdown product of CoA is taurine, via cysteamine, and is an essential nutrient for some carnivorous animals such as cats.

When dietary intakes are low, much of the circulating pantothenate, after filtration in the kidney tubules, is absorbed by the same sodium-dependent active transport process that occurs at most other sites in the body. Secretion into breast milk is proportional to intake and to blood levels of the vitamin; therefore, dietary supplements taken by the lactating mother generally increase the breast milk content of the vitamin.

Metabolic function and essentiality

The biochemical functions, and hence the basis for the dietary requirement of pantothenic acid, arise entirely from its occurrence as an essential component of CoA and of ACP, because the vitamin cannot be synthesized de novo in mammals. Pantothenic acid was isolated as a chick, bacterial, and yeast growth factor, and as an antidermatitis dietary factor, in the 1930s. Use of omega-methyl-pantothenate as a vitamin antagonist in humans resulted in a syndrome of numbness and burning sensation in the extremities, increased sensitivity to insulin, impaired antibody production, and other symptoms.

In addition to the well-established roles of CoA in the degradation and synthesis of fatty acids, sterols, and other compounds synthesized from isoprenoid precursors, there are also a number of acetylation and long-chain fatty acylation processes which require CoA (Pietrocola et al., 2015). The acetylation of amino sugars and many other functions of acetyl-CoA and succinyl-CoA have been known since the 1980s. However, the addition of acetyl or fatty acyl groups to certain proteins in order to modify and control their properties is a more recent discovery. Acetylation of N-terminal amino acid residues occurs in at least half of all the known proteins in higher organisms. The recipients of these acetyl groups are most commonly methionine, alanine, glycine, threonine, or serine. The purposes of this are not entirely clear and may include modifications of hormone function, of ligand-binding and site recognition, of tertiary peptide structure, and of susceptibility to degradation. Acetylation can also involve the amino groups of the side chain of internal lysine residues, notably the histones in the cell nucleus, and the α-tubulin proteins of the cytoplasmic microtubules, which help to determine cell shape and motility. Its role in the synthesis of α-tubulin appears particularly important.

Proteins can also be modified by acylation with certain long-chain fatty acids, notably the 16-carbon saturated fatty acid, palmitic acid, and the 14-carbon saturated fatty acid, myristic acid. Although structurally very similar to each other, these two fatty acids seek entirely different protein locations for acylation and have different functions. They have mainly been explored in viral and yeast proteins, but proteins in higher animals, in organs such as lungs and brain, can also become acylated with palmityl moieties, and enable transport of protein through the Golgi apparatus. Protein acylations may control protein interactions, especially in cell membranes, and palmitoylated proteins are associated with plasma membranes. Signal transduction (e.g., of the human β_2-adrenergic receptor) is one process that appears to be controlled by palmitoylation, and other palmitoylated proteins are structurally important, for example, in the protein–lipid complex of brain myelin. Acylation may be involved in the activation of some hormones and transcription factors. Clearly, these subtle protein modifications, all of which depend on CoA and hence on pantothenic acid, have wide-ranging significance.

Pantothenic acid is an essential component of the diet of all mammalian species studied, namely humans, bovines, pigs, dogs, cats, and rodents, as well as poultry and fish. Pantothenate deficiency signs in animals are nonspecific and vary between species and with age. Deficiency in young animals results in anorexia and impaired growth, and the requirement estimates based on maximum growth rates are between 8 and 15 mg kg^{-1} diet. Rats fed a diet low in pantothenate also exhibit scaly dermatitis, alopecia, hair discoloration and loss, porphyrin-caked whiskers, spastic gait, anemia, leukopenia, impaired antibodies, sex organ disruption, congenital malformations, and adrenal necrosis. Deficient chicks have dermatitis, abnormal feather development, thymus involution, myelin degradation, locomotor abnormalities, neurological symptoms including convulsions, fatty liver, and hypoglycemia. Pigs exhibit dermatitis, intestinal problems, spastic gait, and abnormalities of dorsal root ganglion cells, and several species suffer nerve demyelination. Fish show fused gill lamellae, reproductive failure, clumping of mitochondria, and kidney lesions.

Signs specific for pantothenate depletion are not well characterized for humans. A syndrome that included "burning feet" has been described in tropical prisoner-of-war camps during World War II, and it was said to respond to pantothenic acid supplements; however, this was likely to have been a more complex deficiency. A competitive analog of pantothenate, ω-methyl pantothenate, interferes with the activation of pantothenic acid and produces burning feet symptoms, Reye-like syndrome, cardiac instability, gastrointestinal disturbance, dizziness, paresthesia, depression, fatigue, insomnia, muscular weakness, loss of immune (antibody) function, insensitivity to adrenocorticotrophic hormone, and an increased sensitivity to insulin. Calcium hopantenate, another potential antagonist of the vitamin, has produced some similar effects. Large doses of pantothenate can reverse these changes. One of the earliest functional changes in mildly deficient rats was an increase in serum triacylglycerols and free fatty acids, resulting from impairment of β-oxidation. Paradoxically, CoA levels are relatively resistant to dietary pantothenate deficiency.

CoA is required for Golgi function, and hence protein transport; pantothenate deficiency therefore causes a reduction in the secretion of some proteins. Other metabolic responses to deficiency include a reduction in urinary 17-ketosteroids, a reduction in serum cholesterol, a reduction in drug acetylation, a general reduction in immune response, and an increase in upper respiratory tract infection.

Studies of wound healing and fibroblast growth have indicated that both pantothenic acid and ascorbic acid are involved in trace element distribution in the skin and scars of experimental animals and that pantothenic acid can improve skin and colon wound healing in rabbits. It is not yet known whether these observations are relevant to wound healing in humans. Reports that high-dose pantothenic acid supplements can alleviate some of the symptoms of rheumatoid arthritis or lupus erythematosus have yet to be confirmed.

Requirements

The US/Canada adequate intake (AI) for pantothenic acid was set at 5 mg day^{-1} for adults, rising to 6 mg in pregnancy and 7 mg in lactation; at 4 mg for children aged 9–13 years; at 3 mg for 4–8 years, at 2 mg for 1–3 years and at 1.7–1.8 mg for 0–1 year. There was insufficient evidence to set an estimated average requirement (EAR), a recommended daily allowance (RDA), or a tolerable upper intake level (UL). As for most water-soluble vitamins, maternal blood levels decrease during pregnancy, and the mean daily output of the vitamin in breast milk is of the order of 1.7 mg. Infant formulas should contain at least 2 mg pantothenate per liter (Institute of Medicine, 2006).

In a technical report issued in 2017, the European Food Safety Authority (EFSA) updated the AIs of all nutrients (EFSA 2017). The pantothenic acid AI for adults is set at 5 mg day^{-1}. The AI for adults also applies to pregnant women. For lactating women, an AI of 7 mg day^{-1} is proposed, to compensate for pantothenic acid losses through breast milk. For infants over six months, an AI of 3 mg day^{-1} is proposed by extrapolating from the pantothenic acid intake of exclusively breast-fed infants aged zero to six months, using allometric scaling (body weight to the power of 0.75) and reference body weight for each age group, in order to account for the role of pantothenic acid in energy metabolism, and rounding to the nearest unit. The AIs for children and adolescents are set at 4 and 5 mg day^{-1}, respectively, based on observed intakes in the EU.

There are few studies in communities where intakes are likely to be low; indeed, pantothenic acid is so widely distributed in human foods that it is unlikely that any natural diets with a very low content will be encountered. Some variations in status among communities have been described, but these do not define requirements. In a group of adolescents in the USA, daily pantothenate intakes were approximately 4 mg; total blood pantothenate was in the "normal" range of approximately 350–400 ng mL^{-1}, and intakes were correlated with red cell pantothenate ($r = 0.38$) and with urinary pantothenate ($r = 0.60$), both $P < 0.001$. In adults, these correlations were less strong.

Dietary sources and high intakes

Pantothenate is widely distributed in food; rich sources include animal tissues, especially liver, and yeast, with moderate amounts occurring in whole grain cereals and legumes (see **Table 1**). It is stable during cooking and storage, although some destruction occurs at high temperatures and at pH values below 5 or above 7. Highly processed foods have lower contents than fresh foods, and one-fifth to two-thirds may be lost during freezing, canning, etc. Commercial vitamin supplements containing pantothenate usually use the calcium salt, which is crystalline and more stable than the free acid.

Synthesis by gut microbiota in humans is suspected, and the rarity of diet-induced deficiency has been attributed partly to the likely contributions from gut microbiota.

Pantothenic acid supplements may be beneficial for treatment of rheumatoid arthritis and for enhancement of athletic performance, specifically while running. Pantethine, the disulfide dimer of pantetheine, may have cholesterol-lowering properties and pantothenol may enhance wound healing in animal models. The mechanisms of these reported effects are unclear and they require verification. A homolog of pantothenate, pantoyl γ-aminobutyrate (hopantenate, see section on metabolic function and essentiality), which can act as a pantothenate antagonist, has been used to enhance cognitive function, especially in Alzheimer disease. It acts on γ-amino-butyric acid (GABA) receptors to enhance acetylcholine release and cholinergic function in the brain. Some of the side effects of valproic acid (used in the treatment of epilepsy and bipolar disorder) may be susceptible to alleviation by pantothenate and carnitine supplements.

There is little evidence for pantothenate toxicity at high intakes, even up to 100–500 times the normal intake; at daily intakes of approximately 10 g, there may be mild diarrhea and gastrointestinal disturbance.

Table 1 Pantothenate content of selected foods.

Food	*mg per 100 g wet wt*
Meat, offal, and fish	
Stewed minced beef	0.36
Grilled pork chop	1.08
Calf liver, fried	4.1
Lamb's kidney, fried	4.6
Cod, grilled	0.33
Dairy products	
Cow's milk, full cream	0.58
Cheese, cheddar	0.50
Yogurt (whole milk, plain)	0.50
Boiled chicken's egg	1.25
Human milk	0.25
Fruits	
Apples, eating, flesh and skin	Trace
Oranges, flesh	0.27
Pears, flesh and skin	0.08
Strawberries, raw	0.37
Dried mixed fruit	0.09
Vegetables	
Potatoes, boiled, new	0.50
Carrots, boiled, young	0.18
Brussels sprouts, boiled	0.40
Cauliflower, boiled	0.47
Onions, fried	0.04
Grains, grain products, nuts	
Wholemeal bread	0.63
Rice, boiled, white	0.20
Baked beans in tomato sauce	0.11
Peanuts, plain	2.66

Compiled from data in the McCance and Widdowson's composition of foods integrated dataset, 2021. Public Health England (https://www.gov.uk/government/publications/composition-of-foods-integrated-dataset-cofid).

Conclusions

Pantothenic acid has two coenzyme derivatives of great biological importance. The acyl carrier protein (ACP) is part of the enzyme complex used for fatty acid synthesis and is therefore essential for lipogenesis. Coenzyme A is needed for metabolic activation of all acyl residues, including both fatty acids and acid metabolites originating from the catabolism of carbohydrates and some amino acids. Pantothenic acid is therefore essential in the energetic utilization of all types of macronutrients. Given its abundance in foods, isolated pantothenic acid deficiencies are very rare.

References

Chan, D.I., Vogel, H.J., 2010. Current understanding of fatty acid biosynthesis and the acyl carrier protein. Biochem. J. 430, 1–19.

Combs Jr., G.F., McClung, J.P., 2016. The Vitamins. Fundamental Aspects in Nutrition and Health, fifth ed. Academic Press, London.

Di Meo, I., Carecchio, M., Tiranti, V., 2019. Inborn errors of coenzyme A metabolism and neurodegeneration. J. Inherit. Metab. Dis. 42, 49–56. https://doi.org/10.1002/jimd.12026.

European Food Safety Authority (EFSA), 2017. Dietary Reference Values for Nutrients. Summary Report. https://doi.org/10.2903/sp.efsa.2017.e15121.

Institute of Medicine, 2006. Dietary Reference Intakes: The Essential Guide to Nutrient Requirements. The National Academies Press, Washington, DC. https://doi.org/10.17226/11537.

Kelly, G.S., 2011. Pantothenic acid. Monograph. Altern. Med. Rev. 16, 263–274.

Pietrocola, F., Galluzzi, L., Bravo-San Pedro, J.M., Madeo, F., Kroemer, G., 2015. Acetyl coenzyme A: a central metabolite and second messenger. Cell Metab. 21, 805–821.

Rodwell, V.W., Bender, D.A., Botham, K.M., Kennelly, P.J., Weil, P.A., 2018. Harper's Illustrated Biochemistry, 31st ed. McGraw-Hill Education, New York.

Ross, A.C., Caballero, B., Cousins, R.J., Tucker, K.L., Ziegler, T.R., 2014. Modern Nutrition in Health and Disease, eleventh ed. Wolters Kluwer Health, Lippincott Williams & Wilkins, Philadelphia.

Sanvictores, T., Chauhan, S., 2021. Vitamin B_5 (Pantothenic Acid). StatPearls. StatPearls Publishing, Treasure Island (FL).

Stipanuk, M.H., Caudill, M.A., 2018. Biochemical, Physiological, and Molecular Aspects of Human Nutrition, fourth ed. Saunders, Philadelphia.

Phosphorus

Olga P. García, School of Natural Sciences, Universidad Autónoma de Querétaro, Querétaro, Mexico

© 2023 Elsevier Ltd. All rights reserved.

This is an update of J.J.B. Anderson, Phosphorus: Physiology, Dietary Sources, and Requirements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 28–32, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00225-7.

Introduction 346
Phosphorus requirements and food sources 347
Absorption 348
Phosphorus homeostasis 348
PTH 348
Calcitriol $1,25(OH)_2D_3$ 349
FGF-23 349
Hyperphosphatemia and its consequences 349
Hypophosphatemia and its consequences 350
Conclusion 350
References 350

Key points

- Phosphorus main functions include bone formation, regulation of gene transcription, enzymes activation, energy production and storage (ATP), blood buffer in extracellular fluids, and cellular signal transduction.
- The main dietary sources of phosphorus are high protein foods, dietary supplements, fortified food/water, and phosphorus added during food processing.
- Absorption has two main pathways: a sodium-dependent transcellular pathway, and the predominant sodium-independent paracellular pathway.
- Phosphorus is tightly regulated and homeostasis is maintained by the intestine, the kidney and the bone, and the activity of three hormones: parathyroid hormone (PTH), calcitriol and fibroblast growth factor 23 (FGF-23).
- Hypophosphatemia may be caused by low dietary intake, increased renal excretion, decreased intestinal absorption and a shift from extracellular to intracellular phosphorus causing bone loss.
- Hyperphosphatemia may be caused by decreased phosphorus renal excretion, excessive phosphorus intake and a shift from intracellular to extracellular phosphorus.
- The most common cause of hyperphosphatemia is acute or chronic renal disease. The high concentration of serum phosphorus in chronic kidney disease alters the homeostasis of phosphorus, causing secondary hyperparathyroidism, vascular calcification, and bone resorption.
- Elevated concentrations of phosphorus have been associated with impaired cognitive function, dementia and Alzheimer.
- Phosphorus, its transporters and polyPi may play an important role in cardiovascular, bone and brain health that should be further explored.

Introduction

Phosphorus is an essential nutrient that is available naturally and added in many foods. It is the second most abundant mineral in the human body and approximately 80% is stored in the bone (Calvo and Lamberg-Allardt, 2015, Takeda et al., 2012). It is also present in the human body in the form of phospholipids (cell membrane and ATP), is a component of DNA and RNA, and other proteins and sugars that are phosphorylated (Takeda et al., 2012; Heaney, 2012). Its main functions include bone formation, regulation of gene transcription, enzymes activation, energy production and storage (ATP), blood buffer in extracellular fluids, and cellular signal transduction (Heaney, 2012; Calvo and Lamberg-Allardt, 2015; Takeda et al., 2012). Inorganic phosphate is the main form of serum phosphorus, representing less than 1% of the total inorganic phosphate in the body. It is highly regulated through intestinal absorption, renal excretion and bone formation to maintain a serum concentration between 2.5 and 4.5 mg/dL (Calvo and Lamberg-Allardt, 2015; Takeda et al., 2012).

The consumption of adequate dietary phosphorus is critical for the support of human metabolic functions. Under normal conditions, dietary phosphorus is sufficient to maintain its multiple functions. However, too much phosphorus may contribute to bone loss (with low calcium intake) and an increased risk of cardiovascular disease (CVD), coronary calcification and renal disease

(Fulgoni and Fulgoni, 2021). On the other hand, too little phosphorus with low calcium intake may not adequately maintain bone mass, particularly in the elderly.

This chapter looks into the main food sources of phosphorus, its absorption, homeostasis and consequences of both hypo- and hyperphosphatemia.

Phosphorus requirements and food sources

The Recommended Daily Allowance (RDA) of phosphorus for men and women 19 years and older is 700 mg, and the Estimated Average Requirement (EAR) is 580 mg. The Upper Tolerable Level (UL) is 4000 mg/d for those 19–70 years old and 3000 mg/d for those over 71 years (IOM, 1998). Recently, the European Food Safety Authority (EFSA) established an Acceptable Daily Intake (ADI) of 40 mg of phosphorus/kg body weight/d (Younes et al., 2019). **Table 1** gives representative values of phosphorus in selected foods.

The three main sources of dietary phosphorus are: (1) organic phosphorus contained in high protein foods, (2) inorganic phosphorus from medications, dietary supplements, and fortified food/water, and (3) inorganic phosphorus added during food processing. Approximately 40–60% of the dietary phosphorous is naturally occurring organic phosphorus that comes from animal food sources (milk, milk products, meat, poultry, fish) and legumes and grain products (Calvo and Urivarri, 2013). According to data from the NHANES 2015–2016, the average daily phosphorus intake in the United States among children and teens (2–19 years) is 1237 mg (USDA, 2019). The mean intake of organic phosphorus of adults (>19 years old) in 2015–2016 was 1113 ± 10 mg/d, and the main sources were cheese, pizza, chicken, reduced-fat milk and eggs (Fulgoni and Fulgoni, 2021). The bioavailability of phosphorus from animal food sources is higher than that from grains and legumes. The main form of storing phosphorus in the plants is phytate, and is present in the outer layer of grains and in the seed coats of legumes (Calvo and Urivarri, 2013). The bioavailability of phosphorus from phytate is low, thus, the amount of phosphorus from foods that is absorbed will depend on the total content of phosphorus in the diet and its bioavailability.

Most of the inorganic phosphorus that is added during the processing of many different foods are phosphate salts that have specific functions. These phosphorus-containing food additives are widely used for emulsification, leavening, acidification, and also as acidulants, flavorants, color stablilizers and buffers (Fulgoni and Fulgoni, 2021; Calvo and Urivarri, 2013). They also are used to improve texture and taste, and to reduce cooking time. Thus, the content of phosphorus additives is higher in processed

Table 1 Calcium and phosphorus composition of selected foods.

Food	*Phosphorus mg/serving*
Milk, dairy products and eggs	
Cheese, cheddar, 100 g	458
Cheese, cottage, low fat, 2% milkfat, 1 cup	326
Milk 2% milkfat, 1 cup	252
Yogurt, Greek, plain, nonfat, 150 g	212
Cheese, mozzarella, part skim, 40 g	197
Egg, whole, cooked, hard boiled, 1 large	86
Animal food sources	
Chicken breast, skinless, meat only, 1 piece	419
Salmon, cooked, 85 g	266
Beef, loin, lean, boneless, cooked, 100 g	259
Tilapia, cooked, 1 fillet (87 g)	177
Tuna, light, canned in water, 1 can (107 g)	147
Cereals and legumes	
Lentils, cooked without salt, 1 cup	356
Beans, black, boiled, 1 cup	241
Bread, whole wheat, 1 slice	212
Pasta, whole wheat spaghetti, 1 cup	166
Tortillas, corn, 1 piece	75
Rice, cooked, 1 cup	68
Fruits and vegetables	
Peas, cooked, 1 cup	187
Potatoes, baked, flesh and skin, 1 medium	121
Asparagus, four spears (60 g)	32
Bananas, 1 medium	26
Apple, with skin, 1 medium	17
Watermelon, 1 cup, diced	17

Source: U.S. Department of Agriculture (2019).

foods. It has been reported that these additives contribute approximately 70 g/100 g of extra phosphorus compared with foods that have no phosphorus additives (Fulgoni and Fulgoni, 2021). However, intake of this added inorganic phosphorus is not easy to quantify and it is well-recognized that phosphorus intake may be underestimated. Recently, a novel approach has been used to calculate added phosphorus content in foods using data from the NHANES 2015–2016 (Fulgoni and Fulgoni, 2021). The added phosphorus intake was 155 ± 4.1 mg/d in adults (>19-year-old), representing approximately 12% of the total phosphorus intake. The main sources of added phosphorus were cheese, soft drinks, cakes/pies, rolls/buns, and cookies/brownies (Fulgoni and Fulgoni, 2021). With this approach, it may be easier to estimate total phosphorus intake and associate it with health outcomes.

Absorption

Phosphorus absorption in adults and children is very efficient (65–75%) and an increase in serum phosphorus can be observed within an hour after ingestion of a meal begins.

Absorption of dietary phosphate occurs in the gastrointestinal tract through at least two pathways: the transcellular, sodium-dependent pathway, and the paracellular, sodium-independent pathway. The main apical intestinal transporter in the transcellular absorption of phosphorus is the sodium-dependent phosphate cotransporter 2b (NaPi-2b), which is believed to be responsible for absorption when dietary phosphorus is low (Saurette and Alexander, 2019; Fishban and Nigwekar, 2021). The transporter NaPi-2b is highly regulated by the expression of calcitriol (1,25 $(OH)_2D_3$) and by dietary phosphorus levels. For instance, when phosphorus concentrations are low, calcitriol levels increase, and then, calcitriol increases the expression of NaPi-2b to promote the intestinal absorption of phosphorus (Saurette and Alexander, 2019). On the contrary, when phosphorus concentration is high, the hormone phosphatin fibroblast growth factor 23 (FGF-23) is secreted in the bone, inhibits the synthesis of calcitriol, and thus, phosphorus absorption decreases. This pathway saturates at very low luminal phosphorus concentrations, so it is not very relevant in populations with high phosphorus intake.

The paracellular, sodium-independent pathway, is a passive transport of phosphorus through tight junction complexes between the intestinal cells, using electrical and chemical concentration gradients. These complexes include transmembrane adhesion proteins such as occludins and claudins (Fishban and Nigwekar, 2021; Saurette and Alexander, 2019). The paracellular absorption does not saturate, thus, when luminal phosphorus concentrations are high, absorption increases.

For years, the transcellular system has been thought to be the main pathway for phosphorus absorption, but now it is recognized that the paracellular pathway predominates and is the primary system of phosphate absorption (Saurette and Alexander, 2019).

Gut microbiota contributes with phosphorus absorption by degrading phytate and liberating inorganic phosphorus. In addition, it has been suggested that the microbiota provides short and long chain polyphosphated (polyPi) that are also absorbed and provide another source of phosphorus (Bird and Eskin, 2021). It has been suggested that polyPi, short, medium and long chains, can also be synthesized endogenously, and exert diverse physiological functions, both positive and negative.

Once absorbed, phosphorus transporters are responsible for carrying phosphorus to different tissues and through different cells. These transporters belong to three families of solute carrier proteins (SLC): SLC17, SLC20 and SLC34. The intestinal transporter NaPi-2b, for instance, belongs to the SLC34 family, and is present also in the thyroid gland, lung, mammary glands, among others. Recent research has explored the importance of the SLC10 family, because they are expressed in the brain, liver and heart (Bird and Eskin, 2021). All these transporters contribute to phosphorus homeostasis.

Phosphorus homeostasis

The blood concentration of phosphorus is tightly regulated, but less than other micronutrients, such as calcium. The kidney was believed to be the organ responsible for phosphorus homeostasis. However, phosphorus homeostasis in adults is maintained through urine losses and total phosphorus absorbed, and the amount of phosphorus that is resorbed and deposited in the bone. Therefore, it has been recognized recently that in order to maintain phosphorus homeostasis three organs are needed: the intestine, the kidney and the bone (Calvo and Lamberg-Allardt, 2015). This effective homeostatic action of the intestine, the bone and the kidney produces phosphorus balance (**Fig. 1**). These three organs work together with the activity of three hormones: parathyroid hormone (PTH), calcitriol and fibroblast growth factor 23 (FGF-23) (Bergwitz and Jüppner, 2010).

PTH

High serum phosphorus concentration decreases calcium concentration, which in turn stimulates the parathyroid glands to secrete and synthesize PTH. Stimulation of PTH by high phosphorus concentration can occur independently of low calcium concentration. PTH acts on the kidneys to increase urinary phosphate excretion by inhibiting phosphate reabsorption, on the bone by increasing osteoblast activity, and finally, stimulates the synthesis of $1,25(OH)_2D_3$, which together, lower serum phosphate concentration (Bergwitz and Jüppner, 2010). PTH can also increase phosphorus concentration by increasing its intestinal absorption through its effect on the kidney and on the production of $1,25(OH)_2D_3$, and also by stimulating bone turnover and the release of phosphorus (Ureña and De Brauwere, 2011).

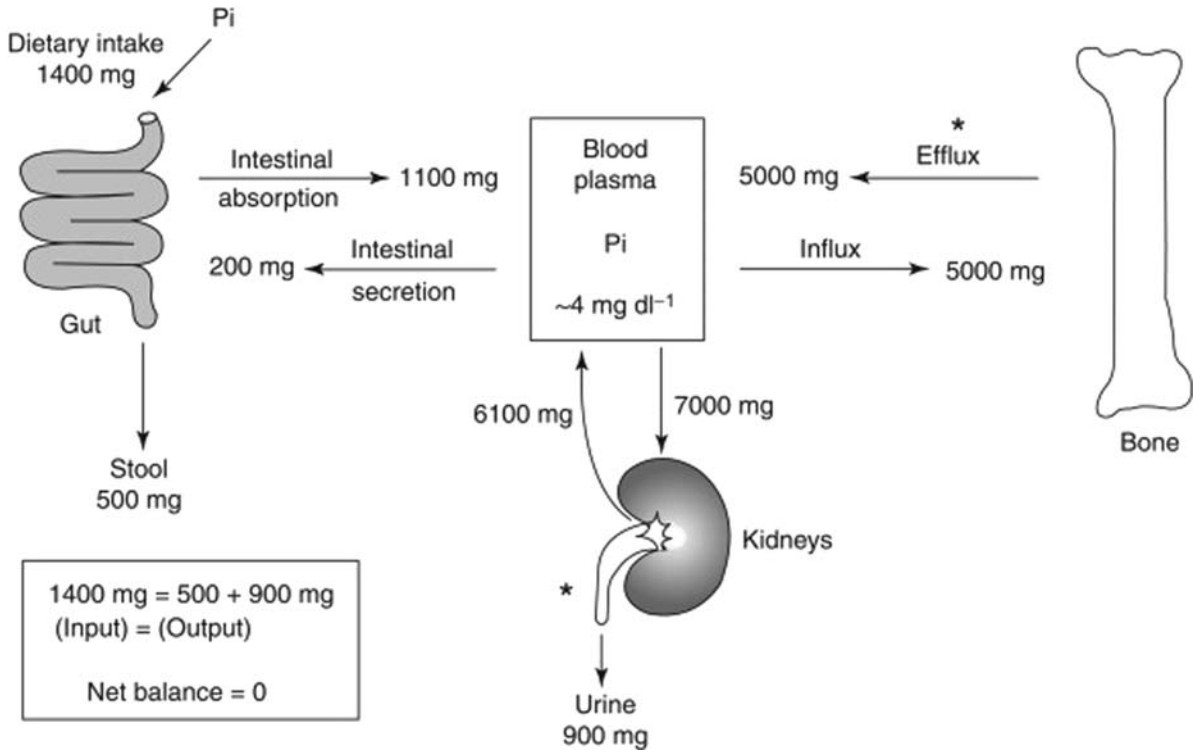


Fig. 1 Phosphorus homeostasis and balance. The intestine, kidneys, and bone are organs involved in phosphate homeostasis. Fluxes of phosphate ions between blood and these organs are shown. Note the high fluxes in and out of bone each day. To convert phosphorus values from g to mmol, multiply by 32.29; from mg dl−1 to mmol l−1, multiply by 0.3229. Steps enhanced by parathyroid hormone. Reproduced with permission from Anderson et al. (2001).

Calcitriol $1,25(OH)_2D_3$

Calcitriol can increase or decrease phosphorus serum concentration through its effect on the bone, the intestine and parathyroid gland. Calcitriol stimulates phosphorus intestinal absorption by increasing the expression of NPT2b, stimulates phosphate renal reabsorption and modulates bone resorption, thus, increasing phosphorus serum levels. On the contrary, $1,25(OH)_2D_3$ may decrease phosphorus concentration by stimulating the synthesis of FBF-23 (Bergwitz and Jüppner, 2010; Ureña and De Brauwere, 2011). High serum phosphate, calcium and FBF-23 levels downregulate the expression of $1,25(OH)_2D_3$, and its expression is upregulated by PTH.

FGF-23

FGF-23 is a hormone secreted in the bone that reduces phosphorus serum concentration by decreasing renal reabsorption and by decreasing $1,25(OH)_2D_3$ synthesis, which decreases intestinal phosphorus absorption (Ureña and De Brauwere, 2011). FGF-23 needs FGF receptors, and Klotho, a transmembrane protein, as a co-receptor in order to lower phosphorus concentrations (Bergwitz and Jüppner, 2010; Lanske and Razzaque, 2014; Ureña and De Brauwere, 2011). Both PTH and $1,25(OH)_2D_3$ can induce FGF-23 expression in the bone, and PTH can also regulate Klotho expression in the kidney (Lanske and Razzaque, 2014).

Hyperphosphatemia and its consequences

Hyperphosphatemia (serum phosphate >4.5 mg/dL) may be caused by three main mechanisms: decreased phosphorus renal excretion, excessive phosphorus intake and a shift from intracellular to extracellular phosphorus. Signs and symptoms of hyperphosphatemia are similar regardless of the cause (Leung and Crook, 2019).

The most common cause of hyperphosphatemia is acute or chronic renal disease. In normal conditions, the kidney of healthy individuals has the capacity to excrete about 67% of their absorbed phosphate via the urine. However, when there is a decline in kidney function, the homeostasis of phosphorus is altered, and this occurs even at the early stages of kidney disease (Stremke and Gallant, 2018). Kidney disease is a progressive disease where glomerular filtration rate (GFR) is reduced, and phosphorus cannot be excreted, causing an increase in serum phosphorus concentration. In early renal disease, PTH secretions increase to maintain normal phosphorus concentrations, and though PTH concentrations are elevated, they remain within the upper limits of the normal range.

Both, serum phosphorus and PTH continue to climb as renal function declines, since more PTH is needed to maintain phosphorus homeostasis. As the disease progresses, the hyperphosphatemia persists as PTH is not able to lower phosphorus serum concentration. Thus, high concentration of serum phosphorus in chronic kidney disease alters the homeostasis of phosphorus, causing secondary hyperparathyroidism, vascular calcification, and bone resorption (Calvo and Urivarri, 2013). The condition where there is altered phosphorus (and calcium) metabolism, increased vascular calcification and bone loss is called chronic kidney disease-mineral bone disorder (CKD-MDB), which significantly increases the risk of bone fractures, cardiovascular disease, and ultimately, death (Leung and Crook, 2019; Stremke and Gallant, 2018).

It has been observed that hyperphosphatemia increases the risk of cardiovascular disease and mortality, even in healthy individuals (Calvo and Urivarri, 2013; Vervolet et al., 2017; Kendrick et al., 2011). Vascular calcification is caused when high concentrations of phosphate bind to calcium to create hydroxyapatite (calcium phosphate) and deposit in the arteries, the atheroma plaque and the cardiac valves, causing medial, intimal and heart valve calcification, respectively (Villa-Bellosta, 2021).

Recently, elevated concentrations of phosphorus have been found in patients with dementia and Alzheimer (Asharaf et al., 2019; Li et al., 2017). Dementia is now recognized as a comorbidity of CKD, where high serum concentrations of phosphorus are found. In addition, FBF-23 is also elevated, which has been indirectly associated with dementia and decline in cognitive functions because of its relationship with cerebral small vessel disease (Wright et al., 2014). Further research is needed to determine if measuring phosphorus concentration may be used as an indicator of cognitive impairment or dementia progression, and also to evaluate the effect a low phosphorus diet may have. In addition, the role polyPi has on the brain and other organs, needs to be further explored.

Hypophosphatemia and its consequences

Hypophosphatemia (serum phosphate <2.5 mg/dL) may be caused by low dietary intake, increased renal excretion, decreased intestinal absorption and a shift from extracellular to intracellular phosphorus (Rudolph and Gonin, 2012; Leung and Crook, 2019). Hypophosphatemia is not common, but it may occur in patients that have undergone a renal transplant, patients with primary or secondary (to vitamin D deficiency) hyperparathyroidism, with protein-energy malnutrition, with intestinal malabsorption, or with alcoholism (Rudolph and Gonin, 2012; Leung and Crook, 2019). In addition, chronic use of anti-acids has been shown to develop hypophosphatemia.

There are usually no clinical signs of mild and moderate hypophosphatemia, though severe hypophosphatemia may cause irritability and respiratory difficulty (Rudolph and Gonin, 2012). Also, long term moderate hypophosphatemia has been known to cause rickets or osteomalacia (Leung and Crook, 2019). Hypophosphatemia can also occur in the elderly, mainly because of low phosphorus intake, that combined with low calcium intake, increases PTH, and hence, increases the risk of bone loss and more severe osteoporosis.

Conclusion

It is well recognized that phosphorus is a mineral widely available from food sources, that is efficiently absorbed and tightly regulated by the intestine, the bone and the kidney. Both hyper and hypophosphatemia have adverse health consequences. Phosphorus, its transporters and polyPi may play an important role in cardiovascular, bone and brain health that should be further explored.

References

Anderson, J.B., Sell, M.L., Garner, S.C., Calvo, M.S., 2001. In: Bowman, B., Russell, R. (Eds.), Present Knowledge in Nutrition, eighth ed. International Life Sciences Institute Press, Washington, DC.
Ashraf, A., Stosnach, H., Parkes, H.G., et al., 2019. Pattern of altered plasma elemental phosphorus, calcium, zinc, and iron in Alzheimer's disease. Sci. Rep. 9, 3147.
Bergwitz, C., Jüppner, H., 2010. Regulation of phosphate homeostasis by PTH, vitamin D, and FGF23. Annu. Rev. Med. 61, 91–104.
Bird, R.P., Eskin, M., 2021. The emerging role of phosphorus in human health. Adv. Food Nutr. Res. 96, 27–88.
Calvo, M.S., Lamberg-Allardt, C.J., 2015. Phosphorus. Adv. Nutr. 6, 860–862.
Calvo, M.S., Uribarri, J., 2013. Contributions to total phosphorus intake: all sources considered. Semin. Dial. 26, 54–61.
Fishban, S.N., Nigwekar, S., 2021. Phosphate absorption and hyperphosphatemia management in kidney disease: a physiology-based review. Kidney Med. 3, 1057–1064.
Fulgoni, K., Fulgoni III., V.L., 2021. Trends in total, added, and natural phosphorus intake in adult Americans, NHANES 1988–1994 to NHANES 2015–2016. Nutrients 13, 2249.
Heaney, R.P., 2012. Phosphorus. In: Erdman, J.W., Macdonald, I.A., Zeisel, S.H. (Eds.), Present Knowledge in Nutrition, tenth ed. Wiley-Blackwell, Washington, DC, pp. 447–458.
Institute of Medicine, 1998. (US) Standing Committee on the Scientific Evaluation of Dietary Reference Intakes. Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride. National Academies Press, Washington, DC, USA.
Kendrick, J., Kestenbaum, B., Chonchol, M., 2011. Phosphate and cardiovascular disease. Adv. Chron. Kidney Dis. 18, 113–119.
Lanske, B., Razzaque, S., 2014. Molecular interactions of FGF23 and PTH in phosphate regulation. Kidney Int. 86, 1072–1074.
Leung, J., Crook, M., 2019. Disorders of phosphate metabolism. J. Clin. Pathol. 72, 741–747.
Li, T., Xie, Y., Bowe, B., Xian, H., Al-Aly, Z., 2017. Serum phosphorus levels and risk of incident dementia. PLoS One 12, e0171377.
Rudolph, E.H., Gonin, J.M., 2012. Disorders of phosphorus metabolism. In: Lerma, E.V., Nissen, A.R. (Eds.), Nephrology Secrets, third ed. Mosby, pp. 551–559.
Saurette, M., Alexander, T., 2019. Intestinal phosphate absorption: the paracellular pathway predominates? Exp. Biol. Med. 244, 646–654.
Stremke, E.R., Hill Gallant, K.M., 2018. Intestinal phosphorus absorption in chronic kidney disease. Nutrients 10, 1364.

Takeda, E., Yamamoto, H., Yamanaka-Okumura, H., Taketani, Y., 2012. Dietary phosphorus in bone health and quality of life. Nutr. Rev. 70, 311–321. https://doi.org/10.1111/j.1753-4887.2012.00473.x.

Ureña, P.A., De Brauwere, D.P., 2011. Three feedback loops precisely regulating serum phosphate concentration. Kidney Int. 80, 443–445.

U.S. Department of Agriculture, Agricultural Research Service, 2019. What We Eat in America, 2015–2016.

Vervloet, M.G., Sezer, S., Massy, Z.A., Johansson, L., Cozzolino, M., Fouque, D., 2017. The role of phosphate in kidney disease. Nat. Rev. Nephrol. 13, 27–38.

Villa-Bellosta, R., 2021. Vascular calcification: key roles of phosphate and pyrophosphate. Int. J. Mol. Sci. 22, 13536.

Wright, C.B., Dong, C., Stark, M., Silverberg, S., Rundek, T., Elkind, M., et al., 2014. Plasma FGF23 and the risk of stroke: the Northern Manhattan Study (NOMAS). Neurology 82, 1700–1706.

Younes, M., Aquilina, G., Castle, L., Engel, K.-H., Fowler, P., Frutos Fernandez, M.J., Furst, P., Gürtler, R., Husøy, T., et al., 2019. Scientific opinion on the re-evaluation of phosphoric acid–phosphates–di-, tri- and polyphosphates (E 338–341, E 343, E450–452) as food additives and the safety of proposed extension of use. EFSA J. 17, e05674.

Physical activity: Beneficial effects

Peter T. Katzmarzyk, Pennington Biomedical Research Center, Baton Rouge, LA, United States

© 2023 Elsevier Ltd. All rights reserved.

Introduction 352
Key terms and definitions 352
The genesis of modern physical activity epidemiology 353
Summary of the evidence: physical activity 353
All-cause mortality 354
Cardiovascular disease 354
Cancer 354
Leisure-time versus occupational physical activity 355
Muscle-strengthening activity 356
Light, moderate and vigorous physical activity 356
Summary of the evidence: sedentary behavior 357
The intersection of physical activity and sedentary behavior 358
Summary of the evidence: physical fitness 358
Cardiorespiratory fitness 358
Muscular fitness 359
Global burden of physical inactivity 359
Conclusion 360
References 360

Key Points

- Physical activity is associated with numerous health benefits;
- Contemporary evidence-based physical activity guidelines call for accumulating 150–300 min of moderate-intensity, or 75–150 min of vigorous-intensity physical activity weekly, while incorporating muscle-strengthening activity at least 2 days per week;
- More than a quarter of adults are not sufficiently active to achieve the associated health benefits;
- Up to 8% of premature mortality and noncommunicable diseases worldwide are directly attributable to physical inactivity.

Introduction

An increasing volume of scientific evidence implicates physical inactivity as a major risk factor for several non-communicable diseases (NCDs) and premature mortality. Indeed, physical inactivity is considered the fourth leading cause of death globally, following closely behind high blood pressure, high blood glucose, and tobacco use (World Health Organization, 2009). Recent data from 1.9 million individuals from 168 countries indicate that 27.5% of adults are not sufficiently active to reap the associated health benefits, and the prevalence of physical inactivity is more than twice as high in high-income countries compared to low income countries (Guthold et al., 2018). The purpose of this article is to explore the health benefits associated with physical activity, and to describe the global health impacts of not meeting the current physical activity guidelines.

Key terms and definitions

Physical activity refers to any bodily movement produced by skeletal muscles that results in energy expenditure (Caspersen et al., 1985). On the other hand, physical inactivity generally refers to a lack of physical activity, often in reference to not meeting physical activity guidelines or recommendations. In a similar vein, the term "insufficiently active" refers specifically to not meeting current physical activity guidelines (accumulating 150–300 min of moderate-intensity, or 75–150 min of vigorous-intensity physical activity weekly, while incorporating muscle-strengthening activity at least 2 days per week) (World Health Organization, 2020).

Sedentary behavior has been defined as any waking behavior characterized by an energy expenditure $\leq$1.5 METs (Metabolic Equivalents of Task) while in a sitting, reclining, or lying posture (Tremblay et al., 2017). Thus, sedentary behavior occurs at the opposite end of the energy expenditure spectrum compared to moderate-to-vigorous physical activity. Furthermore, the postural

component of the definition (i.e., sitting, reclining or lying) identifies it as a potential behavioral target for interventions that is separate from physical activity *per se* (Barone Gibbs et al., 2015).

Physical fitness is a concept that is adjacent to that of physical activity. While physical activity is a behavior, physical fitness is defined as an attained set of characteristics that relates to the ability to perform physical activity, and is determined by a variety of factors, including an individual's level of habitual physical activity, heredity and diet (Bouchard and Shephard, 1994). Although there are several components of physical fitness (i.e., cardiorespiratory, muscular, motor, morphological, metabolic, etc.) this article will focus on two of the most commonly studied components in relation to health outcomes: cardiorespiratory and muscular fitness.

The genesis of modern physical activity epidemiology

The value of physical activity in promoting a sound mind and body was recognized by many ancient cultures, including those in China, India and Greece. Indeed, Hippocrates wrote extensively on the utility of exercise to address a number of afflictions in the fifth century BC, as did Galen in the first century AD (Pate, 2007). However, the systematic study of the health benefits of physical activity using modern research practices began in earnest during the 1950s.

The seminal studies on physical activity and coronary heart disease (CHD) in London Transport workers and British civil servants by Jeremy Morris and colleagues marked the beginning of the modern era of physical activity epidemiology (Morris et al., 1953; Paffenbarger et al., 2001). The initial studies observed that physically inactive bus drivers of London's double-decker buses had higher rates of CHD compared to the more active conductors who climbed the stairs. Morris continued to test his hypothesis, reproducing many of the findings when extending the study to London postmen and less active postal clerks (Morris et al., 1953). Subsequently, a series of investigations conducted by Ralph Paffenbarger, Jr. and colleagues on occupational (Paffenbarger and Hale, 1975) and leisure-time physical activity (Paffenbarger et al., 1986) identified physical inactivity as a powerful independent predictor of CHD and all-cause mortality.

Summary of the evidence: physical activity

There has been an incredible amount of research conducted on the health benefits associated with physical activity over the last 60 years. This evidence has informed the development of evidence-based public health physical activity guidelines in individual countries in addition to global recommendations from the World Health Organization (World Health Organization, 2010, 2020). The research supporting these public health efforts has been summarized in extensive systematic reviews of the literature (2018 Physical Activity Guidelines Advisory Committee, 2018; Bull et al., 2020; Physical Activity Guidelines Advisory Committee, 2008; Powell et al., 2019). **Table 1** provides a summary of the health benefits associated with physical activity in adults.

Table 1 Health benefits associated with physical activity.

Reduced risk of:
- All-cause mortality
- CVD mortality
- CVD (stroke and CHD)
- Hypertension
- Type 2 diabetes
- Depression
- Dementia
- Excessive weight gain
- Site-specific cancers (bladder, breast, colon, endometrium, esophagus, kidney, & stomach)
- Falls and fall-related injuries
- Excessive weight gain during pregnancy
- Gestational diabetes
- Post-partum depression

Improved:
- Cognitive function
- Quality of life
- Sleep
- Physical function in older adults

Adapted from Powell et al. (2019).

A comprehensive review of all health benefits associated with physical activity is beyond the scope of this review; therefore, the focus will be on mortality and major non-communicable diseases (NCDs), including cardiovascular disease (CVD) and cancer.

All-cause mortality

Regular physical activity is associated with a significantly reduced risk of premature all-cause mortality. This robust result has been replicated in numerous prospective cohort studies conducted over the past 60 years. The association between physical activity and all-cause mortality is best summarized in the results of a large pooled analysis of six prospective cohort studies comprising more than 600,000 adults (Moore et al., 2012; Arem et al., 2015). The results demonstrated a reverse J-shaped dose-response association between physical activity and all-cause mortality, where the risk of death was reduced by approximately 30–40% for individuals who were physically active for at least 150–300 min per week compared to inactive individuals. There are several characteristics of the dose-response curve there are common to studies of this type: there is no lower threshold for effect; there is a steep, early slope; about 70% of the benefit is reached by 150 min per week of moderate-intensity physical activity, there is no apparent upper threshold for effect; there is no evidence for increased risk at the greatest amounts of physical activity; and there is no obvious "best amount" (Kraus et al., 2019).

Cardiovascular disease

Cardiovascular disease (CVD) is the leading cause of death worldwide, accounting for 32% of global deaths in 2017 (GBD 2017 Causes of Death Collaborators, 2018). Individuals with higher levels of habitual physical activity have a lower risk of developing and dying from CVD (Kraus et al., 2019). A recent meta-analysis characterized the dose-response association between physical activity and incident CVD based on data from 43 studies for CHD and 33 for stroke (Kyu et al., 2016). **Fig. 1** presents the dose-response association between levels of total physical activity and risk of ischemic CHD and stroke. Those individuals in the highest category of physical activity have a 25% and 26% lower risk of CHD and stroke, respectively compared to individuals in the lowest physical activity category (Kyu et al., 2016).

Cancer

In 2020, there are expected to be more than 19 million cancer cases worldwide and approximately 10 million deaths (International Agency for Research on Cancer (IARC), 2020). The 2008 Physical Activity Guidelines Advisory Committee found strong evidence for associations between physical activity and cancers of the breast and colon (Physical Activity Guidelines Advisory Committee, 2008). Based on a substantial increase in the evidence base in the subsequent decade, the 2018 Physical Activity Guidelines Advisory Committee confirmed these associations, and also found strong evidence for associations between physical activity and cancers of the bladder, endometrium, esophagus, kidney, and stomach (2018 Physical Activity Guidelines Advisory Committee, 2018;

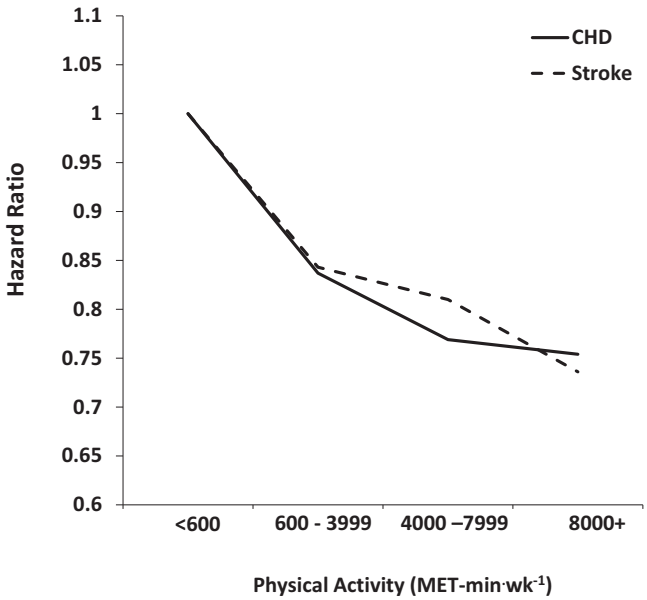


Fig. 1 Relationship between total physical activity and the risk of coronary heart disease (CHD) and ischemic stroke. The figure is drawn from data presented in Kyu et al. (2016) based on a meta-analysis of 43 studies for CHD and 33 studies for stroke.

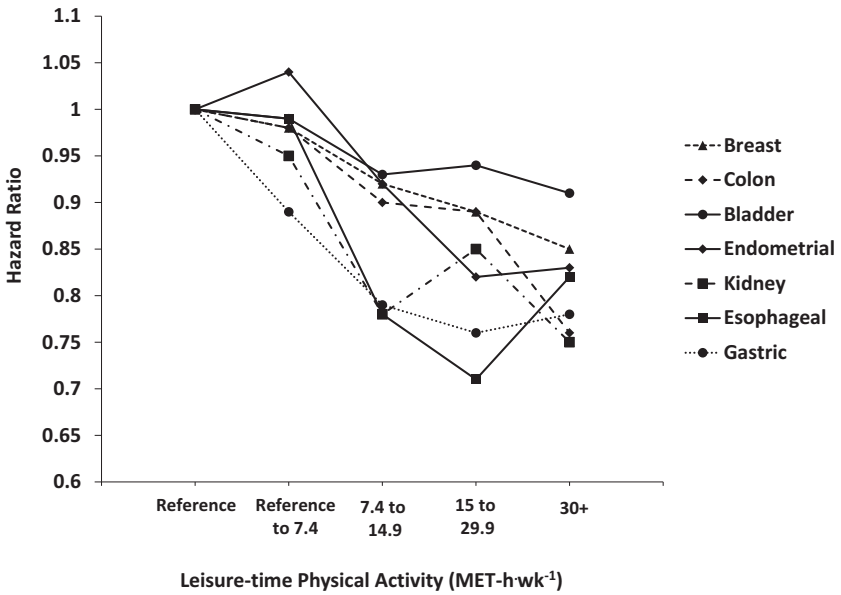


Fig. 2 Associations between leisure-time physical activity and the incidence of several cancers. The figure is drawn from data from a pooled analysis of 9 prospective cohort studies (Matthews et al., 2020).

Mctiernan et al., 2019). Further, there is moderate evidence of an association between physical activity and lung cancer, but this association may be modified by smoking status (McTiernan et al., 2019).

The risks of developing site-specific cancers associated with physical activity were estimated using a pooled analysis of 1.44 million adults from 12 prospective cohort studies (Moore et al., 2016). Comparing the highest (90th percentile) versus lowest (10th percentile) levels of physical activity at baseline, the hazard ratios (HRs) were 0.84 (95% CI: 0.77 to 0.91) for colon cancer, 0.90 (95% CI: 0.87 to 0.93) for breast cancer, 0.87 (95% CI: 0.82 to 0.92) for bladder cancer, 0.79 (95% CI: 0.68 to 0.92) for endometrial cancer, 0.58 (95% CI: 0.37 to 0.89) for esophageal cancer, 0.77 (95% CI: 0.70 to 0.85) for kidney cancer, and 0.78 (95% CI: 0.64 to 0.95) for stomach cancer (Moore et al., 2016). **Fig. 2** shows the dose-response associations between physical activity and the incidence of several cancers, based on a pooled analysis of 9 prospective cohort studies comprising more than 750,000 adults (Matthews et al., 2020). In general, there is a decreasing risk of incident cancer with increasing levels of physical activity.

Leisure-time versus occupational physical activity

Physical activity occurs in a variety of settings, or domains. For example, physical activity can occur during an individual's leisure time, while they are completing household chores, or at work. The majority of the evidence of associations between physical activity and health outcomes has been based on studies of total or leisure-time physical activity, largely due to the nature of the physical activity questionnaires used in large prospective cohort studies. For example, greater leisure-time physical activity is associated with a lower risk of all-cause mortality (Lee et al., 2012), CVD mortality (Cheng et al., 2018), CHD (Sofi et al., 2008), type 2 diabetes (Huai et al., 2016), hypertension (Liu et al., 2017), and several cancers (Matthews et al., 2020).

As described above, the early studies of physical activity and the risk of CHD relied on comparisons of different occupations as proxies for physical activity. For example, studies of bus drivers versus conductors and postal carriers versus clerks provided some of the earliest evidence on the health benefits of physical activity (Morris et al., 1953). A recent umbrella review that summarized the evidence from 17 reviews covering 23 health outcomes found very low to moderate quality evidence that those who engaged in high versus low occupational physical activity had a lower risk of cancers of the colon and prostate, ischemic stroke, CHD, and mental health outcomes such as well-being and life satisfaction (Cillekens et al., 2020). On the other hand, there was also low to very low quality evidence that high levels of occupational physical activity were associated with a greater risk of all-cause mortality in men, depression, anxiety, osteoarthritis, and sleep quality and duration (Cillekens et al., 2020). However, it has been argued that these surprising results may be explained by inadequate classification of occupational demands and lack of proper adjustment for the effects of covariates, particularly from cigarette smoking and low socioeconomic status (Shephard, 2019).

There is a range of physical activity behavior in the population, and the prevalence of physical inactivity is approximately twice as high in high-income countries compared to low-income countries (Guthold et al., 2018). A recent study examined the effects of physical inactivity on mortality across 17 countries that ranged from low-income to high-income, given that the domains (recreational vs. occupational) where physical activity is occurring differ among countries at different levels of development (Lear et al., 2017). The results demonstrated that higher levels of both recreational and non-recreational physical activity were associated

with a lower risk of mortality in adults low-income, middle-income, and high-income countries (Lear et al., 2017). These results demonstrate the universality of physical inactivity as a risk factor for mortality across diverse contexts and settings.

Muscle-strengthening activity

Most of the epidemiological evidence of the health effects associated with physical activity has been derived from studies of physical activity levels accumulated from participation in aerobic activities of different intensities, for example in moderate-to-vigorous physical activity (2018 Physical Activity Guidelines Advisory Committee, 2018). However, public health guidelines consistently call for participation in muscle-strengthening activities, usually for 2 days per week, as part of the physical activity recommendation (World Health Organization, 2020).

Although the effects of resistance exercise and health outcomes such as mortality and NCDs has been less studied, a recent umbrella review of 11 systematic reviews concluded that resistance training was associated with a reduction in all-cause mortality and cardiovascular disease incidence, and an improvement in physical functioning (El-Kotob et al., 2020). Of particular note, a recent meta-analysis showed that resistance training alone or in combination with aerobic exercise training was associated with a significantly lower risk of all-cause mortality; with HRs of 0.79 (95% CI: 0.69 to 0.91) and 0.60 (95% CI: 0.49 to 0.72), respectively (Saeidifard et al., 2019). However, the effects of resistance training on health-related quality of life or cognitive function were less certain (El-Kotob et al., 2020). More research is required to better define the dose and intensity of muscle-strengthening activities that promote optimal health outcomes.

Light, moderate and vigorous physical activity

Early studies of physical activity and health outcomes were focused on measuring moderate-to-vigorous physical activity as the primary exposure. This is largely due to the reliance on self-reported physical activity, and the use of questionnaires that asked participants about their level of participation in a variety of physical activities of moderate or higher intensity ($\geq$3 METs). However, with the advent and application of devices such as accelerometers to assess physical activity, data are now more widely available on the associations between physical activity and health outcomes across a spectrum of activity intensity (light, moderate and vigorous). The associations of total, light and moderate-to-vigorous physical activity with all-cause mortality were explored in a meta-analysis of 8 studies including 36,383 participants (Ekelund et al., 2019). The shape of the dose-response curves is similar across the three exposure variables, and the results indicate that higher levels of physical activity of any intensity are associated with substantially reduced risk for mortality. Compared to the lowest quartile of activity, the HRs for premature mortality in the highest quartile were 0.27 (95% CI: 0.23 to 0.32), 0.38 (95% CI: 0.28 to 0.51), and 0.52 (95% CI: 0.43 to 0.61) for total activity (counts min^{-1}), light intensity activity (min day^{-1}) and MVPA (min day^{-1}), respectively (Ekelund et al., 2019). The effect estimates based on these device-based measurements are greater than are observed in studies that rely on self-report.

A study using device-based measures of physical activity in the US National Health and Nutrition Examination Survey (NHANES) directly addressed the influence of physical activity intensity on the risk of all-cause mortality (Saint-Maurice et al., 2018). The results indicated that the total volume of activity seemed to be the key driver in reducing the risk for all-cause mortality, and the difference in mortality benefits associated with activities of different intensity (i.e., moderate-to-vigorous vs. light) were modest after accounting for the total volume of physical activity (Saint-Maurice et al., 2018). **Fig. 3** presents the joint HRs for light physical activity and moderate-to-vigorous physical activity with all-cause mortality.

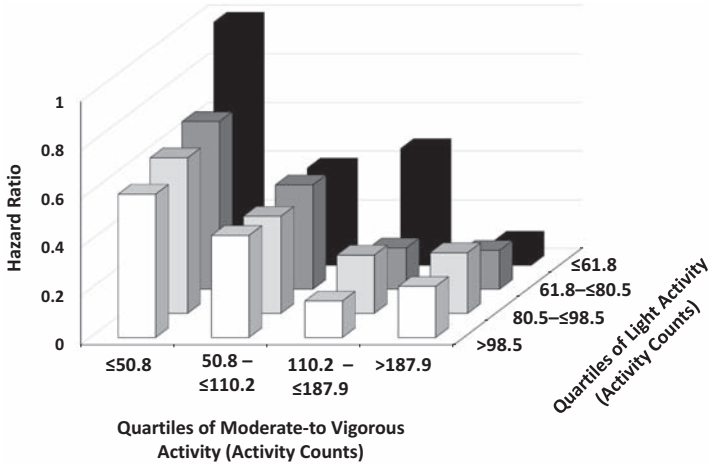


Fig. 3 Joint associations of light and moderate-to-vigorous physical activity with risk of all-cause mortality. Activity counts expressed in thousands. The figure is drawn from analysis of the U.S. National Health and Nutrition Examination Survey in Saint-Maurice et al. (2018).

The results show that higher levels of light activity had a stronger influence on mortality among those doing the least amount moderate-to-vigorous activity, but there was only some evidence of an influence of light activity at higher levels of moderate-to-vigorous intensity activity, such as in the third quartile.

Taken together, the current evidence suggests that there are mortality benefits associated with physical activity of all intensities, including light activity. These results highlight that the benefits of activity can be derived from combining activities of different intensities, based on the preference of the individual.

Summary of the evidence: sedentary behavior

Compared to the large body of evidence on the association between physical activity and health that has accumulated over the last 60 years, the study of sedentary behaviors, such as sitting, and health outcomes, is in its infancy. In the first comprehensive study of sitting and mortality, a clear dose-response relationship between greater daily sitting time and an increased risk of all-cause and CVD mortality was observed in the 12-year mortality follow-up of the Canada Fitness Survey (Katzmarzyk et al., 2009). At baseline, sitting time was classified into 5 groups: almost none of the time, one-fourth of the time, half of the time, three-fourths of the time, and almost all of the time. After adjustment for potential confounders, there was a higher risk of all-cause mortality (HRs: 1.00, 1.00, 1.11, 1.36, 1.54; P for trend <0.0001) and CVD mortality (HR: 1.00, 1.01, 1.22, 1.47, 1.54; P for trend <0.0001) across successively higher levels of sitting time (Katzmarzyk et al., 2009). The following year, similar results were obtained in a 6.6-year follow-up of the Australian Diabetes, Obesity and Lifestyle Study (AusDiab), where there was a significant positive association between television viewing time and mortality from all-causes and from CVD (Dunstan et al., 2010). After adjustment for covariates, the HRs for each 1 h increment in daily television viewing time were 1.11 (95% CI: 1.03 to 1.20) for all-cause mortality and 1.18 (95% CI: 1.03 to 1.35) for CVD mortality (Dunstan et al., 2010).

The results of these early investigations have been replicated in numerous prospective cohort studies in the intervening years (Katzmarzyk et al., 2019). **Fig. 4** presents the results of a harmonized meta-analysis of 8 studies reporting the association between device-measured (accelerometry) sedentary time and risk of premature all-cause mortality (Ekelund et al., 2019). There is a clear dose-response association, with a statistically significant increased risk of mortality observed at more than 9.5 h of daily sedentary time (Ekelund et al., 2019).

The 2018 Physical Activity Guidelines Advisory Committee conducted a systematic review of the health effects associated with sedentary behavior (Katzmarzyk et al., 2019). In addition to all-cause and CVD mortality, they found moderate to strong evidence that higher levels of sedentary behavior are associated with an increased risk of incident CVD, type 2 diabetes, incident endometrial, colon and lung cancer, while there was only limited evidence that sedentary behavior is associated with cancer mortality and weight status (Katzmarzyk et al., 2019). A more recent systematic review expanded the slate of outcomes associated with sedentary behavior and reported that sedentary behavior is also unfavorably associated with cognitive function, depression, function and disability, physical activity levels, and physical health-related quality of life (Saunders et al., 2020).

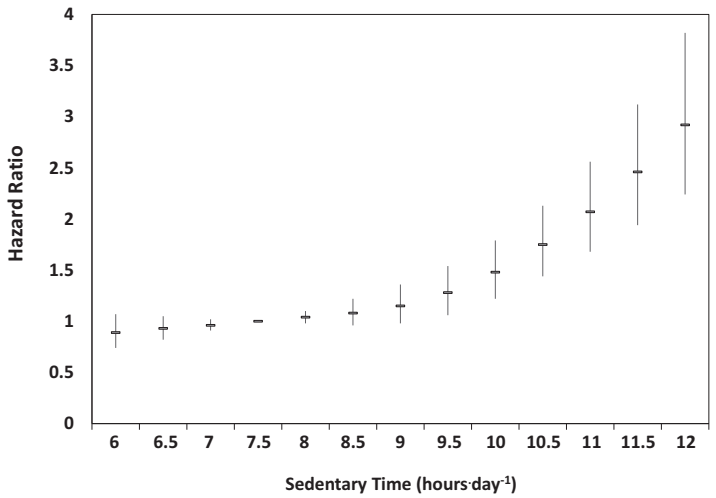


Fig. 4 Dose-response association between sedentary time and risk of all-cause mortality. Covariates in the model included age, sex, wear time, body mass index and socioeconomic position. Error bars are 95% confidence intervals. The figure is drawn from a meta-analysis of data from 8 studies (Ekelund et al., 2019).

The intersection of physical activity and sedentary behavior

While sedentary behavior and MVPA are at opposite ends of the energy expenditure spectrum, they are also two distinct behaviors, and as such, could represent unique behavioral targets for interventions. One can easily imagine a situation where an individual is meeting the guidelines for physical activity (i.e., 150–300 min of moderate-to-vigorous physical activity per week and muscle-strengthening activity on 2 days) but also participates in high levels of excessive sedentary behavior (sitting). The question as to whether high levels of physical activity can attenuate the detrimental effects of sedentary behavior was investigated in a harmonized meta-analysis of over 1 million participants (Ekelund et al., 2016). **Fig. 5** describes the joint association between physical activity and sedentary behavior with all-cause mortality (Ekelund et al., 2016). The figure shows that the risk of mortality associated with higher levels of physical activity decreases within categories of sitting, and that the risk of mortality increases across levels of sitting within physical activity categories. The highest risk category is low physical activity ($\leq$2.5 MET-h wk^{-1}) and high levels of sitting (>8 h day^{-1}) (HR = 1.59; 95% CI: 1.52 to 1.66). Furthermore, the risk associated with sedentary behavior is never significantly elevated in the high physical activity category (white bars; >35.5 MET-h wk^{-1}). The results indicate that ~60–75 min per day of moderate intensity physical activity can eliminate the increased risk of death associated with high sitting time (Ekelund et al., 2016).

Summary of the evidence: physical fitness

As described above, physical fitness can have many facets, including cardiorespiratory, muscular, motor, morphological, and metabolic components (Bouchard and Shephard, 1994). For the purpose of this article the focus will be on the health benefits associated with two main components that are relevant to both performance-related and health-related fitness: cardiorespiratory and muscular.

Cardiorespiratory fitness

The Aerobics Center Longitudinal Study (ACLS) was a prospective observational study of physical activity, physical fitness, and health outcomes among men and women receiving a preventative medical examination, including a graded exercise treadmill test, at the Cooper Clinic in Dallas, TX. The results of an early analysis in the cohort indicated that men and women in the lowest cardiorespiratory fitness quintile were 3.44 (95% CI: 2.05 to 5.77) and 4.65 (95% CI: 2.22 to 9.75) times more likely to die over an average follow-up of 8 years compared with men and women in the upper quintile, respectively (Blair et al., 1989). Several studies have now replicated the results of this seminal study. Indeed, a meta-analysis of 21 studies reported a pooled relative risk (RR) of 1.70 (95% CI: 1.51 to 1.92) for all-cause mortality comparing participants with low versus high cardiorespiratory fitness (Kodama et al., 2009). More recently, the association between cardiorespiratory fitness and mortality was studied in a large cohort of 122,007 consecutive patients undergoing exercise treadmill testing in an academic medical center (Mandsager et al., 2018).The results showed that cardiorespiratory fitness was inversely associated with all-cause mortality without an observed upper limit of benefit; the HR comparing those with low fitness (<25th percentile) to those with high fitness (75th to 97.6th percentile) was 3.90 (95% CI: 3.67 to 4.14) (Mandsager et al., 2018).

The beneficial effects of cardiorespiratory fitness are also observed for CVD. For example, a meta-analysis of 24 studies reported a pooled RR of 1.40 (95% CI: 1.32 to 1.48) for CHD/CVD events comparing participants with low versus high cardiorespiratory fitness (Kodama et al., 2009). Several studies have also demonstrated that cardiorespiratory fitness is associated with a lower risk of stroke (Wang et al., 2020). A meta-analysis of 14 cohort studies documented that high cardiorespiratory fitness was

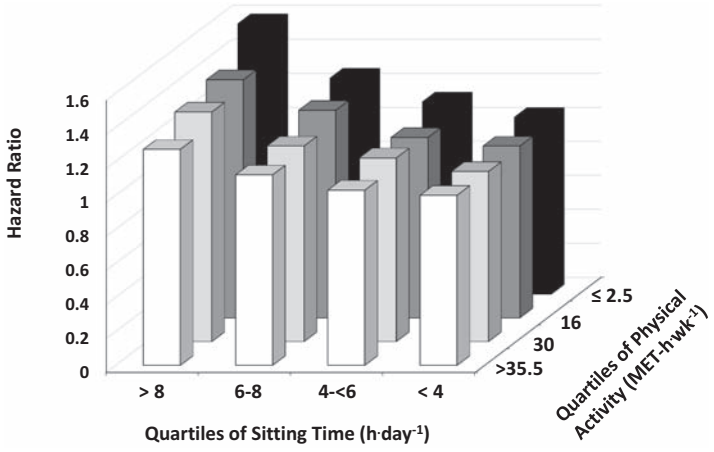


Fig. 5 Joint associations of sitting time and physical activity with all-cause mortality. The figure is drawn from data from a pooled meta-analysis of 1,005,791 individuals (Ekelund et al., 2016).

associated with a 42% lower risk of stroke (RR = 0.58; 95% CI: 0.51 to 0.66) compared to low cardiorespiratory fitness. Furthermore, ischemic stroke (RR = 0.71; 95% CI: 0.54 to 0.93) and hemorrhagic stroke (RR = 0.69; 95% CI: 0.47 to 1.00) demonstrated significant and similar associations with high cardiorespiratory fitness (Wang et al., 2020).

Studies have also documented a strong link between cardiorespiratory fitness and risk of cancer. A meta-analysis of six prospective studies demonstrated a strong, graded, inverse association between cardiorespiratory fitness and total cancer mortality (Schmid and Leitzmann, 2015). Compared to low cardiorespiratory fitness, the RRs for total cancer mortality for moderate and high levels of cardiorespiratory fitness were 0.80 (95% CI: 0.67 to 0.97) and 0.55 (95% CI: 0.47 to 0.65), respectively (Schmid and Leitzmann, 2015). Furthermore, a recent meta-analysis of seven studies showed that compared to low cardiorespiratory fitness, the HRs for moderate and high levels of cardiorespiratory fitness were 0.53 (95%: 0.39 to 0.68) and 0.52 (95% CI: 0.42 to 0.61) for lung cancer, 0.74 (95% CI: 0.55 to 0.93) and 0.77 (95% CI: 0.62 to 0.92) for colorectal cancer, and 0.86 (95% CI: 0.79 to 0.93) and 0.81 (95% CI: 0.75 to 0.87) for all cancer sites, respectively (Pozuelo-Carrascosa et al., 2019).

Moderate and high levels of cardiorespiratory fitness are also protective against developing type 2 diabetes. A recent meta-analysis of 22 studies demonstrated a negative linear association between cardiorespiratory fitness and the development of type 2 diabetes, such that the HR for each 1 MET higher cardiorespiratory fitness was 0.92 (95% CI: 0.92 to 0.92), after statistical adjustment for adiposity and other covariates (Tarp et al., 2019). These results were confirmed in a second meta-analysis of 15 studies which also showed a linear association between higher cardiorespiratory fitness and lower risk of type 2 diabetes, such that the HR associated with every 1 MET increase in cardiorespiratory fitness was 0.90 (95% CI: 0.86 to 0.94) for the general population (Qiu et al., 2019).

Muscular fitness

Muscular fitness is generally assessed using tests of muscular strength and endurance. A commonly used test of muscular strength that can be conducted in the clinic, in a lab, or in the field, is grip strength, measured using a portable hand grip dynamometer. In recent years there has been increasing interest in understanding the health benefits associated with muscular fitness. The results from a large meta-analysis of data from 734,481 participants from 40 studies reported a HR of 0.71 (0.66–0.77) for all-cause mortality comparing highest versus lowest levels of grip strength (Wu et al., 2017). These results are supported by another recent meta-analysis of 1,907,580 participants from 38 studies which demonstrated that higher levels of handgrip strength were associated with a lower risk of all-cause mortality (HR: 0.69; 95% CI: 0.64 to 0.74) compared with lower handgrip strength (Garcia-Hermoso et al., 2018). This study also showed that higher levels of knee extension strength were associated with a lower risk of all-cause mortality (HR: 0.86: 95% CI: 0.80 to 0.93) compared with lower knee extension strength.

In addition to all-cause mortality, there is also evidence that muscular strength is associated with other health outcomes. For example, a large meta-analysis comparing highest versus lowest levels of grip strength reported a HR estimate of 0.61 (0.51–0.74) for incident CVD (1,789,942 participants from 12 studies) and 1.12 (0.83–1.52) for cancer (162,821 participants from 8 studies) (Wu et al., 2017). Another meta-analysis reported a RR for type 2 diabetes of 0.87 (0.81, 0.94) for each standard deviation increase in muscular strength (Tarp et al., 2019).

Fewer studies have examined the association between measures of muscular endurance and health outcomes. For example, data from the longitudinal follow-up of the 1981 Canada Fitness Survey found a significantly higher risk of all-cause mortality in the lowest quartile of sit-ups, compared to the highest quartile, in both men (RR: 2.72; 95% CI: 1.56 to 4.64) and women (RR = 2.26; 95% CI: 1.15 to 4.43) (Katzmarzyk and Craig, 2002). While the analysis from the Canada Fitness Survey did not find an association between push-ups and all-cause mortality (Katzmarzyk and Craig, 2002), a study of 1104 firefighters showed that men who completed more than 40 push-ups had a significantly lower risk of incident CVD compared with those who completed less than 10 push-ups (Incident rate ratio = 0.04; 95% CI: 0.01 to 0.36) (Yang et al., 2019).

Global burden of physical inactivity

Given that physical inactivity is associated with an increased risk of several health outcomes at the level of the individual, it follows that this translates into substantial public health burden. Using data from 1.9 million participants from 168 countries, Guthold and colleagues reported that 27.5% of adults failed to meet the current physical activity recommendations (at least 150 min of moderate-intensity, or 75 min of vigorous-intensity physical activity per week) (Guthold et al., 2018). The prevalence of insufficient physical activity in high-income countries was more than double that of low-income countries, and there has been a significant increase in insufficient physical activity over time in high-income countries (Guthold et al., 2018).

An analysis conducted in 2012 estimated that physical inactivity was responsible for 6% of coronary heart disease, 7% of type 2 diabetes, 10% of breast and colon cancers, and 9% of all-cause deaths globally (Lee et al., 2012). Furthermore, physical inactivity cost health care systems $53.8 billion globally in 2013, of which $31.2 billion was paid by the public sector (Ding et al., 2016). Given the explosion of physical activity research that has been conducted in the last decade, an analysis recently updated the public health burden of physical inactivity to include 15 NCDs and mortality that exhibited strong evidence of association with physical inactivity, as determined by the 2018 U.S. Physical Activity Guidelines Advisory Committee (2018 Physical Activity Guidelines Advisory Committee, 2018). **Fig. 6** presents the population attributable risks (PAR%) from this analysis, which represent the proportion of each of these diseases at the population level that is directly attributable to physical inactivity (Katzmarzyk et al.,

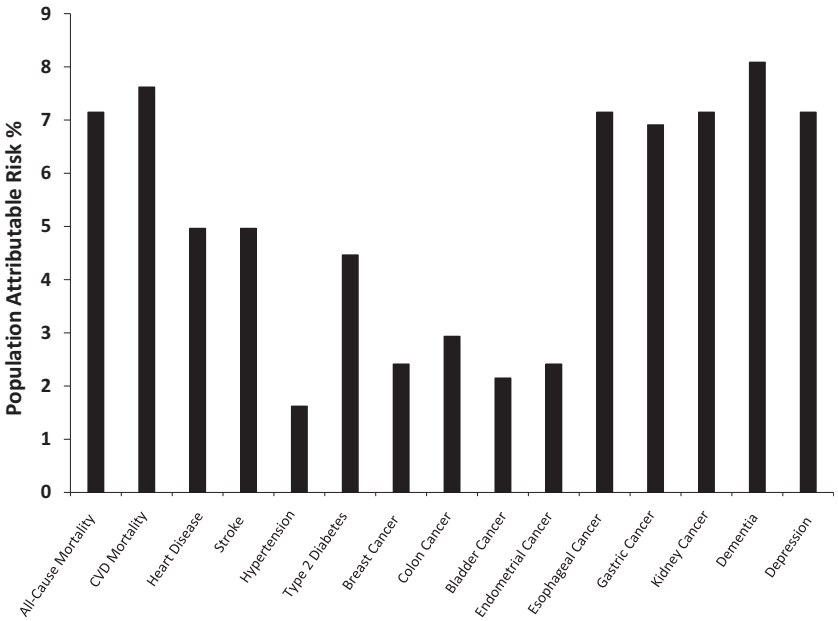


Fig. 6 Global estimates of population attributable risk (PAR%) for mortality and noncommunicable diseases showing strong associations with physical inactivity. The figure is drawn from data presented in Katzmarzyk et al. (2021).

2021). The results demonstrated that up to 8% of these 15 NCDs and mortality is directly attributable to physical inactivity (Katzmarzyk et al., 2021). The PAR% estimates increase across low-, to middle-, to high-income countries in a graded manner, which reflects the greater prevalence of physical inactivity in middle- and high-income countries. However, the percentage of all CVD and all-cause deaths attributable to physical inactivity is highest in middle-income countries (74% and 69%, respectively) because of the larger overall number of deaths in these populations (Katzmarzyk et al., 2021).

Conclusion

There is abundant evidence that physical activity and physical fitness levels are inversely associated with the risk of premature mortality and many NCDs including coronary heart disease, stroke, hypertension, type 2 diabetes, dementia, depression, and cancers of the bladder, breast, colon, endometrium, esophagus, stomach, and kidney. Given the high prevalence of physical inactivity in many countries and regions, the associated public health burden is substantial. The promotion of physical activity should be a priority for public health efforts in addition to clinical practice.

References

2018 Physical Activity Guidelines Advisory Committee, 2018. 2018 Physical Activity Guidelines Advisory Committee Scientific Report. U.S. Department of Health and Human Services, Washington, DC.

Arem, H., Moore, S.C., Patel, A., et al., 2015. Leisure time physical activity and mortality: a detailed pooled analysis of the dose-response relationship. JAMA Intern. Med. 175, 959–967.

Barone Gibbs, B., Hergenroeder, A.L., Katzmarzyk, P.T., et al., 2015. Definition, measurement, and health risks associated with sedentary behavior. Med. Sci. Sports Exerc. 47, 1295–1300.

Blair, S.N., Kohl, H.W., Paffenbarger Jr., R.S., et al., 1989. Physical fitness and all-cause mortality: a prospective study of men and women. J. Am. Med. Assoc. 262, 2395–2401.

Bouchard, C., Shephard, R.J., 1994. Physical activity, fitness and health: the model and key concepts. In: Bouchard, C., Shephard, R.J., Stephens, T. (Eds.), Physical Activity, Fitness and Health. Human Kinestics, Champaign, IL.

Bull, F.C., Al-Ansari, S.S., Biddle, S., et al., 2020. World Health Organization 2020 guidelines on physical activity and sedentary behavior. Br. J. Sports Med. 54, 1451–1462.

Caspersen, C.J., Powell, K.E., Christenson, G.M., 1985. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Publ. Health Rep. 100, 126–131.

Cheng, W., Zhang, Z., Cheng, W., et al., 2018. Associations of leisure-time physical activity with cardiovascular mortality: a systematic review and meta-analysis of 44 prospective cohort studies. Eur. J. Prev. Cardiol. 25, 1864–1872.

Cillekens, B., Lang, M., Van Mechelen, W., et al., 2020. How does occupational physical activity influence health? An umbrella review of 23 health outcomes across 158 observational studies. Br. J. Sports Med. 54, 1474–1481.

Ding, D., Lawson, K.D., Kolbe-Alexander, T., et al., 2016. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. Lancet 388, 1311–1324.

Dunstan, D.W., Barr, E.L., Healy, G.N., et al., 2010. Television viewing time and mortality: the Australian Diabetes, Obesity and Lifestyle Study (AusDiab). Circulation 121, 384–391.

Ekelund, U., Steene-Johannessen, J., Brown, W.J., et al., 2016. Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonized meta-analysis of data from more than 1 million men and women. Lancet 388, 1302–1310.

Ekelund, U., Tarp, J., Steene-Johannessen, J., et al., 2019. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonized meta-analysis. BMJ 366, l4570.
El-Kotob, R., Ponzano, M., Chaput, J.P., et al., 2020. Resistance training and health in adults: an overview of systematic reviews. Appl. Physiol. Nutr. Metabol. 45, S165–S179.
Garcia-Hermoso, A., Cavero-Redondo, I., Ramirez-Velez, R., et al., 2018. Muscular strength as a predictor of all-cause mortality in an apparently healthy population: a systematic review and meta-analysis of data from approximately 2 million men and women. Arch. Phys. Med. Rehabil. 99, 2100–2113 e5.
GBD 2017 Causes of Death Collaborators, 2018. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 392, 1736–1788.
Guthold, R., Stevens, G.A., Riley, L.M., Bull, F.C., 2018. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1.9 million participants. Lancet Glob. Health 6, e1077–e1086.
Huai, P., Han, H., Reilly, K.H., et al., 2016. Leisure-time physical activity and risk of type 2 diabetes: a meta-analysis of prospective cohort studies. Endocrine 52, 226–230.
International Agency for Research on Cancer (Iarc), 2020. Available at: https://gco.iarc.fr/today/data/factsheets/cancers/39-All-cancers-fact-sheet.pdf. (Accessed 4 June 2021).
Katzmarzyk, P.T., Craig, C.L., 2002. Musculoskeletal fitness and risk of mortality. Med. Sci. Sports Exerc. 34, 740–744.
Katzmarzyk, P.T., Church, T.S., Craig, C.L., Bouchard, C., 2009. Sitting time and mortality from all causes, cardiovascular disease, and cancer. Med. Sci. Sports Exerc. 41, 998–1005.
Katzmarzyk, P.T., Powell, K.E., Jakicic, J.M., et al., 2019. Sedentary behavior and health: update from the 2018 Physical Activity Guidelines Advisory Committee. Med. Sci. Sports Exerc. 51, 1227–1241.
Katzmarzyk, P.T., Friedenreich, C., Shiroma, E.J., Lee, I.M., 2021. Physical inactivity and non-communicable disease burden in low-income, middle-income and high-income countries. Br. J. Sports Med. https://doi.org/10.1136/bjsports-2020-103640. Online ahead of print; March 29, 2021.
Kodama, S., Saito, K., Tanaka, S., et al., 2009. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. J. Am. Med. Assoc. 301, 2024–2035.
Kraus, W.E., Powell, K.E., Haskell, W.L., et al., 2019. Physical activity, all-cause and cardiovascular mortality, and cardiovascular disease. Med. Sci. Sports Exerc. 51, 1270–1281.
Kyu, H.H., Bachman, V.F., Alexander, L.T., et al., 2016. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. BMJ 354, i3857.
Lear, S.A., Hu, W., Rangarajan, S., et al., 2017. The effect of physical activity on mortality and cardiovascular disease in 130,000 people from 17 high-income, middle-income, and low-income countries: the PURE study. Lancet 390, 2643–2654.
Lee, I.M., Shiroma, E.J., Lobelo, F., et al., 2012. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 380, 219–229.
Liu, X., Zhang, D., Liu, Y., et al., 2017. Dose-response association between physical activity and incident hypertension: a systematic review and meta-analysis of cohort studies. Hypertension 69, 813–820.
Mandsager, K., Harb, S., Cremer, P., et al., 2018. Association of cardiorespiratory fitness with long-term mortality among adults undergoing exercise treadmill testing. JAMA Netw. Open 1, e183605.
Matthews, C.E., Moore, S.C., Arem, H., et al., 2020. Amount and intensity of leisure-time physical activity and lower cancer risk. J. Clin. Oncol. 38, 686–697.
Mctiernan, A., Friedenreich, C.M., Katzmarzyk, P.T., et al., 2019. Physical activity in cancer prevention and survival: a systematic review. Med. Sci. Sports Exerc. 51, 1252–1261.
Moore, S.C., Patel, A.V., Matthews, C.E., et al., 2012. Leisure time physical activity of moderate to vigorous intensity and mortality: a large pooled cohort analysis. PLoS Med. 9, e1001335.
Moore, S.C., Lee, I.M., Weiderpass, E., et al., 2016. Association of leisure-time physical activity with risk of 26 types of cancer in 1.44 million adults. JAMA Intern. Med. 176, 816–825.
Morris, J.N., Heady, J.A., Raffle, P.A.B., et al., 1953. Coronary heart-disease and physical activity of work. Lancet ii (1053–57), 1111–1120.
Paffenbarger Jr., R.S., Hale, W.E., 1975. Work activity and coronary heart mortality. N. Engl. J. Med. 292, 545–550.
Paffenbarger Jr., R.S., Hyde, R.T., Wing, A.L., Hsieh, C.C., 1986. Physical activity, all-cause mortality, and longevity of college alumni. N. Engl. J. Med. 314, 605–613.
Paffenbarger Jr., R.S., Blair, S.N., Lee, I.M., 2001. A history of physical activity, cardiovascular health and longevity: the scientific contributions of Jeremy N Morris, DSc, DPH, FRCP. Int. J. Epidemiol. 30, 1184–1192.
Pate, R.R., 2007. Historical perspectives on physical activity, fitness and health. In: Bouchard, C., Blair, S.N., Haskell, W.L. (Eds.), Physical Activity and Health. Human Kinetics, Champaign, IL.
Physical Activity Guidelines Advisory Committee, 2008. Physical Activity Guidelines Advisory Committee Report, 2008. U.S. Department of Health and Human Services, Washington, DC.
Powell, K.E., King, A.C., Buchner, D.M., et al., 2019. The scientific foundation for the Physical Activity Guidelines for Americans, 2nd Edition. J. Phys. Activ. Health 16, 1–11.
Pozuelo-Carrascosa, D.P., Alvarez-Bueno, C., Cavero-Redondo, I., et al., 2019. Cardiorespiratory fitness and site-specific risk of cancer in men: a systematic review and meta-analysis. Eur. J. Canc. 113, 58–68.
Qiu, S., Cai, X., Yang, B., et al., 2019. Association between cardiorespiratory fitness and risk of type 2 diabetes: a meta-analysis. Obesity 27, 315–324.
Saeidifard, F., Medina-Inojosa, J.R., West, C.P., et al., 2019. The association of resistance training with mortality: a systematic review and meta-analysis. Eur. J. Prev. Cardiol. 26, 1647–1665.
Saint-Maurice, P.F., Troiano, R.P., Berrigan, D., et al., 2018. Volume of light versus moderate-to-vigorous physical activity: similar benefits for all-cause mortality? J. Am. Heart Assoc. 7, e008815.
Saunders, T.J., Mcisaac, T., Douillette, K., et al., 2020. Sedentary behavior and health in adults: an overview of systematic reviews. Appl. Physiol. Nutr. Metabol. 45, S197–S217.
Schmid, D., Leitzmann, M.F., 2015. Cardiorespiratory fitness as predictor of cancer mortality: a systematic review and meta-analysis. Ann. Oncol. 26, 272–278.
Shephard, R.J., 2019. Is there a "recent occupational paradox" where highly active physically active workers die early? Or are there failures in some study methods? Br. J. Sports Med. 53 (24), 1557–1559.
Sofi, F., Capalbo, A., Cesari, F., et al., 2008. Physical activity during leisure time and primary prevention of coronary heart disease: an updated meta-analysis of cohort studies. Eur. J. Cardiovasc. Prev. Rehabil. 15, 247–257.
Tarp, J., Stole, A.P., Blond, K., Grontved, A., 2019. Cardiorespiratory fitness, muscular strength and risk of type 2 diabetes: a systematic review and meta-analysis. Diabetologia 62, 1129–1142.
Tremblay, M.S., Aubert, S., Barnes, J.D., et al., 2017. Sedentary Behavior Research Network (SBRN)—terminology consensus project process and outcome. Int. J. Behav. Nutr. Phys. Activ. 14, 75.
Wang, Y., Li, F., Cheng, Y., et al., 2020. Cardiorespiratory fitness as a quantitative predictor of the risk of stroke: a dose-response meta-analysis. J. Neurol. 267, 491–501.
World Health Organization, 2009. Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks. World Health Organization, Geneva.
World Health Organization, 2010. Global Recommendations on Physical Activity for Health. World Health Organization, Geneva, Switzerland.
World Health Organization, 2020. WHO Guidelines on Physical Activity and Sedentary Behavior. World Health Organization, Geneva.
Wu, Y., Wang, W., Liu, T., Zhang, D., 2017. Association of grip strength with risk of all-cause mortality, cardiovascular diseases, and cancer in community-dwelling populations: a meta-analysis of prospective cohort studies. J. Am. Med. Dir. Assoc. 18, 551.e17–551.e35.
Yang, J., Christophi, C.A., Farioli, A., et al., 2019. Association between push-up exercise capacity and future cardiovascular events among active adult men. JAMA Netw. Open 2, e188341.

Potassium

L.J. Appel, Johns Hopkins University, Baltimore, MD, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of L.J. Appel, Potassium, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 52–55, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00228-2.

Introduction 362
Acid–base considerations 362
Adverse effects of insufficient potassium 363
Kidney stones and bone demineralization 363
Elevated blood pressure 363
Cardiovascular disease 364
Adverse effects of excess potassium intake 365
Recommended potassium intake, current intake, and dietary sources 365
Conclusion 366
References 366

Key points

- Dietary potassium is one of the major factors that determine blood pressure (BP) levels.
- Insufficient potassium together with excess sodium intake raise BP, an extraordinarily common and important risk factor for cardiovascular disease, particularly stroke, and kidney diseases.
- In addition to high BP, an insufficient intake of potassium leads to greater salt sensitivity, an increased risk of kidney stones and possibly increased bone loss.
- Individuals should strive to increase their consumption of potassium-rich foods, particularly fruits and vegetables.
- Replacement of regular salt with potassium-enriched salt is a promising approach to raise potassium intake and lower sodium intake and BP.

Introduction

The major intracellular cation in the body is potassium, which is maintained at a concentration of approximately 145 mmol/L in intracellular fluid but at much lower concentrations in the plasma and interstitial fluid (3.8–5 mmol/L of extracellular fluid). The high intracellular concentration of potassium is maintained via the activity of the Na^+/K^+-ATPase pump. Because this enzyme is stimulated by insulin, alterations in the plasma concentration of insulin can affect cellular influx of potassium and thus plasma concentration of potassium. Relatively small changes in the concentration of extracellular potassium greatly affect the extracellular/intracellular potassium ratio and thereby affect nerve transmission, muscle contraction, and vascular tone.

In unprocessed foods, potassium occurs mainly in association with bicarbonate-generating precursors such as citrate and, to a lesser extent, with phosphate. In processed foods to which potassium is added and in supplements, the form of potassium is potassium chloride. In healthy people, a large fraction of dietary potassium is absorbed. Most potassium is excreted in urine, whereas the remainder is excreted mainly in feces, with much smaller amounts excreted in sweat (Agarwal et al., 1994). Interestingly, there are large racial differences in the percentage of dietary potassium that is excreted in the urine. Blacks excrete a lower percentage of dietary potassium than whites (e.g., 67% in blacks vs. 74% in whites); this racial difference also differs by concomitant diet (Turban et al., 2008). Because most potassium that is filtered by the glomerulus of the kidney is reabsorbed (70–80%) in the proximal tubule, only a small amount of filtered potassium reaches the distal tubule. The majority of potassium in urine results from secretion of potassium into the cortical collecting duct, a secretion regulated by a number of factors including the hormone aldosterone. An elevated plasma concentration of potassium stimulates the adrenal cortex to release aldosterone, which in turn increases secretion of potassium in the cortical collecting duct.

Acid–base considerations

A diet rich in potassium from fruits and vegetables favorably affects acid–base metabolism because these foods are also rich in precursors of bicarbonate. Acting as a buffer, the bicarbonate-yielding organic anions found in fruits and vegetables neutralize noncarbonic acids generated from meats and other high-protein foods. In the setting of an inadequate intake of bicarbonate precursors,

which is commonplace in contemporary Western diets, excess acid in the blood titrates bone buffer (Frassetto et al., 2001). As a result, bone becomes demineralized, calcium is released, and urinary calcium excretion increases. This state has been termed a "low-grade metabolic acidosis." Increased bone breakdown and calcium-containing kidney stones are adverse clinical consequences of excess diet-derived acids. Diets rich in potassium with its bicarbonate precursors might prevent kidney stones and bone loss. Studies also suggest that a diet rich in bicarbonate precursors might also retard the progression of chronic kidney disease (Scialla et al., 2012). In processed foods to which potassium is added and in potassium supplements, the conjugate anion is typically chloride, which cannot act as a buffer.

Adverse effects of insufficient potassium

The adverse effects of potassium insufficiency depend on the extent of the potassium deficit (Institute of Medicine (US) Panel on Dietary Reference Intakes for Electrolytes and Water, 2005). Severe potassium deficiency, which most commonly results from diuretic-induced potassium losses, is characterized by a serum potassium concentration of less than 3.5 mmol/L, termed hypokalemia. The adverse consequences of hypokalemia are cardiac arrhythmias, muscle weakness, and glucose intolerance. Moderate potassium deficiency, which commonly results from an inadequate dietary intake of potassium, occurs without hypokalemia and is characterized by increased BP, increased salt sensitivity, an increased risk of kidney stones, and increased bone turnover. Through its effects on BP, an inadequate intake of dietary potassium increases the risk of cardiovascular (CV) and kidney diseases, more specifically, stroke, myocardial infarction, heart failure, kidney failure and likely cognitive decline. Independent of its effects on BP, an insufficient intake of potassium might also increase the risk of stroke and perhaps other diseases as well.

Kidney stones and bone demineralization

Because of its effects on acid–base balance, an increased dietary potassium intake might have favorable effects on kidney stone formation. In one large observational study of women (**Fig. 1**), there was a progressive, inverse relationship between greater intake of potassium and incident kidney stones (Curhan et al., 1997). At a median potassium intake of 4.7 g day^{-1} (119 mmol day^{-1}), the risk of developing a kidney stone was 35% less compared to that for women with an intake of <2.0 g day^{-1} (52 mmol day^{-1}). Similar results were evident in men (Curhan et al., 1993). In the one available trial, an intake of approximately 3.6–4.7 g day^{-1} (92–120 mmol day^{-1}) of potassium in the form of potassium citrate reduced the risk of recurrent kidney stones (Barcelo et al., 1993).

Epidemiologic studies have consistently documented that increased potassium intake is associated with greater bone mineral density (Institute of Medicine (US) Panel on Dietary Reference Intakes for Electrolytes and Water, 2005). In trials, supplemental potassium bicarbonate and potassium citrate reduced bone-turnover as manifest by less urinary calcium excretion and by favorable changes in biomarkers of bone turnover (Moseley et al., 2013; Dawson-Hughes et al., 2015). Few trials have examined the effects of potassium on bone density, and none on clinical outcomes related to osteoporosis.

Elevated blood pressure

High levels of potassium intake are associated with reduced BP. Observational data have been reasonably consistent in documenting this inverse relationship, whereas data from individual trials have been less consistent, particularly in trials of nonhypertensive individuals. Several meta-analyses of these trials have each documented a significant inverse relationship such that higher potassium

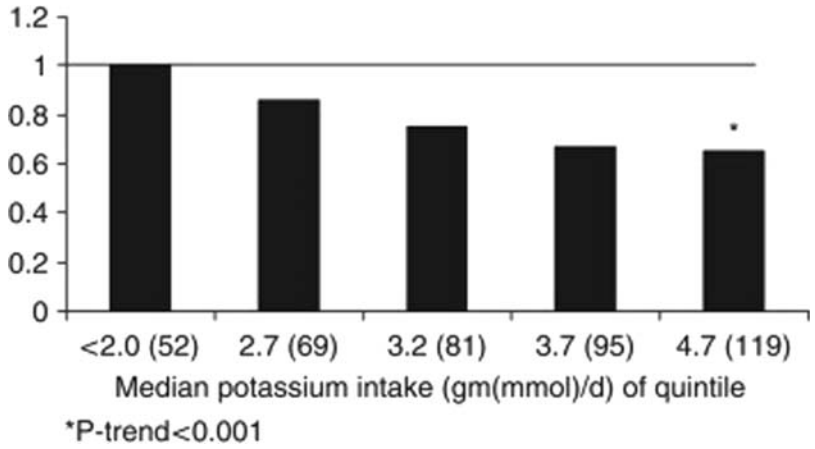


Fig. 1 Relative risk of kidney stones during 12 years of follow-up by quintile of potassium intake in 91 731 women. Data from Curhan et al. (1997).

intake, most often from potassium chloride, lowers BP. In a recent meta-analysis of 18 trials, average net systolic/diastolic BP reductions were 6.4/3.5 mmHg (National Academies of Sciences, 2019). In these meta-analyses, BP reductions are statistically significant in trials that enrolled persons with hypertension, whereas BP reductions in individuals without hypertension are typically nonsignificant. In general, BP reductions from potassium are greater in African Americans than non-African Americans (Whelton et al., 1997). Most of the trials that tested the effects of potassium on BP used pill supplements, typically potassium chloride. In the few available trials, potassium chloride and potassium bicarbonate had similar effects on BP (He et al., 2005). Mechanisms by which potassium affects BP and its sequelae has recently been reviewed (Gritter et al., 2019).

Likewise, potassium-enriched salt substitutes lower BP. In a recent meta-analysis of trials, most conducted in China, potassium-enriched salt substitutes lowered mean systolic/diastolic BP by 5.6/2.9 mmHg, in comparison to regular salt. The effects are most evident in persons with hypertension; in those without hypertension, available evidence is sparse, without clear evidence of BP reduction. Still, in a large, population-based study in Peru, potassium-enriched salt substitutes significantly reduced mean systolic/diastolic BP by 1.3/0.8 mmHg (Bernabe-Ortiz et al., 2020). It is noteworthy that 82% of participants were nonhypertensive at baseline; in this subgroup, use of the salt substitute was associated with a 51% reduced risk of developing hypertension compared with regular salt. Such evidence is consistent with observational studies suggesting that dietary potassium might retard the age-related rise in systolic BP (Rose et al., 1988).

A high potassium intake has been shown to blunt the rise in BP in response to increased salt intake. The term "salt-sensitive BP" applies to those individuals or subgroups who experience the greatest reduction in BP when salt intake is reduced. One metabolic study of 38 healthy, nonhypertensive men (24 African Americans and 14 non-African Americans) investigated the effect of potassium supplementation on the pressor effect of salt loading (5.7 g day^{-1} of sodium (250 mmol)). Before potassium was supplemented, 79% of the African American men and 26% of the non-African American men were termed "salt sensitive," as defined by a salt-induced increase in mean arterial pressure of at least 3 mmHg. There was a progressive reduction in the frequency of salt sensitivity as the dose of potassium was increased. In the African Americans with severe salt sensitivity, increasing dietary potassium to 4.7 g day^{-1} (120 mmol day^{-1}) reduced the frequency of salt sensitivity to 20%, the same percentage as that observed in non-African American subjects when their potassium intake was increased to only 2.7 g day^{-1} (70 mmol day^{-1}) (Morris et al., 1999). A recent review of intervention studies confirms this association and suggests that when potassium intake is high in African-Americans, there is no evidence of racial disparities in salt sensitivity (Kurtz et al., 2021).

Other studies indicate that potassium has greater BP lowering in the context of a higher salt intake and lesser BP reduction in the setting of a lower salt intake (Whelton et al., 1997). Conversely, the BP reduction from a reduced-salt intake is greatest when potassium intake is low (Sacks et al., 2001). These data are consistent with sub-additive effects of reduced-salt intake and increased potassium intake on BP.

Cardiovascular disease

Through its effects on BP, an increased intake of potassium should reduce the occurrence of BP-related cardiovascular disease. Potassium may also have protective effects that are independent of BP reduction. This possibility has been tested in experimental studies conducted in rodents. In a series of animal models, the addition of either potassium chloride or potassium citrate markedly reduced mortality from stroke (Tobian et al., 1985). Interestingly, these reductions occurred in the setting of stable BP. Such data indicate that potassium has both BP-dependent and BP-independent properties that are cardioprotective.

In humans, observational studies often document that at higher levels of potassium intake, there is a reduced risk of subsequent stroke (Aburto et al., 2013). Meta-analyses of such studies have likewise been reasonably consistent. For example, in the Health Professionals Follow-Up Study, there was a significant inverse relationship between baseline potassium intake and stroke after

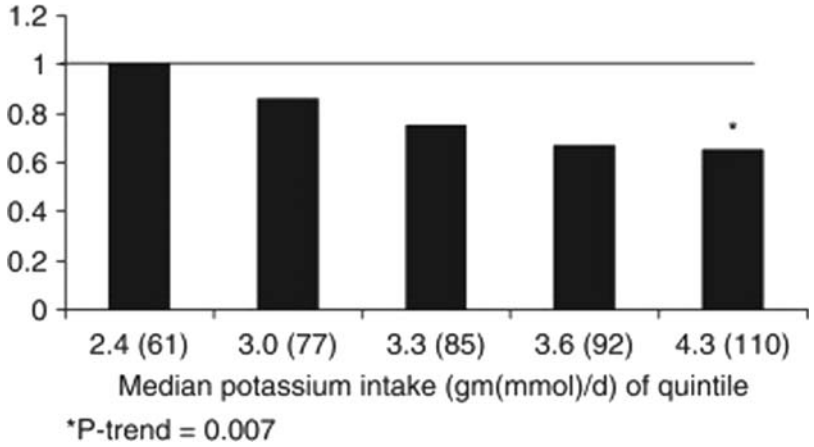


Fig. 2 Relative risk of ischemic stroke by quintile of potassium intake in 43,738 men. Data from Ascherio et al. (1998).

adjustment for established cardiovascular disease risk factors, including BP and caloric intake (Fig. 2) (Ascherio et al., 1998). In this study, a median potassium intake of 4.3 g day^{-1} (110 mmol day^{-1}) was associated with a 41% reduced risk of stroke in comparison to those with a median intake of 2.4 g day^{-1} (61 mmol day^{-1}). Consistent with these studies are other observational studies that have repeatedly documented a reduced risk of stroke from an increased intake of fruits and vegetables. In contrast to the inverse relationship of potassium intake and stroke, the relationship of potassium intake with myocardial infarction and mortality is inconsistent.

In observational studies, clinical events such as stroke are often more strongly associated with the dietary sodium/potassium ratio, than either sodium or potassium alone. As discussed previously, such a relationship is biologically plausible. However, there is also the potential for methodological artifact, specifically, the sodium/potassium ratio is less prone to errors from under- and overreporting in dietary assessment and urinary electrolyte excretion than estimated intakes of sodium and potassium alone.

Recent studies have strengthened the evidence base in support of a higher potassium and reduced sodium intake. First, a major trial documented that replacement of regular salt with a potassium-enriched salt substitute significantly reduced the risk of stroke (14%) and total mortality (12%) in rural China, where potassium intake is low (Neal et al., 2021). Second, a meta-analysis of observational studies in generally healthy adults documented that lower potassium and higher sodium intakes, as estimated from multiple 24 h urine samples, were associated in a dose–response manner with a higher cardiovascular risk (Ma et al., 2021).

Adverse effects of excess potassium intake

In the generally healthy population with normal kidney function, a high-potassium intake from foods, potassium supplements, and potassium-enriched salt substitutes poses no risk because excess potassium is readily excreted in the urine. Under unusual circumstances, an extremely high intake of potassium from supplements can lead to acute toxicity in healthy individuals (Institute of Medicine (US) Panel on Dietary Reference Intakes for Electrolytes and Water, 2005).

However, in individuals whose urinary potassium excretion is impaired, there is concern about the occurrence of hyperkalemia, which can cause cardiac arrhythmias. Drugs that commonly impair potassium excretion are angiotensin converting enzyme inhibitors, angiotensin receptor blockers, and potassium-sparing diuretics. Common medical conditions associated with impaired potassium excretion are diabetes, chronic renal insufficiency, end stage renal disease, severe heart failure, and adrenal insufficiency. Elderly individuals are at increased risk of hyperkalemia because they often have one or more of these conditions or take one or more of the medications that impair potassium excretion.

A contemporary issue is the effect of dietary potassium and potassium-enriched salt substitutes on blood levels of potassium and the occurrence of hyperkalemia in persons with impaired potassium excretion (Greer et al., 2020). The relevance of this issue is highlighted by replacement of regular salt with potassium-enriched salt substitutes as a means to lower BP, an approach which led to significant reductions in stroke and total mortality in a major trial conducted in China (Neal et al., 2021). In the general population without potassium excretion, a meta-analysis of trials documented that 45 mmol/d (1755 mg/d) led to a small increase in mean serum potassium levels of 0.14 mmol/L. Unfortunately, there is sparse evidence on the effects on serum potassium of increased potassium intake from any source in persons with impaired potassium excretion. In the one available trial, which enrolled persons with chronic kidney disease, a higher potassium diet that provided 2300 mg/d (60 mmol/d) more than a lower potassium diet increased mean net serum potassium levels by 0.21 mmol/L, and two participants had confirmed hyperkalemia (serum $K \geq 5.5$ mmol/L) (Turban et al., 2021).

Recommended potassium intake, current intake, and dietary sources

On the basis of available data, a National Academy of Medicine committee set an Adequate Intake for potassium of 3400 mg per day for healthy men and 2600 mg per day for healthy women (National Academies of Sciences, 2019). This level of dietary intake should lower BP levels and have the potential benefits of reducing the adverse effects of salt on BP, reducing the risk of kidney stones, and possibly decreasing bone loss (Institute of Medicine (US) Panel on Dietary Reference Intakes for Electrolytes and Water, 2005). Currently, dietary intake of potassium is lower than these levels. In recent surveys, the mean intake of potassium in the United States was ~3200 mg/day in adult men and ~2400 mg/day in adult women (Bailey et al., 2015). Because African Americans have a relatively low intake of potassium and a high prevalence of high BP and salt sensitivity, this subgroup of the population would especially benefit from an increased potassium intake. Worldwide, dietary potassium intakes by country vary considerably, with a very low, average intake in China and a much higher average intake in Finland (van Mierlo et al., 2010).

Dietary intake surveys typically do not include estimates from salt substitutes and supplements. However, less than 10% of those surveyed in NHANES reported using salt substitutes or a reduced-sodium salt. Because a high dietary intake of potassium can be achieved through diet rather than pills and because potassium derived from foods also comes with bicarbonate precursors, as well as a variety of other nutrients, the preferred strategy to achieve the recommended potassium intake is to consume foods. In addition, replacement of regular salt with potassium-enriched salt substitutes is a promising strategy to reduce sodium intake, not just raise potassium intake.

Dietary sources of potassium, as well as bicarbonate precursors, are fresh fruits, fruit juices, dried fruits, and vegetables. Although meat, milk, and cereal products contain potassium, their content of bicarbonate precursors does not sufficiently balance the amount

Table 1 Foods rich in potassium.

Food	*Portion size*	*Potassium content, g (meq)*
Beans		
Cooked dried beans	1/2 cup	0.4 (10.7)
Lima beans	5/8 cup	0.4 (10.8)
Fruit		
Apple	1 medium	0.1 (2.8)
Apricots	3 medium	0.3 (7.2)
Banana	6 in.	0.4 (9.5)
Cantaloupe	1/4 medium	0.3 (6.4)
Dates	10 pitted	0.6 (16.6)
Orange	1 small	0.3 (7.7)
Peach	1 medium	0.2 (5.2)
Prunes, dried	10 medium	0.7 (17.8)
Raisins	1 tablespoon	0.1 (2.0)
Watermelon	1 slice	0.6 (15.4)
Fruit juices		
Grapefruit	1 cup	0.4 (10.4)
Orange	1 cup	0.5 (12.4)
Pineapple	1 cup	0.4 (9.2)
Tomato	1 cup	0.5 (13.7)
Vegetables		
Corn	1 ear	0.2 (5.0)
Potato	–	–
White	1 boiled	0.3 (7.3)
Sweet	1 boiled	0.3 (7.7)
Tomato	1 medium	0.4 (9.4)
Squash, winter	1/2 cup boiled	0.5 (11.9)
Meats		
Hamburger	1 patty	0.4 (9.8)
Rib roast	2 slices	0.4 (11.2)
Fish (e.g., haddock)	1 medium fillet	0.3 (8.0)
Milk		
Skim milk	8 oz.	0.3 (8.5)
Whole milk	8 oz.	0.4 (9.0)

of acid-forming precursors, such as sulfur amino acids, found in higher protein foods. The typical content of potassium-rich foods is displayed in Table 1. Potassium-enriched, salt substitutes currently available in the marketplace range from 0.4 to 2.8 g/teaspoon (11–72 mmol/teaspoon) of potassium, all as potassium chloride; salt substitutes with no more than 30% potassium avoid the bitter, metallic taste of potassium chloride and are virtually indistinguishable from regular salt.

Conclusion

Potassium is an essential nutrient that is required for normal cellular function. Although humans evolved on diets rich in potassium, contemporary diets are quite low in potassium. An increased intake of potassium from foods should prevent many of the adverse effects of inadequate potassium intake, which are higher BP levels, greater salt sensitivity, increased risk of kidney stones, and possibly increased bone loss. Based on the results of one major trial, replacement of regular salt with potassium-enriched salt reduces the risk of stroke in persons with a low potassium intake. In view of the high prevalence of elevated BP, stroke, and conditions related to bone demineralization (i.e., osteoporosis and kidney stones) in the general population, individuals should strive to increase their consumption of potassium-rich foods, particularly fruits and vegetables.

References

Aburto, N.J., Hanson, S., Gutierrez, H., Hooper, L., Elliott, P., Cappuccio, F.P., 2013. Effect of increased potassium intake on cardiovascular risk factors and disease: systematic review and meta-analyses. BMJ 346, f1378.

Agarwal, R., Afzalpurkar, R., Fordtran, J.S., 1994. Pathophysiology of potassium absorption and secretion by the human intestine. Gastroenterology 107, 548–571.

Ascherio, A., Rimm, E.B., Hernan, M.A., Giovannucci, E.L., Kawachi, I., Stampfer, M.J., Willett, W.C., 1998. Intake of potassium, magnesium, calcium, and fiber and risk of stroke among US men. Circulation 98, 1198–1204.

Bailey, R.L., Parker, E.A., Rhodes, D.G., Goldman, J.D., Clemens, J.C., Moshfegh, A.J., Thuppal, S.V., Weaver, C.M., 2015. Estimating sodium and potassium intakes and their ratio in the American diet: data from the 2011–2012 NHANES. J. Nutr. 146, 745–750.

Barcelo, P., Wuhl, O., Servitge, E., Rousaud, A., Pak, C.Y., 1993. Randomized double-blind study of potassium citrate in idiopathic hypocitraturic calcium nephrolithiasis. J. Urol. 150, 1761–1764.

Bernabe-Ortiz, A., Sal, Y.R.V.G., Ponce-Lucero, V., Cardenas, M.K., Carrillo-Larco, R.M., Diez-Canseco, F., Pesantes, M.A., Sacksteder, K.A., Gilman, R.H., Miranda, J.J., 2020. Effect of salt substitution on community-wide blood pressure and hypertension incidence. Nat. Med. 26, 374–378.

Curhan, G.C., Willett, W.C., Rimm, E.B., Stampfer, M.J., 1993. A prospective study of dietary calcium and other nutrients and the risk of symptomatic kidney stones. N. Engl. J. Med. 328, 833–838.

Curhan, G.C., Willett, W.C., Speizer, F.E., Spiegelman, D., Stampfer, M.J., 1997. Comparison of dietary calcium with supplemental calcium and other nutrients as factors affecting the risk for kidney stones in women. Ann. Intern. Med. 126, 497–504.

Dawson-Hughes, B., Harris, S.S., Palermo, N.J., Gilhooly, C.H., Shea, M.K., Fielding, R.A., Ceglia, L., 2015. Potassium bicarbonate supplementation lowers bone turnover and calcium excretion in older men and women: a randomized dose-finding trial. J. Bone Miner. Res. 30, 2103–2111.

Frassetto, L., Morris Jr., R.C., Sellmeyer, D.E., Todd, K., Sebastian, A., 2001. Diet, evolution and aging—the pathophysiologic effects of the post-agricultural inversion of the potassium-to-sodium and base-to-chloride ratios in the human diet. Eur. J. Nutr. 40, 200–213.

Greer, R.C., Marklund, M., Anderson, C.A.M., Cobb, L.K., Dalcin, A.T., Henry, M., Appel, L.J., 2020. Potassium-enriched salt substitutes as a means to lower blood pressure: benefits and risks. Hypertension 75, 266–274.

Gritter, M., Rotmans, J.I., Hoorn, E.J., 2019. Role of dietary K(+) in natriuresis, blood pressure reduction, cardiovascular protection, and renoprotection. Hypertension 73, 15–23.

He, F.J., Markandu, N.D., Coltart, R., Barron, J., Macgregor, G.A., 2005. Effect of short-term supplementation of potassium chloride and potassium citrate on blood pressure in hypertensives. Hypertension 45, 571–574.

Institute Of Medicine (U.S.) Panel on dietary reference intakes for electrolytes and water, 2005. DRI, Dietary Reference Intakes for Water, Potassium, Sodium, Chloride, and Sulfate. National Academies Press, Washington, D.C.

Kurtz, T.W., Dicarlo, S.E., Pravenec, M., Morris Jr., R.C., 2021. No evidence of racial disparities in blood pressure salt sensitivity when potassium intake exceeds levels recommended in the US dietary guidelines. Am. J. Physiol. Heart Circ. Physiol. 320, H1903–H1918.

Ma, Y., He, F.J., Sun, Q., Yuan, C., Kieneker, L.M., Curhan, G.C., Macgregor, G.A., Bakker, S.J.L., Campbell, N.R.C., Wang, M., Rimm, E.B., Manson, J.E., Willett, W.C., Hofman, A., Gansevoort, R.T., Cook, N.R., Hu, F.B., 2021. 24-Hour urinary sodium and potassium excretion and cardiovascular risk. N. Engl. J. Med. 386 (3), 252–263.

Morris Jr., R.C., Sebastian, A., Forman, A., Tanaka, M., Schmidlin, O., 1999. Normotensive salt sensitivity: effects of race and dietary potassium. Hypertension 33, 18–23.

Moseley, K.F., Weaver, C.M., Appel, L., Sebastian, A., Sellmeyer, D.E., 2013. Potassium citrate supplementation results in sustained improvement in calcium balance in older men and women. J. Bone Miner. Res. 28, 497–504.

National Academies of Sciences, Engineering, and Medicine, 2019. Dietary Reference Intakes for Sodium and Potassium. The National Academies Press, Washington (DC).

Neal, B., Wu, Y., Feng, X., Zhang, R., Zhang, Y., Shi, J., Zhang, J., Tian, M., Huang, L., Li, Z., Yu, Y., Zhao, Y., Zhou, B., Sun, J., Liu, Y., Yin, X., Hao, Z., Yu, J., Li, K.C., Zhang, X., Duan, P., Wang, F., Ma, B., Shi, W., Dl Tanna, G.L., Stepien, S., Shan, S., Pearson, S.A., Li, N., Yan, L.L., Labarthe, D., Elliott, P., 2021. Effect of salt substitution on cardiovascular events and death. N. Engl. J. Med. 385, 1067–1077.

Rose, G., Stamler, J., Stamler, R., et al., 1988. Intersalt: an international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion. Intersalt Cooperative Research Group. BMJ 297, 319–328.

Sacks, F.M., Svetkey, L.P., Vollmer, W.M., Appel, L.J., Bray, G.A., Harsha, D., Obarzanek, E., Conlin, P.R., Miller Iii, E.R., Simons-Morton, D.G., Karanja, N., Lin, P.H., 2001. Effects on blood pressure of reduced dietary sodium and the Dietary Approaches to Stop Hypertension (DASH) diet. DASH-sodium collaborative research group. N. Engl. J. Med. 344, 3–10.

Scialla, J.J., Appel, L.J., Astor, B.C., Miller Iii, E.R., Beddhu, S., Woodward, M., Parekh, R.S., Anderson, C.A., African American Study of Kidney Disease, Hypertension Study Group, 2012. Net endogenous acid production is associated with a faster decline in GFR in African Americans. Kidney Int. 82, 106–112.

Tobian, L., Lange, J., Ulm, K., Wold, L., Iwai, J., 1985. Potassium reduces cerebral hemorrhage and death rate in hypertensive rats, even when blood pressure is not lowered. Hypertension 7, I110–I114.

Turban, S., Juraschek, S.P., Miller Iii, E.R., Anderson, C.A.M., White, K., Charleston, J., Appel, L.J., 2021. Randomized trial on the effects of dietary potassium on blood pressure and serum potassium levels in adults with chronic kidney disease. Nutrients 13.

Turban, S., Miller Iii, E.R., Ange, B., Appel, L.J., 2008. Racial differences in urinary potassium excretion. J. Am. Soc. Nephrol. 19, 1396–1402.

Van Mierlo, L.A., Greyling, A., Zock, P.L., Kok, F.J., Geleijnse, J.M., 2010. Suboptimal potassium intake and potential impact on population blood pressure. Arch. Intern. Med. 170, 1501–1502.

Whelton, P.K., He, J., Cutler, J.A., Brancati, F.L., Appel, L.J., Follmann, D., Klag, M.J., 1997. Effects of oral potassium on blood pressure. Meta-analysis of randomized controlled clinical trials. JAMA 277, 1624–1632.

Protein deficiency

Rajavel Elango, Department of Pediatrics, University of British Columbia, BC Children's Hospital Research Institute, School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

© 2023 Elsevier Ltd. All rights reserved.

Introduction 368
Protein needs 369
Relationship between protein intake and needs 370
Protein quality 370
Adaptation to low protein intakes 371
Protein deficiency and influence of life-stage 372
Children 372
Pregnancy & lactation 372
Elderly 373
Causes of protein deficiency 373
Principles of treatment of protein deficiency 373
Summary and conclusions 374
References 374

Key points

- Understand importance of protein and amino acid intakes in humans
- Understand the dynamic nature of protein metabolism and the relationship between protein intakes and protein needs
- Understand the influence of life-stage on protein deficiency
- Understand the causes and, principles of treatment of protein deficiency

Introduction

"Protein" is derived from the Greek word, "*proteios*", which means of the first rank or position (Elango and Laviano, 2019). Originally coined in 1838, the word was chosen to represent the fundamental nature of protein's role in human nutrition. Protein as a macronutrient is unique for it's source of nitrogen, when compared to the other 2 macronutrients—carbohydrates and fats. The nutritional importance of protein is also because of their constituent amino acids. The 20 α-amino acids which are part of mammalian protein are classified based on their nutritional importance into indispensable (essential) amino acids, conditionally indispensable (essential) amino acids and the dispensable (nonessential) amino acids (**Table 1**). From a dietary point of view, protein quantity, and protein quality of the dietary protein sources, to provide all amino acids in the right balance, are important to ensure normal bodily functions (Pencharz et al., 2016). From a body cellular, tissue and organ point of view all 20 amino acids are needed in different proportions and amounts at any given time to ensure adequate protein synthesis. Protein and individual amino acids are also necessary for several key nitrogen containing compounds including creatine, nucleic acids, nucleotides, neurotransmitters, and hormones such as insulin, growth hormone, etc (**Table 2**).

Table 1 Indispensable, conditionally indispensable and dispensable amino acids for humans.

Indispensable	*Conditionally indispensable*	*Dispensable*
Histidine	Arginine	Alanine
Isoleucine	Cysteine	Aspartic acid
Leucine	Glutamine	Asparagine
Lysine	Glycine	Glutamic acid
Methionine	Proline	Serine
Phenylalanine	Tyrosine	
Threonine		
Tryptophan		
Valine		

 https://doi.org/10.1016/B978-0-12-821848-8.00125-6

Table 2 Examples of roles of amino acids other than protein synthesis.

Precursor amino acids	*End product*	*Function*
Tyrosine	Catecholamines	Stimulation of glycogenolysis, lipolysis
Tryptophan	Serotonin	Vasoconstrictor, stimulates smooth muscle contraction
Aspartate, glutamate	Nucleic acids	Cell division
Methionine, glycine, serine	Methyl group metabolism	DNA methylation
Cysteine, glutamate, glycine	Glutathione	Antioxidant
Arginine	Nitric oxide	Vasodilation, polyamines

Protein needs

Protein in the body is in a dynamic state referred to as protein turnover (Fig. 1), which involves continuous degradation to free amino acids, and re-synthesis of new proteins (Waterlow, 2006). The free amino acids are also constantly degraded and oxidized to carbon dioxide and nitrogenous end products, principally urea and ammonia (Fig. 2). Dietary protein is necessary to replenish these losses of amino acids to maintain protein homeostasis. Furthermore, in children, and other active stages of growth, including pregnancy and lactation there is an increased need from dietary protein to allow new tissue growth. The requirement for dietary protein is therefore composed of two components: maintenance and growth.

Within the context of the above description, the definition of protein needs, as currently accepted by a joint World Health Organization/Food and Agriculture Organization/United Nations University (WHO/FAO/UNU, 2007) consultation is: "the lowest level of dietary protein intake that will balance the losses of nitrogen from the body, and thus maintain the body protein mass, in persons at energy balance with modest levels of physical activity, plus, in children or in pregnant or lactating women, the needs associated with the deposition of tissues or the secretion of milk at rates consistent with good health".

Key points to be considered in the current definition of protein need includes "energy balance", and "physical activity". Energy needs and protein needs are interconnected, and thus if energy needs are not met, protein needs are proportionally increased, and increases the probability of protein deficiency. The influence of energy needs are more during active stages of growth and development (Uauy, 2013).

Current recommendations for protein needs through different life stages are primarily based on nitrogen balance studies in adult humans. All other life-stage needs for protein are based on the concept of maintenance (in adults) plus growth (tissue deposition x efficiency of utilization) (Table 3). Human milk is the optimal source of nutrients for normal, full-term infants throughout the first year of life and is recommended as the sole nutritional source for infants during the first 4–6 months of life. For infants weaned to solid foods, and up to 10 years of age the recommendations are combined for both sexes. For children 10 year and above—18 years, the recommendations are separated by sexes to take the different trajectories of growth. For pregnancy, the recommendations are provided based on trimesters, with maintenance needs plus additional needs for fetal growth which progressively increase with stages of gestation. For lactation, the recommendations are based on milk production rates, combined with the concentration of protein and non-protein nitrogen in human milk. The values are provided based on infants receiving exclusive human milk for the first 6 months post-partum, and partial milk feeding for 6–12 months post-partum (Table 3).

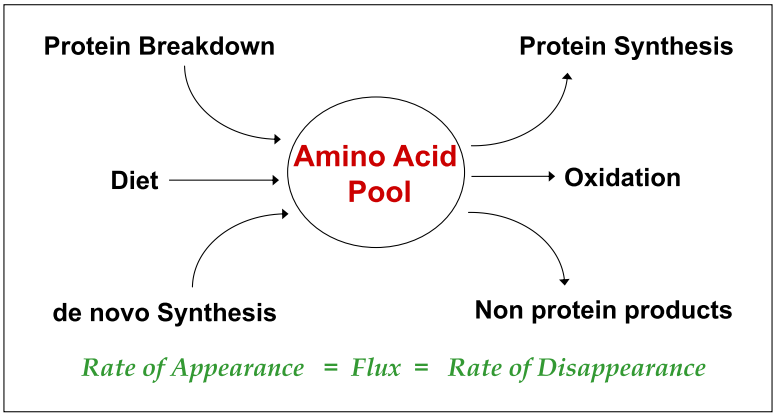


Fig. 1 Protein metabolism is dynamic. Protein is in a constant state of flux, with the rate of appearance and disappearance from the freely available amino acid pool. Amino acid flux is regulated by diet, breakdown, and endogenous (de novo) synthesis on the one hand, and protein synthesis in tissues, oxidation and synthesis of nitrogen containing metabolites on the other.

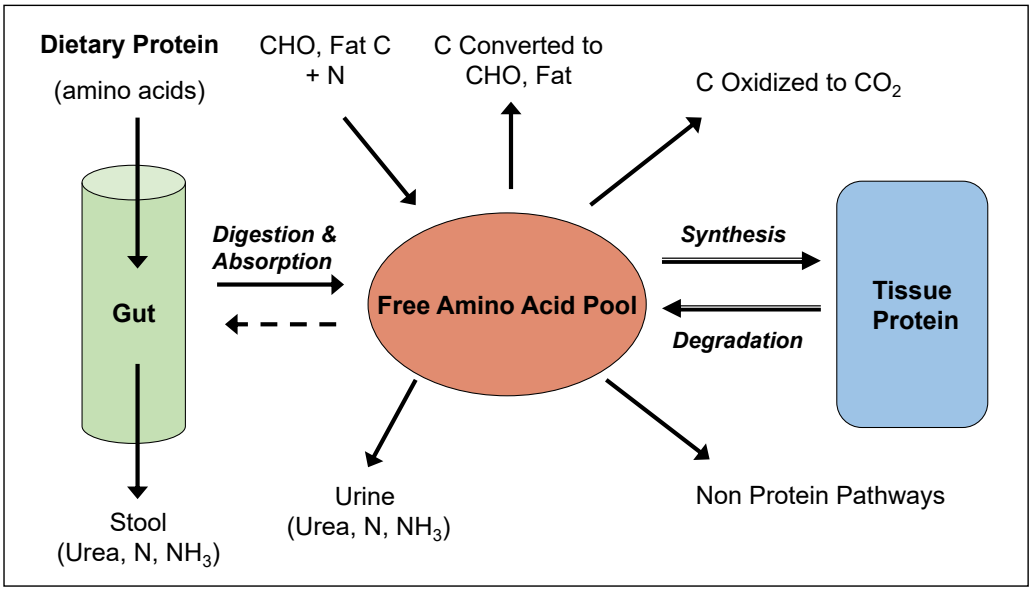


Fig. 2 Overview of protein metabolism. Protein metabolism is influenced by digestion and absorption from the gut to release amino acids in circulation. Protein metabolism, via the carbon skeletons of amino acids interacts with the other macronutrients—carbohydrates (CHO) and fats. Protein synthesis into new tissues and tissue degradation to release amino acids occur constantly. Amino acids are also involved in synthesis of key metabolites which play functional roles. Complete catabolism of amino acids releases the carbon skeleton as CO_2, and the nitrogen is excreted, primarily as urea in stool and urine

These recommendations have been argued to be underestimates based on other end-point criteria, other than nitrogen balance, such as stable-isotope based oxidation studies (Elango et al., 2010). Although from a severe protein deficiency point of view, intakes below these current recommendations would certainly not be sufficient.

Relationship between protein intake and needs

The human body does not have a major store of amino acids, other than skeletal muscle protein, which is not freely available. Thus, in order to maintain the constant state of protein turnover (Fig. 1) a daily intake of protein is required to maintain homeostasis (Fig. 2). Both high or low protein intakes will influence all aspects of protein turnover including protein synthesis, protein degradation, amino acid synthesis, amino acid oxidation and urea production (Young and Marchini, 1990). The rate at which changes occur in each of these inter-related components is regulated by

- protein/nitrogen intake
- degree to which pattern of amino acid intake matches the amino acid need in the body
- balance between indispensable, conditionally indispensable and dispensable amino acid intake
- degree to which energy intake matches energy needs

Protein quality

In the context of dietary protein sources, it is important to note that animal sources such as milk, eggs, meat are considered "complete proteins" due to the balance in amino acid composition (indispensable:dispensable amino acids). On the other hand, plant sources such as legumes/pulses, and cereal grains are considered "incomplete proteins" since they are limiting in one or more indispensable amino acids. Pulses/legumes which are considered the primary proteins in plant-based diets include lentils, beans, chickpeas etc. are limiting in methionine. Cereal grains such as rice, wheat, corn/maize are limiting in lysine. Protein complementation, where a pulse is combined with a cereal in a meal is one way that reduces the impact of the imbalanced amino acid content (Manary et al., 2016). Furthermore, plant-based proteins are also relatively lower in total protein/nitrogen content, are also less digestible compared to animal sources, and have a lower energy density. Recently the FAO has recommended that each individual amino acid be considered a "nutrient". This has huge significance, since the amino acid digestibility, and the proportion that is subsequently bioavailable varies considerably among the amino acids, and can influence the overall protein quality of foods. In the overall context, likelihood of protein deficiency is higher when populations are predominantly consuming plant-based diets (Bandyopadhyay et al., 2020).

Table 3 Protein requirement through the life cycle.[1]

Age	*Average requirement (EAR)*[2]	*Safe level of intake (RDA)*[3]
	(g protein/kg body weight per day)	
Exclusive human milk feeding		
1 month	1.41	1.77
2 months	1.23	1.50
3 months	1.13	1.36
4 months	1.07	1.24
6 months	0.98	1.14
Infants weaned to solid foods, and older children		
0.5 years	1.12	1.31
1 year	0.95	1.14
1.5 years	0.85	1.03
2 years	0.79	0.97
3 years	0.73	0.90
4 years	0.69	0.86
5 years	0.69	0.85
6 years	0.72	0.89
7–10 years	0.75	0.92
Girls		
11 years	0.73	0.90
12 years	0.72	1.89
13 years	0.71	1.88
14 years	0.70	0.87
15 years	0.69	0.85
16 years	0.68	0.84
17 years	0.67	0.83
18 years	0.66	0.82
Boys		
11 years	0.75	0.91
12 years	0.74	0.90
13 years	0.73	0.90
14 years	0.72	0.89
15 years	0.72	0.88
16 years	0.71	0.87
17 years	0.70	0.86
18 years	0.69	0.85
[4]Adults 19 years+	0.66	0.83
Pregnancy	0.66	0.83 + 1, 9, 31 g/d for first, second and third trimester, respectively
Lactation	0.66	0.83 + 21 g/d for 0–6 months postpartum +12.5 g/d for >6 months postpartum

[1]Data from WHO/FAO/UNU (2007).
[2]EAR, Estimated Average Requirement – Calculated from maintenance + growth (rate of protein deposition × efficiency of protein utilization).
[3]RDA, Recommended Dietary Allowance – Calculated from EAR + 2XSD of EAR.
[4]Adult recommendations include both sexes, and older age individuals.

Adaptation to low protein intakes

Within the context of protein homeostasis, the human body does respond in the short-term to a reduced protein intake by altering several mechanisms. Amino acid oxidation rates are reduced quite quickly, with reduced plasma concentrations of amino acids. Whole body protein synthesis and degradation are reduced, although there might be some tissue specific differences, where the liver and muscle response rates differ, due to an acute reduction in protein intakes. The urea cycle whose principal role is to facilitate the removal of excess protein/nitrogen in the diet and its accumulation as ammonia in the body, adapts to a low protein diet by reducing the cycle activity to promote nitrogen conservation (Young and Marchini, 1990). These adaptive responses are considered successful, as it allows the maintenance of a steady state in body homeostasis, although there is a change in function of one or more

mechanism to achieve this. On the other hand, when the restriction of protein continues in the long-term, there will be accommodative responses, where a loss of function as a consequence of low protein occurs. This influence (adaptive vs. accommodative) can result in significant health impacts, and they vary based on life-stage and the presence of disease.

Protein deficiency and influence of life-stage

Children

Traditionally protein deficiency has been associated with young growing children and the term Protein-Energy Malnutrition (PEM). Historically, PEM were defined as marasmus and kwashiorkor. Marasmus was characterized with low intakes of both energy and protein, while kwashiorkor was characterized with intakes of sufficient energy and low protein. It is now recognized that there are several different gradations of protein deficiency including Severe Acute Malnutrition (SAM), Moderate Acute Malnutrition (MAM), and chronic malnutrition—stunting (Jahoor et al., 2008).

Based on the involvement of edema, SAM is further classified as non-edematous (marasmus) and edematous (kwashiorkor and marasmic-kwashiorkor). In both marasmus and marasmic-kwashiorkor wasting is involved. Wasting is defined by WHO as mid-upper arm circumference (MUAC) less than 115 mm or weight-for-height z score (WHZ) less than—3 for ages 6 months–59 months. The edematous forms of malnutrition are characterized by several metabolic perturbations beyond edema, including anorexia, hypopigmentation of skin, hair and liver steatosis. In general, the edematous forms are more difficult to treat and manage. From a protein metabolism point of view, in children with edematous SAM a slower protein breakdown rate is observed with a reduction in plasma amino acid concentrations of indispensable amino acids, particularly of methionine and phenylalanine. Since Methionine and phenylalanine are the pre-cursors of cysteine and tyrosine, respectively, their plasma concentrations are remarkably reduced as well. Cysteine is a key limiting factor for synthesis of glutathione (the body's primary antioxidant); tyrosine is the pre-cursor for synthesis of neurotransmitters and catecholamines (**Table 2**). Deficiency of these two conditionally indispensable amino acids needs to be corrected by dietary supplementation as part of the management of SAM. Methionine deficiency, which is often observed in SAM has also been the focus of new studies. Methionine via S-adenosyl Methionine, is the primary methyl donor for nearly all methylation reactions in the body. As part of the methionine cycle, several methyl (one-carbon metabolites) such as betaine, choline, homocysteine are altered in edematous SAM, and complicates the management of this disorder (May et al., 2022). Non-edematous SAM on the other hand protein breakdown rates are not different and thus children are able to maintain plasma amino acid concentrations.

Globally, MAM defined as a weight-for-age between -3 and -2 z-scores below the median of the WHO child growth standards is also more prevalent. Progression from MAM to SAM is likely if protein and/or energy deficiency continues to persist.

Compared to SAM and MAM, stunting in children is a marker of chronic undernutrition, and is more widespread with nearly a quarter of the world's children affected. Stunting is defined as height-for-age less than -2 z-scores below the median of the WHO child growth standards. Using a large data set and country-level data, it was shown that a significant relationship exists between population risk of dietary protein inadequacy and prevalence of stunting. It was also shown that protein quality of the diet remained significantly related to stunting rates after adjustment for dietary energy, house-hold income, and prevalence of infection. In response to infection protein metabolism changes occurs at the hepatic level with the production of acute phase proteins (APP) (Jahoor et al., 2008). Some APP involved in host defense mechanisms increase in serum concentrations, and are referred to as positive APP (α_1-antitrypsin, haptoglobin, fibrinogen) and some APP involved in transport mechanisms decrease and are referred to as negative APP (albumin, transthyretin, retinol binding protein). These negative APP are usually rich in specific amino acids including cysteine, glycine, serine and thus protein and amino acid deficiency contribute to the further reduction in plasma concentrations.

Pregnancy & lactation

Protein deficiency during the critical stage of pregnancy has negative consequences for both the mother and infant. Low birthweight and preterm birth have both been associated with lower protein intakes during pregnancy (Bandyopadhyay et al., 2020). In turn, these neonates have increased mortality, morbidity and long-term developmental impacts which extend into adulthood. Mothers who themselves are undernourished have increased child birth related consequences, increased risk for preeclampsia, gestational diabetes in the current pregnancy. Furthermore, it also increases the risk for chronic diseases later including type 2 diabetes, hypertension and metabolic syndrome. There is also evidence that protein needs progressively increase with gestation stages, however amino acid needs do not proportionally increase (Elango and Ball, 2016). For example, phenylalanine needs increase early in pregnancy, and continue to increase during later stages. Lysine needs in early gestation is similar to non-pregnancy needs, but increase significantly by later stages. Recently it was also shown that glycine, a dispensable amino acid in adults, is conditionally-dispensable in later stages of pregnancy (Rasmussen et al., 2021).

Lactation stages, especially during the immediate post-partum period is an extremely energy demanding stage. While lactation protein needs are yet to be determined directly, earlier nitrogen balance studies have shown that to maintain positive nitrogen balance, a substantially higher protein intake than current recommendations (**Table 3**) is required. Human milk content is influenced by the maternal diet; however, protein deficiency has not been shown to directly reduce milk output in the short-term. It is likely that maternal endogenous turnover compensates to a certain extent, but this needs to be confirmed. Current evidence does

point to the long-term impact of perinatal and postnatal influence of maternal nutrition on child health through the concept of developmental origins of health and disease, and protein/amino acids have a large regulatory role in this paradigm.

Elderly

Older adults (65 y+) increasingly form a large proportion of the human population and they are also at increased risk of protein-energy malnutrition. Aging is associated with progressive loss of muscle mass, which impacts mobility, loss of function, decreased immune function and increased risk of infection (Wolfe, 2012). Simultaneously there is a loss of appetite, inability to eat a variety of foods and dysphagia. Protein synthesis can be stimulated in older individuals with protein intake; however a higher quantity and quality of protein is required, compared to younger individuals to see a similar effect. This effect referred to as "anabolic resistance" is observed in older adults with a reduced appearance of amino acids in circulation following a meal, and the blunted response in muscle protein synthesis in response to protein ingestion. Combination of resistance exercise, high quality proteins and more protein energy dense meals have shown to increase protein synthesis (Phillips and Martinson, 2019). Leucine, an indispensable amino acid has been the widely studied amino acid in elderly, and there is evidence to suggest that leucine requirements might be considerably higher compared to younger individuals. There are also considerable sex-based differences in responses to protein and amino acid ingestion, and thus contributes to the response to protein supplements in older adults.

Causes of protein deficiency

The primary cause for protein deficiency is usually lack of availability of food or starvation. The reason for this could be multi-fold, due to socio-economic conditions, geo-political issues, season of the year (rainy vs. dry), famine and an emergency (natural/man-made). Food insecurity can exist in both developed and developing countries, and can contribute to protein deficiency.

A deficiency in protein could also occur under conditions of higher physiological demands, due to chronic stress, moderate persistent inflammation. Clinical conditions such as burns, post-surgical individuals and trauma victims also will have elevated protein demands due to a significant increase in protein turnover (both protein synthesis and degradation). In the various forms of pediatric malnutrition protein deficiency could be multifactorial due to the presence of acute infections, environmental toxins and an overall lack of hygiene (Golden, 2010). Furthermore, underlying pathologies such as small intestinal damage due to mucosal atrophy, villous blunting, and intestinal inflammation could impact digestion and absorption of protein and several other nutrients. Parasitic infestation is also common in many populations living in poor sanitary conditions, and presence of parasites has been shown to increase specific amino acid (lysine, leucine) and protein needs.

Principles of treatment of protein deficiency

The key to treating protein deficiency is to ensure a thorough understanding of the underlying cause(s) has been fully examined. This includes an assessment of the physiological stage (life-stage), presence of other clinical symptoms, diagnoses of disease, and any other associated infections. This should be combined with a thorough assessment of food intake, which will allow the characterization of energy, protein, other macro- and micro-nutrient intakes. It is important to note that many key micronutrients, vitamins and minerals accompany dietary protein. Thus, deficiency of protein is likely associated with micronutrient deficiency, which would exacerbate the problem. This aspect is perfectly highlighted with the example of methionine deficiency that was discussed earlier in SAM. As part of the methionine cycle, 4 key B vitamins interact—folate, B12 (cobalamin), pyridoxine and riboflavin. Some of the impacts of methionine deficiency can be reduced if optimal B vitamin status is maintained.

For the treatment of specific forms of child undernutrition, WHO has created guidelines and products available to use. These include F75 (which is considered a starter), F100 (for catch-up growth) which are used for acute management of SAM (Golden, 2010). The F75 has relatively less calories and protein, with the goal being to correct electrolyte imbalance, correct edema, and allow a slow recovery to body homeostasis. The F100 is more dense with energy, calories and other electrolytes/minerals to allow catch up growth to occur and a pathway to recovery in these malnourished children. In a community setting, for these children recovering from SAM, as well as to treat MAM, stunting, Ready to Use Therapeutic Foods (RUTF) are available. These are blends of peanut - milk base, corn-soy, etc which are produced and packaged to be food-safe. The protein quality of these RUTF is an intense area of research, with an emphasis on locally resourced and produced products, such that the treatment is both context relevant and allows the local communities to be self-reliant in managing protein malnutrition.

For older individuals in developed countries, similar energy-nutrient dense easy to drink products are available, although the products could be expensive. Similar to young children, the underlying cause of protein deficiency must be explored, to ensure other pathologic conditions, micro-nutrient deficiencies are corrected, and more often physical activity with some form of resistance exercise will be required to ensure a full recovery from protein deficiency is possible.

Summary and conclusions

In summary, protein deficiency can occur to individuals, primarily during vulnerable stages of growth and development (childhood, pregnancy, lactation) and in the elderly. While the major cause is a reduction in dietary protein intake, overall reduction in food intake, lower protein quality of the foods being consumed, presence of underlying infections, or disease can exacerbate protein deficiency. Treatment and management of protein deficiency in children is primarily based on the type of protein malnutrition, with the SAM forms requiring correction of underlying biochemical, physiological and immunological factors first. During the catch-up phase in children recovering from protein deficiency, in children with chronic undernutrition and stunting, pregnant, lactating mothers and in elderly the focus is on both adequate quantity of protein and quality of the protein source. Therefore, consumption of "high quality" protein rich in all the indispensable amino acids, principally animal sources, or a balanced blend of complementary mixtures of plant proteins are recommended.

References

Bandyopadhyay, S., Shivakumar, N., Kurpad, A.V., 2020. Protein intakes of pregnant women and children in India-protein quality implications. Matern. Child Nutr. 16 (Suppl. 3), e12952.

Elango, R., Ball, R.O., 2016. Protein and amino acid requirements during pregnancy. Adv. Nutr. 7, 839S–844S.

Elango, R., Laviano, A., 2019. From old to new: roles of protein sources and individual amino acids in clinical nutrition. Curr. Opin. Clin. Nutr. Metab. Care 22, 58–59.

Elango, R., Humayun, M.A., Ball, R.O., Pencharz, P.B., 2010. Evidence that protein requirements have been significantly underestimated. Curr. Opin. Clin. Nutr. Metab. Care 13, 52–57.

Golden, M.H., 2010. Evolution of nutritional management of acute malnutrition. Indian Pediatr. 47, 667–678.

Jahoor, F., Badaloo, A., Reid, M., Forrester, T., 2008. Protein metabolism in severe childhood malnutrition. Ann. Trop. Paediatr. 28, 87–101.

Manary, M., Callaghan, M., Singh, L., Briend, A., 2016. Protein quality and growth in malnourished children. Food Nutr. Bull. 37 (Suppl. 1), S29–S36.

May, T., de la Haye, B., Nord, G., Klatt, K., Stephenson, K., Adams, S., Bollinge, r L., Hanchard, N., Arning, E., Bottiglieri, T., Maleta, K., Manary, M., Jahoor, F., 2022. One-carbon metabolism in children with marasmus and kwashiorkor. EBioMedicine 75, 103791.

Pencharz, P.B., Elango, R., Wolfe, R.R., 2016. Recent developments in understanding protein needs—how much and what kind should we eat? Appl. Physiol. Nutr. Metabol. 41, 577–580.

Phillip, S.M., Martinson, W., 2019. Nutrient-rich, high-quality, protein-containing dairy foods in combination with exercise in aging persons to mitigate sarcopenia. Nutr. Rev. 77, 216–229.

Rasmussen, B.F., Ennis, M.A., Dyer, R.A., Lim, K., Elango, R., 2021. Glycine, a dispensable amino acid, is conditionally indispensable in late stages of human pregnancy. J. Nutr. 151, 361–369.

Uauy, R., 2013. Keynote: rethinking protein. Food Nutr. Bull. 34, 228–231.

Waterlow, J.C., 2006. Protein Turnover. CABI, Oxfordshire, OX10 8DE, UK.

Wolfe, R.R., 2012. The role of dietary protein in optimizing muscle mass, function and health outcomes in older individuals. Br. J. Nutr. 108 (Suppl. 2), S88–S93.

WHO/FAO/UNU, 2007. Protein and amino acid requirements in human nutrition. Report of a joint WHO/FAO/UNU Expert Consultation. W. H. O. Tech. Rep. Ser. 935, 103–133.

Young, V.R., Marchini, J.S., 1990. Mechanisms and nutritional significance of metabolic responses to altered intakes of protein and amino acids, with reference to nutritional adaptation in humans. Am. J. Clin. Nutr. 51, 270–289.

Protein digestion and bioavailability

Daniel Tomé, UMR PNCA, AgroParisTech, INRAE, Université Paris-Saclay, Paris, France

© 2023 Elsevier Ltd. All rights reserved.

This is an update of Z.A. Bhutta, K. Sadiq, Protein Digestion and Bioavailability, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 116–122, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00240-3.

Introduction 376
Protein digestion in the intestinal lumen 376
Protein digestion in the stomach 377
Protein digestion in the small intestinal lumen 378
Intestinal epithelial digestion and absorption of peptides and amino acids 379
Brush border membrane and cytoplasmic peptidases 379
Absorption of di- and tripeptides 380
Amino acid transport systems 380
First-pass splanchnic extraction 381
Digestion and fermentation in the distal intestine 382
Amino acid catabolism by the gut microflora 382
Amino acid synthesis by the gut microflora 382
Protein digestibility 383
Fecal and ileal digestibility 383
Apparent and true digestibility 384
Differences in protein digestibility 384
Conclusion 385
References 385
Further reading 386

Key points

- After food ingestion, dietary proteins are subjected to digestion in the gastrointestinal tract, and a part of protein entering the gastrointestinal tract daily is also derived from endogenous sources.
- The purpose of protein digestion is the cleavage of protein into smaller fragments constituted by amino acids, dipeptides, and tripeptides so that they can be absorbed as amino acids made available to the organism to support anabolic and catabolic processes.
- Protein digestion starts in the stomach with the actions of pepsins. In the duodenum and the small intestine protein released from the stomach are subjected to hydrolysis by pancreatic proteolytic enzymes (trypsin, chymotrypsin, carboxypeptidase A and B) and intestinal brush-border enzymes.
- Protein digestion produces amino acids and di- and tripeptides absorbed through transporters present in the brush-border membranes of the enterocytes. Small peptides are further hydrolyzed by peptidases within the cytoplasm, so that products that reach the bloodstream are single amino acids.
- The fraction of protein not digested and absorbed in the small intestine reaches the large intestine where amino acids are not quantitatively absorbed but metabolized by the microbiota. They can be directly incorporated into bacterial cells as building blocks of proteins or can enter catabolic pathways.
- Food protein digestion occurs in the stomach and small intestine and produces amino acids absorbed in the small intestine. In the colon unabsorbed amino acids are mostly metabolized by colonic bacteria and converted to ammonia that can be absorbed.
- The intricate and coordinated system of digestion ensures that under normal conditions, approximately 95% of ingested protein in the intestinal lumen is digested and the amino acid absorbed and made available to the organism to support anabolic and catabolic processes.
- As intestinal amino acid absorption is practically complete at the end of the small intestine, oro-ileal digestibility is preferred to oro-faecal digestibility. As digesta contain nitrogen and amino acids from both exogenous and endogenous, true oro-ileal digestibility is corrected by the endogenous losses.
- Protein digestibility is an important component of protein quality. The digestibility of a protein is dependent on the relative ease with which peptide bonds can be hydrolyzed, and on several other factors including the presence of additional dietary factors such as trypsin inhibitors.

Introduction

Proteins represent a significant part of animal and plant tissues and microorganisms and are the main nitrogenous constituents with approximately 16% nitrogen by weight. The basic structural units of proteins are the amino acids which are characterized by the presence of an amine function, an acid carboxyl group and a specific lateral chain. Protein are constituted by 20 amino acids which are linked together by peptide bonds. Units of two or three amino acid residues are called di- or tripeptides, units of less than 100 amino acid residues are called peptides or polypeptides, and units above 100 amino acid residues are called protein. Dietary proteins are the sources of nitrogen and amino acids, including the 9 indispensable amino acids which cannot be synthesized by humans. They are crucial for the synthesis of body tissues proteins, regulatory proteins, and different amino acid-derived metabolites. The digestion of proteins in the gastrointestinal tract involves a coordinated series of events with sequential processes by which ingested protein are progressively hydrolyzed by proteolytic enzymes, leading to the release of amino acids which are absorbed and transferred into the bloodstream. The intricate and coordinated system of digestion ensures that under normal conditions, approximately 95% of ingested protein in the intestinal lumen is digested and the amino acid absorbed and made available to the organism to support anabolic and catabolic processes.

Protein digestion in the intestinal lumen

The daily dietary protein intakes in different adult populations range from 40 to 100 g day^{-1} that constitute on average approximately 10–20% of daily energy intake. After food ingestion, dietary proteins are subjected to digestion in the gastrointestinal tract, and a part of protein entering the gastrointestinal tract daily is also derived from endogenous sources (Fig. 1). The endogenous luminal proteins include salivary, gastric, biliary, pancreatic, and intestinal secretions for approximately 20–30 g day^{-1}, desquamated villus epithelial cells and mucous proteins for an additional 30 g, and a relatively smaller amount (2 g) derived from plasma proteins leaking into the lumen. Exogenous dietary proteins derived from the food consumed and endogenous proteins are mixed

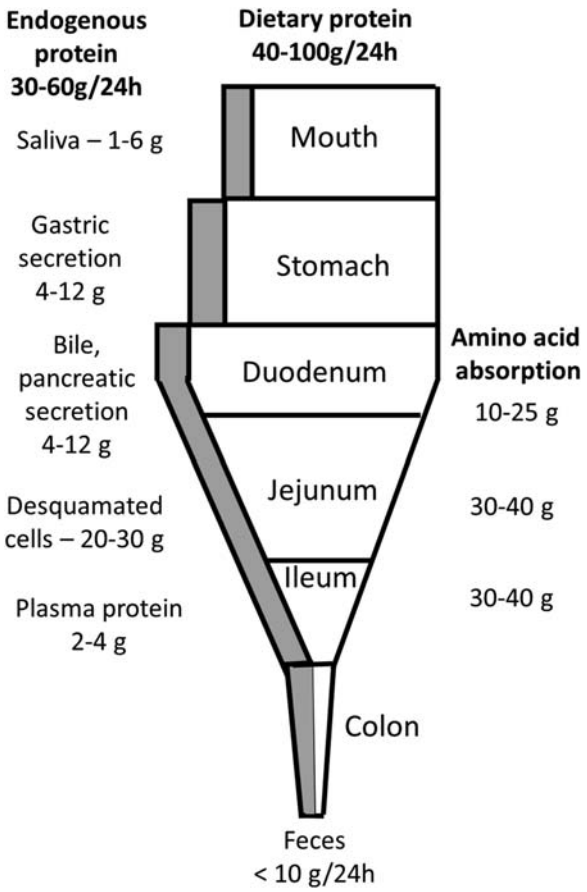


Fig. 1 Protein digestion is a complex process with continuous movements and exchange of protein, amino acids and nitrogen between the gut lumen and the systemic pools.

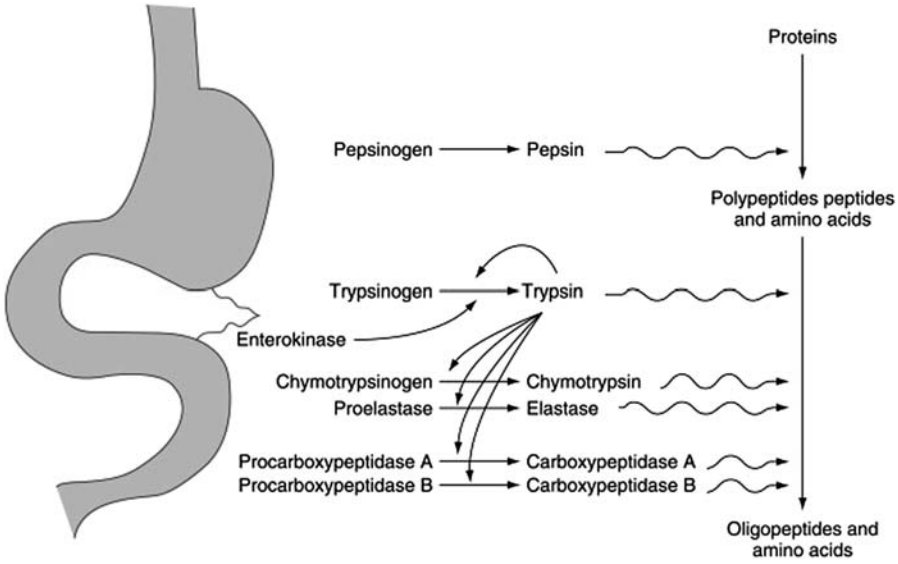


Fig. 2 Cascade of protein hydrolysis in the gastrointestinal tract.

in the intestinal lumen and the daily protein load requiring digestion within the gastrointestinal tract is approximately 100–150 g daily. The purpose of protein digestion is the progressive cleavage of protein into smaller fragments constituted by amino acids, dipeptides, and tripeptides so that they can be absorbed (**Fig. 2** and **Table 1**). Food digestion proceeds into three phases with different regulatory mechanisms, namely the cephalic, gastric, and intestinal phase initiated by the sensory experience of seeing and eating food, the presence of food within the stomach, and the emptying of stomach contents into the small Intestine, respectively.

Protein digestion in the stomach

Protein digestion starts with chewing to mechanically increase the surface area and after swallowing, and contractions of the stomach facilitate gastric mixing and chemical breakdown by gastric acid and pepsins (Bornhorst and Paul Singh, 2014). Digestion

Table 1 Proteolytic enzyme activity in the gastrointestinal tract.

Enzyme	*Precursor*	*Products*	*Catalyst*	*Substrate*	*Action*
Stomach					
Pepsins	Pepsinogens	Polypeptides of diverse sizes and some amino acids	Acid pH	Protein	Hydrolyze bonds between aromatic amino acids (e.g., phenylalanine or amino acid)
Pancreatic proteases					
Trypsin	Trypsinogen	Oligopeptides	Enterokinase	Proteins	Cleaves internal bonds at lysine or arginine amino acids; cleaves other pancreatic proenzymes
			Trypsin	Polypeptides	
Chymotrypsin	Chymotrypsinogen	Oligopeptides	Trypsin	Protein	Cleaves bonds of aromatic or neutral amino acids
				Polypeptides	
Elastase	Proelastase	Oligopeptides	Trypsin	Elastin	Cleaves bonds of aliphatic amino acids (e.g., alanine, glycine, and serine)
				Other proteins	
Carboxypeptidase A	Procarboxypeptidase A	Aromatic amino acids and peptides	Trypsin	Polypeptides at the free *C*-terminal end of the chain	Cleaves aromatic amino acids from *C*-terminal end of protein and peptides
Carboxypeptidase B	Procarboxypeptidase B	Arginine, lysine, and peptides	Trypsin	Polypeptides at the free *C*-terminal end of the chain	Cleaves arginine or lysine from *C*-terminal end of protein and peptides

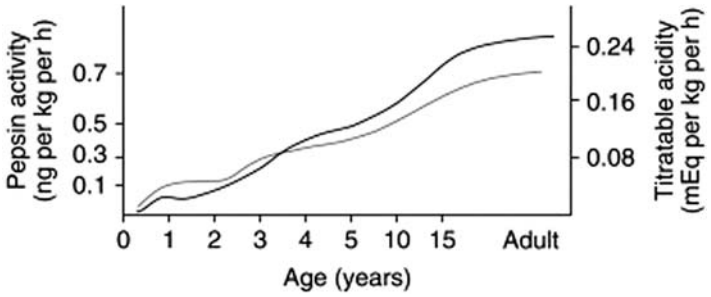


Fig. 3 Postnatal development of gastric acid secretion and titratable acidity. Modified from Koldovsky (1985), with permission from Little Brown and Company.

of proteins in the stomach involves the actions of pepsins, which are secreted as the precursor pepsinogen by the gastric mucosa main cells. The release of pepsinogens is induced by gastrin, histamine, and cholinergic stimulation and pepsinogens are converted to the active form pepsins by the loss of a small basic peptide. Pepsins are most active at a pH of approximately 2 and totally inactive at a pH above approximately 5. So, for pepsins to produce any digestive action on protein, the stomach juices must be acidic. Pepsins have a broad proteolytic specificity, splitting peptide bonds mostly involving phenylalanyl, tyrosyl, and leucyl residues. Immunohistochemistry indicates two distinct forms of pepsinogen, Pepsinogen I only found in acid-secreting regions of the stomach, and pepsinogen II also found in the mucous cells of the oxyntic and pyloric regions of the stomach as well as in the duodenal Brunner's glands. Although these two forms of pepsinogen have slightly different pH optima, their substrate specificity is very similar, and both are rapidly inactivated by the alkaline pH beyond the pylorus. A gelatinase liquefying gelatin is also found in the stomach. The completeness of gastric protein digestion is dependent on several factors, including the rate of gastric emptying, the pH of intragastric contents, and the type of protein ingested. Given the significant buffering capacity of food, it is unlikely that gastric proteolysis plays a major role in protein digestion. This agrees with the observation that neither patients with achlorhydria nor those recovering from major gastric surgery appear to have a major problem with protein digestion. The level of peptic activity and acid production is lower in premature infants and increases in relation to gestational age; pepsin activity increases approximately twofold between infancy and adulthood (**Fig. 3**). There is controversy regarding the presence of rennin (a peptidyl peptide hydrolase) in the stomach of young infants; however, the mild clotting activity in human infants is rapid.

Protein digestion in the small intestinal lumen

The exocrine pancreas secretes an aqueous mixture into the duodenum composed of digestive enzymes, which aid in further degradation of ingested food, and bicarbonate which helps neutralize stomach acid. Neutralization of gastric acid is a critical function of the pancreas as the small intestine mucosa is specialized for nutrient absorption and thus cannot possess a thick protective mucous layer like that of the stomach. Furthermore, pancreatic digestive enzymes are optimally active at basic pH and thus pancreatic bicarbonate secretion is a key requirement for proper digestion.

Pancreatic secretions are primarily regulated by three factors acting synergistically. Acetylcholine released by vagal efferent primarily stimulates synthesis of digestive enzymes by pancreatic acinar cells. Cholecystokinin released by I Cells in the duodenum and jejunum upon entry of food primarily stimulates synthesis of digestive enzymes by pancreatic acinar cells. Secretin released by S Cells of duodenum in response to entry of low pH stomach acid primarily stimulates production of aqueous sodium bicarbonate solution by pancreatic ductal cells, the composition of pancreatic secretions is slightly different during the three phases of food digestion with different regulatory mechanisms.

The cephalic and gastric phases primarily involve vagus nerve stimulation of acinar cells to produce digestive enzymes. Because little aqueous sodium bicarbonate solution is produced by ductal cells these enzymes lie inactive within the pancreatic acini and ducts. By the end of the cephalic and gastric phases, the pancreatic ducts are filled with inactive digestive zymogens ready for washing out into the intestinal lumen by aqueous sodium bicarbonate solution. The intestinal phase involves release of both secretin and cholecystokinin which stimulate the generation of aqueous sodium bicarbonate solution by pancreatic ductal cells that washes out the inactive pancreatic enzymes within the pancreatic ducts into the duodenum where they are activated.

Gastric emptying determines the rate at which the ingested protein is delivered in the duodenum, where segmentation contractions further facilitate luminal hydrolysis by pancreatic proteolytic enzymes, such as trypsin, chymotrypsin, carboxypeptidase A and B, and intestinal brush-border enzymes (Guerra et al., 2012). The pancreatic proteases are secreted as proenzymes and are activated in the lumen. The enteropeptidase (also called enterokinase) released from the brush border membrane removes a hexapeptide from the *N*-terminal end of trypsinogen, converting it to the active form trypsin. Trypsin, in turn, activates the other protease proenzymes and autocatalytically promotes further activation of trypsinogen. The pancreatic endopeptidases trypsin, chymotrypsin, and elastase, primarily split peptide bonds located within the protein molecules resulting in the production of short-chain polypeptides. These are further hydrolyzed by the exopeptidases carboxypeptidase A and B, acting on aromatic/aliphatic C terminals or basic C terminals residues to remove single amino acids, respectively. These pancreatic peptidases cannot hydrolyze peptide bonds with

proline at the *C*-terminal. The product of this coordinated intraluminal digestion by the endopeptidases and exopeptidases is a mixture of neutral and basic amino acids (30%), and peptide with chains varying in length from two to six amino acids (70%). The presence of excess amino acids in the lumen can further limit peptide hydrolysis (product inhibition). The activity of enterokinase is noticeable after 26 weeks of gestation and its activity at term is approximately 10% of that of adults. Although pancreatic trypsin levels are substantial in both preterm and term infants, the secretory response to secretin and pancreozymin stimulation is somewhat blunted at birth compared with that at 2 years of age. However, such comparatively lower levels of protease activity in newborn infants do not appear to limit protein digestion significantly.

Intestinal epithelial digestion and absorption of peptides and amino acids

An important step in the digestion of protein in the small intestine is the cleavage by peptidases at the level of the intestinal brush border (Guerra et al., 2012). Subsequently, amino acids, dipeptides, and tripeptides, are released and taken up across the intestinal mucosa, after which they are absorbed. An important physiological observation is that absorption at the brush-border membrane can occur both as amino acids and as di- and tripeptides. Amino acid absorption is the process of amino acid, dipeptide, and tripeptide uptake from the gastrointestinal lumen. Peptides and amino acids are transported from the lumen to the portal bloodstream through a variety of transporters present in the brush-border and baso-lateral membranes of the enterocytes. Although the end-product of protein digestion is amino acids, small peptides are the dominant form of entry of amino acids into enterocytes, where they are further hydrolyzed into amino acids by peptidases within the cytoplasm and absorbed into the bloodstream (**Fig. 4**). Thus, most products of protein digestion that reach the bloodstream are single amino acids.

Brush border membrane and cytoplasmic peptidases

A range of peptidases are present at the level of the brush border membrane or cytoplasm with the capability of hydrolyzing oligopeptides of up to eight amino acid residues (Tobey et al., 1985) (**Table 2**). These oligopeptidases are synthesized in the rough endoplasmic reticulum of enterocytes and, after transfer through the Golgi apparatus, are transported to the brush border membrane and extruded by exocytosis. There is little posttranslational processing of these peptidases, and they are attached to the brush border membrane by short anchoring pieces. Most oligopeptidases are aminopeptidases, acting at the *N*-terminal amino acid. The brush border proteolysis rate is most rapid for tripeptides and least rapid for dipeptides, whereas the rates of hydrolysis of tetrapeptides and pentapeptides are somewhat intermediate. The brush border peptidases can hydrolyze peptide bonds with proline at the *C*-terminal. Of the cytoplasmic peptidases, the most abundant is a dipeptidase that cleaves neutral dipeptides, whereas the aminotripeptidase has a high specificity toward tripeptides with *N*-terminal amino acids or those containing proline terminally. The brush border peptidases differ in several ways from the cytoplasmic peptidases; the bulk of the hydrolysis of tetrapeptides and longer peptides occurs at the brush order, whereas the converse is true for dipeptidase activity, which is primarily within the

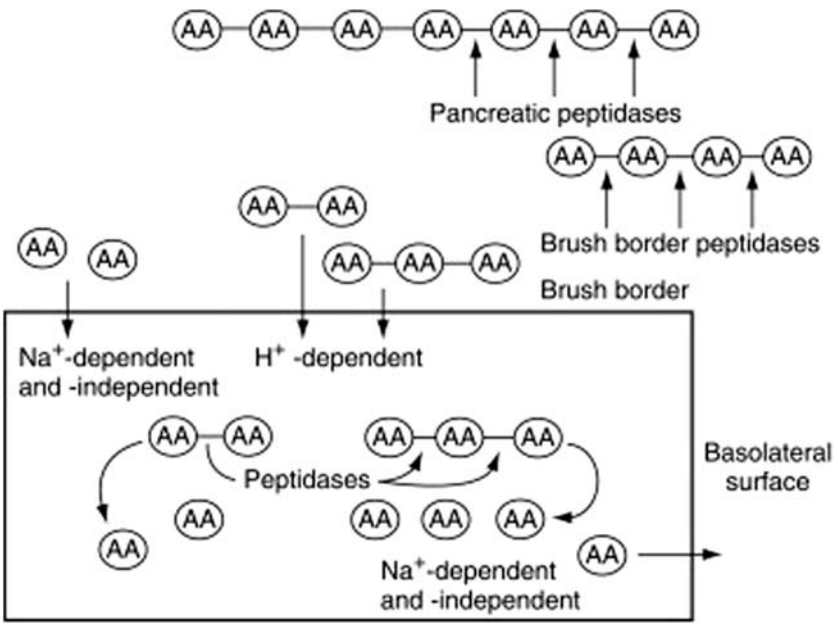


Fig. 4 Small intestinal protein digestion and absorption. Adapted from Shulman (1993).

Table 2 Peptidases present at the brush border membrane and cytoplasm of villous epithelial cells.

Peptidase	*Action*	*Products*
Brush border membrane peptidase		
Amino-oligopeptidases (at least two types)	Cleave amino acids from *C*-terminal of 3–8 amino acid peptides	Amino acids dipeptides
Aminopeptidase A	Cleaves dipeptides with acidic amino acids at *N*-terminal	Amino acids
Aminopeptidase I	Cleaves methionine-containing dipeptides	Amino acids
Aminopeptidase III	Cleaves glycine-containing dipeptides	Amino acids
Dipeptidyl aminopeptidase IV	Cleaves proline-containing peptides with free *C*-terminal	Peptides and amino acids
Carboxypeptidase P	Cleaves proline-containing peptides with free *C*-terminal	Peptides and amino acids
Angiotensin I converting enzyme (ACE) γ-glutamyl transpeptidase	Cleaves γ-glutamyl bonds and transfers	γ-glut amino and/or peptide
Endopeptidases (two, including PABA peptidase)		
Folate conjugase	Cleaves pteroyl polyglutamates	Monoglutamate
Cytoplasmic peptidases		
Endopeptidases (several, including Gly–Leu dipeptidase)	Cleaves most dipeptides	Amino acids
Aminotripeptidase	Cleaves tripeptides	Amino acids
Proline dipeptidase	Cleaves X-Pro bonds in proline-containing dipeptides	Proline and amino acids

PABA, *para*-aminobenzoic acid.

cytoplasm. In general, the cytoplasmic peptidases are more heat labile than brush border peptidases. Very little is known about the developmental aspects of brush border and cytoplasmic proteases. However, the activity of many of these proteases is discernible by 10–16 weeks of gestation and progressively increases during development. In contrast, γ-glutamyl transpeptidase activity decreases with increasing gestational age, but the significance of this transition is unknown.

Absorption of di- and tripeptides

Di- and tripeptides can cross the brush border membrane by a peptide transport system with broad specificity. This transporter (Pept-1) can transport dibasic as well as diacid peptides and peptides consisting of up to three amino acid residues. However, there is some stereospecificity for this transporter because the longer the length of the amino acid side chain on the peptides, the easier the absorption. The transporter system also has greater affinity for dipeptides than tripeptides, and the acidic and basic amino acid residues in dipeptides lower the affinity for the transport system compared with neutral amino acids. In general, the absorption of L-isomers of amino acids in dipeptides is preferred over the D forms. The peptide transport system is coupled to the proton pump system rather than the sodium gradient. The absorption as di- and tripeptides is the major mechanism for absorption of protein-derived amino acids from the human intestinal lumen and is considered a more efficient way of amino acid absorption compared with that of single amino acids (Adibi et al., 1975) (**Fig. 5**). Even when a di- or tripeptide is subject to rapid hydrolysis by brush border peptidases, 30–50% of it is directly absorbed unconverted. The recognition that peptides are the main physiological routes of entry of amino acids into the enterocytes is a point of fundamental importance in the formulation of special protein hydrolyzates and enteral feeds. Several factors may determine the levels of Pept-1, such as insulin, which may stimulate membrane insertion of the transporter from a preformed cytoplasmic pool, and cholera toxin, which decreases the activity of Pept-1 through an increase in the intracellular concentration of cyclic AMP. Once in the absorbing cell, the di- and tripeptides are further hydrolyzed to the constituent amino acids by the cytoplasmic peptidases before absorption. Some small peptides are known to enter the portal blood directly but their relative proportion in comparison to free amino acids is inconsequential. It is also recognized that the intestinal permeability of the preterm and newborn infant may be high, allowing the entry of small amounts of undigested proteins. The maternal antibodies from colostrum can enter the newborn's bloodstream relatively unaltered by a process of endocytosis and subsequent exocytosis. Although the intestinal permeability decreases with age, adults can still absorb larger proteins in abnormal circumstances. However, the predominant form of absorption and presentation of large foreign proteins is through the specialized microfold or M cells overlying the lymphoid Peyer's patches. This mode of absorption of intact proteins or polypeptides, however, is nutritionally insignificant.

Amino acid transport systems

Although some diffusion of amino acids does occur, they are mostly absorbed by active transport systems (Bröer and Fairweather, 2011). Unlike peptides, which are absorbed equally well in both proximal and distal small intestine, amino acids are absorbed more rapidly in the duodenum and jejunum. In contrast to the parsimonious peptide transport system, there are multiple transport mechanisms for various amino acids at both the luminal end and the basolateral membrane of the enterocyte (**Table 3**). At the luminal end, the transporters are mostly located at the villous enterocytes. The villous enterocytes utilize approximately 10% of

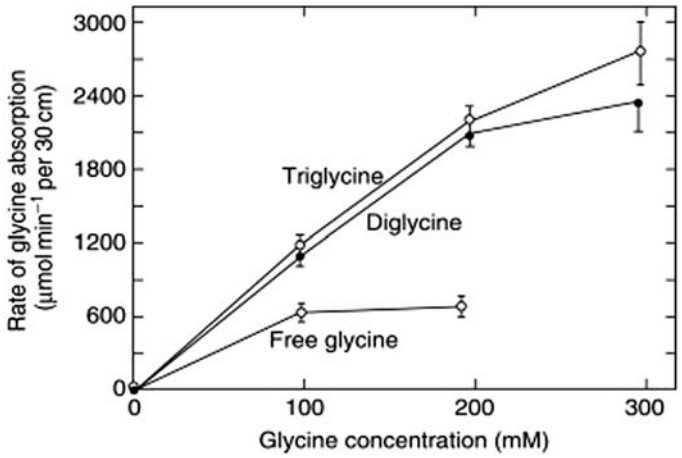


Fig. 5 Rates of glycine absorption (mean ± SEM) from perfusion solutions containing equivalent amounts of glycine in free or peptide form. Reproduced from Adibi et al. (1975), with permission from American Society for Clinical Investigation.

Table 3 Major amino acid transport systems in the intestinal epithelial cells.

Transport system	*Substrates*	*Sodium gradient-dependent*
Brush border membrane		
B	Dipolar α amino acids	+
$B^{0,+}$	Dipolar α amino acids, basic amino acids, cystine	+
$B^{0,+}$	Dipolar α amino acids Basic amino acids, cystine	−
Y^+	Basic amino acids (e.g., lysine), cysteine	−
IMINO	Imino acids (e.g., proline), β-Alanine	+
$X_{G\bar{A}}$	Acidic amino acids (e.g., glutamate, aspartate)	+
B	β-amino acids (e.g., alanine)	+
Basolateral membrane		
L	Broad selectivity	−
A	Broad selectivity	−
ASC	Neutral amino acids (e.g., alanine, serine), cysteine	+
N	Glutamine, histidine, asparagine	+

Modified from Shulman (1993).

the absorbed amino acids for their own protein production, whereas the crypt cells derive their amino acid supply from the portal circulation. There are at least five different sodium-dependent transport systems for amino acid uptake. Most energy-dependent transporters are coupled either to cotransport of Na^+ or Cl^- or to the counter transport of K^+. The sodium-dependent transport can be facilitated by energy derived from Na^+/K^+-exchanging ATPase at the basolateral membrane. An additional system of sodium-independent facilitated diffusion also exists and is predominantly geared toward basic and dipolar α amino acids. These passive transporters are either facilitated transporters or channels. Amino acid transport systems develop *in utero* by the end of the first trimester, whereas peptide transport systems can be demonstrated by the beginning of the second trimester (Lonnerdal, 1994).

First-pass splanchnic extraction

Most of the absorbed amino acids are released into the systemic circulation where they are transported and taken up by peripheral tissues. However, following absorption, a substantial part of the ingested amino acids undergoes first-pass splanchnic extraction, i.e., amino acid uptake and disposal in intestinal and hepatic tissues. A part of amino acids is metabolized during their transcellular transport through the absorptive intestinal cells. This metabolism corresponds to both local utilization of amino acids for protein synthesis and to the production of metabolites from several amino acids including for instance arginine, proline, and cysteine. Some

of these metabolites are used in the intestinal mucosa. It corresponds also to the production of metabolites that are used peripherally outside the intestinal mucosa. For instance, ornithine and citrulline which are not present in proteins can be synthesized in enterocytes from several amino acids present in proteins and play roles in the interorgan metabolism. In addition, the enterocytes use several amino acids (glutamine, glutamate and aspartate) as fuels in the context of a high energy requirement for the cell renewal in the epithelial layer and for nutrient absorption. Of the various amino acids, glutamine appears to have a major role in the nutrition and regeneration of enterocytes, and it is now recognized that in the human intestine the predominant mechanism for assimilation of glutamine dipeptides is absorption as intact dipeptide rather than hydrolysis.

Digestion and fermentation in the distal intestine

The fraction of ingested protein that is not digested and absorbed in the small intestine reaches the large intestine where amino acids are not quantitatively absorbed but metabolized by the microbiota. The contributory role of colonic protein digestion may become particularly important for people with reduced small intestinal function such as short bowel syndrome. Colonic digestion and fermentation are important for energy production. Colonic fermentation may lead to the production of short-chain fatty acids from undigested starch, non-starch polysaccharides, or proteins reaching the colon, providing approximately 5–10% of daily energy requirements. In the first steps of protein catabolism by the intestinal bacteria, these compounds are hydrolyzed by extracellular proteases and peptidases into amino acids and peptides. Provided specific transporters are present, amino acids and peptides are taken up into bacterial cells and undergo different fates. Amino acids can be directly incorporated into bacterial cells as building blocks of proteins or can enter catabolic pathways.

Amino acid catabolism by the gut microflora

The first step in the catabolic fate of amino acids by the gut microflora is transamination or deamination, mostly leading to the corresponding keto acids or saturated fatty acids related to central intermediates that can be easily degraded. Several amino acids released from proteins in the large intestine are precursors for short chain fatty acids synthesis. Many anaerobic bacteria metabolize these amino acids through the fermentation pathways with pyruvate as a central intermediate leading to the production of short chain fatty acids (mainly acetate, propionate, and butyrate), organic acids (mainly formate, lactate and succinate), ethanol and gazes (mainly H2 and CO2). Usually, organic acids do not accumulate since they are rapidly metabolized by other bacterial species to short chain fatty acids. Acetate can be produced by the microbiota from glycine, alanine, threonine, glutamate, lysine, and aspartate, butyrate from glutamate and lysine and propionate from alanine and threonine (Macfarlane and Macfarlane, 2012). The branched-chain fatty acids (isobutyrate, 2-methylbutyrate and isovalerate) are derived from branched-chain amino acids valine, isoleucine, and leucine (Macfarlane, 1991). These branched-chain fatty acids originate exclusively from the breakdown of proteins and are not produced from carbohydrates and are present in lower quantities than short-chain fatty acids in the large intestine content.

Primary amines can be deaminated by the same processes as amino acids and urea is hydrolyzed in carbon dioxide and ammonia. The ammonia generated by the deamination can be utilized as a nitrogen source or excreted (Macfarlane and Macfarlane, 2012). The gazes generated (H2 and CO2) can also be consumed by hydrogenotrophic microorganisms (mainly methanogenic archaea, acetogenic bacteria and sulfate reducing bacteria) to generate methane, acetate (by reductive acetogenesis) and hydrogen sulfide. Acetate can be used as energy source by different types of epithelial cells and by other intestinal bacteria. Sulfides released by the microbiota can be further metabolized by colonocytes while methane is a stable end-product that is not further metabolized in the gut. Amino acids can also be metabolized via decarboxylation leading to the production of amines and polyamines. Deaminases and decarboxylases are amino acid-specific enzymes. In vitro, the production of amino acids deaminases and decarboxylases are favored by alkaline and acidic pH, respectively. Some specific amino acids do not solely undergo these reactions and are the precursors of various metabolic end-products. Tyrosine gives rise to 4-ethylophenol, phenol and p-cresol whereas tryptophane results in the production of indole, skatole, kynunerine. Sulfur-containing amino acids yield to the release of sulfide that can be utilized by colonocytes or directly incorporated in de novo-synthesized amino acids.

Amino acid synthesis by the gut microflora

Bacteria can synthesize de novo some if not all the twenty amino acids required for protein biosynthesis. The human distal gut microbiome is enriched for a variety of clusters of orthologous groups of genes involved in amino acid biosynthesis. The amino acids are formed from metabolic precursors derived from the central metabolism. Pyruvate, oxaloacetate and oxoglutarate are the precursors of 13 of the 20 amino acids and among amino acids, glutamate and glutamine are important intermediates in nitrogen metabolism. The microbiota can recycle amino acids to the host, raising the view that the amino acid exchange between the microbiota and the host can proceed in both directions. The observation that conventional but not germ-free rats can incorporate 15N from 15NH4Cl into body lysine (an amino acid that does not transaminate in mammalian tissues) indicates that 15N-lysine in the host was from microbiota origin (Torrallardona et al., 1996). In the pig, lysine produced by the microbiota is mainly used for protein synthesis in the splanchnic area, intestine and liver. In human volunteers, 15NH4Cl given orally is rapidly incorporated into microbial amino acids and 15N threonine from intestinal microbiota origin appears in the blood plasma even if this

contribution to whole-body threonine metabolism could not be quantified. In human infants, amino acids in plasma can derive from urea after hydrolysis and utilization of nitrogen by the intestinal microbiota but the mechanisms by which amino acids synthetized by the microbiota enter the systemic amino acid pool remains unclear.

Protein digestibility

Protein and amino acid digestibility is a measure of the amount or proportion of dietary protein derived nitrogen and amino acids that are made available to the organism after digestion and absorption. As the capacity of the small intestinal mucosa to absorb amino acids, dipeptides, and tripeptides is generally far greater than the amounts in which they are released in the lumen, protein digestion mainly represents the limiting step in amino acid absorption. Digestibility is determined by measuring the digestive losses at the level of the terminal ileum or in the feces (Fig. 6).

Fecal and ileal digestibility

Dietary protein-derived nitrogen and/or amino acid absorption is assessed by measuring their disappearance in different part of the gastrointestinal. The oldest approach determines nitrogen or amino acid absorption by measuring the difference between oral intake of protein-derived nitrogen or amino acids and nitrogen or amino acids fecal excretion (oro-faecal nitrogen and/or amino acid disappearance).

A major limitation of this oro-faecal approach is that food protein digestion occurs in the stomach and small intestine and produces amino acids absorbed in the small intestine while in the colon amino acids are not absorbed and mostly metabolized by colonic bacteria with the release of ammonia that can be absorbed. While there is some suggestion that amino acids can be absorbed in the large intestine, there is no evidence that this occurs in relevant amounts. The amino acid composition of fecal material is impacted by microbial metabolism in the large intestine, and the oro-faecal disappearance of nitrogen and/or amino acids that include metabolism in the large intestine does not necessarily reflect amino acid absorption by the host organism. To avoid interference from microbial metabolism in the large intestine, the disappearance of ingested nitrogen and/or amino acids is determined at the terminal ileum, i.e., at the end of the small intestine (oro-ileal disappearance). As intestinal amino acid absorption is practically complete at the end of the small intestine, oro-ileal disappearance is considered to represent dietary protein-derived nitrogen amino acid absorption more accurately than oro-faecal disappearance.

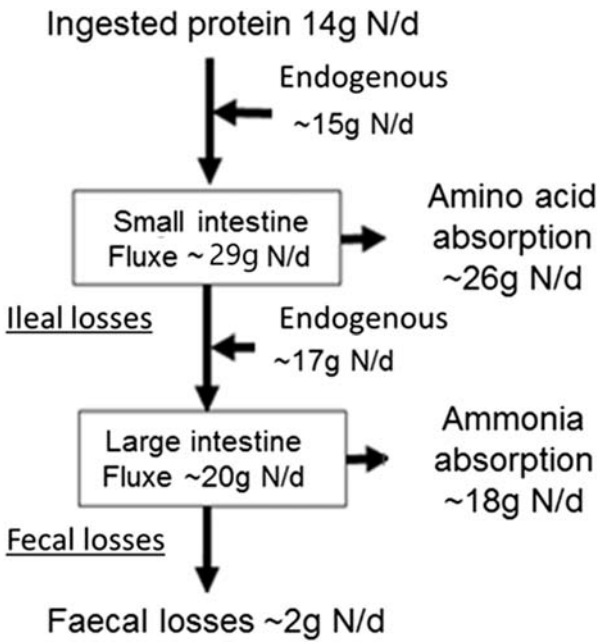


Fig. 6 Digestibility issue of proteins: ileal versus fecal digestibility. Adapted from (WHO, 2007). Digestibility is determined by measuring the digestive losses at the level of the terminal ileum or in the faeces. Food protein digestion occurs in the stomach and small intestine and produces amino acids absorbed in the small intestine. In the colon unabsorbed amino acids are mostly metabolized by colonic bacteria and converted to ammonia that can be absorbed.

Apparent and true digestibility

Another limitation for both fecal and ileal disappearance approaches is that digesta in the feces or terminal ileum contain undigested and/or unabsorbed nitrogen and amino acids from both exogenous and endogenous origin. Moreover, gut endogenous nitrogen and amino acid losses can be distinguished into basal or obligatory and specific losses. The exact estimation of endogenous losses is only possible with isotopic methods. Basal or obligatory losses represent the minimal amount of loss and are not impacted by dietary composition, such as digestive enzyme secretion and epithelial cell turnover, and have been variably estimated to range from 20 mg kg^{-1} day^{-1} in young infant and preschool children to approximately 12 mg kg^{-1} day^{-1} in adults. Specific losses are additional nitrogen and amino acid losses above basal losses that result from stimulation by dietary composition. When oro-fecal or oro-ileal disappearance is not corrected for endogenous losses either in the feces or at the terminal ileum, the terms apparent fecal or ileal digestibility are used. When oro-fecal or oro-ileal disappearance are corrected for basal fecal or ileal endogenous amino acid losses, the terms standardized fecal or ileal digestibility are used. When fecal or ileal digestibility is corrected for total fecal or ileal endogenous losses (i.e., both basal and specific losses), the term true fecal or ileal digestibility is used. The True Ileal Digestibility assay is the best currently available approach to assess amino acid absorption.

$$\text{Apparent digestibility} = [(\text{Dietary} - \text{Total fecal or ileal}) / \text{Dietary}] \times 100$$

$$\text{True digestibility} = [(\text{Dietary} - (\text{Total fecal or ileal} - \text{Total endogenous fecal or ileal})) / \text{Dietary}] \times 100$$

The estimated value of true digestibility of food proteins is dependent on the excretion of gut endogenous nitrogen and amino acid losses. For example, dietary fiber and anti-nutritional factors have been shown to enhance digestive enzyme secretion and epithelial cell turnover and consequently increase specific losses and lowers the true digestibility of the diet protein.

Differences in protein digestibility

Protein digestibility is an important component of protein quality. The amino acid scoring approach has been adopted by FAO/WHO as the preferred index for protein value in human diets (Lee et al., 2016; Shivakumar et al., 2020). The protein digestibility-corrected amino acid score (PDCAAS) and the more recently developed Digestible Indispensable Amino Acid Score (DIAAS) correct the chemical score of the limiting amino acid of a protein by the digestibility of the protein or of the limiting amino acid, respectively. There is strong evidence that true ileal, and not fecal, digestibility is the correct parameter for correction of these amino acid scores (Shivakumar et al., 2020). Uniformly 15N, 13C and/or 2H-labeled dietary proteins from microbes, plants or animal sources are used for assessing ileal digestibility and metabolic utilization of dietary protein-derived amino acids. The direct determination of true ileal nitrogen or amino acid digestibility requires the collection of ileal digesta that can be performed either in rat model, in pig model equipped with a T-cannula, or in human by using naso-ileal intubation methods or by collection of digesta from surgically exteriorized ileum (ileostomates) (Lee et al., 2016; Shivakumar et al., 2020). Minimally-invasive stable isotope signature-based methods for amino acid digestibility were proposed including the indicator amino acid oxidation technique, based on comparison with a mixture of free amino acids, and the dual stable isotope tracer approach. These methods contribute to accumulate values for true protein and amino acid digestibility from human food sources (**Table 4**).

The digestibility of a protein is dependent on the physical shape of the protein and the relative ease with which peptide bonds can be hydrolyzed. Fibrous proteins with long polypeptide chains, such as collagen, keratin, and elastin, are relatively insoluble. In contrast, globular proteins, which are coiled and tightly packed, are comparatively soluble and thereby more digestible. Such proteins are insulin, enzymes, hemoglobin, and albumin. In general, milk and eggs have the highest true digestibility values of approximately 97%, followed by meats, fish, and poultry. Digestibility of protein and amino acid from plant protein sources are usually lower than for animal protein sources and the difference is more important when plant proteins are consumed in the form of complex flour or whole grains. Plant and legumes have protein digestibility values ranging from 75% to 85%. However, some fibrous animal proteins, such as keratin and collagen, are relatively indigestible. A useful approximation is to assume a protein digestibility of 75–80% for diets based on whole grain cereals and vegetables, 95% for diets based on refined cereals and animal proteins, and 85–90% for mixed diets. In the context of infant nutrition, breast milk is extremely well digested, but some proteins, such as secretory IgA, lactoferrin, and α_1-antitrypsin, escape digestion.

Several other factors affect protein digestibility of a diet, including the presence of additional dietary factors such as trypsin inhibitors. The latter may be present in certain foods, such as navy beans and soybeans, and can be largely inactivated by heating, thus improving protein digestibility. Although moderate heating can promote digestibility by promoting breakdown of peptide cross-linkages and inactivation of protease inhibitors in natural food, strong heating, especially in the presence of a carbohydrate or oxidized lipids, may make the protein resistant to enzymatic hydrolysis. The Maillard or browning reaction occurs after high, usually prolonged heating of a protein in the presence of a reducing sugar such as lactose or glucose, resulting in cross-linkages of the sugar with the free side chain of the lysine residues. This may make up to 30% of the lysine biologically unavailable. These changes are of particular importance in situations of marginally sufficient protein intake, in which cooking procedures may further aggravate protein malnutrition. The Maillard reaction is highly influenced by the pH of foodstuffs or other agents. The reduction of pH that may be performed by fermentation in the baking industry lessens the decomposition of lysine and tryptophan in proteins. Fermentation is widely used as a strategy to increase the digestibility of starch and improve the organoleptic properties of weaning foods and cereal-based preparations in developing countries. However, although the impact of fermentation on starch digestion is

Table 4 Illustrative values of protein digestibility in humans.

Protein sources	*True fecal digestibility (%)*	*Protein source*	*True ileal digestibility %*
Eggs	97	Egg	80–97
Milk and cheese	95	Milk	94–99
Meat and fish	94	Meat	80–99
Corn	85	Corn	60–70
Soy flour	86	Soy	75–90
Soybean isolate	94		
Whole wheat	86	Wheat	65–85
Refined wheat	96		
Polished rice	88	Rice	65–85
Peanut butter	95	Peanut	80–90
Beans	78	Beans	75–78
Oatmeal	86	Green pea	70–90
Millet	79	Rapeseed	80–84
Chinese mixed diet	94	Chickpea	57–75
Brazilian mixed diet	78	Yellow pea	72
Guatemalan mixed diet	79	Lupine	80–90
Indian rice and milk diet	87	Mung beans	63
Mixed American diet	96		

Adapted from Torún (1985), WHO FAO UNU (2007), Tome (2012), Gaudichon et al. (2002), Oberli et al. (2015), Fromentin et al. (2012), Kashyap et al. (2018, 2021), Tessier et al. (2020), Devi et al. (2020).

well established, the impact on protein digestibility is variable. Although some studies have suggested an impact on protein quality and digestibility of legumes and finger millet-based foods, other studies suggest that fermentation only modifies the gastric emptying rate and does not significantly affect the level of diet hydrolysis, the endogenous nitrogen stimulation, or the digestibility rate.

Conclusion

Digestion of protein is a complex process with numerous exchanges of protein, amino acid, and nitrogen along the gastrointestinal tract. It is more usually an efficient process leading to the absorption of more than 90% of the ingested protein as amino acid made available to the organism to support anabolic and catabolic process. The efficiency is due for a part to the cascade of numerous and different proteolytic enzyme action and specificity that progressively hydrolyze protein to the end-products small peptides and amino acids. Moreover, the capacity of the small intestinal mucosa to absorb amino acids, dipeptides, and tripeptides is very high and the efficiency of absorption is amplified by the simultaneous transfer of free amino acid and di and tripeptides. In these processes, the limiting step is the digestion of food in the intestinal lumen and as consequence protein digestibility is identified as an important component of food protein quality.

References

Adibi, S., Morse, E.L., Masilamani, S., Amin, P., 1975. Evidence for two different modes of tripeptide disappearance in human intestine. Uptake by peptide carrier systems and hydrolysis by peptide hydrolases. J. Clin. Invest. 56, 1355–1363.

Bornhorst, G.M., Paul Singh, R., 2014. Gastric digestion in vivo and in vitro: how the structural aspects of food influence the digestion process. Annu. Rev. Food Sci. Technol. 5, 111–132.

Bröer, S., Fairweather, S.J., 2011. Amino acid transport across the mammalian intestine. Compr. Physiol. 9, 343–373.

Devi, S., Varkey, A., Dharmar, M., et al., 2020. Amino acid digestibility of extruded chickpea and yellow pea protein is high and comparable in moderately stunted South Indian children with use of a dual stable isotope tracer method. J. Nutr. 150, 1178–1185.

Fromentin, C., Sanders, P., Nau, F., et al., 2012. A pilot study for the intrinsic labeling of egg proteins with 15N and 13C. Rapid Commun. Mass Spectrom. 26, 43–48.

Gaudichon, C., Bos, C., Morens, C., et al., 2002. Ileal losses of nitrogen and amino acids in humans and their importance to the assessment of amino acid requirements. Gastroenterology 123, 50–59.

Guerra, A., Etienne-Mesmin, L., Livrelli, V., Denis, S., Blanquet-Diot, S., Alric, M., 2012. Relevance and challenges in modeling human gastric and small intestinal digestion. Trends Biotechnol. 30, 591–600.

Kashyap, S., Shivakumar, N., Varkey, A., et al., 2018. Ileal digestibility of intrinsically labeled hen's egg and meat protein determined with the dual stable isotope tracer method in Indian adults. Am. J. Clin. Nutr. 108, 980–987.

Kashyap, S., Shivakumar, N., Sejian, V., et al., 2021. Goat milk protein digestibility in relation to intestinal function. Am. J. Clin. Nutr. 113, 845–853.

Koldovsky, O., 1985. Digestion and absorption of carbohydrates, protein and fat in infants and children. Nutr. Pediatr. 253–277.

Lee, W.T., Weisell, R., Albert, J., Tomé, D., Kurpad, A.V., Uauy, R., 2016. Research approaches and methods for evaluating the protein quality of human foods proposed by an FAO Expert Working Group in 2014. J. Nutr. 146, 929–932.

Lonnerdal, B., 1994. Digestibility and absorption of protein in infants. In: Protein Metabolism During Infancy, vol. 33. Raven Press, Vevey, pp. 53–65.

Macfarlane, G.T., Macfarlane, S., 2012. Bacteria, colonic fermentation, and gastrointestinal health. J. AOAC Int. 95, 50–60.
Macfarlane, G.T., 1991. The colonic flora, fermentation and large bowel digestive function. In: The Large Intestine: Physiology, Pathophysiology and Disease, pp. 51–92.
Oberli, M., Marsset-Baglieri, A., Airinei, G., et al., 2015. High true ileal digestibility but not postprandial utilization of nitrogen from bovine meat protein in humans is moderately decreased by high-temperature, long-duration cooking. J. Nutr. 145, 2221–2228.
Shivakumar, N., Jackson, A.A., Courtney-Martin, G., Elango, R., Ghosh, S., Hodgkinson, S., Xipsiti, M., Lee, W.T., Kurpad, A.V., Tomé, D., 2020. Protein quality assessment of follow-up formula for young children and ready-to-use therapeutic foods: recommendations by the FAO Expert Working Group in 2017. J. Nutr. 150, 195–201.
Shulman, R., 1993. Intraluminal digestion and absorption in the small intestine. In: Perinatal and Pediatric Pathophysiology: A Clinical Perspective. Hodder & Stoughton, Boston.
Tessier, R., Khodorova, N., Calvez, J., et al., 2020. 15N and ^{2}H intrinsic labeling demonstrate that real digestibility in rats of proteins and amino acids from sunflower protein isolate is almost as high as that of goat whey. J. Nutr. 150, 450–457.
Tobey, N., Heizer, W., Yeh, R., Huang, T.-I., Hoffner, C., 1985. Human intestinal brush border peptidases. Gastroenterology 88, 913–926.
Tome, D., 2012. Criteria and markers for protein quality assessment-a review. Br. J. Nutr. 108, S222–S229.
Torrallardona, D., Harris, C.I., Coates, M.E., Fuller, M.F., 1996. Microbial amino acid synthesis and utilization in rats: incorporation of 15N from 15NH4Cl into lysine in the tissues of germ-free and conventional rats. Br. J. Nutr. 76, 689–700.
Torún, B., 1985. Proteínas: química, metabolismo y requerimientos nutricionales, es [Proteins: Chemistry, Metabolism and Nutritional Requirements].
WHO, FAO, UNU, 2007. Protein and Amino Acid Requirements in Human Nutrition: Report of a Joint WHO/FAO/UNU Expert Consultation;[Geneva, 9–16 April 2002]. WHO, Geneva.

Further reading

Abubucker, S., Segata, N., Goll, J., et al., 2012. Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS Comput. Biol. 8, e1002358.
Adeola, O., Xue, P., Cowieson, A., et al., 2016. Basal endogenous losses of amino acids in protein nutrition research for swine and poultry. Anim. Feed Sci. Technol. 221, 274–283.
Adibi, S.A., 2003. Regulation of expression of the intestinal oligopeptide transporter (Pept-1) in health and disease. Am. J. Physiol. Gastrointest. Liver Physiol. 285, G779–G788.
Backes, G., Hennig, U., Petzke, K.J., et al., 2002. Contribution of intestinal microbial lysine to lysine homeostasis is reduced in minipigs fed a wheat gluten–based diet. Am. J. Clin. Nutr. 76, 1317–1325.
Barker, H., 1981. Amino acid degradation by anaerobic bacteria. Annu. Rev. Biochem. 50, 23–40.
Blachier, F., Boutry, C., Bos, C., et al., 2009. Metabolism and functions of L-glutamate in the epithelial cells of the small and large intestines. Am. J. Clin. Nutr. 90, 814S–821S.
Bos, C., Juillet, B., Fouillet, H., et al., 2005. Postprandial metabolic utilization of wheat protein in humans. Am. J. Clin. Nutr. 81, 87–94.
Brodin, B., Nielsen, C.U., Steffansen, B., et al., 2002. Transport of peptidomimetic drugs by the intestinal di/tri-peptide transporter, PepT1. Pharmacol. Toxicol. 90, 285–296.
Castagna, M., Shayakul, C., Trotti, D., et al., 1997. Molecular characteristics of mammalian and insect amino acid transporters: implications for amino acid homeostasis. J. Exp. Biol. 200, 269–286.
Chassard, C., Bernalier-Donadille, A., 2006. H2 and acetate transfers during xylan fermentation between a butyrate-producing xylanolytic species and hydrogenotrophic microorganisms from the human gut. FEMS Microbiol. Lett. 254, 116–122.
Chen, L., Li, P., Wang, J., et al., 2009. Catabolism of nutritionally essential amino acids in developing porcine enterocytes. Amino Acids 37, 143–152.
Darragh, A.J., Hodgkinson, S.M., 2000. Quantifying the digestibility of dietary protein. J. Nutr. 130, 1850S–1856S.
Drasar, B., 1974. Some factors associated with geographical variations in the intestinal microflora. Soc. Appl. Bacteriol. Symp. 187–196.
Elsden, S.R., Hilton, M.G., 1978. Volatile acid production from threonine, valine, leucine and isoleucine by clostridia. Arch. Microbiol. 117, 165–172.
Fuller, M.F., Tomé, D., 2005. In vivo determination of amino acid bioavailability in humans and model animals. J. AOAC Int. 88, 923–934.
Gaudichon, C., Mahe, S., Luengo, C., et al., 1996. A 15N-leucine-dilution method to measure endogenous contribution to luminal nitrogen in the human upper jejunum. Eur. J. Clin. Nutr. 50, 261–268.
Gill, S.R., Pop, M., DeBoy, R.T., et al., 2006. Metagenomic analysis of the human distal gut microbiome. Science 312, 1355–1359.
Gorissen, S.H., Trommelen, J., Kouw, I.W., et al., 2020. Protein type, protein dose, and age modulate dietary protein digestion and phenylalanine absorption kinetics and plasma phenylalanine availability in humans. J. Nutr. 150, 2041–2050.
Gottschalk, G., 1979. Nutrition of bacteria. In: Bacterial Metabolism. Springer.
Hellström, P.M., Grybäck, P., Jacobsson, H., 2006. The physiology of gastric emptying. Best Pract. Res. Clin. Anaesthesiol. 20, 397–407.
Hopfer, U., 1987. Membrane transport mechanisms for hexoses and amino acids in the small intestine. In: Physiology of the Gastrointestinal Tract, vol. 2, pp. 1499–1526.
Macfarlane, G., 1995. Microbiological aspects of the production of short-chain fatty acids in the large bowel. In: Physiological and Clinical Aspect of Short-Chain Fatty Acids, pp. 88–106.
Mailliard, M.E., Stevens, B.R., Mann, G.E., 1995. Amino acid transport by small intestinal, hepatic, and pancreatic epithelia. Gastroenterology 108, 888–910.
Metges, C.C., Petzke, K.J., El-Khoury, A.E., et al., 1999. Incorporation of urea and ammonia nitrogen into ileal and fecal microbial proteins and plasma free amino acids in normal men and ileostomates. Am. J. Clin. Nutr. 70, 1046–1058.
Metges, C.C., 2000. Contribution of microbial amino acids to amino acid homeostasis of the host. J. Nutr. 130, 1857S–1864S.
Miller, T.W., Wang, E.A., Gould, S., et al., 2012. Hydrogen sulfide is an endogenous potentiator of T cell activation. J. Biol. Chem. 287, 4211–4221.
Millward, D.J., Forrester, T., Ah-Sing, E., et al., 2000. The transfer of 15N from urea to lysine in the human infant. Br. J. Nutr. 83, 505–512.
Mimoun, S., Andriamihaja, M., Chaumontet, C., et al., 2012. Detoxification of H2S by differentiated colonic epithelial cells: implication of the sulfide oxidizing unit and of the cell respiratory capacity. Antioxid. Redox Signal. 17, 1–10.
Moehn, S., Bertolo, R.F., Pencharz, P.B., et al., 2005. Development of the indicator amino acid oxidation technique to determine the availability of amino acids from dietary protein in pigs. J. Nutr. 135, 2866–2870.
Mortensen, P.B., Clausen, M.R., 1996. Short-chain fatty acids in the human colon: relation to gastrointestinal health and disease. Scand. J. Gastroenterol. 31, 132–148.
Moughan, P.J., Schuttert, G., 1991. Composition of nitrogen-containing fractions in digesta from the distal ileum of pigs fed a protein-free diet. J. Nutr. 121, 1570–1574.
Moughan, P.J., Wolfe, R.R., 2019. Determination of dietary amino acid digestibility in humans. J. Nutr. 149, 2101–2109.
Nordgaard, I., Mortensen, P.B., 1995. Digestive processes in the human colon. Nutrition 11, 37–45.
Nordgaard, I., Mortensen, P., Langkilde, A.M., 1995. Small intestinal malabsorption and colonic fermentation of resistant starch and resistant peptides to short-chain fatty acids. Nutrition 11, 129–137.
Rasmussen, H.S., Holtug, K., Mortensen, P., 1988. Degradation of amino acids to short-chain fatty acids in humans: an in vitro study. Scand. J. Gastroenterol. 23, 178–182.
Shahkhalili, Y., Finot, P.A., Hurrell, R., et al., 1990. Effects of foods rich in polyphenols on nitrogen excretion in rats. J. Nutr. 120, 346–352.
Smith, E.A., Macfarlane, G.T., 1998. Enumeration of amino acid fermenting bacteria in the human large intestine: effects of pH and starch on peptide metabolism and dissimilation of amino acids. FEMS Microbiol. Ecol. 25, 355–368.
Stein, H., Seve, B., Fuller, M., et al., 2007. Committee on terminology to report AAB, et al. Invited review: amino acid bioavailability and digestibility in pig feed ingredients: terminology and application. J. Anim. Sci. 85, 172–180.
Stephen, A.M., Cummings, J., 1980. The microbial contribution to human faecal mass. J. Med. Microbiol. 13, 45–56.

Torrallardona, D., Harris, C., Milne, E., et al., 1995. Site of absorption of lysine synthesised by the gastrointestinal microflora of pigs. In: Publication-European Association for Animal Production, vol. 81, pp. 369–372.
Trommelen, J., Tomé, D., van Loon, L.J., 2021. Gut amino acid absorption in humans: concepts and relevance for postprandial metabolism. Clin. Nutr. Open Sci. 36, 43–55.
van der Wielen, N., Moughan, P.J., Mensink, M., 2017. Amino acid absorption in the large intestine of humans and porcine models. J. Nutr. 147, 1493–1498.
Volpi, E., Mittendorfer, B., Wolf, S.E., et al., 1999. Oral amino acids stimulate muscle protein anabolism in the elderly despite higher first-pass splanchnic extraction. Am. J. Physiol. Endocrinol. Metab. 277 (3).
Watford, M., Lund, P., Krebs, H.A., 1979. Isolation and metabolic characteristics of rat and chicken enterocytes. Biochem. J. 178, 589–596.

Proteostasis and turnover

D.J. Millward, University of Surrey, Guildford, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of D.J. Millward, Protein: Synthesis and Turnover, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 139–146, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00243-9.

Introduction 388
Whole-body proteostasis during growth and development 389
Protein turnover and its regulation 391
Protein synthesis 391
Proteolysis 392
Models and tracer methods for the study of protein turnover 394
Extent and physiological implications of protein turnover 396
The diurnal cycle of feeding and fasting and postprandial protein utilization 397
Conclusion 399
References 399
Further reading 400

Key Points

- Whole-body proteostasis during growth and development is achieved through both within-tissue responses to daily anabolic and catabolic influences of the feeding and fasting cycle and through longer-term interactions within a protein-stat mechanism, involving dietary protein-mediated long-bone length-growth regulating muscle growth through a passive stretch mechanism, with the growth of most other tissues reflecting their metabolic work.
- Protein turnover occurs within a proteostasis network within which protein synthesis and proteolysis systems are the main nutritionally sensitive components.
- Protein synthesis reflects its capacity, provided by the ribosomal and tRNA apparatus necessary to translate mRNA molecules, and its efficiency in terms of the rate of translation per ribosome.
- Nutritionally sensitive proteolysis involves both lysosomal-autophagy and ubiquitin proteasome systems, each of which are capable of selective proteolysis.
- The mTORC1 complex mediates acute nutritional control of protein synthesis and proteolysis, with amino-acids and insulin initiating signal transduction, mediating nucleolar ribosomal RNA synthesis and assembly, cytoplasmic activation of mRNA translation, and inhibition of proteolysis.
- Whole-body and tissue protein-turnover can be studied with isotope tracer techniques measuring both the amino-acid flux through the plasma amino-acid pool or the α-amino pool by ^{15}N-end product methods, and the turnover of tissue proteins.
- Protein turnover in the human adult, about 300 g/day, accounts for a significant fraction of the RMR, a minimum of 14%.
- The balance between changes in proteolysis, protein synthesis and amino-acid oxidation with feeding determines the efficiency of postprandial protein-utilization, with the most efficient mechanism involving the inhibition of proteolysis with minimal increases in protein synthesis and amino-acid oxidation, and with amino-acid signaling initiated from the extracellular pool.

Introduction

About 20.4 K putative protein-coding genes are predicted by the human genome of which about 9 K are expressed in all tissues suggesting these include housekeeping proteins needed to maintain basic cellular structure and function, i.e., ribosomal and associated proteins involved in protein synthesis, enzymes essential for cell metabolism and gene expression, and mitochondrial proteins needed for energy generation, as well as proteins responsible for the structural integrity of the cell. Proteins also vary greatly in abundance, from fewer than 50 copies per cell in the case of certain transcription factors to more than 10 million molecules for histones, cytoskeletal, ribosomal, and myofibrillar proteins. Protein abundance must be carefully controlled to support cell signaling and the proper flux of substrates through metabolic pathways and to allow the stoichiometric assembly of large macromolecular machines, such as ribosomes, the mitochondrial respiratory chain and the myofibrillar contractile apparatus. Proteostasis is the maintenance of an intact proteome within cells, acting through the proteostasis network, (PN) which exerts strict control of the initial production and folding of a protein, its conformational maintenance, control of abundance and subcellular localization, and disposal by degradation (Klaips et al., 2018; Hipp et al., 2019; Pey, 2020) see **Fig. 1**. Various molecular chaperones and co-chaperones are of central importance ensuring correct de novo folding and maintenance of a soluble, nonaggregated state as

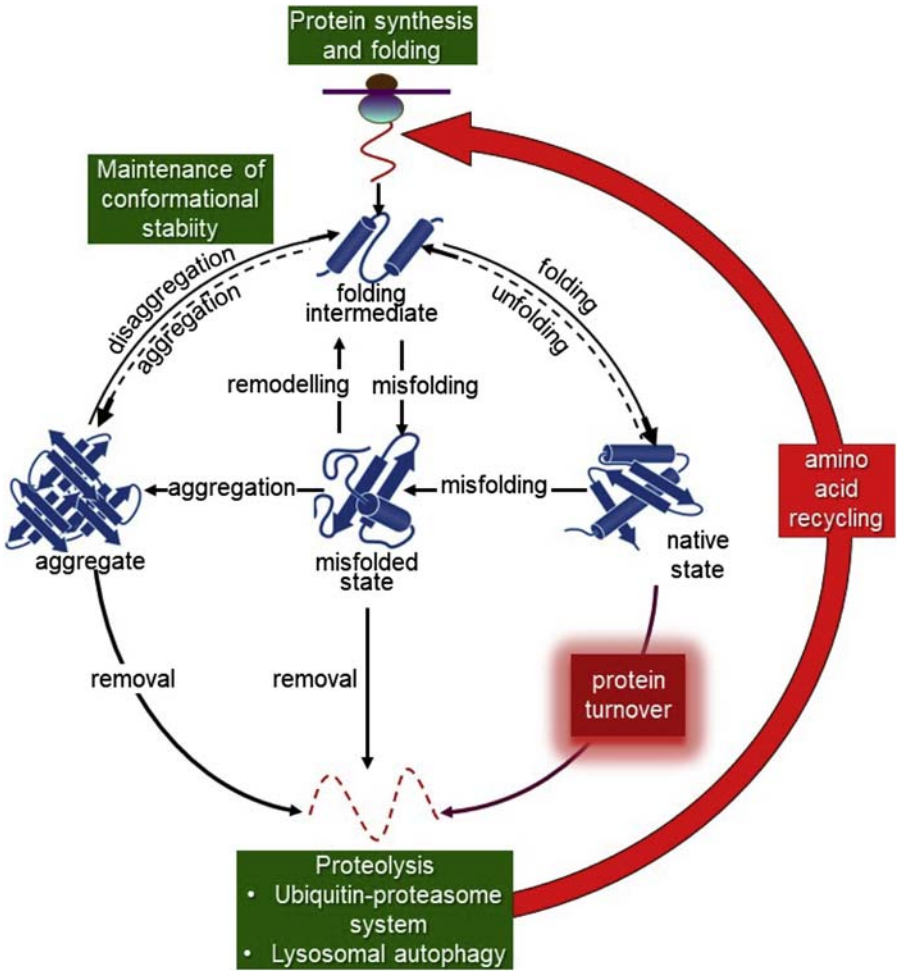


Fig. 1 Scheme showing the proteostasis network in relation to protein turnover. Production of functional proteins requires not only their synthesis by the ribosomal apparatus but also their correct folding and assembly. The proteostasis network achieves this and also minimizes non-productive misfolding or formation of harmful aggregates through remodeling and removal by proteolysis by the ubiquitin-proteasome and lysosomal autophagic systems. Proteolytic removal of native state proteins, identified here as protein turnover, also occurs at variable rates according to both the structural and physicochemical properties of the protein which influence their susceptibility to proteolysis and the type of tissue within which it occurs in terms of the relative abundance of proteolytic systems. Amino acids released by proteolysis are recycled for protein synthesis. It is estimated that the proteostasis network comprises ~2000 proteins in human cells, ~400 associated with protein synthesis and folding, ~300 with maintenance of conformational stability and >1000 with protein breakdown, of which ~300 are molecular chaperones which orchestrate the processes shown.

well as targeting misfolded proteins for degradation or spatial sequestration to protect the rest of the proteome from aberrant interactions. Some of the PN is concerned with environmental stress responses (thermal, pathogen and oxidative stress) which can be triggered and coordinated by proteostasis transcription factors and it is believed that the accumulation of protein aggregates such as occurs in the brain with aging represents a failure of proteostasis. Thus protein turnover, in its classical context of the degradation and resynthesis of cellular proteins can now be described as the consequence of the action of the ~2000 components of the PN specifically involved in protein synthesis, chaperone-mediated folding and proteolysis mainly by the lysosomal-autophagic system, (LAS) and ubiquitin–proteasomal systems (UPS). The term proteostasis will be extended here to also include the whole body, in terms of the regulation of and relationships between proteomes of different tissues and organs.

Whole-body proteostasis during growth and development

The growth of an individual in height and overall shape is mainly a function of bone growth which is genetically determined but subject to a favorable diet and hazard-free environment. During the initial few months of life food intake occurs at very frequent intervals, but after this growth and development adapts to a diurnal cycle of day-time feeding and nocturnal fasting during sleep. This means from early-life the nutritional regulation of whole-body proteostasis during growth and development responds to not only the background diet but also to the diurnal-changes in overall protein balance mediated by the acute daily responses to feeding

and fasting. Such changes have been investigated in detail in relation to the skeletal–muscle mass and the viscera, especially liver. However less is known about the acute nutritional sensitivity of the proteome of organs and tissues like the brain and nervous system although both human and primate studies have shown postnatal brain growth in size and structure is nutritionally sensitive. Certainly neurons contain an active PN which is important in limiting the accumulation of protein aggregates in the aging brain and the UPS in brain has been associated with memory.

The whole body proteome, the fat-free lean-body mass, (FFM), comprises two major compartments, skeletal muscle, accounting for about 40% FFM in young children increasing to over 50% in young adults and exhibiting a sexual dimorphism from puberty, and the non-muscle FFM. In both children and adults, muscle is highly regulated at a fixed maximum genotypic-size relative to overall stature unlike the non-muscle FFM which can expand with increasing protein and food intakes and adiposity (Millward, 1995, 2021). Regulation of the whole body proteome has been described by a protein-stat mechanism, the central feature of which is an interaction between linear growth of bone, protein deposition in skeletal muscle and dietary protein intake (see **Fig. 2**). The first controller is nutritional, the minimum dietary protein intake required to exert a sufficient anabolic-drive, mediated by indispensable amino-acids and the endocrine system acting both on linear bone-growth to allow the genetic programming of its postnatal growth trajectory to be expressed, and on skeletal–muscle to maintain its mass at a specific size set by the linear dimensions of the organism. The second controller is internal, namely linear bone-growth which mediates muscle myofibre-growth capacity, with a passive stretch, mechanotransduction mechanism, acting on muscle stem, (satellite), and other cells which provide extra myonuclei enabling myofibre growth and which modify the architecture of the muscle extracellular matrix (ECM) enveloping each myofibre. The ECM can be visualized as connective tissue "bags" which define maximum myofibre volume and potential protein content, so that myofibre expansion during growth, i.e., "bag" enlargement and "bag" filling is dependent on ECM remodeling (see Millward and Smith, 2019; Millward, 2021). The bone-muscle interaction is two way, in that subsequent to bone length growth driving muscle growth, the force on bone of muscle contraction drives bone strength and mineralization via a mechanostat mechanism. The growth of most other organs is secondary to this main interaction, determined primarily by the level of food and protein intake and the consequent metabolic work of, and functional demand for, the organ and is not specifically limited in size.

The third controller is appetite, controlled by an amino-static appetite mechanism, sensing the effect of the flow of amino-acids into muscle to satisfy the metabolic demand for growth, on the circulating amino-acid pool (Millward, 1995). This allows food protein intake to match capacity for muscle growth and is most obvious in catch-up growth in children and may also accompany saltatory growth-spurts in children. Muscle growth ceases in the absence of passive stretch when bone length-growth ceases at the end of puberty, requiring stretch through resistance-exercise to induce any further growth in adulthood. Such a mechanism would mean that any reduction in bone-length will result in loss of muscle-mass and this may be one explanation of sarcopenia, the loss of

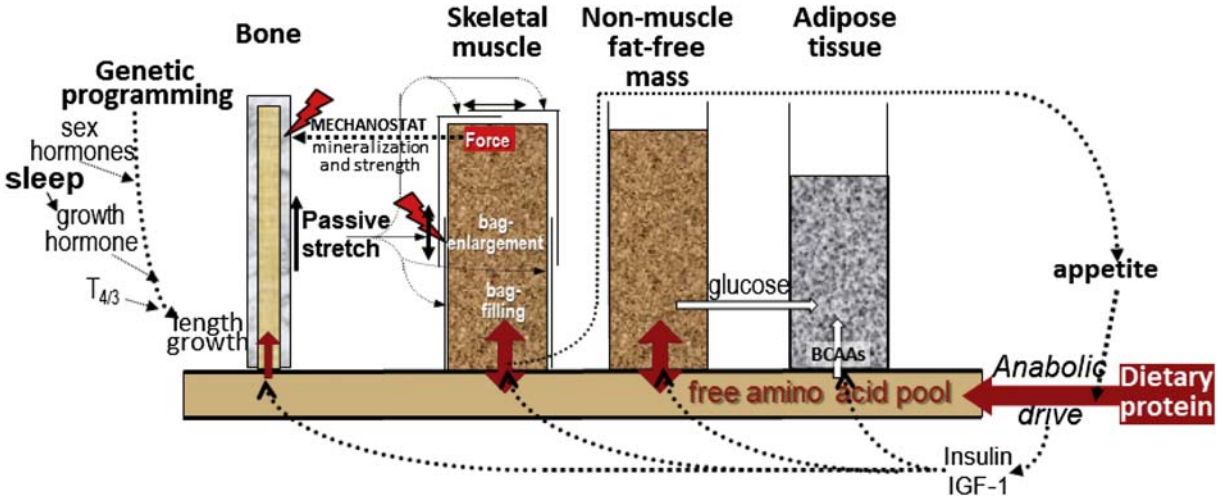


Fig. 2 Scheme showing protein-stat mechanism for coordinated regulation of whole body proteostasis during growth and homeostasis. Whole-body protein content is controlled through an amino-static appetite mechanism, acting primarily to maintain skeletal-muscle mass at a level set by the linear dimensions of the organism. Bone lengthening occurs at rates determined by genetic programming (canalization) mediated by growth hormone, thyroid hormones and by the sex steroids during puberty, acting together with an appropriate anabolic drive deriving from dietary protein and mediated by amino acids, insulin, IGF-1 (and in the rat T3) and other important nutrients like zinc and vitamin D, and provides the regulatory stimulus for protein deposition in all tissues. Net protein deposition during skeletal muscle growth occurs within myofibres which are limited in volume by the extracellular matrix of connective tissue which surrounds individual and groups of myofibres like concentric "bags". Thus, increased myofibre diameter and length during muscle growth requires remodeling of the muscle ECM to enable "bag" enlargement which is mediated by the passive stretching of skeletal muscle subsequent to bone length growth, activating satellite and other cells through mechanotransduction mechanisms. The linkage of bone length to muscle mass allows muscle size to be regulated at a genotypic muscle weight–bone length ratio. Increasing muscle size and force generation acts via the mechanostat to increase bone mineralization and strength commensurate with muscle mass. Any dietary protein-derived amino acid intake in excess of that required for maximal "bag filling" will either expand the non-muscle lean body mass or be oxidized with the carbon skeletons leaving the liver as ketones and glucose, the latter to be taken up in adipose tissue for lipogenesis together with excess branched chain amino acids. Modified from Millward (2021).

appendicular muscle in old age. Although long-bone shortening in old age has not been described in detail, some shortening could accompany loss of mineral mass with osteoporosis. Given the sensitivity of muscle-mass to changes in bone length, (muscle mass varies with length^4), it may be that only very small reductions in length are sufficient to mediate measurable sarcopenia.

Evidence for the limitation of muscle growth in the absence of ECM remodeling comes from in vivo studies of muscle protein synthesis in adults during nutritional stimulation by amino-acids which show a tachyphylaxis while the nutritional stimulation is still present, i.e., the development of an inhibition following the initial stimulation. This is consistent with the inability of the muscle fiber proteome to expand beyond its regulated "bag-full" level though protein feeding alone (Millward and Smith, 2019; Millward, 2021).

Protein turnover and its regulation

Within the PN, protein turnover is mediated by genomic transcription, ribosomal translation, folding and conformational maintenance of the protein's native state and disposal by proteolysis.

Protein synthesis

This can be discussed in terms of its capacity, i.e., ribosomal and tRNA necessary to translate mRNA molecules, and its efficiency in terms of the rate of translation. This is operationally measurable as protein synthesis per unit RNA, and occurs at a rate of about 4 amino-acids/second when fully active and is regulated by nutritional and mechanical factors as discussed below. The PN includes many factors controlling the faithfulness of ribosomal mRNA translation and subsequent protein folding. Nevertheless protein synthesis and folding is imperfect despite the presence of abundant chaperone systems, and it has been estimated that 5–30% of all newly synthesized proteins do not properly fold and need to be targeted for immediate degradation (see Hipp et al., 2019).

Overall, the rates of protein synthesis and degradation within most cells are balanced precisely to avoid rapid growth or atrophy of cellular mass, (proteostasis), although for muscle there is the additional constraint imposed by the ECM on maximum myofibre volume. Fine control is achieved through highly complex multiple signaling systems with signal transduction involving receptor mediated hormone action, cytokines, metabolites which provide energy, and amino-acids, the latter acting both as substrates for and specific activators of protein synthesis. The mTOR complex is at the center of the regulatory network in most cells and comprises a highly conserved serine/threonine kinase which phosphorylates, (activating or inhibiting), other proteins involved in signal transduction pathways regulating cell growth and metabolism. Of the two structurally and functionally distinct multiprotein complexes, mTORC1 and mTORC2, mTORC1 is the important control-center for the nutritional regulation of protein turnover, integrating signals from insulin, amino-acids and the energy status of the cell all of which change after food intake, promoting protein synthesis and inhibiting proteolysis. Thus in the postabsorptive state mTORC1 is inhibited in part because of the low energy state within tissues. The energy state of the cell is sensed as AMP levels by AMP-activated protein kinase (AMPK) which inactivates mTORC1. Recent work has also shown that cyclic AMP levels are a major mechanism for mTORC1 inactivation. Thus adenylyl cyclase, which forms cyclic AMP can be activated through G-protein coupled receptors, which respond to glucagon for example, and cyclic AMP activates a kinase (protein kinase A) which inhibits mTORC1 (Jewell et al., 2019). Even in the presence of amino acids increased cAMP inhibits mTORC1.

Food intake reverses this inhibition of mTORC1. Thus in muscle for example, a meal providing protein and energy which increases plasma amino-acids and insulin, transforms the net catabolic postabsorptive state of decreased protein-synthesis and increased proteolysis, into postprandial anabolism through the combination of amino-acids stimulating protein synthesis, and insulin, inhibiting proteolysis (see Millward and Smith, 2019). The limitation on muscle myofibre expansion by the ECM within the protein-stat model shown in **Fig. 2**, is observable in terms of a "bag-full" signal when the elevation of protein synthesis in muscle by a meal or by amino-acids is terminated after a short period even though amino-acid levels in muscle remain elevated. For the whole body proteome as indicated by measurements of the plasma flux, (as discussed below), feeding responses appear to be less influenced by muscle and more by the splanchnic organs with feeding-induced changes in proteolysis more marked than protein synthesis.

Recent studies of amino-acid sensing mechanisms in a variety of cell types have revealed a complex (not entirely understood) system in which mTORC1 is activated at the lysosomal surface by leucine, arginine, S-adenosylmethionine and glutamine (Jewell et al., 2019; Takahara et al., 2020), see **Fig. 3** which are sensed by both cytosolic and lysosomal amino-acid specific sensors. In fact it appears that a total of 10 amino-acids, (alanine, arginine, asparagine, glutamine, histidine, leucine, methionine, serine, threonine, and valine), can promote mTORC1 activity. However of these amino-acids, leucine is the most rapidly acting. The amino-acid transport system participates in this regulation not only delivering amino-acids as substrates for protein synthesis and activators of mTORC1, but also initiating signal transduction like hormone receptors: i.e., acting as multifunctional "amino-acid transceptors". Thus the Na+/Gln cotransporter in muscle appears to be able to directly signal to mTOR during its transport of Gln into the muscle cell. This is one way in which changes in amino-acids in the EC pool can signal through mTORC1 to activate protein synthesis.

Downstream of mTORC1, p70-S6 kinase (S6K) and 4E-binding protein (4E-BP) are the main effectors of regulation of protein synthesis which is necessarily complex at both transcriptional and translational levels. Transcriptional control of ribosome concentrations in tissues determines the capacity for protein synthesis and in this way controls the overall tissue protein turnover rate and the changes associated with postnatal development. Animal studies have indicated that cellular ribosome concentrations can

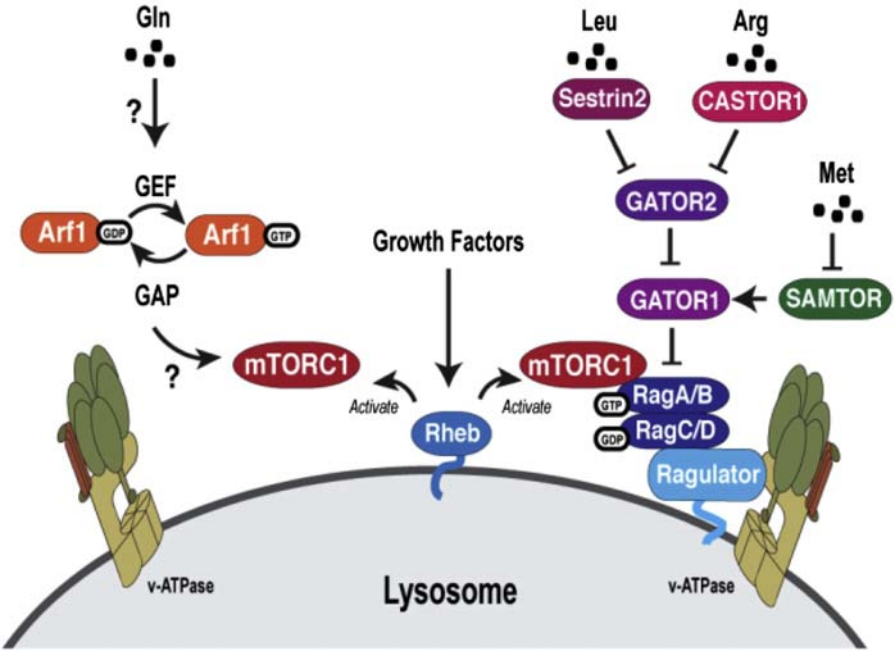


Fig. 3 Scheme showing amino acid activation of mTORC1. Glutamine activates mTORC1 through a Rag GTPase-independent pathway involving Arf 1 (left). Leucine, arginine and methionine activate mTORC1 through a Rag GTPase-dependent pathway (right). From Jewell et al. (2019) copied under the terms of the Creative Commons Attribution License.

change acutely, e.g., during the diurnal cycle of feeding and fasting, and chronically in response to the composition of the background diet, increased functional demand, and hormones such as insulin, thyroid, growth hormone, and the glucocorticoids. Furthermore these influences are tissue specific with glucocorticoids for example increasing hepatic ribosome concentrations, (as part of the hepatic acute phase response), and decreasing ribosome concentrations in muscle. In contrast thyroid hormones increase ribosomes and proteolytic enzymes in both muscle and liver in association with a generalized increase in protein turnover.

Transcriptional control of specific gene expression is rapid and extensive to the extent that changes in specific mRNA concentrations have become surrogate measures of acute changes in rates of synthesis of specific proteins, as for hepatic export-protein synthesis. Thus the down regulation of albumin synthesis in response to lack of dietary protein or during the pro-inflammatory cytokine-mediated acute-phase response is mediated largely through reductions in mRNA for albumin and other hepatic export proteins and increases in mRNA for acute phase proteins.

Initiation of translation is a key target process with reversible phosphorylation known to regulate at least four separate steps of the initiation cycle enabling very rapid changes in protein synthesis. Animal studies indicate that insulin, glucocorticoids, and amino-acids have all been implicated in such regulation. However there is uncertainty still about the detail with major differences between the mechanisms observed in the young rapidly-growing compared with the adult animal. Thus in skeletal muscle in the young rat, an insulin-mediated stimulation of translation occurs while in adult human muscle, insulin is relatively ineffective with amino-acid levels the main stimulatory influence (Millward and Smith, 2019).

A summary of key steps in the regulation of muscle protein synthesis in terms of ribosomal capacity and efficiency by nutritional and mechanical factors is shown in Fig. 4.

Proteolysis

Of the six major proteolytic systems in tissues, lysosomal, Ca2+-dependent, caspase-dependent, ubiquitin–proteasome system, (UPS), metalloproteinases and the mitochondrial Lon protease), the lysosomal-autophagic system, (LAS), and UPS are each under nutritional control and are likely to account for most of whole body protein turnover (WBPT) (see Dikic, 2017).

The LAS, is present in all cells and involves acid hydrolases including proteinases, (cathepsins), within a distinct vacuolar structure. Autophagy-related genes control the LAS which removes misfolded proteins, damaged organelles, and pathogens and native proteins and which includes 3 major categories of autophagy: microautophagy, macroautophagy, and chaperone-mediated autophagy as shown in Fig. 5. Microautophagy and macroautophagy are mainly non-selective although the processes can mediate the selective degradation of mitochondria and ubiquitylated aggregates under particular conditions. Chaperone-mediated autophagy is the selective pathway involving recognition of soluble proteins with a specific amino-acid sequence motif common to many proteins, by a cytoplasmic heat shock protein, HSC70 enabling the substrate unfolding and translocation to the lysosomal lumen via a transmembrane lysosomal associated protein structure (LAMP2A).

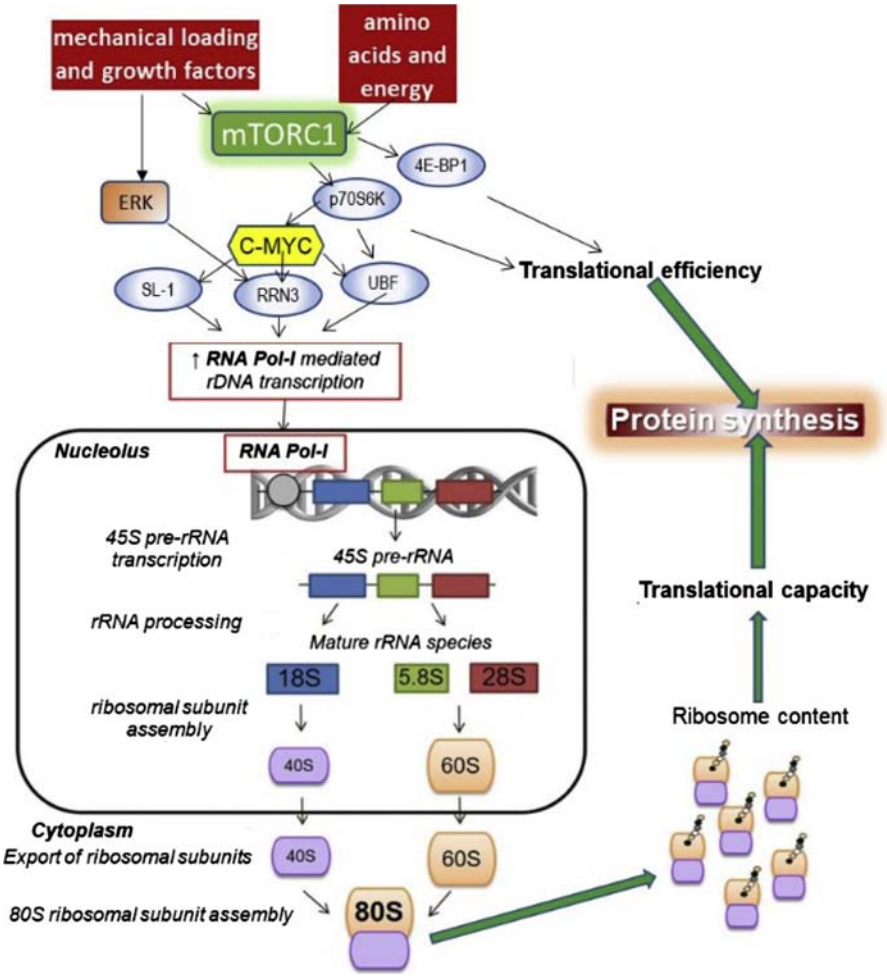


Fig. 4 Flow diagram of regulation of muscle protein synthesis in terms of ribosomal capacity and efficiency by nutritional and mechanical factors. Multiple signaling pathways including extracellular signal-regulated kinase (ERK) and mTORC1 mediate activation of the nucleolar RNA Polymerase I (Pol-I) by a number of factors which form a transcriptional complex, the "Pol I regulon", at the rDNA promoter. The elevated expression of the proto-oncogene transcription factor C-MYC, achieved through mTORC1/S6K1 plays a central role stimulating the Pol II-dependent transcription of a cohort of factors associated with Pol I regulon including RNA polymerase I-specific transcription initiation factor RRN3 and upstream binding factor (UBF). Ribosome biogenesis involves transcription of the 45S ribosomal RNA precursor (45S pre-rRNA), processing of the 45S pre-rRNA into the smaller rRNAs (18S, 5.8S and 28S rRNAs) followed by assembly of these rRNAs and other ribosomal proteins into ribosomal subunits (40S and 60S) which are exported into the cytoplasm. The 80S ribosome is the mRNA translating unit, requiring mTORC1-mediated activation of initiation through p70S6K (p70 kDa ribosomal protein subunit kinase 1) and 4E-BP1 (eukaryotic initiation factor 4E binding protein 1). mTORC1 also regulates RNA Pol-II (which transcribes the ribosomal proteins and other proteins required for transcription) and Pol-III, (which synthesizes small structural RNAs such as 5S rRNA and transfer RNA (tRNA) (not shown). Modified from Millward (2021) under a Creative Commons Attribution 4.0 International License.

The UPS (see Fig. 6) is also widely distributed among tissues. It consists of a cascade of enzymes that activate with ATP, conjugate, and ligate the 76-amino-acid protein ubiquitin onto a lysine residue of a target protein. Because ubiquitin contains several lysine residues itself, both linear or branched chains of polyubiquitin can be assembled. Chains coupled to different ubiquitin lysine residues act as signaling motifs for proteolysis by the 26S proteasome (as well as other processes including the LAS). The 26S particle is a giant multifunctional protease complex consisting of a 19S regulatory particle which includes the "lid" to the 20S core particle within which proteolysis occurs. This recognizes, unfolds, and degrades ubiquitinated proteins in an ATP-dependent process. The molecular determinants of selective proteolysis, i.e., which proteins are targeted for ubiquitination are defined by the E3 ubiquitin ligase enzymes, of which ~700 different species are encoded in the human genome. The 26S proteasome catalyzes proteolysis to peptides averaging about 8 amino-acids long and proteolysis of these peptides is achieved by soluble peptidases.

Thus both the LAS and the UPS can achieve selective turnover and can be assumed to account for most of WBPT although the relative importance of the two systems for WBPT in different tissues has yet to be quantified. As far as their nutritional regulation, early studies focused on hepatic macroautophagy through which hepatic protein mass appeared to be regulated by both insulin and a receptor-mediated, amino-acid dependent inhibitory process, each acting through mTORC1 (Ueno et al., 2012; Meijer et al., 2015). Less is known about which is the dominant system in muscle although some suggest that the LAS is a crucial pathway

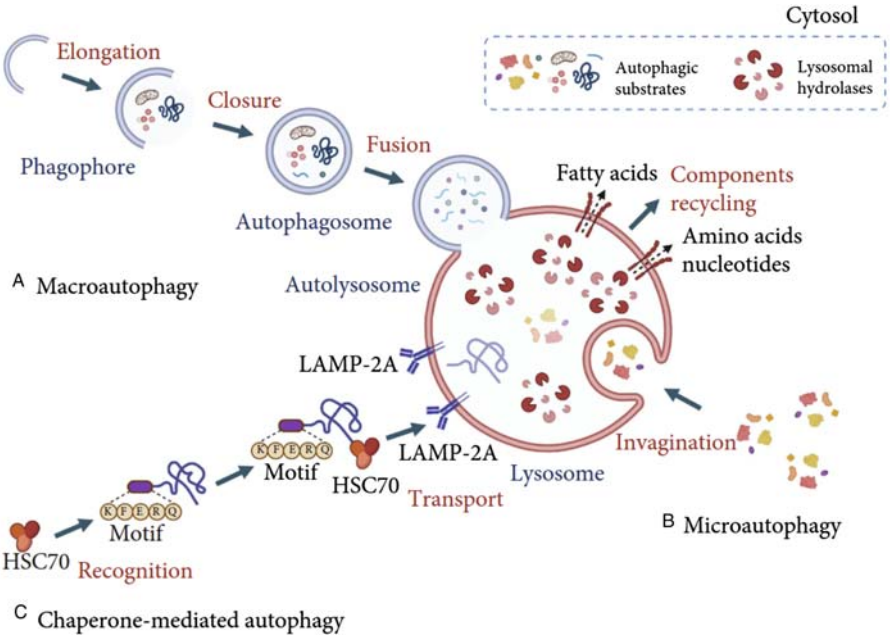


Fig. 5 Scheme showing the three key types of autophagy in eukaryotic cells: macroautophagy, microautophagy, and chaperone-mediated autophagy. Each type delivers cargoes of substrates to the lysosomes for degradation through the different pathway prior to releasing their components as amino acids, fatty acids, and nucleotides back into the cytosol for cellular reuse. Copied from Xu et al. (2021) under a Creative Commons Attribution License.

for muscle mass maintenance with its regulation different from that in liver in that unlike rapid and transient activation of the LAS by starvation in liver, in muscle such induction is sustained (Combaret et al., 2016). The ubiquitin-proteasome system is activated in skeletal muscle during fasting and following glucocorticoid treatment, supporting a role in the physiological regulation of protein turnover. Activation can be achieved by increasing rates of ubiquitination and by phosphorylation of the 19S subunit Rpn6 by protein kinases (VerPlank et al., 2019). This can occur in physiological conditions which increase cAMP and activate protein kinase A including in human muscles following intense exercise, in mouse muscles and liver after a brief fast and in hepatocytes after epinephrine or glucagon. These are the same conditions, (increased cAMP), under which the LAS is stimulated through inhibition of mTORC1 (see Jewell et al., 2019). Thus it may be that in the fasted state, the increased proteolysis in muscle reflects activation of both the LAS and the UPS via similar, coordinately regulated mTORC1-related mechanisms (Zhao and Goldberg, 2016). Another important but poorly understood control mechanism involves changes in cell volume. Thus, swelling acts like a proliferative anabolic signal, inhibiting proteolysis while cell shrinkage is catabolic, stimulating proteolysis. These effects have been shown in liver and there is some evidence for such a mechanism in skeletal muscle (see Millward, 2021). What is clear is that the nutritional regulation of proteolysis in muscle is mainly mediated by insulin, which mediates the post-prandial inhibitory response independently of amino acid availability (see Millward and Smith, 2019).

Models and tracer methods for the study of protein turnover

Study of protein turnover has utilized isotope tracer techniques, radioactive tracers (^{3}H, ^{14}C and ^{35}S) in animals and stable isotopes (^{13}C, ^{15}N, and ^{2}H) in both animal and human studies. Whole-body protein-turnover, (WBPT), can be measured simply as a function of the amino-acid flux through the plasma amino-acid pool during a primed continuous intravenous infusion of trace-labeled amino-acid, e.g., ^{13}C-1 labeled leucine. The infused tracer isotopic enrichment is diluted by unlabeled amino-acid from proteolysis (D) and the diet (I). At isotopic equilibrium (i.e., a plateau or constant degree of labeling of the tracee), the tracer/tracee ratio in relation to the infusion rate, (i), indicates the flux, (Q), which is the appearance into or irreversible loss from the plasma leucine pool. Appearance, Ra, is partitioned into intake, (I), from diet (and tracer infusion if it is significant), and entry from proteolysis of body protein, (D), (i.e., no de novo synthesis of leucine occurs). Loss or disappearance, Rd, is partitioned into protein synthesis, (S), and oxidative catabolism, (O), as $^{13}CO_2$, calculated from measurement of the production of labeled $^{13}CO_2$ in the breath and the labeling of the leucine or its keto acid in the plasma.

$$\text{i.e.,} \quad Q = \text{i} \times \text{tracer dilution} = I + D = S + O$$

This allows the components of protein turnover, (D and S), to be calculated as $S = Q - O$ and $D = Q - I$. With the free leucine pool relatively small and turning over rapidly, isotopic equilibrium can be reached in 2–4 h when an initial priming dose is given as a bolus injection. Using the leucine content of tissue proteins, rates of leucine appearance and loss can be converted into rates of

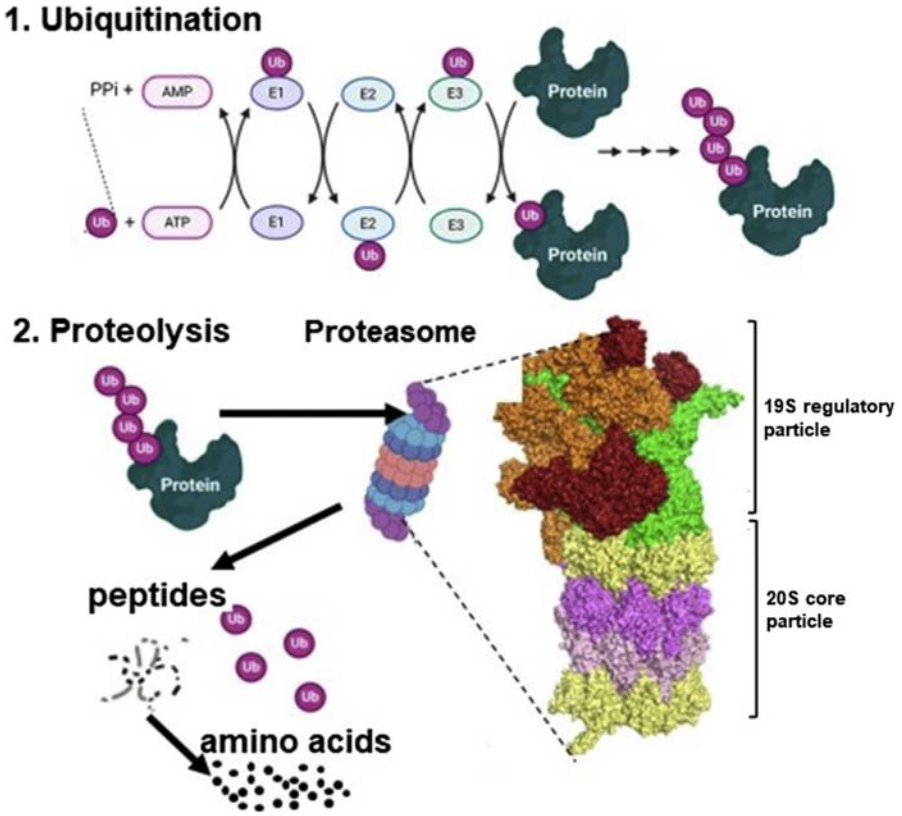


Fig. 6 Scheme showing the ubiquitin-proteasomal system for proteolysis. The Ubiquitin Proteasome System. (1) Ubiquitin, 8.5 kDa, is thiolated to ubiquitin-activating enzyme (E1) and subsequently transthioesterified to a cognate ubiquitin-conjugating enzyme (E2, of which 40 different species are coded in the human genome). E3 ubiquitin ligases (~700 types in two major classes), recognize specific target proteins and facilitate linkage of ubiquitin to the ε-amino group of target lysine residues. (2) Subsequent turnover of ubiquitinated proteins is mediated by the 26S proteasome. This comprises the 19S regulatory particle, composed of lid (orange), ubiquitin receptors (red) and ATPases (green) which sit on top of the heptameric rings (yellow), that constitute the entry portal to the 20S catalytic chamber (magenta). Ubiquitin is released and recycled and peptides released are further degraded to amino acids by soluble peptidase enzymes. Modified from Scholz et al. (2020) under a Creative Commons Attribution License.

whole-body protein synthesis and proteolysis. ^{13}C-1 leucine is especially useful since its essentiality enables D to be calculated; it has a small pool enabling equilibrium to be achieved in a short period; decarboxylation is the first irreversible step in its catabolism releasing $^{13}CO_2$ quantitatively; and its intracellular transamination product α-ketoisocaproate, (α-KIC), rapidly equilibrates with its plasma pool and can be sampled to serve as a measure of the labeling of the intracellular leucine precursor pool. This is particularly important since the problem of definition of isotopic enrichment of the actual precursor amino-acid pool for protein synthesis is the most serious problem in these studies. During leucine infusions the α-KIC enrichment at equilibrium is lower than that of leucine so that the plasma flux calculated from α-KIC enrichment is higher (≈25%).

When tissues can be sampled, as in animal experiments or with muscle biopsies for example, tracer incorporation into protein will indicate tissue rates of protein synthesis, and a variety of approaches have been exploited to study regulation of tissue protein mass. In animal studies, a "flooding large dose" measurement of protein synthesis was introduced in 1980 using L-[$^3H_4$]phenylalanine which proved much more convenient than previous constant intravenous infusions. This enables all tissue phenylalanine precursor pools to become equally labeled very quickly allowing protein synthesis rates in individual tissues to be assessed from isotope uptake into protein during short periods after the dose. While the large dose of [^{3}H] phenylalanine was shown not to influence the rate of protein synthesis in the original rat studies, problems arose when the method was adopted for human use. Thus large doses of [^{13}C] leucine, and several other essential amino-acids were shown to stimulate the rate of protein synthesis whereas the non-essential amino-acids, (arginine, glycine, serine and proline), did not. Thus the flooding dose method with [^{13}C] or [^{15}N] proline was successfully applied to the study of collagen synthesis in muscles, bones, tendons, and ligaments (Millward and Smith, 2019).

Quantification of rates of proteolysis is especially problematical. In animal studies proteolysis can be estimated from measured rates of synthesis in non-growing tissues or, during growth, (G) of tissue protein as $D = S - G$, and with careful design, proteolysis rates can be measured by this method over relatively short periods when protein mass is changing quickly such as during stretch-induced growth of skeletal muscle or endotoxin-induced catabolism. While urinary excretion rates of the post-translationally modified amino-acid 3-methyl histidine was proposed as a measure of myofibrillar protein degradation, since it is excreted quantitatively

in the urine, the main source was shown to be small, rapidly turning-over pools, most likely actin-microfilaments of the cytoskeleton present in most cell-types. This invalidates this approach as far as whole body studies of skeletal muscle turnover are concerned. Nevertheless its release from incubated or perfused muscle can be used to determine myofibrillar protein degradation.

Simultaneous determination of protein synthesis and degradation can be made, in principle at least, from organ tracer balance studies, (i.e., measurements of concentrations and isotopic enrichments of tracer amino-acids across tissues like leg or forearm that have been combined with measurements of 3-methyl-histidine release). Such studies in humans have identified an inhibitory effect of insulin on muscle proteolysis and stimulatory influences of amino-acids on muscle protein synthesis.

All of these methods allow study of turnover of individual amino-acids in protein and measurement of their nonprotein metabolic fate (e.g., oxidation). A different approach is to use ^{15}N glycine, often given orally as a single dose, to study overall amino-nitrogen turnover. Because of nitrogen exchange between amino-acids by transamination, this label acts as a tracer for total free amino nitrogen. The whole body nitrogen flux is estimated from the relative proportion of administered tracer excreted in the end product, which is assumed equal to the relative proportion of the flux which is excreted as total N. This is then resolved into protein synthesis and proteolysis from measurements of N intake and excretion. While simple in concept this approach is metabolically complicated with two urinary end products of nitrogen metabolism, urea and ammonia, each deriving from different pathways and each giving different flux values, explaining why it is used infrequently.

Finally approaches involving heavy water (^2H_2O) have recently re-emerged (see Millward and Smith, 2019), as a suitable tracer to study muscle protein metabolism. Following a bolus of ^2H_2O, ^2H exchange occurs with H atoms in multiple biological substrates including amino-acids. Thus with alanine for example, of its four carbon-hydrogen bonds, an average of 3.7 are exchanged with ^2H, enabling precursor product approaches to measurement of muscle protein synthesis with ^2H_2O and protein bound alanine. Orally administered ^2H_2O, which is relatively inexpensive, has obvious advantages enabling volunteers to be studied in free-living conditions, over extended time periods, providing a cumulative measure of for example myofibrillar protein synthesis measured over 12 days. The potential for measurement of turnover with ^2H_2O in a wide range of macromolecules including ribosomal RNA is an exciting feature of ^2H_2O studies. Also proteome dynamics can be explored employing the recent advances in LC-MS-MS instrumentation coupled with either ^2H_2O studies, uniformly deuterated valine [$^2H_8$]valine or [$^{13}C_6$] lysine to study individual protein turnover rates within the entire proteome in vivo (Hammond et al., 2016; Claydon et al., 2012), or ^{13}C6 arginine and ^{13}C6lysine to study the proteome in cultured dermal fibroblasts isolated from different mammalian species with life spans ranging from 3 to 200 years (Swovick et al., 2021).

The choice of method must depend on the questions asked and circumstances of the subjects under study. ^{13}C carbon labeling is more suited to short-term (e.g., 3 or 4 h–24 h) clinical measurements for which frequent blood and breath sampling is possible. Thus, the efficiency and mechanisms of post-prandial protein utilization during meal feeding has been studied by means of ^{13}C-1 leucine balance and turnover measurements. ^{15}N methods are more suitable for free living subjects and patients, when urine sampling is possible but regular blood and breath sampling are inconvenient. The most famous examples are use of this method in an unassisted Antarctic crossing, and in the space station. Both methods involve many assumptions but in practice the two approaches have been shown to give similar results.

Extent and physiological implications of protein turnover

Early studies of the variability of protein turnover rates between tissues and for different proteins in the same tissues (Waterlow et al., 1978; Bates et al., 1983) demonstrated quite clearly that for an individual protein, its turnover reflected the proteolytic environment within which it occurred, identified as coarse control, with fine control involving individual susceptibility to proteolysis. Thus in the rat, mean half-lives of whole tissue proteins ranged from <1 day for liver to >10 days for skeletal muscle with comparable differences for mixed mitochondrial proteins. Also for mixed myofibrillar-proteins extracted from different muscles half-lives varied between 5.5 days in heart, 9–23 days in different skeletal muscle types in the rat, and 4.5 and 12.8 days in slow and fast contracting muscle in fowl. Within myofibrils actin turnover is half that of myosin heavy chain (MHC). One aspect of the variation in "proteolytic environment" was shown to be cell or nuclear domain size, as indicated by the protein/DNA ratio, with turnover rates highest for small cells and falling as cell or nuclear-domain size increased. This is the case for liver, (small cell), and skeletal muscle, (larger cell), and for muscle during post natal development with the protein/DNA ratio increasing markedly but reaching an equilibrium adult value which is a function of the muscle fiber type. The protein/DNA ratios are higher and turnover is proportionately slower in fast-twitch, compared with slow-contracting tonic muscles. These relationships mean that overall rates of proteolysis per diploid nucleus (per unit DNA) do not vary much but the interaction of the proteolytic systems with their cellular protein substrates does vary. For the liver, its function appears to be best served, (for reasons which are by no means clear), by its dominant lysosomal-autophagic system degrading its cellular proteins rapidly and replacing them. In fact although hepatocyte sizes do vary this involves variation in polyploidy so that protein/DNA ratios do not change. For muscle, its function, speed and duration of its contractions, and its resulting ultrastructure; i.e., highly branching slow tonic muscles or non-branching fast twitch muscles, as seen in classical examples of the anterior and posterior latissimus dorsi muscles of the fowl, (see Millward, 2021), determines the susceptibility of its myofibrils to the UPS, the likely main system involved in muscle proteolysis, with individual proteins like actin less susceptible than MHC. It is to be hoped that the recent advances in stable isotope-proteomic approaches to tissue protein turnover rates (Hammond et al., 2016; Swovick et al., 2018, 2021) can throw more light on these questions.

In the human adult about 300 g of protein turnover occurs each day, (4 g kg^{-1} d^{-1})- that is, three or four times the daily dietary intake. This is less per kg than in infancy with turnover rates of 10–20 g kg^{-1} d^{-1} because of the different body composition in terms of the higher proportion of metabolically active more rapidly turning-over tissues in infancy.

Protein turnover constitutes an appreciable fraction of resting energy expenditure (REE) although the extent of this is not clear, with reports of regressions of fasting WBPT on REE indicating turnover accounting for between 20% (Welle and Nair, 1990), to >50% (Boirie et al., 2001) of REE with no clear reason for the difference. While the former, lower value is more consistent with the stoichiometric calculations based on about 5 mol ATP/GTP per mole of protein turnover, (4 mol per peptide bond with an extra mole for amino-acid transport, RNA turnover and proteolysis), the latter value is discrepant. Thus at an energy cost of ATP synthesis from glucose of ~75 kJ/mol, and a molecular weight of 110 per mole of peptide bond, 5 mol ATP/GTP per mole of protein turnover is equivalent to about 3.41 kJ/g protein turnover or about 14.3% of the adult REE for 300 g/day WBPT. Thus the mismatch with the higher energy cost of ≥50% REE is a factor of ~3.5. Although the ATP consumption by the proteosome is now known to be higher than earlier estimates, (at about 0.5 mol per peptide bond), it is difficult to construct a stoichiometry whereby WBPT accounts for >50% of the REE without including a large number of extra ATP-consuming reactions per peptide bond formed or cleaved. Whatever the actual figure it is clear that WBPT is a very important driver of the REE and it is the case that the fall with age in both protein turnover and metabolic rate from birth to adulthood involve a factor of 3–4 in each case.

As far as any relationship between protein turnover and protein requirements, there is no *a priori* reason for this and little evidence of any. Thus turnover does not consume significant quantities of amino-acids and amino-acid catabolism and oxidation is not known to be linked to turnover. Maintenance protein requirements are assumed to decrease relatively little with age, (<20%) compared with the 3–4 fold fall in turnover.

The diurnal cycle of feeding and fasting and postprandial protein utilization

Overall proteostasis of the whole body proteome is maintained within a diurnal cycle of postprandial protein gains and postabsorptive losses, with human studies, mainly in adults, focusing on the nutritional regulation of these diurnal changes within the whole body (Pacy et al., 1994; Fereday et al., 1998) or in muscle because of its accessibility, (see Millward and Smith, 2019). Thus the stimulation of protein synthesis and inhibition of proteolysis with protein feeding has been examined together with the relative importance of insulin or amino-acids as mediators of the feeding response. One key question is what influences the efficiency of postprandial protein utilization and consequent protein requirements for overall balance? This depends on the balance between post prandial protein deposition and amino acid oxidation.

Because many indispensable amino-acids are potentially toxic and are maintained at very low concentrations in tissues, their rapid and regulated postprandial disposal is important for metabolic homeostasis. After a protein meal two pathways for amino-acid disposal exist. One is disposal through the high-capacity, finely regulated oxidative pathways with rates mainly influenced by their tissue concentrations, (generally similar to the K_m of the rate limiting enzymes), together with substrate activation and covalent enzyme modification. Examples are phenylalanine hydroxylase, and branched-chain α-keto acid dehydrogenase, which are both regulated by substrate binding and reversible phosphorylation and dephosphorylation. The second pathway is net protein deposition. This can be achieved by stimulation of protein-synthesis or inhibition of proteolysis by postprandial hyper-amino-acidemia as described above. However the increased intracellular amino-acid levels which stimulate protein synthesis may also increase amino-acid oxidation rates and while this allows effective removal of amino-acids, this would not be the most efficient protein utilization mechanism. If signal transduction could commence in the extracellular amino-acid pool, i.e., increases in plasma amino-acid concentrations can signal an increase in protein synthesis while intracellular amino-acids need not increase to the same extent, this would minimize any increase in oxidation. There is increasing evidence that this is the case (see Millward and Smith, 2019).

For proteolysis, amino-acids exert an inhibitory influence as described above, and this inhibition will reduce endogenous intracellular amino-acid supply, minimizing any increases in amino-acid levels and amino-acid oxidation and maximizing dietary protein utilization. Furthermore since inhibition of proteolysis and lowering of intracellular amino-acid levels can be achieved by receptor mediated mechanisms involving insulin as well as specific amino-acids, (e.g., leucine), this allows the postprandial increases in plasma amino-acids to mediate sufficient amino-acid transport into cells for protein deposition, without marked increases in intracellular amino-acid levels thus minimizing increases in amino-acid oxidation. Thus as a strategy for mediating postprandial protein utilization, inhibition of proteolysis would be predicted to be more efficient than increasing protein synthesis.

^{13}C-1 leucine studies of WBPT have provided clear experimental support for such a mechanism. In adult subjects habituated to increasing protein intakes from below to twice habitual levels, isoenergetic meals of increasing, (i.e., habitual), protein intake induced a clear inhibition of proteolysis at all levels of protein intake, to an increasing extent with intake, but much less obvious changes in protein synthesis, with slight inhibition or no change at low intakes to stimulation at high intakes (Pacy et al., 1994). The relative importance of the energy/insulin as opposed to protein/amino acid component of the feeding response was established in other ^{13}C-1 leucine feeding studies examining the response to habitual protein intakes fed as small-repeated milk meals (Fereday et al., 1998). These involved three phase, (3*3 h), continuous infusions, i.e., fasting, followed by isoenergetic low, and then habitual protein-feeding. Insulin levels were similar throughout the two feeding phases: i.e., the phase 2 response reflected energy/insulin and the phase 3 complete meal response reflected insulin/energy plus protein/amino acids (Gibson et al., 1996; Fereday et al., 1998). The results are shown in **Fig. 7**. In all 25 subjects studied, the complete meal reduced proteolysis by 46% on

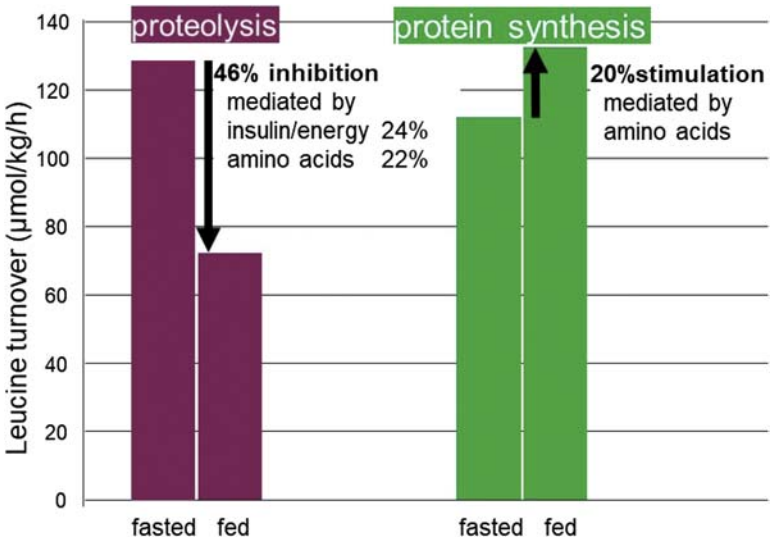


Fig. 7 Scheme showing feeding-induced responses of proteolysis and protein synthesis in the whole body as indicated by leucine kinetics. Studies involved 3 by 3 h continuous, primed constant intravenous infusions of 25 healthy adults, (10 females, 15 males) with ^{13}C-1 leucine, initiated in the fasted state with feeding of frequent small liquid meals providing subjects energy needs, with low (3–6 h) followed by high protein (6–9 h) equivalent to habitual daily protein intakes. A constant insulinemia was maintained during the two feeding phases. Protein kinetics calculated from the plasma flux based on enrichment of plasma ^{13}C-1 α-keto isocaproate at plateau at the end of each 3 h period. Data from Fereday et al. (1998).

average and stimulated protein synthesis by 20% with the amino acids effecting similar changes in proteolysis and synthesis, but with energy/insulin mediating an additional inhibition of proteolysis. Further analysis of the interindividual variability of these responses in terms of the efficiency of post-prandial protein-utilization showed clearly that the most efficient utilization, (minimal increase in amino acid oxidation), involved maximal inhibition of proteolysis by protein feeding with minimal increases in protein synthesis. Thus the efficiency of protein utilization in individuals appears to be determined mainly by the sensitivity of the insulin-mediated inhibition of proteolysis to amino-acid supply.

These and other studies suggests a mechanism indicated in Fig. 8, in which the major target of insulin is inhibition of proteolysis with extracellular amino-acids acting to both enhance the inhibition of proteolysis and stimulate protein synthesis and with intracellular amino-acids stimulating their own oxidation. With extracellular amino-acid levels controlled by diet, and tissue levels controlled by endogenous supply from proteolysis, inhibition of proteolysis will minimize any increase in amino-acid levels, minimize oxidation, and maximize protein utilization. Since protein synthesis and amino-acid oxidation appear to be stimulated in

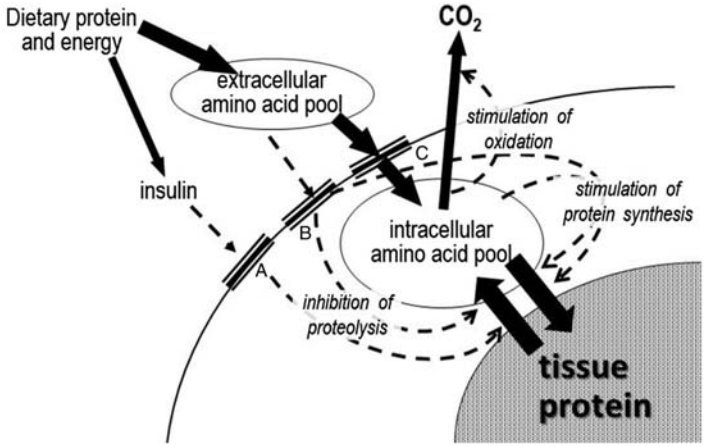


Fig. 8 Scheme for the action of insulin and amino-acid supply on postprandial protein utilization. During the post prandial response to food, insulin exerts an inhibitory influences on proteolysis through a receptor mediated mechanism, a, and extracellular amino-acids signal a stimulation of protein synthesis through b. Intracellular amino-acid levels are regulated by amino-acid transporters, c, and by proteolysis, and their increases stimulate amino-acid oxidation and protein synthesis to some extent in parallel. Maximal inhibition of proteolysis and minimal increases in intracellular amino-acid levels is the optimal response. Modified from Millward et al. (1996).

parallel, the optimum strategy for maximum efficiency of postprandial protein utilization would appear to involve maximal inhibition of proteolysis ensuring minimal postprandial increases in tissue amino-acid levels with a stimulation of protein synthesis mediated from extracellular amino-acid supply.

Clearly these whole body responses reflect weight average responses of different tissues but as indicated above the much higher turnover rate of the splanchnic proteome means that its response to feeding will dominate the whole body response. However although measurements in muscle show a robust feeding response of protein synthesis, extracellular amino acid signaling has been clearly demonstrated in muscle as in the liver. Obviously when amino-acid dietary supply exceeds the capacity for its net deposition, intracellular amino-acids will rise stimulating their disposal via oxidation.

Conclusion

Ideas about proteostasis within the proteostasis network (PN) have emerged in the last decade. Thus the proteomic identification of the thousand or so proteins in the PN involved in the regulation of protein synthesis, folding, transport, and degradation is revealing how cellular health is maintained and shining new light on our understanding of protein turnover. At the level of the whole body proteome we now have a better understanding of the coordination of growth of muscle and stature through mechanotransduction mechanisms which prevent any dietary protein-induced muscle growth in excess of that allowed by growth in stature, a limitation extending throughout the life cycle, at least in the absence of resistance exercise. The importance of amino-acids and insulin in mediating nutritional regulation of protein turnover has been strengthened through new findings about their mTORC1-mediated control mechanisms of protein synthesis and proteolysis. Nevertheless much mystery remains especially in skeletal muscle including how complex protein structures such as the myofibrillar apparatus are assembled and dismantled, how much of this occurs during the daily cycle of gains and losses of tissue protein, and how the constraint to muscle myofibre expansion by its ECM, observed as the tachyphylaxis of the anabolic response to protein feeding, is exerted at the molecular level.

References

Bates, P.C., et al., 1983. Myofibrillar protein turnover. Synthesis of protein-bound 3-methylhistidine, actin, myosin heavy chain and aldolase in rat skeletal muscle in the fed and starved states. Biochem. J. 214, 587–592.

Boirie, Y., Beaufrère, B., Ritz, P., 2001. Energetic cost of protein turnover in healthy elderly humans. Int. J. Obes. 25 (5), 601–605. https://doi.org/10.1038/sj.ijo.0801608.

Claydon, A.J., et al., 2012. Protein turnover: measurement of proteome dynamics by whole animal metabolic labelling with stable isotope labelled amino acids. Proteomics 12 (8), 1194–1206. https://doi.org/10.1002/pmic.201100556.

Combaret, L., et al., 2016. Cellular mechanisms of protein degradation among tissues. In: The Molecular Nutrition of Amino Acids and Proteins, pp. 27–37. https://doi.org/10.1016/b978-0-12-802167-5.00003-7.

Dikic, I., 2017. Proteasomal and autophagic degradation systems. Annu. Rev. Biochem. 86, 193–224. https://doi.org/10.1146/annurev-biochem-061516-044908.

Fereday, A., et al., 1998. Variation in amino acid mediated, insulin activated inhibition of proteolysis determines the efficiency of protein utilization. Clin. Sci. 95, 725–733. https://doi.org/10.1042/cs0950725.

Gibson, N., et al., 1996. Influences of dietary energy and protein on leucine kinetics during feeding in healthy adults. Am. J. Physiol. Endocrinol. Metabol. 270 (2), E282–E291. https://doi.org/10.1152/ajpendo.1996.270.2.E282.

Hammond, D.E., et al., 2016. Proteome dynamics: tissue variation in the kinetics of proteostasis in intact animals. Mol. Cell. Proteomics 15 (4), 1204–1219. https://doi.org/10.1074/mcp.M115.053488.

Hipp, M.S., Kasturi, P., Hartl, F.U., 2019. The proteostasis network and its decline in ageing. Nat. Rev. Mol. Cell Biol. 20, 421–435. https://doi.org/10.1038/s41580-019-0101-y.

Jewell, J.L., et al., 2019. GPCR signaling inhibits mTORC1 via PKA phosphorylation of raptor. eLife 8, 1–26. https://doi.org/10.7554/eLife.43038.

Klaips, C.L., Jayaraj, G.G., Hartl, F.U., 2018. Pathways of cellular proteostasis in aging and disease. J. Cell Biol. 217 (1), 51–63.

Meijer, A.J., et al., 2015. Regulation of autophagy by amino acids and MTOR-dependent signal transduction. Amino Acids 47 (10), 2037–2063. https://doi.org/10.1007/s00726-014-1765-4.

Millward, D., et al., 1996. Post-prandial protein metabolism. Bailliere's Clin. Endocrinol. Metab. 10 (4), 533–549.

Millward, D.J., Smith, K., 2019. The application of stable-isotope tracers to study human musculoskeletal protein turnover: a tale of bag filling and bag enlargement. J. Physiol. 597 (5), 1235–1249. https://doi.org/10.1113/JP275430.

Millward, D.J., 1995. A protein-stat mechanism for regulation of growth and maintenance of the lean body mass. Nutr. Res. Rev. 8 (1), 93–120. https://doi.org/10.1079/nrr19950008.

Millward, D.J., 2021. Interactions between growth of muscle and stature: mechanisms involved and their nutritional sensitivity to dietary protein: the protein-stat revisited. Nutrients 13 (3), 1–65. https://doi.org/10.3390/nu13030729.

Pacy, P.J., et al., 1994. Nitrogen homoeostasis in man: the diurnal responses of protein synthesis and degradation and amino acid oxidation to diets with increasing protein intakes. Clin. Sci. 86 (1), 103–118. https://doi.org/10.1042/cs0860103.

Pey, A.L., 2020. Protein homeostasis and disease. In: Protein Homeostasis Diseases Mechanisms and Novel Therapies. Elsevier Inc., pp. 23–37. https://doi.org/10.1016/B978-0-12-819132-3/00002-6

Scholz, N., et al., 2020. Targeting the ubiquitin system in glioblastoma. Front. Oncol. 10, 1–30. https://doi.org/10.3389/fonc.2020.574011.

Swovick, K., et al., 2018. Cross-species comparison of proteome turnover kinetics. Mol. Cell. Proteomics 17 (4), 580–591. https://doi.org/10.1074/mcp.RA117.000574.

Swovick, K., et al., 2021. Interspecies differences in proteome turnover kinetics are correlated with life spans and energetic demands. Mol. Cell. Proteomics 20, 100041. https://doi.org/10.1074/MCP.RA120.002301.

Takahara, T., et al., 2020. Amino acid-dependent control of mTORC1 signaling: a variety of regulatory modes. J. Biomed. Sci. 27 (1), 1–16. https://doi.org/10.1186/s12929-020-00679-2.

Ueno, T., Ezaki, J., Kominami, E., 2012. Metabolic contribution of hepatic autophagic proteolysis: old wine in new bottles. Biochim. Biophys. Acta Protein Proteonomics 1824 (1), 51–58. https://doi.org/10.1016/j.bbapap.2011.06.019.

VerPlank, J.J.S., et al., 2019. 26S Proteasomes are rapidly activated by diverse hormones and physiological states that raise cAMP and cause Rpn6 phosphorylation. Proc. Natl. Acad. Sci. U. S. A. 116 (10), 4228–4237. https://doi.org/10.1073/pnas.1809254116.

Waterlow, J.C., Garlick, P.J., Millward, D.J., 1978. Protein Turnover in Mammalian Tissues and in the Whole Body. Elsevier/North-Holland, Oxford.
Welle, S., Nair, K.S., 1990. Relationship of resting metabolic rate to body composition and protein turnover. Am. J. Physiol. Endocrinol. Metabol. 258 (6), E990–E998. https://doi.org/10.1152/ajpendo.1990.258.6.E990.
Xu, F., et al., 2021. Modulation of autophagy: a novel "rejuvenation" strategy for the aging liver. Oxid. Med. Cell. Longev. https://doi.org/10.1155/2021/6611126.
Zhao, J., Goldberg, A.L., 2016. Coordinate regulation of autophagy and the ubiquitin proteasome system by MTOR. Autophagy 12 (10), 1967–1970. https://doi.org/10.1080/15548627.2016.1205770.

Further reading

Alfaro, I.E., et al., 2019. Chaperone mediated autophagy in the crosstalk of neurodegenerative diseases and metabolic disorders. Front. Endocrinol. 9 (778), 1–13. https://doi.org/10.3389/fendo.2018.00778.
Bota, D.A., Davies, K.J.A., 2016. Mitochondrial Lon protease in human disease and aging: including an etiologic classification of Lon-related diseases and disorders. Free Radic. Biol. Med. 100, 188–198. https://doi.org/10.1016/j.freeradbiomed.2016.06.031.
Madrigal-Matute, J., Cuervo, A.M., 2016. Regulation of liver metabolism by autophagy. Gastroenterology 150 (2), 328–339. https://doi.org/10.1053/j.gastro.2015.09.042.
Peris-Moreno, D., Taillandier, D., Polge, C., 2020. MuRF1/TRIM63 , master regulator of muscle mass. Int. J. Mol. Sci. 21, 1–39. https://doi.org/10.3390/ijms21186663.
Proud, C.G., 2014. Control of the translational machinery by amino acids. Am. J. Clin. Nutr. 99 (1), 231–236. https://doi.org/10.3945/ajcn.113.066753.
Saxton, R.A., Sabatini, D.M., 2017. mTOR signaling in growth, metabolism, and disease. Cell 168 (6), 960–976. https://doi.org/10.1016/j.cell.2017.02.004.

Riboflavin

Christopher J. Bates[a,*] **and Luis Fontana**[b,c,d], [a] MRC Human Nutrition Research, Cambridge, United Kingdom; [b] Department of Biochemistry and Molecular Biology 2, School of Pharmacy, Granada, Spain; [c] Institute of Nutrition and Food Technology "José Mataix", Biomedical Research Center, Parque Tecnológico Ciencias de la Salud, Granada, Spain; and [d] Instituto de Investigación Biosanitaria ibs.GRANADA, Complejo Hospitalario Universitario de Granada, Granada, Spain

© 2023 Elsevier Ltd. All rights reserved.

This is an update of C.J. Bates, Riboflavin, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 158–165, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00246-4.

Introduction 401
Absorption, transport, and storage 402
Absorption in humans 403
Riboflavin transport at other sites and storage 403
Metabolism and excretion 403
Metabolic functions 405
Essentiality 405
Fatty acid oxidation 406
Iron economy 406
Malaria 406
Cataracts and photoreceptors 406
Interaction with vitamin B_6 406
Effect on folate metabolism 407
Assessment of riboflavin status 407
Urinary excretion 407
The glutathione reductase test 407
Requirements 408
Dietary sources and high intakes 409
Conclusions 410
References 410

Key points

- To obtain an overview of the role of riboflavin in intermediary metabolism.
- To understand the relationship between the chemical structure of this vitamin and its biological function.
- To know the processes of absorption, transport and storage of this vitamin, and the formation of the corresponding coenzymes.
- To have information about the dietary intakes of this vitamin.
- To know the metabolic disorders related to riboflavin deficiency.
- To know the most important food sources of this vitamin.

Introduction

Riboflavin is a derivative of a flavin compound, isoalloxazine. It has a weak basic character, crystallizes in yellowish needles and is soluble in water, but much less than other vitamins of its group (Ross et al., 2014; Stipanuk and Caudill, 2018). It is resistant to heat and oxidation, stable in acid solution, but unstable in alkaline solution, and especially sensitive to ultraviolet light, which destroys it irreversibly, forming lumiflavin, if in an alkaline medium, or lumichrome, if in an acid or neutral medium (Peechakara and Gupta, 2020).

This article will consider the processes of absorption, transport, storage, metabolism and excretion, the functions, the assessment of the nutritional status, the requirements, the dietary sources, and the status measurement of riboflavin.

*CJ Bates passed away on January 2018.

Absorption, transport, and storage

Riboflavin (vitamin B_2) is not synthesized by higher animals, in which it is an absolute dietary requirement for the synthesis of certain essential coenzymes needed for intermediary metabolism (Combs and McClung, 2016; Rodwell et al., 2018). It is transported from food sources in the gastrointestinal tract, across the gut wall into the circulatory system, and thence via the blood to the tissues. This occurs against a concentration gradient, thus retrieving even small amounts from food and from the low concentrations in plasma to the higher concentrations inside cells (Barile et al., 2016).

Three human riboflavin transporters have been cloned and characterized, belonging to the SLC52 family of solute carriers, whose more recent nomenclature is hRFVT1, hRFVT2 and hRFVT3. They are encoded by genes SLC52A1-3, respectively. hRFVT1 is mainly expressed in placenta, intestine and kidney. hRFVT2 is rather ubiquitously expressed but is particularly abundant in nervous tissues and salivary glands. hRFVT3 is most highly expressed in testis but also in prostate, intestine, stomach and pancreas. These different but overlapping expression profiles might explain the vulnerability of certain tissues to mutations in one or more of the SLC52A genes (O'Callaghan et al., 2019).

Gut riboflavin transport system at low (e.g., micromolar) concentrations is temperature and energy-dependent (it is inhibited by inhibitors of adenosine tri phosphates (ATP) production from energy substrates), it becomes saturated as the concentration of riboflavin increases, and it is sodium ion dependent. These characteristics are shared with many other types of small molecules that are actively transported across the gut wall. More specifically for riboflavin, the active transport mechanism involves phosphorylation (to riboflavin phosphate, also known as flavin mononucleotide (FMN)), followed by dephosphorylation back to riboflavin, both steps occurring within the intestinal cells (**Fig. 1**). This is one of a number of strategies that the gut uses to entrap essential nutrients and then transfer them in a controlled manner. A similar strategy is used at other sites in the body, to ensure entrapment of circulating riboflavin by cells whose nascent flavin-dependent enzymes need a regulated supply of the vitamin.

Although the active transport of riboflavin across the gut wall and across other cell membrane barriers within the animal is a saturable process, if pharmacological amounts are present, then the slower and less efficient, but nonsaturable, process of passive absorption predominates. The active transport process is increased in riboflavin deficiency and is decreased if the riboflavin content of the tissues is high, and it involves calcium and calmodulin, but not sodium. Specific riboflavin receptors and a role for the microtubules have been identified.

Although some of the riboflavin in foods is present as the free vitamin, a larger fraction is in the phosphorylated coenzymes FMN and flavin adenine dinucleotide (FAD), and there may also be small amounts of a glucoside. These forms are all efficiently converted into free vitamin by enzymes secreted into the gut lumen and thus become available for absorption. There are also small amounts of covalently bound and thus unavailable forms of riboflavin, present in enzymes such as succinate dehydrogenase (succinate: ubiquinone oxidoreductase EC 1.3.5.1), monoamine oxidase, and gulonolactone oxidase, which cannot be released by the hydrolytic enzymes in the gut. Also unavailable (or very poorly available) in humans is the riboflavin synthesized by the gut microbiota of the large bowel. Certain animal species such as rodents can utilize this riboflavin source by coprophagy.

Analogs of riboflavin have been prepared to explore riboflavin economy. Some of these have riboflavin-like activity; others are inactive, whereas some are antagonists and can cause functional deficiency. These structural variations can affect absorption and the

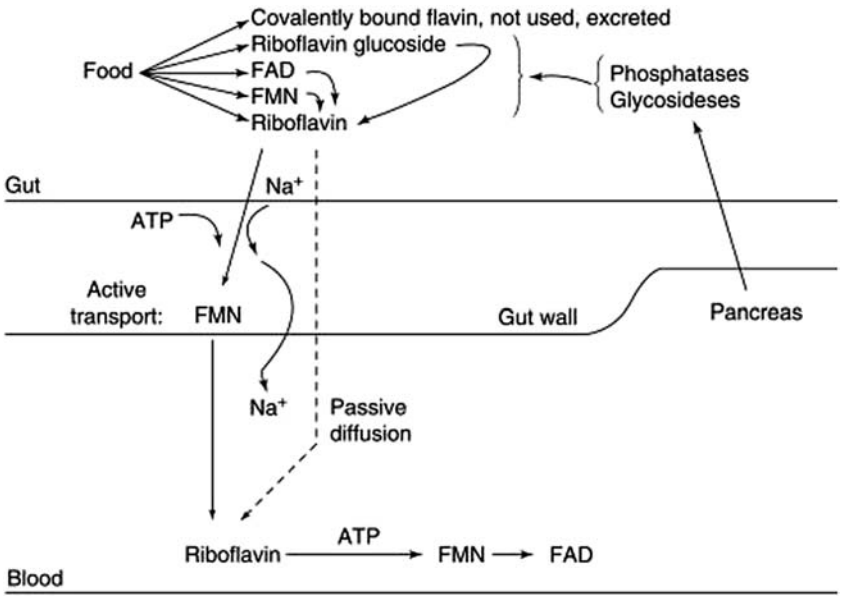


Fig. 1 Characteristics of the absorption process for riboflavin and its coenzymes.

conversion of riboflavin into its coenzyme forms. Certain drugs such as phenothiazines (antipsychotics) have sufficient structural similarity to riboflavin to act as antagonists.

Absorption in humans

Studies of riboflavin absorption by human subjects require a test dose taken by mouth, and a sampling procedure to estimate the amount absorbed and its subsequent fate. The sampling compartment is generally urine, because plasma is unsatisfactory (see the last paragraph of this section and the section on Assessment of riboflavin status below), and fecal sampling is useless because of the synthesis of riboflavin by bacteria in the large bowel. Riboflavin labeled with radioactive or stable isotopes has not yet been widely used in human studies. Instead, most studies have relied on relatively large bolus oral doses of several milligrams of riboflavin, with urinary monitoring over a few hours. Riboflavin can be quantified in urine by its characteristic fluorescence, by a microbiological assay, or high-performance liquid chromatography (HPLC). The duration of exposure in the upper ileum is critical, because this is the region of greatest absorptive efficiency. Slow-release forms of the vitamin do not enhance its absorption, but there does appear to be some absorptive advantage for synthetic lipophilic esters, such as the tetrabutyrate ester, which becomes hydrolyzed to free vitamin during or after absorption. These esters possess beneficial (e.g., antioxidant) properties in some model systems. Food can enhance absorption, possibly by increasing bulk transit time. The efficiency of absorption does not vary markedly with age or sex in humans.

The large intestine is known to possess efficient and specialized carrier-mediated systems that are capable of mediating the absorption of bacterially synthesized B-vitamins, including riboflavin, and this is likely to be especially important for local colonocyte nutrition.

Plasma is of little use as a sampling fluid because redistribution to other tissue sites and urine occurs too rapidly. Although the urinary response to a test dose is preferred, it has the disadvantage that physiological intakes, and especially intakes from poor food sources, cannot be measured accurately. A more sensitive biochemical marker of riboflavin status at low intakes is the index known as the "erythrocyte glutathione reductase activation coefficient" (EGRAC).

Riboflavin transport at other sites and storage

Free riboflavin is trapped as one of its phosphorylated coenzyme forms, which are then associated (and in a few cases covalently linked) to the protein chains of catalytic flavoenzymes. If not covalently linked, the flavin coenzyme can often be released, for example, by extremes of pH. At some locations, such as mature red cells, flavoenzymes such as glutathione reductase (NADPH: oxidized-glutathione oxidoreductase EC 1.6.4.2) may exist partly in their apoenzyme form, i.e., without the flavin coenzyme and therefore without enzyme activity. An increased supply of riboflavin will allow the depleted coenzyme (in this case FAD) to be synthesized so that holoenzyme and activity can be restored.

Different enzymes and different tissue sites differ in the tenacity with which they retain flavin coenzymes during riboflavin deficiency; thus, there is a characteristic pecking order for flavoenzyme protection, reflecting the metabolic importance of the pathways affected. However, there is no repository of unused or nonfunctional riboflavin that can act as a store in times of dietary deficiency. Although some organs (such as the liver) have relatively high concentrations of flavin enzymes, all of the flavin is present as coenzyme in flavin holoenzymes. Each tissue has a characteristic ceiling level of riboflavin at saturation and a floor level characteristic of severe depletion, and these are determined, respectively, by the total amount of apoflavoprotein and the irreducible amounts of the holoenzymes, which cannot be depleted of their cofactor complement.

Riboflavin is secreted into milk, whose concentration is species-specific and dependent on maternal status and intake. There is active transport from the maternal to the fetal circulation in pregnancy, the flavin concentration being greater on the fetal side. Studies from India have identified a riboflavin carrier protein (RCP) present in bird (e.g., chicken) eggs, which is specific for riboflavin, and essential for normal embryological development. If it is rendered ineffective (e.g., by immunoneutralization), embryonic development ceases and the embryo dies. A genetic mutant lacking RCP was infertile. A homologous protein, which was rendered ineffective by the antibody to pure chicken riboflavin carrier protein, was found to occur in several mammalian species, including two species of monkeys, and in humans. Circulating RCP levels and the immunohistochemical staining of RCP in biopsy specimens may provide new markers for breast cancer diagnosis and prognosis. Termination of pregnancy occurred after immunoneutralization of RCP in monkeys. The role(s) of RCP in humans, however, remain controversial, and other, less specific riboflavin binders in blood, including gamma-globulins, also seem to play an important role. Further evidence of the flavin needs of developing embryos has been provided by the demonstration that riboflavin analogs can cause teratogenic changes in the absence of any detectable damage to maternal tissues.

Metabolism and excretion

The riboflavin coenzymes are depicted in Fig. 2. Both coenzymes are much more soluble in water than riboflavin. ATP is a cosubstrate and driving force (in energy terms) for both stages of the conversion of riboflavin into FAD. Some flavoenzymes specifically require FAD, whereas others specifically require FMN. Table 1 lists the broad categories (by two alternative classification options) of flavoenzymes found in living tissues: all revolve around redox processes. The central biochemical reaction of the flavin coenzymes

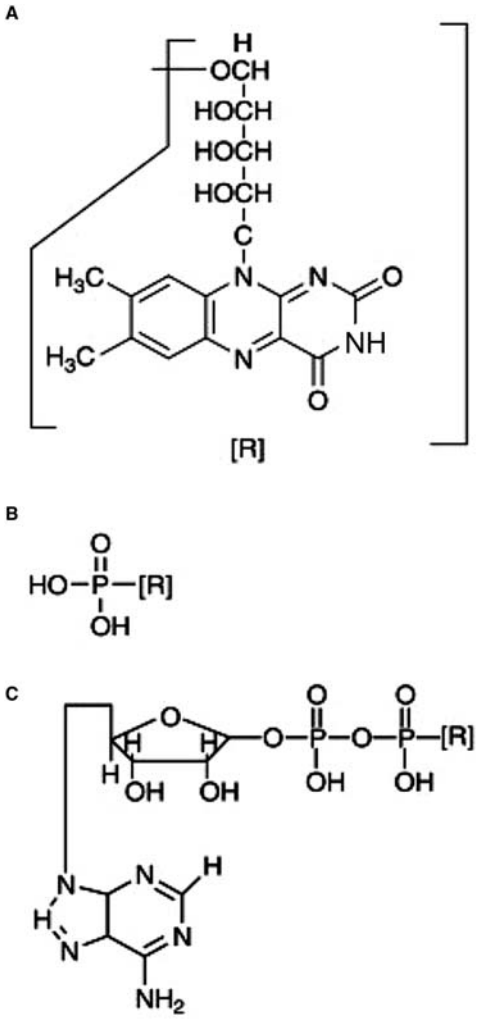


Fig. 2 Structure of riboflavin and its coenzyme derivatives. (A) Riboflavin; (B) riboflavin phosphate (flavin mononucleotide, FMN); and (C) flavin adenine dinucleotide (FAD).

Table 1 Two classification options for flavoenzyme categories.

Category	*Example*
Classification by reaction-type	
Electron transferases	Mitochondrial electron-transfer flavoprotein
Dehydrogenases	Mitochondrial NADH dehydrogenase and succinate dehydrogenase; (cytosolic) glutathione reductase
Dehydrogenase-oxygen reductases (with O_2 reduction to H_2O_2)	Monoamine oxidase
Flavoprotein monooxygenases	Bacterial lactate monooxygenase, microsomal FAD-containing monooxygenase
Classification by flavin-type	
Enzymes with FMN	NADH-FMN oxidoreductase
Enzymes with FAD	Glutathione reductase, methylene tetrahydrofolate reductase
Enzymes with both FMN and FAD	NADPH-cytochrome P-450 reductase
Enzymes with covalently bound flavin (formed by posttranslational flavinylation)	Succinate dehydrogenase

involves the interconversion of the reduced, dihydro form of the flavin ring and the oxidized form. One of the most important sites of action of flavoenzymes in higher animals is the electron transport chain in the mitochondria. Flavins in succinate dehydrogenase and NADH dehydrogenase form essential redox links between the oxidizable energy-rich substrates of aerobic metabolism and the cytochrome chain, leading to molecular oxygen, which can generate approximately 38 mol of ATP per mole of glucose oxidized.

Hormone status can affect riboflavin economy and riboflavin status can affect hormone production. One important control point for riboflavin economy is thyroid hormone status: hypothyroidism leads to lower tissue levels of flavin coenzymes, and hence to inactivation of certain flavoenzymes, thus resembling the effects of dietary riboflavin deficiency. Both flavokinase (ATP: riboflavin 5′-phosphotransferase EC 2.7.1.26) and FAD pyrophosphorylase (ATP: FMN adenyltransferase EC 2.7.7.2) are regulated by thyroid hormone status. In the kidney, the synthesis of flavokinase and hence of flavin coenzymes is controlled by aldosterone.

The amount of absorbed riboflavin that remains in the circulation (in blood plasma) is regulated by glomerular and tubular filtration, and tubular reabsorption, in the kidneys. This is an active, saturable, transport process, with characteristics similar to those of active transport in the gastrointestinal tract. It is mainly responsible for the very sharp and characteristic transition between minimal urinary excretion of riboflavin at low intakes, and a much higher level of excretion, proportional to intake, at higher intake levels. This transition point has been used to define and to measure riboflavin status and requirements and to allow studies of intestinal absorption in vivo. Excretion of riboflavin is affected by some chemicals (such as boric acid, which forms a complex with it) and by certain diseases and hormone imbalances.

In addition to the excretion of unchanged riboflavin, there are small urinary amounts of hydroxylated breakdown products of the vitamin, which arise through normal turnover, either within the tissues of the body or in the gastrointestinal tract from bacterial action, before absorption. The rate of destruction of riboflavin by this turnover pathway is low in all species examined and riboflavin within the mammalian body is efficiently conserved.

Metabolic functions

FMN and FAD are tightly bound to a large number of proteins (flavoproteins or flavoenzymes) that perform electronic transport functions through oxidation-reduction processes. In some cases, these reactions involve the transfer of two electrons, so that both FMN and FAD accept the two hydrogen atoms given up by the substrate. A typical example of these reactions is the succinate to fumarate step in the Krebs cycle, a reaction catalyzed by succinate dehydrogenase (EC 1.3.5.1). Another reaction of this type, of great metabolic interest, is the oxidation of acyl-CoA by acyl-CoA dehydrogenases, which constitutes a fundamental step in the mitochondrial catabolism of fatty acids.

In other cases, these coenzymes can accept a single electron, giving rise to a stable chemical species of the semiquinone type. This class of reactions is essential in mitochondrial electron transport chains (respiratory chains). Specifically, electrons from FADH2 bound to acyl-CoA dehydrogenases are transferred to another flavoprotein, called electron transfer flavoprotein (ETF), by this mechanism. In turn, this protein transfers the electrons to proteins with iron and sulfur that have ETF:ubiquinone oxidoreductase activity (EC 1.5.5.1). Finally, the electron acceptor is ubiquinone.

These reactions also occur in the electron transport chains located in the endoplasmic reticulum (microsomal electron transport chains). There are two main types of these chains. One of them performs the desaturation of fatty acids. This process involves enzymatic proteins (desaturases), cytochromes (cytochrome b5) and flavoproteins. An electronic transport from NADH to oxygen takes place, as in the respiratory chains. The difference is that, in this case, the hydrogens of the substrate are also used to form water, thus leading to the formation of the double bond.

Other microsomal electron transport chains carry out the hydroxylation of molecules with a steroid structure, enabling, for example, the formation of cholesterol and 1,25-dihydroxychalciferol (or calcitriol). This system works analogously to the previous one, although it is a more complicated process. In this case, the source of electrons is NADPH, instead of NADH, and a special cytochrome called cytochrome P-450 is used, in addition to the corresponding enzymatic proteins and flavoproteins. As these are enzymes with very low substrate specificity, this system also performs oxidations on exogenous molecules (xenobiotics), such as alcohol and drugs.

An additional capacity of flavoenzymes is to react directly with oxygen, forming hydrogen peroxide. This is how, for example, xanthine oxidase (EC 1.17.3.2), the enzyme responsible for the formation of uric acid, functions.

Given the great versatility of riboflavin-derived coenzymes, it is not surprising that there are numerous flavin enzymes. Most of them are involved in the oxidative metabolism of nutrients in their different stages. But, in addition, some are part of the antioxidant defense, such as glutathione reductase, or are used in the biosynthesis of other coenzymes, such as pyridoxal phosphate or 5′-methyltetrahydrofolic acid. It is therefore easy to deduce that riboflavin deficiency can influence many biochemical areas.

Essentiality

The following section addresses the biochemical and physiological actions of flavins that are responsible for the characteristic functional effects of riboflavin deficiency (Balasubramaniam et al., 2019; Mahabadi et al., 2020; Mosegaard et al., 2020; Plantone et al., 2021; Saedisomeolia and Ashoori, 2018; Thakur et al., 2017).

Fatty acid oxidation

An early effect of serious metabolic disturbance seen in moderate riboflavin deficiency is a disturbance of fatty acid oxidation. The normal first stage in the spiral process of beta-oxidation of fatty acids within the mitochondria is the removal of two hydrogen atoms from the two carbons located alpha and beta to the activated carboxyl end of the fatty acid chain. The fatty acyl coenzyme A substrate is acted on by one of several fatty acyl CoA dehydrogenase flavoprotein enzymes (e.g., long-chain acyl-CoA:(acceptor) 2,3-oxido-reductase EC 1.3.99.13), each of which is specific for a narrow range of acyl chains. The second stage involves transfer of the electrons via electron-transferring flavoprotein dehydrogenase (ETF: ubiquinone oxidoreductase EC 1.5.5.1, mentioned above) and thence to the mitochondrial cytochrome chain and to oxygen. These flavoenzymes, unlike the flavoenzymes that are linked to carbohydrate oxidation, are highly sensitive to dietary riboflavin depletion. Characteristic disturbances of lipid metabolism therefore arise in riboflavin-deficient tissues and organisms.

Disturbances in fatty acid oxidation by isolated mitochondria, for example, from the livers of deficient animals, have been demonstrated, and one of the most characteristic metabolic changes, observed even in a mild deficiency state in experimental animals, is the appearance of abnormal dicarboxylic acids and their derivatives in the urine. These products may arise because fatty acyl intermediates become diverted away from the usual pathway of mitochondrial beta-oxidation, toward abnormal partial oxidation in the peroxisomes.

Humans normally do not accumulate these urinary products but individuals with an abnormal gene resulting in dicarboxylic-aciduria (as in multiple acyl CoA dehydrogenase deficiency or glutaric aciduria type II) do respond to riboflavin supplements quite frequently, showing a reduction in their excretion and clinical improvement. High-dose riboflavin can thus overcome the genetic abnormality, by providing more coenzyme, thereby ensuring that the residual fatty acid oxidation pathway works at optimum capacity. The accumulation of dicarboxylic acids in urine is characteristic of riboflavin-deficient mammals but not of birds; thus, chick embryos deprived of riboflavin via a genetic lesion affecting riboflavin carrier protein seem to die of hypoglycemic shock, but do not exhibit dicarboxylic-aciduria.

Iron economy

An important interaction of riboflavin with iron economy is indicated by the fact that iron-deficient animals fail to respond to iron supplements if they are also riboflavin deficient and that the redox system involving riboflavin and its coenzymes interacts readily with the redox system between ferric and ferrous iron.

Studies in experimental animals indicate impairment of iron absorption in riboflavin-deficient animals, changes in its distribution between body compartments; and an increase in rate of iron loss from the intestinal mucosa, resulting in impaired retention of the body iron stores. This enhanced rate of iron loss is accompanied by hyperproliferation of crypt cells, and increased cellular transit along the villi, leading to an excessive proportion of immature villi and a reduction in absorptive area. These studies help to explain how a combination of iron deficiency and riboflavin deficiency, which is frequently encountered in human populations in developing countries, may lead to a gradual deterioration of iron status, which is often accompanied by other intestinal lesions and impaired gut function.

Riboflavin enhances the hematological response to iron, and deficiency may account for at least some of the anemia seen in human populations. Unlike iron-deficiency anemia, the anemia of riboflavin deficiency is usually normocytic and normochromic.

Malaria

Low-dietary riboflavin intakes are frequently encountered in malarious areas of the world, and in some studies, biochemical riboflavin deficiency is associated with a reduced level of blood cell parasitemia. Although neither animal nor human studies have indicated that riboflavin deficiency protects from the life-threatening sequelae of malaria, there may be an interaction between the parasite and flavins within cells. Some prophylactic drugs used to prevent malaria infection have riboflavin-like structures.

Cataracts and photoreceptors

Several micronutrients, especially those with antioxidant-type functions in living tissues, might provide some protection against degenerative eye diseases, such as cataract. Animal models, epidemiological studies, and an intervention study in China support the suggestion that good riboflavin status, or riboflavin supplements, may be protective and this possibility deserves further study.

Another intriguing role of flavoproteins in the eye involves a photoreceptor function that synchronizes circadian rhythms with the solar light–dark cycle, acting via cryptochromes 1 and 2, which contain FAD and function as blue light-sensitive photoreceptors.

Interaction with vitamin B_6

Riboflavin and vitamin B_6 are metabolically interrelated. The conversion of pyridoxine or pyridoxamine phosphates into pyridoxal phosphate is catalyzed by a flavoenzyme (pyridoxaminephosphate oxidase EC 1.4.3.5) and a deficiency of riboflavin can, at certain sites, result in a secondary deficiency in vitamin B6-dependent pathways.

Effect on folate metabolism

FAD is an essential coenzyme for 5,10-methylene tetrahydrofolate reductase (EC 1.5.1.20), a key enzyme of the folate activation pathway, catalyzing the interconversion of 5,10-methylene tetrahydrofolate and 5-methyltetrahydrofolate. Of several single nucleotide polymorphisms affecting this enzyme, the best known are the C699T and A1298C variants. The former confers thermolability and lowered reductase activity in the TT homozygote, apparently explained by enhanced loss of the FAD cofactor. Marginal riboflavin status is associated with increased plasma homocysteine levels (possibly predictive of increased vascular disease risk), arising from the reduced activity of this key enzyme in TT subjects. The same polymorphism appears to modulate the risk of some cancers, notably colorectal cancer.

Another polymorphism in the 5,10-methylene tetrahydrofolate reductase gene, C677T, seems to be related to blood pressure variation. Supplementation with riboflavin can lower blood pressure in hypertensive individuals with the TT genotype but not in those with the CC or CT genotype (Horigan et al., 2010).

Assessment of riboflavin status

Assessment of riboflavin status is closely linked to the estimation of its dietary requirement and the monitoring of human populations for intake adequacy. It is often cheaper, easier, and more accurate to collect a sample of blood or urine from an individual and then perform biochemical analyses that determine status, than attempting to measure food intakes, because the latter requires more cooperation from the subject and has the limitations of imprecision and poor applicability of food table riboflavin values.

Riboflavin status estimates are generally based on urinary excretion or measurements of erythrocyte glutathione reductase (NADPH: oxidized-glutathione oxidoreductase EC 1.6.4.2) and its reactivation with FAD in red cell lysates. Other biochemical indices, such as plasma or red cell flavin concentrations, have been less widely used, but their potential may increase with new assay techniques such as capillary electrophoresis with laser-induced fluorescence detection. Functional indices linked to flavin-requiring pathways in vivo are rarely used, except for the investigation of errors of metabolism or of rare diseases. The two principal status tests are as follows:

Urinary excretion

The amount of riboflavin excreted in the urine is negligible at low intakes of the vitamin. As the dietary level increases, there is slow increase to a transition point, above which the slope of the excretion rate increases very sharply and then remains proportional to intake until absorption is saturated. For population studies, it has been found convenient to use the creatinine excretion rate as the denominator, and the suggested interpretation of urinary riboflavin excretion rates is <27 μg riboflavin g^{-1} as deficient; 27–79 μg g^{-1} as low; and >80 μg g^{-1} as acceptable. This index is sufficiently sensitive to distinguish riboflavin requirements between individuals on low-fat, high-carbohydrate diets and the slightly higher requirement associated with high-fat, low-carbohydrate diets. However, one serious drawback of the urinary excretion index is that it is relatively insensitive to intake variations at low to moderate riboflavin intakes. Another is that 24 h urine samples are not easy to collect and excretion rates may fluctuate over short time periods. In addition, metabolic states associated with tissue catabolism may release riboflavin during cell turnover, resulting in increased urinary excretion even when the dietary intake is low.

The glutathione reductase test

A more reliable status index is the degree of unsaturation of the red blood cell enzyme, glutathione reductase (NADPH: oxidized-glutathione oxidoreductase EC 1.6.4.2), with respect to its flavin cofactor, FAD (**Fig. 3**).

Inadequate dietary riboflavin results in low circulating levels and hence a gradual progressive loss of cofactor from this red cell flavoenzyme over a period of several weeks. Because the enzyme protein (apoenzyme) remains intact and is re-activatable by FAD, it is possible to assess riboflavin status by measuring glutathione reductase activity with, and without, its FAD cofactor, in washed red cells. If riboflavin replete, FAD has little effect and the activation coefficient or the ratio of FAD stimulated to unstimulated activity (EGRAC) is between 1.0 and 1.3–1.4. If deficient, FAD produces a larger stimulation and the activation coefficient is much higher. For individuals living in communities with very low intakes of riboflavin and a significant prevalence of clinically recognizable deficiency, activation coefficients as high as 2.0–3.0 are common. In Western countries, few values as high as 2.0 are encountered. However, recent population surveys in the UK have indicated that the proportion of values between 1.3 and 1.8 is considerable across all age ranges. Whether this reflects suboptimal intakes of riboflavin-rich foods, such as cow's milk, remains uncertain.

This blood test is highly sensitive to, and predictive of, the extent of tissue depletion in the range of severe to moderate deficiency. It is robust and requires only a small sample of blood and can be automated by commercial enzyme rate reaction analyzers. When deficient subjects are provided with riboflavin supplements, there is rapid restoration of saturation of the enzyme, and graded supplements can be used to estimate human requirements.

There are minor operational differences among different published versions of the analytical procedure for EGRAC, which result in small between-laboratory differences in the definition of the normal range, and there are external factors that may cause ambiguity of interpretation. One of these is the genetic variant, glucose-6-phosphate dehydrogenase (D-glucose-6-phosphate: NADP+ 1-

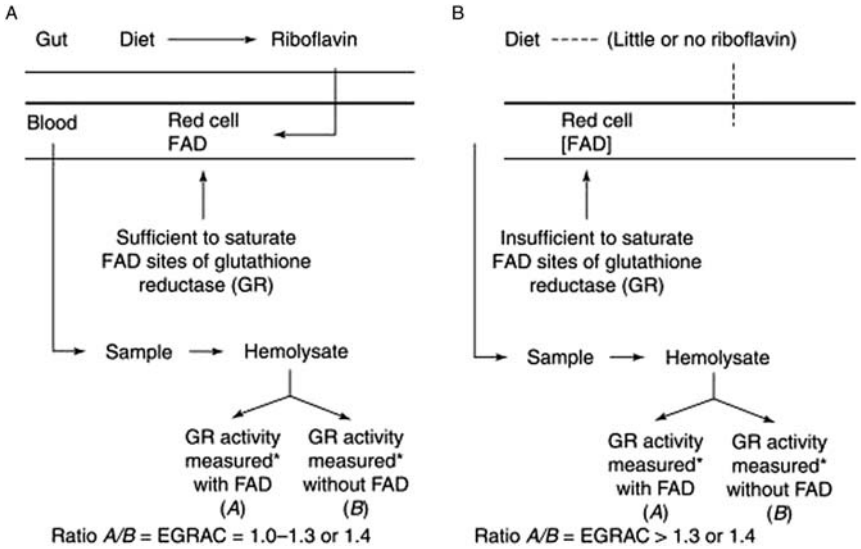


Fig. 3 Basis of the glutathione reductase for riboflavin status: (A) riboflavin sufficient; (B) riboflavin deficient. *Reaction of oxidized glutathione with reduced nicotinamide adenine dinucleotide phosphate.

oxidoreductase EC 1.1.1.49) deficiency. Both homo- and hetero-zygotes are affected, and their erythrocyte glutathione reductase is almost saturated with FAD, even when their tissues are riboflavin deficient. Alternative tests of status, such as HPLC measurement of riboflavin in blood fractions, are then required.

Some groups of individuals have increased requirements for riboflavin. There is, for instance, a progressive increase in requirement during pregnancy, followed by a gradual decrease during lactation. Babies exposed to phototherapy for neonatal jaundice have increased requirements. Oral contraceptives may increase requirements, but the evidence is conflicting. Individuals with inborn metabolic errors leading to dicarboxylic-aciduria and associated clinical abnormalities may have a functional deficiency. Some drugs affect riboflavin status indices, but their need for riboflavin supplements is uncertain.

Requirements

As for all micronutrients, the evidence on which requirement estimates are based can be subdivided into the following broad classes of criteria:

1. Prevention of clinical (pathological) deficiency.
2. Attainment of specified blood levels or tissue stores of riboflavin.
3. Titration to the urinary excretion threshold.
4. Tests based on cofactor saturation of one or more accessible, diet-sensitive, flavin-dependent enzymes, such as erythrocyte glutathione reductase.
5. Optimization of riboflavin-dependent physiological functions.

Of these five classes of criteria, the first has been useful in defining "minimum" requirements, but as a practical test of status, it has several drawbacks. Clinical signs of deficiency in human communities tend to be nonspecific and multifactorial and signs such as angular stomatitis and cheilosis do not always correlate closely with, or respond rapidly to, changes in dietary riboflavin supply or biochemical evidence of deficiency. Factors such as local infection are also likely to be critical.

The use of physiological functional indices in relation to riboflavin deficiency (analogous to dark adaptation for vitamin A; clotting factors for vitamin K, etc.) has not proved possible, because the analogous riboflavin-sensitive physiological processes are insufficiently specific for use in population studies. Of the biochemical indices, urinary excretion and reactivation of erythrocyte glutathione reductase are the most favored for human studies.

In the US, the current recommended dietary allowances (RDAs) are 1.3 mg day^{-1} for men and 1.1 mg day^{-1} for women, increasing to 1.4 mg day^{-1} in pregnancy and 1.6 mg day^{-1} in lactation, with proportional amounts, based on metabolic body weights and growth requirements, for children and adolescents. RDAs are set 20% higher than the estimated average requirement (EAR) for each group (Institute of Medicine, 2006).

According to the European Food Safety Authority (EFSA), new scientific data have become available for adults since the publication of the Scientific Committee for Food (SCF) report in 1993, and therefore updated average requirements (ARs) and population reference intake (PRIs) can be set for adults, children, pregnant and lactating women (European Food Safety Authority, 2017). For adults, EFSA considers that an AR of 1.3 mg day^{-1} (after rounding) can be determined from the weighted mean of riboflavin

intake associated with the inflection point in the mean urinary riboflavin excretion curve in relation to riboflavin intake as reported in four intervention studies in different nonEuropean Union (EU) countries. The EFSA considers that the potential effect of physical activity and of MTHFR 677TT genotype on riboflavin requirement is covered by the data presented from the studies considered, thus is accounted for in the assumed the coefficient of variation (CV) applied to set the PRI for riboflavin. A CV of 10% was used to calculate PRIs from the ARs for adults, i.e., 1.6 mg day^{-1} after rounding, and the same CV was used for all other population groups. EFSA considers that there is no indication of different riboflavin requirement according to sex or between younger and older adults, and sets the same DRV for men and women (without correction per difference in body weight) of all ages. For all infants aged 7–11 months, in the absence of sufficient data to set an AR, the EFSA sets an AI of 0.4 mg day^{-1} based on the estimated intake of riboflavin of exclusively breastfed infants from birth to six months, and upward extrapolation by allometric scaling (on the assumption that riboflavin requirement is related to metabolically active body mass), taking into account the difference in reference body weight.

For children aged 1–17 years, EFSA sets ARs by downward extrapolation from the AR of adults, by allometric scaling (on the assumption that riboflavin requirement is related to metabolically active body mass), applying growth factors and taking into account the differences in reference body weight. EFSA considers unnecessary to set sex-specific ARs and PRIs for boys and girls of all ages. EFSA sets ARs ranging from 0.5 (children aged 1–3 years) to 1.4 mg day^{-1} (children aged 15–17 years) and PRIs ranging from 0.6 (children aged 1–3 years) to 1.6 mg day^{-1} (children aged 15–17 years). For pregnant women, EFSA considers that data are insufficient to estimate the additional needs for dietary riboflavin during pregnancy based on fetal uptake and riboflavin accretion in the placenta during pregnancy. EFSA sets an AR of 1.5 mg day^{-1}, calculated by allometric scaling from the AR for nonpregnant women, considering the mean gestational increase in body weight of 12 kg, and also sets a PRI of 1.9 mg day^{-1}. For lactating women, an additional riboflavin requirement of 0.31 mg day^{-1} is calculated considering the secretion of riboflavin into milk during lactation (0.291 mg day^{-1}), the mean milk transfer during the first six months of lactation in exclusively breastfeeding women (0.8 L day^{-1}), and an absorption efficiency of 95%. An AR of 1.7 mg day^{-1} is calculated by EFSA, considering the additional requirement above the AR of non-lactating women, and a PRI of 2 mg day^{-1} is set for lactating women. Based on data from 13 surveys in nine countries of the EU, riboflavin intake mean estimates ranged across countries from 0.6 to 1.2 mg day^{-1} in infants (<1 year), from 0.9 to 1.4 mg day^{-1} in children aged 1 to <3 years, from 1 to 1.8 mg day^{-1} in children aged 3 to <10 years, and from 1.2 to 2.2 mg day^{-1} in children aged 10 to <18 years. Riboflavin intake mean estimates ranged between 1.4 and 2.2 mg day^{-1} in adults.

Dietary sources and high intakes

Table 2 lists the riboflavin contents of some commonly consumed foods in Western countries. As is the case with most other B vitamins, the richest food sources comprise items such as offal and yeast extract, with meat and dairy products such as milk also providing generous amounts (In the UK, milk intake by children and young adults has tended to decline in recent decades, being affected by changes in government policy on provision of free milk to school children, and by the increasing popularity of manufactured soft drinks). Fruit and vegetables provide modest amounts of riboflavin, and ungerminated grains and seeds, such as nuts, are relatively poor sources. There is an enormous difference in intakes and in status observed between most Western countries, on the one hand, where the dietary intake tends to be relatively generous, and many developing countries, on the other, where the common staples tend to be riboflavin-poor foods such as polished rice. Although riboflavin deficiency is not as life threatening

Table 2 Riboflavin content of selected foods.

Food	*mg per 100 g fresh wt*
Meat, offal, and fish	
Stewed minced beef	0.06
Grilled pork chop	0.18
Calf liver, fried	2.89
Lamb's kidney, fried	3.10
Cod, grilled	0.10
Dairy products	
Cows' milk, full cream	0.23
Cheese, cheddar	0.39
Yogurt (whole milk, plain)	0.27
Boiled chicken's egg	0.47
Human milk	0.03
Fruits	
Apples, eating flesh and skin	0.04
Oranges, flesh	0.03
Pears, flesh and skin	0.04

(Continued)

Table 2 Riboflavin content of selected foods.—cont'd

Food	*mg per 100 g fresh wt*
Strawberries, raw	0.02
Dried mixed fruit	0.05
Vegetables	
Potatoes, boiled, new	0.02
Carrots, boiled, young	0.01
Brussels sprouts, boiled	0.07
Cauliflower, boiled	0.03
Onions, fried	0.03
Grains, grain products, nuts	
White bread	0.07
Wholemeal bread	0.05
Rice, boiled, white	Trace
Cornflakes (kellogg)	1.3
Baked beans in tomato sauce	0.03
Peanuts, plain	0.10
Other	
Marmite (yeast hydrolysate)	11.9
Bovril (beef hydrolysate)	8.5

Compiled from data in the McCance and Widdowson's composition of foods integrated dataset, 2021. Public Health England. (https://www.gov.uk/government/publications/composition-of-foods-integrated-dataset-cofid).

as some other types of malnutrition that are commonly encountered in the Third World, it can nevertheless cause debility, through skin lesions and metabolic dysfunctions. Therefore, optimization of riboflavin nutrition deserves a place in public health improvement programs.

As with most other B vitamins, riboflavin and its cofactors are remarkably nontoxic even at high intakes. The reasons for this are probably the upper limit on absorption and very rapid urinary excretion of any absorbed vitamin that exceeds cellular requirements. Some recent studies have suggested that high-dose riboflavin may benefit certain medical conditions, such as migraines, lactic acidoses, myopathies, and Leigh disease.

Conclusions

Riboflavin plays a fundamental role in the metabolism of carbohydrates, lipids and amino acids through the formation of the coenzymes flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). These coenzymes are intimately linked to the corresponding enzymes (flavoenzymes) that catalyze oxidation-reduction reactions. Therefore, these enzymes are also part of the cellular antioxidant defense. Although the coenzyme derivatives of riboflavin are not directly involved in cell proliferation phenomena, their fundamental importance in metabolism may explain why deficiency in this vitamin is particularly evident in tissues with a more rapid cell turnover, such as skin and epithelia.

References

Balasubramaniam, S., Christodoulou, J., Rahman, S., 2019. Disorders of riboflavin metabolism. J. Inherit. Metab. Dis. 42, 608–619.

Barile, M., Giancaspero, T.A., Leone, P., Galluccio, M., Indiveri, C., 2016. Riboflavin transport and metabolism in humans. J. Inherit. Metab. Dis. 39, 545–557.

Combs Jr., G.F., McClung, J.P., 2016. The Vitamins. Fundamental Aspects in Nutrition and Health, fifth ed. Academic Press, London.

European Food Safety Authority (EFSA), 2017. Dietary Reference Values for Nutrients. Summary Report. https://doi.org/10.2903/sp.efsa.2017.e15121.

Horigan, G., McNulty, H., Ward, M.M., Strain, J.J., Purvis, J., Scott, J.M., 2010. Riboflavin lowers blood pressure in cardiovascular disease patients homozygous for the 677C–>T polymorphism in MTHFR. J. Hypertens. 28, 478–486.

Institute of Medicine, 2006. Dietary Reference Intakes: The Essential Guide to Nutrient Requirements. The National Academies Press, Washington, DC. https://doi.org/10.17226/11537.

Mahabadi, N., Bhusal, A., Banks, S.W., 2020. Riboflavin Deficiency. (2020). StatPearls [Internet]. StatPearls Publishing, Treasure Island (FL).

Mosegaard, S., Dipace, G., Bross, P., Carlsen, J., Gregersen, N., Olsen, R.K.J., 2020. Riboflavin deficiency-implications for general human health and inborn errors of metabolism. Int. J. Mol. Sci. 21, 3847.

O'Callaghan, B., Bosch, A.M., Houlden, H., 2019. An update on the genetics, clinical presentation, and pathomechanisms of human riboflavin transporter deficiency. J. Inherit. Metab. Dis. 42, 598–607.

Peechakara, B.V., Gupta, M., 2020. Vitamin B2 (Riboflavin). StatPearls [Internet]. StatPearls Publishing, Treasure Island (FL).

Plantone, D., Pardini, M., Rinaldi, G., 2021. Riboflavin in neurological diseases: a Narrative review. Clin. Drug Investig. https://doi.org/10.1007/s40261-021-01038-1.

Rodwell, V.W., Bender, D.A., Botham, K.M., Kennelly, P.J., Weil, P.A., 2018. Harper's Illustrated Biochemistry, 31st ed. McGraw-Hill Education, New York.

Ross, A.C., Caballero, B., Cousins, R.J., Tucker, K.L., Ziegler, T.R., 2014. .Modern Nutrition in Health and Disease, eleventh ed. Wolters Kluwer Health, Lippincott Williams & Wilkins, Philadelphia.
Saedisomeolia, A., Ashoori, M., 2018. Riboflavin in human health: a review of current evidences. Adv. Food Nutr. Res. 83, 57–81.
Stipanuk, M.H., Caudill, M.A., 2018. Biochemical, Physiological, and Molecular Aspects of Human Nutrition, fourth ed. Saunders, Philadelphia.
Thakur, K., Tomar, S.K., Singh, A.K., Mandal, S., Arora, S., 2017. Riboflavin and health: a review of recent human research. Crit. Rev. Food Sci. Nutr. 57, 3650–3660.

Selenium

Anna Patricia Kipp, Department of Nutritional Physiology, Institute of Nutritional Sciences, Friedrich Schiller University Jena, Jena, Germany

© 2023 Elsevier Ltd. All rights reserved.

This is an update of C.D. Thomson, Selenium, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 186–192, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00248-8.

Introduction 413
Functions of selenium 413
Selenium homeostasis 414
Nutritional supply of selenium 414
Intestinal absorption and bioavailability 414
Metabolism and distribution 415
Excretion 416
Selenium requirement 416
Selenium deficiency diseases 416
Mutations and inborn errors of selenoprotein biosynthesis and functions 416
Assessment of the selenium status 417
Selenium requirements and recommended dietary intakes 417
Selenium toxicity 417
Selenium and human health 418
Immune function 418
Cancer 418
Cardiovascular disease 418
Conclusion 419
References 419

Key points

- Selenoproteins are the main mediators of selenium effects
- Newly developed reference values for selenium are based on optimizing selenoprotein P plasma levels
- A plant-based diet provides less selenium which is most pronounced in geographic areas with low soil selenium concentrations
- Both selenium deficiency and excess are resulting in impaired health

Glossary

Antioxidants Antioxidants are molecules which can safely interact with free radicals and terminate the chain reaction, thereby inhibiting the oxidation and damage to cells that may play a role in heart disease, cancer and other diseases

Kaschin–Beck disease A selenium-responsive endemic osteoarthropathy with necrosis of joints and epiphysial plate cartilage

Keshan disease An endemic cardiomyopathy responsive to selenium with clinical features of cardiac insufficiency and enlargement, electrocardiographic changes, and fibrosis

Selenium Selenium is a trace element with an atomic mass of 78.96, which is an essential nutrient for good health but required only in very small amounts. Selenium is incorporated into selenoproteins, which have a variety of functions including antioxidant, redox and thyroid function regulation

Selenoproteins Proteins requiring selenium for functionality and containing the selenoamino acid, selenocysteine, at the active site

Single nucleotide polymorphisms A single-nucleotide polymorphism (SNP) is a DNA sequence variation occurring when a single nucleotide—A, T, C, or G—in the genome (or other shared sequence) differs between members of a biological species or paired chromosomes in an individual

Introduction

Selenium is an essential trace element for humans indicating that it needs to be taken up via the food to maintain body's health. The identification of Keshan and Kaschin–Beck diseases as endemic selenium-responsive conditions, occurring in a central belt of China and areas of Russia and Africa, demonstrated conclusively that selenium is not only an essential element for man but also that deficiencies occur naturally. As it is a trace element, small amounts of selenium are sufficient for supporting growth, fertility, immune response and overall health. If provided at too high dosages, selenium becomes toxic because of its accumulation in the body resulting in selenosis. A large body of research over the past decades has provided information on the metabolism of selenium and its importance to human health, and has led to the establishment of recommended dietary intakes based on amounts required for maximal plasma selenoproteins. But the challenge still is to define the "optimal" selenium supply for humans considering additional factors such as health status, age, and stress and to maintain this supply through a healthy diet. The focus of research has now expanded further to possible benefits of selenium in preventing certain types of cancer and cardiovascular disease (CVD), and for the maintenance of an optimal immune system. In addition, rare diseases attributed to mutations in selenoproteins or the selenoprotein synthesis machinery provide further information on the relevance of selenoproteins in humans which are the main functional units of selenium in the body.

Functions of selenium

Selenium functions mainly as a constituent of selenoproteins encoded by 25 genes in humans (Kryukov et al., 2003). In addition to selenoproteins, selenium can be bound to selenium-binding proteins which are less well characterized so far and it can be incorporated as selenomethionine instead of methionine to any regular protein. The latter one, however, mainly take place under conditions of a high selenium supply (see "Selenium Toxicity") (Ferreira et al., 2021).

The selenoproteins have a number of functions but the common denominator is their participation in redox reactions and thus maintenance of redox homeostasis by their antioxidant function (e.g., glutathione peroxidases, thioredoxin reductases). Furthermore, selenoproteins are important for thyroid metabolism (iodothyronine deiodinases), immune and, reproductive function. Selenium is in the active site of all selenoproteins as selenocysteine, which is cotranslationally inserted into proteins through a process described in the Section on "**Metabolism and distribution**" (Schomburg, 2021).

GPX1 was the first selenoprotein to be characterized, and is now known to belong to a family of enzymes with eight members out of which five contain a selenocysteine in the active center while the others have cysteine instead. All of them catalyze the reduction of hydrogen peroxide and other hydroperoxides. The selenocysteine residue is oxidized by the peroxide forming selenenic acid, which is then reduced back to the selenolate by thiols, in most cases glutathione. The GPX isoenzymes differ mainly in their localization. While GPX1 and GPX4 are rather ubiquitously expressed, GPX2 is specifically located in epithelial cells, and GPX6 is only expressed in the olfactory epithelium. In contrast, GPX3 is an extracellular protein which contributes to the selenium concentration in plasma. GPX1 is one of the most highly responsive selenoproteins to changes in selenium status and deficiency. Its hepatic expression is discussed to serve as an indirect storage of selenium in the body. GPX4, also called phospholipid hydroperoxide GPX differs from the other GPXs in that it can metabolize phospholipid hydroperoxides in cell membranes, and thus may protect biomembranes against oxidation. GPX4 is also specifically involved in regulatory processes such as inhibition of lipoxygenases and ferroptosis. In sperm, GPX4 also can be transformed into a structural protein required for sperm maturation.

There are three distinct thioredoxin reductases in humans (TXNRD1, 2, 3). These are NADPH-dependent flavoprotein oxidoreductases that reduce the disulfide of thioredoxin. They support redox-regulated signaling cascades and cell proliferation and may be involved in spermatogenesis, embryonic development, and other redox-related aspects of health and disease including cancer (Roman et al., 2014).

The three iodothyronine 5′deiodinases (DIO1, 2, 3) are involved in synthesis and metabolism of thyroid hormones, which regulate most metabolic functions and are essential for growth and development. The deiodinases catalyze the conversion of thyroxine (T_4) to its active metabolite triiodothyronine (T_3) and severe selenium deficiency results in an increase in plasma T_4 and a decrease in T_3. If selenium and iodine are both deficient in a human population, the thyroid deficiency is more severe (and goiters are larger) than if only iodine is lacking, a situation which occurs in some areas of Central Africa (Schomburg and Kohrle, 2008).

SELENOP is the major selenium transport protein in plasma, providing approximately 50% of total plasma selenium. SELENOP is a glycoprotein which is the only selenoprotein that contains multiple selenocysteine residues. One of them is located in the N-terminal domain as part of a thioredoxin-like motif indicating an antioxidant activity. Up to nine additional selenocysteine residues can be found in the C-terminal domain of human SELENOP which interacts with the APOER2 receptor mediating selenium uptake into organs such as brain, testis, placenta, and bone.

The additional selenoproteins can be categorized based on their intracellular localization resulting in a major group of ER-resident selenoproteins (DIO2, SELENOM, T, F, K, S, N) which are involved in proper protein folding, degradation of misfolded proteins or calcium homeostasis. Another way of classifying selenoproteins is based on the localization of their selenocysteine within a thioredoxin-like fold. These selenoproteins (SELENOF, H, M, V, W) are called redoxins (Roman et al., 2014). Further functions of selenoproteins are described in **Table 1**.

Table 1 Functions of selenoproteins.

Selenoprotein	*Function*
Glutathione Peroxidases (GPX)	Reduction of hydroperoxides
Thioredoxin Reductases (TXNRD)	Oxidoreductase activity having NADPH as a cofactor
Iodothyronine 5′deiodinases (DIO)	Production of active thyroid hormone T3, reverse T3 (rT3) and T2
Selenophosphate Synthetase (SEPHS2)	Synthesis of selenophosphate as part of selenoprotein synthesis
SELENOF	Protein folding, redox regulation
SELENOH	Redox-sensitive DNA binding protein
SELENOI	Ethanolamine phosphotransferase, phospholipid synthesis
SELENOK	Degradation of misfolded proteins, protein palmitoylation
SELENOM	Calcium release from the ER
SELENON	Calcium homeostasis, cofactor for ryanodine receptors
SELENOO	Redox regulation in mitochondria
SELENOP	Selenium transport, redox regulation
SELENOR/Methionine Sulfoxide Reductase B1 (MSRB1)/SELENOX	Reduction of methionine R sulfoxides
SELENOS	Degradation of misfolded proteins, immune response
SELENOT	Calcium homeostasis, neuroendocrine function
SELENOV	Male reproduction
SELENOW	Redox regulation in muscle and heart

Selenium homeostasis

Nutritional supply of selenium

The main forms of selenium in foods are the selenoamino acids, selenomethionine, and selenocysteine. The inorganic forms selenite and selenate are used as supplements or can be found in the drinking water even though the latter only marginally contributes to the overall selenium supply. Selenate is also localized in plant-derived food.

Selenium is not essential for plants, but it is normally taken up and substituted in place of sulfur depending on the selenium concentration of the soil. Thus, dietary intake of selenium via plant-derived food varies greatly with geographic source of foods e.g., the selenium content of cereals and grains grown in soils poor or rich in selenium may vary 100-fold. Brazil nuts are an exceptionally good plant-based sources of selenium (concentrations up to 300 μg selenium per g) because they only grow in the Amazon region in South America which has high selenium soil contents. Some vegetables, such as garlic, mushrooms, and cabbage, can tolerate high selenium soil concentrations better than other plants and thus can accumulate selenium from soil if there are high concentrations. Therefore, these selenium accumulating plants may contain high levels of selenium that are stored as many different selenocompounds, the most prominent ones are selenomethyl selenocysteine and γ-glutamyl selenomethyl selenocysteine. Synthesis of these non-proteinogenic selenocompounds is the plant's attempt to cope with high selenium concentrations and to avoid selenium toxicity which is driven by selenomethionine that is overtly incorporated into proteins instead of methionine.

Selenomethionine is not only present in plant but also in animal foods while selenocysteine is mainly located in animal foods as part of selenoproteins. As most livestock are substituted with selenium (at least in Europe) to fulfill their selenium requirement the selenium supply via animal-derived food items is rather stable and less variable depending on the geographic origin of food production. Rich animal sources of selenium include meat, liver, shellfish, and some other types of fish. Thus, vegetarians and especially vegans take up lower total selenium amounts but a broader variety of selenocompounds as compared to people consuming omnivorous diets. Average daily intakes of selenium range from 10 μg in low-selenium areas of China, to medium intakes in New Zealand or Europe of 40–50 μg to over 200 μg or more in seleniferous areas in Venezuela and parts of the USA (Ferreira et al., 2021).

Intestinal absorption and bioavailability

Selenium is readily absorbed, especially in the small intestine. Selenoamino acids are almost completely taken up by enterocytes (70–90%) in a sodium-dependent, active manner using the methionine transporter in case of selenomethionine and the cysteine transporter for selenocysteine. Selenate is supposed to be absorbed using the solute carrier 26 (SLC26) family of multifunctional anion exchangers or via a paracellular path. Absorption efficiency for selenite is lower and mechanisms are less clear but it might be enhanced by binding to glutathione in the intestinal lumen. Intestinal absorption does not depend on the selenium status, indicating that there is no homeostatic regulation at this stage. In addition, selenium bioavailability is influenced by the amount of protein, fat, and heavy metals in the diet. Many bacteria also have a requirement for selenium including those of the intestinal microbiota. If there is enough selenium taken up by the diet bacteria contribute to the metabolism of selenocompounds which reach the lower gastrointestinal tract. However, under limited selenium supply bacteria and host appear to compete with each other resulting in lower availability of selenium for the host (Ferreira et al., 2021).

Metabolism and distribution

Fig. 1 summarizes the main pathways of interconversion of selenocompounds in mammalian tissues with a focus on liver which is the main organ for selenium homeostasis. Unlike other elements, selenium can be incorporated in two distinct ways into proteins, either as a functional active center (in specific selenoproteins) via a selective incorporation mechanism that ensures selenocysteine insertion, or alternatively by a nonspecific incorporation pathway. In this nonspecific pathway, selenomethionine (or selenocysteine) can enter the general amino acid pool and replace methionine (or cysteine) at random, without apparently conferring any special functional characteristics on recipient proteins. This is followed by a gradual release of selenium upon degradation of these proteins to again enrich the specific incorporation pathway. Inorganic selenium, in contrast, feeds directly into the specific incorporation pathway via selenide. As a result, the form administered also influences levels in tissues and retention of selenium. Thus, selenomethionine is on the one hand more effective in raising blood selenium levels than selenite or selenate but on the other hand is supposed to mediate selenium toxicity at high concentrations.

Both inorganic and organic forms of selenium are transformed to selenide, the precursor of selenophosphate which is used for selenocysteine synthesis to be inserted at the active site of selenoproteins. Selenite and selenate are reduced to selenide, organic selenocysteine is directly lyzed to selenide, and selenomethionine is transformed to selenocysteine and then lyzed to selenide. As an alternative route, selenomethionine can be degraded to methyl selenol (CH_3SeH) by γ-lyase which is further metabolized to selenide. Selenide (-2 oxidation state) is converted to selenophosphate by the selenoprotein selenophosphate synthetase 2 (SPHS2) which is the main precursor for selenocysteine synthesis. In fact, the amino acid selenocysteine is the only amino acid that does not have its own tRNA synthetase. Instead, this amino acid is synthesized on its cognate tRNA in a three step process. First, the $tRNA^{[Ser/Sec]}$ is loaded with serine which becomes phosphorylated. In the third step of the process, the synthesis of selenocysteinyl-$tRNA^{[Ser/Sec]}$ is catalyzed by selenocysteine synthase (SEPSECS).

The selenocysteinyl residue is then incorporated into selenoproteins by the UGA codon, which has the dual function of selenocysteine insertion and termination of protein synthesis. An RNA stem–loop structure, designated the selenocysteine insertion sequence (SECIS) element in the 3′ untranslated region of eukaryotic mRNAs, directs incorporation of selenocysteine into proteins by stabilizing the ribosome and selenoprotein mRNAs. Additional factors including the SECIS binding protein 2 (SECISBP2) are needed to support these processes. The turnover of selenoproteins results in the release of selenocysteine which is again degraded to selenide by selenocysteine β-lyase. In the liver, the highest expression levels are observed for the selenoproteins SELENOP and GPX1. While SELENOP ensures the selenium transport from the liver to other organs, high GPX1 levels within the liver might serve as an indirect selenium storage which can be immediately released upon limited selenium supply.

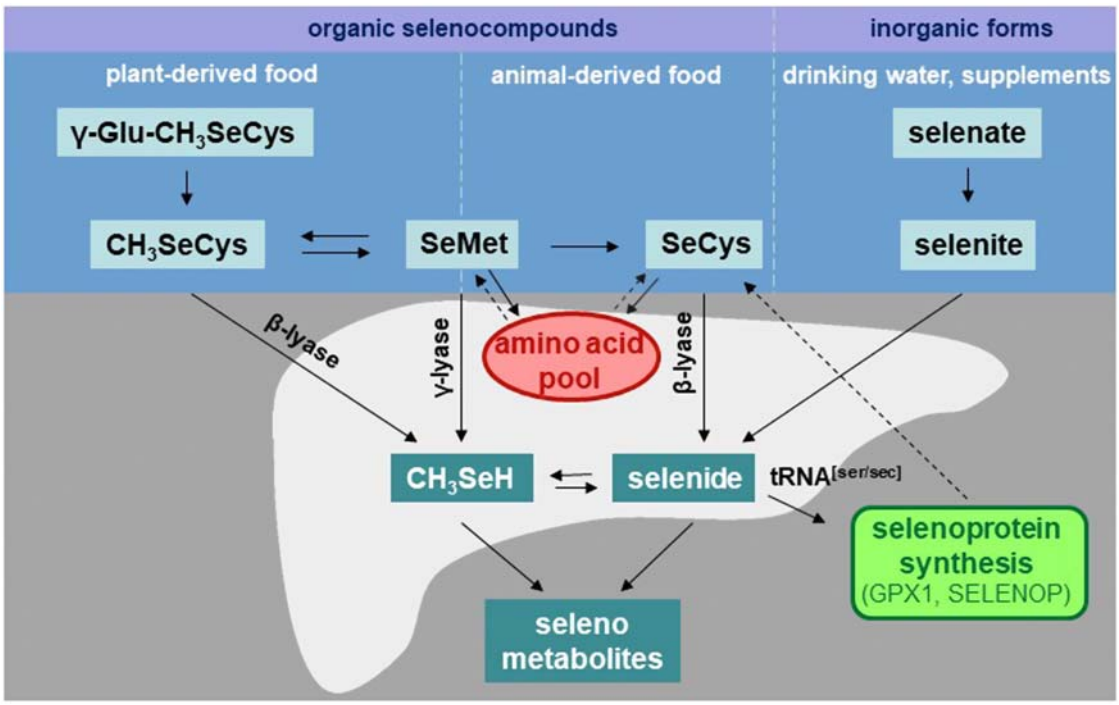


Fig. 1 Ingested selenocompounds and metabolic pathway in human liver. SeMet, selenomethionine; SeCys, selenocysteine; H_2Se, hydrogen selenide; CH_3SeCys, selenomethyl selenocysteine; γ-Glu-CH_3SeCys, γ-glutamyl selenomethyl selenocysteine; CH_3SeH, methyl selenol.

Selenoprotein synthesis is regulated by the selenium supply and there is a hierarchy of expression of individual selenoproteins and of retention of selenium in different organs and tissues. The organ-specific hierarchy under selenium deficiency mostly depends on the APOER2-mediated uptake of SELENOP from the blood maintaining e.g., brain selenium levels at the expense of other organs. Selenium crosses the placenta readily, and breast milk selenium concentration is responsive to changes in maternal selenium intake. The liver is supposed to react most sensitive toward a limited selenium supply with a drop of selenoprotein expression, especially GPX1. Nonspecific incorporation of selenomethionine into proteins contributes to tissue selenium levels but is not immediately available for synthesis of selenoproteins until protein is catabolized. Selenium levels in tissues are influenced by dietary intake, as reflected in the wide variation in blood selenium concentrations of residents of countries with differing soil selenium levels (Hong and Diamond, 2020; Roman et al., 2014).

Excretion

Urine is the main route of selenium excretion, and is thus very important for selenium homeostasis. The major urinary metabolites are selenosugars (e.g., 1 β-methyl seleno-*N*-acetyl-D-galactosamine). Urinary selenium tends to reflect recent intake rather than tissue status. Fecal selenium is mainly unabsorbed selenium. Excess selenium is methylated mainly to trimethylselenonium ion and is excreted via the urine. In addition, volatile dimethylselenide is formed which is excreted in expired air. Urinary trimethylselenonium may be used as a biological marker for excessive selenium doses (Hong and Diamond, 2020).

Selenium requirement

Selenium deficiency diseases

Keshan disease, an endemic cardiomyopathy occurring in low-selenium areas of China and Russia, is associated with low selenium intake and low levels of selenium in blood and hair, and affects mainly children and women of childbearing age. The main clinical features of Keshan disease are cardiac insufficiency and enlargement, electrocardiographic changes, and fibrosis. In 1979 Keshan disease was reported to be responsive to supplementation with sodium selenite and was initially thought to be a simple selenium deficiency. However, some features of Keshan disease (e.g., seasonal variation) cannot be explained solely by very low selenium status. Involvement of a viral factor is likely, as a strain of a Coxsackie B virus has been isolated from infected individuals. This hypothesis is strengthened by the demonstration that Coxsackie B3 virus, in the presence of selenium deficiency in mice, mutates to a more virulent form that impairs heart function.

Another condition that has been associated with severe selenium deficiency is Kaschin–Beck disease, with clinical features of osteoarthropathy and necrosis of joints and epiphysial plate cartilage. Kaschin–Beck disease occurs during preadolescent or adolescent years in rural areas of China, Tibet, and Siberia. However, other factors such as iodine deficiency or presence of mycotoxins may be more important than selenium deficiency.

Severe selenium deficiency in combination with inadequate iodine status contributes to the pathogenesis of myxedematous cretinism. Even mild to moderate selenium deficiency appears to be responsible for initiation and progression of autoimmune thyroid disorders.

Selenium deficiency has also been reported in patients on long-term intravenous nutrition, because of previously negligible amounts of selenium in the fluids. Cardiomyopathy, muscle pain, and weakness in these patients are responsive to selenium supplementation, but are not seen in all patients with low selenium status (Hong and Diamond, 2020).

Mutations and inborn errors of selenoprotein biosynthesis and functions

In 2001, the first selenoprotein human genetic disorder had been described belonging to the group of rare diseases. A congenital rigid spine muscular dystrophy was shown to be caused by mutations in the SELENON gene. Since then, different diseases have been associated to mutations in selenoproteins such as GPX4 and TXNRD1 and 2. Loss of GPX4 results in Sedaghatian disease and those patients die prematurely by cardiorespiratory failure. TXNRD1 mutations cause oxidative stress and concomitant epilepsy. Mutations in TXNRD2 result in different phenotypes, including dilated cardiomyopathy and familial glucocorticoid deficiency.

In addition, mutations in selenoprotein synthesis factors have been identified including SEPSECS, SECISBP2 and TRU-TCA1-1 encoding for $tRNA^{[Ser/Sec]}$. While loss of the two latter ones mainly affects thyroid hormone levels and down-stream target tissues such as bone, inner ear, and muscle, SEPSECS mutations, which result in an overall selenoprotein deficiency, target mainly brain and causes more severe phenotypes. Overall, the current knowledge reveals that comparable organs are affected by mutations of single selenoproteins or factors important for selenoprotein synthesis as described for nutritional selenium deficiency. Accordingly, this strengthens the concept that effects of selenium are mainly mediated by selenoproteins (Fradejas-Villar, 2018).

A growing body of research indicates that next to mutations also single nucleotide polymorphisms (SNPs) in selenoproteins most likely affect their efficiency of selenocysteine incorporation. Thus, SNPs can explain the individual variation of selenoprotein expression in response to selenium consumption, irrespective of the baseline selenium status. In addition, selenoprotein SNPs have been associated with a higher risk of a number of diseases, especially cancer but also autoimmune disease (Schomburg, 2021).

Assessment of the selenium status

Selenium status can be measured by concentrations in plasma or serum (in the following only plasma will be mentioned meaning both), whole blood or erythrocytes, or in platelets, hair, or nails. Plasma selenium reflects short-term status, erythrocyte concentration is a medium-term index, whereas hair and nail concentrations reflect longer-term status. Urinary excretion of selenium can be used to assess daily dietary intake, estimated as twice the daily excretion.

Furthermore, selenoproteins can be analyzed as functional biomarkers for the selenium status. The GPX enzymatic assay in plasma or erythrocytes is another frequently used approach because of the close relationship between plasma GPX3 or erythrocyte GPX1 activity and selenium concentrations. In situations of severe to marginal deficiency, this is a sensitive and responsive index, however, once a selenium supply is achieved, which results in plasma selenium concentrations above 80–90 $\mu g\ L^{-1}$, a plateau of activity is reached that does not respond to further increases in selenium intake. Therefore, if a population exhibits a strong correlation between plasma (or erythrocyte) selenium concentrations and GPX activity, or there is a significant increase in GPX activity after supplementation, this can be taken as evidence of suboptimal selenium status. If there is no correlation or response to supplementation, the population is likely to have adequate selenium intake.

SELENOP accounts for more than 50% of selenium in blood and has been shown to be a reliable marker of selenium status in populations with low-to-moderate selenium status. SELENOP reaches a plateau at higher plasma selenium concentrations of 100–120 $\mu g\ L^{-1}$. Thus, it is supposed to cover a wider concentration range than GPX3 and should be preferentially measured as biomarker together with plasma selenium concentrations. However, hepatic SELENOP synthesis is modulated by additional factors besides selenium resulting in decreased expression under conditions of inflammation, hypoxia or a high intake of eicosapentaenoic acid (EPA). Thus, if there is no correlation between plasma selenium and SELENOP in the concentration range of up to 120 $\mu g\ L^{-1}$ the results need to be interpreted with caution. Plasma selenium concentrations above 120 $\mu g\ L^{-1}$ generally do not correlate with SELENOP anymore because of the decline of selenoprotein biosynthesis rates under conditions of an increasing supraphysiological selenium supply. This plateauing effect on circulating selenoproteins is supposed to efficiently protect from over activation of selenoproteins over a wide intake range and increases the excretion of the surplus of selenium (Burk and Hill, 2009; Xia et al., 2010).

Because of different responses of tissues and selenoproteins to deficient, adequate, or high levels of selenium, conclusions drawn from measurement of one selenoprotein may not apply to all biological functions of selenium. It may be necessary to measure several markers of functional selenium status, in particular those that apply to specific problems associated with suboptimal selenium status. To cover increasing selenium concentrations resulting from a supranutritional intake selenium concentrations are most conclusive (Bornhorst et al., 2018).

Selenium requirements and recommended dietary intakes

The requirement to prevent selenium deficiency is based on comparison of intakes in endemic and nonendemic Keshan disease areas of China. However, dietary reference intakes in most countries are based on an estimate of the intake at which saturation of plasma GPX activity occurs, obtained from studies in China and New Zealand. This estimate indicated a physiological requirement of approximately 45 $\mu g\ day^{-1}$ (Estimated Average Requirement, EAR for USA/Canada), which translated into a recommended dietary allowance (RDA) of 55 $\mu g\ day^{-1}$. An increment of 5 $\mu g\ day^{-1}$ was added for pregnancy and 15 $\mu g\ day^{-1}$ for lactation. In contrast, the World Health Organization (WHO) recommendations are lower (40 $\mu g\ day^{-1}$) based on the premise that full saturation of GPX is unnecessary and two-thirds saturation is probably adequate (Thomson, 2004).

More recent reference values, however, consider the saturation of plasma SELENOP as adequate selenium status and were based on additional supplementation studies measuring this biomarker (Xia et al., 2010). For example, in Australian/New Zealand and Germany/Austria/Switzerland, this results in a recommended daily intake (RDI) of 70 and 60 $\mu g\ day^{-1}$ for males and females, respectively. Preventive aspects were not taken into account for calculating these reference values (Kipp et al., 2015). But the available data also indicate that a plasma selenium concentration of up to 120 $\mu g\ L^{-1}$ which should be achievable by following the RDI may reduce the risk of cancer and CVD while a further increase has no additional protective effect (see sections on "**Cancer**" and "**CVD**").

Selenium toxicity

Toxicity of selenium or selenosis may occur from consuming high-selenium foods grown in seleniferous areas in Venezuela, Colombia, northern USA and Enshi county in China or from uncontrolled usage of supplements. Loss of hair and nails is the most common sign of poisoning, and changes in hair and nails are currently the only diagnostic technique for selenium toxicity. Other overt signs of selenosis include skin damage, mottling of teeth, nerve lesions, nausea, weakness, and diarrhea. Garlic odor on the breath from breathing out dimethylselenide also indicates excessive selenium exposure. Effects of selenium toxicity are seen at chronic daily dietary intakes of above 900 μg. The upper safe limit of dietary intake (UL) was set at 400 $\mu g\ day^{-1}$ by the US/Canadian, Australian/New Zealand and FAO/WHO committees, based on a no-adverse-effect-level of 800 $\mu g\ day^{-1}$ divided by an uncertainty (i.e., safety) factor of 2. The European Food Safety Authority (EFSA) defines a tolerable upper limit for the daily selenium intake of 300 μg. Toxic effects are supposed to be mediated by selenomethionine getting randomly incorporated into proteins and disturbing their proper folding and function (Schomburg, 2021).

Selenium and human health

Immune function

Selenium is important for optimal function of both innate and acquired immune systems, and is involved in defense against bacterial and viral infections. The underlying mechanisms are likely to be related to specific selenoproteins such as SELENOS (**Table 1**) and antioxidant selenoproteins such as GPXs and TXNRDs modulating redox signaling pathways. Accordingly, they are mediating cytokine production with subsequent effects on activation, proliferation, and differentiation of immune cell types such as T cells or natural killer cells. A low selenium status is often observed during acute and chronic infections acting as a vicious cycle to further impair the immune response, a situation also recently described during the COVID-19 pandemic. Under conditions of sepsis, the decline of the selenium status is a mortality risk factor for patients. In HIV-infected individuals, progression to AIDS and decline in T helper (CD4) cell counts are accompanied by a parallel decrease in blood selenium levels. Selenium deficiency appears to increase the probability of mortality in HIV-infected subjects. Selenium supplements can improve several indices of immune function, even in individuals whose selenium status is not severely deficient. This includes stronger vaccine responses and a robust immunity to pathogens.

Studies in mice infected with strains of Coxsackie virus B3 showed that selenium deficiency results in changes in the viral genome which makes them more potent to induce heart lesions. This is relevant to the etiology of Keshan disease, which has been attributed in part to a viral factor. Selenium deficiency also causes mutational changes in another RNA virus, influenza A, and in the protozoan parasites *Trypanosoma cruzi* and *Heligmosomoides polyrus*, enhancing the intensity of infection (Avery and Hoffmann, 2018; Beck et al., 2003).

Cancer

A considerable body of evidence suggests a possible link between increased selenium intakes or status and protection against certain cancers. *In vitro* and animal studies provide evidence for a role of antioxidant selenoproteins to protect from oxidative DNA damage and accordingly to have anticarcinogenic properties during tumor initiation. Evidence from prospective studies investigating the link between low selenium status and increased incidence of cancer at various sites is still weak due to the partially poor quality of the available studies. However, the analysis of samples from the EPIC cohort, conducted in Europe, clearly revealed that low plasma selenium as well as low SELENOP levels were strongly associated with an increased risk to develop colorectal and hepatobiliary cancer. This correlation is less pronounced in areas with higher selenium intake than in Europe. This indicates that selenium plays a role in the primary prevention of cancer development but obviously there is a threshold for this protective effect.

Indeed, intervention studies such as the NPC (Nutritional Prevention of Cancer Trial) und SELECT (Selenium and Vitamin E Cancer Prevention Trial) show that increasing the systemic selenium status up to 120 $\mu g\ L^{-1}$ plasma lowers the risk to develop cancer. Above this border, as indicated by the tertile of participants with the highest initial selenium status, no further protective effects of selenium supplementation could be detected. As the selenium concentration is in the same concentration range as the one needed to optimize SELENOP levels most probably selenoproteins are mainly mediating the anticarcinogenic effects of selenium. Besides this, a direct anticarcinogenic function of methylated selenocompounds such as selenomethyl selenocysteine is still discussed.

In contrast to primary prevention of cancer by selenium, the situation appears to be completely different in cancer patients. There is strong evidence that selenium is preferentially taken up by tumor cells and that these cells profit from being well supplied with selenium because higher selenoprotein levels (e.g., GPX2, TXNRD1, SELENOF) support tumor growth and protect from apoptosis. Indeed, high expression levels of individual selenoproteins within the tumor can predict a shorter survival probability of the patient. Further studies are needed to better understand the function of selenoproteins and their response to the selenium supply during different phases of tumor development. Based on this, recommendations for the selenium intake of cancer patients should be deduced.

Very high supraphysiological selenium concentrations way beyond those needed to modulate selenoprotein expression are studied to specifically kill tumor cells as part of a therapeutic approach. Under these conditions, selenite is supposed to accumulate within tumor cells resulting in a massive oxidative stress to induce tumor cell death. But this pharmacological selenite intervention is far away from any nutrition-related modulation (Kipp, 2020; Rayman, 2012).

Cardiovascular disease

Selenium deficiency results in Keshan disease which is a cardiomyopathy but has also been correlated with an increased risk for heart failure and myocardial infarction. Also, a moderate inverse relationship between plasma selenium and coronary heart disease was demonstrated in several meta-analyses of observational studies. However, attempts to demonstrate a reduction of CVD by selenium intervention have proven disappointing. Clinical trials show no evidence of an effect of selenium supplementation on cardiovascular protection in populations with adequate selenium status. Confirmation of this came from the SELECT trial in which selenium had no beneficial effects on atherosclerosis progression. In contrast, individual selenoproteins appear to be of relevance. For example, low activity of GPX1 in erythrocytes was shown to be a predictor of cardiovascular events in patients with coronary artery disease. GPXs protect against processes relevant to atherosclerosis such as inhibition of LDL oxidation, inhibition of

proatherogenic 15-lipoxygenase, and alteration of expression of adhesion molecules induced by cytokines. Further controlled intervention trials considering modulation of selenoprotein expression are needed to clarify this situation (Shimada et al., 2021; Zhang et al., 2018).

Conclusion

The essential role of selenium in human nutrition and its discrete biochemical functions are of no doubt. Although the number of selenoproteins in the human body is now finite, the extent and diversity of their functions needs to be discovered further. Accordingly, their relationship to known consequences of selenium deficiency or to chronic disease is being clarified. This is being accelerated by genomics research, which indicates that SNPs in selenoproteins may influence susceptibility to the etiology of many diseases. Optimal human intakes of selenium are still a matter of debate because some studies have reported benefits (e.g., anticarcinogenic and immunological effects) when supplements are given, even to populations that appear to be generously supplied with the nutrient, whereas others have identified adverse effects. The distinction between nutritional and pharmacological benefits is unclear, and further trials to determine risk–benefit balance at different intake levels are needed in a range of populations considering age and gender groups. As selenium is only one micronutrient of human nutrition taken up as part of a very complex mixture of a multitude of components interactions with other nutrients and food components need to be taken into account. There appear to be interactions with other essential trace elements such as copper and zinc which are currently under investigation.

References

Avery, J.C., Hoffmann, P.R., 2018. Selenium, selenoproteins, and immunity. Nutrients 10 (9).

Beck, M.A., Williams-Toone, D., Levander, O.A., 2003. Coxsackievirus B3-resistant mice become susceptible in Se/vitamin E deficiency. Free Radic. Biol. Med. 34 (10), 1263–1270.

Bornhorst, J., Kipp, A.P., Haase, H., Meyer, S., Schwerdtle, T., 2018. The crux of inept biomarkers for risks and benefits of trace elements. Trends Anal. Chem. 104, 183–190.

Burk, R.F., Hill, K.E., 2009. Selenoprotein P-expression, functions, and roles in mammals. Biochim. Biophys. Acta 1790 (11), 1441–1447.

Ferreira, R.L.U., Sena-Evangelista, K.C.M., de Azevedo, E.P., Pinheiro, F.I., Cobucci, R.N., Pedrosa, L.F.C., 2021. Selenium in human health and gut microflora: bioavailability of selenocompounds and relationship with diseases. Front. Nutr. 8, 685317.

Fradejas-Villar, N., 2018. Consequences of mutations and inborn errors of selenoprotein biosynthesis and functions. Free Radic. Biol. Med. 127, 206–214.

Hong, L.K., Diamond, A.M., 2020. Selenium. In: Marriott, B.P., Birt, D.F., Stallings, V.A., Yates, A.A. (Eds.), Present Knowledge in Nutrition. Elsevier.

Kipp, A.P., Strohm, D., Brigelius-Flohe, R., Schomburg, L., Bechthold, A., Leschik-Bonnet, E., Heseker, H., German Nutrition, S., 2015. Revised reference values for selenium intake. J. Trace Elem. Med. Biol. 32, 195–199.

Kipp, A.P., 2020. Selenoproteins in regulation of cancer development and progression. In: Berndt, C., Lillig, C.H. (Eds.), Redox Regulation of Differentiation and De-differentiation. CRC Press.

Kryukov, G.V., Castellano, S., Novoselov, S.V., Lobanov, A.V., Zehtab, O., Guigo, R., Gladyshev, V.N., 2003. Characterization of mammalian selenoproteomes. Science 300 (5624), 1439–1443.

Rayman, M.P., 2012. Selenium and human health. Lancet 379 (9822), 1256–1268.

Roman, M., Jitaru, P., Barbante, C., 2014. Selenium biochemistry and its role for human health. Metallomics 6 (1), 25–54.

Schomburg, L., Kohrle, J., 2008. On the importance of selenium and iodine metabolism for thyroid hormone biosynthesis and human health. Mol. Nutr. Food Res. 52 (11), 1235–1246.

Schomburg, L., 2021. Selenium deficiency due to diet, pregnancy, severe illness, or COVID-19-A preventable trigger for autoimmune disease. Int. J. Mol. Sci. 22 (16).

Shimada, B.K., Alfulaij, N., Seale, L.A., 2021. The impact of selenium deficiency on cardiovascular function. Int. J. Mol. Sci. 22 (19).

Thomson, C.D., 2004. Assessment of requirements for selenium and adequacy of selenium status: a review. Eur. J. Clin. Nutr. 58 (3), 391–402.

Xia, Y., Hill, K.E., Li, P., Xu, J., Zhou, D., Motley, A.K., Wang, L., Byrne, D.W., Burk, R.F., 2010. Optimization of selenoprotein P and other plasma selenium biomarkers for the assessment of the selenium nutritional requirement: a placebo-controlled, double-blind study of selenomethionine supplementation in selenium-deficient Chinese subjects. Am. J. Clin. Nutr. 92 (3), 525–531.

Zhang, X., Li, X., Zhang, W., Song, Y., 2018. Selenium and cardiovascular disease: epidemiological evidence of a possible U-shaped relationship. In: Michalke, B. (Ed.), Selenium. Springer.

Sodium: Physiology and dietary sources

Elena Rho and Harald Seeger, Division of Nephrology, University Hospital Zurich, Zurich, Switzerland

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A.R. Michell, Sodium: Physiology, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 200–203, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00251-8.

Introduction 420
Na^+ sources, intake and absorption 421
Role of the skin 422
Osmolarity 422
ECV 422
Sodium excretion via the kidney 422
Proximal tubule 423
Loop of Henle 423
Distal tubule 423
Collecting duct 424
Systemic regulation of ECV and total body sodium 424
Afferent inputs 424
Intravascular baroreceptors 424
Intrarenal mechanisms 424
Hepatic and intestinal regulation 424
Cerebral sensing mechanisms 424
Antinatriuretic effector mechanisms 424
Sympathetic nervous system 424
Renin angiotensin aldosterone system (RAAS) 425
Antidiuretic hormone (ADH) or vasopressin (AVP) 425
Natriuretic effector mechanisms 425
Prostaglandins 425
Kallikrein-kinin system 425
Nitric oxide 425
Atrial natriuretic peptides 425
Conclusions 425
References 426

Key points

- Know the main role of Na^+ in the human body
- Know the main dietary Na^+ sources
- Know what the recommended Na^+-intake is
- Know the role of the kidney in body Na^+ regulation
- Know the role of the skin in Na^+ storage

Introduction

Acting as the osmotic skeleton of extracellular fluid, sodium is the most important cation with key roles in solute transport, membrane excitability and stability of cell volume. Sodium is the major determinant of extracelluar volume (ECV), arterial blood pressure and thus organ perfusion in the body.

The mean body content of sodium in the adult male is 90–100 g or approximately 4200 mmol (Edelman and Leibman, 1959). 40% of Na^+ is located in the extracellular fluid (10% in plasma and 30% in interstitial fluid). The main tissues that serve as Na^+ storage are bone and cartilage, muscle and connective tissue and skin (Bie, 2018). The extracellular Na^+-concentration is 135–145 mmol/L under physiological conditions, while the intracellular Na concentration is ~10 mmol/L. The large difference in Na^+ concentration between inside and outside of the cell is established and maintained by the Na^+/K^+ pump (Na^+/K^+-ATPase), which pumps Na^+ out and K^+ into the cell. The resulting gradient poses the base for the creation of a membrane potential, which

allows for essential activities like neural conduction and myocyte contraction. In addition, the sodium gradient is crucial for secondary active transport processes across membranes.

Having such central roles, it is crucial that the Na^+ concentration in the extracellular and intracellular volume (ICV) remains stable. This is achieved by a sophisticated regulation, that allows to maintain Na^+ balance, even at quite wide ranges of Na^+ intake: from 0.03 to 6 mmol/kg/day (Oliver et al., 1975). The fine regulation is only in minimal part regulated by limiting the absorption of Na^+ in the intestine and mainly mediated by Na^+ reabsorption/excretion in the kidneys. Under physiologic conditions, the kidneys excrete >95% of the ingested sodium. Na^+ has such an important role, that evolution has generally favored mutations that limited sodium deficiency. This is why we are prone to absorb large amounts of Na^+, which we eventually eliminate. In fact, omnivores, especially humans, dogs, and laboratory rats, routinely have dietary sodium intakes well above their nutritional requirement.

In the present chapter the following aspects will be reviewed: Sodium sources, absorption, storage, its physiological role in maintaining osmolarity and ECV as well as its elimination.

Na^+ sources, intake and absorption

Due to the essential role of Na^+ in the vital processes mentioned above, mammals are prone to conserve Na^+ and eliminate the excesses, only if needed. This is why practically all the ingested Na^+ gets absorbed. There is no conclusive scientific evidence for an estimated average requirement (EAR) of sodium hence only recommendations for an adequate intake (AI) exist. Sodium requirements increase with increased loss, such as heavy sweating, diarrhea or vomiting, and with increased demand (pregnancy, breastfeeding).

Salt (NaCl) provides the major source of dietary Na^+. The average NaCl intake in the world is about 10–15 g/day (Powles et al., 2013; McCarron et al., 2013), which is way above the 5 g/day (2g/day of Na) recommended by the World Health Organization (WHO) (World Health O, 2012). The WHO recommendation is based on evidence that an intake above this threshold is *associated* with an increase in blood pressure and adverse cardiovascular outcomes (Hunter et al., 2022). Various studies reported in part controversial outcomes of salt-intake on cardiovascular mortality. However, the two biggest metanalyses about this subject concluded that increased salt intake is associated with increased cardiovascular morbidity and mortality (Strazzullo et al., 2009; Ma et al., 2022).

Interestingly the main source of Na^+ in the western diet are additives that are used in order to preserve food, making the individual effort to reduce salt intake difficult (**Fig. 1**). Population-based approaches might be more effective, as demonstrated in the UK and Finland, where government-based initiatives led to a reduction of salt consumption by 15% and 40%, respectively (He et al., 2014; Laatikainen et al., 2006). However, these kind of efforts are often seen as not cost-effective in the short term and some experts consider salt-reduction to be neither desirable, nor possible (McCarron et al., 2019).

The main sources of Na^+ in the US are: breads and rolls, pizza, sandwiches, cold cuts and cured meats, soups, savoury snacks (chips, crackers, pop-corn etc), chicken, cheese, eggs and their derivatives (U.S. Department of Agriculture ARS, 2016). Vegetable contains in comparison far less salt, which is why evolutionary mechanisms led to a much more pronounced salt appetite in herbivores compared to carnivores. Some herbivores lick mineral salt in order to compensate the missing Na-intake coming from meat and animal products (Denton and Sabine, 1961). Of the five basic tastes: sweet, umami, sour, bitter and salty, the first two are attractive, the third and the fourth are repulsive and the salty taste is the only one which induces a biphasic reaction. It induces an appetitive reaction at concentrations <500 mM and an aversive one at concentrations > 500 mM (Oka et al., 2013). This might be one of the many evolutionary mechanisms protecting against a too low or to high Na^+-intake, which, if massive, might lead to emesis and death (Campbell and Train, 2017).

The absorption of Na^+ takes place mainly in the small intestine and to a smaller extent in the colon. Sodium is absorbed by both passive and secondary active transport. The polarized intestinal cells express transport proteins, which allow movement of Na^+ from the intestinal lumen into the enterocytes. On the basolateral side, the Na^+/K^+-ATPase ("Na^+/K^+ pump") actively pumps Na^+ out of

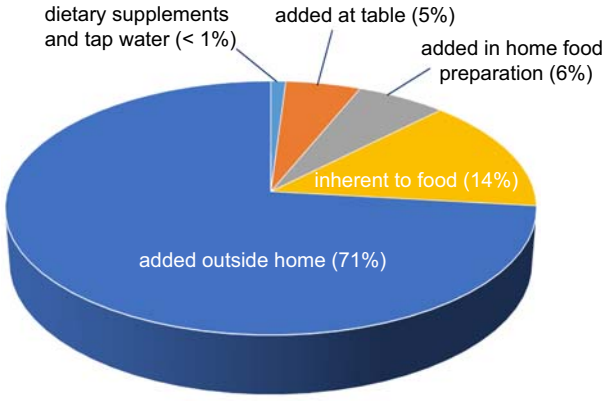


Fig. 1 Proportion of sodium intake from different sources in a typical Western diet. Modified from Harnack et al. (2017).

the enterocyte into the interstitium, from where it enters the circulation. This causes a sodium gradient, which drives secondary active transport processes. The main channel and transport molecules in the gut involved with sodium absorption are NHE (sodium/hydrogen exchanger) 3 and 2, SGLT1 (sodium glucose cotransporter 1), Na^+/amino acid cotransporters (e.g. Slc6, Slc38) Na^+/phosphate cotransporters and ENaC (epithelial sodium channel) in the colon. Parallel to Na^+ absorption H2O absorption occurs through a paracellular pathway.

Under normal conditions without significant sweating, total obligatory Na+ losses are minute and usually do not exceed eight mmol (0.18g) per day. Obligatory sodium loss occurs via urine (~0.2–1.5 mmol/d), skin (~1.1 mmol/d) and feces (~0.4–5.4 mmol/d) (Dahl, 1958).

Role of the skin

For years, it has been assumed that sodium intake has to equal sodium excretion in order to maintain sodium balance. Consequently, a 24h urine sample would reflect sodium intake. This is only partially true. Of course, in order to maintain sodium-balance the Na^+ intake has to be equal to the amount of Na^+ excreted, but Titze et al. (2016) demonstrated that this balance is not reached within 24 h. They investigated how urine sodium correlated with sodium intake in a very controlled environment: in astronauts, who were simulating a mission to mars and were receiving a determined amount of NaCl with their diet: either 3, 6 or 9g/day. Surprisingly, the 24h urine collection could not discriminate between the three groups, but the probability of discriminating increased to 92% if urine was collected for 7 days (Lerchl et al., 2015). They hypothesized that Na^+ has to be stored in the body and that its excretion is not only dependent on intake, but also on a weekly (circaseptan) pattern. Through Na^+ magnetic resonance they could demonstrate that Na^+ gets stored in the skin and that the stores are larger in men, in older and in hypertensive people (Kopp et al., 2013). It has been observed that Na^+ binds to negatively charged glycosaminoglycan (Titze et al., 2004) and accumulates even more at the site of bacterial infection (Jantsch et al., 2015) and it has been hypothesized that this (antibacterial effect) is probably one of the teleological reasons of why Na^+ is stored in the skin.

Osmolarity

The plasma osmolarity expresses the number of osmoles (particles that are osmotically active) in 1 L of plasma. In humans plasma osmolarity in physiologic conditions is 275–299 mmol/L. Sodium accounts for 88% of extracelluar fluid *osmolality* (osmoles of solute per kg of solvent). Considering that cell membranes are freely permeable to water, plasma and intracellular osmolarity are the same. It is essential that osmolarity remains in the above-mentioned range because otherwise the shift of water could lead to shrinkage or swelling of the cells compromising their function and eventually leading to cell death. Plasma osmolarity is regulated by the amount of water in the body. Slight increases in osmolarity elicit thirst and lead to the release of the antidiuretic hormone (ADH), which by increasing the expression of acquaporins on the luminal side of the kidney collecting duct epithelial cells increases the renal reabsorption of free water. Regulation of sodium reabsorption or excretion does not have an effect on plasma osmolarity, but rather on body fluid volume. Na^+ is the main extracellular active osmol and knowing the formula of calculated blood osmolarity makes us understand what an important role Na^+ has in determining it: Posm = 2 [Na^+] + glucose (mmol/L) + urea (mmol/L).

ECV

Being the main osmotic cation Na^+ is also responsible for maintaining the extracellular volume. In virtue of its osmotic activity, water always moves along with sodium. The regulation of Na^+ and ECV is essential to allow life and is tightly regulated through afferent and efferent inputs that, in physiological state, allow the maintenance of ECV in a quite narrow range.

The afferent inputs give information about volume status, which then activate effector mechanisms, that are responsible to increase or decrease ECV, mainly by regulating Na^+ excretion via the kidneys.

In order to facilitate the readers' comprehension, Na^+-excretion in the kidneys will be discussed first and regulatory mechanisms described subsequently.

Sodium excretion via the kidney

It has been known since decades, that the kidney has an essential role in the excretion of sodium and in its regulation. Sodium, being a small cation, is freely filtered in the glomerulum. Every day about 1.7 kg of NaCl, which corresponds to about 30 mol of Na^+, is filtered, but only 0.1–2% of it is actually excreted in the urine. The renal tubules are responsible for the reabsorption of Na^+. The main energetic drive for this process in all tubular segments is the Na^+/K^+-ATPase in the basolateral membrane. The different segments of the tubule contribute to varying degrees and through different transport-mechanisms to this process. A schematic view of sodium reabsorption is presented in Fig. 2.

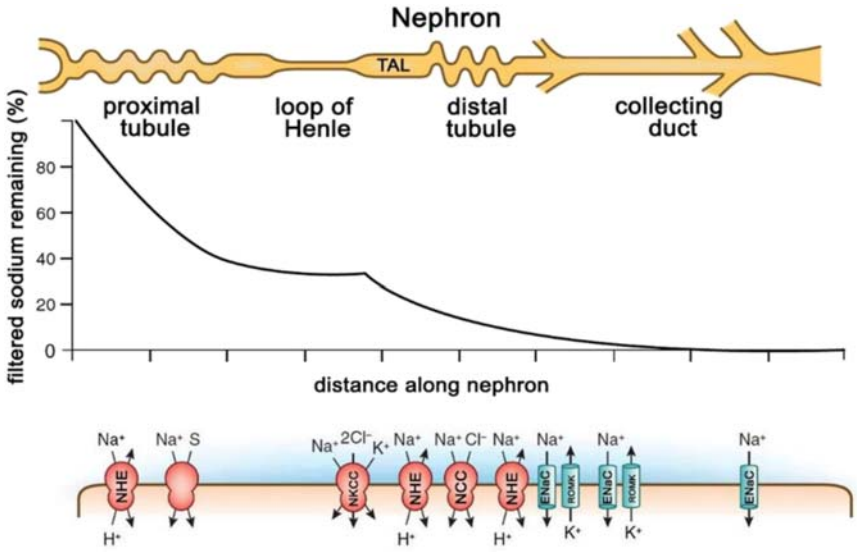


Fig. 2 Sodium reabsorption along the nephron. The percentage of sodium remaining in the ultrafiltrate is displayed along the nephron under circumstances of normal sodium intake. Bulk reabsorption takes place in the proximal tubule and TAL, whereas sodium excretion is regulated in the distal nephron. Symbols below show key Na^+ transporters and channels of each tubular segment. NHE, sodium-hydrogen antiporter; NKCC2, $Na^+K^+2Cl^-$cotransporter; NCC, Na^+Cl^- cotransporter; TAL, thick ascending limb of the loop of Henle; ENaC, epithelial sodium channel; ROMK, renal outer medullary potassium channel. Modified from Palmer and Schnermann (2015).

Proximal tubule

60–70% of the filtered sodium is reabsorbed in the proximal tubule. The whole process is possible largely due to the Na^+/K^+-ATPase. The main transporter through which Na^+ is reabsorbed from the luminal side is the Na^+/H^+ exchanger (NHE) (Alpern, 1985) in the proximal convoluted tubule (PCT). In addition, the negatively charged interstitium serves as a driving force for the paracellular reabsorption of Na^+, which is facilitated by rather leaky tight junctions of the proximal tubule. As matter of fact, the passive reabsorption of Na^+ in the proximal tubule can even exceed the active one, accounting for up to 2/3 of the total (Alpern et al., 1985; Seldin et al., 1991). Other mechanisms that contribute to Na^+-reabsorption in the proximal tubule are cotransporters of Na^+ and other molecules like glucose (mainly through the SGLT2) amino acids, phosphate, lactate and sulfate (Petrovic et al., 2004) (**Fig. 2**).

Factors that upregulate Na^+ reabsorption are cytosolic acidosis, Angiotensin II and sympathetic innervation. Factors that downregulate it are cytosolic alkalosis, dopamine and PTH (parathyroid hormone). Water moves along with sodium, so that in the permeable proximal tubule, H2O gets proportionally reabsorbed and the osmolality at the end of the proximal tubule remains isotonic.

Loop of Henle

The loop of Henle contributes 20–30% of sodium reabsorption. The thin descending limb is water permeable due to the presence of aquaporin-1, but impermeable to solutes. The ascending limb on the contrary is impermeable to water and the main site of Na^+ reabsorption, especially in the thick ascending limb (TAL). In this segment, sodium reabsorption is driven again by the Na^+/K^+-ATPase. On the luminal membrane, Na^+ absorption occurs via the $Na^+/K^+/2Cl^-$ Co-Transporter (NKCC), which is also the target of loop diuretics. The negatively charged interstitium attracts positive cations like Ca^{++}, Mg^{++} and Na^+, which are also reabsorbed via a paracellular way. Also in the loop of Henle, the paracellular pathway is responsible for a relevant proportion (50%) of the Na^+ reabsorption. ADH, PTH, calcitonin, glucagon and adrenaline increase the Na^+ and Cl^+ reabsorption in the TAL. On the contrary, prostaglandin E2 inhibits this process (Higashihara et al., 1979). Moreover, the amount of reabsorbed sodium and chloride is also dependent on the delivered load: the more delivered, the more reabsorbed.

Distal tubule

The distal tubule, like the TAL of the loop of Henle is not permeable to water, but mainly a site of ion reabsorption. Also here the major driving force for sodium reabsorption is the Na^+/K^+-ATPase on the basolateral membrane. By extruding sodium from the cell into the interstitium it creates the gradient that leads to Na^+ absorption from the luminal side. This process happens through the NaCl cotransporter (NCC), which is sensitive to thiazide diuretics.

A very important regulatory mechanism that is elicited by the NaCl concentration in the distal tubule is called tubuloglomerular feedback (TGF). Specific cells of the distal convoluted tubule (macula densa cells), that are in proximity of the glomerulus, belong to

the so-called juxtaglomerular apparatus. These cells by sensing a higher NaCl load can activate a cascade, which leads to the contraction of the afferent arteriole and by doing so to a reduction of the single nephron GFR leading to a decrease in the amount of filtered sodium. In this way, the nephron is protected from an excessive loss of NaCl.

Collecting duct

The collecting duct (CD) has the role of fine-tuning the reabsorption of Na^+ and water. It has three types of cells: principal cells are mainly responsible for water and sodium reabsorption whereas the A and B intercalated cells are involved in acid/base regulation. As described above, water reabsorption is regulated by ADH. The atrial natriuretic peptides (ANP) have the opposite effect, i.e. increase of sodium excretion and diuresis. Na^+ absorption is once again driven by the Na^+/K^+-ATPase. Reduced intracellular Na^+-concentration leads to an increased absorption from the luminal side through the epithelial sodium channel (ENaC), which is the target of the drug amiloride. The mineralcorticoid aldosterone increases the number and activity of ENaC leading to an increase in Na^+ absorption, expansion of ECV, hypertension, hypokalaemia and metabolic alkalosis.

The mechanisms of sodium retention are impressive in their efficiency. Under salt restriction or loss, urine may become practically sodium-free. Urinary sodium levels may fall well beneath 0.5 mmol/L, which is equivalent to a NaCl concentration of 10 mg/L (less than the mean sodium content of tap water). This means that sodium deficiency does not occur with normal renal function as long as food and water supply is adequate. Therefore, sodium deficiency is a consequence of an acquired or hereditary disease (e.g. chronic diarrhea, renal tubulopathy such as Gitelman syndrome) or an adverse effect of a medical treatment (excessive therapy with diuretics).

Systemic regulation of ECV and total body sodium

Afferent inputs

Intravascular baroreceptors

There are many afferent signals providing information about volume status both from the venous and arterial side and from the heart itself. Low-pressure baroreceptors are located in capacitance vessels in the thorax and in the atria. In the case of volume expansion these receptors transmit through the vagus nerve to the medulla oblongata, which decreases the sympathetic tone (Oren et al., 1993). The stretching of the atria causes a decrease in antidiuretic hormone (ADH) and renin secretion, as well as the secretion of ANP and brain natriuretic peptide (BNP), which increase the natriuretic response (Jensen et al., 1998).

On the arterial side, baroreceptors are present in the aorta and carotid arteries. High blood pressure stimulates these baroreceptors, which through the glossopharyngeal and vagus nerve stimulate centers in the medulla, which in turn decrease sympathetic tone and by doing so increase natriuresis (Ludbrook and Ventura, 1996; Potts et al., 1996).

Intrarenal mechanisms

Another important way to regulate Na^+ excretion and the ECV, is to maintain intraglomerular pressure and glomerular filtration rate (GFR) constant. Two important intrarenal feedback mechanisms achieve this aim ("renal autoregulation"). One is the myogenic reflex, which consists of the ability of the afferent arteriole of the glomerulus to dilate or constrict in response to systemic pressure changes. The second one is TGF. Also here, the afferent arteriole regulates glomerular filtration pressure, depending on the amount of delivered NaCl to the macula densa: it will dilate if the delivered NaCl is low thereby increasing filtration and constrict if it is high and by this decrease filtration. This kind of feedback allows to maintain the single nephron GFR in a quite narrow range. A more sustained reduction in NaCl delivery to the macula densa will also cause the secretion of renin by the juxtaglomerular cells, which in turn increases the systemic and renal perfusion pressure.

Hepatic and intestinal regulation

Within the hepatic circulation there are osmoreceptors and NaCl sensitive receptors, which sense the quantity of sodium absorbed by the intestine (Kostreva et al., 1980). In the case of increased stimulation, the release of ADH is increased and sympathetic activity decreases resulting in a natriuric effect. There is also a hepatointestinal reflex by which increased stimulation of these receptors decreases Na^+ uptake from the intestine (Morita et al., 1997).

Cerebral sensing mechanisms

In the central nervous system, there are osmoreceptors, which allow modulating the secretion of ADH and by doing so, to maintain plasma osmolarity in a very narrow range.

Antinatriuretic effector mechanisms

Sympathetic nervous system

As mentioned above, an increase in sympathetic activity is the typical response to a decreased circulating volume. Postganglionic axons originate in the celiac and paravertebral ganglia and innervate the kidneys. The main sympathetic effector mechanisms which lead to sodium retention are:

- Sympathetic nervous system (SNS) activation increases the systemic blood pressure and by doing so also the renal perfusion pressure and the GFR
- SNS stimulation leads to contraction of the efferent arteriole > contraction of the afferent arteriole, leading to an increased intraglomerular pressure and in this way to an increased GFR. The increased intraglomerular hydrostatic pressure leads to a reduced hydrostatic pressure in the peritubular capillaries, increasing the paracellular reabsorption of Na^+ in the proximal tubule
- The SNS increases the renin release by juxtaglomerular cells and has a direct effect on the different segments of the tubule leading to an increased Na^+ reabsorption

Renin angiotensin aldosterone system (RAAS)

The RAAS is the major effector mechanism for ECV maintenance causing Na^+ retention. Renin release is stimulated by hypovolemia or low perfusion pressure, SNS activity and reduced delivery of NaCl to the distal tubule. Renin converts angiotensinogen, a peptide produced by the liver, into angiotensin I, which is then cleaved to angiotensin II. Angiotensin II has direct and indirect antinatriuretic effects. It stimulates the production of the mineralocorticoid aldosterone by the adrenal cortex. Aldosterone increases ENaC expression in the luminal membrane of collecting duct cells thereby leading to Na^+ reabsorption.

Antidiuretic hormone (ADH) or vasopressin (AVP)

In the case of decreased effective circulating volume, venous and arterial baroreceptors stimulate the release of ADH by the hypothalamus. The main effect of ADH is to increase the expression of aquaporin-2 in the collecting duct increasing the reabsorption of free water. It has also an antinatriuretic effect by increasing Na^+ reabsorption in the loop of Henle and in the collecting duct.

Natriuretic effector mechanisms

Prostaglandins

Prostaglandins (PG) derive from the metabolism of arachidonic acid. PGI2 is produced by endothelial cells in the renal cortex and its production is stimulated by renin. PGE2 in the kidney is produced by interstitial and collecting duct cells and its release is stimulated by angiotensin II. Even if their production is triggered by the RAAS they have an antagonistic role i.e. they have a natriuretic function and lead to local vasodilation. Their role is to oppose the effects of RAAS in order to prevent an excess in vasoconstriction, which eventually could lead to ischemia for example in the medulla. PGs exert their functions through the following mechanisms (Higashihara et al., 1979):

- increase of renal blood flow and increased Na^+ delivery to the nephron
- vasodilatation of the efferent arteriole, which reduces the intraglomerular hydrostatic pressure and increases it in the peritubular capillaries reducing the paracellular reabsorption of Na+ in the proximal tubule
- PGs increase blood flow to the medulla and by doing so, diminish the interstitial solute concentration in the medulla, leaving more water in the loop of Henle and by that reducing the luminal Na^+ concentration at this level thereby decreasing its paracellular reabsorption
- They reduce directly the Na reabsorption, probably by diminishing the activity of the Na^+/K^+-ATPase

Kallikrein-kinin system

Renal kallikrein is a protease which produces two kinins that have vasodilatory effects, which leads to a washout of the medullary interstitium and a decreased Na^+ reabsorption.

Nitric oxide

Ntric oxide seems support volume induced natriuresis, which is mediated by its vasodilatory properties.

Atrial natriuretic peptides

Increased stretching of the atria leads to production of the atrial natriuretic peptides: ANP and BNP. Their natriuretic effector mechanisms are:

- decrease in renin, angiotensin II and aldosterone production
- increase in medullary blood flow and thereby washout of the medulla (Davis and Briggs, 1987)
- reduction of Na^+ reabsorption in the medullary collecting duct (Sonnenberg et al., 1986)

Conclusions

The Na^+ content of the body is responsible for maintaining extracellular volume and its homeostasis has to be tightly regulated. The main regulator of body sodium undoubtedly is the kidney. In recent years however, the skin has gained attention as another important regulator via its ability to store sodium in an osmotically inactive form.

Na^+ has an essential role in life. Evolution in non-marine species thus has prioritized mutations that favored Na^+-intake and retention. However, in today's Western societies, humans are exposed to an oversupply of sodium. We are therefore biologically and socially prone to exceed physiologic Na^+-intake. A greater Na^+-intake can lead to an increase in blood pressure and might also intensify the risk for adverse cardiovascular outcomes even though conclusive interventional data is missing. In addition, Na^+ has been implicated as an aggravator of autoimmune inflammatory disorders. Thus, strategies ought to be developed to moderate intake in individuals with high sodium consumption.

References

Alpern, R.J., Howlin, K.J., Preisig, P.A., October 1985. Active and passive components of chloride transport in the rat proximal convoluted tubule. J. Clin. Invest. 76 (4), 1360–1366. https://doi.org/10.1172/JCI112111.

Alpern, R.J., November 1985. Mechanism of basolateral membrane H+/OH-/HCO-3 transport in the rat proximal convoluted tubule. A sodium-coupled electrogenic process. J. Gen. Physiol. 86 (5), 613–636. https://doi.org/10.1085/jgp.86.5.613.

Bie, P., November 1, 2018. Mechanisms of sodium balance: total body sodium, surrogate variables, and renal sodium excretion. Am. J. Physiol. Regul. Integr. Comp. Physiol. 315 (5), R945–R962. https://doi.org/10.1152/ajpregu.00363.2017.

Campbell, N.R.C., Train, E.J., June 23, 2017. A systematic review of fatalities related to acute ingestion of salt. A need for warning labels? Nutrients 9 (7). https://doi.org/10.3390/nu9070648.

Dahl, L.K., June 5, 1958. Salt intake and salt need. N. Engl. J. Med. 258 (23), 1152–1157. https://doi.org/10.1056/NEJM195806052582305.

Davis, C.L., Briggs, J.P., October 1987. Effect of atrial natriuretic peptides on renal medullary solute gradients. Am. J. Physiol. 253 (4 Pt 2), F679–F684. https://doi.org/10.1152/ajprenal.1987.253.4.F679.

Denton, D.A., Sabine, J.R., June 1961. The selective appetite for Na ions shown by Na ion-deficient sheep. J. Physiol. 157, 97–116. https://doi.org/10.1113/jphysiol.1961.sp006708.

Edelman, I.S., Leibman, J., August 1959. Anatomy of body water and electrolytes. Am. J. Med. 27, 256–277. https://doi.org/10.1016/0002-9343(59)90346-8.

Harnack, L.J., Cogswell, M.E., Shikany, J.M., et al., May 9, 2017. Sources of sodium in US adults from 3 geographic regions. Circulation 135 (19), 1775–1783. https://doi.org/10.1161/CIRCULATIONAHA.116.024446.

He, F.J., Brinsden, H.C., MacGregor, G.A., June 2014. Salt reduction in the United Kingdom: a successful experiment in public health. J. Hum. Hypertens. 28 (6), 345–352. https://doi.org/10.1038/jhh.2013.105.

Higashihara, E., Stokes, J.B., Kokko, J.P., Campbell, W.B., DuBose Jr., T.D., November 1979. Cortical and papillary micropuncture examination of chloride transport in segments of the rat kidney during inhibition of prostaglandin production. Possible role for prostaglandins in the chloruresis of acute volume expansion. J. Clin. Invest. 64 (5), 1277–1287. https://doi.org/10.1172/JCI109583.

Hunter, R.W., Dhaun, N., Bailey, M.A., May 2022. The impact of excessive salt intake on human health. Nat. Rev. Nephrol. 18 (5), 321–335. https://doi.org/10.1038/s41581-021-00533-0.

Jantsch, J., Schatz, V., Friedrich, D., et al., March 3, 2015. Cutaneous Na+ storage strengthens the antimicrobial barrier function of the skin and boosts macrophage-driven host defense. Cell Metabol. 21 (3), 493–501. https://doi.org/10.1016/j.cmet.2015.02.003.

Jensen, K.T., Carstens, J., Pedersen, E.B., January 1998. Effect of BNP on renal hemodynamics, tubular function and vasoactive hormones in humans. Am. J. Physiol. 274 (1), F63–F72. https://doi.org/10.1152/ajprenal.1998.274.1.F63.

Kopp, C., Linz, P., Dahlmann, A., et al., March 2013. 23Na magnetic resonance imaging-determined tissue sodium in healthy subjects and hypertensive patients. Hypertension 61 (3), 635–640. https://doi.org/10.1161/HYPERTENSIONAHA.111.00566.

Kostreva, D.R., Castaner, A., Kampine, J.P., May 1980. Reflex effects of hepatic baroreceptors on renal and cardiac sympathetic nerve activity. Am. J. Physiol. 238 (5), R390–R394. https://doi.org/10.1152/ajpregu.1980.238.5.R390.

Laatikainen, T., Pietinen, P., Valsta, L., Sundvall, J., Reinivuo, H., Tuomilehto, J., August 2006. Sodium in the Finnish diet: 20-year trends in urinary sodium excretion among the adult population. Eur. J. Clin. Nutr. 60 (8), 965–970. https://doi.org/10.1038/sj.ejcn.1602406.

Lerchl, K., Rakova, N., Dahlmann, A., et al., October 2015. Agreement between 24-hour salt ingestion and sodium excretion in a controlled environment. Hypertension 66 (4), 850–857. https://doi.org/10.1161/HYPERTENSIONAHA.115.05851.

Ludbrook, J., Ventura, S., May 1996. Roles of carotid baroreceptor and cardiac afferents in hemodynamic responses to acute central hypovolemia. Am. J. Physiol. 270 (5 Pt 2), H1538–H1548. https://doi.org/10.1152/ajpheart.1996.270.5.H1538.

Ma, Y., He, F.J., Sun, Q., et al., January 20, 2022. 24-Hour urinary sodium and potassium excretion and cardiovascular risk. N. Engl. J. Med. 386 (3), 252–263. https://doi.org/10.1056/NEJMoa2109794.

McCarron, D.A., Kazaks, A.G., Geerling, J.C., Stern, J.S., Graudal, N.A., October 2013. Normal range of human dietary sodium intake: a perspective based on 24-hour urinary sodium excretion worldwide. Am. J. Hypertens. 26 (10), 1218–1223. https://doi.org/10.1093/ajh/hpt139.

McCarron, D.A., Geerling, J.C., Alderman, M.H., March 30, 2019. Urinary sodium excretion measures and health outcomes. Lancet 393 (10178), 1294–1295. https://doi.org/10.1016/s0140-6736(19)30232-6.

Morita, H., Yamashita, Y., Nishida, Y., Tokuda, M., Hatase, O., Hosomi, H., March 1997. Fos induction in rat brain neurons after stimulation of the hepatoportal Na-sensitive mechanism. Am. J. Physiol. 272 (3 Pt 2), R913–R923. https://doi.org/10.1152/ajpregu.1997.272.3.R913.

Oka, Y., Butnaru, M., von Buchholtz, L., Ryba, N.J., Zuker, C.S., Feb 28, 2013. High salt recruits aversive taste pathways. Nature 494 (7438), 472–475. https://doi.org/10.1038/nature11905.

Oliver, W.J., Cohen, E.L., Neel, J.V., July 1975. Blood pressure, sodium intake, and sodium related hormones in the Yanomamo Indians, a "no-salt" culture. Circulation 52 (1), 146–151. https://doi.org/10.1161/01.cir.52.1.146.

Oren, R.M., Schobel, H.P., Weiss, R.M., Stanford, W., Ferguson, D.W., June 1993. Importance of left atrial baroreceptors in the cardiopulmonary baroreflex of normal humans. J. Appl. Physiol. 74 (6), 2672–2680. https://doi.org/10.1152/jappl.1993.74.6.2672.

Palmer, L.G., Schnermann, J., April 7, 2015. Integrated control of Na transport along the nephron. Clin. J. Am. Soc. Nephrol. 10 (4), 676–687. https://doi.org/10.2215/CJN.12391213.

Petrovic, S., Barone, S., Weinstein, A.M., Soleimani, M., August 2004. Activation of the apical Na+/H+ exchanger NHE3 by formate: a basis of enhanced fluid and electrolyte reabsorption by formate in the kidney. Am. J. Physiol. Ren. Physiol. 287 (2), F336–F346. https://doi.org/10.1152/ajprenal.00400.2003.

Potts, J.T., Hatanaka, T., Shoukas, A.A., March 1996. Effect of arterial compliance on carotid sinus baroreceptor reflex control of the circulation. Am. J. Physiol. 270 (3 Pt 2), H988–H1000. https://doi.org/10.1152/ajpheart.1996.270.3.H988.

Powles, J., Fahimi, S., Micha, R., et al., December 23, 2013. Global, regional and national sodium intakes in 1990 and 2010: a systematic analysis of 24 h urinary sodium excretion and dietary surveys worldwide. BMJ Open 3 (12), e003733. https://doi.org/10.1136/bmjopen-2013-003733.

Seldin, D.W., Preisig, P.A., Alpern, R.J., March 1991. Regulation of proximal reabsorption by effective arterial blood volume. Semin. Nephrol. 11 (2), 212–219.

Sonnenberg, H., Honrath, U., Chong, C.K., Wilson, D.R., June 1986. Atrial natriuretic factor inhibits sodium transport in medullary collecting duct. Am. J. Physiol. 250 (6 Pt 2), F963–F966. https://doi.org/10.1152/ajprenal.1986.250.6.F963.

Strazzullo, P., D'Elia, L., Kandala, N.B., Cappuccio, F.P., November 24, 2009. Salt intake, stroke, and cardiovascular disease: meta-analysis of prospective studies. BMJ 339, b4567. https://doi.org/10.1136/bmj.b4567.

Titze, J., Shakibaei, M., Schafflhuber, M., et al., July 2004. Glycosaminoglycan polymerization may enable osmotically inactive Na+ storage in the skin. Am. J. Physiol. Heart Circ. Physiol. 287 (1), H203–H208. https://doi.org/10.1152/ajpheart.01237.2003.

Titze, J., Rakova, N., Kopp, C., Dahlmann, A., Jantsch, J., Luft, F.C., July 2016. Balancing wobbles in the body sodium. Nephrol. Dial. Transplant. 31 (7), 1078–1081. https://doi.org/10.1093/ndt/gfv343.

U.S. Department of Agriculture ARS, 2016. Nutrient Intakes From Food and Beverages: Mean Amounts Consumed Per Individual, by Gender and Age, What We Eat in America. NHANES 2013–2014.

World Health Organization, 2012. Effect of Reduced Sodium Intake on Blood Pressure, Renal Function, Blood Lipids and Other Potential Adverse Effects. World Health Organization.

Thiamin: Beriberi

D.I. Thurnham, University of Ulster, Coleraine, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of D.I. Thurnham, Thiamin: Beriberi, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 264–273, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00265-8.

Introduction 429
Epidemiology 429
Etiology 431
Experimental thiamin deficiency in man and measurement of thiamin status 432
Clinical features of beriberi 433
Infantile beriberi 433
Accidental thiamin depletion in infants 433
Alcohol abuse, the Wernicke-Korsakoff syndrome and morbid obesity 435
Management/treatment 436
Lipid-soluble thiamin derivatives 437
Case study 437
Toxicity 438
Beriberi conclusions 438
References 438
Further reading 439

Key points

- There is no storage of thiamin in vivo and turnover is rapid, hence a regular dietary supply of thiamin is needed.
- Refined foods e.g., polished rice or white wheat flour are poor sources of thiamin and should either be fortified with thiamin or the diet enriched by eating with other foods.
- Epidemiologically, beriberi can occur where food eaten is poor, dietary diversity is restricted, physical activity is high, food intake has a high calorie to thiamin ratio, there is fever and infection and occurs more often in men than women.
- Experimental thiamin deficiency is followed after 2–3 weeks by progressively worsening signs and symptoms of disease.
- Acute deficiency of thiamin more often affects the younger and stronger members of a community and symptoms are predominantly cardiac; the "wet" form of beriberi. Usually responds rapidly to thiamin treatment.
- The "dry" form of beriberi affects mainly older people. It predominantly causes neurological impairment and ophthalmoplegia. It may be slow to respond to thiamin treatment and may represent a long-term chronic deprivation of thiamin.
- Ethyl alcohol inhibits thiamin absorption hence alcohol abuse can cause acute beriberi which can further deteriorate to Wernicke's Encephalopathy characterized by cardiac symptoms, peripheral neurological impairment and ophthalmoplegia.
- Severe dietary restriction in the absence of thiamin supplements, e.g., following bariatric surgery, can increase the risk of thiamin deficiency and Wernicke's encephalopathy
- Infantile beriberi characteristically occurs in infants 2–5 months fed solely on breast milk from mothers at risk of thiamin deficiency due to poverty and/or a thiamin-deficient diet.
- The fundamental importance of thiamin as a source of energy can mean that anorexia and illness can implicate thiamin deficiency in a wide variety of Thiamin Deficient Disorders.

Glossary

Ataxic Gait Inability to coordinate leg muscles
Confabulation Description of events that have not happened
Dysesthesia Sensation typically described as painful, itchy, burning or restrictive. It results from nerve damage
Dyspnoe Labored or difficult breathing in sleep (paroxysmal nocturnal dyspnea is a respiratory distress that awakens patients from sleep)
Meteorismus Swelling of the abdominal cavity with gas usually in the large intestine or stomach
Nystagmus Involuntary eye movements
Ophthalmoplegia A paralysis or weakness of one of the muscles that control eye movement

Opisthotonus A type of spasm in which the head and heels arch backward in extreme hyperextension and the body forms a reverse bow
Paresthesia A sensation of numbness or tingling on the skin
Rales (pronounced Ralz) Wet, crackly lung noises heard on inspiration which indicate fluid in the air sacs of the lungs
Wernicke Encephalopathy A syndrome characterized by ataxia, ophthalmoplegia, confusion and impairment of short-term memory

Introduction

Beriberi is caused by a deficiency of thiamin (also called thiamine, aneurin(e), and vitamin B_1). Classic overt thiamin deficiency causes a wide range of Thiamin Deficiency Disorders (TDD) including cardiovascular, cerebral, and peripheral neurological impairment, metabolic, respiratory, musculoskeletal and gastrointestinal systems. The rarity of thiamin deficiency and concurrent symptoms and overlapping signs with other disorders can obscure diagnosis and TDD are frequently misdiagnosed with potentially fatal consequences (McConachie and Haskew, 1988; Smith et al., 2021). The disease emerged in epidemic proportions at the end of the 19th century in Asian and Southeast Asian countries. Its appearance coincided with the introduction of the roller mills that enabled white rice to be produced at a price that poor people could afford. Unfortunately, milled rice is particularly poor in thiamin; thus, for people living in Low and Middle-Income Countries (LMIC) for whom food was almost entirely rice, there was a high risk of deficiency and mortality from beriberi (Platt, 1958; Burgess, 1958). Outbreaks of acute cardiac beriberi still occur, but usually among people who live under restricted conditions due either to poverty or imprisonment when freedom of food choice is limited. The major concern today is subclinical deficiencies in patients with trauma or among the elderly. There is also a particular form of clinical beriberi that occurs in patients who abuse alcohol, known as the Wernicke–Korsakoff syndrome (Victor et al., 1971). Subclinical deficiency may be revealed by reduced blood and urinary thiamin levels, elevated blood pyruvate/lactate concentrations and α-ketoglutarate activity, and decreased erythrocyte transketolase (ETKL) activity. Currently, the *in vitro* stimulation of ETKL activity by thiamin diphosphate (TDP) is the most useful functional test of thiamin status where an acute deficiency state may have occurred and the technology has now been transferred to 96 well plates improving the convenience of the assay (Jones et al., 2020). The stimulation is measured as the TDP effect. Most TDP is in the red cell. Liquid chromatography of thiamin metabolites in plasma extracts will provide an indication of dietary thiamin intakes.

Epidemiology

Beriberi presents in several different clinical forms (**Table 1**). Beriberi became endemic following the introduction of steam-powered rice mills, which enabled milled rice to be produced cheap enough so that almost everybody could afford it and consume it. It was particularly serious at the end of 19th and the beginning of the 20th centuries when seasonal epidemics of wet beriberi occurred with many deaths (Burgess, 1958; Platt, 1958). The disease affected mainly the Chinese and Japanese populations, although outbreaks were reported in India and among settlers in the New World during the long cold winters, and the disease

Table 1 Forms of beriberi in man.

Subclinical beriberi	Identified by transketolase activity or other biochemical tests of thiamin status. May be associated with early subjective symptoms such as anorexia, weakness, dysesthesia, and depression. Responds rapidly to treatment with thiamin.
Wet beriberi	Subacute or cardiac beriberi frequently having muscular pains, edema of feet and legs, enlarged heart, and tachycardia.
	Responds rapidly to treatment with thiamin. Major form and was typically seasonal in endemic areas.
	Acute fulminant type of beriberi in which the main feature is dominated by insufficiency of the heart and blood vessels. Responds rapidly to treatment with thiamin.
Dry beriberi	Chronic, atrophic type of polyneuropathy in which the main features are of a weak wasted person, with painful musculature making walking difficult, impaired sensory nerves and tendon reflexes, and flaccid paralysis of the motor nerves. Poor or no response to treatment with thiamin.
Infantile beriberi	Usually acute wet beriberi. Responds rapidly to treatment with thiamin.
Wernicke–Korsakoff syndrome	Predominantly neurological, affecting walking and vision in most and memory and cardiac function in over 50% of patients. Wernicke or ocular component responds rapidly to treatment but the Korsakoff psychosis responds slowly or not at all.

Modified from Thurnham (1978).

was not necessarily confined to rice-eating populations. Where acute cardiac beriberi occurred, dry beriberi was also present but usually in the older members of the community.

Milled rice has a thiamin concentration that is particularly poor (80 μg/100 g), but social conditions at the time of the large epidemics contributed to the problems. Bonded labor was common, with workers living on the work premises most of the time and paid mainly in the form of rice. In addition, reports at the time suggest that the rice was of uncertain freshness and quality, and that it could be so moldy, matted, and lumpy that it had to be remilled and washed, with a further loss of thiamin. The social conditions prevented natural eating practices because workers had little money to purchase additional food and they were dependent on what they were given. Likewise, badly stored cereals can lose up to 90% of the thiamin content, and toxins associated with mold growth have been implicated in causing sickness that may well precipitate clinical beriberi.

Reports suggest that the acuteness of the outbreak of beriberi and the interrelationship of thiamin deficiency with deficiencies of other nutrients probably had a major role in determining the nature of the pathological changes and lesions produced. For example, it is reported that protein energy malnutrition almost always accompanied subacute beriberi, reflecting the link between impoverishment and the disease. In contrast, it is also suggested that severe beriberi more often affected the more active, stronger, or supposedly better nourished members of the community. The younger, stronger rickshaw puller was most likely to suffer severe beriberi. This enigma may be due to thiamin intakes from a diet containing a high proportion of rice being insufficient to meet the thiamin requirement posed by the higher calorie intakes of the more active community members.

In older literature infantile beriberi appeared to affect the male infant who "tended to be overfed" (Fehily, 1940). Human milk is barely adequate in thiamine (0.23 mg/4.2 MJ) and this may be reduced still further in marginally deficient mothers and thiamine status was further compromised if given supplements of thiamin-poor rice. It is a common habit even today for rural mothers in LMICs to give very young infants, even beginning at 1 week of age, a bolus of masticated rice to supplement the milk intake. In 2009, a study in Vientiane reported that 27% of infants were given water or formula milk before receiving breast milk and premasticated glutinous rice was the first food supplement in 20–48% of infants in the first week of life. In the same study mothers underwent variable periods of dietary restriction in the first 3 months postpartum including exposure to hot beds of embers (Soukaloun et al., 2003; Barennes et al., 2009). The effects of these traditional practices is not known but pyrexia is known to increase energy and thiamine requirements. It was widely observed that nonspecific pyrexia was a precipitating factor for beriberi. A 1 °C rise in body temperature is associated with a 10% increase in basal metabolic rate. It has been suggested that more than half the mild cases of beriberi were associated with a nonspecific bout of fever, and such cases responded less readily to treatment with thiamin.

Parboiled rice is partially cooked before milling, and this prevents beriberi because some of the thiamin is dispersed through the grain (190 μg/100 g). The advantages of this were clearly seen in Malaya, where at the end of 19th century there were large-scale immigrations of young, able-bodied Chinese to work in the tin mines and Indians to work on the rubber estates. In both cases, immigrants often lived in remote regions where there was little opportunity to purchase local food and they were dependent on imported rice. It was the Chinese who, because of their dietary preference for milled rice, died in enormous numbers (Burgess, 1958). However, parboiling may not be completely protective and is not universal through India. In northern regions like Kashmir, the diet consists of polished, unfortified rice and cases of infantile beriberi have been reported. Cases all occurred in exclusively breastfed infants and most of the mothers followed customary dietary restriction (Whitfield et al., 2018). Clinical data of thiamin deficiency was also presented at a recent workshop in November 2019, from Assam in India, and Bhutan in pregnant and lactating women (Smith et al., 2021).

Although ways of avoiding the disease were known to the Japanese navy at the end of the 19th century, since the director general of the medical department had demonstrated that the disease was almost eradicated if the traditional rice diet was supplemented with fish, vegetables, meat, and barley, this information was not widely available, and supplementation was not feasible by the vast majority of people (Sinclair, 1982). It was widely believed that the cause of beriberi was an infection or toxin resulting from bad food. In particular, Pasteur's work on the microbiological cause of infections led many to search for an infectious agent, but none could be consistently identified. The scale of the problem for the colonial powers in Southeast Asia in the latter part of the 19th and early 20th centuries should not be underestimated. Labor was cheap but the death toll posed enormous problems. Extracts from reports at the time are illuminating: In 1887, there were 690 deaths out of 1931 native government officers in Sumatra, infant mortality was 445 per 1000 live births in the Philippines in 1910, and one report stated that there were so many deaths that "there was insufficient earth to bury the corpses" (Burgess, 1958).

The Dutch government sought to resolve the situation by appointing a medical bacteriologist, Christiaan Eijkman, to travel to Indonesia to investigate the problem (Luyken, 1990). Working in Java, he showed within 6 years that beriberi was a nutritional problem and that a paralytic condition closely resembling the polyneuritic symptoms of beriberi could be produced in chickens by feeding them both stale and freshly cooked polished rice. However, it was Funk (1911) who first reported the isolation of a "vital amine" from rice polishings that had anti-beriberi properties. Funk was the first person to coin the word "vitamine" as a substance essential for life. The structure and synthesis of thiamin were reported in 1936 (Williams and Cline, 1936).

Currently, clinical beriberi no longer occurs with the devastating effects of former years. Considerable improvements have occurred in nutrition worldwide, the diversity of foods available, the quality of food due to improved storage methods, and social and economic development. However, in many LMIC countries, especially in Southeast Asia and Africa, sporadic outbreaks do occur, which are usually of the acute, fulminating type of beriberi causing many deaths often in young men aged 20–40 years. An outbreak of acute beriberi occurred in a Gambian village in 1988 at the start of the wet season and killed 22 young adults before it was recognized and treated (Tang et al., 1989). In Thailand, infantile beriberi was reported in Karen refugees in 2003 and, in Laos, infantile beriberi is currently recognized as a public health problem (Soukaloun et al., 2003). Usually, a combination of factors is

responsible, but once the cause is identified, treatment is cheap and readily available and, if given rapidly, tragic circumstances can be averted.

Two iatrogenic causes of subclinical beriberi are known, namely that associated with diuretic treatment (Yui et al., 1980) and one resulting from alcohol abuse. Both are of concern because the use of diuretics is introduced to manage cardiovascular disease, a condition that will deteriorate if thiamin status is impaired, and alcohol abuse can lead to Wernicke–Korsakoff syndrome (Victor et al., 1971), which can have many of the features of both wet and dry beriberi.

Severe multisystem trauma, endotoxemia, or situations in which there is a raised metabolic demand for thiamin, such as pregnancy, thyrotoxicosis, and intercurrent illness or impaired absorption (e.g., alcohol abuse or gastrointestinal disease or resection), can produce subclinical evidence of TDD or more severe life-threatening aspects of beriberi, such as renal and/or cardiovascular failure (McConachie and Haskew, 1988; Whitfield et al., 2018). There are warnings in the literature concerning the necessity of vitamin, and particularly thiamin, supplements when energy metabolism may be increased e.g., following a dextrose-containing intravenous fluid. The elderly may be particularly at risk of subclinical thiamin deficiency. One Belgian study on patients with a mean age of 83 years reported that 40% had a raised TDP effect (>15%), in whom there was a high proportion of Alzheimer's disease, depression, cardiac failure, and falls. The diuretic furosemide was also more frequently taken by the thiamin-deficient patients.

Etiology

The factors associated with the various forms of beriberi are listed in **Table 2**. Beriberi is caused by a lack of thiamin in the diet, but the onset of the disease and the symptoms associated with the disease are influenced by one or more of the other etiological factors. Wet beriberi (i.e., cardiac beriberi) and Wernicke's encephalopathy are conventionally described as acute manifestations of the disease and respond most rapidly to treatment. In contrast, dry beriberi is described as due to a chronic deficiency of thiamin and does not respond well to treatment (Carpenter, 2002). However, experimental acute deficiency studies, which very rapidly produced subjective feelings of malaise and weakness at the slightest exertion, very rarely produced evidence of edema and peripheral pain. These observations suggest that all forms of beriberi are probably preceded by an indeterminate period of chronic thiamin deficiency during which pathophysiological adaptations to the marginal nutritional state occur. Thus, physiological adaptations to the vascular system may well have occurred particularly in those who did heavy physical work and needed to overcome the weakness and malaise imposed by a low thiamin diet. The factor(s) that precipitated the clinical disease may not be thiamin at all. Platt (1958), in his descriptions of beriberi in China, recounts how humid weather and infections such as malaria increased the number of cases of wet beriberi. The extra energy needed to cool the body in hot conditions or fuel the rise in temperature during infection may have imposed a critical burden on energy production that the system could not meet, and beriberi ensued.

However, the increased number of cases associated with heat, humidity, and malaria may also be due to a seasonal decline in the quality of food. A 6–12-fold decline in thiamin content is reported for millet when stored under traditional thatched storage houses in The Gambia (Thurnham et al., 2011), and reports suggest that much of the rice consumed late in the season was not in the best condition. Some of the products introduced by mold growth may possess anti-thiamin properties that impair thiamin bioavailability. Thus, the ratio of thiamin to calories is likely to fall during the agricultural year and to be at its worst when calorie

Table 2 Etiological factors contributing to thiamin deficiency.

Dietary thiamin deficiency	Commonly milled rice
High dietary carbohydrate to fat ratio	Metabolism of carbohydrate requires thiamin, whereas metabolism of fat spares thiamin requirements
Heavy physical activity	Predisposes to beriberi when accompanied by low intake of thiamin
Protein energy malnutrition	Older literature reports PEM sometimes accompanies subacute beriberi indicating importance of impoverished diet
Poor storage conditions for food	Fall 6- to 10-fold in thiamin content of cereals. Molds may accelerate decay as well as increase risk of toxins
Thiaminases	Two known, but only of importance when uncooked foods are consumed
Anti-thiamin factors	Factors in food that chelate with thiamin and potentially reduce bioavailability
Alcohol abuse	Alcohol impairs the active absorption mechanism for thiamin which can be especially important when accompanied by dietary restriction such as that imposed by bariatric surgery
Infection and trauma	Increase requirements for thiamin to support increased carbohydrate metabolism and energy production
Diuretics, long-term use	Accelerate thiamin excretion and appear to block thiamin control mechanism
Seasonal factors	Combination of heavy work load, impoverished diet, and last season's (badly stored) cereals
Post-partum cultural practices and infantile beriberi	Restricting food intake; exposure to hot coals in early weeks following birth
Male sex	Some evidence that men have higher thiamin requirements than women but more likely to be a combination of the first three factors listed

requirements are at their highest for land preparation and weeding. Land preparation also takes place just prior to or at the beginning of the rainy season, when the prevalence of malaria and diarrheal diseases increases.

The thiaminase enzymes destroy thiamin activity by breaking the thiamin molecule into two parts—the pyrimidine and thiazole moieties. Thiaminases are inactivated by cooking; thus, the enzymes are only a problem where certain foods are eaten raw. It has been suggested that in northern Thailand, where consumption of fermented raw fish products is widely practiced, thiamin status may be impaired by these food habits. Even as recently as 2001, marginal thiamin status was reported in more than 50% of women 3 months postpartum despite thiamin supplements of 100 mg/day during pregnancy. The deficiencies were found in Karen refugee women living on the Thai–Burmese border and whose diet contained fermented fish, tea leaves, and betel nuts—substances suspected of containing thiaminases. Polyphenol compounds in tea and many vegetables may also posses anti-thiamin properties and impair bioavailability, but their etiological importance in causing thiamin deficiency is difficult to assess. In Lao and Cambodia, some postpartum cultural traditions, nutrition practices and socio-economic factors may also impair mothers thiamin status and increase the risk of infantile beriberi (Soukaloun et al., 2003; Barennes et al., 2009).

Alcohol is an important factor in causing thiamin deficiency because it inhibits the active transport of thiamin across the gut and when abused it impairs the quality of the diet consumed. Diuretics accelerate the excretion of thiamin and appear to override the renal conservation mechanism. Their use is of potential concern in elderly people whose diet may be poor for other medical reasons and their physicians may be unaware of their need for supplemental nutrients (Ibner et al., 1982; Pepersack et al., 1999).

Both sexes are vulnerable to the effects of thiamin deficiency, but in many of the sporadic outbreaks that have been reported, there appears to have been a male excess. This may be due to higher thiamin requirements in men than women because of their higher lean body mass or to hormonally driven sex differences. However, it is also possible that the cause is due to a higher risk of a thiamin:calorie imbalance in men compared to women. In many rural communities, men traditionally eat first and may satisfy their calorie requirements, whereas their womenfolk make do with the leftovers. Because of their greater physical strength, men frequently do heavier work than women, requiring more energy (i.e., more food to meet their requirements). Thus, men may consume more of the thiamin-depleted cereals in the diet to satisfy calorie needs and in so doing achieve a poorer thiamin:calorie ratio than women.

Experimental thiamin deficiency in man and measurement of thiamin status

In young and healthy non-alcoholic subjects, subjective symptoms appear after 2 or 3 weeks of deficient diet but urinary thiamin will already be falling (**Table 3**) (Brin, 1964). Characteristic early symptoms include anorexia, weakness, dysesthesia, and depression. At this stage, urinary thiamin will be almost zero, ETKL activity depressed, and the TDP effect approximately 15–30%. After 6–8 weeks the only objective signs at rest may be a slight fall in blood pressure and moderate weight loss, although urinary thiamin will now be negligible and the TDP effect $\geq$35%. After 2 or 3 months, apathy and weakness become extreme, calf muscle tenderness develops, and there is loss of recent memory, confusion, ataxia, and sometimes persistent vomiting. Urinary thiamin will be negligible and the TDP effect may be normal (because apo-ETKL is unstable even in vivo), but ETKL activity should be considerably depressed.

Table 3 Effects of thiamin deficiency on urinary thiamin, the erythrocyte transketolase TDP effect, and early clinical symptoms of thiamin deficiency in human volunteers.

Days of deficiency	*Urinary thiamin (μg/day)*[a]	*TDP effect (%)*[a]	*Clinical signs of deficiency following diets containing 150–350 μg thiamin/day*[b]
5	50	0–10	Mostly studies report no signs but one study (360 μg/day) found within 1 week chest pains, extreme lassitude, anorexia, palpitation, and burning feet
10	25	15	
21–28	<25	30	Loss of body weight, anorexia, general malaise, insomnia, increased irritability, fatigue on slightest exertion
30–40	Negligible	≥40	Increased malaise, loss of body weight, intermittent claudication and polyneuritis, bradycardia, peripheral edema, a cardiac enlargement, ophthalmoplegia
>45	10–20	>40	Additional signs of nausea and dizziness appeared
75	10–20		Additional signs of vomiting, low blood pressure, and tenderness of calves

[a]Biochemical data and report of edema and cardiac enlargement from Brin (1964), in which healthy male medical students were fed 200 μg thiamin per day for 6 weeks. TDP effect is a measure of thiamin status obtained by measuring the activity of erythrocyte transketolase in the presence and absence of added thiamin diphosphate.

[b]Clinical signs adapted from several studies. Investigators were impressed by the rapid degree of debility induced by the specific withdrawal of thiamin from the diet. In one group (150 μg/day for 75 days, four female mental patients), the authors reported that the condition more closely resembled 'neurasthenia' than beriberi and noted that edema, cardiac dilation, and peripheral pain characteristic of classic beriberi were all absent (Carpenter, 2002).

The clinical symptoms resulting from experimental thiamin deficiency in man have usually responded rapidly to treatment with thiamin. In one feeding study, however, two mental patients were kept for 110 days on a diet providing 200 μg thiamin daily and 1 mg of thiamin by injection 1 day each week; thus, their overall weekly average was 350 μg/day. They developed a polyneuropathy characterized by defects in the sensory nervous pathways, loss of tendon reflexes, and paralysis of the legs, which took many weeks to respond to large doses of thiamin, and in one case response was still incomplete after 4 months of treatment. The slow cure suggested that degeneration of peripheral nerves had occurred, as is indicated in the dry form of beriberi, in which the neurological lesions are irreversible.

Clinical features of beriberi

Depletion and repletion studies suggest that intakes >300 μg/4.2 MJ are compatible with normal biochemistry and good health, and clinical signs of thiamin deficiency occur at intakes of thiamin below 200 μg/4.2 MJ (1000 kcal). The disease as studied from the 1880s onward in Asians subsisting on white rice began typically with weakness, "wandering pains" in the legs, and lack of feeling in the feet. Some patients then developed edema (the presence of excessive amounts of fluid in the intercellular tissue spaces of the body) of the legs, trunk, and face. In severe cases, sufferers found it increasingly difficult to catch their breath and would die of heart failure. The clinical features of subacute and acute wet beriberi are summarized in **Table 4**. The main form was subacute beriberi, which was typically seasonal in endemic areas. There are reports that the peripheral muscles most severely affected were those most frequently used; thus, in male laborers it was the legs. Aching pain, tightness, and cramps in the calf and associated muscles were usually a first cause of complaint, and pain on squeezing the calves was one of the most useful diagnostic tests for beriberi. In women who performed repetitive tasks involving hands and arms, a loss of sensation in the fingers was frequently a first cause of complaint.

Dry beriberi is essentially a chronic condition showing muscular atrophy and polyneuritis and frequently occurring in older adults. Walking is usually difficult because of the weak wasted and painful musculature, and in the later stages feeding and dressing may also become impossible. When bed-ridden and cachectic (extreme state of malnutrition and wasting), patients become very susceptible to infections. Sensory nervous function is impaired (hypesthesia) almost to the point of anesthesia. Hypesthesia is particularly evident in the extremities and progressively extends over the outer aspects of the legs, thighs, and forearms. Motor nerve disturbances also begin in the extremities and ascend progressively. Flaccid paralysis of the extensor muscles precedes that affecting the flexors and results in "wrist drop" and "foot drop" (**Fig. 1**). Loss of the Achilles tendon reflex usually precedes an impaired patellar reflex.

Infantile beriberi

In the early reports, mainly from S.E. Asian countries, mortality from infantile beriberi mainly affected breast-fed infants between the second and fifth months of life, when solid foods were often first introduced (Fehily, 1940). The introduction of white rice porridges, poor in thiamin, to a rapidly growing child and/or the increased exposure to infections when solids are introduced, may both have exacerbated thiamin deficiency and contributed to infantile beriberi. The onset of the disease was rare in the first month and early signs could be mild and somewhat subjective (e.g., vomiting, restlessness, anorexia, and insomnia). Early signs could progress to subacute infantile beriberi, the acute and usually fatal condition, or a chronic form. Features of acute infantile beriberi are presented in **Table 5**. The subacute form was characterized by slight edema in the form of puffiness, vomiting, abdominal pain, oliguria, dysphagia, and convulsions. In addition, aphonia (soundless cry) was often a feature of subacute infantile beriberi and may have been due to nerve paralysis or edema of vocal cords. Vomiting was also a feature of chronic infantile beriberi and could be accompanied by inanition, anemia, aphonia, neck retraction, opisthotonus, edema, oliguria constipation, and meteorismus (swelling of the abdominal cavity from gas in the intestine). Opisthotonus is a characteristic of acute thiamin deficiency in birds and is described as due to a tetanic spasm in which the spine and extremities are bent backwards (Thurnham, 1978).

In contrast to these earlier descriptions, the features of beriberi illustrated by the Israeli infants who consumed thiamine deficient formula appeared somewhat later than those in the S.E. Asian infants. Age of presentation ranged mainly from 5 to 12 months although there was one infant of 2.5 months. However, clinical symptoms were similar with vomiting, refusal to feed, gastrointestinal, neurological and developmental delay being common features and all 9 infants with a raised TDP effect, presented with infection and or fever (see below).

Accidental thiamin depletion in infants

Western societies are heavily dependent on milk formulas for the nourishment of human infants. Therefore, when a modified form of a well-known formula that only contained one-10th of the reported thiamin content was released onto the market in Israel in the latter half of 2003, it was followed by a wave of hospitalizations of infants in intensive care units, a national state of pandemonium and two infants died of cardiomyopathy (Fattal-Valevski et al., 2005). Analysis of the data revealed only a single link among them, all infants were fed with a reputable, soy-based infant formula (Remedia Super Soya 1) but the contents printed on the label met the stringent standards of the Israeli Department of Health.

Table 4 Common features of wet beriberi.

	Subacute beriberi	*Acute fulminating beriberi*[a]
Digestive system	Anorexia is common; constipation more frequent than diarrhea	Vomiting is common, often with intense thirst
		Liver enlarged and tender and the epigastric region spontaneously painful
Neurological	Aching pain, stiffness, tightness, or cramps in calf or associated muscles	Pupils dilated with anxious expression on face
	Increasing muscular tenderness and weakness with fatigue pains resembling muscular ischemia, especially at night	Aphonia frequently present and patient moans with cries of a special kind as a result of hoarseness produced by paralysis of laryngeal muscles
	Pain on squeezing calves	Reflexes of ankle or knee lost or diminished
	Inability to rise from squatting position without use of hands	
	Diminished reflexes of ankle and knees usually bilaterally	
	Hypesthesia or paresthesia presenting as "pins and needles," numbness particularly over the tibia, formication (like ants running on the skin) or itching	
Cardiac	Edema of feet and legs often appearing first on dorsa of feet and extending up legs but may also appear on back of hands and as puffiness in face	Patients severely dyspneic, have violent palpitations of the heart, are extremely restless, experience intense precordial agony but accessory muscles of respiration on slightly brought into action
	Heart enlarged with tachycardia and bounding pulse	
	Raised venous pressure (see Fig. 2) with percussion sometimes revealing dilation of right auricle and ventricle	Widespread and powerful undulating pulsations visible in the region of the heart, epigastrium, and neck due to a tumultuous heart action
	Heart murmurs if present are usually systolic	Facial cyanosis more marked during inspiration
	Apex beat is downward and outwardly displaced	Pulse is moderately full, regular, even with frequency of 120–150/min
	Neck vein possibly distended showing visible pulsations	A wavelike motion may be felt over the heart
	Dyspnea upon exertion	On percussion, the heart is enlarged both to the left and right but mainly the latter, and the apex beat may reach the axilla
	Palpations, dizziness, and giddiness	
	Extremities possibly cold and pale with peripheral cyanosis but where circulation is maintained, skin warm due to vasodilatation	Raised systolic pressure and low diastolic pressure give the "pistol shot" sound on auscultation over the large arteries
	Electrocardiograms often undisturbed but QRS complex may show low voltage and inversion of T waves indicating disturbed conduction	Rapidly increasing edema may extend from legs to trunk and face with associated pericardial, pleural, and other serous effusions
Urine	Nocturia; no albuminuria	Oliguria or anuria; no albuminuria or glycosuria

[a]The whole picture of acute fulminating beriberi is dominated by insufficiency of heart and blood vessels and this tends to mask all other features of the subacute form, although these are often present and accentuated. Death is accompanied by a systolic pressure falling to 70–80 mm, the pulse becomes thinner, and the veins dilate. The rough whistling respiration deteriorates and rales appear. The patient dies intensely dyspneic but usually fully conscious.
Modified from Thurnham (1978).

The cause of the trouble was revealed when a male infant of 5.5 months was admitted to Sourasky Medical Center with upbeat nystagmus, ophthalmoplegia and vomiting. Wernicke's encephalopathy was suspected and within hours of treatment with thiamin, his condition improved. A brief investigation revealed the infant was fed the same formula. Public health authorities were informed, analyses were done both in Israel and by the manufacturer and the mistake was revealed.

Fortunately, thiamin deficiency was detected in only 20 infants probably because most infants were not fed solely with milk. It is interesting to note the presenting signs and symptoms in these infants, where thiamin deficiency was known to be causal, as it assists detection of sub-clinical beriberi in less-developed countries than Israel, The clinical signs initially reported were constipation, agitation, apathy, vomiting, lack of appetite and later diarrhea, grunting, nystagmus, convulsions and unconsciousness. In 9 patients who met the laboratory criteria for thiamin deficiency (TDP effect 13.8–37.8%) infection was the precipitating factor and 6 of the patients were admitted with fever. Almost all presented with vomiting and gastrointestinal symptoms. Most of the infants were aged between 5 and 12 months and received between 800 and 1100 mL of the formula per day but for a variable amount of time (20–70% of their lives).

Possible long-term neurological consequences in the form persistent cognitive and motor deficits are now being recognized among the Israeli children who survived the deficient formula (Measelle et al., 2021). The most severely affected survivors demonstrated marked intellectual disabilities, seizures, motor disabilities, microcephaly, auditory dysfunction and complete heart block.



Fig. 1 Patient with dry beriberi showing evidence of motor nerve disturbances resulting in a flaccid paralysis of the extensor muscles and "wrist drop" and "foot drop". Courtesy of the late Professor B.S. Platt, formally Head of the Department of Human Nutrition, London School of Hygiene and Tropical Medicine, UK.

Table 5 Features of acute infantile beriberi and frequency of occurrence.

	Features	*Frequency (%)*
Appearance	Pale and cyanotic appearance, edematous, ill-tempered with abdominal distension	40
Voice	Hoarseness	80
	Sometimes groaning	50
Digestive system	Vomiting	80
	Dyspepsia	46
Cardiac	Tachycardia, <200 beats/min	83
	Heart dilated	31
	Femoral sound on auscultation	5
Lungs	Rapid breathing	83
	Accentuation of the 2nd pulmonary sound	
Neurological	Tendon reflex usually increased	74
	Less frequently decreased	26
	Convulsions	17
Urinary	Oliguria	65
Other	Slight fever	50
	Uneasiness	50

Modified from Thurnham (1978).

Not all infants who received the formula displayed clinical affects initially but a variety of language impairments and learning disabilities have subsequently appeared which suggest that the long term deficiency of thiamin may have caused negative cognitive and developmental outcomes. Thus early life exposure to thiamin deficiency can have both acute and potentially serious long-term consequences but also more insidious neurological effects potentially affecting their health and development through life.

Alcohol abuse, the Wernicke-Korsakoff syndrome and morbid obesity

In alcoholic and other malnourished subjects, one of the early signs of thiamin deficiency is anorexia. In alcohol abuse, the overwhelming desire for alcohol may outweigh all other interest in food, leading to generalized malnutrition. Alcohol specifically blocks

Table 6 Clinical features of Wernicke–Korsakoff syndrome and frequency of occurrence.

	Features	*Frequency (%)*[a]
Ocular disorders	Nystagmus (ocular ataxia—rhythmical oscillation of the eyeballs), almost always horizontal and in 50% of cases associated with vertical nystagmus on upward gaze	85
	Paralysis of one or more of the ocular muscles	50
	Sluggish reaction by pupils to light	19
Ataxia (inability to coordinate muscles)	Gait	87
	Legs	20
	Arms	12
	Speech	87
Polyneuropathy	Limbs only affected, mainly the legs only	82
	Of arms and legs	18
	Common symptoms include weakness, paresthesia, pain, loss of tendon reflexes and of sensation and motor power	
	Some cases of foot drop or wrist drop or both	
Cerebral function	Global confusional state, profound disorientation, apathy, deranged perception and of memory, drowsiness, inattentiveness, indifference	56
	Disorder of memory: both retrograde and ante-retrograde amnesia, confabulation	57
	Alcohol abstinence syndrome	16
Cardiac	Tachycardia	51
General medical abnormalities	Disorders of skin and mucous membranes	36
	Redness and/or papillary atrophy of the tongue	29
	Liver disease	60

[a]Percentages based on 188–245 cases.
Modified from Thurnham (1978).

the active absorption of thiamin and alcohol abuse can progress to the potentially fatal condition known as Wernicke–Korsakoff syndrome (Victor et al., 1971). The typical clinical features of Wernicke's encephalopathy comprise ophthalmoplegia, nystagmus (usually horizontal), ataxic gait, and an abnormal mental state that can range from mild delirium to global confusion. Liver disease and tachycardia occur in more than 50% of cases. Korsakoff's psychosis is characterized by a profound amnesia, disorientation, and often confabulation. The clinical features of Wernicke–Korsakoff syndrome are listed in **Table 6**.

A specific group of patients in whom dietary deficiencies can arise are those who undergo bariatric surgery (gastric bypass or also sleeve gastrectomy). The prevalence of morbid obesity has risen to global epidemic proportions and bariatric surgery has been shown to be the most effective treatment to achieve substantial and long-lasting weight loss A complication of these procedures is the appearance of Wernick's encephalopathy (WE) especially in patients who receive no micronutrient supplements and abuse alcohol. Current guidelines for bariatric surgery suggest a preventative multivitamin supplement containing 12 mg thiamin, and for patients showings signs of WE post-surgery, 500 mg parenteral thiamin, 3 times daily until symptoms subside. Rapid correction with thiamin is important to prevent the further progression of WE to the untreatable Korsakoff syndrome (Oudman et al., 2018).

Management/treatment

A wide range of clinical disorders exist where thiamin deficiency is involved and there are potentially serious consequences if it is overlooked during treatment. Hence it is not wrong to add thiamin to a therapeutic regimen even though the deficiency may not have been actively assessed especially in a clinical emergency or an obviously at-risk population.

Patients in whom cardiac and renal signs of thiamin deficiency are identified usually respond well to treatment. The dose given and route used will vary with the seriousness of the deficiency. Intravenous doses as high as 250 mg/day for 14 days and intramuscular doses of 25 mg followed by thrice daily oral doses of 10 mg have been reported for wet beriberi and are followed by a marked increase in urinary output and improvement in cardiac function. Peripheral neuropathy (dry beriberi) is more resistant to treatment. Patients with the ocular signs of Wernicke's disease usually respond to two or three daily injections of 50 mg thiamin. Long-term oral treatment of other manifestations of Wernicke–Korsakoff syndrome with doses up to 50 mg/day is reported, although benefit is variable and considerably influenced by patients' ability to avoid further alcohol consumption. It is unlikely that patients receiving oral thiamin will absorb more than 5–7 mg/day, but in patients likely to abuse alcohol, absorption by passive diffusion of high thiamin doses is the only way to ensured that the patient will receive any thiamin. In addition, as in all patients who show evidence of nutritional deficiency, the likelihood of other coexisting deficiencies should not be overlooked and multinutrient treatment is probably desirable. Finally, it is important to realize that untreated thiamin deficiency can result in sudden death.

Lipid-soluble thiamin derivatives

In recent years, several lipid-soluble derivatives of thiamin have been introduced, of which the best known is benfotiamine (Hammes et al., 2003). Advantages of these compounds appear to be increased absorption of thiamin, but by the diffusion mechanism only, and greatly increased transketolase activity. Transketolase is the rate-limiting enzyme of the nonoxidative branch of the pentose phosphate pathway. Benfotiamine has been shown to be useful for the management of rare genetic disorders in thiamin transport and may also prove useful to prevent damage from diabetic hyperglycemia. One study demonstrated that benfotiamine prevented experimental retinopathy. Diabetic hyperglycemia is accompanied by an increase in the potentially pathogenic glycolytic metabolites glyceraldehyde-3-phosphate and fructose-6-phosphate. Benfotiamine, by increasing transketolase activity, stimulates the pentose phosphate pathway to metabolize these glycolytic intermediates into pentose-5-phosphates and prevent the intracellular increase of potentially toxic products.

Case study

A good example of the specific effect of thiamin in the treatment of beriberi is illustrated by the response of a 29-year-old male who was admitted with an unexplained acute renal failure and had been anuric for 24 h. The physicians' report on his symptoms should be compared with the common clinical features of wet beriberi shown in **Table 4**. The patient's physical state and voice were extremely weak but speech was copious and confused. He complained intermittently of severe central chest and epigastric pain. A central cyanosis was present and he had a respiratory rate of 36 beats per minute. His temperature was normal and peripheries were lukewarm. He had gross generalized edema. The jugular venous pressure became grossly elevated (**Fig. 2**). Pulse rate was 100 beats per minute, regular, and weak at the wrist, although the carotid pulses were visibly bounding. Blood pressure was 80/60. There was a marked parasternal heave present, with a loud pulmonary second heart sound. The chest was clear; the abdomen was obese.

The father reported that the patient's usual beer intake was 6–12 pints daily and his one regular meal was usually no more than a sausage roll or a pie. Prior to admission, for 6 weeks he had felt too tired to go out in the evening, and for 2 weeks he had suffered epigastric discomfort and had eaten nothing. Eight days before admission, he developed painful calf stiffness and he became too weak to go to work. He had a painful dry cough and dyspnoe on the slightest exertion. Finally, confusion, cyanosis, and intermittent vomiting led to admission.

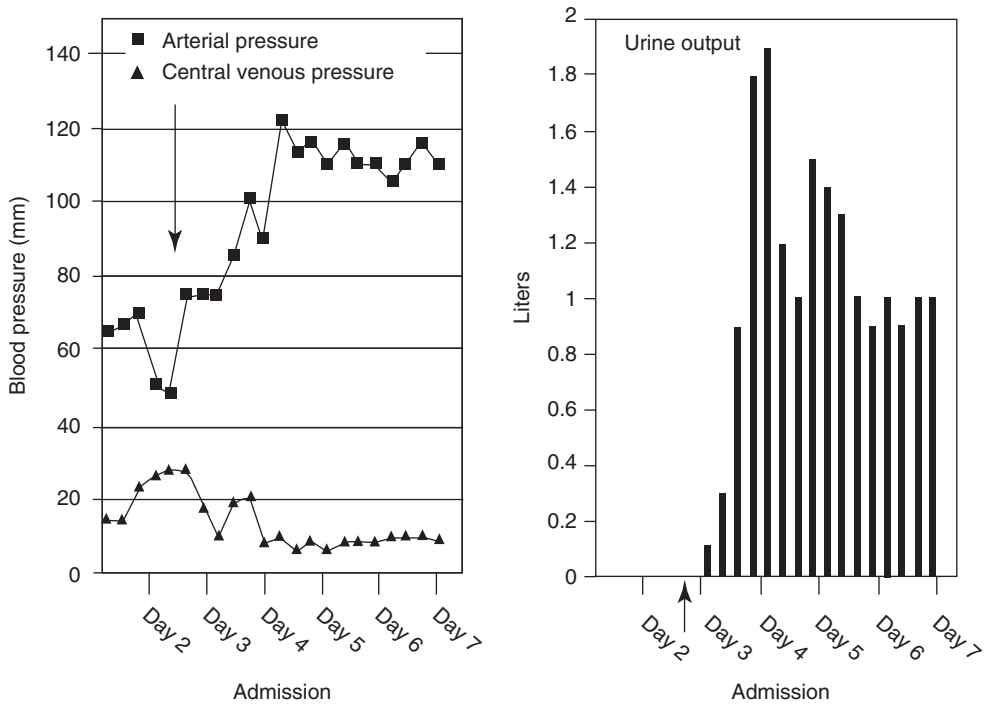


Fig. 2 (Left) Arterial and central venous blood pressure and (right) urine output of a patient who was admitted with unexplained acute renal failure in a very weak physical state and whose speech (although very weak) was copious and confused. The patient was discovered to be a regular beer drinker, consuming 6–12 pints daily, and his usual food intake amounted to no more than a sausage roll or pie. He had become progressively weaker over the past 8 weeks and had eaten nothing at all in the past 2 weeks. After excluding other diagnoses, it was suspected that the patient had fulminant beriberi and he was treated with thiamin after 36 h. The figures display the rapidly increasing arterial pressure, fall in venous pressure, and a rapid resumption in renal function following thiamin treatment. The patient lost ~20 L of urine during the first 7 days in the hospital. Modified from Anderson et al. (1985).

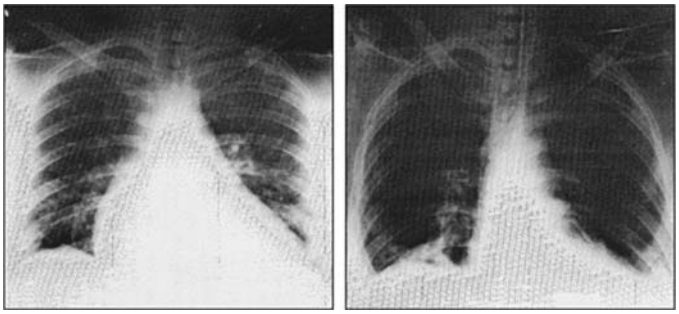


Fig. 3 Chest radiographs of the patient described in **Fig. 2** obtained on admission (left) and 14 days after high-dose, parenteral thiamin treatment (right). On admission, the heart was grossly enlarged, extending downward and to the right, with a cardiothoracic ratio of 0.63. After treatment for 14 days, the cardiothoracic ratio was 0.44. Modified from Anderson et al. (1985).

The first diagnoses considered were myocardial infarction, pulmonary embolism, and overwhelming septicemia, and he was placed on dialysis and received appropriate treatments. His lack of response at 36 h, continuing low systolic pressure (70 beats per minute), increasingly gross hyperdynamic precordial signs, and moribund appearance led to a diagnosis of beriberi. Treatment with intravenous thiamin (250 mg for 14 days) brought about a dramatic response (**Fig. 2**). Within 6 h peripheral pulses were strong, blood pressure had risen to 105 systolic, and central venous pressure had fallen by half. By 12 h the parasternal heave was less marked and diuresis of up to 6 L per day ensued. After 24 h, plasma urea concentration peaked at 50.4 mmol/L and creatinine at 832 μmol/L, and thereafter there was a steady fall over the next 2 weeks during which thiamin treatment continued and dialysis stopped. He lost a net 20 L of fluid over the first 7 days in the hospital and creatinine clearance 3 weeks after admission was 178 mL/min, indicating a return to normal kidney function. Other biochemical abnormalities resolved over the 2 weeks on high-dose thiamin, including the chest radiograph (**Fig. 3**). It is interesting to note, however, that when he was discharged 3 months after admission, he was walking with a caliper because of a right-sided foot drop (**Fig. 1**). The persistence of the foot drop is a further indication of the greater difficulty in reversing neurological consequences, in contrast to the cardiac effects, of thiamin deficiency.

Toxicity

Chronic intakes in excess of 50 mg/kg, or more than 3 g/day, are toxic to adults with a wide variety of clinical signs, including headache, irritability, insomnia, rapid pulse, weakness, contact dermatitis, pruritis, and, in one case, death. Early researchers also indicated that regular administration or contact with thiamin occasionally led to allergic response, contact dermatitis, or hypersensitivity (Sinclair, 1982).

Beriberi conclusions

Beriberi is a deficiency disease caused by a lack of dietary thiamin. Thiamin has a fundamental role in the metabolic generation of energy from carbohydrate. Situations which increase energy needs increase the need for thiamin. An acute insufficiency of thiamin causes cardiac "wet" beriberi which can be rapidly fatal in both adults and infants. A consistently low or chronic intake of thiamin increases the risk of "dry" beriberi; a peripheral neuropathy with motor nerve damage and ophthalmoplegia. Alcohol abuse can cause Wernicke's Encephalopathy which can include all symptoms of the disease including cognitive problems. The rapid effects of a thiamin deficiency and the fundamental role of energy provision, imply that the deficiency may influence the clinical presentation of various morbidities which may not be promptly recognized and treated. Thus thiamin deficiency may present as a number of thiamin deficiency disorders.

See Also: Alcohol: Effects of consumption on diet and nutritional status

References

Anderson, S.H., Charles, T.J., Nicol, A.D., 1985. Thiamine deficiency at a district general hospital: report of five cases. Q. J. Med. 55, 15–32.

Barennes, H., Simmala, C., Odermatt, P., et al., 2009. Postpartum traditions and nutrition practices among urban Lao women and their infants in Vientiane, Lao PDR. Eur. J. Clin. Nutr. 63, 323–331.

Brin, M., 1964. Erythrocyte as a biopsy tissue for functional evaluation of thiamine adequacy. J. Am. Med. Assoc. 187, 762–766.

Burgess, R.C., 1958. VI Special problems concerning beriberi. B. Infantile beriberi. Proceedings of a conference on beriberi, endemic goitre and hypervitaminosis A entitled "Nutritional Disease". Proc. Fed. Assoc. Soc. Exp. Biol. 17 (Suppl. 2), 39–46.

Carpenter, K.J., 2002. Acute versus marginal deficiencies of nutrients. Nutr. Rev. 60, 277–280.

Fattal-Valevski, A., et al., 2005. Outbreak of life-threatening thiamine deficiency in infants in Israel caused by a defective soy-based formula. Pediatrics 115, 233–238.
Fehily, L., 1940. Infantile beriberi in Hong Kong. Caduceus 19, 78–93.
Funk, C., 1911. The chemical nature of the substances that causes polyneuritis in birds induced by a diet of polished rice. J. Physiol. 43, 395–400.
Hammes, H.-P., et al., 2003. Benfotiamine blocks three major pathways of hyperglycemic damage and prevents experimental diabetic retinopathy. Nat. Med. 9, 294–299.
Ibner, F.L., Blass, J.P., Brin, M., 1982. Thiamin in the elderly, relation to alcoholism and to neurological degenerative disease. Am. J. Clin. Nutr. 36, 1067–1082.
Jones, K.S., Parkington, D.A., Cox, L.J., Koulman, A., 2020. Erythrocyte Transketolase Activity Coefficient (ETKAC) Assay Protocol for the Assessment of Thiamine Status. Annals of the New York Academy of Sciences. December 22nd, on line ahead of print.
Luyken, R. (Ed.), 1990. Polyneuritis in Chickens, or the Origin of Vitamin Research. Roche. Basle [First English edition of papers by Christian Eijkman published 1890–1896].
McConachie, I., Haskew, A., 1988. Thiamine status after major trauma. Intensive Care Med. 14, 628–631.
Measelle, J.R., Baldwin, D.A., Gallant, J., Chan, K., et al., 2021. Thiamine supplementation holds neurocognitive benefits for breastfed infants during the first year of life. In: Annals of the New York Academy of Sciences. June 7 on line ahead of print.
Oudman, E., Wijnia, J.W., van Dam, M., Biter, L.U., Postma, A., 2018. Preventing Wernicke encephalopathy after bariatric surgery. Obes. Surg. 28, 2060–2068.
Pepersack, T., Garbusinski, J., Robberecht, J., et al., 1999. Clinical relevance of thiamine status amongst hospitalised elderly patients. Gerontology 45, 96–101.
Platt, B.S., 1958. Epidemiology and clinical features of endemic beriberi. Proceedings of a conference on beriberi, endemic goitre and hypervitaminosis A. Proc. FASEB 17 (Suppl. 2), 3–20.
Sinclair, H.M., 1982. Thiamin. In: Barker, B.M., Bender, D.A. (Eds.), Vitamins in Medicine, vol. 2. W, Heinmann Medical Books Ltd., London, pp. 114–167.
Smith, T.J., Johnson, C.R., Koshy, R., Hess, S.Y., et al., 2021. Thiamine deficiency diseases: a clinical perspective. Ann. N. Y. Acad. Sci. 1498 (1), 9–28.
Soukaloun, D., Kounnavong, S., Pengdy, B., et al., 2003. Dietary and socio-economic factors associated with beriberi in breast fed Lao infants. Ann. Tropical Pediatr. 23, 181–186.
Tang, C.M., Rolfe, M., Wells, J.C., Cham, K., 1989. Outbreak of beri-beri in the Gambia. Lancet 2, 206–207.
Thurnham, D.I., Cathcart, A.E., Livingstone, M.B.E., 2011. A retrospective investigation of thiamin and energy intakes following and outbreak of beriberi in the Gambia. Nutrients 3, 135–151.
Thurnham, D.I., 1978. Thiamin. In: Rechcigl Jr., M. (Ed.), Nutrition Disorders. CRC Press, West Palm Beach, FL, pp. 3–14.
Victor, M., Adams, R.D., Collins, G.H., 1971. The clinical findings. In: Plum, F., McDowell, F.H. (Eds.), The Wernicke–Korsakoff Syndrome. Blackwell, Oxford, pp. 16–34.
Whitfield, K.C., Bourassa, M.W., Adamolekun, B., Bergeron, G., et al., 2018. Thiamine deficiency disorders: diagnosis, prevalence and a roadmap for global control programs. Ann. N. Y. Acad. Sci. 1–41.
Williams, R.R., Cline, J.K., 1936. Synthesis of vitamin B1. J. Am. Chem. Soc. 58.
Yui, Y., Itokawa, Y., Kawai, C., 1980. Furosemide induced thiamine deficiency. Cardiovasc. Res. 14, 537–540.

Further reading

Thurnham, D.I., 2004. An overview of interactions between micronutrients and of micronutrients with drugs, genes and immune mechanisms. Nutr. Res. Rev. 17, 211–240.
Williams, R.R., 1961. Towards the Conquest of Beriberi. Harvard University Press, Cambridge, MA.

Thiamin: Physiology

D.I. Thurnham, Northern Ireland Center for Food and Health, University of Ulster, Coleraine, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of D.I. Thurnham, Thiamin: Physiology, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 274–279, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00266-X.

Introduction 441
Dietary sources of thiamin 441
Low-dose thiamin supplements in lactating Cambodian mothers 443
Absorption and ethyl alcohol 443
Transport, storage and excretion 443
Biological functions 444
Assessment of thiamin status 445
Recommended dietary allowances 446
Drug–nutrient interactions 446
Toxicity 447
Conclusions 447
References 447

Key points

- Thiamin is a water-soluble vitamin with an active turnover in vivo; there is no storage; a regular supply is needed from a variety of food sources.
- Refined foods are generally thiamin-poor, especially rice. There is evidence of the thiamin deficiency disease, beriberi, in some Low and Middle-Income South and SE Asian countries.
- Thiamin supplements can improve breast-milk thiamin in counties where rice is a staple food and potentially lower the risk of infantile beriberi.
- Absorption of thiamin is an active saturable process which is inhibited by ethyl alcohol. There is some diffusion across the gut from high doses.
- Most thiamin in vivo is present as thiamin diphosphate (TDP). TDP acts as a coenzyme in two important dehydrogenase complexes that generate ATP and they are a fundamental source of energy for the whole body.
- The principal symptoms of thiamin deficiency are cardiac and/or neurological but because of the wide variety of presentations, they are now termed Thiamin Deficiency Disorders.
- The measurement of TDP either as the coenzyme for the enzyme transketolase or as the concentration in blood is used to assess thiamin status.

Glossary

Achlorhydria Reduction in gastric acid content

Acidosis Raised concentration of pyruvic and lactic acids in the blood when pyruvate cannot be converted to acetyl CoA for onward metabolism by the tricarboxylic acid cycle and a fall in ATP production stimulates more glycolysis and more pyruvate production

Aleuron layer Layer in the cereal grain occurring below the husk

Beriberi The clinical condition resulting from a lack of dietary thiamin

Cocarboxylase Alternative name for thiamin diphosphate or thiamin pyrophosphate

Diuretic drug Increases production of urine to reduce edema in heart failure

Gut neoplasia Cancer in the gut

Parboiled rice Rice that has been boiled in the husk

Polished rice Rice from which the outer husk and aleurone layer have been removed

Thiamin Essential vitamin; also called thiamine, aneurin(e) and vitamin B1

Thiaminase Enzymes that will inactivate thiamin. Found in a number foods but are inactivated by cooking

Ulcerative colitis Inflammation of the gut accelerating food transit and reducing absorption of thiamin

Wernicke-Korsacoff syndrome A form of thiamin deficiency that occurs in patients who abuse alcohol

Introduction

Thiamin is a water-soluble vitamin and the structure comprises a pyrimidine and a thiazole ring linked by a methylene bridge (Fig. 1). In its metabolically active forms, the hydroxyl group on the thiazole moiety is replaced by one, two, or three phosphate groups to form three phosphorylated coenzymes. A well-nourished human adult body contains approximately 30 mg of thiamin—approximately 80–90% as thiamin diphosphate (TDP), 10% as thiamin triphosphate (TTP), and a small amount of thiamin monophosphate (TMP) and thiamin. Like most water-soluble vitamins, there is no definable store in the body; the only reserves are thiamin coenzymes that are present in most cells in combination with appropriate thiamin-requiring enzymes. The predominant need for thiamin is linked to energy production but there is increasing evidence that thiamin is also needed for additional neurological functions. Thiamin is found in the aleuron layer of cereal grains as well as in animal food products such as liver. Man's desire for high-extraction cereal products in situations in which the diets contained little more than the cereal was a main contributory factor to the scourge of beriberi throughout much of Southeast Asia at the end of 19th and beginning of the twentieth century. Thiamin is relatively unstable and destroyed by poor cooking habits, and it is susceptible to degradation in foods that are not stored properly. Thiamin turnover is also quite rapid, and the absence of stores means that a continuous supply of thiamin is required. So thiamin status can be fairly rapidly impaired by factors affecting intake (e.g., vomiting and alcohol abuse) or excessive excretion (e.g., induced by diuretics). Thus, thiamin deficiency is sometimes a problem in pregnancy, in alcohol abuse, and in the elderly. Seasonal outbreaks can also occur in Low and Middle-Income Countries (LMIC) when energy demand is high, cereals may have been badly stored for many months leading to losses of thiamin and food supplies are restricted (Whitfield et al., 2018).

Dietary sources of thiamin (Paul and Southgate, 1978)

Thiamin is present in most foods but cereal products provide most thiamin for most people in the world, although the source is fundamentally different in developing and more industrialized countries. In the developing world, unrefined cereal grains and/or starchy roots and tubers provide 60–85% of dietary thiamin, whereas most dietary thiamin in industrialized countries is obtained from fortified cereal products. In the United Kingdom, for example, wheat flour is fortified with 2.4 mg thiamin per kilogram and many breakfast cereals contain 30% or more of the daily thiamin requirement per portion. Thiamin is present in greatest amounts in brewers yeast, the germ and aleuron layers of fresh wheat, egg yolk, and mammalian liver. It is also present in meat flesh, particularly pork, and vegetables, nuts, and legumes (Table 1). Milk from both humans (0.49–0.79 μmol/L; 0.23 mg/4.2 MJ (1000 kcal)) and cows (1.18–1.48 μmol/L) is a poor source of thiamin. Thiamin is actively secreted into milk by the lactating mother, and it is of interest that in the thiamin-replete woman the amount of thiamin in human milk is not increased by supplements, but the concentration and of course the volume consumed increase during the first 6 weeks of lactation. See also effects of thiamin supplements on milk thiamin concentrations in women living in areas at risk of thiamin deficiency (below).

Refined foods in general, such as fat, sugar, and alcohol, are poor sources of thiamin. Polished rice is particularly low in thiamin (80 μg/100 g) and is especially important because of its widespread consumption and importance as a source of calories. Cereal grains lose thiamin during refining, but the process of parboiling rice before milling enables most of the thiamin to be retained (190 μg/100 g) since it migrates into the starchy endosperm during the procedure. Proper storage of cereal grains is also important to maintain thiamin activity. Studies in The Gambia, West Africa, found that old season millet, which had been stored under thatch and in high humidity, when consumed 6–8 months later in the middle of the rainy season had thiamin concentrations (11 μg/100 g) that were 6–12 times lower than cooked samples obtained immediately postharvest. Imported rice used in the village likewise only contained 10 μg/100 g at the time of consumption (Thurnham et al., 2011).

Because of the water-soluble properties of thiamin, it can be leached from food during cooking. Thiamin is stable in slightly acid water up to boiling point but is unstable in alkaline solution that oxidizes it quantitatively to thiochrome (Fig. 2). In addition, antithiamin factors in food can accelerate thiamin losses. Paralysis in foxes fed raw carp led to the discovery of the thiaminase enzymes. Two thiaminases are found in food. Thiaminase I is found in fish, shellfish, ferns, and some bacteria and catalyzes a base exchange reaction between thiazole and another base. Thiaminase II is a hydrolytic enzyme that cleaves the vitamin at the methylene bridge and is found mainly in bacteria. The thiaminases are heat labile, so only food that is eaten raw or fermented may loose thiamin during its preparation or in the gastrointestinal tract. There are also heat-stable anti-thiamin factors that are found in ferns, tea, betel

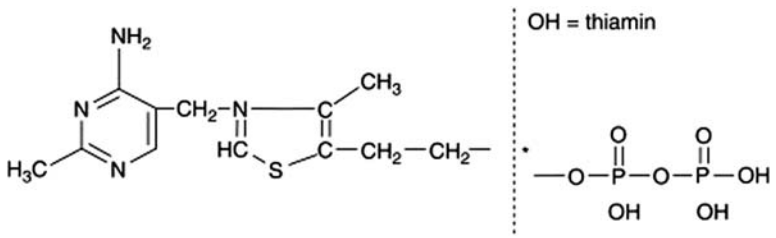


Fig. 1 Thiamin and thiamin diphosphate (asterisk). Thiamin monophosphate and triphosphate are formed by the similar addition of one or three phosphate groups at the asterisk.

Table 1 Thiamin content of common foods.

Food group	*Food item*	*Thiamin content (mg/100 g)*
Bread	Wholemeal	0.26
	White	0.18
	Hovis	0.52
Breakfast cereals	Cornflakes (fortified)	1.8
	Rice krispies	2.3
	Weetabix	1.0
Flour	Wholemeal (100%[a])	0.46
	Brown (85%)	0.42
	White (fortified) (70%)	0.28–0.33
Milk, cheeses		0.03–0.06
Eggs	Cooked (various)	0.07–0.09
	Yolk raw	0.30
Vegetables (cooked)	Various leaf and root types	0.02–0.07
	Dahl, chick peas, green, beans, etc.	0.05–0.14
Pork products	Gammon rashers (lean)	1.0
	Bacon (various)	0.36–0.55
	Pork meat	0.5–0.88
	Liver (stewed)	0.21
Other meats	Beef (various)	0.03–0.09
	Lamb (various)	0.04–0.14
	Lamb liver	0.56
	Chicken (various)	0.04–0.10
	Game	~0.30
Yeast (dried)	Peanuts (fresh)	2.33
Nuts	Peanuts (roasted & salted)	0.90
		0.23

[a]Percentages indicate the level of extraction in flour preparation.

Source: Paul, A.A., Southgate, D.A.T., 1978. *McCance & Widdowson's* the Composition of Food, fourth ed. London: HMSO.

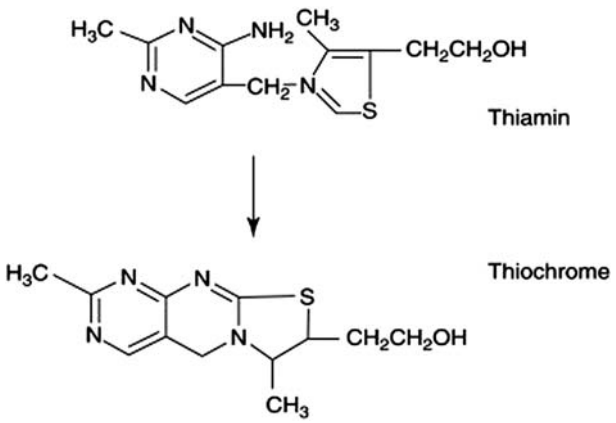


Fig. 2 Structures of thiamin and thiochrome.

nuts, large numbers of plants and vegetables, and some animal tissues. Anti-thiamin factors bind with varying degrees of attachment to thiamin and may or may not interfere with the bioavailability of thiamin. Diphenols, especially those with the hydroxyl groups in the *ortho* position, tend to react to give products that are both thiochrome negative and microbiologically inactive (i.e., thiamin is deactivated). Thus, in areas of Northern and North-Eastern Thailand where tea drinking, chewing fermented tea leaves, chewing betel nuts, and consuming raw/fermented fish are common practices, thiamin deficiency still occurs despite thiamin intakes of 0.44–0.50 mg/4.2 MJ. Similar food habits are also found in Laos but there, the mean thiamin intake of women during the post-partum period was recently reported to be 0.35 mg/4.2 MJ and infantile beriberi is a public health problem (Soukaloun et al., 2003; Barennes et al., 2009).

Low-dose thiamin supplements in lactating Cambodian mothers

Infantile beriberi related mortality is still common in South and SE Asia. Recent studies to investigate the effects of thiamin supplements on milk thiamin concentrations of Cambodian women during the first 6 mo postpartum were found to increase mean human milk thiamin concentrations at 24 weeks by 20, 25 and 36%. The doses given were zero, 1.2, 2.4 and 10 mg/d and all mean thiamin concentrations were significantly higher in the supplemented groups than the placebo but the range in concentration of milk thiamin was so wide that there was no significant difference between the 3 supplements. The objective of the study was to determine the dose of supplement that would equate to 90% of the maximum average human milk total thiamin concentration. The resultant figure based on all data was 2.35 mg/d but in fact the milk thiamine concentrations obtained from the 1.2 mg/d group were no different from those obtained with 2.4 mg/d (Gallant et al., 2021).

The large interindividual variation has been previously reported in thiamin-replete American and Cambodian women but the results may indicate one of the problems of supplementing women with a single nutrient where there may be several other dietary deficiencies interacting with the milk thiamin concentration e.g., riboflavin, iron, and other micronutrients. Furthermore the bolus dose of 10 mg was no more effective than 1.2 or 2.4 mg thiamine probably because active absorption of thiamine is restricted to 2 mg per meal (see Absorption below). A portion of higher amounts can be absorbed by passive diffusion but this may well be small. In conclusion, 1.2 mg/d achieved milk concentrations (1.83 ± 91 μg/L) comparable with adequate concentrations in thiamin-replete mothers and there were significant improvements in thiamin status based on the erythrocyte transketolase stimulation assay of both mothers and infants at 24 weeks from all three doses (Gallant et al., 2021).

Absorption and ethyl alcohol

In food, thiamin occurs mainly as phosphate coenzymes and the predominant form is thiamin diphosphate (TDP; also called thiamin pyrophosphate and cocarboxylase). The phosphate coenzymes are broken down in the gut by phosphatases to give free thiamin for absorption. Thiamin is absorbed mainly from the upper intestine, and less thiamin is absorbed on an empty stomach than when taken with a meal. The latter could be due to the alkaline conditions in the duodenum, which are prevented by the presence of food. Absorption of up to 2 mg per meal occurs by an active saturable process involving a sodium-dependent adenosine triphosphatase and against a concentration gradient. During absorption, thiamin is phosphorylated to the monophosphate ester (TMP). Thiamin is absorbed via the portal venous system. Further phosphorylation to TDP occurs on entry into all tissues. TDP can cross the blood–brain barrier, where a portion is converted to TTP, although even in the brain, TDP is the predominant form of thiamin. A second passive absorption process operates when intakes of thiamin are >5 mg but the maximum that can be absorbed from an oral dose is 2–5 mg (Friedemann et al., 1948; Hoyumpa et al., 1977).

The active process of absorption is impaired by ethyl alcohol. For example, 55% of a 5 mg dose of orally-administered, labeled thiamin was recovered over 72 h in healthy adults, but this was reduced by 25–40% if they were previously given 1.5–2 g alcohol/kg. In people with fatty livers who had previously been abusing alcohol, mean thiamin absorption was reduced by 60%. However, the passive absorption of thiamin is not inhibited by alcohol, nor does it block entry of thiamin into the liver or interfere with thiamin metabolism in the tissues. Absorption of thiamin may also be reduced by gastrointestinal disturbances, such as vomiting and diarrhea, ulcerative colitis, and neoplasia, and in patients with hepatic disease and achlorhydria.

Transport, storage and excretion

Thiamin with some TMP (19–75 nmol/L) circulates in the blood bound to albumin. When the binding capacity of plasma albumin is exceeded, or thiamin is in excess of tissue needs, it is rapidly excreted in the urine. Most thiamin in erythrocytes is present as TDP principally bound to the enzyme transketolase. Likewise, in most other tissues, there is very little free thiamin and it is mostly present as TDP (90%) in coenzymes bound to respective enzymes and a smaller amount of TTP (10%) in nervous tissues. The concentration of thiamin in specific tissues is of the order of 2–3 μg/g for heart muscle; 1 μg/g for brain, liver, and kidney; and 0.5 μg/g in skeletal muscle. Thiamin supplements can increase these concentrations slightly and prolonged febrile illnesses are likely to reduce them. Thiamin is mainly excreted intact in the urine but there are small amounts of thiochrome (**Fig. 2**) and other thiazole and pyrimidine metabolites. A linear relationship exists between intake and excretion of thiamin until intake falls to an amount approaching minimum requirements when excretion decreases rapidly indicating a renal conservation mechanism (Sinclair, 1982; Whitfield et al., 2018).

There is concern that the long-term use of diuretics in the management of chronic congestive heart failure (CHF) may impair thiamin status and, as a consequence, impair myocardial function (Seligmann et al., 1991; Suter and Vetter, 2000). The diuretic drug furosemide has been the subject of much attention. In healthy volunteers, a dose-dependent increase in urine flow accompanied by an increase in the urinary thiamin excretion rate have been demonstrated. In furosemide-treated patients, the concomitant presence of thiamin in the urine and biochemical deficiency of thiamin from measurements in blood has been shown. These results suggest that furosemide treatment can override the renal conservation mechanism. In one study, 23 patients with chronic CHF receiving 80–240 mg furosemide daily for 3–14 months were studied along with 16 age-matched controls without heart failure and not taking diuretics. No subjects in either group were identified as consuming inadequate thiamin intake or having increased

thiamin requirements. However, biochemically, 21 of the 23 CHF patients and 2 of the controls were biochemically thiamin deficient. Furthermore, 5 of the CHF patients were treated with intravenous thiamin (100 mg thiamin HCl twice daily for 7 days). Biochemical thiamin status normalized and echocardiographic assessment of left ventricular ejection fraction (the proportion of blood ejected per beat) increased in 4 of the 5 patients. Because no other changes were made in the patients' therapeutic regimen, the results suggest that the improvement in cardiac contractility was due to the correction of the thiamin deficiency.

Biological functions

Thiamin functions as the coenzyme TDP in the metabolism of carbohydrates and branched-chain amino acids (α-keto-isocaproic, α-keto-β-methyl valeric, and α-keto-isovaleric acids). In association with Mg^{2+} ions, TDP is important (1) in various dehydrogenase complexes for the oxidation of α-keto acids (pyruvate, α-ketoglutarate, and the branched-chain α-keto acids) and (2) in the formation of α-ketols among the hexose and pentose phosphates catalyzed by transketolase (EC 2.2.1.1). Thus, a deficiency of thiamin has severe consequences for energy generation and amino acid interconnections, and these have important links with lipid metabolism, cell replication, and neural activity (Whitfield et al., 2018).

Two principal dehydrogenase complexes that require the participation of TDP are pyruvate dehydrogenase, which generates acetyl-CoA, and the oxidative decarboxylation of α-ketoglutarate to succinyl-CoA (**Fig. 3**). Pyruvate dehydrogenase is situated at the junction of the glycolysis pathway, where it enters the tricarboxylic acid cycle. Acetyl-CoA is a key source of energy for mitochondrial oxidation and the production of adenosine triphosphate (ATP) as well as an important precursor in lipid metabolism. The impaired functioning of pyruvate dehydrogenase leads to a lactic acidosis, with increased concentrations of serum pyruvate and/or lactate especially as a result of exercise. The lactate acidosis can be explained by the fact that ATP depletion stimulates glycolysis, thus generating more pyruvate. As pyruvate concentrations increase, lactate dehydrogenase converts some of the pyruvate to lactate, producing the lactic acidosis. The increases in these compounds formed the basis of the earliest biochemical test for thiamin deficiency, which was later made more reproducible by taking the blood soon after moderate exercise (e.g., climbing a measured number of steps).

Many features of beriberi indicate that thiamin plays an important role in neural tissues (Sinclair, 1982; Bender, 1984). TTP is specifically found in nervous tissues, but although this triphosphorylated metabolite of thiamin has been known for approximately 30 years, its precise role is still in doubt. TDP in the dehydrogenase complexes is undoubtedly also required for normal function. Some of the earliest biochemical studies on the brain documented abnormalities in the oxidative metabolism of glucose and

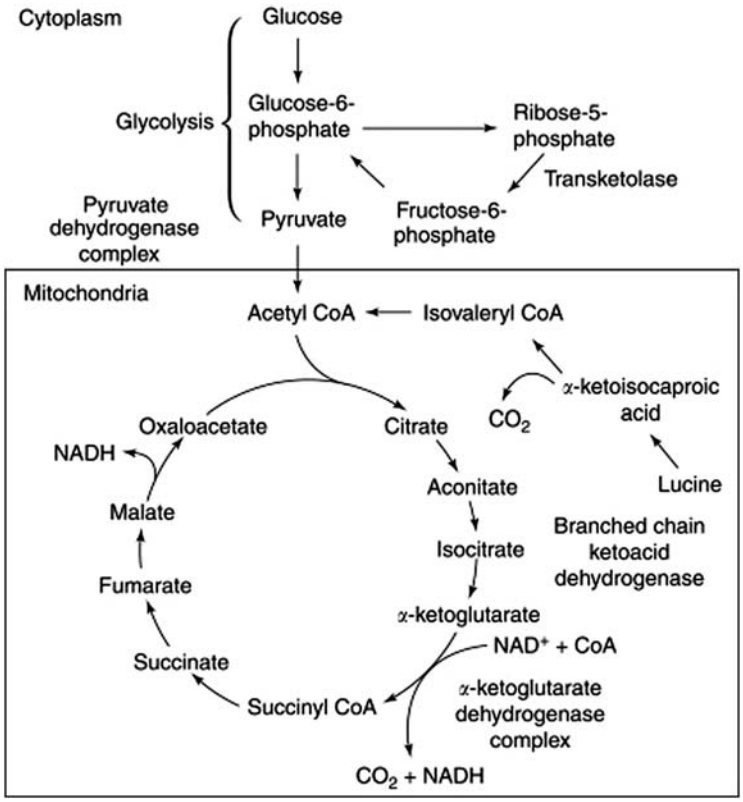


Fig. 3 The four principal sites of action of thiamin diphosphate coenzyme in carbohydrate metabolism.

a disruption in energy supply may underlie many of the neurochemical changes and structural lesions associated with thiamin deficiency. For example, acetyl-CoA produced by pyruvate dehydrogenase is a precursor of the parasympathetic transmitter molecule acetylcholine, but the obligatory requirement of glucose as an energy source for nervous tissue indicates the essentiality of TDP. Likewise, the cytosolic enzyme transketolase is also present in nervous tissue, and as a key enzyme in the hexose monophosphate shunt (HMS) it may be important in minimizing oxidant stress. The HMS generates NADPH, which is required to maintain the antioxidant compound, glutathione, in the reduced state. Thus, thiamin has an important role in maintaining redox homeostasis of tissues.

The cellular and subcellular localization of the enzymes responsible for metabolism of thiamin phosphates in nervous tissues may indicate possible sites of action of the specific metabolites. Thiamin that enters the brain is phosphorylated by thiamin pyrophosphokinase to form TDP. The concentration of thiamin phosphates is 3 or 4 times higher in neurons than in neuroglia, and the activity of thiamin diphosphatase (TDPase), which converts TDP to TMP, is 20 times higher in neurons than neuroglia. Thiamin monophosphatase is only detected in neuroglia. Within the neuron, TDPase is mostly localized in the microsomal fraction. Thiamin triphosphatase (TTPase), which converts TTP to TDP, is particularly enriched in presynaptic terminals. Stimulation of nerves or treatment with certain neuroactive drugs result in decreases in TDP and particularly TTP in the nerve, with an increase in free TMP in the surrounding fluid. It is postulated that TTP plays an essential role in nerve transmission involving a gating mechanism for sodium and potassium ion transport via the specific ATPase. Some evidence for this comes from patients with Leigh's disease (pathologically similar to Wernicke–Korsakoff syndrome), in whom severe neurological disease is accompanied by a deficiency in TTP but normal TDP concentrations.

The well-documented role of mitochondria in programmed cell death and the importance of thiamin for oxidative stability have stimulated investigators to examine brain thiamin homeostasis in neurodegenerative diseases (Gibson and Zhang, 2002). Diminished thiamin-dependent processes, abnormal metabolism, and oxidative stress accompany the neurodegeneration of Alzheimer's disease (AD), Huntington's disease, Wernicke–Korsakoff syndrome, progressive supranuclear palsy, and the adult-onset neurodegenerative diseases that are caused by genes containing variable numbers of DNA-triplets, otherwise known as CAG repeats, within their coding regions. Abnormalities in the thiamin-dependent processes have also been linked with thiamin-responsive maple syrup urine disease, Leigh's disease (a subacute necrotizing encephalomyelopathy), sudden infant death syndrome, cerebellar degeneration, thiamin-responsive anemia, ataxia, and disorders of energy metabolism including pyruvate dehydrogenase deficiency. The extent to which disturbances in thiamin metabolism are a cause or a consequence of the disease process is still under examination. The wide-ranging associations of thiamin deficiency with metabolic diseases justifies the term thiamin deficiency disorders (TDD) which has recently been introduced (Whitfield et al., 2018).

Assessment of thiamin status

Thiamin status can be assessed using methods that measure thiamin or its metabolites in plasma, erythrocytes and urine (**Table 2**). Samples are acidified to stabilize the thiamin and precipitate any protein. Usually, thiamin is oxidized to thiochrome (**Fig. 2**) using cyanogen bromide in alkaline solution and measured by fluorescence with or without chromatography. Concentrations of thiamin in urine and plasma tend to reflect dietary intake, being high when intake is adequate and low when dietary sources are poor. Erythrocyte thiamin is mainly in the form of the coenzyme TDP, which can be extracted from washed erythrocytes, derivatized as described previously, and quantified by high-performance liquid chromatography. The most useful functional test, however, is the erythrocyte transketolase (ETKL) stimulation test, which measures enzyme activity with and without added TDP. The reference range for ETKL activity in well-nourished, thiamin-adequate people is reported to be 570–830 mU/g hemoglobin. The stimulation test measures the proportion of the TKL apoenzyme in red cell homogenate (i.e., the proportion that is not bound to TDP and

Table 2 Biochemical assessment of thiamin status.

Test	*Acceptable*	*Marginal risk*	*High risk*
Urinary thiamin (μmol/mol creatinine)[a]			
1–3 years	>66	45–66	<45
4–6 years	>45	32–45	<32
Adults	>25	10–25	<10
Erythrocyte transketolase activity[c]			
Activity coefficient	<1.11	1.11–1.25	>1.25
TDP effect (%)	<11	11–25	>25
Red cell thiamin concentrations (nmol/L^{-1})	749 ± 196		560[b]
Whole blood thiamin concentrations (nmol/L^{-1})[c,d]	166–266		<133

[a]Converted from μg/g using creatinine the factor (×0.376).
[b]Based on a decrease of 25% in red cell thiamin diphosphate (TDP).
[c]Reproduced from Thurnham (1985).
[d]Reproduced from Gibson (2005).

represents the degree of thiamin deficiency). Studies have shown that results from the urinary assay for thiamin agree reasonably well with those obtained by the ETKL stimulation test. The ETKL stimulation test can now be done using 96-well plates which will considerably improve the speed of measurement of thiamin status (Jones et al., 2021).

One of the reasons for the usefulness of the ETKL stimulation test is that sensitivity is still good even in the presence of thiamin deficiency. In all other measurements of thiamin status, as deficiency approaches, the quantity of thiamin or its metabolites diminishes in the biological fluid. Low concentrations of a product are usually more difficult to measure and precision deteriorates, or the amount of sample has to be increased to provide sufficient material to detect. In contrast with the ETKL stimulation test, in an acute thiamin deficiency, ETKL activity is maintained and only the amount of TDP decreases, so the test becomes more sensitive. However, in chronic thiamin-deficient states, the apoenzyme of ETKL is reported to be unstable in vivo, and in the absence of the coenzyme, concentrations of the apoenzyme decrease, with the result that *in vitro* stimulation may show normal thiamin status and basal ETKL activity may be a better indicator of status than the stimulation test. Thus, in situations in which chronic thiamin deficiency is suspected as a result of a long-term marginal thiamin intake, alcohol abuse, or use of diuretics for many months, one or more of the concentration tests may be useful as an adjunct to the stimulation test.

Certain precautions should be taken in handling samples for thiamin analysis. Urine should be acidified to avoid degradation and stored below $-20\,^{\circ}C$. Heparinized whole blood should be collected and immediately put on ice. For total erythrocyte TDP measurements, cells are separated from plasma within 2 h when possible, washed in saline, and diluted 1:1 with saline prior to acidification. Centrifugation of the acidified mixture provides a clear extract that can be stored for no more than 5 days at $4\,^{\circ}C$ or longer at $\leq -20\,^{\circ}C$. Washed red cells are also used for the ETKL assay. Duplicate tubes of the red cells in saline suspension with and without added TDP are mixed and can be stored at $-70\,^{\circ}C$ prior to enzymatic analysis of ETKL activity. Even at $-70\,^{\circ}C$, however, storage should be for no more than a few weeks. The ETKL apoenzyme is unstable, and even in the tubes to which TDP has been added, if mixing did not thoroughly expose all apoenzyme to the added coenzyme, deterioration will occur and results will be unreliable.

Recommended dietary allowances (Department of Health, 1991; WHO/FAO, 2002)

Quantifying thiamin requirements is based on a variety of biochemical data (Sauberlich et al., 1979). Early results indicated that a thiamin intake of 0.4 mg/day on a low-energy intake was close to the absolute minimum requirement. Epidemiological evidence suggested that beriberi occurred when the intake of thiamin was $<$0.2 mg thiamin per 4.2 MJ (1000 kcal); however, when 0.188 mg/4.2 MJ was fed to sedentary elderly men for 2 years, no indisputable alteration in clinical state occurred. Thiamin requirements are strongly influenced by physical activity and at higher energy intakes with liquid formula diets containing 11.76 and 15.12 MJ (2800 and 3600 kcal), there was good agreement between thiamin excretion and ETKL stimulation to interpret thiamin status at different levels of thiamin intake. Increasing intake from 0.2 to 0.23 mg/4.2 MJ moved first the urinary excretion and then ETKL activation out of the deficient range. Both measurements were normalized at intakes of 0.3 mg/4.2 MJ, and to allow for variance the recommended nutrient intake adopted by the Department of Health in the United Kingdom was 0.4 mg/4.2 MJ. This amount is recommended for all groups of the population since additional needs in pregnancy and lactation are met by increased energy intakes. It was recommended that formula feed should contain not less than 0.3 mg/4.2 MJ.

Women are less affected by beriberi than are men even when they are consuming the same diet, but there is no consistent indication that men have greater needs than women. Differences between the sexes that may affect susceptibility to beriberi need further investigation (e.g., the amount of food eaten by the sexes when supplies are short or of poor quality, metabolic responses to infection during illness, and differences in energy requirements). The close association between thiamin metabolism and carbohydrate metabolism means that thiamin requirements are determined by basal metabolic rate (BMR) and physical activity. BMR of men is slightly higher than that of women of the same weight, but total energy expenditure can vary 1.4–2.5 times BMR depending on physical activity.

Drug–nutrient interactions

Mention has already been made of the influence of alcohol and diuretics on thiamin status. Oral contraceptives are reported to have no effects on thiamin status (Thurnham, 2004). Some recent publications have drawn attention to transporter-mediated drug nutrient interactions in humans which may cause potentially serious adverse events (Vora et al., 2020; Enogieru et al., 2021). Genetic mutations of the thiamine transporters SLC19A3 and SLC19A2 which encode thiamine transporters 2 and 1 respectively can be clinically important particularly in vulnerable populations e.g. individuals with alcoholism. Vora et al. (2020) identified 28 oral drugs, including metformin (used in treatment of type-2 diabetes) where intestinal concentrations were predicted to inhibit thiamin absorption. Likewise, others report a case of thiamin-responsive megaloblastic anemia caused by mutations of the SLC19A2 gene. These workers identified 10 drugs that inhibited SLC19A2-mediated thiamin transport but only erythromycin (antibiotic used to treat a number of bacterial infections) was active at clinically relevant concentrations (Enogieru et al., 2021).

Toxicity

High intakes of thiamin administered orally are nontoxic. The rapidly saturable thiamin absorption mechanism limits the amount taken up from a single dose to ~2.5 mg, and thiamin present in excess of protein binding capacity is excreted. However, there are reports of toxicity from chronic intakes in excess of 50 mg/kg or >3 g/day with a wide variety of clinical signs, including headache, irritability, insomnia, rapid pulse, weakness, rapid pulse contact dermatitis, pruritus, and, in one case, death (Sinclair, 1982).

Conclusions

Thiamin is needed to support all metabolic functions necessary for life. In the absence of a regular supply of thiamin e.g., when deprived by restricted feeding, a lack of food choice, anorexia etc. or, there is increased requirement for thiamin as in fever or infection or increased physical activity or when treated with a glucose load for medical reasons, there is a risk of insufficient thiamin which can impair physical activity, cardiac function and, long term, has neurological consequences. Epidemiologically, those low and middle-income populations that use polished rice as their staple food are at greatest risk of chronic thiamin deficiency and infants may be at the greatest risk of deficiency if they are solely fed with breast milk in the first 6 months of life. Food fortification with thiamin can address this problem as in many developed countries, and governments of LMIC should investigate similar schemes with local foods.

See Also: Carbohydrates: Regulation of metabolism; Cereal grains; Drug-nutrition interactions; Thiamin: Beriberi

References

Barennes, H., Simmala, C., Odermatt, P., et al., 2009. Postpartum traditions and nutrition practices among urban Lao women and their infants in Vientiane, Lao PDR. Eur. J. Clin. Nutr. 63, 323–331.

Bender, D.A., 1984. B vitamins in the nervous system. Neurochem. Int. 6, 297–321.

Department of Health, 1991. Dietary Reference Values for Food Energy and Nutrients for the United Kingdom. HMSO, London Report on Health and Social Subjects No. 41.

Enogieru, O.J., Koleske, M.L., Vora, B., Ngo, H., et al., 2021. The effects of genetic mutations and drugs on the activity of the thiamine transporter, SLC19A2. Am. Assoc. Pharm. Sci. J. 23 article 35.

Friedemann, T.E., Kmieciak, T.C., Keegan, P.K., Blum Sheft, B., 1948. The absorption, excretion and destruction of orally-administered thiamin by human subjects. Gastroenterology 11, 101114.

Gallant, J., Chan, K., Green, T.J., Wieringa, F.T., et al., 2021. Low dose thiamine supplementation of lactating Cambodian mothers improves human milk thiamine concentrations: a randomized control trial. Am. J. Clin. Nutr. 114, 90–110.

Gibson, G.E., Zhang, H., 2002. Interactions of oxidative stress with thiamine homeostasis promote neurodegeneration. Neurochem. Int. 40, 493–504.

Assessment of the status of thiamin, riboflavin and niacin. In: Gibson, R.S. (Ed.), 2005. Principles of Nutritional Assessment. Oxford University Press, Oxford, pp. 545–574. Chap. 20.

Hoyumpa, A.M.J., Nichols, S.G., Wilson, F.A., Schenker, S., 1977. Effect of ethanol on intestinal (Na,K) ATPase and intestinal thiamin transport in rats. J. Lab. Clin. Med. 90, 1086–1095.

Jones, K.S., Parkington, D.A., Cox, L.J., Koulman, A., 2021. Erythrocyte transketolase activity coefficient (ETKAC) assay protocol for the assessment of thiamine status. Ann. N. Y. Acad. Sci. 1498 (1), 77–84.

Paul, A.A., Southgate, D.A.T., 1978. *McCance* & *Widdowson's* the Composition of Food, fourth ed. HMSO, London.

Sauberlich, H.E., Herman, Y.F., Stevens, C.O., Herman, R.H., 1979. Thiamin requirements of the adult human. Am. J. Clin. Nutr. 32, 2237–2248.

Seligmann, H., Halkin, H., Rauchfleisch, S., et al., 1991. Thiamine deficiency in patients with congestive heart failure receiving long-term furosemide therapy: a pilot study. Am. J. Med. 91, 151–155.

Sinclair, H.M., 1982. Thiamin. In: Barker, B.M., Bender, D.A. (Eds.), Vitamins in Medicine, fourth ed. Heinemann Medical Books, London, pp. 114–167.

Soukaloun, D., Kounnavong, S., Pengdy, B., et al., 2003. Dietary and socio-economic factors associated with beriberi in breastfed Lao infants. Ann. Trop. Paediatr. 23, 181–186.

Suter, P.M., Vetter, W., 2000. Diuretics and vitamin B_1: are diuretics a risk factor for thiamin malnutrition? Nutr. Rev. 58, 319–323.

Thurnham, D.I., Cathcart, A.E., Livingston, M.B.E., 2011. A retrospective investigation of thiamin and energy intakes following an outbreak of beriberi in the Gambia. Nutrients 3, 135–151.

Thurnham, D.I., 1985. Interpretation of biochemical measurements of vitamin status in the elderly. In: Kemm, J. (Ed.), Vitamin Deficiency in the Elderly. Prevalence, Clinical Significance and Effects on Brain Function. Pergamon Press, London, pp. 46–67.

Thurnham, D.I., 2004. An overview of interactions between micronutrients and between micronutrients with drugs, genes and immune mechanisms. Nutr. Res. Rev. 17, 211–240.

Vora, B., Green, E.A.E., Khun, N., Bellgren, F., et al., 2020. Drug-nutrient interactions: discovering prescription drug inhibitors of the thiamine transporter ThTR-2 (SLC19A3). Am. J. Clin. Nutr. 111, 110–121.

Whitfield, K.C., Bourassa, M.W., Adamolekun, B., Bergeron, G., et al., 2018. Thiamine deficiency disorders: diagnosis, prevalence and a roadmap for global control programs. Ann. N. Y. Acad. Sci. 1–41.

WHO/FAO, 2002. Human Vitamin and Mineral Requirements; Report of a Joint FAO/WHO Expert Consultation. WHO/FAO, Rome.

Ultratrace elements

Forrest H. Nielsen, Research Nutritionist Consultant, Grand Forks, ND, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of F.H. Nielsen, Trace and Ultratrace Elements, Reference Module in Biomedical Sciences, Elsevier, 2014, ISBN 9780128012383, https://doi.org/10.1016/B978-0-12-801238-3.00235-X.

Introduction 448
Aluminum 449
Arsenic 449
Boron 450
Bromine 451
Cadmium 452
Fluorine 452
Germanium 452
Lead 453
Lithium 453
Molybdenum 453
Nickel 454
Rubidium 454
Silicon 455
Strontium 455
Tin 456
Vanadium 456
Summary 456
References 457

Key points

- Most ultratrace elements once suggested to be essential based mostly on an apparent beneficial actions in animals and essential functions in lower forms of life have not been established as essential in humans when using the rigorous definition of essentiality, which is: required to prevent death or complete the life cycle.
- Boron, bromine and silicon are ultratrace elements for which there is substantial evidence they will be found to be essential.
- Fluoride and lithium are elements recognized as having beneficial biological activity in supra nutritional amounts, but not enough evidence to be considered nutritionally essential.
- Molybdenum is an essential ultratrace element of limited nutritional concern.
- Nickel and vanadium may have beneficial biological activity in nutritional amounts.

Introduction

In the early 20th century, scientists could qualitatively detect small amounts of several mineral elements in living organism. In these reports, these elements were described as being present in "traces" or "trace amounts," and became known as trace elements. Sophisticated analytical techniques have resulted in the accurate determination of many mineral elements, some at very low concentrations in living organisms. Some of these elements are essential, that is, having a defined biochemical function that is needed to prevent death or complete the life cycle. Most mineral elements are nonessential fortuitous reminders of our geochemical origin or indicators of environmental exposure. However, some trace elements not established as essential have been found to have beneficial effects in humans and animals in nutritional or supra nutritional amounts. All the trace elements are toxic when intakes are excessive.

In the 1980s, the term "ultratrace element" appeared in the nutritional literature. The definition given for ultratrace elements for humans was an element with an established, estimated, or hypothesized requirement of less than 1 mg day^{-1} or indicated by micrograms per day. Using this definition for ultratrace elements resulted in three recognized essential mineral elements required by humans fitting into the ultratrace category; these elements are iodine, molybdenum, and selenium. Cobalt probably could be considered an ultratrace element. However, it is only required in the form of vitamin B_{12}, and thus its requirement is usually presented in this form. Iodine and selenium, in addition to chromium, an ultratrace element that has not been established as essential, are discussed elsewhere in this encyclopedia.

In the last 30 years of the 20th century, numerous ultratrace elements received some limited recognition as being essential based on a less rigorous definition of essentiality. The essentiality definition was that a dietary deficiency of a substance had to consistently and adversely change a biological function from optimal, and this change was preventable or reversible by physiological or nutritional amounts of the substance. For the ultratrace elements, this definition became questionable because of concerns about possible suboptimal diets and environments in studies reporting possible essentiality. Instead of overcoming a suboptimal function, it was considered possible that mineral element supplementation was having a pharmacological effect such as alleviating a pathology caused by a poor diet or environment, substituting for an another essential nutrient provided in a deficient amount, or having an effect on intestinal microorganisms in a manner that was beneficial to the host. Minerals that once were suggested to be essential in ultratrace amounts based mostly on an apparent beneficial action in animals and essential functions in lower forms of life included aluminum, arsenic, boron, bromine, cadmium, fluoride, germanium, lead, lithium, nickel, rubidium, silicon, strontium, tin, and vanadium. These elements will be the focus of this article. Also included is the established essential ultratrace element molybdenum which is given very little nutritional attention apparently because a human nutritional deficiency has not been unequivocally identified other than in an individual nourished by total parenteral nutrition.

Aluminum

In the 1990s, it was reported that a dietary deficiency of aluminum in goats increased abortions, depressed growth, caused incoordination and weakness in the hind legs, and decreased life span (Angelow et al., 1993). It was also reported that aluminum deficiency depressed growth in chicks (Carlisle and Curran, 1993). Neither of these reports have been confirmed. In both these studies, the diet and environment were not optimal, which suggests that the effect of aluminum may have been something other than that of an essential function. There is no question that aluminum can be bioactive. *In vitro* studies show that it can activate the enzyme adenylate cyclase, enhance calmodulin activity, stimulate DNA synthesis, and upregulate the parathyroid calcium-sensing receptor (Nielsen, 1996). However, aluminum is not required to perform these functions in living organisms. Most evidence indicates that for humans the body attempts to minimize the exposure of aluminum. Intestinal absorption is estimated to be 0.005% of intake. Any absorbed aluminum entering circulation is excreted by the kidney or becomes complexed to cellular components such citrate, glutamic acid and entities containing phosphate molecules without disrupting cellular biochemistry. Thus, although it is found in the body in ultratrace amounts, there is insufficient evidence to support aluminum as being essential or indicate it having beneficial bioactivity in humans, which makes it inappropriate to give a suggested beneficial intake. If aluminum has a beneficial activity, the typical daily dietary intake of aluminum of 2–25 mg likely provides an adequate amount to accomplish it.

Aluminum toxicity is not of concern for healthy individuals. However, high intakes of aluminum by susceptible individuals (e.g., impaired kidney function, low-birth weight infants) can lead to aluminum accumulation in various organs including bone, parathyroid, brain, liver, and kidney. Bypassing natural barriers to the entry of aluminum in the body such as intravenous infusion of parenteral solutions high in aluminum also is another risk for aluminum accumulation in these organs. This accumulation can lead to pathology such as osteomalacia, hypoparathyroidism, and microcytic anemia (Klein, 2019), and perhaps exacerbate neurodegenerative disease such as Alzheimer's disease (Exley and Mold, 2019). For most healthy individuals, an aluminum intake of 125 mg day^{-1} should not result in toxicological consequences.

Arsenic

Arsenic is unquestionably a bioactive element in higher animals and humans. However, most studies of its bioactivity have been directed toward demonstrating that some forms of arsenic in amounts found in food and water are toxic. The concept that arsenic has beneficial activity in nutritional intakes is not well accepted, although numerous reports have purported that feeding a diet low in arsenic (e.g., <12 μg kg^{-1} diet for rats, chicks, and hamsters, and <35 μg kg^{-1} for goats and minipigs) resulted in an arsenic deficiency in animals that was overcome by low microgram quantities of supplemental arsenic (Uthus, 1992; Anke, 2005). The most consistent signs of the reported arsenic deprivation signs were depressed growth and abnormal reproduction characterized by impaired fertility and increased perinatal mortality. Death with myocardial damage also was reported for lactating goats. Many of these obtained deficiency findings were affected by various dietary manipulations and perhaps environmental conditions. For example, many reported signs of arsenic deprivation were changed and generally enhanced by nutritional stressors that affected sulfur amino acid or labile methyl metabolism. Because the "deprivation" findings might be the result of arsenic supplementation alleviating a suboptimal status caused by factor/s other than an arsenic deficiency, the findings are not sufficient to support essentiality. However, they indicate that arsenic could be beneficial in ultratrace amounts.

Interestingly, although considered to be carcinogenic, some studies indicate that certain amounts and forms arsenic may be beneficial through reducing the risk for cancer. In one human study, both low exposure (<50 μg L^{-1}) and high exposure (>100 μL^{-1}) in drinking water were associated with increased cancer incidence (Kayajanian, 2003). A similar finding was obtained by a study involving rats fed diets containing 0.1, 0.5, or 50 μg kg^{-1} arsenic with dimethylhydrazine induced aberrant colon crypts.

More aberrant crypts and aberrant crypt foci, a relative DNA hypomethylation, and increased activity of DNA methyltransferase was found in rats fed the 0.1 and 50 μg kg^{-1} diets than those fed the 0.5 μg kg^{-1} diet (Uthus and Davis, 2005). Another human study found decreased serum arsenic concentrations in individuals undergoing hemodialysis treatment was correlated with cancer, in addition to vascular diseases and central nervous system injury (Mayer et al., 1993). Consistent with arsenic sometimes having beneficial bioactivity against cancer is the finding that pharmacological amounts of arsenic trioxide are effective treatments for some forms of cancer, especially promyelocytic leukemia (Liu et al., 2021).

A plausible mechanism for the bioactivity of arsenic is an effect on the methylation of metabolically or genetically important molecules. This suggestion is supported by the finding that global DNA methylation in cultured Caco-2 cells was significantly decreased when the culture contained low (25 μg L^{-1}) or high (175 μg L^{-1}) arsenic than when it contained 100 μg L^{-1} (Davis et al., 2000).

Dietary Reference Intakes (DRIs) have not been set for arsenic. Several studies indicate that human arsenic intakes range from 20 to 70 μg day^{-1}. This intake is similar to intakes found beneficial in animals and epidemiological studies. Animal data suggest that a beneficial intake of arsenic could be 12–25 μg day^{-1} (Nielsen, 1996).

Despite the findings above, any measurable intake arsenic not in the arsenobetaine form is mostly considered a toxicological concern. This occurs although humans have enzymes that convert absorbed inorganic arsenic into a methylated form that is readily excreted in the urine (Thomas et al., 2007). Organic forms of arsenic that are readily absorbed, such as arsenobetaine and arsenocholine that is mostly converted to arsenobetaine, are found in significant amounts in seafoods; arsenobetaine passes through the body without transformation and is virtually nontoxic (Nielsen, 2006). A standard known as the reference dose (RfD; lifetime exposure that is unlikely to cause adverse health effects) of 0.3 μg kg^{-1} body weight per day, or 21 μg day^{-1} for a 70 kg human has been suggested for inorganic arsenic (Chappell et al., 1994). This intake conforms with the beneficial intakes indicated above.

The finding that arsenic might have beneficial effects in ultratrace amounts suggests that the human response curve to arsenic intakes is U or J shape where increasing or decreasing, respectively, very low or excessive intakes would be beneficial. Only limited concern about low intakes is needed because most reported intakes through water and diet exceed this amount, especially if the diet is high in cereal grains. On the other hand, excessive intakes of inorganic arsenic, especially trivalent arsenic (arsenite), is a significant global concern. Signs of subacute and chronic excessive intakes of inorganic arsenic include various dermatoses (hyperpigmentation, hyperkeratosis, desquamation, and hair loss) liver damage characterized by jaundice, portal cirrhosis, and ascites; peripheral neuritis; and hematopoietic depression. Excessive intakes have also been linked to increased risk for skin, lung, liver, kidney, and bladder cancers (Stoeppler, 2004). Excessive intakes inducing such pathologies are found in specific areas of countries where drinking water is high in inorganic arsenic, including Argentina, Bangladesh, Bolivia, Chile, China, Hungary, India, Mexico, Nepal, Romania, Taiwan, Thailand, Peru, United States, and Vietnam (Stoeppler, 2004).

Consuming foods from plants high in inorganic arsenic because of uptake from arsenic contaminated soil and water also can result in excessive intakes of arsenic. Of major concern is rice which is grown on soils throughout the world that favor the uptake of inorganic arsenic to amounts 10–20 times that usually are found in grains such as wheat. Rice in many parts of the world is a food staple and rice and rice products are commonly consumed in other countries in significant quantities other than just rice such as in ready-to-eat cereals, baked goods, rice cakes, and rice-based snacks in the daily diet. Of special concern are foods for infants and toddlers whose first foods are often rice-based. The European Commission (European Food Safety Authority et al., 2021) has given a limit of 100 μg/kg for inorganic arsenic in rice-based foods for infants and young children. The United States Food and Drug Administration in 2016 proposed a limit of 100 μg/kg arsenic in cereal and baby food. It should be noted that arsenic in rice varies greatly based on the variety and where grown, so not all rice and its products are a substantial source of arsenic. For example, rice grown in Asia has been found to be higher in inorganic arsenic than rice grown in the United States. Moreover, if rice and its products are not a major component of the diet, inorganic trivalent arsenic intake likely would not exceed suggested safe dietary intakes (Uthus, 1994).

Boron

Boron has been shown to be essential for the completion of the life cycle for organisms in all phylogenetic kingdoms. Frogs and zebra fish are animals for which boron has been found essential. Boron has been found to be needed for cleavage of zebra fish zygotes to the two and four cell preblastula stage of development (Eckhert and Rowe, 1999), and for the normal morphogenesis of the frog *Xenopus laevis* (Fort et al., 2002). Boron has not been established as nutritionally essential for humans because it does not have a defined biochemical function. However, an increasing number of cell culture, animal, and human studies indicate that boron in nutritional amounts has beneficial effects in human health.

The beneficial effects of boron have been suggested to occur because boric acid and borates form ester complexes several biologically important sugars, especially ribose (Hunt, 2002). Through the formation of these esters, boron may modify the functions of biochemical entities containing ribose such as S-adenosylmethionine, diadenosine phosphate, oxidized nicotinic adenine dinucleotide (NAD^+), and the NAD^+ metabolite cyclic adenosine diphosphate c(ADP) ribose (cADPR), which are involved in bone formation and maintenance, cardiovascular health, cancer risk, and neurological function (Nielsen and Eckhert, 2020). Boron has been shown to beneficially affect these processes in animals and humans.

Boron also has been suggested to be beneficially bioactive through forming complexes with phosphoinositide, glycoproteins, and glycolipids that contain *cis*-hydroxyl groups in membranes (Nielsen and Meacham, 2011). This apparently is the basis for boron influencing cellular membrane transport in plants (Bolaños et al., 2004). Thus, boron might have beneficial effects in animals and humans through affecting membrane function such that it influences transmembrane signaling or transmembrane movement of anions, cations, and regulatory molecules. This may be the basis for boron supplementation after boron deprivation beneficially affecting numerous cell and organ functions.

Boron supplementation of animals fed diets made very low in boron (generally <0.2 mg kg^{-1}) and cells in culture media (<0.02 ng mL^{-1}) has been found to improve trabecular and alveolar bone growth and maintenance (Nielsen, 2012). Among the cell culture studies, it was found that four mineralization genes and five genes involved in cell proliferation and differentiation were upregulated by boron in osteoblastic cells. Boron also has been found to have a regulatory effect on tuftelin 1 gene that is of primary importance in formation of mineralized structures such as enamel and alveolar bone (Hakki et al., 2021). Cell culture studies have shown the boron can activate transcription factor 4 (ATF4) which is required for the differentiation of osteoblasts, osteogenesis, and thus bone remodeling (Nielsen and Eckhert, 2020).

Since 2008, numerous studies have shown that boron can modulate the response to inflammatory stress. These studies were with animals and cultured cells with induced oxidative or inflammatory stress, and supplementation and epidemiological of individuals with oxidative or inflammatory stress. Because inflammatory and oxidative stress are considered risk factors for atherosclerosis and cardiovascular disease, boron has been implicated as being a beneficial element in respect to these pathologies. In cultured human prostate cells, boron acts as a reversible noncompetitive inhibitor of cADPR leading to decreased endoplasmic Ca^{2+} (Kobylewski et al., 2016). The increased Ca^{2+} activates nuclear factor erythroid 2 like 2 (Nrf2) which increases mRNA antioxidant response genes. Nrf2 directly regulates the induction of *anti*-inflammatory genes and indirectly counteracts inflammation by modulating reactive oxygen and nitrogen species (Yamada and Eckhert, 2019). Several human studies have shown that boron supplementation reduced serum C-reactive protein (CRP) in individuals with concentrations >3.0 mg L^{-1}, which is considered an indicator of inflammatory stress (Nielsen, 2020). For example, in a double-blind, placebo-controlled clinical study of healthy individuals with serum CRP concentrations >3.0 mg L^{-1}, supplements of 3 or 6 mg day^{-1} of boron as calcium fructoborate significantly decreased the serum concentrations of inflammatory markers CRP, interlukin-6, and monocyte chemoattractant protein-1 (Rogoveanu et al., 2015).

Modulation of inflammatory and oxidative stress also might be the basis for some of the other beneficial bioactivity reported for boron (Nielsen, 2012). Occasional reports have indicated that boron supplementation can alleviate the symptoms of arthritis. Boron has been inversely associated with prostate, cervical, lung, and breast cancer. Boron has been found to inhibit the growth or proliferation of some types of cultured cancerous prostate cells.

Findings indicating boron can affect cell signaling and calcium ion cellular transport, release and utilization likely explains reports that boron increases insulin sensitivity in chicks and rats, exacerbates gross bone abnormalities in chicks and rats fed low vitamin D diets, made oocytes responsive to progesterone in frogs, and enhanced the estrogen response in rats and humans (Nielsen, 2014). Alteration in cell membrane function also might be the basis for the finding that boron supplementation after boron deprivation resulted in electroencephalograms indicating less drowsiness and increased mental alertness, improved motor speed and dexterity skills, and improved attention and short-term memory in humans (Nielsen, 2014).

Neither a Recommend Dietary Allowance (RDA) nor an AI has been established for boron. Human depletion-repletion experiments found a response to boron supplementation after the consumption of a diet supplying only 0.2–0.4 mg day^{-1}. Human findings indicate that a fasting serum concentration <34 mg L^{-1} and dietary intake of <1.0 mg day^{-1} could be an indication that a person would benefit from an intake of boron of ≥ 1.0 mg day^{-1}. Good sources of boron are foods and drinks obtained from plants, including fruits, leafy vegetables, and nuts. Avocados, dried fruits such as raisins, peanuts, pecans, prune juice, grape juice, wine, and chocolate powder are especially rich sources of boron.

The Upper Intake levels (ULs) set by various official bodies indicate boron is a relatively nontoxic element. The United States ULs are, for children aged 1–3 years, 3 mg day^{-1}; and aged 4–8 years, 6 mg day^{-1}. The UL for adolescents aged 9–18 years is 17 mg day^{-1}, and for adults, 20 mg day^{-1} (Institute of Medicine, Food and Nutrition Board, 2001). The World Health Organization, International Program on Chemical Safety (1998) has indicated that a safe UL would be 0.4 mg kg^{-1} body weight or $\sim$28 mg day^{-1} for a 70 kg person. The European Food Safety Authority (2013) established an Acceptable Daily Intake for boron at 0.16 mg day^{-1}, or $\sim$11.2 mg day^{-1} for a 70 kg person. The past use of boric acid and borates as food preservatives provides further evidence of the low order of toxicity for boron.

Bromine

There is evidence that indicates bromine could be considered an essential ultratrace element because a defined biochemical function that occurs in humans has been identified in *Drosophila*; bromide deficiency is lethal to *Drosophila* (McCall et al., 2014). Bromide has been identified as a required co-factor for the peroxidasin-catalyzed formation of sulfilimine crosslinks of the collagen IV scaffold in basement membranes needed for tissue development (McCall et al., 2014). A bromoester, 2-octyl 4-bromo-3-oxobutanoate, in cerebrospinal fluid (Gribble, 1999) has been found to induce rapid eye movement sleep and to have *anti*-cholesterase activity. Insomnia exhibited by some hemodialysis patients has been suggested to be caused by a bromide deficiency (Oe et al., 1981). One laboratory has reported bromide deficiency depresses growth, fertility, hematocrit, hemoglobin, and life expectancy, and increases abortions in goats (Anke et al., 1993b). These findings have not been confirmed in another laboratory. Bromide also

can have beneficial activity by affecting essential halogen metabolism. Bromide has been found to alleviate growth retardation caused by hyperthyroidism in mice and chicks (Huff et al., 1956; Bosshardt et al., 1956), and to substitute for part of the chloride requirement of chicks (Leach, 1963). Bromine apparently is essential to lower forms of life. A bromoperoxidase has been found in marine algae that oxidizes bromide to facilitate the formation of the carbon-bromine bond in bromine-containing compounds (Vilter, 1995), which apparently are essential to the algae.

Because it has not been firmly established an essential ultratrace element, neither a RDA nor an AI has been established for bromide. If bromide is required by humans, the requirement likely would be very low because of its limited number of functions in the body. A low requirement similar to iodine or $<300\ \mu g\ day^{-1}$ would be easily met by the typical human bromine intake of 2–5 mg day^{-1}. Bromine ingested as bromide has a low order of toxicity for humans and thus is not of toxicological concern in nutrition. Pharmacological bromide intakes of 500 mg–1.0 g day^{-1} have been found to cause toxicity.

Cadmium

Very limited evidence exists to indicate that cadmium can be a beneficial element for humans. In the 1970s, it was reported at scientific conferences that cadmium supplementation to rats fed $>4\ \mu g\ kg^{-1}$ diet (Schwarz and Spallholz, 1979) and goats fed $<20\ \mu g\ kg^{-1}$ diet (Anke et al., 1978) improved growth. These findings have not been confirmed in a peer reviewed publication. Other isolated reports that suggest cadmium can have positive bioactivity include the finding a cadmium-containing carbonic anhydrase in a marine diatom (Lane et al., 2005). Cadmium can substitute for zinc to maintain optimal growth in marine organisms, which may be related to cadmium having nutrient-like behavior in oceans. If cadmium has any beneficial action in humans and higher animals, it will have to occur at extremely low amounts. Cadmium has a long half-life in the body. Thus, it does not take much of an elevated intake to result in an accumulation that leads to pathology in some organs, especially the kidney. In addition to renal dysfunction, high cadmium intakes have been associated with hypertension, some types of cancer and osteomalacia. The European Food Safety Authority has concluded that a tolerable weekly intake for cadmium is $2.5\ \mu g\ kg^{-1}$ body weight, or about $25\ \mu g\ day^{-1}$ for a 70 kg person. The typical daily intake is about $20\ \mu g\ day^{-1}$, which would be a safe intake and would easily provide any beneficial effects of cadmium.

Fluorine

There are no reported unequivocal or specific signs of fluoride deficiency that provide evidence that it is an essential nutrient. Except for unconfirmed studies with goats showing that fluoride deficiency decreased life expectancy and caused pathological changes in the kidney and endocrine organs (Anke et al., 1991; Avtsyn et al., 1993), evidence that has been used to suggest a need for fluoride have come from studies using pharmacological intakes. Pharmacological or supra nutritional doses of fluoride have been shown to prevent tooth caries and bone loss in humans (Jenkins, 1990); improve fertility, hematopoiesis and growth in iron-deficient mice and rats (Messer et al., 1974; Wegner et al., 1976); and prevent phosphorus-induced nephro-calcinosis in rats (Fransbergen et al., 1991).

Although fluoride should not be considered an essential element, it still is accepted as an important beneficial bioactive element for humans. On the basis that it can reduce dental caries without adverse effects, the United States and Canada set AIs for fluoride (Institute of Medicine, Food and Nutrition Board, 1997). These are per day: infants 0–6 months, 0.01 mg and 6–12 months, 0.5 mg; children 1–3 years, 0.7 mg and 4–8 years, 1 mg; adolescents 9–13 years, 2 mg and 14–18 years, 3 mg; and adult women, 3 mg and in men 4 mg. These are supra nutritional intakes that do not generally result in undesirable mottling of teeth. The prevention of dental caries should not be considered an essential function of fluoride because they are not preventing a nutritional deficiency.

Chronic fluoride toxicity through excessive intake mainly through drinking water and industrial exposure has been reported throughout the world. Chronic intakes through water and food providing not much more than 2.0 mg day^{-1} can result in barely discernible dental fluorosis or mottled enamel; much higher amounts can result in stained and pitted enamel. Crippling skeletal fluorosis apparently occurs in individuals who ingest 10–25 mg day^{-1} for 7–20 years. The UL (mg day^{-1}) is for ages 0–6 months, 0.7; 7–12 months, 0.9; 1–3 years, 1.3, 4–8 years, 2.2; and >8 years, 10 mg (Institute of Medicine, Food and Nutrition Board, 1997).

Germanium

Most evidence indicating that germanium has beneficial biological activity has been obtained with supra nutritional or pharmacological amounts. Compared to diets containing about 1.0 mg kg^{-1}, a germanium supplement of 30 mg kg^{-1} altered bone and liver mineral composition and decreased tibial DNA in rats (Seaborn and Nielsen, 1994), and a germanium dioxide supplement of 10 mg kg^{-1} stimulated growth in rats (Venugopal and Luckey, 1978) and chicks (Li et al., 1993). These findings suggest that the beneficial activity of germanium occurs not through an essential function but through substituting or enhancing the function of another essential nutrient or through modification of them at the intestinal level. Pharmacological amounts of some organic germanium compounds have been promoted in the dietary supplement industry as having *anti*-tumor activity in addition to having therapeutic effects on rheumatoid arthritis, osteoarthritis, and osteoporosis (Cho et al., 2020). These effects have been attributed to the

organic supplements having *anti*-inflammatory and immune stimulating effects. Although the published reports indicate that inorganic and some organic germanium compounds may have beneficial activity in higher animals and humans, they do not provide evidence that germanium is an essential nutrient.

Although germanium is thought to have a low order of toxicity because of its diffusible state and rapid elimination from the body, inappropriate intakes can cause toxicity. Rats fed 5 μg mL^{-1} of water for 36 months exhibited impaired kidney and liver function. Although organic forms of germanium touted in the supplement industry are less toxic than inorganic forms, some individuals consuming high amounts of organic germanium supplements contaminated with inorganic germanium have died from kidney failure (Tao and Bolger, 1997). Until more knowledge is obtained about the intakes of various forms of germanium that are toxic, they probably should not routinely greatly exceed those found in a typical diet. The typical daily dietary intake of germanium is 0.4–1.5 mg.

Lead

One research group has suggested that lead may be essential in ultratrace amounts because 30 ng g^{-1} versus 30 μg g^{-1} in the diet resulted in apparent deficiency signs in pigs (Kirchgessner et al., 1991) and rats (Reichlmayr-Lais and Kirchgessner, 1991). These signs were depressed growth; anemia; elevated serum cholesterol, phospholipids, and bile acids; disturbed iron metabolism; decreased liver glucose triacylglycerols, LDL-cholesterol and phospholipids; increased liver cholesterol; and altered blood and liver enzymes. Most of these findings have not been confirmed by another laboratory. Another group did report that lead alleviated iron deficiency in young rats (Uthus and Nielsen, 1988). Thus, many of the deficiency signs reported for lead might have been the result of a beneficial effect of on a disordered iron metabolism or deficiency, not through an unfilled essential function.

Although lead may be beneficial at low intakes under certain conditions, the concern about lead toxicity inhibits it from being recommended at any intake. Lead is considered a significant environmental pollutant because the past use of lead in paints, plumbing, food can solder, and fuel additives have resulted in it still being present in the environment. When blood concentrations reach 10 μg dL^{-1}, lead adversely affects bone and mental development and blood pressure (Agency for Toxic Substances and Disease Registry, 2021). Anemia, nephrotoxicity, and more overt neurological impairments occur when concentrations exceed 30 μg dL^{-1}.

Lithium

In the early 1990s, two reports of induced lithium deficiency appeared. Rats deprived of lithium exhibited depressed fertility, birthweight, litter size, and weaning weight (Pickett and O'Dell, 1992). Goats exhibited depressed fertility, birthweight, and life span in addition to alteration in activity of several liver and blood enzymes (Anke et al., 1990). These findings are not sufficient to ascribe lithium as essential, but they do indicate that lithium could have beneficial biological action in nutritional amounts. In vitro findings supporting this conclusion are lithium stimulating growth of some cultured cells (Rybak and Stockdale, 1981), and having insulin mimetic action (Rossetti et al., 1990). Lithium is best known as a pharmacological agent in treating manic-depressive psychosis. Its ability to affect mental function might be the basis for the finding that the incidence of violent crimes is lower in areas with high-lithium drinking water (Schrauzer and Shrestha, 1990; Schrauzer et al., 1992).

The amount of lithium needed to be beneficial in animal experiments suggest that less than 25 μg day^{-1} would provide these benefits to humans, which is much less than the usual dietary intake of 200–600 μg day^{-1}. Lithium is not a particularly toxic element, but the therapeutic dose (~500 mg day^{-1}) needed to treat psychiatric disorders is close to concentrations to that inducing mild toxicity signs. These signs include gastrointestinal disturbances, muscular weakness, tremor, drowsiness, and a dazed feeling. Severe toxicity results in coma, muscle tremor, convulsions, and death.

Molybdenum

The evidence for the essentiality of molybdenum is substantial and conclusive. Molybdenum functions as a cofactor in enzymes aldehyde oxidase, sulfite oxidase, xanthine oxidase, and mitochondrial amidoxime-reducing component (Belaidi and Schwarz, 2016). However, only an isolated human case of nutritional molybdenum deficiency has been reported (Abumrad et al., 1981). A patient with Chron's disease receiving long-term parenteral nutrition exhibited high methionine and low uric acid in blood, high oxypurines and low uric acid in urine, and very urinary sulfate excretion. The patient had mental disturbances that progressed to coma. Intravenous administration of ammonium molybdate alleviated the mental disturbances and reversed the sulfur handling and uric acid defects. Other than this case, there have been no reports of a molybdenum deficiency induced through a low dietary intake. However, rare mutations in genes causing encoding errors involved in molybdenum cofactor synthesis can result in a severe deficiency syndrome (Belaidi and Schwarz, 2016). These genetic mutations result in the loss of all molybdenum cofactor enzymes. The deficiency usually diagnosed shortly after birth results in a marked loss of white matter in the brain. Biochemical changes include elevated urinary sulfite and S-sulfocysteine, hypouricemia, elevated plasma S-sulfonated transthyretin and deficient

molybdoenzyme activity in fibroblasts. Most patients die in childhood; some survive only a few days. Mild cases have been reported, perhaps because of a low residual activity of a mutant protein.

Although a dietary molybdenum deficiency in a healthy individual has not been reported, DRIs have been set (Institute of Medicine, Food and Nutrition Board, 2001) for this element based on balance studies performed in 1995. These studies found that adults achieve molybdenum balance on an intake of 25 μg day^{-1} (Turnland et al., 1995). Because some foods might inhibit the bioavailability of molybdenum, a bioavailability of 75% was used to set an EAR of 34 μg day^{-1} for adults. The RDA was set as the EAR+30% to get 45 μg day^{-1} for adults. DRIs for children were extrapolated from the adult values using metabolic weight ($kg^{0.75}$). The RDAs set in μg day^{-1} were for infants 0–6 months, 2, and 7–12 months, 3; children 1–3 years, 17, and 4–8 years, 22; for adolescents 9–13 years, 34, and 14–8 years, 34 (females) and 43 (males).

Pulses, especially various forms of beans, are rich sources of molybdenum. Whole grains, cereals, and organ meats also are good sources. Moderate sources include dark leafy and Brassica vegetables and milk and milk products. Daily intakes have been found to usually exceed the RDAs for molybdenum (Novotny and Peterson, 2018). In the United States, mean intakes of 76 μg day^{-1} for women and 109 μg day^{-1} for men have been reported. Higher mean intakes in other countries have been found, including 225 μg day^{-1} In Japan and 275 μg day^{-1} in France.

Molybdenum is a relatively low toxicity element, which results in a lack of toxicity data for humans. Most relevant toxicity data have come from nonruminant animal studies that used milligram doses instead of nutritional microgram doses of molybdenum. Based largely on animal findings, an UL of 2000 μg (2.0 mg) day^{-1} was set for adults in the United States (Institute of Medicine Food and Nutrition Board, 2001). The UL (in μg day^{-1}) set for children was: 1–3 years, 300, and 4–8 years, 600; for adolescents 9–13 years, 1100, and 14–18 years, 1700.

Nickel

Nickel is essential for plants and some bacteria. In these lower forms of life, nickel is an essential component of eight enzymes; seven of these enzymes are involved in the use or production of the gases ammonia, hydrogen, carbon monoxide, carbon dioxide, methane and oxygen (Ragsdale, 2009). The other enzyme converts methylglyoxal to lactate (Fabiano et al., 2015). A mammalian nickel-dependent enzyme has not been identified. Thus, nickel has not been established as an essential nutrient for higher animals and humans. However, nickel deprivation studies with several experimental animal models have shown that nickel has beneficial biological activity, and possibly an essential function (Nielsen, 2006). Reported effects of deprivation include impaired reproduction (decreased conception rate and sperm production and motility), impaired bone health (decreased strength and altered composition), altered carbohydrate and lipid metabolism (increased plasma lipids and decreased serum glucose), impaired iron status and utilization, altered thyroid metabolism, exacerbated vitamin B_{12} deficiency with high blood homocysteine, exacerbated renal damage and high blood pressure induced by a high-salt diet, and impaired special senses of vision, olfaction, and taste.

Several suggestions have been given for the beneficial biological activity of nickel (Nielsen, 2021). These include affecting the formation or action of a functional gaseous molecule, performing actions normally done by iron, influencing intracellular calcium content and its signaling, and affecting the intestinal microbiome. Support for these suggestions include the following (Nielsen, 2020). Nickel has a stabilizing effect on hypoxia-inducible factor-1α protein and activates hypoxia-inducible gene expression, which affects glucose metabolism and osteogenesis. Nickel also potently induces the activity of heme oxygenase that produces carbon monoxide, an activator of guanylate cyclase that produces cyclic guanylate monophosphate (cGMP). The cGMP signal transduction system has a role in vision, taste, smell, blood pressure control, kidney function, and sperm motility, all which are affected by nickel deprivation. The essential role of nickel in some microorganisms that may exist in the intestinal microbiome possibly results in the formation or function of substances that are beneficial to higher animals and humans.

Neither RDAs or AIs have been established for nickel. Based on animal findings, a beneficial intake for humans could be <100 μg day^{-1} and has been suggested to be as low as 25–35 μg day^{-1} (Nielsen, 2021). Most individuals achieve this intake because typical daily dietary intakes for nickel are 70–400 μg day^{-1}. Foods of plant origin are generally high in nickel while foods of animal origin are low. Rich sources of nickel include chocolate, nuts, dried beans, and peas, and grains. Diets high in these foods could supply >900 μg day^{-1}.

Life-threatening toxicity of nickel through oral intake is unlikely because intestinal absorption with food is low and absorbed nickel is rapidly excreted in the urine. Nickel has little tendency to accumulate in tissues of animals and humans. Based on extrapolation form animal studies, the ULs set for nickel as soluble nickel salts (mg day^{-1}) in the United States are for children aged 1–3 years, 0.2; 4–8 years, 0.3; and 9–13 years, 0.6; and for adults, 1.0 (Institute of Medicine, Food and Nutrition Board, 2001).

Rubidium

In the 1990s, it was reported that rubidium supplementation alleviated depressed food intake, milk production, growth, and life expectancy in goats fed a diet low in rubidium (Anke et al., 1993a). Also, it was reported that rubidium supplementation of rats fed a low rubidium diet exhibited tissue mineral concentrations in a manner suggesting an effect on potassium, phosphorus, calcium, and

magnesium (Yokoi et al., 1996). Neither of these reports have been confirmed. Rubidium can substitute for some roles performed by potassium. This might have been the basis for the limited reports about the beneficial effects of rubidium, which do not support it being essential. Based on animal experiments, any beneficial intake of rubidium would be less than 200 μg day^{-1}, which would be easily met by the typical daily dietary intake of 1–5 mg day^{-1}. Rubidium is a relatively non-toxic element and thus is not of toxicological concern from the nutritional point of view. Increasing the normal body content of rubidium (0.36 g) by 50–100 times does not have any noticeable adverse effect.

Silicon

Silicon is nutritionally essential for some lower forms of life, including diatoms, radiolarians, some sponges, and some plants (Carlisle, 1997). Silicon has not been established as essential for humans because it lacks a clearly defined biochemical function, and a deficient intake has not shown to prevent the life cycle. However, cell culture, animal, and human studies indicate that silicon in nutritional amounts can be beneficial, especially for bone and connective tissue function. Beneficial effects of silicon likely occur because it easily forms stable complexes with polyols that have at least four hydroxyl groups (Kinrade et al., 1999). Such polyols include hexosamines, and ascorbate used to form glycosaminoglycans, mucopolysaccharides, and collagen, which are involved in connective tissue formation and stabilization and bone formation. The complexing of silicon with hydroxylated molecules in the bone growth region apparently attracts calcium to form apatite. After calcification, silicon is lost, which explains the finding that silicon is much higher in under-mineralized than mineralized bone (Jugdaohsingh et al., 2015). Silicon in nutritional amounts may also be beneficial by inhibiting detrimental effects of other minerals such as aluminum, which has been associated with impaired cognitive function and bone health. This apparently occurs through the formation of non-metabolizable aluminosilicates.

Human studies indicating that silicon in nutritional amounts can have beneficial effects include the finding that hip bone density was higher in men and premenopausal women with intakes of >40 mg day^{-1} than those with intakes <14 mg day^{-1} (Jugdaohsingh et al., 2004). Also, energy-adjusted silicon intakes have been found to be negatively associated with urinary markers of bone resorption and positively associated with a marker of bone formation in a study where the lowest quartile of silicon intake was 16 mg day^{-1} and the highest quartile was 31.5 mg day^{-1} (Macdonald et al., 2012). Support for silicon having a beneficial effect on connective tissue is the finding that a silicon supplement of 10 mg day^{-1} as choline-stabilized orthosilicic acid for 20 weeks improved photodamaged skin and decreased hair and nail brittleness (Spector et al., 2008).

Animal and human data have been judged too limited to establish DRIs for silicon. In the Framingham Offspring Cohort study (Jugdaohsingh et al., 2004), the highest hip bone mineral density occurred in premenopausal women and men, respectively, with quintiles ranging from 30.2 to 63.2 mg day^{-1} and 34.4–118.0 mg day^{-1} compared to the lowest quintile intakes ranging from 7.1 to 16.7 mg day^{-1} and 18.8 mg day^{-1}. These findings, in addition to extrapolation from animal data and questionable balance data suggest that a beneficial intake of silicon may be near 25 mg day^{-1} (Nielsen, 2006). The highest concentrations of silicon are found in cereals and cereal products, especially those less refined. Substantial amounts are also found in beans, spinach, dried fruits, bananas, and red lentils. The silicon in barley and hops is solubilized during the brewing process, which makes beer a rich source of silicon. Silicon is added to processed foods for anticaking, thickening, and stabilizing purposes. These forms of silicon are not readily absorbed by the gastrointestinal tract and should not be considered providing beneficial intakes of silicon. There are no reports of silicon toxicity through excessive dietary intake for people with normal renal function.

Strontium

There is no conclusive evidence that strontium is essential for higher animals and humans. Strontium deprivation has been reported to depress growth, impair calcification of bones and teeth, and increase dental caries in rats and guinea pigs (Rygh, 1949), but these findings have not been confirmed. The findings might be explained by strontium, like fluoride, in supplemental or pharmacological amounts having beneficial effects on teeth and bone (Marie et al., 2001). Strontium ranelate in pharmacological amounts has been found to promote bone growth, prevent fractures in both spine and hip, and increase bone mineral density in the spine and femur of humans (O'Donnell et al., 2006). It is thought that the beneficial effects of strontium result from replacing or substituting for calcium in some of its biological roles.

Unfortunately, strontium ranelate in therapeutic doses increases the risk for venous thromboembolism, pulmonary embolism, and myocardial infarction, which has restricted its use (O'Donnell et al., 2006). There is no evidence that strontium provided by other forms in nutritional amounts by dietary means evoke such toxicological effects.

DRIs have not been established for strontium. Because beneficial effects of strontium have been found using non-nutritional amounts, it is difficult to give a dietary intake recommendation for this element. However, because strontium has been associated with bone and teeth health, assuring the consumption of foods relatively rich in strontium, such as whole grains and unpeeled fruits and vegetables would be prudent action. The typical daily dietary intake of strontium of 1.5–3 mg day^{-1} probably provides the possible benefits of nutritional amounts of strontium.

Tin

In 1970, it was reported that tin deprivation depressed growth in rats (Schwarz et al., 1970); the diet used in this study was subsequently found to be inadequate in other nutrients including riboflavin. In 1990, another report appeared in which rats fed 17 ng g^{-1} tin versus 1.99 μg g^{-1} tin in the diet exhibited depressed growth, response to sound, and feed efficiency (Yokoi et al., 1990). The tin deprivation of the rats also altered copper, iron, manganese and zinc concentrations in the heart, muscle, and tibia. These findings have not been subsequently confirmed. Thus, the limited data available do not support tin essentiality and make it uncertain whether tin can have beneficial biological activity in nutritional amounts in humans. In addition, it is difficult to indicate a beneficial intake. However, any such intake is likely to be easily met by the usual dietary intake of 1–40 mg day^{-1} by humans. Tin is a relatively nontoxic element. However, because 50 mg day^{-1} of tin was found to detrimental to zinc and copper metabolism, routine intakes near this amount probably should be avoided.

Vanadium

Vanadium is essential for some lower forms of life (algae, seaweeds, a lichen, and a fungus) where it is a component of the enzymes bromoperoxidase, iodoperoxidase, and chloroperoxidase (Crans et al., 2004). Vanadium reacts with H_2O_2 to form a dioxygen species, which is an intermediate in the formation of the carbon-halogen bond. Vanadium has not been found to meet the criteria for essentiality for higher animals or humans but does have beneficial effects in supra nutritional or pharmacological and perhaps nutritional intakes. Animals deprived of vanadium (2–10 ng kg^{-1} diet) responded to vanadium supplements of 1–2 μg kg^{-1} diet. The deprivation signs in goats were swollen joints, skeletal deformations, and decreased life span (Anke et al., 1989) and in rats were altered thyroid hormone metabolism, impaired reproduction an altered bone morphology (Uthus and Nielsen, 1990; Nielsen, 1998). These findings have not been subsequently confirmed. The supra nutritional or pharmacological beneficial action of vanadium that has received the most attention is its ability to mimic insulin (Marzban and McNeill, 2003). The inhibition of protein tyrosine phosphatase enzymes may be the basis for many of the reported beneficial responses to nutritional and supra nutritional intakes of vanadium (Hulley and Davison, 2003). However, vanadium affects other phosphorous associated enzymes including glucose-6-phosphate dehydrogenase, adenosine triphosphatase, phosphodiesterases, and phosphomutases in metabolic diseases (Treviño et al., 2019).

No studies have defined an intake at which nutritional intakes would have a beneficial effect in humans. As a result, neither an RDA nor AI has been established for vanadium. Based on animal findings, a daily intake of 10 μg probably would provide any nutritional benefits of vanadium. The typical daily dietary intake of vanadium is 12–30 μg day^{-1}. Good sources of vanadium include grains, mushrooms, parsley, and shellfish. An UL has been set for vanadium in the United States (Institute of Medicine, Food and Nutrition Board, 2001). Based on evidence for renal toxicity in animals, the UL for adults was set at 1.8 mg day^{-1}; no UL was set for other age groups.

Summary

The current knowledge about the essentiality, biochemical function, beneficial biological activity, possible deprivation signs, beneficial or required intakes, toxicity and nutritional importance of 16 ultratrace elements are presented. Emerging evidence indicates that several of these elements, including boron, molybdenum, nickel, silicon, vanadium, have beneficial bioactivity in nutritional amounts. This bioactivity may have ameliorative or preventive effects on chronic diseases such as cardiovascular disease, cancer, diabetes, and osteoporosis. Some elements, including fluoride and lithium in supra nutritional amounts have beneficial bioactivity. Other ultratrace elements, including bromine, germanium, molybdenum, rubidium, strontium, and tin might have beneficial activity in nutritional amounts that are normally provided by the typical diet. Although there is evidence for beneficial activity in ultratrace amounts for elements aluminum, arsenic, cadmium, and lead, they are of more concern about their toxicity. Thus, based on usual or typical intakes and strength of evidence for beneficial activity, only boron, molybdenum, nickel, silicon, and vanadium are ultratrace elements that presently should receive attention about achieving intakes needed for optimal beneficial effects. Because of reports that it is not uncommon for individuals not achieving beneficial intakes of boron (Nielsen and Meacham, 2011) and silicon (Jugdaohsingh et al., 2004), most attention should be directed toward boron and silicon. These individuals most likely have diets low in fruits, vegetables, and whole grains. If dietary modification cannot achieve beneficial intakes, these amounts of boron and silicon could be achieved by using supplements. Individuals with diets containing recommended amounts of pulses, nuts, and grains should have no concern about achieving beneficial intakes of molybdenum, nickel, and vanadium. For those with difficulty in consuming those types of foods, multi-mineral and vitamin supplements are available that provide molybdenum in recommended amounts and small amounts of nickel and vanadium.

References

Abumrad, M.M., Schneider, A.J., Steel, D., Rogers, L.S., 1981. Amino acid intolerance during prolonged total parenteral nutrition reversed by molybdate therapy. Am. J. Clin. Nutr. 34, 2551–2559.

Agency for Toxic Substances and Disease Registry, 2021. Toxicological Profile for Lead. Available at. www.atsdr.cdc.gov. (Accessed 24 June 2021).

Angelow, L., Anke, M., Groppel, B., Glei, M., Müller, M., 1993. Aluminum: an essential element for goats. In: Anke, M., Meissner, D., Mills, C.F. (Eds.), Trace Elements in Man and Animals. TEMA 8. Verlag Media Touristik, Gersdorf, pp. 699–704.

Anke, M., Groppel, B., Krause, U., 1991. Fluorine deficiency in goats. In: Momcilovic, B. (Ed.), Trace Elements in Man and Animals 7. IMI, Zagreb, pp. 26:28–26:29.

Anke, M., Hennig, A., Groppel, B., Partschefeld, M., Grün, M., 1978. The biochemical role of cadmium. In: Kirchgessner, M. (Ed.), Trace Elements in Man and Animals-3. Tech Univ Munchen, Freising-Weihenstephen, pp. 540–548.

Anke, M., Groppel, B., Gruhn, K., Langer, M., Arnhold, W., 1989. The essentiality of vanadium for animals. In: Anke, M., Baumann, W., Braunlich, H., et al. (Eds.), 6th International Trace Element Symposium, 1989, vol. 1. Friedrich-Schiller-Universit, Jena, pp. 17–27.

Anke, M., Arnhold, W., Groppel, B., Krause, U., 1990. The biological importance of lithium. In: Schrauzer, G.N., Klippel, K.F. (Eds.), Lithium in Biology and Medicine. VCH Publishers, Weinheim, pp. 148–167.

Anke, M., Angelow, L., Schmidt, A., Gürtler, H., 1993a. Rubidium: an essential element for animal and man? In: Anke, M., Meissner, D., Mills, C.F. (Eds.), Trace Elements in Man and Animals, TEMA 8. Verlag Media Touristik, Gersdorf, pp. 719–723.

Anke, M., Groppel, B., Angelow, L., Dorn, W., Drusch, S., 1993b. Bromine: an essential element for goats. In: Anke, M., Meissner, D., Mills, C.F. (Eds.), Trace Elements in Man and Animals. TEMA 8. Verlag Media Touristik, Gersdorf, pp. 737–738.

Anke, M., 2005. Recent progress in exploring the essentiality of the non-metallic ultratrace element arsenic to the nutrition of animals and man. Biomed. Res. Trace Elem. 16, 188–197.

Avtsyn, A.P., Anke, M., Zhavoronkov, A.A., et al., 1993. Pathological anatomy of the experimentally-induced fluorine deficiency in she-goats. In: Anke, M., Meissner, D., Mills, C.F. (Eds.), Trace Elements in Man and Animals. TEMA 8. Verlag Media Touristik, Gersdorf, pp. 745–746.

Belaidi, A.A., Schwarz, G., 2016. Molybdenum cofactor in humans: health, disease, and treatment. In: Collins, J.E. (Ed.), Molecular, Genetic, and Nutritional Aspects of Major and Trace Elements. Academic Press, San Diego, pp. 399–410.

Bolaños, L., Lukaszewski, K., Bonilla, I., Blevins, D., 2004. Why boron? Plant Physiol. Biochem. 42, 907–912.

Bosshardt, D.K., Huff, J.W., Barnes, R.H., 1956. Effect of bromine on chick growth. PSEBM 92, 219–221.

Carlisle, E.M., Curran, M.J., 1993. Aluminum: an essential element for the chick. In: Anke, M., Meissner, D., Mills, C.F. (Eds.), Trace Elements in Man and Animals. TEMA 8. Verlag Media Touristik, Gersdorf, pp. 695–698.

Carlisle, E.M., 1997. Silicon. In: O'Dell, B.L., Sunde, R.A. (Eds.), Handbook of Nutritionally Essential Minerals. Marcel Dekker, New York, pp. 603–618.

Chappell, W.R., Abernathy, C.O., Cothern, C.R. (Eds.), 1994. Arsenic Exposure and Health. Science and Technology, Northwood.

Cho, J.M., Chae, J., Jeong, S.R., et al., 2020. Immune activation of bio-germanium in a randomized, double-blind, placebo-controlled clinical trial with 130 human subjects: therapeutic opportunities from new insights. PLoS One 15, e0240358.

Crans, D.C., Smee, J.J., Gaidamauskas, E., Yang, L., 2004. The chemical and biochemistry of vanadium and the biological activities exerted by vanadium compounds. Chem. Rev. 104, 849–902.

Davis, C.D., Uthus, E.O., Finley, J.W., 2000. Dietary selenium and arsenic affect DNA methylation in vitro in Caco- cells and in vivo in rat liver and colon. J. Nutr. 130, 2903–2909.

Eckhert, C.D., Rowe, R.I., 1999. Embryonic dysplasia and adult retinal dystrophy in boron-deficient zebrafish. J. Trace Elem. Exp. Med. 12, 213–219.

European Food Safety Authority (EFSA), Arcella, A., Cascio, C., Ruiz, J.A.G., 2021. Chronic dietary exposure to inorganic arsenic. EFSA J. 19, e06380. https://doi.org/10.2903/j.efsa.2021.6380.

European Food Safety Authority, 2013. Scientific opinion on the re-evaluation of boric acid (E284) and sodium tetraborate (E285) as food additives. Eur. Food Saf. J. 11, 3407.

Exley, C., Mold, M.J., 2019. Aluminum in human brain tissue: how much is too much? J. Biol. Inorg. Chem. 24, 1279–1282.

Fabiano, C.C., Tezotto, T., Favarin, J.L., Polacco, J.C., Mazzafera, P., 2015. Essentiality of nickel in plants: a role in plant stresses. Front. Plant Sci. 6, 754.

Fort, D.J., Rogers, R.L., McLaughlin, D.W., Sellers, C.M., Schlekat, C.L., 2002. Impact of boron deficiency on *Xenopus laevis*: a summary of biological effects and potential biochemical roles. Biol. Trace Elem. Res. 90, 117–142.

Fransbergen, A.J., Lemmens, A.G., Beynen, A.C., 1991. Dietary fluoride, unlike bromide and iodide, counteracts phosphorus-induced nephrocalcinosis in female rats. Biol. Trace Elem. Res. 31, 71–78.

Gribble, G.W., 1999. The diversity of naturally occurring organobromine compounds. Chem. Soc. Rev. 28, 335–346.

Hakki, S.S., Bozkurt, B.S., Hakki, E.E., Nielsen, F.H., 2021. Boron as boric acid induces mRNA expression of the differentiation factor tuftelin in pre-osteoblastic MC3t3-E1 cells. Biol. Trace Elem. Res. 199, 1534–1543.

Huff, J.W., Bosshardt, D.K., Miller, O.P., Barnes, R.H., 1956. A nutritional requirement for bromine. PSEBM 92, 216–219.

Hulley, P., Davison, A., 2003. Regulation of tyrosine phosphorylation cascades by phosphatases: what the actions of vanadium teach us. J. Trace Elem. Exp. Med. 16, 281–290.

Hunt, C.D., 2002. Boron-binding biomolecules: a key to understanding the beneficial physiologic effects of dietary boron from prokaryotes to humans. In: Goldbach, H.E., Rekasem, B., Wimmer, M.A., et al. (Eds.), Boron in Plant and Animal Nutrition. Kluwer Academic/Plenum Publishers, New York, pp. 21–36.

Institute of Medicine, Food and Nutrition Board, 1997. Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride. National Academy Press, Washington D.C.

Institute of Medicine, Food and Nutrition Board, 2001. Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc. National Academy Press, Washington D.C.

Jenkins, G.N., 1990. The metabolism and effects of fluoride. In: Priest, N.D., Van De Vyver, F.L. (Eds.), Trace Metals and Fluoride in Bones and Teeth. CRC Press, Boca Raton, pp. 141–173.

Jugdaohsingh, R., Tucker, K.L., Qiao, N., et al., 2004. Dietary silicon intake is positively associated with bone mineral density in men and premenopausal women of the Framingham offspring cohort. J. Bone Miner. Res. 19, 297–307.

Jugdaohsingh, R., Pedro, L.D., Watson, A., Powell, J.J., 2015. Silicon and boron differ in their localization and loading in bone. Bone Rep. 1, 9–15.

Kayajanian, G., 2003. Arsenic. Cancer, and thoughtless policy. Ecotoxicol. Environ. Saf. 55, 139–142.

Kinrade, S., Del Nin, J.W., Schach, A.S., et al., 1999. Stable five- and six-coordinated silicate anions in aqueous solution. Science 285, 1542–1545.

Kirchgessner, M., Plass, D.L., Reichlmayr-Lais, A.M., 1991. Lead deficiency in swine. In: Momcilovic, B. (Ed.), Trace Elements in Man and Animals 7. IMI, Zagreb, pp. 11:20–11:21.

Klein, G.L., 2019. Aluminum toxicity to bone: a multi system effect? Osteoporos Sarcopenia 5, 2–5.

Kobylewski, S.E., Henderson, K.A., Yamada, K.E., Eckhert, C.D., 2016. Activation of EIF2α/ATF4 and ATF6 pathways in DU-145 cell by boric acid at the concentration reported in men at the US mean boron intake. Biol. Trace Elem. Res. 176, 278–293.

Lane, T.W., Saito, M.A., George, G.N., et al., 2005. A cadmium enzyme from a marine diatom. Nature 435, 42.

Leach Jr., R.M., 1963. Studies on chloride deficiency in chicks. J. Nutr. 81, 193–199.
Li, J.F., Kirchgessner, M., Steinruck, U., 1993. Growth promoting effect and toxicity of germanium in chickens. Archiv fur Geflugelkunde 57, 205–210.
Liu, G., Song, Y., Li, C., et al., 2021. Arsenic compounds: the wide application and mechanisms applied in acute promyelocytic leukemia and carcinogenic toxicology. Eur. J. Med. Chem. 221, 113519. https://doi.org/10.1016/j.ejmech.2021.113519. Epub 2021 May 4.
Macdonald, H.M., Hardcastle, A.C., Jugdaohsingh, R., et al., 2012. Dietary silicon interacts with oestrogen to influence bone health: evidence from the Aberdeen Prospective Osteoporosis Screening Study. Bone 50, 681–687.
Marie, P.J., Ammann, P., Boivin, G., Rey, C., 2001. Mechanism of action and therapeutic potential of strontium in bone. Calcif. Tissue Int. 69, 121–129.
Marzban, L., McNeill, J.H., 2003. Insulin-like actions of vanadium: potential as a therapeutic agent. J. Trace Elem. Exp. Med. 16, 253–267.
Mayer, D.R., Kosmus, W., Pogglitsch, H., Mayer, D., Beyer, W., 1993. Essential trace elements in humans. Serum arsenic concentrations in hemodialysis patients in comparison to healthy controls. Biol. Trace Elem. Res. 37, 27–38.
McCall, A.S., Cummings, C.F., Bhave, G., et al., 2014. Bromine is an essential element for assembly of collagen IV scaffolds in tissue development and architecture. Cell 157, 1380–1392.
Messer, H.H., Armstrong, W.D., Singer, L., 1974. Essentiality and function of fluoride. In: Hoekstra, W.G., Suttie, J.W., Ganther, H.E., Mertz, W. (Eds.), Trace Elements Metabolism in Animals-2. University Park Press, Baltimore, pp. 425–437.
Nielsen, F.H., Eckhert, C.D., 2020. Boron. Adv. Nutr. 11, 461–462.
Nielsen, F.H., Meacham, S.L., 2011. Growing evidence for human health benefits of boron. J. Evid. Based Complement. Altern. Med. 16, 169–180.
Nielsen, F.H., 1996. Other trace elements. In: Ziegler, E.E., Filer Jr., L.J. (Eds.), Present Knowledge in Nutrition, seventh ed. ILSI Press, Washington D.C, pp. 353–377.
Nielsen, F.H., 1998. The nutritional essentiality and physiological metabolism of vanadium in higher animals. In: Tracey, A.S., Crans, D.C. (Eds.), Vanadium Compounds. Chemistry, Biochemistry, and Therapeutic Applications. ACS Symposium Series, vol. 711. American Chemical Society, Washington D.C, pp. 297–307.
Nielsen, F.H., 2006. Boron, manganese, molybdenum, and other trace elements. In: Bowman, B.A., Russell, R.M. (Eds.), Present Knowledge in Nutrition, ninth ed. International Life Sciences Institute, Washington D.C., pp. 506–526
Nielsen, F.H., 2012. Manganese, molybdenum, boron, chromium, and other trace elements. In: Erdman Jr., J.W., MacDonald, I., Zeisel, S. (Eds.), Present Knowledge in Nutrition, tenth ed. Wiley-Blackwell, Oxford, United Kingdom, pp. 586–607.
Nielsen, F.H., 2014. Update on human health effects of boron. J. Trace Elem. Med. Biol. 28, 383–387.
Nielsen, F.H., 2020. Manganese, molybdenum, boron, silicon, and other trace elements. In: Marriott, B.P., Birt, D.F., Stallings, V.A., Yates, A.A. (Eds.), Present Knowledge in Nutrition, 11th ed. 1. Elsevier Academic Press, London, pp. 485–500.
Nielsen, F.H., 2021. Nickel. Adv. Nutr. 12, 281–282.
Novotny, J.A., Peterson, C.A., 2018. Molybdenum. Adv. Nutr. 9, 272–273.
O'Donnell, S., Cranney, A., Wells, G.A., Adachi, J.D., Reginster, J.Y., 2006. Strontium ranelate for preventing and treating postmenopausal osteoporosis. Cochrane Database Syst. Rev. (3), CE005326. https://doi.org/10.1002/14651858.CD005326,pub2.
Oe, P.L., Vis, R.D., Meijer, J.H., et al., 1981. Bromine deficiency and insomnia in patients on dialysis. In: McC Howell, J., Gawthorne, J.M., White, C.L. (Eds.), Trace Element Metabolism in Man and Animals, TEMA-4. Australian Academy of Science, Canberra, pp. 526–529.
Pickett, E.E., O'Dell, B.L., 1992. Evidence for dietary essentiality of lithium in the rat. Biol. Trace Elem. Res. 34, 299–319.
Ragsdale, S.W., 2009. Nickel-based enzyme systems. J. Biol. Chem. 284, 18571–18575.
ReichImayr-Lais, A.M., Kirchgessner, M., 1991. Lead—an essential trace element. In: Momcilovic, B. (Ed.), Trace Elements in Man and Animals 7. IMI, Zagreb, pp. 35:1–35:2.
Rogoveanu, O.C., Mogoşanu, G.D., Bejenaru, C., et al., 2015. Effects of calcium fructoborate on levels of C-reactive protein, total cholesterol, low-density lipoprotein, triglycerides, IL-1β, IL-6, and MCP-1: a double-blind, placebo-controlled clinical study. Biol. Trace Elem. Res. 163, 124–131.
Rossetti, L., Giaccari, A., Klein-Roggenhaar, E., Vogel, L.R., 1990. Insulinomimetic properties of trace elements and characterization of their *in vivo* mode of action. Diabetes 39, 1243–1250.
Rybak, S.M., Stockdale, F.E., 1981. Growth effects of lithium chloride in BALB/c 3T3 fibroblasts and Madin-Darby canine kidney epithelial cells. Exp. Cell Res. 136, 263–270.
Rygh, O., 1949. Recherches sur les oligo-éléments. I. De L'importance du strontium, du barium et du zinc. Bull. Soc. Chem. Biol. 31, 1052–1061.
Schrauzer, G.N., Shrestha, K.P., 1990. Lithium in drinking water and the incidence of crimes, suicides, and arrests related to drug additions. Biol. Trace Elem. Res. 25, 105–113.
Schrauzer, G.N., Shrestha, K.P., Flores-Arce, M.F., 1992. Lithium in scalp hair of adults, students, and violent criminals. Effects of supplementation and evidence for interactions of lithium with vitamin B-12 and with other trace elements. Biol. Trace Elem. Res. 34, 161–176.
Schwarz, K., Spallholz, J.E., 1979. The potential essentiality of cadmium. In: Bolck, F., Anke, M., Schneider, H.J. (Eds.), Cadmium-Symposium. Freidrich-Schiller Universitat, Jena, pp. 188–194.
Schwarz, K., Milne, D.B., Vinyard, E., 1970. Growth effects of tin compounds in rats maintained in a trace element-controlled environment. Biochem. Biophys. Res. Commun. 40, 22–29.
Seaborn, C.D., Nielsen, F.H., 1994. Effects of germanium and silicon on bone mineralization. Biol. Trace Elem. Res. 42, 11–18.
Spector, T.D., Calomme, M.R., Anderson, S.H., et al., 2008. Choline-stabilized orthosilicic acid supplementation as an adjunct to calcium/vitamin D3 stimulates markers of bone formation in osteopenic females: a randomized, placebo-controlled trial. BMC Muscoskel. Disord. 9, 85. https://doi.org/10.1186/1471-2474-9-85.
Stoeppler, M., 2004. Arsenic. In: Merian, E., Anke, M., Ihnat, M., Stoeppler, M. (Eds.), Elements and Their Compounds in the Environment, Occurrence, Analysis, and Biological Relevance, Nonmetals, Particular Aspects, second ed., vol. 3. Wiley-VCH, Weinheim, pp. 1321–1364.
Tao, S.H., Bolger, P.M., 1997. Hazard assessment of germanium supplements. Regul. Toxicol. Pharmacol. 25, 211–219.
Thomas, D.J., Li, J., Waters, S.B., et al., 2007. Arsenic (+3 oxidation state) methyltransferase and the methylation of arsenicals. Exp. Biol. Med. 232, 3–13.
Treviño, S., Diaz, A., Sánchez-Lara, et al., 2019. Vanadium in biological action: chemical, pharmacological aspects, and metabolic implications in diabetes mellitus. Biol. Trace Elem. Res. 188, 68–98.
Turnland, J.R., Keyes, W.R., Peiffer, G.L., 1995. Molybdenum absorption, excretion, and retention studied with stable isotopes in young men during depletion and repletion. Am. J. Clin. Nutr. 61, 1102–1109.
Uthus, E.O., Davis, C.D., 2005. Dietary arsenic affects dimethylhydrazine-induced aberrant crypt formation and hepatic global DNA methylation and DNA methyltransferase activity in rats. Biol. Trace Elem. Res. 103, 133–146.
Uthus, E.O., Nielsen, F.H., 1988. Effects in rats of iron on lead deprivation. Biol. Trace Elem. Res. 16, 155–163.
Uthus, E.O., Nielsen, F.H., 1990. Effect of vanadium, iodine and their interaction on growth, blood variables, liver trace elements and thyroid status indices in rats. Magnesium Trace Elem. 9, 219–226.
Uthus, E.O., 1992. Evidence for arsenic essentiality. Environ. Geochem. Health 14, 55–58.
Uthus, E.O., 1994. Estimation of safe and adequate daily intake for arsenic. In: Mertz, W., Abernathy, C.O., Olin, S.S. (Eds.), Risk Assessment of Essential Elements. ILSI Press, Washington DC, pp. 273–282.
Venugopal, B., Luckey, T.D., 1978. Metal toxicity in mammals. In: Chemical Toxicity of Metals and Metalloids, vol. 2. Plenum Press, New York, pp. 175–180.
Vilter, H., 1995. Vanadium-dependent haloperoxidases. In: Sigel, H., Sigel, A. (Eds.), Metal Ions in Biological Systems, Vanadium and Its Role in Life, vol. 31. Marcel Dekker, New York, pp. 325–362.

Wegner, M.E., Singer, L., Ophaug, R.H., Mafil, S.G., 1976. The interrelationship of fluoride and iron in anemia. PSEBM 153, 414–418.
World Health Organization, International Programme on Chemical Safety, 1998. Environmental Health Criteria 204 Boron. World Health Organization, Geneva.
Yamada, K.E., Eckhert, C.D., 2019. Boric acid activation of eIF2α and Nrf2 is PERK dependent: a mechanism that explains how boron prevents DNA damage and enhances antioxidant status. Biol. Trace Elem. Res. 188, 2–10.
Yokoi, K., Kimura, M., Itokawa, Y., 1990. Effect of dietary tin deficiency on growth and mineral status in rats. Biol. Trace Elem. Res. 24, 22–29.
Yokoi, K., Kimura, M., Itokawa, Y., 1996. Effect of low dietary rubidium on plasma biochemical parameters and mineral levels in rats. Biol. Trace Elem. Res. 51, 199–208.

Vitamin A: Deficiency and interventions

Amanda C. Palmer and Keith P. West, Jr., Department of International Health, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of K.P. West, Vitamin A: Deficiency and Interventions, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 323–332, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00274-9.

Introduction 460
- Vitamin A deficiency 461
- VA deficiency disorders (VADD) 461
- Xerophthalmia 461
  - Night blindness (XN) 461
  - Conjunctival xerosis (X1A) and Bitot's spots (X1B) 462
  - Corneal xerophthalmia (X2/X3) 462
- Other VADD: infection, anemia and poor growth 463
  - Infection 463

Epidemiology 463
- Person (high-risk groups) 464
- Place (geographic clustering) 464
- Time (periodicity) 465
- Causal agents (diet and infection) 465
  - Breastfeeding and diet 466
  - Infection 466
- Impact of interventions 467
  - On vitamin A status 467
  - On xerophthalmia 467
  - On mortality and morbidity 467
- Management 468
  - Treatment 468
  - Prevention 469

References 469

Key points

- Vitamin A (VA) deficiency results from inadequate intake of foods rich in preformed VA or its provitamin A (proVA) carotenoid precursors, often combined with inadequate breastfeeding and frequent infections.
- Approximately 30% of young children and ~15% of pregnant women in low-to-middle income countries are affected by VA deficiency.
- VA deficiency increases the risk of severe infection and consequent mortality and, when severe, can cause xerophthalmia—the eye disease of VA deficiency.
- VA deficiency can be prevented by several, complementary interventions: semi-annual supplementation with a large, oral dose of VA; fortification of staple food items; biofortification of staple crops with carotenoids; promotion of adequate and extended breastfeeding; and dietary diversification.

Introduction

Vitamin A (VA) is an essential nutrient in the human diet. Deficiency can develop when its intake is chronically low and, when sufficiently severe, manifest as eye disease. Night blindness, the mildest ocular stage of xerophthalmia, has been recognized since antiquity, as depicted in *bas-relief* on the wall of the Egyptian pyramid in Saqqura. Hippocrates in the 4th century BC recognized and treated the condition with animal liver (a major dietary source of VA). In the 18th and 19th centuries, eye signs of VA deficiency were linked to a poor diet, as cod liver oil emerged for treatment, but the active compound remained unknown. After years of animal research, "fat soluble factor A" was described in 1913, an ether-soluble compound in butter and egg yolk essential for

growth, general health and vision; thus, launching a century of discovery, leading to the vitamin's chemical synthesis, advances in learning its metabolic, morphological, epithelial, immune, hematopoietic and osteoid functions, and a global pursuit to assess and prevent VA deficiency and its public health consequences.

Vitamin A deficiency

Multiple methods are available to assess VA status (Tanumihardjo et al., 2016). Deficiency in populations is most often determined by plasma concentrations of retinol, or its carrier retinol binding protein (RBP4), or less often via indirect estimation of hepatic or total body nutrient stores, that likely precede functional consequences of deficiency. In uncomplicated hypovitaminosis A, plasma or serum retinol is homeostatically controlled until body (primarily liver) stores become lower, after which the plasma retinol concentration declines. Plasma retinol also decreases in response to an acute inflammatory response or chronic inflammation, in parallel with raised concentrations of positive acute phase proteins, increased tissue VA delivery, reduced hepatic mobilization via retinol-binding protein, and increased urinary loss. Plasma retinol gradually normalizes during recovery from infection in the presence of adequate hepatic stores. If VA is deficient, infection can further deplete stores. VA deficiency is generally diagnosed at a serum retinol concentration below 0.70 μmol L^{-1} (or 20 μg dL^{-1}), below which often 15% to >50% of concentrations lie among individuals sampled in a VA-deficient population, compared to <2% in well-nourished societies. A serum retinol level <0.35 μmol L^{-1} (10 μg dL^{-1}) indicates severe deficiency, while concentrations between 0.70 and 1.05 μmol L^{-1} is often characterized as marginal or low status. Decrements in serum retinol below 0.70 μmol L^{-1} can be expected to increase the severity of xerophthalmia and infectious illness. Breast milk retinol may also serve as an indicator, with low milk retinol serving as a measure of both maternal status and intake inadequacy of breastfed infants. Dried blood retinol, determined to provide reasonably comparable data to serum, has periodically been collected and analyzed to assess population prevalence of VA deficiency. A test, occasionally deployed, that indirectly assesses total liver retinol reserves, called a modified (from an earlier version) relative dose response test, involves measuring the dehydroretinol (VA_2-acetate, a different form of VA) concentration in serum collected 5 h following a small challenge dose. As the orally administered form of VA_2 is released from the liver carried by RBP4 that accrues in the liver during VA deficiency, a high concentration relative to serum retinol provides a qualitative indication of a population that is depleted. Increasingly, stable isotopes of VA (e.g., ^{2}H- or ^{13}C-labeled retinyl acetate) are being employed to assess total body stores of VA of both individuals and groups. Retinol isotope dilution is a method that involves the administration of a small dose of the isotope, followed by a blood draw 4–14 days later to measure the ratio of labeled to unlabeled VA.

VA deficiency disorders (VADD)

VA deficiency disorders (VADD) can be defined as pathophysiological consequences attributed, in part or whole, to in adequate VA nutriture, as illustrated in **Fig. 1**. As VA is essential to maintain normal retinal function, and in regulating cellular proliferation, differentiation and energy utilization, VADD can range from deficiency-induced tissue, organ or system dysfunction to disease conditions. Given VA is essential for photoreceptor function, night blindness represents a VADD. As effects of deficiency readily occur in rapidly dividing, bipotential cells such as epithelial linings leading to metaplasia and keratinization of mucosal surfaces, disorders include conjunctival and corneal xerosis, as well as metaplasia and mucociliary and immune defects throughout the respiratory, genitourinary and gastrointestinal tracts that may increase risk of infection (Sommer and West, 1996). Exposure to deficiency during the embryonic and fetal periods may effect epigenetic changes that permanently alter developing cell populations and organ systems that may impair lung, innate and adaptive immune or other organ systems with adverse health effects later in life.

Xerophthalmia

Conjunctival and corneal epithelia deprived of VA undergo keratinizing metaplasia (Sommer, 1995; Sommer and West, 1996). Columnar epithelial cells on the ocular surface become squamous and mucus-producing goblet cells disappear, providing the histopathologic mechanisms for deficiency-induced xerotic (drying) changes to the ocular surfaces. VA deficiency is also required for rod vision in dim light. VA deficiency-induced night blindness often occurs with histopathologic changes on the ocular surface. Thus, night blindness and clinical eye signs both are listed under one xerophthalmia classification scheme (**Table 1**).

Night blindness (XN)

VA, as retinaldehyde, is an essential photosensitive pigment in rod cells of the retina that respond to light (become "bleached") by releasing the VA ligand from the protein rhodopsin, thereby initiating neural impulses to the brain that permits vision under conditions of low illumination. The utilization and recycling of VA in this process is known as the visual (or retinoid) cycle. Hypovitaminosis A restricts rhodopsin production that, in turn, raises the scotopic (low light) visual threshold. Gradually, a perceptive threshold is reached that leads to recognition of XN, the earliest symptom of xerophthalmia. It is marked by an inability to move about in the dark. Young children between 1 and 5 years of age and pregnant women appear to be at greatest risk of XN. Where VA deficiency is endemic, there is often a local term for XN that translates into "evening" or "twilight" blindness or "chicken

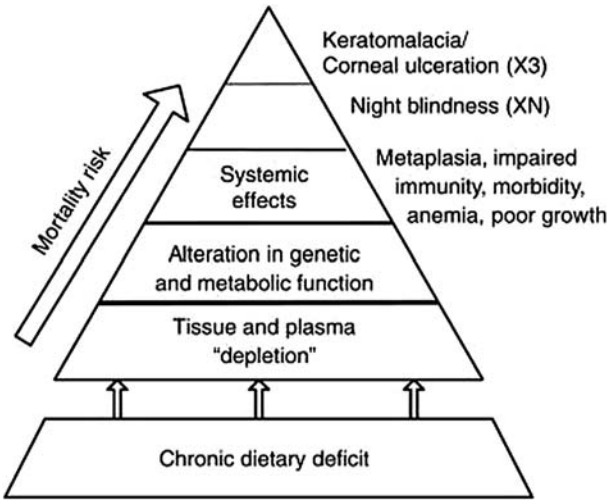


Fig. 1 Concept of VADD, due primarily to underlying chronic dietary deficit in preformed VA and proVA carotenoids. Reproduced from West (2002).

Table 1 World Health Organization (WHO) Classification and minimum prevalence criteria for xerophthalmia and VA deficiency as a public health problem.

Definition (code)	*Minimum prevalence (%)*	*Highest-risk period*
Children 1–5 years of age		
Night blindness (XN)	1.0	2–6 yr
Conjunctival xerosis (X1A)	–	–
Bitot's spots (X1B)	0.5	2–6 yr
Cornea xerosis (X2)/corneal ulceration (X3A)/ keratomalacia (X3B)	0.01	1–3 yr
Xerophthalmic corneal scar (XS)	0.05	>1 yr
Deficient serum retinol (<0.70 μmol L^{-1})	15.0	<5 yr
Pregnant/lactating women		
XN during most recent pregnancy	5.0	Third trimester
Low serum retinol (<1.05 μmol L^{-1})	20.0	Third trimester

Source: Adapted from Sommer and Davidson (2002). Alpha-numeric characters in ( ) denote the WHO classification scheme for xerophthalmia.

eyes" (chickens lack rod cells and also cannot see at night), making the condition readily detectable by history. XN typically disappears within 24–48 h of VA treatment, while gestational XN resolves spontaneously shortly after child birth and expulsion of the placenta, likely due to reduced metabolic demands for VA.

Conjunctival xerosis (X1A) and Bitot's spots (X1B)

Early xerosis of the conjunctiva can be detected by filter paper impression cytology, showing distorted, enlarged, and noncontiguous sheaths of epithelial cells, and disappearance of goblet cells. In advanced VA deficiency, confluent xerosis appears clinically as a dry, unwetable surface of the bulbar conjunctiva (X1A). The affected areas are usually overlaid with superficial white, cheesy, or foamy patches of triangular or oval shape that consist of desquamated keratin and bacteria (often *Bacillus xerosis*). These are known as Bitot's Spots (designated X1B). They are nearly always bilateral, found temporal (and, in more advanced cases, also nasal) to the corneal limbus and is a sign more reliably diagnosed than X1A. X1B are not blinding but are reflective of chronic moderate-to-severe systemic depletion of VA (Sommer and West, 1996).

Corneal xerophthalmia (X2/X3)

Corneal xerophthalmia is an ocular manifestation of severe VA deficiency. The earliest corneal lesions appear as superficial punctate defects, evident with a slit lamp, that with advanced deficiency become more numerous and concentrated. The cornea is considered xerotic (X2) when the keratopathy covers large areas of its surface rendering a hazy, nonwetable, lusterless, and irregular appearance on handlight examination. Stromal edema may be present. In more severe cases, thick, elevated X2 plaques may form. Usually both eyes are affected. X3A can be sharply demarcated, round, or oval defects that are usually shallow but may also perforate the cornea. Healed ulcers form a leukoma (scar) or adherent leukoma if the iris has plugged the perforated ulcer. Most

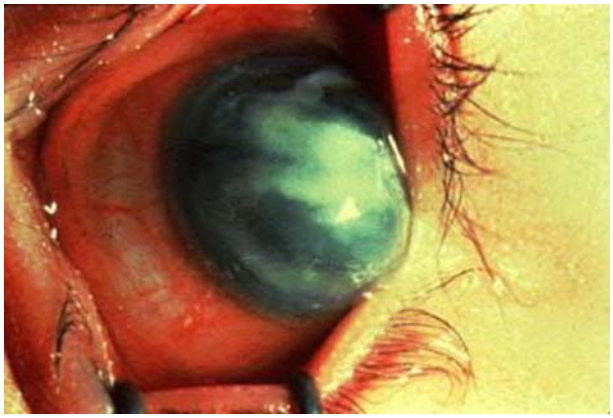


Fig. 2 Keratomalacia. Reproduced from Sommer (1995).

ulcers occur peripheral to the visual axis and, thus, may not threaten central vision if promptly treated. X3B refers to a full-thickness softening and necrosis of the corneal stroma that can cause protruding, opaque, yellow-to-gray lesions (**Fig. 2**). These may reduce or slough off leaving a descemetocele following VA treatment. X3B usually impairs vision in the involved eye although the degree of visual loss depends on the location, thickness, and extent of corneal necrosis and the resultant scar. Owing to the generally malnourished and ill state of children with corneal xerophthalmia, fatality of hospitalized cases ranges from 4% to 25%.

Other VADD: infection, anemia and poor growth

Infection

A synergism exists between hypovitaminosis A and infection, each exacerbating the other, representing a classic "vicious cycle." In this context, infection may be considered both a cause of VA deficiency and, in terms of severity and sequellae, a consequence, or "disorder." Xerophthalmia or severe hyporetinolemia have been consistently associated in cross-sectional assessment with higher risks of diarrhea, fever, and other infections though an acute phase reaction to infection can lower the circulating retinol level and directionality can be difficult to establish.

VA deficiency presumably raises risk of infection by compromising "barrier" epithelial defenses and impairing regulatory innate, cell-mediated and antibody-mediated immune mechanisms. Retinoic acid is also involved in suppressing pathological and autoimmune responses to inflammation, which could underlie its role in controlling severity of infection (Stephensen and Lietz, 2021). Epidemiological corollary evidence exists. For example, Indonesian preschoolers with mild xerophthalmia were twice as likely to develop acute respiratory infection and three times more likely to develop diarrhea over subsequent 3–6-month periods. VA-deficient children are also more likely to die, as seen in Indonesia (**Fig. 3**), where risk of mortality rose with increased severity of mild eye signs (Sommer and West, 1996; Palmer et al., 2017). Measles is an illness that can deplete VA nutriture which, in malnourished children, can lead to corneal xerophthalmia, more severe complications, and a high risk of death. VA deficiency may exacerbate severity of middle ear infection that may consequently increase the risk of hearing loss (Schmitz et al., 2012).

Anemia and poor growth

Children with xerophthalmia tend to be anemic relative to peers without eye disease. VA supplementation can often improve hematological indicators of iron status and reduce anemia. Mechanisms involved in this interaction are not clear but could involve enhanced iron absorption, storage, and transport as well as direct effects on hematopoiesis in the presence of adequate iron stores.

VA deficiency decelerates growth in animals and has been observed to be associated with both stunting and wasting malnutrition in children, possibly reflecting plausible roles for the vitamin in osteogenesis and energy metabolism. Trials, however, have failed to show consistent effects of VA supplementation on child growth, possibly due to variations in the extent of infectious morbidities, seasonality in dietary protein and energy adequacy across populations studied, exclusion of xerophthalmic (most VA deficient) children in growth studies, or other nutritional deficiencies limiting growth.

Epidemiology

The epidemiology of VA deficiency is mostly understood in relation to hyporetinolemia and xerophthalmia (**Table 1**), especially the more prevalent, milder stages of XN and X1B. Preschool children and women of reproductive age are at highest risk of being VA

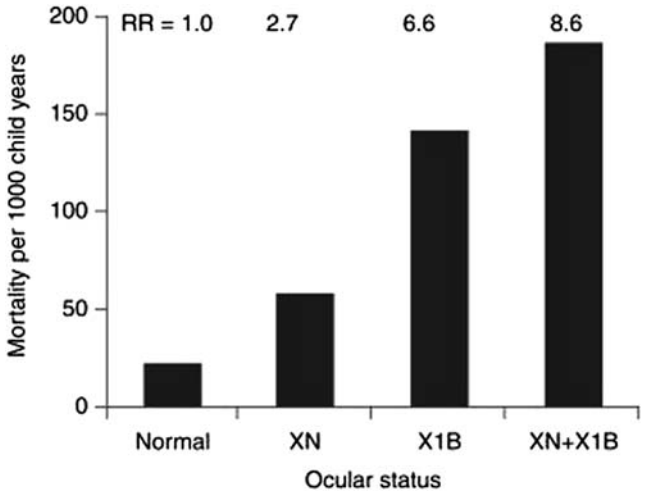


Fig. 3 Risk of mortality among ~3500 Indonesian preschool children by ocular status at the outset of each 3-month interval. RR, relative risk of mortality. Adapted from Sommer et al. (1983).

deficient, affecting an estimated ~30% an ~15%, respectively, based on population-based estimates of serum retinol (or surrogate RBP4) concentrations $<0.70\ \mu mol\ L^{-1}$ (or $20\ \mu g\ dL^{-1}$), respectively, best understood in terms of how deficiency distributes by person, place, and time (World Health Organization, 2009; Stevens et al., 2015).

Person (high-risk groups)

Despite uncertainty about an appropriate cutoff for hyporetinolemia in the 1st months of life, infants in all settings are born with low VA stores. Infants and young children rely on VA intake from breast milk and complementary foods to build stores. If these are inadequate, as they are in many low-income settings, deficiency can persist throughout early childhood. Most recent global estimates suggest that VA deficiency (serum retinol $<0.70\ \mu mol\ L^{-1}$) afflicts approximately 30% of preschool-aged children in the developing world. An estimated 1% of these exhibit XN.

The current extent of xerophthalmia is less known, as few national ophthalmological surveys have been conducted since the 1990s, although case and local area reports serve as reminders that VA deficiency can recur. Milder stages of xerophthalmia typically affect children beyond the 1st year of age, as breast milk is replaced by less nutritious food from the household diet. Boys tend to be at higher risk, possibly reflecting gender differences in dietary practices. Risk of corneal xerophthalmia (**Fig. 2**) peaks in the second through 4th years of life, typically following epidemics of acute infection such as severe measles. Informally, it appears that numbers of severe xerophthalmia cases globally have dropped in high-risk populations, likely due to the broad reach of semiannual VA supplementation, which reduces risk of corneal disease by ~90%, and decades of steadily increasing coverage of vaccination against measles (Sommer and West, 1996). Effects of the COVID-19 pandemic on corneal xerophthalmia are unknown, though could be a concern amid widespread suspension of measles vaccinations and calls in recent years to limit semi-annual VA supplementation campaigns.

VA deficiency can affect women, especially during pregnancy, in undernourished settings, appearing to be driven by heightened nutrient demands of gestation. Globally, some 19 million pregnant women are estimated to be VA deficient (serum retinol concentration $<0.70\ \mu mol\ L^{-1}$), of whom nearly 10 million are estimated to become night blind (World Health Organization, 2009). Maternal XN not only identifies women at high risk of hyporetinemia but also anemia, wasting malnutrition, offspring mortality, and maternal morbidity and mortality (**Fig. 4**). Widespread maternal VA deficiency has been observed to coexist with HIV infection and AIDS throughout sub-Saharan Africa. However, modest, absent, or inconsistent effects of maternal VA supplementation on infant and maternal health, survival, or HIV transmission suggest the deficiency may be a consequence more than determinant of disease severity. Host-HIV competition for VA has also been implicated from intervention trials in Africa.

Place (geographic clustering)

Southern Asia and sub-Saharan Africa remain regions with the greatest burden of VA deficiency (**Fig. 5**) (World Health Organization, 2009; Stevens et al., 2015). Within a region, risk of deficiency can vary by local risk factors, especially, where diets are chronically inadequate and infectious disease burden high (Palmer et al., 2017). Clustering of xerophthalmia has been observed in surveys in Africa and Asia, where preschool children have been shown to incur a ~2-fold higher risk of xerophthalmia in villages where another child has been diagnosed compared to villages free of previous xerophthalmia (**Table 2**). More striking is a 7–13-fold higher risk of xerophthalmia described in siblings of cases compared to children at home with no history of xerophthalmia (Katz et al., 1993). Maternal XN and childhood xerophthalmia often coexist in household and community clusters, where spatial clustering appears to arise mostly from a shared, inadequate diet.

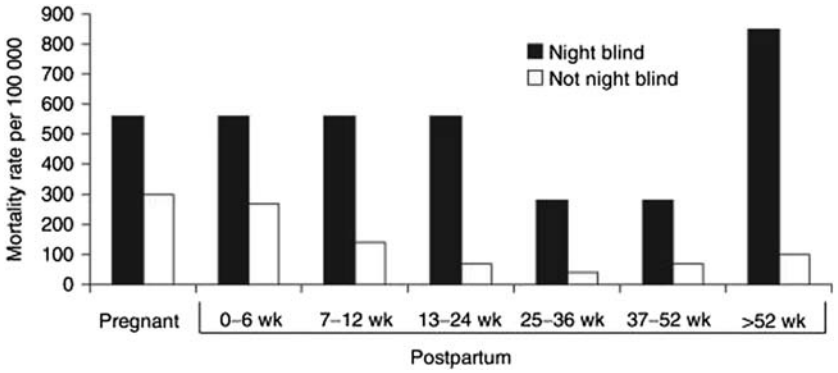


Fig. 4 Mortality rates of rural Nepalese women (per 100,000 pregnancies) during and for up to 2 years following pregnancy according to whether mothers experienced night blindness ($n = 361$) or not ($n = 3052$) during pregnancy. Reproduced from Christian et al. (2000).

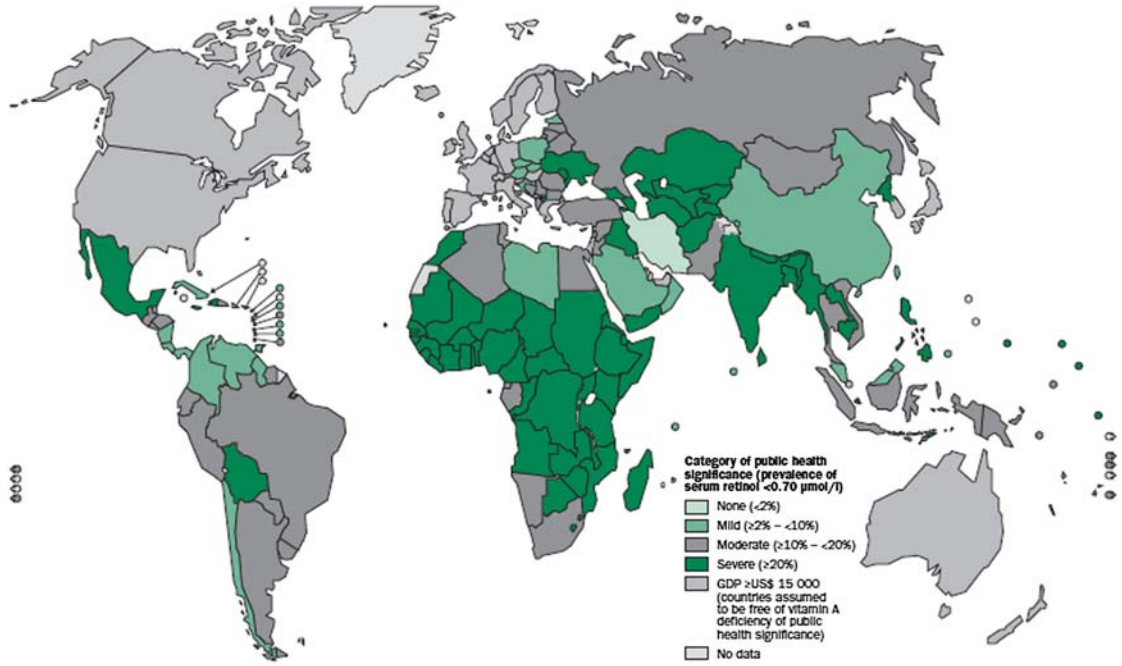


Fig. 5 Global geographic distribution and public health significance of the prevalence of VA deficiency, based on serum retinol concentrations <0.70 µmol L^{-1} in preschool-aged children. Reproduced from the World Health Organization (2009).

Time (periodicity)

Occurrence of xerophthalmia can follow predictable, though not parallel, seasonal peaks in different parts of the world, appearing to emerge from a convergence of factors. In South Asia, for example, absent VA interventions mild xerophthalmia peaks in the hot dry/early monsoon seasons (April–July), following a postharvest growth spurt in the preceding cool dry season. Peak risk also coincides with a general scarcity of proVA-rich vegetables and fruits and a seasonal rise in diarrhea, respiratory infections and measles, that interestingly tends to decline abruptly midway through the beta-carotene-rich mango season. Periodicity, where it exists, can help identify not only causes but specific times of the year for targeting interventions.

Causal agents (diet and infection)

Deficiency results from consuming a diet chronically inadequate in VA in relation to need, that can be exacerbated by infection. A low dietary fat intake (e.g., <5% of calories) may restrict absorption of proVA carotenoids from vegetables and fruits and thus also predispose certain poor populations to deficiency.

Table 2 Age-adjusted village and household odds ratios for risk of xerophthalmia among preschool children.

	Malawi		*Zambia*		*Indonesia*		*Nepal*	
	n	*OR*	N	*OR*	n	*OR*	N	*OR*
Village	50	1.2 (1.0–1.5)	110	1.7 (0.9–3.2)	460	1.8 (1.4–2.2)	40	2.3 (1.6–3.4)
Household	2899	7.3 (3.2–16.7)	2449	7.9 (3.5–17.8)	16,337	10.5 (7.0–15.7)	2909	13.2 (6.0–29.0)

Source: Adapted from Katz et al. (1993). Numbers of children <6 years of age in each country: Malawi ($n = 5441$); Zambia ($n = 4316$); Indonesia ($n = 28,586$); and Nepal ($n = 4764$). Figures are pairwise odds ratios (OR) based on alternating logistic regression (95% confidence intervals).

Breastfeeding and diet

Breast milk is an infant's most important initial dietary source of VA. Commonly, breast milk from marginally nourished mothers contains ~500 μg of retinol activity equivalents (RAE) per liter, thereby delivering 325 μg RAE per day to infants typically consuming ~650 mL day^{-1}. An "adequate intake" of 400–500 μg RAE has been set as a dietary guide for infants, making this common level of VA intake from breast milk marginal during the 1st year and, beyond infancy, marginally above an estimated deficient intake threshold of 210 μg RAE, but usually sufficient to prevent clinical signs. Studies in Asia and Africa show breastfed infants and toddlers to be 65–90% less likely to develop xerophthalmia than nonbreast-fed peers of the same age. In Malawi, children who developed xerophthalmia had begun weaning ~1 month earlier and completed breastfeeding ~6 months earlier than nonxerophthalmic children. Even among breastfed children, in rural Nepal a higher daily frequency of feeding was associated with added protection against xerophthalmia.

The choice of foods offered to complement breast feeding may also affect risk of VA deficiency. Indonesian preschoolers were at a 2-to 6 fold higher risk of xerophthalmia if food sources of VA such as dark green leaves, mango or papaya, egg, meat or fish with liver, and milk, and other dairy products were not routinely given during their 1st year of complementary feeding, likely initiating a dietary exposure pattern that increases throughout the preschool years. Across undernourished regions, less frequent intakes of preformed VA and proVA-rich foods are observed by children with a history of xerophthalmia (Fig. 6). Similar dietary inadequacy is more likely to be reported among women with, versus without, a history of XN during a previous pregnancy.

Infection

Evident as a vicious cycle, infection can be viewed as a cause of VA deficiency, consequent to impaired VA absorption, increased metabolism, and urinary losses of the vitamin (Stephensen and Lietz, 2021). Prospective studies show that severe infections such as measles, chicken pox, diarrhea, and acute respiratory illness decrease serum retinol, as well as apparent hepatic retinoid stores, and increase risk of xerophthalmia. In some settings, severe measles has been observed to increase by more than 13-fold the risk of precipitating corneal xerophthalmia. In Indonesia, young children with diarrhea and acute respiratory infections were twice as likely to develop mild xerophthalmia (XN or X1B) than disease-free children. Similar temporality between infection and incident XN have been observed in undernourished pregnant women. Explanations for a role of infection as a cause of VA deficiency include decreased absorption of VA, impaired retinol transport, increased metabolic requirements, greatly increased renal excretion during the acute-phase response, and immunodeficiency.

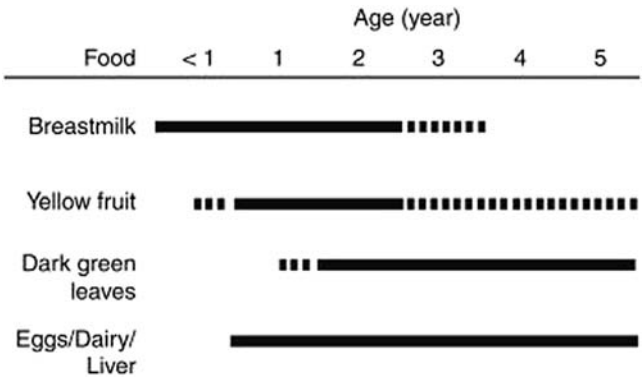


Fig. 6 Foods that protect against xerophthalmia in early childhood, based on numerous studies. Dark line, strong evidence; dashed line, suggestive evidence. Reproduced from Sommer and West (1996).

Impact of interventions

VA deficiency can be prevented through direct supplementation, fortification of common food items in factories or at "point of consumption" (in the home) and other food-based approaches that include biofortification of staple crops with proVA carotenoids, home gardening, small animal husbandry, and nutrition education. Evaluations to date—mostly of supplementation, and both industrially fortified and biofortified foods—have focused on VA status, xerophthalmia, mortality, and morbidity.

On vitamin A status

The impact of VA interventions on population status can vary by indicator, dosage, and mode of delivery, as well as population status and health, including severity of deficiency and dominant infectious diseases (e.g., malaria, HIV, possibly COVID). Despite its limitations in reflecting liver reserves, except at their lowest concentrations, serum retinol, historically to the present day, remains the most widely employed indicator of VA status. Serum concentrations of retinol binding protein (RBP4), the major nutrient carrier in circulation may be substituted, with limitations, for retinol status. Methods that rely on use of stable isotopes or time-dependent response-to-therapy to VA supplementation can approximate body nutrient stores, valuable for metabolic nutrition research, are more rarely deployed for population use due to technical difficulty, requisite assumptions, cost and, thus, sample size.

A single, large oral dose of VA (210 μmol, 60 mg RAE, or 200,000 IU) as routinely given semiannually to young children in undernourished populations of low-middle income countries can elevate serum retinol to near adequacy for up to a few weeks to months before declining (Palmer et al., 2012). Reasons for the transient response to a supplemental bolus remain poorly understood. Routine intake of a third to full recommended dietary allowance of VA through fortified food items, such as vegetable oil, sugar and wheat flour, gradually improves and sustains adequate serum and breast milk retinol concentrations and liver stores, assessed indirectly. Regular consumption of proVA, either as natural food sources (dark green leaves, yellow vegetables and fruits) or via biofortified crops, generally improves VA status in undernourished populations, although not to adequacy, as the response can be affected by many factors including food matrix, methods of storage and preparation, amounts of preformed VA and fat in the diet, gut integrity and function, host protein energy and VA status and genetic predisposition (e.g., single nucleotide polymorphisms in beta-carotene oxygenase 1, the intestinal enzyme responsible for cleaving beta-carotene to retinaldehyde that subsequently forms retinol metabolites). Despite variable responses, and multisectoral challenges in effecting food-based interventions, there is unanimity in the public health nutrition community in striving toward assuring a steady, adequate and safe dietary supply of VA and proVA foods to improve VA status of populations.

On xerophthalmia

A large dose VA given to preschool children as a capsule or in syrup (in India) every 6 months is ~90% efficacious in preventing both corneal and noncorneal xerophthalmia. Prophylactic failure (~10% non-response or recurrence) may reflect inadequacy of dosage for some children who are severely VA-deficient or become ill. Xerophthalmia, on the other hand, virtually disappears in societies routinely consuming adequate dietary VA through fortified foods. Supervised treatment with proVA-rich vegetables and fruit has been reported to cure or improve noncorneal xerophthalmia.

On mortality and morbidity

A synergism between infection and VA deficiency leading to more severe morbidity and risk of death has been shown experimentally in animals for the past century. In the 1980s and 90s, the efficacy of VA in reducing preschool child mortality was firmly established by eight controlled, community trials designed and executed to assess this outcome. In six trials, children 6 months to 6 years of age received an oral supplemented with 200,000 IU of VA every 4–6 months. A half-dose was provided to children 6–11 months of age. One study, in India (Vijayaraghavan et al., 1990), provided a small weekly dose of 15,000 IU to young children and another, in Indonesia, supplied one-third of a recommended dietary allowance of VA to children in treatment villages via fortified monosodium glutamate product (flavor enhancer) that was routinely added to meals (Muhilal et al., 1988). Six of the eight trials showed reductions of 19%–54% in child mortality beyond either 6 or 12 months of age, with meta-analyses of data from all eight trials revealing an overall mortality reduction of 23%–34% (Beaton et al., 1993), depending on inclusion criteria.

Causes of death likely impacted by VA include severe measles, diarrhea, dysentery, and febrile illnesses. In Ghana, where child mortality was reduced 19%, VA supplementation also significantly decreased childhood clinic visits for illness by 12% and hospitalization rates for severe disease by 38%; among children hospitalized over the course of the trial, diarrheal illness was less severe among VA versus placebo recipients (Ghana VAST Study Team, 1993). In a trial in Papua New Guinea (Shankar et al., 1999), VA supplementation reduced febrile, falciparum parasitemic episodes by 30% accompanied by lowered risk of an enlarged spleen, suggesting VA may plausibly reduce malaria case fatality. VA has consistently been shown to reduce measles case fatality, by ~50%. A lack of effect on fatality from acute lower respiratory infection (unrelated to measles) has been a consistent and perplexing finding across trials, reasons for which remain uncertain (Stephensen and Lietz, 2021). Overall, data suggest VA has little impact on prevalence of common childhood morbidities, but rather holds the potential to attenuate severity of infections, supporting the effect of the nutrient on mortality in undernourished populations.

Over the decades regularly updated meta-analyses have added data from nearly another 40 trials of VA on mortality and infectious disease outcomes in child populations of widely varying risks, including Australia, China and India, among other countries in

South Asia, Africa and Latin America (Imdad et al., 2022). While adding studies not designed, powered or executed to assess child mortality can dilute the overall effect size, these analyses continue to sustain evidence VA interventions can reduce severe infection and child mortality in undernourished settings.

Other trials have extended the assessment of VA supplementation shortly after birth on infant mortality, requiring different designs than for children. In all, five trials in Southern Asia (Indonesia, Bangladesh, India and Pakistan) and six in sub-Saharan Africa (Tanzania, Zimbabwe, Ghana and Guinea Bissau) have examined the efficacy of a 50,000 IU dose of VA in oil, expressed from a capsule into the mouth of newborns (Neonatal Vitamin A Supplementation Evidence Group, 2019). In Southern Asia, newborn VA significantly consistently reduced mortality in the first half of infancy, by an average of 13% (risk ratio of 0.87) but had no significant effect in African settings. Plausible explanations for the observed impact may include rapid maturational effects of VA on immature, newborn intestinal, airway and systemic immunity that could enhance resistance to infection months later, especially in cultures where prelacteal feeding may impair intestinal immunity. Reasons the for a lack of effect in Africa remain speculative, with overall differences in maternal VA deficiency, more common in Asian settings, and mediating effects of HIV, which was endemic in some African populations, among them.

A third high risk group among whom VA interventions may affect health and survival is pregnant women. An initial trial, in rural Nepal where maternal XN routinely affected ~10% of pregnant women, women were provided a weekly capsule containing the equivalent of a recommended dietary allowance of VA, either preformed or as proVA beta-carotene, or placebo prior to conception through the 1st few years postnatally. Maternal VA or beta-carotene supplementation lowered all-cause, pregnancy-related mortality through twelve weeks postpartum by ~40%, plausibly mediated by improved resistance to severe infections, as may occur in the puerperium, and hematopoietic effects leading to less severe complications related to less anemia (West et al., 1999). Infant mortality was only lower in mothers prone to developing night blindness. Interestingly, when offspring were followed up between 10 and 13 years of age, children born to VA-supplemented mothers exhibited improved lung function (Checkley et al., 2010), expressed by forced expiratory volume at 1 s and forced vital capacity measures, and improved immunity as indicated by greater natural antibody concentrations than children born to mothers receiving placebo during pregnancy (Palmer et al., 2015). Despite the multiple effects of antenatal VA supplementation on mothers and children in Nepal, pregnancy intervention trials in Bangladesh and Ghana failed to reduce maternal mortality (Kirkwood et al., 2010; West et al., 2011), although risk of bacterial vaginosis was markedly lower among VA supplemented mothers in Bangladesh. Discrepant findings on maternal survival may have been related to lower maternal risks of mortality, population differences in VA deficiency and general nutritional status, as well as better primary health care systems in the latter two country settings.

Management

Treatment

WHO recommendations for treating clinical VA deficiency have remained the same for 25 years (Sommer, 1995; Ross, 2002). Children with xerophthalmia and measles should be treated immediately with oral, high-potency VA (200,000 IU), summarized in **Table 3**, and provided other supportive nutritional and medical therapy, as indicated. Corneal lesions should be topically treated with a suitable antibiotic (e.g., tetracycline or chloramphenicol) to prevent bacterial infection. Corneal xerophthalmia typically improves with VA treatment within a week, and resolves within 4 weeks, depending on size, thickness and location of the lesion, and nutritional and health status of the patient. Corneal xerosis (X2) typically resolves without sequelae; ulceration (X3A) will typically heal leaving a scar (leukoma), with location, size and depth in relation to the visual axis being key determinants of vision. Healed keratomalacia (X3B), mostly leaves the cornea severely scarred, possibly shrunken (phthis) or bulging (staphyloma), and the eye blind. Mild xerophthalmia is non-blinding. XN typically disappears within 24–48 h of treatment with VA. Most Bitot's spots start to resolve within 2–5 days of VA treatment, and disappear within 2 weeks, though in older children X1B (underlying keratinized metaplasia) may persist for months, for unclear reasons. Importantly, any signs of xerophthalmia may recur if individuals become VA-deficient again.

Table 3 VA treatment and prevention schedules.

		Prevention	
Age	*Treatment at diagnosis*	*Dosage*	*Frequency*
<6 mo	50,000 IU	50,000 IU	Once within 3 days after birth
6–11 mo	100,000 IU	100,000 IU	Every 4–6 months
12–59 mo	200,000 IU	200,000 IU	Every 4–6 months
Women	By severity of eye signs	200,000 IU	2 doses 24 h apart ~6 weeks after delivery

Source: Reproduced from Ross (2002) and Sommer (1995). Treat all cases of xerophthalmia and measles on days 1 and 2; give an additional dose for xerophthalmia on day 14. For severe malnutrition give 1 dose on day 1. For women of reproductive age, give 200,000 IU only for corneal xerophthalmia on days 1, 2, and 14; for night blindness or X1B give 10,000 IU per day or 25,000 IU per week for >3 months. Newborn prevention schedule is based on authors' recommendation for newborns in Southern Asia, where multiple trials have yielded an average reduction of 13% in infant mortality, but not in Africa where no effect has been observed (Neonatal Vitamin A Supplementation Evidence Group, 2019).

Severely wasted children should be given a single large oral dose (200,000 IU) for prophylaxis. It is also judicious to treat children with severe diarrhea, dysentery, respiratory infection, and measles a single, large oral dose of VA. Large-dose VA is indicated for women of reproductive age with corneal disease. For XN, smaller daily (10,000 IU) or weekly (25,000 IU) doses are recommended for at least 3 months.

Prevention

The above health consequences, or disorders, of VA deficiency can be prevented via direct supplementation of target groups, food fortification at production facilities or by adding nutrient-dense powders with VA to meals at home, proVA carotenoid biofortification of staple crops, and other agricultural or dietary approaches, including gardening and education programs that encourage exclusive and extended breast-feeding and improved dietary quality (Palmer et al., 2017). Indirectly, public health interventions that reduce infection, inflammation and environmental stress can help to improve VA status in a population.

Administration of large-dose, oral VA (200,000 IU), adjusted to age (**Table 3**), on a ~6 monthly basis is a common preventive in most undernourished societies. A half-dose is dispensed to infants 6–11 months of age. Periodically providing a large dose of VA is thought to increase nutrient liver stores from where it is mobilized into circulation, as needed. Supplements can be provided during routine health care (e.g., for growth monitoring, immunization, and other extension services) or more extensively and systematically on a regular (e.g., semiannual) "universal" basis, especially through national campaigns often called Child Health Days, which routinely achieve 80% or greater coverage. Nearly five decades of distributing billions of large-dose VA supplements for prevention, coupled with marked declines in mortality attributed to the intervention, attests to the acceptance, likely continued effectiveness, and safety of this approach. Gastrointestinal side effects may occur in ~5% of recipient children.

Commercial fortification of foods with VA is a longer-term strategy, aiming to increase dietary intakes and sustainably increase VA status by offering a quarter to a full day's recommended allowance of VA. Food vehicles currently being used include sugar, vegetable oils, nonfat milk powder, wheat flour (for bread or noodles), and nonrefrigerated margarine. These share the common traits of being technically fortifiable at planned levels and consumed within a range that could be both effective in deficient groups while remaining safe for nutritionally adequate or overconsuming segments of a population. The decline of VA deficiency as a nutritional concern in Central America is largely attributed to successful sugar fortification in that region. Programs in Africa have had more limited success, with evaluations underscoring the need for strict quality control procedures at the production level to achieve targeted VA levels.

Biofortification of major staple grains and tubers, such as maize, cassava, and sweet potatoes, with proVA carotenoids has the potential to improve VA status in deficient populations, or at the minimum offer a "dietary safety net" that delivers a minimum amount of bioconvertable carotenoid to prevent deficiency during years of economic or agricultural downturn. Dietary diversification is widely held to be the most culturally appropriate and potentially sustainable approach to preventing VA deficiency. Although pilot trials show efficacy of a variety of dietary approaches in improving VA intake and status, data on effectiveness and cost of population food-based interventions remain lacking. Dietary intakes can be improved through home and school gardening initiatives, nutrition education, and social marketing of locally available food sources of VA. However, effective dietary change requires a thorough understanding of local cultural, food system, and behavioral factors that increase the risk of VA deficiency.

References

Beaton, G.H., Martorell, R., Aronson, K.J., Edmonston, B., McCabe, G., Ross, A.C., Harvey, B., 1993. Effectiveness of Vitamin A Supplementation in the Control of Young Child Morbidity and Mortality in Developing Countries. Administrative Committee on Coordination–Subcommittee on Nutrition (ACC/SCN), Geneva, Switzerland. Nutrition Policy Discussion Paper No. 13.

Checkley, W., West Jr., K.P., Wise, R.A., Baldwin, M.R., Wu, L., LeClerq, S.C., Christian, P., Katz, J., Tielsch, J.M., Khatry, S., Sommer, A., 2010. Maternal vitamin A supplementation and lung function in offspring. N. Engl. J. Med. 362 (19), 1784–1794.

Christian, P., West Jr., K.P., Khatry, S.K., Kimbrough-Pradhan, E., LeClerq, S.C., Katz, J., Shrestha, S.R., Dali, S.M., Sommer, A., 2000. Night blindness during pregnancy and subsequent mortality among women in Nepal: effects of vitamin A and beta-carotene supplementation. Am. J. Epidemiol. 152 (6), 542–547.

Ghana VAST Study Team, 1993. Vitamin A supplementation in Northern Ghana: effects on clinic attendances, hospital admissions, and child mortality. Lancet 342 (8862), 7.

Imdad, A., Mayo-Wilson, E., Haykal, M.R., Regan, A., Sidhu, J., Smith, A., Bhutta, Z.A., 2022. Vitamin A supplementation for preventing morbidity and mortality in children from six months to five years of age. Cochrane Database Syst. Rev. 3, CD008524.

Katz, J., Zeger, S.L., West Jr., K.P., Tielsch, J.M., Sommer, A., 1993. Clustering of xerophthalmia within households and villages. Int. J. Epidemiol. 22 (4), 709–715.

Kirkwood, B.R., Hurt, L., Amenga-Etego, S., Tawiah, C., Zandoh, C., Danso, S., Hurt, C., Edmond, K., Hill, Z., Ten Asbroek, G., Fenty, J., Owusu-Agyei, S., Campbell, O., Arthur, P., ObaapaVit, A.T.T., 2010. Effect of vitamin A supplementation in women of reproductive age on maternal survival in Ghana (ObaapaVitA): a cluster-randomised, placebo-controlled trial. Lancet 375 (9726), 1640–1649.

Muhilal, Permeisih, D., Idjradinata, Y.R., Muherdiyantiningsih, Karyadi, D., 1988. Vitamin A-fortified monosodium glutamate and health, growth, and survival of children: a controlled field trial. Am. J. Clin. Nutr. 48 (5), 1271–1276.

Neonatal Vitamin A Supplementation Evidence Group, 2019. Early neonatal vitamin A supplementation and infant mortality: an individual participant data meta-analysis of randomised controlled trials. Arch. Dis. Child. 104 (3), 217–226.

Palmer, A.C., Darnton-Hill, I., West Jr., K.P., 2017. Vitamin A deficiency. In: de Pee, S., Taren, D., Bloem, M.W. (Eds.), Nutrition and Health in a Developing World. Humana Press, Cham, Switzerland.

Palmer, A.C., Schulze, K.J., Khatry, S.K., De Luca, L.M., West Jr., K.P., 2015. Maternal vitamin A supplementation increases natural antibody concentrations of preadolescent offspring in rural Nepal. Nutrition 31 (6), 813–819.

Palmer, A.C., West Jr., K.P., Dalmiya, N., Schultink, W., 2012. The use and interpretation of serum retinol distributions in evaluating the public health impact of vitamin A programmes. Publ. Health Nutr. 15 (7), 1201–1215.

Ross, D.A., 2002. Recommendations for vitamin A supplementation. J. Nutr. 132 (9 Suppl. 1), 2902S–2906S.

Schmitz, J., West Jr., K.P., Khatry, S.K., Wu, L., Leclerq, S.C., Karna, S.L., Katz, J., Sommer, A., Pillion, J., 2012. Vitamin A supplementation in preschool children and risk of hearing loss as adolescents and young adults in rural Nepal: randomised trial cohort follow-up study. BMJ 344, d7962.

Shankar, A.H., Genton, B., Semba, R.D., Baisor, M., Paino, J., Tamja, S., Adiguma, T., Wu, L., Rare, L., Tielsch, J.M., Alpers, M.P., West Jr., K.P., 1999. Effect of vitamin A supplementation on morbidity due to *Plasmodium falciparum* in young children in Papua New Guinea: a randomised trial. Lancet 354 (9174), 203–209.

Sommer, A., 1995. Vitamin A Deficiency and Its Consequences: A Field Guide to Detection and Control. World Health Organization, Geneva, Switzerland, p. 78.

Sommer, A., West Jr., K.P., 1996. Vitamin A Deficiency: Health, Survival and Vision. Oxford University Press, New York.

Sommer, A., Davidson, F.R., 2002. Assessment and control of vitamin A deficiency: the Annecy Accords. J. Nutr. 132 (9 Suppl. 1), 2845S–2850S.

Sommer, A., Tarwotjo, I., Hussaini, G., Susanto, D., 1983. Increased mortality in children with mild vitamin A deficiency. Lancet 2 (8350), 585–588.

Stephensen, C.B., Lietz, G., 2021. Vitamin A in resistance to and recovery from infection: relevance to SARS-CoV2. Br. J. Nutr. 126 (11), 1663–1672.

Stevens, G.A., Bennett, J.E., Hennocq, Q., Lu, Y., De-Regil, L.M., Rogers, L., Danaei, G., Li, G., White, R.A., Flaxman, S.R., Oehrle, S.P., Finucane, M.M., Guerrero, R., Bhutta, Z.A., Then-Paulino, A., Fawzi, W., Black, R.E., Ezzati, M., 2015. Trends and mortality effects of vitamin A deficiency in children in 138 low-income and middle-income countries between 1991 and 2013: a pooled analysis of population-based surveys. Lancet Glob. Health 3 (9), e528–536.

Tanumihardjo, S.A., Russell, R.M., Stephensen, C.B., Gannon, B.M., Craft, N.E., Haskell, M.J., Lietz, G., Schulze, K., Raiten, D.J., 2016. Biomarkers of Nutrition for Development (BOND)-vitamin A review. J. Nutr. 146 (9), 1816S–1848S.

Vijayaraghavan, K., Radhaiah, G., Prakasam, B.S., Sarma, K.V., Reddy, V., 1990. Effect of massive dose vitamin A on morbidity and mortality in Indian children. Lancet 336 (8727), 1342–1345.

West Jr., K.P., Christian, P., Labrique, A.B., Rashid, M., Shamim, A.A., Klemm, R.D., Massie, A.B., Mehra, S., Schulze, K.J., Ali, H., Ullah, B., Wu, L.S., Katz, J., Banu, H., Akhter, H.H., Sommer, A., 2011. Effects of vitamin A or beta carotene supplementation on pregnancy-related mortality and infant mortality in rural Bangladesh: a cluster randomized trial. JAMA 305 (19), 1986–1995.

West Jr., K.P., Katz, J., Khatry, S.K., LeClerq, S.C., Pradhan, E.K., Shrestha, S.R., Connor, P.B., Dali, S.M., Christian, P., Pokhrel, R.P., Sommer, A., 1999. Double blind, cluster randomised trial of low dose supplementation with vitamin A or beta carotene on mortality related to pregnancy in Nepal. The NNIPS-2 Study Group. BMJ 318 (7183), 570–575.

West Jr., K.P., 2002. Extent of vitamin A deficiency among preschool children and women of reproductive age. J. Nutr. 132 (9), 2857S–2866S.

World Health Organization, 2009. Global prevalence of vitamin A deficiency in populations at risk 1995–2005. In: WHO Global Database on Vitamin A Deficiency. World Health Organization, Geneva, Switzerland, p. 116.

Vitamin A: Physiology, dietary sources and requirements

A. Catharine Ross, Department of Nutritional Sciences, The Pennsylvania State University, University Park, PA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A.C. Ross, Vitamin A: Physiology, Dietary Sources, and Requirements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 333–339, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00273-7.

Introduction 472
Major molecules in vitamin A biochemistry and physiology 472
Metabolites formed through oxidation 473
Polar compounds 473
Transport 473
Retinol-binding protein (RBP4) 473
Cellular retinoid binding proteins 474
Nuclear retinoid receptors 474
Absorption and metabolism 474
Intestinal metabolism 474
Retinoic acid in the intestine 475
Hepatic vitamin A uptake, storage and release 475
Plasma concentrations 475
Relationship of plasma retinol to liver retinol concentration 476
Metabolic disturbances that affect plasma retinol 476
Hypervitaminosis A 476
Vitamin A kinetics 476
Tissue retinoid metabolism 476
Physiological actions 477
Vision 477
Vitamin A and the cornea 477
Functions in cell differentiation 478
Immunity 478
Dietary sources and nutritional equivalency 478
Units of nutritional activity 478
Recommended Dietary Allowances (RDA) and tolerable upper intake levels (UL) for vitamin A 478
Conditions of excess 480
Hypervitaminosis A 480
Excessive carotene intake 480
Vitamin A supplementation in public health 480
Conclusion/summary/outlook 480
References 481

Key points

- Both preformed vitamin A (retinol) and provitamin A (carotenoids) serve as the nutritionally necessary forms of vitamin A
- Metabolism occurring in the intestine, liver, and peripheral organs is essential to produce the bioactive forms of vitamin A
- The retina of the eye utilizes a unique form of vitamin A, 11-*cis*-retinal, to form the visual pigment rhodopsin, required for vision
- Both vitamin A deficiency and excess have profound consequences on embryonic development, as well as affecting the immune system and other essential physiological processes

Glossary

Retinol-binding protein The transport protein for vitamin A (retinol) in plasma
Retinol Activity Equivalent (RAE) The unit of dietary vitamin A quantification

Retinoic acid The active metabolite of vitamin A in most tissues of the body
Visual cycle The reactions that recycle retinol and 11-*cis*-retinal in the rods and cones of the retina

Introduction

Vitamin A (retinol) physiology comprises reactions that take place in the intestine (absorption), liver (uptake, storage, metabolism, secretion and elimination), and numerous target organs and epithelial cells throughout the body, especially the eye, immune system, and reproductive systems. Metabolism in the small intestine converts dietary forms of vitamin A, either retinyl esters or beta-carotene, into functionally active metabolites. In the retina, specific metabolic reactions generate 11-*cis*-retinal for vision. All-*trans*-retinoic acid is produced and catabolized by well-controlled metabolic processes in nearly all tissues. As vitamin A is not naturally water-miscible, its metabolism depends on various binding proteins that increase solubility and direct the vitamin A compounds to specific enzymes for metabolism. The nuclei of cells in nearly all tissues contain nuclear retinoic acid receptors that function to regulate vitamin A homeostasis and the many physiological functions of vitamin A throughout the body. This chapter elucidates these processes and discusses the nutritional values of various molecules in the vitamin A family, as well as the physiological consequences of a deficiency or excess of vitamin A.

Vitamin A (retinol) is a fat-soluble micronutrient that is essential for the life of all vertebrates. Its many functions include roles in vision, immunity, skin health, and reproduction. "Fat-soluble A" was discovered in 1913 as a dietary factor required for growth and survival, and shown to be present in very small amounts in the lipid fraction of certain foods of animal origin, including eggs, butter, whole milk, and fish liver, all of which contain preformed vitamin A in the form of retinol and its esters (Wolf, 1996). These precursors are metabolized in vivo into compounds that, in turn, exert the biological effects of vitamin A. A second form of vitamin A was soon discovered in certain foods of plant origin, especially green and deep-yellow vegetables, and characterized as beta-carotene. These "provitamin A" molecules can be cleaved and converted to retinol during intestinal absorption (von Lintig, 2012). Therefore, the nutritional requirement for vitamin A can be met by either preformed retinol, provitamin A carotenoids, or a mixture of both, providing vitamin A in forms that are useful to carnivorous, herbivorous, and omnivorous mammals (Ross and Harrison, 2014).

The true biological activities attributed to vitamin A are due to metabolites that are formed intracellularly in a well-regulated series of processes (Dawson, 2000). After retinol is absorbed and distributed to tissues, its activation takes place in step-wize oxidative reactions. In the case of provitamin A carotenoids, they must first be cleaved and then enzymatically reduced to produce retinol before they can undergo activation. Two of the best known metabolites are crucial to two of the major physiological functions of vitamin A: 11-*cis*-retinaldehyde (retinal), which is a component of the visual pigment required for vision, rhodopsin (Saari, 2016); and all-*trans*-retinoic acid, a carboxylic acid derivative, which is required in essentially all tissues for the regulation of gene expression, and thereby exerts a wide range of biological effects (Kedishvili, 2016; Ghyselinck and Duester, 2019).

In addition to the natural forms of vitamin A, there are now numerous synthetic compounds, termed retinoids (although the term is often used inclusively for the natural as well as synthetic compounds). Some retinoids are used as drugs, mainly in the treatment of certain disorders of the skin and cancers. Synthetic retinoids must be used with great caution and monitoring due to their potential for causing serious side effects.

Major molecules in vitamin A biochemistry and physiology

Vitamin A and its metabolites comprise numerous molecules that differ in isomeric form and/or oxidation state. The major physiological forms are retinol, including its long-chain fatty acid esters, retinal, and retinoic acid. Additional polar metabolites of retinol and retinoic acid are often formed as end-products.

All-*trans*-Retinol is considered the parent molecule of the vitamin A family (**Fig. 1**). It is comprised of a beta-ionone ring, a methyl-substituted conjugated side chain, and a terminal hydroxyl functional group. Both all-*trans* and *cis* isomers of retinol and retinoic acid have been isolated from human tissues. In general, the all-*trans* forms are the most abundant. However, in the

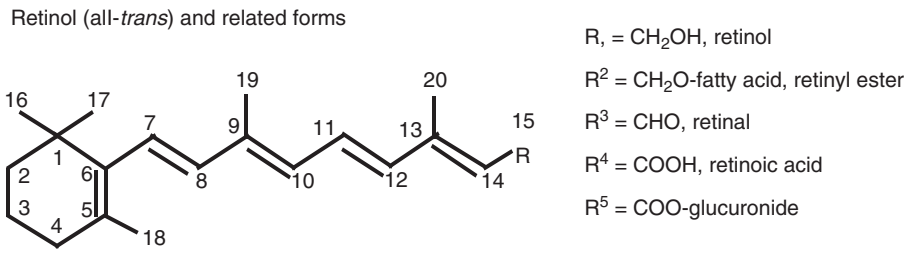


Fig. 1 Structure of all-*trans*-retinol and several related forms.

rods and cones of the retina, 11-*cis*-retinal is the most abundant form, acting as an essential light-sensing component of rhodopsin (see section on **Vision**). Tissues and plasma may also contain low concentrations of 9-*cis* and 13-*cis* isomers. In foods and most of the body's tissues where vitamin A is stored, retinol is esterified with a long-chain fatty acid such as palmitic acid to form retinyl ester, an extremely hydrophobic form that is usually contained within a lipid droplet or lipoprotein. Retinol in nutritional supplements may also be in the esterified form, which improves stability. Vitamin A compounds are sensitive to light and oxygen, so supplements should be kept out of light in closed containers.

Some naturally-occurring, variant forms of retinol are present in certain foods and human tissues, generally in small amounts, these include vitamin A_2, (3,4 didehydroretinol) found in tissues of freshwater fish and produced metabolically in human skin, and α (alpha)-retinol, a structural and less active isomer of all-*trans*-retinol.

Metabolites formed through oxidation

Oxidative metabolism is essential for the physiological functions of vitamin A (Kedishvili, 2016). **Fig. 2** illustrates key processes in retinol metabolism. Retinol is oxidized within cells to generate retinal, which is generally a transient metabolite that can either be reduced back to generate retinol, allowing for cycling between these two molecular forms of vitamin A, or it can be further oxidized in an irreversible reaction to form retinoic acid. All-*trans*-retinoic acid is the major acidic form of vitamin A and the form with widespread effects on gene expression. 9-*cis*-Retinoic acid has also been proposed to function in gene regulation although its in vivo activity remains unproven. 13-*cis*-Retinoic acid another *cis* isomer found naturally in plasma and skin, as well as being used clinically.

Polar compounds

Polar metabolites of retinol or retinoic acid are formed by the addition of hydroxyl or keto groups to the ionone ring structure, generally at ring positions 4 or 18, or by formation of epoxides such as 5,6 epoxy-retinoic acid (Isoherranen and Zhong, 2019). Retinol, retinoic acid and their more polar metabolites may be conjugated with glucuronic acid, forming retinyl and retinoyl-beta-glucuronide. These derivatized forms generally have lower biological activity and a short half-life in vivo; they are readily eliminated in bile and urine.

Transport

Due to the low aqueous solubility of most retinoids, their transport in plasma and cells depends on their association with specific proteins (reviewed in Ross and Harrison, 2014; Blaner et al., 2016; Honarbakhsh et al., 2021). Major forms include:

Retinol-binding protein (RBP4)

RBP (gene name RBP4) is the major transporter of retinol in plasma. RBP is a 21-kDa protein of the lipocalin family, it is synthesized mostly in the liver (see later). It is filtered from plasma in the renal glomerulus, and recovered to a significant extent by binding to a multi-ligand receptor, megalin, which facilitates reuptake. Each RBP molecule binds noncovalently with a single molecule of all-

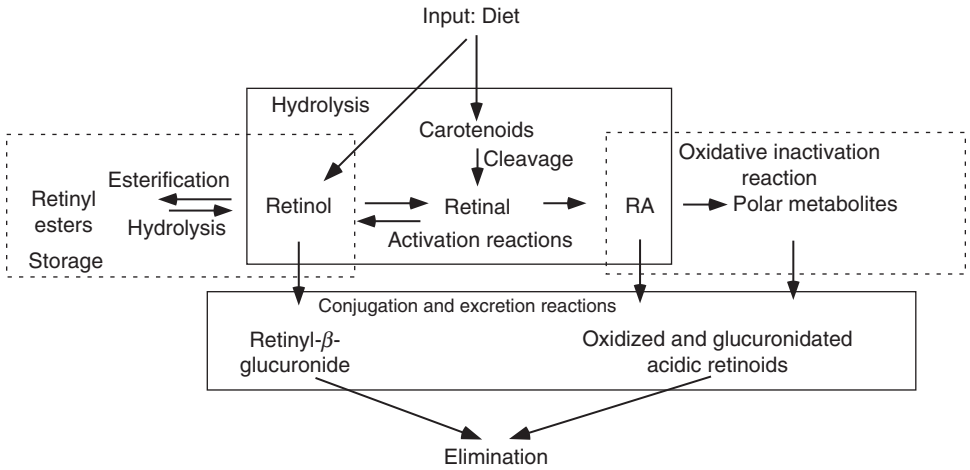


Fig. 2 Schematic of principal reactions of vitamin A metabolism.

trans-retinol; this complex is referred to as holo-RBP. Holo-RBP itself circulates in non-covalent association with another co-transport protein, transthyretin (TTR), to form a molecular complex of retinol-RBP-TTR having a molecular weight of about 75 kDa. TTR is also a transporter for thyroxine.

Cellular retinoid binding proteins

Several proteins referred to as cellular retinoid-binding proteins are members of the gene family of fatty acid binding proteins. These proteins exist in the cytosol of various cell types, each specific protein differing in its preferred ligand and cell-type distribution. The two most abundant forms of cellular retinoid binding proteins are CRBP-I and CRBP-II, each of which binds a molecule of all-*trans*-retinol, with CRBP-I predominant in liver, eye, and several other tissues, and CRBP-II predominant in the enterocytes of the small intestine. CRBP-II can also bind retinal. Similar proteins that are specific for binding all-*trans*-retinoic acid are called CRABP-I and -II. The CRBPs and CRABPs function as chaperones that confer aqueous solubility on their lipophilic retinoid and facilitate their trafficking to specific enzymes that catalyze their metabolism. In the retina, (see **Vision**), there are additional specialized retinoid binding proteins for transporting retinal, which is abundant in the retina, within and between cells of the retina.

Nuclear retinoid receptors

The nuclear retinoid receptor proteins RAR and RXR are specialized forms of retinoid chaperone proteins that reside mainly in the nucleus of cells where they function as regulators of gene transcription (Ghyselinck and Duester, 2019). The RAR (RARα, β and γ) and RXR (RXRα, β and γ) possess a ligand binding domain and a DNA binding domain; these domains determine their specificity with respect to the ligand they prefer to bind, and the specific DNA sequences, termed RARE, that they bind to in retinoid-responsive genes. The binding of all-*trans*-RA to the ligand-binding domain of RARs and, potentially, of 9-*cis*-retinoic acid to the ligand-binding domain of RXRs, alters the receptor's conformation and hence its interactions with transcriptional regulatory complexes. Additionally, numerous post-translational protein modifications of the RAR and RXR have been described which further modulate their gene-regulatory activity. Several hundred genes have been shown to be responsive to retinoic acid in a manner mediated by nuclear receptor signaling.

Absorption and metabolism

Intestinal metabolism

Numerous similarities exist between the absorption of vitamin A and of dietary fat (Ross and Harrison, 2014; Blaner et al., 2016; Honarbakhsh et al., 2021). Beginning in the lumen of the small intestine, dietary retinyl esters must be emulsified with bile salts and phospholipids, incorporated into lipid micelles, and hydrolyzed to free retinol before retinol itself can be absorbed into the mucosa (**Fig. 3**). Hydrolysis is catalyzed by retinyl ester hydrolase enzymes secreted in pancreatic juice or bound to the brush border of duodenal and jejunal enterocytes. Efficient absorption of vitamin A requires that dietary fat is consumed concomitantly, as would

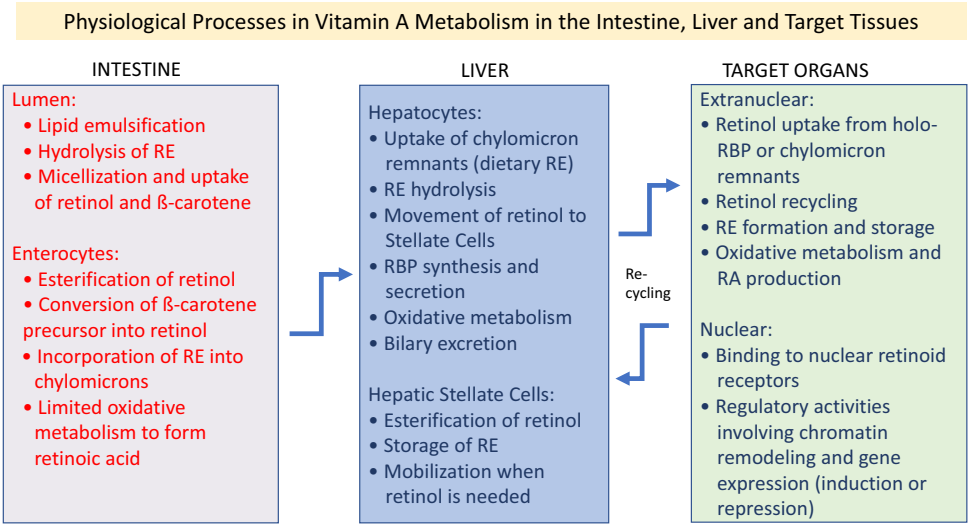


Fig. 3 Legend to **Fig. 3**. Major physiological processes in the intestine, liver and various target organs, which include the skin, immune, respiratory, reproductive and immune systems, and epithelial cells throughout the body. RE, retinyl ester, RBP, retinol-binding protein.

typically occur in most meals. Once released, the retinol molecules diffuse into the enterocytes and are bound to CRBP-II, then chaperoned to the esterifying enzyme lecithin:retinol acyltransferase (LRAT) before becoming incorporated as retinyl esters within the lipid core of the chylomicrons, the intestinally-produced lipoproteins that transport newly absorbed dietary fat to the circulation. Animals lacking LRAT absorb vitamin A inefficiently. Usually, about 70–90% of intestinal vitamin A is absorbed, and this fraction is not significantly down-regulated when vitamin A consumption increases. Intakes of vitamin A that are acutely very high or chronically excessive are still highly absorbable and thus can result in hypervitaminosis A (toxicity). Conversely, conditions that cause fat malabsorption may reduce the bioavailability of vitamin A.

Beta-Carotene is absorbed less efficiently than retinol, on the order of only 9–22%, and, moreover, efficiency falls as the ingested dose increases. Like preformed vitamin A, the provitamin A carotenoids also require dietary fat for their absorption, but, unlike retinyl esters, they are absorbed into the enterocytes intact and then undergo metabolism within these cells. Beta-Carotene is first cleaved at its central 15, 15′ double bond by carotene monooxygenase (BCO-I), generating retinal which is then rapidly metabolized to retinol; CRBP-II facilitates this process as well as the esterification of retinol by LRAT (Von Lintig, 2012). An analogous enzymatic process involving acentric cleavage of beta-carotene yields a variety of products referred to as apo-carotenals, only some of which may be converted into retinol. About one-third of beta-carotene in the human small intestine escapes cleavage and is absorbed intact, becoming incorporated unchanged into chylomicrons and then exchanging onto serum lipoproteins that transport carotenoids in plasma. The central cleavage enzyme BCO-1 is feedback regulated at the transcriptional level by retinoic acid such that, when vitamin A is abundant and its metabolism generates higher levels of retinoic acid, the retinoic acid activates nuclear receptor signaling which down-regulates the production of BCO-1, reducing beta-carotene cleavage and the production of retinol (von Lintig et al., 2020).

Retinoic acid in the intestine

Retinoic acid is produced in small quantities during the absorption of vitamin A and beta-carotene. It is either absorbed into the portal vein and transported to the systemic circulation or used locally in its gene-regulatory capacity by epithelial and immune system cells within the intestinal epithelium and submucosal layers.

Hepatic vitamin A uptake, storage and release

The liver plays no fewer than 5 essential roles in the metabolism of vitamin A: (1) uptake of the dietary retinyl esters carried in partially metabolized chylomicrons (remnants); (2) conversion of dietary retinyl esters into new retinyl esters for storage in hepatic stellate cells (HSC); (3) production of RBP and release of holo-RBP to the circulation; (4) oxidative catabolism of excess retinoids (often taken up from plasma) into polar forms that can be excreted; (5) biliary excretion of various retinoids into the intestinal tract for removal from the body (Ross and Harrison, 2014; Blaner et al., 2016). Firstly, the retinyl esters contained in chylomicron remnants are cleared rapidly from plasma, with the greatest uptake into liver parenchymal cells (hepatocytes) (**Fig. 3**), wherein the ester bond is hydrolyzed and after which retinol is transferred into hepatic HSC. The HSC are enriched in CRBP-I which facilitates the esterification of retinol by LRAT, forming the major products retinyl palmitate and retinyl stearate. These esters, along with other lipids, are stored in the numerous small lipid droplets present within the cytoplasm of the HSC. When vitamin A nutriture is adequate, greater than 90% of the body's total vitamin A resides in HSC as retinyl esters. Interestingly, a similar process involving smaller quantities of retinol exists in several extrahepatic tissues that also contain cells resembling HSCs, suggesting the presence of a network of vitamin A-storing cells throughout the body.

When peripheral tissues need retinol, retinyl esters in HSCs are hydrolyzed. Exactly which metabolic signals initiate this hydrolytic process is not clear. However, it has been shown that the liberated retinol combines with newly synthesized RBP and is secreted through the Golgi apparatus into plasma. Either shortly before or after the secretion of holo-RBP into plasma, the complex binds noncovalently with a tetramer of TTR to form the larger holo-RBP-TTR transport complex described earlier. The liver's role in oxidative metabolism will be considered later.

Plasma concentrations

In the fasting condition, holo-RBP is the major form of circulating vitamin A. The concentration of plasma retinol is maintained within a narrow range of around about 2 μmol/L in adults, and somewhat lower in children. RBP is normally 80–90% saturated with retinol. Apo-RBP (i.e., RBP without retinol) has reduced affinity for TTR and the free apo-RBP protein is readily filtered in the kidneys and lost in urine.

In the postprandial state, a significant proportion of total plasma retinol is present as retinyl ester in the lipid-rich chylomicrons described above, or associated with higher-density lipoproteins. This postprandial peak in concentration of retinyl esters depends directly on the amount of dietary vitamin A that was consumed. They are typically clearly within 6–8 h after their consumption. When blood specimens for clinical analysis are collected in the fasting state, postprandial retinyl esters will be missed.

Several other retinoids are present in plasma in minor concentrations, including retinoic acid (all-*trans* and some *cis* isomers), ring-oxidized forms of retinoids, and glucuronides thereof.

Relationship of plasma retinol to liver retinol concentration

The quantitative relationship between plasma vitamin A and liver vitamin A has been studied in humans and animals (Tanumihardjo et al., 2016). As noted above, nearly steady-state levels of plasma retinol are generally maintained even though hepatic vitamin A concentrations often vary widely. In general, the body's vitamin A reserves are sufficient to maintain plasma retinol at its normal level; however, if intake is not sufficient and liver vitamin A reserves fall below a concentration of about 20–30 μg retinol/g liver, then the secretion of holo-RBP from the hepatocytes into plasma is compromised, and apo-RBP levels fall. This decline occurs slowly, over weeks or months, and continues until nearly all of the vitamin A in liver is used up. As plasma retinol falls, a state of marginal vitamin A deficiency occurs first, which may be detected biochemically but isn't apparent otherwise. Clinical signs of vitamin A deficiency take longer to appear and will begin to become manifest after the point where liver reserves have been exhausted and plasma retinol concentrations have become too low to maintain the normal functions of vitamin A in peripheral tissues.

If retinol is consumed again, either by consuming vitamin A-rich foods or by taking a vitamin A-containing supplement, retinol will be rapidly absorbed, taken up into the liver, and secreted back into plasma as holo-RBP. Whereas depletion is slow, repletion is rapid such that plasma retinol levels return to near normal in a matter of hours.

Metabolic disturbances that affect plasma retinol

Several disturbances are known to affect the levels of retinol and RBP in plasma. These include intestinal conditions causing fat and fat-soluble vitamin malabsorption; deficiencies of protein or calories that are severe enough to affect the synthesis of RBP and TTR in liver; and inflammation, wherein the cytokines associated with tissue damage, including interleukin-6, decrease the synthesis of RBP and TTR in the liver and therefore plasma levels fall. In inflammatory states, laboratory values for plasma retinol must be interpreted cautiously because it may appear that the person is deficient in vitamin A, whereas the true cause of hyporetinolemia is inflammation (Rosales et al., 1996). Other causes of impaired uptake, transport and metabolism of vitamin A include excessive alcohol consumption, which affect metabolism in the liver and thereby the delivery of retinol to target tissues by RBP. Although it is much more rare, inadvertent exposure to environmental toxicant, has also been shown to significantly reduce plasma retinol concentrations (Grignard et al., 2020).

Hypervitaminosis A

This term is used to describe the condition caused by excessive consumption, which, if high enough, causes overt toxicity. Hypervitaminosis A is dose-dependent and may develop slowly due to ingestion of too much vitamin A over a period of time, or it may be very acute after intake of extremely large doses (Tanumihardjo et al., 2016). Biochemical findings include liver levels >300 μg retinol/g liver, and the presence of retinyl esters in plasma lipoproteins in the fasting state. Manifestations of severe hypervitaminosis A are noted below. In order to provide dietary guidance to the public for the prevention of nutrient-related toxicity, the Tolerable Upper Intake Level, known as the UL, has been established (see later section on Recommended Dietary Allowances).

Vitamin A kinetics

RBP and TTR each have a relatively short half-life in the circulation, ~0.5 and 2–3 d, respectively. Due to their rapid turnover, the maintenance of stable plasma concentrations requires continuous synthesize and secretion of holo-RBP by the liver.

Based on kinetic studies, each molecule of retinol circulates through the plasma compartment several times before becoming irreversibly degraded (see **Tissue retinoid metabolism**, below). More retinol passes through plasma per day than is actually degraded, indicative of extensive recycling. In contrast to retinol, RBP itself is not significantly recycled, implying that new RBP molecules must be synthesized in order for the continued recycling of retinol from peripheral tissues back to liver. Several extrahepatic tissues including kidney and adipose, contain RBP mRNA at levels about ~5–10% that in liver, and thus may be active in RBP production. The kidney plays a critical role in the recycling and conservation of retinol as, following the glomerular filtration of holo-RBP the protein binds to megalin, the multi-ligand receptor mentioned earlier which is present on the surface of renal tubular epithelial cells. After uptake, retinol crosses the epithelium and is returned to plasma. This conservation mechanism is important for efficient utilization of vitamin A, for when megalin is lacking RBP and retinol are lost in the urine (Nielsen et al., 2016).

Adipose tissue may be another source of plasma RBP. Additionally, adipose-derived RBP has been reported to have adipokine-like activity. Overall, numerous studies have shown that the body is very efficient at conserving retinol. In contrast, it is relatively inefficient in degrading and eliminating excess retinoids.

Tissue retinoid metabolism

Vitamin A is distributed on holo-RBP to tissues throughout the body (Ross and Harrison, 2014). The uptake mechanism is best elucidated for the retina, where a receptor for holo-RBP known as Stra6 is present on the plasma membrane surface of retinal pigment epithelial (RPE) cells. Some other tissues also express the Stra6 gene. However, as the liver does not appreciably express Stra6 yet is involved in retinol recycling, this suggests there must be other uptake mechanisms for the reuptake of retinol into the liver.

Numerous organs play a role in retinoid uptake and oxidation (Kedishvili, 2016). As noted earlier, the active forms of vitamin A are 11-*cis*-retinal (in the eye only) and all-*trans*-retinoic acid, occurring in many tissues. Several enzymes of the alcohol dehydrogenase, the short-chain dehydrogenase/reductase, and aldehyde dehydrogenase gene families participate in these oxidative reactions. Retinoic acid is generally present in tissues at nanomolar concentrations, far lower than that of retinol–suggesting that its formation is closely regulated and that much of the retinol taken up into peripheral tissues is released back to plasma and recycled. The half-life of retinoic acid is short, on the order of minutes to a few hours. Some tissues, such as brain, derive most of their retinoic acid through their own metabolism while other tissues such as liver obtain retinoic acid mostly by uptake from plasma.

As noted above, polar metabolites of both retinol and retinoic acid are formed by oxidation of their ring structure, usually at carbon 4, to yield hydroxy and oxo metabolites. The process is partially auto-regulated by retinoic acid itself, as retinoic acid through its nuclear receptors induces the expression of cytochrome P450 enzymes of the CYP26 family, which are retinoic acid 4-hydroxylases (Isoherranen and Zhong, 2019). Normally, tissue retinoid levels are maintained at low concentrations by the balance between retinoid production (activation) and retinoid oxidation (inactivation). Nevertheless, it is possible for homeostatic mechanisms to be overwhelmed when vitamin A or other retinoids are consumed in excess (see section on Hypervitaminosis A and retinoid toxicity).

Physiological actions

Vision

The retina functions as the light-sensing organ of the body (**Table 1**). It is organized into several specific cell layers. The layer of RPE cells proximal to the choroid capillaries takes up retinol from RBP. Some of this retinol is esterified by LRAT into retinyl esters for local storage in the RPE, where these esters will ultimately be used in hydrolysis and isomerization processes to generate11-*cis*-retinal (Saari, 2016). The rod and cones cells comprise the next layer of photoreceptor cells, just in front of the RPE cells. The photoreceptor cells contain the highest concentrations of 11-*cis*-retinal, which is bound to specific opsin proteins. Axonal-like processes projecting from the photoreceptor cells are connected to those from the next cell layer comprised of neuronal ganglion cells that signal to the brain. As 11-*cis*-retinal is needed by the photoreceptor cells, it is transported from the RPE to the photoreceptor cells by a novel interstitial retinoid-binding protein, IRBP. After uptake by the rods and cones, 11-*cis*-retinal combines covalently with opsin in rods to generate the visual pigment rhodopsin, and, similarly, with iodopsin proteins in cones to form red-, green- and blue–sensitive pigments. These photoreceptor pigments (e.g., rhodopsin) are very efficient in absorbing visible light. The outer segment portion of the rod cell is densely packed with membrane disks that contain some 10^8 molecules of rhodopsin per cell. Much research has been conducted to elucidate a specific visual cycle which constitutes a recycling process in which 11-*cis*-retinal is converted by photons of light into all-*trans*-retinal, which then is transported back to the RPE and regenerated by an enzymatic cascade into 11-*cis*-retinal again. This regeneration process is referred to as the visual cycle, and is related to the physiological process of dark adaptation. Because dark adaptation is relatively slow (on the order of minutes) as compared to the very rapid process of photoisomerization, retinyl esters stored in the RPE are quickly recruited through hydrolysis and isomerization to generate new 11-*cis*-retinal, which is passed by IRP to the rods where rhodopsin is regenerated. However, when the RPE's supply of retinyl esters is low, then the visual cycle slows. The clinical outcome of slow visual pigment regeneration is referred to as night blindness, which is one of the first clinical signs of vitamin A deficiency (see Chapter on Hypovitaminosis A).

Vitamin A and the cornea

The cornea of the eye is composed of epithelial cells that also require vitamin A, but in this case, it is in the form of retinoic acid for maintenance of normal cell differentiation (**Table 1**). Although the cornea is avascular, holo-RBP is present in the lacrimal glands and tears, and this vitamin A is likely to provide retinol for the local formation of retinoic acid. In states of prolonged retinoid deficiency, the mucin-secreting goblet cells within the cornea lose their function, resulting in corneal xerosis and Bitôt spots (foamy

Table 1 Forms and functions of vitamin A in vision.

	Part of eye	
Characteristic	Retina	Cornea
Major active form of retinoid	11-*cis*-retinal	All-*trans*-retinoic acid
Major cell types	Photoreceptors (rods and cones); Retinal pigment epithelium	Epithelial cells and mucus-secreting goblet cells
Major function	Light sensing and neural transmission	Barrier and antimicrobial
Signs of retinoid deficiency	Night blindness	Xerosis, Bitôt spots

deposits of cells and bacteria, usually at the outer quadrants of the eye), which indicate severe vitamin A deficiency (see chapter on Hypovitaminosis A). Persons with these signs need repletion with vitamin A immediately to prevent further damage resulting in life long blindness.

Functions in cell differentiation

The entire body's epithelial tissues (skin, respiratory tract, immune system, reproductive organs, etc.) are sensitive to vitamin A, both deficiency and excess (Tanumihardjo et al., 2016). The systemic effects of vitamin A deficiency include dryness of the skin (follicular hyperkeratosis), loss of goblet cells in the trachea and respiratory tract, as well as the cornea as mentioned earlier; the morphological appearance is often of a flattening, dry epithelia referred to as squamous metaplasia, (sometimes with keratinization). The hematopoietic system is also affected which results in multiple impairments in the immune response (see below). Reproduction is abnormal, including impaired spermatogenesis in vitamin A deficient males and abnormal fetal development, which is manifest in both vitamin A deficiency and excess. During embryonic development, retinoids are required from the post-gastrulation stage and formation of the neural tube, continuing throughout the stages of organogenesis (Ghyselinck and Duester, 2019; Schubert and Gibert, 2020). Retinoic acid concentration is very tightly controlled as it is a key determinant of the expression of several developmentally important genes, particularly *Hox* genes that are critical for the formation of the body pattern, and genes necessary for formation of the circulatory and nervous systems. Both a deficiency of vitamin A and an excess of vitamin A and an excess of either natural or synthetic retinoids are known to cause developmental defects of the head, face, limbs, and organ systems.

Immunity

Well-regulated retinoid signaling is necessary for the maintenance of normal immune cell populations and immune responses (Larange and Cheroutre, 2016). A lack of vitamin A results in reductions in mucosal epithelial lymphocyte numbers and functions, reduced T cell counts and altered T-cell subsets, and abnormalities in natural killer cells and macrophages. The production of the cytokines that are required for the normal regulation of T-cell immunity and B-cell antibody production becomes dysregulated. Often, antibody responses are low. Mucosal immunity affecting the intestine is impaired and gut infections are altered (Chai et al., 2021). The function of the lungs and body's ability to generate antibody responses are impaired (Penkert et al., 2020). Although specific studies of vitamin A deficiency and SARS-2/Covid-19 infection are not yet available, the similarities between infections caused by the SARS-2 virus and measles virus have been reviewed, and the data suggest that adequate vitamin A is an important factor for maintaining the mucosal and systemic immune functions necessary to combat both of these viruses (Stephensen and Lietz, 2021). As noted earlier, inflammation, which is common in many infectious and non-infectious diseases, is itself a cause of hyporetinolemia.

Dietary sources and nutritional equivalency

Preformed vitamin A is present at highest concentration in liver and fish oils; it is present at lower concentrations in non-organ meats and eggs. Vitamin A, usually as retinyl palmitate for improved stability, is also used in food fortification and supplement preparation. In the US, about two-thirds of vitamin A consumption is in the form of preformed vitamin A consumed in foods of animal origin, and the rest is consumed as carotenoids. However, the proportions vary widely depending on dietary patterns. In low-income countries, provitamin A is the major source. Importantly, a variety of dietary patterns can provide adequate vitamin A nutriture.

Units of nutritional activity

Because vitamin A exists in multiple forms and since the efficiency of utilization of some of its forms–including carotenoids–is lower than that of preformed vitamin A, it is necessary to have a unit that expresses *the total bioactivity of vitamin A*. The currently accepted nutritional equivalents, established in 2001 by the Institute of Medicine (2001), are expressed in terms of Retinol Activity Equivalents (RAE), where 1 μg RAE has equivalent activity to 1 μg of all-*trans*-retinol. It is now considered that, on average, carotenoids must be ingested in the following amounts to provide the equivalent nutritional value of 1 μg of all-*trans*-retinol (1 RAE):

* 2 μg of supplemental beta-carotene (in an oily, easily absorbed solution)
* 12 μg of beta-carotene, or 24 μg of α-carotene or beta-cryptoxanthin in fruits and vegetables (due to association of carotenoids with the food matrices, and therefore reduced digestibility)

A still older unit, the international unit (IU), is sometimes still in use. This unit does not account for newer knowledge about carotenoid bioavailability and equivalency. One IU is defined as 0.3 μg of all-*trans*-retinol.

Recommended Dietary Allowances (RDA) and tolerable upper intake levels (UL) for vitamin A

RDA for the US and Canada were updated in 2001 (Institute of Medicine, 2001). Guidelines were also established for a Tolerable Upper Intake Level (UL). The RDA and UL for vitamin A for various life stages are listed in **Table 3**. The UL is defined as the highest intake of a nutrient that is likely to pose no risk of adverse health effects in nearly all healthy individuals. It is important to recognize that for vitamin A the UL applies specifically to *preformed* vitamin A (e.g., retinol, but not carotenoids), including intake from foods, fortified foods, and supplements (**Table 2**). The IOM committee set the UL based on epidemiological evidence of increased risk of

Table 2 Food sources of vitamin A.

Food	*Micrograms (mcg) RAE per serving*	*Percent DV**
Beef liver, pan fried, 3 ounces	6582	731
Sweet potato, baked in skin, 1 whole	1403	156
Spinach, frozen, boiled, ½ cup	573	64
Pumpkin pie, commercially prepared, 1 piece	488	54
Carrots, raw, ½ cup	459	51
Ice cream, French vanilla, soft serve, 1 cup	278	31
Cheese, ricotta, part skim, 1 cup	263	29
Herring, Atlantic, pickled, 3 ounces	219	24
Milk, fat free or skim, with added vitamin A and vitamin D, 1 cup	149	17
Cantaloupe, raw, ½ cup	135	15
Peppers, sweet, red, raw, ½ cup	117	13
Mangos, raw, 1 whole	112	12
Breakfast cereals, fortified with 10% of the DV for vitamin A, 1 serving	90	10
Egg, hard boiled, 1 large	75	8
Black-eyed peas (cowpeas), boiled, 1 cup	66	7
Apricots, dried, sulfured, 10 halves	63	7
Broccoli, boiled, ½ cup	60	7
Salmon, sockeye, cooked, 3 ounces	59	7
Tomato juice, canned, ¾ cup	42	5
Yogurt, plain, low fat, 1 cup	32	4
Tuna, light, canned in oil, drained solids, 3 ounces	20	2
Baked beans, canned, plain or vegetarian, 1 cup	13	1
Summer squash, all varieties, boiled, ½ cup	10	1
Chicken, breast meat and skin, roasted, ½ breast	5	1
Pistachio nuts, dry roasted, 1 ounce	4	0

*DV = Daily Value. FDA developed DVs to help consumers compare the nutrient contents of foods and dietary supplements within the context of a total diet. The DV for vitamin A is 900 mcg RAE for adults and children age 4 years and older, where 1 mcg RAE = 1 mcg retinol, 2 mcg beta-carotene from supplements, 12 mcg beta-carotene from foods, 24 mcg alpha-carotene, or 24 mcg beta-cryptoxanthin. FDA does not require food labels to list vitamin A content unless vitamin A has been added to the food. Foods providing 20% or more of the DV are considered to be high sources of a nutrient, but foods providing lower percentages of the DV also contribute to a healthful diet.
Source: Office of Dietary Supplements, https://ods.od.nih.gov/factsheets/VitaminA-HealthProfessional/#h3 (Last accessed March 18, 2022).

Table 3 Recommended Dietary Allowances (RDA) for vitamin A in micrograms (μg) Retinol Activity Equivalents (RAE) and International Units (IUs), and Tolerable Upper Intake Levels (UL, μg retinol/day), for children and adults.

Age (years)	*Children*	*Men*	*Women*	*Pregnancy*	*Lactation*
1–3	300 μg or 1000 IU				
4–8	400 μg or 1333 IU				
9–13	600 μg or 2000 IU				
14–18		900 μg or 3000 IU	700 μg or 2330 IU	750 μg or 2500 IU	1200 μg or 4000 IU
19+		900 μg or 3000 IU	700 μg or 2330 IU	770 μg or 2565 IU	1300 μg or 4335 IU
UL, μg retinol/day					
1–3	600 μg				
4–8	900 μg				
9–13	1700 μg				
14–18		3000 μg	2800 μg	2800 μg	2800 μg
19+		3000 μg	3000 μg	3000 μg	3000 μg

RDA, μg RAE/day.
Source: Office of Dietary Supplements, https://ods.od.nih.gov/factsheets/VitaminA-HealthProfessional/#h3 (Last accessed March 18, 2022).

birth defects with intakes >10,000 IU/day, as well as on minimizing the risk of liver damage due to hypervitaminosis A. For children, the RDA and UL values were scaled down based on body weight.

For consumer information purposes, the labels present on food items use the term % Daily Value (%DV), which is a less quantitative means to conveniently compare the nutritional value of various foods. It provides the relative contribution that a serving of that food makes to daily vitamin A intake, assuming a reference intake. Unlike the RDA and UL, the %DV is not scaled to different values for adults and children.

Conditions of excess

Hypervitaminosis A

Hypervitaminosis A is a rare but serious condition, where the severity depends on dose and frequency of intake (Tanumihardjo et al., 2016). It can arise acutely, within hours or days, after consumption of very large amounts of preformed vitamin A, causing serious illness and death. Or, it can develop slowly due to a persistent intake of lower but still excessive amounts of preformed vitamin A. Carotenoids are not a cause of vitamin A toxicity. Most clinical cases of toxicity are the result of excessive intake of retinol-containing supplements, or in patients taking prescription retinoids. The clinical signs of vitamin A toxicity include nausea and vomiting, headache, dizziness, blurred vision, lack of muscular coordination, abnormal liver function, and pain in weight-bearing bones and joints. The offspring of women of child-bearing age who consume excess vitamin A (over the UL on a regular basis), or use retinoids for therapeutic reasons, are at increased risk of birth defects. For this reason, the Food and Drug Administration (FDA) closely regulates the prescription of medicinal retinoids (FDA, 2012). Similarly, dietary supplements that contain amounts of vitamin A over the UL, which may be advertised as having health effects, should be avoided by women of child-bearing age, as the safety of intakes about the UL is not certain. Diets high in vitamin A-rich foods, such as liver and sausages prepared with organ meats, if consumed frequently, could pose a risk to developing fetus. As teratogenic effects on the fetus occur very early in development, even before pregnancy may be detected, women capable of becoming pregnant should avoid high intakes of these foods (Van den Berg et al., 1996).

There is no antidote for vitamin A toxicity, and little that can be done besides eliminating the intake of vitamin A or use of retinoids, and waiting for the body to clear the excess to the extent possible. Therefore, prevention is imperative. The overconsumption of preformed vitamin A, including that in supplements, should be avoided. As noted above, the UL, which applies to only preformed vitamin A, is meant to provide guidance to the public and health care workers for what is a safe upper level of intake on a chronic basis.

Excessive carotene intake

Individuals who consume large amounts of carotenoid-rich foods or juices may develop yellowing of the skin (carotenodermia), especially in fatty tissues and the palms of the hands. This condition is considered benign and the yellow color will gradually subside after intake of carotene is reduced to a normal level. Regarding beta-carotene supplementation and cancer, although results have been somewhat controversial, the data from large clinical trials of high-dose supplemental beta-carotene in men who smoked showed adverse effects, with higher incidence of lung cancer in the supplement arm of the trial. It is therefore prudent for persons who smoke to avoid taking extra supplemental beta-carotene. Based on these studies, the "AREDS" formula to reduce risk of age-related macular degeneration was modified to eliminate beta-carotene, substituting other non-provitamin A carotenoids instead (Gorusupudi et al., 2017). Diets high in beta-carotene have not been shown to have adverse effects, and it is difficult to obtain very high intakes from dietary sources. Additionally, as noted above, the efficiency of absorption of dietary beta-carotene is physiologically regulated.

Vitamin A supplementation in public health

Even at the present time, vitamin A deficiency is a significant public health concern, focused in low- and middle-income countries. A report (Stevens et al., 2015) described that despite improvements in vitamin A status overall during the past 3 decades, vitamin A deficiency remains prevalent in south Asia and sub-Saharan Africa. In public health settings such as in low-income parts of the world where children are at risk of developing vitamin A deficiency, large doses of vitamin A are often administered on a periodic basis as prophylaxis, or vitamin A may be administered therapeutically in cases of severe illnesses such as measles (Bello et al., 2011). Vitamin A supplementation in children ages 6 months to 5 years, living in at-risk settings, has significantly reduced disease severity and child mortality (Sommer, 2008) (see Chapter on Hypovitaminosis A).

Conclusion/summary/outlook

As reviewed in this chapter, the human body makes use of both preformed vitamin A (retinol and its esters) and provitamin A (mainly beta-carotene) to produce the physiologically necessary forms of vitamin A, particularly 11-*cis*-retinal and all-*trans*-retinoic acid. Through the extensive processing of these dietary precursors, a vitamin A-adequate diet can be achieved through a variety of dietary patterns. These processing reactions begin during intestinal metabolism, when retinol is absorbed and beta-carotene is absorbed and then cleaved. They continue in the liver and in the many peripheral organs that utilize vitamin A, mainly through reactions that are part of oxidative metabolism. The transport of retinol between tissues is mediated by holo-RBP, the retinol component of which is extensively recycled. The kidneys are especially important in recapturing retinol from the renal filtrate and returning it to plasma. Nearly all cells of the body contain nuclear retinoic acid receptors that target the actions of vitamin A in gene expression. The retina is unique in utilizing 11-*cis*-retinal, which, in protein-bound form as rhodopsin, is required for vision. Vitamin A deficiency is a public health problem due to poor nutrition, linked to poor immune outcomes and loss of vision.

In contrast, vitamin A excess occurs much less commonly, is often due to excessive use of supplements, and can be very acute. Nutritional recommendations such as the RDA are set to provide an adequate allowance of vitamin A for all its physiological functions and to provide for some storage in tissues, whereas the UL is set as public health guidance to avoid excessive intake and the occurrence of hypervitaminosis A and, in its most severe form, vitamin A toxicity.

See Also: Carotenoids: Health effects; Carotenoids: Chemistry, sources and physiology

References

Bello, S., Meremikwu, M.M., Ejemot-Nwadiaro, R.I., Oduwole, O., 2011. Routine vitamin A supplementation for the prevention of blindness due to measles infection in children. Cochrane Database Syst. Rev. CD007719, 2011.

Blaner, W.S., Li, Y., Brun, P.J., Yuen, J.J., Lee, S.A., Clugston, R.D., 2016. Vitamin A absorption, storage and mobilization. Subcell. Biochem. 81, 95–125.

Chai, Z., Lyu, Y., Chen, Q., Wei, C.-H., Snyder, L.M., Weaver, V., Sebastian, A., Albert, I., Li, Q., Cantorna, M.T., Ross, A.C., 2021. RNAseq studies reveal distinct transcriptional response to vitamin A deficiency in small intestine versus colon, uncovering novel vitamin A-regulated genes. J. Nutr. Biochem. 98, 108814.

Dawson, M.I., 2000. The importance of vitamin A in nutrition. Curr. Pharm. Des. 6 (3), 311–325.

FDA, 2012. The iPLEDGE Program. The Guide to Best Practices for the iPLEDGE Program. Reference ID: 3136456). https://www.fda.gov/media/80023/download. (Accessed 2 September 2021).

Ghyselinck, N.B., Duester, G., 2019. Retinoic acid signaling pathways. Development 146 (13), dev167502, 4.

Gorusupudi, A., Nelson, K., Bernstein, P.S., 2017. The Age-Related Eye Disease 2 Study: micronutrients in the treatment of macular degeneration. Adv. Nutr. 8 (1), 40–53, 17.

Grignard, E., Hakansson, H., Munn, S., 2020. Regulatory needs and activities to address the retinoid system in the context of endocrine disruption: the European viewpoint. Reprod. Toxicol. 93, 250–258.

Honarbakhsh, M., Ericsson, A., Zhong, G., Isoherranen, N., Zhu, C., Bromberg, Y., Van Buiten, C., Malta, K., Joseph, L., Sampath, H., Lackey, A.I., Storch, J., Vetriani, C., Chikindas, M.L., Breslin, P., Quadro, L., 2021. Impact of vitamin A transport and storage on intestinal retinoid homeostasis and functions. J. Lipid Res. 13 (62), 100046.

Institute of Medicine, 2001. Dietary reference intakes for vitamin A, vitamin K, arsenic, boron, chromium, copper, iodine, iron, maganese, molybdenum, nickel, silicon, vanadium, and zinc. In: Food and Nutrition Board. National Academy Press, Washington, DC.

Isoherranen, N., Zhong, G., 2019. Biochemical and physiological importance of the CYP26 retinoic acid hydroxylases. Pharmacol. Ther. 204, 107400.

Kedishvili, N.Y., 2016. Retinoic acid synthesis and degradation. Subcell. Biochem. 281, 127–161.

Larange, A., Cheroutre, H., 2016. Retinoic acid and retinoic acid receptors as pleiotropic modulators of the immune system. Annu. Rev. Immunol. 20 (34), 369–394.

Nielsen, R., Christensen, E.I., Birn, H., 2016. Megalin and cubilin in proximal tubule protein reabsorption: from experimental models to human disease. Kidney Int. 89 (1), 58–67.

Penkert, R.R., Cortez, V., Karlsson, E.A., Livingston, B., Surman, S.L., Li, Y., Ross, A.C., Schultz-Cherry, S., Hurwitz, J.L., 2020. Vitamin A corrects tissue deficits in diet-induced obese mice and reduces influenza infection after vaccination and challenge. Obesity 28 (9), 1631–1636.

Rosales, F.J., Ritter, S.J., Zolfaghari, R., Smith, J.E., Ross, A.C., 1996. Effects of acute inflammation on plasma retinol, retinol-binding protein, and its mRNA in the liver and kidneys of vitamin A-sufficient rats. J. Lipid Res. 37 (5), 962–971.

Ross, A.C., Harrison, E.H., 2014. Vitamin A and carotenoids: nutritional aspects. In: Zempleni, J., Suttie, J.W., Gregory III, J.F., Stover, P.J. (Eds.), Handbook of Vitamins, fifth ed. CRC Press, Boca Raton, FL, pp. 2–49.

Saari, J.C., 2016. Vitamin A and vision. Subcell. Biochem. 81, 231–259.

Schubert, M., Gibert, Y., 2020. Retinoids in embryonic development. Biomolecules 10 (9), 1278.

Sommer, A., 2008. Vitamin A deficiency and clinical disease: an historical overview. J. Nutr. 138, 1835–1839.

Stephensen, C.B., Lietz, G., 2021. Vitamin A in resistance to and recovery from infection: relevance to SARS-CoV2. Br. J. Nutr. 1–10.

Stevens, G.A., Bennett, J.E., Hennocq, Q., Lu, Y., De-Regil, L.M., Rogers, L., et al., 2015. Trends and mortality effects of vitamin A deficiency in children in 138 low-income and middle-income countries between 1991 and 2013: a pooled analysis of population-based surveys. Lancet Glob. Health 3, e528–e536.

Tanumihardjo, S.A., Russell, R.M., Stephensen, C.B., Gannon, B.M., Craft, N.E., Haskell, M.J., Lietz, G., Schulze, K., Raiten, D.J., 2016. Biomarkers of nutrition for development (BOND)-Vitamin A review. J. Nutr. 146 (9), 1816S–1848S.

Van den Berg, H., Hulshof, K.F., Deslypere, J.P., 1996. Evaluation of the effect of the use of vitamin supplements on vitamin A intake among (potentially) pregnant women in relation to the consumption of liver and liver products. Eur. J. Obstet. Gynecol. Reprod. Biol. 66, 17–21.

von Lintig, J., Moon, J., Lee, J., Ramkumar, S., 2020. Carotenoid metabolism at the intestinal barrier. Biochim. Biophys. Acta Mol. Cell Biol. Lipids 1865 (11), 158580.

Von Lintig, J., 2012. Provitamin A metabolism and functions in mammalian biology. Am. J. Clin. Nutr. 96 (5), 1234S–1244S.

Wolf, G., 1996. A history of vitamin A and retinoids. FASEB J. 10 (9), 1102–1107.

Vitamin B_{12}: Physiology, dietary sources, and requirements

R. Green, University of California, Davis, CA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an reproduction of R. Green, Vitamin B12: Physiology, Dietary Sources, and Requirements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 351–356, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00056-8, with revisions made by the Editor.

Introduction 482
Biochemistry and metabolic functions 482
Physiology 484
Causes, mechanisms, and effects of cobalamin deficiency 485
Diagnosis of cobalamin deficiency 486
Inborn errors of cobalamin metabolism 487
Further reading 488

Glossary

Cobalamins Compounds that are collectively known as vitamin B_{12} and contain cobalt in a corrin ring. The cobalt can be attached to one of several sidegroups, including cyanide. Cyanocobalamin is the stable form of the vitamin manufactured for use as supplements and a food fortificant

Macrocytosis When red blood cells (erythrocytes) are larger than normal

Transcobalamins Carrier proteins which bind and transport vitamin B_{12}

Introduction

The cobalamins are a group of closely related and interconvertible compounds with a complex structure that are collectively known by the common name of vitamin B_{12}. Recommended biochemical nomenclature restricts the term "vitamin B_{12}" for the particular form of cobalamin known as cyanocobalamin. All cobalamins belong to the broader family of corrinoids, which share the characteristic of consisting of a planar four-membered pyrrole ring (corrin ring) containing a central cobalt atom. Cobalamins are distinguished from other corrinoids by possessing both alpha (lower) and beta (upper) axial ligands that are attached to the central cobalt atom (Fig. 1). The lower ligand consists of a base (5,6-dimethylbenzimidazole) attached to a sugar (ribose), which in turn is attached to a phosphate and an amino-propyl group that ultimately is tethered back to the corrin ring. In the naturally occurring cobalamins the upper ligand is variably a cyano-, hydroxo-, aquo-, methyl-, or adenosylgroup, giving rise to the correspondingly named chemical forms of the vitamin. Of these, methylcobalamin and deoxyadenosylcobalamin are the forms that function as coenzymes for metabolic reactions. All chemical forms of B12 are highly sensitive to destruction by light; cyanocobalamin, as a synthetic naturally not occurring form, is a more stable form and is therefore used in therapeutic preparations. Hydroxo- or aquocobalamin are intermediates during the synthesis of the coenzyme forms. Other forms including sulphito-, nitrito-, and glutathionyl derivatives of cobalamin have also been described, but their precise role in metabolism is not known.

Biochemistry and metabolic functions

Only two reactions in humans and other animals are known to require cobalamin (Fig. 2). One, isomerization of methylmalonyl CoA, which requires deoxyadenosylcobalamin, catalyzed by the enzyme methylmalonyl CoA mutase and is mitochondrial. The other reaction is the transmethylation of homocysteine by 5-methyl-tetrahydrofolate to methionine, catalyzed by the enzyme methionine synthase (*N5*-methyl homocysteine methyl transferase), which requires methylcobalamin as coenzyme and is located in the cytosol. It is through their essential roles in this important metabolic reaction that cobalamin and folate interact and are linked with respect to their importance in nutrition. In addition, there are major similarities in the effects of their deficiencies in humans.

Considering this "metabolic crossroad" for the two vitamins, it may be pointed out that without adequate supplies of both nutrients, the synthesis of methionine and its derivative *S*-adenosylmethionine (SAM) is disrupted, with consequent profound effects on normal cellular function. Methionine is an essential amino acid and normal supply depends critically on recycling through the remethylation pathway (Fig. 3). Moreover, SAM is the universal methyl donor, essential for more than 100 transmethylation reactions involving amino acid, nucleotide, neurotransmitter, and phospholipid metabolism as well as detoxification reactions.

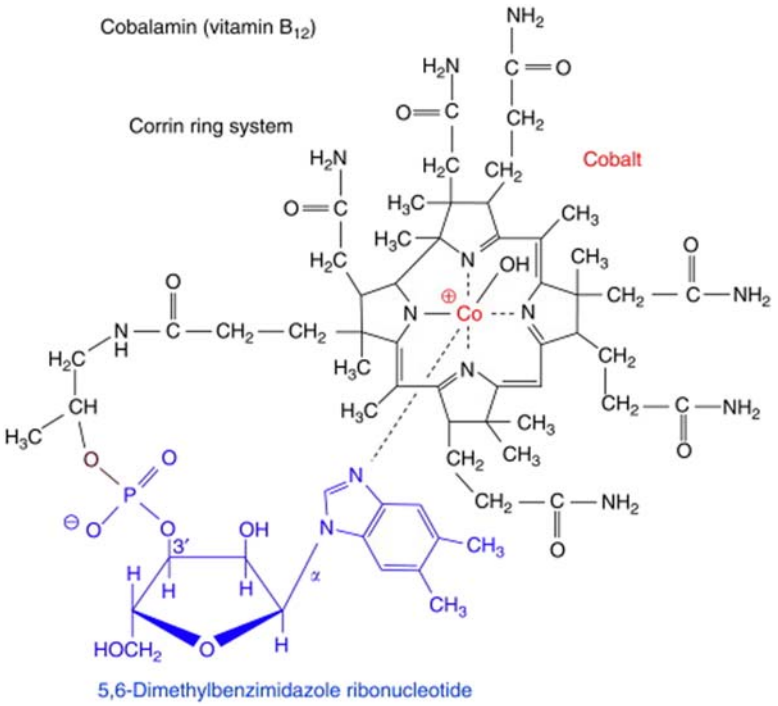


Fig. 1 Chemical structure of cobalamin.

Apart from methionine the other product of the methionine synthase reaction, which is almost completely irreversible, is tetrahydrofolate (THF). This constitutes the first step by which folate enters bone marrow and other cells from plasma, for its conversion into the various intracellular forms of reduced folate containing a series of one-carbon substituents (Fig. 3). The active forms of these folate congeners are all polyglutamated by an enzyme, folate polyglutamate synthetase, which cannot use methyl-THF as substrate. THF is the obligate substrate for polyglutamate addition. Consequently, when the methionine synthase reaction is blocked as a result of cobalamin deficiency, there is "THF starvation". Methyl-THF accumulates in the plasma, although intracellular folate concentrations fall due to failure of formation of the critical intracellular folate polyglutamates because of "methyl-folate trapping." This theory explains the abnormalities of folate metabolism, which occur in cobalamin deficiency (high concentrations of serum folate, low red blood cell folate), and also why the anemia that occurs in cobalamin deficiency will temporarily or partially respond to folic acid in large doses. The explanation of why the serum cobalamin falls in folate deficiency may also be related to impairment of the methionine synthase reaction resulting from reduced formation of methylcobalamin, the predominant circulating form of cobalamin in plasma.

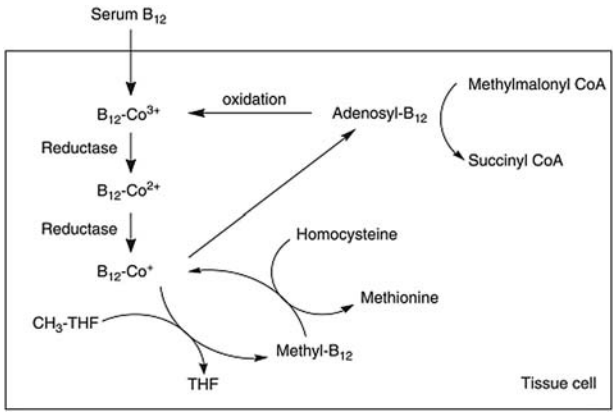


Fig. 2 Mammalian intracellular reactions requiring cobalamin. Conversion of methylmalonyl CoA to succinyl CoA is mitochondrial and requires cobalamin in the form of adenosyl-B_{12}. Conversion of homocysteine to methionine is cytosolic and requires cobalamin in the form of methyl-B_{12}.

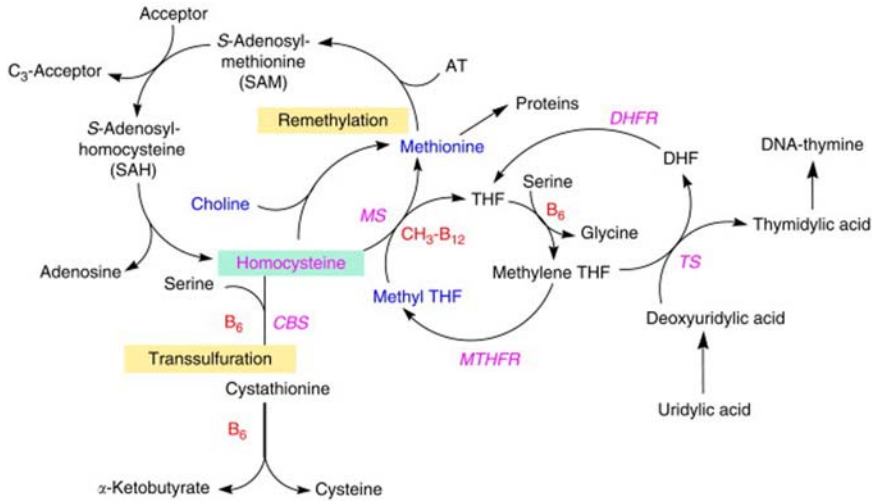


Fig. 3 The remethylation and transsulfuration pathways of homocysteine metabolism showing the role of cobalamin in the form of methyl-B_{12} (CH_3-B_{12}) in the production of methionine. Other B vitamins (B_6, folate) and required enzymes for related pathways are shown. MS, methionine synthese; MTHFR, methylene tetrahydrofolate reductase; TS, thymidylate synthase; DHFR, dihydrofolate reductase; CBS, cystathionine beta synthase; THF, tetrahydrofolate.

Physiology

The recommended dietary allowance (RDA) for cobalamin in adults, proposed by the Food and Nutrition Board of the Institute of Medicine in 1997, is 2.4 μg/d for men and women, 2.6 μg/d and 2.8 μg/d during pregnancy and lactation, respectively. Cobalamins do not occur in plants and are synthesized by certain bacteria, fungi, and algae, which constitute the ultimate source of all cobalamins found in nature. Cobalamins enter the food chain through herbivorous animals that harbor cobalamin-producing microorganisms in their upper gastrointestinal tract (e.g., the "first stomach" of ruminants). Consumption of the meat or products of these animals supplies cobalamin in the diet for other animals. Dietary sources of cobalamin in humans are restricted to meat, poultry, fish, shellfish, eggs, and dairy products. Cobalamin is relatively resistant to destruction by cooking and depends besides others also on the cooking technique (frying leads to the highest losses), unlike the heat labile folates. On account of the exceedingly small daily requirement for cobalamin, in the order of 2–3 μg, and the relatively large body store of the vitamin (3000–5000 μg in individuals in developed countries), complete absence of intake or absorption of cobalamin is preceded by a long lag of up to 3–5 years before depletion of body cobalamin stores reaches a critical point that begins to result in the manifestations of cobalamin deficiency. This is not the case in developing countries where the onset of depletion may be much more rapid because of initially lower stores. Still, complete lack of dietary intake of cobalamin is somewhat rare and occurs only in strict vegans who shun all animal foods, including dairy products and eggs.

Cobalamin absorption is a complex mechanism that consists of several steps and involves several protein chaperones and receptors, defects of which can result in reduced or absent uptake of dietary cobalamin. The related and continuous processes of ingestion, digestion, and absorption of cobalamin comprising the assimilation of the vitamin are arbitrarily divided into six steps. During the first step of mastication and swallowing of food, dietary cobalamin becomes mixed with a binding protein derived from saliva belonging to the family of cobalamin-binding proteins known as haptocorrins. Cobalamin in foods is generally complexed to proteins that must first be digested to release the bioavailable vitamin. In the second step release of cobalamin takes place largely in the stomach, under the influence of gastric hydrochloric acid and proteolytic digestion by pepsin. It is during this process and in the acid environment of the stomach that salivary haptocorrin preferentially binds and protects food cobalamin. Another specific cobalamin-binding protein, known as intrinsic factor, is secreted by the parietal cells of the stomach, but is unable to bind the cobalamin still tightly complexed to haptocorrin. During the third step, which occurs in the duodenum, cobalamin is released from its complex with haptocorrin through the combined effects of pancreatic bicarbonate, which neutralizes the gastric acid, and the proteolytic action of the enzymes trypsin and chymotrypsin that digest haptocorrin and thus enable the binding of the free cobalamin by gastric intrinsic factor. In the fourth step, the intrinsic factor-cobalamin complex, having traversed the full length of the small intestine, arrives at the luminal surface of the terminal ileum. There, it comes in contact with specialized receptors. In the presence of calcium, the complex attaches to the receptor consisting of two distinct proteins coded by the genes cubulin and amnionless that is necessary to complete the assimilation process. Both proteins are essential for the internalization of the intrinsic factor-B_{12} complex through the process of receptor-mediated endocytosis. Through this process, cobalamin together with intrinsic factor is escorted by the receptor and taken into lysosomes. Here the intrinsic factor-cobalamin complex is released and intrinsic factor is degraded through the action of acid hydrolysis by lysosomal peptidases. The final fifth step is poorly understood but first

Table 1 Properties of human plasma cobalamin binding proteins.

	Haptocorrins (TC I + III)	*Transcobalamin (TC II)*
Source	Granulocytes	Endothelial cells
Transport functions	Storage, excretion of B_{12} analogs, antimicrobial	Cellular B_{12} uptake
Binding specificity	Low specificity, binds B_{12} analogs	Binds B_{12} with higher specificity
Membrane receptors	Nonspecific asialoglycoprotein receptors on hepatocytes	Specific receptors on most cells
Saturation	High (mainly "holo")	Low (mainly "apo")
Fraction of total B_{12}	70–90%	10–30%
Plasma clearance	Slow ($t_{1/2}$~10 days)	Rapid ($t_{1/2}$~6 min)
Molecular weight	60,000	38,000–45,000

involves the release from lysosomes and then the metabolism of cobalamin to its methyl and deoxyadenosyl derivatives. It is currently believed that cobalamin enters the plasma in the form of methylcobalamin. The assimilation of food containing vitamin B_{12} is a lengthy process, as evidenced by the 6–8 h taken for orally administered cobalamin to first appear in the plasma and several additional hours for the process to be completed. Recent evidence indicates that cobalamin leaves the cell through an exit portal that is part of the ABC drug transport system, ABCC1 (also known as the Multidrug Resistance Protein, MRP1), present in the basolateral membrane of the intestinal epithelium as well as in other cells.

After cobalamin enters the plasma via the MRP1 it becomes bound to the cobalamin binding protein, transcobalamin (previously known as transcobalamin II to distinguish it from transcobalamins I and III which, together with the salivary cobalamin binding protein and other cobalamin binders present in secretions are now referred to collectively as the haptocorrins). The properties of transcobalamin and the haptocorrins are summarized in **Table 1**. The fraction of cobalamin bound to transcobalamin accounts for only 10–30% of the total plasma cobalamin. The major residual fraction of the plasma cobalamin is attached to haptocorrin. The function of haptocorrins is not known, but rapidly proliferating cells including bone marrow precursors can obtain cobalamin only from transcobalamin. Consequently, the critical fraction of the serum cobalamin is the transcobalamin-bound portion, known as holotranscobalamin. Conditions that alter the amount or distribution of cobalamin on these binding proteins can critically affect delivery and transport. Therefore, conditions that lead to an increase in haptocorrins, such as chronic granulocytic leukemia, characterized by markedly increased numbers of granulocytes (haptocorrins are produced in granulocytes), can give rise to an apparently normal serum cobalamin level even in patients who have severe underlying cobalamin deficiency, through redistribution of reduced body cobalamin stores. Conversely, insufficient holotranscobalamin can result in cobalamin deficiency even if the total serum cobalamin level is apparently normal. This occurs in infants and children affected by congenital transcobalamin deficiency, which is associated with severe megaloblastic anemia. Levels of transcobalamin may be affected by a number of factors. Lowering of holotranscobalamin can result in tissue cobalamin deficiency with a normal total serum cobalamin level. Now that sufficiently sensitive and robust methods are available to measure holotranscobalamin levels in serum, several studies have reported that there is a good inverse correlation between serum holotranscobalamin concentration and serum levels of the metabolites methylmalonic acid and homocysteine, which are described below.

Causes, mechanisms, and effects of cobalamin deficiency

There are several causes of cobalamin deficiency that range in severity and frequency of occurrence. These are summarized in **Table 2**. In general, causes of cobalamin deficiency can be divided into those caused by absent or markedly reduced dietary intake and those caused by malabsorption, either gastric or ileal. The most frequent cause of severe cobalamin deficiency is pernicious anemia, caused by an autoimmune destruction of the gastric mucosa with consequent failure of intrinsic factor production. Less common causes include chemical inactivation and inherited defects in cobalamin absorption or metabolism.

In all situations resulting from impairment of cobalamin absorption, the time to onset of deficiency depends on several factors, including the size of the body store, the extent of impairment of absorption (partial or complete), and, in diseases like pernicious anemia and others affecting the small intestine, the rate of progression of the disease. In general, however, cobalamin deficiency resulting from malabsorption develops sooner than is the case in the dietary deficiency encountered among vegans. This difference may be explained by the existence of a considerable enterohepatic recirculation of cobalamin. Biliary cobalamin is efficiently reabsorbed in vegans compared with patients with pernicious anemia or other forms of malabsorption, because the intrinsic factor-dependent mechanism is intact.

Deficiency of cobalamin, when severe, affects all rapidly growing (DNA-synthesizing) tissues. After the marrow, the next most affected tissues are the epithelial cell surfaces of the gastrointestinal tract (mouth, stomach, and the small intestine). Affected cells are large, with increased numbers of multinucleated and apoptotic dying cells. The gonads are also affected and infertility occurs in patients with cobalamin deficiency. Cobalamin deficiency may cause bilateral peripheral neuropathy or degeneration

Table 2 Causes of vitamin B_{12} deficiency.

1. Dietary[a]
   a. Veganism or very low intake of animal source foods
2. Gastric
   a. Atrophic gastritis and food B_{12} malabsorption
   b. Autoimmune gastritis/gastric atrophy (classical pernicious anemia)
   c. Extensive gastric disease or resection (including bariatric surgery)
3. Ileal
   a. Extensive ileal disease (Crohn disease, inflammatory bowel disease, tuberculous enteritis), or resection for these diseases
   b. Luminal disturbances (chronic pancreatic disease and gastrinoma) and parasites (giardiasis, bacterial overgrowth, and fish tapeworm).
4. Chemical/drug
   a. Nitrous oxide
   b. PAS, metformin, colchicine, proton pump inhibitors, H2-receptor antagonists
5. Congenital/inherited
   a. Intrinsic factor deficiency/defect ("Juvenile" pernicious anemia).
   b. Intrinsic factor receptor deficiency/defect (Immerslund-Gräsbeck disease).
   c. Transcobalamin (TC) II deficiency or polymorphisms[a]
   d. Cobalamin mutants (C-G).

[a]Suboptimal cobalamin status caused by lowered intake or TC polymorphisms can predispose to more rapid onset of deficiency when other pathological causes of cobalamin deficiency occur.

(demyelination) of the posterior and pyramidal tracts of the spinal cord and, less frequently, atrophy of the optic nerve or cerebral symptoms. Cobalamin-deficient patients typically display sensory disturbances (paraesthesiae), muscle weakness, difficulty in walking and sometimes dementia, psychotic disturbances or visual impairment. Long-term nutritional cobalamin deficiency in infancy leads to poor brain development and impaired intellectual development. The effects of cobalamin deficiency on the blood and on the nervous system may occur separately or in combination and their severity is often inversely rather than directly correlated. The biochemical basis for cobalamin neuropathy, however, remains obscure. Its occurrence in the absence of methylmalonic aciduria in TCII deficiency, and in monkeys given the anesthetic agent nitrous oxide, suggests that the neuropathy is related to a defect in homocysteine-methionine conversion. Accumulation of *S*-adenosylhomocysteine in the brain, resulting in inhibition of transmethylation reactions has been suggested as the mechanism.

Psychiatric disturbance is common in cobalamin deficiencies. Like the neuropathy, this has been attributed to a failure of the synthesis of SAM, due to reduced conversion of homocysteine to methionine. SAM is needed for methylation of biogenic amines (e.g., dopamine), as well as of proteins, phospholipids, and neurotransmitters in the brain.

Diagnosis of cobalamin deficiency

Cobalamin deficiency is suspected in individuals who display the typical manifestations of deficiency of the vitamin as described in the Section above, Causes, Mechanisms and Effects of Cobalamin Deficiency. In addition to the symptoms that may be experienced related to anemia (easy fatigue, shortness of breath, palpitations) and neuropathy (sensory and motor disturbances and memory loss) there are features that may be detected by a physician, including skin pallor (from anemia), abnormalities in neurological examination (sensory loss, abnormal balance and reflexes, and mental changes), and epithelial changes (skin pigmentation and smooth tongue). On the basis of any combination of such changes, cobalamin deficiency may be suspected but confirmation is necessary using laboratory tests because other conditions may give rise to effects that closely resemble cobalamin deficiency. The need to confirm suspected cobalamin deficiency applies also to individuals who have abnormalities in their blood count measurements.

The standard screening test for cobalamin deficiency consists of direct measurement of circulating levels of cobalamin. Serum levels less than 150 pmol L^{-1} (<200 ng/L) are considered deficient and levels 150–250 pmol L^{-1} are considered borderline. Serum or plasma cobalamin concentration can be measured in several ways and this has evolved from early microbial growth assays through competitive binding assays that were first radioisotopic and are now enzyme linked or based on chemiluminescence detection. The sensitivity and specificity of these assays are imperfect, such that measurement of serum cobalamin levels does not always detect the presence of deficiency, nor does the finding of a low serum cobalamin always connote true deficiency. There are several reasons for this including the distribution of cobalamin between the binding proteins in circulation (Table 1), imperfections in the assays for its measurement, and various poorly understood factors relating to exchange of cobalamin between cellular and circulatory compartments. Regarding the distribution of cobalamin between plasma-binding proteins, because transcobalamin is responsible for cobalamin delivery to cells, the fraction of the total cobalamin that is associated with transcobalamin (holoTC), even though small in comparison with the haptocorrin-associated fraction, is more likely to be indicative of cobalamin status than is the total serum cobalamin. Serum vitamin B12 assays measure the sum of vitamin B12 bound to haptocorrin (i.e., holohaptocorrin) and the fraction bound to transcobalamin (i.e., the holotranscobalamin). However, only the fraction of vitamin B12 bound to

Table 3 Laboratory identification of cobalamin deficiency.

Test	*Finding*	*Major limitations*
Serum/plasma cobalamin concentration	Low (<150 pmol L^{-1})	Normal levels in some deficient subjects; slight to moderately low levels may not connote deficiency
Serum/plasma holotranscobalamin (holo TC II)	Low (<35 pmol L^{-1})	Holo-TC assays show a rather wide coefficient of variation according to the used assay methods and thus variable interpretation of the results; assay interference due to autoantibodies or vitamin B12 supplements
Serum/plasma or urine methylmalonic acid	Raised (>350 nmol L^{-1})	Considerable inter-assay variability, levels raised in renal insufficiency, increased levels during pregnancy
Plasma homocysteine	Raised (>12 μmol L^{-1})	Levels raised in folate and in vitamin B_6 deficiencies, renal insufficiency, hypothyroidism

transcobalamin is taken up by the cells to cover their metabolic requirements (thus often named as "active B12"), suggesting that this biochemical parameter would represent a more suitable index. Despite still ongoing controversy in clinical practice the measurement of holotranscobalamin is favored as the first line assay, although serum/plasma levels as well as the holotranscobalamin levels do show a rather large gray zone for a clear interpretation of the results so that the measurement of the concentration of methylmalonic acid as a second line assay becomes often inevitable.

Since there is still no gold standard to assess vitamin B12 nutriture in a fully reliable and cost-efficient manner a combination of parameters might be applied. Cobalamin deficiency can be identified indirectly, based on the detection of raised levels of compounds in the blood or urine that require adequate tissue levels of cobalamin for their metabolic disposal. The compounds most commonly measured for identification of possible cobalamin deficiency are methylmalonic acid and homocysteine. These are the substrates in two cobalamin-dependent reactions shown in Fig. 2. Because of the identification of these metabolic roles for cobalamin, it has been apparent that deficiency of cobalamin or disturbances in its metabolism would result in accumulation of these substances and clinical assays for these metabolites is now available. Of the two compounds, elevation of the levels of methylmalonic acid is more specific for identification of cobalamin deficiency; however, renal insufficiency can cause raised levels of methylmalonic acid in the blood. In addition to cobalamin deficiency, several other conditions also can cause raised homocysteine levels in the blood, including deficiencies of folate and vitamin B6, lowered levels of thyroid hormone, and renal insufficiency.

Table 3 shows the idealized usefulness of the various tests commonly employed for the detection of cobalamin deficiency. Despite the lack of a reliable single parameter approach to assess deficiency, the main clinical issue is to assure high plasma/serum levels of vitamin B12 (>300 pmol L^{-1}) in any patient. This pragmatic approach is supported by recent evidence that the presence of a "high-folate-low-vitamin B12" status (as is frequently occurring due to the fortification of cereals with folate) might be a correctable cause of vitamin B12 depletion.

Inborn errors of cobalamin metabolism

There are several known but rare inherited molecular defects resulting in absence or structural defects of proteins required for normal absorption, transport, or metabolism of cobalamin. These include the intestinal binding proteins, gastric intrinsic factor and its ileal receptor complex cubulin and amnionless, the plasma binders transcobalamin and haptocorrin, the enzymes that are required for conversion of cobalamin to its coenzymatically active methyl and deoxyadenosyl forms, and enzyme complexes involved in the catalysis of the two cobalamin-dependent reactions responsible for conversion of homocysteine to methionine and methylmalonate to succinate, respectively. Individuals who inherit a defective gene from each parent for any one of the proteins that are critical for cobalamin metabolism suffer from varying degrees of impairment of normal cobalamin-related status (see Table 4), closely mimicking the various manifestations of cobalamin deficiency described above. These disorders usually become manifest at an early age.

Table 4 Inherited disorders affecting cobalamin metabolism and their effects.

Cobalamin protein	*Effects of deletion or mutation*
Intrinsic factor	Cobalamin malabsorption (juvenile pernicious anemia)
Cubulin/amnionless complex	Selective malabsorption of cobalamin, autosomal recessive megaloblastic Anemia (MGA1, Immerslund-Gräsbeck disease)
Transcobalamin	Severe cobalamin deficiency
Haptocorrin	No apparent abnormality
Cobalamin reducing and activating enzymes (mut^+ and mut-, cobalamin mutants C-G)	Varying degrees of disruption in one or both cobalamin-dependent pathways

See Also: Folate/folic acid; Homocysteine

Further reading

Allen, L.H., Miller, J.W., de Groot, L., Rosenberg, I.H., Smith, A.D., Refsum, H., Raiten, D.J., December 2018. Biomarkers of nutrition for development (BOND): vitamin B-12 review. J. Nutr. 148 (Suppl. 4), 1995S–2027S. https://doi.org/10.1093/jn/nxy201.

Carmel, R., Green, R., Rosenblatt, D.S., Watkins, D., 2003. Update on cobalamin, folate, and homocysteine. Hematology 62–81.

Green, R., 2011. Indicators for assessing folate and vitamin B_{12} status and for monitoring the efficacy of intervention strategies. Am. J. Clin. Nutr. 94, 666S–672S.

Green, R., Miller, J.W., 2007. Vitamin B_{12}. In: Zempleni, J., Rucker, R.B. (Eds.), Handbook of Vitamins. CRC Press, Boca Raton, FL, p. 413.

Guéant, J.-L., Guéant-Rodriguez, R.-M., Kosgei, V.J., Coelho, D., 2022. Causes and consequences of impaired methionine synthase activity in acquired and inherited disorders of vitamin B_{12} metabolism. Crit. Rev. Biochem. Mol. Biol. 57 (2), 133–155. https://doi.org/10.1080/10409238.2021.1979459.

Miller, J.W., July 2018. Proton pump inhibitors, H2-receptor antagonists, metformin, and vitamin B-12 deficiency: clinical implications. Adv. Nutr. 9 (4), 511S–518S. https://doi.org/10.1093/advances/nmy023.

Selhub, J., Miller, J.W., Troen, A.M., Mason, J.B., Jacques, P.F., January 2022. Perspective: the high-folate–low-vitamin B-12 interaction is a novel cause of vitamin B-12 depletion with a specific etiology—a hypothesis. Adv. Nutr. 13 (1), 16–33. https://doi.org/10.1093/advances/nmab106.

Sobczyńska-Malefora, A., Delvin, E., McCaddon, A., Ahmadi, K.R., Harrington, D.J., 2021. Vitamin B12 status in health and disease: a critical review. Diagnosis of deficiency and insufficiency—clinical and laboratory pitfalls. Crit. Rev. Clin. Lab Sci. 58 (6), 399–429. https://doi.org/10.1080/10408363.2021.1885339.

Stabler, S.P., 2020. Vitamin B12. In: Marriott, B.P., Birt, D.F., Stallings, V.A., Yates, A.A. (Eds.), Present Knowledge in Nutrition. Volume 1: Basic Nutrition and Metabolism, eleventh ed. Elsevier Inc.

Stabler, S.P., Allen, R.H., 2004. Vitamin B_{12} deficiency as a world-wide problem. Annu. Rev. Nutr. 24, 299–326.

Vitamin B_6

P.B. Mills[a], **E.J. Footitt**[b], **and P.T. Clayton**[a], [a] Genetics and Genomic Medicine, Great Ormond Street Institute of Child Health, University College London, London, United Kingdom; and [b] Department of Paediatric Metabolic Medicine, Great Ormond Street Hospital NHS Trust, London, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of D.A. Bender, Vitamin B6: Physiology, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 340–350, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00275-0.

Introduction 490
Chemical structures of B_6 vitamers and dietary sources 490
Digestion and absorption 490
Bacterial metabolism 490
Metabolism and transport 491
Cellular homeostasis 492
Reactions catalyzed by PLP enzymes 492
Glutamate decarboxylase (GAD) 493
Aromatic amino acid decarboxylase 493
Glycine and serine metabolism, 1-carbon and folate metabolism 493
Sulfur amino acid metabolism 493
Tryptophan metabolism 494
Glyoxylate metabolism 494
Glucose homeostasis 494
Heme synthesis 494
Reactivity of PLP 494
Roles of B_6 vitamers other than as an enzyme cofactor 494
Dietary deficiency 495
Assessment of vitamin B_6 status 495
Recommended daily intake 495
Dietary constituents that can cause PLP deficiency 496
Drug-induced deficiency 496
Toxicity of vitamin B_6 496
Pyridoxine 497
Acute toxicity 497
Chronic toxicity 497
Pyridoxal phosphate 497
Genetic disorders affecting vitamin B_6 metabolism and other disorders which may be treated with vitamin B_6 498
The role of vitamin B_6 in common diseases 498
Inflammation 499
Cardiovascular disease 499
Infection 500
Cancer 500
Parkinson's disease 500
Conclusion/summary/outlook 500
Acknowledgments 501
References 501

Key points

To explain:

- How the various forms of vitamin B_6 in the diet are converted to pyridoxal phosphate (PLP) in our organs and tissues
- How PLP acts as a cofactor for numerous enzymes, particularly those involving amino acids
- How dietary constituents and drugs that bind PLP can increase the amount of B_6 an individual needs to avoid symptoms of PLP deficiency
- How inborn errors of metabolism can have a similar effect.

To show how dietary deficiency of B_6 and the effects of drugs and inborn errors of metabolism have taught us that:
- The principal effects of PLP deficiency are neurological including seizures, movement disorder and polyneuropathy
- Deficiency can, however, also cause other disorders such as anemia
- These effects can be traced to reduced activity of PLP-dependent enzymes

To indicate to the reader that:
- Vitamin B_6 has functions in the body in addition to its role in supplying the PLP cofactor for PLP-dependent enzymes
- PLP plays a role in many common diseases including infection and inflammation, cancer and neurodegenerative disorders such as Parkinson disease

Introduction

Vitamin B_6 refers collectively to a group of related compounds, namely pyridoxine (PN), pyridoxamine (PM) and pyridoxal (PL), free, or as phosphates or glycoside esters. Most unicellular organisms and plants are capable of synthesizing vitamin B_6 de novo. *Escherichia coli* synthesizes PLP via the deoxyxylulose-5-phosphate (DXP)-dependent pathway utilizing erythrose-4-phosphate, glyceraldehyde-3-phosphate and pyruvate. Other bacteria, fungi, and plants use an alternative pathway that does not involve DXP, referred to as the DXP-independent pathway (Raschle et al., 2009). Animals, however, are not able to synthesize vitamin B_6 and are therefore reliant on the forms available in their diet, despite a contribution from gut microbiota, or that are recycled from protein turnover via a salvage pathway. One of the forms of vitamin B_6, pyridoxal 5′-phosphate (PLP), is catalytically active functioning as a cofactor for more than 4% of all known cellular enzymes. Increasingly, it is being recognized that B_6 vitamers also have nonenzymatic roles; having been shown to be effective antioxidants, regulating activity of steroid hormone receptors and modulating gene expression, and playing a role in brain glucose regulation. Because of such a wide spectrum of functions, vitamin B_6 is implicated in many common and rare disorders. A severe deficiency of PLP is rare and manifests as an epilepsy, believed to occur as a result of neurotransmitter imbalance. Milder deficiencies are more common, it has been suggested that 24–31% of the elderly have a B_6 intake below the estimated average requirement (EAR) (Porter et al., 2016). These have been implicated in disordered sleep, behavior, cardiovascular function and a loss of hypothalamus pituitary control of hormone excretion as well has having been reported to be associated with cognitive deficits and structural brain changes. Conversely, very high levels of vitamin B_6 may have toxic effects and cause peripheral neuropathy or induce seizures.

Chemical structures of B_6 vitamers and dietary sources

B_6 vitamers (Fig. 1) are derivatives of 2-methyl-3-hydroxy-5-hydroxymethyl-pyridine. The C4 substituent of the pyridine ring is an aldehyde group in pyridoxal, a hydroxymethyl group in pyridoxine and an aminomethyl group in pyridoxamine. The 5′ alcohol group is esterified to phosphate in pyridoxal phosphate (PLP), pyridoxine phosphate (PNP) and pyridoxamine phosphate (PMP) and linked to glucose in pyridoxine β-glucoside (PNG). The principal forms of vitamin B_6 in meat and fish are PLP and PMP, in breast milk PLP and pyridoxal and, in fruit and vegetables pyridoxine, PNP and PNG. PNG provides about 15% of the total vitamin B_6 in a mixed diet, however, the bioavailability of PNG is only about 50–60% that of pyridoxine ((Mackey et al., 2004); see Digestion and Absorption below). PLP in the diet may be bound to the ε-amino group of lysine residues proteins as a Schiff base (aldimine) and excessive heating of food can lead to reduction of the aldimine to produce phosphopyridoxyllysine (Gregory, 1980); this was implicated in dietary deficiency of B_6 in infants (see below). Vitamin B_6 can be synthesized by the gut flora and this may make a significant contribution to our B_6 intake (see below).

Digestion and absorption

Prior to absorption phosphorylated B_6 vitamers must be dephosphorylated by intestinal phosphatases and PNG must also be hydrolyzed to release pyridoxine. The latter reaction is catalyzed by two enzymes, the brush border membrane β-glucosidase, lactase phlorizin hydrolase and the cytosolic β-glucosidase, pyridoxine 5′-β-D-glucosidase (Mackey et al., 2004). Individuals with relative lactase deficiency may be at risk for a degree of pyridoxine deficiency (Mackey et al., 2002).

Bacterial metabolism

The gut flora is capable of synthesizing vitamin B_6 that can be used the host; indeed it has been suggested that the gut microbiota may be able to synthesis enough pyridoxine to supply 86% of the recommended dietary intake of B_6 although this needs to be

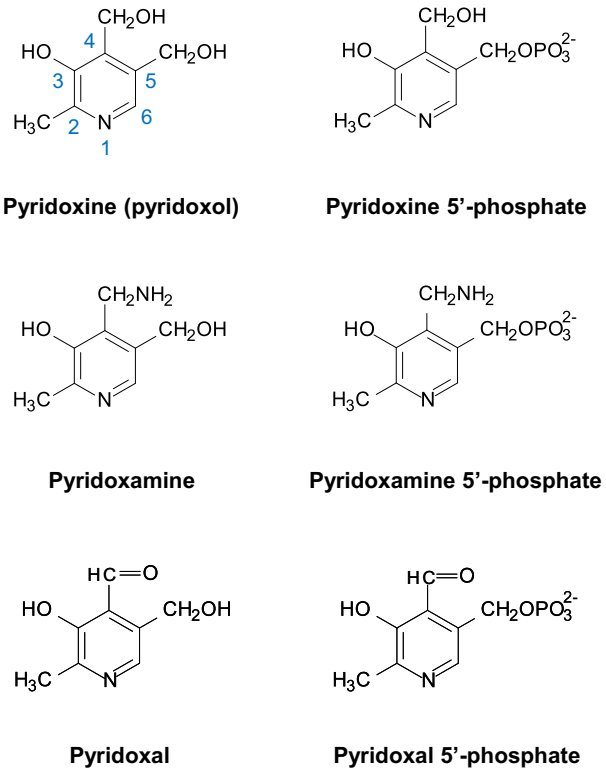


Fig. 1 Chemical structures of the B_6 vitamers. The B6 vitamers are based on a pyridine ring structure. Pyridoxine, pyridoxamine and pyridoxal differ by the substituent at the 4-position and when phosphorylated at position 5 form pyridoxine 5′-phosphate, pyridoxamine 5′-phosphate and pyridoxal 5′-phosphate, respectively. Numbering of the ring positions is shown in blue for pyridoxine.

verified experimentally (Magnúsdóttir et al., 2015). However, some bacteria also catabolize B_6 vitamers (McCulloch et al., 2010) and this may become important when individuals with inborn errors of metabolism are treated with very high doses of PLP given orally.

Metabolism and transport

Pyridoxal, pyridoxamine and pyridoxine absorbed from the gut are mainly converted to PLP in the liver (Fig. 2), although this can occur in other tissues. Phosphorylation by pyridoxal kinase (PK) produces PLP, PMP and PNP and conversion of PMP and PNP to PLP is catalyzed by pyridox(am)ine 5′-phosphate oxidase (PNPO) which utilizes flavin mononucleotide (FMN) as cofactor (Wilson et al., 2019). PLP produced in the liver is exported into the circulation bound to lysine 190 of albumin. Studies have also shown (Albersen et al., 2013) that B_6 vitamers can be converted to PLP by a Caco-2 cell line, an in vitro intestinal model derived from a colon carcinoma, with PL being the predominant vitamer excreted.

PK and PNPO are both inhibited by PLP. Regulation of enzyme activities also occurs at the transcriptional level (at least in experimental animals). Three transcription factors, DBP (albumin D-site-binding protein), HLF (hepatic leukemia factor), and TEF (thyrotroph embryonic factor), increase transcription of PK leading to a circadian rhythm of PK expression and PLP levels in mice (Gachon et al., 2004). Knockout of these 3 transcription factors leads to brain PLP deficiency and epilepsy. Transcription of PNPO can be affected by knockout of a non-coding microRNA, miR324 and this also leads to increased hippocampal excitability (Hayman et al., 2021).

When the intake of B_6 vitamers exceeds the body requirement, PLP is catabolized, mainly in the liver, by phosphatases with subsequent conversion of the pyridoxal to pyridoxic acid which is excreted in the urine. It is not certain whether the oxidation of pyridoxal to pyridoxic acid is catalyzed by an aldehyde dehydrogenase or an aldehyde oxidase.

The main form of B_6 in plasma is PLP bound to albumin. To supply B_6 to the brain, PLP must dissociate from albumin, and undergo dephosphorylation to pyridoxal which then crosses the blood brain barrier and or the blood CSF barrier of the choroid plexus (Wilson et al., 2019). Dephosphorylation is catalyzed by the ectoenzyme tissue non-specific alkaline phosphatase (TNSALP) which is tethered to the blood facing membrane by a glycophosphatidylinositol (GPI anchor). Once inside neurons and glia, pyridoxal is rephosphorylated by pyridoxal kinase. PLP regenerated from pyridoxal, by pyridoxal kinase, in the choroid plexus can be

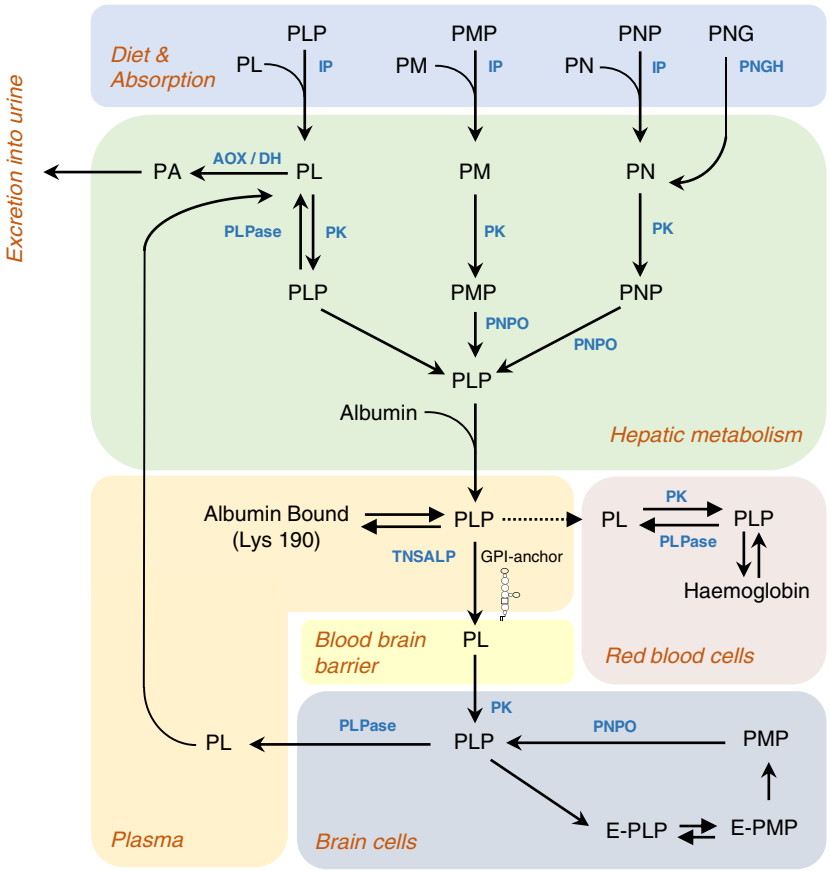


Fig. 2 Interconversion of dietary B_6 vitamers to the active cofactor, pyridoxal 5′-phosphate. Enzymes involved in the interconversion given in blue text. Pyridoxal 5′-phosphate (PLP); pyridoxamine 5′-phosphate (PMP); pyridoxal (PL); pyridoxine (PN); pyridoxine 5′-phosphate (PNP); pyridoxine-5′-β-D-glucoside (PNG); intestinal phosphatases (IP); pyridoxal kinase (PK); pyridox(am)ine 5′-phosphate oxidase (PNPO); tissue non-specific alkaline phosphatase (TNSALP); pyridoxal-phosphatase (PLPase); GPI anchor (glycosylphosphatidylinositol anchor); AOX/DH (Aldehyde oxidase [Mo cofactor]/β-NAD dehydrogenase); E-PLP (enzyme-PLP); E-PMP (enzyme-PMP).

secreted into the CSF bound to albumin. Uptake of B_6 from the CSF requires dephosphorylation by TNSALP. Supply of B_6 to other tissues of the body from plasma PLP is also dependent on TNSALP.

The body conserves B_6 by means of a salvage pathway. PMP generated when PLP-dependent enzymes catalyze a partial reaction (e.g., the first half reaction of a transamination) or when they are catabolized can be converted back to PLP by PNPO. Free vitamers can be rephosphorylated by pyridoxal kinase (Wilson et al., 2019).

Cellular homeostasis

All cells in the body require PLP bound to their PLP-dependent enzymes, however, concentrations of free PLP over 10 μM can inhibit enzymes (Christmann-Franck et al., 2007) and to avoid unwanted reactions of PLP the free PLP is maintained at 1 μM (di Salvo et al., 2011). Mechanisms for controlling the free PLP concentration are thought to include: (i) inhibition of pyridoxal kinase and PNPO by PLP; (ii) transfer of newly formed PLP from either enzyme to apo-B_6-enzymes by direct channeling; (iii) hydrolysis of excess PLP by phosphatases; (iv) binding of PLP to abundant intracellular proteins e.g., glycogen phosphorylase in muscle and hemoglobin in red cells; (v) binding of PLP to the pyridoxal phosphate homeostatic protein (PLPHP aka PLP-binding protein, PLPBP, PROSC). Deficiency of the latter was found to be a cause of impaired cellular PLP homeostasis and B_6-dependent epilepsy in 2016 (Darin et al., 2016) but the exact function of PLPHP is not known.

Reactions catalyzed by PLP enzymes

PLP enzymes in the body, in their resting state, have the PLP bound covalently via an imine bond to the ε-amino group of the active site lysine forming a Schiff base or "internal aldimine" (Eliot and Kirsch, 2004). The largest group of PLP-catalyzed reactions act on

amino acid substrates. The amino group of the substrate displaces the lysine ε-amino group forming a new aldimine with the PLP ("external aldimine/external Schiff base"). This increases the reactivity of the α, β and γ carbons of the amino acid. The reactions that are catalyzed at the α-carbon include transamination, decarboxylation, racemization and elimination and replacement of an electrophilic group. Reactions catalyzed at the β− and γ− carbons include eliminations and replacements (Eliot and Kirsch, 2004).

The notable exception to the general rule that PLP-catalyzed reactions occur through the formation of an external aldimine of PLP and an amino acid is glycogen phosphorylase. This enzyme catalyzes the transfer of a phosphate group to glycogen generating glucose-1-phosphate (Takagi et al., 1982).

The consequences of reduced activity of some specific PLP enzymes and PLP-dependent pathways are discussed below:

Glutamate decarboxylase (GAD)

This PLP enzyme is required for the synthesis of GABA, the major inhibitory neurotransmitter. Deficiency of the GAD65 isoform in rats causes low GABA levels in the brain, epilepsy and premature lethality (Kakizaki et al., 2021). Animal models with PLP deficiency causing seizures have low brain GABA levels (Ciapaite et al., 2020; Johnstone et al., 2019; Pena et al., 2017). In children with pyridoxine-dependent epilepsy there is evidence of low GAD activity and low brain and CSF GABA (Gospe et al., 1994).

Aromatic amino acid decarboxylase

This PLP enzyme catalyzes the production of dopamine from L-Dopa and of serotonin from 5-hydroxytryptophan. Isolated AADC deficiency leads to deficiency of dopamine and serotonin most easily detected by reduced CSF concentrations of their metabolites, homovanillic acid (HVA) and 5-hydroxyindoleacetic acid (5HIAA) respectively. The clinical phenotype is a severe movement disorder with particularly prominent oculogyric crises (Hyland et al., 1992; Saudubray et al., 2016). Inborn errors of metabolism causing PLP deficiency can also demonstrate low CSF HVA and 5HIAA and the dominant feature of the phenotype can be a movement disorder resembling AADC deficiency (Johnstone et al., 2019).

Glycine and serine metabolism, 1-carbon and folate metabolism

The glycine decarboxylase component of the glycine cleavage enzyme is a PLP-dependent enzyme. Isolated deficiency of the glycine cleavage enzyme (GCE) causes non-ketotic hyperglycinaemia which presents with neonatal hypotonia, severe seizures and raised plasma and CSF glycine. Thus, it is likely that reduced glycine decarboxylase activity contributes to the high plasma and CSF glycine, and possibly to the seizures seen in PLP deficiency. Serine reacts with tetrahydrofolate in a reaction catalyzed by the PLP-dependent enzyme, serine hydroxymethyl transferase (SHMT), to produce glycine and 5,10-methylenetetrahydrofolate. Serine levels in plasma are elevated in B_6 deficiency; mathematical modeling suggests that this is due more to reduction in the activity of GCE than SHMT (Nijhout et al., 2009).

When neuronal cells are cultured in B_6-deficient medium, their content of serine and glycine is reduced as is their content of 5-methyltetrahydrofolate. This suggests that the PLP deficiency is affecting the PLP-dependent steps involved in serine synthesis from glycolytic intermediates and this together with the direct effect on GCE and SHMT is causing impaired function of the folate cycle (Ramos et al., 2017). In some individuals with inborn errors of metabolism leading to PLP deficiency, folinic acid-responsive seizures have been described (Johnstone et al., 2019; Gallagher et al., 2009). In *C. elegans* rendered profoundly PLP deficient by ensuring that the *E.coli* that formed the diet as well as the gut flora could not synthesize B_6, dysfunction of the folate cycle and 1-carbon metabolism was clearly demonstrated (Scott et al., 2017).

It has been suggested that the neuropathology of folate and B_{12} disorders could be caused by deficiency of S-adenosyl-methionine (Bottiglieri et al., 1994); the same may be true of features of B_6 disorders such as peripheral and optic neuropathies.

Sulfur amino acid metabolism

B_6 deficiency can contribute to hyperhomocysteinemia. The PLP-dependent enzyme cystathionine β-synthase provides one route for metabolism of homocysteine i.e., conversion to cystathionine. Individuals with cystathionine β-synthase deficiency have considerably elevated plasma homocysteine levels and in about 50% the hyperhomocysteinemia can be effectively treated with large doses of pyridoxine (Saudubray et al., 2016). However reduced activity of CBS may be only one factor that leads to hyperhomocysteinemia in B_6 deficiency. The folate cycle generates 5-methyl-tetrahydrofolate which is required for the B_{12}-dependent conversion of homocysteine to methionine and as indicated above the folate cycle may be impaired in B_6 deficiency. Folic acid supplementation can ameliorate the hyperhomocysteinemia induced in rats by B_6 deficiency and methionine supplementation (Yamamoto et al., 2012). Finally, PLP is required for cystathionine γ-lyase activity and PLP deficiency can lead to raised plasma cystathionine, however, this may not contribute to disease as individuals with cystathionine γ-lyase deficiency are asymptomatic (Saudubray et al., 2016).

Cysteine supplies the sulfur in the iron/sulfur clusters that are essential for oxidative phosphorylation and this is achieved through the action of cysteine desulfurase, a PLP-dependent enzyme. Individuals with cysteine desulfurase deficiency have defective oxidative phosphorylation (Farhan et al., 2014) and inborn errors of metabolism that cause severe PLP deficiency can show marked lactic acidosis, for which cysteine desulfurase deficiency may be one cause.

Tryptophan metabolism

The catabolism of tryptophan and its conversion to nicotinamide require the PLP-dependent enzyme, kynureninase, which converts 3-hydroxykynurenine to 3-hydroxyanthranilic acid. If 3-hydroxykynurenine builds up it is converted to xanthurenic acid. PLP deficiency leads to increased excretion of xanthurenic acid especially if the individual is given a tryptophan load. Deficiency of nicotinamide is probably unusual in PLP deficiency, but it has been described in isoniazid-induced deficiency (Prabhu et al., 2021).

Glyoxylate metabolism

The B_6-dependent enzyme alanine glyoxylate aminotransferase catalyzes the conversion of alanine and glyoxylate to pyruvate plus glycine. Reduced activity of this enzyme leads to increased production of oxalate increasing the risk of formation of oxalate stones in the urinary tract. This can occur in B_6-deficient rats (Nishijima et al., 2003).

Glucose homeostasis

Hypoglycaemia has been described in inborn errors of metabolism causing PLP deficiency. The following enzyme deficiencies may play a role: glycogen phosphorylase (rate limiting step for release of glucose from glycogen); transaminases and other PLP dependent reactions of amino acids (required for gluconeogenesis from amino acids); AADC deficiency (impaired catecholamine synthesis) and others.

Heme synthesis

PLP is needed by the enzyme that catalyzes the first step in heme synthesis i.e., δ-aminolaevulinic acid synthase. Isolated deficiency of this enzyme causes a sideroblastic anemia which can be effectively treated with pyridoxine, plus folic acid (Saudubray et al., 2016). Anemia is a feature of PLP deficiency of various causes and deficiency of δ-aminolaevulinic acid synthase may contribute.

Reactivity of PLP

The reactive aldehyde group of PLP allows it to act as the cofactor for enzyme-catalyzed reactions as indicated above. However some reactions of PLP have adverse or potentially adverse effects if the reacting molecule is present in sufficiently high concentration (see **Table 1** for small molecules). PLP can also have adverse effects on proteins if present at high concentration (Kajita et al., 2013; Christmann-Franck et al., 2007). Finally, as an endogenous aldehyde, PLP may in some circumstances damage chromosomes and mutate stem cells (Garaycoechea et al., 2018).

Roles of B_6 vitamers other than as an enzyme cofactor

B_6 vitamers can act as antioxidants; they can quench singlet oxygen at rates comparable to vitamin C and E (Bilski et al., 2000).

PLP can affect the expression and action of steroid hormone receptors (Tully et al., 1994). It can inhibit the binding of the transcription factor HNF1 to DNA (Oka et al., 2001). The action of a corepressor for many nuclear factors, RIP140, is enhanced by conjugation of Lys613 to PLP (Huq et al., 2007). This may affect the function of RIP141 in adipocyte differentiation.

PLP can act as an antagonist of ATP at the purinergic receptor P2X7 (Thériault et al., 2014). Activation of P2X7 receptors by ATP is thought to be a mechanism for the genesis of epilepsy in neuroinflammation (Beamer et al., 2017). Antagonism by PLP may explain why it acts as an anticonvulsant even when there is no evidence of PLP deficiency.

In the immune system, PLP binds tightly to the D1 domain of CD4 and this may interfere with the reaction of CD4 with MHC II (Salhany and Schopfer, 1993).

Using Drosophila as a model system, Marzio et al. (2014) obtained evidence that PLP plays a crucial role in maintenance of the integrity of the genome. Mutations in the gene encoding pyridoxal kinase (*dPdxk*) produce chromosome aberrations (CABs) in larval neuroblasts; this can be prevented by PLP. CABs can also be induced by PLP antagonists such as 4-deoxypyridoxine hydrochloride; penicillamine, cycloserine and isoniazid also increase the frequency of CABs. The authors suggest that, since CABs are precursors of cancer, low levels of PLP may predispose to initiation of cancer. In the same Drosophila model, Marzio et al. (2014) found that $dPdxk^1$ mutants have increased levels of glucose in their hemolymph attributable to insulin resistance. They suggested that this might indicate a susceptibility to diabetes and that DNA damage is caused by high glucose levels in PLP deficient cells.

Table 1 Molecules that react with PLP.

Molecule	*Reason for build-up*	*Consequence of reduced PLP*
Δ^1-piperideine-6-carboxylate (P6C)	ALDH7A1 deficiency (OMIM 266100)	Pyridoxine-responsive epilepsy
Δ^1-pyrroline-5-carboxylate (P5C)	ALDH4A1 deficiency (OMIM 239510)	Pyridoxine responsive seizures
Sulfite and P6C	Molybdenum cofactor deficiency (OMIM 603708)	Pyridoxine-responsive seizures
1-Amino-D-proline (Linatine)	Ingestion of linseed/oil	Seizures in rats
Isoniazid	Treatment of tuberculosis	Peripheral neuropathy. Increased urinary xanthurenic acid Pellagra (decreased nicotinamide synthesis from tryptophan) Overdose: Seizures, lactic acidosis
D-cycloserine	Treatment of tuberculosis	Seizures Psychosis
Hydralazine	Treatment of hypertension[a]	Peripheral neuropathy
Monomethylhydrazine	Poisoning with *Gyromitra* mushrooms (false morel)[b]	Seizures, coma
Penicillamine	Treatment of Wilson disease[c]	Optic neuropathy Epilepsy Increased urinary xanthurenic acid
L-Dopa	Treatment of Parkinson disease	Reduced PLP and reduced L-Dopa may both contribute to reduced effectiveness of L-Dopa Contribution to peripheral neuropathy[d]

[a]Raskin, N.H., Fishman, R.A., 1965. Pyridoxine-deficiency neuropathy due to hydralazine. N Engl. J. Med. 273, 1182–1185.
[b]Lheureux, P., Penazola, A., Gris, M., 2005. Pyridoxine in clinical toxicology: a review. Eur. J. Emerg. Med. 12, 79-85.
[c]Gibbs, K., Walshe, J.M., 1969. Interruption of the tryptophan-nicotinic acid pathway by penicillamine-induced pyridoxine deficiency in patients with Wilson's disease and in experimental animals. Ann. N. Y. Acad. Sci. 166(1).
[d]Romagnolo, A., Merola, A., Artusi, C.A., Rizzone, M.G., Zibetti, M., Lopiano, L., 2018. Levodopa-induced neuropathy: a systematic review. Mov. Disord. Clin. Pract. 6(2):96–103. 158–169.

Dietary deficiency

The main symptoms of dietary B_6 deficiency in infancy are neurological. In 1954, Coursin described 300 infants who were fed on an infant formula that had been heated excessively and only contained 60 μg/L of available B_6; heat converts protein bound PLP to pyridoxyllysine. After between 6 weeks and 4 months the unfortunate infants developed a triad of hyperacusis, irritability and convulsions; these symptoms worsened when protein intake was increased and improved with a high carbohydrate diet. All symptoms were abolished by increasing the dietary vitamin B_6 intake.

Adults fed diets deficient in B_6 also develop neurological symptoms—confusion, irritability and convulsions associated with an abnormal electroencephalogram—but they also develop a microcytic anemia and skin and mouth symptoms (eczema, seborrheic dermatitis, cheilosis, glossitis and angular stomatitis (Berger et al., 2022).

Dietary deficiency of vitamin B_6 will respond to treatment with oral supplementation. This is readily available and 50–100 mg for one to two weeks is safe (Berger et al., 2022). The no-observed-adverse-effect level (NOAEL) is 100 mg/day.

Assessment of vitamin B_6 status

Historically a number of biochemical parameters have been used to assess B_6 status. These have included measurement of erythrocyte alanine and aspartate transaminases before and after addition of PLP and measurement of urinary excretion of xanthurenic acid and kynurenic acid after a tryptophan load. Nowadays, vitamin B_6 status is usually assessed by direct measurement of plasma PLP and urinary pyridoxic acid. An increase in dietary B_6 intake leads to an immediate increase in pyridoxic acid excretion, plasma PLP levels rise more slowly and plateau at about 7–10 days after the increase in intake. It has been suggested that urinary pyridoxic acid excretion is an index of recent B_6 intake whereas plasma PLP is a marker of body stores. However, it is also useful to look at a range of indirect biomarkers that indicate the effect of B_6 status on individual enzymes/pathways e.g., kynurenine pathway, one-carbon metabolism, transsulfuration (cystathionine), and glycine decarboxylation (serine and glycine) (Ueland et al., 2015). Normal values of plasma PLP are 20–200 nmol/L (5–50 μg/L) (Berger et al., 2022).

Recommended daily intake

The Reference Nutrient Intake (RNI, European Food Safety Authority) and Recommended Dietary Allowance (RDA, United States), i.e., estimates of the amount of B_6 that is enough to ensure requirements of 97.5% of the population are being met, have been

Table 2 Recommended dietary allowance of vitamin B_6 for infants, children and adolescents.

Age	*Male*	*Female*
0–6 months[a]	0.1 mg	0.1 mg
7–12 months[a]	0.3 mg	0.3 mg
1–3 years	0.5 mg	0.5 mg
4–8 years	0.6 mg	0.6 mg
9–13 years	1.0 mg	1.0 mg
14–18 years	1.3 mg	1.2 mg

[a]Adequate intake.
Institute of Medicine (1998).

determined by depletion-repletion studies using either plasma PLP or the tryptophan load as measures of B_6 sufficiency. Because PLP is central to the reactions of amino acid catabolism, it is not surprising that vitamin B_6 depletion was shown to occur more rapidly in individuals on a high protein intake compared to those on a low intake (Canham et al., 1969).

The recommended daily allowance (RDA) for vitamin B_6 is 1.3–1.7 mg/day, irrespective of gender for 14–70-year-olds, although one study has suggested that the over 60s may have a requirement of 1.96 mg for men and 1.9 mg for women (Ribaya-Mercado et al., 1991). In pregnant women this can increase to 2 mg/day. The upper advised limit for pyridoxine intake is 80 mg/day for 14–18 year olds; 100 mg/day for those over 19 years old (Berger et al., 2022). The RDA for infants, children and adolescents is presented in **Table 2** (Institute of Medicine, 1998).

Dietary constituents that can cause PLP deficiency

Ginkgo nuts contain 4-O-methyl pyridoxine which is an inhibitor of pyridoxal kinase. The seizures induced by eating too many Ginkgo nuts can be effectively treated with pyridoxine (Hasegawa et al., 2006).

Sulfite is present in many foods and in wine as a preservative. As it forms a complex with PLP, it may reduce the availability of B_6 (Gould and Russell, 2003).

Chicks and turkey poults show poor growth and signs of B vitamin deficiency when reared on linseed meal. The signs of vitamin deficiency can be prevented by the addition of 20 ppm of pyridoxine to the linseed-containing diet (Kratzer and Williams, 1948). Klosterman et al. identified the vitamin B_6 antagonist as 1-amino-proline ("linatine") (Klosterman et al., 1967). Administration of linatine to rats causes seizures (Sasaoka et al., 1976). Many people consume linseed (flax seed) as a dietary supplement. However, the known side effects in man (gastrointestinal and clotting) do not include seizures.

Poisoning with *Gyromitra* mushrooms (False morel) can cause refractory seizures because they contain monomethyl hydrazine which forms a complex with PLP (Lheureux et al., 2005).

Drug-induced deficiency

Drugs can lead to deficiency of PLP either by forming a complex with it or by inhibiting an enzyme required for its synthesis from dietary B_6 vitamers and its recycling.

Examples of drugs forming a complex with PLP are included in **Table 1**. Consequences include peripheral neuropathy with isoniazid and hydralazine (Raskin and Fishman, 1965), seizures and psychosis with D-cycloserine, optic neuropathy and seizures with penicillamine (Gibbs and Walshe, 1969). These side effects can be prevented by coadministration of pyridoxine. Treatment of Parkinson's disease with L-Dopa and carbidopa can lead to peripheral neuropathy and B_6 deficiency is a contributory factor (Romagnola et al., 2018). An example of a drug that inhibits synthesis and recycling of PLP is theophylline, a non-competitive inhibitor of pyridoxal kinase (Ubbink et al., 1990). An overdose of theophylline can induce seizures and pyridoxine treatment has been shown to be beneficial in animals (Glenn et al., 1995).

Toxicity of vitamin B_6

Similar to Vitamin B_6 deficiency states, the predominant symptoms of B_6 toxicity are neurological. Toxicity of vitamin B_6 is not observed through dietary intake alone, however it is reported following the intake of supraphysiological doses of both pyridoxine and pyridoxal phosphate (PLP) used for various indications in clinical practice and in some instances, individuals self-treating with over-the-counter supplements.

Pyridoxine

Acute toxicity

In animal studies very high doses of pyridoxine (2–6 g/kg) induce ataxia and seizures and may be lethal (Unna, 1940). Administration of such high doses is not reported in human studies, however B_6 appears to have low acute toxicity in man with doses up to 357 mg/kg used for a short time period in isoniazid overdose without any adverse events (Lheureux et al., 2005).

Chronic toxicity

Typical symptoms of chronic pyridoxine toxicity include peripheral sensory neuropathy which manifests as symmetrical paresthesia in the extremities and difficulty in walking due to ataxia and disequilibrium (Albin et al., 1987). Other symptoms reported less frequently in a large study of patients include hyperesthesia, muscle weakness, fasciculations and bone pains (Dalton and Dalton, 1987).

Clinical and electrophysiological evidence of sensory neuropathy is observed in healthy volunteers and patients taking pyridoxine for a variety of indications. A single case reports evidence of sensory and motor neuropathy following a prolonged course of high intake (1 g/day for 10 years) (Morra et al., 1993). In the majority of cases, symptoms are reversible on stopping pyridoxine intake (although residual nerve damage remains in some patients) (Schaumburg et al., 1983) and shows a clear relationship to pyridoxine dose and in some instances to length of administration (Cohen and Bendich, 1986; Bender, 1999). Regular clinical examination and nerve conduction studies are recommended for all patients receiving long term high dose pyridoxine therapy.

Rarely, high doses of pyridoxine have been reported to cause serious life-threatening complications such as convulsions and apnea which have both been reported in the neonate. It appears that this age group may respond differently to large doses of pyridoxine and it is recommended that the lowest possible doses should be used alongside adequate monitoring in this age group.

Ames et al. (2020) report a newborn infant treated with 100 mg pyridoxine daily (31 mg/kg/d) for one week followed by 200 mg daily (54 mg/kg/d) for a second week as part of the protocol for management of newborn screen positive for classical homocystinura. After approximately 2 weeks of pyridoxine treatment the infant developed dyspnea, cyanosis and apnea requiring resuscitation and intensive care admission. She was also found to have elevated liver transaminase and creatine kinase (CK). No other cause for her deterioration was found and her condition improved within 72 h of pyridoxine cessation and did not recur when she was recommenced on a low dose pyridoxine (10 mg/kg).

Another case of pyridoxine toxicity is described in a neonate whose mother received pyridoxine antenatally and was then treated with pyridoxine from birth (33 mg/kg/day) due to familial risk of pyridoxine dependent epilepsy—ALDH7A1-deficiency (Hartmann et al., 2011). From day 14 the infant began to feed poorly and had abnormal breathing. On Day 15 he required admission to hospital for repeated apnea and seizure activity with abnormal EEG. A further increase in pyridoxine dose to 60 mg/kg/day resulted in the need for ventilatory support. He was noted to have elevated CK, AST with plasma PLP grossly elevated >3000 μg/L (normal range up to 18 μg/L). Treatment with vitamin B_6 was stopped when ALDH7A1-deficiency was excluded, and the infant made a full recovery.

The mechanism of pyridoxine toxicity is incompletely understood. Studies of chronic, long-term pyridoxine administration in animals have shown the potential neurotoxicity of vitamin B_6 with doses of 200 mg/kg causing ataxia, peripheral neuropathy and muscle weakness. Histological examination demonstrates widespread neuronal damage with loss of myelin and degeneration of sensory fibers in peripheral nerves, dorsal columns of the spinal cord and descending tract of the trigeminal nerve (Phillips et al., 1978). Studies in dogs (Yun et al., 2020) demonstrate neuropathological changes in dorsal root ganglia following large doses of pyridoxine (150 mg/kg) which are reversible. Neurotoxicity of injected pyridoxine may be enhanced by a protein deficient diet. This is perhaps due to decreased protein binding in serum and decreased urinary excretion of the toxin secondary to oliguria which is evident in these animals because of reduced water consumption (Levine and Saltzman, 2004).

A recent in vitro study showed that pyridoxine can induce cell death in a dose dependent manner and that it also inhibits pyridoxal-phosphate dependent enzymes, thus partly explaining why symptoms of pyridoxine toxicity are similar to those of deficiency (Vrolijk et al., 2017).

Pyridoxal phosphate

Less is known about PLP toxicity as clinical experience is comparatively limited. Hammen et al. (1998) report a newborn with intractable epilepsy who showed an increase in seizure frequency and EEG alterations after administration of vitamin B_6. Wang et al. (2005) also describe a paradoxical seizure increase during treatment with PLP. PLP may also cause tonic-clonic convulsions in immature mice possibly secondary to degeneration of GABA-ergic neurotransmission (Ishioka et al., 1995).

PLP at a dose of 1800 mg/day induced hepatotoxicity in a child with homocystinuria (Yoshida et al., 1985) and liver cirrhosis has developed in at least two children with pyridoxamine 5′-phosphate oxidase (PNPO) deficiency while on treatment with oral PLP (Coman et al., 2016; Sudarsanam et al., 2014). The mechanism of toxicity is unknown however, preparation of PLP immediately prior to administration may prevent deterioration of the vitamin to toxic intermediates in UV light on standing.

Table 3 Disorders in which vitamin B_6 preparations may be used therapeutically where metabolites accumulate that inactivate pyridoxal 5′-phosphate.

Disorder (OMIM reference)	*Enzyme*	*Gene; location*	*Vitamin B_6 preparation; dose*	*Clinical aim or biochemical effect*
Pyridoxine dependent epilepsy (266100/107323)	α-Aminoadipic semialdehyde dehydrogenase EC 1.2.1.31	*ALDH7A1* 5q31	PN; 50–100 mg IV single dose(s), 5–15 mg/kg/d oral maintenance	Cessation and prevention of seizures and improvement of IQ
Hyperprolinaemia type II (239510/606811)	L-Δ^1-Pyrroline-5-carboxylic acid dehydrogenase EC 1.5.1.12	*ALDH4A1* 1p36	PN; 50 mg/d, subsequent reduction to 10 mg/d	Cessation and prevention of seizures (particularly during intercurrent infection)

PN - pyridoxine.

Genetic disorders affecting vitamin B_6 metabolism and other disorders which may be treated with vitamin B_6

There are several rare genetic metabolic disorders which affect vitamin B_6 metabolism in differing ways, ultimately resulting in a common endpoint; cellular deficiency of pyridoxal phosphate, the active form of vitamin B_6 (Tables 3 and 4).

Due to the critical role of PLP as an enzyme cofactor in the central nervous system, these disorders manifest with a neurological phenotype and the majority present with epilepsy early in life, often soon after birth. These disorders are characteristically responsive to treatment with vitamin B_6 (pyridoxine or pyridoxal phosphate) and lifelong treatment is required.

Table 5 summarizes other disorders which may benefit from treatment with vitamin B_6 whereby the treatment aim is to augment the residual enzyme activity of a B_6 dependent enzyme.

The role of vitamin B_6 in common diseases

While a severe isolated deficiency of vitamin B_6 is rare, suboptimal levels have been reported in smokers and alcoholics, users of oral contraceptives and some drugs, pregnant women, and in individuals with chronic renal failure, diabetes and inflammatory bowel disease. Low plasma PLP is associated with an increased risk of more common chronic disorders including cardiovascular diseases, stroke, diabetes and cancer and has also been linked to rheumatoid arthritis. Inflammation is believed to play a pivotal role in the pathogenesis or progression of these conditions.

Table 4 Disorders in which vitamin B_6 preparations (pyridoxine or pyridoxal 5′-phosphate) may be used therapeutically where there is inadequate production of the active cofactor (in the correct location) due to an inborn error affecting B_6 interconversion or due to disruption in cellular PLP homeostasis.

Disorder (OMIM reference)	*Enzyme/protein*	*Gene; location*	*Vitamin B_6 preparation; dose*	*Clinical aim or biochemical effect*
Pyridoxal phosphate dependent epilepsy (603287/610090)	Pyridox(am)ine 5′-phosphate oxidase EC 1.4.3.5	*PNPO*; 17q21.32	PLP; 30–50 mg/kg/d orally In some PN; 6–55 mg/kg/d	Cessation and prevention of seizures
Hypophosphatasia (infantile) (241500/171760)	Tissue nonspecific alkaline phosphatase EC 3.1.3.1	*ALPL/TNSALP/TNAP*; 1p36.1-34	PN; 100 mg IV PLP; 30 mg/kg/d	Cessation and prevention of seizures
Hyperphosphatasia (239300/610274)	Phosphatidyl-inositol glycan class V	*PIGV*; 1p36.11	PN; 100 mg/day	Cessation of seizures, paradoxical change in electroencephalogram
Pyridoxal phosphate-binding protein deficiency (604436/617290)	Pyridoxal phosphate binding protein	*PLPBP*; 8p1123	PLP is superior treatment for seizure control; 10–45 mg/kg/d	Cessation of seizures
Pyridoxal kinase deficiency (179020/618511)	Pyridoxal kinase EC 2.7.1.35	*PDXK*; 21q22.3	PLP; 50 mg/day	Improve polyneuropathy—improve motor function and neuropathic pain

PLP—pyridoxal 5′-phosphate, PN—pyridoxine.

Saudubray, J.M., Baumgartner, M.R., Walter, J., 2016. Inborn metabolic diseases: diagnosis and treatment. sixth ed. Berlin, Heidelberg, New York: Springer-Verlag.

Table 5 Disorders in which vitamin B_6 preparations (pyridoxine or pyridoxal 5′-phosphate) may be used therapeutically to augment residual enzyme activity in inborn errors affecting a PLP-dependent enzyme.

Disorder (OMIM reference)	*Enzyme*	*Gene; location (B_6 responsive mutations)*	*Vitamin B_6 preparation; dose*	*Clinical aim or biochemical effect*
Homocystinuria (236200) (B_6 responsive subgroup)	Cystathionine β-synthase EC 4.2.1.22	*CBS;* 21q22.3 (I278T, 114V, R266K, 336H, K384E, L539S)	[a]PN; 100 mg/d	Biochemical parameters normalize. Prevention of thromboembolic events
Gyrate atrophy of the choroid and retina (258870)	Ornithine δ-aminotransferase EC 2.6.1.13	*OAT;* 10q26 (V332M, A226V, 318K)	PN; 500–1000 mg/d	Decrease in plasma ornithine. Unknown effect on chorioretinal degeneration
Aromatic amino acid decarboxylase deficiency (608643/107930)	Aromatic L-amino acid decarboxylase EC 4.1.1.28	*AADC;* 7p11	[b]PN; 400–800 mg/day [c]PLP; 200 mg/day	Improvement of Parkinsonian movement disorder
Pyridoxine responsive anemia (X-linked sideroblastic anemia) (300751/301300)	Δ-aminolevulinate δ-ALA synthase EC 2.3.1.37	*ALAS2;* Xp11.21	PN; 50–400 mg/d	Resolution of anemia
Primary hyperoxaluria type I (259900/604285)	Liver-specific alanine/glyoxylate aminotransferase EC 2.6.1.44	*AGXT;* 2p37.3 (c.508G>A)	PN; 5–10 mg/kg/day	Reduction/normalization of hyperoxaluria. Urinary tract stone formation and subsequent renal failure reduced
Cystathioninuria (219500/607657)	γ-cystathionase EC 4.4.1.1	*CTH;* 1p31.1	PN; 100 mg/d	Reduction of cystathioninaemia/uria
Phosphoserine aminotransferase deficiency (610936/610992)	Phosphoserine aminotransferase EC 2.6.1.52	*PSAT1;* 9q21.31	PN; 120 mg/d	No effect observed in single case reported
McArdles's disease; glycogen storage Disease type V (232600/608455)	Muscle glycogen phosphorylase EC 2.4.1.1	*PYGM;* 11q13	PN; 50–100 mg/d	Reduced exercise intolerance and cramp

PN—pyridoxine, PLP—pyridoxal 5′-phosphate.

[a]Ensure patient is folate replete.

[b]Effective to augment L-dopa treatment (not as single therapeutic agent). Some cases report no improvement.

[c]In twin patients being treated with tranylcypromine, pergolide and pyridoxine, substitution of pyridoxine with pyridoxal phosphate led to improved symptom control.

Saudubray, J.M., Baumgartner, M.R., Walter, J., 2016. Inborn metabolic diseases: diagnosis and treatment. sixth ed. Berlin, Heidelberg, New York: Springer-Verlag.

Inflammation

The anti-inflammatory properties of vitamin B_6 are involved in maintaining a normal immune response (Du et al., 2020). B_6 helps prevent excessive inflammation in macrophages by inhibiting lipopolysaccharide-induced expression of COX-2 and iNOS thereby suppressing the NF-KB signaling and mitogen-activated protein kinase (MAPK) signaling pathways which when activated cause release of inflammatory factors e.g., interleukin-1 beta, tumor necrosis factor alpha, Il-6 and nitric oxide. The exact mechanisms involved are unclear although plasma and liver PLP show an inverse association with inflammatory markers (Ueland et al., 2017). A possible mechanism is mobilization of B_6 to sites of inflammation so that it is available as a cofactor for enzymes involved in pathways that modulate the immune response. Recently it has been shown that B_6 supplementation enhances the activity of sphingosine-1-phosphate (S1P) lyase, a PLP-dependent enzyme, reducing accumulation of S1P and suppressing excessive inflammation. S1P affects the activation of NF-KB signaling and mitogen-activated protein kinase (MAPK) signaling pathways. Other relevant B_6-dependent inflammatory pathways include the kynurenine, serine, glycine and transulfuration pathways (Ueland et al., 2017). Vitamin B_6 also has antioxidant properties, which are likely to also play a role in its anti-inflammatory properties, hydroxyl and amine substituents of the pyridine ring directly reacting with peroxy radicals (Contestabile et al., 2020).

Cardiovascular disease

Several studies have shown that vitamin B_6 deficiency is associated with atherosclerosis, stroke and thrombosis and is a risk factor for cardiovascular outcomes and mortality in coronary artery disease. Prospective studies have also shown that a higher dietary intake of vitamin B_6 is inversely associated with cardiovascular disease (CVD) and that high-risk patients scheduled for a coronary artery bypass graft benefitted from oral administration of PLP. However, epidemiological studies have yielded controversial results, with meta-analysis of several randomized control trials having suggested that vitamin supplements, including B_6, are ineffective at preventing cardiovascular outcomes (Kumrungsee et al., 2022; Stach et al., 2021; Jeon and Park, 2019).

Proposed cardioprotective mechanisms of vitamin B_6 include: (i) lowering of homocysteine, a risk factor for CVD and stroke, through the B_6-dependent transsulfuration pathway; (ii) lowering of chronic inflammation through modulation of the B_6-dependent kynurenine pathway, kynurenine-related metabolites may have beneficial anti-inflammatory effects; (iii) suppression of NF-KB activation (as described above) and (iv) inhibition of inflammation through B_6 inhibiting the nucleotide-binding oligomerization domain-like receptor family pyrin domain-containing 3 (NLRP3) inflammasome. Vitamin B_6 may also supress oxidative stress, by increasing levels of cardiac imidazole dipeptides and regulate Ca^{2+} influx into cells via voltage-gated ATP-dependent purinergic receptors (Kumrungsee et al., 2022; Stach et al., 2021).

Infection

Vitamin B_6 is involved in the production of interleukins and T lymphocytes. A deficiency of B_6 is associated with lower immune function and increased susceptibility to viral infection. It leads to decreased production of IL-2 and increase in IL-4. In conditions of chronic inflammation there is an inverse association of B_6 with IL-6 and TNF-α levels (Stach et al., 2021). In mice, vitamin B_6 had beneficial effects against pneumonia, inhibiting macrophage activation and preventing lipopolysaccharide-induced systemic inflammation (Kumrungsee et al., 2020). Mice fed a B_6-deficient diet were more susceptible to murine pneumonia virus.

Cancer

A deficiency of B_6 not only leads to an increased risk of developing certain types of tumors, including those affecting the lungs and gastrointestinal tract, but is also observed in patients with other forms of cancer. The latter may be due to increased biosynthesis requirements of proliferating tumor cells, with PLP being involved as a cofactor in several biosynthetic pathways. Enzymes involved in B_6 metabolism are differentially expressed in cancer tissues (Contestabile et al., 2020). PNPO expression is upregulated significantly in most cancer types and has been proposed as a prognostic factor, although is down-regulated in adrenocortical and kidney renal clear cell carcinomas, acute myeloid leukemia and testicular germ cell tumors (Zhang et al., 2022). PDXK is also upregulated in non-small cell lung cancer and is abundantly expressed in myeloid leukemia cells and when depleted causes anti-proliferative effects (Contestabile et al., 2020). However, the role of vitamin B_6 in cancer is multifaceted as it can also stimulate an immune response and it is an effective inhibitor of the formation of advanced glycation end products. It also functions as an antioxidant protecting DNA from damage, having singlet oxygen quench capacity comparable to antioxidants vitamin C and E (Shen, 2015). The hydroxyl and amine substituents on the pyridine ring of vitamin B_6 directly react with peroxy radicals (Contestabile et al., 2020). B_6 is also involved in the generation of glutathione, a key regulator of redox state within cells, through its role as a cofactor in the transulfuration pathway. However, it can also lead to elevated homocysteine levels, which in turn result in ROS generation. In *Drosophila*, a deficiency of PLP causes chromosome aberrations which can be rescued by PLP suggesting that B_6 is also important for maintaining chromosomal integrity (Contestabile et al., 2020).

Parkinson's disease

While the number of studies that have looked at vitamin B_6 intake and Parkinson's Disease (PD) are limited, meta-analysis suggests that higher dietary intake may be associated with a decreased risk of Parkinson's disease (Shen, 2015). Patients with PD have also been reported to have increased levels of homocysteine which may lead to dopaminergic cell death. Given that PLP is required as a cofactor by two of the enzymes in homocysteine metabolism, increased B_6 intake may decrease plasma homocysteine and the risk of PD. PLP is also required as a cofactor for the final step in dopamine synthesis where L-DOPA is converted to dopamine by the enzyme aromatic L-amino acid decarboxylase; PD is characterized by the degeneration of dopaminergic neurons resulting in decreased levels of the neurotransmitter dopamine. Pyridoxine has been shown recently to induce synthesis of GSH, an important antioxidant in the brain which is generated from homocysteine, in a PD mouse model and to reduce dopaminergic neuron loss (Wei et al., 2020). Besides its role as a cofactor, the antioxidant properties of B_6 may also contribute to the protective effects proposed.

Symptomatic treatment of PD is usually achieved using L-DOPA, which is typically co-administered with a DOPA decarboxylase inhibitor, such as carbidopa, to prevent peripheral breakdown. Plasma vitamin B_6 levels should be monitored in PD patients receiving treatment with high levodopa (L-DOPA)/carbidopa doses as carbidopa binds covalently and irreversibly to PLP and can cause B_6 depletion (Rojo-Sebastián et al., 2020).

Conclusion/summary/outlook

B_6 vitamers in the diet are converted to intracellular pyridoxal phosphate (PLP) which is the cofactor for at least 70 enzyme-catalyzed reactions. Within the cell, the concentration of PLP must be kept low to avoid unwanted reactions of this reactive cofactor while at the same time ensuring its efficient transfer to apoenzymes. This intracellular homeostasis is the subject of current research.

Deficiency of PLP in the brain can give rise to seizures (probably attributable, at least in part, to reduced activity of glutamate decarboxylase and hence reduced levels of the inhibitory neurotransmitter g-aminobutyric acid [GABA], and to a movement disorder (probably due to reduced activity of aromatic amino acid decarboxylase leading to low levels of dopamine).

In the peripheral nervous system, lack of PLP can cause a sensory-motor neuropathy. Paradoxically ingestion of large doses of pyridoxine also causes a neuropathy, perhaps because pyridoxine or pyridoxine phosphate inhibits PLP-dependent enzymes.

In the hematopoietic system, PLP deficiency can cause anemia because heme synthesis is disrupted by reduced activity of D-aminolaevulinic acid synthase.

PLP- enzymes play important roles in inflammation and the immune response and a B_6 deficient diet predisposes to viral pneumonia in the mouse.

B_6 vitamers have roles in addition to supplying PLP to PLP-dependent apo-enzymes. They have antioxidant properties, bind to some transcription factors and corepressors, affect the expression and function of steroid hormone receptors and contribute to genome stability.

The optimum personalized intake of B_6 for individuals with a common disease cannot easily be determined currently but should be a focus of future research. Deficiencies of specific PLP-dependent enzymes can, in some cases, be very effectively treated with supra-physiological doses of B_6. Disorders of the synthesis and recycling of PLP may also respond dramatically to treatment with pyridoxine or pyridoxal phosphate.

Deficiency of PLP because of insufficient B_6 in the diet is rare probably because, at least after infancy, the gut flora can synthesize significant amounts of B_6. PLP deficiency can, however, occur if we ingest something that binds PLP e.g., large amounts of Gingko nuts or hydralazine.

Acknowledgments

P.M and P.C. are supported by funding from the NIHR Great Ormond Street Hospital Biomedical Research Center. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health.

References

Albersen, M., et al., 2013. The intestine plays a substantial role in human vitamin B6 metabolism: a Caco-2 cell model. PLoS One 8 (1), e54113.
Albin, R.L., Albers, J.W., Greenberg, H.S., et al., 1987. Acute sensory neuropathy-neuronopathy from pyridoxine overdose. Neurology 37 (11), 1729–1732.
Ames, E.G., Scott, A.J., Pappas, K.B., et al., 2020. A cautionary tale of pyridoxine toxicity in cystathionine beta-synthase deficiency detected by two-tier newborn screening highlights the need for clear pyridoxine dosing guidelines. Am. J. Med. Genet. 182 (11), 2704–2708.
Beamer, E., Fischer, W., Engel, T., 2017. The ATP-Gated P2X7 Receptor as a target for the treatment of drug-resistant epilepsy. Front. Neurosci. 11, 21.
Bender, D.A., 1999. Non-nutritional uses of vitamin B6. Br. J. Nutr. 81 (1), 7–20.
Berger, M.M., Shenkin, A., Amrein, K., et al., 2022. ESPEN micronutrient guideline. Clin. Nutr. S0261–5614 (22), 00066–00068.
Bilski, P., Li, M.Y., Ehrenshaft, M., Daub, M.E., Chignell, C.F., 2000. Vitamin B6 (pyridoxine) and its derivatives are efficient singlet oxygen quenchers and potential fungal antioxidants. Photochem. Photobiol. 71 (2), 129–134.
Bottiglieri, T., Hyland, K., Reynolds, E.H., 1994. The clinical potential of ademetionine (S-adenosylmethionine) in neurological disorders. Drugs 48 (2), 137–152.
Canham, J.E., Baker, E.M., Harding, R.S., Sauberlich, H.E., Plough, I.C., 1969. Dietary protein-its relationship to vitamin B6 requirements and function. Ann. N. Y. Acad. Sci. 166 (1), 16–29.
Christmann-Franck, S., Fermandjian, S., Mirambeau, G., Der Garabedian, P.A., 2007. Molecular modelling studies on the interactions of human DNA topoisomerase IB with pyridoxal-compounds. Biochimie 89 (4), 468–473.
Ciapaite, J., Albersen, M., Savelberg, S.M.C., et al., 2020. Pyridox(am)ine 5'-phosphate oxidase (PNPO) deficiency in zebrafish results in fatal seizures and metabolic aberrations. Biochim. Biophys. Acta 1866 (3), 165607.
Cohen, M., Bendich, A., 1986. Safety of pyridoxine–a review of human and animal studies. Toxicol. Lett. 34 (2–3), 129–139.
Coman, D., Lewindon, P., Clayton, P., Riney, K., 2016. PNPO deficiency and cirrhosis: expanding the clinical phenotype? JIMD Rep. 25, 71–75.
Contestabile, R., di Salvo, M.L., Bunik, V., Tramonti, A., Vernì, F., 2020. The multifaceted role of vitamin B6 in cancer: Drosophila as a model system to investigate DNA damage. Open Biol. 10 (3), 200034.
Coursin, D.B., 1954. Convulsive seizures in infants with pyridoxine-deficient diet. J. Am. Med. Assoc. 154 (5), 406–408.
Dalton, K., Dalton, M.J., 1987. Characteristics of pyridoxine overdose neuropathy syndrome. Acta Neurol. Scand. 76 (1), 8–11.
Darin, N., Reid, E., Prunetti, L., et al., 2016. Mutations in PROSC disrupt cellular pyridoxal phosphate homeostasis and cause vitamin-B6-dependent epilepsy. Am. J. Hum. Genet. 99 (6), 1325–1337.
di Salvo, M.L., Contestabile, R., Safo, M.K., 2011. Vitamin B(6) salvage enzymes: mechanism, structure and regulation. Biochim. Biophys. Acta 1814 (11), 1597–1608.
Du, X., Yang, Y., Zhan, X., et al., 2020. Vitamin B6 prevents excessive inflammation by reducing accumulation of sphingosine-1-phosphate in a sphingosine-1-phosphate lyase-dependent manner. J. Cell Mol. Med. 24 (22), 13129–13138.
Eliot, A.C., Kirsch, J.F., 2004. Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. Annu. Rev. Biochem. 73, 383–415.
Farhan, S.M., Wang, J., Robinson, J.F., et al., 2014. Exome sequencing identifies NFS1 deficiency in a novel Fe-S cluster disease, infantile mitochondrial complex II/III deficiency. Mol. Genet. Genomic Med. 2 (1), 73–80.
Gachon, F., Fonjallaz, P., Damiola, F., et al., 2004. The loss of circadian PAR bZip transcription factors results in epilepsy. Genes Dev. 18 (12), 1397–1412.
Gallagher, R.C., Van Hove, J.L., Scharer, G., et al., 2009. Folinic acid-responsive seizures are identical to pyridoxine-dependent epilepsy. Ann. Neurol. 65 (5), 550–556.
Garaycoechea, J.I., Crossan, G.P., Langevin, F., et al., 2018. Alcohol and endogenous aldehydes damage chromosomes and mutate stem cells. Nature 553 (7687), 171–177.
Gibbs, K., Walshe, J.M., 1969. Interruption of the tryptophan-nicotinic acid pathway by penicillamine-induced pyridoxine deficiency in patients with Wilson's disease and in experimental animals. Ann. N. Y. Acad. Sci. 166 (1), 158–169.
Glenn, G.M., Krober, M.S., Kelly, P., McCarty, J., Weir, M., 1995. Pyridoxine as therapy in theophylline-induced seizures. Vet. Hum. Toxicol. 37 (4), 342–345.
Gregory III, J.F., 1980. Effects of epsilon-pyridoxyllysine bound to dietary protein on the vitamin B-6 status of rats. J. Nutr. 110 (5), 995–1005.

Gospe, S.M., Olin, K.L., Keen, C.L., 1994. Reduced GABA synthesis in pyridoxine-dependent seizures. Lancet 343, 1133–1134.
Gould, G.W., Russell, N.J., 2003. Sulfite. In: Russell, N.J., Gould, G.W. (Eds.), Food Preservatives. Springer, Boston, MA, pp. 85–101.
Hammen, A., Wagner, B., Berkhoff, M., Donati, F., 1998. A paradoxical rise of neonatal seizures after treatment with vitamin B6. Eur. J. Paediatr. Neurol. 2 (6), 319–322.
Hayman, D.J., Modebadze, T., Charlton, S., et al., 2021. Increased hippocampal excitability in miR-324-null mice. Sci. Rep. 11, 10452.
Hartmann, H., Fimgerhut, M., Jakobs, C., Plecko, B., 2011. Status epilepticus in a neonate treated with pyridoxine because of a familial recurrence risk for antiquitin deficiency: pyridoxine toxicity. Dev. Med. Child Neurol. 53 (12), 1150–1153.
Hasegawa, S., Oda, Y., Ichiyama, T., Hori, Y., Furukawa, S., 2006. Ginkgo nut intoxication in a 2-year-old male. Pediatr. Neurol. 35 (4), 275–276.
Hyland, K., Surtees, R.A., Rodeck, C., Clayton, P.T., 1992. Aromatic L-amino acid decarboxylase deficiency: clinical features, diagnosis, and treatment of a new inborn error of neurotransmitter amine synthesis. Neurology 42 (10), 1980–1988.
Huq, M.D., Tsai, N.P., Lin, Y.P., Higgins, L., Wei, L.N., 2007. Vitamin B6 conjugation to nuclear corepressor RIP140 and its role in gene regulation. Nat. Chem. Biol. 3 (3), 161–165.
Institute of Medicine Food and Nutrition Board, 1998. Dietary Reference Intakes: Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline. National Academy Press, Washington, DC.
Ishioka, N., et al., 1995. In vivo modification of GABAA receptor with a high dose of pyridoxal phosphate induces tonic-clonic convulsion in immature mice. Neurochem. Int. 26 (4), 369–373.
Johnstone, D.L., Al-Shekaili, H.H., Tarailo-Graovac, M., et al., 2019. PLPHP deficiency: clinical, genetic, biochemical, and mechanistic insights. Brain 142 (3), 542–559.
Jeon, J., Park, K., 2019. Dietary vitamin B6 intake associated with a decreased risk of cardiovascular disease: a prospective cohort study. Nutrients 11 (7), 1484.
Kajita, R., Goto, T., Lee, S.H., Oe, T., 2013. Aldehyde stress-mediated novel modification of proteins: epimerization of the N-terminal amino acid. Chem. Res. Toxicol. 26 (12), 1926–1936.
Kakizaki, T., Ohshiro, T., Itakura, M., et al., 2021. Rats deficient in the GAD65 isoform exhibit epilepsy and premature lethality. Faseb J. 35 (2), e21224.
Klosterman, H.J., Lamoureux, G.L., Parsons, J.L., 1967. Isolation, characterization, and synthesis of linatine. A vitamin B6 antagonist from flaxseed (*Linum usitatissimum*). Biochemistry 6 (1), 170–177.
Kratzer, F.H., Williams, D.E., 1948. The relation of pyridoxine to the growth of chicks fed rations containing linseed oil meal. J. Nutr. 36 (2), 297–305.
Kumrungsee, T., Zhang, P., Chartkul, M., Yanaka, N., Kato, N., 2020. Potential role of vitamin B6 in ameliorating the severity of COVID-19 and its complications. Front. Nutr. 29 (7), 562051.
Kumrungsee, T., Peipei, Z., Yanaka, N., Suda, T., Kato, N., 2022. Emerging cardioprotective mechanisms of vitamin B6: a narrative review. Eur. J. Nutr. 61 (2), 605–613.
Levine, S., Saltzman, A., 2004. Pyridoxine (vitamin B6) neurotoxicity: enhancement by protein-deficient diet. J. Appl. Toxicol. 24 (6), 497–500.
Lheureux, P., Penazola, A., Gris, M., 2005. Pyridoxine in clinical toxicology: a review. Eur. J. Emerg. Med. 12, 79–85.
Mackey, A.D., Henderson, G.N., Gregory III, J.F., 2002. Enzymatic hydrolysis of pyridoxine-5'-beta-D-glucoside is catalyzed by intestinal lactase-phlorizin hydrolase. J. Biol. Chem. 277 (30), 26858–26864.
Mackey, A.D., McMahon, R.J., Townsend, J.H., Gregory III, J.F., 2004. Uptake, hydrolysis, and metabolism of pyridoxine-5'-beta-D-glucoside in Caco-2 cells. J. Nutr. 134 (4), 842–846.
Magnúsdóttir, S., Ravcheev, D., de Crécy-Lagard, V., Thiele, I., 2015. Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. Front. Genet. 20 (6), 148.
Marzio, A., Merigliano, C., Gatti, M., Vernì, F., 2014. Sugar and chromosome stability: clastogenic effects of sugars in vitamin B6-deficient cells. PLoS Genet. 10 (3), e1004199.
McCulloch, K.M., Mukherjee, T., Begley, T.P., Ealick, S.E., 2010. Structure determination and characterization of the vitamin B6 degradative enzyme (E)-2-(acetamidomethylene) succinate hydrolase. Biochemistry 49 (6), 1226–1235.
Morra, M., et al., 1993. Sensory and motor neuropathy caused by excessive ingestion of vitamin B6: a case report. Funct. Neurol. 8 (6), 429–432.
Nijhout, H.F., Gregory, J.F., Fitzpatrick, C., et al., 2009. A mathematical model gives insights into the effects of vitamin B-6 deficiency on 1-carbon and glutathione metabolism. J. Nutr. 139 (4), 784–791.
Nishijima, S., Sugaya, K., Morozumi, M., Hatano, T., Ogawa, Y., 2003. Hepatic alanine-glyoxylate aminotransferase activity and oxalate metabolism in vitamin B6 deficient rats. J. Urol. 169 (2), 683–686.
Oka, T., Sugitatsu, H., Nordin, H., et al., 2001. Pyridoxal 5'-phosphate inhibits DNA binding of HNF1. Biochim. Biophys. Acta 1568 (3), 189–196.
Pena, I.A., Roussel, Y., Daniel, K., et al., 2017. Pyridoxine-dependent epilepsy in zebrafish caused by Aldh7a1 deficiency. Genetics 207 (4), 1501–1518.
Phillips, W.E., et al., 1978. Subacute toxicity of pyridoxine hydrochloride in the beagle dog. Toxicol. Appl. Pharmacol. 44 (2), 323–333.
Porter, K., Hoey, L., Hughes, C.F., Ward, M., McNulty, H., 2016. Causes, consequences and public health implications of low B-vitamin status in ageing. Nutrients 8 (11), 725.
Prabhu, D., Dawe, R.S., Mponda, K., 2021. Pellagra a review exploring causes and mechanisms, including isoniazid-induced pellagra. Photodermatol. Photoimmunol. Photomed. 37 (2), 99–104.
Ramos, R.J., Pras-Raves, M.L., Gerrits, J., et al., 2017. Vitamin B6 is essential for serine de novo biosynthesis. J. Inherit. Metab. Dis. 40 (6), 883–891.
Raschle, T., Speziga, D., Kress, W., et al., 2009. Intersubunit cross-talk in pyridoxal 5'-phosphate synthase, coordinated by the c terminus of the synthase subunit. J. Biol. Chem. 284 (12), 7706–7718.
Raskin, N.H., Fishman, R.A., 1965. Pyridoxine-deficiency neuropathy due to hydralazine. N. Engl. J. Med. 273, 1182–1185.
Ribaya-Mercado, J.D., Russell, R.M., Sahyoun, N., Morrow, F.D., Gershoff, S.N., 1991. Vitamin B-6 requirements of elderly men and women. J. Nutr. 121 (7), 1062–1074.
Rojo-Sebastián, A., González-Robles, C., de Yébenes, J.C., 2020. Vitamin B6 deficiency in patients with Parkinson disease treated with levodopa/carbidopa. Clin. Neuropharmacol. 43 (5), 151–157.
Romagnolo, A., Merola, A., Artusi, C.A., Rizzone, M.G., Zibetti, M., Lopiano, L., 2018. Levodopa-induced neuropathy: a systematic review. Mov. Disord. Clin. Pract. 6 (2), 96–103.
Salhany, J.M., Schopfer, L.M., 1993. Pyridoxal 5'-phosphate binds specifically to soluble CD4 protein, the HIV-1 receptor. Implications for AIDS therapy. J. Biol. Chem. 268 (11), 7643–7645.
Sasaoka, K., Ogawa, T., Moritoki, K., Kimoto, M., 1976. Antivitamin B-6 effect of 1-aminoproline on rats. Biochim. Biophys. Acta 428 (2), 396–402.
Saudubray, J.M., Baumgartner, M.R., Walter, J., 2016. Inborn Metabolic Diseases: Diagnosis and Treatment, sixth ed. Springer-Verlag, Berlin, Heidelberg, New York.
Schaumburg, H., et al., 1983. Sensory neuropathy from pyridoxine abuse. A new megavitamin syndrome. N. Engl. J. Med. 309 (8), 445–448.
Scott, T.A., Quintaneiro, L.M., Norvaisas, P., et al., 2017. Host-microbe co-metabolism dictates cancer drug efficacy in *C. elegans*. Cell 169 (3), 442–456.
Shen, L., 2015. Associations between B vitamins and Parkinson's disease. Nutrients 7 (9), 7197–7208.
Stach, K., Stach, W., Augoff, K., 2021. Vitamin B6 in health and disease. Nutrients 13 (9), 3229.
Sudarsanam, A., et al., 2014. Cirrhosis associated with pyridoxal 5'-phosphate treatment of pyridoxamine 5'-phosphate oxidase deficiency. JIMD Rep. 17, 67–70.
Takagi, M., Fukui, T., Shimomura, S., 1982. Catalytic mechanism of glycogen phosphorylase: pyridoxal(5')diphospho(1)-alpha-D-glucose as a transition-state analogue. Proc. Natl. Acad. Sci. U. S. A. 79 (12), 3716–3719.
Thériault, O., Poulin, H., Thomas, G.R., et al., 2014. Pyridoxal-5'-phosphate (MC-1), a vitamin B6 derivative, inhibits expressed P2X receptors. Can. J. Physiol. Pharmacol. 92 (3), 189–196.
Tully, D.B., Allgood, V.E., Cidlowski, J.A., 1994. Modulation of steroid receptor-mediated gene expression by vitamin B6. Faseb J. 8 (3), 343–349.
Ubbink, J.B., Delport, R., Bissbort, S., Vermaak, W.J., Becker, P.J., 1990. Relationship between vitamin B-6 status and elevated pyridoxal kinase levels induced by theophylline therapy in humans. J. Nutr. 120 (11), 1352–1359.
Ueland, P.M., Ulvik, A., Rios-Avila, L., Midttun, Ø., Gregory, J.F., 2015. Direct and functional biomarkers of vitamin B6 status. Annu. Rev. Nutr. 35, 33–70.
Ueland, P.M., McCann, A., Midttun, Ø., Ulvik, A., 2017. Inflammation, vitamin B6 and related pathways. Mol. Aspect. Med. 53, 10–27.

Unna, J.C., 1940. Studies of the toxicity and pharmacology of vitamin B6. J. Pharmacol. Exp. Therapeut. 70, 400–407.
Vrolijk, M.F., et al., 2017. The vitamin B6 paradox: supplementation with high concentrations of pyridoxine leads to decreased vitamin B6 function. Toxicol. Vitro 44, 206–212.
Wang, H.S., et al., 2005. Pyridoxal phosphate is better than pyridoxine for controlling idiopathic intractable epilepsy. Arch. Dis. Child. 90 (5), 512–515.
Wei, Y., Lu, M., Mei, M., et al., 2020. Pyridoxine induces glutathione synthesis via PKM2-mediated Nrf2 transactivation and confers neuroprotection. Nat. Commun. 11 (1), 941.
Wilson, M.P., Plecko, B., Mills, P.B., Clayton, P.T., 2019. Disorders affecting vitamin B6 metabolism. J. Inherit. Metab. Dis. 42 (4), 629–646.
Yamamoto, K., Isa, Y., Nakagawa, T., Hayakawa, T., 2012. Folic acid fortification ameliorates hyperhomocysteinemia caused by a vitamin B_6-deficient diet supplemented with L-methionine. Biosci. Biotechnol. Biochem. 76 (10), 1861–1865.
Yoshida, I., Sakaguchi, Y., Nakano, M., Yamashita, F., Hitoshi, T., 1985. Pyridoxal phosphate-induced liver injury in a patient with homocystinuria. J. Inherit. Metab. Dis. 8 (2), 91.
Yun, S., Kim, W., Kang, M.S., et al., 2020. Neuropathological changes in dorsal root ganglia induced by pyridoxine in dogs. BMC Neurosci. 21 (1), 11.
Zhang, L., Li, X., Zhang, J., Xu, G., 2022. Prognostic implication and oncogenic role of PNPO in pan-cancer. Front. Cell Dev. Biol. 9, 763674.

Vitamin C

David A. Bender[a] **and Luis Fontana**[b,c,d], [a] Department of Structural and Molecular Biology, University College London, London, United Kingdom; [b] Department of Biochemistry & Molecular Biology 2, School of Pharmacy, Granada, Spain; [c] Institute of Nutrition & Food Technology "José Mataix", Biomedical Research Center, Parque Tecnológico Ciencias de la Salud, Granada, Spain; and [d] Instituto de Investigación Biosanitaria ibs.GRANADA, Complejo Hospitalario Universitario de Granada, Granada, Spain

© 2023 Published by Elsevier Ltd.

This is an update of D. A. Bender, Vitamin C: Physiology, Dietary Sources, and Requirements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 363–369, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00019-2.

Introduction **505**
Chemical structure **505**
Absorption, transport, and storage **505**
Metabolism and excretion **506**
Metabolic functions of ascorbic acid **507**
Copper-containing hydroxylases 507
2-Oxoglutarate-linked, iron-containing hydroxylases 507
The role of ascorbate in iron absorption 508
Inhibition of nitrosamine formation 508
Antioxidant and prooxidant actions of ascorbate 508
Assessment of vitamin C status **509**
Requirements **509**
Requirements estimated from maintenance of the body pool of ascorbate 510
Dietary sources and high intakes **510**
Vitamin C and gene expression **511**
Vitamin C and demethylation of DNA 511
Vitamin C and demethylation of histones 512
Vitamin C and disease **512**
Scurvy 512
Common cold 513
Cardiovascular disease 513
Cancer 513
Cataracts 513
Asthma 513
Pneumonia 513
Pregnancy complications 513
Conclusions **513**
References **514**

Key points

- To learn the processes of absorption, transport, storage, metabolism and excretion of vitamin C.
- To understand the metabolic functions of ascorbic acid.
- To have information about the requirements and the assessment of vitamin C status.
- To know the main dietary sources of this vitamin.
- To know the recent discovered role of ascorbate as an epigenetic regulator.
- To recognize the relationship between ascorbate and disease.

Glossary

Carnitine β-Hydroxy-γ-trimethylammonium butyrate, biosynthesized by methylation of lysine, important for the transport of fatty acids into the mitochondrion for oxidation

Epigenetics The study of how cells control gene activity without changing the DNA sequence. "Epi-" means on or above in Greek, and "epigenetic" describes factors beyond the genetic code. Epigenetic changes are modifications to DNA that regulate whether genes are turned on or off

Flavoprotein Any enzyme (or other protein) with a prosthetic group formed from riboflavin (vitamin B_2) or one of its derivatives
Glutathione A tripeptide, γ-glutamyl–cysteinyl–glycine that acts as a major redox cofactor, being oxidized to yield the hexapeptide lined by a disulfide bridge
Granulocytes A class of white blood cells characterized by conspicuous cytoplasmic granules
Hydroxymethylglutaryl CoA reductase The first and rate-limiting enzyme of cholesterol biosynthesis. The target of statin drugs for the treatment of hypercholesterolemia
Hypercholesterolemia Elevated blood concentration of cholesterol
Leukocytes White blood cells of various kinds
Teleost fish Bony, as opposed to cartilaginous, fishes

Introduction

Ascorbic acid is a vitamin (vitamin C) for only a limited number of species: human beings and other primates, bats, the guinea pig, a number of birds, and teleost fishes. In other species ascorbic acid is not a vitamin but is an intermediate in glucuronic acid catabolism, and its rate of synthesis bears no relation to physiological requirements for ascorbate (Combs and McClung, 2016). Species for which ascorbate is a vitamin lack the enzyme gulonolactone oxidase (EC 1.1.3.8) and have an alternative pathway for glucuronic acid metabolism, via reduction and the pentose phosphate pathway (Bender, 2003; Rodwell et al., 2018).

Ascorbic acid functions as a relatively nonspecific, radical-trapping antioxidant, and also reduces the tocopheroxyl radical formed by oxidation of vitamin E. It has a specific metabolic function as the coenzyme for dopamine β-hydroxylase and peptidyl glycine hydroxylase, and it is required to maintain the iron of 2-oxoglutarate-dependent hydroxylases in the reduced state. It also enhances the absorption of inorganic iron from foods (Ross et al., 2014; Stipanuk and Caudill, 2018).

Chemical structure

The term vitamin C encompasses all compounds with the biological activity of L-ascorbic acid (2,3-enediol gulonic acid or 2-oxo-L-threo-hexono-1,4-lactone-2,3-enediol). It is a chemically simple compound although it presents an atypical structure, whose empirical formula is C6H8O6. It is a lactonic derivative of hexuronic acid and corresponds to an oxidized form of glucose. Specifically, it is a δ-ketolactone of six carbon atoms showing a five-membered lactone ring and a bifunctional enediol group with an adjacent carbonyl group. The enediol group is essential for its biological activity (**Fig. 1**).

Absorption, transport, and storage

In species for which ascorbate is not a vitamin, intestinal absorption is by sodium-independent facilitated diffusion, whereas in human beings and guinea pigs there is sodium-dependent active transport of the vitamin at the brush border membrane, with a sodium-independent mechanism at the basolateral membrane. Dehydroascorbate is absorbed passively in the intestinal mucosa, and is reduced to ascorbate before transport across the basolateral membrane (Bender, 2003).

At intakes up to about 100 mg day^{-1}, 80–95% of dietary ascorbate is absorbed, falling from 50% of a 1 g to 25% of a 6 g and 16% of a 12 g dose. Unabsorbed ascorbate is a substrate for intestinal bacterial metabolism.

Ascorbate and dehydroascorbate circulate in the bloodstream both in free solution and bound to albumin. Approximately 5% of plasma vitamin C is normally in the form of dehydroascorbate. Ascorbate enters cells by sodium-dependent active transport; dehydroascorbate is transported by insulin-dependent glucose transporters and is accumulated intracellularly by reduction to ascorbate. In poorly controlled diabetes mellitus, tissue uptake of dehydroascorbate is impaired because of competition by glucose, as well as insulin resistance or lack of insulin, which is required for activity of the transporters, and there may be functional deficiency of vitamin C despite an apparently adequate intake.

Approximately 70% of blood-borne ascorbate is in plasma and erythrocytes (which do not concentrate the vitamin from plasma). The remainder is in white cells, which have a marked ability to concentrate ascorbate; mononuclear leukocytes achieve 80-fold concentration, platelets 40-fold, and granulocytes 25-fold, compared with the plasma concentration.

There is no specific storage organ for ascorbate; apart from leukocytes (which account for 10% of total blood ascorbate), the only tissues showing a significant concentration of the vitamin are the adrenal and pituitary glands. Although the concentration of ascorbate in muscle is relatively low, skeletal muscle contains much of the body pool of 5–8.5 mmol (900–1500 mg) of ascorbate.

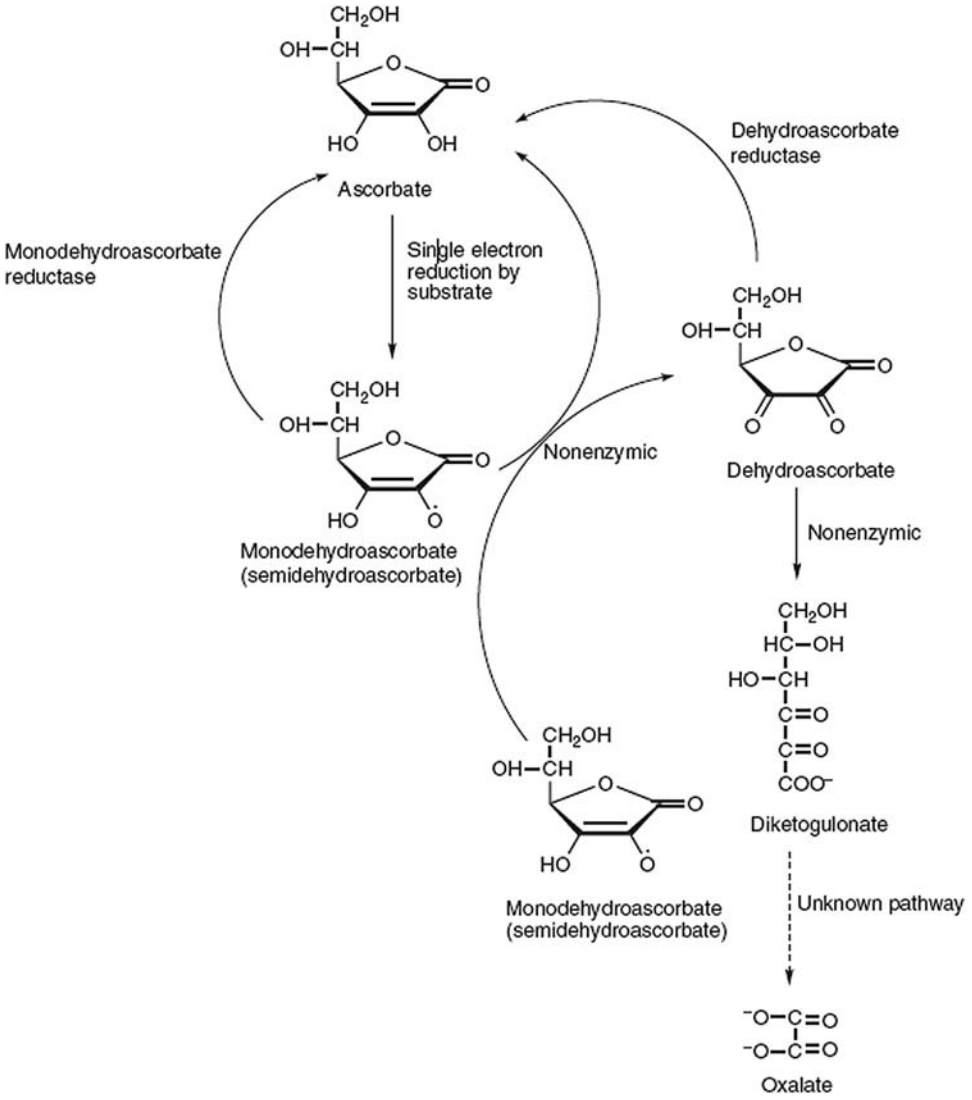


Fig. 1 The metabolism of ascorbate.

Metabolism and excretion

As shown in Fig. 1, oxidation of ascorbic acid proceeds by a one-electron process, forming monodehydroascorbate, which disproportionates to ascorbate and dehydroascorbate. Most tissues also contain monodehydroascorbate reductase (EC 1.6.5.4), a flavoprotein that reduces the radical back to ascorbate. In the cytosol dehydroascorbate is reduced by either reduced nicotinamide adenine dinucleotide (NADH) or reduced nicotinamide adenine dinucleotide phosphate (NADPH)-dependent enzymes; in mitochondria it is reduced by complex 3 of the electron transport chain. It is also reduced nonenzymatically by reaction with glutathione. At neutral pH dehydroascorbate is unstable and is rapidly hydrated to diketogulonate, but little is lost in this way.

Both ascorbate and dehydroascorbate are filtered at the glomerulus, then reabsorbed by the same transporters as are involved in intestinal absorption; the sodium-dependent vitamin C transporter for ascorbate and the glucose transporter for dehydroascorbate. When glomerular filtration exceeds the capacity of the transport systems, at a plasma concentration of ascorbate above about 85 μmol L^{-1}, the vitamin is excreted in the urine in amounts proportional to intake.

Approximately 1.5% of dietary ascorbate is excreted as oxalate; this may be a factor in development of oxalate renal stones. The pathway of oxalate formation from ascorbate is not known, and it is likely that there is nonenzymic cleavage of diketogulonate to oxalate. Ascorbate also increases the intestinal absorption of dietary oxalate.

Metabolic functions of ascorbic acid

Ascorbic acid has specific and well-defined roles in two classes of enzymes: copper-containing hydroxylases and the 2-oxoglutarate-linked, iron-containing hydroxylases. It also increases the activity of a number of other enzymes *in vitro*—a nonspecific reducing action rather than reflecting a metabolic function of the vitamin. In addition, ascorbic acid has a number of less specific effects due to its action as a reducing agent and oxygen-radical quencher. There is also evidence that ascorbate has a role in regulating the expression of connective tissue protein (and some other) genes (Bender, 2003).

Copper-containing hydroxylases

Dopamine β-hydroxylase (EC 1.14.17.1) is a copper-containing enzyme involved in the synthesis of the catecholamines noradrenaline and adrenaline from tyrosine in the adrenal medulla and central nervous system. The active enzyme contains Cu^{+}, which is oxidized to Cu^{2+} during the hydroxylation of the substrate; reduction back to Cu^{+} specifically requires ascorbate, which is oxidized to monodehydroascorbate. In the chromaffin granules, monodehydroascorbate is reduced by transmembrane electron transport via cytochrome b_{561}, with electrons provided by ascorbate in the cytosol. The resultant monodehydroascorbate in the cytosol is reduced back to ascorbate by a mitochondrial outer membrane reductase.

More than half of the peptide hormones undergo postsynthetic modification to form a carboxyl terminal amide, which is essential for biological activity. One function of this amidation is to render the peptides more hydrophobic, and so enhance receptor binding. The amide group is derived from a glycine residue that is to the carboxyl side of the amino acid that will become the amidated terminal of the mature peptide. The reaction is catalyzed by peptidyl glycine hydroxylase (peptidyl α-amidase, EC 1.14.17.3).

The first step in the reaction is hydroxylation of the glycine residue to yield hydroxyglycine. The hydroxylase is a copper-containing oxygenase that uses ascorbate to reduce the two copper ions to Cu^{+}; these then activate oxygen, with incorporation of one atom into hydroxyglycine and reduction of the other to water. This is followed by the cleavage of the peptide bond, with amidation of the amino acid to the amino side of the hydroxyglycine residue, and release of glyoxylate. These two activities occur in a single bifunctional protein.

2-Oxoglutarate-linked, iron-containing hydroxylases

A number of iron-containing hydroxylases (listed in **Table 1**) share a common reaction mechanism, in which hydroxylation of the substrate is linked to decarboxylation of 2-oxoglutarate. Ascorbate is required for the activity of all of these enzymes, but it does not function as either a stoichiometric substrate or a conventional coenzyme (which would not be consumed in the reaction).

Proline and lysine hydroxylases are required for the postsynthetic modification of collagen, and proline hydroxylase also for the postsynthetic modification of osteocalcin in bone and the Cl_q component of complement, as well as a number of other proteins that have a collagen-like domain.

Aspartyl β-hydroxylase catalyzes hydroxylation of aspartyl and asparaginyl residues in epidermal growth factor (EGF) and a number of other proteins that have an EGF-like domain, including several of the vitamin K-dependent blood clotting factors.

Trimethyllysine and γ-butyrobetaine hydroxylases are required for the synthesis of carnitine, and many of the early signs of scurvy may be due to carnitine deficiency.

The best studied of this class of enzymes is procollagen proline hydroxylase; it is assumed that the others follow essentially the same mechanism. As shown in **Fig. 2**, the first step is the binding of oxygen to the enzyme-bound iron, followed by attack on the 2-oxoglutarate substrate, resulting in decarboxylation to succinate, leaving a ferryl radical at the active site of the enzyme. This catalyzes the hydroxylation of proline, restoring the free iron to undergo further reaction with oxygen.

It has long been known that ascorbate is oxidized during the reaction, but not stoichiometrically with hydroxylation of proline and decarboxylation of 2-oxoglutarate. The purified enzyme is active in the absence of ascorbate, but after some 5–10 s (approximately 15–30 cycles of enzyme action) the rate of reaction falls. The loss of activity is due to a side-reaction of

Table 1 Vitamin C dependent, 2-oxoglutarate-linked hydroxylases.

Aspartyl β-hydroxylase	EC 1.14.11.16
Histone demethylase	EC 1.14.11.27
p-hydroxyphenylpyruvate hydroxylase	EC 1.14.11.27
Procollagen lysine hydroxylase	EC 1.14.11.4
Procollagen proline 3-hydroxylase	EC 1.14.11.7
Procollagen proline 4-hydroxylase	EC 1.14.11.2
Pyrimidine deoxynucleotide dioxygenase	EC 1.14.11.3
Thymidine dioxygenase	EC 1.14.11.10
Thymine dioxygenase	EC 1.14.11.6
Trimethyllysine hydroxylase	EC 1.14.11.8
γ-Butyrobetaine hydroxylase	EC 1.14.11.1

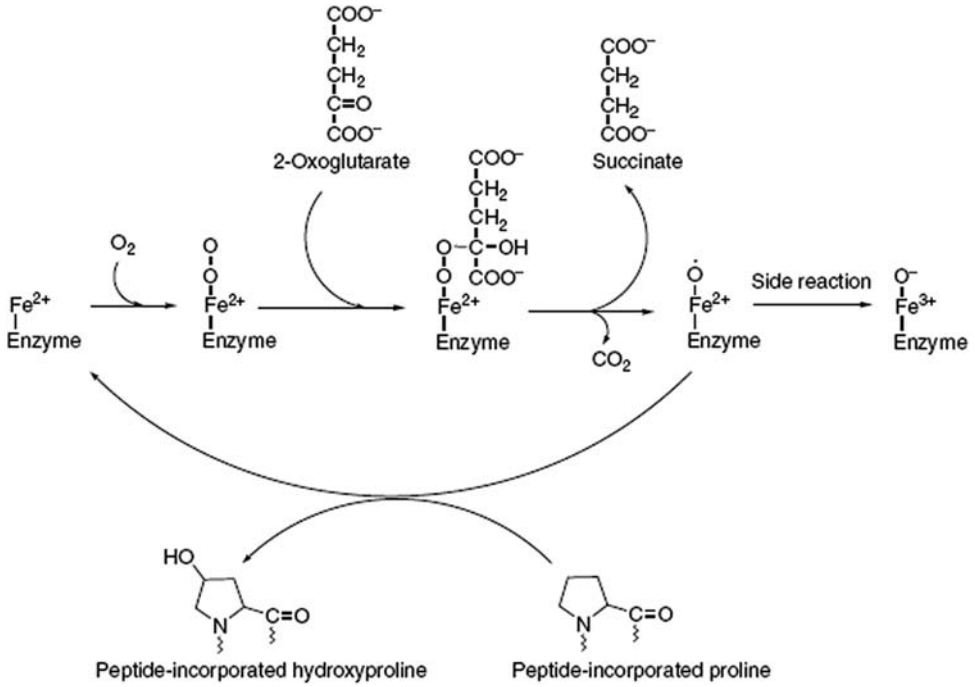


Fig. 2 The reaction of procollagen proline hydroxylase.

the highly reactive ferryl radical in which the iron is oxidized to Fe^{3+}, which is catalytically inactive—so-called uncoupled decarboxylation of 2-oxoglutarate. Activity is only restored by ascorbate, which reduces the iron back to Fe^{2+}.

The role of ascorbate in iron absorption

Inorganic dietary iron is absorbed as Fe^{2+}, and not as Fe^{3+}; ascorbic acid in the intestinal lumen not only maintains iron in the reduced state but also chelates it, increasing absorption considerably. A dose of 25 mg of vitamin C taken together with a meal increases the absorption of iron approximately 65%, whereas a 1 g dose gives a ninefold increase. This is an effect of ascorbic acid present together with the test meal; neither intravenous administration of vitamin C nor supplements several hours before the test meal affects iron absorption, although the ascorbate secreted in gastric juice should be effective. This is not a specific effect of ascorbate; a variety of other reducing agents including alcohol and fructose also enhance the absorption of inorganic iron. In addition, ascorbate is the electron donor for the intracellular ferric reductase in intestinal mucosal cells.

Inhibition of nitrosamine formation

Oral bacteria can reduce nitrate to nitrite which, under the acidic conditions of the stomach, can react with amines in foods to form carcinogenic *N*-nitrosamines. In addition to dietary sources, a significant amount of nitrate is formed endogenously by the metabolism of nitric oxide—1 mg per kg bodyweight per day (about the same as the average dietary intake), increasing 20-fold in response to inflammation and immune stimulation, and nitrate is secreted in saliva.

Ascorbate reacts with nitrite forming NO, NO_2, and N_2, so preventing the formation of nitrosamines. In addition to ascorbate in foods, there is considerable secretion of ascorbate in the gastric juice, and inhibition of gastric secretion for treatment of gastric ulcers, as well as reducing vitamin B_{12} absorption, also inhibits this presumably protective gastric secretion of ascorbate.

However, although ascorbate can deplete nitrosating compounds under anaerobic conditions, the situation may be reversed in the presence of oxygen. Nitric oxide reacts with oxygen to form N_2O_3 and N_2O_4, both of which are nitrosating reagents, and can also react with ascorbate to form NO and monodehydroascorbate. It is thus possible for ascorbate to be depleted, with little or no significant effect on the total concentration of nitrosating species.

Antioxidant and prooxidant actions of ascorbate

Chemically, ascorbate is a potent reducing agent, both reducing hydrogen peroxide and also acting as a radical-trapping antioxidant, reacting with superoxide and a proton to yield hydrogen peroxide, or with the hydroxy radical to yield water. In each case the product is monodehydroascorbate, which, as shown in Fig. 1, undergoes dismutation to ascorbate and dehydroascorbate. In studies

of ascorbate depletion in men there is a significant increase in abnormalities of sperm deoxyribonucleic acid (DNA), suggesting that vitamin C may have a general, nonspecific radical-trapping antioxidant function.

Ascorbate also acts to reduce the tocopheroxyl radical formed by the oxidation of vitamin E in cell membranes and plasma lipoproteins. It thus has a vitamin E sparing antioxidant action, coupling lipophilic and hydrophilic antioxidant reactions.

The antioxidant efficiency of ascorbate is variable. From the chemistry involved, it would be expected that overall 2 mol of tocopheroxyl radical would be reduced per mole of ascorbate, because of the reaction of 2 mol of monodehydroascorbate to yield ascorbate and dehydroascorbate. However, as the concentration of ascorbate increases, the molar ratio decreases, and only at very low concentrations of ascorbate it tends toward the theoretical ratio. This is because, as well as its antioxidant role, ascorbate can be a source of hydroxyl and superoxide radicals.

At high concentrations, ascorbate can reduce molecular oxygen to superoxide, being oxidized to monodehydroascorbate. Both Fe^{3+} and Cu^{2+} ions are reduced by ascorbate, again yielding monodehydroascorbate; the resultant Fe^{2+} and Cu^{+} are reoxidized by reaction with hydrogen peroxide to yield hydroxide ions and hydroxyl radicals. Thus, as well as its antioxidant role, ascorbate has prooxidant action; the net result will depend on the relative rates of formation of superoxide and hydroxyl radicals by autooxidation and metal-catalyzed reactions of ascorbate, and the trapping of these radicals by ascorbate.

It seems likely that the prooxidant actions of ascorbate are of relatively little importance *in vivo*. Except in cases of iron overload there are almost no transition metal ions in free solution, they are all bound to proteins, and because the renal transport system is readily saturated, plasma and tissue concentrations of ascorbate are unlikely to rise to a sufficient extent to lead to significant radical formation.

Assessment of vitamin C status

The early method of assessing vitamin C nutritional status was by testing the extent of saturation of the body's reserves, by giving a test dose of 500 mg (2.8 mmol), and measuring the amount excreted in the urine. In a subject with high status, more or less all of the test dose is recovered over a period of 5–6 h.

More sensitive assessment of status is achieved by measuring the concentration of the vitamin in whole blood, plasma, or leukocytes (Bender, 2003; Card, 2019). Criteria of adequacy are shown in **Table 2**. The determination of ascorbate in whole blood is complicated by nonenzymic oxidation of the vitamin by hemoglobin, and most studies rely on plasma or leukocyte concentrations of ascorbate.

A problem arises in the interpretation of leukocyte ascorbate concentrations because of the different capacity of different classes of leukocytes to accumulate the vitamin. Granulocytes are saturated at a concentration of approximately 530 pmol per 10^6 cells, whereas mononuclear leukocytes can accumulate 2.5-times more ascorbate. A considerable mythology has developed to the effect that vitamin C requirements are increased in response to infection, inflammation, and trauma, based on reduced leukocyte concentrations of ascorbate in these conditions. However, the fall in leukocyte ascorbate can be accounted for by an increase in the proportion of granulocytes in response to trauma and infection (and hence a fall in the proportion of mononuclear leukocytes). Total leukocyte ascorbate is not a useful index of vitamin C status without a differential white cell count.

There is increased formation of 8-hydroxyguanine (a marker of oxidative radical damage) in DNA during (short term) vitamin C depletion, and the rate of removal of 8-hydroxyguanine from DNA by excision repair, and hence its urinary excretion, is affected by vitamin C status. This suggests that measurement of urinary excretion of 8-hydroxyguanine may provide a biomarker of optimum status, as a basis for estimating requirements.

Requirements

Ascorbate requirements are estimated from the plasma and leukocyte concentrations. Although the minimum requirement for ascorbate is firmly established, there are considerable differences between the reference intakes published by different national and international authorities. Depending on the chosen criteria of adequacy, and assumptions made in interpreting experimental

Table 2 Plasma and leukocyte ascorbate concentrations as criteria of vitamin C nutritional status.

		Deficient	*Marginal*	*Adequate*
Whole blood	mmol L^{-1}	<17	17–28	>28
	mg L^{-1}	<3.0	3.0–5.0	>5.0
Plasma	mmol L^{-1}	<11	11–17	>17
	mg L^{-1}	<2.0	2.0–3.0	>3.0
Leukocytes	pmol per 10^6 cells	<1.1	1.1–2.8	>2.8
	μg per 10^6 cells	<0.2	0.2–0.5	>0.5

results, it is possible to produce arguments in support of reference intakes ranging from 30 to 100 mg day^{-1}. Average ascorbate requirements published by The Institute of Medicine of the National Academies (USA) (2006) and the European Food Safety Administration (2017) appear on Table 3.

Requirements estimated from maintenance of the body pool of ascorbate

An alternative approach to estimating requirements is to determine the fractional rate of catabolism of total body ascorbate; an appropriate intake would then be that required to replace losses and maintain the body pool.

Clinical signs of scurvy are seen when the total body pool of ascorbate is below 1.7 mmol (300 mg). The pool increases with intake, reaching a maximum of approximately 8.5 mmol (1500 mg) in adults—114 μmol (20 mg) per kg bodyweight. The basis for the 1989 United States Recommended Daily Amount (US RDA) of 60 mg was the observed mean fractional turnover rate of 3.2% of a body pool of 20 mg per kg bodyweight per day, with allowances for incomplete absorption of dietary ascorbate and individual variation.

It has been argued that a total body pool of 5.1 mmol (900 mg) is adequate; it is threefold higher than the minimum required to prevent scurvy, and there is no evidence that there are any health benefits from a body pool greater than 600 mg. The observed body pool of 8.5 mmol in depletion/repletion studies was found in subjects previously consuming a self-selected diet, with a relatively high intake of vitamin C, and therefore might not represent any index of requirement. Assuming a total body pool of 5.1 mmol and catabolism of 2.7% per day, allowing for efficiency of absorption and individual variation gives a reference intake of 40 mg day^{-1}.

Because the fractional turnover rate was determined during a depletion study, and the rate of ascorbate catabolism varies with intake, it has been suggested that this implies a rate of 3.6% per day before depletion. On this basis, and allowing for incomplete absorption and individual variation, various national authorities arrive at a reference intake of 80 mg.

The rate of ascorbate catabolism is affected by intake, and the requirement to maintain the body pool cannot be estimated as an absolute value. A habitual low intake, with a consequent low rate of catabolism, will maintain the same body pool as a habitual higher intake with a higher rate of catabolism.

Dietary sources and high intakes

It is apparent from the list of rich sources of vitamin C in Table 4 that the major determinant of vitamin C intake is the consumption of fruits and vegetables; deficiency is likely in people whose habitual intake of fruit and vegetables is very low. However, clinical signs of deficiency are rarely seen in developed countries. The range of intakes by healthy adults reflects fruit and vegetable consumption: the 2.5 percentile intake is 19 mg day^{-1} (men) and 14 mg day^{-1} (women), whereas the 97.5 percentile intake from foods (excluding supplements) is 170 mg day^{-1} (men) and 160 mg day^{-1} (women). Smokers may be at increased risk of deficiency; there is some evidence that the rate of ascorbate catabolism is twofold higher in smokers than in nonsmokers.

Intakes in excess of approximately 80–100 mg day^{-1} lead to a quantitative increase in urinary excretion of unmetabolized ascorbate, suggesting saturation of tissue reserves. It is difficult to justify a requirement in excess of tissue storage capacity.

A number of studies have reported low ascorbate status in patients with advanced cancer—perhaps an unsurprising finding in seriously ill patients. One study has suggested, on the basis of an uncontrolled open trial, that 10 g daily doses of vitamin C resulted in increased survival. Controlled studies have not demonstrated any beneficial effects of high dose ascorbic acid in the treatment of

Table 3 Average requirements of vitamin C by life stage groups.

The American Institute (USA)			*EFSA*		
Life stage	*Males*	*Females*	*Life stage*	*Males*	*Females*
1–3 y	13	13	1–3 y	15	15
4–8 y	22	22	4–6 y	25	25
9–13 y	39	39	7–10 y	40	40
14–18 y	63	56	11–14 y	60	60
≥19 y	75	60	15–17 y	85	75
			≥18 y	90	80
Pregnancy					
≤18 y		66			
19–50 y		70			
Lactation			Lactation		145
≤18 y		96			
19–50 y		100			

Results are expressed in mg day^{-1}.
Taken from: (1) European Food Safety Authority (EFSA) (2017). (2) Institute of Medicine (2006).

Table 4 Vitamin C content of selected foods.

Food	*mg 100 g^{-1}*	*Food*	*mg 100 g^{-1}*
Blackcurrants	200	Artichokes, globe	2
Oranges	52	Potatoes	14
Orange juice	48	Avocados	5
Strawberries	57	Leeks	3
Grapefruit	35	Lemons	53
Melon	16	Okra	21
Green peppers	120	Peas	24
Red peppers	225	Raspberries	19
Sweet potato	23	Tomato juice	8
Loganberries	35	Plantain, green	15
Spinach	29	Blackberries	7
Red currants	40	Tomatoes	22
White currants	40	Bananas	9
Pineapple	53	Cauliflower	56
Brussels sprouts	115	Beans, broad	32
Mangoes	26	Cabbage	48
Satsumas	42	Nectarines	3
Tangerines	42	Parsnips	17
Turnips	17	Rhubarb	6
Gooseberries	26		
Broccoli	79		
Swedes	31		
Spring greens	180		

Compiled from data in the McCance and Widdowson's Composition of Foods Integrated Dataset (2021).

advanced cancer (see section "**Vitamin C and disease**"). On the contrary, large doses of vitamin C supplements have been shown to cause diarrhea, nausea, vomiting, heartburn, stomach cramps, headache and nephrolithiasis.

High doses of ascorbate are popularly recommended for the prevention and treatment of the common cold. The evidence from controlled trials is unconvincing, and meta-analysis shows no evidence of a protective effect against the incidence of colds. There is, however, consistent evidence of a beneficial effect in reducing the severity and duration of symptoms (see section "**Vitamin C and disease**"). This may be due to the antioxidant actions of ascorbate against the oxidizing agents produced by, and released from, activated phagocytes, and hence a decreased inflammatory response.

Vitamin C and gene expression

The first genes whose transcription was shown to be modulated (particularly up-regulated) by ascorbic acid were procollagen and other extracellular matrix genes. Later on, ascorbic acid was shown to revert in part the Charcot-Marie-Tooth 1A disease in transgenic mice. This disease is due to the overexpression of a major myelin gene, *PMP22*, in humans and ascorbic acid treatment lowered PMP22 expression in the mice (Belin et al., 2010). Results in humans, however, indicate that ascorbic acid does not improve the course of the disease, according to the last available Cochrane systematic review (Gess et al., 2015).

Other genes up-regulated by ascorbic acid in mice implanted with cancer cells are Raf kinase inhibitory protein (RKIP) and annexin A5. Among the genes whose transcription is down-regulated by ascorbate are those belonging to the families of tRNA synthetases and translation initiation factors (Belin et al., 2010).

Recent discoveries have provided evidence that ascorbate is an important epigenetic regulator through various mechanisms, which has repercussions on genomic stability.

Vitamin C and demethylation of DNA

In mammals, DNA methylation in the form of 5′methylcytosine (5 mC) is the major form of DNA modification and plays a key role in development and in the control of gene transcription under normal and pathologic conditions. *De novo* 5′methylation of cytosine is induced by DNA methyltransferase 3A (DNMT3A) and DNMT3B; 5 mC is maintained by methyltransferase DNMT1. DNA methylation is a reversible process and 5 mC can be converted back to unmodified cytosine (C) by TET (Ten-Eleven Translocation) proteins (TETs).

The mechanism of TET-meditated demethylation is complex: 5 mC is converted to C by TET mediated oxidation to 5-hydroxy methylcytosine (5hmC), 5-formylcytosine (5 fC) and 5-carboxylcytosine (5caC), followed by the process of active modification-

active removal (AM-AR): by excision of 5 fC or 5caC-mediated thymine DNA glycosylase (TDG) coupled with base excision repair (BER) or the process of active modification-passive dilution (AM-PD): replication-dependent dilution of 5hmC, 5 fC or 5caC.

Three TET enzymes have been discovered in mammalian cells, TET1, TET2 and TET3, all possessing DNA-demethylating activity. TET enzymes are iron (II)/α-ketoglutarate (Fe(II) α-KG)-dependent dioxygenases (α-KGDD) and have a typical structure. They contain a carboxy-terminal core catalytic domain that comprises a cysteine-rich domain and double-stranded β-helix domain (DSBH). Within the DSBH domain, there are some essential catalytic residues, interacting with 2OG and Fe (II): upon cofactor binding, molecular oxygen oxidizes Fe (II) in the catalytic pocket and induces the oxidative decarboxylation of 2OG and substrate oxidation. Within the DSBH domain there is low complexity insert, whose deletion increases the activity of the catalytic domain. Finally, TET1 and TET3, but not TET2, have a DNA-binding domain called the CXXC domain, composed by two Cys4-type zinc finger motifs and located at the amino-terminal region. Vit C increases the enzymatic activity of TET enzymes. Thus, various studies have shown that ascorbate drives DNA methylation in cultured cells in a TET-dependent manner. For instance, In embryonic stem cell cultures, incubation in the presence of 0.25 mM ascorbate elicited a marked loss of 5 mC levels, associated with a moderate increase of 5hmC levels (Mastrangelo et al., 2018; Brabson et al., 2021).

Two mechanisms have been proposed to explain the stimulatory effect of ascorbate on TET enzymatic activity. A first set of studies suggests that ascorbate acts as an enzyme cofactor, able to directly interact with the catalytic domain of TET proteins to increase their enzymatic activity; furthermore, Vit C could promote TET folding to improve the recycling of Fe (II). This hypothesis is supported by the observation that antioxidants other than ascorbate have no effects on TET activity in cultured cells (Mastrangelo et al., 2018).

The second hypothesis challenges the role of ascorbate as a cofactor of TET enzymes and proposes that the stimulatory role of vitamin C on TET activity is related to its capacity to promote the reduction of Fe^{3+} to Fe^{2+}. This conclusion is also supported by the observation that redox-active quinone stimulate TET activity by reducing iron-free Fe^{3+} to Fe^{2+} (Mastrangelo et al., 2018).

Interestingly, physiologic ascorbate concentrations (i.e., 0.1 mM) increase 5hmC content in human melanoma cell lines to levels comparable to those observed in normal melanocytes and reduced the malignancy of these cells, without affecting their proliferation. Ascorbate at these low and almost physiologic concentrations (100–200 μM) elicited apoptosis of some melanoma cell lines through a mechanism involving epigenetic downregulation of clusterin, responsible in turn for Bax activation and Bcl-XL sequestering in the mitochondria, with consequent induction of apoptosis (Mastrangelo et al., 2018).

Vitamin C and demethylation of histones

In addition to DNA methylation, also histone molecules at the levels of lysine and arginine residues can be methylated. Histone methylation is a key event at the level of the epigenetic mechanisms of chromatin control and can resulted in either gene activation or in gene repression. Histone methylation is a dynamic process and involves methyltransferases, such as MLL, that operate histone methylation and several demethylases that demethylate mono-, di- and tri-methylated histones. Two groups of histone demethylases have been identified: lysine-specific histone demethylases (LSD1 and LSD2), promoting the demethylation of mono- and di-methylated lysine residues; and JmjC domain-containing histone demethylases, promoting the demethylation of mono-, di- and tri-methylated histone lysine/arginine residues (Mastrangelo et al., 2018; Brabson et al., 2021).

Vitamin C accelerates and enhances somatic cell reprogramming (generation of induced pluripotent stem cells) with exogenous factors, through a mechanism not simply ascribable to its antioxidant activity alone. This effect of ascorbate is mediated by its capacity to enhance histone demethylase function of Jhdm 1a/1b (Mastrangelo et al., 2018).

Vitamin C also promotes pluripotency of human induced pluripotent stem cells through the induction of increased expression of the histone demethylase JARID1A, and facilitates dopamine neuron differentiation in fetal mid brain through modulation of TET1 and JMJD3 activity (Mastrangelo et al., 2018).

Finally, vitamin C can participate as a cofactor to enhance and maintain the activity of other α-KGDD family members, such as hypoxia inducible factor (HIF) prolyl hydroxylases, and AlkB homologs (ALKBHs), in addition to TET proteins, that in combination can influence genomic stability (Brabson et al., 2021).

Vitamin C and disease

Vitamin C has been and is the subject of much research, and has been implicated in the cure and prevention of numerous diseases. The truth of the matter is that, according to Cochrane Database Systematic Reviews, there is not scientific evidence nowadays to support the use of ascorbate as a preventive or therapeutic agent with the exception of scurvy.

Scurvy

Observations on scurvy first appeared in Egyptian medical scrolls 3500 years ago, and continued through to the discovery of vitamin C and the modern research on the physiological role of ascorbic acid. The observations of great navigators during the 15th and 16th centuries, when scurvy plagued ships' crews, played an important role in clarifying scurvy's etiology. Among the personalities in the history of the disease, James Lind and Albert Szent-Györgyi are most noteworthy, the first for conducting the first clinical trial on the

treatment of scurvy with lemon and orange juices, and the second for discovering and identifying vitamin C (Magiorkinis et al., 2011).

Common cold

The most recent systematic reviews and meta-analyses conclude that the consumption of vitamin C does not prevent the incidence of common cold, so routine vitamin C supplementation is not justified. Yet vitamin C may be useful for people exposed to brief periods of severe physical exercise. Regular supplementation trials have shown that vitamin C reduces the duration of colds, but this was not replicated in the few therapeutic trials that have been carried out. Nevertheless, given the consistent effect of vitamin C on the duration and severity of colds in the regular supplementation studies, and the low cost and safety, it may be worthwhile for common cold patients to test on an individual basis whether therapeutic vitamin C is beneficial for them (Gómez et al., 2018; Hemilä and Chalker 2013).

Cardiovascular disease

The evidence available on the effect of vitamin C supplementation and the risk of cardiovascular disease is of low- and very low-quality. Currently, there is no evidence to suggest that vitamin C supplementation reduces the risk of cardiovascular disease in healthy participants and those at increased risk, but evidence is limited to one trial of middle-aged and older male physicians from the USA (Al-Khudairy et al., 2017).

Cancer

Well-designed randomized controlled trials have shown no beneficial effect of vitamins A, C, D and E supplements for the prevention of lung cancer and lung cancer mortality in healthy people. On the contrary, vitamin C increases lung cancer incidence in women (Cortés-Jofré et al., 2020). Vitamin C supplements do not exert significant effects on gastrointestinal cancers (Bjelakovic et al., 2008).

Cataracts

There is no evidence from randomized controlled trials that supplementation with vitamin C prevents or slows the progression of age-related cataracts (Mathew et al., 2012).

Asthma

There is no indication currently that vitamin C can be recommended as a therapeutic agent in asthma (Milan et al., 2013).

Pneumonia

The current evidence is too weak to advocate prophylactic use of vitamin C to prevent pneumonia in the general population. Nevertheless, therapeutic vitamin C supplementation may be reasonable for pneumonia patients who have low vitamin C plasma levels because its cost and risks are low (Padhani et al., 2021).

Pregnancy complications

The data do not support routine vitamin C supplementation alone or in combination with other supplements for the prevention of fetal or neonatal death, poor fetal growth, preterm birth or pre-eclampsia (Rumbold et al., 2015).

Conclusions

Vitamin C is widely distributed in nature, but it is found mainly in vegetable foods, where occurs naturally in two interconvertible chemical forms: ascorbic acid (reduced form) and dehydroascorbic acid (oxidized form). Both forms have similar biological action. Although its biological functions have been known for a long time, its role in the regulation of gene expression and its epigenetic effects are more novel.

Vitamin C has been implicated in the cure and prevention of numerous diseases. However, the scientific evidence available nowadays only clearly supports the use of this vitamin to prevent scurvy and reduce the duration and severity of colds.

References

Al-Khudairy, L., Flowers, N., Wheelhouse, R., Ghannam, O., Hartley, L., Stranges, S., Rees, K., 2017. Vitamin C supplementation for the primary prevention of cardiovascular disease. Cochrane Database Syst. Rev. 3, CD011114. https://doi.org/10.1002/14651858.CD011114.pub2.

Belin, S., Kaya, F., Burtey, S., Fontes, M., 2010. Ascorbic acid and gene expression: another example of regulation of gene expression by small molecules? Curr. Genom. 11, 52–57. https://doi.org/10.2174/138920210790217936.

Bender, D.A., 2003. Ascorbic acid (vitamin C). In: Bender, D.A. (Ed.), Nutritional Biochemistry of the Vitamins, second ed. Cambridge University Press, New York, pp. 357–384.

Bjelakovic, G., Nikolova, D., Simonetti, R.G., Gluud, C., 2008. Antioxidant supplements for preventing gastrointestinal cancers. Cochrane Database Syst. Rev. 3, CD004183. https://doi.org/10.1002/14651858.CD004183.pub3.

Brabson, J.P., Leesang, T., Mohammad, S., Cimmino, L., 2021. Epigenetic regulation of genomic stability by vitamin C. Front. Genet. 12 (12), 675780. https://doi.org/10.3389/fgene.2021.675780.

Card, D.J., 2019. Methods for assessment of vitamin C. In: Harrington, D. (Ed.), Laboratory Assessment of Vitamin Status. Academic Press, London, pp. 301–316. https://doi.org/10.1016/B978-0-12-813050-6.00013-9.

Combs Jr., G.F., McClung, J.P., 2016. The Vitamins. Fundamental Aspects in Nutrition and Health, fifth ed. Academic Press, London.

Cortés-Jofré, M., Rueda, J.R., Asenjo-Lobos, C., Madrid, E., Bonfill Cosp, X., 2020. Drugs for preventing lung cancer in healthy people. Cochrane Database Syst. Rev. 3, CD002141. https://doi.org/10.1002/14651858.CD002141.pub3.

European Food Safety Authority (EFSA), 2017. Dietary Reference Values for Nutrients. Summary Report. https://doi.org/10.2903/sp.efsa.2017.e15121.

Gess, B., Baets, J., De Jonghe, P., Reilly, M.M., Pareyson, D., Young, P., 2015. Ascorbic acid for the treatment of Charcot-Marie-Tooth disease. Cochrane Database Syst. Rev. 12, CD011952. https://doi.org/10.1002/14651858.CD011952.

Gómez, E., Quidel, S., Bravo-Soto, G., Ortigoza, A., 2018. Does vitamin C prevent the common cold? Medwave 18, e7235. https://doi.org/10.5867/medwave.2018.04.7236. PMID: 30113569.

Hemilä, H., Chalker, E., 2013. Vitamin C for preventing and treating the common cold. Cochrane Database Syst. Rev. 1, CD000980. https://doi.org/10.1002/14651858.CD000980.pub4.

Institute of Medicine, 2006. Dietary Reference Intakes: The Essential Guide to Nutrient Requirements. The National Academies Press, Washington, DC. https://doi.org/10.17226/11537.

Magiorkinis, E., Beloukas, A., Diamantis, A., 2011. Scurvy: past, present and future. Eur. J. Intern. Med. 22, 147–152. https://doi.org/10.1016/j.ejim.2010.10.006.

Mastrangelo, D., Pelosi, E., Castelli, G., Lo-Coco, F., Testa, U., 2018. Mechanisms of anti-cancer effects of ascorbate: cytotoxic activity and epigenetic modulation. Blood Cells Mol. Dis. 69, 57–64. https://doi.org/10.1016/j.bcmd.2017.09.005.

Mathew, M.C., Ervin, A.M., Tao, J., Davis, R.M., 2012. Antioxidant vitamin supplementation for preventing and slowing the progression of age-related cataract. Cochrane Database Syst. Rev. 6, CD004567. https://doi.org/10.1002/14651858.CD004567.pub2.

McCance and Widdowson's Composition of Foods Integrated Dataset, 2021. Public Health England. (https://www.gov.uk/government/publications/composition-of-foods-integrated-dataset-cofid).

Milan, S.J., Hart, A., Wilkinson, M., 2013. Vitamin C for asthma and exercise-induced bronchoconstriction. Cochrane Database Syst. Rev. 10, CD010391. https://doi.org/10.1002/14651858.CD010391.pub2.

Padhani, Z.A., Moazzam, Z., Ashraf, A., Bilal, H., Salam, R.A., Das, J.K., Bhutta, Z.A., 2021. Vitamin C supplementation for prevention and treatment of pneumonia. Cochrane Database Syst. Rev. 11, CD013134. https://doi.org/10.1002/14651858.CD013134.pub.

Rodwell, V.W., Bender, D.A., Botham, K.M., Kennelly, P.J., Weil, P.A., 2018. Harper's Illustrated Biochemistry, 31st ed. McGraw-Hill Education, New York.

Ross, A.C., Caballero, B., Cousins, R.J., Tucker, K.L., Ziegler, T.R., 2014. Modern Nutrition in Health and Disease, eleventh ed. Wolters Kluwer Health, Lippincott Williams & Wilkins, Philadelphia.

Rumbold, A., Ota, E., Nagata, C., Shahrook, S., Crowther, C.A., 2015. Vitamin C supplementation in pregnancy. Cochrane Database Syst. Rev. 9, CD004072. https://doi.org/10.1002/14651858.CD004072.pub3.

Stipanuk, M.H., Caudill, M.A., 2018. Biochemical, Physiological, and Molecular Aspects of Human Nutrition, fourth ed. Saunders, Philadelphia.

Vitamin C: Deficiency states

CJ Bates, MRC Human Nutrition Research, Cambridge, UK

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of CJ Bates, Vitamin C: Deficiency States, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 357–362, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00020-9.

Scurvy: The History and Discovery of Vitamin C 515
Degradation, Turnover, and Factors that Induce Increased Requirements for Vitamin C 517
Well-Established Metabolic Functions of Vitamin C that are Impaired by Deficiency 517
Measurement of Vitamin C Status; Biochemical Tests for Adequacy and Deficiency 518
Occurrence of Low Intakes and Poor Biochemical Status in Present-Day Societies 519
Further Reading 520

Glossary

Clinical deficiency Often results from severe prolonged tissue (nutrient) deficiency, which may in turn result from inadequate (nutrient) intake, but may also arise from impaired absorption, increased turnover and/or excretion; increased tissue demand, etc. Deficiency signs are directly observable; symptoms can be elicited by testing, and the response of these to intervention (e.g., nutrient supplementation) can help to confirm a diagnosis of deficiency.

Deficiency (clinical or biochemical) Arises when the amounts of an essential nutrient in (tissues of) the body falls below a critical minimum level. Biochemical (tissue) depletion then occurs, and may be followed by clinical (i.e., pathological) signs and symptoms of deficiency. Deficiency of an essential nutrient in the diet may lead to tissue- and clinical-deficiency, but impaired absorption, increased losses, increased tissue demands, etc., may also lead to a functional (tissue) deficiency, even when the diet-content is adequate.

Nutrient (e.g., vitamin) status Commonly assessed by measuring the concentration of the nutrient in an accessible body fluid such as serum or urine, or else the functionality of an enzyme or biochemical pathway (functional status). Published 'normal ranges' enable the results to be classified as (e.g.,) deficient, low, normal, or high.

Scurvy: The History and Discovery of Vitamin C

Scurvy is traditionally associated with long sea voyages, which often lasted for several years; the seamen's diets were confined to whatever could be stored at room temperature for long periods. In the absence of refrigeration, their diets typically consisted of dried biscuits and other dry cereal foods (wheat flour and oatmeal), salted meat, dried peas, cheese, butter, and ale, i.e., whatever could be dried and preserved in adverse tropical climates. The signs and symptoms described in classical accounts of scurvy, written long before its cause was understood, included lassitude, swollen joints, putrid and bleeding gums, failure of wound healing and the opening of old wounds and sores, intradermal bleeding due to capillary fragility, heart failure, and sudden death (Table 1). Although nowadays we distinguish the symptoms of true scurvy (now known to be produced specifically by vitamin C deficiency) from conditions such as beriberi (thiamin deficiency, associated with edema of the lower limbs), vitamin A deficiency (associated with night-blindness and corneal lesions), and rickets (vitamin D deficiency caused mainly by a lack of exposure to sunlight in children), in the older literature these conditions were often not recognized as distinct. Signs and symptoms of scurvy occurred on land in times of siege or during prolonged military campaigns where accesses to fresh foods were severely restricted. Although some medical practitioners and political leaders became convinced that scurvy had a dietary cause, and that its cure or prevention were possible by including fresh plant materials such as scurvy grass, decoctions of evergreen needles, etc., in the diet (Table 2), many believed, right up to the beginning of the twentieth century, that factors such as 'foul vapours or infections were to be blamed. Indeed, before the recognition and discovery of essential micronutrients in the early years of the twentieth century, which confirmed that small amounts of certain complex organic molecules are needed in the diet to maintain health; it was thought that food is needed to supply only energy, protein, certain minerals, and water.

Although treatments for scurvy stretch back over many centuries, a definitive dietary cure is attributable to James Lind, whose controlled intervention trial on board HMS Salisbury, followed by 'A Treatise of the Scurvy' in 1753, provided decisive evidence that persuaded the British Admiralty to insist on the inclusion of citrus fruit regularly in the British naval diet (Table 2). Lind showed that a 'rob' or decoction of oranges could rapidly cure the disease, and his discovery rapidly brought about dramatic reductions in the incidence and mortality due to scurvy.

In the first decade of the twentieth century, two Norwegian scientists, Holst and Frohlich, found that guinea pigs, like humans, reproducibly become scorbutic if fed a diet that lacks the 'antiscorbutic principle' which was, by now, known to be abundantly

Table 1 Signs, symptoms, and biochemical indicators of vitamin C deficiency

Signs of clinical scurvy:
Bleeding gums and looseness of teeth.
Skin petechiae (intradermal bleeding due to capillary fragility); perifollicular hemorrhages; larger sheet hemorrhages, especially of the skin of the limbs and trunk; hematomas; purpura; bruising.
Ecchymoses (failure of eruption of hair from hair follicles); hyperkeratosis of hair follicles.
Breakdown and bleeding from previously healed wounds.
Swollen painful joints, with effusions and arthralgia; hemarthrosis; myalgia
Edema (especially pedal); dyspnea; dry eyes and mouth.
Abnormal long bone development in children; swelling ('beading') of rib cage; swelling of long-bone joints.
Intracranial hemorrhages in children; 'bulging eyes'; gums often affected if teeth already erupted.
Clinical symptoms of scurvy:
Fatigue, lassitude, joint pain, 'pithed frog' appearance in babies, for example, after prolonged feeding of condensed milk.
Abnormal X-ray appearance of long bones in children, especially at the epiphyses; subperiosteal bleeding; osteolysis; osteonecrosis.
Capillary fragility, as measured by the Hess test (pressure cuff on limb); nonspecific anemia in many cases.
Heart failure in severe cases, potentially leading to death.
Biochemical indicators of vitamin C deficiency:
Plasma vitamin C (ascorbate) concentrations <11 μmol/l
Buffy coat vitamin C concentrations <15 μg/10^8 cells
Suggested, but not yet verified, functional tests, for example, abnormalities of hydroxylation and cross-links of nascent collagen; of carnitine formation, of vitamin E recycling; of purine oxidation, etc.

Table 2 Historical timeline of investigations into scurvy and discovery of vitamin C (ascorbic acid)

Tudor England: Leaders on certain long naval voyages successfully made use of plant foods (e.g., scurvy grass; young pine needles) as an *ad hoc* remedy for the prevalent scourge of scurvy.
Mid-eighteenth century: James Lind (a Scottish physician) carried out the first controlled trial of scurvy remedies at sea; he demonstrated efficacy of citrus fruit (oranges, lemons) and published 'Treatise of the Scurvy', 1753. Sir Gilbert Blane (end of eighteenth century) caused British naval diets to be supplemented with citrus (lemons, limes), thereby reducing prevalence of scurvy and greatly improving sailors' health – although limes eventually proved less effective than lemons, and scurvy reappeared as a result of failure to realise this.
Nineteenth century: Emergence of 'Barlow's disease' – scurvy in babies mainly fed with condensed milk.
Early twentieth century: Holst and Frohlich developed a guinea-pig model of scurvy, permitting investigation of the antiscorbutic potency of various foods.
1929: Albert Szent-Gyorgyi in Cambridge isolated 'hexuronic acid' from adrenal extracts, which proved identical to 'ascorbic acid', or 'vitamin C', the elusive antiscorbutic substance present in citrus fruit and plants, then being studied by Charles Glen King in Pittsburgh. 'Vitamin C' was thus isolated and named.
1930s: Norman Haworth and Tadeus Reichstein achieved chemical synthesis of vitamin C (ascorbic acid) from common sugars (glucose, galactose), thereby proving its chemical structure and elucidating its unique chemical and physiological (nutritional) properties.
1940s: Wartime studies showed that 10 mg vitamin C per day is sufficient to prevent and cure clinical scurvy in adult males.
Mid-twentieth century: Biochemical studies in several laboratories elucidated multiple roles of vitamin C: (1) as a cofactor for certain key enzymes, involved in specific oxidation and hydroxylation pathways of intermediary metabolism; (2) as a redox-modulator and provider of protection against certain types of (especially oxygen-radical-mediated) oxidative damage.

present in fresh fruit and vegetables. Few other mammals share this unusual dietary requirement, which, we now know, is solely attributable to the lack of a key enzyme, L-gulonolactone oxidase, which is needed in the multistep biosynthetic pathway for the synthesis of ascorbic acid from common sugars such as glucose and galactose. This animal model of dietary-induced scurvy provided a very important tool for the study of the disease.

Paradoxically, crystalline ascorbic acid (vitamin C) was first isolated, not from a well-established antiscorbutic plant source such as fruit or green leaves, but instead from an animal source, namely adrenal glands, where high concentrations of the vitamin are also found. Indeed, the original motivation for the isolation of the crystalline material (by Albert Szent-Gyorgyi in Gowland Hopkins' laboratory in Cambridge) was a serendipitous accident, arising out of his attempts to isolate a new adrenal hormone. Charles Glen King in Pittsburgh then showed that this easily-oxidized sugar derivative (initially named 'hexuronic acid') was indeed the long-sought antiscorbutic principle, i.e., vitamin C or L-ascorbic acid.

The vitamin was isolated in 1928, and its chemical structure was proven by de novo synthesis from common sugars a few years later (Table 2).

Degradation, Turnover, and Factors that Induce Increased Requirements for Vitamin C

The instability of vitamin C in air, and especially in neutral or alkaline aqueous solution, is attributable to the fact that in the presence of oxygen or other oxidizing agents it readily undergoes two successive one-electron oxidation steps to produce (reversibly, see below) another unstable product, dehydroascorbate, which readily undergoes an irreversible lactone ring-hydrolysis to yield 2,3-diketogulonic acid. Thus the vitamin is readily destroyed, both in foods during storage and (at a lower rate because of efficient recycling mechanisms) in the body. Diketogulonic acid is one of several degradation products of vitamin C that cannot be reconverted to the vitamin and are further degraded to stable excretory products, such as oxalic acid, by oxidative catabolism. Of all the micronutrients that are essential for human health and survival, vitamin C is the most easily destroyed during drying and other traditional methods of preserving food. Citrus fruits contain other organic acids that inhibit this process of oxidation by lowering the pH of the fruit juice. This enables them, and extracts of them, to preserve at least some of their vitamin content for several weeks and even months of storage and thereby helps them to prevent and cure scurvy.

It remains largely a mystery why some people succumb to scurvy after a modest period of very low intake, whereas others survive for much longer. It has been speculated that some people may be able to produce all of the enzymes of the vitamin C synthetic pathway, including gulonolactone oxidase. However, this now seems unlikely and it is more probable that the retention and recycling mechanisms for the vitamin are more efficient in some people than in others. For example, smokers have a higher turnover of endogenous vitamin C than nonsmokers, mainly because of the free-radical oxidant species in cigarette smoke. People with infections also have increased vitamin C turnover, which is associated with the liberation of pro-oxidant substances (such as hypochlorous acid) that are used by the body to kill bacteria. Some people have genetic variants of the vitamin C transporters (see the following section); others have isoforms of certain blood proteins such as haptoglobins, both of which may be associated with relatively low levels of vitamin C in the blood. Very occasionally, there arise nonlethal mutations of vitamin C-dependent pathways whose abnormalities can be treated with high vitamin C intakes. A well-characterized example is Ehlers–Danlos syndrome, type VI, which is associated with impaired collagen lysyl hydroxylation and presents a variety of clinical and biochemical connective-tissue (collagen-related) defects. However, much more research is needed to determine, which of many possible genetic and environmental factors modulate the turnover of vitamin C in the body and to determine individual requirements and hence relative resistance to scurvy. Although 100–200 mg of the vitamins per day is needed to approach saturation of the tissues of humans, the amount needed to prevent or cure scurvy is less than 10 mg day^{-1}, as was shown by experiments involving prolonged periods of feeding with depleted diets in the middle of the twentieth century (**Table 2**). Today, overt clinical scurvy is rare. It is occasionally seen in refugee camps or in elderly people with poor diets that are devoid of the usual sources of the vitamin. The latter high-risk group contains many individuals who are unable to chew fresh fruit and vegetables because of poorly-fitting dentures or poor gastric tolerance of acidic or fibrous foods (see final section).

An essential dietary requirement for vitamin C is shared by only a small number of vertebrates, including most primates, guinea pigs and agoutis, and some birds and fishes. Most mammals synthesize the vitamin in their livers from hexose sugars; birds synthesize it in their kidneys. The final enzyme in the pathway, L-gulonolactone oxidase, has been lost in several unrelated species, suggesting a vulnerable and easily mutated locus on the genome. Presumably this mutation was neutral or advantageous during the natural selection of man's ancestors, when human and related-primate diets were rich in plant sources of the vitamin.

Well-Established Metabolic Functions of Vitamin C that are Impaired by Deficiency

Studies of guinea pigs (and other species that require a dietary source of vitamin C) have revealed that, when deprived of the vitamin, characteristic lesions of growing bones, failure of wound-healing of skin and bones, capillary defects, and other lesions arise, all of which point to a failure of the new synthesis of or repair processes for, connective tissues and especially the protein collagen, which is the major extracellular protein and comprises a third of all the protein in the body (**Table 2**). As the biochemical pathway of collagen biosynthesis became better understood, during the middle years of the twentieth century, it became clear that hydroxylated amino-acids, comprising two different hydroxylated forms of proline and one of lysine, occurred uniquely in collagen. These were not coded by the genome or inserted by the amino-acid-assembly machinery of the cell but instead were created by 'post-translational' amino-acid hydroxylation processes that took place after the nascent procollagen polypeptide chain had been synthesized on the polysomal messenger RNA. Some of the prolyl residues of the procollagen molecule were then hydroxylated to hydroxyprolyl residues, and some of the lysyl residues were hydroxylated to hydroxylysyl residues. The hydroxylated prolyl residues are essential for subsequent collagen triple-helix formation and hence for the secretion of nascent collagen; the hydroxylated lysyl residues form part of the essential pyridinoline-type crosslinks that stabilize the collagen fibers, especially those in bone. In the absence of sufficient vitamin C, these hydroxylation reactions rapidly fail, because the ferrous iron at the active center of the 'mixed function oxidase' enzymes that catalyze them is rapidly inactivated by oxidation. Vitamin C, specifically, is needed to keep the essential ferrous residues at the hydroxylase-enzyme active centers in the reduced, active, form. In the absence of the vitamin, these enzymes are inactivated only after a few cycles of hydroxylation. Hydroxyproline formation in the C1q component of complement, an important component of the immune system, is also vitamin C-dependent.

The essential function of vitamin C in collagen maturation can go a long way toward explaining many of the clinical lesions of scurvy (**Table 1**). However, the vitamin may also act directly on the transcription and translation of collagen mRNA and on the synthesis of other parts of the cell machinery that are needed for the formation of normal connective tissues.

Vitamin C plays a cofactor-like role in the reactions of several other enzymes that split molecular oxygen, notably members of the group of enzymes that are classified as 'mixed-function oxidases.' Two enzymes containing ferrous iron that are involved in carnitine biosynthesis (trimethyl lysine hydroxylase and γ-butyrobetaine hydroxylase) fall into this category. Aspartate β-hydoxylase, which is needed for the postsynthetic modification of protein kinase C, also requires vitamin C. Another enzyme that requires vitamin C is the copper-containing dopamine β-hydroxylase, and in the reaction that it catalyzes, ascorbic acid reduces cupric to cuprous copper at the active site. Peptidyl glycine hydroxylase (peptidyl α-amidase) is also a copper-containing enzyme requiring vitamin C as cosubstrate. Vitamin C can increase the activities of several other enzymes, usually by a nonspecific reducing or protective action that is also shared by some other cellular reductants. Newly identified roles for the vitamin include cell-signaling; nucleic acid and histone dealkylation, and proteoglycan deglycanation (e.g., via turnover of glycipan-1, which is involved in cell growth and differentiation). One such newly identified ascorbate-dependent reaction is a dioxygenase-dependent hydroxylation of prolyl and asparaginyl residues in the α-subunit of hypoxia-inducible transcription factor 1 (HIF-1). The asparagine-hydroxylation modulates interaction with other activators, whereas the proline-hydroxylation targets HIF-1 for destruction in proteasomes; the unhydroxylated HIF-1 increases expression of most glycolytic enzymes and two glucose transporters (which also transport dehydroascorbate).

During its function, ascorbic acid is oxidized in two successive reversible one-electron steps, and most, if not all, of its essential biological actions are centered around this redox cycle. The first oxidation product is the free-radical form of the vitamin, which is known variously as 'monodehydroascorbate,' 'semidehydroascorbate,' or 'ascorbate free radical' (AFR). Although this intermediate shares with most other free radicals, the properties of having a relatively short half life and a high degree of chemical reactivity, it is more stable than many other free radicals, contrasting with the highly reactive and damaging radicals such as hydroxyl or superoxide radical that arise from molecular oxygen. By reacting with, and thus quenching, these damaging oxygen free radicals, ascorbate can act as a free-radical chain terminator and can thereby protect vulnerable macromolecules such as DNA, lipids, and proteins from oxidative damage by free-radical chain reactions. Such reactions would otherwise cause extensive damage, including genetic damage (to DNA), the formation of potentially atherogenic oxidized lipids, and oxidative inactivation of enzymes. For this reason, ascorbic acid is thought to possess important 'protective' antioxidant properties that are not directly connected with its other cofactor-like or cosubstrate-like roles in enzyme reactions. Ascorbate probably also protects host tissues against damage by oxidants such as hypochlorous acid that are produced in the normal course of bacterial-killing action of white cells.

The second one-electron oxidation step in ascorbate oxidation produces dehydroascorbate from the free-radical intermediate AFR. Both of these oxidized forms can be recycled to ascorbate either by nonenzymatic reactions with glutathione as the reductant (electron acceptor) or by pyridine nucleotide-dependent enzymatically catalyzed reactions. Thus, the two sequential one-electron oxidation steps from ascorbate to dehydroascorbate are fully reversible *in vivo*. However, the subsequent spontaneous nonenzymatic reaction comprising of hydrolysis of the 1,4-lactone ring is not reversible, so that the product of this reaction, diketogulonic acid, has no provitamin activity. Normally, approximately 3% of the vitamin C in the body is degraded every day and this loss must be replaced from the diet. Nevertheless, many weeks at or near zero intake are usually needed to reach scorbutic levels.

An important aspect of functional adequacy of nutrients is their efficient transport across cell membranes and hence between tissue pools. Reduced ascorbic acid is transported by two distinct sodium–vitamin C transporters, SVCT1, which is found mainly in epithelial tissue such as intestine, liver, and kidney and SVCT2, which is widely distributed among tissues. Naturally-occurring genetic variants affecting SVCT1 have recently been shown to affect circulating concentrations of the vitamin in human populations, and thus, presumably, vitamin C status and requirements. The oxidized form, dehydroascorbate, is transported by the glucose transporter systems, GLUT1 and GLUT2, and glucose and other sugars that share these transporters compete with it, a fact that is relevant for some diseases, such as diabetes, in which glucose transport and concentrations are abnormal.

Measurement of Vitamin C Status; Biochemical Tests for Adequacy and Deficiency

In species (such as humans) that cannot synthesize vitamin C in their bodies, the vitamin concentration in tissues and blood compartments (plasma, erythrocytes, and white blood cells) varies characteristically with the dietary intake of the vitamin. Because the blood-compartment concentrations mirror the concentrations in most other cells and tissue compartments, tissue vitamin C status can be monitored by measuring the concentration in plasma or blood, even though the blood intracellular concentrations are generally lower than those in most tissues. The concentration ratios between extracellular and various intracellular compartments are determined by active transport systems (SVCT 1 and 2, see previous section), that concentrate the vitamin inside many cell types. At high intakes of the vitamin, the intestinal absorption process is overwhelmed, so that some of the ingested vitamins remain unabsorbed and is destroyed in the lower intestine by intestinal bacteria. The maximum steady-state level in plasma can be temporarily exceeded following a high bolus intake, but the excess vitamin is rapidly excreted in the urine once the renal threshold for filtration and reabsorption is exceeded. These safety mechanisms limit the maximum concentration of the vitamin to which the tissues are normally exposed.

For many years, the best biochemical measure of vitamin C status was considered to be the buffy coat, or total white-cell concentration of the vitamin (**Table 1**), expressed as micrograms or micromoles per 10^8 white cells, the cell count in the assay sample being estimated by an electronic cell counter. This status index varied predictably with total body vitamin C stores during controlled (e.g., animal) depletion studies. However, in practice it has proved to be a difficult test to use in human studies and especially in surveys, as it requires complex laboratory operations to be performed immediately after collecting the blood. It is also difficult to harmonize

this test between laboratories, and, because it measures the average vitamin C content across several different white-cell types, whose individual proportions and relative vitamin C contents may vary considerably, its interpretation was not always straightforward. In addition, infection affects the values obtained. For all of these reasons, this assay has fallen out of favor and is now rarely used. The concentration of the vitamin in erythrocytes or whole blood is not an ideal alternative, partly because hemoglobin can catalyze the oxidative destruction of the vitamin *in vitro* and partly because erythrocyte concentrations do not mirror other body compartments in a simple manner.

Serum or plasma vitamin C has therefore become the most commonly used status assay. To avoid short-term fluctuations caused by recent bolus intakes from food or supplements, it is preferable to collect an overnight-fasting blood sample. Because the vitamin is extremely easily oxidized, the sample must be carefully preserved unless the assay can be performed immediately. The usual approach is to add freshly prepared metaphosphoric acid, usually at between 2 and 5% w/v, which precipitates plasma proteins, chelates transition-metal ions, and provides a protective acidic environment of a suitable pH. If stored, the samples must be kept at a low temperature, for example, at −25 °C for not more than a week or two or at −80 °C for up to 1–2 years. There are many alternative physicochemical and chemical assay methods for measuring vitamin C in extracts of plasma or serum. These include (1) the measurement of its chemical reducing action on reducible dyes such as dichlorophenol indophenol or (2) the formation of either a colored osazone, or a fluorescent derivative with orthophenylene diamine, after conversion to dehydroascorbate. Quantitation by absorbance or by electrochemical detection after separation by high-performance liquid chromatography is favored by many workers. This procedure has the advantage of being relatively specific (i.e., free from most forms of interference) and highly sensitive, but it is more time-consuming than the simpler nonchromatographic methods. Different methods may differ with respect to their specificity and their sensitivity to problems of interference as well as in the precautions that are needed to avoid oxidative destruction of the vitamin during the assay. Careful validation and robust quality-control procedures are essential.

Plasma or serum levels below 11 μmol l^{-1} (<0.2 mg per 100 ml) are considered to be evidences of biochemical deficiency, and if this is severe and prolonged, the risk of clinical deficiency, i.e., scorbutic signs and symptoms, gradually increases. Intakes below 20 mg day^{-1} are likely to result in plasma levels in this range. Studies of human volunteers in the middle of the twentieth century showed clearly that an intake of 10 mg vitamin C per day in a healthy adult is sufficient to prevent clinical scurvy, and this small amount is also sufficient to cure scorbutic signs and symptoms (**Table 2**).

Assay methods based on urinary excretion of vitamin C have been used to study status, but they are too cumbersome and difficult to interpret to be useful in population studies. There are no well-established functional assays available to define vitamin C status and requirements at present. An older method known as the 'Hess test,' which measures relative capillary fragility under pressure or suction (**Table 1**), is useful only if subclinical scurvy is present and is rarely attempted today. Studies of collagen crosslinks or oxidative damage to macromolecules such as DNA or lipids may yield evidences about functional status in the future, but this remains a research challenge and is not yet an available option for routine studies or surveys.

Occurrence of Low Intakes and Poor Biochemical Status in Present-Day Societies

Although scurvy is rare, biochemical evidence of poor vitamin C status is not uncommon in certain high-risk groups in different human populations. Studies in The Gambia in West Africa, for instance, have shown that there is a regular seasonal cycle of availability of foods rich in vitamin C, with a good availability in the dry season alternating with a severe shortage during the rainy season. Plasma, buffy-coat, and breast-milk concentrations are all, on average, adequate in the dry season but are severely reduced during the rains. Functional and health-related parameters also deteriorate during the rains, but it has so far proved to be difficult to

Table 3 Prevalence of low vitamin C intakes and low plasma vitamin C concentrations in Britain at the end of the twentieth century

Age group	*LRNI(mg day^{-1})*	*Intake less than LRNI*	*Less than 11 μmol l^{-1} plasma vitamin*
Preschool 1.5–4.5 years	8	8/723=1.1%	24/723=3.3%
Young people 4–18 years			
4–10 years	8	1/423=0.2%	6/422=1.4%
11–14 years	9	0/307=0%	4/307=1.3%
15–18 years	10	1/271=0.4%	8/271=3.0%
Adults 19–64 years			
19–24 years	10	1/212=0.5%	11/151=7.3%
25–34 years	10	0/429=0%	8/307=2.6%
35–49 years	10	0/571=0%	8/414=1.9%
50–64 years	10	0/512=0%	8/366=2.2%
Adults 65 years and above			
Free-living 65–79 years	10	8/606=1.3%	88/606=14.5%
Free-living 80+years	10	7/274=2.5%	45/274=16.4%
Institution-living	10	2/248=0.8%	98/248=39.5%

The LRNI is the Lower Reference Nutrient Intake, deemed to be sufficient for only that 2.5% of the population who have the smallest requirements, and a plasma concentration of 11 μmol l^{-1} is the cut-off for biochemical deficiency.

reverse this deterioration by vitamin C supplements alone. Therefore robust evidence of health consequences of this seasonal availability cycle has not yet been obtained.

From recent surveys in the UK, **Table 3** shows the prevalence of low intakes of vitamin C (estimated from the proportion of participants receiving less than the lower reference nutrient intake (LRNI), which is the amount deemed to be sufficient for only a few people in a population group, namely the 2.5% with the lowest requirements). Also shown in **Table 3** is the prevalence of plasma concentrations below the lower cut-off in normality, set at 0.2 mg dl^{-1} or 11 μmol l^{-1}. This is shown for several subgroups of the British population of different ages, from data collected in three nationally representative population surveys during the decade 1990s. It is clear from these results that a very few people were getting less than the LRNI for vitamin C over a 4 days or 7 days period of weighed-intake estimates of their diets. Low plasma levels were likewise relatively uncommon in the younger age groups; however, they were more common in older people and were especially prevalent, at almost 40%, in older people living in institutions such as nursing homes. Many of these relatively low plasma levels seen in frail older people are likely to be caused by factors other than very low intakes of the vitamin, such as reduced efficiency of the vitamin C transporters, especially SVCT1 or increased vitamin turnover. In the UK, unlike The Gambia, there was relatively little evidence of a major seasonal variation in vitamin C intake or status at the end of the twentieth century.

Vitamin C absorption does not appear to be abnormally low in healthy older people. However, the multiple pathologies associated with old age (and with debility at any age) are associated with increased turnover of the vitamin. Older people with very low levels of vitamin C are at higher risk of dying sooner than those with high levels, although short-term vitamin supplements generally fail to reverse this increased risk. It thus appears that vitamin C status can act as a barometer of health as well as being a marker of adequacy of vitamin C intake. Further research is needed to determine the key mechanisms that affect the rate of vitamin C turnover and its control in different age groups and different metabolic states. Because frail older people are at high risk of developing pressure sores and of needing surgery for a variety of ailments, there seems to be a potential public-health advantage in protecting vitamin C stores, especially in this vulnerable older age group.

Further Reading

Bown, SR, 2005. Scurvy. Thomas Dunne Books, St Martin's Griffen, New York, pp. 1–254.

Carpenter, KJ, 1986. The History of Scurvy and Vitamin C. Cambridge University Press, Cambridge.

Institute of Medicine, 2000. Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids. National Academy Press, Washington, DC, pp. 95–185.

Johnston, CS, 2006. Vitamin C. In: Bowman, BA, Russell, RM (Eds.), Present Knowledge in Nutrition, 9th edn. ILSI Press, Washington, DC, pp. 233–241. ch. 17.

Johnston, CS, Steinberg, FM, Rucker, RB, 2007. Ascorbic acid. In: Zempleni, J, Rucker, RB, McCormick, DB, et al. (Eds.), Handbook of Vitamins, 4th edn. CRC Press, Boca Raton, USA, pp. 489–520. ch. 17.

Kivirikko, KI, Myllyla, R, 1982. Posttranslational enzymes in the biosynthesis of collagen: Intracellular enzymes. Methods in Enzymology 82, 245–304.

Levine, M, Katz, A, Padayatty, S, 2006. Vitamin C. In: Shils, ME, Shike, M, Ross, AC, et al. (Eds.), Modern Nutrition in Health and Disease. Lippincott Williams and Wilkins, Baltimore, pp. 507–524.

Michels, AJ, Hagen, TM, Frei, B, 2010. A new twist on an old vitamin: human polymorphisms in the gene encoding the sodium-dependent vitamin C transporter. American Journal of Clinical Nutrition 92, 271–272.

Olmedo, JM, Yannias, JA, Windgassen, EB, et al., 2006. Scurvy: A disease almost forgotten. International Journal of Dermatology 45, 909–913.

Packer, L, Fuchs, J, 1997. Vitamin C in Health and Disease. Marcel Dekker Inc, New York.

Pinnell, SR, Krane, SM, Kenzora, JE, et al., 1972. A heritable disorder of connective tissue. Hydroxylysine-deficient collagen disease. New England Journal of Medicine 286, 1013–1020.

Sato, P, Udenfriend, S, 1978. Studies on vitamin C related to the genetic basis of scurvy. Vitamins and Hormones 36, 33–52.

Vitamin D: Physiology, clinical applications, dietary sources, and requirements

M.F. Holick, Boston University Medical Center, Boston, MA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of M.F. Holick, Vitamin D: Physiology, Dietary Sources, and Requirements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 370–382, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00276-2.

Introduction 521
Origin and structure of vitamin D 523
Production of vitamin D in the skin 524
Absorption, metabolism, and excretion of vitamin D 526
Biologic functions of vitamin D on calcium metabolism 526
Evaluation for and consequences of vitamin D deficiency 526
Nonskeletal consequences of vitamin D deficiency 528
Recommended dietary intake of vitamin D 529
Recommendations for adequate intake for ages 0–12 months 529
Recommendations for adequate intake for ages 1–18 years 530
Recommended adequate intake for ages 19–70+ years 531
Recommendations for adequate intake for pregnancy and lactation 531
Healthy vitamin D intakes 531
Tolerable upper intake levels and vitamin D intoxication 531
Conclusion 532
References 533
Further reading 534
Suggested Reading 534

Key points

- Understand how vitamin D is produced in the skin during sunlight exposure.
- Appreciate that a variety of factors including time of day, season, latitude, skin pigmentation and aging can have a dramatic influence on the cutaneous production of vitamin D.
- Recognize a major cause of rickets is due to vitamin D deficiency.
- Increase knowledge about the many noncalcemic benefits of vitamin D for reducing risk of chronic illnesses including autoimmune disorders, type I and type 2 diabetes, neurocognitive function, some deadly cancers, neurocognitive dysfunction and acute infection including COVID 19.
- Understand the definition of vitamin D deficiency, insufficiency and sufficiency and vitamin D requirements for all 3 groups.
- To convert ng mL^{-1} to SI unit nmol L^{-1} multiply ng mL^{-1} by 2.5

Introduction

Vitamin D is a fat-soluble vitamin that is recognized for its importance for bone health. Vitamin D is neither a vitamin nor a nutrient. It is a hormone because exposure to sunlight can produce vitamin D in the skin (Wacker and Holick, 2013). There are two major forms of vitamin D (Holick, 2007). Vitamin D_2 originates from ultraviolet irradiated yeasts, mushrooms and plants. Vitamin D_3 is produced in the skin during sun exposure and is also found in oily fish including salmon, mackerel and herring. We take vitamin D for granted because it was casual exposure to sunlight that provided most humans with their vitamin D requirement when they were hunter gatherers. Due to our civilized lifestyle and concerns about skin cancer we no longer are able to get an adequate amount of vitamin D from sun exposure. The fortification of milk and other foods including some margarines and cereals with vitamin D have eradicated vitamin D deficiency rickets as a significant health problem for most children in the US and countries that practice this fortification process (Holick, 2006). India has now begun a fortification program where cooking oil and milk are fortified with vitamin D_2. Although there is some debate as to whether vitamin D_2 is as effective as vitamin D_3 in raising blood concentrations of 25(OH)D several studies have reported that physiologic doses of either form will raise blood concentrations of 25(OH)D to essentially the same degree (Holick et al., 2011).

It is now recognized that both children and adults worldwide are at risk for developing vitamin D deficiency and sufficiency. The Endocrine Society Guidelines defines deficiency and insufficiency as a serum concentration of 25-hydroxyvitamin D [25(OH)D] of <20 ng mL^{-1} (<50 nmol L^{-1}) and 21–29 ng mL^{-1} (51–74 nmol L^{-1}) respectively. Vitamin D sufficiency is defined as a serum 25(OH)D of between 30 (75 nmol L^{-1}) and 100 (250 nmol L^{-1}) ng mL^{-1} (nmol L^{-1}). The guidelines suggested that vitamin D intoxication is not usually seen until a circulating serum concentration of 25(OH)D is greater than 150 ng mL^{-1} (375 nmol L^{-1}) (Holick et al., 2011). In 2010 the Institute of Medicine recommended that to achieve a circulating blood level of 25(OH)D of at least 20 ng mL^{-1} required the daily intake of 400 IU day^{-1} for children 0–1 year, 600 IU day^{-1} for children over 1 year and all and adults <70 years old, and 800 IU day^{-1} for those >70 years (Ross et al., 2011). The Endocrine Society recommended that to achieve and maintain a circulating serum concentration of 25(OH)D of at least 30 ng mL^{-1} (75 nmol L^{-1}) required the daily ingestion of 400–1000 IUs, 600–1000 IUs and 1500–2000 IUs for infants up to 1 year of age, children 1 year of age and older and all adults of normal weight respectively. Obese adults (BMI >30) require 2–3 times more vitamin D to satisfy their requirement compared to a normal weight adult (**Table 1**) (Holick et al., 2011).

Once vitamin D_3 is formed in the skin or vitamin D_2 and/or vitamin D_3 are ingested from the diet or supplements, vitamin D (D represents D_2 and/or D_3) enters the bloodstream and travels to the liver and kidneys where it is hydroxylated on carbons 25 and 1 to form [25(OH)D] and 1,25-dihydroxyvitamin D [$1,25(OH)_2D$] respectively (Holick, 2007; Bouillon et al., 2019). 25-Hydroxyvitamin D is the major circulating form of the vitamin that is measured to determine the vitamin D status of patients (Holick et al., 2011). $1,25(OH)_2D$ is the biologically active form of vitamin D that is responsible for maintaining calcium homeostasis and bone health (Holick, 2007). It is now recognized that vitamin D deficiency may increase the risk of many acute and chronic diseases, including some deadly cancers including cancer of the breast, prostate, and colon, type I diabetes mellitus, multiple sclerosis, rheumatoid arthritis, neurocognitive dysfunction including Alzheimer's disease and depression, type 2 diabetes, cardiovascular disease and infectious diseases including influenza and COVID-1 (Grant, 2002; Bouillon et al., 2019; Charoenngam and Holick, 2020; Kaufman et al., 2020; Charoenngam et al., 2021; Seal et al., 2022).

Table 1 Sources of vitamin D_2 and vitamin D_3.

Source	*Vitamin D content (IU = 25 ng)*
Natural sources	
Cod liver oil	~400–1000 IU/tsp vitamin D_3
Salmon, fresh wild caught	~600–1000 IU/3.5 oz vitamin D_3
Salmon, fresh farmed	~100–250 IU/3.5 oz vitamin D_3, vitamin D_2
Salmon, canned	~300–600 IU/3.5 oz vitamin D_3
Sardines, canned	~300 IU/3.5 oz vitamin D_3
Mackerel, canned	~250 IU/3.5 oz vitamin D_3
Tuna, canned	~236 IU/3.5 oz vitamin D_3
Shiitake mushrooms, fresh	~100 IU/3.5 oz vitamin D_2
Shiitake mushrooms, sun dried	~1600 IU/3.5 oz vitamin D_2
Egg yolk	~20 IU/yolk vitamin D_3 or D_2
Sunlight/UVB radiation	~20,000 IU equivalent to exposure to 1 minimal erythemal dose (MED) in a bathing suit. Thus, exposure of arms and legs to 0.5 MED is equivalent to ingesting~3000 IU vitamin D_3
Fortified foods	
Fortified milk	100 IU/8 oz usually vitamin D_3
Fortified orange juice	100 IU/8 oz vitamin D_3
Infant formulas	100 IU/8 oz vitamin D_3
Fortified yogurts	100 IU/8 oz usually vitamin D_3
Fortified butter	56 IU/3.5 oz usually vitamin D_3
Fortified margarine	429/3.5 oz usually vitamin D_3
Fortified cheeses	100 IU/3 oz usually vitamin D_3
Fortified breakfast cereals	100 IU/serving usually vitamin D_3
Pharmaceutical sources in the US	
Vitamin D_2 (ergocalciferol)	50,000 IU/capsule
Drisdol (vitamin D_2) liquid	8000 IU/cc
Supplemental sources	
Multivitamin	400, 500, 1000 IU vitamin D_3 or vitamin D_2
Vitamin D_3	400, 800, 1000, 2000, 5000, 10,000, and 50,000 IU

Modified from Holick (2007) Vitamin D deficiency. *New England Journal of Medicine* 357: 266–281.

Origin and structure of vitamin D

As the industrial revolution began to take hold in Northern Europe in the fifteenth century, it was quickly associated with a new disease that caused severe growth retardation and bony deformities in young children (Fig. 1) (Holick, 2006). This disease was commonly known as rickets or "English disease" and plagued the children of the industrialized cities in Europe and North America for more than 250 years. Although Sniadecki in 1822 and Palm in 1890 both recognized that it was lack of exposure to sunlight that was the likely cause of rickets in children, Huldschinsky, in 1919, was the first to prove that exposure of the skin to ultraviolet radiation could cure rickets. Within 2 years, Hess and Unger reported that exposure of several rachitic children to sunlight was adequate for curing this bone-deforming disease.

Steenbock and Black and Hess independently recognized that exposure of animals and their food to ultraviolet radiation imparted antirachitic activity. This led to the recommendation for the ultraviolet irradiation of foods as a means of fortifying them with vitamin D. This resulted in the addition of provitamin D to milk followed by ultraviolet irradiation. As soon as it was possible to commercially synthesize vitamin D_2 from yeast in large quantities, it was added directly to milk and other foods (Wacker and Holick, 2013).

The first vitamin D was isolated from the irradiation of the yeast sterol ergosterol (Fig. 3). This vitamin D was thought to be identical to that produced in the skin of animals and humans. However, studies revealed that when vitamin D produced from yeast was fed to chickens, they were unable to utilize it and developed rickets. When chickens were fed natural vitamin D from fish liver oil, rickets was prevented. This led to the conclusion that vitamin D originating from yeast was different from that in fish liver oil and animal and human skin. In 1937, this mystery was solved when the structure of provitamin D from pig skin was determined. A



Fig. 1 This is a typical presentation of a child with rickets. The child is suffering from severe muscle weakness, has bony deformities including bowed legs, and knob-like projects in the middle of his ribcage called the rachitic rosary. Reproduced from Fraser D and Scriver CR (1979) Disorders associated with hereditary or acquired abnormalities of vitamin D function: Hereditary disorders associated with vitamin D resistance or defective phosphate metabolism. In: De Groot LJ. et al. (eds.) *Endocrinology*, pp. 797–808. New York: Grune and Stratton.

structural analysis revealed that provitamin D derived from ergosterol differed from that derived from pig skin. The provitamin D (ergosterol; provitamin D_2) that came from yeast had a double bond between carbons 22 and 23 and a methyl group on carbon 24. The provitamin D in animal skin had a side chain that was identical to cholesterol, that is, it did not contain either a double bond or methyl group on carbons 22–23 and 24, respectively, and was identified as 7-dehydrocholesterol (provitamin D_3) (**Fig. 2**). The vitamin Ds generated from ergosterol and 7-dehydrocholesterol were called ergocalciferol (vitamin D_2) and cholecalciferol (vitamin D_3), respectively (Wacker and Holick, 2013).

Production of vitamin D in the skin

During exposure to sunlight, the ultraviolet B photons with energies between 290 and 315 nm are absorbed by provitamin D_3 (7-dehydrocholesterol) in the skin. This absorption results in a photolysis of the B-ring of provitamin D_3 resulting in the formation of previtamin D_3 (**Fig. 3**). However, because previtamin D_3 is thermodynamically unstable, it quickly undergoes an isomerization (rearrangement) of its triple bond system to form vitamin D_3. This isomerization process is enhanced in skin cells because the previtamin D_3 is synthesized in the cell membrane, which restricts its movement thereby accelerating the transformation of previtamin D_3 to vitamin D_3. Once vitamin D_3 is formed in the skin cell membrane, it is no longer restricted in its movement and freely translocates into the extracellular space to find its way into the dermal capillary bloodstream where it is bound to a specific vitamin D-binding protein (**Fig. 3**).

An increase in skin pigmentation and zenith angle of the sun (change in latitude, season, and time of day) and the topical application of a sunscreen can markedly diminish or even prevent the production of vitamin D_3 in the skin. Over the age of ~65 years, there is a three- to fourfold decline in the synthetic capacity of the skin to produce vitamin D_3. Excessive exposure to sunlight cannot cause vitamin D_3 intoxication because once previtamin D_3 and vitamin D_3 are made in the skin, excessive quantities are rapidly destroyed by sunlight (**Fig. 3**) (Holick, 2007; Wacker and Holick, 2013).

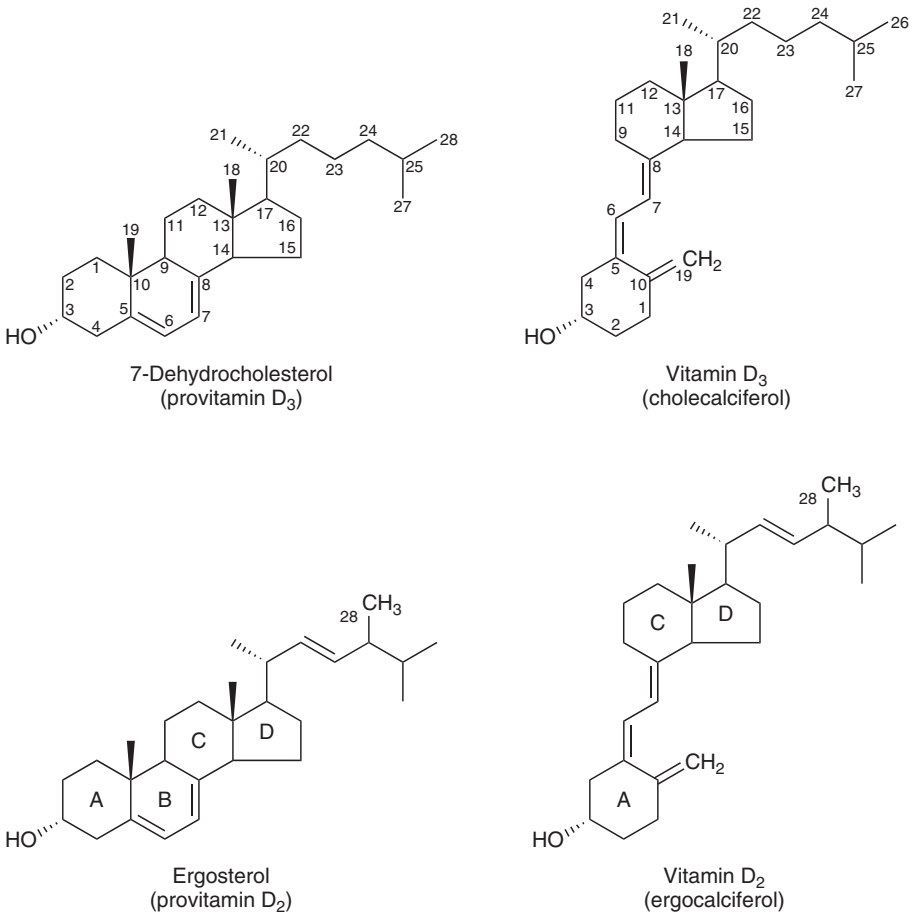


Fig. 2 Structures for 7-dehydrocholesterol (provitamin D_3), ergosterol (provitamin D_2), vitamin D_3 (cholecalciferol), and vitamin D_2 (ergocalciferol). The carbons are numbered and the ring systems are labeled. Reproduced with permission from Holick MF Copyright 2007.

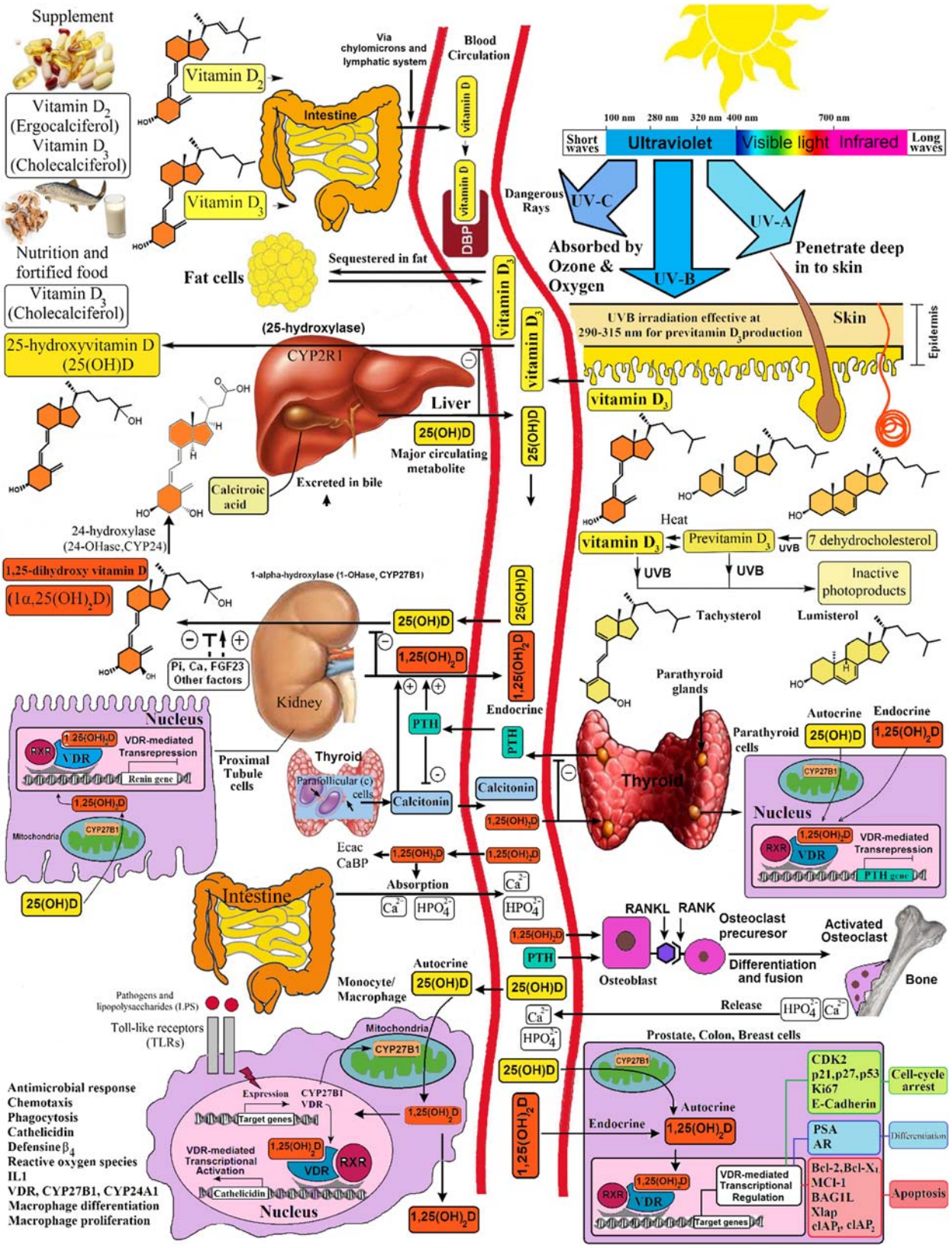


Fig. 3 Schematic representation of the synthesis and metabolism of vitamin D for skeletal and nonskeletal function. 1-OHase = 25-hydroxyvitamin D-1a-hydroxylase; 24-OHase = 25-hydroxyvitamin D-24-hydroxylase; 25(OH)D = 25-hydroxyvitamin D; 1,25(OH)2D = 1,25-dihydroxyvitamin D; CaBP = calcium-binding protein; CYP27B1, Cytochrome P450-27B1; DBP = vitamin D binding protein; ECaC = epithelial calcium channel; FGF-23 = fibroblast growth factor-23; PTH = parathyroid hormone; RANK = receptor activator of the NF-kB; RANKL = receptor activator of the NF-kB ligand; RXR = retinoic acid receptor; TLR2/1 = Toll-like receptor 2/1; VDR = vitamin D receptor; vitamin D = vitamin D2 or vitamin D3. Copyright Holick (2013), reproduced with permission.

Absorption, metabolism, and excretion of vitamin D

Vitamin D (vitamin D without a subscript represents either vitamin D_2 or D_3) is fat soluble and, therefore, once ingested vitamin D_2 and vitamin D_3 are incorporated into the chylomicron fraction and absorbed in the small intestine into the lymphatic system. Both dietary vitamin D_2 and vitamin D_3, and cutaneous vitamin D_3 enter the circulation and are bound to a specific α_1-globulin known as the vitamin D-binding protein (DBP). It is believed that this protein acts as a buffering system whereby it helps maintain circulating concentrations of 25(OH)D. The bound form of 25(OH)D can enter into the renal tubular cells via the membrane receptor megalin. Once inside the renal tubule 25(OH)D is dissociated from DBP and enters the mitochondria to be metabolized to 1,25-dihydroxyvitamin D [$1,25(OH)_2D$](**Fig. 3**) (Hossein-nezhad and Holick, 2013).

Neither vitamin D_2 nor vitamin D_3 possess any intrinsic biologic activity on calcium metabolism. They both require a hydroxylation on carbon 25 to form 25(OH)D (**Fig. 4**). When given as a 1000 or 2000 IU supplement, vitamin D_2 is as effective as vitamin D_3 in raising serum 25(OH)D levels. 25(OH)D is the major circulating form of vitamin D, and at physiologic concentrations, it too has little biologic activity on calcium metabolism. It must undergo a hydroxylation on carbon 1 in the kidney to form $1,25(OH)_2D$, the biologically active form of vitamin D (**Fig. 3**). The metabolism of 25(OH)D to $1,25(OH)_2D$ is tightly regulated by parathyroid hormone (PTH), fibroblast growth factor 23 (FGF23), and serum phosphorus levels (**Fig. 5**). PTH and low serum phosphorus levels increase the production of $1,25(OH)_2D$, whereas FGF23 suppresses its production (DeLuca, 1988; Holick, 2007; Hossein-nezhad and Holick, 2013; Charoenngam et al., 2021).

25(OH)D and $1,25(OH)_2D$ act as substrate for a 24-hydroxylase (an enzyme that attaches a hydroxyl on carbon-24), which is found in the kidneys and other target tissues for $1,25(OH)_2D$. Once $1,25(OH)_2D$ is hydroxylated on carbon 24, this is the first step in its degradation to a water-soluble acid, calcitroic acid (**Fig. 4**). Vitamin D and calcitroic acid are excreted in the bile (Holick, 2007; Hossein-nezhad and Holick, 2013).

Biologic functions of vitamin D on calcium metabolism

$1,25(OH)_2D$ interacts with a specific nuclear receptor that is commonly known as the vitamin D receptor (VDR) and is one of the many members of the super family of steroid hormone receptors that includes retinoic acid, thyroid hormone, glucocorticoids, and sex steroids. Once $1,25(OH)_2D$ interacts with the VDR, the complex forms a heterodimer with retinoic acid X receptor (RXR) (**Fig. 3**). This new complex sits on specific segments of vitamin D responsive genes known as vitamin D responsive elements (VDREs) to either increase or decrease transcriptional activity of the vitamin D-sensitive genes such as osteocalcin, calcium binding protein (calbindin), PTH, calcium channel and osteonectin (**Fig. 3**) (Hossein-nezhad and Holick, 2013).

In the intestine, $1,25(OH)_2D$ enhances the absorption of dietary calcium (from ~15% to 30–40% in adults) and phosphorus (from ~60 to ~80%) across the microvilli of the small intestinal absorptive cells (**Fig. 3**). $1,25(OH)_2D$ also interacts with osteoblasts to stimulate the expression for receptor activator NFκB ligand (RANKL) in the bone to initiate the transformation of monocytes into mature osteoclasts (**Fig. 3**). Thus, $1,25(OH)_2D_3$ regulates serum calcium levels by enhancing the efficiency of intestinal calcium absorption and stimulating resorption of calcium from the bone (Holick, 2007; Hossein-nezhad and Holick, 2013).

There are a variety of other tissues including the brain, gonads, pancreas, stomach, activated T and B lymphocytes, monocytes, macrophages, and skin that have a nuclear VDR. Although the exact physiologic function of $1,25(OH)_2D$'s interaction with these VDRs is not well understood, it is known that *in vivo* and *in vitro* $1,25(OH)_2D_3$ can inhibit proliferation and induce terminal differentiation of various normal and tumor cells including normal human keratinocytes. This is the reason why activated vitamin D compounds are now routinely used for the treatment of the hyperproliferative skin disorder psoriasis (Wacker and Holick, 2013). It has been estimated that at least 200 and as many as 2000 genes are influenced directly or indirectly by $1,25(OH)_2D_3$ (Carlberg and Haq, 2018; Shirvani et al., 2019; Charoenngam and Holick, 2020).

Evaluation for and consequences of vitamin D deficiency

Vitamin D deficiency in infants and children causes rickets. As a child becomes vitamin D deficient, this results in a decrease in the efficiency of intestinal calcium absorption. There is a decline in blood-ionized calcium, which causes the parathyroid glands to produce and secrete more PTH. PTH tries to conserve calcium by enhancing tubular reabsorption of calcium in the kidneys. However, in the face of developing hypocalcemia, which could disturb neuromuscular function and a wide variety of metabolic and cellular processes, the body calls upon $1,25(OH)_2D$ and PTH to stimulate the expression of RANKL in osteoblasts to mobilize monocytic stem cells to become functional osteoclasts, which, in turn, mobilize calcium from the skeleton (**Fig. 3**) (Aksenes and Aarskog, 1982; Holick, 2006; Hossein-nezhad and Holick, 2013). In addition, PTH causes a loss of phosphorus into the urine causing hypophosphatemia (low blood phosphate concentration). Thus, in early vitamin D deficiency, the serum calcium is normal; it is the low serum phosphorus that causes the extracellular $CaXPO_4$ to be too low for normal mineralization of bone matrix. This causes a disruption in the orderly sequence of events in the differentiation of hypertrophied chondrocytes in the epiphyseal plates resulting in their disorganization causing a widening of the epiphyseal plates (end of long bones). PTH causes a demineralization of the skeleton, and newly laid down collagen matrix is unable to be mineralized All of these in concert cause bony deformities seen in rachitic children (**Fig. 1**) (Holick, 2006; Hossein-nezhad and Holick, 2013, Charoenngam and Holick, 2020).

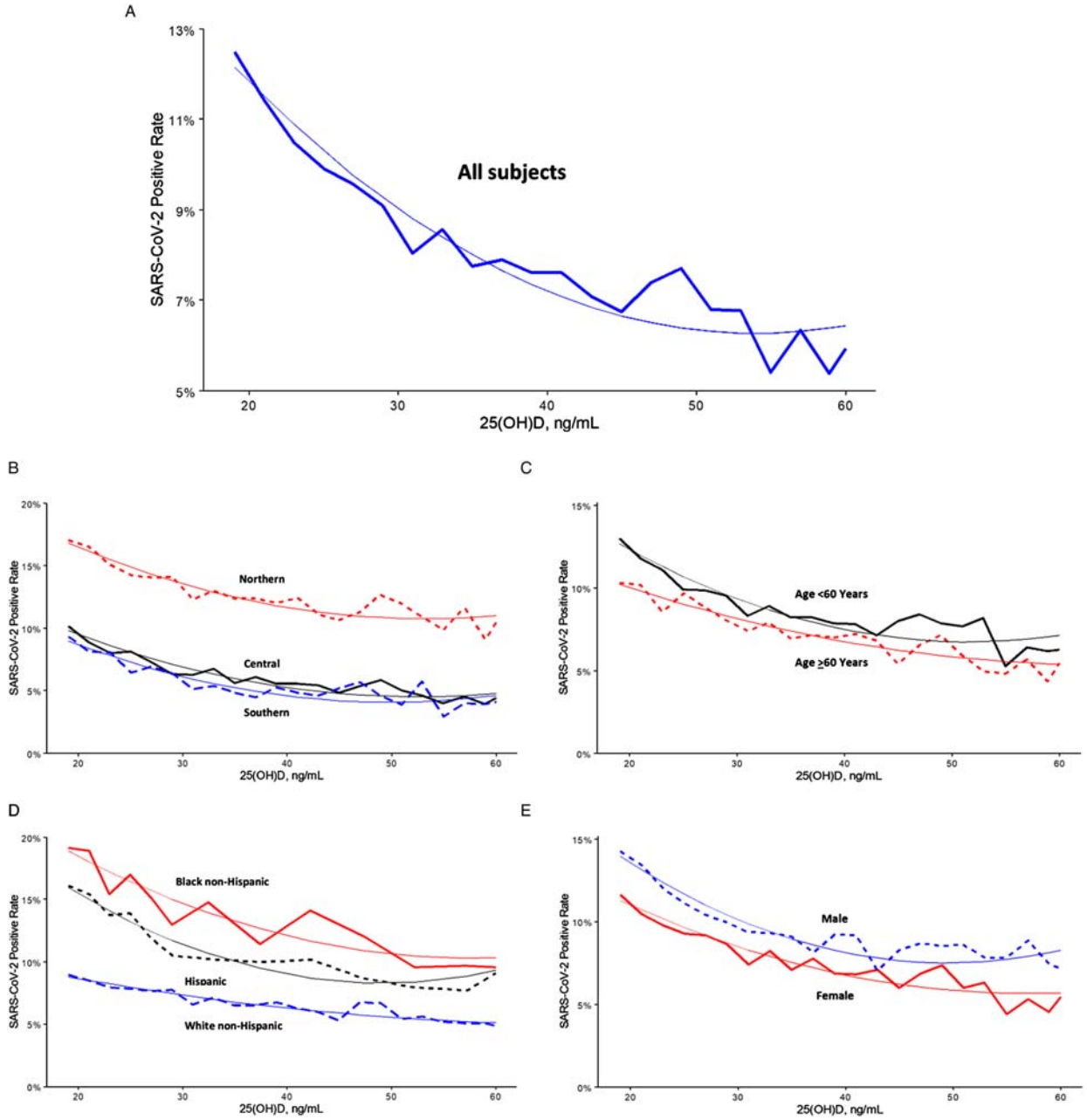
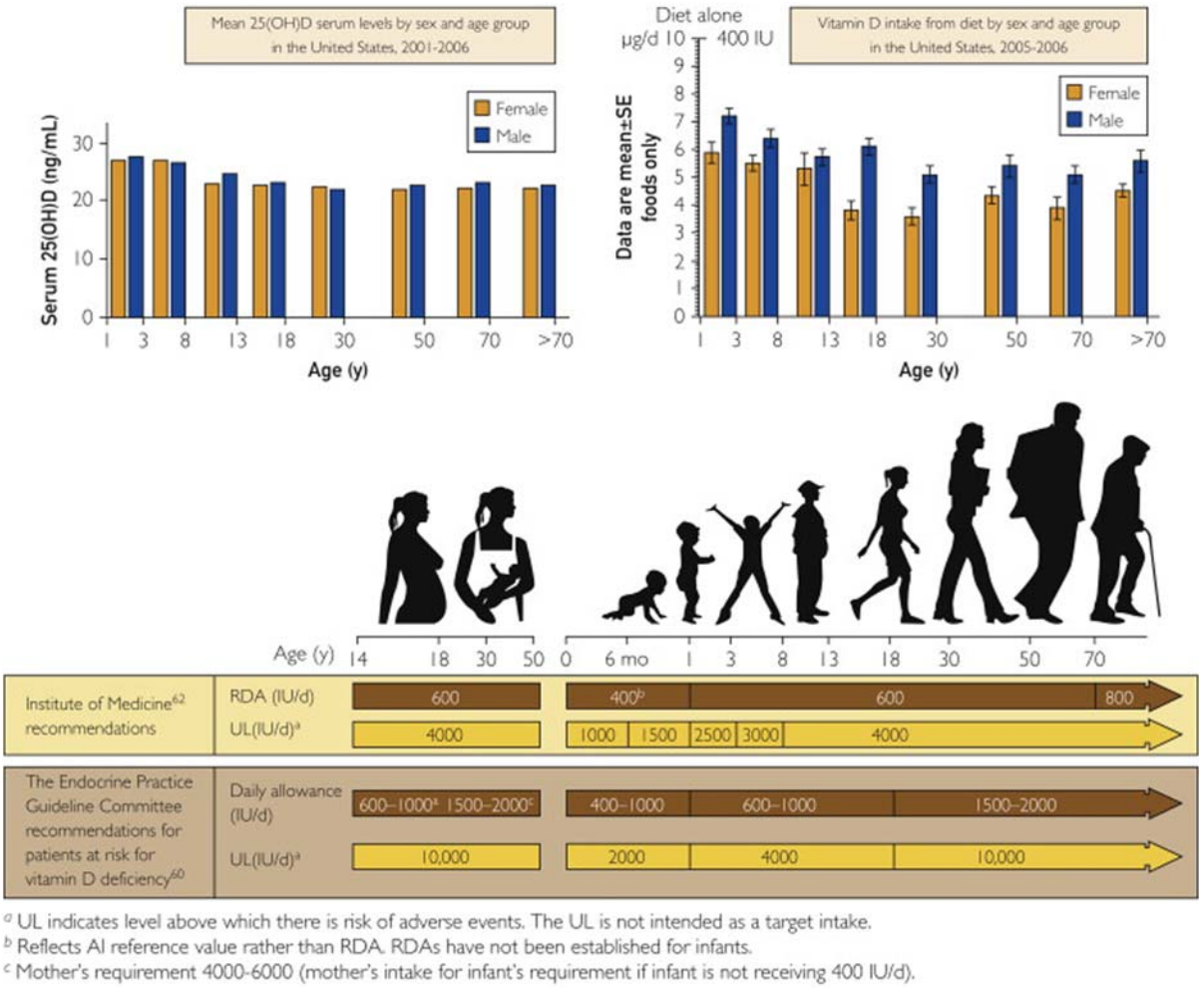


Fig. 4 SARS-CoV-2 nucleic acid amplification test positivity rates and circulating 25(OH)D levels in all subjects (A) and stratified by latitude region (B), predominately Black non- Hispanic, Hispanic and White non-Hospanic zip codes (C), age group (D), and sex (E). Smooth lines represent the weighted second order polynomial regression fit to the data associating circulating 25(OH)D levels (*x*-axis) and SARS-CoV-2 positivity rates (*y*-axis). 25(OH)D ¼ 25-hydroxyvitamin D; SARS-CoV-2 ¼ severe acute respiratory distress syndrome coronavirus 2. Copyright Kaufman et al. (2020) reproduced with permission.

Once the epiphyseal plates are closed later in adolescence, vitamin D deficiency can no longer cause bone deformities. Instead, there is an inability to mineralize newly deposited bone matrix leading to wide osteoid seams within the trabecular and cortical bone causing the bone disease commonly known as osteomalacia. Although osteomalacia is not associated with bone deformities it is associated with aching throbbing bone pain and muscle weakness that can be misdiagnosed as fibromyalgia and chronic fatigue syndrome respectively. In addition, the secondary hyperparathyroidism that results from vitamin D deficiency results in the mobilization of precious calcium stores from the skeleton thereby exacerbating bone loss and causing osteoporosis. This can increase a person's risk for fracture (Holick, 2007; Hossain and Holick, 2013; Bouillon et al., 2019). A study by Priemel et al., (2010)



[a] UL indicates level above which there is risk of adverse events. The UL is not intended as a target intake.
[b] Reflects AI reference value rather than RDA. RDAs have not been established for infants.
[c] Mother's requirement 4000-6000 (mother's intake for infant's requirement if infant is not receiving 400 IU/d).

Fig. 5 Vitamin D intakes recommended by the Institute of Medicine and the Endocrine Practice Guidelines Committee. 25(OH)D = 25-hydroxyvitamin D; AI = adequate intakes; RDA = recommended dietary allowance; SE = standard error; UL = tolerable upper intake level. Copyright Holick (2013), reproduced with permission.

reported that in a large cohort of adults ages 20 and above who had died unexpectedly that they found no evidence of osteomalacia when the circulating concentration of 25(OH)D was at least 30 ng mL^{-1} (75 nmol L^{-1}).

The hallmark for determining the vitamin D status is the measurement of the circulating concentration of 25(OH)D. The 25(OH)D is low or undetectable in vitamin D deficiency and markedly elevated in vitamin D intoxication. The Endocrine Society Practice Guidelines on Vitamin D defined vitamin D deficiency and insufficiency as a 25(OH)D of <20 ng mL^{-1} (<50 nmol L^{-1}) and 21–29 ng mL^{-1} (51–74 nmol L^{-1}) and vitamin D sufficiency is defined as a serum 25(OH)D of between 30 (75 nmol L^{-1}) and 100 (250 nmol L^{-1}) ng mL^{-1} (nmol L^{-1}) respectively. Vitamin D intoxication is normally not observed until the blood level of 25(OH)D > 150 ng mL^{-1} (250 nmol L^{-1}). Measurement of $1,25(OH)_2D$ is of little value for determining the vitamin D nutritional status because its synthesis is tightly regulated. Indeed, as a person becomes vitamin D deficient, there is an increase in the secretion of PTH which, in turn, increases the production of $1,25(OH)_2D$. Thus, early in vitamin D deficiency one can see a normal fasting serum calcium, low-normal to low phosphorus, low 25(OH)D, and elevated PTH, $1,25(OH)_2D$ and alkaline phosphatase. In chronic vitamin D deficiency, all the above are seen with the exception that serum calcium and $1,25(OH)_2D$ are low-normal or low (Holick, 2007).

Nonskeletal consequences of vitamin D deficiency

As early as 1941, it was appreciated that if you lived at higher latitudes in the US you were at higher risk of dying of cancer (Apperly, 1941). A multitude of epidemiologic studies clearly show that if you live at higher latitudes and are more prone to vitamin D deficiency, then you are at higher risk of dying of colon, prostate, breast, ovarian, and a variety of other cancers (Grant, 2002; Hosseinnezhad and Holick, 2013; Wacker and Holick, 2013). It is also known that living at higher latitudes increases risk of having high

blood pressure and heart disease as well as autoimmune diseases including multiple sclerosis and type I diabetes (Wacker and Holick, 2013).

Essentially every cell and organ in the body requires vitamin D, that is, they all have a VDR. It is also known that most tissues in the body can activate vitamin D (Hossain and Holick, 2013; Bouillon et al., 2019; Charoenngam and Holick, 2020). Thus, maintaining adequate concentrations of 25(OH)D in the circulation of at least 30 ng mL^{-1} may be necessary for various organs including colon, breast, and macrophages and dendritic cells to convert it to $1,25(OH)_2D$, which in turn can help regulate various genes responsible for cell growth and differentiation (**Fig. 3**) (Hossein-nezhad and Holick, 2013; Bouillon et al., 2019; Shirvani et al., 2019; Charoenngam and Holick, 2020). This could be the explanation for how vitamin D sufficiency is protective against many common cancers (Grant, 2002; Hossein-nezhad and Holick, 2013). The immune cells also recognize $1,25(OH)_2D_3$. This may explain why children who had received 2000 IU of vitamin D a day during their first year decreased their risk of developing type I diabetes by 88%. A study in Japanese children who received 1200 IU vitamin D_3/d from December through March had a 42% reduced risk for developing influenza A infection compared to children who received a placebo pill (Hypponen et al., 2001; Urashima et al., 2010; Bouillon et al., 2019; Charoenngam and Holick, 2020). A study of more than 191,000 COVID 19 positive patients revealed a 54% reduced risk for those patients who had a blood level of 25(OH)D of 34 ng mL^{-1} compared to patients with a blood level of less than 20 ng mL^{-1} (Kaufman et al., 2020). The benefit of having less infection continued up to a blood level of 55 ng mL^{-1} (**Fig. 4**). A major consequence of being infected with COVID 19 is the cytokine storm which increases risk for morbidity and mortality. Several studies have reported that patients who had a blood concentration of 25(OH)D of at least 30 ng mL^{-1} significantly reduced risk of unconsciousness, intensive care admission and mortality (Charoenngam et al., 2021; Seal et al., 2022). Increasing intake of vitamin D and sun exposure has been associated with decreased risk of developing multiple sclerosis, rheumatoid arthritis, and even Crohn's disease (Moan et al., 2008; Wacker and Holick, 2013).

The relationship of vitamin D to cardiovascular disease is finally being understood. $1,25(OH)_2D$ inhibits the production of the blood pressure hormone renin. It also alters cardiomyocyte growth and modulates the inflammatory response of atherosclerosis. Vitamin D deficiency has been associated with a 50% increased risk for having a heart attack and stroke, and 80% increased risk for developing peripheral vascular disease. When African-American teenagers were given 2000 IU vitamin D_3/d for 4 months, they had a significant reduction in arterial wall stiffness, a prelude to hypertension and artherosclerosis when compared to teenagers receiving 400 IU vitamin D_3/d for four months (Melamed et al., 2008; Dong et al., 2010; Dobnig et al., 2008; Hossain and Holick, 2013; Dudenkov et al., 2018).

Recommended dietary intake of vitamin D (Fig. 5)

Vitamin D is very rare in foods naturally, with the exception of fatty fish, some fish liver oils, and mushrooms exposed to UVR (**Fig. 3**). Milk and some dairy products in the United States are fortified with vitamin D_3/8 oz or serving. Some orange juice and other juice products are fortified with calcium and 100 IU of vitamin D_3/8 oz (Holick, 2007). In India milk and cooking oil is fortified with vitamin D_2. Multivitamin preparations that contain vitamin D are a good source of vitamin D as are pharmaceutical preparations (**Table 2**).

In 2010, the Institute of Medicine and the National Academy of Sciences reviewed the recommended dietary intake for calcium and vitamin D. The recommended dietary allowance (RDA) was defined as the daily intake level that is sufficient to meet nutrient requirements for nearly all (97–98%) individuals in life-stage and gender group. The RDA was meant to apply to individuals and not groups. When sufficient scientific evidence was not available to calculate an estimated average requirement (EAR), that is, a nutrient value that was estimated to meet the requirement defined by a specified indicator of adequacy in 50% of individuals in a life-stage and gender group, the Committee recommended using an adequate intake (AI). The AI is based on the observation of experimentally determined approximations of average nutrient intake by a defined population or subgroup that appears to sustain a defined nutritional state such as normal circulation nutrient values or growth. Because sunlight played such an important role in providing humans with their vitamin D requirement and, therefore, was a variable that was difficult to quantify in studies in infants that were reviewed by the Committee, it was concluded that an AI rather than an RDA should be used for vitamin D (**Fig. 5**, **Table 2**) (Ross et al., 2011). At the same time the Endocrine Society formed a committee to evaluate how to treat and prevent vitamin D deficiency as it relates to skeletal health. The Endocrine Society Guidelines on Vitamin D was published in 2011 and the recommendations are included in **Table 2** (Holick et al., 2011).

Recommendations for adequate intake for ages 0–12 months

It is well documented that human and cows' milk has very little vitamin D naturally (Hollis and Wagner, 2004). Human milk contains on average between 10 and 50 IU L^{-1} (0.25–1.25 μg). This is dependent on the mother's exposure to sunlight and her vitamin D intake. Several studies have suggested that infant intakes of vitamin D of between 8.5 (340 IU) and 15 μg (600 IU) day^{-1} would provide the maximum effect on their linear growth. A study in infants from Northern China (40–47° N) found that vitamin D supplements of 2.5 (100 IU), 5 (200 IU), or 10 μg (400 IU) day^{-1} resulted in 36, 29, and 2% of the infants being vitamin D deficient with 25(OH)D levels of less than 25 nmol L^{-1} (10 ng mL^{-1}). None of the infants, however, had overt manifestations of rickets i.e bony deformities. Chinese infants from two southern cities (22° N and 30° N) maintained normal vitamin D status on as little as 2.5 μg (100 IU) day^{-1} of vitamin D (Holick, 2006; Hossain and Holick, 2013).

Table 2 Adequate intakes (AI), daily allowance and tolerable upper limit (UL) for Vitamin D.

	IOM recommendations				*The current endocrine society practice guidelines*	
Life stage group	*AI*	*EAR*	*RDA*	*UL*	*Daily allowance (IU day^{-1})*	*UL (IU)*
Infants (months)						
0–6	~400 IU ~(10 μg)	–	–	1000 IU (25 μg)	400–1000	2000
6–12	~400 IU ~(10 μg)	–	–	1500 IU (38 μg)	400–1000	2000
Children (years)						
1–3	–	C	600 IU (15 μg)	2500 IU (63 μg)	600–1000	4000
4–8	–	400 IU (10 μg)	600 IU (15 μg)	3000 IU (75 μg)	600–1000	4000
Males (years)						
9–13	–	400 IU (10 μg)	600 IU (15 μg)	4000 IU (100 μg)	1500–2000	4000
14–18	–	400 IU (10 μg)	600 IU (15 μg)	4000 IU (100 μg)	1500–2000	4000
19–30	–	400 IU (10 μg)	600 IU (15 μg)	4000 IU (100 μg)	1500–2000	10,000
31–50	–	400 IU (10 μg)	600 IU (15 μg)	4000 IU (100 μg)	1500–2000	10,000
51–70	–	400 IU (10 μg)	600 IU (15 μg)	4000 IU (100 μg)	1500–2000	10,000
>70	–	400 IU (10 μg)	800 IU (20 μg)	4000 IU (100 μg)	1500–2000	10,000
Females (years)						
9–13	–	400 IU (10 μg)	600 IU (15 μg)	4000 IU (100 μg)	1500–2000	4000
14–18	–	400 IU (10 μg)	600 IU (15 μg)	4000 IU (100 μg)	1500–2000	4000
19–30	–	400 IU (10 μg)	600 IU (15 μg)	4000 IU (100 μg)	1500–2000	10,000
31–50	–	400 IU (10 μg)	600 IU (15 μg)	4000 IU (100 μg)	1500–2000	10,000
51–70	–	400 IU (10 μg)	600 IU (15 μg)	4000 IU (100 μg)	1500–2000	10,000
>70	–	400 IU (10 μg)	800 IU (20 μg)	4000 IU (100 μg)	1500–2000	10,000
Pregnancy (years)						
14–18	–	400 IU (10 μg)	600 IU (15 μg)	4000 IU (100 μg)	1500–2000	10,000
19–30	–	400 IU (10 μg)	600 IU (15 μg)	4000 IU (100 μg)	1500–2000	10,000
31–50	–	400 IU (10 μg)	600 IU (15 μg)	4000 IU (100 μg)	1500–2000	10,000
Lactation* (years)						
14–18	–	400 IU (10 μg)	600 IU (15 μg)	4000 IU (100 μg)	1500–2000	10,000
19–30	–	400 IU (10 μg)	600 IU (15 μg)	4000 IU (100 μg)	1500–2000	10,000
	–					

Recommended adequate intakes (AI), estimated average requirement (EAR), recommended dietary allowance (RDA) and tolerable upper limit (UL) by the Institute of Medicine (IOM) and Dr. Holick's recommendation for Daily Allowance and safe Upper Limit (UL) for vitamin D for children and adults who are not obtaining adequate vitamin D from sun exposure and who are at risk for vitamin D deficiency.

Modified from Holick et al. (2011) Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* 96(7): 1911–1930.

There was a seasonal variation of vitamin D status of infants when they were fed human milk only and did not receive vitamin D supplements; their 25(OH)D levels decreased in the winter due to less exposure to sunlight. However, this decrease did not occur in infants receiving a vitamin D supplement of 10 μg (400 IU) day^{-1} beginning at 3 weeks of age (Aksenes and Aarskog, 1982; Holick, 2006; Kumar et al., 2009; Grober et al., 2013).

Therefore, based on the available literature, it was concluded that a minimum intake of 2.5 μg (100 IU) day^{-1} of vitamin D was adequate to prevent rickets. However, at this intake and in the absence of sunlight, infants are at risk for developing hypovitaminosis D; therefore, it was recommended by the Institute of Medicine (IOM) that an adequate intake of 10 μg day^{-1} (400 IU) was prudent for infants during their first year of life (Ross et al., 2011). The American Academy of Pediatrics emphasized the importance of providing breast fed infants with 400 IUs daily since there is little if any vitamin D in human breast milk. An alternative is for mothers who are breast-feeding their infant to ingest 6400 IUs daily. This amount of intake places enough vitamin D in her milk to satisfy her infant's requirement. The Endocrine Society Practice Guidelines on Vitamin D recommended infants should receive 600–1000 IUs for maximum bone health (Holick et al., 2011).

Recommendations for adequate intake for ages 1–18 years

There are no studies in the scientific literature that systematically evaluated the influence of different amounts of vitamin D on either serum 25(OH)D or bone mineral content in this age group. Sunlight exposure is very important for this age group to obtain its required vitamin D. In South Africa, children aged 1–8 years of mixed race showed no evidence of vitamin D deficiency (Holick et al., 2011).

During puberty, there is a need to increase the efficiency of dietary calcium absorption in order to satisfy the rapid growth of the skeleton. As a result, there is an increase in the metabolism of 25(OH)D to $1,25(OH)_2D$. Because the blood levels of $1,25(OH)_2D$ are approximately 1000 times less than 25(OH)D, this increase in metabolism does not appear to increase the requirement of vitamin D for either boys or girls between the ages of 8 and 18 years. Girls of 9–17 years in Lebanon who took 2000 IU day^{-1} for 1 year had better bone density and muscle strength compared to girls who received 400 IU vitamin D_3 per day. Therefore, based on the available literature, it appears that children between 1 and 18 years obtain some of their vitamin D from exposure to sunlight. The IOM recommended they ingest 15 μg (600 IU) day^{-1}. The Endocrine Society Practice Guidelines on Vitamin D recommended 600–1000 IUs daily for maximum bone health (Ross et al., 2011; Holick et al., 2011).

Recommended adequate intake for ages 19–70+ years

There is only sparse literature regarding the roles that sunlight and diet play in maintaining an adequate vitamin D status for men and women in this age group. This age group depends on sunlight for some of its vitamin D requirement (Wacker and Holick, 2013). There was strong evidence-based literature that demonstrated a decrease in the circulating concentration of 25(OH)D, and an increase in the PTH level correlated with an increased risk of skeletal fractures in both the hip and spine in this age group. Studies in both men and women supplemented with 10–25 μg day^{-1} of vitamin D demonstrated reduced bone resorption, increased bone mineral content, and a decrease in vertebral and nonvertebral fractures. An evaluation of 333 ambulatory Caucasian women (mean age 58 ± 6 years) found that serum PTH concentrations were elevated in the winter (between March and May) in women consuming less than 5.5 μg (220 IU) day^{-1} of vitamin D. There was no seasonal variation in serum PTH concentrations when vitamin D intakes were greater than 5.5 μg (220 IU) day^{-1}. When bone loss was evaluated between seasons in women (62 ± 0.5 years) who had a usual vitamin D intake of 2.5 μg day^{-1}, a dietary supplement of 10 μg day^{-1} decreased spinal and hip-bone density loss. An analysis of the skeletons of German adults at autopsy revealed no evidence of vitamin D deficiency osteomalacia for those who had a 25(OH)D > 30 ng mL^{-1}. Regardless of exposure to sunlight, it was estimated by the IOM that 15 μg (600 IU) day^{-1} meets this age group's needs for bone health. They recommended for adults over the age of 70 years that they required 800 IUs daily (Ross et al., 2011). The Endocrine Society Practice Guidelines on Vitamin D recommended for all adults 19 years and above require the same amount of vitamin D for maximum bone health which was 1500–2000 IUs daily. They also recommended that for obese adults they require 2–3 times more vitamin D to satisfy their vitamin D requirement. This is because vitamin D be fat soluble and is diluted in the body fat and therefore not directly bioavailable to the body (Holick et al., 2011).

Recommendations for adequate intake for pregnancy and lactation

Although there is an increase in the metabolism of 25(OH)D to $1,25(OH)_2D$, during the last trimester of pregnancy and during lactation, the IOM concluded there is nothing in the evidence-based literature to suggest that there is an increased vitamin D requirement for pregnant and lactating women. Therefore, it was recommended that the RDA of vitamin D for pregnancy and lactation follows that recommended for their age group, that is, 15 μg (600 IU) day^{-1}. However, a study in Boston reported that 81% newborns and 76% of their mothers had a 25(OH)D < 20 ng mL^{-1} even though they took a prenatal vitamin supplement containing 400 IU and drank two glasses of milk per day thereby ingesting 600 IU day^{-1} (Lee et al., 2007). Several reports have suggested that pregnant women who maintain a serum concentration of 25(OH)D of at least 30 ng mL^{-1} reduced their risk of preeclampsia, premature births and need for a Cesarean section. The Endocrine Society Practice Guidelines recommended that pregnant and lactating women require this same amount of vitamin D as all adults (Holick et al., 2011).

Healthy vitamin D intakes

There have been a multitude of studies during the past decade that suggest that the RDAs recommended for vitamin D may still be inadequate if there is no exposure to sunlight. Based on a multitude of studies, it is reasonable for children and adults to increase their vitamin D intake to the concentrations recommended by the Endocrine Society Practice Guidelines (**Table 1**). In order to maintain a blood level of 25(OH)D > 30 ng mL^{-1} and not to exceed 100 ng mL^{-1} for all the potential health benefits of vitamin D (**Fig. 6**) (Holick et al., 2011; Charoenngam and Holick, 2020).

Tolerable upper intake levels and vitamin D intoxication

An excessive intake of vitamin D can lead to vitamin D intoxication. This is characterized by a marked increase in serum concentration of 25(OH)D that is usually greater than 150 ng mL^{-1} (375 nmol L^{-1}) and is associated with hypercalciuria and hypercalcemia. There is a compensatory decrease in serum concentrations of PTH and often an increase in serum concentrations of phosphate. This can lead to soft tissue calcification of the kidneys (nephrocalcinosis) and arteries and increased risk of kidney stones. The safe upper limits for vitamin D, as recommended by the IOM and the Endocrine Society Guidelines on Vitamin D, are found in **Table 1** (Holick, 2007; Ross et al., 2011; Holick et al., 2011).

Vitamin D intoxication usually occurs when a person ingests more than 100,000 IU of vitamin D daily for at least several months. A person does not need to be concerned about becoming vitamin D intoxicated if they take a multivitamin that contains

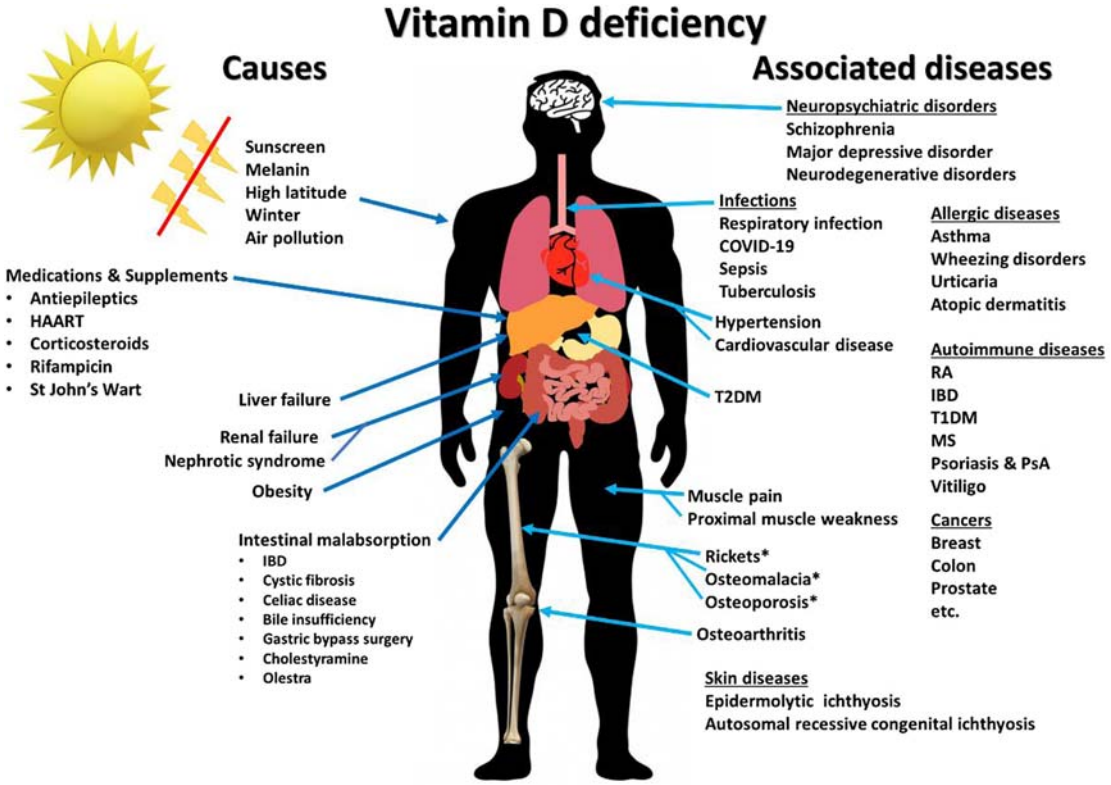


Fig. 6 Summary of causes of vitamin D deficiency and diseases and disorders associated with vitamin D deficiency. Abbreviation: HARRT: highly active antiretroviral therapy; IBD: inflammatory bowel diseases; MS: multiple sclerosis; PsA: psoriatic arthritis; T1DM: type 1 diabetes mellitus; T2DM: type 2 diabetes mellitus; RA: rheumatoid arthritis. Copyright Holick (2020) reproduced with permission. "*" denotes diseases that are direct consequences of vitamin D deficiency.

1000 IU of vitamin D or vitamin D supplementation as recommended by the Endocrine Society drink a quart of milk that contains 400 IU of vitamin D, and is exposed to sunlight. The IOM recognized that vitamin D is not as toxic as once believed and raised the upper limit (UL) for older children and adults up to 4000 IU day^{-1}. Based on the evidence-based literature, these recommendations are conservative. Several studies have reported that healthy adults ingesting up to 10,000 IUs daily for at least 6 months did not have any toxicity and maintained serum concentrations of 25(OH)D less than 100 ng mL^{-1} (Shirvani et al., 2019). A Canadian study reported healthy adults in Canada ingesting up to 20,000 IUs daily demonstrated no toxicity (Ekwaru et al., 2014). Based on all of the evidence the Endocrine Society Practice Guidelines on Vitamin D recommended a UL of 4000 and 10,000 IU day^{-1} for children and adults respectively (**Table 2**) (Holick et al., 2011).

Conclusion

Vitamin D deficiency continues to be a major health problem worldwide (**Fig. 7**) that is due to a variety of causes (**Fig. 6**). Sunlight exposure was the major source of vitamin D for our hunter gatherer forefathers since very few foods naturally contained vitamin D. A study of the Maasai and Hadzebe who are hunter gatherers revealed that their mean blood concentration of 25(OH)D was 46 ng mL $^{-1}$ (115 nmol L^{-1}) (Luxwolda et al., 2013). In order for an adult to achieve this level would require ingesting 4000–5000 IUs daily (Ekwaru et al., 2014). Studies have shown that healthy adults who have a starting blood concentration of 25(OH)D of between 15 and 20 ng mL^{-1} require 100 IUs to raise the blood level by approximately 0.6–1 ng mL^{-1}. A multitude of studies have reported that maintaining a blood concentration of 25(OH)D of at least 30 ng mL^{-1} and ideally as recommended by the Endocrine Society between 40 and 60 ng mL^{-1} reduces risk of preeclampsia, premature births, autoimmune disorders, cardiovascular disease, neurocognitive dysfunction including Alzheimer's disease, type 2 diabetes and infectious diseases (Holick, 2010; Holick et al., 2011; Hossein-nezhad and Holick, 2013; Płudowski et al., 2013; Charoenngam and Holick, 2020). In the era of COVID 19 several studies have reported significant reductions in infectivity, morbidity and mortality associated with this highly infectious and deadly viral infection (Charoenngam and Holick, 2020; Kaufman et al., 2020, Charoenngam et al., 2021; Seal et al., 2022). It is unrealistic for most of the world's population to obtain an adequate amount of vitamin D from dietary sources and sensible sun exposure. Although in the lay press it is suggested that 5–15 min of sun exposure will produce enough vitamin D, this is highly unlikely unless you are fair skinned in a bathing suit and exposed to sunlight at noontime near the equator. Time of

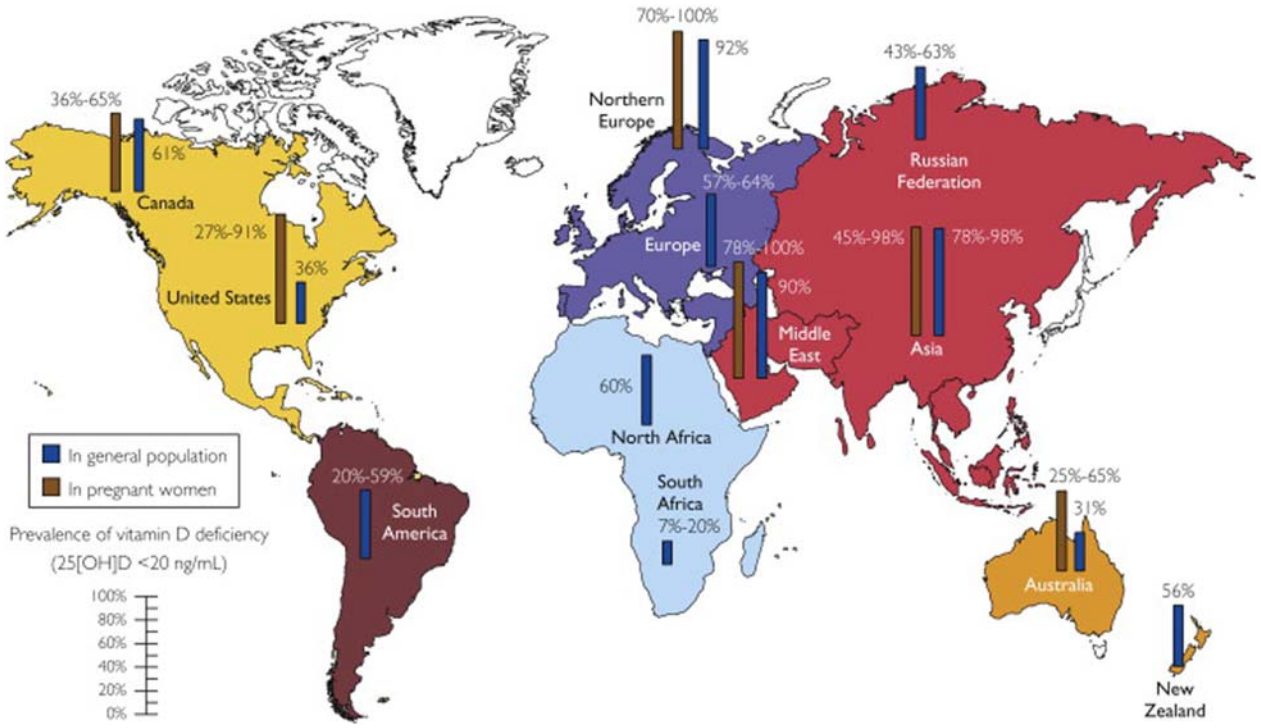


Fig. 7 Reported incidence of vitamin D deficiency defined as a 25-hydroxyvitamin D (25[OH]D) level below 20 ng mL^{-1} around the globe in pregnant women and the general population. To convert 25(OH)D values to nmol L^{-1}, multiply by 2.496. Copyright Holick (2013), reproduced with permission.

day, season, latitude, altitude, weather conditions, skin pigmentation and sunscreen use all influence the skin's production of vitamin D during sun exposure (Wacker and Holick, 2013). Therefore, institution of food fortification programs as has been done for almost 100 years in the United States and Canada and more recently in India, Sweden and Finland should be encouraged. To guarantee vitamin D sufficiency with all of its potential health benefits following the Endocrine Society Guidelines on Vitamin D is prudent.

See Also: Calcium; Lactation: Dietary requirements; Pregnancy: Nutrient requirements

References

Aksnes, L., Aarskog, D., 1982. Plasma concentrations of vitamin D metabolites in puberty: per morning and Effect of sexual maturation and implications for growth. J. Clin. Endocrinol. Metab. 55, 94–101.

Apperly, F.L., 1941. The relation of solar radiation to cancer mortality in North America. Cancer Res. 1, 191–195.

Bouillon, R., Marcocci, C., Carmeliet, G., et al., 2019. Skeletal and extraskeletal actions of vitamin D: current evidence and outstanding questions. Endocr. Rev. 40 (4), 1109e1151.

Carlberg, C., Haq, A., 2018. The concept of the personal vitamin D response index. J. Steroid Biochem. Mol. Biol. 175, 12–17. https://doi.org/10.1016/j.jsbmb.2016.12.011.

Charoenngam, N., Holick, M.F., 2020. Immunologic effects of vitamin D on human health and disease. Nutrients 12, 2097.

Charoenngam, N., Shirvani, A., Holick, M.F., March 2021. Vitamin D and its potential benefit for the COVID-19 pandemic. Endocr. Pract. 17. PMID: 33744444.

DeLuca, H., 1988. The vitamin D story: a collaborative effort of basic science and clinical medicine. Federal Proc. Am. Soc. Exp. Biol. 2, 224–236.

Dobnig, H., Pilz, S., Scharnagl, H., et al., 2008. Independent association of low serum 25-hydroxyvitamin D and 1,25-dihyroxyvitamin D levels with all-cause and cardiovascular mortality. Arch. Intern. Med. 168 (12), 1340–1349.

Dong, Y., Stallmann-Jorgenson, I.S., Pollock, N.K., et al., 2010. A 16-week randomized clinical trial of 2,000 IU daily vitamin D_3 supplementation in black youth: 25-hydroxyvitamin D, adiposity, and arterial stiffness. J. Clin. Endocrinol. Metab. 95 (10), 4584–4591.

Dudenkov, D.V., Mara, K.C., Petterson, T.M., Maxson, J.A., Thacher, T.D., 2018. Serum 25-hydroxyvitamin D values and risk of all-cause and cause-specific mortality: a population-based cohort study. Mayo Clin. Proc. 93, 721–730. https://doi.org/10.1016/j.mayocp.2018.03.006.

Ekwaru, J.P., Zwicker, J.D., Holick, M.F., Giovannucci, E., Veugelers, P.J., 2014. The importance of body weight for the dose response relationship of oral vitamin D supplementation and serum 25-hydroxyvitamin D in healthy volunteers. PLoS One. https://doi.org/10.1371/journal.pone.0111265.

Grant, W.B., 2002. An estimate of premature cancer mortality in the US due to inadequate doses of solar ultraviolet-B radiation. Cancer 70, 2861–2869.

Grober, U., Spitz, J., Reichrath, J., Kisters, K., Holick, M.F., 2013. Vitamin D: update 2013: from rickets prophylaxis to general preventive healthcare. Derm. Endocrinol. e2-331–e2-347.

Holick, M.F., 2006. Resurrection of vitamin D deficiency and rickets. J. Clin. Investig. 116 (8), 2062–2072.

Holick, M.F., 2007. Vitamin D deficiency. N. Engl. J. Med. 357, 266–281.

Holick, M.F., 2010. The Vitamin D Solution. Hudson Street Press, New York.

Holick, M.F., Binkley, N.C., Bischoff-Ferrari, H.A., Gordon, C.M., Hanley, D.A., Heaney, R.P., et al., 2011. Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline. J. Clin. Endocrinol. Metab. 96 (7), 1911–1930.
Hollis, B.W., Wagner, C.L., 2004. Vitamin D requirements during lactation: high-dose maternal supplementation as therapy to prevent hypovitaminosis D for both the mother and the nursing infant. Am. J. Clin. Nutr. 80, 1752S–1758S.
Hossein-nezhad, A., Holick, M.F., 2013. Vitamin D for health: a global perspective. Mayo Clin. Proc. 88 (7), 720–755.
Hypponen, E., Laara, E., Jarvelin, M.-R., Virtanen, S.M., 2001. Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study. Lancet 358, 1500–1503.
Kaufman, H.W., Niles, J.K., Kroll, M.H., Bi, C., Holick, M.F., 2020. SARS-CoV-2 positivity rates associated with circulating 25-hydroxyvitamin D levels. PLoS One 15 (9), e0239252. http://refhub.elsevier.com/S1530-891X(21)00087-2/sref92.
Kumar, J., Muntner, P., Kaskel, F.J., Hailpern, S.M., 2009. Prevalence and Associations of 25-hydroxyvitamin D deficiency in US children: NHANES 2001–2004. Pediatrics 124 (3).
Lee, J.M., Smith, J.R., Philipp, B.L., Chen, T.C., Mathieu, J., Holick, M.F., 2007. Vitamin D deficiency in a healthy group of mothers and newborn infants. Clin. Pediatr. 46, 42–44.
Luxwolda, M.F., Kuipers, R.S., Kema, I.P., van der Veer, E., Dijck-Brouwer, D.A.J., Muskiet, F.A.J., 2013. Vitamin D status indicators in indigenous populations in East Africa. Eur. J. Nutr. 52, 1115–1125. https://doi.org/10.1007/s00394-012-0421-6.
Melamed, M.L., Muntner, P., Michos, E.D., et al., 2008. Serum 25-hydroxyvitamin D levels and the prevalence of peripheral arterial disease: results from NHANES 2001 to 2004. Arterioscler. Thromb. Vasc. Biol. 28 (6), 1179–1185.
Moan, J., Porojnicu, A.C., Dahlback, A., Setlow, R.B., 2008. Addressing the health benefits and risks, involving vitamin D or skin cancer, of increased sun exposure. Proc. Natl. Acad. Sci. U. S. A 105 (2), 668–673.
Płudowski, P., Karczmarewicz, E., Bayer, M., Carter, G., Chlebna-Sokół, D., Czech-Kowalska, J., Dębski, R., Decsi, T., Dobrzańska, A., Franek, E., Głuszko, P., Grant, W.P., Holick, M.F., Yankovskaya, L., Konstantynowicz, J., Książyk, J.B., Księżopolska-Orłowska, K., Lewiński, A., Litwin, M., Lohner, S., Lorenc, R.S., Łukaszkiewicz, J., Marcinowska-Suchowierska, E., Milewicz, A., Misiorowski, W., Nowicki, M., Povoroznyuk, V., Rozentryt, P., Rudenka, E., Shoenfeld, Y., Socha, P., Solnica, B., Szalecki, M., Tałałaj, M., Varbiro, S., Żmijewski, M.A., 2013. Practical guidelines for the supplementation of vitamin D and the treatment of deficits in Central Europe – recommended vitamin D intakes in the general population and groups at risk of vitamin D deficiency. Endokrynol. Pol. 64 (4), 319–327.
Priemel, M., von Domarus, C., Klatte, T.O., et al., 2010. Boner mineralization defects and vitamin D deficiency: Histomorphometric analysis of Iliac crest bone biopsies and circulating 25-hydroxyvitamin D in 675 patients. J. Bone Miner. Res. 25 (2), 305–312.
Ross, A.C., Manson, J.E., Abrams, S.A., Aloia, J.F., Brannon, P.M., Clinton, S.K., et al., 2011. The 2011 report on dietary reference intakes for calcium and vitamin D from the Institute of Medicine: what clinicians need to know. J. Clin. Endocrinol. Metab. 96 (1), 53–58.
Seal, K.H., Bertenthal, D., Carey, E., Grunfeld, C., Bikle, D., Lu, C.M., 2022. Association of Vitamin D status and COVID-19-related hospitalization and mortality. J. Intern. Med. https://doi.org/10.1007/s11606-021-07170-0.
Shirvani, A., Kalajian, T.A., Song, A., Holick, M.F., 2019. Disassociation of vitamin D's calcemic activity and non-calcemic genomic activity and individual responsiveness: a randomized controlled double-blind clinical trial. Sci. Rep. 9, 17685. https://doi.org/10.1038/s41598-019-53864-1.
Urashima, M., Segawa, T., Okazaki, M., Kurihara, M., Wada, Y., Ida, H., 2010. Randomized trial of vitamin D supplementation to prevent seasonal influenza A in schoolchildren. Am. J. Clin. Nutr. 91, 1255–1260.
Wacker, M., Holick, M.F., 2013. Sunlight and vitamin D: a global perspective for health. Dermatoendocrinol 5, 51–108. https://doi.org/10.4161/derm.24494.

Further reading

IOM Institute of Medicine, 2011. Dietary Reference Intakes for Calcium and Vitamin D. The National Academies Press, Washington DC.
Malabanan, A., Veronikis, I.E., Holick, M.F., 1998. Redefining vitamin D insufficiency. Lancet 351, 805–806.

Suggested Reading

Holick, M.F., 2010. The Vitamin D Solution. Hudson Street Press, New York.

Vitamin D: Role in chronic and acute diseases

Ana I. Álvarez-Mercado[a,b], **María Dolores Mesa**[a,b], **and Ángel Gil**[a,b,c], [a] Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology "José Mataix", Center of Biomedical Research, University of Granada, Granada, Spain; [b] Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain; and [c] CIBEROBN (Physiopathology of Obesity and Nutrition Network CB12/03/30038), Institute of Health Carlos III (ISCIII), Madrid, Spain

© 2023 Elsevier Ltd. All rights reserved.

Introduction 535
Vitamin D and chronic diseases 536
Obesity 536
Metabolic syndrome 537
Type 2 diabetes mellitus 538
Cancer 538
Rheumatoid arthritis 538
Osteoporosis 539
Chronic obstructive pulmonary disease 539
Inflammatory bowel disease 539
Vitamin D and acute diseases 540
Vitamin D and the immune system 540
Vitamin D and respiratory bacterial infections 541
Vitamin D and virus infections 541
Vitamin D and the human immunodeficiency virus infection 541
Vitamin D and critically ill patients 542
Vitamin D and SARS-CoV-2 infection 543
Conclusions 543
References 543
Further reading 544

Key points

- To know the role of vitamin D in the prevention and treatment of non-communicable chronic diseases.
- To understand how vitamin D deficiency is related to obesity, insulin resistance and metabolic syndrome, type 2 diabetes, cancer, rheumatoid arthritis, osteoporosis, chronic pulmonary obstructive disease, and inflammatory bowel disease.
- To know the function of vitamin D in the prevention and treatment of acute infectious diseases of bacterial and viral origins, namely respiratory infections.
- To understand the role of vitamin D in the prevention of HIV and SARS-Cov2 infections
- To know how vitamin D status affects the course of critical illness diseases

Introduction

Vitamin D is involved in bone health by promoting calcium absorption in the gut and maintaining serum calcium and phosphate concentrations, and by its action on bone growth and reorganization through osteoblasts and osteoclasts cells. Moreover, during the last three decades, novel actions of vitamin D have been discovered. Indeed, active vitamin D also regulates cell proliferation and differentiation and has a key role in the responses of the immune and nervous systems. Current effects of vitamin D include xenobiotic detoxification, oxidative stress reduction, neuroprotective functions, antimicrobial defense, immunoregulation, *anti*-inflammatory/anticancer actions, and cardiovascular benefits (Gil et al., 2018).

Several systematic reviews and meta-analyses have shown links between serum vitamin D levels and non-communicable diseases. In line with these findings, a wide number of studies have suggested associations of vitamin D deficiency (VDD) with respiratory tract infections, osteoporosis, and other chronic and metabolic diseases such as obesity, metabolic syndrome, type 2 diabetes mellitus (T2DM), cancer, rheumatoid arthritis (RA), and inflammatory bowel disease (IBD). **Fig. 1** shows a summary of the benefits of vitamin D supplementation on main non-communicable chronic diseases.

Hence, clinical trials addressed to evaluate the efficacy of administration of vitamin D and its metabolites for the treatment of both chronic and acute diseases are of great interest, although variable outcomes are being reported. In this concern, evidence shows

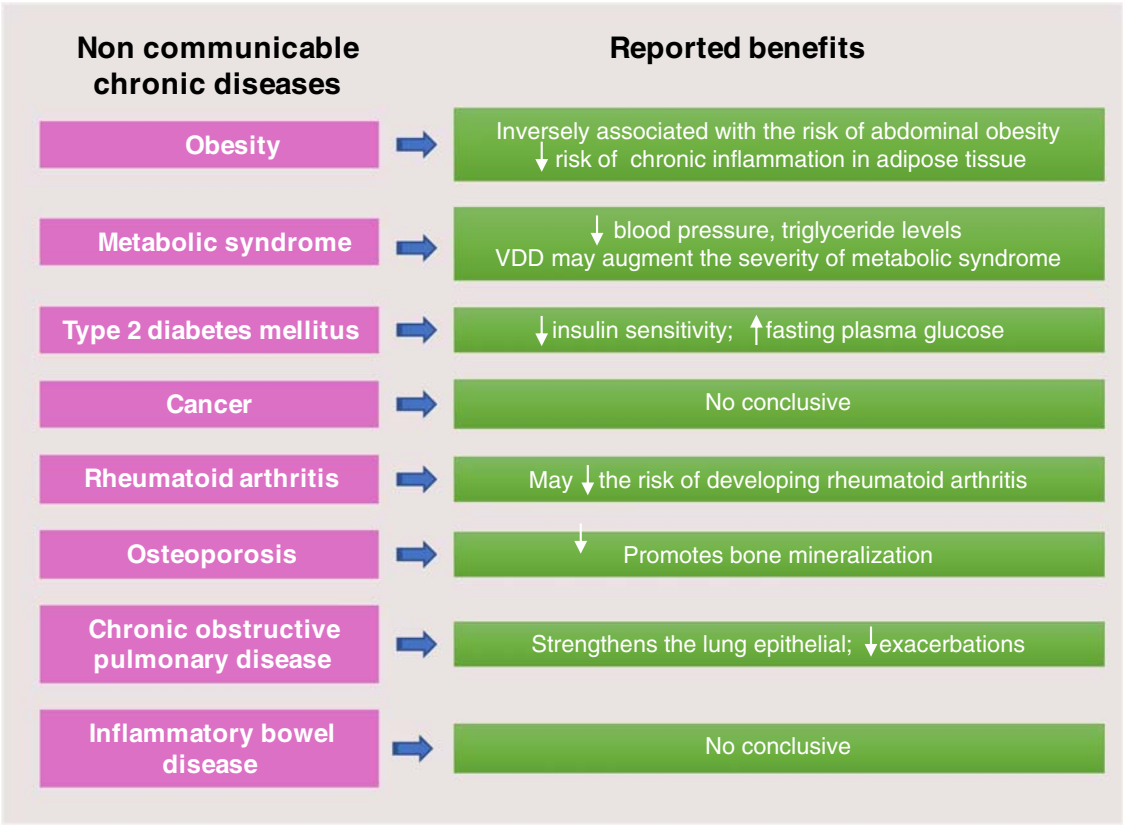


Fig. 1 Summary of reported benefits of vitamin D in non-communicable chronic diseases.

a notorious inter-individual difference in gene expression in human peripheral blood mononuclear cells in response to vitamin D supplementation, which suggests that some individuals might present a higher grade of benefits from vitamin D supplementation than others (Charoenngam, 2021) (**Table 1**).

Vitamin D also exerts important actions in the clinical course of infectious and other acute diseases, particularly respiratory bacterial infections, tuberculosis, and virus infections, e.g., those generated by human immunodeficiency and SARS-CoV-2 (COVID-19) viruses.

In humans, the recommended daily dietary allowance of vitamin D is 400–800 IU depending on age and sex. The circulating 25-hydroxy-vitamin D [25(OH)D] is the most used biomarker of VDD recommended by clinical guidelines and 1,25 dihydroxy-vitamin D [1,25 $(OH)_2D$], named calcitriol, the active form-of vitamin D. Other biomarkers and immune assay methods are also being evaluated and compared, such as bioavailable and free 25(OH)D, 24,25 dihydroxy-vitamin D [24,25$(OH)_2D$], other vitamin D metabolites, vitamin D binding protein or parathyroid hormone (Ganmaa et al., 2021).

According to the guidelines of the American Endocrine Society, serum levels of 25(OH)D below 20 ng/mL (50 nmol/L) are considered as VDD, while 25(OH)D serum levels between 21 and 29 ng/mL (52.5–72.5 nmol/L) are defined as vitamin D insufficiency (Holick et al., 2011). To maintain adequate levels in the preferred range of 40–60 ng/mL (100–150 nmol/L) (Ganmaa et al., 2021; Charoenngam, 2021) and thus avoid the risk of VDD, it is advisable to increase the intake of vitamin D and have adequate exposure to sunlight. However, it remains controversial what is the optimal serum level of 25(OH)D. For this, some professional societies recommend higher vitamin D intakes, and, in consequence, physicians sometimes prescribe more than 4000 IU to compensate for VDD. This can be biologically explained by: (1) the fact that the vitamin D receptor is expressed in the majority of human tissues; (2) vitamin D levels in northern latitudes are far lower than the hominids evolved in the equatorial Africa area and (3) the stimulation of vitamin D receptor by calcitriol alter the expression of over two hundred genes to support a large range of physiological responses with the potential to protect against the development of several pathologies (Ganmaa et al., 2021).

Vitamin D and chronic diseases

Obesity

Evidence shows that obesity negatively regulates circulating vitamin D levels. In fact, obesity decreases the detectable serum levels of 25(OH)D through the sequestration of vitamin D in body fat tissue (since vitamin D is a fat-soluble vitamin) or reduce skin synthesis of vitamin D because of limited outdoor activity and sun exposure (Al Anouti et al., 2020).

Table 1 Doses of vitamin D recommended or administered in interventions studies with reported benefits in chronic diseases.

Disease	*Dose*
Obesity/MetS	Obese adults and the elderly obese adults and the elderly (BMI 30+ kg/m^2): (BMI 30+ kg/m^2):
T2DM	1250–1500 μg/week (prevention)
Cancer	400 IUs or 2000 IU/day (reduced cancer mortality but not cancer incidence)
Rheumatoid arthritis	25–75,000 IU (decrease in fatigue severity and in pain)
Osteoporosis	2000 IU/day
COPD	Keeping the <25 nmol/L but not higher levels reduced the rate of moderate/severe exacerbations
IBD	5000–10,000 IU/day

BMI, body mass index; COPD, chronic obstructive pulmonary disease; IBD, inflammatory bowel disease; IU, international units; MetS, metabolic syndrome; T2DM, type 2 diabetes *mellitus*.

Moreover, alternations in vitamin D metabolism in obese subjects manifesting low serum levels of 25(OH)D is well-recognized, while weight reduction with loss of adipose tissue is associated with improvement in circulating 25(OH)D. Additionally, patients with obesity and intestinal malabsorption require a two to three times higher intake of vitamin D to maintain the same serum 25(OH)D concentrations (Charoenngam, 2021). It has been also demonstrated that abdominal obesity is more prevalent in those individuals with lower serum vitamin D levels. By contrast, results from observational studies aimed to investigate the relationship between VDD and the risk of central obesity are inconsistent. However, a meta-analysis of epidemiologic studies revealed that serum vitamin D level was inversely associated, in a dose-response manner, with the risk of abdominal obesity, particularly in adults. Furthermore, vitamin D reduced the risk of chronic disease and chronic inflammation in adipose tissue. In addition, dose-response analysis showed that every 25 nmol/L increments in serum vitamin D were related to an 8% reduced risk of abdominal obesity, 10% decreased central adiposity risk in representative populations, and 13% lower risk of metabolic syndrome.

The entire mechanisms throughout vitamin D affect the lipid profile is still undeciphered, even though observational and interventional studies report conflicting evidence. Moreover, it has been suggested that the association between vitamin D and metabolic disorders may be confounded by obesity rather than being a causal relationship. Accordingly, usual chronic inflammatory processes in obese patients might decrease 25(OH)D levels and, at the same time, affect several metabolic parameters (Al Anouti et al., 2020).

Metabolic syndrome

Metabolic syndrome is known as one of the most important risk factors of T2DM and cardiovascular disease and can increase the risk of myocardial infarction and stroke two-fold. According to the National Cholesterol Education Program's Adult Treatment Panel III (NCEP-ATP III) criteria of metabolic syndrome has six main components, namely obesity, dyslipidemia, high blood pressure, insulin resistance or glucose intolerance, pro-inflammatory state, and prothrombotic state (Ganmaa et al., 2021). Although this disease is a well-known and serious public health burden, the metabolic syndrome does not have any direct and ultimate treatment due to its multifactorial nature. In consequence, clinical management to reduce the risk factors is the most common intervention.

Results from randomized control trials (RCTs) suggest that vitamin D supplementation positively impacts blood pressure and abdominal obesity. However, most trials have revealed that vitamin D supplementation has no effect in reducing myocardial cardiovascular events, heart attack, death from cardiovascular disease, or in the treatment of chronic heart failure. In this concern, 1-year supplementation with 4000 IU/day vitamin D3 did not affect cardiovascular disease, lipid profiles, or C-reactive protein levels in T2DM patients, but reduced triglyceride levels. Another trial reported that supplementation with 150,000 IU bolus of vitamin D every 3 months was unable to alter inflammatory markers and lipids in adults with metabolic syndrome. In addition, a Cochrane review including 159 RCTs suggested that vitamin D supplementation may reduce all-cause mortality compared with placebo or no intervention; by contrast, vitamin D supplementation had no significant effect on mortality related to cardiovascular events (Ganmaa et al., 2021).

Despite the populations recruited for the study of the role of vitamin D in metabolic syndrome vary by sex, age, and country, data from observational studies highlight specific associations between vitamin D and individual components of this disease such as obesity, dyslipidemia and blood pressure, as well as with metabolic syndrome as a complete entity.

VDD or low levels of 25(OH)D is related to a higher risk of this disease. Additionally, suboptimal levels of vitamin D may augment the severity of the metabolic syndrome. Concentrations of 25(OH)D are lower in patients with metabolic syndrome than in those without it. The prevalence of metabolic syndrome is reduced by half if individuals have high 25(OH)D concentrations. Specifically, vitamin D might modulate the atherogenic components of metabolic syndrome (Al Anouti et al., 2020).

Considering that inflammation plays a key role in the development of the metabolic syndrome, and under the premise that the *anti*-inflammatory effect of vitamin D would decrease the risk of this disease, a dose-response analysis reported that an increase of 25 nmol/L in plasma 25(OH)D was associated with a 13% lower risk of metabolic syndrome.

Type 2 diabetes mellitus

T2DM is a worldwide disease that grows in parallel with other diseases affecting multi-body systems. Thus, it is mandatory to develop strategies to treat T2DM effectively, maintaining glucose homeostasis to avoid complications such as diabetic nephropathy, peripheral neuropathy, and retinopathy.

Vitamin D has *anti*-inflammatory action, inhibiting cytokine production, which has an important role in suppressing the chronic low-grade inflammation present in T2DM. Vitamin D regulates insulin secretion by binding to vitamin D receptors present on pancreatic beta cells. Vitamin D increases insulin sensitivity by upregulating the expression of insulin receptors and binding to the vitamin D response element present in the human insulin receptor gene promoter. It also affects fatty acid metabolism in insulin-responsive tissues through the activation of its transcription factor. Hence, the role of vitamin D in glucose metabolism and fuel homeostasis is supported by several observational studies revealing an inverse relationship seen between vitamin D and T2DM.

Although not many studies address the relationship between insulin sensitivity and vitamin D, one study reported that serum 25(OH)D concentration is responsible for 21.2% of the variation of the insulin sensitivity index. In addition, serum 25(OH)D concentration accounted for 8.2% of the variation of beta-cell function.

Data from meta-analyses of RCTs showed favorable effects of vitamin D intervention in T2DM non-obese individuals [BMI < 30] but not in those with BMI ≥ 30. Furthermore, supplementation with vitamin D in pre-diabetic subjects prevented progression to T2DM, improved insulin sensitivity, decreased insulin resistance and systemic inflammation, and lowered fasting plasma glucose (Ganmaa et al., 2021).

Since vitamin D affects different organs and tissues in patients with T2DM a systematic review was conducted to evaluate the effect of vitamin D supplementation in glycemic homeostasis and its impact on the T2DM patients. This study was focused on how this vitamin influences effectively, maintaining glucose homeostasis to avoid its complications. The authors found an inverse relationship between vitamin D levels and neuropathy and diabetic retinopathy.

Cancer

Biological functions of vitamin D include modulation of the immune system and *anti*-carcinogenic effects. The association of VDD with cancer (along with other effects), has been described because of its potential effect on cell differentiation and the suppression of cell proliferation. Consequently, recent studies have assessed the association between serum vitamin D and the risk of some types of cancer.

In a meta-analysis of RCTs, the authors reported that vitamin D supplementation significantly reduced total cancer mortality but did not reduce total cancer incidence (Ganmaa et al., 2021). On the other hand, a dose-response meta-analysis demonstrated that each five nmol/L increase in blood vitamin D levels was associated with a 6% decrease in the risk of breast cancer, while a 400 IU/day increase in vitamin D intake was not significantly correlated to this type of cancer. By contrast, other dose-response meta-analyses found that both serum vitamin D and vitamin D intake were inversely related to colorectal cancer. Further, another meta-analysis of RCTs showed a protective effect of vitamin D supplementation on cancer incidence and mortality. Additionally, a significant reduction in metastatic or fatal cancers has been reported in men $\leq$50 years and women $\leq$55 years (Ganmaa et al., 2021).

Given these controversial results, to date, neither 1,25(OH)2D nor its analogs have ever been successfully developed as a strategy to treat or prevent any type of cancer.

Rheumatoid arthritis

RA is a chronic autoimmune condition resulting in synovial inflammation around joints, progressively leading to cartilage and bone destruction. Results from multiple observational studies have evidenced a link between a low level of serum 25(OH)D and the presence and/or severity of several rheumatic diseases (Charoenngam, 2021). Indeed, vitamin D is believed to play a role in modulating RA's pathogenesis and disease activity, based on the actions of 1,25(OH)2D on the adaptive immune response that suppresses the proliferation and activity of T helper 1 cells (Th1) and Th17, and enhances the T regulatory cells (Treg) activity. Furthermore, genomic studies have shown that certain polymorphisms of the gene encoding vitamin D receptor and vitamin D binding protein are associated with susceptibility to RA (Charoenngam, 2021). An inverse correlation between circulating vitamin D levels and RA incidence and disease activity is also known.

Despite these promising results, unfortunately, evidence from clinical trials demonstrating the impact of any form of vitamin D supplementation on most rheumatic diseases has not been established yet. Moreover, it is still unclear whether the association between vitamin D and these conditions are causal or more likely explained by confounders and reverse causation, such as limited physical activity or corticosteroid administration (Charoenngam, 2021).

Concerning studies about dose-response, it is recommendable that patients with rheumatic diseases should maintain a serum 25(OH)D level of at least 30 ng/mL (75 nmol/L) to prevent osteomalacia, secondary osteoporosis and fracture, and possibly 40–60 ng/mL (100–150 nmol/L) to reach the higher benefit of vitamin D (Charoenngam, 2021). In treatment-naive RA patients, a significant negative association was observed between vitamin D levels and disease activity parameters. Notably, the association of vitamin D with the incidence and severity of RA is very well supported by evidence. However, due to heterogeneity in dosages and durations of supplementation, robust evidence supporting vitamin D supplementation in ameliorating clinical outcomes is still needed. In sum, observational studies suggest that increasing vitamin D intake to raise serum 25(OH)D may reduce the risk of

developing RA. However, there is no demonstration from a reliable clinical trial that vitamin D supplementation can reduce the risk of RA. In addition, there is moderate evidence that vitamin D supplements or the oral administration of 1,25(OH)2D can mitigate RA severity (Charoenngam, 2021).

Osteoporosis

Vitamin D regulates the absorption of calcium and phosphorus and, thus, is universally accepted as an essential vitamin for bone strength and as a promotor of the immune system function; it has been reported to have an *anti*-inflammatory role and established benefits in osteoporosis and osteomalacia.

Accordingly, VDD is a well-recognized health problem and contributes to bone loss and calcium dysregulation, which causes or aggravates osteoporosis. 1,25(OH)2D regulates calcium and phosphate homeostasis by acting on the small intestine, kidneys, and bones. It passively promotes bone mineralization by inducing intestinal absorption of calcium and phosphate and renal tubular calcium reabsorption, which helps maintain adequate calcium-phosphate crystallizing in the collagen matrix.

However, multiple studies failed to demonstrate any benefit from vitamin D supplementation in patients with osteoporosis, and a systematic review and meta-analysis also were unable to confirm any beneficial effect on bone density or fracture prevention. Additionally, placebo-control RTCs revealed a threshold effect of vitamin D with no benefit observed on the subjects with baseline 25(OH)D level $\geq$75 nmol/L (30 ng/mL). Furthermore, possible detrimental effects on bone mineral density were observed in subjects who received a higher dose of vitamin D (250 μg or 10,000 IU daily) with a mean 25(OH)D of 200 nmol/L or 80 ng/mL.

On the other hand, a maternal vitamin D osteoporosis study reported that daily vitamin D supplementation in pregnant women did not increase offspring whole-body bone mineral content above that of the placebo group (Ganmaa et al., 2021).

Chronic obstructive pulmonary disease

Chronic obstructive pulmonary disease (COPD) is related to high mortality and morbidity worldwide, and its exacerbations cause significant morbidity, mortality, impaired quality of life, and costs (Lokesh et al., 2021). The main pathogenesis associated with COPD development is inflammation, oxidative stress, protease–antiprotease imbalance, and lung rebuilding (Ganmaa et al., 2021).

Having this into account, supplementation with vitamin D is particularly interesting due to its various effects on lungs, tissue remodeling, reduction of pro-inflammatory cytokines, and beneficial modulation of both innate and adaptive immune systems (Lokesh et al., 2021). In fact, lower vitamin D levels have been related to the regulation of typical characters associated with COPD (i.e., higher expression of proteases, modulation of inflammation and extracellular matrix turnover, and increased oxidative stress), their pathogenesis and severity. Some evidence suggests that vitamin D strengthens the lung epithelial. Accordingly, a recent meta-analysis of RCTs suggested that vitamin D supplementation may safely and substantially reduce the rate of moderate/severe COPD exacerbations in patients with baseline 25(OH)D levels $<$25 nmol/L, improve lung function and acute exacerbation. However, barrier studies focused on investigating a potential causal effect remain limited, and randomized trials report conflicting results for vitamin D supplementation and prevention (Ganmaa et al., 2021) Moreover, most of the studies focused on studying the relationship of vitamin D with COPD, its severity and exacerbations were performed in areas with sub-optimal sunlight and lack of sunlight throughout the year, or in cities where most residents spend time indoors (i.e., homes or offices). In consequence, reported data aimed to relate the association of vitamin D with the severity of COPD and its exacerbation in populations that are both adequately exposed to sunlight or keep higher levels of physical activity through the life course are quite limited (Lokesh et al., 2021).

On the other hand, it is deeply needed to perform longitudinal studies to evaluate whether young healthy subjects exposed to risk factors such as smoking or biomass fuel with low levels of vitamin D would present a higher risk of developing COPD, as well as whether vitamin D supplementation in these subjects would have the ability to prevent or delay the development of this disease (Lokesh et al., 2021).

Inflammatory bowel disease

IBD is a chronic inflammatory condition of the gastrointestinal tract that includes ulcerative colitis and Crohn's disease (Myint et al., 2020). It has been observed that samples from colonic biopsies of patients with IBD present a decreased concentration of vitamin D receptors and have a higher level of pro-inflammatory cytokines such as TNF-α and IL-1β. The most important IBD pharmacological strategy is to reduce inflammation. Due to the immunomodulatory properties that possess vitamin D, it is considered a very appropriate candidate for modifying risk factors and preventing the development of IBD, as well as ameliorating its severity. In fact, vitamin D protects the intestinal mucosa by increasing the expression of the proteins responsible for tight junction creation and affects epithelial integrity by the inhibition of the intestine epithelial apoptosis, which contributes to intestinal inflammation and the healing process.

However, although epidemiological and clinical observational studies have demonstrated an association between vitamin D and IBD, it is necessary to identify whether VDD is a consequence or a cause of IBD (Myint et al., 2020). Despite being limited by issues with trial design such as small sample size, uncertain use of control groups, inconsistent definitions of VDD, and/or absence of clinical outcomes, vitamin D supplementation trials suggest a variable association between VDD and IBD activity (Myint et al., 2020). However, the mentioned inconsistencies indicate that data from trials to evaluate dosing strategies for treatment are not sufficiently reliable.

Vitamin D and acute diseases

As mentioned previously, calcitriol is involved in the regulation of cell proliferation and differentiation and has a key role in the inflammatory and immune system response. This regulation confers an important role in the defense against bacterial and virus infections, which is mediated by its capacity to activate the innate defense response and to exert an *anti*-inflammatory action on the adaptive response, resulting in an overall immunotolerance effect (Gil et al., 2018).

The European Food Safety Authority (EFSA) considers vitamin D, within other vitamins and minerals, to be essential for the normal growth and functioning of the immune system. Impaired nutritional status or deficiencies of this vitamin is associated with increased risk and severity of many different types of infections. However, the immunomodulatory role of vitamin D against infections is complex and varies according to the nature of the pathogen. Moreover, the effectiveness of vitamin D is uncertain and appears to be highly dependent on the genetic polymorphisms of its receptor and epigenetic modifications. Fig. 2 shows a summary of the association of vitamin D with the main acute diseases.

Vitamin D and the immune system

Innate immunity is the first defense against invading microorganisms, including bacteria, viruses, fungi and protozoa. These mechanisms recognize microorganisms or their products and trigger a cascade of events that will eliminate and/or destruct the invading agents through the release of cytokines and antimicrobial peptides. Adaptive immunity is the second defense mechanism that mediates an antigen-specific immune response through antigen presentation cells (APC), namely dendritic cells (DCs), and the antigen recognition cells, T and B lymphocytes. Their activation causes the production of various cytokines and antibodies and induces cell killing.

The roles of vitamin D in the regulation of immunity are particularly well known. Immune system cells express the vitamin D receptors (VDR) when activated by external stimuli, indicating that vitamin D is especially involved in regulating their defense mechanisms. The binding of vitamin D to its cellular VDR activates immunity surveillance signaling pathways that modulate gene expression of proteins, such as cathelicidins (LL37) and defensins, involved in inflammation, oxidation, cell proliferation and differentiation, in apoptosis, and the processes of autophagy and bacterial destruction of infected macrophages (Gil et al., 2018).

On the other hand, related to the antimicrobial defense, vitamin D can inhibit the Th1, Th2 and Th17 adaptive response directly on lymphocytes via VDR signaling, or indirectly through paracrine signaling on APC. Vitamin D decreases the maturation of DCs and their ability to present antigens and alters the profile of Th and Treg. Specifically, vitamin D inhibits Th1, Th17, and Th9 cells development and leads to immune tolerance, and suppresses T-cell-mediated inappropriate proinflammatory mediators production. On the other hand, vitamin D enhances differentiation and proliferation of Th2 and Treg cells, which stimulate the release of *anti*-inflammatory cytokines (IL-4, IL-5, and IL-10) to be a significant mechanism by which vitamin D could be advantageous in autoimmune diseases by promoting an *anti*-inflammatory vs. inflammatory response.

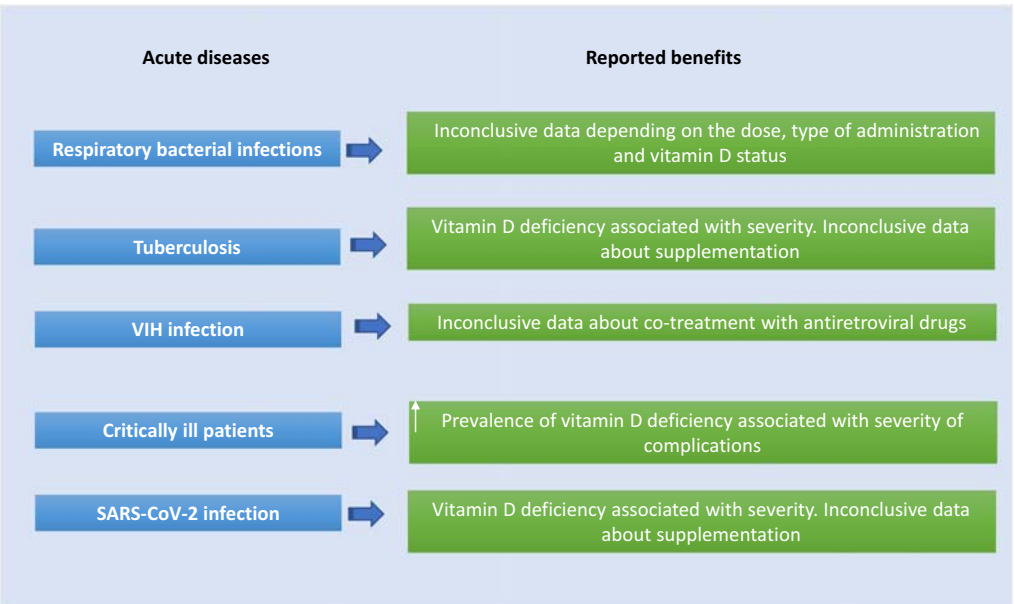


Fig. 2 Summary of reported benefits of vitamin D in acute diseases.

Vitamin D and respiratory bacterial infections

A large number of studies have associated a deficit of vitamin D blood levels and increased risk and severity of acute respiratory infections in children, adolescents and adults. Randomized controlled trials exploring the potential of vitamin D to prevent acute respiratory infections have yielded mixed results. Meta-analysis has the potential to identify factors that may explain this heterogeneity. Vitamin D supplementation is safe, and it protects against acute respiratory infections overall. Very deficient individuals and those receiving continuous supplements and not bolus doses are the ones who experienced the benefits. Incorporation of additional individual participant data from ongoing trials will be useful to increase the knowledge in these fields.

Tuberculosis is one of the most common and devastating infectious diseases worldwide. Several systematic reviews have found low vitamin D blood levels in patients with tuberculosis compared to healthy controls. These works have concluded that VDD may be a risk factor for tuberculosis disease progression. Current studies have shown that vitamin D plays a significant role in the host immune defense against *Mycobacterium tuberculosis*, but clinical trials reported inconsistent results. In 2021, a systematic review and meta-analysis have concluded an association between low levels of vitamin D and tuberculosis infections in adults and children. However, this analysis showed considerable heterogeneity of included studies, and therefore, the findings should be applied with caution (Kafle et al., 2021).

VDD can predict the risk of tuberculosis in a dose-dependent manner and is more likely a risk factor for tuberculosis than its consequence. But unfortunately, the beneficial effect of vitamin D supplementation for treating *M. tuberculosis* infection in vitamin D-deficient patients is not still clear, and it seems that several factors such as genetic and the mode of administration, among others, may influence clinical outcomes. Intervention clinical trials have reported that vitamin D supplementation has no clinical benefits on tuberculosis therapy. However, it reduced the time to sputum culture conversion in patients with tt genotype of the TaqI vitamin D receptor gene polymorphism and improved the multidrug-resistant tuberculosis sputum culture conversion. Another meta-analysis has concluded that vitamin D administration together with *anti*-tuberculosis treatment may be well tolerated and effective, improving sputum smear conversion rate and chest radiological appearance, while exhibiting an inflammation resolution effect. In addition, meta-analyses of vitamin D supplementation in the prevention of tuberculosis and other acute respiratory conditions support the efficacy of daily low dose supplementation but not after intermittent bolus dosing. Indeed, the optimal dose to achieve benefits remains unclear; anyhow, it is suggested to continue studies since vitamin D has shown a synergistic benefit on the immune system. Therefore, long-term prospective cohort studies in tuberculosis endemic countries should be conducted to understand better the causal relationship between VDD and this disease as well as its possible therapeutical application.

1,25D3 upregulates human cathelicidin from monocytes/macrophages infected with *M. tuberculosis*, resulting in autophagy, and upregulates nitric oxide synthase, suppressing mycobacterial growth. A meta-analysis has revealed a significant difference in vitamin D and cathelicidin LL-37 blood levels among tuberculosis and healthy subjects. It seems that active pulmonary tuberculosis infection is associated with hypovitaminosis D and elevated blood cathelicidin concentrations. In contrast, in local tissue lesions, cathelicidin LL-37 expression was lower in tuberculosis than in healthy subjects. Therefore, the mechanism involved in vitamin D-mediated immune regulation against *M. tuberculosis* needs to be further investigated.

Vitamin D and virus infections

Virus infections have also been associated with VDD. Children with hand-foot-and-mouth disease have very low vitamin D levels, associated with a poor prognosis. Patients with atopic dermatitis are susceptible to microbial infection due to the decreased production of cathelicidin and other antimicrobial peptides, and they can be improved by vitamin D supplementation. In these cases, the participation of vitamin D in the regulation of the secretion of antimicrobial peptides is also altered. This occurs in cathelicidins (LL37) and defensins, molecules that through different mechanisms prevent the entry of viruses and their activity, by interacting with the viral coatings and limiting its replication directly or indirectly by modulating the immune cells, migration, activation, proliferation and differentiation, specifically neutrophils, monocytes-macrophages, and T lymphocytes. These antimicrobial peptides are also able to regulate autophagy and apoptosis of infected cells, thus contributing to the reduction of viral load. Indeed, vitamin D supplementation in healthy subjects appears to confer some protection against virus respiratory tract infection in healthy adults from the USA and Canada but not in other world regions. However, these conclusions cannot be generalized for all types of virus infections, and their potential effect must be demonstrated. For example, associations between vitamin D and herpesviruses remain inconclusive and further studies are needed in the general population.

Indeed, vitamin D-modulated immune response against viral infection is mediated by its VDR, which acts as a transcription factor modulating the expression of genes triggering the response against viruses. To date, six major VDR polymorphisms (*Cdx, A1012G, FokI, BsmI, ApaI,* and *TaqI*) have been studied in the context of viral infection susceptibility. Reported studies show controversial results, probably due to statistical lack of power and population genetic differences. In this sense, *FokI* polymorphism is a relevant variant capturing the association of VDR polymorphisms with viral infections.

Vitamin D and the human immunodeficiency virus infection

Human immunodeficiency virus (HIV) infection is a heavy burden worldwide. Observational studies have reported a high prevalence of VDD among people living with HIV compared with the general population. Low 25(OH)D is common in diverse HIV-infected populations and is an independent risk factor for clinical and virologic failure. It has been associated with increased HIV mortality, although it may also be influenced by other factors such as older age, lower body mass index, lower latitude, male sex and antiretroviral treatments.

The disagreement regarding the effect of antiretroviral therapy drugs on vitamin D metabolism is still unresolved. The Prospective Evaluation of Antiretrovirals in Resource Limited Settings (PEARLS) study (ACTG5175) is a large-scale, randomized controlled trial in diverse populations in many different settings from four continents. This study investigated whether simplified antiretroviral regimens (once daily) were as effective as standard twice-daily regimens, which may make it more suitable for low-income settings, and decrease adverse effects. It was observed that baseline VDD was associated with diminished CD4 recovery after combined antiretroviral therapy initiation, and that impaired CD4 recovery may contribute to the poor clinical outcomes observed in individuals with VDD (Ewald et al., 2019). Prospective studies assessing the potential benefit of vitamin D supplementation among HIV patients that initiate a combined antiretroviral therapy are lacking and therefore warranted, to demonstrate if vitamin D recommendation policies should be part of routine clinical practice.

Vitamin D and critically ill patients

It has been suggested that vitamin D insufficiency is a risk factor in intensive care and plays an essential role in infectious, immunologic, neurologic, cardiovascular, and respiratory complications. VDD has been hypothesized not only to be common but also to represent a potentially modifiable risk factor for greater illness severity and outcomes during critical illness. Observational studies have demonstrated an association between VDD and increased risk of morbidity and mortality in critically ill patients. Cohort studies and pilot trials have suggested promising beneficial effects of vitamin D replacement in critical illness, at least in patients with severe VDD. However, the results are inconclusive.

Vitamin D insufficiency is reported in up to 77% of critically ill patients. It is associated with increased mortality, length of stay in an intensive care unit (ICU) and hospital, as well as with respiratory disorders in prolonged ventilated patients (Langlois et al., 2018). In critically ill children, the status and the effect of VDD on the outcome are still unclear. In 2017, a systematic review and meta-analysis stated that approximately 50% of critically ill children have VDD (blood total 25(OH)D concentration under 50 nmol/L) at the time of pediatric ICU (PICU) admission. VDD was also associated with greater illness severity, multiple organ dysfunction, and mortality in the PICU setting. Another meta-analysis has confirmed that 25(OH)D deficiency is prevalent in critically ill children at PICU admission and seems to be associated with higher cardiovascular sequential organ failure assessment and pediatric risk of mortality III scores, sepsis, length of hospital stays, and duration of mechanical ventilation.

However, clinical trials are not consistent in confirming the beneficial effects of improving vitamin D status on patient outcomes. Studies are needed to determine if improving low 25OHD levels (to a level of about 30 ng/mL) is associated with reduced mortality and can improve the functional discharge status of ICU patients. Some studies have stated that vitamin D administration might be associated with reducing mortality without significant adverse events in critically ill patients. However, other studies concluded that vitamin D administration did not improve clinical outcomes and was not significantly associated with reduced mortality or with the length of ICU stay, although statistical imprecision could be explained by the sparse number of trials (Langlois et al., 2018). Therefore, the causal association between VDD and worse outcomes of the critically ill needs further investigations, and large multicenter randomized trials are necessary to conclusively establish the potential beneficial effect of vitamin D supplementation. Indeed, vitamin D supplementation does not provide additional advantages over placebo for critically ill patients and is not associated with reduced all-cause mortality in critically ill patients. Therefore, large-scale prospective studies are needed to validate these findings.

VDD has been related to the risk of sepsis. Different meta-analyses have evaluated the relationship between serum 25(OH)D at admission and mortality risk and concluded that severe VDD might be independently associated with increased mortality in septic adult patients and in septic children compared to those without sepsis. Even though a high prevalence of VDD was found in sepsis, it was not associated with greater severity of illness or other clinical outcomes.

A burn is a severe form of injury associated with severe altered pathophysiological immune-inflammatory responses. In addition, burn patients are a group of critically ill patients in which vitamin D skin synthesis is compromised. Therefore, burn patients suffer common complications that overlap with those reported by patients with VDD. Burn patients with low vitamin D are more susceptible to higher complication rates, including sepsis, pneumonia, cardiovascular complications, and graft loss. However, the literature regarding vitamin D status and its influence on clinical outcomes remains insufficient in this type of patient.

Sufficient vitamin D concentrations and vitamin D supplementation may be of benefit in burn-injured patients since this supplementation is the only means to avoid vitamin D insufficiency in burn victims. However, as in other critically ill situations, the adequate dose, formulation, and route of administration remain unknown, and there is limited data on the impact of vitamin D status on clinical outcomes. Indeed, the high incidence of low serum D25 levels 1 year after major burn injury indicates prolonged compromise of vitamin D metabolism. Therefore, continued treatment with vitamin D3 beyond the acute phase postburn is recommended to improve the abnormal blood levels and associated co-morbidities in adults and children.

Another critical situation that a vitamin D deficient state may impair is patients with traumatic injuries, in which inflammation is a consequence of the trauma. For example, traumatic brain injury is the most frequent trauma and a leading cause of injury-related death and disability, in which inflammation processes influence severity and mortality incidences. VDD may induce impaired immune responses and increase the risk of infections in these patients, and therefore, vitamin D intervention has been proposed as a good tool for preventing these complications. To date, there is no actual data on the effectiveness of vitamin D for the improvement of immune function in traumatic brain injury patients. 25(OH)D level measured within 24 h after admission to the trauma

ICU was unrelated to clinical outcomes. However, patients with increased 25(OH)D levels after 7 days of hospitalization had better clinical outcomes than those with decreased levels.

Since vitamin D supplements are inexpensive and safe, their use could potentially improve clinical outcomes in all critical ill situations by reducing inflammation and infection-associated morbidity and mortality rates. However, more clinical interventional trials are mandatory to ascertain its effectiveness and the appropriateness of its prescription.

Vitamin D and SARS-CoV-2 infection

Special mention should be made of the effects of vitamin D against SARS-CoV2 infection. Vitamin D inadequacy may be involved in the mechanisms of SARS-CoV-2 infection and in potential risk factors for disease propagation or control of coronavirus disease 2019 (COVID-19). Several studies have been performed, but the inconsistency of results, due to heterogeneity of studies design and patients, do not allow to establish final conclusions. A Cochrane systematic review has concluded that at this moment, we cannot know whether vitamin D helps prevent death from COVID-19. However, it may reduce the need for assisted ventilation, although the evidence is still uncertain. This effect is most notable in vitamin D-deficient COVID-19 subjects (Stroehlein et al., 2021).

The discussion is open regarding the relationship between blood levels of calcitriol and SARS-CoV-2 infection severity. Most of the COVID-19 patients suffered from VDD or insufficiency. In addition, there is about a three times higher chance of getting infected with SARS-CoV-2 among vitamin D-deficient subjects, and about five times higher probability of developing the severe disease associated with VDD. Another meta-analysis has found strong evidence of low blood D3 as a predictor rather than just a side effect of this infection. Other meta-analysis has associated VDD with the risk of SARS-CoV-2 infection and with the severity of the disease but did not find any association with mortality rates.

Lower vitamin D levels have been related to key altered clinical and biochemical parameters during SARS-CoV-2 infection. In this type of critical patients, a vitamin D deficit has been recognized at admission and further deterioration after three days of stay. Given the different responses of the 25OHD3 and 25OHD2 forms, it would be useful to monitor them on the evolution of these critically ill patients. A meta-analysis has stated that relations between VDD and ICU admission, pulmonary complications, hospitalization, and inflammation were inconsistent and insufficient since although studies were heterogeneous in methodological and statistical approach, most of them showed a significant relation between 25(OH)D and SARS-CoV-2 infection, COVID-19 composite severity, and mortality. Regarding infection, caution should be taken for the interpretation of the results, due to intrinsic study limitations. Although the current findings indicate a potential role of vitamin D in improving COVID-19 severity in hospitalized patients, no significant difference with vitamin D supplementation on major health-related outcomes in COVID-19 patients has been found, and more robust data from randomized controlled trials are needed to verify its effects on mortality.

As it occurs with tuberculosis and other conditions, it is hypothesized that the administration mode may influence the outcomes, i.e., low daily doses may be more useful than intermittent bolus, and there is an urgent need for well-designed and sufficiently powered randomized controlled trials to address this matter.

Conclusions

Vitamin D plays a key role in the regulation and maintenance of innate immunity. A poor vitamin D status is related to noncommunicable chronic diseases, namely obesity, metabolic syndrome, T2DM, cancer, RA, COPD, and IBD. Also, a low vitamin D status is associated with an increased incidence of acute respiratory illnesses of both bacterial and viral origin. Likewise, low 25(OH)D levels are associated with a worse prognosis of patients with acute illness. Well-designed randomized intervention clinical trials addressed to evaluate the efficacy of administration of vitamin D and its metabolites for the treatment of both chronic and acute diseases are needed.

References

Al Anouti, F., et al., 2020. Effects of vitamin D supplementation on lipid profile in adults with the metabolic syndrome: a systematic review and meta-analysis of randomized controlled trials. Nutrients 12 (11), 3352.

Charoenngam, N., 2021. Vitamin D and rheumatic diseases: a review of clinical evidence. Int. J. Mol. Sci. 22 (19), 10659.

Ewald, E.R., et al., 2019. Effect of baseline micronutrient and inflammation status on CD4 recovery post-cART initiation in the multinational PEARLS trial. Clin. Nutr. 38 (3), 1303–1309.

Ganmaa, D., et al., 2021. Vitamin D, respiratory infections, and chronic disease: review of meta-analyses and randomized clinical trials. J. Intern. Med. 291 (2).

Gil, Á., Plaza-Diaz, J., Mesa, M.D., 2018. Vitamin D: classic and novel actions. Ann. Nutr. Metab. 72 (2), 87–95.

Holick, M.F., et al., 2011. Evaluation, treatment, and prevention of vitamin D deficiency: an endocrine society clinical practice guideline. J. Clin. Endocrinol. Metab. 96 (7), 1911–1930.

Kafle, S., et al., 2021. Association of vitamin D deficiency with pulmonary tuberculosis: a systematic review and meta-analysis. Cureus 13 (9), E17883.

Langlois, P.L., et al., 2018. Vitamin D supplementation in the critically ill: a systematic review and meta-analysis. Clin. Nutr. 37 (4), 1238–1246.

Lokesh, K.S., et al., 2021. Vitamin D deficiency is associated with chronic obstructive pulmonary disease and exacerbation of COPD. Clin. Respir. J. 15 (4), 389–399.

Myint, A., Sauk, J.S., Limketkai, B.N., 2020. The role of vitamin D in inflammatory bowel disease: a guide for clinical practice. Expet Rev. Gastroenterol. Hepatol. 14 (7), 539–552.

Stroehlein, J.K., et al., 2021. Vitamin D supplementation for the treatment of COVID-19: a living systematic review. Cochrane Database Syst. Rev. 5 (5), CD015043.

Further reading

Acen, E.L., et al., 2021. Impact of vitamin D status and cathelicidin antimicrobial peptide on adults with active pulmonary TB globally: a systematic review and meta-analysis. PLoS One 16 (16), E0252762.
Chang Villacreses, M.M., et al., 2021. Conundrum of vitamin D on glucose and fuel homeostasis. World J. Diabetes 12 (9), 1363–1385.
Griffin, G., et al., 2021. "Perspective: vitamin D supplementation prevents rickets and acute respiratory infections when given as daily maintenance but not as intermittent bolus: implications for COVID-19". Clin. Med. 21 (2), E144–E149.
He, M., et al., 2021. Vitamin D deficiency relation to sepsis, paediatric risk of mortality III score, need for ventilation support, length of hospital stay, and duration of mechanical ventilation in critically ill children: a meta-analysis. Int. J. Clin. Pract. 75 (4), E13908.
Kazemi, A., et al., 2021. Association of vitamin D status with SARS-CoV-2 infection or COVID-19 severity: a systematic review and meta-analysis. Adv. Nutr. 12 (5), 1636–1658.
Kikut, J., et al., 2021. Diet supporting therapy for inflammatory bowel diseases. Eur. J. Nutr. 60 (5), 2275–2291.
Martineau, A.R., et al., 2019. Vitamin D supplementation to prevent acute respiratory infections: individual participant data meta-analysis. Health Technol. Assess. 23 (2), 1–44.
Peng, L., et al., 2020. Association between Vitamin D supplementation and mortality in critically ill patients: a systematic review and meta-analysis of randomized clinical trials. PLoS One 15 (12), E0243768.
Verma, S., et al., 2021. Vitamin D deficiency: concern for rheumatoid arthritis and COVID-19? Mol. Cell. Biochem. 476 (12), 4351–4362.
Wang, Y., et al., 2021. Increased risk of vitamin D deficiency among HIV-infected individuals: a systematic review and meta-analysis. Front. Nutr. 8, 722032.
Zakhary, C.M., et al., 2021. Protective role of vitamin D therapy in diabetes mellitus type II. Cureus 13 (8), E17317.

Vitamin E | metabolism and requirements

Maret G. Traber, Linus Pauling Institute, Oregon State University, Corvallis, OR, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of M.G. Traber, Vitamin E: Metabolism and Requirements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 383–389, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00278-6.

Introduction 546
General description and scientific name 546
Vitamin E actions 547
Antioxidant activity 547
Biologic activity 547
Molecular function 548
Vitamin E bioavailability 548
Absorption and plasma transport 548
Plasma α-tocopherol concentrations and tissue delivery 549
Plasma vitamin E kinetics 550
Vitamin E catabolism 550
Human vitamin E deficiency 551
Recommended intake levels 551
Dietary vitamin E 551
Vitamin E supplements 552
Other vitamin E forms 552
Precautions and adverse reactions 552
Over-dosage 552
Increased tendency to bleed 552
Adverse effects of pharmaceutical drugs on vitamin E status 553
Chronic disease prevention 553
Conclusion 554
References 554

Key points

- Vitamin E is a potent fat-soluble antioxidant
- Vitamin E functions to stop lipid peroxidation
- Vitamin E deficiency occurs in humans with gene defects in the α-tocopherol transfer protein (α-TTP) or with fat malabsorption syndromes
- α-TTP and vitamin E catabolism function to maintain the body's preference for α-tocopherol
- Vitamin E deficiency dysregulates energy metabolism, the methionine cycle, and thiol status
- α-TTP and α-tocopherol are necessary for neurologic development in experimental models
- Vitamin E | Metabolism and Requirements

Glossary

α-tocopherol transfer protein The hepatic α-TTP preferentially facilitates secretion of α-tocopherol, specifically *2R*-α-tocopherols, not other tocopherols or tocotrienols, from the liver into the plasma

Ataxia with Vitamin E Deficiency Humans with a defect in the α-TTP gene, who are not supplemented with vitamin E, display a syndrome called "Ataxia with Vitamin E Deficiency" (AVED)

Fat Malabsorption Syndromes Fat malabsorption syndromes are various disorders that cause an inability to absorb dietary fat. These can include cystic fibrosis, short bowel syndrome or cholestatic liver disease, as examples

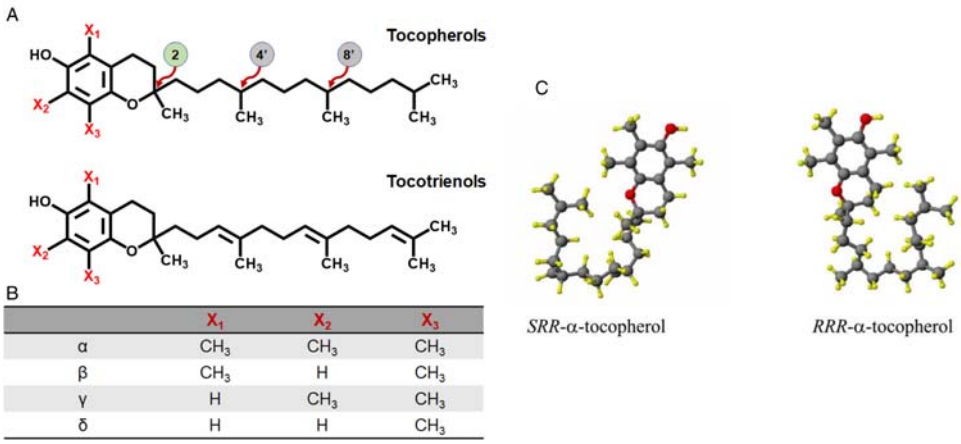
Introduction

Vitamin E was discovered in 1922 because it prevented fetal resorption in rats (Evans and Bishop, 1922). Vitamin E is the most potent, fat-soluble antioxidant in vivo. There are at least 8 different molecular forms with vitamin E antioxidant activity, yet the body preferentially retains α-tocopherol. This preference was recognized in the 2000 Dietary Reference Intakes (DRIs) for vitamin E, which recommended that only α-tocopherol, not the other forms, meets human vitamin E requirements (Food and Nutrition Board and Institute of Medicine, 2000). This recommendation was based on the observations that α–tocopherol is maintained in the circulation by the hepatic α-tocopherol transfer protein (α–TTP). Abnormalities in the α–TTP gene leads to vitamin E deficiency in humans, a syndrome called "Ataxia with Vitamin E Deficiency" [(AVED) OMIM #277460] (Schuelke, 2005 May 20 [Updated 2016 Oct 13]). This demonstration of human vitamin E deficiency solidified the role of α-TTP as a critical determinant of plasma α-tocopherol concentrations. α-TTP is also the key factor in the discrimination between α-tocopherol and other forms of vitamin E, as well as between natural and synthetic α-tocopherols. The other important factor in determining the retained form of vitamin E is metabolism. Specifically, catabolism increases excretion of non-α–tocopherol vitamin E forms, as well as excess α–tocopherol and 2*S*-α-tocopherol (defined below).

General description and scientific name

Dietary components with vitamin E antioxidant activity include α–, β–, γ–, and δ–tocopherols and α–, β–, γ–, and δ–tocotrienols. These compounds all have a chromanol ring with a varying (one to three) number of methyl groups and have either a saturated, phytyl tail (tocopherols) or an unsaturated tail (tocotrienols). α–Tocopherol and α–tocotrienol have three methyl groups, β– and γ– have two, and δ– has one.

The naturally occurring form of α-tocopherol is called *RRR*-α-tocopherol; or on supplement labels, *d*–α-tocopherol; or more formally, 2,5,7,8-tetramethyl-2*R*-(4′*R*,8′*R*,12 trimethyltridecyl)-6-chromanol (**Fig. 1**). Chiral carbon-centers are located at α-tocopherol positions 2, 4′ and 8′ in the side chain. In naturally occurring α-tocopherol, these carbons are in the *R*-conformation, but can be in either the *R*- or the *S*-conformation in synthetic. Chemical α–tocopherol synthesis produces an equal mixture of eight different stereoisomers (*RRR*, *RSR*, *RRS*, *RSS*, *SRR*, *SSR*, *SRS*, *SSS*), or more formally 2,5,7,8-tetramethyl-2*RS*-(4′*RS*,8′*RS*,12 trimethyltridecyl)-6-chromanol. The first letter of the three-letter combination is the 2 position, which is the most important for biologic activity. Thus, only half of the synthetic α-tocopherol is in the "active" 2*R*-α-tocopherol conformation. The dietary reference intakes (DRIs) for vitamin E are given in units of mg 2*R*-α-tocopherol (**Table 1**, see below for discussion). These same units are used for the updated food and supplement labeling, where the daily value or "reference dietary intake" (RDI) is 15 mg (Food and Drug Administration, 2016). To indicate that synthetic α–tocopherol is a racemic mixture, it is called *all-rac*-α-tocopherol, or on supplement labels, *dl*-α–tocopherol. **Table 2** lists the factors to convert IU to mg 2*R*-α-tocopherol. For example, if a vitamin E supplement is labeled 400 IU and it is *dl*-α-tocopheryl acetate, then 400 times 0.45 equals 180 mg 2*R*–α–tocopherol, but if it is labeled *d*-α-tocopheryl acetate, then 400 times 0.67 equals 268 mg 2*R*–α–tocopherol.



	X_1	X_2	X_3
α	CH_3	CH_3	CH_3
β	CH_3	H	CH_3
γ	H	CH_3	CH_3
δ	H	H	CH_3

Fig. 1 Vitamin E structures. (A) The tocopherols have a saturated, phytyl tail and the tocotrienols have an unsaturated tail. Chiral carbon-centers are located at α-tocopherol positions 2, 4′ and 8′ in the side chain. Both tocopherols and tocotrienols have a chromanol ring with a varying (one to three) number of methyl groups. The location of the methyl groups on the chromanol ring is indicated by an X_n. (B) The table shows either a methyl group (CH_3) or a hydrogen (H) at the X location for each of the four tocopherols or tocotrienols. For example, both α–tocopherol and α–tocotrienol have three methyl groups. (C) The α–tocopherol stereochemical structure of natural *RRR*-α-tocopherol [2,5,7,8-tetramethyl-2*R*-(4′*R*,8′*R*,12 trimethyltridecyl)-6-chromanol] is shown on the right. Chemical α–tocopherol synthesis produces an equal mixture of eight different stereoisomers (*RRR*, *RSR*, *RRS*, *RSS*, *SRR*, *SSR*, *SRS*, *SSS*), or more formally 2,5,7,8-tetramethyl-2*RS*-(4′*RS*,8′*RS*,12 trimethyltridecyl)-6-chromanol. *SRR*-α-tocopherol is shown with the S-conformation at the 2 position.

Table 1 Estimated average α−tocopherol requirements (EARs), recommended dietary allowances (RDAs), and average intakes (AIs) in mg/day for Adults and Children.

Life stage	*EAR*	*RDA*	*AI*
0–6 months			4
7–12 months			6
1–3 years	5	6	
4–8 years	6	7	
9–13 years	9	11	
14–18 years	12	15	
Adult (male or female)	12	15	
Pregnant	12	15	
Lactation	16	19	

Adapted from Food and Nutrition Board and Institute of Medicine (2000).

Table 2 Factors to convert IU vitamin E on supplement labels to mg 2*R*-α-tocopherol.

	mg/IU[a]
***all rac*-α-tocopherol and esters**[b]	
dl-α-tocopheryl acetate	0.45
dl-α-tocopheryl succinate	0.45
dl-α-tocopherol	0.45
***RRR*-α-tocopherol and esters**[b]	
d-α-tocopheryl acetate	0.67
d-α-tocopheryl succinate	0.67
d-α-tocopherol	0.67

[a]Multiply the IU times the indicated factor to obtain the mg active vitamin E (2*R*-α−tocopherol).

[b]Note when the esters are synthesized the weight per IU is adjusted to include the molecular weight of the ester form. Thus, to calculate mg α−tocopherol, this adjustment is not included in the calculation.

Adapted from Food and Nutrition Board and Institute of Medicine (2000).

Vitamin E actions

Antioxidant activity

Vitamin E protects polyunsaturated fatty acids within membrane phospholipids, lipid droplets or in plasma lipoproteins. During lipid peroxidation, a peroxyl radical forms in a lipid milieu. In the absence of vitamin E, lipid peroxidation is a chain reaction that continuously produces more radicals. When vitamin E is present in a phospholipid-containing membrane, for example, the peroxyl radical is 1000 times more likely to attack the vitamin E molecule than a polyunsaturated fatty acid (**Fig. 2**). The vitamin E hydroxyl group on the chromanol ring reacts with the peroxyl radical to form the corresponding lipid hydroperoxide and tocopheroxyl radical. The lipid hydroperoxide can be detoxified by other antioxidant systems, such as glutathione peroxidase. However, the vitamin E forms a radical itself, which could re-initiate lipid peroxidation. The tocopheroxyl radical has a number of possible fates. The tocopheroxyl radical can react with another radical to form non-reactive products. Alternatively, it can be further oxidized to the tocopheryl quinone, a two-electron oxidation product. Another possibility is "vitamin E recycling", where the tocopheroxyl radical is restored to its unoxidized form by other antioxidants such as vitamin C or ubiquinol, or with thiols, such as glutathione. This process will deplete these other antioxidants. For this reason, it is important to maintain a good intake of other dietary antioxidants. Additionally, in a closed system, such as a vitamin E deficient zebrafish embryo, other metabolic pathways also become depleted in an effort to maintain phosphatidyl choline, choline, methionine and other methyl donors, as well as glucose-dependent energy systems, such as nicotinamide adenine dinucleotide phosphate [NAD(P)H], (McDougall et al., 2017; Zhang et al., 2021).

Biologic activity

Biologic activity is a term that has been used historically to indicate a disconnect between vitamin E antioxidant activities and in vivo activities. Observations in rodent experiments carried out in the 1930s, formed the basis for determining the "biologic activity" of vitamin E. Although the various vitamin E forms have similar structures and antioxidant activities, they differ in their abilities to prevent or reverse specific vitamin E deficiency symptoms (e.g., fetal resorption, muscular dystrophy and encephalomalacia). α-Tocopherol with three methyl groups and a free hydroxyl group on the chromanol ring with the phytyl tail meeting the ring in the *R*-orientation (**Fig. 1**) has the highest biological activity. This specific structural requirement for biological, but not chemical, activity is dependent upon the α-TTP binding pocket structure (Min et al., 2003) and its hepatic function (Kono and Arai,

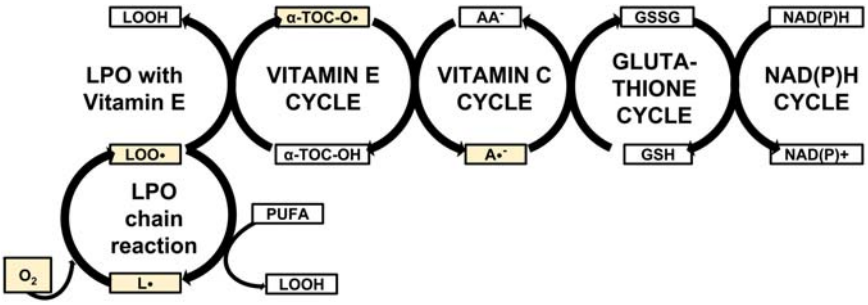


Fig. 2 Lipid peroxidation and vitamin E. During lipid peroxidation (LPO), a peroxyl radical (LOO•) forms when carbon centered radical (L•) reacts with oxygen (O_2, a di-radical). In the absence of vitamin E, LPO is a chain reaction that continuously produces more radicals because LOO• reacts with polyunsaturated fatty acids (PUFA) to generate a lipid hydroperoxide (LOOH) and another lipid radical. When vitamin E is present (shown here as α-tocopherol, α-TOC-OH), the LOO• is 1000 times more likely to attack α-TOC-OH than the PUFA. The vitamin E hydroxyl group (OH) on the chromanol ring reacts with the peroxyl radical to form the corresponding LOOH and tocopheroxyl radical (α-TOC-O•). The LOOH can be detoxified by other antioxidant systems, such as glutathione peroxidase, so long as glutathione (GSH) is available as a reductant. The α-TOC-O• can be restored to its unoxidized form by other antioxidants such as vitamin C (ascorbic acid, AA^-), which is oxidized to the ascorbyl radical ($A•^-$). The $A•^-$ is then reduced by GSH, in this example. GSH is oxidized to GSSG, which can be reduced by GSH reductase using the reduced form of nicotinamide adenine dinucleotide phosphate [NAD(P)H], as a co-factor. Adapted from Cadenas et al. (2016).

2015), as discussed below. α-TTP maintains plasma, and indirectly tissue, α–tocopherol concentrations. The antioxidant functions (described above) are not limited to α-tocopherol, but because the other forms are at much lower concentrations in vivo, α-tocopherol is the major lipid soluble antioxidant in the body.

Molecular function

In addition to antioxidant activity, there have been claims that there are specific α-tocopherol-dependent functions that normalize cellular functions, are involved in cell signaling or in gene regulation. However, most of the information in this area has been obtained from in vitro studies in cell culture experiments. Various studies in vivo have demonstrated that α–tocopherol has no specific molecular transduction role in addition to its antioxidant activity, as reviewed (Blaner et al., 2021). However, α–tocopherol is reportedly necessary for intracellular vesicle formation (Nell et al., 2007), which appears to be important for membrane repair (Howard et al., 2011).

Knockdown of α–TTP in zebrafish is embryonically lethal and α–tocopherol deficiency also causes 80% of the embryos to die, while sufficient embryos during the same 5 day period become swimming fish. The α–TTP gene (*Tttpa*) is highly expressed in the zebrafish embryo, especially in the developing nervous system (Head et al., 2020). α–Tocopherol deficiency causes malformation of the brain and spinal cord, with defects reminiscent of neural tube defects, while simultaneously energy metabolism, thiol status and methylation pathways are dysregulated (**Fig. 3**) (Head et al., 2021; Zhang et al., 2021). During α–tocopherol deficiency, various metabolic pathways become dysregulated, as a result of downstream consequences of replacing peroxidized phospholipids. These responses appear to be under mTOR (mechanistic target of rapamycin) control (Head et al., 2021).

Vitamin E bioavailability

Absorption and plasma transport

Intestinal absorption of vitamin E is dependent upon normal processes of fat absorption. Specifically, both biliary and pancreatic secretions are necessary for solubilization of vitamin E in mixed micelles containing bile acids, fatty acids and monoglycerides. α–Tocopheryl acetates (or other esters) from vitamin E supplements, are hydrolyzed by pancreatic esterases to release α-tocopherol prior to absorption. Various cholesterol transporters [scavenger receptor class B type I (SR-BI), CD36 molecule (CD36), NPC1-like transporter 1 (NPC1L1), and ATP-binding cassettes A1 and G1 (ABCA1 and ABCG1)] are involved in vitamin E uptake from micelles into enterocytes (Reboul, 2019). Following uptake, vitamin E is largely retained within the intestinal cells until the next meal (Traber et al., 2019b). During fat consumption, lipid droplets are rapidly formed inside enterocytes (Soayfane et al., 2016), and retained rather than being immediately secreted in chylomicrons (Khalifeh-Soltani et al., 2016). The next meal stimulates chylomicron secretion from enterocytes that have retained vitamin E and other lipid soluble compounds. Importantly, the various vitamin E forms accumulate in enterocyte lipid droplets, then these droplets coalesce with nascent chylomicrons to enrich their contents with vitamin E, and in the process vitamin E absorption is facilitated. Similar processes occur during absorption of cholesterol (Beaumier-Gallon et al., 2001).

Once in the circulation, chylomicron triglycerides are hydrolyzed by lipoprotein lipase. During chylomicron catabolism in the circulation, free fatty acids are released, and vitamin E is non-specifically transferred both to tissues and to other circulating lipoproteins. It is not until the vitamin E-containing chylomicrons reach the liver that discrimination between the various dietary

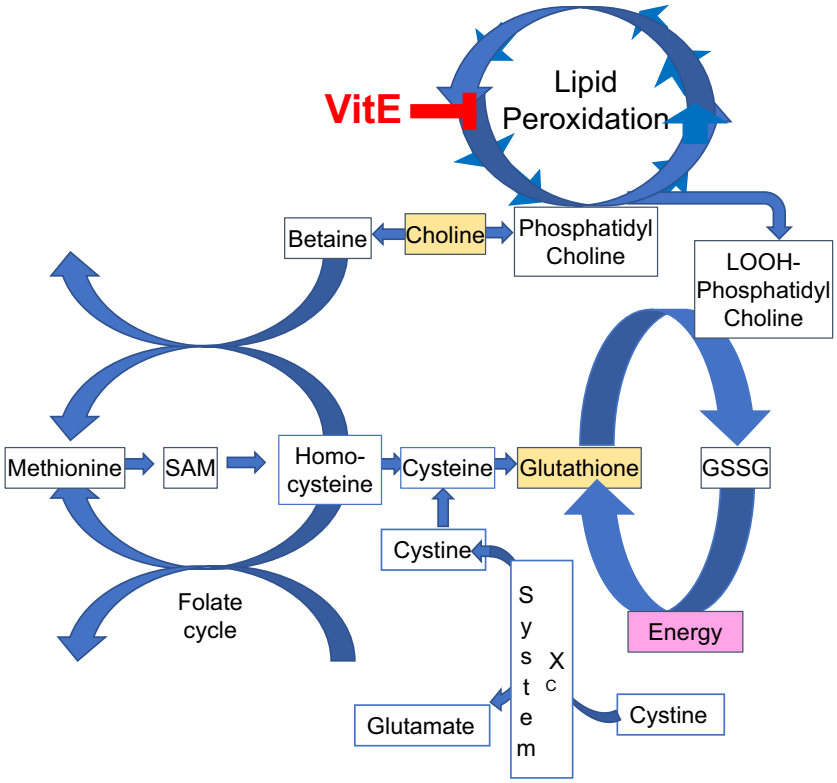


Fig. 3 Impact of lipid peroxidation on other systems. During vitamin E deficiency insufficient antioxidant protection increases lipid hydroperoxides (e.g., LOOH-phosphatidyl choline). Glutathione peroxidase-4 uses glutathione (GSH) to reduce the LOOH. But the regeneration of the oxidized glutathione (GSSG) is through either (A) NADPH-mediated enzymatic reduction, or (B) via de novo GSH synthesis from cysteine. Studies using vitamin E-deficient zebrafish embryos have shown that they experience increased lipid peroxidation, depletion of phosphatidyl choline with docosahexaenoic acid (DHA-PC) and dysregulated phospholipid metabolism (McDougall et al., 2016, 2017). Choline and its oxidation product betaine became dysregulated, leading to impacts on the methionine cycle, S-adenosyl methionine (SAM) and methylation, as well as impacts on cysteine status. The increased need for GSH increases the requirement for cysteine for GSH synthesis. Thus, inadequate vitamin E also impacts the methionine cycle as well as the system Xc-, which exchanges glutamate for extracellular cystine. Adapted from Zhang et al. (2021).

vitamin E forms occurs (Fig. 4). The hepatic α-TTP preferentially facilitates secretion of α-tocopherol, specifically *2R*-α-tocopherols, not other tocopherols or tocotrienols, from the liver into the plasma in very low-density lipoproteins (VLDLs). In the circulation, VLDLs are catabolized to low-density lipoproteins (LDL are also known as the "bad cholesterol" because high LDL levels are associated with increased heart disease).

During lipolysis of both chylomicrons and VLDL there is transfer of vitamin E to HDL. Estimates using a vitamin E-enriched lipid emulsion showed that the half-life of α-tocopherol transfer to HDL during lipolysis in humans was 3 min (Traber et al., 2019b). Although all vitamin E forms are transferred during chylomicron lipolysis, during VLDL lipolysis because they are α-tocopherol enrched, all of the circulating lipoproteins become enriched with α-tocopherol.

Plasma α-tocopherol concentrations and tissue delivery

Plasma α-tocopherol concentrations in normal adult humans range from 11 to 37 μmol/L (Ford et al., 2006). When plasma lipids are taken into account the lower limits of normal are 1.6 μmol α-tocopherol/mmol lipid or 2.5 μmol α-tocopherol/mmol cholesterol. If lipid concentrations are extraordinarily high or low, then correction for lipid levels are helpful to determine adequacy of vitamin E status because α-tocopherol is transported in plasma lipoproteins. Additionally, α-tocopherol concentrations in erythrocytes, adipose tissue or even peripheral nerve have been used to assess vitamin E status.

Vitamin E is delivered to tissues by three mechanisms: transfer from triglyceride-rich lipoproteins during lipolysis, as a result of tissue lipoprotein uptake by various receptors that mediate lipoprotein uptake, and as a result of vitamin E exchange between lipoproteins or tissues. The regulation of tissue vitamin E is not well understood, but α–tocopherol is the predominant form in tissues as a result of its dominance in plasma. All evidence shows that no vitamin E specific-carrier protein is needed for plasma vitamin E transport, but rather vitamin E is non-specifically transported in lipoproteins. An advantage of vitamin E transport in lipoproteins is that easily oxidizable lipids are protected by the simultaneous transport of this lipid-soluble antioxidant. Thus, as peroxidizable lipids are taken up by tissue, the tissues simultaneously acquire a lipid soluble antioxidant.

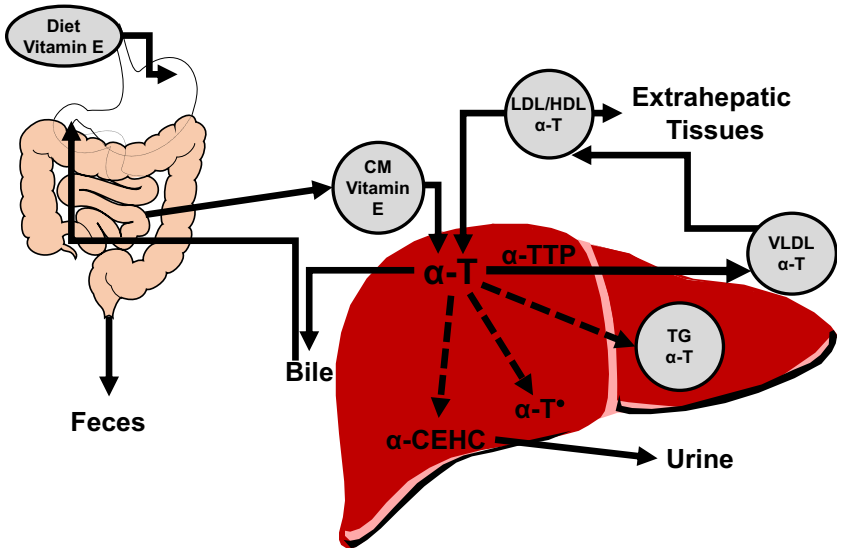


Fig. 4 Vitamin E absorption and trafficking. Intestinal absorption of dietary vitamin E is dependent upon normal processes of fat absorption including biliary and pancreatic secretion. Following vitamin E uptake by enterocytes, nascent chylomicrons (CM) enriched with vitamin E are secreted into the lymphatics, then into the circulation. During chylomicron catabolism by lipoprotein lipase in the circulation, free fatty acids are released, and vitamin E is non-specifically transferred both to tissues and to other circulating lipoproteins. The CM-remnants are taken up by hepatocyte receptor mediated processes. In the liver the hepatic α-TTP preferentially facilitates secretion of α-tocopherol, specifically *2R*-α-tocopherols, into the plasma in very low-density lipoproteins (VLDLs). In the circulation, VLDLs are catabolized to low-density lipoproteins (LDL). During lipolysis of both chylomicrons and VLDL there is transfer of vitamin E to HDL. During VLDL lipolysis, all of the circulating lipoproteins become enriched with α-tocopherol. Hepatic α-tocopherol (α-T) has several possible fates: (A) α-TTP can facilitate its transfer to VLDL; (B) it can be excreted in bile; (C) it can be catabolized to α-CEHC, which can be secreted into plasma, transported to the kidney and excreted in urine; (D) it can be oxidized by a peroxyl radical to α-T•, but is likely rapidly reduced by other antioxidants, or (v) it may remain in the liver in lipid (triglyceride, TG) droplets. Adapted from Traber et al. (2019a).

Plasma vitamin E kinetics

Vitamin E absorption gives an estimate of how much of consumed vitamin E is available to the body. The dual isotope method to measure fractional α-tocopherol absorption is based on the ratio of the plasma concentrations of the administered oral α-tocopherol to IV (equals 100%). Using this technique, fractional α-tocopherol absorption in healthy women is 55% ± 3% (mean ± SEM; n = 10) (Traber et al., 2019b). Higher fractional absorption (~80%) has been estimated using radioactive tocopherol with the balance method (oral minus excreted) (Novotny et al., 2012). Dietary fat (40% vs. 0% fat in the meal) had no effect on α-tocopherol absorption. However, when labeled α-tocopheryl acetate was used as the test dose, fat was necessary for absorption (Bruno et al., 2006). These data suggest that vitamin E esters in food or supplements should be consumed with a fat-containing meal for optimal absorption.

Half-life gives an estimate of how long half of a dose of labeled vitamin E remains in the plasma. The apparent half-life of *RRR*-α-tocopherol in plasma of normal subjects is approximately 48 h, while that of *SRR*-α-tocopherol or γ-tocopherol are only 15 h and the tocotrienols are less than 4 h. More recent estimates suggest that the *RRR*-α-tocopherol half-life is ~30–34 h (Mah et al., 2015; Traber et al., 2019b). α-Tocopherol half-life is slowed in persons with hyperlipidemia, likely because lipoprotein catabolism is slowed. Additionally, bioavailability is less in persons with metabolic syndrome or with fatty liver disease, which may exacerbate oxidative stress associated with the increased inflammation in these disorders.

Vitamin E catabolism

α-TTP is the body's key to maintaining plasma α-tocopherol concentrations, but the xenobiotic catabolism of non-α-tocopherol vitamin E forms and "excess" α-tocopherol serves to increase their excretion and thus decrease their concentrations. Cytochrome P450 4F2 (CYP4F2) initiates vitamin E catabolism by omega-hydroxylation of the sidechain to form the long chain, 13′-hydroxy (13′-OH) catabolite (Sontag and Parker, 2002, 2007). The 13′-OH catabolite undergoes ω-oxidation to form the 13′-carboxy (13′-COOH) catabolite, which then undergoes several rounds of β-oxidation to form carboxyethyl hydroxychromanol (CEHC) (Sontag and Parker, 2007). α− and γ−tocopherols, as well as α− and γ−tocotrienols, are metabolized to α− and γ−CEHCs (2,5,7,8-tetramethyl- and 2,7,8-trimethyl-2-(2′carboxyethyl)-6-hydroxychromanols), respectively. About 1% of a dose of α-tocopherol or tocotrienol, or 5% of a dose of γ−tocopherol or tocotrienol is excreted in the urine as CEHC. Vitamin E catabolism is a key pathway for the regulation vitamin E status and essential for depleting the body of non-α–tocopherol forms.

Tissues involved in vitamin E catabolism based on their CYP4F2 expression include the liver and kidney (Hirani et al., 2008), and intestine (Wang et al., 2007; Uehara et al., 2015). α-Tocopherol catabolism was quantitated using a dual isotope technique. An intravenous, deuterium-labeled α-tocopherol that was administered as an oil:water emulsion to mimic chylomicrons and a differently deuterium-labeled α-tocopherol was consumed (Traber et al., 2019b). Remarkably, more than 90% of the catabolism to urinary α-CEHC took place in the liver when the experimental conditions mimicked normal meal eating. However, when the participants fasted for 12 h following the "test" breakfast, both the intestine and liver played equal roles in α-tocopherol catabolism. These data suggest that delay in intestinal secretion of chylomicrons can markedly increase vitamin E catabolism and thus decrease bioavailability.

Human vitamin E deficiency

The frequency of human vitamin E deficiency is very rare. There have been reports of vitamin E deficiency symptoms in persons with protein-calorie malnutrition. Dietary changes such as decreasing fat intakes, substituting fat-free foods for fat-containing ones and increased reliance of meals away from the home, have resulted in decreased consumption of α−tocopherol-containing foods.

Erythrocyte fragility, hemolysis and anemia were described as vitamin E deficiency symptoms in various animals fed diets devoid of vitamin E. Additionally, studies in experimental animals have shown that a deficiency of both selenium (a required component of glutathione peroxidases) and vitamin E causes a more rapid and severe onset of debilitating deficiency symptom. Deficiency of both vitamins E and C should also cause more severe antioxidant deficiency symptoms, but most animals make their own vitamin C, so this interaction has been difficult to demonstrate in animals. However, when guinea pigs were fed a vitamin E deficient diet and then made vitamin C deficient, the results were catastrophic, resulting in the animals' death within weeks of feeding the vitamin C deficient diet.

In contrast to experimental vitamin E deficiency in rodents, in humans the major vitamin E deficiency symptom is a peripheral neuropathy characterized by the degeneration of the large caliber axons in the sensory neurons. Anemia has also been described.

Vitamin E deficiency was first described in children with fat malabsorption syndromes, principally abetalipoproteinemia, cystic fibrosis and cholestatic liver disease. Subsequently, humans with severe vitamin E deficiency with no known defect in lipid or lipoprotein metabolism were described to have a defect in the α-TTP gene. This syndrome is called "Ataxia with Vitamin E Deficiency" or AVED. The neurologic defects in hypobetalipoproteinemia or chylomicron retention disease are also attributed to vitamin E deficiency.

The neurologic abnormalities can be halted and, in some cases, reversed by vitamin E supplements. Enormous daily supplemental α-T amounts (>100 mg/kg body weight) given long-term can overcome the lack of apoB-lipoproteins in abetalipoproteinemia (Zamel et al., 2008) and prevent oxidative damage (Granot and Kohen, 2004). Reportedly, vitamin E supplementation even allows for normal pregnancy despite apparently low circulating α-T concentrations (Ferreira et al., 2014). It is likely that high density lipoproteins (HDL) can fulfill this gap in apoB-containing lipoproteins (Anwar et al., 2007).

Similarly, supplemental α-T (1000 mg/day) can prevent progression of neurologic defects in AVED (Di Donato et al., 2010); one patient has been reported to be stable for over 30 y (Kohlschutter et al., 2020).

Patients with fat malabsorption due to impaired biliary secretion generally do not absorb orally administered vitamin E. These patients are treated with special forms of vitamin E, such as α-tocopheryl polyethylene glycol succinate that spontaneously form micelles, obviating the need for bile acids.

Recommended intake levels

In 2000, the Food and Nutrition Board of the Institute of Medicine, National Academy of Sciences published the dietary reference intakes (DRIs) for Vitamin C, Vitamin E, Selenium and the Carotenoids. Their recommendations for vitamin E appear in **Table 1**.

The requirements for vitamin E intakes are based primarily on long-term (5–7 years) depletion and repletion studies in humans. Serum α−tocopherol concentrations and corresponding hydrogen peroxide-induced erythrocyte hemolysis were determined at various intervals. Serum α–tocopherol concentrations necessary to prevent in vitro erythrocyte hemolysis in response known levels of α–tocopherol intake in subjects who had undergone experimentally-induced vitamin E deficiency were used to determine estimated average requirements (EARs). The recommended dietary allowances (RDAs) are levels that represent the daily α–tocopherol intakes required to ensure adequate nutrition in 95%–97.5% of the population and are an overestimation of the level needed for most people in any given group.

Dietary vitamin E

Vitamin E can be readily obtained from food. Generally, the richest sources are vegetable oils. Wheat germ oil, safflower oil, and sunflower oil contain predominantly α–tocopherol, while soy and corn oils contain predominantly γ-tocopherol. All of these oils are polyunsaturated. Monounsaturated oils, such as olive or canola oils, also contain predominantly α-tocopherol. Whole grains and nuts are also excellent sources of vitamin E. Most fruits and vegetables, although rich in water-soluble antioxidants, are *not* excellent sources of vitamin E. The FDA defines an excellent source as a usual serving having 20% of the reference daily intake

(RDI) (Code of Federal Regulations (eCFR), 2017a). The RDI for vitamin E is 15 mg α-tocopherol, where 1 mg α-tocopherol = 1 mg *RRR*-α-tocopherol = 2 mg *all rac*-α-tocopherol (Code of Federal Regulations (eCFR), 2017b).

Vitamin E supplements

Most vitamin E supplements and food fortificants contain *all rac*-α-tocopherol, but can contain mixtures of tocopherols or tocotrienols. Supplements often are sold as esters, which protect α–tocopherol from oxidation. These esters can be acetates, succinates or nicotinates of α–tocopherol. Either the natural stereoisomer (*RRR*-α-tocopherol) or the synthetic (*all rac*-α-tocopherol) can be sold as an ester, e.g., *d*- or *dl*-α–tocopheryl acetate, respectively.

Other vitamin E forms

It is often assumed that γ−tocopherol can substitute for α−tocopherol with an efficiency of 10%. However, functionally γ−tocopherol, as well as other non-α−tocopherol forms, are not equivalent to α−tocopherol, cannot be converted by the body to α-tocopherol, and are rapidly excreted from the body.

Precautions and adverse reactions

A number of meta-analyses of outcomes from vitamin E supplementation trials have been reported. Some of these report increased calculated risk of mortality with vitamin E supplementation, while others do not, despite using a similar assortment of clinical trial data. To date, there are no mechanistic studies demonstrating how vitamin E supplements cause to increased mortality. The US Preventive Services Task Force (Fortmann et al., 2013) concluded "trials of vitamin E supplementation showed mixed results and altogether had no overall effect on cancer, CVD, or all-cause mortality". The documented assortment of various adverse consequences from α-tocopherol supplement intervention trials include increased bleeding (Lee et al., 2005), heart failure (Lonn et al., 2005), hemorrhagic stroke (Sesso et al., 2008; Schurks et al., 2010) and increased risk of prostate cancer (Klein et al., 2011).

Over-dosage

The 2000 Food and Nutrition Board of the Institute of Medicine, National Academy of Sciences recommended 1000 mg as an upper limit (UL) of all forms of α-tocopherol in supplements taken by adults 19 years and older, including pregnant and lactating women. UL were set for children and adolescents by adjusting the adult limit on the basis of relative body weight. **Table 3** gives the α-tocopherol UL by age group. No UL was set for infants due to lack of adequate data. The 2000 Food and Nutrition Board did recommend that food be the only source of vitamin E for infants. However, a UL of 21 mg/day was suggested for premature infants with birth weights of 1.5 kg, based on the adult UL.

The vitamin E UL was set for supplements because it is almost impossible to consume enough α–tocopherol-containing foods to achieve a daily 1000 mg intake for prolonged periods of time. The UL was defined for all forms of α–tocopherol, not just the 2*R*-forms, because all of the forms in *all rac*-α–tocopherol are absorbed and delivered to the liver. The appropriate conversion factors are *different* from those shown in **Table 2**, and necessary to estimate the UL for supplements containing either *RRR*- or *all rac*-α–tocopherol supplements. The UL amounts given in IU are shown in **Table 4**. The UL for *RRR*-α–tocopherol is apparently higher because each capsule of *RRR*-α–tocopherol contains less α–tocopherol than does one containing *all rac*-α–tocopherol.

Increased tendency to bleed

High vitamin E intakes are associated with an increased tendency to bleed and this was the symptom that the Food and Nutrition Board used to set the UL. It is not known if increased bleeding is a result of decreased platelet aggregation caused by an inhibition of protein kinase C by α-tocopherol, some other platelet related mechanism, or decreased clotting due to a vitamin K and E interaction causing abnormal blood clotting. Notably, the Women's Health Study reported that nosebleeds were an adverse effect of

Table 3 Upper limits (UL) for α–tocopherol intakes.

Age (years)	*UL (mg/day)*
1–3	200
4–8	300
9–13	600
14–18	800
>19	1000

Adapted from Food and Nutrition Board and Institute of Medicine (2000).

Table 4 Upper limits (UL) reported in IU for α–tocopherol-containing supplements.

	Number of IU that equal 1000 mg
***all rac*-α-tocopherol and esters**	
dl-α-tocopheryl acetate	1100
dl-α-tocopheryl succinate	1100
dl-α-tocopherol	1100
***RRR*-α-tocopherol and esters**	
d-α-tocopheryl acetate	1500
d-α-tocopheryl succinate	1500
d-α-tocopherol	1500

Adapted from Food and Nutrition Board and Institute of Medicine (2000).

vitamin E supplements (39,876 women > or = 45 years of age, 600 IU *RRR*-α–tocopherol or placebo on alternate days for 10 y) (Glynn et al., 2007). Additionally, they found that vitamin E supplements decrease the risk of venous thromboembolism (blood clots). Individuals who are deficient in vitamin K or who are on anticoagulant therapy are at increased risk of uncontrolled bleeding. Patients on anticoagulant therapy should be monitored when taking vitamin E supplements to insure adequate vitamin K intakes.

Adverse effects of pharmaceutical drugs on vitamin E status

Drugs intended to promote weight loss by impairing fat absorption, such as Orlistat or sucrose polyester, can also impair vitamin E and other fat-soluble vitamin absorption. Therefore, multivitamin supplementation is recommended when taking these drugs. Vitamin supplements should be taken with meals at times other than when these drugs are taken to allow adequate absorption of the fat-soluble vitamins.

Chronic disease prevention

Importantly, vitamin E's potential role in preventing or ameliorating chronic diseases associated with oxidative stress leads us to ask whether vitamin E supplements might be beneficial. For many vitamins, when "excess" amounts are consumed, they are excreted and provide no added benefits. Antioxidant nutrients may, however, be different. Heart disease and stroke, cancer, chronic inflammation, impaired immune function, Alzheimer's disease—a case can be made for the role of oxygen free radicals in the etiology of all of these disorders, and even in aging itself. Do antioxidant nutrients counteract the effects of free radicals and thereby ameliorate these disorders? And if so, do large antioxidant supplements have beneficial effects beyond "required" amounts? The 2000 DRI report on Vitamin C, Vitamin E, Selenium and Carotenoids stated that there was insufficient proof to warrant advocating supplementation with antioxidants. But they also stated that the hypothesis that antioxidant supplements might have beneficial effects was promising. Even 20 years later, this remains a very controversial area in vitamin E research.

One approach that has been successful in demonstrating vitamin E effectiveness in mitigating chronic disease risk is the identification of patients with increased levels of oxidative stress. One such strategy, called pharmacogenomics by its proponents, identified that diabetic patients with the haptoglobin 2-2 genotype is associated with increased risk of cardiovascular disease (CVD). These patients were found to have benefit from vitamin E supplementation.

Another approach is to evaluate large populations, then follow their chronic disease progression. The Alpha-Tocopherol, Beta-Carotene Cancer Prevention trial had 29,092 male participants. Men with higher serum α-tocopherol at baseline had significantly lower all-cause mortality and significantly decreased mortality from cardiovascular disease, heart disease, stroke, cancer, respiratory disease, and other causes, with risk reductions from 17% to 47% for the highest versus lowest quintile (Huang et al., 2019). Conversely, low α-tocopherol intakes increase serum γ-tocopherol concentrations. It is interesting that serum γ-tocopherol concentrations in a subset of 3904 men and 4461 women from the Multiethnic Cohort Study, were associated with increased mortality.

Specific interventions where vitamin E supplements have had a benefit include slowing the progression of macular degeneration (Age-Related Eye Disease Study Research, 2001; Evans and Lawrenson, 2017), slowing the decline of Alzheimer disease (Farina et al., 2017) and in the treatment of fatty liver disease (NAFLD, NASH) in both adults (Sanyal et al., 2010; Sato et al., 2015) and children (Lavine et al., 2010). Interestingly, all of these disorders occur in tissues that express α-TTP, suggesting that these tissues may be more susceptible to damage by lipid peroxidation. This concept may be true for the liver where excess lipid droplets may prevent normal α-tocopherol trafficking by α-TTP (Podszun et al., 2020; Violet et al., 2020).

Conclusion

Vitamin E is a nutrient that is necessary to be obtained from the diet and better quality diets have higher vitamin E contents and are associated with better health (Gicevic et al., 2021). Nonetheless, in the enthusiasm for vitamin E supplements to prevent, or perhaps treat chronic diseases, the role of vitamin E as a required nutrient is frequently overlooked. In this regard, recent studies in zebrafish embryos highlight the necessity of α-tocopherol to protect the nervous system in the developing embryonic, vertebrate animal. This model system also highlights the need for α-TTP for trafficking α-tocopherol even before a liver has developed. In this regard, it is also interesting that plants synthesize a variety of molecules with vitamin E antioxidant activity, but animals have developed not only a preference for α-tocopherol, but a strong xenobiotic system to rapidly catabolize non-α-tocopherols.

References

Age-Related Eye Disease Study Research Group, 2001. A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E, beta carotene, and zinc for age-related macular degeneration and vision loss: AREDS report no. 8. Arch. Ophthalmol. 119, 1417–1436. https://doi.org/10.1001/archopht.119.10.1417.

Anwar, K., Iqbal, J., Hussain, M.M., 2007. Mechanisms involved in vitamin E transport by primary enterocytes and in vivo absorption. J. Lipid Res. 48, 2028–2038. https://doi.org/10.1194/jlr.M700207-JLR200.

Beaumier-Gallon, G., Dubois, C., Senft, M., et al., 2001. Dietary cholesterol is secreted in intestinally derived chylomicrons during several subsequent postprandial phases in healthy humans. Am. J. Clin. Nutr. 73, 870–877. https://doi.org/10.1093/ajcn/73.5.870.

Blaner, W.S., Shmarakov, I.O., Traber, M.G., 2021. Vitamin A and Vitamin E: will the real antioxidant please stand up? Annu. Rev. Nutr. 41, 7.1–7.27. https://doi.org/10.1146/annurev-nutr-082018-124228.

Bruno, R.S., Leonard, S.W., Park, S., Zhao, Y.Y., Traber, M.G., 2006. Human vitamin E requirements assessed with the use of apples fortified with deuterium-labeled alpha-tocopheryl acetate. Am. J. Clin. Nutr. 83, 299–304.

Cadenas, E., Packer, L., Traber, M.G., 2016. Antioxidants, oxidants, and redox impacts on cell function—a tribute to Helmut Sies. Arch. Biochem. Biophys. 595, 94–99. https://doi.org/10.1016/j.abb.2015.11.012.

Code of Federal Regulations (eCFR), 2017a. Federal Register U.S. Food and Drug Administration. 21CFR101.54. Retrieved Date 8/06/2021 from. https://ecfr.federalregister.gov/current/title-21/chapter-I/subchapter-B/part-101/subpart-D/section-101.54.

Code of Federal Regulations (eCFR), 2017b. Federal Register U.S. Food and Drug Administration. 21CFR101.9. Retrieved Date 8/06/2021 from. https://ecfr.federalregister.gov/current/title-21/chapter-I/subchapter-B/part-101/subpart-A/section-101.9#p-101.9(c)(9).

Di Donato, I., Bianchi, S., Federico, A., 2010. Ataxia with vitamin E deficiency: update of molecular diagnosis. Neurol. Sci. 31, 511–515. https://doi.org/10.1007/s10072-010-0261-1.

Evans, H.M., Bishop, K.S., 1922. On the existence of a Hitherto unrecognized dietary factor essential for reproduction. Science 56, 650–651. https://doi.org/10.1126/science.56.1458.650.

Evans, J.R., Lawrenson, J.G., 2017. Antioxidant vitamin and mineral supplements for slowing the progression of age-related macular degeneration. Cochrane Database Syst. Rev. 7, CD000254. https://doi.org/10.1002/14651858.CD000254.pub4.

Farina, N., Llewellyn, D., Isaac, M.G., Tabet, N., 2017. Vitamin E for Alzheimer's dementia and mild cognitive impairment. Cochrane Database Syst. Rev. 1, CD002854. https://doi.org/10.1002/14651858.CD002854.pub4.

Ferreira, F., Patel, V., Matts, S., 2014. A successful spontaneous pregnancy in abetalipoproteinemia: Amsterdam or the art of vitamin replacement? BMJ Case Rep. 2014. https://doi.org/10.1136/bcr-2014-206754.

Food and Drug Administration, 2016. Food labeling: revision of the nutrition and supplement facts labels. Final rule. Fed. Regist. 81 (103), 33742–33999.

Food and Nutrition Board and Institute of Medicine, 2000. Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids. National Academy Press, Washington (DC). https://www.nap.edu/catalog/9810/dietary-reference-intakes-for-vitamin-c-vitamin-e-selenium-and-carotenoids.

Ford, E.S., Schleicher, R.L., Mokdad, A.H., Ajani, U.A., Liu, S., 2006. Distribution of serum concentrations of alpha-tocopherol and gamma-tocopherol in the US population. Am. J. Clin. Nutr. 84, 375–383. https://doi.org/10.1093/ajcn/84.1.375.

Fortmann, S.P., Burda, B.U., Senger, C.A., et al., 2013. Vitamin, Mineral, and Multivitamin Supplements for the Primary Prevention of Cardiovascular Disease and Cancer: A Systematic Evidence Review for the U.S. Preventive Services Task Force. Agency for Healthcare Research and Quality. Evidence Report No. 108. No.

Gicevic, S., Tahirovic, E., Bromage, S., Willett, W., 2021. Diet quality and all-cause mortality among US adults, estimated from National Health and Nutrition Examination Survey (NHANES), 2003–2008. Publ. Health Nutr. 24, 2777–2787. https://doi.org/10.1017/S1368980021000859.

Glynn, R.J., Ridker, P.M., Goldhaber, S.Z., Zee, R.Y., Buring, J.E., 2007. Effects of random allocation to vitamin E supplementation on the occurrence of venous thromboembolism: report from the Women's Health Study. Circulation 116, 1497–1503. https://doi.org/10.1161/CIRCULATIONAHA.107.716407.

Granot, E., Kohen, R., 2004. Oxidative stress in abetalipoproteinemia patients receiving long-term vitamin E and vitamin A supplementation. Am. J. Clin. Nutr. 79, 226–230. https://doi.org/10.1093/ajcn/79.2.226.

Head, B., La Du, J., Tanguay, R.L., Kioussi, C., Traber, M.G., 2020. Vitamin E is necessary for Zebrafish nervous system development. Sci. Rep. 10, 15028. https://doi.org/10.1038/s41598-020-71760-x.

Head, B., Ramsey, S.A., Kioussi, C., Tanguay, R.L., Traber, M.G., 2021. Vitamin E deficiency disrupts gene expression networks during Zebrafish development. Nutrients 13, 468. https://doi.org/10.3390/nu13020468.

Hirani, V., Yarovoy, A., Kozeska, A., Magnusson, R.P., Lasker, J.M., 2008. Expression of CYP4F2 in human liver and kidney: assessment using targeted peptide antibodies. Arch. Biochem. Biophys. 478, 59–68. https://doi.org/10.1016/j.abb.2008.06.025.

Howard, A.C., McNeil, A.K., McNeil, P.L., 2011. Promotion of plasma membrane repair by vitamin E. Nat. Commun. 2, 597. https://doi.org/10.1038/ncomms1594.

Huang, J., Weinstein, S.J., Yu, K., Mannisto, S., Albanes, D., 2019. Relationship between serum alpha-tocopherol and overall and cause-specific mortality. Circ. Res. 125, 29–40. https://doi.org/10.1161/CIRCRESAHA.119.314944.

Khalifeh-Soltani, A., Gupta, D., Ha, A., et al., 2016. Mfge8 regulates enterocyte lipid storage by promoting enterocyte triglyceride hydrolase activity. JCI Insight 1, e87418. https://doi.org/10.1172/jci.insight.87418.

Klein, E.A., Thompson Jr., I.M., Tangen, C.M., et al., 2011. Vitamin E and the risk of prostate cancer: the selenium and vitamin E cancer prevention trial (SELECT). JAMA 306, 1549–1556. https://doi.org/10.1001/jama.2011.1437.

Kohlschutter, A., Finckh, B., Nickel, M., Bley, A., Hubner, C., 2020. First recognized patient with genetic vitamin E deficiency stable after 36 years of controlled supplement therapy. Neurodegener. Dis. 20, 35–38. https://doi.org/10.1159/000508080.

Kono, N., Arai, H., 2015. Intracellular transport of fat-soluble vitamins A and E. Traffic 16, 19–34. https://doi.org/10.1111/tra.12231.

Lavine, J.E., Schwimmer, J.B., Molleston, J.P., et al., 2010. Treatment of nonalcoholic fatty liver disease in children: TONIC trial design. Contemp. Clin. Trials 31, 62–70. https://doi.org/10.1016/j.cct.2009.09.001. S1551-7144(09)00150-5 [pii].

Lee, I.M., Cook, N.R., Gaziano, J.M., et al., 2005. Vitamin E in the primary prevention of cardiovascular disease and cancer: the Women's Health Study: a randomized controlled trial. JAMA 294, 56–65. https://doi.org/10.1001/jama.294.1.56.

Lonn, E., Bosch, J., Yusuf, S., et al., 2005. Effects of long-term vitamin E supplementation on cardiovascular events and cancer: a randomized controlled trial. JAMA 293, 1338–1347. https://doi.org/10.1001/jama.293.11.1338.

Mah, E., Sapper, T.N., Chitchumroonchokchai, C., et al., 2015. alpha-Tocopherol bioavailability is lower in adults with metabolic syndrome regardless of dairy fat co-ingestion: a randomized, double-blind, crossover trial. Am. J. Clin. Nutr. 102, 1070–1080. https://doi.org/10.3945/ajcn.115.118570.

McDougall, M.Q., Choi, J., Stevens, J.F., et al., 2016. Lipidomics and H2(18)O labeling techniques reveal increased remodeling of DHA-containing membrane phospholipids associated with abnormal locomotor responses in alpha-tocopherol deficient zebrafish (danio rerio) embryos. Redox Biol. 8, 165–174. https://doi.org/10.1016/j.redox.2016.01.004.

McDougall, M., Choi, J., Kim, H.K., et al., 2017. Lethal dysregulation of energy metabolism during embryonic vitamin E deficiency. Free Radic. Biol. Med. 104, 324–332. https://doi.org/10.1016/j.freeradbiomed.2017.01.020.

Min, K.C., Kovall, R.A., Hendrickson, W.A., 2003. Crystal structure of human alpha-tocopherol transfer protein bound to its ligand: implications for ataxia with vitamin E deficiency. Proc. Natl. Acad. Sci. U. S. A. 100, 14713–14718. https://doi.org/10.1073/pnas.2136684100.

Nell, S., Bahtz, R., Bossecker, A., et al., 2007. PCR-verified microarray analysis and functional in vitro studies indicate a role of alpha-tocopherol in vesicular transport. Free Radic. Res. 41, 930–942. https://doi.org/10.1080/10715760701416988.

Novotny, J.A., Fadel, J.G., Holstege, D.M., Furr, H.C., Clifford, A.J., 2012. This kinetic, bioavailability, and metabolism study of RRR-alpha-tocopherol in healthy adults suggests lower intake requirements than previous estimates. J. Nutr. 142, 2105–2111. https://doi.org/10.3945/jn.112.166462.

Podszun, M.C., Alawad, A.S., Lingala, S., et al., 2020. Vitamin E treatment in NAFLD patients demonstrates that oxidative stress drives steatosis through upregulation of de-novo lipogenesis. Redox Biol 37, 101710. https://doi.org/10.1016/j.redox.2020.101710.

Reboul, E., 2019. Vitamin E intestinal absorption: regulation of membrane transport across the enterocyte. IUBMB Life 71, 416–423. https://doi.org/10.1002/iub.1955.

Sanyal, A.J., Chalasani, N., Kowdley, K.V., et al., 2010. Pioglitazone, vitamin E, or placebo for nonalcoholic steatohepatitis. N. Engl. J. Med. 362, 1675–1685. https://doi.org/10.1056/NEJMoa0907929.

Sato, K., Gosho, M., Yamamoto, T., et al., 2015. Vitamin E has a beneficial effect on nonalcoholic fatty liver disease: a meta-analysis of randomized controlled trials. Nutrition 31, 923–930. https://doi.org/10.1016/j.nut.2014.11.018.

Schuelke, M., 2005. Ataxia with vitamin E deficiency. In: Adam, M.P., Ardinger, H.H., Pagon, R.A., et al. (Eds.), GeneReviews®. University of Washington, Seattle, Seattle, WA. Retrieved August 2021 from. https://www.ncbi.nlm.nih.gov/books/NBK1241/.

Schurks, M., Glynn, R.J., Rist, P.M., Tzourio, C., Kurth, T., 2010. Effects of vitamin E on stroke subtypes: meta-analysis of randomised controlled trials. BMJ 341, c5702. https://doi.org/10.1136/bmj.c5702.

Sesso, H.D., Buring, J.E., Christen, W.G., et al., 2008. Vitamins E and C in the prevention of cardiovascular disease in men: the Physicians' Health Study II randomized controlled trial. JAMA 300, 2123–2133. https://doi.org/10.1001/jama.2008.600.

Soayfane, Z., Terce, F., Cantiello, M., et al., 2016. Exposure to dietary lipid leads to rapid production of cytosolic lipid droplets near the brush border membrane. Nutr. Metab. 13, 48. https://doi.org/10.1186/s12986-016-0107-9.

Sontag, T.J., Parker, R.S., 2002. Cytochrome P450 omega-hydroxylase pathway of tocopherol catabolism. Novel mechanism of regulation of vitamin E status. J. Biol. Chem. 277, 25290–25296. https://doi.org/10.1074/jbc.M201466200.

Sontag, T.J., Parker, R.S., 2007. Influence of major structural features of tocopherols and tocotrienols on their omega-oxidation by tocopherol-omega-hydroxylase. J. Lipid Res. 48, 1090–1098. https://doi.org/10.1194/jlr.M600514-JLR200.

Traber, M.G., Buettner, G.R., Bruno, R.S., 2019a. The relationship between vitamin C status, the gut-liver axis, and metabolic syndrome. Redox Biol. 21, 101091. https://doi.org/10.1016/j.redox.2018.101091.

Traber, M.G., Leonard, S.W., Ebenuwa, I., et al., 2019b. Vitamin E absorption and kinetics in healthy women, as modulated by food and by fat, studied using 2 deuterium-labeled alpha-tocopherols in a 3-phase crossover design. Am. J. Clin. Nutr. 110, 1148–1167. https://doi.org/10.1093/ajcn/nqz172.

Uehara, S., Murayama, N., Nakanishi, Y., et al., 2015. Immunochemical quantification of cynomolgus CYP2J2, CYP4A and CYP4F enzymes in liver and small intestine. Xenobiotica 45, 124–130. https://doi.org/10.3109/00498254.2014.952800.

Violet, P.C., Ebenuwa, I.C., Wang, Y., et al., 2020. Vitamin E sequestration by liver fat in humans. JCI Insight 5, e133309. https://doi.org/10.1172/jci.insight.133309.

Wang, M.Z., Wu, J.Q., Bridges, A.S., et al., 2007. Human enteric microsomal CYP4F enzymes O-demethylate the antiparasitic prodrug pafuramidine. Drug Metab. Dispos. 35, 2067–2075. https://doi.org/10.1124/dmd.107.016428.

Zamel, R., Khan, R., Pollex, R.L., Hegele, R.A., 2008. Abetalipoproteinemia: two case reports and literature review. Orphanet J. Rare Dis. 3, 19. https://doi.org/10.1186/1750-1172-3-19.

Zhang, J., Head, B., Leonard, S.W., et al., 2021. Vitamin E deficiency dysregulates thiols, amino acids and related molecules during zebrafish embryogenesis. Redox Biol 38, 101784. https://doi.org/10.1016/j.redox.2020.101784.

Vitamin E physiology and health effects

Jean-Marc Zingg, Miller School of Medicine, University of Miami, Miami, FL, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of P.A. Morrissey, M. Kiely, Vitamin E: Physiology and Health Effects, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 390–397, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00279-8.

Introduction 556
Chemistry of the eight natural vitamin E analogs 557
Only one of the eight vitamin E analogs, *RRR*-αT, is essential 558
Natural and synthetic analogs of vitamin E (αT) 559
Vitamin E metabolites 561
Dietary sources 562
Interaction of vitamin E with MUFA and PUFA 562
Molecular mechanisms of vitamin E absorption, metabolism and excretion 563
Vitamin E deficiency 564
Molecular mechanisms of vitamin E action 565
Gene expression arrays, proteomics, metabolomics and lipidomics 565
Vitamin E and disease prevention 566
Polymorphisms relevant for vitamin E action 566
Summary and conclusion 567
References 567

Key points

- Only one of the eight natural vitamin E analogs (α-tocopherol, αT) is essential for humans
- The other seven vitamin E analogs reach low micromolar concentrations only in the postprandial state, after which they are metabolized and eliminated
- Vitamin E is the most active lipid-soluble chemical antioxidant protecting poly/mono-unsaturated fatty acids and cholesterol against damage induced by free radicals
- Vitamin E analog specific interactions with membranes, signaling enzymes, transcription factors, transport and structural proteins lead to genome-wide changes of gene expression
- Polymorphisms in genes related to vitamin E transport and action may help to identify sub-populations of individuals at risk for low vitamin E status susceptible to a personalized approach of disease prevention

Introduction

This year nears 100 years since Herbert McLean Evans and Katherine Julia Bishop discovered a fat-soluble dietary constituent essential for reproduction in rats later named vitamin E (α-tocopherol (αT)). Initial studies mainly focused on the chemical structure of vitamin E, its synthesis, its antioxidant properties, and its requirements as an essential dietary component in studies with rodents. In the last 20–30 years, with the advance of modern molecular biology, genetics and biochemistry, the mechanisms of vitamin E uptake into the body, its metabolism and more recently its regulatory roles in cells have been elucidated. In animals, these studies firmly established the essential role of vitamin E in preventing reproductive failure and symptoms of neuro/muscular degeneration; in humans, clear evidence for its essential function came after the seminal discovery of a rare genetic disease with very low plasma levels of vitamin E, ataxia with vitamin E deficiency (AVED), with similar symptoms as observed in vitamin E deficient animals.

AVED is caused by mutations in a single gene, the alpha-tocopherol transfer protein (αTTP), that is essential for enriching vitamin E in plasma to a level required for normal development and health. Whereas sufficient vitamin E is present in an equilibrated diet to sustain its essential functions, certain diseases and conditions impair the absorption of vitamin E from foods and supplementation is indicated. Health effects with vitamin E supplementation beyond the essential have extensively been investigated in animal studies, with however often equivocal results in human studies. These findings highlight the complexity of vitamin E action in humans, that depends on many factors that are difficult to control including interactions with other dietary components, individual-specific uptake and metabolism and the presence of genetic polymorphisms. Recent developments on the regulatory effects of natural and synthetic vitamin E analogs on genome wide gene expression and metabolomics emphasize the importance of maintaining life-long adequate levels of vitamin E for development and disease prevention.

Chemistry of the eight natural vitamin E analogs

In photosynthetic organisms including higher plants, cyclization and differential methylation of precursors is leading to eight vitamin E analogs (the α-, β-, γ-, δ-tocopherols (αT, βT, γT, δT) and α-, β-, γ-, δ-tocotrienols (αTT, βTT, γTT, δTT)) (Fig. 1) (Mene-Saffrane, 2017).

The different chemical structure and the consequent differential interaction with membranes, organelles, and proteins (enzymes, transcription factors, transporters, structural proteins) are thought to be responsible for the vitamin E analog specific uptake, distribution and cellular effects.

Chemically, the eight vitamin E analogs act as free radical chain breaking molecules via the chromanol hydroxy group mostly when embedded in cellular membranes, lipid vesicles and lipoproteins and protect the membrane components (unsaturated lipids, cholesterol) from oxidation. The chemical antioxidant activity of the eight natural vitamin E analogs is similar, albeit at the molecular level some differences in the chemical, physical and biological effects can be distinguished that depend on the assay used as well as on the solvent, microenvironment and presence in membrane, micelles or lipoproteins (Miyazawa et al., 2019; Wagner et al., 2004). *In vitro*, the efficiency of inhibition of lipid peroxidation by tocopherols has been determined to be within the same range with $\alpha > \gamma \approx \beta > \delta$.

The chromane ring is methylated at various locations and differences in methylation not only defines the four tocopherol and tocotrienol analogs (α, β, γ, δ) but also leads to different affinity to αTTP resulting in preferential enrichment and retention of fully methylated *RRR*-αT (Fig. 1). Moreover, differential methylation of each vitamin E analog affects the interaction with enzymes, structural proteins, transporters, transcription factors and membrane lipids believed to be the main reason for the differential regulatory effects on signaling and gene expression that can occur by mechanisms independent of acting as a chemical antioxidant. Chemically, αT is more stable than the other analogs since the reactive hydroxy group is located between two methyl groups so that the unpaired electron can delocalize over the fully substituted chromanol ring after reacting with a lipid peroxide thus slowing its chemical reactivity, until it is reduced and regenerated by vitamin C (L-ascorbic acid) and glutathione or fully oxidized to alpha-tocopheryl quinone (αTQ). In γT (and also βT and δT) the aromatic ring can react at the unmethylated sites with peroxynitrite forming 5-NO_2-γT and the corresponding metabolite, 5-NO_2-γ-CEHC (Wagner et al., 2004), but unlike the αT phenoxyl radical, 5-NO_2-γT is not regenerated. Peroxynitrite is formed by macrophages during inflammation and scavenging by γT suggests that it may act as a better anti-inflammatory agent – in particular in skin where the levels of γT are generally high, but possibly even at the lower concentrations reached in plasma. Accordingly, neointima formation induced by vascular injury is reduced by γT by reducing nitrosative stress in insulin resistant rats, whereas the effects seen with αT are weaker.

The hydrophobic side chain present in all vitamin E analogs (for tocopherols, saturated derived from phytyl pyrophosphate; for tocotrienols, unsaturated derived from geranylgeranyl pyrophosphate) is responsible for its preferential location in the lipid phase (lipoprotein, membranes, vesicles, hydrophobic pockets of proteins). The tocotrienols (TT) have a more pronounced antioxidant effects since they are more homogenously distributed in membranes facilitating recycling of the chromanoxyl radical. In

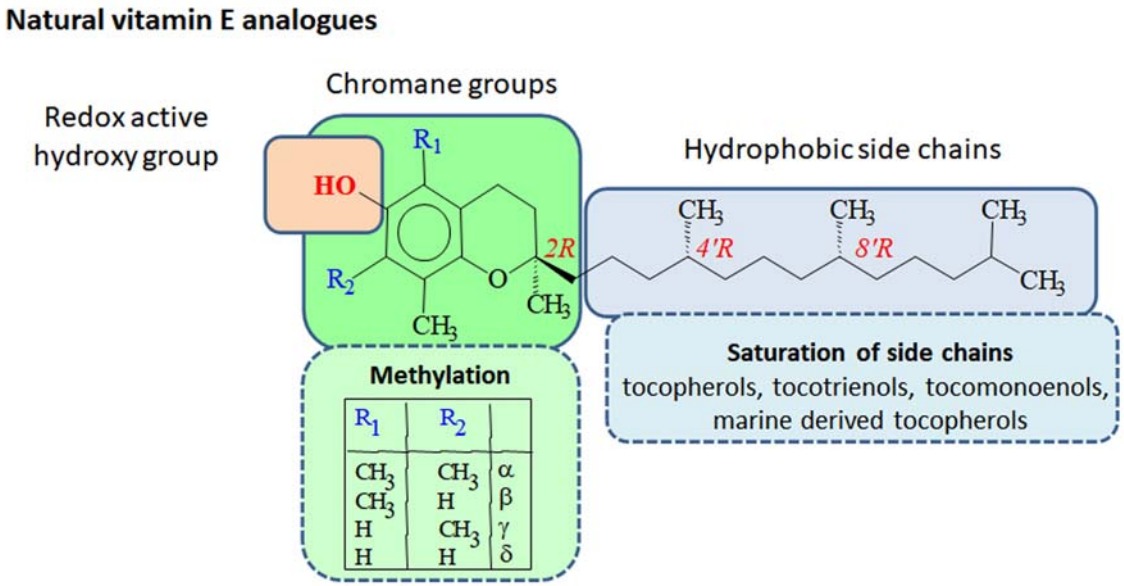


Fig. 1 Chemical structures of natural vitamin E. Natural vitamin E is a composite of three structural and functional entities: (1) the hydrophobic side chain that anchors vitamin E in the plasma membrane, vesicles and hydrophobic pockets of proteins (for tocopherols, a saturated phytyl side chain natural in the *RRR*-configuration; for tocotrienols and marine derived tocopherols, an unsaturated isoprenoid side chain); (2) the chromane group modified by differential methylation (α-, β-, γ-, δ-tocopherols and tocotrienols), and (3) the hydroxy group responsible for the antioxidant activity.

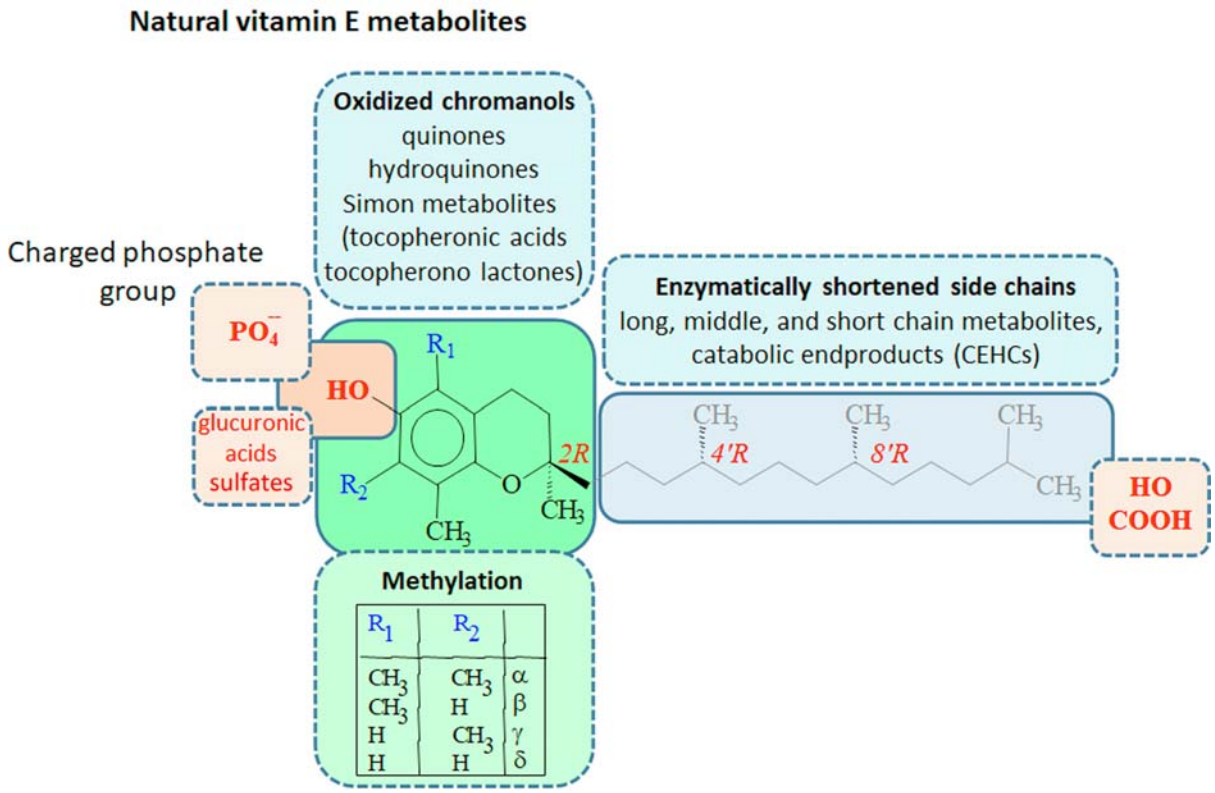


Fig. 2 Natural vitamin E metabolites. Vitamin E metabolites are generated from natural vitamin E either enzymatically or oxidatively. Enzymatically generated vitamin E metabolites become first hydroxylated at the hydrophobic side chain and then are progressively shortened to produce long-, intermediate- and short-chain metabolites until the metabolic end products (CEHCs) are formed. These enzymatic modifications of vitamin E lead to increasingly more water-soluble metabolites but still maintaining the antioxidant activity of the chromanol group known for the intact molecules. Thus, metabolism of vitamin E may liberate activities associated with the chromanol moiety (antioxidant, signaling molecule) and facilitate their movement into different cellular compartments as well their secretion by the kidney. In contrast, intracellular distribution of intact vitamin E is facilitated by several lipid transporters and lipid exchange reactions (by αTTP, TAP1/2/2 or SEC14L2/3/4, respectively, NPCL1, NPC1, NPC2), or possibly can occur after enzymatic modification with polar residues (e.g., hydroxy, phosphate, sulfate, glucuronide). Oxidative metabolites contain oxidized chromanols (quinones, hydroquinones), and the so-called Simon metabolites (tocopheronic acids, tocopheronolactones). For better elimination by the kidney, the metabolites become modified by glucuronidation or sulfation.

membranes the vitamin E analogs are the main molecules that act as chain breaking antioxidants able to scavenge reactive oxygen (ROS) and reactive nitrogen species (RNS) with consequent reduction of the random formation and accumulation of damaged molecules such as membrane lipids, proteins and nucleic acids and consequent cellular dysfunction and death. Accordingly, the level of vitamin E may affect the generation of oxidized lipids in membranes and thus indirectly modulate their signaling function as active lipid mediators.

The hydrophobic side chains of the vitamin E analogs become hydroxylated and progressively shortened upon enzymatic metabolism, increasing the solubility of the metabolites but still maintaining the antioxidant activity of the chromanol group known for the intact molecules (Fig. 2).

Thus, metabolism of vitamin E may "liberate" activities associated with the chromanol moiety (antioxidant, signaling molecule) and facilitate their movement into different cellular compartments. In contrast, mobilization of intact vitamin E from membranes and lipid droplets and intracellular distribution requires lipid exchange reactions mediated by lipid transport proteins (e.g., αTTP, TAP1/2/3 also named SEC14L2/3/4, respectively, NPCL1, NPC1, NPC2), or possibly is facilitated after enzymatic modification with polar residues (e.g., with hydroxy, phosphate, sulfate, glucuronide).

Only one of the eight vitamin E analogs, *RRR*-αT, is essential

Selective recognition by αTTP in the liver and enrichment of plasma and tissue with one analog, *RRR*-αT, is believed to be the main reason for its much higher biological potency (α >> γ > δ > β) (Azzi, 2018), and since it fulfills all essential functions, it has been recognized as the analog that acts as vitamin E (Food and Nutrition Board, 2000). This is best demonstrated by comparing αT to αTT that both have the same chromanol head group but only αT is enriched in plasma and considered to be essential, whereas αTT not

despite having *in vitro* a comparable or more pronounced antioxidant and cellular activity. Upon supplementation and after a postprandial peak at about 4–8 h, excess αT and all the other tocopherols and tocotrienols become metabolized by the liver and eliminated, whereas αT remains elevated even in the fasted state, with an average steady-state plasma concentration of ~23 μM ($t_{1/2}$–29 h). It is unknown whether the not-retained tocopherol analogs and their metabolites have some essential cellular function at their nanomolar to low micromolar concentrations reached in plasma, beyond a possible "E" function in support to αT.

Preferential retention of *RRR*-αT may also be the result of evolutionary selection of the fully methylated vitamin E analog forming a more stable tocopheroxyl radical that upon full oxidation to alpha-tocopherol quinone/hydroquinone (αTQ/αTHQ) has non-arylating electrophilic properties, whereas oxidation of βT, γT and δT leads to the arylating desmethyl analogs (βTQ, γTQ, δTQ) with mutagenic, cytotoxic and apoptotic properties. Moreover, αTHQ may be the vitamin E analog with higher activity to prevent ferroptosis and neurodegenerative dysfunction that occurs in the brain during vitamin E deficiency, by reducing the non-heme iron of 15-lipoxygenase from its active Fe^{3+} to an inactive Fe^{2+} state.

Whereas intake of vitamin E is generally regarded as safe (RDA for humans is 15 mg/day, supplements are usually between 200 and 400 mg/day, and the upper tolerable limit is 1000 mg/day), adverse effects have been reported to occur above 400 mg/day (Hathcock et al., 2005). Increased lipids in liver have been reported as one of the symptoms of hypervitaminosis E in rats and mice. Similarly, long-term intake of a 2% tocotrienol mixture showed some highly proliferative liver lesions and a reduction of the survival rate in rats. More insight into the biological reasons for selecting only αT to become enriched in the body during evolution of higher organisms may come from an engineered αTTP, that instead of αT preferentially recognizes γT (γTTP).

It should be emphasized at this point, that albeit natural *RRR*-αT harbors all the essential vitamin E functions, numerous experiments *in vitro* and *in vivo* have shown vitamin E associated regulatory effects and cellular responses with all natural vitamin E analogs and their contribution to disease prevention is under investigation. *In vivo* a vitamin E analog specific response has been mostly attributed to their different bioavailability and rate of metabolism. Intriguingly, rapid conversion of all the non-*RRR*-αT analogs into metabolites may increase their concentration in plasma to a level required for their action (Birringer and Lorkowski, 2019). However, even when used at equal concentration in cultured cells the eight vitamin E analogs and their metabolites often show differences suggesting additional regulatory mechanisms that go beyond that of scavenging free radicals.

Natural and synthetic analogs of vitamin E (αT)

Albeit eight different vitamin E analogs are present in nature, only *RRR*-αT is being considered as "the vitamin E" since it is enriched in the human body and carries all the to date known essential functions (Azzi, 2018). The natural tocopherols contain a phytyl side chain with three chiral centers which naturally occur in the *RRR* configuration, whereas the tocotrienols have an unsaturated isoprenoid side chain (**Fig. 1**). In commercially available products, natural vitamin E is present either as *RRR*-αT (formerly called *d*-αT), or as a mixture of naturally occurring tocopherols and tocotrienols (all as *RRR*-stereoisomers). In synthetic vitamin E products, as result of chemical synthesis, the side chain consists of all possible combinations of *R* and *S* side-chain stereoisomers occurring at equal amounts (e.g., synthetic racemic vitamin E or *all-rac*-αT, formerly called *dl*-αT) (**Fig. 3**).

For use in supplements and cosmetics stabilized vitamin E derivatives that are esterified at the chromanol hydroxy group and thus are not susceptible to oxidation have been synthesized (e.g., α-tocopheryl acetate (αTA), α-tocopheryl succinate (αTS), α-tocopheryl phosphate (αTP)) having advantages for processing, storage and absorption. Depending on the application, these stabilized vitamin E analogs may have different solubility, transport and metabolism. Stabilized esters of αT are mostly converted to the natural forms by pancreatic, intestinal or epidermal esterases and can be considered to be pro-vitamins since they ultimately perform the same function in the body as the natural vitamin E.

The activity of the natural and synthetic vitamin E analogs has been tested and compared using an assay preventing reproductive failure in rats, that was developed based on the original description of vitamin E as an essential nutrient for reproduction. The natural-source *RRR*-αT vitamin E has a 1.36-fold greater biological potency compared to synthetic *all-rac*-αT assumed to be mainly the result of the better retention of *RRR*-αT by αTTP in the liver. Based on the rat fetal resorption assay, the units definition (United States Pharmacopoeia (USP)) of the different vitamin E analogs has been (1 USP of vitamin E = 1 mg of *all-rac*-αTA, or 0.67 mg *RRR*-αT, or 0.74 mg *RRR*-αTA) (Food and Nutrition Board, 2000). One mg of the most commonly used synthetic form of vitamin E, *all rac*-αTA, is regarded as 1 international unit (IU), and the potency of natural form of αT is set to be equal to 1.49 IU. Compared to that, the biological activity of γT is about 1/10th of that of αT.

In situations where ester hydrolysis is inefficient, the stabilized αT derivatives may perform different cellular functions that are based on their different chemical and physical properties. Some of these derivatives, such as αTS, may exert a dual function, first as anticancer pro-vitamin E in the non-hydrolyzed form, and second as vitamin E, in the hydrolyzed form (Neuzil, 2002). Within mesenchymal stem and progenitor cells, intact αTA is thought to maintain primitive cells and attenuate mitochondrial oxygen consumption by mimicking hypoxia. Recent findings indicate that conversion of αTA to αT may less occur in lungs upon inhaling vapors from e-cigarettes containing αTA that was present as vape-cartridge additive (DiPasquale et al., 2020). Increased levels of intact αTA are reportedly detected in electronic-cigarette or vaping product use-associated lung injury (EVALI). Although the mechanisms by which intact αTA may affect lungs in cases of EVALI are still unclear it may act as linactant and disrupt the pulmonary surfactant and/or contribute to the formation of macrophages foam cells detected in bronchoalveolar lavages and lungs after inhaling e-cigarette vapors from products containing Δ9-tetrahydrocannabinol (THC) and αTA.

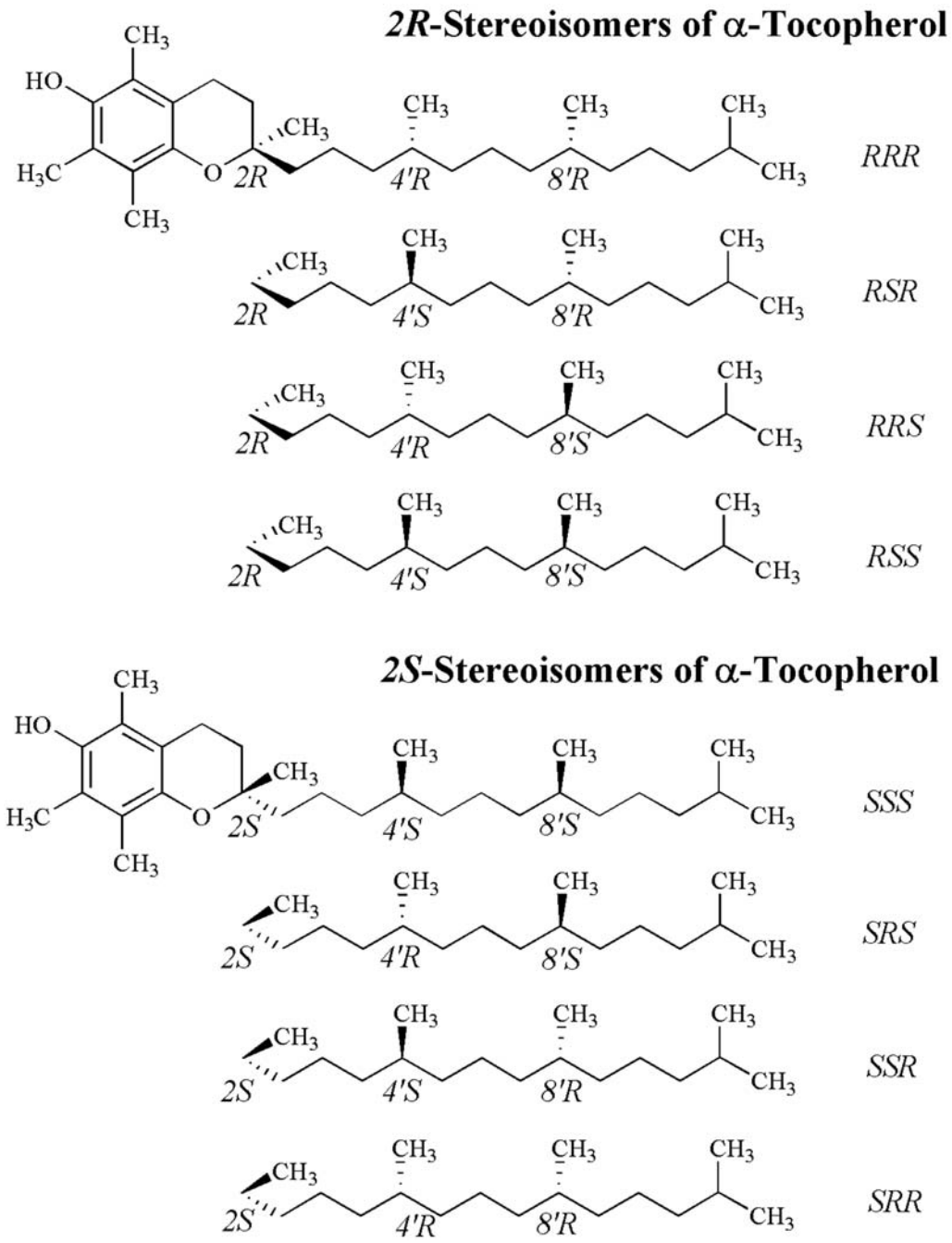


Fig. 3 Natural and synthetic analogs of vitamin E (αT). As result of chemical synthesis, synthetic vitamin E products consist of all possible combinations of *R* and *S* side-chain stereoisomers occurring at equal amounts (e.g., synthetic racemic vitamin E or *all-rac*-αT, formerly called dl-αT). In the liver, αTTP recognizes preferentially *RRR*-αT and less also *RSR*, *RRS* and *RSS*-side chain isomers (the so-called *2R*-αT-stereoisomers) for enrichment in very low-density lipoproteins (VLDL), whereas the so-called *2S*-αT-stereoisomers are metabolized and excreted.

Intact αTP can affect signaling and gene expression and modulates cellular events ranging from proliferation, survival/apoptosis, lipid uptake and metabolism, phagocytosis, long term potentiation, cell migration, telomere maintenance and angiogenesis. The bioavailability as an intact molecule is generally low, in skin, only about 3.1%–4.8% of αTP were reported to be converted to αT within 24 h, albeit other studies in other cell types suggested more rapid conversion. Similarly, conversion of αTA to αT was not observed on the surface or in the horny layers of human skin, while up to 50% was converted in the underlying skin. αTP show improved penetration into epidermis and can form nanocarriers for delivery of molecules across the skin, such as drugs (e.g., caffeine, oxycodone), vitamins (vitamins, including αT and αTP itself, vitamin D3), fatty acids (omega-3 fatty acid docosahexaenoic acid (DHA)) and other molecules such as carnosine, insulin and coenzyme Q10.

Vitamin E metabolites

Broadly, two different types of vitamin E metabolites can be distinguished reflecting the mechanism by which they are generated in the body (enzymatically and oxidatively) (Fig. 2). Enzymatically generated vitamin E metabolites are produced when their level exceeds the capacity of αTTP to export them from the liver. For αT, which is selectively recognized by αTTP and specifically sorted for incorporation into very low-density lipoproteins (VLDL), much higher micromolar concentrations need to be reached for metabolism to occur when compared to the other seven tocopherols and tocotrienols that become metabolized and excreted already at nanomolar to low micromolar plasma concentrations (reviewed in (Zingg and Azzi, 2004)). However, even in αTTP knockout mice non-αT analogs are preferentially metabolized suggesting increased recogniton by metabolic enzymes as mechanism for their more efficient elimination.

The proposed pathway of enzymatic metabolism of the vitamin E analogs proceeds first via ω-oxidation of the side-chains catalyzed by the cytochrome P_{450} enzymes CYP3A and CYP4F2 in the endoplasmic reticulum leading to 13′-hydroxychromanols (13′-OHs), and then β-oxidation in peroxisomes and mitochondria leading to long chain metabolites (LCM), e.g., 13′-carboxychromanols (13′-COOHs), that are further β-oxidized to intermediate- and short-chain metabolites, e.g., the carboxymethylbutyl hydroxychromans (CMBHCs), and finally to the metabolic end products, the carboxyethyl hydroxychromans (CEHC) (Birringer and Lorkowski, 2019).

With increasing intake and after a threshold of plasma αT has been exceeded, α-CEHC excretion in urine is augmented and the intact chromane structure suggests that it is enzymatically generated and not derived from αT that has reacted as antioxidant. Metabolism of δT leads to δ-CEHC, and that of γT to γ-CEHC, which has been identified in human urine and proposed to act as a natriuretic factor. For better excretion, these increasingly more water-soluble metabolites become further conjugated with glucuronide or sulfate.

Mechanistically, metabolism and elimination occur by induction of cytochrome P_{450} enzymes via activation of the pregnane X receptor (PXR) (reviewed in (Traber, 2004)). The extent by which the chromanol hydroxy group is exposed may determine the strength of activation of PXR, with αT being the weakest, whereas βT, γT, δT and the tocotrienols being stronger. Interestingly, the tocopherol metabolic products do not activate PXR by themselves suggesting that vitamin E transport proteins may be required to facilitate transport of intact vitamin E analogs to PXR in the nucleus and/or to cytochromes P_{450} enzymes in the endoplasmic reticulum and in mitochondria. Similar requirements for intracellular transport of other hydrophobic molecules have been described, such as for squalene (by supernatant protein factor (SPF), also known as TAP1/SEC14L2), vitamin A/retinoids (by intracellular retinoid-binding proteins such as cellular retinol-binding protein (CRBP), cellular retinoic acid-binding protein (CRABP) and cellular retinal-binding protein (CRALBP)), for fatty acids (by several fatty acid binding proteins (FABP)), phospholipids (phospholipid transfer protein (PLTP)), or cholesterol and cholesteryl esters (triacylglycerols or cholesteryl ester transfer protein (CETP), oxysterol-homology binding protein 4 (OSH4p)) (Kono and Arai, 2015). Since activation of PXR mediates metabolism of many drugs by inducing cytochromes P_{450} enzymes (e.g., CYP3A) and ATP-binding cassette (ABC) transporters, the vitamin E metabolism may be influenced by other drugs and nutrients including the different vitamin E analogs themselves. In fact, some inhibitors of the CYP3A family, like sesamin and ketoconazole, inhibit the formation of γ-CEHC, explaining the increased serum γT levels in humans after dietary intervention with sesame oils. On the other hand, activators of CYP3A such as rifampicin can lead to higher formation of α-CEHC in HepG2 cells.

Oxidatively generated vitamin E metabolites are generated upon scavenging free radicals by vitamin E. The so-called "Simon metabolites" (tocopheronic acid and tocopheronolactone) are excreted in the urine as glucuronides or sulfates. These metabolites have a shortened side chain and an opened, oxidized, chromane structure that is often quoted to demonstrate their antioxidant function *in vivo*. A marked increase of these metabolites is observed in the urine of healthy volunteers after daily intake of 2–3 g *all rac*-αT. Likewise, in children with type 1 diabetes, the presence of conjugated α-tocopheronolactone (sulfate and glucuronide) in urine has been suggested as a biomarker for oxidative stress.

The nitrated metabolite of γT, NO_2-γ-CEHC, can be generated from γT after nitration to 5-NO_2-γT and subsequent metabolism. However, since 5-NO_2-γT is apparently inefficiently incorporated into liver cells, it is not detected in urine while γ-CEHC is abundant and it is to date unknown whether it exerts any cellular activity at higher concentrations. In smokers, αT supplementation reduces circulating γT in part by increasing its metabolism and competition with uptake with consequent lower 5-NO_2-γT formation by scavenging of RNS.

As recently critically reviewed, some of the metabolites act as bioactive molecules that modulate the activity of transcription factors, membrane channels and enzymes even at the low nanomolar concentrations reached in plasma and in the intestine (low micromolar concentrations can be reached after supplementation) (Birringer and Lorkowski, 2019). Among the activities of metabolites (LCM, CEHC) are: natriuretic activity of γ-CEHC, anti-inflammatory and antioxidative activities by inhibiting several enzymes (e.g., cyclooxygenase-1 and 2 (COX-1/2), prostaglandin E2 (PGE2) synthase, 5-, and 12-lipoxigenases (5-, 12-LOX), inducible nitric oxide synthetase (iNOS), leukotriene synthase). Furthermore, gene regulatory activities relevant for lipid homeostasis including upregulation of CD36 scavenger receptor/fatty acids transporter (CD36/FAT) and the lipid droplet-associated protein perilipin 2 (PLIN2) have been observed. At higher micromolar concentrations, some metabolites can induce apoptosis in cancer cell lines. In prostate cancer cells, inhibition of cyclin D1 expression by γ-CEHC leads to inhibition of cell proliferation, and since the inhibitory effect of γ-CEHC is competed for by α-CEHC non-antioxidant mechanism have been suggested.

Dietary sources

The eight natural vitamin E analogs are exclusively synthesized in photosynthetic organisms including higher plants. Significant amounts are present in all green tissues and also in seeds, where the eight vitamin E analogs are present in different relative amounts (reviewed in (Mene-Saffrane, 2017; Zingg and Azzi, 2004)). The major source of vitamin E in the human diet are olive and vegetable seed oils (corn, soybean, safflower, palm) that contain different relative amounts of each vitamin E analog (Wagner et al., 2004; Jiang, 2014). Sunflower seed oil and corn oil are excellent sources of αT, whereas corn and soybean oil contain the highest amounts of γT (Jiang, 2014). Olive, sunflower, canola oil and corn oil are classified as a good food source of tocopherols, whereas animal products are generally a relatively poor source of vitamin E. Tocotrienols are mainly present in oils from palm, wheat germ, barely, and rice bran. Grape seed oils contain αTT and γTT and annatto/achiote seeds contain predominantly δTT. Palm oil also contains α-tocomonoenol, and some marine organisms contain marine derived tocopherol (MDT) having a single unsaturated bond at the end of the phytyl side chain as a possible adaption to maintain membrane fluidity in cold water. Other sources of tocopherols are nuts, seeds, whole grain and wheat germ (200–1000 mg/kg), and to a lesser extend fruits (with the exception olives (a stone fruit)) and vegetables. Preferential intake of corn oil leads to higher intake of γT and in the US, high intake of soybean oil leads to more intake of γT and δT in Asia, whereas preferential intake of olive and sunflower oils in Europe leads to higher intake of αT (Wagner et al., 2004; Zingg and Azzi, 2004). Adherence to certain diets and cooking habits (e.g., Mediterranean, Asian) are thought to influence the relative intake of vitamin E analogs with possible consequence for human health.

Interaction of vitamin E with MUFA and PUFA

As the main lipid-soluble antioxidant, vitamin E stabilizes membranes by itself as a structural lipid and protects other unsaturated lipids such as polyunsaturated and monosaturated fatty acids (PUFA and MUFA, respectively) by scavenging lipid peroxyl radicals and acting as a chain-breaking antioxidant that can be regenerated by vitamin C and glutathione. With a typical dietary PUFA intake, the vitamin E requirements have been estimated to range between 12 and 20 mg (Raederstorff et al., 2015), an amount that is as determined by the Institute of Medicine within the range of the recommended dietary allowance (RDA) of 15 mg vitamin E (Food and Nutrition Board, 2000). The relative level of vitamin E and PUFA appears to be particularly important when PUFA are increased in the diet by supplementation (Raederstorff et al., 2015; Zingg and Meydani, 2019). It is to date unknown whether scavenging of peroxynitrite by γT could reduce the formation of nitro-fatty acids from PUFA, such as conjugated linoleic acid (CLA) that have emerged as important regulators of vascular and inflammatory functions.

The n-3 and n-6 polyunsaturated fatty acids (omega-3 or omega-6 PUFA, respectively) are nutritionally essential since they cannot be synthesized *de novo* from 2-carbon fragments. As a result of their unsaturated double bonds, PUFA are susceptible to chemical reactions with ROS and RNS. PUFA in phospholipids not only influence membrane fluidity, curvature and the properties of membrane microdomains, but increase also the risk for chain reactions of lipid peroxidation with consequent membrane destabilization and cellular dysfunction. Supplementation with vitamin E decreases the urinary level of F2-isoprostane (a biomarker for oxidized arachidonic acid, an n-6 PUFA) in *ApoE*$^{-/-}$ atherosclerotic mice that have elevated levels in urine, plasma, and vascular tissue. However, in humans very high doses of vitamin E over a prolonged time were required to reduce F2-isoprostane in plasma. Thus, vitamin E and PUFA interact with each other as a pair in which vitamin E protects PUFA and excess PUFA "consume" vitamin E, and a high PUFA/vitamin E ratio is generally considered as disadvantageous (Raederstorff et al., 2015; Zingg and Meydani, 2019).

In cells, both vitamin E and PUFA have redox-independent regulatory functions, mostly after being metabolized to active lipid mediators via binding to specific enzymes and receptors involved in modulating signal transduction and gene expression relevant for immune, inflammatory, developmental and metabolic pathways (Zingg and Meydani, 2019). In recent studies that investigate the role of PUFA and vitamin E in the brain, gene regulatory effects of vitamin E and PUFA have been analyzed using genome-wide gene expression analysis and metabolomics (Traber, 2020). In 1 year old zebrafish (*Danio rerio*) with vitamin E deficiency and limiting dietary long-chain PUFA, a protective antioxidant effect of vitamin E on PUFA in brain has been demonstrated. However, of the 155 lipids surveyed in brain extracts by lipidomics, only four phospholipids (PL) were significantly different suggesting additional regulatory mechanisms.

Interestingly, as revealed by metabolomics vitamin E supplementation (400 mg/d *RRR*-αTA, for 4 wks) increased the levels of a number of lysophosphatidylcholine (LPC) species (16:0, 18:0, 18:1, 18:2, 20:3, 22:6) in human plasma. Some LPC species are selectively transported across the blood brain barrier (BBB) by Mfsd2a, a member of the major facilitator superfamily of proteins, and furthermore converted into the n-3 fatty acid docosahexaenoic acid (DHA), suggesting that their increased plasma presence in response to vitamin E may indirectly lead to changes in DHA-mediated signaling and gene expression in the brain. The importance of adequate uptake of LPC-DHA by Mfsda2a for brain growth and function was demonstrated both in zebrafish and in humans in which mutations in Mfsda2a were associated with a microcephaly syndrome.

In the retina of vitamin E deficient αTTP-knockout mice, an age-related decrease of n-3 PUFA was detected that was more pronounced when compared to controls, suggesting enhanced PUFA peroxidation and accelerated retinal degenerative damage. Alternatively, vitamin E may act indirectly by increasing LPC-DHA levels in plasma with consequent increased transport of LPC-DHA in the eye relevant for the development of photoreceptor discs.

Taken together, the efficiency of uptake, transport and metabolism of vitamin E and PUFA, their interaction and their consequent relative levels in cells and tissues may act as important determinants for both physiological and patho-physiological cellular functions and therefore influence the risk for a number of diseases such as atherosclerosis, inflammation, chronic heart failure, rheumatoid arthritis, retinal degeneration, allergies and neurodegenerative disorders.

Molecular mechanisms of vitamin E absorption, metabolism and excretion

Differential transport, retention and metabolism of natural αT over the other 7 vitamin E analogs is believed to be the main reason for the essential role of only αT (Fig. 4) (Kono and Arai, 2015).

The absorption of the vitamin E analogs in the intestine is affected by several dietary factors (food matrix, fat, micronutrients) that facilitate emulsification and mixed micelles formation, and in the case of stabilized αTA is dependent on hydrolysis by intestinal and pancreatic esterases. Passage across the intestinal epithelium follows a similar route as fatty acids, and recent research has highlighted the involvement of scavenger receptors such as CD36 and SR-BI, as well as intracellular lipid transporters such as Niemann-Pick -C1 like 1 (NPCL1), fatty acids binding proteins (FABPs) and tocopherol associated proteins (TAP1/2/3) (Reboul, 2018). The eight natural vitamin E analogs and also synthetic *all-rac*-αT are taken up across the intestinal epithelium with similar efficiency ranging from 10 to 80%. Distribution to peripheral cells of these vitamin E analogs coming from the intestine is mediated by chylomicrons (and in part also from high density lipoproteins (HDL)), and chylomicron remnants bring the remainder to the liver. In the liver, αT (preferentially *RRR*, but less also *RSR*, *RRS* and *RSS*-side chain isomers, the so-called *2R*-αT-stereoisomers (Fig. 3)) are recognized by αTTP, retained and incorporated into VLDL, leading to an up to 50 fold enrichment of αT in plasma.

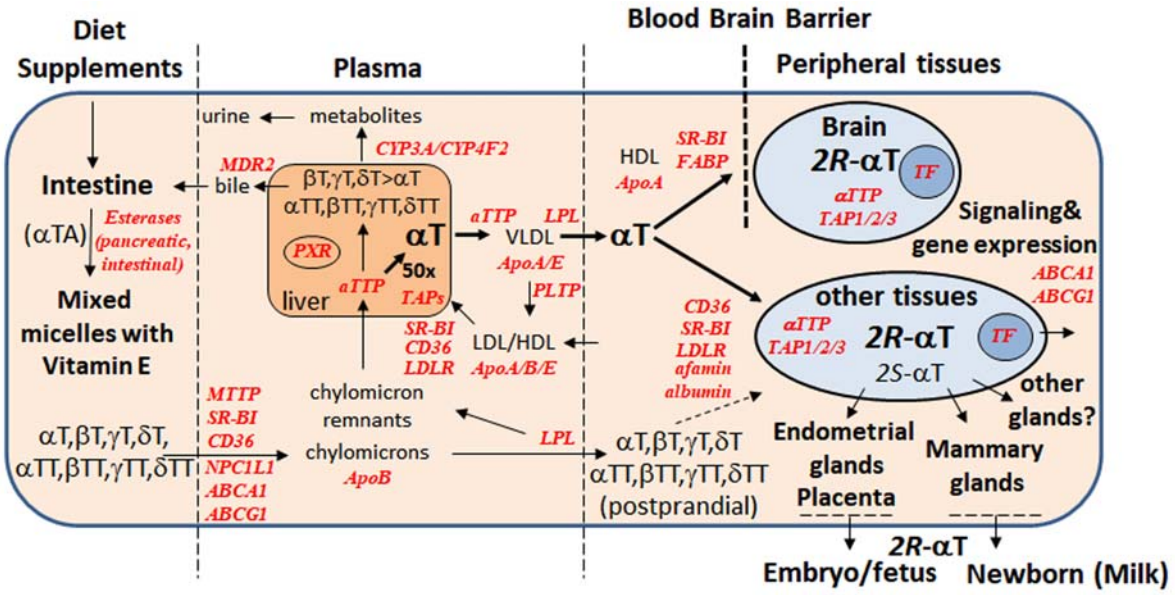


Fig. 4 Uptake, distribution, metabolism and secretion of different natural and synthetic vitamin E analogs and the most important genes involved. The eight natural vitamin E analogs (αT, βT, γT, δT and αTT, βTT, γTT, δTT) are present in the diet in different quantities, and upon formation of mixed micelles are absorbed by the intestine with similar efficiency. Intestinal and pancreatic esterases convert the stabilized vitamin E form (αTA) present as precursor in many supplements into αT before it can be taken up by the intestine. Transport of the eight vitamin E analogs across the intestinal epithelium is mediated by transport proteins (MTTP, NPC1L1, ABCA1/G1, SR-BI, CD36), after which they are secreted within chylomicrons to the blood stream. Lipoprotein lipase (LPL) releases the vitamin E analogs from chylomicrons to the cells in peripheral tissues. Several proteins (SR-BI, CD36, LDLR, afamin, albumin) facilitate uptake of the vitamin E analogs into cells in which they are transported by αTTP, TAP1/2/3 or SEC14L2/3/4, respectively, and secreted by ABCA1/G1. The liver takes up vitamin E from chylomicron remnants and LDL/HDL via LDLR, SR-BI and CD36. In the liver, αT (preferentially *RRR*-αT, but also all the *2R*-αT stereoisomers) is selectively recognized by αTTP, enriched (~50x) and incorporated into very low-density lipoproteins (VLDL), whereas the other vitamin E analogs (βT, γT, δT, excess *RRR*-αT, the *2S*-αT stereoisomers, and the four tocotrienols) are metabolized by CYP3A/CYP4F2 and secreted in urine and bile involving MDR2. Uptake and transport is most efficient for the natural *RRR*-αT stereoisomer, less for the 2*R*-αT stereoisomers and the other seven vitamin E analogs, whereas the synthetic *2S*-αT stereoisomers are inefficiently recognized by αTTP and become metabolizes in the liver. The *2R*-αT stereoisomers are preferentially transported to the brain (blood brain barrier), to the embryo/fetus (endometrial glands, placenta) and to the newborn (mammary glands). In tissues, the vitamin E analogs differently affect several signal transduction pathways leading to changes in the activity of several transcription factors (TF). Polymorphisms within these genes involved in uptake, transport, distribution and metabolism can play a role in vitamin E bioavailability and bioactivity and thus can also influence signaling and gene expression.

After a postprandial peak, the other seven vitamin E analogs, excess αT and the so-called *2S*-αT-stereoisomers (**Fig. 3**) are metabolized and excreted, explaining their differential efficiency of cellular uptake, transport, intracellular distribution, and conversion to different metabolites. Metabolism of γT in αTTP knockout mice is still higher than that of αT suggesting increased recogniton by metabolic enzymes and not differential transport as reason for metabolic discrimination of non-αT analogs, albeit binding to αTTP protected to some degree γT from being metabolized *in vitro*.

Some non-αT analogs, such as γT occur in relatively high amounts in adipose tissue, muscle and skin, suggesting passive partition into lipid-rich environments after post-prandial increase of these analogs in plasma or possibly to some extent also selective transport. In these tissues, γT may have regulatory effects, for example on inflammatory processes as it inhibits cyclooxygenase-2 (COX2) activity, leading to decreased prostaglandin E2 (PGE2) production. Likewise, in a mouse model for Parkinson's disease, γT prevented damage to dopaminergic neurons by treatment with 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) better when compared to αT. When vascular and lymphatic endothelial cells were exposed to physiological levels of αT, γT, or δT (0–40 μM), no toxicity was observed but δT was more potent to limit cell behaviors related to inflammation/angiogenesis.

Intestinal epithelial cells may be more exposed to the different vitamin E analogs present in the diet. In adult and fetal-derived intestinal epithelial cell lines, the peroxyl radical-induced membrane oxidation and inflammatory response were prevented by the vitamin E analogs in the order δT > γT > αT. Interestingly, at concentrations of 50–100 μM apoptosis-induced cytotoxicity was increased by δT and γT, whereas αT was not cytotoxic. These results point out an important role of the non-essential dietary vitamin E analogs during their absorption by the intestine, at least until they become metabolized and eliminated by the liver.

Vitamin E transport to the brain involves recognition of HDL by SR-BI expressed in the microvascular endothelial cells of the BBB (**Fig. 4**) (Lee and Ulatowski, 2019). Vitamin E is incorporated into HDL either by exchange from LDL by the phospholipid-transport protein and cholesterol-ester transport proteins (PL-TP, CE-TP, respectively), or to a lower degree by transport from cells via ABC transporters A1 and G1 (ABCA1/G1). It is likely that the preference of the BBB for HDL limits the access of the non-αT analogs and non-natural stereoisomers to the brain, as found in infant rhesus macaques, rats as well as humans. Accordingly, during mouse neurodevelopment, the brain preferentially acquires the *2R*-αT over *2S*-αT stereoisomers (Traber, 2020).

Vitamin E secretion by endometrial glands and transplacental transport are important for proper development of the embryo and fetus and may involve selective recognition of *RRR*-αT by αTTP and TAP1/SEC14L2 that are expressed at the placental interface. By comparing vitamin E levels in maternal and umbilical cord plasma, preferential transport of natural *RRR*-αT over synthetic *S*-αT was observed, albeit measurable levels of *S*-αT were also detected.

In human milk, *RRR*-αT is the predominant stereoisomer suggesting selective transport and secretion by epithelial cells of the mammary glands. In fact, several vitamin E transport and metabolic genes (αTTP, TAP1/SEC14L2, CYP4F2) are expressed in bovine as well as human mammary gland tissues and are possibly involved in selective transport to milk to provide vitamin E to the newborn. Whether a similar selective transport occurs in other glands (e.g., sebaceous, salivary or lacrimal glands) needs to be determined.

Vitamin E deficiency

Mutations in a number of genes involved in vitamin E uptake and distribution leading to defective transport of αT are responsible for primary and secondary vitamin E deficiency disorders in humans. Vitamin E plasma levels below ~12 μM are regarded as deficient, with an average normal level of about 23 μM. In humans severe vitamin E deficiency with plasma αT concentrations below ~2 μM and full neurological symptoms is a rare familial disease usually resulting from mutations of the α-tocopherol transfer gene (αTTP) leading to ataxia with vitamin E deficiency (AVED). The symptoms in these patients include ataxia, peripheral neuropathy, dysarthria, vibratory and proprioceptive sensory loss, loss of neurons, retinal atrophy, massive accumulation of lipofuscin in neurons and retinitis pigmentosa. Supplementation with high doses of vitamin E (up to 2000 mg/day) can prevent the symptoms of AVED.

More frequent than severe vitamin E deficiency is vitamin E insufficiency as result of suboptimal dietary vitamin E supply and inefficient uptake and distribution of vitamin E. A low efficiency of vitamin E uptake can occur with certain diseases, like cystic fibrosis, abetalipoproteinemia, chronic cholestatic liver disease, short-bowel syndrome, chronic pancreatitis, progressive systemic sclerosis, or several other lipid malabsorption syndromes as well as patients with total parenteral nutrition. In these cases, the symptoms can be prevented/reversed by supplemental vitamin E.

The human plasma average αT concentrations (22–28 μM) are about 10–100 times higher than that of the not-retained analogs (γT (~2.5 μM), δT (~0.3 μM), tocotrienols (~1–5 μM)). It can be assumed that at these nanomolar and low micromolar concentrations the "non-essential" vitamin E analogs either support the function of αT or exert regulatory effects on their own. Vitamin E levels in plasma may not be the only relevant measure for vitamin E adequacy, since the main biological function is thought to occur in tissues (Traber, 2014). The tissues with the highest αT contents are adipose tissue (150 μg/g tissue) and the adrenal glands (132 μg/g tissue), other organs (e.g., brain, kidney, heart or liver) contain between 7 and 40 μg/g tissues, whereas erythrocytes contain only 2 μg/g tissue. *RRR*-αT is usually the analog with the highest concentration in tissues, with some exceptions such as skin, adipose tissue and muscle where γT can also be high (reviewed in (Zingg and Azzi, 2004)). It remains to be elucidated whether these differences in the relative amounts of the different tocopherols are the result of tissue specific mechanisms for enrichment and/or storage and are involved in a tissue-specific regulatory functions.

Molecular mechanisms of vitamin E action

Over the last decades, both, the essential function of vitamin E (αT) as well as non-essential, associated functions of the other vitamin E analogs has been extensively studied. However, most studies addressing the molecular mechanisms of action have been done in cell culture models and it remains to be resolved whether they represent physiological events occurring also *in vivo* and to what degree they are responsible for the essentiality of vitamin E. In cultured cells and at equal concentration, not all of the regulatory effects can be explained by an antioxidant action of the eight vitamin E analogs and alternative, non-antioxidant roles have been proposed (reviewed in (Zingg and Azzi, 2004)). Differential regulatory effect of each vitamin E analog on signaling and gene expression are believed to be the consequence of specific binding of each vitamin E analog with different strenght to enzymes, transport and structural proteins, and transcription factors and/or result from vitamin E induced alterations of physical and structural properties of membrane lipid domains in which it is embedded (reviewed (Zingg, 2019)). Activity differences between natural vitamin E analogs observed in animal and human studies indicate that their regulatory effects on signaling and gene expression observed in cells also play a role *in vivo*.

The activity of several signal transduction enzymes is affected in a vitamin E analog specific manner leading to modulation of gene expression and cellular behavior, including proliferation, apoptosis, ferroptosis, auto-phagocytosis, survival, secretion, adhesion, migration, inflammation, immunity, senescence, tumorigenesis, metastasis and differentiation. In part as a consequence of modulating signaling enzymes, the activity of a number of transcription factors is modulated in a vitamin E analog specific manner (reviewed by (Rimbach et al., 2010)). Entire regulatory networks are affected by vitamin E suggesting that in many cases the observed regulatory effects may only represent secondary events, and a primary site for vitamin E signaling still needs to be identified. A selection of molecular mechanisms and targets by which the vitamin E analogs may affect signal transduction and gene expression is summarized in **Table 1** (reviewed in (Galli et al., 2016; Rimbach et al., 2010; Mocchegiani et al., 2014; Jiang, 2017; Zingg, 2019)).

Gene expression arrays, proteomics, metabolomics and lipidomics

With the advancement of novel techniques such as genome-wide gene expression arrays, RNAseq, proteomics, metabolomics and lipidomics new insights into the regulatory effects of vitamin E and its analogs have been achieved. Whereas initially it was thought that natural *RRR*-αT and synthetic *all-rac*-αT essentially share an identical transcriptional activity, e.g., when assessed in cultured HepG2 cells, a number of recent studies indicate that some differences can be detected *in vivo* in particular when genome-wide regulatory effects are assessed (reviewed in (Kim and Han, 2019)). Differential expression of a number of genes was observed in mouse T-cells supplemented with *RRR*-αTA and synthetic *all-rac*-αTA as well as *RRR*-αTA and *RRR*-γTA.

Table 1 Molecular mechanisms by which vitamin E modulates signal transduction and gene expression.

Chemically scavenging reactive oxygen and nitrogen species (ROS and RNS, respectively), and prevention of peroxidation of lipids (PUFA, MUFA) in membranes and of their active lipid mediators
Redox regulation of signal transduction enzymes (e.g., protein kinase C (PKC) isoforms, protein phosphatase 2A (PP2A), protein tyrosine phosphatase 1B (PTP1B)), and transcription factors (e.g., nuclear factor erythroid-derived 2-like 2 (NRF2), nuclear factor kappa B (NFκB))
Changing plasma membrane properties, such as membrane stability, curvature, fluidity, permeability, composition, membrane microdomains (lipid rafts/non-rafts), modulation of Ca^{2+} influx and Ca^{2+}-mediated signaling, membrane repair
Direct binding to and modulation of enzymes involved in signal transduction and in production of active lipid mediators (e.g., protein kinase C alpha (PKCα), phospholipase A2 (PLA2), 5-, 12-, 15-lipoxygenase (5-, 12-, 15-LOX), cyclooxygenase 1 and 2 (COX-1, COX-2))
Direct binding to transport proteins involved in signal transduction (e.g., CD36/FAT scavenger receptor/fatty acids transporter, tocopherol associated proteins 1, 2 and 3 (TAP1/SPF/SEC14L2, TAP2/SEC14L3, TAP3/SEC14L4, respectively), α-tocopherol transfer protein (αTTP))
Modulation of membrane-protein interaction and protein translocation to the plasma membrane (e.g., protein kinase C alpha (PKCα), pleckstrin homology domain leucine-rich repeat protein phosphatase, isoform 1 (PHLPP1), diacylglycerol kinase alpha (DAGKα), phosphatidylinositol-3-kinases alpha and gamma (PI3Kα, PI3Kγ), protein kinase B (PKB/Akt), phospholipase A2 (PLA2), phospholipase C (PLC))
Modulation of expression and cell surface exposition and function of membrane receptors such as CD36 and SR-BI scavenger receptor expression
Modulation of transport and conversion of lipids to signaling mediators (DHA), competition with uptake of hydrophobic vitamins A, D, and K and modulation of their regulatory function, modulation of mitochondrial fatty acids desaturases
Modulation of the activity of a number of transcription factors by various mechanisms (e.g., peroxisome proliferator-activated receptor gamma (PPARγ), nuclear factor erythroid-derived 2-like 2 (NRF2), nuclear factor kappa B (NFκB), signal transducer and activator of transcription factor 3 (STAT3), RAR-related orphan receptor alpha (RORα), hypoxia-inducible factor 1 alpha (Hif1α), estrogen receptor beta (ERβ), pregnane X receptor (PXR))
Conversion of vitamin E to active metabolites and lipid mediators (e.g., carboxyethylhydroxychromans (CEHC) and long-chain CEHC precursors (LC-CEHC), α-tocopherol quinone (αTQ), α-tocopherol hydroquinone (αTHQ), 5-NO_2-γT), α-tocopheryl phosphate (αTP)
Modulation of cell proliferation, inflammation and apoptosis in cancer cells by modulating the activity of the MAPK and PI3K/Akt signaling pathways and of cyclooxygenase 1 and 2 (COX-1, COX-2) and of 5-lipoxygenase (5-LOX)
Modulation of *de novo* cholesterol synthesis by inhibition of β-hydroxy-β-methylglutaryl coenzyme A (HMG-CoA) reductase by different vitamin E analogs (tocotrienols > tocopherols)
Conversion of cytochrome c into a peroxidase by α-tocopheryl phosphate, induction of free radicals production and modulation of signaling

These genome-wide regulatory effects of different vitamin E analogs on gene expression may also lead to changes in the production of metabolites as detected by metabolomics and lipidomics, although direct interaction of vitamin E with metabolic enzymes may also contribute. Using non-targeted metabolomics, transcription factors involved in energy metabolism (PGC1α, PPARα, PPARγ) were lower and CD36/FAT was higher in liver of vitamin E deficient rats, possibly leading to the observed decrease of glucose and increase in creatine, phosphocholine, and betaine. Similarly, in vitamin E deficient zebrafish (*Danio rerio*), energy metabolism and cognitive functions were dysregulated with increased oxidative stress resulting in elevated PUFA oxidation (especially DHA), altered brain phospholipids and lysophospholipids composition, mitochondrial dysfunction, depleted choline/methyl-donor and glucose levels and altered thiols and amino acids, and lipidomics suggested protective antioxidant effects of vitamin E. In humans, metabolomic profiling revealed increased levels of a number of LPC species in human plasma after vitamin E supplementation (400 mg/d *RRR*-αTA, for 4 wks) (Traber, 2020).

Transcriptomic profiling (RNAseq) identified over 1000 differentially expressed genes in SR-BI deficient mice with and without failure to close the neural tube during embryonic development. SR-BI plays a role in transport of vitamin E across the BBB and across the maternal-fetal interface, and mouse embryos lacking SR-BI fail to close the neural tube and show cephalic neural tube defects that can be prevented by vitamin E supplementation. In this study, using gene regulatory network analysis, the androgen receptor was identified as a candidate transcription factor regulating genes associated with the expression of several genes involved in neural development and lipid metabolism. Similarly, single-cell RNA-seq revealed changes in signaling in mechanosensitive dorsal root ganglion neurons with vitamin E deficiency involving phospholipase C (PLC), protein kinase C (PKC), diacylglycerol kinase (DAGK), inositol-3-phosphate (IP3), diacylglycerol, and Ca^{2+} that may explain the symptoms of sensory neuropathy.

Vitamin E and disease prevention

In addition to acting as an essential molecule for the prevention of diseases caused with vitamin E deficiency, animal models as well as humans studies have assessed the possible preventive effects of supplementation with vitamin E and its analogs against several diseases including atherosclerosis, non-alcoholic steatohepatitis (NASH), certain types of cancer, inflammation, respiratory infections (influenza, pneumonia), asthma, allergy, fibrotic diseases, diseases of the eye and neurodegenerative diseases such as Alzheimer's or Parkinson's disease (reviewed in (Hathcock et al., 2005; Jiang, 2017)). These studies are based on the observation that intake of vitamin E is generally below the RDA (Traber, 2014) and that vitamin E levels can be increased by extra dietary supplementation with no adverse effects (up to a certain level, e.g., 200–400 mg/day). However, whereas the beneficial effect of vitamin E have been mostly confirmed in animal studies, in humans the situation is less clear. Epidemiological studies indicating preventive effects of vitamin E against cardiovascular events, neurodegenerative disease, macular degeneration, and cancer were in general not confirmed by larger clinical intervention studies, with to date best evidence for a lower risk for diseases such as NASH, CVD, inflammation and immune-modulation in the elderly and prevention of infection, but with a higher risk for prostate cancer (reviewed in (Borel and Desmarchelier, 2016; Mocchegiani et al., 2014; Sozen et al., 2019; Nagashimada and Ota, 2019; Lewis et al., 2019)). In some meta-analyses of clinical studies, increased all-cause mortality was observed with higher doses of vitamin E supplementation, although this was not confirmed in later studies. Several factors may explain the often mixed outcome of these vitamin E supplementation studies, including the high levels of vitamin E already present at baseline, the various doses and duration of supplementation, the diet composition and presence of phytochemicals, micronutrients and other vitamin E analogs, environmental and patho-physiological circumstances that affect the level and action of vitamin E to various degree such as inflammation, infection, smoking or UV irradiation, and most importantly the presence of specific vitamin E related polymorphisms in the analyzed population.

Vitamin analogs that do not reach high concentrations in plasma may reach over time sufficient amounts in peripheral tissues or they may be converted to metabolites able to reduce the risk for diseases. In particular, these analogs may reach higher levels in the gastrointestinal tract and prevent intestinal inflammatory, autoimmune and allergic events or colon cancer. Moreover, these vitamin E analogs could modulate cells in the proximity of tissues that interface with the external environment (the epithelial surface of the skin, the gastrointestinal mucosa, the eye, and the respiratory system). As an example, the tocotrienols (TT) decreased scratching behavior, dermal thickening, and the serum histamine in a mouse model of atopic dermatitis. In IgE-sensitized and dinitrophenyl-BSA (DNP-BSA) stimulated rat RBL-2H3 mast cells, γTT and δTT suppressed degranulation via inhibition of protein kinase C (PKC). In a model for canine atopic dermatitis with mastoparan-stimulated canine mastocytoma cells (C2), vitamin E (*RRR*-αT, 100 μM) inhibited histamine, prostaglandin D2 (PGD2), and chymase release but the signaling pathways involved have not been resolved in detail.

Polymorphisms relevant for vitamin E action

Polymorphisms in genes involved in uptake and distribution, oxidation, metabolism and action may contribute to the variability of individuals to respond to vitamin E and explain the often mixed outcomes of supplemenation studies (**Fig. 4**) (Zingg et al., 2008; Mocchegiani et al., 2014; Borel and Desmarchelier, 2016). Among the genes relevant for vitamin E uptake and distribution, polymorphisms in αTTP, TAP, CD36/FAT, SR-BI and ABCA1/G1 have been identifed that influence the plasma and possibly tissue levels of vitamin E. Polymorphisms in apolipoprotein E (apoE) that are well known to influence the plasma lipid profile and inflammatory cytokine expression have been associated with higher levels of vitamin E in plasma but lower levels in tissues. Vitamin E

metabolic genes, such as LPL, MRP2, PXR, CYP3A4 and CYP4F2 are highly polymorphic (100 bp/coding SNP) and likely to influence inter-individual variation of vitamin E levels.

Some of these polymorphic genes (e.g., TAP1/2/3, CD36, PXR, PPARγ, SREBP2) play also a role in modulating signal transduction and gene expression and thus may affect the cellular responsiveness to vitamin E. The differential responsiveness to the anti-inflammatory effects of vitamin E has been associated with polymorphisms detected in cytokines genes. However, only few polymorphisms have been directly associated with regulatory effects of vitamin E and it remains to be further elucidated to what degree the presence of these polymorphisms translate into altered risks for diseases such as atherosclerosis, cancer and metabolic syndrome.

The elevated risk for prostate cancer observed in the SELECT study upon vitamin E supplementation has been associated with polymorphisms in αTTP or TAP1, most likely as result of regulatory effects of these genes on vitamin E levels as well as on signaling and gene expression in the prostate. Alternatively, the elevated risk in response to vitamin E can be explained by modulating the biosynthesis of cholesterol/steroids by supernatant protein factor (SPF, also named TAP1) able to regulate squalene epoxidase or HMG-CoA reductase activity. Increased oxidative damage as result of activation of phase I activating cytochrome P_{450} (CYP) enzymes involved in metabolism of polycyclic aromatic hydrocarbons (PAH) such as benzo[a]pyrene by high doses of vitamin E (*all-rac*-αTA) may also contribute to an elevated risk in the prostate.

The best studied polymorphism relevant for the bioavailability and bioactivity of vitamin E has been identifed in the haptoglobin (Hp) gene (Somer and Levy, 2020). There are two Hp alleles, the Hp1 and Hp2, leading to 3 genotypes in the population, Hp1-1, Hp1-2, and Hp2-2 (estimated occurrence 15–18%, 46% and 38%, respectively). Haptoglobin is an important plasma protein relevant for the binding and clearing of free hemoglobin (Hb), a toxic oxidant molecule released during intravascular destruction of erythrocytes. Diabetic patients with the Hp2-2 allele have an up to five times higher risk for cardiovascular disease (CVD) since clearance of the Hp2-2/Hb complex by CD163 receptor on monocytes/macrophages is impaired resulting in increased oxidative stress and reduced plasma vitamin E and C levels. In fact, as revealed by retrospective analysis of the HOPE and WHS studies as well as of the prospective ICARE study, vitamin E-supplemented diabetic subjects with Hp2-2 may have a reduced risk for CVD complications (Somer and Levy, 2020). Thus, albeit the number of patients analyzed in theses studies is relatively small, at least in diabetic patients the Hp2-2 polymorphism may be an important determinant for the preventive effects of vitamin E against CVD complications. Interestingly, reduction of oxidative stress may not be the only explanation for the preventive effects since the combination of vitamin E (αT) with vitamin C (L-ascorbic acid) did not increase or even mitigated the protective effect of αT on HDL oxidation.

The importance of certain SNPs on vitamin E status has been revealed in several candidate gene association studies (CGAS) and genome-wide association studies (GWAS) (reviewed in (Borel and Desmarchelier, 2016)). However, the readouts from these studies may be limited and only the most relevant gene variants may be detected, since often more than one SNP may be responsible and other regulatory mechanisms can have an influence (e.g., epigenetics, copy number variants, insertions/deletions, diet composition). Nevertheless, such studies may identify sub-populations of individuals at risk for low vitamin E uptake and molecular action that can be corrected using a personalized approach of vitamin E supplementation as a strategy to increase the benefits of vitamin E, as exemplified with diabetic subjects and the Hp2-2 allotype.

Summary and conclusion

The diet contains eight natural vitamin E analogs but only one of them (αT) has been selected during evolution of higher organisms to become essential. After a transient post-prandial peak, the other seven vitamin E analogs are metabolized and eliminated, but numerous *in vitro* and *in vivo* studies indicate that they can perform important vitamin E associated regulatory functions. Adequate levels of vitamin E in the body are maintained by first allowing access to all eight analogs by the intestinal epithelium and then by selective retention of mainly *RRR*-αT by the liver, and by specific metabolism of all the other tocopherols, tocotrienols, and of excess αT. At a molecular level, natural and synthetic vitamin E analogs differently interact with membranes, enzymes and structural proteins and differently affect signaling and gene expression. Vitamin E analog specific cellular responses may also be the consequence of different efficiency of their cellular uptake and transport, intracellular distribution by tocopherol binding proteins, metabolism and secretion. Preferential transport across cellular barriers of *RRR*-αT is particularly relevant in the placenta, mammary glands and the brain. In that respect the brain represents a special organ, as it is separated by microvascular cells of the BBB with preferential transport of natural vitamin E. In the brain as well as in other organs, the balance between vitamin E with PUFA and their interactions are important during development and for the prevention of diseases associated with low vitamin E levels. Polymorphisms in genes related to vitamin E transport and action have been identified and may help to identify sub-populations of individuals at risk for low vitamin E status that can be corrected using a personalized approach of vitamin E supplementation.

References

Azzi, A., 2018. Many tocopherols, one vitamin E. Mol. Aspect. Med. 61, 92–103.

Birringer, M., Lorkowski, S., 2019. Vitamin E: regulatory role of metabolites. IUBMB Life 71, 479–486.

Borel, P., Desmarchelier, C., 2016. Genetic variations involved in vitamin E status. Int. J. Mol. Sci. 17.

Dipasquale, M., Gbadamosi, O., Nguyen, M.H.L., Castillo, S.R., Rickeard, B.W., Kelley, E.G., Nagao, M., Marquardt, D., 2020. A mechanical mechanism for vitamin E acetate in E-cigarette/vaping-associated lung injury. Chem. Res. Toxicol. 33, 2432–2440.

Food And Nutrition Board, I. O. M, 2000. Vitamin E. In: Press, N.A. (Ed.), Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids (2000). Institute of Medicine (IOM), Washington, D.C.

Galli, F., Azzi, A., Birringer, M., Cook-Mills, J.M., Eggersdorfer, M., Frank, J., Cruciani, G., Lorkowski, S., Ozer, N.K., 2016. Vitamin E: emerging aspects and new directions. Free Radic. Biol. Med. 102, 16–36.

Hathcock, J.N., Azzi, A., Blumberg, J., Bray, T., Dickinson, A., Frei, B., Jialal, I., Johnston, C.S., Kelly, F.J., Kraemer, K., Packer, L., Parthasarathy, S., Sies, H., Traber, M.G., 2005. Vitamins E and C are safe across a broad range of intakes. Am. J. Clin. Nutr. 81, 736–745.

Jiang, Q., 2014. Natural forms of vitamin E: metabolism, antioxidant, and anti-inflammatory activities and their role in disease prevention and therapy. Free Radic. Biol. Med. 72, 76–90.

Jiang, Q., 2017. Natural forms of vitamin E as effective agents for cancer prevention and therapy. Adv. Nutr. 8, 850–867.

Kim, H.K., Han, S.N., 2019. Vitamin E: regulatory role on gene and protein expression and metabolomics profiles. IUBMB Life 71, 442–455.

Kono, N., Arai, H., 2015. Intracellular transport of fat-soluble vitamins A and E. Traffic 16, 19–34.

Lee, P., Ulatowski, L.M., 2019. Vitamin E: mechanism of transport and regulation in the CNS. IUBMB Life 71, 424–429.

Lewis, E.D., Meydani, S.N., Wu, D., 2019. Regulatory role of vitamin E in the immune system and inflammation. IUBMB Life 71, 487–494.

Mene-Saffrane, L., 2017. Vitamin E biosynthesis and its regulation in plants. Antioxidants 7.

Miyazawa, T., Burdeos, G.C., Itaya, M., Nakagawa, K., Miyazawa, T., 2019. Vitamin E: regulatory redox interactions. IUBMB Life 71, 430–441.

Mocchegiani, E., Costarelli, L., Giacconi, R., Malavolta, M., Basso, A., Piacenza, F., Ostan, R., Cevenini, E., Gonos, E.S., Franceschi, C., Monti, D., 2014. Vitamin E-gene interactions in aging and inflammatory age-related diseases: implications for treatment. A systematic review. Ageing Res. Rev. 14, 81–101.

Nagashimada, M., Ota, T., 2019. Role of vitamin E in nonalcoholic fatty liver disease. IUBMB Life 71, 516–522.

Neuzil, J., 2002. Alpha-tocopheryl succinate epitomizes a compound with a shift in biological activity due to pro-vitamin-to-vitamin conversion. Biochem. Biophys. Res. Commun. 293, 1309–1313.

Raederstorff, D., Wyss, A., Calder, P.C., Weber, P., Eggersdorfer, M., 2015. Vitamin E function and requirements in relation to PUFA. Br. J. Nutr. 114, 1113–1122.

Reboul, E., 2018. Vitamin E intestinal absorption: regulation of membrane transport across the enterocyte. IUBMB Life 71 (4).

Rimbach, G., Moehring, J., Huebbe, P., Lodge, J.K., 2010. Gene-regulatory activity of alpha-tocopherol. Molecules 15, 1746–1761.

Somer, S., Levy, A.P., 2020. The role of haptoglobin polymorphism in cardiovascular disease in the setting of diabetes. Int. J. Mol. Sci. 22.

Sozen, E., Demirel, T., Ozer, N.K., 2019. Vitamin E: regulatory role in the cardiovascular system. IUBMB Life 71, 507–515.

Traber, M.G., 2004. Vitamin E, nuclear receptors and xenobiotic metabolism. Arch. Biochem. Biophys. 423, 6–11.

Traber, M.G., 2014. Vitamin E inadequacy in humans: causes and consequences. Adv. Nutr. 5, 503–514.

Traber, M.G., 2020. Brain-E, does it equate to brainy? J. Nutr. 150, 3049–3050.

Wagner, K.H., Kamal-Eldin, A., Elmadfa, I., 2004. Gamma-tocopherol–an underestimated vitamin? Ann. Nutr. Metab. 48, 169–188.

Zingg, J.M., Azzi, A., 2004. Non-antioxidant activities of vitamin E. Cur. Med. Chem. 11, 1113–1133.

Zingg, J.M., Meydani, M., 2019. Interaction between vitamin E and polyunsaturated fatty acids. In: Weber, E.A. (Ed.), Vitamin E in Human Health. Springer Nature Switzerland AG, Switzerland.

Zingg, J.M., Azzi, A., Meydani, M., 2008. Genetic polymorphisms as determinants for disease-preventive effects of vitamin E. Nutr. Rev. 66, 406–414.

Zingg, J.M., 2019. Vitamin E: regulatory role on signal transduction. IUBMB Life 71, 456–478.

Vitamin K

M.K. Shea and S.L. Booth, USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of X. Fu, S.L. Booth, Vitamin K, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 398–403, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00281-6.

Introduction 569
Forms 570
Sources 570
Absorption and metabolism 571
Functions 572
Enzymatic 572
Non-enzymatic 573
Vitamin K dietary requirements 574
Vitamin K status, health & disease 574
Vitamin K status biomarkers 574
Vitamin K status in the newborn 574
Vitamin K status and chronic disease 575
Vitamin K and the gut microbiome 575
Conclusion 575
References 576

Key points

- Vitamin K is a term that includes phylloquinone and 10 menaquinone forms.
- Whereas phylloquinone is well characterized in the food supply, menaquinones have not been systematically analyzed, which limits their use in dietary assessment.
- Phylloquinone absorption is lipid-driven whereas less is known about absorption of menaquinones.
- All vitamin K forms serve as an enzyme cofactor and support gamma-carboxylation of vitamin K-dependent coagulation proteins. Multiple vitamin K-dependent extra-hepatic proteins exist but less is known about their role in human health.
- Non-enzymatic roles for vitamin K are emerging but the research is still in its infancy.
- Initial associations between vitamin K status and chronic diseases such as osteoporosis and cardiovascular disease have not been substantiated in randomized clinical trials.
- Because there are major gaps in knowledge regarding the relative intake and metabolism of the multiple dietary vitamin K forms, current dietary recommendations are approximations. It is unlikely that gut-derived menaquinones substantially contribute to the dietary requirement.

Introduction

Vitamin K is a family of fat-soluble vitamers that share a common chemical structure and a common function. The discovery of vitamin K dates back to 1929, when Danish biochemist Henrik Dam observed hemorrhaging in chicks that were fed a diet free of fat and cholesterol. Dam deduced that removing lipid from the diet also removed an essential fat-soluble compound, which he termed "vitamin K" for koagulation (as spelled in German and Scandinavian languages). The structure of vitamin K was confirmed nearly a decade later by American biochemist, Edward Doisy, and in 1943, Dam and Doisy shared the Nobel Prize in Physiology or Medicine for their work in vitamin K (Suttie, 2009).

Of all the fat-soluble vitamins, vitamin K is the least studied in terms of its role in human nutrition. In part, this is driven by the assumption that vitamin K's role is to support normal coagulation, and among the general adult population, vitamin K-related coagulation disorders are rare. Until recently, many assumed that the human requirement for vitamin K was partially met by production of vitamin K by gut bacteria, hence was not considered a shortfall nutrient. There is no single validated biomarker for measuring overall vitamin K status, which has given rise to considerable confusion in the scientific literature regarding the interpretation of observational studies examining the associations between vitamin K status and health outcomes beyond that of normal coagulation. The purpose of this review is to provide an overview of vitamin K based on current scientific knowledge, its sources, biological functions, and role in human health and disease.

Forms

The term vitamin K refers to 11 known forms. Each of the 11 naturally occurring forms of vitamin K have a common 2-methyl-1,4-naphthoquinone ring with a prenylated side chain. It is the length and saturation of the side chain that differs among the forms. Phylloquinone (historically referred to as vitamin K1) has a side chain containing four isoprenoid units, three of which are saturated. Menaquinones, of which there are at least 10 naturally occurring forms, have side chains with four to thirteen isoprenoid units, which are mostly unsaturated. The number of isoprenoid units in the side chain differentiates the menaquinone forms. For example, menaquinone-7 (MK7) has a side chain containing 7 isoprenoid units, whereas the side chain of menaquinone-4 (MK4) has 4 units (**Fig. 1**). Menaquinones are collectively referred to as vitamin K2, which has fostered some confusion in the literature because not all menaquinones are the same in terms of origin and function. Therefore, for the purpose of this review, we will only refer to the individual menaquinones to avoid further confusion. Menadione (historically known as vitamin K3) is a synthetic vitamin K precursor comprised of the naphthoquinone ring without a prenylated side chain.

Sources

Phylloquinone is found in all photosynthetic plants, serving as an electron acceptor during photosynthesis. Its concentration is highly correlated with chlorophyll, which confers the green color of leaves. Hence, primary dietary sources of phylloquinone are green leafy vegetables and vegetable oils. Mixed dishes that are made with certain vegetable oils, such as soybean, canola or rapeseed, and olive, also contain appreciable amounts of phylloquinone (**Tables 1** and **2**) (Harshman et al., 2017). Most menaquinones are produced by bacteria, and in the food supply, MKs are found primarily in fermented foods. Because different bacteria can produce different MKs, the MK contents of fermented foods depend on the bacteria used in the fermentation process. For example, natto, a fermented soy food commonly consumed in some regions of Japan, is high in MK7 because the bacteria natto bacillus produces MK7. However not all bacteria produce menaquinones so it cannot be assumed that all fermented foods are rich in menaquinone content. MKs are found in some dairy foods and meats, in highly variable amounts that are also dependent on the fermentation process. MK contents of dairy foods are influenced by fat content of the food, with low fat dairy containing lower amounts of MKs. Being bacterially synthesized, MKs are also abundant in the gut microbiome, albeit in highly variable amounts. MK4 is unique among the MKs. Although it is produced in trace amounts by bacteria, it is found mainly in poultry and pork because it is synthesized from menadione, which is abundant in animal feed in certain geographical regions.

The USDA Food Data Central provides the phylloquinone contents of >40,000 foods in the US food supply. The database also provides the MK4 content of ~1200 foods, with more being incorporated. Food composition databases in Australia and the Netherlands include phylloquinone and menaquinones. Some databases, such as the Swedish and Finnish food composition databases,

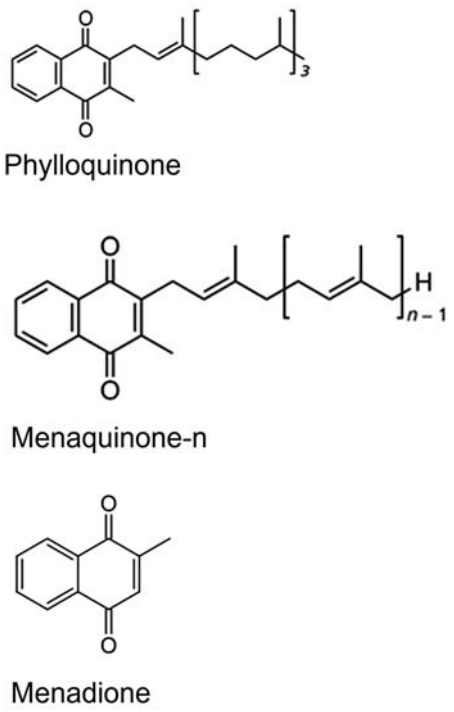


Fig. 1 Forms of vitamin K.

Table 1 Vitamin K contents of foods.

Vitamin K form	*Food*	*Amount per serving*
Phylloquinone	Collard greens, raw	157 μg/cup
	Spinach, raw	145 μg/cup
	Romaine lettuce	48 μg/cup
	Green peas, raw	36 μg/cup
	Soybean oil	26 μg/tablespoon
	Mayonnaise	23 μg/tablespoon
	Lasagna, cheese	21 μg/cup
	Olive oil	8.4 μg/tablespoon
	Stir fried rice, meatless	4.7 μg/cup
	Safflower oil	1 μg/tablespoon
	Corn oil	0.3 μg/tablespoon
Menaquinone-4	Pepperoni	35.4 μg/3 oz
	Hot dog, 5 inch	17.1 μg/hot dog
	Beef, cooked	0.17 μg/3 oz
	Chicken, cooked	35 μg/3 oz
	Cheddar cheese	11 μg/cup
	Cream cheese	1.3 μg/tablespoon
Menaquinone-7	Natto	360 μg/40 g

list vitamin K values without specification as to the individual vitamin K forms being measured. However, caution must be taken when comparing the phylloquinone and menaquinone contents of foods across databases because different analytical methods and sampling plans have been used for quantification. Moreover, since the MK content of the food supply is not yet well characterized, and food composition databases that differentiate the contents of individual MKs (other than MK4) are lacking, studies reporting intakes of specific MKs should be interpreted carefully.

Absorption and metabolism

Phylloquinone is the major circulating form of vitamin K and has large inter- and intra-individual variances in response to intake. The factors contributing to this variance in circulating phylloquinone are still not well-characterized, although recent advances indicate that multiple lipid-related factors influence this large variability. Phylloquinone is absorbed from the diet following pathways of most dietary lipids, involving bile acids, pancreatic juices, mixed micelle formation, and uptake into enterocytes (Shearer and Newman, 2008). Whereas it was historically assumed that phylloquinone had passive absorption, current evidence suggests the Niemann-Pick C1-like 1 (NPC1L1), scavenger receptor class B type I (SR-BI), and a cluster of differentiation 36 (CD36) proteins all could be involved in the active transport of phylloquinone into the enterocyte. In the enterocyte, phylloquinone is then packaged into triglyceride-rich lipoproteins (TRLs) with other lipids and is exported from the enterocyte to the lymph for entry into circulation at the thoracic duct (Shearer and Newman, 2008).

Table 2 Characterized vitamin K dependent proteins.

Protein	*Primary action*	*Tissue location*
Prothrombin (factor II)	Pro-coagulant liver	Liver
Factor VII	Pro-coagulant liver	Liver
Factor IX	Pro-coagulant liver	Liver
Factor X	Pro-coagulant liver	Liver
Protein C	Anti-coagulant liver	Liver
Protein S	Anti-coagulant liver	Liver
Protein Z	Anti-coagulant liver	Liver
Osteocalcin	Bone mineralization	Bone
Matrix gla protein	Calcification inhibitor	Bone, cartilage, other soft tissues
Gla-rich proteins	Calcification regulator	Universal
Gas 6	Ligand for tyrosine kinase Axl	Universal

Phylloquinone is predominantly transported with TRLs in circulation (Ellis et al., 2019), and its absorption seems to be a predominantly lipid-driven effect and not dependent on existing vitamin K status or non-dietary factors, such as age, sex, BMI, or percent body fat (Ellis et al., 2019). Through use of stable isotope tracers, it has been shown that phylloquinone absorption from food varies, in part, according to the meal composition or food matrix. Consistent with absorption of other fat-soluble vitamins, there is higher absorption of phylloquinone when consumed with fats. However, unlike other fat-soluble vitamins that are stored in the body, vitamin K is rapidly excreted. Vitamin K is catabolized to 5-carbon and 7-carbon aglycones, which can be measured in urine.

Less is known about the absorption and transport of MKs. In contrast to phylloquinone, menaquinones are not typically detected in circulation. Absorption of intestinally-synthesized MKs is questionable because the colon, where gut bacteria produce MKs, lacks the bile salts needed for fat-soluble nutrient absorption. Dietary MKs are presumed to be absorbed following pathways of other lipid-soluble nutrients (Shearer and Newman, 2008) so this relative absence of MKs in circulation erroneously led the vitamin K field to assume that phylloquinone was the predominant form in the diet. However, with new mass spectrometry techniques that have enabled more sensitive quantification of MKs in foods, it is now evident that menaquinones are as, if not more, abundant than phylloquinone in the human diet. This presents a conundrum in our understanding of vitamin K metabolism that has given rise to multiple competing theories regarding similarities and differences in the absorption and metabolism of the different vitamin K forms.

Reliance on the measurement of the parent form of vitamin K in circulation may result in an oversimplification in our understanding of vitamin K metabolism. There is also a lack of systematic terminology used in the literature so circulating measures of the parent vitamin form can be referred to as measures of absorption, transport, bioaccessibility and/or bioavailability, which further contributes to the confusion about the interpretation of the data. For example, based on the results of one study which compared post-prandial concentrations of phylloquinone and MK7 in blood of 15 healthy volunteers after ingestion of 1 mg of each vitamer dissolved in corn oil, in which MK7 had higher circulating concentrations compared to phylloquinone (Schurgers et al., 2007), some have claimed superior bioavailability of MK7. An alternative interpretation is that uptake of MK7 by tissues is relatively lower compared to phylloquinone. Unfortunately, no study has compared multiple forms of vitamin K in equimolar doses in humans, which limits the ability to interpret the relative absorption and bioavailability of individual vitamin K. Furthermore, stable isotope tracer studies have not yet been conducted in humans with menaquinones.

This concept of differences in absorption and transport and relative bioavailability among individual vitamin K forms has recently been challenged by the results of an animal experiment. Vitamin K tissue concentrations were measured in mice after they were fed a diet containing equimolar doses of isotopically labeled phylloquinone, MK4, MK7, or MK9 for one week. Surprisingly, this study found no differences in the absorption or transport of the different vitamin K forms studied. Furthermore, MK4 was the primary form of vitamin K in most tissues. As demonstrated through stable isotope technology, all the dietary forms of vitamin K were similarly converted to MK4, and the naphthoquinone ring of the MK4 originated from the dietary phylloquinone or menaquinone form. Except for the liver and intestine, in which the parent vitamin K form was predominant, MK4 was the only form of vitamin K detected in extra-hepatic tissues (Ellis et al., 2022). The conversion to MK4 involves side-chain cleavage, which liberates menadione, followed by transfer a geranylgeranyl group, via the prenyltransferase UBIAD1 in target tissues (Nakagawa et al., 2010). That labeled MK4 was detected in the intestine of mice fed all the forms of labeled vitamin K, including MK4, indicates MK4 biosynthesis begins in the intestine and side chain cleavage is not specific to any particular form of vitamin K. However, the exact mechanism underlying side chain cleavage remains to be identified. Until such studies are complete, our knowledge of menaquinone absorption and transport remains speculative.

Functions

Enzymatic

Vitamin K's main function is as an enzymatic co-factor for the post-translational carboxylation of certain γ-carboxyglutamic acid (gla)-containing proteins, commonly referred to as vitamin K dependent (VKD) proteins. A vitamin K cycle is responsible for VKD protein carboxylation (**Fig. 2**). Dietary vitamin K (the quinone form) is reduced by the vitamin K epoxide reductase (VKOR), which then recycles the oxidized 2,3-epoxide vitamin K, for the cycle to continue. The reduced form of vitamin K is required for the post-translational glutamate carboxylase (ggcx) (**Fig. 2**). The glutamate carboxylase is responsible for the addition of a second carboxyl group to the glutamate amino acid residues, resulting in protein functionality (Suttie, 2009).

Vitamin K antagonists (i.e., warfarin) target the VKOR enzyme, thus inhibiting vitamin K recycling, which reduces the amount of vitamin K available for carboxylation of coagulation factors II, VII, IX, X and protein C and S (Suttie, 2009). A second gene, VKORC1L1, has been found to support the activity of VKOR, respond to vitamin K antagonists, and is present in extra-hepatic tissues. However, VKORC1L1 is unique in that it appears to be 50-fold more resistant to the antagonists, which may explain the low susceptibility of extra-hepatic tissues to vitamin K antagonism and a concomitant lack of effects on vitamin K-dependent proteins.

While the most studied VKD proteins are clotting factors synthesized in the liver, vitamin K and VKD proteins are also present in extrahepatic tissues, including bone, cartilage, and vascular tissue. Osteocalcin, which is produced by osteoblasts during bone formation, is the predominate non-collagenous protein in bone and is vitamin K dependent. Once carboxylated, osteocalcin has a strong affinity for binding calcium in the bone matrix. However, the precise function of carboxylated osteocalcin in bone is

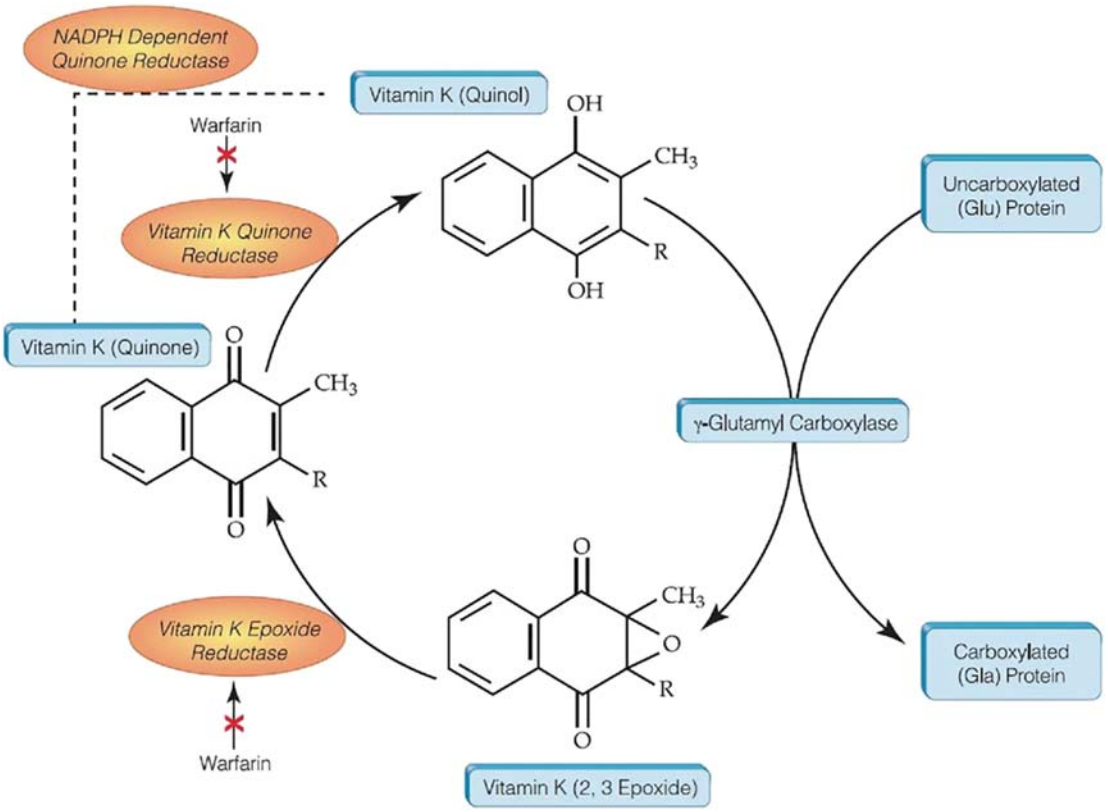


Fig. 2 Carboxylation of Vitamin K–dependent Proteins. The vitamin K hydroquinone is reduced to the vitamin K quinol, which serves as a cofactor to the γ-glutamyl carboxylase enzyme that carboxylates vitamin K–dependent proteins. As a result, the quinol is oxidized to vitamin K epoxide, which is reduced back to the quinone form (Shea and Booth, 2019).

unclear. Matrix gla protein (MGP) and Gla rich protein (GRP, also known as unique cartilage matrix-associated protein (UCMA)) are VKD proteins present in vascular tissue and cartilage, where they are involved in inhibiting mineralization. Additional VKD proteins have been identified in multiple extra-hepatic tissues (Suttie, 2009). However, little is known about the relevance of their carboxylation status because antibodies that differentiate the carboxylated from undercarboxylated forms of these proteins have not been developed.

Phylloquinone and menaquinones are capable of functioning as an enzyme co-factor, but the relative efficacy of the different vitamin K forms in carboxylation is not well understood (Shea et al., 2021). Some have proposed differentiating phylloquinone and menaquinone-dependent proteins. This proposition is based on phylloquinone being present in higher amounts in the liver, where clotting proteins are synthesized, while MK4 being more abundant in extra-hepatic tissues. However, menaquinones are also present in the liver (Suttie, 2009), and recent evidence indicates conversion of phylloquinone and different menaquinone forms to MK4 in the liver and extrahepatic tissues (Ellis et al., 2022). Therefore, the proposition that some VKD proteins require phylloquinone and others require menaquinones is not supported by current data.

Non-enzymatic

Vitamin K is reported to have anti-inflammatory properties via pathways independent of carboxylation. In cell-culture, pretreatment of macrophages with phylloquinone, MK4, and MK7 suppressed the LPS induced production of IL6. In a follow-up experiment that only tested MK4, this effect was found to be through the NFkB signaling pathway. In osteoblast and osteoclast precursor cells, MK7 treatment suppressed the production of pro-inflammatory cytokines, also via the NFkB signaling pathway. In microglia-derived cells, MK4 suppressed the LPS-induced production of IL6, TNFa, IL1B and Cox2 and NFkB nuclear translocation Microglia are involved in neuroinflammation and MK4 is the main form of vitamin K found in the brain.

MK4 is a ligand for the steroid and xenobiotic nuclear receptor (SXR) (also known as pregnane X receptor, PXR) and can regulate gene expression. In culture experiments using bone-derived cells, genes involved in bone formation and collagen accumulation (Ichikawa et al., 2006) were regulated by MK4 via SXR activation. In animal models, PXR deletion resulted in a phenotype consistent with osteopenia In this experiment, the PXR, which is highly expressed in the liver and intestine where it functions in xenobiotic metabolism, was deleted systemically, so it is possible systemic effects were involved. Together these observations suggest vitamin K

is involved in cell-signaling and gene expression, but whether this function of vitamin K is relevant to human health or disease remains to be determined.

Vitamin K dietary requirements

The current Adequate Intakes (AI) for vitamin K in North America are set at 90 μg/d and 120 μg/d for women and men respectively, 19 years and older. The European Food Safety Administration (EFSA) recommends 70 μg/d for all adults ≥18 years old. These recommendations are based on phylloquinone. In the US, current phylloquinone intakes approximate the AI, but the actual range of intake is wide and varies across age groups, geographic regions, and racial/ethnic groups (Harshman et al., 2017). On average, 43% of adult men and 63% of adult women consume the recommended AI for vitamin K, but the percentage of individuals meeting or exceeding the AI declines with age (Harshman et al., 2017). There are currently no dietary requirements specific to menaquinone intakes.

Vitamin K status, health & disease

Vitamin K status biomarkers

Nutritional status is commonly estimated using objectively measured biomarkers. In contrast to other micronutrients, there is not a single measure that is considered a gold-standard biomarker of vitamin K status. Instead, multiple measures are recommended to capture overall status, as reviewed in detail elsewhere (Shea and Booth, 2016). Phylloquinone, the main circulating form of vitamin K, changes in response to intake and serum/plasma phylloquinone has been used to rank individual's vitamin K status in population- and clinic-based studies globally (Shea and Booth, 2016). External quality assurance programs are available to standardize circulating phylloquinone assays and monitor inter-laboratory variation (Card et al., 2009), and indeed any laboratory measuring circulating phylloquinone should actively participate in this program to ensure accuracy of the reported measures. Menaquinones are not typically detected in circulation so have not been measured in many population-based studies. Pilot KEQAS schemes for MK4 and MK7 are available, which laboratories seeking to measure circulating MKs should participate in. Similar quality assurance programs for other MKs are not available, hence accuracy of these measures is uncertain.

The undercarboxylated fractions of three VKDPs are measurable in circulation. PIVKA (undercarboxylated prothrombin, or protein induced in vitamin K absence—factor II) increases in response to dietary vitamin K depletion or warfarin treatment. However, PIVKA does not generally reflect variation in vitamin K intakes in generally healthy individuals, which limits its utility as an indicator of vitamin K status in population-based studies. Serum undercarboxylated osteocalcin (ucOC) is considered a functional indicator of vitamin K status in bone and plasma dephospho-undercarboxylated MGP ((dp)ucMGP), a functional indicator of vitamin K status in vascular tissue or other tissues that use MGP. Both ucOC and (dp)ucMGP decrease in response to vitamin K supplementation and inversely correlate with circulating phylloquinone (Shea and Booth, 2016). In addition to vitamin K availability for carboxylation, the amount of undercarboxylated OC and MGP in circulation depends on synthesis of the protein. To reflect vitamin K status more accurately, ucOC and (dp)ucMGP should be corrected for the total amount of protein in circulation, which itself can vary based on factors independent of vitamin K . Validated assays are available to quantify total OC in circulation and ucOC is typically expressed as a ratio or percent of total OC (i.e., %ucOC). However, validated assays that quantify total circulating MGP are not commercially available, thereby limiting the use of (dp)ucMGP as a sole biomarker of vitamin K status (Shea and Booth, 2016).

Vitamin K status in the newborn

Infants have low vitamin K stores due to poor placental transfer of vitamin K, low concentrations of vitamin K in breastmilk, and possibly insufficient vitamin K from gut microbiota This renders newborns, particularly those born premature, at risk for vitamin K deficiency bleeding (VKDB) (Shearer, 2009). There is a two-pronged approach to address this potential deficiency: (1) prophylactic administration of vitamin K; and (2) phylloquinone fortification of infant formula. Prophylaxis, the dose and mode of which varies by country, confers effective protection for the first week of life (Sankar et al., 2016), after which feeding, and potentially vitamin K produced by the infant gut microbiota, contribute to the vitamin K status of the infant. Countries that do not offer vitamin K prophylaxis at birth have persistently higher rates of VKDB, although the data are incomplete in areas without a public health infrastructure. In the United States, there has been an episodic decline in parental approval for the vitamin K prophylactic shot at birth, in part due to persisting misinformation regarding its safety, and has resulted in an increase in VKDP cases. The daily intake of exclusively breastfed infants averages 1–2 μg/day phylloquinone, whereas formula-fed infants have an average intake of around 50 μg/day phylloquinone (Shearer, 2009). In preterm infants who received both vitamin K prophylaxis and phylloquinone supplementation through parenteral nutrition feeds, the presence of metabolites in serum and excretion of a less-metabolized vitamin K catabolite in urine were suggestive of a possible overload of both vitamin K recycling and metabolic pathways. These data were recently confirmed in a recent study of infant-mother dyads, in which high concentrations of unmetabolized phylloquinone were measured in feces of formula-fed infants. These observations raise the question that in the context of prophylaxis, what are the nutritional vitamin K requirements of the infant?

Vitamin K status and chronic disease

While evidence is accumulating that vitamin K-dependent mechanisms underlie several chronic diseases, currently osteoporosis and cardiovascular disease are the most-studied.

The carboxylation of osteocalcin in bone provides a mechanistic link between vitamin K and skeletal health. A 2006 meta-analysis identified 13 vitamin K supplementation randomized clinical trials (RCTs) that evaluated BMD and 7 RCTs that evaluated fracture as outcomes (Cockayne et al., 2006). The summarized results suggested benefit of vitamin K supplementation with respect to bone mineral density. It was later determined that the results of several of the early RCTs were compromised by allegations of research misconduct and indeed these trials have been retracted in the literature. In 2019, the meta-analysis was updated to exclude the discredited trials and add new trials completed since 2006. The updated results did not support a beneficial effect of vitamin K supplementation with respect to bone mineral density (Mott et al., 2019). There was some indication of a potential protective effect with respect to clinical fracture, but data were too sparse to draw definitive conclusions. Most trials provided 45 mg of MK4, which is used pharmaceutically as an anti-osteoporotic medication in some Asian countries, such as Japan. This dose is not attainable by diet alone, no matter what form of vitamin K is consumed. Nutrient supplementation is most likely to benefit individuals with lower nutrient status at baseline, yet none of the available vitamin K supplementation RCTs screened individuals for low baseline vitamin K status. It is also possible that certain segments of the population may derive more benefit from vitamin K supplementation than others. However, it is unlikely that vitamin K supplementation will have a meaningful effect on bone health in individuals with replete calcium and vitamin D status Characteristics related to bone quality, such as bone material properties or microarchitecture, effect fracture risk independent of bone mineral density. It is possible vitamin K and osteocalcin are involved in fracture through mechanisms not reflected by BMD. However, corroborative research is needed in this area.

Coronary artery calcification (CAC) is a subclinical manifestation of atherosclerotic CVD and higher CAC is associated with a higher risk for clinical CVD events, such as myocardial infarction. MGP and GRP can inhibit arterial calcification when carboxylated, which requires vitamin K. Several observational studies have associated higher circulating (dp)ucMGP and/or lower circulating phylloquinone (both reflective of lower vitamin K status) with less CAC and a lower risk for CVD. However, as reviewed in detail elsewhere (Shea et al., 2021), not all studies agree. Some have suggested MK7 has unique cardioprotective properties based mainly on observational studies which reported higher MK intakes were associated with less CAC and a lower CVD risk. A recently completed systematic review and meta-analysis identified nine RCTs that tested the effect of vitamin K supplementation on subclinical CVD, the majority of which evaluated vascular or valvular calcification and also reported changes in (dp)ucMGP (Vlasschaert et al., 2020). Six trials provided MK7 and three provided phylloquinone supplements, with follow-up ranging from 6 months to 3 years. Both MK7 and phylloquinone supplementation consistently reduced plasma (dp)ucMGP, although the magnitude of the reduction varied considerably (Shea et al., 2021). However, reductions in plasma (dp)ucMGP were not accompanied by reductions in arterial calcification, thereby challenging the clinical importance of lowering plasma (dp)ucMGP, at least with respect to CVD. Individuals with chronic kidney disease (CKD) frequently develop arterial calcification and are reported to have lower vitamin K status and so were thought to particularly benefit from vitamin K supplementation. However, the results of RCTs that evaluated the effect of vitamin K supplementation on vascular calcification in patients with CKD have been mainly null. The available evidence when considered altogether does not support a benefit of vitamin K with respect to CVD, even among high-risk patient populations.

Vitamin K and the gut microbiome

Bacteria residing in the human gut produce menaquinones, but the relevance of gut-derived menaquinones to the vitamin K function and health of the human host is still speculative. Within the gut microbiota, menaquinones play critical roles in bacterial energy metabolism as electron carriers in microbial respiration, yet some bacterial taxa have lost key genes in menaquinone, Menaquinones secreted by some human gut bacterial species act as a growth factor for neighboring bacteria with incomplete menaquinone synthesis pathways, suggesting menaquinones may be an important commensal factor in the human gut microbiota, so gut-derived menaquinones may also influence human health indirectly through modulation of the gut microbial composition.

In female mice, dietary vitamin K deficiency altered the intestinal menaquinone profile, but dietary supplementation with phylloquinone, MK4 or MK9 did not. Similar trends were observed in males, but statistical significance was not reached (Ellis et al., 2021). These findings suggest the amount, rather than the form, of vitamin K in the diet influences intestinal microbial composition. In humans, there appear to be two dominant menaquinone profiles in the gut microbiome, such that individuals tend to either have greater abundance of *Prevotella* spp. and be rich in MK5 and MK11-13, or have greater abundance of *Bacteroides* spp. and be rich in MK9-10. These menaquinone profiles do not appear to change in response to dietary interventions not designed to alter vitamin K intake nor is it known what the implications are for these groupings with respect to overall vitamin K status. More research is needed to elucidate the role, if any, of gut-derived menaquinones in human health.

Conclusion

Vitamin K remains one of the few essential nutrients for which there are insufficient data from which to generate a dietary requirement. Whereas its biochemical role as an enzyme cofactor is well characterized with respect to supporting normal coagulation, and

among the general adult population, vitamin K-related coagulation disorders are rare. Vitamin K prophylaxis is also widely and successfully used to prevent vitamin K deficiency bleeding among newborns who are vitamin K deficient. Unfortunately, there is no single validated biomarker for measuring overall vitamin K status, which has led to an abundance of conflicting theories regarding the relative contribution of the multiple dietary vitamin K forms to health outcomes beyond that of normal coagulation. The production of vitamin K by gut bacteria does not appear to be a major contributor to vitamin K status of the host but may have other roles critical to gut health. There is indeed growing evidence that vitamin K has biochemical and physiological roles beyond that of an enzyme cofactor, However, there are still many gaps in knowledge that need be addressed prior to establishing dietary requirements for vitamin K.

References

Card, D.J., Shearer, M.J., Schurgers, L.J., Harrington, D.J., 2009. The external quality assurance of phylloquinone (vitamin K(1)) analysis in human serum. Biomed. Chromatogr. 23, 1276–1282.

Cockayne, S., Adamson, J., Lanham-New, S., Shearer, M.J., Gilbody, S., Torgerson, D.J., 2006. Vitamin K and the prevention of fractures: systematic review and meta-analysis of randomized controlled trials. Arch. Intern. Med. 166, 1256–1261.

Ellis, J.L., Fu, X., Al Rajabi, A., Grusak, M.A., Shearer, M.J., Naumova, E.N., Saltzman, E., Barger, K., Booth, S.L., 2019. Plasma response to deuterium-labeled vitamin K intake varies by TG response, but not age or vitamin K status, in older and younger adults. J. Nutr. 149, 18–25.

Ellis, J.L., Karl, J.P., Oliverio, A.M., Fu, X., Soares, J.W., Wolfe, B.E., Hernandez, C.J., Mason, J.B., Booth, S.L., 2021. Dietary vitamin K is remodeled by gut microbiota and influences community composition. Gut Microb. 13, 1–16.

Ellis, J.L., Fu, X., Karl, J.P., Hernandez, C.J., Mason, J.B., Debose-Boyd, R.A., Booth, S.L., 2022. Multiple dietary vitamin K forms are converted to tissue menaquinone-4 in mice. J. Nutr. 152, 981–993.

Harshman, S.G., Finnan, E.G., Barger, K.J., Bailey, R.L., Haytowitz, D.B., Gilhooly, C.H., Booth, S.L., 2017. Vegetables and mixed dishes are top contributors to phylloquinone intake in US adults: data from the 2011-2012 NHANES. J. Nutr. 147, 1308–1313.

Ichikawa, T., Horie-Inoue, K., Ikeda, K., Blumberg, B., Inoue, S., 2006. Steroid and xenobiotic receptor SXR mediates vitamin K2-activated transcription of extracellular matrix-related genes and collagen accumulation in osteoblastic cells. J. Biol. Chem. 281, 16927–16934.

Mott, A., Bradley, T., Wright, K., Cockayne, E.S., Shearer, M.J., Adamson, J., Lanham-New, S.A., Torgerson, D.J., 2019. Effect of vitamin K on bone mineral density and fractures in adults: an updated systematic review and meta-analysis of randomised controlled trials. Osteoporos. Int. 30, 1543–1559.

Nakagawa, K., Hirota, Y., Sawada, N., Yuge, N., Watanabe, M., Uchino, Y., Okuda, N., Shimomura, Y., Suhara, Y., Okano, T., 2010. Identification of UBIAD1 as a novel human menaquinone-4 biosynthetic enzyme. Nature 468, 117–121.

Sankar, M.J., Chandrasekaran, A., Kumar, P., Thukral, A., Agarwal, R., Paul, V.K., 2016. Vitamin K prophylaxis for prevention of vitamin K deficiency bleeding: a systematic review. J. Perinatol. 36 (Suppl. 1), S29–S35.

Schurgers, L.J., Teunissen, K.J., Hamulyak, K., Knapen, M.H., Vik, H., Vermeer, C., 2007. Vitamin K-containing dietary supplements: comparison of synthetic vitamin K1 and natto-derived menaquinone-7. Blood 109, 3279–3283.

Shea, M.K., Booth, S.L., 2016. Concepts and controversies in evaluating vitamin K status in population-based studies. Nutrients 8, 8.

Shea, M.K., Booth, S.L., 2019. Vitamin K, vascular calcification, and chronic kidney disease: current evidence and unanswered questions. Curr. Dev. Nutr. 3, nzz077.

Shea, M.K., Berkner, K.L., Ferland, G., Fu, X., Holden, R.M., Booth, S.L., 2021. Perspective: evidence before enthusiasm-A critical review of the potential cardiovascular benefits of vitamin K. Adv. Nutr. 12, 632–646.

Shearer, M.J., 2009. Vitamin K deficiency bleeding (VKDB) in early infancy. Blood Rev. 23, 49–59.

Shearer, M.J., Newman, P., 2008. Metabolism and cell biology of vitamin K. Thromb. Haemost. 100, 530–547.

Suttie, J.W., 2009. Vitamin K in Health and Disease. CRC Press, Taylor and Francis Group, Boca Raton FL.

Vlasschaert, C., Goss, C.J., Pilkey, N.G., Mckeown, S., Holden, R.M., 2020. Vitamin K supplementation for the prevention of cardiovascular disease: where is the evidence? A systematic review of controlled trials. Nutrients 12, 2909.

Zinc: Deficiency disorders and prevention programs

S.Y. Hess, University of California, Davis, CA, USA

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of S.Y. Hess, Zinc: Deficiency Disorders and Prevention Programs, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 431–436, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00287-7.

Causes of Zinc Deficiency **577**
Inadequate Dietary Zinc Intake 577
Other Causes of Zinc Deficiency 578
High Physiological Requirements 579
Prevalence of Zinc Deficiency in Developing Countries **579**
Consequences of Zinc Deficiency in Developing Countries: Evidence Derived from Zinc Supplementation Trials **580**
Child Growth 580
Morbidity and Mortality 580
Pregnancy 580
Zinc Intervention Strategies **581**
Preventive Zinc Supplementation 581
Therapeutic Zinc Supplementation in the Treatment of Diarrhea 581
Food Fortification with Zinc 581
Mass Fortification of Staple Foods 581
Targeted Fortification 582
Dietary Diversification and Modification 582
Biofortification **582**
Further Reading **582**

Glossary

Bioavailability The proportion of the nutrient content of a food or meal that is absorbed and retained.

Phytate : zinc molar ratio The molar ratio between phytate (molecular weight 660 g mol^{-1}) and zinc (molecular weight 65.4 g mol^{-1}) and is used to estimate the bioavailability of zinc in food and mixed diets.

Phytic acid Myo-inositol hexaphosphate, the principal storage form of phosphorus in many plants and the major inhibitor of zinc absorption (also described as phytate).

Zinc A trace element essential for the human body as it required for the activity of >100 enzymes involved in most major metabolic pathways.

Zinc deficiency Lack of sufficient zinc to meet the physiological requirement.

Causes of Zinc Deficiency

As with other micronutrient deficiencies, four main factors are responsible for the development of zinc deficiency in lower-income countries: inadequate dietary zinc intake; poor zinc absorption from high-phytate, plant-based diets; disease states that either induce excessive losses or impair utilization of zinc; and physiological states that increase zinc requirements, such as the periods of rapid growth during childhood and pregnancy.

Adequate zinc nutrition is essential for human health because of zinc's critical structural and functional roles in multiple enzyme systems that are involved in gene expression, cell division and growth, and immunologic and reproductive functions.

Inadequate Dietary Zinc Intake

Inadequate dietary intake of absorbable zinc is one of the major causes of zinc deficiency. Animal-source foods, in particular shellfish, small whole fish, beef, and organ meats such as liver and kidney, are rich sources of zinc. Plant-source foods, such as most fruits and vegetables including green leaves, and starchy roots and tubers, have relatively low zinc content. Although whole grains, nuts, and legumes have moderate to high zinc content, these foods also contain large quantities of phytate (phytic acid or myo-inositol hexaphosphate), the most potent identified dietary inhibitor of zinc absorption.

Plants synthesize phytate, which occurs in highest concentrations in seeds and to a lesser extent in vegetative plant parts. Phytate forms chelates with zinc and other minerals, making these minerals less available for absorption. The inhibitory effect of phytate on zinc absorption appears to follow a dose-dependent response, and the phytate : zinc molar ratio can be used to estimate the proportion of absorbable zinc. The phytate : zinc molar ratio of foods or diets is calculated as follows, where 660 equals the molecular weight of phytate and 65.4 the molecular weight of zinc:

$$\frac{\text{mg phytate}/660}{\text{mg zinc}/65.4}$$

The zinc and phytate content, and the phytate : zinc molar ratio in some foods are shown in **Table 1**. If information on the phytate content of the diet cannot be calculated, then diets can be categorized as having low or average zinc bioavailability based on certain dietary characteristics. For example, unrefined cereal and/or legume-based diets generally have phytate : zinc ratios >18, which is associated with relatively low zinc bioavailability. In contrast, mixed diets containing higher amounts of animal-source foods and less plant-source foods, or refined plant-based diets generally have phytate : zinc ratios between 4 and 18, which are associated with higher zinc bioavailability.

Other Causes of Zinc Deficiency

Under normal physiological conditions, zinc is secreted into the intestine in large quantities together with digestive juices, but is largely reabsorbed. Diarrhea may not only lead to a reduced absorption of dietary zinc during the episode due to decreased intestinal transit time, but may also cause an increase in the loss of endogenous zinc. Given the important role of the intestine in

Table 1 The average content of zinc and phytate, and the phytate : zinc molar ratio in uncooked foods

Food	*Zinc (mg per 100 g)*	*Phytate (mg per 100 g)*	*Phytate : zinc molar ratio*
Cereals			
Corn	1.8	800	44
Pasta	0.7	282	40
Rice (milled)	1.1	352	32
Wheat or whole-wheat bread	2.9	845	29
White bread	0.9	30	3
Nuts and legumes			
Lentils/mung beans	1.3	358	27
Peanuts	3.3	1760	53
Peas	2.9	1154	39
Red beans	2.9	1629	56
Roots and tubers			
Cassava	0.3	54	18
Potato	0.3	81	27
Sweet potato	0.5	50	10
Vegetables			
Cabbage	0.1	0	–
Green leaves	0.2	42	21
Onion	0.2	0	–
Tomato	0.1	6	6
Fruits			
Banana	0.2	0	–
Coconut	1.1	324	29
Orange	0.1	0	–
Mango	0.0	20	–
Animal-source foods			
Beef	3.0	0	–
Chicken	1.3	0	–
Eggs	1.1	0	–
Fish	0.5	0	–
Milk	0.4	0	–
Pork	1.9	0	–

regulating dietary zinc absorption, and the secretion and reabsorption of endogenous zinc during digestion, conditions that affect the health or integrity of the intestine, such as tropical enteropathy, could interfere with the adequate maintenance of zinc balance. The contribution of these conditions to zinc deficiency in lower-income countries requires investigation.

High Physiological Requirements

In accordance with age and physiological status, some population groups have increased daily physiological requirements for absorbed zinc. During growth and pregnancy, the incorporation of zinc into newly synthesized tissue requires relatively larger amounts of zinc daily. Similarly, the amount of zinc transferred from mother to infant in breastmilk must be added to the lactating women's physiological requirement for absorbed zinc. These increased needs for zinc add to the challenge of acquiring sufficient amounts of absorbable zinc from the food supply. Those groups with higher zinc requirements and who are thus at elevated risk of zinc deficiency include: infants (particularly those born prematurely), young children, children recovering from severe malnutrition or diarrhea, adolescents, and pregnant and lactating women. Some evidence exists for the occurrence of zinc deficiency among each of these groups in lower-income country settings.

The elderly may also be at elevated risk of zinc deficiency, due to a decline in zinc intakes and possibly a reduction in the absorption of dietary zinc. However, evidence for zinc deficiency among the elderly has thus far only been reported from industrialized countries; studies are needed among elderly populations in lower-income countries.

Prevalence of Zinc Deficiency in Developing Countries

Guidelines on the assessment of population zinc status were recently published following a consensus conference convened by the World Health Organization (WHO), the United Nations Children's Fund (UNICEF), the International Atomic Energy Agency (IAEA), and the International Zinc Nutrition Consultative Group (IZiNCG). The three main types of zinc status assessment that were considered included biochemical, dietary, and functional methods. Because so little information is available from nationally representative surveys on the prevalence of low serum zinc concentration or inadequate dietary zinc intake, current estimates of the extent of risk of zinc deficiency must rely on the prevalence of stunting among children less than 5 years of age. Approximately 30% of children less than 5 years of age worldwide are stunted (height-for-age Z-score<−2 SD with respect to the distribution of the reference population data). WHO recommends a prevalence of stunting greater than 20% of the population to indicate a public health concern. The highest prevalence rates of stunting (>30%) are observed in countries in sub-Saharan Africa, South Asia, Southeast Asia, and Central America. Intermediate prevalence rates (20–30%) are found in the Andean countries, some Central American countries, Southern Africa, and some countries in North Asia. As zinc deficiency is not the only factor affecting children's growth, assessment of dietary zinc intake and serum zinc levels should be used to confirm the risk of zinc deficiency in these countries. (Figure 1)

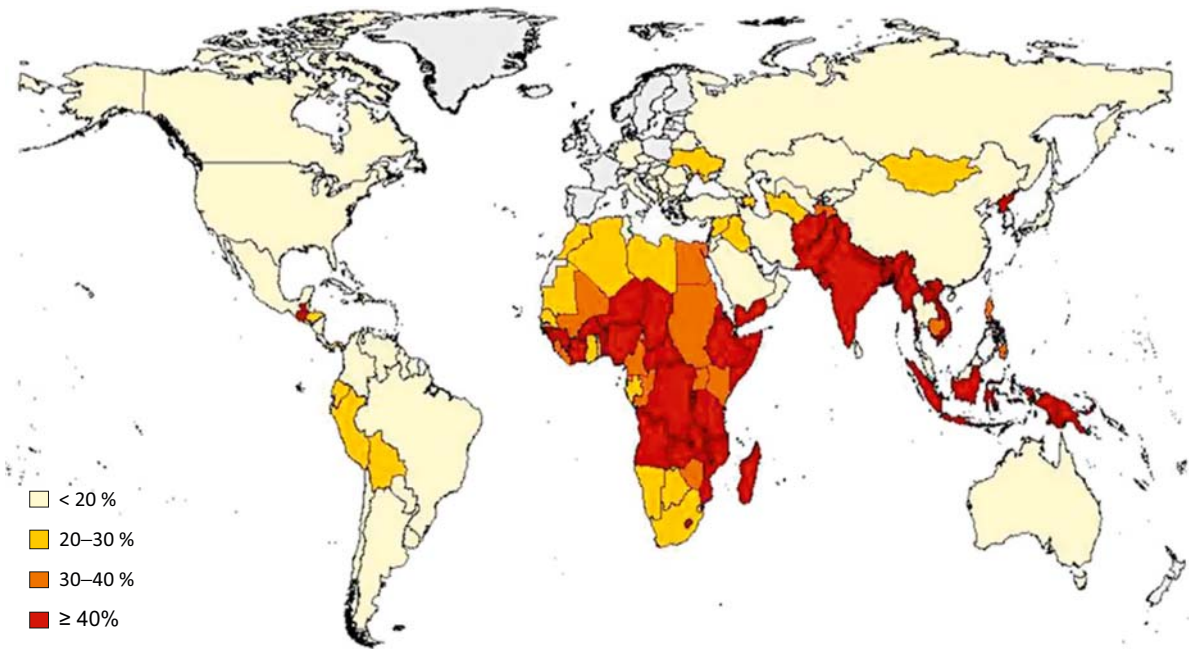


Figure 1 Prevalence of nutritional stunting in children under 5 years of age. Data derived from WHO or most recent Demographic Health Surveys. http://www.izincg.Org/stunning

Consequences of Zinc Deficiency in Developing Countries: Evidence Derived from Zinc Supplementation Trials

In the context of developing country settings, present knowledge on the health consequences of zinc deficiency has been almost entirely derived from community-based trials of zinc supplementation among populations at possible risk of zinc deficiency. In these trials, individuals in the study population are randomly allocated to receive either a zinc supplement, usually in the form of tablets or syrups, or the same supplement format without zinc (i.e., placebo). The condition under study is then monitored for a given period (typically for 2 months to 1 year), and the occurrence of or change in the condition is compared between the zinc-supplemented group and the corresponding control group. Given that several other nutritional and environmental factors can influence the health conditions hypothesized to occur with zinc deficiency, such studies have been essential in demonstrating unequivocally the causal role of zinc deficiency in these conditions among human populations. The following section provides an overview of the population groups at elevated risk of zinc deficiency, and the health consequences associated with zinc deficiency, as concluded from these studies.

Child Growth

Zinc plays an important role in child growth. Several mechanisms may be involved, including the role of zinc in the transcription and translation of genetic material and perhaps more importantly, the regulatory role of zinc in the endocrine system, which controls growth (i.e., the growth hormone–somatomedin axis). Specifically, zinc status is associated with the concentration of circulating insulin-like growth factor-1, the principal growth factor that controls early childhood growth. Among populations where growth restriction occurs, both height and weight gain have improved following supplemental zinc. Stimulation of linear growth appears to be the primary response, while the increase in body weight likely reflects the synthesis of lean tissue, such as bone, cartilage, and muscle, associated with linear growth. This is evident because, in general, weight does not increase independently of improved height in response to zinc supplementation.

According to a recent meta-analysis including 37 studies in infants, preschool, and older prepubertal children, zinc supplementation produces a small, but significant increase in linear growth and weight gain. Zinc deficiency has been demonstrated to be an important limiting factor to the growth of children across a wide range of geographical settings in developing regions. However, it should be noted that not all the studies have demonstrated a significant, positive effect of zinc on growth. Possible explanations for this include: the prevalence or severity of growth stunting in the study communities was low, zinc status was adequate, or deficiencies of other growth-limiting nutrients coexisted thus preventing a positive effect of zinc on growth. The latter situation may also explain the observation in some studies of a transient effect of zinc on growth.

Low-birth-weight infants (<2.5 kg) may have additional needs for zinc, presumably to facilitate their rapid postnatal catch-up growth. Some benefits of supplemental zinc to growth have been observed among low-birth-weight infants in the first 6 months of life.

Severely malnourished infants and children have exhibited improved rates of weight gain, height gain, or synthesis of lean tissue when supplemental zinc has been included in their usual rehabilitation treatment regimen. In these recovering children, zinc has been shown to augment the deposition of lean tissue by increasing protein synthesis. Thus, the inclusion of zinc among other micronutrients is recommended by WHO in the treatment of severely acute malnourished children.

Morbidity and Mortality

Zinc is involved in many aspects of the immune system, contributing both to specific and nonspecific immune functions. Zinc has an important role in both the prevention and treatment of diarrhea, which may be mediated both through functions in immune competence and maintenance of the integrity of the intestine. Preventive zinc supplementation reduces the incidence of diarrhea by approximately 20% among children in lower-income countries, although the current evidence indicates that this beneficial effect of zinc may be limited to children greater than 12 months of age. Zinc also has therapeutic benefits for recovery from diarrheal infections. Overall, supplemental zinc provided to children during either acute or persistent diarrhea leads to a reduction in the duration and possibily the severity of the episode. It has been recommended that zinc should be used in the management of acute diarrhea, in conjunction with oral rehydration salt solution. Zinc deficiency appears to also be associated with an increased incidence of pneumonia. Present evidence from studies that diagnosed acute lower respiratory tract infections (ALRI) based on counting respiratory rate or a physician's examination indicates that preventive zinc supplementation reduces the incidence of pneumonia and ALRI in children by about 21%. Some studies also suggest that zinc supplementation reduces the incidence of malaria, although this remains uncertain because of the limited amount of evidence that is currently available. Based on recent reviews on the impact of zinc on mortality, zinc supplementation reduces the overall mortality rate in children by 6–9%. Similarly to the impact of zinc on diarrhea incidence, the benefits seem to be limited to children above 12 months of age in which the mortality reduction is approximately 18%.

Pregnancy

Zinc requirements during pregnancy have been estimated from the zinc content of accrued tissues during pregnancy. In addition to the zinc transferred to the fetus, zinc is deposited in the placenta, amniotic fluid and uterine, and mammary tissue. The estimated

total additional zinc needed for pregnancy is approximately 100 mg. Few firm conclusions can be made as to the consequences of zinc deficiency during pregnancy on maternal, fetal, and infant health. Observational studies in human populations suggest that maternal zinc deficiency during pregnancy may cause adverse pregnancy outcomes for the mother and the fetus. However, results from zinc supplementation trials have been inconsistent and therefore difficult to interpret. This may be partly attributed to an inadequate study design or failure to consider the zinc status of the women studied. Nevertheless, a recent meta-analysis of supplementation trials indicates a 14% reduction in premature delivery among zinc-supplemented women.

Zinc Intervention Strategies

There are numerous zinc intervention strategies available and all have their strengths and weaknesses. While some approaches are considered short-term solutions, others may require a longer period until successful and effective implementation is achieved. In some cases, a combination of these intervention strategies may be needed to ensure the prevention of zinc deficiency in the most vulnerable population groups.

Preventive Zinc Supplementation

As described above, most of the evidences on the beneficial effects of zinc are derived from randomized controlled trials of preventive zinc supplementation and benefits include improved growth, reduced incidence of diarrhea, ALRI, and reduced all-cause mortality. Evidence to date shows no significant adverse effects on indicators of iron and copper status. It has been stated previously that zinc needs to be provided on a daily basis for an extended period of time, although weekly supplementation may also be beneficial. IZiNCG recommends a daily dose of 5 mg zinc for young children. Zinc supplementation is a short-term strategy that relies heavily on the availability of zinc supplements and individual compliance. The challenges for scaling up zinc supplementation programs are similar to those faced by other programs that attempt to procure and distribute nutritional supplements or medicines in lower-income countries. Several existing delivery platforms have been identified, which could be used for delivering preventive zinc supplements with or without other micronutrients. As these programs are being developed, monitoring and evaluation should be included to allow the required assessment of the program's effectiveness.

Therapeutic Zinc Supplementation in the Treatment of Diarrhea

WHO and UNICEF recommend that zinc supplementation should be included as a component in the treatment of all cases of diarrhea. It is recommended to provide children with 20 mg per day of supplemental zinc for 10–14 days (10 mg per day for infants under six months old) along with oral rehydration salt solution and continued feeding. The aim is that the recommendations become routine practice both in the home and health-care facility and that caretakers will act quickly at the first sign of diarrhea. Efforts are underway to reinforce national diarrhea programs and scale up the inclusion of therapeutic zinc supplementation in many countries.

A recent analysis confirmed that there is strong evidence that therapeutic zinc supplementation during diarrhea decreases the duration of diarrheal episodes by ~0.5 day. To what extent the dose of supplemental zinc provided during diarrhea episodes can prevent zinc deficiency is uncertain. Although it is believed that therapeutic zinc supplementation has the potential to prevent future diarrhea episodes following diarrhea treatment, present evidence is weak.

Food Fortification with Zinc

Food fortification is increasingly recognized as an effective approach to improve population's micronutrient status. Available absorption studies clearly show that zinc fortification can increase dietary zinc intake and total daily zinc absorption. Most studies also indicate that adding zinc to food does not adversely affect the sensory properties of the food or the absorption of other micronutrients, such as iron. However, despite the positive effect of zinc fortification on total zinc absorption, the impact as a public health intervention remains unknown.

There are different types of food fortification: foods fortified that are widely consumed by the general population (mass fortification), foods fortified for specific population subgroups, such as complementary foods for young children or rations for displaced populations (targeted fortification), and foods fortified voluntarily by the manufacturers and available in the market place (market-driven fortification).

Mass Fortification of Staple Foods

As described above, mass fortification is the addition of one or more micronutrients to foods commonly consumed by the general public, such as cereals, condiments and milk. Zinc fortification of cereal food staples (wheat flour, maize flour, or rice) is not yet widely practiced, but many new flour fortification programs are beginning to include zinc. Currently, in four countries– Indonesia, Mexico, Jordan, and South Africa–fortification of wheat flour with zinc is mandatory. Thirteen countries include zinc in voluntary wheat flour fortification programs, and five countries have recently proposed new programs that would include zinc. WHO recently adopted guidelines for fortification of wheat flour with zinc. The recommended levels range from 30 to 100 parts per million (ppm)

depending on the extraction of the wheat flour and the estimated per capita flour consumption, and the amount of zinc and phytate in the rest of the diet. Although several fortification compounds are generally recognized as safe (GRAS) by the US Food and Drug Administration, zinc oxide is most widely used in food fortification due to its low cost.

Targeted Fortification

Infants from 6 to 24 months of age in lower-income countries are especially vulnerable to deficiencies of zinc and iron because: (1) rapid rate of growth during this period imposes relatively high requirements of these nutrients, (2) breastmilk intakes decrease with age, and the zinc and iron contents of human milk decline to low levels in the second semester post-partum, and (3) most home-available complementary foods are limited in the amounts and bioavailability of these nutrients. Reviews of the theoretical nutrient requirements during the period of complementary feeding indicate that it is very difficult for infants in lower-income countries to meet their requirements for zinc and iron from home-available preparations. One approach is to fortify complementary foods to help meet the physiological needs of young children. However, recent analyses of commercially available fortified complementary foods revealed that the zinc content is in many cases insufficient. Various strategies for 'home' fortification, or point-of-use fortification (POUF), have been developed to ensure adequate micronutrient intakes by infants and young children. These types of products include micronutrient powders, crushable tablets, and lipid-based nutrient supplements, which are added to the complementary food at the time of consumption. To date, there is only inconsistent evidence that zinc fortified complementary foods or POUF can improve zinc status, as measured by biochemical or functional indicators. Whether this is due to the low bioavailability of zinc from cereal-based complementary foods (due to the high phytic acid concentration) or due to other reasons is uncertain.

Dietary Diversification and Modification

The most desirable approach to eliminate zinc deficiency will be to ensure access to diets with adequate zinc content and bioavailability. Dietary diversification and modification have the potential to prevent deficiencies of zinc and other micronutrients. The best strategy for enhancing the zinc content of household diets is to promote the consumption of meat, poultry and fish, all good sources of readily-available zinc. This is because beef, pork, lamb, and liver have a higher content of readily absorbed zinc (3.0 to 6.8 mg of zinc/100 g) than poultry (~1.1 to 2.7 mg of zinc/100 g), eggs (~1.0 to 1.3 mg of zinc/100 g), dairy products (~0.3 to 1.0 mg of zinc/100 g), or finfish (~0.3 to 0.7 mg of zinc/100 g for flesh only; ~3.2 mg of zinc/100 g for whole, soft-boned fish with bones).

Several household food preparation and processing methods can be used to reduce the phytate content of diets based on cereals and legumes. These methods are based on the enzymatic hydrolysis of phytic acid to lower inositol phosphates that occurs during germination and fermentation. A combination of dietary strategies involving increased consumption of animal-source foods and phytate reduction is the preferred method to enhance both the content and bioavailability of zinc in the diets in rural areas of lower-income countries. To be effective, such strategies must be integrated with ongoing national agriculture, food, nutrition, and health education programs and implemented using a participatory approach to ensure their acceptability, adoption, and sustainability.

Promotion and support of appropriate breastfeeding practices should also be considered among the recommended dietary strategies to enhance the zinc status of infants and young children, for two reasons: breastmilk is an important source of bioavailable zinc, and breastfeeding protects against diarrhea, which causes excessive zinc losses. Public health programs targeting young children should consider the promotion of the three key recommendations: (1) early initiation of breastfeeding, (2) exclusive breastfeeding to 6 months, and (3) continued breastfeeding to 24 months.

Biofortification

Biofortification is an agricultural strategy that aims to increase the content of selected micronutrients, including zinc, in staple foods such as rice, wheat, maize, pearl millet, and others. Biofortification of staple foods can be achieved through the following processes: conventional breeding, by selecting for genotypes with the highest micronutrient content observed for that crop; use of genetic modifications, such as gene insertions or induced mutations; and use of agronomic practices, such as applications of zinc-containing fertilizers. When consumed, biofortified staple foods would lead to improved adequacy of zinc intakes and hence a reduced risk of dietary zinc deficiency, among those who currently have high rates of inadequate intakes. Although the feasibility and efficacy of biofortification to prevent zinc deficiency still needs to be evaluated through efficacy trials, theoretical analyses indicate that it has a potential to improve zinc intakes in adults and children.

Further Reading

de Benoist, B., Darnton-Hill, I., Davidsson, L., Fontaine, O., Hotz, C., 2007. Conclusions of the joint WHO/UNICEF/IAEA/IZiNCG interagency meeting on zinc status indicators. Food and Nutrition Bulletin 28, S480–S484.

Bhutta, Z.A., Ahmed, T., Black, R.E., et al., 2008. What works? Interventions for maternal and child undernutrition and survival. Lancet 371, 417–440.

Brown, K.H., Baker, S.K., IZiNCG Steering Committee, 2009. Galvanizing action: Conclusions and next steps for mainstreaming zinc interventions in public health programs. Food and Nutrition Bulletin 30, S179–S184.

Brown, K.H., Hambidge, K.M., Ranum, P., the Zinc Fortification Working Group, 2010. Zinc fortification of cereal flours: current recommendations and research needs. Food and Nutrition Bulletin 31, S62–S74.

Brown, K.H., Rivera, J.A., Bhutta, Z., et al., 2004. International Zinc Nutrition Consultative Group (IZiNCG) technical document #1. Assessment of the risk of zinc deficiency in populations and options for its control. Food and Nutrition Bulletin 25, S99–203.

Dewey, KG, Adu-Afarwuah, S, 2008. Systematic review of the efficacy and effectiveness of complementary feeding interventions in developing countries. Maternal and Child Nutrition 1, 24–85.

Gibson, R.S., Anderson, V.P., 2009. A review of interventions based on dietary diversification or modification strategies with the potential to enhance intakes of total and absorbable zinc. Food and Nutrition Bulletin 30, S108–S143.

Gibson, R.S., Bailey, K.B., Gibbs, M., Ferguson, E.L., 2010. A review of phytate, iron, zinc, and calcium concentrations in plant-based complementary foods used in low-income countries and implications for bioavailability. Food and Nutrition Bulletin 31, S134–S146.

Hess, S.Y., Brown, K.H., 2009. Impact of zinc fortification on zinc nutrition. Food and Nutrition *Bulletin* 30, S79–S107.

Hess, S.Y., King, J.C., 2009. Effects of maternal zinc supplementation on pregnancy and lactation outcomes. Food and Nutrition Bulletin 30, S60–S78.

Hotz, C., 2009. The potential to improve zinc status through biofortification of staple food crops with zinc. Food and Nutrition Bulletin 30, S172–178.

International Zinc Nutrition Consultative Group (IZiNCG) – http://www.izincg.org

Walker, C.L., Black, R.E., 2010. Zinc for the treatment of diarrhoea: effect on diarrhoea morbidity, mortality and incidence of future episodes. International Journal of Epidemiology 39 (Supplement 1), i63–i69.

World Health Organization and UNICEF (2004) Clinical Management of Acute Diarrhea. Geneva: WHO/Unicef joint statement. WHO/FCH/CAH/04.7.

World Health Organization, 2009. Recommendations on wheat and maize flour fortification. Meeting Report: Interim Consensus Statement. World Health Organization, Geneva, Switzerland.

Zinc: Physiology, dietary sources, and requirements

H.C. Freake, University of Connecticut, Storrs, CT, USA
K. Sankavaram, Howard University College of Medicine, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of H.C. Freake, K. Sankavaram, Zinc: Physiology, Dietary Sources, and Requirements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 437–443, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00286-5.

Introduction **584**
History of Zinc as a Nutrient **585**
Chemistry of Zinc **585**
Zinc in Foods **585**
Control of Zinc Homeostasis **585**
Zinc Absorption 585
Transport and Distribution 586
Excretion 587
Zinc Biochemistry **587**
Homeostasis 587
Zinc Enzymes 588
Zinc Transcription Factors 588
Zinc Physiology **588**
Growth 589
Immune Function 589
Reproduction 589
Nervous System 589
Antioxidant Defense System 590
Macronutrient Metabolism 590
Human Zinc Deficiency **590**
Mild 590
Severe 590
Zinc Toxicity **590**
Assessment **591**
Recommended Intakes **591**
Further Reading **591**

Glossary

Acrodermatitis enteropathica A genetic disorder leading to zinc deficiency, associated with mutations in the ZIP4 transporter and impaired zinc absorption.
Metallothionein A small, cysteine-rich zinc binding protein that modulates intracellular zinc homeostasis and acts as an antioxidant.
MTF-1 A zinc-regulated transcription factor.
ZIP transporters A family of zinc transporters responsible for moving zinc into the cytoplasm. Also known as SLC39.
ZnT transporters A family of zinc transporters responsible for moving zinc out of the cytoplasm. Also known as SLC30.

Introduction

Zinc is moderately abundant in nature, ranking 23rd of the elements. Of the trace elements in the body, it is second only to iron, but unlike iron, it has only a single redox state. Together with its size and charge characteristics, this has led to its widespread use within proteins of the body. The number of zinc proteins is unknown but growing, including several hundred enzymes and many more nuclear proteins that regulate gene expression. Additional proteins are responsible for zinc homeostasis. Binding sites and functions of zinc within some of these proteins are well understood, but for others this is less clear. In particular, the links between these biochemical roles of zinc within proteins and its physiological functions are often obscure. The range of physiological functions of zinc is broad and can be observed in all tissues of the body. In general, zinc is required for DNA synthesis, cell division and

growth, for protein synthesis and macronutrient metabolism, and for the development and function of most body systems. Lack of an appropriate assessment tool makes it difficult to estimate deficiency prevalence, but undiagnosed marginal zinc deficiency may be a concern.

History of Zinc as a Nutrient

The essentiality of zinc for bacterial growth has been known for almost 150 years. Later, it was shown to be required by plants and then in 1934 for the rat. Because of its broad distribution in the food supply, human zinc deficiency was initially thought to be unlikely. However, in the early 1960s Prasad and others in Iran described a syndrome of dwarfism and lack of sexual development in teenage boys and young adults. The Iranian young men consumed a diet based on unleavened bread with very little animal protein and also ate large amounts of clay (geophagia). They were anemic and responded to treatment with ferrous sulfate, coupled with a more balanced diet including animal protein. The other symptoms also resolved but it seemed unlikely that lack of iron itself was responsible. Prasad then moved to Egypt, where he encountered a similar syndrome. These patients were not geophagic, but ate mostly bread and beans and also were infested with schistomiasis and hookworm. Zinc deficiency was documented in these individuals and treatment with zinc was shown to be more effective at increasing growth rates than either iron or a diet with animal protein. Thus dietary zinc deficiency was demonstrated, presumably due to impaired absorption because of the high fiber and phytate contents of the diet.

Chemistry of Zinc

The conjunction of chemical properties of zinc underlies its biological significance. It is a relatively small ion (atomic number=30) and carries a positive charge of two. It attracts electrons as a strong Lewis acid and this property can be important in its catalytic functions. It has relatively flexible coordination geometry and while binding its ligands with high affinity, exhibits rapid rates of exchange that can facilitate chemical reactions and biological processes. Its single redox state, in contrast to iron or copper, eliminates danger of oxidative damage. Although other trace elements share some of these properties, none share them all. This is what makes zinc so valuable for protein structure and function.

Zinc in Foods

Zinc is associated with proteins in the body and is found associated with proteins in food. Protein rich foods tend to be good sources (Table 1). However, there is great variability, ranging from egg whites, which have almost no zinc, up to oysters at 400 mg kg^{-1}. Legumes and grains have moderate zinc content, though refinement results in large losses, whereas fruits and vegetables are poor sources. Zinc bioavailability is quite variable, due to other food components eaten at the same time. A principal concern is phytate, which renders zinc unavailable. On the other hand, animal proteins appear to enhance zinc absorption, perhaps because amino acids derived from them keep zinc in a soluble form. Enriched breakfast cereals can make significant contributions to zinc intake.

Control of Zinc Homeostasis

The size and charge characteristics of zinc mandate the use of carriers to traverse biological membranes. Two families of transporters have been described and partially characterized. The ZIP family (also called Solute Carrier 39; SLC39), functions to move zinc into the cytoplasm, either from outside of the cell or from subcellular compartments. The second group of transporters, the ZnT family (SLC30), is responsible for zinc egress from the cytoplasm. Fourteen transporters belonging to the ZIP and 10 belonging to the ZnT family have been identified. Cellular zinc homeostasis is tightly regulated by these transporters in normal and pathological states. For example, the normally zinc-rich prostate exhibits zinc loss with malignancy that is associated with the down regulation of ZIP transporters. ZnT-1 is localized to plasma membranes and functions as a cellular efflux protein. ZnT-2 transports zinc into storage vesicles under conditions of high cellular zinc. Collectively, the ZIP and ZnT proteins are likely to underlie the homeostatic control of zinc distribution around the body.

Zinc Absorption

The absorption, distribution, and excretion of zinc are shown in Figure 1. Overall, approximately 20–40% of consumed zinc is absorbed, depending on bioavailability within the particular food source. Zinc is absorbed by both saturable and nonsaturable processes, with the greatest rates of absorption occurring in the jejunum. Absorption is adjusted to meet needs, being proportionately increased in deficiency states and reduced when intake is high. Zinc status is reflected by the intestinal concentration of the zinc binding protein, metallothionein (MT). MT may trap zinc within the epithelial cells, causing it to be lost to feces as the cells are

Table 1 Dietary sources of zinc

Food	*Zinc content mg kg^{-1} raw weight*
Oysters	400
Beef, lean	50
Pork	26
Chicken	
Breast	8
Leg	18
Salmon	4
Egg	
Whole	11
White	0.3
Milk	
Whole	4
Cheese	
Cheddar	31
Wheat	
Whole flour	29
White flour	7
Rice	
Brown	20
Polished	12
Breakfast cereals, fortified	100–500
Kidney beans	27
Lentils	36
Potatoes	3
Broccoli	4
Apples	0.4

sloughed off. This may be part of the explanation of how zinc absorption is adjusted to meet needs. Acrodermatitis enteropathica is an autosomal recessive condition of zinc malabsorption, which can lead to severe deficiency. Gene mutations that lead to this condition have been identified in *ZIP4*, which encodes one of the zinc transporters. This protein has been localized to the apical membrane of intestinal epithelial cells and, given the severity of symptoms associated with its inactivation, appears to be necessary for normal zinc absorption. The zinc efflux protein ZnT-1 is found at the basolateral membrane and so likely promotes passage of zinc out of the intestine. Acrodermatitis enteropathica can be treated with large doses of zinc, supporting the existence of paracellular transport at high intake levels. A large amount of zinc is secreted into the gut from the pancreas and intestine (**Figure 1**). Malabsorption syndromes can lead to a failure to reabsorb these endogenous secretions and a rapid loss of body zinc.

Transport and Distribution

The zinc plasma pool is relatively small, representing only approximately 0.1% of total body zinc. It circulates bound to albumin and α-2-macroglobulin, with approximately 3% complexed with amino acids. Approximately five-fold greater amounts of zinc are found in whole blood, with erythrocytes accounting for approximately 75% of the total. However, approximately 85% of erythrocyte zinc is complexed within carbonic anhydrase and therefore does not exchange easily. Egress of zinc from the circulation across endothelial cells and into tissues of the body is not well understood. Uptake in association with albumin has been suggested but members of the ZIP family of transporters are likely to play a role here. The tissue distribution of zinc is relatively uniform. All cells require the mineral and no cell stores it. The concentration of zinc in the adult human is approximately 0.5 μmol g^{-1}, giving a total body content of approximately 2 g. More than half is found in skeletal muscle and approximately 30% in bone. The bone pool appears to be more labile than the muscle pool and this has been used as an index of zinc status in experimental animals. Liver represents another labile pool. It receives dietary zinc from the portal circulation and contains approximately 5% of body zinc. At the cellular level, approximately 30–40% zinc is present in the nucleus whereas, 50% of zinc is distributed among the cytoplasm, organelles, and specialized vesicles. The remaining zinc is associated with membranes.

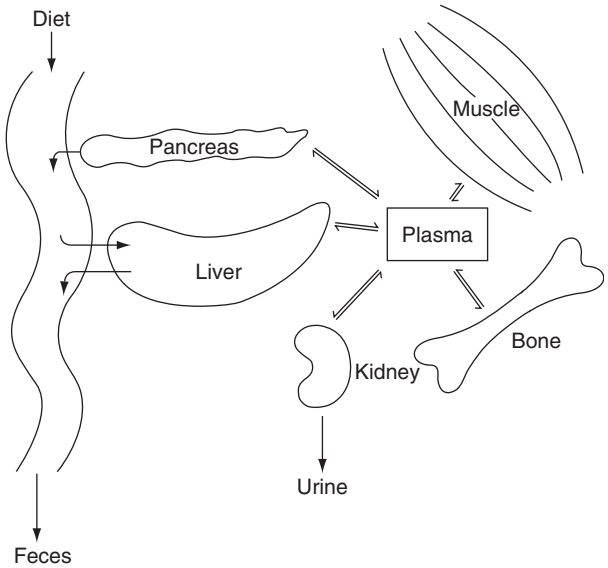


Figure 1 Whole body zinc homeostasis. Zinc in the intestine comes from the diet and endogenous secretions. A portion is absorbed, but much is lost in the feces, which constitute the major site of excretion. Absorbed zinc passes through the liver and then to the general circulation. Zinc is distributed throughout the body with muscle and bone constituting the largest pools. A minor but controlled amount of zinc is lost in the urine.

Excretion

Zinc is lost from the body primarily through the feces (Figure 1). Feces contain unabsorbed dietary zinc, zinc contained within intestinal epithelial cells, which have been sloughed off, and endogenous secretions into the gut from the pancreas, the gall bladder, and the cells lining the gastro–intestinal tract. The endogenous secretions and the extent to which they are reabsorbed can be controlled and represent an important homeostatic mechanism for regulating zinc status. Zinc losses in urine are relatively minor, but do respond to extremes of intake to help maintain homeostasis. Shed skin cells, sweat, hair, menstrual blood, and semen represent additional routes of loss.

Zinc Biochemistry

Zinc homeostasis and action involves an intimate association of the mineral with proteins. These include membrane transporters responsible for the absorption of zinc in the gut and its passage into and out of cells and subcellular organelles, transport and delivery proteins, both in the circulation and within cells, sensing proteins that adjust homeostasis and function according to zinc availability and then a large range of proteins to which zinc is ultimately delivered. Two major classes of these latter proteins are the enzymes and transcription factors.

Homeostasis

The interaction of zinc with its transporters has not been well characterized, though transmembrane domains have been identified thought to be responsible for the transport function. Free concentrations of zinc within the cell appear to be extremely low and may not constitute a sufficient pool for the supply of zinc to its protein ligands. This implies the existence of delivery proteins, a role suggested for MT, which has been shown to transfer zinc to apoenzymes *in vitro*. MT is a small protein that is unusually cysteine-rich and can bind seven atoms of zinc. It may influence the subcellular distribution and availability of zinc, because its own distribution varies. For example, the nuclear content of MT varies with the cell cycle. MT expression is regulated by zinc and also by a range of other signals including glucocorticoids, interleukins, and cAMP. In addition, its zinc binding activity is influenced by the cellular redox state. For example, an increase in the glutathione disulfide/glutathione ratio results in release of zinc from MT and thus its availability for other proteins.

Investigation of the mechanism whereby zinc regulates the expression of MT led to the discovery of the single protein known to act as a zinc sensor within mammalian cells, MTF-1 (metal response element (MRE)-binding transcription factor-1). MTF-1 binds to MREs in the promoter region of MT and other genes and regulates their expression. The ability of MTF-1 to localize to the nucleus and bind to its target genes is dependent on its zinc content. Thus, an increase in cellular zinc concentrations results in greater MTF-1 activity and elevated expression of its target genes. In addition to MT, which will bind more zinc, these include ZnT-1, which will transport zinc out of the cell. Both actions serve to buffer the increase in cytosolic zinc (Figure 2).

Zinc Enzymes

The three-dimensional structures of more than 200 zinc-containing enzymes have now been characterized, and many more have been identified. All six International Union of Biochemistry classes are represented. In some cases (e.g., carbonic anhydrase), zinc is a direct participant in the catalytic function of the enzyme. The zinc atom is coordinated by three amino acids from the enzyme and a molecule of water at the active site. Other enzymes (e.g., protein kinase C) have structural zinc sites, where the metal binds four amino acids within the protein and ensures appropriate folding for bioactivity. For nitric oxide synthase, zinc has been found to serve a bridging function between two separate polypeptides, stabilizing a biologically active larger complex. A selection of zinc enzymes is listed in **Table 2** and help to illustrate the wide variety of metabolic functions requiring zinc.

Zinc Transcription Factors

There are many zinc enzymes but even more transcription factors utilize zinc. Variable numbers of zinc atoms are each coordinated by four cysteine or histidine residues to stabilize a DNA binding structure. A search of the human genome has revealed over 1000 genes (approximately 3% of those identified) containing these characteristic zinc finger domains. An important class of zinc finger transcription factors is the steroid/thyroid receptor superfamily, responsible for mediating the biological response to a wide range of hormonal and metabolic signals, including retinoic acid and vitamin D. These all have nine conserved cysteine residues in the DNA binding region, eight of which are coordinated by two atoms of zinc. Loss of zinc from these sites would interrupt biological function but it is not clear that this ever happens in a physiological context. Gene expression profiling has been used to assess the genome-wide response to changing zinc availability in different tissues, including intestine, liver, and cells of the immune system. The gene products identified as zinc-sensitive by these approaches amount to approximately 5% of the expressed genes within a tissue. MTF-1 is likely to mediate some but not all of these changes and other transcription factors whose activity is dependent on zinc may soon be found.

Zinc Physiology

The enormous range of biochemical roles for zinc predicts broad effects on physiological function. The physiological roles of zinc may be further extended by secondary effects mediated by altered food intake and effects on the function of other nutrients. Although the physiological roles for zinc are well described, it is important to note that the connections between the biochemistry and physiology of zinc remain unclear. Thus the specific zinc-sensitive biochemical step leading to altered physiology is often unknown. The broad spread of zinc through the body at the organ, cellular and even protein levels suggests that the function of most systems is dependent on zinc. Its physiological roles become manifest under the circumstance of deficiency and that framework will be used to discuss principal functions here.

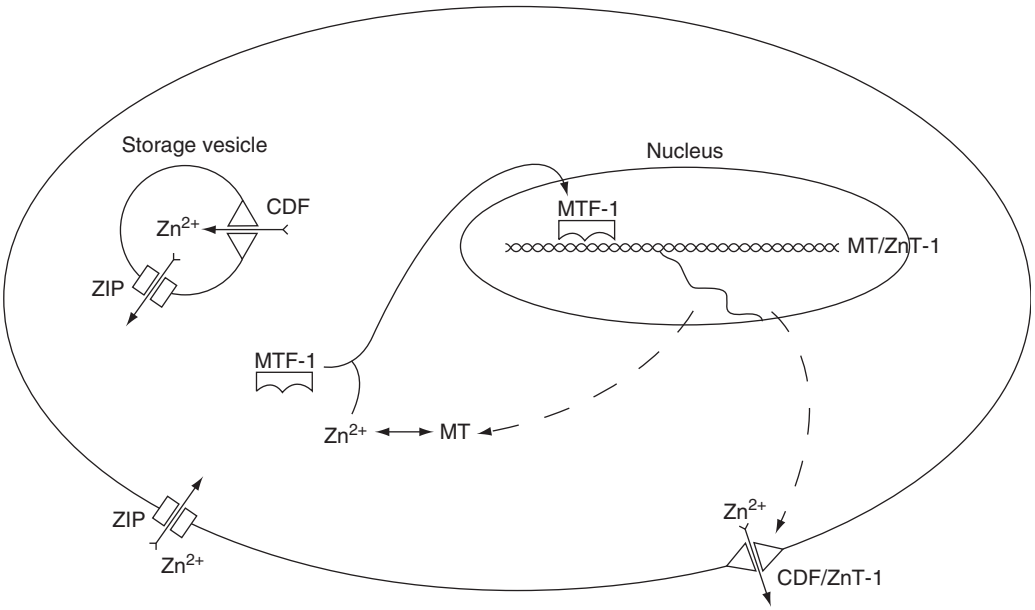


Figure 2 Cellular zinc homeostasis. Zinc is delivered to the cytoplasm from either the extracellular space or vesicles within the cell by members of the ZIP family of transporters. A rise in cellular zinc results in activation and nuclear translocation of MTF-1. In the nucleus, MTF-1 regulates transcription of a set of target genes including MT and ZnT-1. MT will bind zinc and ZnT-1 transports zinc out across the plasma membrane. MT may govern delivery of zinc to other proteins within the cell. Other members of the ZnT family transport zinc into vesicles.

Table 2 Examples of mammalian zinc-dependent enzymes

Enzyme	*Function*
RNA polymerase	Transcription, synthesis of mRNA
Carboxypeptidase A	Protein digestion in intestine
Protein kinase C	Signal transduction
Carbonic anhydrase	Respiration/buffering/hydration of CO_2
Cytochrome *c* oxidase	Respiration/electron transport chain
Alcohol dehydrogenase	Ethanol metabolism
Superoxide dismutase	Inactivation of free radicals
Nitric oxide synthase	Signaling, vasodilation
Angiotensin converting enzyme	Blood pressure regulation/activation of angiotensinogen

Growth

The requirement of zinc for growth of numerous organisms ranging from bacteria to humans is well established. Growth failure is an early consequence of zinc deficiency in experimental animals. Numerous processes seem to contribute to the growth failure. Experiments with animals have shown that zinc deficiency leads to a drop in food intake, though the use of pair-fed controls demonstrates clear effects of zinc deficiency beyond feeding behavior. Multiple effects of zinc deficiency on the somatotrophic axis have been demonstrated, notably a reduction in circulating concentrations of insulin-like growth factor-1 (IGF-1). Again, this is only part of the story because administering exogenous IGF-1 does not restore the growth failure of zinc deficiency. Growth of cultured cells is dependent on media zinc. DNA synthesis is interrupted and production of thymidine kinase mRNA is diminished by removal of zinc, but again this is only a partial explanation. The IGF-1 signaling pathway within cells also seems to be affected. Zinc is also required for wound healing, presumably due to related processes.

Immune Function

The immune system is particularly sensitive to zinc deficiency. Lymphopenia and thymic atrophy are observed and both cell-mediated and antibody-mediated responses are reduced. As with growth, multiple mechanisms appear to be at play. In addition to generalized effects on DNA synthesis, zinc deficiency appears to induce apoptosis, resulting in a loss of B and T cell precursors within the bone marrow. Thymulin, a zinc-dependent enzyme that stimulates the development of T cells within the thymus, may be involved. Production of cytokines by mononuclear cells is also reduced by zinc deficiency. These effects can be of clinical significance. Infections occur more frequently in individuals with acrodermatitis enteropathica and reduced immune function is accompanied by zinc deficiency in several other conditions, including sickle cell anemia and various gastro–intestinal disorders.

Zinc has been used therapeutically in individuals with immune-compromised diseases. In the United States, zinc lozenges have become popular as a treatment for the common cold. Zinc supplementation has also been used as a supporting therapy for patients with HIV and AIDS, herpes simplex viruses, and hepatitis C infections. A shortening of cold duration, improved phagocytosis and increase in the number of Th cells, and a reduced frequency of opportunistic infections was observed in patients with HIV infection. However, results from controlled trials of zinc treatment with these diseases have been variable. Treatment effectiveness may depend on initial zinc status, with greater success being seen in individuals with marginal, undetected zinc deficiency.

Reproduction

The original description of zinc deficiency in humans included lack of pubertal development. Spermatogenesis is a zinc-dependent process. Seminal fluid is particularly rich in zinc and the sperm appears to accumulate zinc from this source before ejaculation. Zinc is also crucial for normal fetal development and deficiency leads to abnormalities in humans and animals. Maternal zinc deficiency has also been linked with pregnancy-associated morbidity, including preterm delivery.

Nervous System

The brain is one of the sites shown to be particularly sensitive to zinc deficiency during fetal development, with neural tube defects and other disorders being found. Although this work was performed with animals, a similar relationship appears likely with humans. Zinc is distributed throughout the brain, but greater concentrations are found within the hippocampus. Here a brain-specific transporter, ZnT-3, concentrates zinc in vesicles within glutamatergic neurons. It is co-secreted with the neurotransmitter and appears to serve as a modulator of neurotransmission. Very high concentrations of zinc (>100 μM) are found within the synaptic cleft during this process. In addition, brain injury, resulting from ischemia or trauma causes the release of massive amounts of zinc, which is thought to be responsible for the resultant cell death.

Antioxidant Defense System

Although not an antioxidant itself, there are several ways in which zinc participates in the antioxidant defense system of the body, with important implications for health. It can bind to thiol groups in proteins, making them less susceptible to oxidation. By displacing redox-reactive metals such as iron and copper from both proteins and lipids it can reduce metal-induced hydroxyl radical formation and thus protect the macromolecules. Copper/zinc superoxide dismutase is an important antioxidant enzyme, which contains zinc and whose activity is impaired in the deficient state. Zinc depletion affects the activity of reactive oxygen species (ROS) and reactive nitrogen species (RNS) scavengers such as glutathione, ascorbic acid, uric acid, and α-tocopherol. The role of zinc in inducing MT has already been mentioned and this protein scavenges hydroxyl radicals. Increased oxidative stress results in the release of zinc from MT, presumably making it more available for other proteins. However, although zinc deficiency has been linked to greater oxidative stress, excess zinc can also lead to the same condition. So although MTs possess intrinsic antioxidant properties, there is evidence that in some circumstances, MTs may also serve as a source of injurious zinc release. Recent findings also suggest that nitrosative stress can act as a critical trigger for zinc mobilization. Nitric oxide (NO) or peroxy nitrite (–ONOO) interact with MT and promote zinc release both *in vitro* and *in vivo*.

The likelihood of increased oxidative stress under conditions of zinc deficiency suggests a potential anticarcinogenic role for this mineral. Indeed, zinc deficiency has been shown to result in DNA damage. The tumor suppressor gene, p53, which is frequently mutated in human cancers, is a zinc-containing transcription factor whose expression is also dependent on zinc. Dysregulation of cellular zinc homeostasis is linked not only to oxidative stress but also mitochondrial dysfunction and activation of apoptotic pathways.

Macronutrient Metabolism

Many of the enzymes of intermediary metabolism contain zinc and deficiency affects all macronutrients. Protein synthesis, as well as DNA and RNA synthesis requires zinc. Insulin is secreted from the pancreas and circulates in association with zinc. This secretion is diminished under conditions of deficiency, leading to impaired glucose metabolism. Lipid metabolism is also affected, with zinc deficiency associated with reductions in circulating high-density lipoprotein.

Human Zinc Deficiency

In addition to dietary inadequacy, there are several other routes that lead to zinc deficiency. Acrodermatitis entero-pathica, the genetic disorder of zinc malabsorption has already been mentioned. Other, more generalized malabsorption syndromes (e.g., celiac disease) can also lead to zinc deficiency. Deficiency has also resulted from inappropriate intravenous feeding and use of chelation therapy. Children are likely to be particularly at risk for zinc deficiency, because of its involvement in growth.

Mild

Given the difficulty of assessing marginal impairments in zinc status, effects of deficiency can often only be verified by a response to treatment. Growth provides a good example of this. Children in Denver, Colorado who were of low height for their age showed increased growth rates in response to zinc supplementation, whereas zinc had no effect in children of normal height. In addition to growth, improvements in immune function, taste and smell acuity, and reproductive function have been noted with zinc supplementation.

Severe

Severe human zinc deficiency has been well characterized by the original descriptions in the Middle East and in patients with acrodermatitis enteropathica. The symptoms of mild deficiency are continued and exaggerated. Thus stunting can be extreme and is accompanied by delayed sexual maturation and impotence. Characteristic skin lesions are found, originating around the mouth and nose but becoming widespread as deficiency develops. Diarrhea is also present. Deficits in taste and smell are accompanied by anorexia and other behavioral changes, including increased irritability and impaired cognitive function. Eye pathologies similar to those seen with vitamin A deficiency are observed.

Zinc Toxicity

Toxicity of zinc from food sources has not been reported and seems unlikely because absorption is homeostatically regulated. Acute gastro–intestinal symptoms and headaches from ingestion of amounts approximately 10–20-fold higher than the recommended intakes have been described. Chronic ingestion of these large amounts has been shown to impair immune response and lipoprotein metabolism. However, the key danger from excessive zinc intake is reduced copper status. This is likely due to a zinc-induced blockage of copper absorption and in fact is clinically useful in individuals with Wilson's disease, a condition of copper toxicity. In the United States, an Upper Limit of 40 mg day^{-1} has been set for adults, based on the threat to copper status. The popularity

Table 3 Recommended intakes for zinc

			FAO/WHO reference nutrient intake		
			Bioavailabilty		
Age group		*US and Canadian recommended dietary allowance*	*High*	*Moderate*	*Low*
Children (1–3 years)		3	2.4	4.1	8.3
Adolescents (14–18 years)	Female	9	4.3	7.2	14.4
	Male	11	5.1	8.6	17.1
Adults (>19 years)	Female	8	3.0	4.9	9.8
	Male	11	4.2	7.0	14.0
Pregnant women	3rd trimester	11	6.0	10.0	20.0
Lactating women	0–3 months	12	5.8	9.5	19.0

of zinc lozenges for treatment of the common cold could lead to circumstances where this intake could be exceeded. Thus use of these treatments should be limited in duration.

Assessment

The prevalence of marginal zinc deficiency in human populations is unknown because of the lack of a good means for assessing zinc status. Measurement of plasma zinc is straightforward, but it does not serve as a reliable indicator of status. Plasma zinc is a quantitatively minor pool that can be easily influenced by minor shifts in tissue zinc. Plasma concentrations do not fall with dietary intake, except at very low intakes. Plasma zinc can also be affected by other factors unrelated to zinc status (e.g., time of day, stress, infection). Cellular components of blood can be assayed, but erythrocyte concentrations of zinc are maintained in deficient states and variable results have been found with leukocytes. Hair zinc concentrations may reflect available zinc but will also be dependent on rate of hair growth.

Several different zinc-dependent enzymes have been investigated as potential markers of zinc status but none has proved reliable. MT in blood cells has been suggested as a useful indicator of zinc status, either assayed at the protein of mRNA levels. MT expression is likely to be regulated by factors other than zinc and therefore may lack the specificity required of a good indicator. The gene array approaches that have recently been used to determine the global effects of zinc deficiency within a tissue would appear to offer hope for the identification of an appropriate functional marker of zinc status.

Recommended Intakes

In the absence of a reliable index of zinc status, both the US Food and Nutrition Board and the Food and Agriculture Organization/World Health Organization (FAO/WHO) expert committee used the factorial method to estimate human zinc requirements. Under this approach, routes of zinc loss from the body are estimated and summed. The requirement is then set as the amount of dietary zinc required to make up for these losses, making appropriate assumptions about bioavailability. For children, an additional amount is added to account for the zinc requirements for growth.

As shown in **Table 3**, the FAO/WHO give three sets of recommendations, depending on the zinc bioavailability of the diet. The US figures fall between those given for moderate and low availability diets. Both groups also set Upper Limits for intake, based largely on the risk of impairing copper status. These values are similar (40 mg US, 45 mg FAO/WHO, for adults).

Further Reading

de Benoist, B., Darnton-hill, I., Davidsson, L., Fontaine, O., Hotz, C., 2007. Conclusions of the joint WHO/UNICEF/IAEA/IZiNCG interagency meeting on zinc status indicators. Food and Nutrition Bulletin 28, S480–S484.

Brown, K.H., Peerson, J.M., Allen, L.H., Rivera, J., 2002. Effect of supplemental zinc on the growth and serum zinc concentrations of pre-pubertal children: A meta-analysis of randomized, controlled trials. The American Journal of Clinical Nutrition 75, 1062–1071.

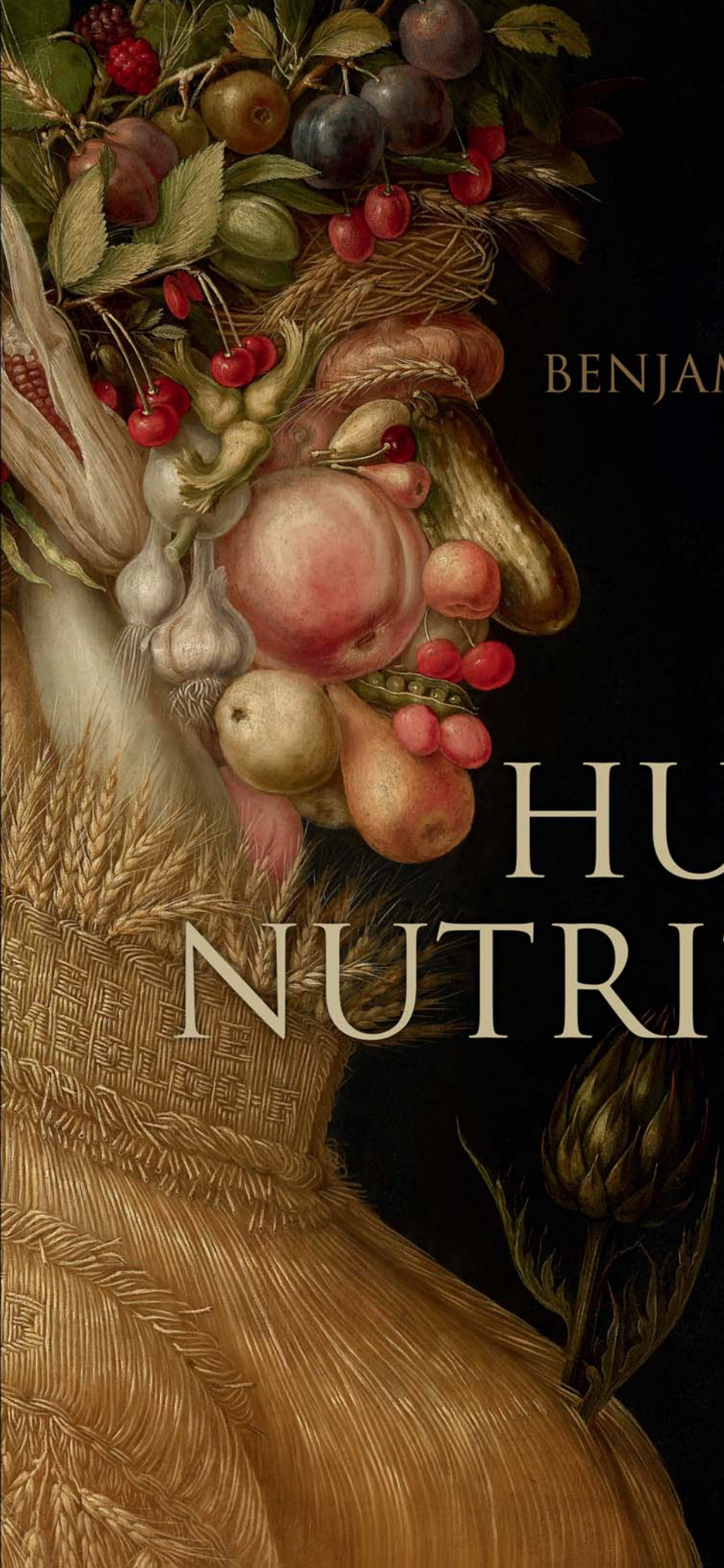
FAO/WHO (2002) Human vitamin and mineral requirements. *Report of a Joint FAO/WHO Expert Consultation*, *Bangkok, Thailand*, pp. 257–270. Rome: Food and Nutrition Division, FAO.

Haase, H., Rink, L., 2009. Functional significance of zinc-related signaling pathways in immune cells. Annual Review of Nutrition 29, 133–152.

Institute of Medicine, 2001. Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium and Zinc. National Academy Press, Washington, pp. 442–501.

Krezel, A., Hao, Q., Maret, W., 2007. The zinc/thiolate redox biochemistry of metallothionein and the control of zinc ion fluctuations in cell signaling. Archives of Biochemistry and Biophysics 463, 188–200.

Laity, J.H., Andrews, G.K., 2007. Understanding the mechanisms of zinc-sensing by metal-response element binding transcription factor-1 (MTF-1). Archives of Biochemistry and Biophysics 463, 201–210.
Lichten, L.A., Cousins, R.J., 2009. Mammalian zinc transporters: Nutritional and physiologic regulation. Annual Review of Nutrition 29, 153–176.
MacDonald, R.S., 2000. The role of zinc in growth and cell proliferation. Journal of Nutrition 130, 1500S–1508S.
Mills, C.F. (Ed.), 1989. Zinc in Human Biology. Springer-Verlag, London.
Prasad, A.S., 1991. Discovery of human zinc deficiency and studies in an experimental human model. The American Journal of Clinical Nutrition 53, 403–412.
Prasad, A.S., 2009. Zinc: Role in immunity, oxidative stress and chronic inflammation. Current Opinion in Clinical Nutrition and Metabolic Care 12, 646–652.
Sekler, I., Sensi, S.L., Hershfinkel, M., Silverman, W.F., 2007. Mechanism and regulation of cellular zinc transport. Molecular Medicine 13, 337–343.
Tubek, S., 2007. Zinc supplementation or regulation of its homeostasis: Advantages and threats. Biological Trace Element Research 119, 1–9.
Vallee, B.L., Falchuk, K.H., 1993. The biochemical basis of zinc physiology. Physiological Reviews 73, 79–118.



EDITOR IN CHIEF

BENJAMIN CABALLERO

VOLUME TWO

Encyclopedia of HUMAN NUTRITION

FOURTH EDITION

ENCYCLOPEDIA OF HUMAN NUTRITION

FOURTH EDITION

This page intentionally left blank

ENCYCLOPEDIA OF HUMAN NUTRITION

FOURTH EDITION

EDITOR IN CHIEF

Professor Benjamin Caballero,

Emeritus Professor, Bloomberg School of Public Health and School of Medicine,
Johns Hopkins University
Baltimore, MD, United States

VOLUME 2

Section 2: Molecular Mechanisms for the Interaction of Nutrients and Health

Edited by Noel W. Solomons

Center for the Studies of Sensory Impairment, Aging, and Metabolism (CeSSIAM),
Guatemala City, Guatemala, Guatemala

Section 3: Diet Composition

Edited by Anura Kurpad

Department of Physiology, St John's Medical College, Bengaluru, India

ELSEVIER

AMSTERDAM ● BOSTON ● HEIDELBERG ● LONDON ● NEW YORK ● OXFORD
PARIS ● SAN DIEGO ● SAN FRANCISCO ● SINGAPORE ● SYDNEY ● TOKYO
Academic Press is an imprint of Elsevier

ACADEMIC PRESS

Academic Press is an imprint of Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge MA 02139, United States

Copyright © 2023 Elsevier Ltd. All rights reserved

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers may always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN 978-0-12-821848-8

For information on all publications visit our website at http://store.elsevier.com



Publisher: Oliver Walter
Acquisitions Editors: Emma Hayes and Michael Nicholls
Content Project Managers: Richard Berryman and Sindhuja Vadlamani
Associate Content Project Manager: N. Kiruthigadevi
Designer: Matthew Limbert

CONTENTS OF VOLUME 2

Contributors to Volume 2 *ix*

Editor Biographies *xv*

Preface *xix*

MOLECULAR MECHANISMS FOR THE INTERACTION OF NUTRIENTS AND HEALTH

Appetite: Psychobiological and behavioral aspects 1
RJ Stubbs, GS Finlayson, and JE Blundell

Effects of diet on behavior 11
EL Gibson, MW Green, and SC Dyall

Brown adipose tissue: Implications for human health 25
Martijn Boost, Borja Martinez-Tellez, Angel Gil, and Jonatan R Ruiz

Chrononutrition 36
Francisco Javier Ruiz-Ojeda and Marta Garaulet

Cytokines: Metabolic and nutritional aspects 48
RF Grimble and PC Calder

Diet and eating behavior: Appetite control and satiety 57
M Dalton, NJ Buckland, and JE Blundell

Dietary modulation of inflammation 67
PC Calder

Drug-nutrition interactions 79
Joseph I Boullata

Eicosanoids: Prostaglandins, leukotrienes, thromboxanes and related compounds 93
Ángel Gil and Luis Fontana

Electrolytes: Acid–base balance 104
PB Mark, KK Stevens, and AG Jardine

Energy metabolism 113
Malcolm Watford

Energy: Adaptation 127
Abdul G Dulloo

Hunger 137
K Beaulieu and JCG Halford

Hydration and thirst 143
J Leiper

Indirect calorimetry 151
Michele N Ravelli, Timothy C Shriver, Dale A Schoeller, and Chad Cook

Microbiota of the intestine: Biology and physiological functions 161
Julio Plaza-Diaz, Francisco Javier Ruiz-Ojeda, and Angel Gil

Microbiota of the intestine: Dietary interactions 168
Julio Plaza-Diaz and Angel Gil

Nutriepigenetics 176
Concepción M Aguilera, Augusto Anguita-Ruiz, and Ángel Gil

Nutrigenetics 188
Ángel Gil, Concepción M Aguilera, and Augusto Anguita-Ruiz

Nutrigenomics 197
Ángel Gil, Concepción M Aguilera, and Augusto Anguita-Ruiz

Protectins, resolvins and maresins 209
Ángel Gil and Luis Fontana

Organic: Biochemical mechanisms and regulation of vitamins and vitamin-like cofactors 215
Robert B Rucker, Francene M Steinberg, and Winyoo Chowanadisai

DIET COMPOSITION

Alcohol: Absorption, metabolism, and physiological effects 250
R Rajendram, RJ Hunter, and VR Preedy

Alcohol: Effects of consumption on diet and nutritional status 266
R Rajendram

Beverages and health 276
Liwei Chen

Biofortification: A primer on nutrient enriched crops 285
Mduduzi NN Mbuya, Valerie M Friesen, and Lynnette M Neufeld

Cereal grains 293
Sumathi Swaminathan

Dietary fiber: Physiological effects and health outcomes 306
Ian T Johnson

Fatty acids: Metabolism 316
PA Watkins

Food composition data 2021 330
F Grande and UR Charrondiere

Food fortification: Technological aspects 339
O Dary, M Guamuch-Castañeda, and JO Mora

Food safety heavy metals and metalloids 351
Fernando Gil and Pablo Olmedo

Food safety: Bacterial contamination 362
AH Simonne

Food safety: Pesticides 375
Antonio F Hernández

Fructose: Metabolism and health effects 389
Chrysanthi Stylianopoulou

Fruits and vegetables 397
María Dolores Ruiz-López and Belén García-Villanova Ruiz

Glucose and its polymers: Chemistry, sources, digestion and metabolism 412
Julie Miller Jones

Glycemic index 422
J Brand-Miller, G Frost, and A Dornhorst

Legumes in human health and nutrition 430
Sayra N Serrano-Sandoval, Daniela Guardado-Felix, and Janet A Gutiérrez-Uribe

Lycopenes and related compounds 438
G Tang

Meal size and frequency: Effect on absorption and metabolism 445
FE Lithander

Meat, poultry, and meat products: Nutritional value 451
PA Lofgren

Mediterranean diet 458
Emilio Martinez de Victoria Muñoz and Angel Gil

Micronutrient supplementation: Programmatic issues 467
Melissa F Young, Lynnette M Neufeld, Sara Hendrix, and Usha Ramakrishnan

Milk and dairy products 479
Javier Fontecha, Manuela Juárez, and Ángel Gil

Mycotoxins: Toxicity, occurrence, risk assessment and prevention 492
L Manyes and G Font

Nutraceuticals: Health effects and clinical applications 501
Oliver Chen and Eunice Mah

Nuts and seeds 513
J Salas-Salvadó and M Pascual-Compte

Organic foods 523
Léon Guéguen and Gérard Pascal

Phytochemicals: Classification and occurrence 530
A Cassidy and C Kay

Phytochemicals: Health effects of proanthocyanidins and related compounds 539
Yi Lin, Ningping Zhan, and Dejian Huang

Probiotics and prebiotics 545
Jose M Saavedra

Protein quality and sources 559
Daniel Tomé

Protein: Requirements and role in diet 568
DJ Millward

Salt, hypertension and cardiovascular outcomes 579
Tilman B Drüeke and Paolo M Suter

Sucrose: Dietary sucrose and disease 596
Philipp A Gerber

Supplementation: Dietary supplements 606
SS Percival

Sustainable diets: Their definition, measurement and promotion 612
Shauna M Downs, Alexandra Payne, and Wiktoria Staromiejska

Sweeteners: Sensory properties, digestion, consumption trends, and health effects 624
Richard D Mattes, Sze-Yen Tan, and Robin M Tucker

Traditional and indigenous foods for food security and sovereignty 639
Emily V Merchant and James E Simon

Trans-fatty acids: Health effects, recommendations, and regulations 648
SK Gebauer and DJ Baer

Ultra-processed foods 653
Paula dos Santos Leffa

Vegetarian diets 663
J Dwyer and J Harvey

Whole grains and chronic disease risk 675
Caleigh M Sawicki, Nicola M McKeown, Shatabdi Goon, and Shilpa N Bhupathiraju

CONTRIBUTORS TO VOLUME 2

Concepción M Aguilera
Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology "José Mataix", Center of Biomedical Research, University of Granada, Armilla, Spain; Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain; and CIBEROBN (Physiopathology of Obesity and Nutrition Network CB12/03/30038), Institute of Health Carlos III (ISCIII), Madrid, Spain

Augusto Anguita-Ruiz
Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology "José Mataix", Center of Biomedical Research, University of Granada, Armilla, Spain; Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain; CIBEROBN (Physiopathology of Obesity and Nutrition Network CB12/03/30038), Institute of Health Carlos III (ISCIII), Madrid, Spain; and Barcelona Institute for Global Health, ISGlobal, Barcelona, Spain

DJ Baer
US Department of Agriculture, Beltsville, MD, United States

K Beaulieu
School of Psychology, University of Leeds, Leeds, United Kingdom

Shilpa N Bhupathiraju
Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States; and Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, United States

JE Blundell
Appetite Control and Energy Balance Research Group, School of Psychology, Faculty of Medicine and Health, University of Leeds, Leeds, United Kingdom

Martijn Boost
Department of Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, the Netherlands

Joseph I Boullata
Clinical Nutrition, Penn Medicine, Clinical Nutrition Support Services, Hospital of the University of Pennsylvania, Philadelphia, PA, United States

J Brand-Miller
University of Sydney, Sydney, NSW, Australia

NJ Buckland
Department of Psychology, University of Sheffield, Sheffield, United Kingdom

PC Calder
School of Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

A Cassidy
Norwich Medical School, University of East Anglia, Norwich, UK

UR Charrondiere
(Retired) Regional Office for Latin America and the Caribbean, Food and Agriculture Organization of the United Nations, Santiago, Chile

Liwei Chen
Department of Epidemiology, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA, United States

Oliver Chen
Biofortis Research, Addison, IL, United States; and Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, United States

Winyoo Chowanadisai
Oklahoma State University, Stillwater, OK, United States

Chad Cook
Body Science Center, LLC, Gurnee, IL, United States

M Dalton
School of Psychology, Leeds Trinity University, Leeds, United Kingdom

O Dary
United States Agency for International Development (USAID), Washington, DC, United States

A Dornhorst
Imperial College London, London, United Kingdom

Paula dos Santos Leffa
Nutrition Research Group, Federal University of Health Sciences, Porto Alegre, Brazil

Shauna M Downs
Rutgers School of Public Health, Newark, NJ, United States

Tilman B Drüeke
CESP, Hôpital Paul Brousse, Paris-Sud University (UPS), Versailles Saint-Quentin-en-Yvelines University (Paris-Ile-de-France-Ouest University, UVSQ), Villejuif, France

Abdul G Dulloo
Department of Medicine, Physiology Unit, University of Fribourg, Fribourg, Switzerland

J Dwyer
Frances Stern Nutrition Center, Tufts Medical Center and Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, United States

SC Dyall
Roehampton University, London, UK

GS Finlayson
Appetite Control and Energy Balance Research Group, School of Psychology, Faculty of Medicine and Health, University of Leeds, Leeds, United Kingdom

G Font
Department of Preventive Medicine and Public Health, Food Science, Toxicology and Forensic Medicine, Faculty of Pharmacy, University of Valencia, Spain

Luis Fontana
Department of Biochemistry and Molecular Biology 2, School of Pharmacy, Granada, Spain; Institute of Nutrition and Food Technology "José Mataix", Biomedical Research Center, Parque Tecnológico Ciencias de la Salud, Granada, Spain; and Instituto de Investigación Biosanitaria ibs.GRANADA, Complejo Hospitalario Universitario de Granada, Granada, Spain

Javier Fontecha
Department of Bioactivity and Food Analysis, Food Lipid Biomarkers and Health Group, Institute of Food Science Research, CIAL (CSIC-UAM), Superior Council of Scientific Investigations and Autonomous University of Madrid, Madrid, Spain

Valerie M Friesen
Global Alliance for Improved Nutrition (GAIN), Geneva, Switzerland

G Frost
Imperial College London, London, United Kingdom

Marta Garaulet
Department of Physiology, University of Murcia, Murcia, Spain; and IMIB-Arrixaca, Murcia, Spain

Belén García-Villanova Ruiz
Department of Nutrition and Bromatology, Universidad de Granada, Granada, Spain

SK Gebauer
US Food and Drug Administration, College Park, MD, United States

Philipp A Gerber
Endocrinology, Diabetology and Clinical Nutrition, University Hospital Zurich (USZ), University of Zurich (UZH), Zurich, Switzerland

EL Gibson
Roehampton University, London, UK

Ángel Gil
Department of Biochemistry and Molecular Biology 2, School of Pharmacy, Granada, Spain; Institute of Nutrition and Food Technology "José Mataix", Biomedical Research Center, Parque Tecnológico Ciencias de la Salud, Granada, Spain; Instituto de Investigación Biosanitaria ibs.GRANADA, Complejo Hospitalario Universitario de Granada, Granada, Spain; and CIBEROBN (Physiopathology of Obesity and Nutrition Network CB12/03/30038), Institute of Health Carlos III (ISCIII), Madrid, Spain. Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology "José Mataix", Center of Biomedical Research, University of Granada, Granada, Spain; and Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain

Fernando Gil
Facultad de Medicina (PTS), Dpto. Medicina Legal y Toxicología, Universidad de Granada, Granada, Spain

Shatabdi Goon
Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

F Grande
Food and Agriculture Organization of the United Nations, Sao Paulo, Brazil

MW Green
Aston University, Birmingham, UK

RF Grimble
School of Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

M Guamuch-Castañeda
Institute of Nutrition of Central America and Panama (INCAP), Guatemala, Guatemala

Daniela Guardado-Felix
Programa Regional de Posgrado en Biotecnología, Facultad de Ciencias Químico Biológicas, Universidad Autónoma de Sinaloa, FCQB-UAS, Culiacán, Sinaloa, Mexico

Léon Guéguen
French Academy of Agriculture, Paris, France

Janet A Gutiérrez-Uribe
Tecnologico de Monterrey, The Institute for Obesity Research, Monterrey, Nuevo León, Mexico; and Tecnologico de Monterrey, Escuela de Ingeniería y Ciencias, Campus Puebla, Reserva Territorial Atlixcáyotl, Puebla, Pue, Mexico

JCG Halford
School of Psychology, University of Leeds, Leeds, United Kingdom

J Harvey
Frances Stern Nutrition Center, Tufts Medical Center and Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, United States

Sara Hendrix
Hubert Department of Global Health, Emory University, Atlanta, GA, United States

Antonio F Hernández
Department of Legal Medicine and Toxicology, University of Granada School of Medicine, Granada, Spain; Biomedical Research Institute of Granada (Instituto de Investigación Biosanitaria) ibs.GRANADA, Granada, Spain; and Consortium for Biomedical Research in Epidemiology & Public Health (CIBER en Epidemiología y Salud Pública), CIBERESP, Madrid, Spain

Dejian Huang
Department of Food Science and Technology, National University of Singapore, Queenstown, Singapore

RJ Hunter
The Barts Heart Center, St Bartholomew's Hospital, Barts Health NHS Trust and Queen Mary, University of London, London, United Kingdom

AG Jardine
University of Glasgow, Glasgow, UK

Ian T Johnson
Quadram Institute, Norwich, United Kingdom

Julie Miller Jones
St. Catherine University, St. Paul, MN, United States

Manuela Juárez
Department of Bioactivity and Food Analysis, Food Lipid Biomarkers and Health Group, Institute of Food Science Research, CIAL (CSIC-UAM), Superior Council of Scientific Investigations and Autonomous University of Madrid, Madrid, Spain

C Kay
Norwich Medical School, University of East Anglia, Norwich, UK

J Leiper
Loughborough University, Loughborough, UK

Yi Lin
Department of Food Science and Technology, National University of Singapore, Queenstown, Singapore

FE Lithander
Trinity College Dublin, Dublin, Ireland

PA Lofgren
Oak Park, IL, USA

Eunice Mah
Biofortis Research, Addison, IL, United States

L Manyes
Department of Preventive Medicine and Public Health, Food Science, Toxicology and Forensic Medicine, Faculty of Pharmacy, University of Valencia, Spain

PB Mark
University of Glasgow, Glasgow, UK

Borja Martinez-Tellez
Department of Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, the Netherlands

Emilio Martinez de Victoria Muñoz
Department of Physiology, Institute of Nutrition and Food Technology "José Mataix", Biomedical Research Center, University of Granada, Health Technological Park, Granada, Spain

Richard D Mattes
RDM Department of Nutrition Science, Department of Public Health, Purdue University, West Lafayette, IN, United States

Mduduzi NN Mbuya
Global Alliance for Improved Nutrition (GAIN), Washington, DC, United States

Nicola M McKeown
Department of Health Sciences, Sargent College of Health and Rehabilitation Sciences, Boston University, Boston, MA, United States

Emily V Merchant
New Use Agriculture and Natural Plant Products Program, Department of Plant Biology, Rutgers University, New Brunswick, NJ, United States; and Center for Agricultural Food Ecosystems, The New Jersey Institute for Food, Nutrition and Health, Rutgers University, New Brunswick, NJ, United States

DJ Millward
University of Surrey, Guildford, United Kingdom

JO Mora
Private Consultant, Washington, DC, United States

Lynnette M Neufeld
Food and Nutrition Division, Food and Agricultural Organization of the United Nations (FAO), Rome, Italy

Pablo Olmedo
Facultad de Medicina (PTS), Dpto. Medicina Legal y Toxicología, Universidad de Granada, Granada, Spain

Gérard Pascal
French Academy of Agriculture, Paris, France

M Pascual-Compte
Universitat Rovira i Virgili, Tarragona, Spain

Alexandra Payne
Amplify Studio, Urban Planning, Brooklyn, NY, United States

SS Percival
University of Florida, Gainesville, FL, USA

Julio Plaza-Diaz
Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada; Department of Biochemistry and Molecular Biology II, School of Pharmacy, University of Granada, Granada, Spain; and Instituto de Investigación Biosanitaria IBS.GRANADA, Complejo Hospitalario Universitario de Granada, Granada, Spain

VR Preedy
Department of Nutrition and Dietetics, Life Course Sciences, Faculty of Life Science and Medicine, King's College London, London, United Kingdom

R Rajendram
Department of Medicine, King Abdulaziz Medical City, King Abdullah International Medical Research Center, Ministry of National Guard-Health Affairs, Riyadh, Saudi Arabia; and College of Medicine, King Saud Bin Abdulaziz University for Health Sciences, Ministry of National Guard-Health Affairs, Riyadh, Saudi Arabia

Usha Ramakrishnan
Hubert Department of Global Health, Emory University, Atlanta, GA, United States

Michele N Ravelli
Department of Neurology, University of Wisconsin, Madison, WI, United States; and Biotechnology Center, University of Wisconsin, Madison, WI, United States

Robert B Rucker
University of California, Davis, Davis, CA, United States

María Dolores Ruiz-López
Department of Nutrition and Bromatology, Universidad de Granada, Granada, Spain

Francisco Javier Ruiz-Ojeda
Department of Biochemistry and Molecular Biology II, School of Pharmacy, University of Granada, Granada, Spain; Instituto de Investigación Biosanitaria IBS.GRANADA, Complejo Hospitalario Universitario de Granada, Granada, Spain; Institute of Nutrition and Food Technology "José Mataix", Center of Biomedical Research, University of Granada, Granada, Spain; and RG Adipocytes and Metabolism, Institute for Diabetes and Obesity, Helmholtz Diabetes Center at Helmholtz Center Munich, Munich, Germany

Jonatan R Ruiz
Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain; and Department of Physical Education and Sports, Faculty of Sport Sciences, University of Granada, Granada, Spain

Jose M Saavedra
Johns Hopkins University School of Medicine, Baltimore, MD, United States

J Salas-Salvadó
Department of Biochemistry & Biotechnology, School of Medicine, IISPV, Rovira i Virgili University, Reus, Spain

Caleigh M Sawicki
Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States; and Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, United States

Dale A Schoeller
Biotechnology Center, University of Wisconsin, Madison, WI, United States; and Nutrition Science, University of Wisconsin, Madison, WI, United States

Sayra N Serrano-Sandoval
Tecnologico de Monterrey, The Institute for Obesity Research, Monterrey, Nuevo León, Mexico

Timothy C Shriver
Biotechnology Center, University of Wisconsin, Madison, WI, United States

James E Simon
New Use Agriculture and Natural Plant Products Program, Department of Plant Biology, Rutgers University, New Brunswick, NJ, United States; and Center for Agricultural Food Ecosystems, The New Jersey Institute for Food, Nutrition and Health, Rutgers University, New Brunswick, NJ, United States

AH Simonne
Food Safety and Food Quality, University of Florida, Family, Youth and Community Science, Gainesville, FL, United States

Wiktoria Staromiejska
Rutgers School of Public Health, Newark, NJ, United States

Francene M Steinberg
University of California, Davis, Davis, CA, United States

KK Stevens
University of Glasgow, Glasgow, UK

RJ Stubbs
Appetite Control and Energy Balance Research Group, School of Psychology, Faculty of Medicine and Health, University of Leeds, Leeds, United Kingdom

Chrysanthi Stylianopoulou
Nutrition Services, Veria, Greece

Paolo M Suter
Clinic for Endocrinology and Clinical Nutrition, University Hospital, Zurich, Switzerland

Sumathi Swaminathan
St John's Research Institute, Bangalore, Karnataka, India

Sze-Yen Tan
Institute for Physical Activity and Nutrition (IPAN), School of Exercise and Nutrition Sciences, Deakin University, Burwood, VIC, Australia

G Tang
Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University, Boston, MA, USA

Daniel Tomé
UMR PNCA, AgroParisTech, INRAE, Université Paris-Saclay, Paris, France

Robin M Tucker
Department of Food Science and Human Nutrition, Michigan State University, East Lansing, MI, United States

Malcolm Watford
Department of Nutritional Sciences, School of Environmental and Biological Sciences, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

PA Watkins
Kennedy Krieger Institute, and Johns Hopkins University School of Medicine, Baltimore, MD, United States

Melissa F Young
Hubert Department of Global Health, Emory University, Atlanta, GA, United States

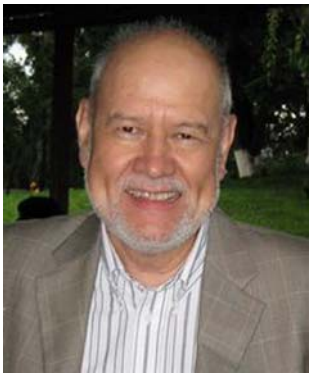
Ningping Zhan
Department of Food Science and Technology, National University of Singapore, Queenstown, Singapore

This page intentionally left blank

EDITOR BIOGRAPHIES

Editor in Chief

Professor Benjamin Caballero
Emeritus Professor, Bloomberg School of Public Health and School of Medicine, Johns Hopkins University, Baltimore, MD, United States.



Dr. Caballero is Professor Emeritus at the Department of International Health, Bloomberg School of Public Health, with joint appointment at the Department of Pediatrics, School of Medicine, Johns Hopkins University. He obtained his MD from the University of Buenos Aires, Argentina, and his PhD (in neuroendocrine regulation) from MIT, in Cambridge, Massachusetts. He started his academic career at Boston Children's Hospital, Harvard Medical School, and subsequently became the Founding Director of the Center for Human Nutrition at Johns Hopkins University.

Dr. Caballero has focused his research on child nutrition and health in developing countries. In particular, he has explored the combination of undernutrition and overweight that has become increasingly prevalent in low- and middle-income countries.

He is currently a member of the Council of the International Union of Nutritional Sciences. He has served on the Food and Nutrition Board of the US National Academy of Medicine and on a number of expert panels, including the Dietary Reference Intakes Committee, the Expert Panel on Macronutrient Requirements, and the Childhood Obesity Task Force. He was also a member of the U.S. Dietary Guidelines for Americans Advisory Committee, of the Scientific Advisory Board of the Food and Drug Administration, and of advisory committees of the National Institutes of Health and the Department of Agriculture. He is a Fellow of the American Society for Nutrition and of the Royal Society of Medicine (UK), and a member of the Spanish Academy of Nutritional Sciences.

He is the Editor-in-Chief of the *Encyclopedia of Food Sciences and Nutrition*, a 10-volume work on food production, consumption, and biological effects. He is also Editor-in-Chief of the *Encyclopedia of Human Nutrition*, which received the Book of the Year Award from the British Medical Association. His *Guide to Dietary Supplements* summarizes the current scientific basis for the use of mineral and vitamin supplements. He also co-edited a widely used textbook on human nutrition, *Modern Nutrition in Health and Disease*.

Section Editors

Section 1: *The Foundations of Human Nutrition*

Professor Angel Gil

Honorary President of the Iberoamerican Nutrition Foundation (FINUT), Emeritus Professor, Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology, Centre of Biomedical Research, University of Granada, Campus de la Salud, Avda. del Conocimiento, 18100 Armilla, Granada, Spain.



Angel Gil is an Emeritus Professor at the University of Granada and a Full Professor in the Department of Biochemistry and Molecular Biology. He is Doctor Honoris Causa at the Autonomous University of Nuevo León, Monterrey, México, and a visitant Professor of the University of Chile, Santiago de Chile, and University of San Miguel de Tucumán, Argentina. He got a major degree in Biology (1973) and PhD in Biochemistry and Molecular Biology (1978).

Prof. Gil is an internationally recognized authority in Food and Nutrition: His expertise extends from the study of human milk composition to the molecular effects of food bioactive compounds and probiotics and the design and development of novel products for infant and clinical nutrition. He conducted pioneering and innovative research, leading 7 international, 27 national, and more than 50 projects and 120 contracts; he has taught since 1981, supervising more than 50 PhD students.

Prof. Gil has several areas of interest that include evaluating the role of dietary nucleotides in early life and the development of infant nutrition products. Besides, the isolation, identification, and description of the mechanism of action of probiotics and the metabolic, molecular, and genetic factors involved in obesity and the early onset of metabolic syndrome (MS) in childhood; and the design, development, and evaluation of enteral clinical nutrition products. What describes Prof. Gil best is the variety of fields and problems he has faced during his professional carrier and his significant ability to combine his knowledge and expertise in Food Science and Human Biochemistry. This has allowed him to design, develop, innovate, and evaluate exclusive products for Human Nutrition, which are demonstrated in his published articles and his patents' impact.

The multi- and interdisciplinary nature of his work is reflected in the variety of international journals in which he has published 546 articles. Also, he has published 28 books and about 180 book chapters. His five volumes *Treatise of Nutrition*, 3rd Edition, Ed. Medica Panamericana, 2017, with more than 3500 pages, is the "bedside" book for the study of Nutritional Sciences in Spain and all Latin American countries.

He has also been the Chairman of the International Union of Nutritional Sciences (IUNS) 21st International Congress of Nutrition (2013) and the Executive Director of the 23rd International Congress of Nutrition (2017) and has been engaged in the organization of other renowned international congresses. He is a member of prestigious international and national nutrition societies and Honorary President of the Iberoamerican Nutrition Foundation (FINUT), a nonprofit organization promoted by the IUNS, in which the main goal is to contribute to the formation of young scientists in Food and Nutrition in the setting of Iberoamerica. He has received 42 National and International Awards for his contribution to Nutrition and Food Science, among them, the Class Fellow 2022 of the American Society of Nutrition; the Sir David Cuthbertson Lecture Award of the European Society of Clinical Nutrition and Metabolism for scientific achievement in clinical nutrition on 2021; the Award "Granada, City of Science and Innovation" 2021 to the Scientific Career; the Gregorio Marañón Award 2018 to the best Spanish Scientist in the field of Food Science and Nutrition; the Institute Danone Spain Award 2017; the Award of the Spanish Federation of Dairy Industries, 2015; the Nutra Excellence Award 2014, Nutra India Summitt; the UIB Honorary Award of 2013; and the NAOS Strategy Prize 2012, Special recognition for his extensive professional experience in the field of nutrition and obesity, Spanish Ministry of Health, Social Services and Equality (AESAN).

Section 2: *Molecular Mechanisms for the Interaction of Nutrients and Health*

Dr. Noel W. Solomons

Center for the Studies of Sensory Impairment, Aging, and Metabolism (CeSSIAM), Guatemala City, Guatemala, Guatemala



Noel W. Solomons was Assistant and Associate Professor from 1977 to 1984 in the former Department of Nutrition and Food Science of MIT. A resident of Guatemala since 1975, he cofounded the Center for Studies of Sensory Impairment, Aging and Metabolism (CeSSIAM) in 1985, where he has been Scientific Director ever since. He is an Adjunct Professor at Tufts University in both the Friedman School of Nutrition Science and Policy and the Department of Public Health and Community Medicine.

Dr. Solomons received his MD degree from Harvard Medical School and subsequently undertook clinical and research training in infectious diseases and in gastroenterology and clinical nutrition at the Universities of Pennsylvania and University of Chicago, respectively. He was Editor-in-Chief of the *Food and Nutrition Bulletin* from 2016–2020. He has 364 publications indexed on PUBMED. In addition, he has edited 2 books and contributed over 100 articles, reviews, editorials, and commentaries in nonindexed venues and over 50 book chapters. These are dedicated to the scientific and academic interests of his career including: clinical nutrition; human growth and body composition; lactose maldigestion; dietary intake, nutritional status, intestinal absorption, and food fortification related to various micronutrients (vitamins, trace elements, and essential fatty acids); complementary feeding; nutrition in aging and chronic disease; and the interaction of malnutrition and infection. Also, he has supervised doctoral dissertations for 12 PhD candidates from the USA, Canada, Germany, Spain, the UK and the Netherlands through CeSSIAM.

Section 3: *Diet Composition*
Professor Anura Kurpad
Department of Physiology, St John's Medical College, Bengaluru, India



Anura Kurpad works at St. John's Medical College, Bengaluru. He received his MBBS and MD from St. John's, and his PhD from Bangalore University in 1992. He was a Postdoctoral Fellow at the Rowett Research Institute and at Cambridge University, UK, and a visiting Scientist at MIT, Cambridge, United States. He is an elected Fellow of the Royal College of Physicians (London), Fellow of the Indian National Academy of Medical Sciences, Indian Academy of Sciences, and International Union of Nutritional Sciences. His interests are in human, clinical, and public health aspects of nutrition, applied throughout the life cycle. His research focuses on the physiology and clinical aspects of human energy and protein requirements and metabolism, with more recent interests in micronutrient status and metabolism. He is an Associate Editor of *The American Journal of Clinical Nutrition*. He was the President of the Nutrition Society of India from 2012 to 2016 and is the current President of the Asia Pacific Clinical Nutrition Society. He has been, and is, on many national and international advisory bodies, including being the Chair of the Indian Nutrient Requirements Committee.

Section 4: *Disorders Directly Related to Inadequate Nutrient Intake*
Dr. Katherine L. Tucker
Department of Biomedical and Nutritional Sciences, Center for Population Health, University of Massachusetts Lowell, Lowell, MA, United States



Katherine L. Tucker, PhD, is University Distinguished Professor of Nutritional Epidemiology in the Department of Biomedical and Nutritional Sciences, and Director of the Center for Population Health, at the University of Massachusetts Lowell. She holds an adjunct appointment at the University of Massachusetts Medical School. She received her PhD from Cornell University and her undergraduate degree from the University of Connecticut, both in nutritional sciences. Between these degrees, she spent 2 years as a Peace Corps volunteer in the Philippines. Before joining UMass Lowell, she was at the USDA Human Nutrition Research Center on Aging at Tufts University, and McGill University. Dr. Tucker has contributed to more than 450 articles in scientific journals. Her research focuses on dietary intake and risk of chronic disease, including osteoporosis, cognitive decline, obesity, metabolic syndrome, and heart disease, with an emphasis on health disparities. She is the PI of the Boston Puerto Rican Health Study, an ongoing cohort study, to examine the roles of diet, health behaviors, stress, and genetic predisposition in relation to chronic conditions, including heart disease, cognitive decline, and bone health; and is actively involved as a scientific advisor for the NHLBI Jackson Heart Study. She served two terms on the Food and Nutrition Board of the National Academies of Science and Engineering. She is a Fellow of the American Society for Nutrition (ASN), the Gerontological Society of America, and the American Society for Bone Mineral Research and is currently the Editor-in-Chief of *Advances in Nutrition*, the international review journal of the ASN, and Senior Editor of the forthcoming 12th edition of the textbook, *Modern Nutrition in Health and Disease*.

Section 5: *Nutrition in Disease States*
Dr. Paolo M. Suter
Department of Endocrinology, Diabetology and Clinical Nutrition, University Hospital, Zurich, Switzerland.



Paolo M. Suter, MD, MS, is a Professor of Medicine presently affiliated with the Department of Endocrinology, Diabetology and Clinical Nutrition at the University Hospital Zurich (Switzerland). He received his MD at the University of Zurich and an MS in Nutrition at Tufts University (Boston, USA). He specialized in Internal Medicine and was a Faculty Member at the Medical Policlinic of the University Hospital Zurich, where he was directing the well-known Hypertension and Obesity Outpatient Consultation. His research activities focused on vitamin nutriture in the elderly, alcohol metabolism and obesity, blood pressure and hypertension, as well as nutrition and lifestyle in non-communicable disease prevention. During his clinical work he focused on the ideal combination of pharmacological therapy with nonpharmacological strategies and on the therapy and prevention of diseases especially hypertension and obesity. Besides many research publications and review articles he authored a widely used textbook entitled *Checkliste Ernährung* (*Checklist Nutrition*). He served on different national and international boards in the area of his expertise.

Section 6: *Nutrition in the Life Cycle*
Professor Lawrence J. Cheskin
Department of Nutrition and Food Studies, George Mason University, Fairfax, VA, United States
And
Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, United States



Lawrence J. Cheskin, MD, FACP, FTOS, is Professor and Chair of the Department of Nutrition and Food Studies at George Mason University in Fairfax, VA, and an Adjunct Faculty Member of the Department of Medicine at the Johns Hopkins School of Medicine in Baltimore, MD, United States. He received his medical degree from Dartmouth and was a Postdoctoral Fellow in gastroenterology and liver diseases at Yale University-New Haven Hospital. Dr. Cheskin's work is at the intersection of public health nutrition and clinical medicine and has dedicated his career to research, education, and program building in the service of combating obesity, through both treatment and prevention. He founded the Johns Hopkins Weight Management Center, a multidisciplinary clinical research and treatment program. He was also Director of Clinical Research of the Global Obesity Prevention Center at Johns Hopkins and directed its Pilot Studies Core, which evaluated systems-focused proposals worldwide to study such areas as school policies and the built environment in preventing childhood obesity. He is co-PI of the Mason Cohort Study of entering college freshmen and has authored over 240 peer-reviewed journal articles and 8 books. He is committed to mentoring the next generation of public health and clinical scholars dedicated to nutrition and obesity.

Section 7: *Nutrition Epidemiology*
Dr. Manuel Franco
Associate Professor at the Department of Surgery, Medical and Social Sciences, University of Alcalá, Madrid, Spain, and adjunct faculty at the Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States.



Manuel Franco, MD, PhD, Epidemiologist, is an Associate Faculty at the University of Alcalá and an Adjunct Faculty at the Johns Hopkins Bloomberg School of Public Health.

He is the Principal Investigator of the European Research Council Starting Grant Heart Healthy Hoods (http://www.hhhproject.es) studying urban characteristics in relation to eating patterns, physical activity levels, smoking, and alcohol consumption. He is also the PI of the Participatory Project on food in the cities: Photovoice Villaverde (https://youtu.be/VIiFggKzVas).

Manuel Franco trained in Medicine both in Madrid and Berlin. As a Fulbright Scholar he joined the Johns Hopkins Bloomberg School of Public Health Department of Epidemiology to obtain his PhD and Postdoctoral Fellowship in the fields of social epidemiology and urban health.

His work focuses on the prevention of chronic diseases and their major risk factors as nutrition. He has published 94 international peer-reviewed articles and has led 15 studies as PI raising over 2.9 million € in competitive international research bids. His methodological interests include the measurement of urban characteristics related to chronic diseases and nutrition, the use of mixed methods, and the conduction of participatory action research methods.

Section 8: *Global aspects of Human Nutrition*
Professor Daniel J. Hoffman
Department of Nutritional Sciences and director of the Program in International Nutrition at Rutgers University, New Brunswick, NJ, United States



Daniel J. Hoffman, PhD, is a Professor in the Department of Nutritional Sciences and Director of the Program in International Nutrition at Rutgers University, New Brunswick, NJ, United States. He graduated from St. John's University, Collegeville, MN, United States, and received his PhD in Human Nutrition from the Tufts University Friedman School of Nutrition Science and Policy, Boston, MA, United States. For his doctoral training, Dr. Hoffman studied the metabolism of children living in the shantytowns of São Paulo, Brazil. He then completed a Postdoctoral Fellowship at Columbia University, New York, NY, United States, before joining the faculty at Rutgers University.

Dr. Hoffman's research focuses on the long-term impact of undernutrition early in life on body composition and metabolism later in life. This work is complemented by research on sustainable methods to improve diet and growth in lower- and middle-income countries, including Brazil, Mexico, Kenya, South Sudan, and Vietnam. He also directs the Center for Childhood Nutrition Research at the New Jersey Institute for Food, Nutrition, and Health where members conduct interdisciplinary research on childhood growth and development. Dr. Hoffman is an Associate Editor of the *Food and Nutrition Bulletin* and a Section Editor for the *Annals of Human Biology*. His other responsibilities include serving on the editorial board of the *Journal of Nutrition*, serving in leadership positions for the American Society for Nutrition, and advising international and national agencies on matters related to childhood nutrition. Aside from his academic life, Dr. Hoffman is a published documentary photographer whose portfolio includes work on the life of Roma in Europe, urban landscape of São Paulo, Brazil, gentrification of Times Square in New York City, and punk music as the "American Mosaic."

PREFACE

By the middle of last century, the science of nutrition had identified most of the essential nutrients and had provided evidence to propose specific dietary intake recommendations for many diet constituents, with the practical aim of preventing nutrient deficiencies. As chronic, noncommunicable diseases such as cardiovascular disease, cancer, etc., began to emerge as an important causal factor for disability and early death, scientists turned their attention to the potential effects of nutrients and diet patterns on chronic disease risk. Pioneering studies by Burkitt, Keys, Breslow, and others were followed by a large number of studies on the role of dietary patterns and constituents on certain chronic diseases. Many important studies were completed over the second part of the century, providing the evidence to support specific dietary recommendations to reduce disease risk.

The 21st century ushered the next transition in nutrition science, this time centered on the interrelationships between nutrients, dietary patterns, and the human genome. Over the past few decades, advances in our understanding of the human genome and on the molecular tools to explore it have permitted to probe those interactions in increasing detail. In turn, findings from nutrient–gene interaction studies have informed population-wide and clinical and metabolic studies, further advancing our understanding of the effects of diets on human health at the molecular level. This understanding of the links between genotype, phenotype, and nutrient/dietary intake became a key contributor to the emerging area of personalized nutrition/precision medicine.

All those phases of research emphasis, to different degree, continue to exist today and result in a vast, multidisciplinary, ever-expanding amount of information reaching the peer-reviewed literature. This massive amount of information needs to be organized and summarized in a way that makes it accessible to experts, teachers, and, as much as possible, the general public. This has been and continues to be the goal of the *Encyclopedia of Human Nutrition* since its first edition, over 20 years ago.

Such an ambitious task can only be achieved by the collective work of many people. We all have experienced the challenge of writing an article that combines focus and relevance with conciseness, so we are very appreciative of the work of our contributors. Their effort was backed up by an excellent editorial board, which reviewed and provided feedback on every manuscript. Finally, we must acknowledge the outstanding support of the Major Reference Works division at Elsevier. A publication like this *Encyclopedia* has a lot of moving parts, and it is a great privilege to be able to concentrate on the content, knowing that the other parts of the process are in the hands of excellent professionals.

We hope that this book will help satisfy the need for accurate and concise information to the many students and professionals who are committed to use nutrition science as a tool to improve people's quality of life.

Benjamin Caballero, MD, PhD
Editor In Chief

This page intentionally left blank

Appetite: Psychobiological and behavioral aspects

R.J. Stubbs, G.S. Finlayson, and J.E. Blundell, Appetite Control and Energy Balance Research Group, School of Psychology, Faculty of Medicine and Health, University of Leeds, Leeds, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of R.J. Stubbs, J.E. Blundell, Appetite: Psychobiological and Behavioral Aspects, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 108–115, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00017-9.

Introduction 2
The nature of feeding behavior and appetite control 2
Learned appetites, satieties, and feeding behavior 2
Methodological issues 4
Measuring hunger 4
Measuring food intake 4
Measuring feeding behavior 5
Sensory stimulation and food hedonics 5
Sensory stimuli and body weight 6
Sensory versus nutritional determinants of intake 6
Emotional stimuli and over-consumption 6
Meal patterns, appetite, and energy balance 7
Social and situational influences on feeding behavior 7
Cognitive and social cues 7
Diet composition and appetite 8
How do macronutrients and energy density affect satiety? 8
Diet composition and satiety to prevent weight gain 8
Conclusion 9
References 10

Key points

- Review the nature of human eating behavior and appetite control
- Discuss the role of learning in acquired eating behavior traits.
- Consider methodological issues relevant to the measurement of appetite control and energy balance.
- Discuss the role of environmental factors (meal patterns, social and situational influences) on human appetite control.
- Summarize the role of diet composition in appetite control and weight management.

Glossary

Appetite The subjective expression of willingness or motivation associated with qualitative selection and quantitative *ingestion* of specific foods during an ingestive event. Appetites are specific to certain foods. Appetite is not necessarily solely related to situations of nutritional depletion and can be influenced by a number of physiological and non-physiological factors. Appetites are often learned and frequently sensory specific

Hunger The subjective sensation and associated processes that can be described as a general motivation to eat. An increase in subjective hunger usually predicts meal initiation in *ad libitum* feeding subjects. It does not necessarily predict type or amount of food eaten

Food hedonics Liking (an experience of pleasure) and wanting (anticipatory motivation), are distinct hedonic processes with dissociable neural pathways that are thought to serve as a basis for animals (including humans) to acquire through learning, eating behaviors that lead to the acquisition of energy and nutrients. They may be influenced by the food, the *physiological state* of the organism, and the environment in which food and subject interact

Liking The sensory pleasure elicited by contact with food contributing to the hedonic motivation to consume (wanting)

Wanting The motivation to consume a specific food, manifesting explicitly (craving/desire) or implicitly (food cue responsiveness)

Satiation Processes during a meal that generate the negative feedback leading to its termination (within-meal inhibition) (strengthened by meal volume and weakened by palatability)
Satiety The motivational process of not wanting to eat and often expressed as degree of satisfaction and/or fullness. This bears some reciprocal relationship to hunger. Some authors (Le Magnen, Blundell) argue that satiation and satiety are mechanistically distinct, the former determines meal size; the latter meal frequency

Introduction

Human eating behavior is complex and influenced by physiological, environmental, genetic factors and learned experience. Eating behavior studies commonly focus on assessment of food intake and its physiological, psychological and environmental determinants (Blundell and Stubbs, 2003). Often, these studies are very short-term and have limited capacity to explain mechanisms affecting longer-term energy balance. Because overnutrition is increasingly prevalent and undernutrition is still a significant problem, it is important to consider how appetite control relates to energy balance. It is also necessary to consider some important methodological issues in appetite research and environmental factors (including the composition of the diet) that influence energy intake and longer-term energy balance (Stubbs and Finlayson, 2018).

The nature of feeding behavior and appetite control

Mammalian feeding occurs regularly and intermittently and despite a general lack of conscious nutritional knowledge on the part of the animal, usually appears to match *energy intake* (EI) and nutrient intakes with requirements. How is this achieved? The common explanation is that appetite, EI or feeding behavior are regulated to ensure that physiological requirements are met. However, there is a lack of direct evidence for this regulation. Neither feeding behavior, nor appetite are regulated in a strictly physiological sense because (1) they are not held constant within certain narrow limits and (2) feeding responses are not an inevitable response to an altered physiological signal or need. Feeding behavior is responsive to a number of induced states such as pregnancy, cold exposure, *growth and development*, emotions, and *weight loss*. These responses have often been cited as an evidence of a system that is regulated through homeostatic, symmetrical negative feedback loops. It is probable that some aspects of body size and composition are regulated and changes in feeding behavior are functionally coupled to those regulatory processes. Evidence suggests that the regulation of human energy balance is largely asymmetric. Human eating behavior has evolved in resource-limiting environments to defend against energy deficits (weight loss) rather than excess energy intake (weight gain) (Stubbs and Turicchi, 2021).

Hunger and satiety often have a large learned, anticipatory component rather than solely being the direct consequences of unconditioned physiological signals *per se*, such as reduced *gastrointestinal content*. Such physiological events can act as important cues for feeding but they do not necessarily directly determine that behavior.

The mechanism by which feeding behavior is coupled to physiological events (and other events) is the process of learning. To understand feeding behavior, hunger, and satiety processes, the mechanism by which learning links feeding behavior to physiological, sensory, nutritional, situational, and other learning cues, must be appreciated. This mechanism is termed as associative conditioning of preferences, appetites, and satieties (Blundell and Stubbs, 2003; Mela and Rogers, 2013).

Learned appetites, satieties, and feeding behavior

Animals and humans learn (or become conditioned) to associate a given food with the physiological consequences of having ingested it. They associate certain proximal stimuli such as the smell, color, taste, or texture of a food (the conditioning stimulus), with a set of sensations that are directly felt (sensory afferent inputs), in relation to the external stimulus and to the endogenous changes such as physiological and neuroendocrine responses to food. The physiological changes that occur as a result of ingesting the food are termed as the unconditioned stimulus. The individual forms a learned or conditioned association between the conditioning stimulus and the unconditioned stimulus (the detectable consequence of eating), which informs them of the sensory and physiological consequences of ingesting that food (Blundell and Stubbs, 2003; Mela and Rogers, 2013). This process is summarized in **Fig. 1**. Conditioned or learned associations are most efficiently established if the food is sensorially distinct, if there is a significant detectable postingestive consequence of ingesting the food, or if a training or learning schedule is encountered (e.g., by repeated exposure to the food under similar conditions). Learning is facilitated by social interaction.

As regards with the notion of *appetite regulation*, a problem arises when foods are constructed to look and taste like foods with a different composition. For sometime, after the initial exposure to the food, subjects will respond to it in a manner that is determined not by immediate exposure to the food but by what they have learned during the previous period of exposure to the similar foods upon which the learning was originally based. Only if the food produces a very large unconditioned stimulus will this

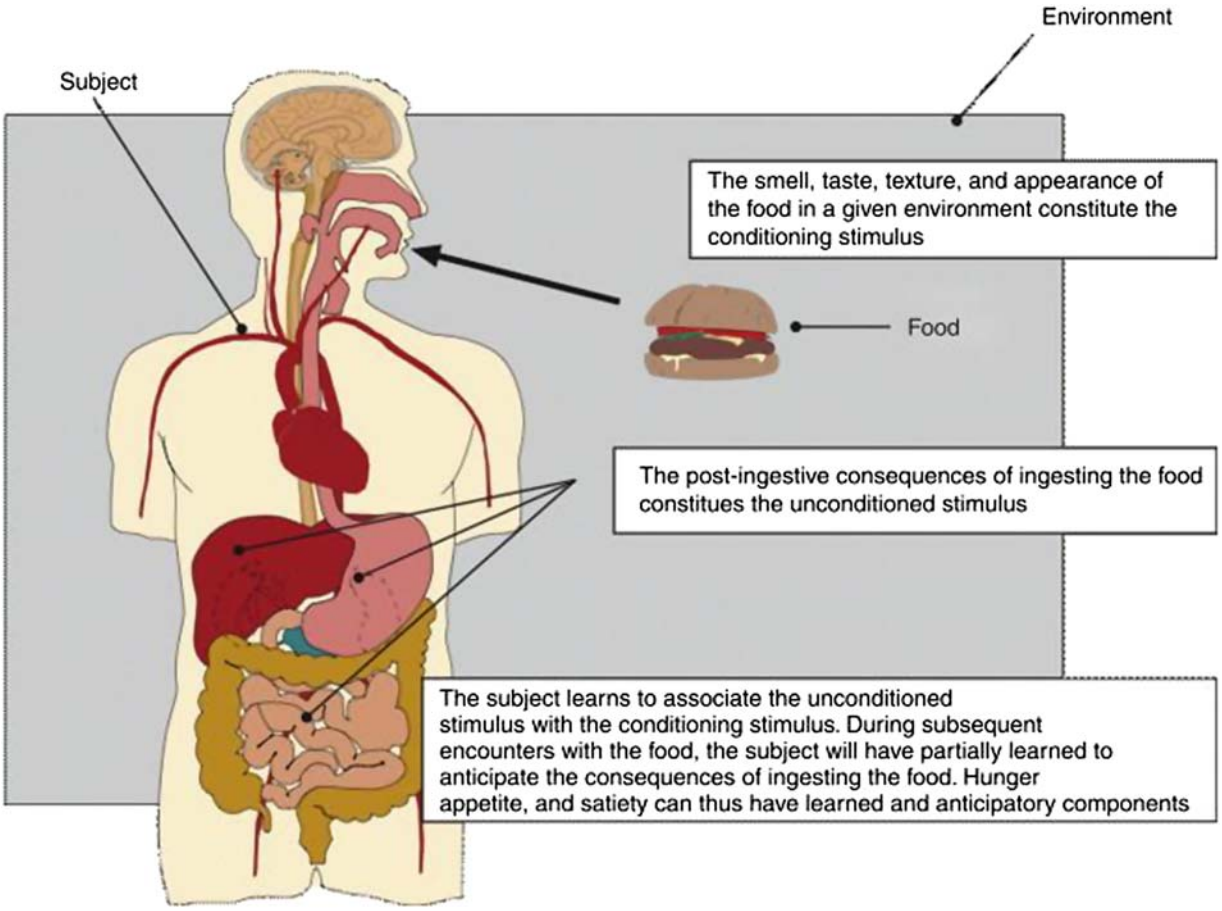


Fig. 1 The process by which the subject learns to associate the postingestive consequences of eating with the food eaten and the environment in which it was eaten. Environmental influences can vary in strength from negligible effect to influences so strong that they can constitute the major factor determining a subject's subsequent response to that food.

previously learned response be instantly over-ridden. This raises the possibility that the use of food mimetics (e.g., high intensity sweeteners) may disrupt stable patterns of learned feeding behavior in consumers at large.

The above view of the nature of feeding behavior has implications for the way the appetite system functions in relation to the development of obesity. Physiological models of body weight (energy balance) regulation suggest that feeding behavior is geared to the regulation of a stable body weight through homeostatic negative feedback. Obesity has therefore been seen as a consequence of defects in the regulation of physiology and behaviors related to energy balance. The evidence from behavioral studies suggests the following: (1) feeding behavior is inherently more responsive to decreases rather than increases in body weight; (2) current secular trends in body weight suggest that, over time, it is very easy to increase body weight, which infers body weight is not tightly regulated, at least with reference to weight gain (for instance, in the National Health and *Nutrition* Examination Survey (NHANES) *dietary surveys* of American adults, average weight increases by 0.2–2.0 kg per year), and (3) there is very limited evidence of clear lean/obese differences in feeding behavior of a type that suggest defects in a regulatory system. For example, evidence suggests that people living with obesity tend to select a diet that is higher fat, energy dense and containing more highly processed foods, which itself facilitates over-consumption. However, the tendency to select fat cannot be viewed as a defect in *physiological regulation, but rather a difference in food preferences that subsequently impact body weight* (Blundell and Stubbs, 2003; Mela and Rogers, 2013; Stubbs and Turicchi, 2021). There are interesting individual differences in response to food-based reward stimuli according to body weight status. Individuals with obesity are more reactive to food cues i.e., their attention is more easily grabbed and held by these cues. In a hungry state, these "attention grabbing" effects become more pronounced in participants with overweight/obesity compared to lean. High BMI also correlates with cravings while dieting and subjective cravings in individuals with overweight is associated with food cue reactivity. Similarly, increased BMI has been associated with more frequent craving, and craving specifically for high fat foods along with increased intake of them. It may be far more profitable to attempt to understand how feeding responds to environmental and endogenous stimuli and which of these responses are functional and adaptive, and which are not (Finlayson et al., 2007). In this context it is important to remember that evolution has selected us to optimize resources in uncertain environments, bank surplus energy, and compensate for energy deficits. Under conditions of environmental uncertainty there is little need

to evolve regulatory systems that protect from weight gain. Furthermore, accumulating body fat provides the ecological advantage of maintaining functional integrity and survival under conditions of food shortage. As a species, humans may have never had the time nor circumstance to evolve patterns of feeding behavior that are protective against weight gain because such evolved behaviors would have little functional or evolutionary advantage. Throughout the development of civilization humans have increasingly manipulated their food production and consumption environment to suit these biological design specifications. This promotes weight gain. Supermarket society is *Optimal Foraging Theory* taken to its logical conclusion. We do not tightly regulate energy balance. Human eating behavior is designed by evolution to feast in times of plenty as protection from famine. Modern industrial marketing strategies play on the constraints our species has evolved with, as an opportunistic *forager*. These considerations influence the methodological approaches that attempt to investigate how feeding behavior responds to endogenous and environmental influences (Stubbs and Tolkamp, 2006; Stubbs and Turicchi, 2021).

Methodological issues

Measuring hunger

Hunger is a subjectively expressed construct that people use to express a motivation to eat. The most appropriate measure of hunger is its subjective expression at a given time. This is achieved by asking subjects to mark a *visual analog scale*, which takes the form of a straight line with two extreme representations of hunger anchored at either end. It is most useful to track changes in subjective hunger over time and in relation to feeding events, diet composition, or physiological parameters. Hunger itself exhibits a large learned component (see the Section on **Learned appetites, satieties, and feeding behavior**, above) as reflected by the fact that most of the variation in a person's subjectively expressed hunger is accounted for by time. If hunger is plotted against time in Western subjects feeding *ad libitum* then it generally exhibits 3 peaks and troughs, which broadly correspond to the 3 main meal times of a Western feeding schedule (**Fig. 2**). Although subjective hunger is a relatively poor proxy for the amount eaten it is a reasonably good predictor of when eating will occur. It is important to recognize that hunger can be influenced by a large number of factors and so a search for a specific unitary hunger signal is likely to prove fruitless. Thus a large survey of over 600 men, women, boys, and girls could find no clear constellation of traits, sensations, or characteristics that typified hunger. A variety of hormones and drugs, the sight and smell of food, its perceived palatability, timing and social situation can all influence hunger (Blundell et al., 2010).

Measuring food intake

Appetite is specific to foods, exhibits wide inter-subject variability, and tends to decline for a specific food as that food is eaten, leading to selection of other foods. Appetite is therefore, said by Le Magnen to be sensory specific (Magnen, 2012). The sensory specificity of appetite has been shown to relate *inter alia* to the postingestive consequences (satiation and satiety) of having ingested a food. The most objective measure of appetite for a given food in a specific experimental situation is therefore the amount of that food that a subject chooses to eat. Appetite is not rigidly determined by physiological signals *per se* although they may greatly influence it. Both the palatability of a food and the appetite for it tend to co-vary and are often increased subsequent to a period of negative energy balance. Two examples are dieting and illness, both of which lead to lowered intake and a subsequent rebound in appetite. As discussed above, the appetite for a food will be learned on the basis of the consequences of having ingested that food on previous eating occasions. Because of this it is possible to use covertly manipulated foods to deceive subjects into behaving in a manner largely determined by prior learned experience. If this were not so, such deception would be impossible because the

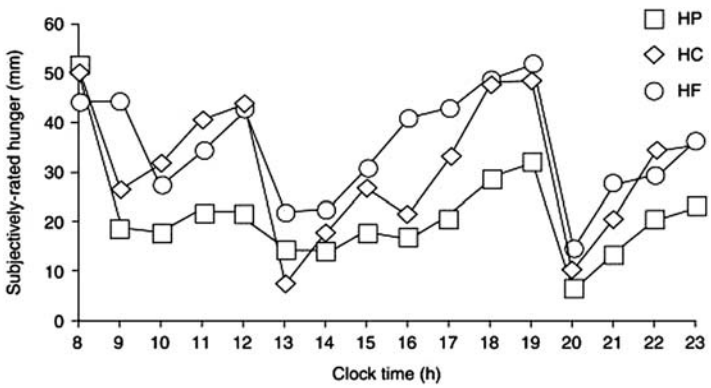


Fig. 2 Subjective hunger tracked during waking hours in six subjects feeding on isoenergetically dense high-protein, high-fat, and *high-carbohydrate diets*. Subjects exhibit the three peaks and troughs of hunger that typify the Western feeding schedule.

physiological signals produced by the sensorially similar, yet nutritionally different food, would immediately translate (through physiological signals) into behavioral compensation. This fact has implications for consumers because food technology can now dissociate the sensory and nutritional properties of foods, which may undermine the learned basis of food intake control in some people.

Measuring feeding behavior

There are many different techniques that can be used to measure feeding behavior in a number of different environmental situations. These techniques are used in naturalistic and/or laboratory studies of feeding. Generally in the laboratory, specific aspects of the appetite and energy balance system are manipulated at the cognitive, sensory, gastrointestinal, or even the metabolic level, for example, by deceiving subjects about the energy content of foods, altering the sensory variety of the diet, changing energy expenditure through systematic alterations in physical activity, administering *nasogastric infusions* or *parenteral infusions*, respectively. A number of other manipulations can be achieved. Because the environment can have such a large influence on feeding behavior, it is important to consider a particular influence on feeding in several environmental contexts. For instance, the effects of fat on *energy intak* have been considered in several laboratory and real life contexts. Under most sets of circumstances increasing the energy density of foods using *dietary fat* appears to be a risk for excess intake. The relationship between the experimental context and how it influences the investigations made is depicted in **Fig. 3** (Blundell et al., 2010; Blundell and Stubbs, 2003; Stubbs et al., 1998).

Sensory stimulation and food hedonics

Food "liking" and "wanting" are emerging constructs in a conceptual approach to appetite where psychological processes of affect and motivation can be viewed as major influences on food intake. Liking and wanting achieve importance in light of the recognition of the contrast between homeostatic and hedonic processes that control eating.

The liking and wanting constructs stem from research exploring the neural basis of palatability and addictive behavior. With principle focus on distinct dopamine and opioid pathways in the brain, the research suggests that processes of liking and wanting can be separately manipulated to produce patterns of behavior that are either exclusively affective or motivational in conjunction with a food stimulus. Liking and wanting are thought to reflect processes that can operate without conscious awareness. This means that they have implicit components (Finlayson et al., 2007). However, their explicit counterparts express themselves subjectively in the form of hedonic feelings from the ingestion of a specific food (i.e., explicit liking) and the intent or desire to consume a specific food (i.e., explicit wanting). Under normal circumstances, explicit liking ("I like this") is closely associated with explicit wanting ("I want this"). However, there is also evidence to suggest that wanting can be "irrational"; i.e., when implicit wanting for a food is greater than explicit wanting, and not proportional to experienced or expected liking (Finlayson et al., 2007).

Liking and wanting appear to have separate and disproportionate roles in promoting overconsumption. In terms of liking, some individuals at risk of weight gain may experience an exaggerated hedonic response to palatable foods, so that foods are enjoyed

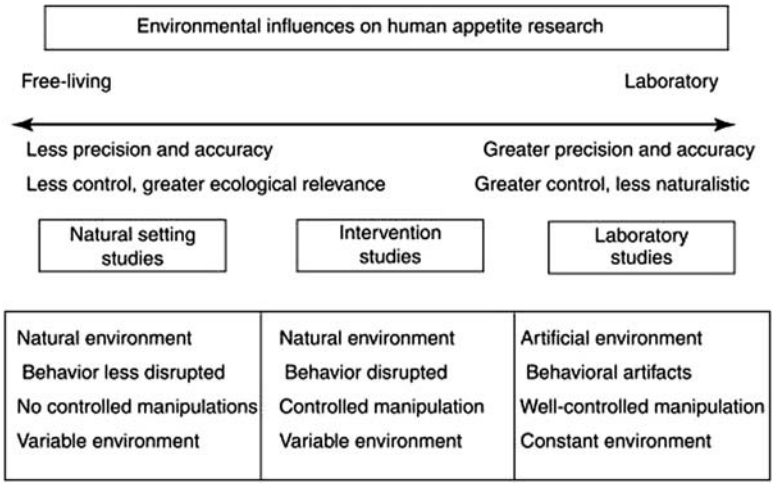


Fig. 3 The constraints and limitations that the experimental environment places on studies of human feeding. In general the environment ranges from totally free living, which is realistic but very difficult to make measurements into the laboratory where measurements are easy but may be contaminated by artifacts due to the artificiality of the laboratory surroundings.

more and therefore eaten in greater amounts for longer periods of time. Conversely, susceptible individuals may have a diminished ability to experience pleasure from food and therefore consumption of palatable food is driven up to satisfy an optimum level of stimulation. Processes of wanting may also bring about vulnerability to weight gain through increased reactivity toward cues signaling the availability of food. Moreover, a reduced ability to resist the motivation to eat when satiated may promote non-homeostatic overconsumption. A widely held notion is that wanting rather than liking may be the crucial process in maintaining an obese state. Moreover, food liking is often a rather stable characteristic within an individual and appears relatively uninfluenced by increasing weight status. The implication is that liking may be important in establishing the motivational properties of food, but once these are retained it is the upregulation of wanting in an obesogenic environment "insensitivity to homeostatic signals but over-reactivity to external cues" that promotes overconsumption by influencing what and possibly how much is eaten from moment to moment (Finlayson et al., 2007).

Sensory stimuli and body weight

It was originally proposed that obese subjects are more susceptible to external stimuli such as sensory stimuli than lean subjects who were more reliant upon internal physiological cues. This predicted that in a Western context the numerous external food stimuli would promote excess EI in susceptible individuals. As can be appreciated from much of this text, the interplay between "external" and "internal" signal is much more complex and interactive than was initially supposed. Nevertheless, given the multiplicity of afferent inputs that can influence feeding it is possible that some subjects have learned to base feeding predominantly on some cues rather than others. However, dividing these cues into simply internal and external sources is perhaps oversimplified. Current questionnaires which attempt to characterize subject's responsiveness to food attempt to dissociate "externality," "restraint," and "emotionality." It is becoming increasingly clear that emotional and reward-based responses to food are critically important in facilitating over-consumption. A theoretical framework that integrates these eating behavior traits into coherent domains that have methodological and predictive validity is still needed (Mela and Rogers, 2013).

Systematic differences have been found in sensory preference profiles, but not perceptions of intensity of various tastes between lean and obese subjects. Work has suggested that subjects with a history of *weight fluctuation* have an enhanced sensory preference for high-fat stimuli that are sweet. It has also been shown that sensory preference for fats is associated with degree of overweight in adults and children. This may be important. Although there is little evidence that sensory stimuli alone promote positive energy balances, there is evidence that people select foods they prefer. Preferential selection of energy-dense foods can lead to excess EI without any apparent change in the amount of food eaten. Thus sensory factors are likely to play a role in the selection of foods, which are conducive to weight gain (Blundell and Stubbs, 2003; Mela and Rogers, 2013).

Sensory versus nutritional determinants of intake

The major problems with the concept of sensory preference or liking as determinants of *hyperphagia* and obesity in humans are that (1) there is little direct evidence for this effect *per se* (because the appropriate experiments are very difficult to do) and (2) both animals and humans appear to acquire sensory preferences for foods dense in readily available energy. Dissociation of the sensory characteristics and postingestive consequences of ingesting a food becomes difficult and perhaps artificial. It seems that at the present time the data from animal studies tends to suggest that maximal sensory preference for a food or diet is achieved when the sensory stimulus is reinforced by the metabolic consequences that form part of the satiety sequence. Indeed, there is controversial evidence that ingestion of one sensory stimulus (sweetness) without the associated nutrient (carbohydrate) promotes ingestion of energy shortly afterward. The contribution of sensory and nutritional determinants of feeding is still poorly understood in humans (Blundell and Stubbs, 2003; Mela and Rogers, 2013).

Emotional stimuli and over-consumption

As an intensely gregarious species we are bound together by numerous relationships ranging from mother–offspring bonds developed at birth, through kin relations to culturally defined positions in any given social structure. Positioning in social structures is also common among the primates and is inextricably intertwined with access to resources, safety (e.g., predatory avoidance), and food. This is critically important to the survival of groups of animals under ecological conditions. Humans have not lost the neuroanatomical bases of our emotions and behavior with which we evolved. Hence, food is as, if not more emotionally important today as it has always been during our evolutionary heritage. However, the social and emotional meanings of food have changed. Food is still important for sharing social emotions and creating social bonds through ceremonies, rituals, and traditions. Food is a major source of reward for achievements or simply as treats. Interestingly, in children, there is evidence that more energy-dense foods are more preferred. More energy-dense foods are also generally cheaper than less energy-dense but nutritionally balanced foods, making the *fast food* outlet phenomenon self perpetuating. Food is a major source of soothing for emotional distress in others, for example, when parents comfort children. In the context of this discussion it is notable that food is also a major source of reward, soothing, and comfort in binge eaters, and is a common theme that emerges, in people struggling to control their weight. Thus, the emotional drives and social facilitators of excess energy intake are powerful, if not more powerful than ever in modern Western society (De Castro, 1997; Finlayson et al., 2007; Mela and Rogers, 2013).

Meal patterns, appetite, and energy balance

The effect of meal patterns on appetite and energy balance is also an unresolved issue. It has been noted that snacking and commercially available *snack* food are often believed to elevate EI. However, there is considerably less evidence that meal or snack patterns contribute to the development of obesity. It is important to note at this point that the relationship between a meal and a snack relates to timing and size of ingestive events in meal feeding animals. In nonhuman species (and indeed humans) that engage in numerous small feeding bouts throughout their diurnal cycle there is little if any distinction between a meal and a snack. Meal-feeding animals are conditioned to ingest the majority of their EI in a few large ingestive events in their diurnal cycle, at approximately the same time points. Under these conditions, a snack can be defined as an energetically small, intermeal ingestive event. To avoid confusion with a common use of the word to describe a certain type of "commercially available food," we use the phrase "commercially available snack foods" to describe those specific foods. Commercially available snack foods tend to differ from the rest of the diet as they are more energy dense, high in fat and carbohydrate and low in protein and usually contain a large fraction of their edible mass as dry matter. They are by no means the only food eaten as a small inter-meal ingestive event by many people at large.

The evidence in relation to meal patterns, appetite, EI, and body weight is indirect and fragmentary. On aggregate, *cross-sectional studies* tend to support no, or a negative, relationship between meal frequency and BMI. However it has been convincingly argued that examinations of the relationship between snacking and energy balance in free-living subjects are extensively flawed by misreporting, misclassification of meals and snacks, and potentially by reverse causality. Under these conditions it is difficult to draw clear conclusions about the effects of snacking in cross-sectional studies. It is therefore important to conduct controlled laboratory interventions over a number of days in humans. These studies suggest that in the short-to-medium term adding mandatory snacks to the diet leads to over-consumption. This effect is most pronounced in those who do not habitually snack and least pronounced in those who do. It is also of note that rats tend to be "snackers" and Western humans tend to be meal feeders. The rat tends to adjust EI by varying meal frequency; the human by varying meal size. However, if rats are *meal fed*, they learn to adjust EI by varying meal size. Humans placed in time isolation begin to adjust intake by varying meal frequency. These comparisons illustrate the fact that adjustment of intake to energy or nutrient requirements occurs within a conditioned time framework, which itself is variable depending on the conditioning environment. Despite large changes in the pattern of feeding, EI can still be adjusted to satisfy requirements.

Social and situational influences on feeding behavior

There are a number of social and situational influences on food intake in humans. In general, the shorter the time period of measurement, the greater the effect of situational and social influences. Thus there are a large number of factors that can influence single meal size in humans. These factors are summarized in **Fig. 4**.

Time of day appears to influence meal size in that the amount eaten and the EI increases on going from breakfast, to lunch, to the evening meal. Meal size also increases across the feeding period in rats. It has been suggested that this occurs in learned anticipation of the energy requirements in the fasting period (night for humans, day for rats). Meal size and EI tend to be greater at weekends than on weekdays, in Western adults. Meal size also varies as a power function of the number of people present at a meal. DeCastro has termed this effect "social facilitation" of feeding. Social facilitation and daily routine account for much of this effect.

Seasonality can influence feeding. A number of studies also suggest that EI, meal size, and eating rate are all elevated in the autumn. In one particular study hunger was associated with meal size in winter and spring, but not so clearly in the autumn (Mela and Rogers, 2013; Stubbs, 2014).

Cognitive and social cues

Throughout the 1960s and 1970s a large number of behavioral studies examined the effect of cognitive and social cues (perceived energy content of foods, salience of cues, eating behavior of others present) on feeding in relation to the externality hypothesis. Although a large number of studies found that so called external cues do relate to short-term feeding behavior, a large number of others did not. However, the presence of external cues alone does not reliably predict how much food people will eat. Neither does the presence of external cues always relate to lean/obese differences in feeding patterns. Some of these differences in relation to cognitive and social cues are better explained in relation to dietary restraint. Restraint is a term used to describe people who are attempting to limit or reduce their body weight by means of cognitive energy restriction (dieting). In doing so it is proposed that they are placing their motivation in relation to feeding at odds with physiological feeding stimuli. Placing cognition at odds with physiological drives can result in pathologies of eating since the normal "regulatory" processes are cognitively undermined. Furthermore it is argued that restraint will increase the probability that a person will break a diet. It has been shown that an intervention (usually a preload) that breaks the rules of restraint, almost paradoxically induces a greater intake. This phenomenon has been termed counter-regulation. This effect is cognitive, because it can be induced by deceiving a restrained eater into believing that a *preload* was high in *calories* (Mela and Rogers, 2013). Because the concept of restraint has predictable behavioral outcomes it is a useful tool in characterizing different people with respect to their feeding behavior. However, it is now generally

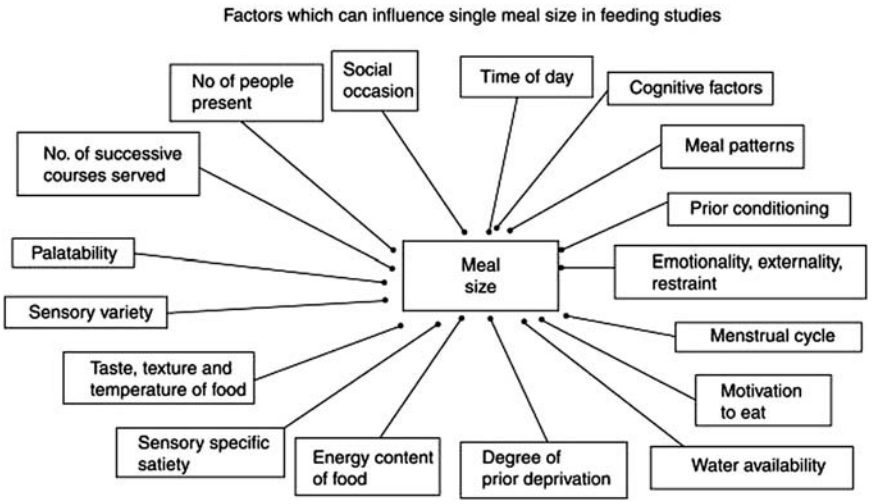


Fig. 4 Major factors known to affect single meal size in humans. In general, the shorter the time interval of measurement, the greater the influence these factors have on the observed feeding behavior.

accepted that restraint is not a unitary construct and people who attempt to lose weight and who show flexible restraint are more successful than those who are rigid in their patterns of cognitive restraint.

It is useful to consider the role of dual process models of behavior in the context of energy balance (weight change). Models of behavior and behavior change have focused historically on social cognition (e.g., beliefs, intentions, attitudes and decisions), emphasizing pathways of reasoned action in which pre-decisional motivation leads to the formation of intentions and the implementation of those intentions as volitional action. Automatic processes (emotions, desires, habits resulting from associative learning and physiological states) may also impact eating behavior and behavior change. These processes tend to be relatively rapid, impulsive (less conscious) and habitual in comparison to the slow, deliberative processes of motivation and cognitive self-regulation. Furthermore, in the context of appetite and energy balance behaviors, the development of self-regulatory behavior change is effortful, particularly in the face of physiological responses to weight loss, while unconscious or automatic components of EB behaviors are rapid and effortless. Physiological mediators of homeostatic and hedonic appetitive drives, and changes in physical activity that are triggered by weight loss may feed into such automatic process of behavior change to undermine self-regulation of eating behavior.

Diet composition and appetite

Food and nutrient *ingestion* influence human appetite through multiple feedbacks at several levels, which can be traced through the processes of food location, ingestion, digestion, absorption, and metabolism. Satiety is therefore maintained by a functional sequence or cascade of sequential physiological events that reinforce each other. Removing parts of a food or nutrient's effects on this sequence will therefore diminish its impact on satiety.

How do macronutrients and energy density affect satiety?

Because the control of food intake in humans is imprecise, alterations in the energy density of the diet can lead to changes in *energy intake*, even over quite prolonged periods. Macronutrients are not all equal when it comes to how satisfied one feels after eating them. Protein has the greatest effect on satiety, followed by carbohydrate, followed by fat. Studies using smaller *preloads* (typically of 250 kcal or less) do not always find this effect. Although fat is the most energy-dense macronutrient it is also the least satisfying. The effect macronutrients have on satiety is illustrated in Fig. 5. Foods that are high in protein include lean meat, fish, beans, peas, *lentils*, *mycoprotein*, *tofu*, and eggs. Foods rich in carbohydrates and fiber include wild rice, potatoes, cereals and *oats*, *bread*, beans, peas, lentils, and couscous. Fat-rich foods include *savory snacks*, pies, biscuits, cakes, cream, and most cheese (Stubbs et al., 2010).

Diet composition and satiety to prevent weight gain

Any *weight reduction* strategy that focuses on one simple aspect of diet alone will have limited effects on *weight control*. Dietary *monotherapies* are very limited in what they will achieve. Several aspects of diet composition can be used simultaneously to influence

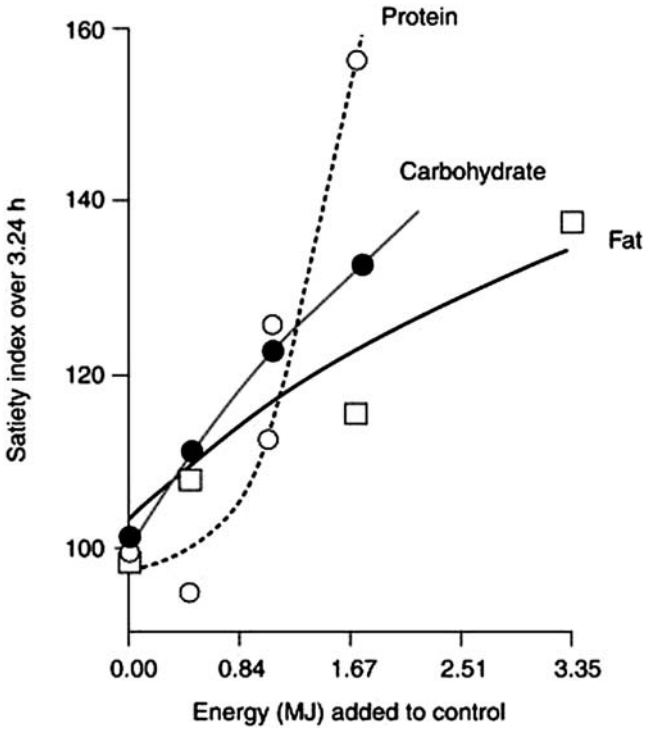


Fig. 5 Effect of increasing energy content of macronutrient loads on satiety Index subjectively expressed over 3.25 h.

satiety and help people navigate to a healthy body weight. In other words, we should be taking a package definition of what constitutes a reasonable dietary approach to weight management and that should be integrated with evidence-based behavior change approaches. Dietary factors that help with self-management of eating behavior include a low fat content, tolerably low energy density of ~1.1–1.3 kcal g^{-1}, a high water content; avoiding where possible, caloric beverages, fiber content should be tolerably high (which parenthetically is not much more than 20–30 g per day). Fiber holds water in foods, which lowers energy density, and decreases the rates at which those foods are digested. To enhance satiety, protein content of the diet can be relatively, tolerably, and reasonably high (probably not exceeding 25–30% of an energy-reduced diet, for example, a diet that is followed to induce a negative energy balance. This translates into less than 20–25% of energy requirements from protein). *Protein intake* from red and processed meats should be limited. The carbohydrate content depends very much on type. There is a good deal in the literature that suggests sweet, short-chain carbohydrates can elevate *energy intake*, especially when combined with fat. This combination is most common in commercial processed and s*nack foods*. The orosensory properties of foods need to be maintained to encourage people to select those foods, so that they become a practical option for the development of healthy eating habits (Stubbs et al., 2010).

Conclusion

Human appetite is influenced by both physiological and psychological cues which themselves are shaped by the environment. Appetite control and energy balance regulation are geared to defend against weight loss and less tightly geared to defend against weight gain. Many of the environmental eating-related cues that are part of modern society also act as stimuli that promote higher levels of food and energy intake, reinforced by an interaction of the sensory properties of foods with liking and wanting circuits. Because many aspects of eating behavior are learned through associative conditioning they are often habitual and not under immediate, direct conscious control. These factors facilitate weight gain in modern environments, but may also provide some potential opportunities for obesity prevention and long-term management of body weight. Gaining weight and excess adipose tissue across the adult lifespan requires little conscious effort. Losing weight and preventing weight regain requires considerable effort and weight loss attempts are often subject to subsequent weight regain. Because most of energy intake and less than half of energy expenditure is behavior, appetite control and self-management of eating behaviors are major targets for many weight management interventions. The effects of behavior change interventions on weight management are still relatively modest and our understanding of the factors that disrupt and undermine self-management of eating behaviors is limited. These factors include physiological resistance to weight loss, gradual compensatory changes in eating and physical activity and reactive processes related to stress, emotions, rewards and desires that influence eating to meet psychological needs. In the future it is likely that more interdisciplinary research between the areas of appetite control, energy balance and behavior change research, together with methodological developments in these areas

will lead to more effective interventions to improve weight and health outcomes in the general population. Better matching evidence-based intervention content to the needs of individuals and better objective tracking of appetite and energy balance behaviors may improve outcomes in the future.

See Also: Dietary intake measurement: Methodology; Food choice: Behavioral aspects

References

Blundell, J., DE Graaf, C., Hulshof, T., Jebb, S., Livingstone, B., Lluch, A., Mela, D., Salah, S., Schuring, E., Van Der Knaap, H., 2010. Appetite control: methodological aspects of the evaluation of foods. Obes. Rev. 11, 251–270.

Blundell, J.E., Stubbs, J., 2003. Diet composition and the control of food intake in humans. In: Handbook of Obesity. CRC Press.

De Castro, J.M., 1997. How can energy balance be achieved by free-living human subjects? Proc. Nutr. Soc. 56, 1–14.

Finlayson, G., King, N., Blundell, J.E., 2007. Liking vs. wanting food: importance for human appetite control and weight regulation. Neurosci. Biobehav. Rev. 31, 987–1002.

Magnen, J., 2012. Neurobiology of Feeding and Nutrition. Academic Press.

Mela, D.J., Rogers, P.J., 2013. Food, Eating and Obesity: The Psychobiological Basis of Appetite and Weight Control. Springer.

Stubbs, J., Whybrow, S., Lavin, J., 2010. Dietary and lifestyle measures to enhance satiety and weight control. Nutr. Bull. 35, 113–125.

Stubbs, R., 2014. Controlling appetite and food intake by regulating eating frequency and timing. In: Managing and Preventing Obesity: Behavioural Factors and Dietary Interventions, vol. 149.

Stubbs, R., Finlayson, G., 2018. Improving laboratory studies of human eating behaviour: energy balance and real-world considerations. Appetite 130, 334–335.

Stubbs, R., Johnstone, A., O'reilly, L., Poppitt, S., 1998. Methodological issues relating to the measurement of food, energy and nutrient intake in human laboratory-based studies. Proc. Nutr. Soc. 57, 357–372.

Stubbs, R., Tolkamp, B., 2006. Control of energy balance in relation to energy intake and energy expenditure in animals and man: an ecological perspective. Br. J. Nutr. 95, 657–676.

Stubbs, R.J., Turicchi, J., 2021. From famine to therapeutic weight loss: hunger, psychological responses, and energy balance-related behaviors. Obes. Rev. 22, e13191.

Effects of diet on behavior

E.L. Gibson, Roehampton University, London, UK
M.W. Green, Aston University, Birmingham, UK
S.C. Dyall, Roehampton University, London, UK

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of E.L. Gibson, M.W. Green, S.C. Dyallo, Effects of Diet on Behavior, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 129–141, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00022-2.

Introduction **12**
Effects of Meals **12**
Meal Timing 13
Breakfast 13
Midday Meal 14
Evening Meal 14
Snacks 14
Meal Size 14
Meal Composition 14
Carbohydrate Versus Protein 14
Effects of Dietary Fat 16
Carbohydrates, Stress, Mood, and Mental Function **17**
Susceptibility to Mood Enhancement by Diet 17
Sugars and Opioids 17
Glucose, Mood, and Mental Function **17**
Hyperactivity and Antisocial Behavior 18
Micronutrients and Mental Function 20
Lipids 20
Cholesterol 20
Essential Fatty Acids 21
Food Deprivation 21
Functional and Pharmacological Components of Foods and Drinks 23
Conclusion **23**
Further Reading **23**

Glossary

Arachidonic acid AA, is the trivial name for cis-5,8,11,14-eicosatetraenoic acid (20:4n-6). This polyunsaturated fatty acid can be biosynthesised from the omega-6 essential fatty acid, linoleic acid, and is abundant in the phospholipid membranes of several tissues, including the brain. It is a precursor of several important signaling molecules such as eicosanoids.

Attention Deficit Hyperactivity Disorder ADHD is a behavioral disorder, mainly of children, that includes a group of symptoms such as short attention span, easily distracted, restlessness, constant fidgeting, and impulsiveness. ADHD is estimated to affect 3–9% of children and about 2% of adults: it is normally diagnosed between the ages of 3 and 7 years. There is no cure, but the disorder can be managed by a combination of behavior therapy and drug therapy, in particular the use of psychostimulant drugs that enhance activity of the neurotransmitters, noradrenaline (norepinephrine) and dopamine. These may induce more focussed attention while lessening impulsivity and limiting behavioral repertoires. However, the perceived overprescription of such drugs to children remains a controversial topic.

Docosahexaenoic acid DHA, 22:6n–3, cervonic acid. This polyunsaturated fatty acid is abundant in fish oils (e.g., tuna oil) and is a significant component of the membrane phospholipids of most tissues, especially the brain, sperm, and the retina of the eye. It can be biosynthesised from the omega-3 essential fatty acid, α-linolenic acid, but the rate of conversion is very low.

Eicosanoids Metabolites of 20-carbon polyunsaturated fatty acids, such as arachidonic acid (AA), and eicosapentaenoic acid (EPA). They possess both autocrine and paracrine signaling roles, and have actions in many biological processes such as, inflammation, fever, regulation of blood pressure, immune system modulation, and control of reproductive processes.

Eicosapentaenoic acid: EPA, 20:5n–3, timnodonic acid. This polyunsaturated fatty acid is present in most fish oils, and is found at low levels in tissue membrane phospholipids. It is a precursor of some eicosanoids, which tend to antagonise the actions of arachidonic acid derived eicosanoids.

Essential fatty acids Polyunsaturated acids of the omega-6 and omega-3 families, which are essential for life and good health. The only true essential fatty acids are the omega-6 and omega-3 fatty acids, linoleic acid and α-linolenic acid, respectively. The other omega-6 and omega-3 fatty acids can be derived from these respective precursors, albeit at low levels.

Hypothalamic–pituitary–adrenal axis The HPA axis is a representation of a cascade of hormonal signaling known to be sensitive both to stress and nutritional state. Such events alter activity of selective nerve cells in the hypothalamus (a collection of nuclei in the base of the brain), which release corticotrophin releasing hormone (CRH). This in turn activates cells in the adjacent anterior pituitary gland to release adrenocorticotrophic hormone (ACTH) into the blood circulation. ACTH is carried to the adrenal glands, near the kidneys, where cells are stimulated to release cortisol, a glucocorticoid steroid hormone, into the circulation. Cortisol has anti-inflammatory and glucose counter-regulatory activity.

Neurotransmitters Chemical messengers that transfer neuronal activity signals across the synapse (microscopic gap) between one nerve cell and another. Electrical activity in the presynaptic nerve cell causes release of pockets (vesicles) of neurotransmitter into the synapse from where they reach receptors on the postsynaptic cell, whose activation may either excite or inhibit activity in the postsynaptic nerve cell. Neurotransmitters discussed here include the monoamine group, serotonin (5-hydroxytryptamine, 5-HT), adrenaline (epinephrine), noradrenaline (norepinephrine) and dopamine. This group is synthesized from essential amino acids (i.e., that must be obtained from the diet): tryptophan for serotonin, and tyrosine for the others. This is of particular interest here because rate of synthesis of these neurotransmitters is sensitive to availability of their precursor amino acids, and thus to the latter's availability from the diet. Another transmitter group discussed is opioid peptides, derived from larger precursor proteins, but synthesis of these is not sensitive to diet.

Omega-3 polyunsaturated fatty acid Also called ω-3, or n-3 fatty acids. These are named after their shared structural motif of having their terminal double bond three carbons from the methyl (or omega) end of the carbon chain. They are derived from α-linolenic acid, by carbon chain-elongation and desaturation and include eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA).

Omega-6 polyunsaturated fatty acid Also called ω-6 (omega-6) acids. These are named after their shared structural motif of having their terminal double bond six carbons from the methyl (or omega) end of the carbon chain. They are derived from linoleic acid, by carbon chain-elongation and desaturation, and include arachidonic acid (AA).

Introduction

Effects of diet on behavior have long been topics of folklore superstition and popular mythology, and more recently the subject of rigorous, and not so rigorous, scientific study. Most research into dietary effects on human behavior has assessed changes in mood or mental function after eating (or drinking), or after fasting. Typical measures of mental function include tests of reaction time, attention, memory, problem solving ability and intelligence (intelligence quotient; IQ). In addition, research has addressed effects of diet on disturbed behavior, including Attention Deficit Hyperactivity Disorder (ADHD) in children, antisocial behavior and aggression, mental illness such as depression, and dementia (see **Table 1**).

Clearly, chronic malnutrition can seriously affect behavior by impairing brain development, and acutely by denying sufficient nutrients for optimal cognitive function. However, this article concentrates on dietary effects on behavior that are not the result of chronic malnutrition, nor of pharmacologically active ingredients of the diet such as alcohol or caffeine. Rather, they arise from more subtle effects of variation in nutrient intake within the normally nourished population.

Effects of Meals

The commonest way in which food can affect behavior is the change in mood and arousal that occurs from before to after eating a meal. This might sound trite, but it is not trivial: this general meal effect is probably the most reliable example of an effect of diet on behavior. Many animals, including human beings, tend to be aroused, alert, and even irritable when hungry. This encourages their search for food. However, their mental efforts become distracted by this task, to the detriment of other behaviors. After eating a satiating meal, we and other animals become calm, lethargic, and may even sleep.

Nevertheless, even this seemingly straightforward phenomenon can be distorted, and can vary across individuals and situations. The impact of a food or drink will depend on the person's initial state. For example, thirsty people improved their vigilance when allowed to drink water, whereas when people were asked to drink when not thirsty, their performance deteriorated. Numerous experiments have shown that manipulation of the structure of meals results in variation in postprandial changes in mood and mental function. One obvious facet of meals that has been investigated is what is eaten, i.e., nutrient composition; the other two main aspects of meal structure that have been studied are meal timing and meal size. Of course, the effect of a meal on appetite also represents a behavioral effect, but this aspect is covered elsewhere in this encyclopedia.

Besides any nutritional effects, two other influences on behavior are also known to interact with attempts to measure dietary effects on behavior. First, most people are very habitual in their choice of food, and size and timing of meals. As a result, they

Table 1 Examples of nutritional variables known or suspected to affect behavior, mood, and cognition

Category	*Variable description*
Food restriction	Early life undernutrition Chronic semistarvation Dieting to lose weight Short-term fasting (e.g., missing a meal)
Meal effects	Pre to postmeal changes Meal timing (e.g., morning, afternoon, night) Meal size Macronutrient composition (acute and chronic effects) Breast milk
Amino acids	Neurotransmitter precursors (e.g., tryptophan, tyrosine, phenylalanine) Phenylketonuria
Sugars	Sucrose (dietary intake) Glucose (supplement; tolerance)
Lipids	Essential fatty acids: Arachidonic acid (omega-6 PUFA) Eicosapentaenoic acid (omega-3 PUFA) Docosahexaenoic acid (omega-3 PUFA)
Micronutrients	Iodine Iron Selenium B-vitamins: B1, B6, B12, Folate Vitamin C Vitamin E Zinc
Diabetes	Acute effects of hypoglycemia Chronic effects
Pharmacological	Caffeine Alcohol Nutraceuticals (e.g., plant compounds)

have learned a set of beliefs and expectations about the impact of their habitual dietary regime. Therefore, particularly in short-term tests, these expectations may over-ride or mitigate physiological changes. Dietary experiences that differ from a person's habitual eating could lead their behavior to change through cognitive rather than (or as well as) physiological influences.

Secondly, there are circadian rhythms and sleep–wake cycles in arousal and performance, which complicate interpretation of meal effects, as we discuss in the next section on meal timing.

Meal Timing

Does the timing of a meal in the day make a difference to any effects on behavior? In other words, do any behavioral effects differ between breakfast, midday, and evening meals, or mid-morning and afternoon snacks?

Breakfast

The potential effects of breakfast on performance and well-being continue to attract much interest, not least from industry, especially concerning performance of schoolchildren. Pollitt and colleagues have argued that children are likely to be more susceptible to the effects of fasting than adults, due to greater brain metabolic demands relative to glycogenic and gluconeogenic capacity. The numerous studies in this area have produced inconsistent results, which is partly attributable to variation in populations studied, their nutritional status, and designs used. There is a consensus that breakfast is more likely than not to benefit schoolchildren's performance, particularly if the children are already nutritionally vulnerable and have mental abilities with room for improvement. Moreover, breakfasts achieving slower release of glucose into the blood may be more effective in sustaining performance over the morning than those allowing rapidly absorbed glucose.

In all of us, there is a tendency for levels of arousal and alertness to rise during the morning, reaching a peak near midday. Some evidence suggests that breakfast may help to control this arousal, so that attention can be successfully focused on the task in hand. Conversely, omitting breakfast may increase autonomic reactivity, leading to less focused attention. This effect could explain one finding that children without breakfast showed better recall of objects to which they had not been asked to attend: such attention to irrelevant stimuli is also known to occur with increased anxiety. Furthermore, increasing hunger is likely to be distracting.

Less attention has been paid to effects of breakfast in adults. However, there are several studies of effects of giving breakfast to students that show a benefit on spatial and verbal recall tasks 1–2 h later, compared to missing breakfast. Interestingly, attention-based and reaction-time tasks were not improved by breakfast, and a logical reasoning task was even slightly impaired. Perhaps those tests benefit more from mild arousal, which could be acutely reduced by some breakfasts. These studies did not determine whether performance later in the morning would be affected by breakfast. Differential effects of breakfast content and size will be discussed below.

Midday Meal

Several studies have demonstrated a drop in performance after the midday meal, particularly for vigilance tasks requiring sustained attention. However, this 'postlunch dip' may not simply be an effect of eating, because vigilance has also been found to decline from later morning to early afternoon in subjects not eating lunch. That is, there is an underlying circadian rhythm in performance that is confounded with the effect of a midday meal. In fact, using noise stress to arouse subjects during a midday meal prevented any decline in performance due to the meal. It has also been shown that the more anxious one is feeling before lunch, the less one will experience any postlunch dip in performance. In support of this, another study found that subjects scoring highly on a personality measure of extraversion and low on neuroticism were more likely to be affected by postlunch dip. These are examples of the importance of individual differences and context on meal effects.

Evening Meal

There are few studies of effects of eating later in the day, although there has been some interest in effects of meals during nightshifts. Accuracy of performance declines with eating during a nightshift, but unlike lunch, premeal anxiety levels had no effect. One study in students of effects of eating a large freely chosen evening meal found little evidence for consistent changes in performance relative to missing the meal. Despite this, the students who omitted the meal reported feeling more feeble and incompetent and less outgoing than those who had eaten.

Snacks

One study specifically addressed whether an afternoon snack (approximately 1–1.2 MJ, 240–290 kcal, of yoghurt, or confectionery) eaten 3 h after lunch (or no lunch) would affect task performance. A beneficial effect of the snack was found on memory, arithmetic reasoning and reaction time 15–60 min later. The comparison was with performance after a 'placebo' zero-energy drink (participants were unaware of energy content). This rather different placebo does not preclude effects due to differences in sensory experience and expectations. Moreover, whether or not lunch had been eaten beforehand had little effect on the outcomes, suggesting that any nutritional effects must be due to acute impact of the snack, irrespective of prior nutritional state. It is known that snacks of this size eaten after a meal have only a small effect on blood glucose, although insulin rises sufficiently to inhibit lipolysis and suppress the release of plasma free fatty acids later in the postprandial period.

The authors reported that these performance benefits from an afternoon snack were not found with a snack taken late morning. The most likely reason is that the beneficial effect depends on the decline in alertness normally occurring during the afternoon.

Other studies have found differential effects of macronutrient content of snacks; these are discussed below.

Meal Size

This topic has been little studied regarding behavioral effects, perhaps because there are a number of methodological difficulties and an absence of theory. For example, what counts as a large or small meal? Should the difference be in terms of absorbed energy, or weight or volume eaten, or even consumption time? If the former, then behavioral outcomes would need to be measured with sufficient delay for differences in energy absorption to be discriminable. Moreover, the influence of expectations and habit might confound experimental nutritional differences.

Two studies in adults found that large lunches (at least 4 MJ, 1000 kcal) impaired vigilance relative to eating small or medium-sized lunches. There was also evidence that this effect depended on the meal size being different from that habitually consumed. In adolescents, a larger breakfast (2.6 MJ, 634 kcal on average) resulted in poorer vigilance but better short-term memory 3 h later, compared to after a smaller breakfast (1.6 MJ, 389 kcal on average). Thus, there is some evidence that vigilance is adversely affected by a large meal.

Meal Composition

Carbohydrate Versus Protein

The effects of varying the nutrient composition of meals have been studied extensively, and rather more for mood than performance. This is largely because of evidence that plasma and brain levels of precursor amino acids for synthesis of monoamine

neurotransmitters (chemicals responsible for signaling between nerve cells), strongly implicated that mood disorders, can depend on carbohydrate:protein ratios in the diet. Synthesis of the neurotransmitter serotonin (or 5-hydroxytryptamine; 5-HT) depends on dietary availability of the precursor essential amino acid, tryptophan (TRP), due to a lack of saturation of the rate-limiting enzyme, tryptophan hydroxylase, which converts TRP to 5-hydroxytryptophan (see **Figure 1**). An important complication is that TRP competes with several other amino acids, the large neutral, primarily branched-chain, amino acids (LNAA), for the same transport system from blood to brain. If the protein content of a meal is sufficiently low, such as 5% (or less) total energy as protein, then relatively few amino acids will be absorbed from the food in the gut. At the same time, insulin will stimulate tissue uptake of competing amino acids from the circulation, and the plasma ratio of TRP to those of amino acids (TRP/LNAA) will rise, favoring more TRP entry to the brain. Conversely, a high-protein meal, which would be less insulinogenic, results in absorption of large amounts of competing amino acids into the blood, especially the branched-chain amino acids, leucine, isoleucine, and valine. On the other hand, TRP is scarce in most protein sources, and is readily metabolized on passage through the liver: thus, the plasma ratio of TRP to competing amino acids falls after a protein-rich meal. Indeed, the protein-induced reduction in plasma TRP ratio often seems to be more marked than any carbohydrate-induced rise. Such effects also depend on the interval since, and nutrient content of the last meal.

This evidence is particularly relevant to dietary effects on mood and arousal, because 5-HT has long been implicated in sleep, as well as affective disorders such as depression and anxiety. However, cognitive performance might also be affected, given the known role of 5-HT in responsiveness to environmental stimuli and stressors, impulsivity, and information processing. Importantly, there is evidence that dietary availability of TRP can influence brain function in humans: for instance, feeding a TRP-free diet, which considerably reduced plasma TRP (and so could be expected to impair 5-HT function) induced depression in previously recovered depressives or in people with a genetic predisposition to depression. Furthermore, a TRP-free drink has been shown to impair performance on tests of visuospatial and visual discrimination learning, as well as memory. In addition, TRP depletion enhances accuracy of predicting events associated with negative consequences, and reduces accuracy of emotional face recognition, supporting the theory that 5-HT normally biases attention in favor of positive events. Typically, manipulations that should increase 5-HT synthesis and release, such as consuming TRP-rich proteins, produce opposite behavioral effects.

There is evidence that people feel calmer and more sleepy after snacks or meals rich in carbohydrate but virtually free of protein (an unusual situation) than after protein-rich meals with little carbohydrate. This is compatible with changes in 5-HT function, but these studies did not determine whether this is due to an increase in 5-HT after the carbohydrate-rich meal, or a decrease after the protein meal, which could prevent the postprandial sleepiness. Furthermore, adding more than 5 or 6% protein (of total energy) to the carbohydrate meal has been shown to prevent the increased synthesis of central 5-HT, relative to fasted levels, in both rats and people (see **Figure 2**). Also, even pure carbohydrate does not appear to induce sleepiness in everyone.

Another difficulty in comparing effects of carbohydrate and protein intake is that relative changes in mood and performance might be due to protein-induced raised plasma tyrosine (TYR), the precursor amino acid for synthesis of the catecholamine neurotransmitters (adrenaline, noradrenaline, dopamine), which also competes with LNAA for entry into the brain. In catecholamine systems where the neurones are firing rapidly, acute physiological increases in brain TYR, for example, by feeding a high-protein diet, can raise TYR hydroxylation rate and catecholamine turnover. Such systems include dopaminergic neurones involved in

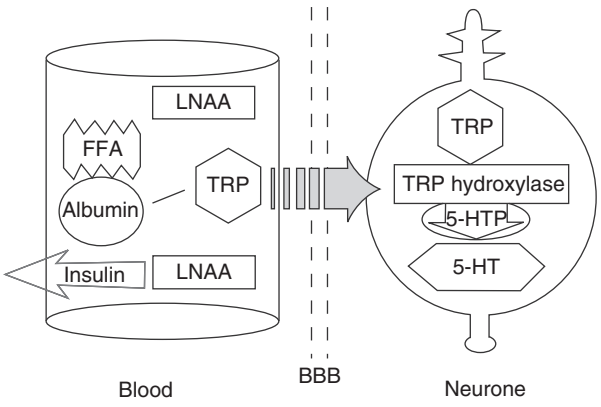


Figure 1 Diagram representing pathways to the synthesis of the neurotransmitter 5-hydroxytryptamine (5-HT, serotonin) from the precursor essential amino acid, tryptophan (TRP). TRP is taken up by neurones from blood, but its passage across the blood brain barrier (BBB) is in competition with another group of essential amino acids known as the large neutral amino acids (LNAA). Thus, the ratio of TRP to total LNAA (TRP:LNAA) determines how much TRP enters the brain. Most TRP is normally bound to albumin in plasma, so not available for uptake into the brain. However, after a carbohydrate-rich low-protein meal, increased release of insulin raises levels of free fatty acids (FFA) in plasma, and these displace TRP from albumin. In addition, insulin promotes tissue uptake of the LNAA from plasma. Hence, the TRP:LNAA ratio increases and more TRP enters the brain. Increased availability of TRP in neurones drives greater synthesis of 5-HT because the rate-limiting enzyme, TRP hydroxylase, which converts TRP to the intermediate 5-hydroxytryptophan (5-HTP), is not fully saturated.

arousal, attention, and motivation. Nevertheless, high-protein meals in human beings do not always raise the plasma TYR/LNAA ratio, depending on nutritional status or time of day.

Differential effects on performance have been seen with less extreme variations in protein and carbohydrate intake. For example, a lunch of 55% energy as protein and 15% as carbohydrate produced faster responses to peripheral stimuli, but greater susceptibility to distraction, compared to eating the reverse proportions of protein and carbohydrate. Sleepiness was not affected by macronutrient composition in that study. However, with these protein:carbohydrate ratios, the plasma TRP/LNAA ratio could still be lowered by the protein-rich meal relative to the carbohydrate-rich one, even if TRP/LNAA does not rise from premeal levels after a carbohydrate-rich meal with much more than 5% protein (Figure 2).

A delay of at least 2 h after eating may be necessary to allow neurotransmitter precursor changes to influence behavior. Earlier effects may be related to changes in glucose availability, and levels of insulin and counter regulatory hormones such as adrenaline, glucagon, and cortisol. These changes could underlie recent results with breakfasts of 20:80, 50:50, and 80:20% protein:carbohydrate ratios (1.67 MJ, 400 kcal). A measure of central attention improved initially after the carbohydrate-rich breakfast, but then later improved after the protein-rich ones; the opposite was found for peripheral attention. This study also found that the 80% protein breakfast produced the best short-term memory performance approximately 1–2 h after eating, but not at 3.5 h.

Effects of Dietary Fat

Most studies of effects of fat have varied its level with that of carbohydrate, while keeping protein constant and so allowing equicaloric meals. Comparisons have been made for low-fat (e.g., 11–29% of energy as fat), medium-fat (e.g., 45%), and high-fat (e.g., 56–74%) breakfasts, midmorning, and midday meals, as well as intraduodenal infusions of lipid or saline. On balance, high-fat meals appear likely to increase subsequent fatigue and reduce alertness and attention, relative to high-carbohydrate/low-fat meals. However, there are inconsistencies relating to changes in specific moods and effects of meal timing: for instance, feelings of drowsiness, confusion, and uncertainty were found to increase after both low- and high-fat lunches but not after a medium-fat lunch. One possibility is that mood may be adversely affected by meals that differ substantially in macronutrient composition from habitual ones. An alternative is that similar mood effects could be induced (albeit by different mechanisms) by high carbohydrate in one meal, and high fat in the other: for example, 1.67 MJ (400 kcal) drinks of pure fat or carbohydrate taken in the morning both increased an objective measure of fatigue relative to a mixed-macronutrient drink, although the two single-nutrient drinks had opposite effects on plasma TRP/LNAA ratios; this is of course an unusual situation and may not generalize to normal mixed-nutrient meals.

In many of these studies, the meals were designed to disguise variation in fat level from participants. It is therefore possible that effects on mood may have resulted from discrepancies between subjects' expectations of certain postingestive effects, and the actual effects that resulted from neurohormonal responses to detection of specific nutrients in the duodenum and liver. A case in point may be the increase in tension, 90 min postlunch, with increasing fat intake, reported by predominately female subjects: this might reflect an aversive reaction to (unexpected) fat-related postingestive sensations.

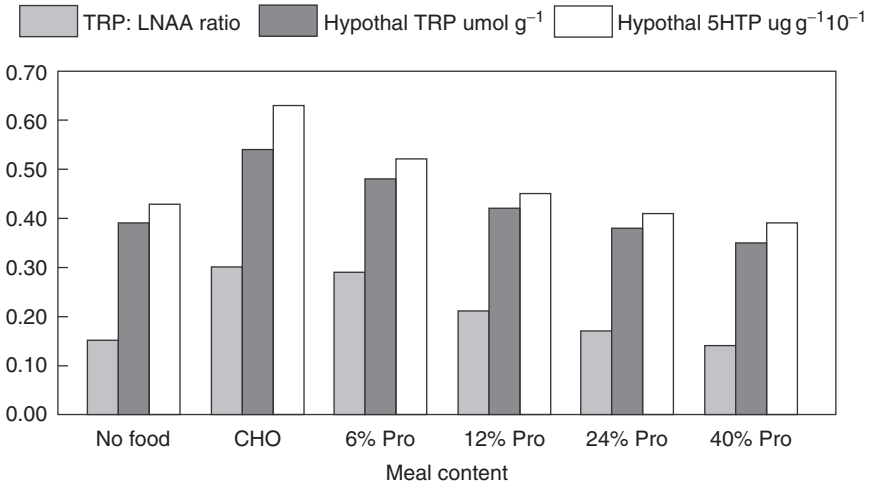


Figure 2 The data show the effect in rats of no meal, a carbohydrate meal with no protein, and one with increasing amounts of protein, on (a) the plasma ratio of tryptophan to the large neutral amino acids (TRP:LNAA) with which TRP competes for entry across the blood brain barrier (crosshatched bars) (b) levels of TRP (μmol g^{-1}) measured in the hypothalamus of rat brain (hatched bars) (c) the levels of 5-hydroxytryptophan (μg g^{-1} 10^{-1}), an intermediate precursor of serotonin synthesis, in the hypothalamus. The rise in TRP entering the brain after a carbohydrate meal drives increased serotonin synthesis, but this effect is progressively inhibited by increasing protein content. The figure is based on data obtained from Fernstrom MH and Fernstrom JD (1995) Brain tryptophan concentrations and serotonin synthesis remain responsive to food consumption after the ingestion of sequential meals. *American Journal of Clinical Nutrition* 61: 312–319, with permission from American Society for Nutrition.

Postprandial declines in arousal can be quite noticeable 2.5–3 h after high-fat meals, but fat in midmorning meals seems more sedating than at lunchtime, which might relate to expectations. By comparison, when lipid was infused directly into the duodenum, a decline in alertness was apparent much sooner, by 30–90 min after the meal. These effects of fat may result from increased release of the gastric regulatory hormone, cholecystokinin. However, in a study comparing ingestion of pure fat, carbohydrate, and protein (1.67 MJ, 400 kcal, at breakfast), measures of memory, attention, and reaction time deteriorated more after carbohydrate and protein than after fat. This beneficial effect of fat was attributed to the demonstrated relative absence of glycemic and hormonal (insulin, glucagon, and cortisol) perturbations in the 3 h following fat ingestion.

Carbohydrates, Stress, Mood, and Mental Function

Susceptibility to Mood Enhancement by Diet

The possibility that a carbohydrate-rich low-protein meal could raise 5-HT function gave rise to the proposal that some depressed people may self-medicate by eating carbohydrate, so leading to increased 5-HT release in a manner reminiscent of effects of antidepressant drugs, which enhance aspects of 5-HT function by inhibiting removal of 5-HT from the synaptic cleft between nerve cells. For the most part, however, early behavioral and pharmacological evidence for such a phenomenon was not very convincing.

Nevertheless, recent research provides some further support for beneficial effects of carbohydrate-rich/protein-poor meals on mood and emotion in some people. When participants were divided into high or low stress-prone groups, as defined by a questionnaire, carbohydrate-rich/protein-poor meals before a stressful task were found to block task-induced depressive feelings and release of the glucocorticoid stress hormone, cortisol, but only in the high stress-prone group. This finding was replicated using high- versus low-TRP containing proteins (alpha-lactalbumin and casein, respectively). It was argued that, because stress increases 5-HT activity, the poor stress-coping of this sensitive group might indicate a deficit in 5-HT synthesis that is improved by this dietary intervention.

There is another link between macronutrient intake, stress, and mood. Chronic dysfunction of the stress-sensitive hormone, cortisol, and its controlling hypothalamic pituitary adrenal (HPA) axis, is associated with depression and anxiety, as well as abdominal obesity. Moreover, protein-rich meals that prevent a meal-induced fall in arousal also stimulate release of cortisol in unstressed people, and the size of this effect is correlated positively with poor psychological wellbeing. Chronically, a carbohydrate-rich diet is associated with better overall mood state and lower average plasma cortisol than a high-protein diet. Acutely, a carbohydrate preload, but not protein or fat load, enhances cortisol release during stress. This may be related to findings from both human and animal research suggesting that eating carbohydrate-rich and perhaps high-fat foods can help restore normal HPA axis function and glucocorticoid stress responses. Raised levels of cortisol in stressed people contribute to insulin resistance, which in turn promotes abdominal obesity. However, insulin resistance may increase the likelihood that high-carbohydrate/low-protein foods would raise brain TRP and 5-HT levels, because of increased levels of plasma fatty acids, which result in more unbound TRP in plasma. Conversely, it has also been found that high baseline cortisol predicts induction of depression by dietary depletion of TRP. This might underlie recent findings that insulin resistant people are less prone to suicide and depression, both of which are believed to be increased by low 5-HT function. Similarly, patients with Seasonal Affective Disorder show increased insulin resistance in the winter, together with a greater predilection for sugar-rich foods. Unfortunately, despite this protective effect, insulin resistance is a substantial risk to health by promotion of cardiovascular disease.

Sugars and Opioids

Endogenous opioids are released during stress, and are known to be important for adaptive effects such as resistance to pain. They are also involved in motivational and reward processes in eating behavior, such as stimulation of appetite by palatable foods. Perhaps the best evidence for opioid involvement in an interaction between stress and eating is the finding that, in animals and human infants, the ingestion of sweet and fatty foods, including milk, alleviates crying and other behavioral signs of stress. Recently, this effect was shown to depend on sweet taste rather than calories, as non-nutritive sweeteners also reduce crying. This stress-reducing effect can be blocked by opioid antagonists. The conclusion that adults select sweet fatty foods for opioid-mediated relief of stress is tempting, but remains speculative. Also, such behavior would need to be explained in the context of stress itself enhancing endogenous opioid release.

Glucose, Mood, and Mental Function

The possibility that ingesting glucose could alter mood and improve mental function has generated considerable research interest. However, there is only space here to summarize and interpret the key findings and controversies. The interest in glucose arises from two observations: (1) that the primary source of energy for brain function is glucose, and (2) that mental function and mood deteriorate when blood glucose concentration falls below basal physiological levels, i.e., hypoglycemia (<3.6 mmol l^{-1}).

The first observation must be qualified by a recent evidence that (1) in times of metabolic demand, the brain can also use lactate very effectively as an energy source, and (2) the brain contains significant stores of glycogen in specialized cells called astrocytes, which can be metabolized for energy by neighboring neurones. Nevertheless, in rats, extracellular glucose levels in a specific region

of the brain critical for memory, the hippocampus, decline to a greater extent during more demanding memory tasks, and this decline is prevented by a systemic glucose load.

As for hypoglycemia, this is rarely induced by normal food, although large amounts of sugar-rich drinks on an empty stomach might do so in some people. Yet, many studies of effects of glucose use a method similar to the Oral Glucose Tolerance Test (OGTT), in which fasted patients drink aqueous solutions containing 50–75 g of some form of glucose. This is not meant to be a normal nutritional manipulation, but a test of glucoregulation. Associations have been reported between rapid and substantial declines in blood glucose after OGTTs and aggressive thoughts and behavior: however, this might be mediated by greater counter-regulatory hormone release.

In studies comparing sugar-rich drinks with zero-energy sweet placebos, many find no effect on mood, but some report a rise in subjective energy within an hour, followed by increased calmness. In children, controlled studies failed to support the popular myth that sugar is excitatory: again, it had either no effect or was calming. However, it is worth noting that some adults, and probably children, are especially sensitive to rapid drops in blood glucose, showing counter-regulatory hormone release and 'hypoglycemic symptoms' even though actual hypoglycemic levels of glucose are not reached.

It may be that beneficial effects of glucose ingestion only become consistently apparent when demands are placed on mental function or when there is a compromised nutritional state such as food deprivation or a metabolic disorder. The findings on glucose and cognitive performance can be summarized as follows:

- The majority of studies that administered a glucose drink found subsequent improvements in some performance compared to placebo, particularly on tests of short-term memory or vigilance tasks that require a large component of 'working memory'.
- Improvements in performance can be associated with rising or falling blood glucose, even independently of consuming a glucose load.
- Young healthy subjects require more demanding tasks than the elderly to detect a beneficial effect of glucose load.
- Associations between performance and glucose may be mediated by individual glucoregulatory efficiency.
- Both glucoregulation and performance are influenced by hormones sensitive to stressful or arousing cognitive tasks, such as adrenaline and cortisol. Emotion-dependent learning resists improvement by glucose.
- Personality, stress-sensitivity, and task involvement can influence glucose uptake and disposal, and also the effects of glucose on cognition.

However, one important pattern does emerge that memory performance is worse in poorer than better glucoregulators (see **Figure 3**). This is true not just for elderly patients but among a healthy student population too, especially if the task is sufficiently demanding. Peak blood glucose predicts poor memory performance in elderly patients, whether or not a glucose load has been given before testing. This relationship between raised glucose levels and poor memory performance could underlie a recent finding that a snack with a high glycemic index (greater plasma glucose rise) resulted in poorer memory performance 2–3 h later, relative to a low glycemic index snack. Even so, it seems that a moderate glucose load can lessen the memory deficit present in young and old poor glucoregulators (with little consistency or no effect in good glucoregulators).

One reason why poor glucoregulation predicts poor memory performance may be that glucose intolerance is associated with higher basal and stress-induced cortisol secretion. Raised cortisol is known to impair memory, probably by an action on hippocampal neurons, including inhibition of glucose uptake. However, the substantial rise in insulin induced by a glucose load in poor glucoregulators may overcome the negative impact of cortisol in some cases: hyperinsulinemia induced independently of hyperglycemia has been shown to ameliorate memory deficits in patients with Alzheimer-type dementia.

Two other mechanisms might explain the ability of a glucose load to improve performance in subsequent challenging tasks. One is an increase in sympathetic activation by the glucose load: adrenaline is known to enhance memory. The other is increased synthesis and release of the neurotransmitter acetylcholine during challenging tasks: acetylcholine is also known to be critically involved in learning and memory, and is synthesized from dietary choline and acetyl CoA, which is a by-product of glucose metabolism. These complex interactions between glucose ingestion and brain function are illustrated in **Figure 4**.

Docosahexaenoic acid (DHA), a long-chain polyunsaturated fatty acid (PUFA) of the omega-3 series may also be an important regulator of brain energy metabolism and glucose uptake. For example, feeding rats an omega-3 PUFA deficient diet decreased brain DHA levels and reduced brain glucose uptake by 30–40% and cytochrome-c oxidase activity by 20–40% in the fronto–parietal cortex, hippocampus, and suprachiasmatic nucleus, compared to animals on the control diet. The level of expression of the GLUT-1 glucose transporters were also decreased in endothelial cells, although GLUT-3 levels were unaffected in neurons. Impairment in brain (especially hippocampal) glucose uptake is a common feature of age-related cognitive decline, and treatment with DHA may provide a potential therapeutic approach to support the maintenance of brain energy metabolism.

Hyperactivity and Antisocial Behavior

In children, there is an increasing frequency of diagnosis of ADHD, a condition characterized by inattention, impulsive, and disruptive behavior, learning difficulties, and increased levels of gross motor activity and fidgeting. Also, prevalence of food allergies and intolerance has been increasing. Perhaps it is not surprising that dietary explanations and treatments for ADHD have been sought regularly for several decades, given theories of allergic reactions or intolerance to food additives, ingredients in chocolate, and even refined sugar (often grouped as the 'Feingold Theory', after an early instigator of unproven dietary intervention). There has also been a longstanding interest in the possibility that antisocial behavior in children and adults might in part result from poor nutrition,

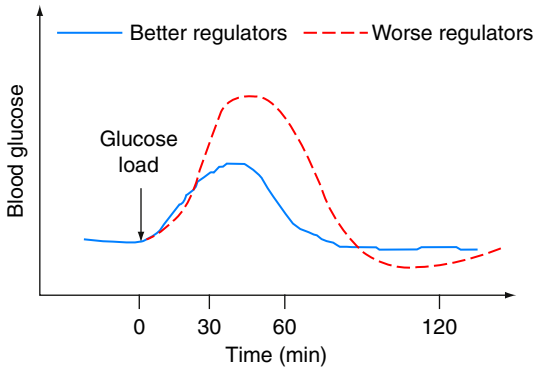


Figure 3 The graph represents a model of changes in blood glucose levels after a glucose load in people who are either good regulators (solid line) or poor regulators (broken line) of blood glucose. The increased peak and delayed recovery of blood glucose in worse regulators suggests glucose intolerance and insulin resistance. In such people, there may be a brief period of mild hypoglycemia before return to baseline levels. This type of difference in glucoregulation has been demonstrated in both young and elderly people without diabetes. Poor glucoregulation predicts poor cognitive performance in challenging tests, especially those involving memory.

although early studies were poorly designed. A behavioral effect of sugar and of many additives has by and large not been supported by controlled studies: however, determining unequivocally whether behavior of young children is affected by specific dietary components is difficult to achieve. ADHD may be associated with disrupted eating behavior and poor nutrition, so that removal of a number of nutrient deficiencies might improve behavior. In addition, parents or unqualified health professionals may devise unsuitable dietary regimes that can raise the risk of undernutrition. As a result, there is little consensus as to what in the diet may or may not provoke disturbed behavior in children, other than that only a small minority of children are likely to be affected. Nevertheless, a recent British study, in which children were given a collection of food colorings and preservatives, or placebo in drinks, found deterioration in behavior reported by parents for both hyperactive and normal children given the additives, which seemed unrelated to allergic history. This effect was not detectable in a clinical setting. Clearly, a definitive answer awaits more research.

Specific deficiencies or imbalances in blood levels of essential fatty acids, particularly omega-3 polyunsaturated fatty acids (PUFAs), have consistently been shown in developmental disorders such as ADHD, dyslexia, dyspraxia, and autistic spectrum disorders. Indeed the severity of ADHD correlates with the level of DHA deficiency in plasma, although the nature of this relationship is

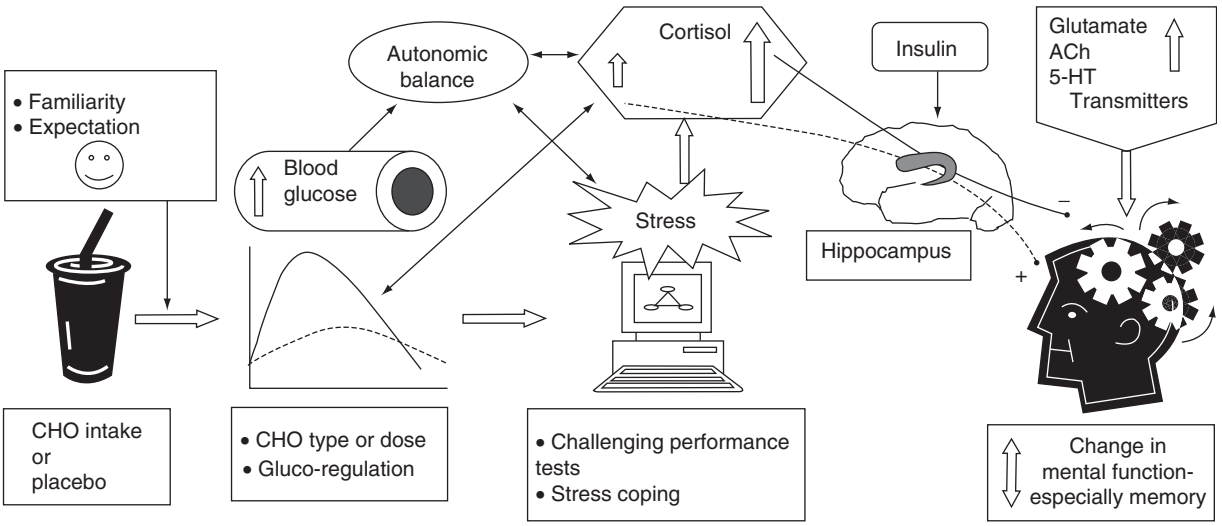


Figure 4 A summary diagram of putative pathways linking carbohydrates and mental function, with various modulatory influences. One explanation for variability in findings is that some individuals, especially if poor glucoregulators, may be much more susceptible to manipulations of mood and mental function by glucose availability, particularly when brain function is increased by a stressful challenge. However, associations between glucoregulatory ability and mental performance may be more directly mediated by differential release of cortisol than by changes in glucose availability. It is suggested that this neuroendocrine consequence of glucose absorption and stressful tasks might help explain the narrow dose–response function, and why slower glucose delivery, for example, by low-GI foods, has been reported to benefit cognition compared to faster glucose delivery. Identifying sensitive individuals, and the relevant nutritional, physiological, and psychological parameters, remains a potentially fruitful topic of research in this area. Reproduced with permission from Gibson EL (2007) Carbohydrates and mental function: Feeding or impeding the brain? *Nutrition Bulletin* (Supplement 32), 71–83.

unclear. There is a strong rationale for investigating the effects of essential fatty acid supplementation in children with ADHD; however, so far few randomized controlled trials have been undertaken, and results have been inconclusive. For example, two double-blind trials showed no benefits following DHA treatment, whereas another, which provided a mixed omega-3 and omega-6 PUFA supplement found significant improvements in symptoms. Given the safety and tolerability of essential fatty acids, these compounds offer an intriguing prospect as a potentially new therapeutic approach and the evidence so far strongly supports the case for further research.

Other effects of essential fatty acids on cognition are discussed below.

Micronutrients and Mental Function

There has been, over the years, an increasing body of evidence suggesting that vitamin and mineral status is significantly related to both brain development in childhood and the degree of cognitive decline experienced as we age. Indeed, it is certainly the case that deficiencies of some vitamins are associated with negative neurological symptoms such as neural tube defects. The work examining vitamin and mineral supplementation comprises both cohort studies and nutritional interventions and has generated much confusing and contradictory data. To a large degree, this confusion and contradiction is dependent on a number of factors such as the methodological rigor of each study, the measures of cognitive function used and the precise nutrient being studied.

Early work in this area concentrated on the notion that supplementing the diet of schoolchildren with multi-vitamin supplements would improve both their IQ scores and academic achievement. This work was controversial and marked by a number of deficiencies, such as any clear indication as to whether participants were actually nutritionally compromised before treatment, difficulties in determining, which if any, of the vitamins in the cocktail were producing effects, and the lack of any clear hypotheses regarding mechanisms responsible. The consensus now is that supplementation will have a benefit on cognitive development and IQ (especially non-verbal) in a minority of children who are not otherwise adequately nourished. In Britain and the USA at least, there is particular concern that a significant proportion of adolescents, especially girls, are deficient in iron. There is good evidence that iron deficiency contributes to poor cognitive ability, perhaps in association with low vitamin C status, which has also been linked to reduced cognitive function. Iron is known to be essential for synthesis and function of neurotransmitters, such as dopamine, noradrenaline, and serotonin (5-HT). Selenium is another mineral, which may be important for brain function, and low levels of which have been associated with cognitive decline and depressed mood in the European population. There is also evidence that zinc deficiency is associated with problematic cognitive development in children and that dietary supplementation with zinc leads to cognitive improvements, relative to nonsupplemented controls.

Much recent work, however, has concentrated on the use of vitamins in the treatment of age-related cognitive decline and dementia and, to varying degrees, is more scientifically rigorous than the earlier work. The overwhelming majority of the experimental work has targeted the action of two groups of micronutrients; anti-oxidant and B-complex vitamins. The work concerning the effects of antioxidant vitamins, although showing some promise with correlational studies in that levels of these vitamins (vitamin E most consistently) are associated with function in a range of cognitive domains, is more contradictory when one considers the clinical intervention trials. The work on B- complex vitamins is, however, more consistent and supported by a strong hypothetical basis. This relies on the role of vitamins B_{12} and folate in methylation of membrane phospholipids and neurotransmitters, and in breaking down the toxic sulfur-amino acid homocysteine. High levels of homocysteine are now considered by some to be a far greater risk factor for the development of coronary and vascular problems than high levels of cholesterol. Elevated levels of homocysteine may be a cause of minor ischemic events, which cumulatively, lead to a degradation of cognitive function due either to sub-clinical deficiencies of, or problems with the absorption of, B-complex vitamins. Indeed, a large number of studies has consistently demonstrated relationships between homocysteine levels, B-complex vitamin levels and neuropsychological task performance. The number of direct intervention trials, which have supplemented the diet of the elderly with B-complex vitamins is, however, small. Although some studies have shown no net benefit of supplementation on the cognitive function of the elderly, a larger number of studies have shown a stabilization of cognitive function and reduction of homocysteine levels to result from B-complex vitamin supplementation. As with the antioxidant vitamins, however, these studies must be interpreted with a degree of caution because they use differing dosages, periods of supplementation, and measures of neuropsychological function.

Lipids

Lipids are another nutrient category, which has attracted a good deal of research interest in terms of their possible effects on psychological function. The main nutrients studied fall into three groups; these being cholesterol, omega-3 and omega-6 essential fatty acids, and phospholipids. In general, the theoretical basis underpinning the effects (or lack of) these nutrients on psychological function relates to how their relative concentrations affect cell membrane fluidity. The rigidity of lipid bilayers of cell membranes is thought to be essential for neurotransmitter function by maintaining maximum exposure of receptors at the synaptic cleft between neurons.

Cholesterol

The interest in cholesterol as a substance that is related to psychological well-being stems back to the 1980s. During this period, a number of epidemiological studies found that individuals with low cholesterol levels were more prone to aggressive behavior,

risk of suicide, and violent death. In addition, it was also found that nonhuman primates increased their incidence of aggressive behavior when kept on a low cholesterol diet. In terms of neuropsychological function, a number of studies have found associations between cholesterol levels and choice reaction time or memory function. Two of these studies to date have sought actively to reduce cholesterol levels by means of pharmacological or dietary means, both finding that lowering cholesterol produced small but statistically significant impairments in memory and attention. Conversely, however, a number of studies have also demonstrated that high cholesterol levels are also a significant risk factor for the development of Alzheimer Dementia. One mechanism for these negative effects of cholesterol lowering may be loss of rigidity in neural cell walls, thereby decreasing the relative exposure of serotonin (5-HT) membrane receptors at the synaptic cleft and impeding 5-HT signal transmission. Interestingly, omega-3 PUFA have been shown to reduce the cholesterol content of neuronal membranes.

Essential Fatty Acids

The central nervous system is highly enriched in long-chain polyunsaturated fatty acid (PUFA) of the omega-6 and omega-3 series, specifically arachidonic acid and docosahexaenoic acid (DHA), respectively. The presence of these fatty acids as structural components of neuronal membranes influences cellular function both directly, through effects on membrane properties, and also by acting as a precursor pool for lipid-derived messengers. For example, DHA alters membrane properties and thereby influences membrane proteins, such as receptors, ion channels, and enzymes. It alters dopaminergic, serotoninergic, and cholinergic neurotransmission, regulates signal transduction pathways, production of eicosanoids and other lipid-derived messengers, regulates transcription factors and gene expression, and stimulates hippocampal neurogenesis and neurite outgrowth. A summary of some of the potential mechanisms of action of omega-3 PUFA is shown in **Figure 5**.

An adequate intake of omega-3 PUFA is essential for optimal visual function and neural development. Fetal development has a high requirement for essential fatty acids, especially DHA. Maternal DHA levels influence infant levels, such that higher maternal levels produce higher levels in the neonate. Higher level of maternal plasma DHA during pregnancy correlates with more mature neonatal sleep-state patterns, which is a measure of maturity in the central nervous system. Such effects probably underlie evidence that higher intake of seafood during pregnancy is associated with improved social and cognitive function in the offspring, and may contribute to the advantage to cognitive development seen for breast versus formula-fed infants.

There is increasing evidence that an increased intake of the long-chain omega-3 PUFA, eicosapentaenoic acid (EPA), and DHA, may confer benefits in a variety of psychiatric and neurological disorders, and in particular neurodegenerative conditions. For example, patients suffering from depression typically exhibit reduced DHA levels in plasma, and several cross-sectional studies show associations between seafood consumption and rates of both depression and bipolar disorder, suggesting that low omega-3 PUFA levels may be a factor in the etiological of these disorders. Furthermore, a recent meta-analysis of trials involving patients with major depressive disorder and bipolar disorder has provided evidence that omega-3 PUFA supplementation reduces symptoms of depression.

It is beyond the scope of this article to discuss all of the evidence however, a number of lines of research suggest that an elevated intake of EPA and DHA may confer benefits in depression, schizophrenia, Alzheimer's disease, and Parkinson's disease. However, the mechanisms underlying these beneficial effects are still poorly understood. It is very likely that EPA and DHA operate through a number of different, potentially overlapping mechanisms, involving various cellular targets. These PUFA can act as endogenous agonists for transcription factors such as the retinoid receptors, and the peroxisome-proliferator activated receptors. Neuronal excitability can be modulated by PUFA through their effects on sodium and calcium channels or the activation of the TREK potassium channels, which are two-pore potassium channels abundantly expressed in the brain. Furthermore, it cannot be ruled out that their effects may also involve specific fatty acid receptors, such as the recently identified GPR40 receptor, which has a widespread expression in the central nervous system. Because many neurological conditions share common features, such as excitotoxicity, oxidative stress, and inflammation, which appear to be modified by omega-3 PUFA, this may explain the therapeutic potential of EPA and DHA across disease boundaries. However, the relative importance of certain mechanisms and targets may vary depending on the condition considered.

Membrane-bound DHA is also a positive modulator of biosynthesis of the membrane phospholipid phosphatidylserine in neuronal tissues. It has been suggested that the observed anti-apoptotic (i.e., cell protective) effect of DHA is due at least in part to the DHA-induced phosphatidylserine accumulation. A further potential benefit of DHA-induced phosphatidylserine accumulation may be to prevent the age-related decline in cognitive ability, because phosphatidylserine supplementation has been shown to improve age-related decline in cognitive ability and memory. Certainly, epidemiological evidence links diets higher in PUFA to saturated fat ratio to improved cognitive function in the elderly.

Food Deprivation

There is evidence to suggest psychological effects of undernutrition. In severe cases such as anorexia nervosa, neuropsychological function is impaired primarily as a result of structural changes in brain anatomy resulting from starvation. Evidence that undernutrition is associated with psychological problems in those not suffering eating disorders was first hinted in the Minnesota Study of Semi-Starvation in the 1950s. Volunteers who were kept on a half-calorie intake diet for a period of months reported mood swings, increased irritability, poorer memory, and an inability to concentrate. Although these self-reported effects were not supported by objective testing, the lack of a nondeprived control group means that this lack of effect could have been masked by a practice effect.

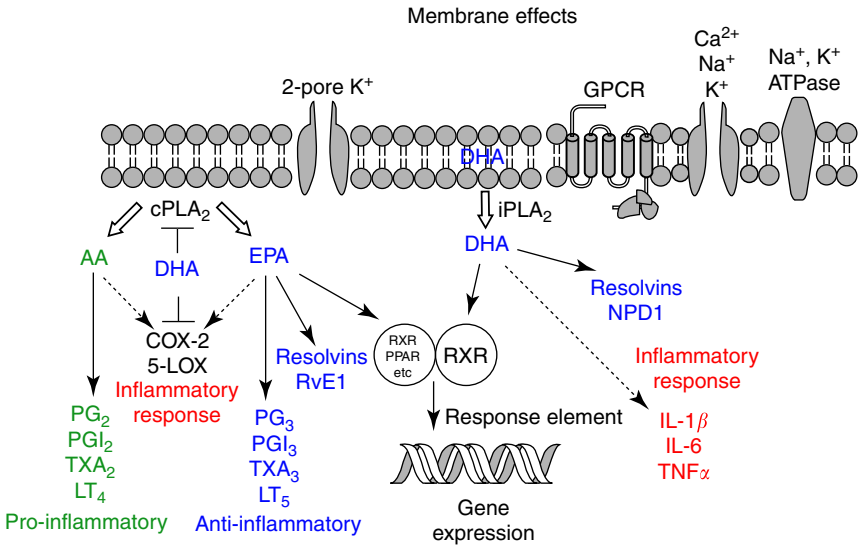


Figure 5 EPA and DHA potentially operate through a variety of overlapping mechanisms of action. These are related to direct actions on plasma membranes, altered inflammatory response and control of gene expression. The effects on membrane bound proteins such as ion channels, G-protein coupled receptors (GPCR), and the Na^+, K^+ ATPase appear to relate to alterations to the biophysical properties of the cell membrane. Alterations in inflammatory response are mediated through competition between AA and EPA for eicosanoid biosynthetic enzymes, with a high EPA content favouring the production of EPA derived anti-inflammatory mediators, such as series 3 prostaglandins, prostacyclins and thromboxanes, and series 5 leukotrienes. Nonesterified EPA and DHA are also the precursors of anti-inflammatory resolvins, such as RvE1 and NPD1, respectively. Nonesterified EPA and DHA may also regulate gene expression *via* transcription factors, such as retinoid and peroxisomal proliferator signaling pathways. Solid arrows indicate positive effects, flat arrow-heads inhibition, dotted arrows competition, and open arrows phospholipase A_2-induced release from the cell membrane. Abbreviations: 2-pore K^+, 2-pore potassium channel; Ca^{2+}, L-type calcium channel; $cPLA_2$, cytosolic PLA_2; $iPLA_2$, Ca^{2+}-independent PLA_2; GPCR, G-coupled protein receptor; K^+, K_v, and K_{irr} channels; LT, leukotriene; Na^+, voltage-gated sodium channel; PG, prostaglandin; PGI, prostacyclin; PPAR, peroxisomal proliferator-activated receptor; RXR, retinoid X receptor and TXA, thromboxane.

Dieting to lose weight is one of the most common food choice related behaviors in the Developed World and it has been consistently associated with negative psychological consequences such as preoccupation with body shape and depression. In addition, a number of investigators have also found that dieting to lose weight is associated with impairments in cognitive function, with dieters performing more poorly than nondieters on measures of reaction time, immediate memory, and the ability to sustain attention. This is unlikely to be due to pre-existing differences between individuals who happen to be dieting or not dieting at the time of testing because, within the same individuals, performance is poorer when dieting than when not dieting. It is unlikely that these effects are due to the gross physical effects of food deprivation because experimentally induced food deprivation of varying lengths fails to produce a comparable impairment in task performance, in addition to the poorer task performance being found amongst dieters who claim not to have actually lost any weight over the course of the diet.

Rather than being a function of food deprivation per se, the poorer task performance amongst current dieters appears to be a function of the preoccupying concerns with hunger and body shape, which are characteristic of dieters. Indeed, the impairments in task performance amongst dieters appear to be comparable in both structure and magnitude found to result from the preoccupying concerns of the characteristic of clinical depression and anxiety disorders. Specifically, the primary deficit appears to be a reduction in the amount of available working memory capacity, working memory being the primary cognitive system which serves to allocate processing capacity to ongoing cognitive operations. A threshold hypothesis has been formulated to account for this phenomenon. Nondieting, highly restrained eaters are characterized by an enduring, trait concern with body shape, which consumes a certain amount of working memory capacity (explaining why nondieting restrained eaters perform at a level intermediate to that of current dieters and unrestrained eaters). When they decide to diet, they then experience preoccupations with food and an increased desire to eat, this extra drain on working memory capacity reaches a point where sufficient capacity is unavailable to maintain task performance. Support for this hypothesis can be seen in a study in which highly restrained nondieters were instructed to imagine eating their favorite food or their favorite holiday while performing a reaction time task. When imagining their favorite food, but not their favorite holiday, restrained nondieters performed as poorly as current dieters on the reaction time task.

Although evidence seems to be mounting that the poor cognitive function of current dieters is due to psychological and not biological factors, work still continues to examine some of the more subtle possible biological mechanisms which, may underlie the effects. One possible mechanism is that a low dietary intake of the amino acid TRP (the precursor for 5-HT) leads dieters to have impaired serotonergic function. However, analysis of the urine of dieters for the 5-HT metabolite, 5-hydroxyindoleacetic acid (5-HIAA), found no evidence for this. Another possibility (not yet investigated) is that, by avoiding eating red meat, dieters

experience mild iron deficiency, with deleterious consequences for hemoglobin status, brain oxygen supply, and neurotransmitter function.

The types of dieter studied so far are those who attempt to lose weight in an unsupported, unsupervised manner. Comparisons between this type of dieter and those who attempt to lose weight in the context of an organized weight loss group reveal dramatic differences. Those who diet as part of a group do not show the impairments in task performance typical of unsupported dieters. In addition, unsupervised dieters display an elevated stress response after one week of attempted weight loss (as measured by salivary cortisol levels) whereas supported dieters do not. It would appear, therefore, that the poor performance characteristic of unsupervised dieting is a result of the stress associated with this type of weight loss attempt and that the psychological manifestation of this stress is the preoccupying thoughts outlined above. The motivational processes underlying weight loss also appear to be mediating factors in any cognitive processing deficits observed, because it has been found that individuals dieting for esthetic reasons display impaired cognitive function, whereas those dieting for other, health related, reasons do not.

Functional and Pharmacological Components of Foods and Drinks

There is growing interest, particularly in the food and beverage industry, in developing foods and drinks with functional properties (nutraceuticals) attractive to the consumer. These include effects on behavior, such as improvements in cognitive function, mood, and physical performance. Components of interest include caffeine, herbal extracts such as ginkgo biloba and panax ginseng, micronutrients, essential fatty acids, amino acids, and carbohydrates. In the context of caffeine and glucose, however, there is some evidence to support the view that any beneficial effects of these substances on psychological function is, at least in part, due to an expectancy regarding their potential effects. There is some support for beneficial effects of these components, but they are not reviewed further here.

Conclusion

The scientific understanding of dietary effects on behavior has begun to move in from the fringes of respectability, indeed sufficiently to attract substantial commercial interest. Advances in nutritional and neuropsychological knowledge, experimental design, and sensitivity of measures of behavior and brain function have produced replicable findings in some areas to mollify earlier skepticism. New understanding of the impact of nutrition on brain function, and predictors of individual susceptibility, has also allowed reinterpretation of old data. Promising areas with encouraging developments in understanding include the interactions between macronutrients, stress, and mood disorders, and the effects of vitamins, minerals, and lipids on cognition, dementia, and psychiatric disorders. Some findings, including recent awareness of poorly nourished sectors of the population, suggest useful interventions. Nevertheless, research in this field is at an early stage, and the coming years should bring further revelations on the link between diet and behavior. With industrial backing, few may escape the consequences.

Further Reading

Awad, N., Gagnon, M., Desrochers, A., Tsiakas, M., Messier, C., 2002. Impact of peripheral glucoregulation on memory. Behavioral Neuroscience 116, 691–702.

Benton, D., 2002. Carbohydrate ingestion, blood glucose, and mood. Neuroscience and Biobehavioral Reviews 26, 293–308.

Berr, C., 2002. Oxidative stress and cognitive impairment in the elderly. Journal of Nutrition Health and Aging 6, 261–266.

Bruinsma, K.A., Taren, D.L., 2000. Dieting, essential fatty acid intake, and depression. Nutrition Reviews 58, 98–108.

Dyall, S.C., Michael-Titus, A.T., 2008. Neurological benefits of omega-3 fatty acids. Neuromolecular Medicine 4, 219–235.

Eskelinen, M.H., Ngandu, T., Helkala, E.L., et al., 2008. Fat intake at midlife and cognitive impairment later in life: A population-based CAIDE study. International Journal of Geriatric Psychiatry 23, 741–747.

Falkingham, M., Abdelhamid, A., Curtis, P., et al., 2010. The effects of oral iron supplementation on cognition in older children and adults: A systematic review and meta-analysis. Nutrition Journal 9, 4–20.

Fernstrom, J.D., 1994. Dietary amino acids and brain function. Journal of the American Dietetic Association 94, 71–77.

Freemantle, E., Vandal, M., Tremblay-Mercier, J., et al., 2006. Omega-3 fatty acids, energy substrates, and brain function during aging. Prostaglandins Leukotrienes, and Essential Fatty Acids 75 (3), 213–220.

Gibson, E.L., 2007. Carbohydrates and mental function: Feeding or impeding the brain? Nutrition Bulletin (Supplement 32), 71–83.

Gibson, E.L., Green, M.W., 2002. Nutritional influences on cognitive function: Mechanisms of susceptibility. Nutrition Research Reviews 15, 169–206.

Green, M.W., 2001. Dietary restraint and craving. In: Hetherington, M.M. (Ed.), Food Cravings and Addiction. Leatherhead Food RA, Leatherhead.

Hoyland, A., Dye, L., Lawton, C.L., 2009. A systematic review of the effect of breakfast on the cognitive performance of children and adolescents. Nutrition Research Reviews 22, 220–243.

Hoyland, A., Lawton, C.L., Dye, L., 2008. Acute effects of macronutrient manipulations on cognitive test performance in healthy young adults: A systematic research review. Neuroscience and Biobehavioral Reviews 32, 72–85.

Innis, S.M., 2007. Dietary (*n*-3) fatty acids and brain development. Journal of Nutrition 137, 855–859.

Lamport, D.J., Lawton, C.L., Mansfield, M.W., Dye, L., 2009. Impairments in glucose tolerance can have a negative impact on cognitive function: A systematic research review. Neuroscience and Biobehavioral Reviews 33, 394–413.

Markus, C.R., Olivier, B., Panhuysen, G.E.M., et al., 2000. The bovine protein alpha-lactalbumin increases the plasma ratio of tryptophan to the other large neutral amino acids, and in vulnerable subjects raises brain serotonin activity, reduces cortisol concentration, and improves mood under stress. American Journal of Clinical Nutrition 71, 1536–1544.

Richardson, A.J., 2004. Clinical trials of fatty acid treatment in ADHD, dyslexia, dyspraxia and the autistic spectrum. Prostaglandins Leukotrienes and Essential Fatty Acids 70 (4), 383–390.

Riggs, K., Spiro, A., Tucker, K, Rush, D., 1996. Relations of vitamin B_{12}, vitamin B_6, folate, and homocysteine to cognitive performance in the Normative Aging Study. American Journal of Clinical Nutrition 63, 306–314.

Reynolds, E.H., 2002. Folic acid, ageing, depression, and dementia. British Medical Journal 324, 1512–1515.

Rogers, P.J., 2001. A healthy body, a healthy mind: Long-term impact of diet on mood and cognitive function. Proceedings of the Nutrition Society 60, 135–143.

Sontrop, J, Campbell, M.K., 2006. Omega-3 polyunsaturated fatty acids and depression: A review of the evidence and a methodological critique. Preventative Medicine 42 (1), 4–13.

Brown adipose tissue: Implications for human health

Martijn Boost[a], **Borja Martinez-Tellez**[a], **Angel Gil**[b,c,d], **and Jonatan R. Ruiz**[c,e], [a] Department of Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, the Netherlands; [b] Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology "José Mataix", Center of Biomedical Research, University of Granada, Granada, Spain; [c] Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain; [d] CIBEROBN (Physiopathology of Obesity and Nutrition Network CB12/03/30038), Institute of Health Carlos III (ISCIII), Madrid, Spain; and [e] Department of Physical Education and Sports, Faculty of Sport Sciences, University of Granada, Granada, Spain

© 2023 Published by Elsevier Ltd. All rights reserved.

Introduction 26
Obesity is a global problem 26
Brown adipose tissue as a target to combat obesity 26
The discovery of brown adipose tissue 27
Brown adipose tissue discovered in rodents 27
Early investigations of human brown adipose tissue 27
The re-discovery of active brown adipose tissue in live humans 28
The physiology of brown adipose tissue 28
Adaptive thermogenesis by brown adipose tissue 28
Adaptive thermogenesis in response to cold exposure 28
Mitochondrial metabolism is uncoupled in brown adipocytes 29
Brite adipocytes can be recruited in response to cold exposure 29
Brown adipose tissue activity in humans 29
Retrospective analysis of brown adipose tissue activity 29
Brown adipose tissue detection by cold exposed ^{18}F-FDG PET/CT scan 30
Dysfunction of brown adipose tissue in obese individuals 30
Brown adipose tissue glucose uptake may be independent of thermogenic capacity 30
Alternative PET tracers to quantify brown adipose tissue activity in humans 30
The metabolic significance of brown adipose tissue 30
Brown adipose tissue is metabolically significant 31
Contribution of human brown adipose tissue to cold-induced thermogenesis 31
Brown adipose tissue as an endocrine organ 31
Potential novel batokines have been recently identified 32
A subpopulation of acetate-secreting brown adipocytes may regulate thermogenesis 33
Brown adipose tissue as a therapeutic target 33
Activation of brown adipose tissue to treat obesity 33
Activation of brown adipose tissue by β-adrenergic receptor agonists 33
Other pharmacological approaches to the activation of brown adipose tissue 33
Adipose tissue browning by peroxisome proliferator-activated receptor-γ agonists 34
Future perspectives 34
New methods for to assessment of brown adipocyte tissue mass and function 34
Batokines as a therapeutic target 34
Transplantation of brown adipocyte tissue 35
Conclusion 35
References 35

Key points

- To know the physiology of brown adipose tissue
- To know the metabolic significance of brown adipose tissue
- To know ways to measure brown adipose tissue activity in humans
- To understand mechanisms by which brown adipose tissue may exert metabolic effects
- To understand the potential of brown adipose tissue as a therapeutic target to treat obesity

Glossary

Batokines Specific molecules secreted by brown adipose tissue into the bloodstream which signal to metabolically active organs such as skeletal muscle, liver and brain

Brite Brown-in-white adipocytes, also known as beige adipocytes. Brite cells have a multilocular morphology, are enriched in mitochondria, and also express UCP1. They are peculiar in that they share characteristics with white and brown adipocytes

Brown adipose tissue (BAT) It is a lipid-containing tissue, which produces heat in response to cold exposure via a procedure known as thermogenesis

Capsaicinoids Capsaicinoids are alkaloids found in chili peppers that are thought to enhance energy expenditure when ingested

Glucagon-like peptide-1 receptor (GLP-1R) It is a receptor protein found on beta cells of the pancreas and neurons of the brain. It is involved in the control of blood sugar levels by enhancing insulin secretion. Stimulation of the GLP-1R receptor was found to enhance BAT thermogenesis in mice.

NIRS Near-infrared spatial resolved spectroscopy is a simple, non-expensive, and non-invasive method to measure tissue oxygenation in vivo. It measures different optical properties of the tissue based on oxygen-dependent absorption changes, and it allows the calculation of the tissue saturation index and the concentrations of total hemoglobin, oxyhemoglobin, and deoxyhemoglobin

Positron emission tomography (PET) It is a nuclear medicine procedure based on the measurement of positron emission from radiolabeled tracer molecules

UCP-1:Uncoupling protein 1 (UCP1) The UCP1 is a member of a superfamily of homologous proteins formed by the mitochondrial metabolite transporters. It is devoted to adaptive thermogenesis, a specialized function performed by brown adipocytes. Whereas the family of mitochondrial metabolite carriers comprises ~40 members, UCP1 is the only member able to translocate protons through the inner membrane of brown adipocyte mitochondria. By this process, UCP1 uncouples respiration from ATP synthesis and therefore provokes energy dissipation in the form of heat while, also stimulating high levels of fatty acid oxidation

^{18}F-fluorodeoxyglucose (^{18}F-FDG) It is a radiolabeled sugar (glucose) molecule. Chemically, it is 2-deoxy-2-[^{18}F]fluoro-D-glucose, a glucose analog, with the positron-emitting radionuclide fluorine-18 substituted for the normal hydroxyl group at the C-2 position in the glucose molecule. The uptake of ^{18}F-FDG by tissues is a marker for the tissue uptake of glucose, which in turn is closely correlated with certain types of tissue metabolism. Imaging with ^{18}F-FDG PET is used to determine sites of abnormal glucose

Introduction

Obesity is a global problem

The global prevalence of obesity has doubled over the past three decades, affecting over 600 million adults worldwide in 2015. Obesity is characterized by the accumulation of excess body fat and is defined as a body mass index (BMI) equal to or over 30 kg/m^2. It increases the risk of diseases such as type 2 diabetes mellitus, hypertension, stroke, dementia and several cancers, and contributes to a decline in quality of life and life expectancy. The fundamental cause of obesity is a chronic positive energy balance caused by overnutrition. The imbalance between calories consumed and calories spent then leads to the deposition of excess energy as triglycerides in white adipose tissue (WAT). The main three elements influencing the energy balance are diet, activity and metabolism.

Weight-loss interventions commonly target the first two elements, diet and activity, through dietary caloric restriction and exercise. However, these weight-loss interventions are frequently not successful in the long term. The development of drugs to treat obesity has also been mainly focused on weight loss, but many of the drugs have had limited success and some have been withdrawn due to harmful side effects. The third element influencing the energy balance, metabolism, is heavily influenced by heritable factors. Less effort has been spent in involving this element in weight-loss interventions. However, the relatively recent re-discovery of brown adipose tissue (BAT) in humans has ignited an interest in increasing metabolic energy expenditure to combat obesity (Nedergaard et al., 2007).

Brown adipose tissue as a target to combat obesity

BAT is a lipid-containing tissue, which produces heat in response to cold exposure via a procedure known as thermogenesis. Brown adipocytes are essential to thermogenesis due to their unique expression of uncoupling protein (UCP)-1, which uncouples the movement of protons across the inner mitochondrial membrane from the synthesis of adenosine triphosphate (ATP). As a result of the uncoupling, all energy produced by the mitochondria is released as heat. For this reason, thermogenesis represents one

mechanism through which energy can be expended without exercise, and the stimulation of BAT thermogenesis represents a possible approach to increase energy expenditure in humans.

Hence, strategies able to activate BAT are of great interest to the scientific community. As such, this article aimed is to provide a framework of knowledge to support the development of new strategies by reviewing the historical and current knowledge of BAT and its implications for human health. Therefore, we will first describe the discovery of active BAT in humans, followed by the current knowledge of the physiology of human BAT. Subsequently, the metabolic activity of BAT in humans will be considered, and possible alternative functions of BAT in humans beyond its metabolic activity will be highlighted. Several existing pharmacological approaches to the activation of BAT for the treatment of obesity will also be described. Finally, we will discuss future perspectives on the stimulation of BAT, which have the potential to lead to new strategies for the treatment of obesity.

The discovery of brown adipose tissue

Brown adipose tissue discovered in rodents

BAT was first discovered in small mammals as early as 1551 by the Swiss physician Konrad Gesner, who found BAT in hibernating marmots (Fig. 1). He distinguished brown adipocytes from white by their characteristic brown color, which we now know originates from the high density of iron-containing mitochondria in the cells. The second main histological characteristic of brown adipocytes is their multilocular appearance. This appearance is in contrast to white adipocytes, which contain only a single lipid droplet.

The role of the tissue remained unclear until the second half of the twentieth century. Thermogenic activity was found in rodent interscapular fat during *in vitro* experiments in 1961. The role of BAT in thermoregulatory heat production was directly demonstrated three years later in experiments with rodents by recording the temperature increase in the interscapular fat in response to cold stimulus.

Early investigations of human brown adipose tissue

The discovery of its function renewed interest in the tissue, and in 1966 the functioning of BAT in humans was investigated by studying tissue samples collected during necropsies. BAT was found to be widely distributed through the body in infants. It was found in adipose deposits including the interscapular area, under the clavicles, around the neck, near the pancreas and kidneys and in para-aortic deposits. In tissue samples of newborn human infants, the brown adipocytes were observed to decrease in fat content in the hours directly after birth, suggesting the burning of fat in thermogenic activity similar to that observed in rodents.

Contrary to finding in rodents, the presence of the tissue in humans was found to gradually disappear during adolescence, particularly from more peripherally situated areas, while persisting in more deeply situated areas until as late as the eighth decade. It was inferred that the collective function of the tissue was to form a thermogenic jacket, particularly during the first years of life when other heat regulation mechanisms have not yet matured sufficiently to protect against cold. It was assumed that BAT gradually disappeared in later stages of life as the thermogenic capacity of the jacket was no longer required.

However, an investigation of cadaver tissue of winter outdoor workers in 1981 found evidence of greater BAT presence in adult humans adapted to cold. BAT was found in the neck and the pericardium of outdoor workers, compared to indoor workers who had none. The amount of BAT was greater in outdoor workers who were exposed to cold recently, as the outdoor worker who perished in February had more multilocular adipose tissue than the worker who died in August. Similarly, retired farmers also showed a decrease in BAT after their return to warmer conditions. These findings suggest that humans can retain BAT during the later stages of life if

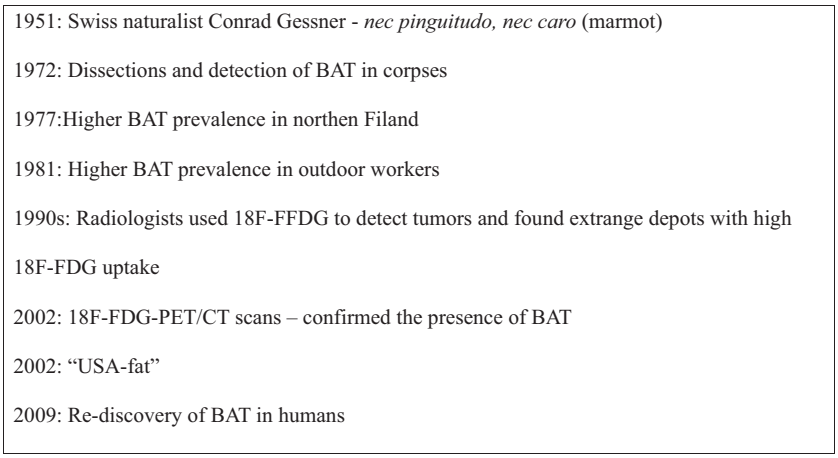


Fig. 1 The Discovery of brown adipose tissue in humans.

sufficiently exposed to cold, but that BAT gradually disappears in warm environments. It also implied that BAT activity can be enhanced by recent cold exposure.

The role of thermogenesis in human metabolism remained controversial. Until its discovery in live adult humans in the 21st century, the general belief was that adult humans only possess vestigial amounts of BAT and that its metabolic effects in humans are negligible. However, the physiology of BAT continued to be studied in rodents. Thermogenesis was found to be defective in several animal models of obesity, and it was theorized that BAT plays an important role in rodents in maintaining leanness through promoting energy expenditure.

The re-discovery of active brown adipose tissue in live humans

Evidence of BAT in living humans was finally discovered in 2002. The tissue was detected on fluorine-18 fluorodeoxyglucose (FDG) positron emission tomography (PET) scans, which had been introduced mainly for oncological imaging. ^{18}F-FDG uptake represents glucose uptake and is normally present in the brain, the kidneys and the urinary collecting system, in various malignancies including carcinomas, lymphomas and melanomas, as well as in inflammatory sites and exercised muscle. A specific symmetrical pattern of ^{18}F-FDG uptake in the shoulder, neck and thoracic spine region appearing on the PET scans of several cancer patients was reported in 1996. The presence of the pattern was unusual; however, the resolution of PET scans was insufficient to pinpoint the exact origin of the signal, and at first it was attributed to muscular uptake. In 2002, it was investigated the pattern by means of co-registered ^{18}F-FDG PET and computed tomography (CT), which can visualize the density of tissues with a higher spatial resolution. The combination of functional imaging by PET with the images from CT allowed exact anatomical localization of the ^{18}F-FDG uptake pattern, and it became clear that the pattern originated not from muscles, but from adipose tissue. Due to its anatomical location, it became briefly known as Uptake in the Supraclavicular Area (USA) - fat, although it was hypothesized that it represented active BAT.

After that, questions arose why it had taken so long to recognize BAT in the ^{18}F-FDG uptake pattern, as there was ample evidence of BAT in humans. However, at the start of the 21st century, it was still the general belief that adult humans only possess vestigial amounts of BAT. Evidence for BAT in adult humans had been largely ignored, until the discovery that the tissue is metabolically active and can be visualized by ^{18}F-FDG PET/CT scan, which challenged the accepted dogma.

In 2007, Nedergaard et al. published an authoritative review on the previous ^{18}F-FDG PET/CT findings. The observations of metabolically active fat on ^{18}F-FDG PET scans were published in nuclear medicine journals since the uptake of ^{18}F-FDG by BAT was considered a disturbing complication during the diagnosis of tumors. However, with the review by Nedergaard et al. the findings were viewed in a positive light, and BAT was hypothesized to be of metabolic significance. Although further investigation of BAT in live humans would require dedicated studies, the evidence gathered from PET/CT so far demonstrated that the knowledge of the physiology of BAT in animal models could also be relevant to humans.

The physiology of brown adipose tissue

The discovery of active BAT in living humans led to the hypothesis that BAT was of metabolic significance. A concerted effort was made to investigate the physiology of human BAT.

Adaptive thermogenesis by brown adipose tissue

Mammals generate heat and increase their energy expenditure in response to cold exposure in a process known as adaptive thermogenesis. The two main tissues known to be responsible for adaptive thermogenesis are skeletal muscles, which contribute to adaptive thermogenesis by shivering in response to cold exposure, and BAT, which contributes by non-shivering thermogenesis (NST). However, the contribution of BAT to NST remains unclear. A substantial difference between rodents and human adaptive thermogenesis is the ratio of energy expenditure between shivering and non-shivering thermogenesis, as the contribution of skeletal muscle is much greater in humans compared to rodents (Blondin et al., 2015a,b).

Adaptive thermogenesis in response to cold exposure

Cold exposure is detected by the brain, leading to the activation of the sympathetic nervous system which innervates both BAT and skeletal muscle. NST in BAT is activated in response to adrenaline and noradrenaline, which are the main effectors of the sympathetic nervous system. Evidence of this interaction was gathered from mice lacking adrenaline and noradrenaline as a consequence of knocking out the gene that encodes dopamine beta-hydroxylase. The mice were found to be intolerant to cold, as they were unable to activate thermogenesis in BAT. Similar intolerance to cold was observed in mice lacking the β1, β2 and β3-adrenergic receptors. The observations obtained from animal models were validated by the results of PET/CT scans in humans. Cold exposure also increases ^{18}F-FDG uptake by BAT, which could be fully suppressed by keeping the subject warm during the procedure. The activation of human BAT is similarly sympathetically stimulated, as demonstrated by early observations when patients scheduled for PET scans were administered the β-blocker diazepam as a treatment for nervousness. In response, ^{18}F-FDG uptake by BAT was inhibited.

In response to the sympathetic signal, adaptive thermogenesis is achieved by increasing mitochondrial respiration. During respiration, mitochondria catabolize nutrients and synthesize adenosine triphosphate (ATP) from adenosine diphosphate (ADP), producing heat as a byproduct. As mitochondrial respiration continues, cellular ADP stores deplete, and ATP will need to be consumed to resume heat production. Several mechanisms are known through which cells can use the synthesized ATP to generate additional heat. One mechanism is the activation of a futile cycle, where two metabolic pathways run simultaneously in opposite directions. An example of a futile cycle is the simultaneous activity of the glycolysis and gluconeogenesis metabolic pathways: when the futile cycle is complete there is no overall effect other than the dissipation of energy as heat. A second mechanism for using ATP is employed by skeletal muscle cells. ATP is converted to ADP by myosin when it releases actin during muscle relaxation. During shivering, muscle contraction and relaxation result in the consumption of ATP and the generation of heat.

An alternative method of heat production is present in brown adipocytes. Instead of producing and using ATP during mitochondrial respiration, the production of ATP is bypassed altogether. In normal mitochondria, the production of ATP is invariably linked to respiration, as ATP is produced in a sequence of coupled reactions. This signifies that the catabolism of a nutrient always results in the generation of a fixed amount of NADH and $FADH_2$, which in turn results in a fixed number of protons moving across the inner mitochondrial membrane and the production of a fixed number of ATP units by ATP synthase. In brown adipocytes, the production of ATP is uncoupled from mitochondrial metabolism, so that nutrients can be catabolized and directly converted to heat without requiring the activation of other metabolic pathways.

Mitochondrial metabolism is uncoupled in brown adipocytes

Brown adipocytes possess specific proteins to uncouple the production of ATP from mitochondrial respiration. The uncoupling is the result of a proton leak in the mitochondria of brown adipocytes, which is catalyzed by uncoupling protein (UCP)-1. In normal mitochondria, nutrient catabolism eventually results in a proton gradient across the inner mitochondrial membrane. Protons can only move across the membrane through channels in ATP synthase, resulting in the synthesis of ATP. However, the presence of UCP-1 in the mitochondria of brown adipocytes allows protons to leak through the membrane and bypass ATP synthase, which allows the energy of the proton movement to be directly converted to heat. Current evidence indicates that BAT thermogenesis is fully dependent on mitochondrial uncoupling by UCP-1. Two homologs of UCP-1 have been identified: UCP-2 and UCP-3. However, no indication of uncoupled brown adipocyte mitochondria was found in UCP-1-ablated mice, even though UCP-2 and UCP-3 expression was unchanged, and the absence of UCP-1 in mice led to low cold tolerance.

The activity of UCP-1 in brown adipocytes is regulated by β-adrenergic receptor stimulation by the sympathetic nervous system. Activation of the β-adrenergic receptor results in an increase in cellular cAMP, which activates lipolysis through protein kinase A. The increase in free fatty acids generated by lipolysis acutely stimulates UCP-1 activity and increases thermogenesis.

Brite adipocytes can be recruited in response to cold exposure

Rodents can recruit additional UCP-1 expressing adipocytes as part of the long-term adaptation to cold. Although cold-induced adipocytes express UCP-1 and other functional thermogenic genes, they were found to lack transcription factors associated with classic brown adipocytes. The newly discovered adipocytes were named recruitable or inducible BAT, or beige or "brite" (brown-in-white) adipocytes, as they are located in WAT but were capable of expressing brown adipocyte markers such as UCP-1 during cold acclimation. These cells are thought to originate from white adipocytes, which transdifferentiate into beige adipocytes after stimulation of the β3-adrenergic receptor in a process known as the "browning" of fat. The transdifferentiation of white adipocytes has been found to occur in major WAT depots in mice, including the anterior subcutaneous and abdominopelvic depots. Mice were also found to be capable of recruiting brown adipocytes from precursor cells in the intrascapular BAT depot; however, this process is ontogenically different from fat browning.

Cells with an expression profile similar to murine beige adipocytes were also found in humans, which indicates that fat browning occurs in humans as well. Histological analysis of the BAT deposits detected in ^{18}F-FDG PET/CT scans showed that human BAT consists partially or even completely of beige adipocytes, which suggests that fat browning may play an important role in cold adaptation in humans.

Brown adipose tissue activity in humans

After the physiology of BAT was elucidated, it was hypothesized that BAT activation could potentially prevent obesity. The next step was to investigate the presence of BAT in humans and its activity during cold exposure.

Retrospective analysis of brown adipose tissue activity

Since ^{18}F-FDG PET/CT was the only available noninvasive technique to measure BAT activity in humans, it became the most widespread method and remains so currently. As the method had previously been in use in oncology, a large number of ^{18}F-FDG PET/CT scans were available to be retrospectively analyzed. Several retrospective studies reported a prevalence of BAT in humans of around 5–10% (36–38), which suggested that BAT has a limited role in metabolism. However, a review by Lee et al. revaluated PET scans

from several studies and concluded that the prevalence of BAT is higher than the previous publications indicated (Lee et al., 2010). The detection of BAT was found to be particularly hard to reproduce; among patients with detectable BAT, a second PET/CT scan was only positive in 13.3% of cases.

Brown adipose tissue detection by cold exposed ^{18}F-FDG PET/CT scan

The retrospective studies also identified a significant correlation between the probability of BAT detection and outdoors temperature at the time of the scan. Since BAT activity was known to depend on the environmental temperature in animal models, ^{18}F-FDG PET/CT was used to investigate BAT activity during cold exposure in humans. BAT was found to be detectable in 23 out of 24 subjects during cold exposure, but not under thermoneutral conditions. BAT could also be detected in obese subjects during cold exposure, which suggests that BAT is present in the obese and activation of BAT can be induced. Increased BAT activity in response to cold exposure was subsequently reported in several studies (Chen et al., 2016). The successful detection of BAT proved to be partially dependent on the environmental temperature and the exact method of cold exposure prior to the scan (Chen et al., 2016). Measurement of BAT activity was found to be most successful when using a personalized protocol aimed at maximizing non-shivering thermogenesis by water cooling.

Dysfunction of brown adipose tissue in obese individuals

BAT glucose uptake is lower in obese compared to lean individuals. It was assumed that BAT thermogenesis was impaired in obese individuals because BAT glucose uptake had earlier been found to correlate with BAT activity in healthy individuals.

The ^{18}F-FDG PET/CT method visualizes glucose uptake, which is generally understood to imply metabolic activity. However, it can be hypothesized that glucose uptake by BAT does not reflect an increase in metabolic rate in the tissue. This hypothesis is supported by evidence from cultured rodent brown adipocytes, where glucose uptake is directly adrenergically stimulated, even in the absence of UCP-1. However, conflicting evidence from live mice indicates the opposite, as BAT was unable to take up glucose following noradrenaline administration if UCP-1 was knocked out. As additional support for the hypothesis that glucose uptake and BAT activity are disconnected, BAT glucose was found to be predominantly utilized for liponeogenesis in rats, and BAT is known to predominantly utilize fatty acids, not glucose, as a fuel source.

Brown adipose tissue glucose uptake may be independent of thermogenic capacity

Older age, higher BMI and type 2 diabetes were all previously associated with defective BAT due to lower glucose uptake by BAT. In 2015, Blondin et al. investigated whether BAT thermogenesis is truly impaired in individuals with type 2 diabetes, compared to both age-matched controls and young healthy control subjects (Blondin et al., 2015a,b). They used ^{18}F-FDG PET/CT to monitor BAT glucose uptake during cold stimulation in subjects with type-2 diabetes. ^{18}F-FDG was supplemented with two other radiotracers: ^{18}F-fluoro-thiaheptadecanoic acid (FTHA), a non-esterified fatty acid (NEFA) tracer and ^{11}C-acetate, a measure for oxidative activity. As expected, the observed volume of ^{18}F-FDG-positive BAT, as well as glucose uptake per BAT volume, were significantly lower in both the subjects with type-2 diabetes and the obese age-matched controls compared to the healthy control subjects. However, after cold exposure, the three subject groups showed similar increases in oxidative activity and NEFA uptake per BAT volume. BAT effectively metabolized fatty acids in response to cold exposure, despite a reduction in ^{18}F-FDG-positive BAT volume and glucose uptake in individuals with type 2 diabetes. These results imply that BAT oxidative metabolism is not blunted by obesity or type-2 diabetes as previously thought, but that the reduced uptake of ^{18}F-FDG by BAT is a consequence of insulin resistance.

Alternative PET tracers to quantify brown adipose tissue activity in humans

These findings also demonstrate that ^{18}F-FDG PET/CT may not be an adequate measure of BAT activity in obese individuals, as it could underestimate BAT activity. Uptake of FTHA may theoretically be a better marker for BAT activity, but the fatty acid tracker is of limited use on its own as it offers little contrast between active BAT and WAT. During cold exposure and in lean individuals, uptake of fatty acids such as FTHA by BAT is also rather slow compared to ^{18}F-FDG. This makes fatty acid tracers a suboptimal choice for the imaging of BAT activity. The third tracer used in the investigation by Blondin, ^{11}C-acetate, as well as ^{15}O-labelled O_2 allow for the estimation of tissue oxidative metabolism. However, both isotopes are cyclotron-produced and have a very short half-life, so their use is restricted to research institutes with access to an in-house cyclotron.

The metabolic significance of brown adipose tissue

After the presence of active BAT in humans was demonstrated, its metabolic significance was assessed. Studies with animal models have shown that the presence and activation of BAT provides significant health benefits for rodents. The absence of BAT activity quickly results in obesity in animal models. The benefits of active BAT were subsequently investigated in humans.

Brown adipose tissue is metabolically significant

Alike in animal models, the presence of BAT was found to have health benefits in humans. Individuals with detectable ^{18}F-FDG uptake by BAT had significantly lower glucose levels and body weight compared to individuals without detectable BAT. In agreement with these findings, Becher et al. (2021) found similar correlations after reviewing 134,529 ^{18}F-FDG PET/CT scans of 52,487 patients. Individuals with detectable BAT showed a significantly lower prevalence of type 2 diabetes, dyslipidemia, cardiovascular disease, coronary artery disease, cerebrovascular disease, congestive heart failure and hypertension. Additionally, while obesity was negatively correlated with detectable BAT, obese individuals who retained detectable BAT appeared to be protected against cardiovascular disease. In a follow-up study of 19,019 ^{18}F-FDG PET/CT scans the presence of detectable BAT was found to correlate with an improved metabolic profile, as the presence of detectable BAT was associated with lower white blood cell count and blood triglycerides and higher high-density lipoprotein. This indicates that the positive effects can be partially explained by the healthier body fat distribution associated with BAT. These findings suggest that active BAT might be beneficial for obese individuals. It is possible that BAT can be reactivated by cold exposure, which in turn can protect against type 2 diabetes and cardiovascular disease.

Contribution of human brown adipose tissue to cold-induced thermogenesis

Several studies aimed to investigate BAT energy expenditure in response to a cold stimulus. In one study, cold stimulation resulted in a 1.8 fold increase in whole-body energy expenditure, although the contribution of BAT was not assessed. However, other findings were less optimistic. In a study BAT energy expenditure was measured separately by ^{15}O-labelled O_2 PET imaging. The increase in BAT energy expenditure in response to a cold stimulus was found to be relatively low at less than 20 kcal/day. In a similar study by U Din et al., seven healthy subjects were scanned by PET/CT for the uptake of ^{15}O-labelled O_2, ^{15}O-labelled H_2O and FTHA at room temperature and during acute cold exposure. Whole-body energy expenditure was measured by performing indirect calorimetry simultaneously with PET scans. Cold exposure resulted in an increase of 2–47% in whole-body energy expenditure, and oxygen consumption was likewise increased. However, the change in BAT-specific energy expenditure in response to cold exposure was estimated to be 4 ± 3 kcal/day, which accounted for only 1% of the total change in whole-body energy expenditure. Meanwhile, the increase in energy expenditure by skeletal muscle in the field of view of the PET scanner correlated with the increase in whole-body energy expenditure. These results suggest that skeletal muscle in the vicinity of BAT, rather than BAT itself, is the main contributor to the increase in energy expenditure in response to cold exposure. These findings are in line with an earlier investigation, where glucose uptake in response to cold exposure was found to be more than one order of magnitude greater in skeletal muscle compared to BAT.

Brown adipose tissue as an endocrine organ

As previously shown, BAT activity is thought to play a role in lipid metabolism. However, combustion of fatty acids by BAT likely plays only a minor role, as energy expenditure of activated BAT was found to be relatively low compared to whole-body energy expenditure. Recent evidence suggests that BAT may instead indirectly affect lipid metabolism through its endocrine function. This hypothesis is based on the observation that the transplantation of BAT could improve glucose tolerance, increase insulin resistance and reduce body weight in high-fat diet-induced obese insulin-resistant mice. These results surpass what could be expected from the thermogenic activity of the transplanted tissue. Instead, BAT likely influences systemic signaling through the secretion of brown adipokines, so called batokines (**Fig. 2**). The genes encoding proteins secreted by BAT have been identified as part of the pattern of genes expressed by BAT in response to cold stimulation.

Fibroblast growth factor 21 (FGF21) was among the first batokines discovered to be secreted by BAT in mice. FGF21 has been suggested to have an autocrine role in stimulating the browning of adipose tissue, as it has been shown to increase expression of UCP-1 and other thermogenic genes in murine WAT, and FGF21-deficient mice are impaired in their ability to adapt to chronic cold exposure. Experiments with human adipocytes *in vitro* support this theory, as treatment with FGF21 was able to induce UCP-1 expression, respiratory uncoupling and heat production in white adipocytes. FGF21 secretion seems to be induced by cold exposure along with the myokine irisin, which suggests that these autocrine agents play a role in the long-term adaptation to cold by the stimulation of fat browning.

Secretion of FGF21 by BAT may also influence cardiac health. FGF21 was found to be expressed by the liver and the heart in rodents, and expression is elevated in the heart of animal models of induced cardiac hypertrophy and myocardial infarction. Secretion of FGF21 by the heart was shown to protect against cardiac hypertrophy in mice, suggesting FGF21 to be a cardioprotective molecule secreted by the heart in response to stress. BAT secretion of FGF21 was also found to provide cardiac protection during hypertension in mice. A similar interaction could be expected in humans, considering the previously reported beneficial effects of BAT activity on human cardiovascular health. FGF21 expression levels are increased in the failing human heart, and FGF21 was found to act as an antioxidant factor in the heart by inducing the expression of several enzymes involved in the removal of reactive oxygen species in cardiomyocytes. These findings suggest a cardioprotective role for FGF21 secreted by the heart, which may also apply to FGF21 secreted by BAT, as BAT has been shown to contribute significantly to systemic FGF21 levels, particularly during cold exposure.

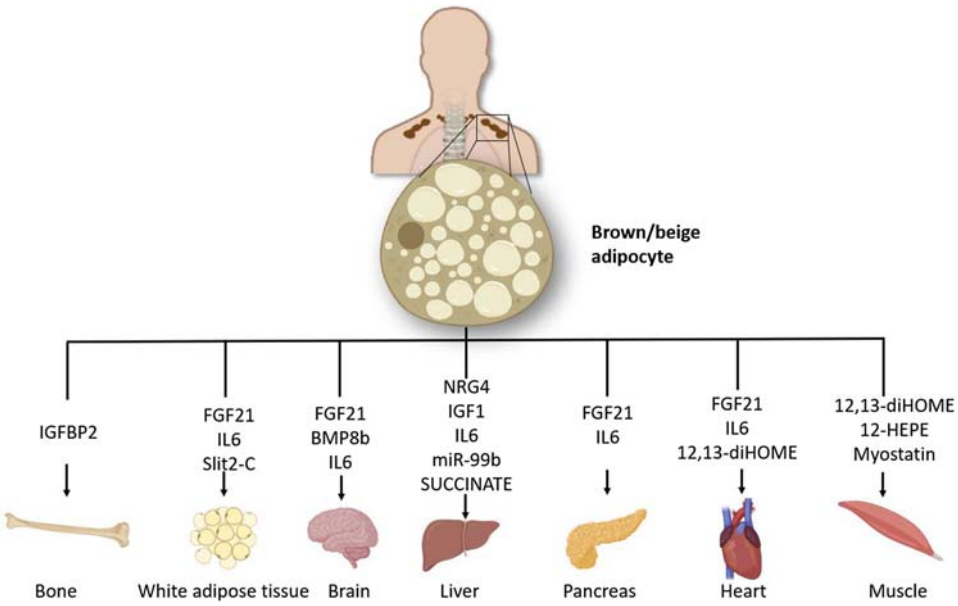


Fig. 2 Endocrine factors released by brown and beige adipocytes might signal to distinct organs, such as bone, white adipose tissue, brain, liver, pancreas heart or skeletal muscle. BMP8b: bone morphogenetic protein 8b; FGF21: Fibroblast growth factor 21; IL6: Interleukin 6; Slit 2-C: Slit Guidance Ligand 2; NRG4: Neuregulin 4; 12,13-diHOME: 12,13-dihydroxy-9*Z*-octadecenoic acid; 12-HEPE: 12-Hydroxyeicosapentaenoic acid.

A second batokine implicated in cold adaptation during long-term cold exposure is Interleukin-6 (IL-6). IL-6, along with the cytokines interleukin-1 and tumor necrosis factor alpha, has previously been found to stimulate heat production and induce fever by acting on the brain during the systemic inflammatory response. However, it was also found to be important for cold-induced thermogenesis. In mice, IL-6 was shown to be required for the browning of WAT in response to long-term cold exposure, as IL-6 knockout mice were found to have a blunted induction of UCP-1 protein expression in WAT after cold exposure. In humans, blocking IL-6 signaling as a treatment for rheumatoid arthritis results in weight gain, which suggests a reduction in energy expenditure through blunted adaptive thermogenesis similar to the observations in rodents. IL-6 was also shown to be directly involved in fat browning in humans. Secretion of IL-6 was observed in *ex vivo* differentiating beige adipocytes, but not in classical brown adipocytes. Blocking of the IL-6 receptor by specific antibody resulted in the downregulation of UCP-1 and other brown adipocyte marker genes in the differentiating cells, and led to morphological changes that are characteristic of white adipocytes. Based on these observations, IL-6 secretion by beige adipocytes may be part of autocrine positive feedback that stimulates the differentiation of adipocytes toward a beige phenotype to increase the capacity for thermogenesis.

Another identified batokine is myostatin, which is also known to be secreted by skeletal muscle. The main function of myostatin was found to be the inhibition of muscle growth. In mice, myostatin loss of function leads to muscle hypertrophy and the suppression of body fat accumulation, as well as WAT browning. In humans, increased secretion and expression of myostatin are associated with obesity. The effect of myostatin was found to be inhibited by follistatin, which can be secreted by skeletal muscle, the liver and BAT. Follistatin is a promotor of skeletal muscle growth that was also found to directly stimulate murine WAT browning *in vitro*. Myostatin and follistatin expression are both known to be influenced by exercise. In humans, the concentration of circulating follistatin acutely increases during exercise. Furthermore, eight weeks of concurrent training resulted in a significant decrease in myostatin serum concentration, while serum follistatin was increased. The same study also observed significant improvements in weight, muscle mass and body composition. From these results, it can be hypothesized that the secretion of myostatin and follistatin facilitates crosstalk between skeletal muscle and BAT, resulting in muscle growth and fat browning after exercise. In mice, myostatin secretion by BAT was found to significantly impact exercise capacity. If the same holds true for humans, BAT activity can potentially influence human energy metabolism by controlling skeletal muscle function.

Potential novel batokines have been recently identified

Possible novel batokines have been discovered by performing high-sensitivity mass-spectrometry-based proteomics on the cell media of brown and white adipocytes. A total of 101 proteins were exclusively discovered in the secretome of brown adipocytes; the method was validated by the discovery of vascular endothelial growth factor A (VEGF-A) among them, which has been previously reported as being secreted by developing BAT. Certain growth factors, granulins and hepatoma-derived growth factor (HDGF), were found to be more abundant in brown compared to white adipocyte cell media, and are possible novel batokines. These growth factors have not been associated with BAT before, though granulins have been associated with the mediation of systemic insulin resistance in mice and humans. Among the proteins secreted by BAT was also mammalian ependymin-related protein 1 (EPDR1), which has been suggested to act as a lipid transporter. After further investigation, human EPDR1 knockout brown

adipocytes were found to display blunted thermogenic activity, and EPDR1 knockout mice had a deficiency in thermogenic capacity and a pronounced accumulation of body fat. These data suggests that EPDR1 plays a role in the development of brown adipocytes. However, injection of EPDR1 also resulted in an increase in energy metabolism in mice without an increase in BAT activity, though no long-term effects of EPDR1 injection on metabolism were observed. This suggests that EPDR1 is not only involved in the development of functional thermogenic adipocytes, but also acts in an endocrine fashion. The pathway through which EPDR1 acts remains to be explored.

A subpopulation of acetate-secreting brown adipocytes may regulate thermogenesis

It has been reported a previously unknown population of adipocytes in mice that can regulate the activity of other adipocytes. A small subpopulation of both mature white and brown adipocytes was found to express a set of genes including the marker gene Cyp2e1 and Aldh1a1. The cells were first thought to represent recruited brown adipocytes, but this was discredited by the observation that the population increased under thermoneutral conditions. Instead, the cells were found to inhibit thermogenesis in brown adipocytes through the secretion of acetate by acting on G-protein-coupled receptor 43 (GPR43) on the surface of brown adipocytes.

Acetate is related to insulin sensitivity before, but the findings were inconclusive. Supporting the findings of Sun et al., circulating acetate has been negatively associated with peripheral insulin sensitivity in obese individuals. However, a beneficial effect of orally ingested acetate on insulin sensitivity was reported in rodent studies, and colonically administered acetate was found to increase energy expenditure and fasting fat oxidation in obese humans. Maybe circular and colonic acetate may play different roles in the regulation of metabolism. It can also be proposed that acetate produced by the gut microbiome is not entirely cleared by the liver, and may enter the circulation to interfere in thermoregulation by endocrine brown adipocytes. However, current data remains inconclusive and elucidating the role of acetate in obesity will require further research.

Brown adipose tissue as a therapeutic target

BAT activity may increase energy expenditure and could have cardiometabolic positive effects through the secretion of batokines (Fig. 2). Thus, pharmacological approaches to the activation of BAT are of interest.

Activation of brown adipose tissue to treat obesity

BAT is an attractive therapeutic target for the treatment of obesity. Activation of BAT increases energy expenditure, which could lead to a significant reduction in body weight if the effects are sustained. Activation of BAT through short-term cold exposure has also been shown to increase insulin sensitivity, and active BAT may have a beneficial effect on lipid metabolism and cardiovascular health, as described in the previous section. Although daily cold exposure is effective in increasing BAT activity, it may be challenging to adhere to a lifestyle involving acute cold exposure. Consequently, current research aims to develop pharmacological compounds that can activate BAT. Among the proposed pharmacological agents are β-adrenergic receptor (β-AR) agonists, which may be able to activate BAT through the adrenergic sympathetic pathway, and peroxisome proliferator-activated receptor-γ (PPAR-γ) agonists, which may increase BAT thermogenic capacity.

Activation of brown adipose tissue by β-adrenergic receptor agonists

The activation of BAT through the adrenergic sympathetic pathway was first studied in rodents. Three β-adrenergic receptors were identified: β1-AR, β2-AR and β3-AR. Stimulation of β3-AR was found to increase energy expenditure in mice (Berbée et al., 2015), although β3-AR agonists failed to produce clinical effects on weight loss in humans. As these studies were performed before the development of ^{18}F-FDG PET/CT imaging, it was unclear if β3-AR agonists were capable of stimulating BAT in humans. Interest in β3-AR agonists reignited in 2015, when the new β3-AR agonist mirabegron was approved to treat overactive bladder. Stimulation with mirabegron was found to increase lipid oxidation, supraclavicular skin temperature and a decrease in BAT fat fraction and an increase in BAT activity as measured by ^{18}F-FDG PET/CT. However, there were also reports of unwanted cardiovascular responses, which led to the suspicion that high doses of mirabegron were not selective to the β3-AR, and that the responses were the result of general β-AR stimulation. This hypothesis was proven to be correct when a high dose of mirabegron was found to increase heart rate and WAT lipolysis, which are β1- and β2- AR-dependent responses. Furthermore, the activation of BAT was found to rely on stimulation of β2-AR. *In vitro* stimulation of brown adipocytes with the highly selective β2-AR agonist formoterol increased uncoupled respiration, while the β3-AR agonist mirabegron had no such effect (Blondin et al., 2020). The previously reported clinical results of the administration of mirabegron were attributed to its activation of the β2-AR at high doses. The effect of β2-AR stimulation on BAT activity in humans has yet to be investigated.

Other pharmacological approaches to the activation of brown adipose tissue

Products that have been proposed to stimulate BAT activity through means other than the activation of the adrenergic sympathetic pathway include capsaicinoids, the glucagon-like peptide-1 receptor (GLP-1R) agonist exenatide, the dipeptidyl peptidase-4 (DPP4)

inhibitor sitagliptin and secretin. Capsaicinoids are alkaloid found in chili peppers that are thought to enhance energy expenditure when ingested. Capsaicinoid analogs, known as capsinoids, have been found to be able to activate BAT, and the combination of capsinoids with cold exposure has a synergistic effect on BAT development. This suggests the activation of BAT by cold exposure can be enhanced by the consumption of capsinoids.

Stimulation of the GLP-1R receptor was found to enhance BAT thermogenesis in mice. In humans, the GLP-1R agonist exenatide was found to have no effect on energy expenditure in obese individuals. However, exenatide has an anorexic effect that could lead to a decrease in energy expenditure in the short term, which could obfuscate potential increases in BAT thermogenesis. Following this hypothesis, the GLP-1R agonist liraglutide was found to increase energy expenditure when observed during a 1-year study period. Furthermore, administration of exenatide was recently shown to increase BAT glucose uptake during cold exposure as measured by ^{18}F-FDG PET/CT scan, which is thought to represent an increase in BAT volume. The enhanced BAT activity in response to exenatide is thought to be mediated by an increase in sympathetic output to BAT; whether other GLP-1R agonists can similarly increase BAT activity remains to be investigated.

Sitagliptin is another promising agent that was found to stimulate BAT activity by increasing UCP-1 expression in adipose tissue in mice. However, administration of sitagliptin failed to increase ^{18}F-FDG uptake by BAT in obese humans, although improvements in glucose tolerance and lipid metabolism were observed. It was hypothesized that insulin resistance obfuscated the effects of sitagliptin by affecting ^{18}F-FDG uptake. As such, BAT stimulation by sitagliptin could be more accurately assessed in the future by using fatty acid PET/CT trackers.

The hormone secretin has also been found to activate BAT in mice and humans. Secretin is a peptide hormone secreted by the duodenum that is known to regulate food intake by acting on the hypothalamus. It was recently shown to be able to activate BAT and stimulate meal-associated thermogenesis in humans, independent of sympathetic innervation. It is hypothesized that the postprandial activation of meal-associated thermogenesis by secretin induces satiation and reduces feeding through a gut-BAT-brain axis. The potential of secretin in reducing food intake will have to be assessed in future clinical trials.

Adipose tissue browning by peroxisome proliferator-activated receptor-γ agonists

Thiazolidinediones are a class of PPAR-γ agonists primarily used to increase insulin sensitivity for the treatment of type-2 diabetes. Thiazolidinediones primarily act on adipose tissue and have been shown to induce UCP-1 expression in human adipose-derived stem cells. The browning of adipose tissue in response to PPAR-γ agonists was also observed in rodents. However, the PPAR-γ agonist-induced increase in BAT mass was not associated with an elevation in thermogenesis; an increase in energy expenditure occurred only after subsequent pharmacological stimulation of the adrenergic sympathetic pathway by β-AR agonists. This observation led to the hypothesis that adipose tissue could be primed by treatment with PPAR-γ agonists, followed by the activation of BAT by β-AR agonists. However, a later study found that PPAR-γ activation impairs BAT activation, likely caused by the inhibition of β-AR expression by PPAR-γ agonists, and the combined treatment did not yield significant results in humans. Furthermore, thiazolidinediones and other PPAR-γ agonists are associated with an increased risk of cardiovascular events. It is currently unclear if the benefits of treatment with PPAR-γ agonists can outweigh the risks.

Future perspectives

Several new strategies for the detection and activation of BAT can be considered. In the future, these strategies may lead to the effective treatment of obesity and cardiovascular disease through the activation of BAT.

New methods for to assessment of brown adipocyte tissue mass and function

The accurate quantification of BAT is a challenge, as no specific imaging biomarker has been reported. Although ^{18}F-FDG PET/CT is the most commonly used method to measure BAT activity, its application has limits. As such, there has been a search for novel imaging probes that can consistently detect BAT. A potential strategy is the use of a BAT-specific PET/CT tracker, such as ^{11}C-PBR28, which is a translocator-specific ligand that can be used to assess BAT mass under thermoneutral conditions. An alternative to PET/CT is near-infrared spectroscopy (NIRS), which can be used to non-invasively measure changes in light absorption in the tissue. NIRS can be employed to estimate BAT mass under thermoneutral conditions by measuring tissue oxygenation, and optically active probes can also allow for the imaging of metabolites such as fatty acids. NIRS could be a useful non-invasive strategy for the assessment of BAT activity, although further validation is necessary (Acosta et al., 2019).

Batokines as a therapeutic target

The beneficial effects of BAT on lipid metabolism and cardiometabolic health can be partially attributed to its secretion of batokines. These endocrine factors could represent potential targets for the treatment of metabolic disorders. For example, the administration of FGF21 mimetic antibodies was able to reduce body weight and improvement in cardiometabolic parameters in obese humans. This suggests that the pharmacologic activation of pathways targeted by batokines can achieve positive effects associated with active BAT.

Transplantation of brown adipocyte tissue

BAT transplantation has considerable beneficial effects in rodent models of obesity. Transplantation reverses obesity in genetically obese mice, and BAT-deficient mice experience increases in energy expenditure and insulin sensitivity when transplanted with BAT from healthy mice. A different approach does not rely on a healthy BAT donor mouse, but instead involves a tissue-grafting strategy where WAT is converted to BAT through *ex vivo* browning and reimplanted. This strategy could potentially be used to treat obesity in humans lacking active BAT.

Conclusion

The activation of BAT represents a potential approach for the treatment of obesity and cardiovascular disease in humans. A possible strategy for the treatment of obesity is the sustained activation of BAT through pharmacological compounds such as β2-AR agonists to achieve weight loss. Future strategies should also consider the endocrine function of BAT, through which it may protect against type 2 diabetes, cardiovascular disease and heart failure. Batokines represent potential targets for the treatment of metabolic disorders. The stimulation of BAT by means of pharmacological agents may lead to the effective treatment of obesity and cardiovascular disease, but further research is warranted.

References

Acosta, F.M., Berchem, J., Martinez-Tellez, B., Sanchez-Delgado, G., Alcantara, J.M.A., Ortiz-Alvarez, L., Hamaoka, T., Ruiz, J.R., 2019. Near-infrared spatially resolved spectroscopy as an indirect technique to assess Brown adipose tissue in young women. Mol. Imag. Biol. 21 (2), 328–338.

Berbée, J.F., Boon, M.R., Khedoe, P.P., Bartelt, A., Schlein, C., Worthmann, A., et al., 2015. Brown fat activation reduces hypercholesterolaemia and protects from atherosclerosis development. Nat. Commun. 6, 6356.

Becher, T., Palanisamy, S., Kramer, D.J., Eljalby, M., Marx, S.J., Wibmer, A.G., et al., 2021. Brown adipose tissue is associated with cardiometabolic health. Nat. Med. 27 (1), 58–65.

Blondin, D.P., Labbé, S.M., Noll, C., Kunach, M., Phoenix, S., Guérin, B., et al., 2015a. Selective impairment of glucose but not fatty acid or oxidative metabolism in Brown adipose tissue of subjects with type 2 diabetes. Diabetes 64 (7), 2388–2397.

Blondin, D.P., Labbé, S.M., Phoenix, S., Guérin, B., Turcotte, É.E., Richard, D., et al., 2015b. Contributions of white and brown adipose tissues and skeletal muscles to acute cold-induced metabolic responses in healthy men. J. Physiol. 593 (3), 701–714.

Blondin, D.P., Nielsen, S., Kuipers, E.N., Severinsen, M.C., Jensen, V.H., Miard, S., et al., 2020. Human Brown adipocyte thermogenesis is driven by β2-AR stimulation. Cell Metabol. 32 (2), 287–300.e7.

Chen, K.Y., Cypess, A.M., Laughlin, M.R., Haft, C.R., Hu, H.H., Bredella, M.A., et al., 2016. Brown adipose reporting criteria in imaging STudies (BARCIST 1.0): recommendations for standardized FDG-PET/CT experiments in humans. Cell Metabol. 24 (2), 210–222.

Lee, P., Greenfield, J.R., Ho, K.K., Fulham, M.J., 2010. A critical appraisal of the prevalence and metabolic significance of brown adipose tissue in adult humans. Am. J. Physiol. Endocrinol. Metab. 299 (4), E601–E606.

Nedergaard, J., Bengtsson, T., Cannon, B., 2007. Unexpected evidence for active brown adipose tissue in adult humans. Am. J. Physiol. Endocrinol. Metab. 293 (2), E444–E452.

Chrononutrition

Francisco Javier Ruiz-Ojeda[a] **and Marta Garaulet**[b,c], [a] Department of Biochemistry and Molecular Biology II, School of Pharmacy, University of Granada, Granada, Spain; [b] Department of Physiology, University of Murcia, Murcia, Spain; and [c] IMIB-Arrixaca, Murcia, Spain

© 2023 Elsevier Ltd. All rights reserved.

Introduction 37
Circadian system and chronodisruption 37
How does the circadian system work? 37
What is chronodisruption? 38
Chronodisruption, obesity, and metabolic dysfunction 39
Mutations in experimental animals related to obesity and metabolic dysfunction 39
Genetic variations of clock genes in human obesity and metabolic dysfunction 39
Genetic variations associated with obesity 39
Genetic variations interact with several behaviors for obesity 40
Clock genes and body weight loss during obesity treatments 40
Other pathologies 40
Epigenetic mechanism of clock genes in obesity 40
Adipose tissue as a circadian organ in obesity 42
The circadian system and nutrition (chrononutrition) 43
Timing of food intake as an external synchronizer of the biological clock 43
Organs and systems involved in chrononutrition 45
Types of food timing interventions and effect on obesity and metabolic dysfunction 45
Calorie restriction and intermittent fasting 45
Timing of food intake in nutrigenetics 46
Conclusions 47
References 47

Key points

- Chronodisruption: a serious disturbance of the internal temporal order of the biochemical, physiological and behavioral circadian rhythms
- Epigenetics: science that studies the heritable regulation of gene expression without change in nucleotide sequence.
- Nutrigenetics: science that aims to study the effect of genetic variations on the interaction between diet and disease.
- Single-nucleotide polymorphism (SNP): genetic variations in a single nucleotide at a specific position in the genome. These variations are present in more than 1% of the general population.

List of abbreviations

AMPK 5′-prime-AMP-activated protein kinase
BDNF brain derived neurotrophic factor
BHLHE40 basic helix-loop-helix family member e40
BMAL1 aryl hydrocarbon receptor nuclear translocator-like
CLOCK clock circadian regulator
CREB1 cAMP responsive element binding protein 1
CRY1 cryptochrome circadian regulator 1
CRY2 cryptochrome circadian regulator 2
FBXL3 F-box and leucine rich repeat protein 3
GLUT2 glucose transporter 2
HSL hormone-sensitive lipase
LC3A microtubule associated protein 1 light chain 3 alpha
MIF macrophage migration inhibitory factor
MTNR1B melatonin receptor 1B

mTOR mechanistic target of rapamycin kinase
NR1D1 nuclear receptor subfamily 1 group D member 2
PER1 period circadian regulator 1
PER2 period circadian regulator 2
PLIN1 perilipin 1
PPARG peroxisome proliferator activated receptor gamma
PRKAG2 protein kinase AMP-activated non-catalytic subunit gamma 2
REV-ERB-α (nuclear receptor subfamily 1, group D, member 2 alpha)
RORA RAR related orphan receptor A
SGLT-1 solute carrier family 5 member 1
SIRT1 sirtuin 1

Introduction

What we eat plays an important role in our health, however, it is well established that when we eat might be crucial as well, regardless the number of calories (i.e., total energy intake). Our physiology changes during the day, and the optimal synchronization of behavioral and physiological events is an essential mechanism in humans and other species. The circadian system acts as a synchronizer and is affected by other external synchronizers such as the changes between light and darkness, eating and fasting, or activity and resting, and it influences sleep, body temperature, hormones, and appetite. Consequently, several hormones that are related to obesity (cortisol, leptin, adiponectin, insulin, among others) display circadian rhythmicity (Garaulet and Madrid, 2010).

When all this system is altered, we talk of chronodisruption (CD), which is related to several disturbances in many systems and organs of our body. The study of how nutrition relates to our circadian system has been recently called Chrononutrition, which involves energy intake distribution along the day, meal frequency and regularity, duration of the eating and the fasting periods, and the relative importance of these factors in metabolic health and the risk of chronic diseases. Evidence from both animal and human research reveals adverse metabolic consequences of circadian disruption. Conversely, aligning food intake to periods of the day when circadian rhythms in metabolic processes are optimized for nutrition may be effective for improving metabolic health. The importance of the "when" we eat is tied to our internal 24 h biological timing system, the circadian clock, regulating the metabolic processes across the body.

In our modern society, CD may be produced by external situations that are relatively common such as jet-lag, shift work, night light pollution or overnight recreational activities (social jet lag). Other factors are endogenous or internal, and may produce CD by alterations of the core machinery of the molecular circadian clock. BMAL1, PER2 and CLOCK, among others clock proteins, have a specific role in our physiology as well as in the circadian molecular clock. To date, several studies performed in mutant animals have demonstrated that several mutations in clock genes are related to obesity and metabolic disturbances. In humans, different association studies have been focused to elucidate the genetic variations in one SNP that underlie differences in the vulnerability to diseases. Furthermore, epigenetic changes such as DNA methylation at different CpG sites (i.e., regions of DNA where a cytosine nucleotide is followed by a guanine nucleotide) of *CLOCK* are associated with obesity and obesogenic behaviors, indicating that eating behaviors could help to modulate our destiny. Hence, nutrigenetics may help to change our eating habits in order to change our behavior and our adiposity. These association studies between clock genes and several behaviors will be explained in the following lines. In this article, we will describe how the circadian system works and how chronodisruption may aggravate body weight gain and increase adiposity. Besides, we describe how the timing of food intake is an external synchronizer of the biological clock and why eating at a "wrong" timing may have metabolic consequences. We will explain which are the organs and systems involved in chrononutrition, and describe different types of food timing interventions and which are their potential effects on obesity and metabolic dysfunction. Finally, we will describe the particular role of food timing in nutrigenetics and its contribution to the development of new strategies for personalized nutrition programs in overweight and obese people.

Circadian system and chronodisruption

How does the circadian system work?

The circadian rhythm is a natural, internal process that regulates the sleep–wake cycle, which is repeated on each rotation of the Earth, roughly every 24 h. This system allows the organism to anticipate and adapt to the predictable rhythmic changes in the environment (e.g., light/dark cycle). In mammals, endogenous circadian rhythms are produced by a multi-oscillator system composed of the central circadian pacemaker, located in the hypothalamic suprachiasmatic nucleus (SCN), as well as by peripheral clocks in virtually every organ, tissue, and cell (Fig. 1). The SCN is primarily entrained by light through direct photic inputs received from the retina that are transmitted to the SCN via the retinohypothalamic tract. It functions to synchronize peripheral clocks via

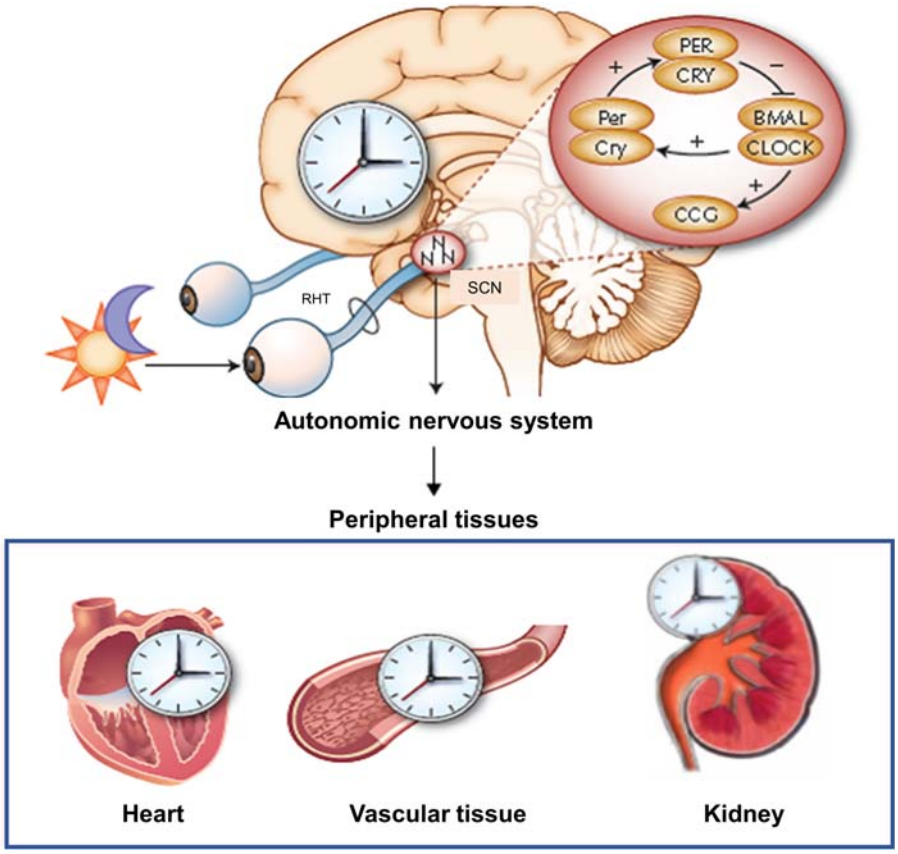


Fig. 1 Schematic representation of how the endogenous circadian system works. Hypothalamic suprachiasmatic nucleus (SCN) is entrained by light through direct photic inputs from retina to the SCN via retinohypothalamic tract (RHT). It synchronizes through autonomic nervous system to peripheral tissues such as heart, vascular tissues, kidney, among others.

a combination of neuronal, behavioral, body temperature, and endocrine outputs (Garaulet et al., 2020). Light is the dominant "zeitgeber" (time giver) for the SCN oscillator, which in turn orchestrates rhythms in the peripheral organs/tissues at appropriate phases. At the epicenter of the molecular complex that constitutes the circadian clock are the core transcription factors CLOCK and BMAL1 that orchestrate the transcription of their own repressors, period (PER) and cytochrome (CRY), forming a self-regulated feedback loop. The internal timing system, or circadian clock, coordinates the rhythmic regulation of physiological and behavioral processes in mammals. Therefore, chronobiology comprises different aspects related to the internal clock and its genes and proteins such as the study of (a) the external synchronizers (i.e., light, food intake, or exercise); (b) several hormones and tissues highly related to the clock; and (c) physiological aspects such as nutrient processing, sleep/awakening, body temperature, hormone secretion, tissue repair and many others that when altered may be related to disease.

What is chronodisruption?

CD is a relevant disturbance of the circadian organization of physiology, endocrinology, metabolism and behavior, which links light, biological rhythms and the development of several diseases, being melatonin a key biological intermediary. The terms "circadian disruption" or "disruption of circadian rhythms" refer to the fact that 24 h rhythms can become desynchronized and that this may have adverse effects on health. In the modern lifestyle, factors such as shift work, night eating syndrome and sleep disorders can disrupt the inner biological rhythm and impair synchronization between the internal clock and the metabolic rhythm. In humans, even 2 h of "social jetlag", a disruption to the typical sleep wake cycle, is associated with adverse endocrine, behavioral and cardiovascular risk factors in apparently healthy individuals, further illustrating the significance of the circadian regulation on the biological and behavioral rhythms. Nowadays, it is widely demonstrated the existence of tight connections between circadian-related factors and obesity, diabetes, heart and metabolic diseases, psychosocial mental disorders such as depression and anxiety and certain types of cancer.

Chronodisruption, obesity, and metabolic dysfunction

Mutations in experimental animals related to obesity and metabolic dysfunction

Circadian rhythms are widely characterized at a molecular level and they are generated through the expression of several clock genes. Studies performed in experimental animals with specific mutations in different clock genes have demonstrated a relationship between these mutations, further failures in the circadian system, and metabolic alterations and diseases. Indeed, a genetic mutation in the homozygous mutant mice for *clock*, displayed obesity, adipocyte hypertrophy and metabolic syndrome. In addition to that, *Bmal1* (−/−) mice had reduced lifespan (mean lifespan of knockout animals was 37 weeks vs. 120 weeks for wild type animals) and displayed various symptoms of premature aging, including sarcopenia, alteration of the percentage of lymphocytes and impaired vision among others. Furthermore, *Bmal1* (−/) mice have suppressed diurnal variations in glucose and triglycerides and abolished gluconeogenesis while. Together with these examples, many other mutations in clock genes have been related with metabolic diseases, such as those in *Per2* knockout mice, that have shown to display obesity when fed a high-fat diet (HFD); or those in *Cry1/2*−/− mice that display increased insulin secretion and lipid storage in adipose tissue and elevation of proinflammatory cytokines.

Genetic variations of clock genes in human obesity and metabolic dysfunction

Genetic variations associated with obesity

Even though mutations are rare in humans, genetic variations are present in more than 1% of the population and are characterized by variations in one SNP at some specific position of the genome. As predicted by Turek et al. (2005) in the homozygous mutant mice for *clock*, in humans, genetic variants at *CLOCK* are associated with obesity, especially with abdominal adiposity, and this gene plays a relevant role in the development of metabolic syndrome, type 2 diabetes (T2D) and cardiovascular disease (CVD). Similarly, Garaulet and Madrid (2010) revealed several *CLOCK* SNPs (rs3749474, rs4580704, and rs1801260 (3111T > C)) associated with metabolic syndrome and body mass index (BMI), energy intake, and different variables related to obesity (Fig. 2). In fact, the individuals carrying the minor alleles ate more quantity (and energy), ate more fat, and were more obese. On the other hand, minor allele carriers (A) of *CLOCK* rs4580704 showed decreased risk of developing diabetes (31% lower) and hypertension (46% lower) than non-carriers. Besides, one genetic variation in *CLOCK* rs1801260 (3111T > C) was associated with increased obesity and with lower weight loss effectiveness during dietary treatments for obesity. Some of these associations are functionally explained by the presence of a polymorphism involving a change in the structure of the mRNA leading to a change in gene expression (Garaulet and Madrid, 2010).

As predicted by studies in knock-out *Per2*−/− mice which displayed obesity when fed a high fat diet, in humans several SNPs in *PER2* (rs2304672C > G and rs4663302C > T) have been associated with abdominal obesity and obesogenic behaviors. In particular, *PER2* rs2304672C > G minor allele carriers G (6% of the population) showed several obesogenic behaviors such as an

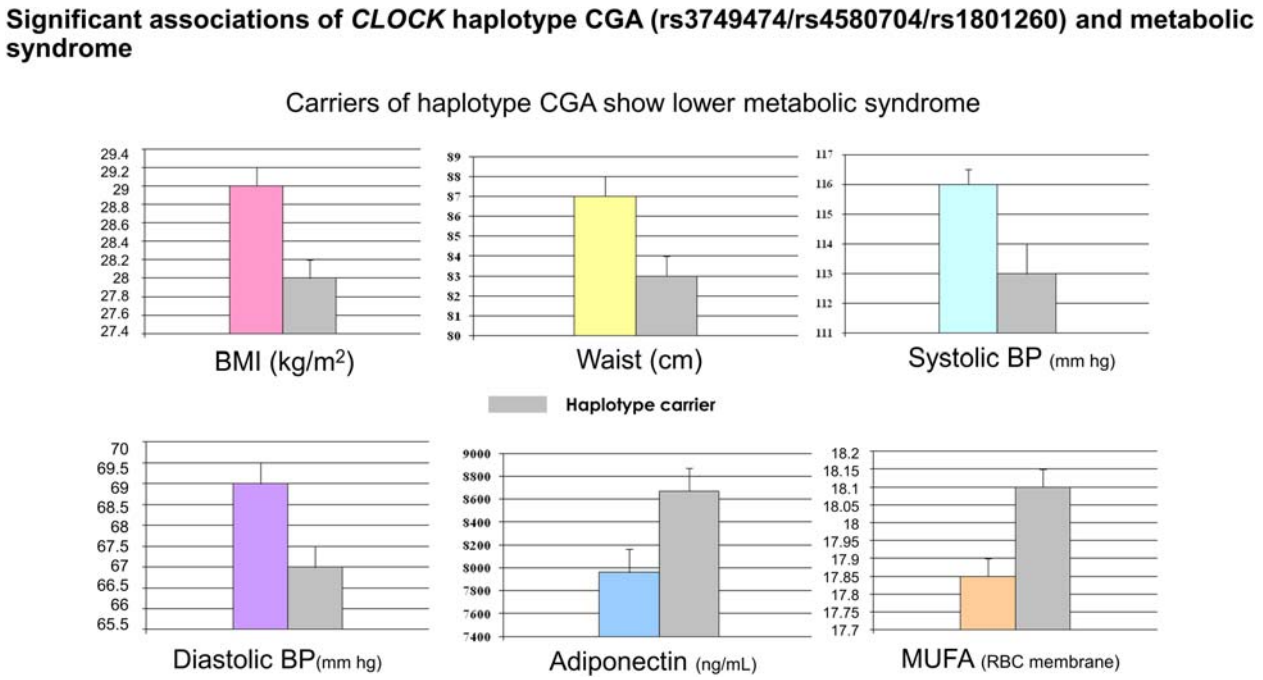


Fig. 2 Significant associations of CLOCK haplotype CGA (rs3749474/rs4580704/rs1801260) and metabolic syndrome. BP, blood pressure; FA, fatty acid; MUFA, monounsaturated fatty acid; RBC, red blood cell.

increased attrition of the weight-loss treatment, increased frequency of snacking, stress while dieting, eating while bored and skipping breakfast, when compared with non-carriers C (Garaulet and Madrid, 2010). Other results from Garaulet and Gómez-Abellán (2014) group also showed that different genetic variants of the clock such as rs2314339 in *REV-ERB-α* were also associated with obesity and this association replicated in two independent populations: a Mediterranean population from Spain and a North American population. *REV-ERB-α* is the major regulator of rhythmic *BMAL1* transcription and it is considered to be the molecular link between the positive elements of the clock (*CLOCK* and *BMAL1*) and the negative (*PER* and *CRY*). The already mentioned genetic variant at *REV-ERB-α* (rs2314339) has been associated with obesity through a decrease in physical activity, differently to other clock genetic variants, which were associated with obesity through changes in dietary intake. These results were also predicted by studies in mutant mice for *Rev-erbα*$^{-/-}$, which displayed reduced spontaneous locomotor activity.

Genetic variations interact with several behaviors for obesity

Clock genes may also interact with behaviors for obesity, some examples are those behaviors directly related to emotional eating. In this way, those emotional eating individuals who attended to a weight-loss program and who carried the risk allele C at *CLOCK* 3111T > C had difficulties to lose weight during the intervention. However, those carriers of the risk allele C that were not emotional eaters, together with those non-risk allele carriers T (independently to their emotional eating profile) had no difficulties in losing weight. These findings help to understand that we can reduce the deleterious effect of a genetic variant if we change the "how" we eat.

In addition to genetic variants at *CLOCK*, subjects carrying minor alleles at *SIRT1* and *CLOCK loci* displayed a higher resistance to weight loss and a lower weight loss rate as compared with homozygotes for both major alleles (**Fig. 3**). Besides, carrying the minor alleles for both variants associated with a lower intake of carbohydrates and monounsaturated fats, and a higher intake of saturated fats, suggesting a less adherence to Mediterranean diet patterns among carriers.

In order to achieve a better understanding of the relationship between circadian rhythms and obesity, Bandín et al. (2013) evaluated changes in circadian rhythmicity in C carriers of *CLOCK* 3111T > C SNP as compared to T carriers, by continuously recording body temperature (wrist temperature, WT), actimetry and position during one week. Interestingly, risk carriers (C) displayed a less marked circadian rhythm than TT carriers and a delayed acrophase that characterizes "evening-type" subjects (Bandín et al., 2013).

Clock genes and body weight loss during obesity treatments

One of the main problems in weight loss treatments is the dramatic inter-individual variability in response to the treatment. It is believed that an elucidation of the genetic component may help to predict weight loss effectiveness. Some studies carried out in monozygotic twins analyzed genetic factors for effectiveness of weight loss programs, other studies emphasized the familial aggregation in the ability to lose weight; and the role that parental obesity plays in this matter. In this context, Garaulet et al. (2011) detected in a Spanish overweight and obese population a significant interaction between the *PPARγ* Pro12Ala polymorphism and fat intake for total weight loss. When total fat intake was high, Ala12-carriers exhibited a significantly lower weight loss than major allele carriers indicating that *PPARG2* genotype could be used as a predictive gene to improve the weight loss effectiveness. Clock genes are also related to body weight loss. Indeed, *CLOCK* 3111T > C (rs1801260) associated with the effectiveness of a weight loss program, demonstrating that C carriers are more resistant to lose weight than TT homozygotes. These data suggest that *CLOCK* SNPs may predict weight loss in response to a low-energy diet.

Other pathologies

In addition to obesity and weight loss management, circadian rhythms regulate key biological processes influencing several metabolic pathways and therefore affecting the metabolic risk for T2D and CVD. For example, a significant association has been observed between the *CLOCK*-rs4580704 SNP and the incidence of T2D and CVD in 3671 non-T2D participants of PREDIMED study, with G allele carriers showing decreased incidence (dominant model) compared with CC homozygotes. Besides, the protection was higher in the Mediterranean diet intervention group compared to control, and only those T2D subjects carrying the *CLOCK*-rs4580704 SNP had higher risk of stroke, suggesting that core clock genes may significantly contribute to increased CVD risk in T2D.

Epigenetic mechanism of clock genes in obesity

Epigenetics is the study of heritable phenotype changes that do not involve alterations in the DNA sequence. In particular, DNA methylation is an epigenetic mechanism that modifies the function of genes and affects gene expression. Aberrant DNA methylation has been found to be associated with various complex human diseases, including obesity. With regards to the circadian system, DNA methylation at different CpG sites of *CLOCK* are more frequent in women with obesity and have been associated to several obesogenic behaviors such as an increased frequency of snacking, eating when bored or eating from great packages. Similarly, Milagro et al. (2012) demonstrated significant associations between the methylation levels of several CpGs located in *CLOCK* with the metabolic syndrome, with difficulties in weight loss and with obesity. In this study, performed in 60 women, it was demonstrated that the methylation degree in several CpG sites of *CLOCK*, such as CpG 1, 5, 6 and 8, increased with obesity. More importantly, the "how" we eat was related to methylation levels at *CLOCK*. The authors observed that those patients who tended to snack more frequently had 12 times higher methylation levels in *CLOCK* CpG1 than those who did not usually snack; whereas those individuals who tended to eat when bored or to eat from great packages had 9 to 19 times higher methylation levels, respectively (**Table 1**). These increases in methylation suggest a suppression of *CLOCK* expression, which has also been related to obesity. Furthermore,

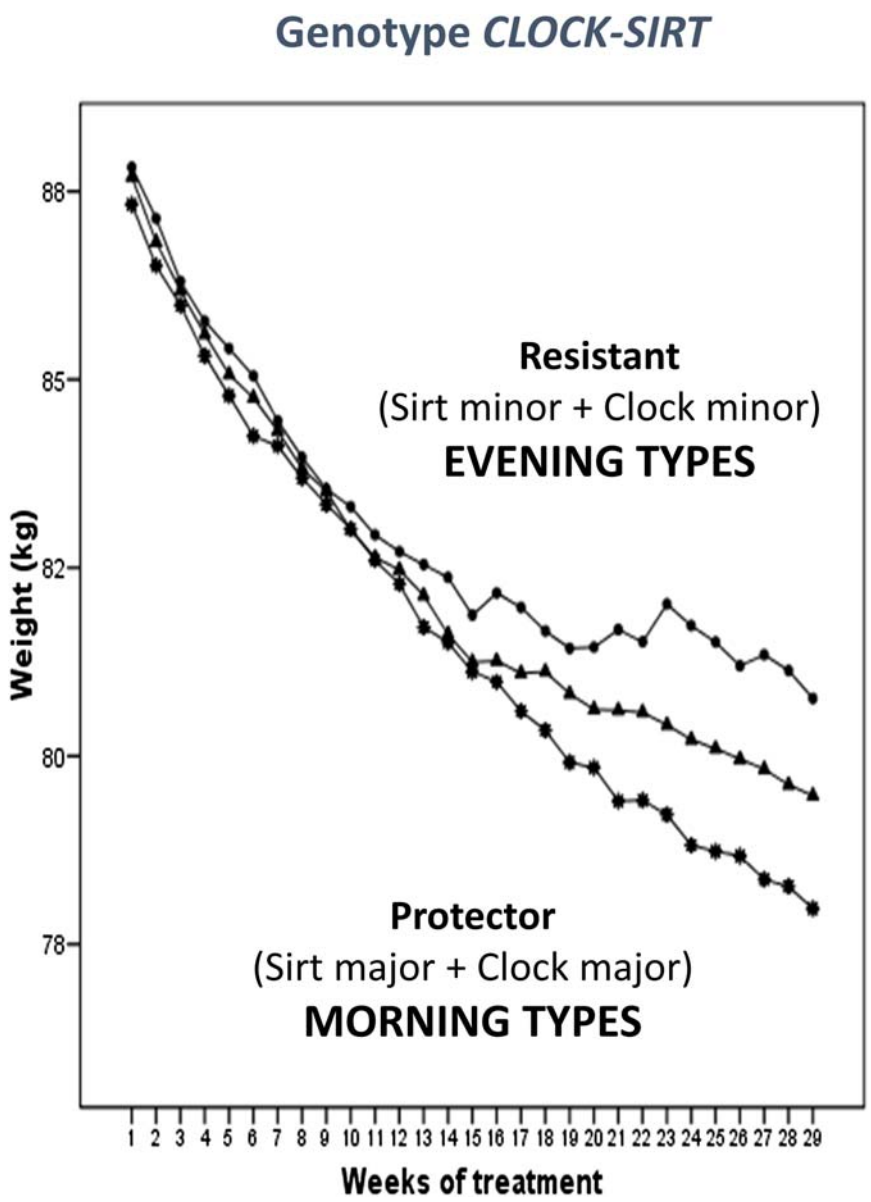


Fig. 3 Weight loss and Clock genes. Weight loss progression during 30 weeks of treatment in subjects of different SIRT1-CLOCK combined genotypes after adjusting for sex, age and study center. SIRT1 minor and CLOCK minor in evening types for subjects resistant in total weight loss (kg). SIRT1 major and CLOCK major in morning types for subject's protector. Differences between the groups ($P < 0.050$).

Table 1 Effect of different behaviors in methylation levels of *CLOCK* CpG1, which shows the importance of "how we eat".

CLOCK *CPG1 (highly methylated)*	*Methylation levels*	P-*value*
Snacking frequency	12x	0.026
Eat fast	9x	0.08
Eat when bored	3x	0.008
Eat from great packages	19x	0.004

Source: Milagro et al. (2012).

research from the same group showed that a weight loss intervention based on a Mediterranean diet modifies the methylation pattern of *BMAL1*, *CLOCK*, and *NR1D1*, and changes in the methylation levels of *BMAL1*, because of the dietary treatment, associated with a reduction in metabolic risk parameters, i.e., serum lipids.

More recently, Monti et al. (2021) found a significant association between the exposure to airborne atmospheric pollutants and methylation at several clock genes in individuals with overweight and obesity. In particular, short-term exposure to airborne particulate matter induced higher methylation of *CLOCK*, *CRY1*, *CRY2*, and *PER3*, and lower of *PER2* in the blood of people with obesity. These findings indicate that BMI may be a variable that might influence the effect of air pollutants on clock genes methylation, clock genes expression, and therefore may affect circadian rhythms in obesity (Monti et al., 2021). Another study found different DNA methylation patterns at six circadian rhythm pathways genes (*RORA*, *PER3*, *BHLHE40*, *FBXL3*, *CREB1* and *PRKAG2)* correlated with BMI, metabolic profiles and dietary intakes, revealing potential associations of DNA methylation profiles at circadian genes with obesity.

A recent study showed a relationship between obesity, depression and epigenetic changes (more DNA methylation) in clock genes *CRY*1 and *CRY2* in women affected by overweight or obesity, establishing an association between BMI and depressive symptoms through clock genes.

Since childhood obesity has dramatically increased in the last decades, a study identified several CpG sites in *PER3*, among other genes, that were differentially methylated between obese and non-obese children, suggesting a role for DNA methylation concerning development of childhood obesity. Overall, epigenetics changes in several clock genes such as DNA methylation are associated with obesity and obesogenic behaviors, indicating that chronobiology is related to obesity though epigenetics.

Adipose tissue as a circadian organ in obesity

The discovery of peripheral clocks in different tissues such as the white adipose tissue, kidney or liver, links those tissues to chronobiology, playing a crucial role in the regulation of the whole body metabolism. These peripheral clocks seem to regulate the rhythmicity of at least 10% of the genes expressed within each tissue, and their discovery has revolutionized our understanding of the physiological processes in the body (Froy and Garaulet, 2018) (**Fig. 4**).

As already discussed in this article, animal studies have shown that a high fat diet may contribute to the development of obesity and insulin resistance via alterations in the circadian period of locomotor activity rhythms and changes in the oscillation of clock genes in adipose tissue. In this context, *CLOCK*, *BMAL1*, *PER1*, *PER2*, *CRY1 and REV-ERBα* play a dual role in regulating the core clock mechanism, as well as adipose tissue metabolism, and link circadian rhythms with lipogenesis and lipolysis.

CD occurs when the synchronization between external environmental cues and internal physiological processes are lost. This can either result in a total loss of rhythmicity, in a reduction in rhythm amplitudes, or in phase differences between the SCN and peripheral clocks. As already discussed earlier in this article, experiments performed in clock gene mutants (*CLOCK*, *BMAL1*, *CRYs*, *PER2* and *REV-ERBα*) and knockouts have demonstrated the most compelling linkage between metabolic disorders and the circadian clock (Froy and Garaulet, 2018). For example, a combination of the *clock* (delta19) mutation with the leptin knockout (*ob*/*ob*) leads

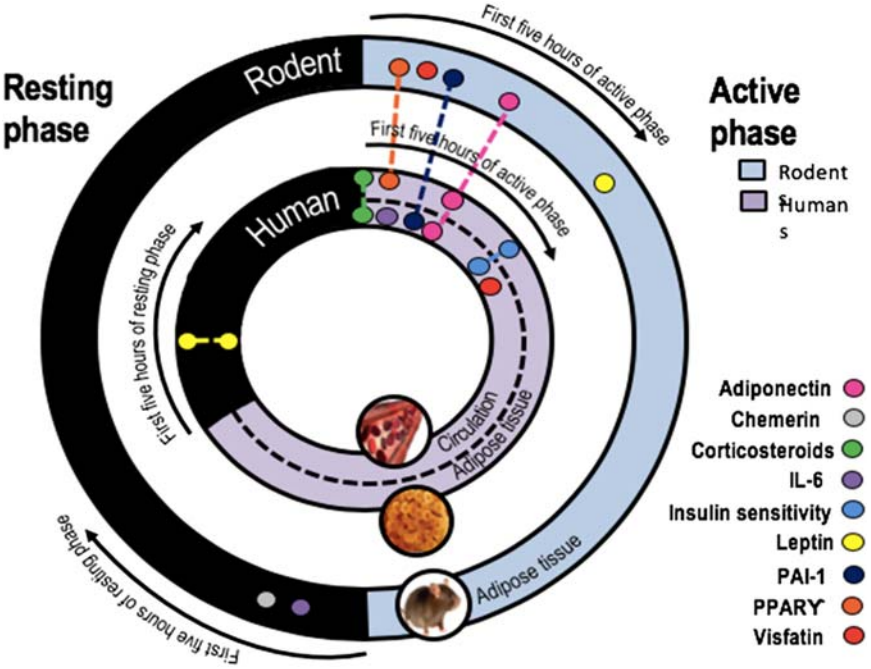


Fig. 4 Temporal expression of adipocytokines in rodents and humans. The peak expression of each adipokine is depicted. Representation of the acrophase or timing of the maximum levels of adipokine expression in human adipose tissue: comparison with human circulating levels and with rodent adipose tissue mRNA levels. Data are represented as the 5 h of active and resting phase in both humans and rodents, as this is the timing when most adipokines peak. IL-6, interleukin 6; PAI-1, plasminogen activator inhibitor 1; PPAR, peroxisome proliferator–activated receptor.

to obesity, reiterating the contribution of clock disruption to the obese phenotype. Similarly, *Per2* knockout mice under high fat diet feeding developed obesity compared with wild-type mice. Adipocyte-specific deletion of *Bmal1* results in obesity in mice and leads to changes in the expression of hypothalamic neuropeptides that regulates appetite, emphasizing the role of the adipocyte clock in the temporal organization of energy regulation.

In humans, glucose tolerance varies with the time of the day and plasma glucose excursions in response to meals, oral glucose, and intravenous glucose are markedly higher in the evening and at night than in the morning. Garaulet et al. (2011) demonstrated that both the negative elements (*PER2* and *CRY1*) and the positive elements (*CLOCK* and *BMAL1*) of the clock showed circadian rhythmicity in *ex vivo* adipose tissue explants and oscillated independently of the SCN for at least two circadian cycles after surgery. Moreover, insulin sensitivity, which is related to body fat accumulation, also shows an endogenous circadian rhythm in subcutaneous adipose tissue, with maximum sensitivity at noon and minimum at midnight. Regarding fat mobilization in white adipose tissue, a recent study has shown circadian rhythms in the hormone-sensitive lipase (HSL) activity in human adipose tissue, which allows a better understanding of the intricate relationships between food timing, fasting duration and body fat regulation (Arredondo-Amador et al., 2020) (**Fig. 5**). Overall, great evidence supports the important role of adipose tissue as a circadian organ in obesity that regulates the whole-body metabolism.

The circadian system and nutrition (chrononutrition)

Timing of food intake as an external synchronizer of the biological clock

Food is one external synchronizer of our peripheral clocks. The primary role of the circadian clock is to entrain the organism to the environmental cues and changes in timing of food may lead to an uncoupling of peripheral oscillators from the central pacemaker. Therefore, unusual feeding time can induce a disruption of the circadian system, which might produce unhealthy consequences in humans. For example, changes in the timing of food intake may alter the circadian rhythmicity of many hormones involved in metabolism, such as insulin, glucagon, adiponectin, corticosterone, leptin, chemerin, lipocain and visfatin. A study reported that the times during which subjects were awake and eating during their biological night resulted in multiple metabolic alterations including increased concentration of both glucose and insulin. Moreover, night leptin values in plasma were decreased, which could influence energy balance. It is known that in humans, plasma leptin levels display circadian rhythms that results from the daily variations in food intake and the endogenous clock.

The time and number of meals fluctuate greatly from culture to culture and through time. Certainly, the timing of food intake is a modifiable behavior that may influence energy regulation and consequently the risk of obesity. The "three meals a day" pattern, which consists of one intake in the morning, another at noon, and another at night, helps to maintain the internal temporal order of the circadian system. However, the current 24 h society in which we live causes us to frequently abandon these patterns, not only for example, by the shift work and jet-lag, but also, and especially in young people, due to social jet-lag, the discrepancy in a person's sleep pattern between the weekday and the weekend, which can cause a person to feel "jet lagged" or tired and fatigued. While social

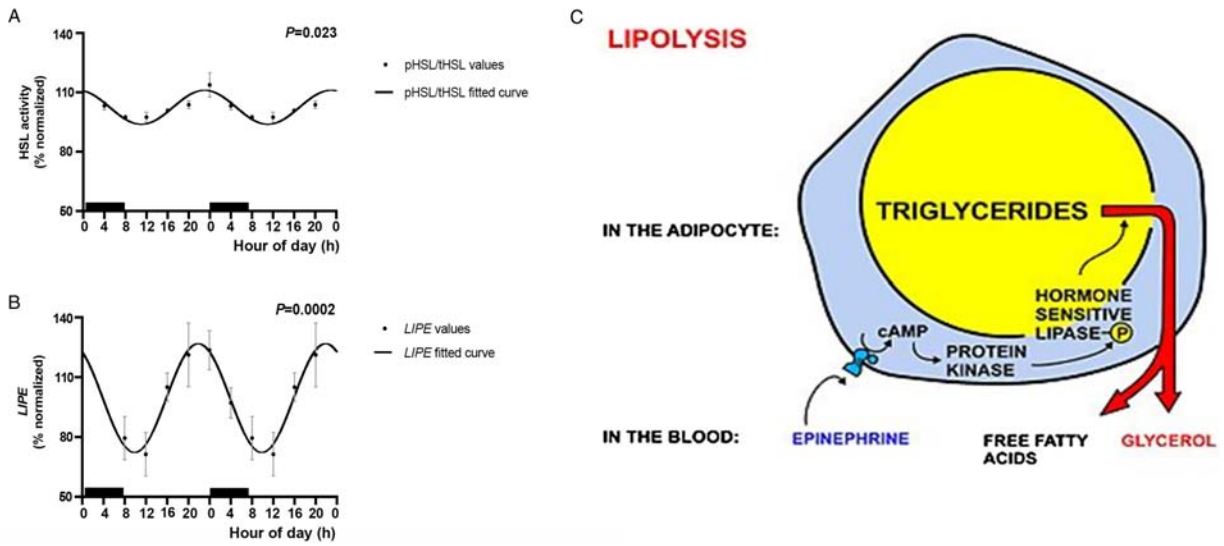


Fig. 5 HSL activity and LIPE expression in subcutaneous AT, circadian rhythms, and phase map. Timing is represented both in circadian time (CT, bottom x-axis) and relative local clock time (top *x*-axis). (A) presence of significant circadian rhythms in HSL activity (pHSL/tHSL) in the total population ($n = 18$). HSL activity is represented in double plotted graphs of a 24 h sinusoidal curve. Data (% normalized) are reported as means ± standard error of the mean SEM) and represented by black dots. The solid line represents the 24 h sinusoidal curve of the total population; (B) Rhythmic expression of LIPE in adipose tissue explants. Data (% normalized) are reported as means ± SEM and represented by black dots. The solid line represents the 24 h sinusoidal curve of the total population. (C) Schematic representation of fat mobilization in adipocyte by lipolysis.

jet lag can affect anyone, the problem is particularly common in teenagers due to the growing demand for study and even leisure and pleasure.

Lunch timing: The first study performed in humans about the timing of meals, was focused in the main meal of the day that is lunch in Spain. Garaulet et al. (2013) evaluated the role of food timing in weight-loss effectiveness in 420 individuals who followed a 20-week weight-loss treatment and they were grouped in early eaters and late eaters. In fact, late lunch eaters lost less weight and displayed a slower weight-loss rate during the treatment. Though energy intake, dietary composition, appetite hormones and sleep duration were similar between groups, the late eaters were more evening types, had less energetic breakfasts and skipped this meal-time more often.

Midpoint of intake: More recently, in a cohort of 3362 subjects (ONTIME clinical trial), Dashti et al. (2020) demonstrated that late eating is associated with cardiometabolic risk factors and reduced efficacy of a weight-loss intervention in a Spanish cohort. They observed that late eaters had higher BMI, higher concentrations of triglycerides, lower insulin sensitivity and higher concentrations of the satiety hormone leptin in the morning (Dashti et al., 2020). In this study, authors used the "Midpoint of intake", as an indicator of food timing. Midpoint of meal intake was determined by calculating the midpoint between the weekly averages for breakfast and dinner times. The reason was that this variable and was an endogenous marker of food intake.

Eating at night: Wang et al. (2014) demonstrated that while energy intake in the morning was not associated with obesity, those who consumed ≥33% of daily energy intake in the evening were two-fold more likely to be obese than morning eaters (**Fig. 6**). Thus, eating more of the day's total energy intake at midday is associated with a lower risk of being overweight/obese. Consequently, lighter changes in meal timing such as the distribution of caloric intake across the normal wake episode, appear to influence not only obesity, but also the success of weight loss therapy.

Breakfast: Indeed, it has been shown in a 12-week experimental study that subjects assigned to high caloric intake during breakfast lost significantly more body weight than those assigned to high caloric intake during the dinner. Furthermore, a study performed by measuring meal timing relative to sleep timing as a proxy marker of circadian time, showed that a higher percent of total daily energy intake consumed during the morning window was associated with lower odds of being overweight or obese. This association was stronger in people with an earlier chronotype. Authors also found that for the earlier chronotype a higher percent of sugar and fiber intake during the morning window was associated with a substantial decrease in the odds of being obese (Xiao et al., 2019). Therefore, it seems that breakfast intake may affect body weight and energy balance. In the last years, the proportion of adults who skip breakfast has increased from 14% to 25%, raising the question of whether there could be a causal relationship between breakfast skipping and obesity. The study demonstrated through Mendelian randomization (MR), the existence of causal links between genetically determined breakfast skipping and higher BMI.

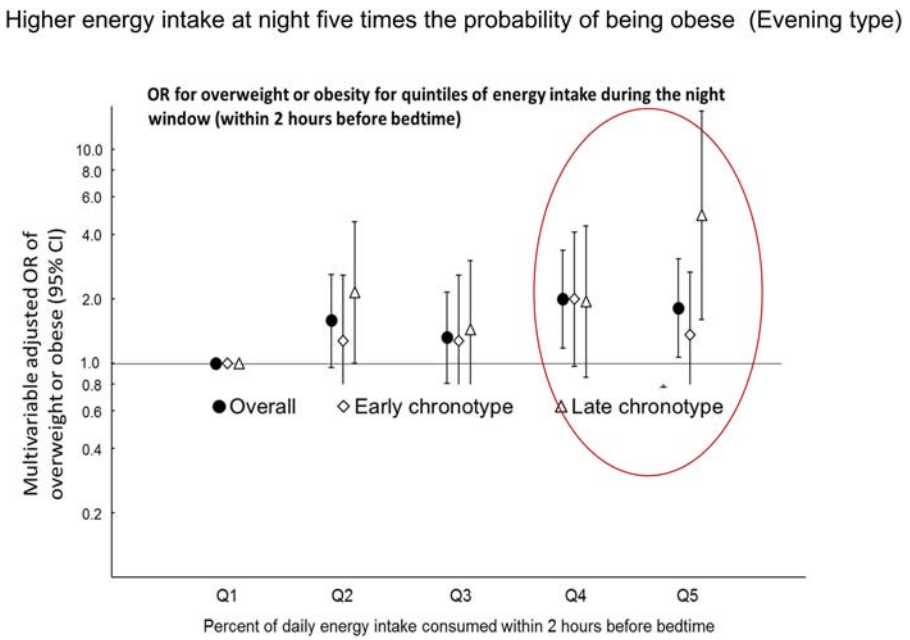


Fig. 6 Associations between percent of total energy consumed during the night window and the odds of being overweight or obese in the overall study population and as divided according to sleep timing. Earlier chronotype was defined as a chronotype earlier than the median (3:04 a.m.), while the later chronotype was defined as a chronotype later than the median. The model was adjusted for age, sex, race/ethnicity, total time in bed, chronotype (in the overall population), total steps per day, duration of sedentary time, and total daily energy intake. All quintiles are based on the overall population. Abbreviations: CI confidence interval, OR odds ratio.

Organs and systems involved in chrononutrition

Different organs and tissues related to food intake present a large number of genes that display oscillations in their expression and encode for important regulators of carbohydrates, lipids and proteins metabolism (Garaulet and Madrid, 2010). These organs are involved in digestive processes as motility, digestion and absorption of nutrients, or in the post absorptive process.

The digestive process begins in the mouth where food is chewed and mixed with saliva, which contains enzymes that begin the chemical process of digestion and, in fact, the salivary α-amylase, which is involved in the chemistry degradation of polysaccharides (starch) to disaccharide such as maltose, is a circadian enzyme. A study has suggested that salivary α-amylase levels may be associated with the individual chronotype being lower in the morning-type than in the evening-type subjects. Although digestion begins in the mouth, most of digestive processes are performed in the stomach and intestine with the help of the pancreas. Nowadays, it is well known that a large number of digestive and intestinal enzymes are expressed in a circadian manner and are synchronized by food.

In stomach: the most important processes related to digestion are those connected to the mechanical digestion. In this sense, it has been demonstrated that the rate of gastric emptying changes during the day, and it shows lower rates after evening meals than after morning meals. Eating late during the day or having a high caloric intake during dinner may be associated with an impaired digestion. Moreover, the stomach is highly related to the control of food intake though ghrelin, one orexigenic hormone secreted by stomach and popularly known as the "hunger hormone". Circulating levels of ghrelin increase before meals and decrease after food intake. Changes in the timing of food intake may modify ghrelin 24 h rhythmicity, and consequently may alter the physiological control of hunger, influencing total energy intake and as a result, weight loss.

Intestine: Absorptive processes in intestine may also be modulated rhythmically. An example is the absorption rhythm of glucose due to the action of several transporters (SGLT-1 in enterocytes and GLUT-2 in the basolateral membrane) which show circadian rhythms. Several authors have revealed that in rats fed *ad libitum*, glucose absorption displays 24 h changes, being lower during the morning hours and higher during the night. Some intestinal enzymes also show synchronization with the feeding rhythm, this is the case of intestinal disaccharidases, especially maltase and sucrase in rats. Their enzymatic activity increases around 1 h before feeding and falls 3 h afterward and it is synchronized by the feeding time. In addition, expression of islet-specific genes such as *GLUT-2* and *MIF* are modulated in a circadian manner, suggesting a potential circadian regulation of pancreatic beta-cells and their implication in insulin production and secretion.

Liver: In relation to systemic metabolism, liver is the most important organ implicated in the post absorptive phase of digestion. Most of the liver genes, which encode enzymes or regulatory proteins involved in food processing, display rhythmic expression, for example glycogen metabolism. Glycogen storage changes throughout the day coinciding with changes in the activity of the key enzymes in its storage, such as glycogen synthase, glycogen phosphorylase and glucose-6-phosphatase. Other examples include cholesterol 7α-hydroxylase (limiting enzyme in the synthesis of bile acids); transcription factors governing fatty acid metabolism, such as PPARs; metabolic enzymes involved in cholesterol metabolism; a number of cytochrome P450 enzymes involved in detoxification and elimination of food components (e.g., coumarin hydroxylase) and many more. Considering that the major function of circadian oscillators in the liver is the coordination of physiological needs during digestive processes, changes in food timing may strongly alter the function of the liver. Besides, the transcriptional regulation of some genes implicated in the hepatic bile acid formation of rats (that consume most of their food during the night) is severely altered when food is provided during the morning. Indeed, in nocturnal animals, feeding during the day inversed the phase of several circadian oscillators in hepatocytes.

Types of food timing interventions and effect on obesity and metabolic dysfunction

Calorie restriction and intermittent fasting

Long-term calorie restriction (CR) without malnutrition has been proposed as one of the most consistent dietary interventions to improve health and extend survival across different species. Thus, changes in fasting time and duration, nutrient composition, circadian rhythms, age of interventional onset, and sex of the individuals are all variables that can affect the effectiveness and outcomes of CR. In the last years, evidence suggests that the length (or duration) of fasting, the dietary composition, and the individual chronotype contribute to the effects of CR. In rodents, CR increases health and survival, but most of the studies have involuntarily introduced changes in fasting time that could be affecting results. CR shows low adherence for most humans, therefore, alternative feeding methods have been developed in an attempt to provide the same benefits as CR (Duregon et al., 2021).

Intermittent fasting (IF): or periodic energy restriction includes eating patterns in which periods of fasting are followed by periods of *ad libitum* feeding (1–3 days per week). Among the forms of IF includes: (1) periodic fasting (PF), which is an alternative to CR consisting of long but less frequent fasting periods. One form is the fasting mimicking diet (FMD), which combines a plant-based, very-low-caloric diet provided continuously for 4 days, twice a month. FMD after 3-month study has been shown to be effective to decrease blood glucose levels and IGF-1 levels, and increase circulating ketones. (2) The 5:2 intermittent fasting (fasting 2 days each week) is another IF type and lead to greater reduction in body fat and improvement in insulin sensitivity in overweight women after 3-month study than a regular energy reduced diet (DER). The success of these 5:2 intermittent fasting may be related not only to the adherence to the 2 d restriction, but also to the spontaneous restriction of energy intake on non-restricted days. Indeed, authors indicate that the intakes on these days were comparable with those of the DER group. Furthermore, authors discuss that "observation suggests that intermittent dieting does not lead to disordered eating and overconsumption on non-restricted days, however other studies have shown that this could depend on the type of person i.e., emotional eaters or individual chronotype etc." (3)

Alternate-day fasting (ADF) is a continuous sequence of a fast day (100% energy restriction) followed by and a feed day (*ad libitum* food consumption) where it shows reduction in body weight, visceral fat and "browning" of WAT in rodents. Moreover, ADF improves insulin resistance and glucose tolerance in rodents for T2D. (4) Time-restricted feeding (TRF, also known as time-restricted eating, TRE), which is defined as the provision of food for up to 12 h during the active phase, and this restriction restores circadian rhythms and imparts pleiotropic metabolic benefits in animal models. Thus, TRF is a form of intermittent fasting that limits the daily food consumption to a period of 4–12 h, which induces a fasting window of 12–20 h per day. However, TRF does not require caloric restriction and it requires a consistent daily eating window.

It has been shown that TRF protects mice against obesity under HFD feeding, insulin resistance, hepatic steatosis and inflammation. Mechanistically, TRF can reprogram the circadian clock in the fasting state via AMPK and in the fed state via mTOR, and TRF increases the amplitude of *cry1* and *per1* in liver of mice fed chow and restored the amplitude of *bmal1*, *cry1*, *per2* and *Rev-erbα* in mice fed HFD. In addition, TRF reduces the pyruvate carboxylase and glucose 6-phosphatase, and increases glucokinase during the active phase, decreasing the glucose production by the liver. TRF also reduces genes related to lipid metabolism in liver during the active phase, reducing the lipid storage and increasing triglyceride hydrolysis. Moreover, TRF protected in knockouts mice (whole-body *Cry1;Cry2* and in liver-specific *Bmal1* and Rev-erbα/β) from HFD-induced weight gain, glucose intolerance, hepatic steatosis, and dyslipidemia, providing evidence that daily rhythms in the fasting and feeding cycle is enough to maintain metabolic homeostasis, independently of circadian clocks.

In humans, TRF have been shown to reduce body weight and fat mass, to improve glucose tolerance, and to reduce blood pressure in humans in those with overweight or obesity. Moreover, TRF have shown to reduce plasma triglycerides and inflammatory markers. A recent clinical trial in overweight adults have shown lower glucose levels, increases ketones, cholesterol and gene expression in the whole blood of *SIRT1* and autophagy gene *LC3A* in the morning, while in the evening, there is an increased in the expression of *mTOR* (major nutrient sensing protein), and a tend to increase brain-derived neurotropic factor (*BDNF*). Besides, TRF also altered diurnal levels of cortisol, and increased the expression of circadian clock genes *BMAL1*, *CRY1*, *CRY2*, and *RORA* in the morning. In the evening, there was a decreased in gene expression of *PER1*, and increased in *CRY1*, *and RORA* in blood, indicating a different circadian pattern under TRF. In addition, TRF in overweight men affects the rhythmicity of serum and muscle metabolites and regulates several amino acid transporter genes and metabolites in muscle, without perturbing core clock gene expression (Lundell et al., 2020). Hence, TRF is recognized as simple and well-tolerated diet with beneficial health effects based on chrononutrition. Nevertheless, the studies to date in humans are limited in size and duration, and the effectiveness and principles acceptability of TRF in the general population remains unclear. Furthermore, one of the key questions related to TRF is the timing of fasting. In this sense, results of a study performed in human adipose tissue culture, which analyzed the circadian changes in the hormone-sensitive lipase (HSL) activity, show that at least 12 h of fasting during the nighttime is able to duplicate the activity of HSL and stimulate body fat mobilization from adipose tissue. Therefore, authors conclude that an easy way to follow a TRF could be advancing dinner timing (for example at 8 p.m.) and delaying breakfast timing (for example to 8 a.m.) achieving a 12 h fasting during night, the natural timing that we are in fasting, (i.e., when we are sleeping). This will facilitate TRF without major changes in individual habits or family and social life.

Commonly, IF represents a dietary approach of interest to reduce body weight and to improve the health by delaying many age-related diseases while improving metabolic markers independently from weight loss. In addition, IF reduces glycemia, improves insulin sensitivity, and decreases systemic inflammatory marker.

Timing of food intake in nutrigenetics

Nutrigenetics points to understanding how the genetic background of an individual impact on the diet. This novel science involves the heterogeneous response of genetic variants to nutrients and dietary components and may be useful in developing nutraceuticals, which may be designed for specific genetic subtypes. Concerning chronobiology, different SNPs of our circadian system interact with food intake and various behaviors for obesity and metabolic risk variables. *CLOCK* polymorphisms interact with dietary intake of different fatty acids subgroups (for example saturated, or MUFA) to modulate several metabolic syndrome traits such as glucose and insulin sensitivity. One example is the protective effect of the minor allele (G) at rs4580704 in the presence of higher MUFAs intake (more than 13% of total energy intake). Afterward, the same group demonstrated an interaction of the *CRY1* gene and several characteristics of diet for insulin resistance both in a Mediterranean and in a North American population. Authors showed that a higher intake of carbohydrate was associated with higher HOMA-IR and lower QUICKI, only in those individuals homozygous for the minor allele CC (rs2287161) of *CRY1* gene. This finding supports the link between the circadian system and glucose homeostasis in order to develop personalized nutrition programs to prevent insulin resistance in CC carriers of *CRY1* rs2287161. Garaulet et al. (2015) also showed that Melatonin Receptor 1B (*MTNR1B*) rs10830963 risk variant G worsens the deleterious effect of melatonin on glucose tolerance in humans. These findings suggest the importance of genotyping for personalized nutrition programs in subjects consuming melatonin, or in late eaters who are exposed to high levels of melatonin when eating late at night and are carrying allele G for *MTNR1B* (Garaulet et al., 2015) (**Table 2**).

The timing of food intake (for example eating lunch late) may also interact with our genes for obesity or weight loss effectiveness. Indeed, studies have demonstrated that eating late the main meal of the day, that is lunch in Spain, may be deleterious but not for everybody. Indeed, the deleterious effect only affect a genetic subtype of *PLIN1* (14994A > T variant). Interestingly, among AA carriers, eating late was associated with less weight loss, whereas the time of eating did not influence weight loss among TT carriers.

Table 2 Effects of melatonin on glucose metabolism.

	Low melatonin *Circadian … light* *…day … exposure*	*High melatonin* *Circadian … exogenous* *…night … melatonin*
Fasting	β cell rest↓	β cell rest↑
Postprandial	β cell function insulin sensitivity↑	β cell function insulin sensitivity↓

Source: Garaulet et al. (2020).

These results contribute to the development of new strategies for personalized nutrition programs in overweight and obese people based on genetics (Garaulet et al., 2016).

All of these studies show how our habits can interact with our genes and reduce or increase the effect of a specific genetic risk variant. Consequently, as we cannot change our genome, we can change our habits in order to improve the health and prevent the disease.

Conclusions

In this article, we summarize that changes in "how we eat", "what we eat" and other daily behaviors such as the timing of food intake can improve our health due to the interaction with our genome. In this sense, experimental studies in mutant's animals and observational studies in humans revealed the association and putative function of clock genes in relation to obesity and weight loss. Furthermore, eating is an external synchronizer of our circadian system and the "when" we eat might play an important role in obesity and metabolic dysfunction. Thus, we have described the impact of time-related eating patterns on weight-loss treatment, and the differences between early eaters and late eaters. Therefore, changes in our behaviors toward early meals may improve the weight-loss efficiency during the treatment. Finally, epigenetics shows that our behaviors can help to change our genome and decrease the risk of several metabolic diseases.

References

Arredondo-Amador, M., Zambrano, C., Kulyté, A., Luján, J., Hu, K., Sánchez de Medina, F., Scheer, F.A.J.L., Arner, P., Ryden, M., Martínez-Augustin, O., Garaulet, M., 2020. Circadian rhythms in hormone-sensitive lipase in human adipose tissue: relationship to meal timing and fasting duration. J. Clin. Endocrinol. Metab. 105, e4407–e4416.

Bandín, C., Martinez-Nicolas, A., Ordovás, J.M., Ros Lucas, J.A., Castell, P., Silvente, T., Madrid, J.A., Garaulet, M., 2013. Differences in circadian rhythmicity in CLOCK 3111T/C genetic variants in moderate obese women as assessed by thermometry, actimetry and body position. Int. J. Obes. 37, 1044–1050.

Dashti, H.S., Gómez-Abellán, P., Qian, J., Esteban, A., Morales, E., Scheer, F.A.J.L., Garaulet, M., 2020. Late eating is associated with cardiometabolic risk traits, obesogenic behaviors, and impaired weight loss. Am. J. Clin. Nutr. 113, 154–161.

Duregon, E., Pomatto-Watson, L., Bernier, M., Price, N.L., de Cabo, R., 2021. Intermittent fasting: from calories to time restriction. Geroscience 43 (3), 1083–1092.

Froy, O., Garaulet, M., 2018. The circadian clock in white and brown adipose tissue: mechanistic, endocrine, and clinical aspects. Endocr. Rev. 39, 261–273.

Garaulet, M., Gómez-Abellán, P., Rubio-Sastre, P., Madrid, J.A., Saxena, R., Scheer, F.A., 2015. Common type 2 diabetes risk variant in MTNR1B worsens the deleterious effect of melatonin on glucose tolerance in humans. Metabolism 64, 1650–1657.

Garaulet, M., Madrid, J.A., 2010. Chronobiological aspects of nutrition, metabolic syndrome and obesity. Adv. Drug Deliv. Rev. 62, 967–978.

Garaulet, M., Gómez-Abellán, P., 2014. Timing of food intake and obesity: a novel association. Physiol. Behav 134, 44–50.

Garaulet, M., Ordovás, J.M., Gómez-Abellán, P., Martínez, J.A., Madrid, J.A., 2011. An approximation to the temporal order in endogenous circadian rhythms of genes implicated in human adipose tissue metabolism. J. Cell Physiol 226 (8), 2075–2080.

Garaulet, M., Gómez-Abellán, P., Alburquerque-Béjar, J.J., Lee, Y.C., Ordovás, J.M., Scheer, F.A., 2013. Timing of food intake predicts weight loss effectiveness. Int. J. Obes. 37 (4), 604–611.

Garaulet, M., Qian, J., Florez, J.C., Arendt, J., Saxena, R., Scheer, F., 2020. Melatonin effects on glucose metabolism: time to unlock the controversy. Trends Endocrinol. Metabol. 31, 192–204.

Garaulet, M., Vera, B., Bonnet-Rubio, G., Gomez-Abellan, P., Lee, Y.C., Ordovas, J.M., 2016. Lunch eating predicts weight-loss effectiveness in carriers of the common allele at PERILIPIN1: the ONTIME (Obesity, Nutrigenetics, Timing, Mediterranean) study. Am. J. Clin. Nutr. 104, 1160–1166.

Lundell, L.S., Parr, E.B., Devlin, B.L., Ingerslev, L.R., Altintaş, A., Sato, S., Sassone-Corsi, P., Barrès, R., Zierath, J.R., Hawley, J.A., 2020. Time-restricted feeding alters lipid and amino acid metabolite rhythmicity without perturbing clock gene expression. Nat. Commun. 11, 4643.

Milagro, F.I., Gómez-Abellán, P., Campión, J., Martínez, J.A., Ordovás, J.M., Garaulet, M., 2012. CLOCK, PER2 and BMAL1 DNA methylation: association with obesity and metabolic syndrome characteristics and monounsaturated fat intake. Chronobiol. Int. 29 (9), 1180–1194.

Monti, P., Iodice, S., Tarantini, L., Sacchi, F., Ferrari, L., Ruscica, M., Buoli, M., Vigna, L., Pesatori, A.C., Bollati, V., 2021. Effects of PM exposure on the methylation of clock genes in a population of subjects with overweight or obesity. Int. J. Environ. Res. Publ. Health 18.

Turek, F.W., Joshu, C., Kohsaka, A., Lin, E., Ivanova, G., McDearmon, E., Laposky, A., Losee-Olson, S., Easton, A., Jensen, D.R., Eckel, R.H., 2005. Obesity and metabolic syndrome in circadian Clock mutant mice. Science 308 (5724), 1043–1045.

Wang, J.B., Patterson, R.E., Ang, A., Emond, J.A., Shetty, N., Arab, L., 2014. Timing of energy intake during the day is associated with the risk of obesity in adults. J. Hum. Nutr. Diet 27, 255–262.

Xiao, Q., Garaulet, M., Scheer, F., 2019. Meal timing and obesity: interactions with macronutrient intake and chronotype. Int. J. Obes. 43, 1701–1711.

Cytokines: Metabolic and nutritional aspects

R.F. Grimble and P.C. Calder, School of Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of R.F. Grimble, Cytokines: Nutritional Aspects, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 423–429, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00066-0.

Introduction 48
Some cytokines affect metabolism 49
Role of cytokines in disease 51
The interaction of nutrition and cytokines 53
Influence of fatty acids on inflammatory cytokine production 54
Interaction of inflammatory cytokines with amino acid and protein metabolism 55
Modulation of inflammatory cytokine production by micronutrients 55
Conclusions and perspectives 56
References 56

Key points

- Cytokines are low molecular weight proteins released mainly from cells of the immune system
- Cytokines are released in response to microbes and tissue injury
- Cytokines act in an autocrine or paracrine manner
- Cytokines include the interleukins, interferons, stem cell growth factors, interferons, tumor necrosis factors and adipokines
- Tumor necrosis factor (TNF)-α, interleukin (IL)-1 and IL-6 have metabolic effects—they induce fever, anorexia, altered metabolism and mobilization of substrates
- Inflammation induces oxidative stress and oxidative stress induces inflammation
- Although part of the host defense mechanism, inflammation can be damaging to the host and must be controlled
- Excessive or uncontrolled inflammation is associated with many diseases including frank inflammatory conditions such as rheumatoid arthritis and also age-related chronic diseases including obesity, type-2 diabetes and atherosclerosis; these are all associated with elevated TNF-α, IL-1 and IL-6
- Long chain n-3 fatty acids can decrease production and action of TNF-α, IL-1 and IL-6
- Through decreasing oxidative stress, antioxidants can decrease production and action of TNF-α, IL-1 and IL-6

Introduction

Cytokines are a broad family of low molecular weight proteins which are released, mainly, though not exclusively, from cells of the immune system, in response microbes and tissue injury, which might be physical, oxidative or metabolic. Cytokines induce a state of inflammation in the body and modulate the activity of the immune system. Although mainly thought of as immune modulators, cytokine production is not restricted to cells in the immune system but fibroblasts, epithelial and endothelial cells, adipocytes, myocytes and specialized tissues, such as the ovary, all produce cytokines. Although largely influencing immune function, including the inflammatory response, a number of cytokines act as growth factors and lead to the proliferation and differentiation of a wide range of cell types in the body. Cytokines act via receptors and generally act in an autocrine or paracrine fashion. Many cytokines are found circulating in the bloodstream. Cytokines include the interleukins, the interferons, many stem cell growth factors, the chemokine family, the tumor necrosis factor family and the adipokines. There is no single cytokine classification system and cytokines are often grouped in different ways depending upon origin or function. For example, monokines are cytokines released from monocytes, lymphokines are cytokines released from lymphocytes and adipokines are cytokines released primarily from adipose tissue. Chemokines are cytokines that mediate the chemically-induced movement of cells known as chemotaxis or chemoattraction. Representative examples of cytokines are detailed in **Table 1**. While all cytokines influence the functions of cells of the immune system (see **Table 1**), some also exert metabolic effects upon the host. These include the three classic proinflammatory cytokines, interleukin (IL)-1, IL-6, and tumor necrosis factor (TNF)-α (Beutler and Cerami, 1986; Dinarello, 1988; Heirich et al., 1990).

Table 1 Selected cytokines, their main sources and their functions. Note that the list is not exhaustive.

Cytokine	*Main sources*	*Main function*
IL-1β	Monocytes/Macrophages	Cell proliferation, differentiation and apoptosis; induces cyclooxygenase-2; causes fever and anorexia; promotes acute phase protein synthesis
IL-2	Helper T cells	Proliferation and differentiation of type 1 helper T cells; maturation of $CD8^+$ T cells
IL-4	Helper T cells	Proliferation and differentiation of type 2 helper T cells; promotes IgE production
IL-5	Type 2 helper T cells	Proliferation of B lymphocytes; activation of eosinophils
IL-6	Monocytes/Macrophages	Proliferation of B lymphocytes and IgA and IgG1 production; promotes acute phase protein synthesis Induces acute phase responses
IL-8	Macrophages	Chemoattractant
IL-10	Monocytes/Macrophages; T cells; B cells; natural killer cells; neutrophils; Eosinophils; mast cells	Inhibits production of IFN-γ, IL-1, IL-6, TNF-α; inhibits antigen presentation
IL-12	Monocytes/Macrophages	Promotes type 1 helper T cell responses; induces IFN-γ
GCSF	Monocytes	Promotes maturation of myeloid cells
MCSF	Monocytes	Promotes maturation of monocytes
GMCSF	Activated macrophages; T-cells	Promotes proliferation, differentiation and survival of monomyelocyte cells
IFN-α and IFN-β	Monocytes/Macrophages; virally-infected cells	Inhibits viral replication
IFN-γ	Type 1 helper T cells; natural killer cells	Activation of type 1 helper T cells; activation of macrophages for proliferation and antigen presentation; production of TNF-α and IL-1
TNF-α	Monocytes/Macrophages	Stimulates generalized immune activation as well as tumor necrosis; causes feber and anorexia; promotes acute phase protein synthesis
TGF-β	Platelets	Immunoinhibitory but stimulates IgA production and connective tissue growth and collagen formation

Abbreviations: GCSF, granulocyte colony stimulating factor; GMCSF, granulocyte-macrophage colony stimulating factor; IFN, interferon; Ig, immunoglobulin; IL, interleukin, MCSF, monocyte colony stimulating factor; TGF, transforming growth factor; TNF, tumor necrosis factor.

Some cytokines affect metabolism

Widespread changes occur as a result of pro-inflammatory cytokine production and action (**Fig. 1**). These seem to be designed to create a hostile environment for pathogens, but are also harmful to the host, especially if they are sustained (Soeters and Grimble, 2009). Cytokines orchestrate metabolic responses to ensure that during infection, nutrients are provided for the immune system as a result of wasting of peripheral tissues (Douglas and Shaw, 1989). For example, amino acids released as a consequence of increased proteolysis in muscle, skin, and bone, provide substrate for the synthesis of proteins produced as part of the immune response (e.g., antibodies, acute phase proteins). Glutamine, released from muscle, and glucose derived from increased hepatic gluconeogenesis using amino acids as the substrate, are major sources of energy for the immune system. Likewise increased lipolysis in adipose tissue provides fatty acids as metabolic fuels, substrates for regulatory molecules, and building blocks for cell membranes (Grunfeld and Feingold, 1992). In addition, proinflammatory cytokines cause insulin resistance (Grimble, 2002; Wiser et al., 2013) which results in an elevation of plasma triacylglycerol concentrations through a combination of effects: by enhancing free fatty acid efflux from adipose tissue, raising hepatic triacylglycerol synthesis, and inhibiting entry of triacylglycerol fatty acids into adipose tissue by inhibition of adipose tissue lipoprotein lipase activity. One interesting effect of raised plasma triacylglycerols is that bacterial endotoxins are sequestered onto the lipoproteins carrying the triacylglycerols, so protecting the host from the deleterious consequences of endotoxin and also targeting them for elimination by the liver. Zinc, an important cofactor in RNA and DNA synthesis, is released from peripheral tissues by inflammatory cytokines, incorporated into the zinc transporting protein metallothionein in liver and kidney, and subsequently utilized by the immune system. Fever and loss of appetite (anorexia) often occur as a result of increased pro-inflammatory cytokine production due to a direct action of these cytokines on the hypothalamus. Pro-inflammatory cytokines also modify metabolism in the liver. Large increases in the rates of gluconeogenesis, glycogen breakdown, and urea synthesis occur. Blood glucose, urea, and triacylglycerol concentrations may rise. Paradoxically, however, metabolism of xenobiotics is decreased owing to a reduction in the activity of

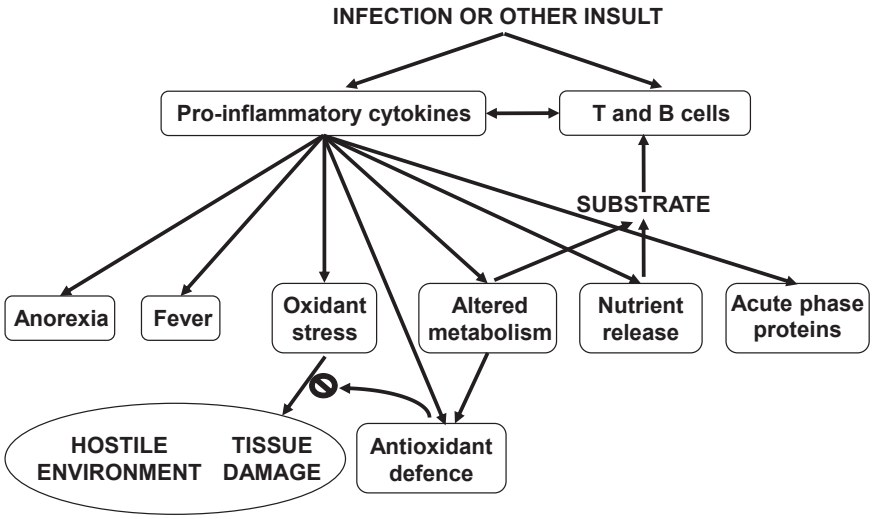


Fig. 1 Central role of pro-inflammatory cytokines in the host response to infection. The dark circle with a diagonal line indicates inhibition.

cytochrome P450 enzymes. The profile of export proteins synthesized by the liver is changed: synthesis of albumin is reduced, and the synthesis of a group of proteins closely associated with inflammation called acute-phase proteins is increased (Heinrich et al., 1990). Acute-phase proteins are multifunctional and include ceruloplasmin (an antioxidant and copper transport protein), C-reactive protein (improves macrophage activity), fibrinogen (for blood clotting), complement proteins (for enhanced phagocytosis and pathogen destruction), and metallothionein (a zinc transport protein). Thus, pro-inflammatory cytokines not only affect the immune system, but also alter metabolism in the liver, adipose tissue, muscle and brain, to control the whole body response to infection or insult. All of the metabolic changes may be purposeful in creating a situation in which substrate is more closely tailored to the requirements of the immune system than would occur from the vagaries of habitual dietary intake.

In addition to the changes in metabolism described above, the antioxidant defenses of the body are strengthened by increases in the activities of superoxide dismutase, catalase, glutathione peroxidase, and glutathione reductase, and by increases in the hepatic synthesis of the reduced form of glutathione (GSH) (Grimble, 1994). The liver thus becomes the main focus for the synthesis of molecules for the nutrition, support and direction of the immune system, and for the protection of the body from the adverse effects of cytokine action. Indeed when the ability of the liver of patients with sepsis (a severe clinical form of inflammation induced by infection), to extract amino acids from the circulation was assessed, it was found that patients who subsequently died had only half of the ability compared with patients who survived.

A number of molecules synthesized in enhanced amounts when cytokines are produced are part of complex feedback systems which act to limit cytokine production and effects (Fig. 2). These include corticosteroids, GSH and some acute-phase proteins which

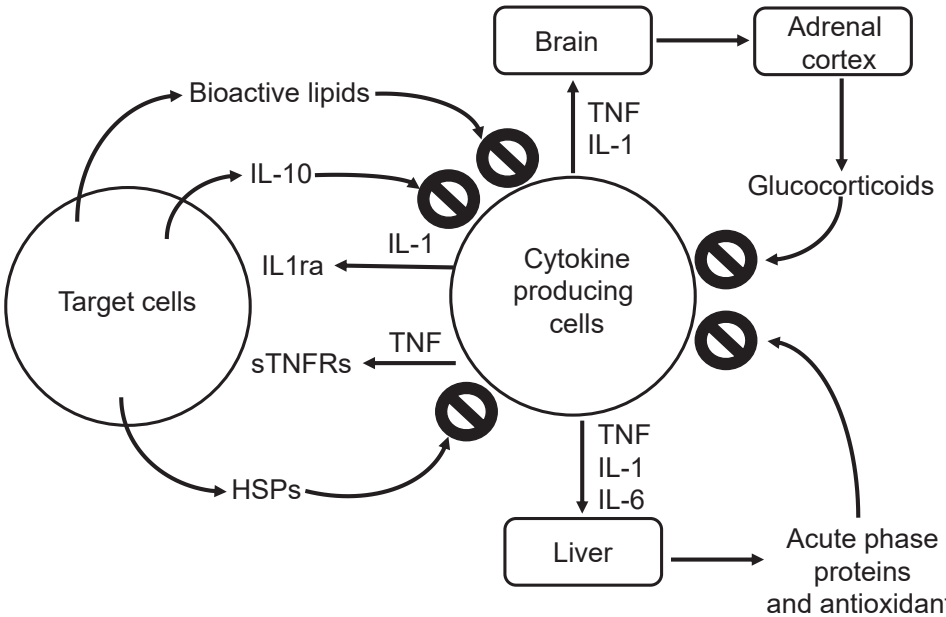


Fig. 2 Some of the feedback control systems for pro-inflammatory cytokine production and action. HSPs, heat shock proteins; IL, interleukin; Il-1ra, IL-1 receptor antagonist; sTNFRs, soluble TNF receptors; TNF, tumor necrosis factor; the dark circle with a diagonal line indicates inhibition.

suppress cytokine production, and cytokine receptor antagonist molecules for IL-1 and TNF which inhibit their action. Other molecules also moderate cytokine actions. *Anti*-inflammatory cytokines such as IL-10 and some heat shock proteins exert an *anti*-inflammatory influence in the latter stages of the inflammatory response. This down-regulation of inflammation, once the infectious agent has been defeated, is important for survival since the inflammatory process depletes the body of resource. The balance between the pro-inflammatory and *anti*-inflammatory process is of key importance for return to homeostasis and for survival.

Role of cytokines in disease

Despite the importance of cytokines in protecting the host from pathogens, these cytokines may have damaging and even lethal effects upon the host (Grimble, 1996; Calder et al., 2009; Soeters and Grimble, 2009). Thus, the response of the host to a pathogen may play as significant a part in the demise of the host as the effects of the pathogen itself. This damaging aspect has sometimes been called the "cytokine storm". This is why the feedback processes described in **Fig. 2** are so important as an infection runs its course. Cytokines can also play a major part in the tissue damage that is seen in chronic inflammatory disease in which no infective agent is involved (e.g., rheumatoid arthritis). Excessive or inappropriate cytokine production has been associated with increased morbidity and mortality in a wide range of diseases and conditions in which inflammation plays a part. These include diseases where the immune system is clearly interacting with invading pathogens, such as in malaria, meningitis, sepsis, and acquired immunodeficiency syndrome and conditions such as asthma, inflammatory bowel disease, rheumatoid arthritis, and cancer where inflammatory disease develops without obvious involvement of pathogens (**Fig. 3**). Furthermore proinflammatory cytokines are involved in the progression of disease processes such as plaque development in atherosclerosis and demyelination in multiple sclerosis and Alzheimer's disease. They are now even known to be involved in changes in adipose tissue as obesity progresses (Tilg and Moshen, 2006).

Damage may also be caused to the host by release of free radicals and other oxidant molecules from phagocytic cells in response to an inflammatory stimulus including IL-1 and TNF. Furthermore, oxidant molecules up regulate production of IL-1, TNF and IL-8, among other inflammatory factors, by activation of the transcription factor, nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) (**Fig. 4**). The factor is normally held quiescent in the cytoplasm owing to attachment to an inhibitory component (IκB). In the presence of either inflammatory stimuli or oxidants, IκB is phosphorylated and dissociates from the rest of NFκB; the now free NFκB migrates to the nucleus and induces the transcription of a large range of genes encoding proteins associated with the inflammatory process (**Fig. 4**). These include genes encoding inflammatory cytokines, acute phase proteins, adhesion molecules, and enzymes including cyclooxygenase 2, inducible nitric oxide synthase and some of the matrix metalloproteinases. Hence antioxidants, which reduce the production of or quench oxidants, can have an *anti*-inflammatory action (**Fig. 4**).

The fact that insulin resistance and disordering of lipid metabolism occur in obesity, diabetes mellitus, and during the inflammatory response has led to the investigation of the possibility that inflammation and obesity are somehow interlinked. As adipose tissue expands it becomes infiltrated by monocytes which differentiate into macrophages. It is now known that other immune cells

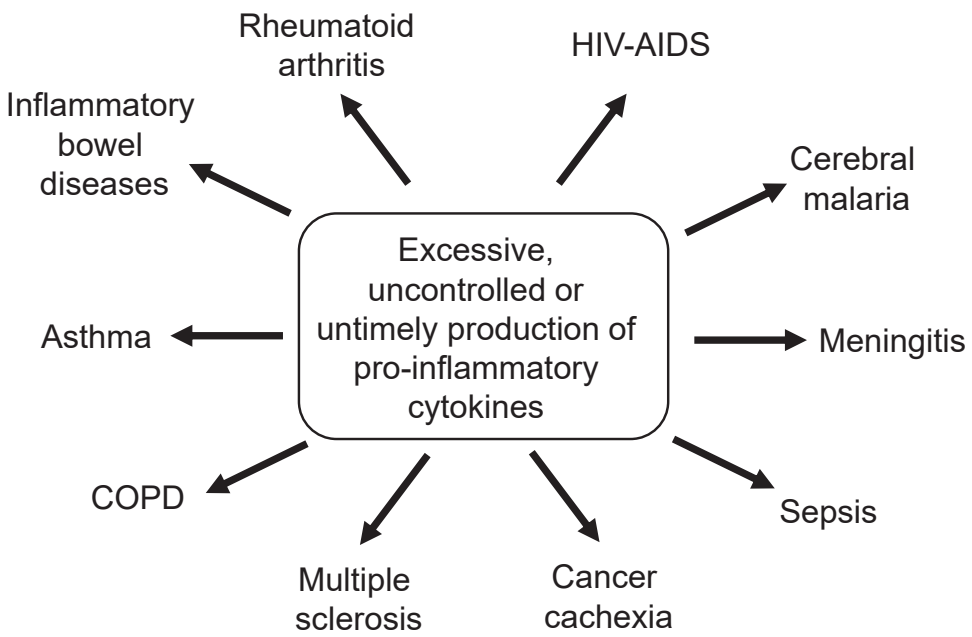


Fig. 3 The central role of excessive, uncontrolled or untimely pro-inflammatory cytokine production in human disease. COPD, chronic obstructive pulmonary disease; HIV-AIDS, human immunodeficiency virus acquired immunodeficiency syndrome.

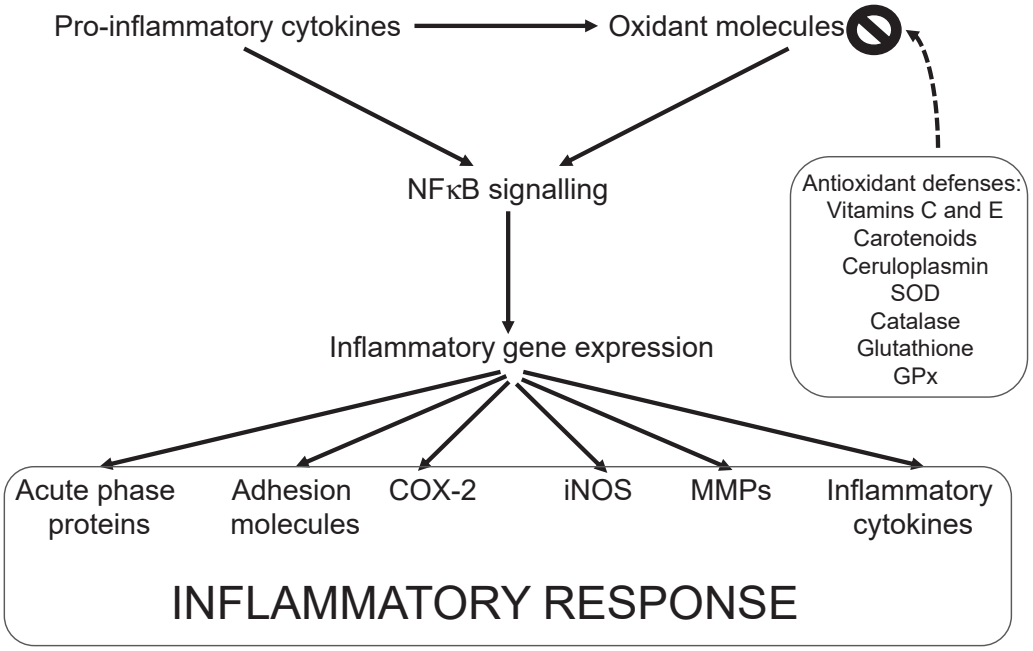


Fig. 4 Interplay between pro-inflammatory cytokines and oxidants in initiating NFκB signaling leading to upregulation of multiple components of the inflammatory response. COX, cyclooxygenase; iNOS, inducible nitric oxide synthase; MMPs, matrix metalloproteinases; NFκB, nuclear factor kappa-light-chain-enhancer of activated B cells; the dark circle with a diagonal line indicates inhibition.

including T and B lymphocytes also infiltrate adipose tissue in obesity. The immune cells interact with adipocytes and an inflammatory environment is created (Hotamisligl et al., 1993; Tilg and Moschen, 2006; Sethi and Hotamisligl, 2021). This most likely contributes to the insulin resistance that occurs in obesity. For example, TNF-α results in insulin insensitivity, indirectly by stimulating stress hormone production and directly by sustained induction of SOCS-3, which has been shown to decrease insulin-induced insulin receptor substrate 1 tyrosine phosphorylation and its association with the p85, the regulatory subunit of phosphatidylinositol-3 kinase, and by negative regulation of PPAR gamma, an important insulin-sensitizing nuclear receptor. Adipose tissue produces and releases inflammatory cytokines including TNF-α and IL-6 and the adipokines leptin and adiponectin; the latter two have opposing actions to one another. Both immune/inflammatory cells and adipocytes produce TNF-α and IL-6 while it is adipocytes that produce leptin and adiponectin. In humans, plasma leptin concentrations are linearly related to body mass index and body fat mass. Thus, compared with people of healthy weight, those living with obesity have higher plasma concentrations of TNF-α, IL-6, leptin, and C-reactive protein and lower concentrations of adiponectin (Calder et al., 2011). This increase in systemic inflammation in obesity may be partly responsible for the co-morbidities that develop in obesity (**Fig. 5**).

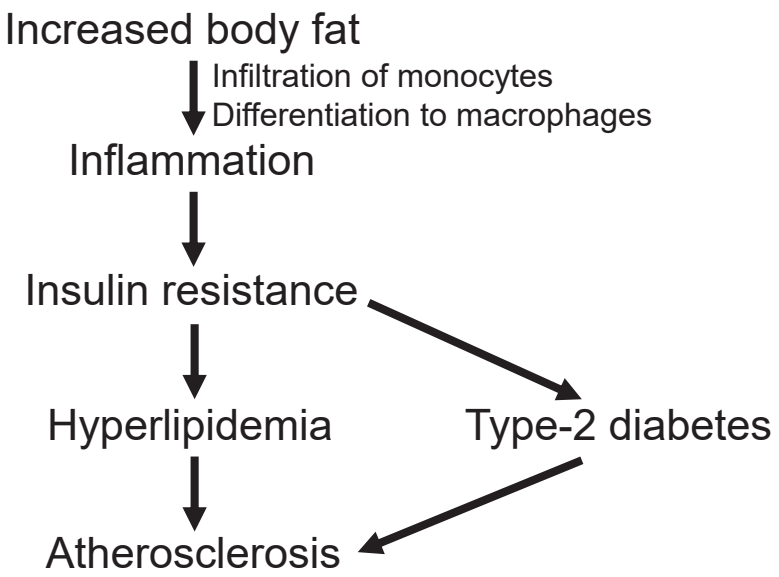


Fig. 5 Inflammation creates a link between increased body fat (obesity) and the cardiometabolic co-morbidities of obesity.

It is also important to note that as people age they can develop a state of chronic low-grade inflammation, termed inflammaging, which is characterized by elevated levels of pro-inflammatory cytokines, such as IL-6, and of C-reactive protein in the plasma (Calder et al., 2017). Inflammaging may be a link to age-related changes in muscle mass and increases in non-communicable diseases (Fig. 6).

The interaction of nutrition and cytokines

Proinflammatory cytokines exert widespread effects on metabolism, involving alterations in lipid, carbohydrate, and protein metabolism (see earlier). In addition there are substantial changes in micronutrient metabolism (see earlier). There is a bidirectional interaction with oxidant stress, that is mitigated by antioxidants (see earlier). A number of intracellular signaling pathways are activated by the actions of cytokines on target cells; these include the cyclic AMP and protein kinase C pathways and pathways leading to production of prostaglandins (PGs) and leukotrienes (LTs) from polyunsaturated fatty acids (PUFAs). Thus, there are many levels at which nutrients and nutrient intake can modify the intensity and characteristics of the cytokine response to inflammatory stimuli (Grimble, 1996). The ability of nutrients to modify inflammation has been used in the treatment of diseases with an inflammatory basis, such as rheumatoid arthritis. The interaction between nutritional status and inflammation is also important in public health by determining the effects of infection on growth and well-being of populations with a poor nutrient intake.

The earliest indications that nutritional status could affect cytokine biology came from studies on malnourished hospital patients, where white blood cells had a reduced capacity to produce cytokines. The high mortality rates in these patients suggested the importance of cytokines in the process of recovery from injury and infection. Protein supplements improved cytokine production and decreased the mortality rate. Because these observations were made, a large number of studies have been conducted in

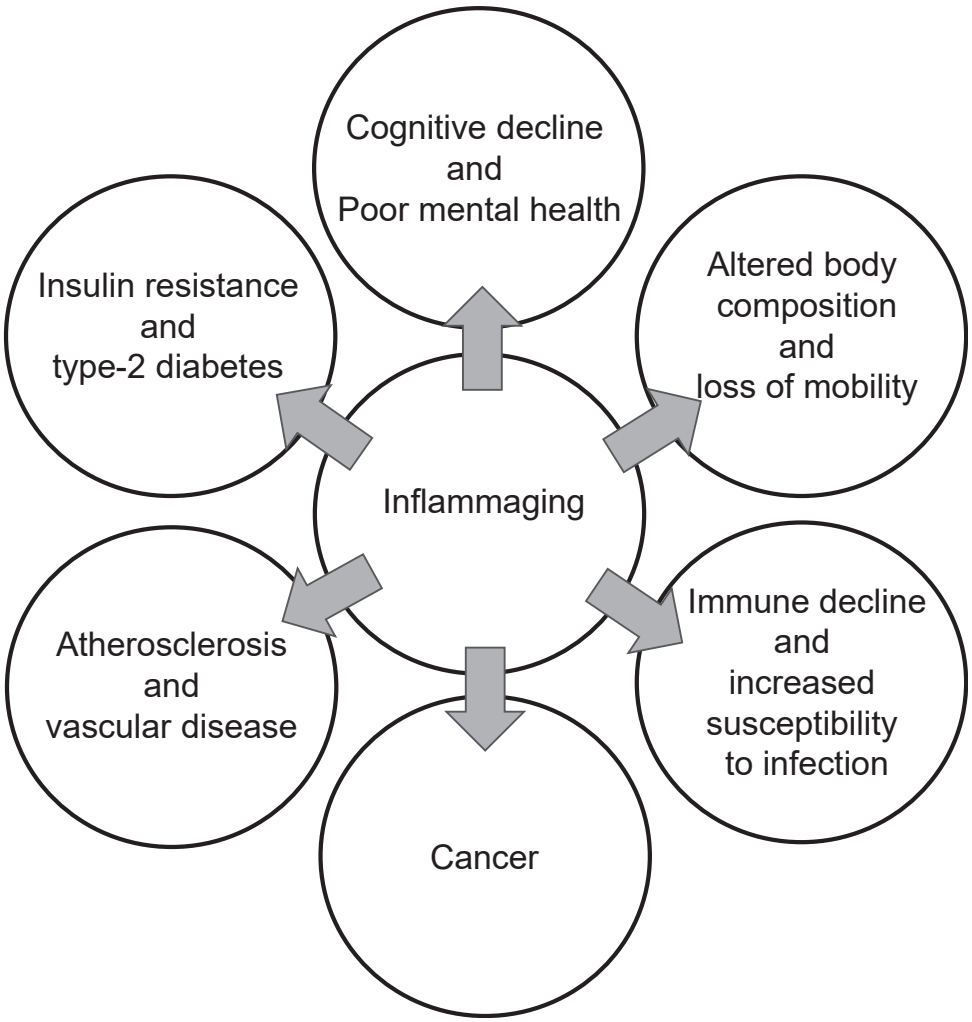


Fig. 6 The central role of inflammaging in chronic diseases of aging.

animals and humans which show that fatty acids, amino acids, and micronutrients change the ability of mammals to produce and respond to a number of cytokines including IL-1, IL-6 and TNF.

Influence of fatty acids on inflammatory cytokine production

The production and actions of proinflammatory cytokines are influenced by dietary fatty acids acting through several different, but perhaps interlinked, mechanisms. Central to these mechanisms is the incorporation of the fatty acids into the cell membrane bilayer (Fig. 7). Subsequently, membrane fluidity may be changed, the formation of signaling platforms called lipid rafts in the membrane may be affected, and the types and amounts of lipid mediators including PGs and LTs being produced may be altered (Calder, 2020a). In addition, there seem to be direct effects by which different fatty acids influence inflammatory cytokine production. For example, several fatty acids, but particularly the n-3 PUFA docosahexaenoic acid (DHA), can bind directly to a cell surface G-protein coupled receptor (GPR) called GPR120 on macrophages which initiates signaling that dampens activation of NFκB, so reducing production of TNF-α, IL-1 and IL-6. Furthermore, there is some evidence that certain saturated fatty acids, including lauric, myristic and palmitic acids, bind to the endotoxin receptor (toll-like receptor 4) to directly promote inflammatory signaling activating NFκB, although this action is disputed by some. With regard to lipid rafts, DHA seems to be the most potent raft disruptor in inflammatory cells and this has been linked to less activation of NFκB. The n-3 PUFAs eicosapentaenoic acid (EPA) and DHA act as activators of PPAR-γ which physically interferes with NFκB translocation to the nucleus. Normally, the main substrate for synthesis of PGs and LTs is the n-6 PUFA arachidonic acid (Calder, 2020b). In general, these lipid mediators have pro-inflammatory effects and they are linked with many inflammatory conditions. However, different mediators produced from arachidonic acid can have different effects and sometimes these can be *anti*-inflammatory. For example, PGE_2, which has many pro-inflammatory actions, actually inhibits the production of TNF-α and IL-1. This is, most likely, an example of the self-regulating nature of the inflammatory response referred to earlier. EPA and DHA decrease production of PGs and LTs from arachidonic acid and give rise to separate families of lipid mediators. EPA is a substrate to PGs and LTs with a different structure to those performed from arachidonic acid and in general these are weakly bioactive. EPA also gives rise to E-series resolvins while DHA gives rise to D-series resolvins, protectins and maresins. The resolvins, protectins and maresins are together termed specialized

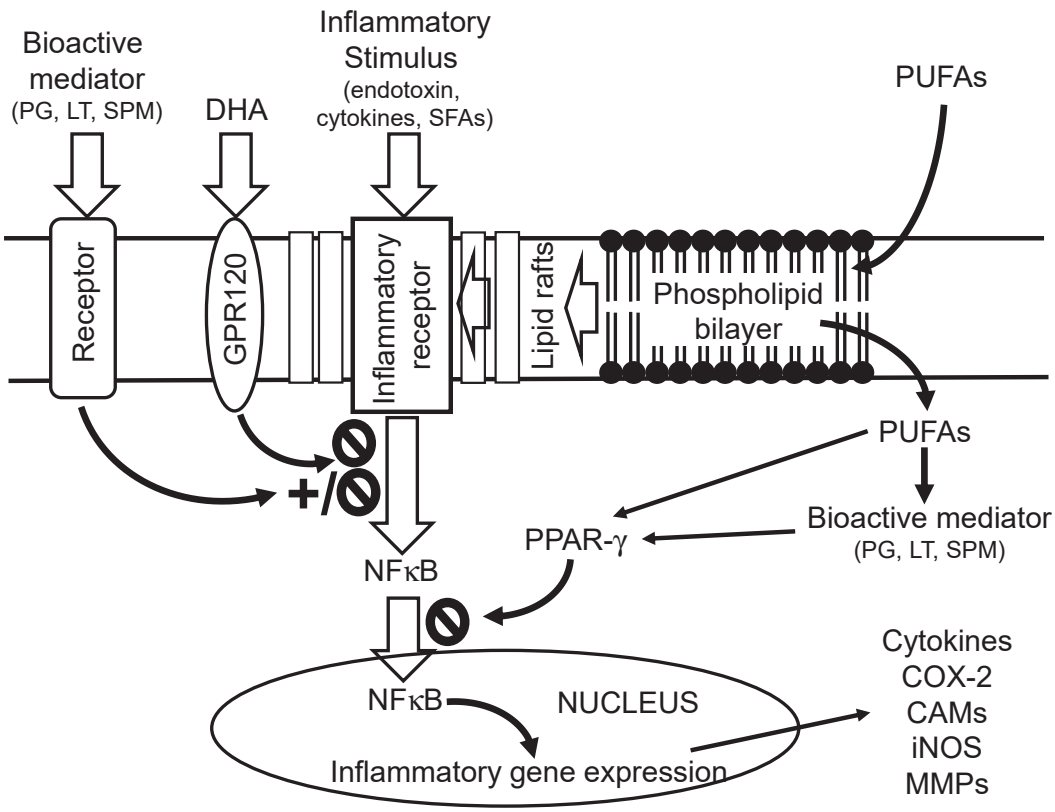


Fig. 7 Summary of the mechanisms by which fatty acids interact with inflammatory processes, including pro-inflammatory cytokine signaling and production. CAMs, cell adhesion molecules; COX, cyclooxygenase; DHA, docosahexaenoic acid; GPR, G protein coupled receptor; iNOS, inducible nitric oxide synthase; LT, leukotriene; MMPs, matrix metalloproteinases; NFκB, nuclear factor kappa-light-chain-enhancer of activated B cells; PG, prostaglandin; PPAR, peroxisome proliferator activated receptor; SFAs, saturated fatty acids; SPM, specialized pro-resolving mediator; the dark circle with a diagonal line indicates inhibition; + indicates activation.

pro-resolving mediators (SPMs) (Chiang and Serhan, 2020). All these bioactive lipid mediators act via cell surface receptors. The SPMs possess *anti*-inflammatory and inflammation resolving capabilities including the ability to decrease the activation of NFκB and the production of TNF-α, IL-1 and IL-6. Thus, in their own right and through their derived mediators EPA and DHA have *anti*-inflammatory and inflammation resolving actions, including the ability to decrease production of the classic pro-inflammatory cytokines. Linked to this, supplements of EPA and DHA have been shown to be of benefit in patients with inflammatory diseases especially arthritis. These effects of n-3 PUFAs might also relate to their ability to reduce risk of chronic non-communicable diseases of aging, especially cardiovascular disease.

Interaction of inflammatory cytokines with amino acid and protein metabolism

Substantial increases in protein synthesis occur as the result of infection. It has been estimated that approximately 45 g of protein is required each day to produce and maintain the increased quantities of white blood cells, acute phase proteins and antibodies in an infected individual. This demand will have a considerable impact on the availability of amino acids for other processes in the body that involve protein synthesis. In children this has an inhibitory effect on growth, which is worsened by repeated bouts of infection. Understandably infection during pregnancy or lactation can have an adverse impact on offspring growth and development. Output of amino acids from skeletal muscle, skin, and bone provides substrate for the synthesis of cells and proteins associated with the response to infection and other inflammatory insults, as indicated earlier. However, the supply may not always match demand, as is evident from the decrease in plasma concentrations of a number of amino acids. In particular, reductions occur in the concentrations of a metabolically related group of amino acids, including glycine, serine, and taurine. All three are metabolically related with the sulfur amino acids. Glycine and serine, together with the sulfur amino acids, are found in high concentrations in many compounds associated with the immune and inflammatory response, most notably comprising 66% of glutathione, 56% of metallothionein, and up to 25% of many acute-phase proteins. Experimental studies have shown that the production of cytokines, acute phase proteins, and glutathione is influenced by the adequacy of both protein and sulfur amino acid intake. The partitioning of cysteine into glutathione and proteins in the liver may change if dietary sulfur amino acid intake becomes inadequate. This phenomenon is due to the biochemical properties of rate limiting enzymes in the two pathways. Although the Km for γ-glutamyl cysteine synthetase (rate limiting for GSH synthesis) is 0.35 mM, that for amino acid activating enzymes (rate limiting for protein synthesis) is only 0.003 mM. This biochemical characteristic means that GSH synthesis will fall below maximal rates at much higher intracellular cysteine concentrations than protein synthesis. Thus, at low sulfur amino acid intakes, antioxidant defenses will become compromised, permitting oxidative and inflammatory stress. Low concentrations of GSH in tissues may have implications for the extent of inflammatory processes in the individual. In animal studies, decreased lung GSH concentrations are associated with accumulation of inflammatory cells in tissues. In studies on patients with human immunodeficiency virus infection given *N*-acetyl cysteine to improve GSH status, a decrease in plasma IL-6 concentrations was noted indicating a reduction in inflammation.

Modulation of inflammatory cytokine production by micronutrients

Micronutrients play widespread and complex roles in the response to infection and trauma. They are incorporated into structures that are synthesized in increased amounts during the inflammatory response and into components of antioxidant defense, and also modulate immune function. Trace elements are present in several acute-phase proteins and enzymes associated with antioxidant defense. These proteins include metallothionein (Zn), ceruloplasmin (Cu), superoxide dismutases (Mn, Cu, Zn), and glutathione peroxidase (Se). Deficiencies in copper impair the ability to increase superoxide dismutase and ceruloplasmin activities in response to inflammatory stimuli. Deficiencies in zinc impair the ability to increase metallothionein synthesis; furthermore, zinc deficiency has potent suppressive effects on many aspects of the immune response including lymphocyte proliferation. Iron status may influence inflammation and immune function in a number of ways. Normally iron is tightly bound to transport proteins such as transferrin and ferritin. However, following tissue damage and infections, such as malaria which may destroy red blood cells, free iron may be released and exerts a proinflammatory effect by catalyzing free radical production. The latter effect may activate NFκB and up-regulate cytokine production. Indeed, iron dextran infusion has been shown to exacerbate inflammatory symptoms in rheumatoid arthritis, while desferrioximine, an iron chelator, suppresses TNF and IL-1 production by macrophages. Iron deficiency also decreases the ability of such cells to produce cytokines. Impairment of immunological defense is commonly found in iron-deficient animals and human populations including defects in T cell proliferation and in the ability of macrophages to engulf and kill bacteria. The latter defect may relate to the role of iron as part of the NADPH oxidase complex that is responsible for the respiratory burst and generation of hydroxyl radicals that kill bacteria. Myeloperoxidase, which generates hypochlorous acid for bacterial killing is a heme protein which is decreased in activity by iron deficiency.

Vitamins also exert a number of effects upon cytokine biology. These effects may relate to the roles which some of these nutrients play as antioxidants and growth factors. Rats deficient in vitamin E exhibit an enhanced inflammatory response to endotoxin; addition of the vitamin to the diet will suppress this effect. In healthy subjects and smokers, a daily dose of 600 IU of vitamin E for 4 weeks reduced the ability of white blood cells to produce TNF and IL-1. Cigarette smoking enhances cytokine production and raises acute-phase protein concentrations. The extent of the elevation is inversely related to vitamin E status. Strenuous exercise results in a small increase in plasma concentrations of IL-1 and IL-6; vitamin E supplementation can prevent this effect. Vitamin A also influences cytokine production, although the mechanism underlying the effect is unclear. Macrophages taken from Indian children who had received a supplement of 100,000 IU of retinol produced seven times the quantity of IL-1 produced by cells

from children who had not received supplementation. The effect may be more pharmacological than nutritional in nature. Mice given vitamin A, at a dose that was 16 times their requirement, had macrophages which produced twice as much IL-1 upon stimulation than cells from unsupplemented animals. Cells of the immune system express the vitamin D receptor and some cells involved in inflammation, including macrophages, can produce the active form of vitamin D (1,25-dihydroxyvitan D_3). 1,25-dihydroxyvitamin D_3 has been shown to have *anti*-inflammatory effects, for example promoting a shift of macrophages and T lymphocytes from a pro-inflammatory to an *anti*-inflammatory phenotype, with decreased production of pro-inflammatory cytokines and increased production of IL-10. Vitamin C, being an *anti*-oxidant, may also decrease pro-inflammatory cytokine production. In accordance with this, incubation of white blood cells from people with severe community-acquired pneumonia with vitamin C decreased oxidative stress and inflammatory cytokine production.

Conclusions and perspectives

The objective of the response of the body to infection is to disadvantage and destroy invading organisms, but at the same time to protect healthy tissues from the damaging influence of agents produced during the response. Cytokines play a central role, being part of the host defense system and bringing about changes in host metabolism aimed at supporting the immune response. The inflammatory cytokines affect the immune response as well as metabolism in tissues including the liver, muscle, adipose tissue and brain. Some of the symptoms of infection including fever and anorexia are due to pro-inflammatory cytokines. Through their metabolic effects they can impair growth, which can become serious in children if the inflammation is prolonged or repeated too frequently. There is a close interrelationship between proinflammatory cytokines and pro-oxidant molecules, with inflammation inducing oxidative stress and oxidative stress inducing inflammation. Ultimately these processes need to be controlled, balanced and limited. The complex antioxidant system, supported through nutrition (e.g., trace elements, sulfur amino acids), and the potent pro-resolving mediator system, again supported through nutrition (n-3 PUFAs), are both central to the control and self limitation of inflammation.

The essence of survival of an individual or species lies in the ability to prioritize physiological processes, particularly those processes which exert a large metabolic demand. Thus, at various times throughout the life cycle, mammals will focus metabolic processes upon achieving growth, the construction of placenta and fetus, the synthesis of milk components, or the repulsion of invasion by pathogens. For the infected individual, the marshaling of resources to combat the invading pathogen must assume a priority over all other physiological events. These other physiological processes can resume once the invasion has been repulsed and the damage done by the invader repaired. Hence excessive prolongation of the inflammatory response is not desirable.

The production of cytokines and other molecules associated with the inflammatory process carries risks of damage to the host as well as a survival advantage. The risk to the host is minimized by a sophisticated range of feedback control systems and synthesis of substances which protect the host. As discussed above, nutrient intake modulates cytokine biology and the control and protective systems. A wide range of nutrients modulate cytokine biology at the level of production and sensitivity of target tissues. As a consequence of the modulation, the extent of depletion of nutrient stores and the risk of damage during the inflammatory response will be changed. The extent of tissue depletion and risk to the host will thus range from mild and transient in nature, to severe, chronic, or lethal in effect.

References

Beutler, B., Cerami, A., 1986. Tumor necrosis factor as two sides of the same biological coin. Nature 320, 584–588.
Calder, P.C., 2020a. n-3 PUFA and inflammation: from membrane to nucleus and from bench to bedside. Proc. Nutr. Soc. 79, 404–416.
Calder, P.C., 2020b. Eicosanoids. Essays Biochem. 64, 423–441.
Calder, P.C., Albers, R., Antoine, J.M., Blum, S., Bourdet-Sicard, R., Ferns, G.A., Folkerts, G., Friedmann, P.S., Frost, G.S., Guarner, F., Løvik, M., Macfarlane, S., Meyer, P.D., M'Rabet, L., Serafini, M., van Eden, W., van Loo, J., Vas Dias, W., Vidry, S., Winklhofer-Roob, B.M., Zhao, J., 2009. Inflammatory disease processes and interactions with nutrition. Br. J. Nutr. 101, S1–S45.
Calder, P.C., Ahluwalia, N., Brouns, F., Buetler, T., Clement, K., Cunningham, K., Esposito, K., Jonsson, L.S., Kolb, H., Lansink, M., Marcos, A., Margioris, A., Matusheski, N., Nordmann, H., O'Brien, J., Pugliese, G., Rizkalla, S., Schalkwijk, C., Tuomilehto, J., Warnberg, J., Watzl, B., Winklhofer-Roob, B.M., 2011. Dietary factors and low-grade inflammation in relation to overweight and obesity. Br. J. Nutr. 106 (Suppl. 3), S5–S78.
Calder, P.C., Bosco, N., Bourdet-Sicard, R., Capuron, L., Delzenne, N., Doré, J., Franceschi, C., Lehtinen, M.J., Recker, T., Salvioli, S., Visioli, F., 2017. Health relevance of the modification of low grade inflammation in ageing (inflammageing) and the role of nutrition. Ageing Res. Rev. 40, 95–119.
Chiang, N., Serhan, S.N., 2020. Specialized pro-resolving mediator network: an update on production and actions. Essays Biochem. 64, 443–462.
Dinarello, C.A., 1988. Biology of IL1. FASEB J. 1, 108–115.
Douglas, R.G., Shaw, J.H.F., 1989. Metabolic response to sepsis and trauma. Br. J. Trauma 76, 115–122.
Grimble, R.F., 1994. Nutritional antioxidants and the modulation of inflammation: the theory and the practice. Crit. Care Med. 2, 175–185.
Grimble, R.F., 1996. The interaction between nutrients, pro-inflammatory cytokines and inflammation. Clin. Sci. 91, 121–130.
Grimble, R.F., 2002. Inflammatory status and insulin resistance. Curr. Opin. Clin. Nutr. Metab. Care 5, 551–559.
Grunfeld, C., Feingold, K.R., 1992. Tumour necrosis factor, interleukin 1 and interferon induce changes in lipid metabolism as part of host defence. Exp. Biol. Med. 200, 214–227.
Heinrich, P.C., Castell, J.V., Andus, T., 1990. Interleukin 6 and the acute phase response. Biochem. J. 265, 621–636.
Hotamisligil, G.S., Shargill, N.S., Spiegelman, B.M., 1993. Adipose expression of tumor necrosis factor-alpha: direct role in obesity-linked insulin resistance. Science 259, 87–91.
Sethi, J.K., Hotamisligil, G.S., 2021. Metabolic messengers: tumour necrosis factor. Nat. Metab. 3, 1302–1312.
Soeters, P.B., Grimble, R.F., 2009. Dangers and benefits of the cytokine mediated response to injury and infection. Clin. Nutr. 28, 583–596.
Tilg, H., Moschen, A.R., 2006. Adipocytokines: mediators linking adipose tissue, inflammation and immunity. Nat. Rev. Immunol. 6, 772–783.
Wiser, V., Moschen, A.R., Tilg, H., 2013. Inflammation, cytokines and insulin resistance: a clinical perspective. Arch. Immunol. Ther. Exp. 61, 119–125.

Diet and eating behavior: Appetite control and satiety

M. Dalton[a], **N.J. Buckland**[b], **and J.E. Blundell**[c], [a] School of Psychology, Leeds Trinity University, Leeds, United Kingdom; [b] Department of Psychology, University of Sheffield, Sheffield, United Kingdom; and [c] School of Psychology, University of Leeds, Leeds, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

Introductory paragraph 57
Appetite and satiety (and satiation) 58
Appetite as a complex system 58
The diet – behavior relationship: humans are omnivores 58
The drive to eat and food selection 60
Foods and the satiety cascade 60
Hedonic processes of appetite control 60
Food hedonics influence satiation and satiety 62
Individual variability in appetite control and satiety responsiveness 62
Applications of satiety to societal challenges 63
The impact of ultra-processed foods on satiety 64
Conclusion 65
References 65
Further reading 66

Key points

- The expression of human appetite emerges from a complex bio-psychological system
- The omnivorous nature of humans makes food choice a key feature of appetite control
- Satiety (short term inhibition of eating and hunger) arises from the physiological and psychological effects of food consumption.
- The Drive to Eat (hunger) is under separate control from the choice of what to eat.
- The hedonic aspect of food choice is composed of the processes of liking and wanting for foods which together modulate the processes of satiety.
- There is huge individual variability (physiological and psychological) in the satiety response to different foods which allows the identification of satiety phenotypes (people who vary in the strength of the satiety response).
- The amount of physical activity (and energy expended) a person does, or the degree of sedentariness, can influence the strength of satiety and the amount of food consumed.
- The joint action of the drive to eat (arising from physiological energy requirements) together with satiety adjust the amount of food eaten, and therefore strongly impact energy balance, weight control and obesity.
- The modern technological environment, including issues of processing of foods, food security and sustainability, has a critical impact on the expression of appetite and the strength of satiety.

Introductory paragraph

There has never been a more important period in which to examine appetite and satiety. The world is currently experiencing a pandemic of the respiratory virus COVID-19, adding to the more longstanding so-called pandemic of obesity. Both pandemics can be treated by modifying behavior and the environment in which it occurs (in addition to medical therapies). For obesity, it is relevant that humans are omnivores with the capability to eat a huge range of materials irrespective of their composition and nutritional value. Obesity is intimately linked to the diet available to people, and to the patterns of eating behavior associated with that diet. Given the negative health consequences of obesity, it is essential that this article addresses the implications of appetite and satiety for the control of body weight.

It is important to recognize that behavior and diet occupy separate domains in the psychobiology of human functioning, but they are inextricably linked. The behavior of eating is the agency through which nutrients enter the body and exert their effects on physiology and metabolism. This means that any factors that influence the behavior of eating have the potential to influence the impact of diet on health. As Rozin et al. (1998) has pointed out "Because behavior is so central to nutrition, the behavioral sciences play an especially important role in the understanding of what we eat and why we eat it. The study of what is in food is extremely important, but all of this knowledge amounts to little if we cannot understand how people eat what is good for

them and avoid what will do them harm". We need knowledge of what is in food and how it affects the body, but we also need an understanding of what makes people eat some foods and not others, and what makes them start and stop eating. This is the essence of the study of appetite.

Diets influence the amount and type of food consumed through behavior which is controlled in a large part by the processes of appetite and satiety. The importance of this is highlighted by the fact that behavior (eating) is the only route through which nutrients enter the body to determine growth and development, and to maintain health. These behavioral processes are not simple linear cause-effect associations; they are complex interactions between physiology and the environment. A starting point for our work will be the recognition of appetite as a representation of a complex system (see **Fig. 1**) in which the choice of which food to consume is influenced by multiple factors. This article will not attempt to review or summarize everything that has been published on this topic. Our job as authors is to provide a conceptual framework and a structure to allow an understanding of appetite and satiety and to explain principles in the control of food intake. In this role we concur with the expressed view that "the purpose of all explanation must be, ultimately, the illumination of the chaotic world with which we are actually surrounded" (Midgley, 1978).

Appetite and satiety (and satiation)

Appetite is not the opposite of satiety; rather satiety is an important component of appetite control. Appetite refers to the sum of processes that influence food consumption. These include: the drive to eat, food selection, food hedonics, psychological traits, factors that influence the initiation of eating, its duration and termination, and the frequency of eating episodes. Appetite, therefore, forms a system rather than a single entity. Satiety is one important component of the appetite system and can be defined as the inhibition of hunger and eating following a meal. Whereas satiation is the term that refers to the termination of a meal and determines meal size. The post-prandial (after-meal) period is characterized by a sequence of physiological events in the gastro-intestinal tract including gastric distention and the progressive release of enzymes and hormones that control the digestion of nutrients and their management during absorption and eventual assimilation into the body. These physiological events reflect the nature of the foods in the diet consumed, and they are often referred to as satiety signals since they accompany the state of satiety. There is no agreement on whether these signals are biomarkers of satiety or are the actual causes of satiety. However, it can be concluded that there is no single unique satiety signal. The inhibition of eating that is the essence of satiety is therefore associated with a complex pattern of gastrointestinal physiological activity.

Because satiety is an inhibitory process it has a key role in controlling the level of hunger and the amount of food consumed; and therefore, has the capacity to prevent (or to permit) overconsumption that potentially influences body weight. For this reason, weak satiety is regarded as a major factor that can lead to the development of obesity. Conversely, specific (functional) foods or drugs that intensify satiety can inhibit eating and potentially lead to weight loss.

Appetite as a complex system

As noted above, appetite refers to the sum of processes that influence food consumption. Since humans are omnivores there is a huge variety of human eating patterns on the planet, which draws attention to the degree of complexity that arises from this heterogeneity. This complexity was originally recognized in a conceptual model that viewed obesity as a complex problem arising from many interlinked proximal and causal factors. Instead of obesity being regarded as the outcome of a single dominant causal agency (or several such single agencies) acting in a linear fashion, obesity was conceived as the outcome of multiple interrelated factors in continuous interactions (see **Fig. 1**). These factors could be organized into groups that were each composed of multiple elements including factors that described the psychological world of people, the food environment at the level of the consumer that could be traced to food manufacturing and procurement, the physical environment, the biological system, and the social and cultural milieu.

While this conceptual map was developed for understanding the factors that contribute to obesity, it equally describes the multiple factors that influence human appetite. The map is composed of more than 300 variables and over 100 feedback loops across 7 domains (see **Fig. 1**). At the center of the map is the core or engine upon which the surrounding interplay of variables has been focused. The core ultimately brings about adjustments in patterns of behavior and consequent changes in body composition, in particular fat mass, which in turn are crucial for appetite control. Moreover, the map makes clear that every domain contains elements that directly or indirectly influence eating behavior. It is therefore appropriate that this concept can be extended from the "Obesity Systems Map" to an "Appetite Systems Map" (**Fig. 1**).

The diet – behavior relationship: humans are omnivores

The fact that humans are omnivores is of huge significance for both behavior and nutrition. Humans are not restricted in their food habits to the same extent as herbivores or carnivores, and consequently they can consume a wide range of nutritional materials. Of course, this ability has been of enormous evolutionary significance and has enabled humans to colonize a wide variety of environments and habitats. Just as different groups of humans can exist on widely divergent types of foods in different parts of the world,

Foresight

Obesity System Map

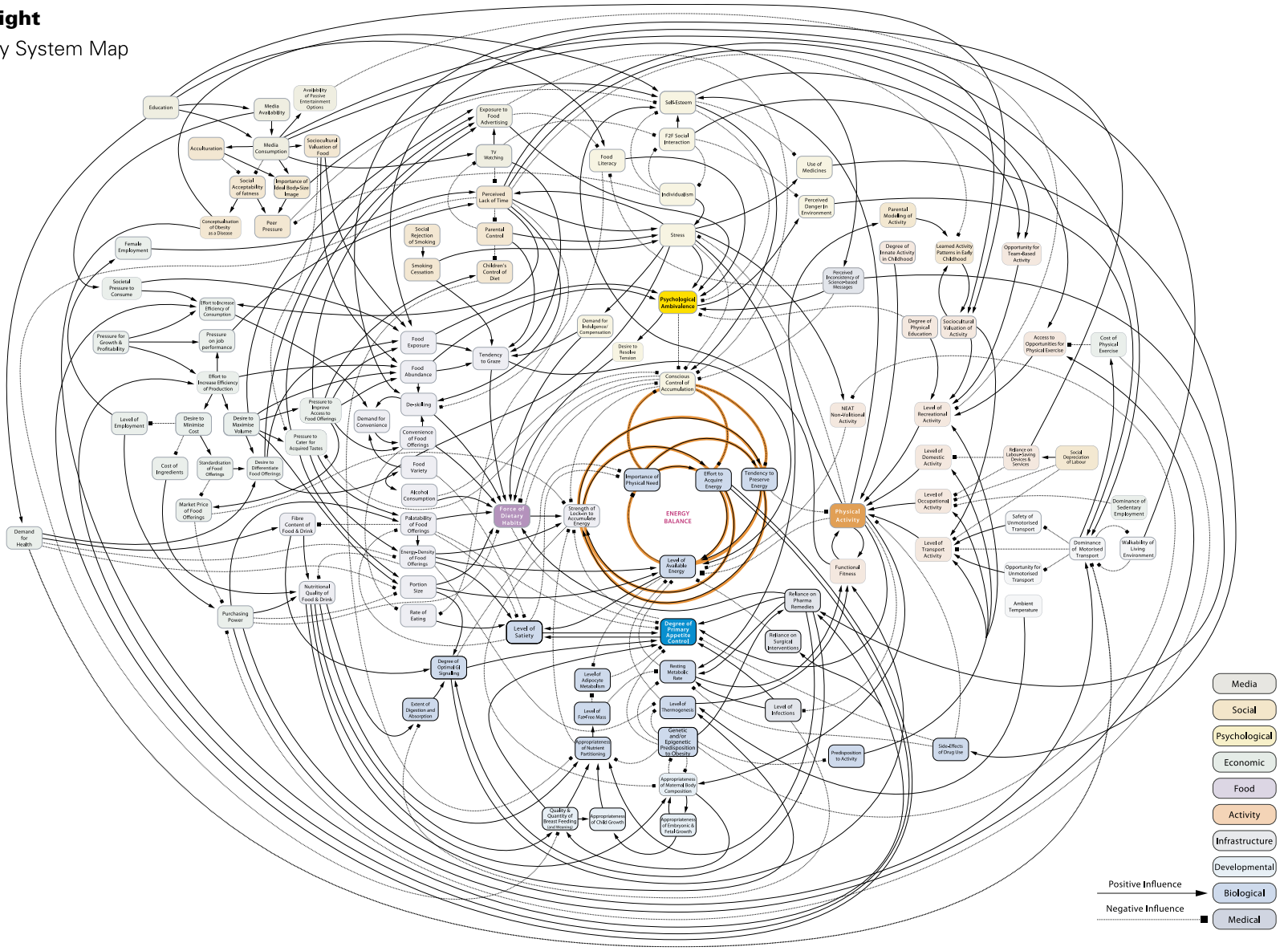


Fig. 1 The Obesity Systems Map identifies many interlinked proximal and causal factors associated with the development of obesity. The map is composed of more than 300 variables and over 100 feedback loops across seven domains. In addition to understanding the factors that contribute to the development of obesity, the map can also be used to better understand the systems underlying appetite control and therefore may equally be termed the "Appetite Systems Map".

the patterns of behavior that bring these nutrients into the mouth can also differ widely. It can be appreciated that developing a science that encompasses such complexity is a daunting proposition. A science has therefore developed around a more restricted range of environments and behavioral types focusing on the nutrition and behavioral types relevant to technologically industrialized societies in which we live, and to the preoccupations of people living in these societies. Over the last 50 years the issue of obesity has provided a dominant framework for understanding the intimate link between behavior (appetite control) and diet. This is relevant when considering that behavior can be seen as the agency that mediates in meeting two nutritional demands; namely what to eat, and how much to eat. Both are important for health, and it means that appetite control is a central feature of the relationship between behavior and diet.

The drive to eat and food selection

The distinction between how much to eat and what to eat reflects the difference between two aspects of appetite control: the drive to eat and food selection. The problem of what to eat arises because of a combination of our omnivorous nature and the abundance of foods in the environment. This is the issue of food choice and involves the conscious or automatic selection among potential edible materials. Interestingly this food choice is not strongly programmed biologically but is dependent upon factors such as geography, climate, religion, ethnicity, and culture. The issue of how much to eat has always been conceptualized in regard to homeostatic principles of energy requirements of the body, with a stronger link to biology.

Traditionally the control of appetite has been linked to the regulation of body fat in the so-called "adipocentric" hypothesis in which the maintenance of fat stores has been regarded as a major influence on food intake. However, in the last decade much evidence has accumulated to show that the drive to eat is strongly associated with the lean mass (fat-free mass) of the body and with the metabolic activity of the lean mass. Specifically, the energy requirements generated by high metabolic rate organs such as the heart, liver, brain, and kidneys (as well as skeletal muscle) create an energy demand that drives food consumption (Hopkins and Blundell, 2017). The relationship between the resting metabolic rate and energy intake has been demonstrated in new-born babies, adolescents, adults with and without obesity, and in groups of older people. These observations give rise to the idea that energy expenditure drives energy intake, and reflects fundamental ideas proposed by human nutritionists more than 70 years ago which remained largely ignored for half a century. This concept is of evolutionary significance since it means that food intake is driven by the need to maintain vital organs in the body responsible for growth and survival. The biological basis of this drive to eat generates what can be termed "homeostatic hunger". Alongside this drive to eat related to lean body tissue, fat tissue has an inhibitory function on the drive to eat mediated by pathways in the brain involving leptin. However, these biological processes determining the strength of the drive to eat are completely unrelated to processes responsible for food selection. In turn, food selection reflects an intimate relationship between the diet and behavior and has a major influence on the processes of satiety and satiation.

Foods and the satiety cascade

Across the planet the omnivorous habit ensures that humans eat thousands of different foods that comprise an array of distinct dietary patterns. In turn, the foods selected influence the degree of satiation and satiety experienced. The processes responsible can be described in the concept of the "Satiety Cascade". **Fig. 2** depicts a model of the satiety cascade, which provides a framework to understand the mechanisms involved in the short-term control of eating behavior. The satiety cascade demonstrates the distinction between satiation and satiety, and shows how episodic signals arising from sensory, cognitive, post-ingestive, and post-absorptive processes influence the size and frequency of meals. The processes of satiation control meal size through their effect on the duration and termination of an eating episode. These processes, along with the nutritional composition of food being consumed, determine the amount of energy consumed during an eating episode. Following the termination of an eating episode, the motivation to eat is suppressed for a measurable period by the physiological effects of the ingested foods, especially in the stomach, and the hormones released from the gastro-intestinal tract during digestion and absorption of foods (see **Box 1 and 2**).

The question of whether certain foods or diets possess nutritional properties that are beneficial to enhancing satiety and therefore promoting greater control of appetite has received considerable attention in recent years. While evidence suggests that it is possible to develop diets (and meals) that are beneficial to improving control over satiation and satiety, and by extension to support weight los, it is a separate (and perhaps more concerning) issue whether people can be motivated to eat such a diet over a prolonged period. In an obesogenic environment, intentions to gain control of eating are often undermined by pervasive pressures to overconsume and the rewarding aspects of food (i.e., food hedonics). Therefore, in the absence of considerable changes to the food environment, it is unlikely that any single food or diet will be able to exert a prolonged effect on the processes depicted in the satiety cascade.

Hedonic processes of appetite control

The hedonic system of appetite control accounts for eating behavior motivated by the appeal of food in the external environment and the sensory appreciation of food as it is consumed. In today's obesogenic environment, food is a reliable source of pleasure and food hedonics have an important role in the initiation, maintenance, and termination of an eating episode. It is well established that food hedonics is more than simply liking the taste of food or experiencing pleasure from it. Research on non-human animals has

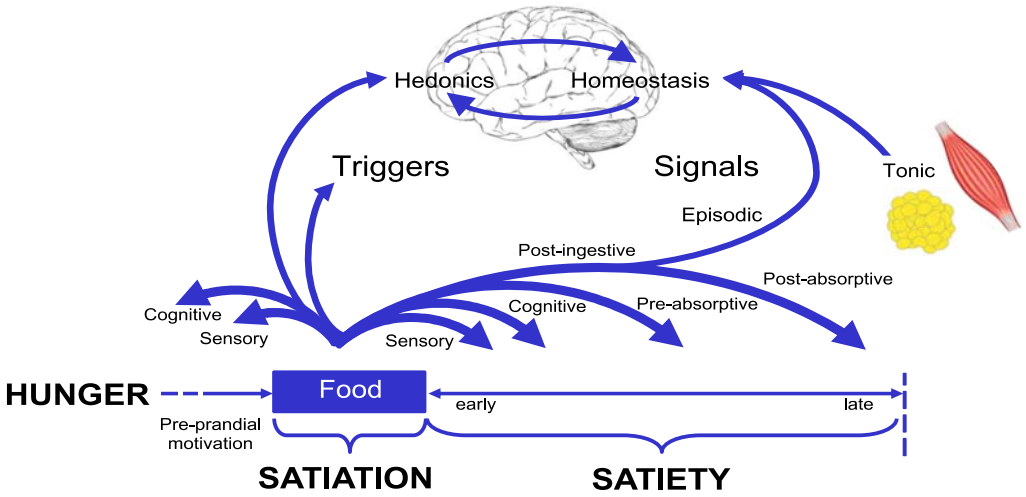


Fig. 2 The satiety cascade depicts how the size of a meal and the interval between meals is influenced by the processes of satiation and satiety. The cascade also demonstrates the interaction between the homeostatic and hedonic influences on the processes of satiation and satiety. Adapted from Blundell and Finlayson (2008).

Box 1 Which features of food cause satiety?

Following eating the feeling of fullness (satiety) is generated by several characteristics of the foods consumed including the volume, weight, sensory characteristics (taste, texture, mouthfeel), pleasantness, visual aspects, nutritional composition, non-nutritional components (e.g., fiber) and packaging/labeling.

Consequently, satiety is brought about by the integrated effects of multiple components of the food eaten. Much research has attempted to identify those specific features of foods that have the most potent impact on satiety. These features are of importance to the food industry in developing foods to control hunger and intensify the sensation of fullness. There are indications that a high protein and fiber content can enhance satiety. However, energy density is a major factor, with low energy dense diets generating stronger sensations of satiety. Currently, much interest is focused upon the effects on satiety of highly processed or ultra-processed foods with a suggestion that such foods may weaken appetite control rather than promote satiety.

Box 2 A note on physiological satiety signals

Since the first investigations of satiety, it has been believed that the effects of food composition are mediated by post-prandial physiological responses. These involve changes in gastric distention, laxation, and emptying together with the release of gastro-intestinal peptides including cholecystokinin (CCK), glucagon-like peptide (GLP-1), peptide YY (PYY), insulin, and others. For many years, CCK was regarded as the unique single satiety signal. It should be noted that all these peptides have other important roles in the management of food within the body through the processes of digestion and absorption, such as the slowing of gastric emptying and the release of bile for the emulsification of fat. Therefore, an effect on satiety may be secondary to these other functions. It remains a point of debate whether gut peptides are biomarkers or the true causes of satiety. Since different foods may have similar effects on satiety but generate quite distinct physiological profiles, it appears that there is no unique pattern underlying satiety. The same degree of satiety may be associated with different post-prandial physiological changes. There is considerable interest in the post-prandial physiological effects of highly- and ultra-processed foods.

demonstrated that the brain structures underpinning food hedonics comprise dissociable affective and motivational subcomponents, termed "liking" and "wanting", respectively. The "liking" component of food hedonics, refers to the subjective experience of pleasure derived from the sensory perception of food and is associated with the release of endogenous opioids acting on clusters of neurons termed "hedonic hotspots" within the brain. The "wanting" component refers to the motivational value of the food and is associated with the release of dopamine in the brain's mesolimbic pathway. In studies of human appetite behavior, liking and wanting for food are often viewed as explicit subjective states that refer to the everyday understanding of these terms. Specifically, liking is understood as an appreciation of the sensory characteristics of food that create its hedonic impact while wanting may describe a subjective state of desire or craving. People tend to be good at estimating and reporting their liking for food, however, they are often unable to accurately gauge their implicit wanting for food (i.e., why they are drawn to or are craving one food over another) (see **Box 3**).

Box 3 Measuring food hedonics in human appetite research

Central to the understanding of food hedonics is the distinction between Liking and Wanting. The approach to measuring liking and wanting in human behavioral studies has incorporated explicit reports of subjective experience to measure liking alongside more implicit assessment of wanting. Liking responses to food can be explicitly reported using VAS ratings by asking individuals to report "how pleasant would it be to taste some of this food right now?" in response to images of food or asking individuals to report how pleasant a food tastes in response to consumed food. Techniques to assess implicit wanting include those that assess people's willingness to expend effort to obtain a target food over an alternative and those that assess people's choice between food cues with a time-critical approach related response, where the speed of the response is interpreted as a measure of motivational value (Finlayson et al., 2007).

Palatable food can disrupt appetite control by counteracting the inhibitory effects of short-term satiety signals to increase food intake. Specifically, palatable food appears able to stimulate an orexigenic drive that can override homeostatic signals of satiety and lead to overconsumption. After food is ingested, its sensory qualities are registered by cognitive and sensory processes before it is swallowed. When a highly palatable food is consumed, the reward pathways within the brain are activated leading to the release of the neurotransmitters, dopamine, and endogenous opioids. The reward pathways have connections with the hypothalamus, which stimulate the release of hunger peptides such as NPY and orexins and inhibit the release of satiety peptides such as insulin, leptin, and cholecystokinin. Therefore, when food is highly palatable, an impulse to eat is generated, mediated by hedonic stimulation rather than biological need. This drive can be termed "hedonic hunger". For this reason, it has been suggested that hedonic processes may exert a more powerful influence over food consumption than homeostatic mechanisms. However, the homeostatic and hedonic systems of appetite control can be considered, to an extent, inseparable as the neural connections between them permit interactions that influences the overall organization of eating behavior.

Food hedonics influence satiation and satiety

The effect of liking and wanting on delaying satiation and therefore increasing the amount consumed has been well established over the last four decades. One mechanism by which liking and wanting delay satiation is through an increase in hunger at the start of, and during the early stages of a meal when the processes of liking and wanting are most likely to overlap with the biological drive for energy. This increase in hunger corresponds with a faster rate of eating and increased energy intake. These findings have been well established in the appetite laboratory-based literature and taken together suggest that liking and wanting for food can counteract the inhibitory effect of short-term satiety signals.

Outside of the laboratory conditions, the influence of liking on satiation may be less important when compared to other factors. For example, de Castro (2001) examined factors that influenced meal size and food choice in a sample of participants who recorded their food intake in their natural environments over 7-days. They found that liking for food, or its palatability, explained only 4% of the variance in total food intake after other environmental, psychological, and social factors were taken into consideration. This finding does not minimize the importance of palatability as a driver of energy intake, but highlights that outside of the laboratory, where individuals are free to choose what they eat from a wide range of palatable options, it is highly likely that they will choose to eat something that they like. Therefore, under free-living conditions, liking for food may be less important than food wanting, as food wanting is more vulnerable to being triggered by environmental cues such as availability, advertising, and marketing strategies.

Evidence suggests that greater wanting and increased food craving undermine appetite control by weakening the processes of satiety and promote overeating and weight gain through increased meal frequency, episodes of opportunistic eating (i.e., snacking, grazing behaviors), and may feature in certain forms of disordered eating. Conversely, research examining predictors of weight loss in people following a 1-year weight management program, found that increased improvements in craving control at week 8 was predictive of greater weight loss success over 1-year (Dalton et al., 2017). This suggests that interventions that target food wanting and experience of craving to enhance feelings of control over food may be beneficial for long-term weight management. This has implications in countries where environmental cues to eat are plentiful in the form of pervasive advertising, ultimately undermining attempts to improve appetite control and reduce body weight. This suggests there is a need to move toward more environmental-based interventions for weight management, rather than interventions that focus on the personal responsibility of the individual to make "good" or "small" choices.

Individual variability in appetite control and satiety responsiveness

The Obesity Systems Map (Fig. 1) identifies several factors that influence the expression of human appetite, which may render an individual vulnerable to passive overconsumption and weight gain. The diversity in susceptibility to weight gain and obesity can be characterized in terms of a person's genetics, central or peripheral physiology, psychology and behavioral profile and their physical activity and food environment. More than five decades of research has identified numerous features of the westernized environment and associated diet that interact with factors to increase vulnerability to overconsumption and subsequent weight gain. However,

Box 4 A note on measuring hunger

Hunger is a distinct conscious feeling that reflects the drive to eat. It is a subjective sensation that can be converted into a measurable quantitative variable by the use of the Visual Analogue Scale (VAS). The line is anchored at each end by a statement about the strength of hunger. A person marks the line at a point corresponding to the strength of their sensation. This instrument can be used to measure hunger before, during and after a meal to track the course of satiety.

The scale has been validated to show that the ratings correlate well with the amount of food eaten. Three other scales are usually used to provide a more complete representation of appetite sensations: fullness, desire to eat and prospective consumption (how much food do you think you can eat?). Over the course of a whole day, tracking hunger reflects the pattern of eating and the type of food consumed and can clearly identify differences between high and low energy dense diets. See **Fig. 3**.

not everyone in a westernized environment lives with overweight and obesity, and we can observe individuals habitually consuming quite distinctive diets and maintaining a body weight within the "normal" range and similarly consuming similar diets but displaying wide variations in body weight. Therefore, not everyone is vulnerable to weight gain, and excluding monogenic forms of obesity, it is unlikely that one single factor can account for why some are more vulnerable than others, which has implications for appetite control and the prevention and treatment of obesity.

One approach to characterize individual susceptibility is through the identification and characterization of phenotypes. Of relevance to the current article, is understanding how differences in recognizing and responding to signals of satiety may underlie susceptibility to overconsumption and weight gain. In controlled laboratory settings, appetite sensations have been shown to be a valid and reliable method for measuring the subjective motivation to eat (see **Box 4**). However, not everyone reports a good relationship between their sensations of appetite (i.e., hunger and fullness) and their eating behavior, and a weakened satiety response to food may contribute to impaired appetite control. Individuals exhibiting a weakened satiety response may be referred to as the "low satiety phenotype" (Drapeau et al., 2013). The low satiety phenotype has largely been observed in people with obesity, but evidence suggests that individuals with a healthy weight may also exhibit a weakened satiety response to food therefore making them vulnerable to future weight gain. Research examining the low satiety phenotype has demonstrated that it is characterized by increased disinhibition and hunger scores, lower levels of craving control, greater food wanting and increased energy intake under laboratory conditions (Drapeau et al., 2013). With regards to weight management, research has found that people low in satiety responsiveness lose less weight, have smaller reductions in central adiposity, report lower craving control and more difficulty in adhering to the dietary program compared to those high in satiety responsiveness.

Applications of satiety to societal challenges

With increasing rates of overweight and obesity, including those seen during the COVID-19 lockdowns, a high proportion of individuals report engaging in weight management attempts that involve dietary restrictions. Indeed, the US weight management industry is worth an estimated $72.7 billion per year. However, sustained body weight reductions are challenging, and most individuals regain weight lost within relatively short periods of time.

Hunger is one of the main challenges reported for sustained engagement in weight management attempts (Stubbs et al., 2012). As such, targeting satiety is one approach to support appetite control in individuals engaging in weight management attempts. Foods that promote satiety have generated large amounts of interest among academics and industry and have received high consumer demand. While technological modifications to food properties associated with satiety are currently under development (Hetherington et al., 2013), diets rich in low energy dense foods can improve appetite control and support weight management attempts. **Fig. 3** shows the profile of subjective appetite sensations reported in a sample of women engaged in a weight management program who, over two separate days and under laboratory conditions, consumed either a low ($\leq$0.8 kcal/g) or high ($\geq$2.5 kcal/g) energy dense diet over 24 h (Buckland et al., 2018). Hunger, desire to eat and prospective consumption were significantly lower, and fullness significantly higher following a low energy dense breakfast and lunch compared to high energy dense meals, and these differences in appetite sensations between energy density conditions were sustained throughout post-meal intervals in the morning and afternoon. These appetite sensations corresponded with an average total day energy reduction of 1057 kcal on the low versus the high energy dense day. Furthermore, within this study the participants who were engaged in a weight management program that advocated ad libitum intake of low energy dense foods, lost significantly greater amounts of weight compared to participants engaged in a calorie-restrictive weight management program. While the energy density manipulations in this study were relatively large, they illustrate the powerful role that energy density has on appetite control, and the potential implications for weight management. Other experimental studies with smaller energy density manipulations and observational studies, have also shown energy density to be beneficial for appetite control and weight management, including under free-living natural settings.

Traditionally, weight management programs deliver one program for all. However, given the role of individual differences in appetite control, personally tailored weight management programs may be more effective. Recent big data approaches have integrated genetic, metabolic, physiological and psychological information to develop evidence-based algorithms to inform personalized dietary approaches. Additionally, pilot work has developed a personalized intervention that assesses appetitive traits (e.g. low

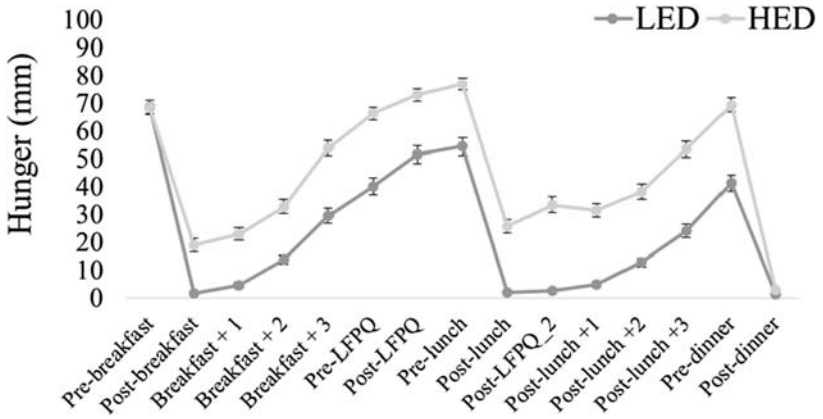


Fig. 3 Hunger profiles reported at pre- and post-meals, and at inter-meal intervals in a sample of women who consumed either a day of low- (LED) or high-energy dense meals (HED) under laboratory conditions. Hunger ratings were significantly higher after a HED breakfast compared to a LED breakfast, and this difference remained throughout the day until post-evening meal (dinner). Similar results were reported for other subjective sensations of appetite (desire to eat, prospective consumption and fullness). Values with +1, +2 and +3 refer to the number of hours after a meal was consumed. "LFPQ" refers to a behavioral task that was completed during the lunch session. Image sourced from Buckland et al. (2018).

satiety responsiveness) and delivers trait-specific support to improve appetite control. Such personalized approaches to weight management are in early phases, and more high-quality trials are needed to confirm their effectiveness. However, provided sufficient resources are available and these strategies are cost-effective, personalized approaches offer a promising avenue that recognizes the role of individual differences in appetite control. Importantly the challenge of dealing with overconsumption is equaled by the challenge of managing the widespread sedentariness and low physical activity that characterizes many modern societies. This is doubly relevant since the interaction between energy expenditure and energy intake means that the degree of physical activity has implications for appetite control (see **Box 5**).

In addition to weight management, a further societal challenge relevant to appetite control and diet, which will be briefly covered here, is climate change. Given the high protein content of meat, it is highly satiating and is often incorporated into dietary weight management programs. Yet, meat is resource intensive, and animal livestock contribute 14.5% of human-made greenhouse gas emissions (GHGE). Estimates show that a 50% reduction in meat intake would result in a 25%–40% reduction in GHGE linked with food production. As such, current meat-rich diets need to change. Promisingly, there has been increasing numbers of individuals reporting a reduction in meat and a shift to a favourably viewed flexitarian diet. However, the challenge now is to achieve widespread reductions in meat intake. For a national shift toward reduced meat diets, meat-free alternatives (e.g., pulses, plant-based, and cultured meat) need to be palatable and satiating. While research is growing on this topic, it remains limited. Some studies have reported that compared to meat-based proteins, meat-free alternatives result in similar or favourable appetite sensations, energy intake, appetite hormones and weight loss. However, most studies are limited to acute designs in the laboratory. Therefore, more research, and a systematic review of the available studies is currently needed. Importantly, methods from appetite research can play a key role in ensuring research on sustainable diets is conducted rigorously to allow research findings to inform societal applications for environmentally sustainable diets.

The impact of ultra-processed foods on satiety

As discussed, low-energy dense foods can improve appetite control and reduce energy intake. A key aspect of low-energy dense foods is that they tend to be minimally processed. The NOVA food classification system comprises four categories that differ according to their level of processing ranging from foods that are unprocessed or minimally processed (group 1) to foods that are defined as being ultra-processed (group 4). Ultra-processed foods (UPFs) often contain high amounts of sugar, sodium, and unhealthy types

Box 5 A note on physical activity and appetite: Sedentariness and Satiety

A body of research now shows that there is an intimate interaction between energy expenditure and energy intake. This means that the control of appetite and the amount of food consumed is related to the degree of physical activity performed (and the energy expended). Appetite is better regulated when activity level is high. In contrast satiety is weakened when people are in a sedentary state; this allows overconsumption to occur. Many studies have shown that people in advanced technological cultures spend 60–70% of the working day in a sedentary state (low energy expenditure and often in a sitting posture), and that this is correlated with the degree of adiposity (amount of body fat). Therefore, a sedentary lifestyle is a risk factor for weak satiety and overconsumption leading to an increase in body fat.

of fat and they are often hyperpalatable. The level of processing food undergoes has been associated with differences in glycaemic index and satiating capacity. Specifically, the more processed a food is, the higher the glycaemic response and the lower its satiety potential. Further to this, greater consumption of UPFs has been associated with differences in energy intake and weight gain. A two-week, crossover randomized control trial compared the effect of consuming a diet primarily comprising of UPFs compared to a diet that contained no processed foods. Findings showed that increased consumption of UPFs was associated with greater caloric intake and weight gain of just under 2 lbs in two weeks, while the no processed foods diet was associated with weight loss of a similar amount. Interestingly, participants in this trial did not rate the UPF diet as more palatable compared to the unprocessed diet and neither diet produced differences in subjective sensations of appetite. The UPF diet was associated with greater eating rate (17 kcal/min compared to 7.4 kcal/min). Within the food industry there has been a shift toward the reformulation of food products in response to evidence that UPFs are fueling the obesity pandemic. Reformulation of products has acted to reduce the amount of fat or sugar within a product, however the removed ingredient is often replaced with more sugar or artificial sweeteners, respectively, making the reformulated products no healthier, but often more expensive, than the original. This reformulation approach has been described as "damage limitation" to avoid more evidence-based approaches that include restrictions on food advertising, availability and pricing policies.

It is apparent that the consumption of minimally processed or unprocessed foods can improve appetite control and support weight management attempts by reducing the energy density of the diet and increasing feelings of satiety. Taken together these findings suggest that restricting the consumption, advertisement, and reliance on UPFs and promoting healthy, less processed meals and foods may be a more effective treatment and prevention strategy for obesity. To this end, in 2019, a Framework Convention on Food Systems was proposed and recommended by the Lancet Commission to tackle "Big Food, Big Soda, and Big Alcohol" and the obesity pandemic using lessons learned from tobacco control and reduction of smoking behavior. Understanding the effect that processed foods have on satiety, energy intake and markers of health is an important focus for future human appetite research. Evidence to date is primarily based on correlational research and while the NOVA classification system is commonly used there is no agreed upon definition for what constitutes a high- or ultra-processed food (Gibney and Forde, 2022).

Conclusion

This article has been written in 2022 at a particular moment relevant to diet and eating behavior. The pandemic of obesity which has existed for over 20 years shows no sign of abating. The world is beginning to emerge from a pandemic of COVID-19, which has been associated with general increases in unhealthy food habits and in the level of obesity. The food supply in many parts of the world is characterized by heavily processed foods or ultra-processed foods which compromise healthy eating. Food companies seek further ways in which the palatability of foods can be enhanced. Wars in various parts of the planet threaten the world's food supply and will cause food shortages. There is a strong movement toward the development of sustainable diets and the promotion of food resilience. While there is much interest in personalized nutrition, for some people food is unpredictable and precarious. All of these situations have implications for the control of human appetite; simultaneously appetite control has implications for the way in which people can deal with these situations. In this article we have tried to show how a description of the processes responsible for appetite and satiety can help our understanding of the relationship between diet and behavior under challenging environmental conditions, and how this understanding can be applied to change the environment to support healthy and sustainable food choices.

References

Blundell, J., Finlayson, G., 2008. Mechanisms and biomarkers of appetite control. Agro Food Industry Hi-Tech, 19 (5), 18–20.

Buckland, N.J., Camidge, D., Croden, F., Lavin, J.H., Stubbs, R.J., Hetherington, M.M., Blundell, J.E., Finlayson, G., 2018. A low energy–dense diet in the context of a weight-management program affects appetite control in overweight and obese women. J. Nutr. 148, 798–806.

Dalton, M., Finlayson, G., Walsh, B., Halseth, A., Duarte, C., Blundell, J., 2017. Early improvement in food cravings are associated with long-term weight loss success in a large clinical sample. Int. J. Obes. 41, 1232–1236.

Drapeau, V., Blundell, J., Gallant, A., Arguin, H., Després, J.-P., Lamarche, B., Tremblay, A., 2013. Behavioural and metabolic characterisation of the low satiety phenotype. Appetite 70, 67–72.

Finlayson, G., King, N., Blundell, J.E., 2007. Is it possible to dissociate "liking" and "wanting" for foods in humans? A novel experimental procedure. Physiol. Behav. 90, 36–42.

Gibney, M.J., Forde, C.G., 2022. Nutrition research challenges for processed food and health. Nat. Food 3, 104–109.

Hetherington, M., Cunningham, K., Dye, L., Gibson, E., Gregersen, N., Halford, J., Lawton, C., Lluch, A., Mela, D., Van Trijp, H., 2013. Potential benefits of satiety to the consumer: scientific considerations. Nutr. Res. Rev. 26, 22–38.

Hopkins, M., Blundell, J.E., 2017. Energy metabolism and appetite control. In: Appetite and Food Intake: Central Control, second ed.

Midgley, M., 1978. Beast and Man: The Roots of. Human Nature. Cornell Univer.

Rozin, P., Dow, S., Moscovitch, M., Rajaram, S., 1998. What causes humans to begin and end a meal? a role for memory for what has been eaten, as evidenced by a study of multiple meal eating in amnesic patients. Psychol. Sci. 9, 392–396.

Stubbs, J., Brogelli, D., Pallister, C., Avery, A., Mcconnon, A., Lavin, J., 2012. Behavioural and motivational factors associated with weight loss and maintenance in a commercial weight management programme. Open Obes. J. 4.

Further reading

Berry, S.E., Valdes, A.M., Drew, D.A., Asnicar, F., Mazidi, M., Wolf, J., Capdevila, J., Hadjigeorgiou, G., Davies, R., AL Khatib, H., 2020. Human postprandial responses to food and potential for precision nutrition. Nat. Med. 26, 964–973.

Berthoud, H.-R., Münzberg, H., Morrison, C.D., 2017. Blaming the brain for obesity: integration of hedonic and homeostatic mechanisms. Gastroenterology 152, 1728–1738.

Blundell, J., 2017. The contribution of behavioural science to nutrition: appetite control. Nutr. Bull. 42, 236–245.

Blundell, J.E., Gibbons, C., Beaulieu, K., Casanova, N., Duarte, C., Finlayson, G., Stubbs, R.J., Hopkins, M., 2020. The drive to eat in homo sapiens: energy expenditure drives energy intake. Physiol. Behav. 219, 112846.

Buckland, N.J., Camidge, D., Croden, F., Myers, A., Lavin, J.H., Stubbs, R.J., Blundell, J.E., Finlayson, G., 2019. Women with a low-satiety phenotype show impaired appetite control and greater resistance to weight loss. Br. J. Nutr. 122, 951–959.

Dalton, M., Hollingworth, S., Blundell, J., Finlayson, G., 2015. Weak satiety responsiveness is a reliable trait associated with hedonic risk factors for overeating among women. Nutrients 7, 7421–7436.

Finlayson, G., Dalton, M., 2012. Hedonics of food consumption: are food "liking" and "wanting" viable targets for appetite control in the obese? Curr. Obes. Rep. 1, 42–49.

Hall, K.D., Ayuketah, A., Brychta, R., Cai, H., Cassimatis, T., Chen, K.Y., Chung, S.T., Costa, E., Courville, A., Darcey, V., 2019. Ultra-processed diets cause excess calorie intake and weight gain: an inpatient randomized controlled trial of ad libitum food intake. Cell Metabol. 30, 67–77. e3.

Mela, D.J., 2006. Eating for pleasure or just wanting to eat? Reconsidering sensory hedonic responses as a driver of obesity. Appetite 47, 10–17.

Rolls, B.J., 2017. Dietary energy density: applying behavioural science to weight management. Nutr. Bull. 42, 246–253.

Dietary modulation of inflammation

P.C. Calder, School of Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of D.H. Hwang, Dietary Modulation of Inflammation, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 74–78, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00161-6.

Introduction 67
The inflammatory response 68
Inflammation and disease 69
The importance of resolution of inflammation 70
Modulation of inflammation by diet and dietary components 70
General mechanisms by which dietary components could influence inflammation 70
Dietary patterns and inflammation 72
Specific foods and inflammation 72
Fatty acids and inflammation 72
Saturated fatty acids 72
Trans fatty acids 72
Polyunsaturated fatty acids 72
Antioxidant micronutrients and inflammation 74
Polyphenols and inflammation 74
The gut microbiota and inflammation 74
Postprandial inflammation 75
Dietary inflammatory indexes 75
Conclusions and perspectives 76
References 77

Key points

- Inflammation is a normal part of host defense against pathogens and is involved in removal of cellular debris and wound repair
- Inflammation involves a number of different cell types, cellular interactions and chemical mediators
- Cells involved in inflammation include neutrophils, monocytes, macrophages, eosinophils and mast cells but on-going inflammation can involve T and B lymphocytes
- Chemical mediators of inflammation include lipids (e.g., prostaglandins, leukotrienes), proteins (e.g., cytokines, chemokines), reactive oxygen species (e.g., superoxide) and amino acid derivatives (e.g., histamine)
- Oxidative stress is part of the inflammatory response
- The acute phase protein C-reactive protein (CRP) is often used as a biomarker of inflammation
- Inflammation is normally self-resolving and resolution of inflammation is an active process
- Failure to resolve inflammation can result in damage to host cells and tissues and can lead to disease
- Diet and dietary components can modulate inflammation
- Healthy dietary patterns and foods including whole grains, nuts and seeds, fruits and vegetables and fish reduce inflammation
- Omega-3 polyunsaturated fatty acids, antioxidant vitamins and many polyphenols reduce inflammation
- The gut microbiota is linked to inflammation and foods or supplements that promote a healthy diverse gut microbiota reduce inflammation
- There is an increase in inflammation after eating a meal (post-prandial inflammation) that is exaggerated by glucose or saturated fatty acids in the meal

Introduction

Inflammation is a normal defense mechanism that protects the host from infection and other insults. Inflammation initiates the killing of pathogens and is involved in removal of cellular debris as well as tissue repair processes, helping to restore homeostasis at infected or damaged sites. The classic signs of inflammation are redness, swelling, heat, pain, and loss of function. Inflammation involves interactions among many different cell types, movement of cells between body compartments and the production of, and

responses to, many different chemical mediators (Calder et al., 2009). When an inflammatory response does occur, it is normally well regulated and does not cause excessive damage to the host; it is also self-limiting and resolves rapidly. Controlled properly, regulated inflammatory responses are essential to remaining healthy and maintaining homeostasis. Loss of tolerance to normally benign exposures (e.g., food, harmless microbes) or an impairment in regulatory processes (e.g., resolution) can lead to pathologic inflammation (Calder et al., 2009). If this becomes excessive, irreparable damage to host tissues and disease can occur. Typically, diseases or conditions with a well-recognized inflammatory component are treated with general or specific anti-inflammatory pharmaceuticals. Because dietary components may influence various parts of the inflammatory response, diet may play a role in predisposing to inflammatory conditions and altered diet or intake of specific dietary components may be useful in the prevention or therapy of such conditions. This article describes inflammation and the inflammatory response, including resolution, and describes evidence that dietary patterns, specific foods and food components can influence inflammatory processes, and indicates the likely mechanisms of action of these food components.

The inflammatory response

Inflammation is largely orchestrated by leukocytes (neutrophils, monocytes/macrophages, eosinophils, and mast cells) designed to eliminate pathogens and to heal tissue injury using multiple chemical mediators (chemokines, cytokines, adhesion molecules, lipid mediators, growth factors, proteases). Inflammation may be classified as acute or chronic (Calder et al., 2013). Acute inflammation is the initial response of the body to a harmful stimulus and is achieved by the increased movement of plasma and leukocytes (especially granulocytes which include neutrophils, eosinophils and mast cells) from the blood into the injured tissues. A cascade of biochemical events propagates and matures the inflammatory response, involving the local vascular system, the immune system, and various cells within the injured tissue. Prolonged inflammation, known as chronic inflammation, leads to a progressive shift in the type of cells present at the site of inflammation (e.g., monocytes and macrophages become more involved) and is characterized by simultaneous destruction and healing of the tissue from the inflammatory process. Chronic inflammation can eventually involve T lymphocytes and B lymphocytes, the former shaping and sustaining the nature of the on-going inflammatory response and the latter producing antibodies specific to the particular response (Calder et al., 2009, 2013). The features of acute and chronic inflammation are compared in **Table 1**.

Common to both acute and chronic inflammation are an afferent phase, in which the presence of a "foreign material" is "sensed" by some types of cell and an efferent phase, in which an inflammatory response is generated to eliminate the perceived hostile intruder. The "sensing" of pathogens is carried out by pattern-recognition receptors (PRRs) on the leukocytes that recognize so-called microbe-associated molecular patterns (MAMPs). Toll-like receptors (TLRs) are an example of PRRs. The purpose of the inflammatory response to harmful microorganisms is obvious, and the response is beneficial and necessary to protect the integrity of the body, so long as it does not become unnecessarily destructive or long lasting. Inflammation caused by non-pathogenic agents, which might include damaged host tissues, can also be beneficial, but it may also have negative health effects, especially if it has a long duration. Irrespective of the cause of the inflammation, the response involves four major events: (1) an increase in blood supply to the site of inflammation; (2) increased capillary permeability caused by opening of junctions between endothelial cells, permitting plasma and larger molecules, not normally capable of traversing the endothelium, to do so, thus delivering soluble mediators to the site of inflammation; (3) leukocyte migration from the capillaries into the surrounding tissue and then to the site of inflammation; this process is promoted by release of chemoattractants from the site of inflammation and by the up-regulation of adhesion molecules on the endothelium.; (4) release of mediators from leukocytes at the site of inflammation - these may include lipid mediators (e.g., prostaglandins, leukotrienes), protein mediators (e.g., cytokines, chemokines), reactive oxygen species (e.g., superoxide), amino acid derivatives (e.g., histamine), and enzymes (e.g., matrix proteases), depending on the cell type, the nature of the inflammatory stimulus, the anatomic site, and the stage during the inflammatory response (Calder et al., 2009, 2013). Some of the inflammatory mediators may escape the inflammatory site into the circulation and from there exert

Table 1 Features of acute and chronic inflammation.

	Acute	*Chronic*
Causative agent	Pathogens, injured tissues	Continued exposure to pathogens or loss of tolerance to harmless pathogens, other environmental agents, or self
Major cell types involved	Neutrophils and other granulocytes (eosinophiles, mast cells), mononuclear cells (monocytes, macrophages)	Mononuclear cells (monocytes, macrophages, T lymphocytes, B lymphocytes)
Primary mediators	Vasoactive amines (histamine), bioactive lipids (prostaglandins, leukotrienes), chemokines	Cytokines, chemokines, growth factors, reactive oxygen species, hydrolytic enzymes
Onset	Immediate	Delayed
Duration	Hours to days	May be months or years; may be intermittent depending upon exposures
Possible outcomes	Resolution, abscess formation, chronic inflammation	Tissue destruction, fibrosis, necrosis, disease

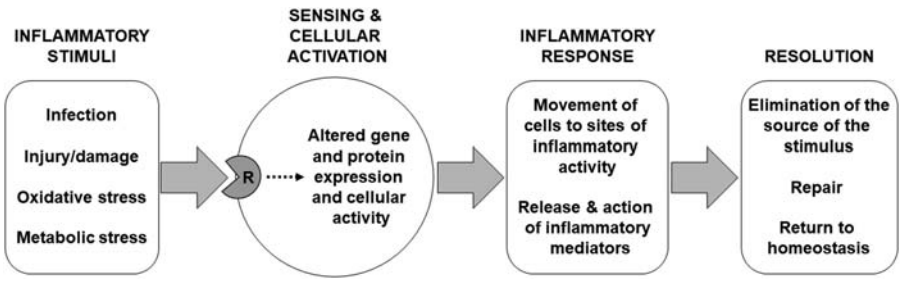


Fig. 1 Overview of the inflammatory response. Inflammatory stimuli trigger an inflammatory response after they are "sensed" often by receptors. This initiates cell signaling processes leading to cellular activation, some immediate cell responses and altered gene and protein expression. Cells move to sites of inflammatory action and there is release of, and responses to, a plethora of inflammatory mediators. These aim to eliminate the source of initial trigger, and then during resolution to remove any cell debris and to repair any local tissue damage.

systemic effects. For example, the cytokine interleukin-6 (IL-6) induces hepatic synthesis of the acute phase protein C-reactive protein (CRP), and the cytokine tumor necrosis factor-α (TNF-α) elicits metabolic effects within skeletal muscle, adipose tissue, and bone. Serum or plasma CRP concentration is often used as a biomarker of inflammation. Concentrations of CRP are greatly increased during infection or in people with an active chronic inflammatory disease and are easily measured by enzyme-linked immunosorbent assays. However, in uninfected healthy people, CRP concentrations are low and require measurement using highly sensitive assays; this is sometimes referred to as "highly sensitive CRP", but it is important to note that this term simply refers to CRP measured using a highly sensitive assay. Using such assays, it has been identified that CRP concentrations (and other circulating markers of inflammation) are elevated with aging (Calder et al., 2017), in those living with obesity (Calder et al., 2011) and in those with cardiometabolic disease; furthermore elevated CRP concentrations are predictive of future cardiovascular disease (Libby and Ridker, 2004).

Fig. 1 provides an overview of the inflammatory response.

Inflammation and disease

Unresolved inflammation can damage host tissues and is a recognized contributor to the pathology of many conditions (Calder et al., 2009, 2013). In some cases, such as critical illness, rheumatoid arthritis, inflammatory bowel diseases, asthma and psoriasis, the central contribution of inflammation to the pathology of the condition is well recognized. Individuals with these conditions have heavy infiltration of inflammatory cells at the site of disease activity (e.g., joints, intestinal mucosa, lungs, skin), and they have elevated concentrations of inflammatory mediators at those sites and in the systemic circulation. These conditions are treated with varying levels of success by anti-inflammatory drugs. In other cases, such as atherosclerosis and obesity, the role of inflammation has emerged more recently (Hansson, 2005; Tilg and Moschen, 2006), and its contribution to the pathologic features alongside the many other relevant factors is less clear. Individuals with these conditions also show infiltration of inflammatory cells at the site of disease activity (e.g., blood vessel wall, adipose tissue) and have moderately elevated levels of inflammatory mediators and CRP in the systemic circulation (Calder et al., 2009, 2011, 2013). Although inflammation-induced tissue damage occurs in an organ-specific manner (e.g., joints, gut, lungs, skin, blood vessel wall, adipose tissue) in different diseases or conditions, there is some commonality among the responses seen in the different organs (Calder et al., 2009, 2013). In general, the inflammatory response observed is normal, but it occurs in the wrong context; this relates to loss of barrier function (epithelial or endothelial), inappropriate triggering (i.e., a response to a normally benign stimulus equivalent to a loss of tolerance), lack of down-regulation (resolution) to control the response, and tissue destruction with a loss of function. In some cases, the inflammation is the result of exogenous triggers such as microbes or allergens. In other cases, it is the result of loss of tolerance to self, as seen in autoimmune conditions, while in other case it is secondary to tissue damage caused by endogenous molecules such as oxidized low-density lipoprotein. Most, if not all, chronic inflammatory diseases are characterized by overproduction of cytokines (e.g., TNF-α, IL-1β, IL-6), chemokines (e.g., IL-8, monocyte chemoattractant protein (MCP)-1), eicosanoids (prostaglandin E_2, 4-series leukotrienes), and matrix metalloproteinases. Elevated levels of these mediators act to amplify the inflammatory process (e.g., by attracting further inflammatory cells to the site) and contribute to tissue damage and to the clinical symptoms observed. Many of these mediators are positively regulated through the transcription factor nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) (Kumar et al., 2004) or through cytoplasmic protein complexes called inflammasomes (Blevins et al., 2022), and some are negatively regulated through peroxisome proliferator-activated receptors (PPARs) (Wang et al., 2016). Fig. 2 provides an overview of these processes. Entry of inflammatory cells to sites of inflammatory activity is facilitated by up-regulation of adhesion molecules on the endothelium, a process that is promoted by inflammatory cytokines and by a range of inflammatory triggers, frequently acting through NF-κB. The continuous process of tissue injury, healing, and repair, in response to the release of cytokines, chemokines, and growth factors by infiltrating inflammatory cells, as well as resident tissue cells, results in tissue remodeling.

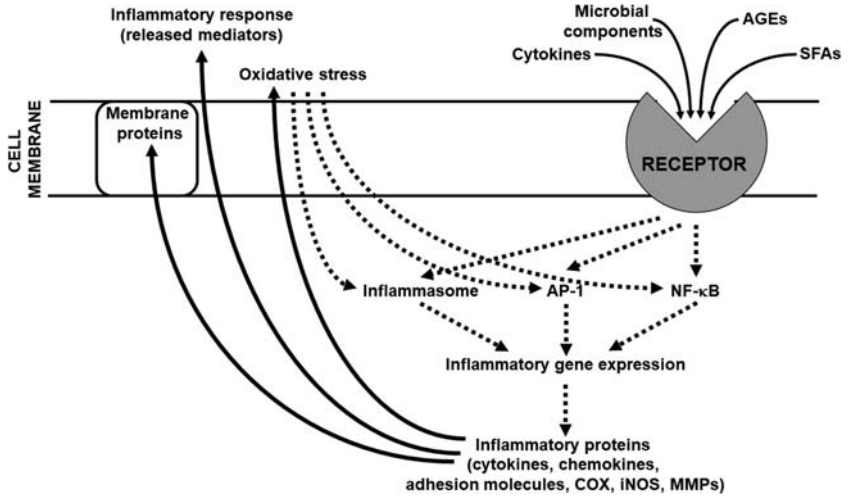


Fig. 2 A generalized overview of inflammatory signaling. A range of inflammatory stimuli are sensed, often by receptors such as pattern recognition receptors, and initiate cell signaling pathways. These pathways lead to activation of various transcription factors, including nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) and activator protein 1 (AP-1), and inflammasomes. Inflammatory gene expression is upregulated and inflammatory proteins (including cytokines, chemokines, adhesion molecules, cyclooxygenase (COX), inducible nitric oxide synthase (iNOS) and matrix metalloproteinases (MMPs)) are produced. Oxidative stress can activate these same inflammatory pathways while the inflammatory response induces production of oxidants creating oxidative stress. AGEs, advanced glycated end products; SFAs, saturated fatty acids.

The importance of resolution of inflammation

As indicated earlier, the inflammatory response can be damaging to host tissues and if this is not controlled pathology and disease can ensue. Therefore, timely resolution of heightened inflammatory responses is critical to prevent such collateral tissue damage. In fact, inflammation is usually self-limiting. It is normally resolved after its purpose is achieved; however, loss of resolution can lead to chronic inflammation, promoting pathogenic processes that lead to the development of major chronic diseases. The self-regulation of inflammation involves the activation of negative feedback mechanisms such as the secretion of anti-inflammatory cytokines (e.g., IL-10), inhibition of proinflammatory signaling cascades, shedding of receptors for inflammatory mediators, and activation of regulatory cells (Barnig et al., 2019; Panigrahy et al., 2021). One interesting aspect of resolution is generation of negative regulators of PRRs (Hedl and Abraham, 2013). These negative regulators include soluble decoy TLRs that interfere with ligand binding or receptor dimerization, and intracellular negative regulator molecules that interfere with recruitment of downstream signaling molecules, or stimulate degradation of signaling molecules. Many of these negative regulators are induced by ligands of PRRs suggesting that PRRs can activate both pro-inflammatory and anti-inflammatory/pro-resolving signals in negative feedback and temporal manners. Another endogenous pathway to promote resolution of inflammation involves the production of lipid mediators, mainly, though not exclusively, from omega-3 polyunsaturated fatty acids (PUFAs) (Chiang and Serhan, 2020). Eicosapentaenoic acid (EPA), docosapentaenoic acid (DPA) and docosahexaenoic acid (DHA) give rise to mediators collectively termed specialized pro-resolving mediators (SPMs) that include resolvins, protectins and maresins (Fig. 3). SPMs seem to be among the most potent and most important pro-resolution molecules.

Modulation of inflammation by diet and dietary components

General mechanisms by which dietary components could influence inflammation

There are a number of mechanisms by which dietary components might influence the inflammatory response. The most obvious would be where a dietary component acts as a direct trigger to inflammation. Perhaps the best example of this is in food allergy where the individual has lost tolerance to (i.e., become immunologically sensitized to) a food antigen; in this case the antigen is termed an allergen (Tedner et al., 2022). Common foods that can cause allergy are cows' milk, hens' egg, peanuts and shellfish. In the case of food allergy, the allergen is recognized by the immune system and there is a rapid and often strong inflammatory response manifested in the skin, the oral cavity, the gastrointestinal tract and, where the allergy has progressed to asthma, the lungs. This reaction may even be fatal with death from anaphylactic shock. Some aspects of the allergic response can be treated by antihistamines while others require agents targeted at other mediators such as leukotrienes, while anaphylactic shock requires treatment with adrenaline. The most obvious way for a sensitized individual to prevent the occurrence of an allergic response is to avoid the food responsible, although some approaches to immunotherapy seem to be a success. Food avoidance can have adverse nutritional, social and psychological impacts. Another example of an inflammatory response to a food component is seen in celiac disease where the trigger is gluten (Iverson and Sollid, 2022). This is not, strictly speaking, an allergy because the immune-

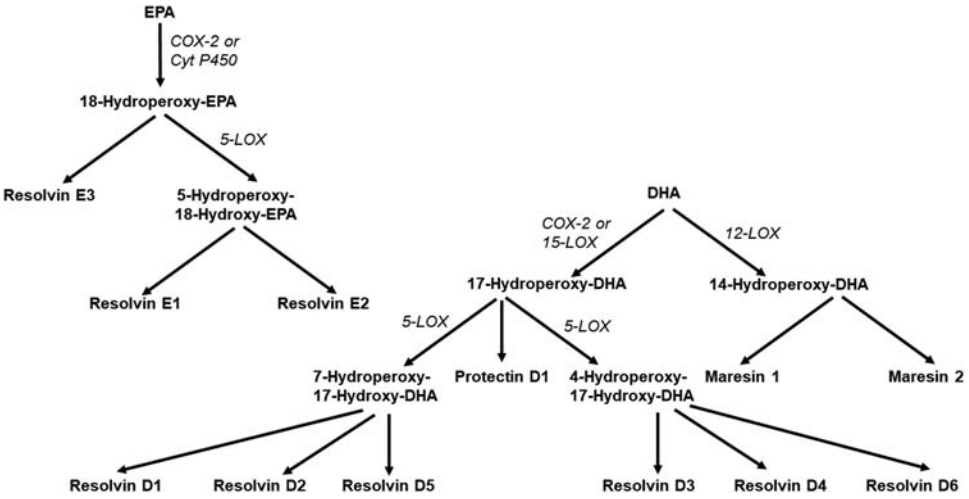


Fig. 3 Outline of the pathways of biosynthesis of specialized pro-resolving mediators (resolvins, protectins and maresins) produced from the omega-3 fatty acids eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA). COX, cyclooxygenase; LOX, lipoxygenase.

inflammatory response to gluten is different from the response to an allergen. Clearly in food allergy and celiac disease a food component is acting as a direct trigger to the inflammatory response. There is emerging evidence that other food components can also act as inflammatory triggers; one example is some saturated fatty acids which have been demonstrated to bind to TLR-4, the PRR for bacterial lipopolysaccharide, so activating the NF-κB system resulting in an inflammatory response driven mainly by monocytes and macrophages (Hwang et al., 2016). The response that occurs is immunologically very different from the allergic response, and so this should not be regarded as a food allergy, and its intensity is relatively modest, although it may be clinically important over a prolonged period of time.

Advanced glycation end products (AGEs) can be formed in the body by oxidative reactions sometimes promoted by high availability of glucose and by oxidative stress. AGEs can also be derived from food, being formed during cooking and food processing procedures, especially the use of high temperatures. AGEs bind to cell surface receptors (RAGEs), which can facilitate their removal but can also lead to cellular activation, the latter triggering proinflammatory signaling pathways causing activation of transcription factors such as NF-kB and consequent upregulation of inflammatory processes (Twarda-Clapa et al., 2022). This could provide one link between oxidative stress, inflammatory processes and disease. Studies in both healthy subjects and people living with diabetes have reported a positive association between intake of AGEs from the diet and biomarkers of inflammation such as CRP.

Beyond acting as direct triggers to inflammation, food components may have a range of other effects that act to modulate the intensity of inflammatory triggers or the response to inflammatory triggers (Fig. 4). As already mentioned, oxidative stress can activate inflammation partly through the NF-κB system (conversely inflammation induces oxidative stress by promoting generation of reactive oxygen species). Thus, dietary components that favor oxidative stress (e.g., AGEs, oxidized lipids) could promote inflammation while dietary components that reduce oxidative stress (e.g., antioxidant vitamins, carotenoids, polyphenols) could reduce inflammation. Some dietary components can regulate inflammatory signaling by interfering with membrane or cytoplasmic processes that drive activation of NF-kB or the inflammasome. PUFAs are important in this regard. It is important to note that some chemical mediators of inflammation are formed from nutrients once those nutrients are assimilated into cells; well known examples are histamine which is produced from histidine and prostaglandins and leukotrienes which are produced from the omega-6 PUFA arachidonic acid.

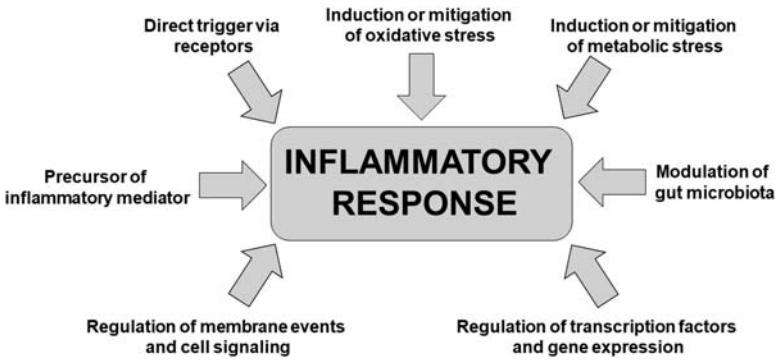


Fig. 4 Overview of the ways in which dietary components can influence inflammation.

Dietary patterns and inflammation

Greater adherence to the traditional Mediterranean diet (rich in fruits, vegetables, whole grains, legumes, nuts, fish, and low-fat dairy products, with moderate consumption of wine, and with olive oil as the principal source of fat) is associated with lower blood concentrations of inflammatory markers (e.g., IL-6, CRP) in healthy people (Chrysohoou et al., 2004; Fung et al., 2005). Intervention studies demonstrate that consuming a Mediterranean diet decreases inflammation in healthy subjects, in obese subjects, and in subjects with high cardiovascular risk (Esposito et al., 2004; Estruch et al., 2006; Dai et al., 2008; Schwingshackl and Hoffmann, 2014). Dietary patterns consistent with vegetarianism have been associated with lower concentrations of inflammatory markers in the bloodstream compared with nonvegetarian diets (Purschwitz et al., 2001; Szeto et al., 2004). The healthy eating index, alternate healthy eating index and diet quality sore, which are all composite scores reflecting a healthy diet, have all been inversely associated with inflammatory biomarkers such as CRP and IL-6 in several studies (Ford et al., 2005; Fung et al., 2005; Fargnoli et al., 2008). Using data from the Nurses' Health Study, a dietary pattern that was significantly associated with higher concentrations of several inflammatory markers was identified; this pattern was high in sugar-sweetened soft drinks, refined grains, diet soft drinks, and processed meat but low in wine, coffee, cruciferous vegetables, and yellow vegetables (Schulze et al., 2005). In contrast to a prudent duet, which was associated with lower concentrations of several inflammatory markers, a Western diet pattern was associated with higher concentrations of these same markers (Lopez-Garcia et al., 2004a).

Specific foods and inflammation

Observational studies have reported inverse associations between intake of whole grains, nuts and seeds, fruits and vegetables, fish, and tea and certain blood biomarkers of inflammation, often including CRP and IL-6 (see Calder et al., 2011 and 2017 for references). Regular consumption of small doses of dark chocolate decreased markers of inflammation in healthy subjects (di Giuseppe et al., 2008). Intervention studies with whole grain foods are inconsistent with regard to the effect on inflammation, although some do report a reduction (see Calder et al., 2011 and 2017 for references). Interventions with fruits and vegetables as a food group have been successful at reducing the blood concentrations of inflammatory markers (see Calder et al., 2011 and 2017 for references). However, studies focusing on a single variety of vegetable or fruit have been inconsistent. Soy protein appears not to affect circulating markers of inflammation. Intervention trials of drinking black tea, green tea, or coffee have not yielded a consistent picture regarding a possible anti-inflammatory effect (see Calder et al., 2011 for references).

Fatty acids and inflammation

Saturated fatty acids

Experimental studies consistently show that several saturated fatty acids, including lauric (12:0), myristic (14:0) and palmitic (16:0), can directly trigger inflammation via TLRs (Hwang et al., 2016). However, human observational studies investigating the link between saturated fatty acid intake and inflammatory markers have produced inconsistent findings. An intervention study involving feeding diets rich in stearic acid (18:0) or in lauric, myristic and palmitic acids to healthy men for 5 weeks showed higher concentrations of CRP, IL-6 and sE-selectin compared with feeding a diet enriched in oleic acid (18:1n-9) (Baer et al., 2004).

Trans fatty acids

Experimental studies consistently shown that trans fatty acids, especially those of industrial origin, promote inflammation through a combination of effects on cell membranes, intracellular signaling pathways and NF-κB activation (Valenzuela et al., 2019). Data from the Nurses' Health Study indicated positive associations between the intake of trans fatty acids and the concentrations of all six inflammatory markers assessed, including CRP, IL-6 and three soluble adhesion molecules (Lopez-Garcia et al., 2005). In a 5 week intervention in healthy men, a trans fatty acid-enriched diet resulted in higher CRP, IL-6 and sE-selectin concentrations than diets rich in oleic acid, stearic acid or the combination of lauric, myristic and palmitic acids (Baer et al., 2004).

Polyunsaturated fatty acids

The key link between fatty acids and inflammation is that eicosanoids, which act as mediators and regulators of inflammation, are generated from 20-carbon PUFAs of both the omega-6 and omega-3 families (Calder, 2020a; Christie and Harwood, 2020). Because inflammatory cells typically contain a high proportion of the omega-6 PUFA arachidonic acid (20:4n-6) and low proportions of other 20-carbon PUFAs, arachidonic acid is usually the major substrate for eicosanoid synthesis. Eicosanoids include prostaglandins, leukotrienes and other oxidized PUFA derivatives, and are generated by reactions catalyzed by cyclooxygenase (COX), lipoxygenase (LOX) and cytochrome P450 enzymes. At least two COX enzymes and several LOX and cytochrome P450 enzymes are expressed in different cell types, according to different conditions; between them, they produce a range of mediators involved in modulating the intensity and duration of inflammatory responses. These mediators have cell- and stimulus-specific sources and frequently have opposing effects. Thus, the overall physiologic (or pathophysiologic) outcome depends on the cells present, the

nature of the stimulus, the timing of eicosanoid generation, the concentrations of different eicosanoids generated, and the sensitivity of target cells and tissues to the eicosanoids generated. Many of the actions of arachidonic acid-derived eicosanoids are pro-inflammatory; those eicosanoids are central to the pathology of several inflammatory diseases including food allergy and asthma and several anti-inflammatory drugs target the synthesis or action of eicosanoids. For these reasons omega-6 PUFAs are generally regarded as pro-inflammatory. While that may be correct in some settings, it is probably an over-simplification. In fact, observational studies have not always reported the anticipated positive relationship between omega-6 PUFA intake or status and inflammatory biomarkers (Innes and Calder, 2018).

The amount of arachidonic acid in inflammatory cells can be decreased by increased consumption of marine omega-3 PUFAs (eicosapentaenoic acid [EPA] and docosahexaenoic acid [DHA]) found in seafood, especially oily fish, and in fish oil supplements (see Calder, 2015 for references). Thus, less substrate is available for synthesis of inflammatory eicosanoids from arachidonic acid. EPA is also able to act as a substrate for COX, LOX and cytochrome P450 enzymes and produces eicosanoids with a structure slightly different from those formed from arachidonic acid. The functional significance of this finding is that the mediators formed from EPA are typically less potent than those formed from arachidonic acid. Further to this, as previously mentioned, both EPA and DHA, and also DPA, give rise to SPMs, which exert potent anti-inflammatory and inflammation-resolving actions. Thus, one anti-inflammatory mechanism of action of marine omega-3 PUFAs is antagonism of production of inflammatory eicosanoids from arachidonic acid coupled with the generation of less potent EPA-derived eicosanoids and anti-inflammatory SPMs from EPA and DHA. Altered eicosanoid and SPM profiles may have downstream effects because these lipid mediators regulate production of inflammatory cytokines. However, eicosanoid-independent effects of marine omega-3 PUFAs also seem likely (Calder, 2015, 2020b). These PUFAs have been shown to decrease activation of the proinflammatory transcription factor NF-κB and the inflammasome, to activate the anti-inflammatory transcription factor PPAR-γ, and to alter key structural and functional aspects of the plasma membrane that reduce TLR-4-mediated inflammatory signaling. In addition, inflammatory macrophages express a G protein-coupled receptor GPR120, through which DHA has been demonstrated to exert anti-inflammatory actions. As a result of these actions, marine omega-3 PUFAs have been shown to reduce leukocyte chemotaxis, adhesion molecule expression, and production of inflammatory cytokines and to increase production of the anti-inflammatory cytokine IL-10. In experimental studies EPA and DHA are able to counter the pro-inflammatory effects of bacterial lipopolysaccharides and saturated fatty acids.

Fig. 5 provides an overview of the mechanisms of anti-inflammatory action of omega-3 PUFAs.

A number of observational studies demonstrated an inverse association between intake or status of marine omega-3 PUFAs and circulating concentrations of several inflammatory markers including CRP, IL-6 and several soluble adhesion molecules (Yli-Jama et al., 2002; Pischon et al., 2003; Lopez-Garcia et al., 2004b; Ferrucci et al., 2006). Intervention studies with marine omega-3 PUFAs showed reduced production of inflammatory eicosanoids and cytokines by isolated inflammatory cells (see Calder, 2015 and 2020b for references). The clinical impact of the anti-inflammatory and inflammation resolving actions of omega-3 PUFAs have been tested in patients with various inflammatory conditions. The strongest evidence of benefit has been in patents with rheumatoid arthritis when high doses of EPA and DHA are used (Miles and Calder, 2012; Abdulrazaq et al., 2017). Anti-inflammatory actions of plant omega-3 PUFA α-linolenic acid appear to require its conversion to the more biologically active EPA (Baker et al., 2016, 2020).

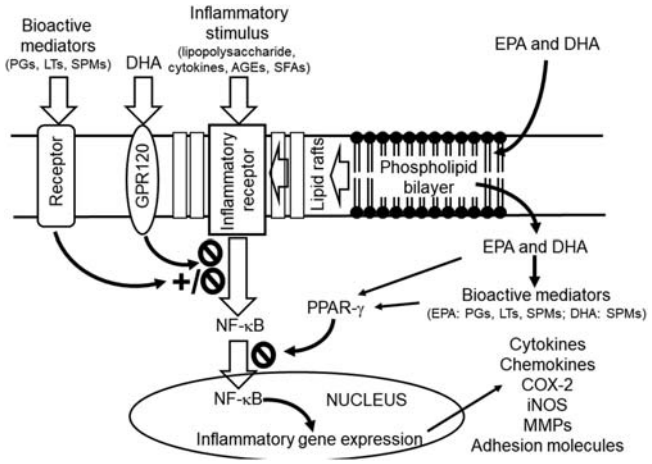


Fig. 5 An overview of the mechanisms of anti-inflammatory action of omega-3 PUFAs. AGEs, advanced glycated end products; COX, cyclooxygenase; DHA, docosahexaenoic acid; GPR, G protein coupled receptor; iNOS, inducible nitric oxide synthase; LTs, leukotrienes; MMPs, matrix metalloproteinases; NF-κB, nuclear factor kappa-light-chain-enhancer of activated B cells; PGs, prostaglandins; PPAR, peroxisome proliferator activated receptor; SFAs, saturated fatty acids; SPMs, specialized pro-resolving mediators; the dark circle with a diagonal line indicates inhibition; + indicates activation.

Antioxidant micronutrients and inflammation

There is a bidirectional interaction between oxidative stress and inflammation. The generation of oxidants (e.g., superoxide radicals, hydrogen peroxide) is part of the host inflammatory response. Oxidants can damage components of host cells. In turn, oxidants and oxidized cell components, acting through transcription factors such as NF-κB, induce inflammation. Thus, dietary components that contribute to oxidative stress (e.g., AGEs and oxidized lipids that result from heating cooking oils to high temperature) could promote inflammatory responses, whereas dietary components that inhibit or quench oxidative stress (various antioxidants) could decrease the strength of inflammatory responses. Antioxidant vitamins act to reduce exposure to oxidants and thereby decrease activation of NF-κB and subsequent production of inflammatory cytokines, eicosanoids, and so on. Inflammatory cells maintain high intracellular concentrations of vitamin C, although these concentrations become depleted during acute activation. In vitro, vitamin E exerts a range of anti-inflammatory actions including decreasing the production of pro-inflammatory cytokines and eicosanoids. Observational studies reported inverse relationships between intake or status of vitamin C, vitamin E and various carotenoids and several markers of inflammation (Kritchevsky et al., 2000; Erlinger et al., 2001; van Herpen-Broekmans et al., 2004; Wannamethee et al., 2006; Wang et al., 2008; Helmersson et al., 2009). Some studies of vitamin C or vitamin E supplements report that they lower CRP and inflammatory cytokines (Devaraj and Jialal, 2000; Block et al., 2004) but findings are not consistent (see Calder et al., 2011 for references).

Polyphenols and inflammation

Polyphenols are secondary metabolites of plants. They include flavanones, flavones, flavanols, and flavonols. In vitro studies suggest that many polyphenols have anti-inflammatory activity through several mechanisms, including decreasing eicosanoid production through inhibition of phospholipase A_2, COX and LOX; inhibition of inducible nitric oxide synthase; and inhibition of inflammatory cytokine production. These effects often seem to involve inhibition of activation of key proinflammatory transcription factors such as NF-κB and activator protein-1. A number of polyphenols have benefits in animal models of inflammatory processes. Observational data in humans relating intake of polyphenols with inflammatory markers are inconsistent (see Calder et al., 2011 for references). Human intervention studies investigating the effect of polyphenols on markers of inflammation have often focused on polyphenol-rich foods. For example, the effects of red wine, chocolate and berries on inflammatory markers are believed to result from their constituent polyphenols (see Calder et al., 2011 for references). However, bioavailability of polyphenols is poor, and circulating concentrations are low and are often much lower than those used in the in vitro experiments that demonstrated strong anti-inflammatory effects. Nevertheless, polyphenols may act to regulate inflammation via effects on the gut microbiota, effects which do not require them to be systemically bioavailable.

The gut microbiota and inflammation

There are believed to be important interactions between the gut microbiota and inflammatory cells present within and beyond the gut wall; these interactions can be chemical or may be by direct cell-to-cell contact. Chemicals involved include short chain fatty acids produced by certain microbes that regulate the activity of inflammatory cells in the gut wall. However, the effects of the gut microbiota extend beyond the gut wall and there is evidence that the microbiota plays a role in defining both local (i.e., gut wall) and systemic inflammatory responses (Samuelson et al., 2015; Verdu et al., 2015). Gut dysbiosis is described in many inflammatory conditions, an observation that is often interpreted to support the role of the microbiota in regulating inflammation. However, reverse causality cannot always be ruled out.

Dietary patterns, specific foods and some dietary components that result in less inflammation often promote a healthy, diverse gut microbiota and it is possible that such patterns/foods/components have dual effects acting both via the gut microbiota and systemically (Wu et al., 2011; Gentile and Weir, 2018; Valdes et al., 2018).

Probiotics are "live microorganisms which when administered in sufficient amounts confer a health benefit to the host" (Hill et al., 2014). Lactic acid-producing bacteria, including lactobacilli and bifidobacteria, are the most commonly used probiotics, although other bacterial genera and some yeasts are also used as probiotics. Several probiotics successfully preserve epithelial barrier function by induction of mucin secretion and maintenance of tight junctions (Ohland and MacNaughton, 2010; Ulluwishewa et al., 2011). This decreases translocation of agents such as bacterial lipopolysaccharides, which can trigger systemic inflammation, from the gut lumen into the bloodstream. Some probiotic bacteria antagonize NF-κB activation in the gut epithelium and thereby decrease production of proinflammatory cytokines, while some interact with gut-residing dendritic cells to induce their maturation and secretion of IL-10, which favors the induction of regulatory T cells that keep inflammation under control (Thomas and Versalovic, 2010). Hence, probiotic bacteria are considered to be anti-inflammatory (Spaiser et al., 2015).

A prebiotic is "a substrate that is selectively utilized by host microorganisms conferring a health benefit" (Gibson et al., 2017). Prebiotics are typically carbohydrates that escape digestion by mammalian enzymes in the small intestine but are metabolized by microbial enzymes in the colon, often to short chain fatty acids. Commonly used prebiotics include fructooligosaccharides and galactooligosaccharides, but it is now recognized that other dietary components including plant polyphenols and marine omega-3 PUFAs have prebiotic actions (Gibson et al., 2017). Prebiotics typically promote growth of lactobacilli and bifidobacteria and hence

they can have effects similar to these probiotic organisms. Some interventions with prebiotic oligosaccharides in humans have been reported to lower the concentrations of inflammatory markers including CRP (Dehghan et al., 2014).

Postprandial inflammation

During the hours after consumption of a meal there is an increase in blood concentrations of a number of inflammatory mediators including cytokines, chemokines and soluble adhesion molecules, an increase in the number of activated monocytes in the circulation and an increase in expression of some PRRs on monocytes including TLR-2 and -4 (Hansen et al., 1997; Burdge and Calder, 2005). These observations suggest that meal components may have a direct pro-inflammatory action soon after eating the meal. One possibility is that food may contain bacterial lipopolysaccharides that translocate across the gut barrier inducing as transient state of inflammation (Erridge et al., 2007). A second possibility is that the inflammation may be linked to oxidative stress induced by meal components (Burdge and Calder, 2005). Whatever the cause, post-prandial inflammation is believed to increase risk of cardiovascular disease (Bell et al., 2008). Whether postprandial inflammation is due to carbohydrates or fats has been studied. Inducing a state of acute hyperglycemia by infusing glucose resulted in increased blood concentrations of TNF-α, IL-6 and IL-18 and these were shown to be causally related as experimentally induced oscillations in glucose concentration were mirrored by oscillations in the concentrations of the three cytokines (Esposito et al., 2002). Providing a bolus oral dose of glucose increased monocyte activation, assessed as NF-κB activation and TNF-α gene expression, after 1, 2 and 3 h (Aljada et al., 2006), although another study saw no effect of either glucose or fructose on blood concentrations of several cytokines, CRP or three soluble adhesion molecules (Mah et al., 2011). Both these studies reported changes in markers of oxidative stress. A high-fat meal but not a high carbohydrate meal, increased blood concentrations of TNF-α, IL-6 and two adhesion molecules, an effect that was prevented by including vitamins E and C in the meal (Nappo et al., 2002), again suggesting that increased oxidative stress is the mechanism behind postprandial inflammation. Another study reported that consumption of a high-fat meal was associated with an increase in IL-18 and a decrease in adiponectin concentration in healthy subjects and in patients with type 2 diabetes (Esposito et al., 2003). Interestingly, IL-18 concentration decreased after consumption of a high-carbohydrate, high-fiber meal (Esposito et al., 2003). In another study, both a high-fat meal or glucose resulted in increased concentrations of soluble adhesion molecules in controls and diabetic subjects and combining the high-fat meal with glucose produced a greater overall increase in these markers (Ceriello et al., 2004). Several studies failed to find an effect of meals high in complex carbohydrates on the postprandial concentrations of soluble adhesion molecules and pro-inflammatory cytokines (Nappo et al., 2002; Esposito et al., 2003). Since glucose intake is associated with inflammation but complex carbohydrate intake is not, the magnitude of the change in pro-inflammatory cytokine and soluble adhesion molecule concentrations may be determined by the glycemic response and the associated dynamic changes to lipid and carbohydrate metabolism following a meal. With regards to the effect of a high fat meal on post-prandial inflammation, there is some evidence that it is saturated fatty acids that are responsible (Alayon et al., 2018; Monfort-Pires et al., 2018).

Dietary inflammatory indexes

The appreciation that chronic low-grade inflammation contributes to risk of cardiovascular, metabolic and cognitive disease and that these diseases are linked to aging and to diet has led to attempts to classify diets according to the extent to which they are likely to promote inflammation. A number of such classification systems have been proposed. The "dietary inflammatory index" (DII) is a literature-derived nutrient-based index based on published associations of 45 dietary components (energy, macronutrients, fiber, micronutrients, plant polyphenols, herbs, spices) which are each weighted (positively for increasing inflammation and negatively for decreasing inflammation) and blood concentrations of six inflammatory biomarkers (CRP, TNF, IL-1β, IL-6, IL-4 and IL-10); a higher score indicates a more pro-inflammatory diet (Shivappa et al., 2014; Hebert et al., 2019). The "dietary inflammation score" (DIS) is a literature-derived food and supplement-based score based on biological plausibility and published associations of intakes of 18 foods or food groups (e.g., leafy greens and cruciferous vegetables, tomatoes, apples and berries, deep yellow or orange vegetables and fruit, legumes, fish, high fat dairy, low fat dairy, nuts, red meat, added sugars, refined grains and starchy vegetables) and micronutrient supplement use which are each weighted (positively for increasing inflammation and negatively for decreasing inflammation) and blood concentrations of four inflammatory biomarkers (CRP, IL-6, IL-8 and IL-10); a higher score indicates a more pro-inflammatory diet (Byrd et al., 2019). The "anti-inflammatory dietary index" (AIDI) is a literature-derived food-based score based on the associations of intakes of 18 foods or food groups (e.g., total fruits and vegetables, whole grain breads, dry fruits, legumes, nuts, unprocessed red meat) which are each weighted (negatively for increasing inflammation and positively for decreasing inflammation) and plasma CRP concentrations; a higher score indicates a more anti-inflammatory diet (Kaluza et al., 2018). The "empirical dietary inflammatory pattern score" (EDIP) is based on regression models predicting the association of intakes of foods and food groups which are each weighted (positively for increasing inflammation and negatively for decreasing inflammation) and blood concentrations of three inflammatory biomarkers (CRP, IL-6, TNF receptor 2); a higher score indicates a more pro-inflammatory diet (Tabung et al., 2016, 2017). EDIP, DIS and AIDI are all based on whole foods, although they are derived differently, while DII is mainly nutrient based. Although high correlations have been reported between DII and DIS in at least three cohorts, there are low-to-moderate correlations between EDIP and DIS and low correlations between EDIP and DII. EDIP and AIDI seem to be similar although they have not been correlated in a single study.

Conclusions and perspectives

Inflammation is a stereotypical physiologic response to infection and tissue injury. It initiates pathogen killing as well as tissue repair processes and helps to restore homeostasis at infected or damaged sites. Acute inflammatory reactions are usually self-limiting and resolve rapidly. Resolution involves the activation of negative feedback mechanisms such as the secretion of immunoregulatory cytokines (e.g., IL-10 and transforming growth factor-β) and pro-resolving lipid mediators (resolvins, protectins and maresins), inhibition of proinflammatory signaling cascades, receptor shedding, and activation of regulatory cells. Inflammatory responses that fail to regulate themselves can become chronic and contribute to the initiation, perpetuation and progression of disease. This has been well recognized for "high grade" chronic inflammation where the inflammatory response is central to the pathology and the disease is treated with anti-inflammatory medications. However, in the last two decades "low grade" chronic inflammation has been recognized to underpin atherosclerosis, cognitive decline and loss of muscle mass and to be linked with obesity playing a causal role in its co-morbidities. Dietary choices are likely to be an important factor in determining the extent of low grade inflammation and this is likely one reason why diet is a risk factor for these diseases.

A healthy eating pattern characterized by consumption of whole grains, nuts and seeds, fruits and vegetables and fish is associated with reduced inflammation, a finding suggesting candidate anti-inflammatory dietary components. Indeed, dietary indexes of inflammation have been defined based upon the known impact of foods or food components on circulating inflammatory markers. These observations indicate that diet plays a role in setting inflammatory tone and in determining risk of at least some diseases that involve inflammation. Among food components, there is good evidence that marine omega-3 PUFAs, antioxidant vitamins, plant polyphenols and those components that beneficially modify the gut microbiota all reduce inflammation (**Fig. 6**). On the other hand, saturated fatty acids, simple sugars, AGEs and oxidized lipids all increase inflammation (**Fig. 6**). Dietary components that enter the circulation act to influence inflammation through a small number of common mechanisms: they may bind to PRRs to trigger inflammation; they promote or mitigate oxidative stress; they act at the membrane of cytosolic level to promote or mitigate inflammatory signaling for example via NF-κB; they act as substrates for synthesis of chemicals involved in the inflammatory response (see **Fig. 4**). The gut microbiota is important in determining host inflammatory tone, at both the gut and systemic levels. Thus, foods and food components that modify the gut microbiota making it more or less diverse/healthy can influence inflammation. The composition of this microbiota can be modified by intake of prebiotics or of probiotics. Both modify inflammation and there is some evidence for improvements in certain inflammatory diseases. However, the effects of probiotics are strain and species dependent (Kekkonen et al., 2008) and this is still an area of some uncertainty.

Three areas stand out as being of particular interest in the link between the diet and inflammation and will remain a focus of research over the next period. The first is the role of the gut microbiota in determining inflammatory tone, the extent to which foods and dietary components influence inflammation via the microbiota and the mechanisms involved. The second is the role of post-prandial inflammation in determining cardiometabolic disease risk, the extent to which foods and dietary components influence post-prandial inflammation and the mechanisms involved. The third is the role of diet on promoting or inhibiting resolution of inflammation: tremendous progress has been made in a relatively short time on identifying novel lipid mediators produced from EPA, DPA and DHA that are involved in resolution of inflammation. However, little is still known about these mediators in the human context (Calder, 2020c) and it is not clear what other diet-related factors could be important in regulating resolution of inflammation.

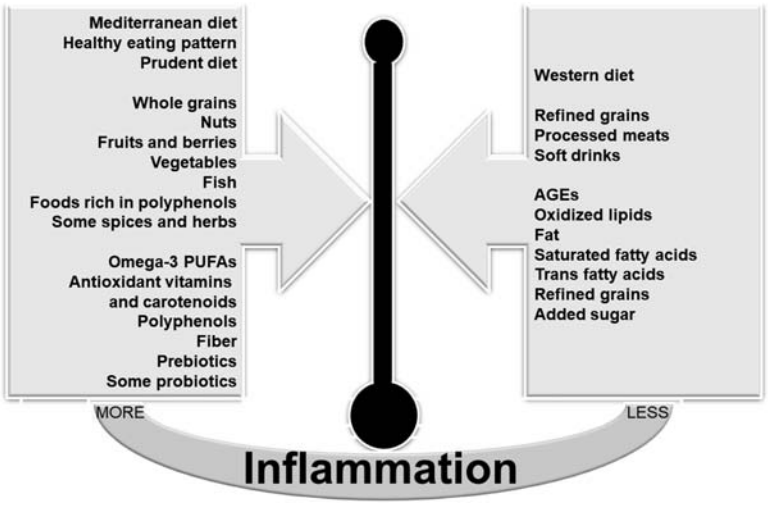


Fig. 6 The comparative effects of dietary patterns, specific foods and individual food components on inflammation. The dietary patterns, foods and food components on the left push the pendulum of inflammation to the right (i.e., toward less inflammation) while the dietary pattern, foods and food components on the right push the pendulum of inflammation to the left (i.e., towards more inflammation). AGEs, advanced glycated end products; PUFAs, polyunsaturated fatty acids.

References

Abdulrazaq, M., Innes, J.K., Calder, P.C., 2017. Effect of ω-3 polyunsaturated fatty acids on arthritic pain: a systematic review. Nutrition 39–40, 57–66.

Alayón, A.N., Rivadeneira, A.P., Herrera, C., et al., 2018. Metabolic and inflammatory postprandial effect of a highly saturated fat meal and its relationship to abdominal obesity. Biomedica 38, 93–100.

Aljada, A., Friedman, J., Ghanim, H., et al., 2006. Glucose ingestion induces an increase in intranuclear nuclear factor kappaB, a fall in cellular inhibitor kappaB, and an increase in tumor necrosis factor alpha messenger RNA by mononuclear cells in healthy human subjects. Metabolism 55, 1177–1185.

Baer, D.J., Judd, J.T., Clevidence, B.A., Tracy, R.P., 2004. Dietary fatty acids affect plasma markers of inflammation in healthy men fed controlled diets: a randomized crossover study. Am. J. Clin. Nutr. 79, 969–973.

Baker, E.J., Miles, E.A., Burdge, G.C., Yaqoob, P., Calder, P.C., 2016. Metabolism and functional effects of plant-derived omega-3 fatty acids in humans. Prog. Lipid Res. 64, 30–56.

Baker, E.J., Valenzuela, C.A., De Souza, C.O., et al., 2020. Comparative anti-inflammatory effects of plant- and marine-derived omega-3 fatty acids explored in an endothelial cell line. Biochim. Biophys. Acta 1865, 158662.

Barnig, C., Bezema, T., Calder, P.C., et al., 2019. Activation of resolution pathways to prevent and fight chronic inflammation: lessons from asthma and inflammatory bowel disease. Front. Immunol. 10, 1699.

Bell, D.S., O'Keefe, J.H., Jellinger, P., 2008. Postprandial dysmetabolism: the missing link between diabetes and cardiovascular events? Endocrinol. Pract. 14, 112–124.

Blevins, H.M., Xu, Y., Biby, S., Zhang, S., 2022. The NLRP3 inflammasome pathway: a review of mechanisms and inhibitors for the treatment of inflammatory diseases. Front. Aging Neurosci. 14, 879021.

Block, G., Jensen, C., Dietrich, M., et al., 2004. Plasma C-reactive protein concentrations in active and passive smokers: influence of antioxidant supplementation. J. Am. Coll. Nutr. 23, 141–147.

Burdge, G.C., Calder, P.C., 2005. Plasma cytokine response during the postprandial period: a potential causal process in vascular disease? Br. J. Nutr. 93, 3–9.

Byrd, D.A., Judd, S.E., Flanders, W.D., et al., 2019. Development and validation of novel dietary and lifestyle inflammation scores. J. Nutr. 149, 2206–2218.

Calder, P.C., 2015. Marine omega-3 fatty acids and inflammatory processes: effects, mechanisms and clinical relevance. Biochim. Biophys. Acta 1851, 469–484.

Calder, P.C., 2020a. Eicosanoids. Essays Biochem. 64, 423–441.

Calder, P.C., 2020b. n-3 PUFA and inflammation: from membrane to nucleus and from bench to bedside. Proc. Nutr. Soc. 79, 404–416.

Calder, P.C., 2020c. Eicosapentaenoic and docosahexaenoic acid derived specialised pro-resolving mediators: concentrations in humans and the effects of age, sex, disease and increased omega-3 fatty acid intake. Biochimie 178, 105–123.

Calder, P.C., Albers, R., Antoine, J.M., et al., 2009. Inflammatory disease processes and interactions with nutrition. Br. J. Nutr. 101 (Suppl. 1), S1–S45.

Calder, P.C., Ahluwalia, N., Brouns, F., et al., 2011. Dietary factors and low-grade inflammation in relation to overweight and obesity. Br. J. Nutr. 106 (Suppl. 3), S5–S78.

Calder, P.C., Ahluwalia, N., Albers, R., et al., 2013. A consideration of biomarkers to be used for evaluation of inflammation in human nutritional studies. Br. J. Nutr. 109 (Suppl. 1), S1–S34.

Calder, P.C., Bosco, N., Bourdet-Sicard, R., et al., 2017. Health relevance of the modification of low grade inflammation in ageing (inflammageing) and the role of nutrition. Ageing Res. Rev. 40, 95–119.

Ceriello, A., Assaloni, R., Da Ros, R., et al., 2004. Effect of atorvastatin and irbesartan, alone and in combination, on postprandial endothelial dysfunction, oxidative stress, and inflammation in type 2 diabetic patients. Circulation 111, 2518–2524.

Chiang, N., Serhan, C.N., 2020. Specialized pro-resolving mediator network: an update on production and actions. Essays Biochem. 64, 443–462.

Christie, W.W., Harwood, J.L., 2020. Oxidation of polyunsaturated fatty acids to produce lipid mediators. Essays Biochem. 64, 401–421.

Chrysohoou, C., Panagiotakos, D.B., Pitsavos, C., Das, U.N., Stefanadis, C., 2004. Adherence to the Mediterranean diet attenuates inflammation and coagulation process in healthy adults: the ATTICA Study. J. Am. Coll. Cardiol. 44, 152–158.

Dai, J., Miller, A.H., Bremner, J.D., et al., 2008. Adherence to the Mediterranean diet is inversely associated with circulating interleukin-6 among middle-aged men: a twin study. Circulation 117, 169–175.

Dehghan, P., Pourghassem Gargari, B., Asghari Jafar-badi, M., 2014. Oligofructose-enriched inulin improves some inflammatory markers and metabolic endotoxemia in women with type 2 diabetes mellitus: a randomized controlled clinical trial. Nutrition 30, 418–423.

Devaraj, S., Jialal, I., 2000. Alpha tocopherol supplementation decreases serum C-reactive protein and monocyte interleukin-6 levels in normal volunteers and type 2 diabetic patients. Free Radic. Biol. Med. 29, 790–792.

di Giuseppe, R., Di Castelnuovo, A., Centritto, F., et al., 2008. Regular consumption of dark chocolate is associated with low serum concentrations of C-reactive protein in a healthy Italian population. J. Nutr. 138, 1939–1945.

Erlinger, T.P., Guallar, E., Miller III, E.R., Stolzenberg-Solomon, R., Appel, L.J., 2001. Relationship between systemic markers of inflammation and serum beta-carotene levels. Arch. Intern. Med. 161, 1903–1908.

Erridge, C., Attina, T., Spickett, C.M., Webb, D.J., 2007. A high-fat meal induces low-grade endotoxemia: evidence of a novel mechanism of postprandial inflammation. Am. J. Clin. Nutr. 86, 1286–1292.

Esposito, K., Nappo, F., Marfella, R., et al., 2002. Inflammatory cytokine concentrations are acutely increased by hyperglycemia in humans: role of oxidative stress. Circulation 106, 2067–2072.

Esposito, K., Nappo, F., Giugliano, F., et al., 2003. Meal modulation of circulating interleukin 18 and adiponectin concentrations in healthy subjects and in patients with type 2 diabetes mellitus. Am. J. Clin. Nutr. 78, 1135–1140.

Esposito, K., Marfella, R., Ciotola, M., et al., 2004. Effect of a mediterranean-style diet on endothelial dysfunction and markers of vascular inflammation in the metabolic syndrome: a randomized trial. JAMA 292, 1440–1446.

Estruch, R., Martínez-González, M.A., Corella, D., et al., 2006. Effects of a Mediterranean-style diet on cardiovascular risk factors: a randomized trial. Ann. Intern. Med. 145, 1–11.

Fargnoli, J.L., Fung, T.T., Olenczuk, D.M., et al., 2008. Adherence to healthy eating patterns is associated with higher circulating total and high-molecular-weight adiponectin and lower resistin concentrations in women from the Nurses' Health Study. Am. J. Clin. Nutr. 88, 1213–1224.

Ferrucci, L., Cherubini, A., Bandinelli, S., et al., 2006. Relationship of plasma polyunsaturated fatty acids to circulating inflammatory markers. J. Clin. Endocrinol. Metab. 91, 439–446.

Ford, E.S., Mokdad, A.H., Liu, S., 2005. Healthy Eating Index and C-reactive protein concentration: findings from the National Health and Nutrition Examination Survey III, 1988–1994. Eur. J. Clin. Nutr. 59, 278–283.

Fung, T.T., McCullough, M.L., Newby, P.K., et al., 2005. Diet-quality scores and plasma concentrations of markers of inflammation and endothelial dysfunction. Am. J. Clin. Nutr. 82, 163–173.

Gentile, C.L., Weir, T.L., 2018. The gut microbiota at the intersection of diet and human health. Science 362, 776–780.

Gibson, G.R., Hutkins, R., Sanders, M.E., et al., 2017. Expert consensus document: the International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. Nat. Rev. Gastroenterol. Hepatol. 14, 491–502.

Hansen, K., Sickelmann, F., Pietrowsky, R., Fehm, H.L., Born, J., 1997. Systemic immune changes following meal intake in humans. Am. J. Physiol. 273, R548–R553.

Hansson, G.K., 2005. Inflammation, atherosclerosis, and coronary artery disease. N. Engl. J. Med. 352, 1685–1695.

Hébert, J.R., Shivappa, N., Wirth, M.D., Hussey, J.R., Hurley, T.G., 2019. Perspective: the Dietary Inflammatory Index (DII)-lessons learned, improvements made, and future directions. Adv. Nutr. 10, 185–195.
Hedl, M., Abraham, C., 2013. Negative regulation of human mononuclear phagocyte function. Mucosal Immunol. 6, 205–223.
Helmersson, J., Arnlöv, J., Larsson, A., Basu, S., 2009. Low dietary intake of beta-carotene, alpha-tocopherol and ascorbic acid is associated with increased inflammatory and oxidative stress status in a Swedish cohort. Br. J. Nutr. 101, 1775–1782.
Hill, C., Guarer, F., Reid, G., et al., 2014. Expert consensus document. The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. Nat. Rev. Gastroenterol. Hepatol. 11, 506–514.
Hwang, D.H., Kim, J.A., Lee, J.Y., 2016. Mechanisms for the activation of Toll-like receptor 2/4 by saturated fatty acids and inhibition by docosahexaenoic acid. Eur. J. Pharmacol. 785, 24–35.
Innes, J.K., Calder, P.C., 2018. Omega-6 fatty acids and inflammation. Prostaglandins Leukot. Essent. Fatty Acids 132, 41–48.
Iverson, R., Sollid, L.M., 2022. The immunobiology and pathogenesis of celiac disease. Annu. Rev. Pathol. https://doi.org/10.1146/annurev-pathmechdis-031521-032634, 2022, in press.
Kaluza, J., Harris, H., Melhus, H., Michaëlsson, K., Wolk, A., 2018. Questionnaire-based anti-inflammatory diet index as a predictor of low-grade systemic inflammation. Antioxidants Redox Signal. 28, 78–84.
Kekkonen, R.A., Lummela, N., Karjalainen, H., et al., 2008. Probiotic intervention has strain-specific anti-inflammatory effects in healthy adults. World J. Gastroenterol. 14, 2029–2036.
Kritchevsky, S.B., Bush, A.J., Pahor, M., Gross, M.D., 2000. Serum carotenoids and markers of inflammation in nonsmokers. Am. J. Epidemiol. 152, 1065–1071.
Kumar, A., Takada, Y., Boriek, A.M., Aggarwal, B.B., 2004. Nuclear factor-kappaB: its role in health and disease. J. Mol. Med. 82, 434–448.
Libby, P., Ridker, P.M., 2004. Inflammation and atherosclerosis: role of C-reactive protein in risk assessment. Am. J. Med. 116 (Suppl. 6A), 9S–16S.
Lopez-Garcia, E., Schulze, M.B., Fung, T.T., et al., 2004a. Major dietary patterns are related to plasma concentrations of markers of inflammation and endothelial dysfunction. Am. J. Clin. Nutr. 80, 1029–1035.
Lopez-Garcia, E., Schulze, M.B., Manson, J.E., et al., 2004b. Consumption of (n-3) fatty acids is related to plasma biomarkers of inflammation and endothelial activation in women. J. Nutr. 134, 1806–1811.
Lopez-Garcia, E., Schulze, M.B., Meigs, J.B., et al., 2005. Consumption of trans fatty acids is related to plasma biomarkers of inflammation and endothelial dysfunction. J. Nutr. 35, 562–566.
Mah, E., Noh, S.K., Ballard, K.D., et al., 2011. Postprandial hyperglycemia impairs vascular endothelial function in healthy men by inducing lipid peroxidation and increasing asymmetric dimethylarginine:arginine. J. Nutr. 141, 1961–1968.
Miles, E.A., Calder, P.C., 2012. Influence of marine n-3 polyunsaturated fatty acids on immune function and a systematic review of their effects on clinical outcomes in rheumatoid arthritis. Br. J. Nutr. 107 (Suppl. 2), S171–S184.
Monfort-Pires, M., Crisma, A.R., Bordin, S., Ferreira, S.R.G., 2018. Greater expression of postprandial inflammatory genes in humans after intervention with saturated when compared to unsaturated fatty acids. Eur. J. Nutr. 57, 2887–2895.
Nappo, F., Esposito, K., Cioffi, M., et al., 2002. Postprandial endothelial activation in healthy subjects and in type 2 diabetic patients: role of fat and carbohydrate meals. J. Am. Coll. Cardiol. 39, 1145–1150.
Ohland, C.L., MacNaughton, W.K., 2010. Probiotic bacteria and intestinal epithelial barrier function. Am. J. Physiol. 298, G807–G819.
Panigrahy, D., Gilligan, M.M., Serhan, C.N., Kashfi, K., 2021. Resolution of inflammation: an organizing principle in biology and medicine. Pharmacol. Ther. 227, 107879.
Pischon, T., Hankinson, S.E., Hotamisligil, G.S., et al., 2003. Habitual dietary intake of n-3 and n-6 fatty acids in relation to inflammatory markers among US men and women. Circulation 108, 155–160.
Purschwitz, K., Rassoul, F., Reuter, W., et al., 2001. [Soluble leukocyte adhesion molecules in vegetarians of various ages]. Zeitschrift für Gerontologie und Geriatrie 34, 476–479.
Samuelson, D.R., Welsh, D.A., Shellito, J.E., 2015. Regulation of lung immunity and host defense by the intestinal microbiota. Front. Microbiol. 6, 1085.
Schulze, M.B., Hoffmann, K., Manson, J.E., et al., 2005. Dietary pattern, inflammation, and incidence of type 2 diabetes in women. Am. J. Clin. Nutr. 82, 675–684.
Schwingshackl, L., Hoffmann, G., 2014. Mediterranean dietary pattern, inflammation and endothelial function: a systematic review and meta-analysis of intervention trials. Nutr. Metab. Cardiovasc. Dis. 24, 929–939.
Shivappa, N., Steck, S.E., Hurley, T.G., Hussey, J.R., Hébert, J.R., 2014. Designing and developing a literature-derived, population-based dietary inflammatory index. Publ. Health Nutr. 17, 1689–1696.
Spaiser, S.J., Culpepper, T., Nieves Jr., C., et al., 2015. Lactobacillus gasseri KS-13, Bifidobacterium bifidum G9-1, and Bifidobacterium longum MM-2 ingestion induces a less inflammatory cytokine profile and a potentially beneficial shift in gut microbiota in older adults: a randomized, double-blind, placebo-controlled, crossover study. J. Am. Coll. Nutr. 34, 459–469.
Szeto, Y.T., Kwok, T.C., Benzie, I.F., 2004. Effects of a long-term vegetarian diet on biomarkers of antioxidant status and cardiovascular disease risk. Nutrition 20, 863–866.
Tabung, F.K., Smith-Warner, S.A., Chavarro, J.E., et al., 2016. Development and validation of an empirical dietary inflammatory index. J. Nutr. 146, 1560–1570.
Tabung, F.K., Smith-Warner, S.A., Chavarro, J.E., et al., 2017. An empirical dietary inflammatory pattern score enhances prediction of circulating inflammatory biomarkers in adults. J. Nutr. 147, 1567–1577.
Tedner, S.G., Asarnoj, A., Thulin, H., et al., 2022. Food allergy and hypersensitivity reactions in children and adults-a review. J. Intern. Med. 291, 283–302.
Thomas, C.M., Versalovic, J., 2010. Probiotics-host communication: modulation of signaling pathways in the intestine. Gut Microb. 1, 148–163.
Tilg, H., Moschen, A.R., 2006. Adipocytokines: mediators linking adipose tissue, inflammation and immunity. Nat. Rev. Immunol. 6, 772–783.
Twarda-Clapa, A., Olczak, A., Białkowska, A.M., Koziołkiewicz, M., 2022. Advanced glycation end-products (AGEs): formation, chemistry, classification, receptors, and diseases related to AGEs. Cells 11, 1312.
Ulluwishewa, D., Anderson, R.C., McNabb, W.C., et al., 2011. Regulation of tight junction permeability by intestinal bacteria and dietary components. J. Nutr. 141, 769–776.
Valdes, A.M., Walter, J., Segal, E., Spector, T.D., 2018. Role of the gut microbiota in nutrition and health. BMJ 361, k2179.
Valenzuela, C.A., Baker, E.J., Miles, E.A., Calder, P.C., 2019. Eighteen-carbon trans fatty acids and inflammation in the context of atherosclerosis. Prog. Lipid Res. 76, 101009.
van Herpen-Broekmans, W.M., Klöpping-Ketelaars, I.A., Bots, M.L., et al., 2004. Serum carotenoids and vitamins in relation to markers of endothelial function and inflammation. Eur. J. Epidemiol. 19, 915–921.
Verdu, E.F., Galipeau, H.J., Jabri, B., 2015. Novel players in coeliac disease pathogenesis: role of the gut microbiota. Nat. Rev. Gastroenterol. Hepatol. 12, 497–506.
Wang, L., Gaziano, J.M., Norkus, E.P., et al., 2008. Associations of plasma carotenoids with risk factors and biomarkers related to cardiovascular disease in middle-aged and older women. Am. J. Clin. Nutr. 88, 747–754.
Wang, S., Dougherty, E.J., Danner, R.L., 2016. PPARγ signaling and emerging opportunities for improved therapeutics. Pharmacol. Res. 111, 76–85.
Wannamethee, S.G., Lowe, G.D., Rumley, A., Bruckdorfer, K.R., Whincup, P.H., 2006. Associations of vitamin C status, fruit and vegetable intakes, and markers of inflammation and hemostasis. Am. J. Clin. Nutr. 83, 567–574.
Wu, G.D., Chen, J., Hoffmann, C., et al., 2011. Linking long-term dietary patterns with gut microbial enterotypes. Science 334, 105–108.
Yli-Jama, P., Seljeflot, I., Meyer, H.E., et al., 2002. Serum non-esterified very long-chain PUFA are associated with markers of endothelial dysfunction. Atherosclerosis 164, 275–281.

Drug-nutrition interactions

Joseph I. Boullata, Clinical Nutrition, Penn Medicine, Clinical Nutrition Support Services, Hospital of the University of Pennsylvania, Philadelphia, PA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of K.G. Conner, Drug–Nutrient Interactions, In Encyclopedia of Human Nutrition (Third Edition), edited by Benjamin Caballero, Academic Press, 2013, https://doi.org/10.1016/B978-0-12-375083-9.00080-5.

Introduction 79
Definitions, recognition and clinical relevance of interactions 79
Mechanisms 80
Food–meal 82
Food—diet component 83
Nutritional status on drug disposition 85
Drug-induced alteration in nutrition 86
Metabolic status 86
Overall nutritional status 86
Specific nutrients 87
Summary 88
References 88

Key points

- Interactions between drugs and nutrition occur and may complicate patient care.
- A meal, specific food or diet component, or nutritional status may influence the physiologic disposition and clinical effect of a drug.
- A medication may influence metabolic profile, overall nutritional status, or status of an individual nutrient.
- The presence of a potential interaction requires close monitoring and management as needed.

Introduction

Medication use is a part of life for many people who make regular use of prescription drugs, non-prescription drugs, and supplements in pharmaceutical form. The oral bioavailability and clinical effect of these drugs are influenced by a variety of factors which together can determine outcome. Although medication-related fatalities occur, less severe outcomes of medication use predominate. These adverse effects include unexpected reactions and drug interactions. In some cases, the adverse reactions and the interactions overlap with or are directly related to nutrition.

Although drug-drug interactions are commonly recognized by clinicians, drug-nutrition interactions are seldom as widely familiar to clinicians, but no less important in their ability to influence patient outcome. Evaluation of a drug may identify multiple influences on nutrition. In turn nutrition plays a role in drug disposition and effect. Whether interactions are positive or negative, they need to be better recognized, understood, predicted and managed as necessary.

This article will define drug-nutrition interactions, explain the different subtypes of interaction and describe their mechanisms using clinical examples. The goal is to provide an appreciation of potential interactions, using some examples, with a focus on mechanistic characteristics rather than attempt exhaustive lists.

Definitions, recognition and clinical relevance of interactions

Clinicians widely appreciate the influence of interactions between medications on a drug's disposition and clinical effect. Drug interactions with nutrition are no less important than drug-drug interactions (Boullata, 2013). Drug-nutrition interactions result from a physical, chemical, physiological or pathophysiological relationship between a medication and one or more nutrients, food in general, specific foods or food components, nutritional status or metabolic profile (Santos and Boullata, 2005; Boullata, 2013; Boullata and Hudson, 2012; Chan, 2013). Therefore, drug-nutrition interactions can influence the anticipated effect of the drug as well as the patient's nutritional status (Gezmen-Karadağ et al., 2018). As with any interaction, one element of the relationship is considered the "perpetrator" (precipitating factor) while the other is the "victim" (object) of the interaction. The precipitating factor may be the drug, or a nutrient, food or nutritional status, and then another component serves as the targeted object of the interaction.

Table 1 Types of drug-nutrition interaction.

Precipitating factor	*Object of the interaction*	*Interaction term*
Meal	Drug	Food-drug
Diet component[a]	Drug	Diet-drug, nutrient-drug
Nutritional status	Drug	Nutritional status-drug
Drug	Nutritional status	Drug-nutrition (status)
Drug	Metabolic profile	Drug-metabolic (profile)
Drug	Nutrient	Drug-nutrient

[a]Includes those found in dietary supplement products.

Therefore, several subtypes of drug-nutrition interaction exist. They are individually described by their precipitant-and-object as "food-drug" interactions or "drug-nutrient" interactions, among others (**Table 1**) (Boullata, 2013; Boullata and Hudson, 2012). When classified into subtypes, the systematic approach to identifying, recognizing, and understanding the many interactions is more inclusive and allows evolving interactions to be classified as data become available (Boullata and Hudson, 2012; Boullata, 2013). Despite often being overlooked, drug-nutrition interactions are considered of significant public health relevance (Péter et al., 2017).

Some of the classic drug-nutrition interactions described decades ago are often recognized in isolation rather than within the context of a classification scheme. Examples include the influence of isoniazid on vitamin B_6 metabolism (Biehl and Vilter, 1954) and the influence of iron on tetracycline absorption (Neovonen et al., 1970). Additionally described have been the impact of malnutrition on drug metabolism (Krishnaswamy, 1978) and the influence of drugs on nutrient disposition (Roe, 1985). The effects of food on drug absorption as a result of intraluminal interactions were often the most clinically recognized of the drug-nutrition interactions (Welling, 1977). These food-drug interactions are the most likely interactions to appear in regulator-approved drug product labeling. Even so, using the definition above, many drug products have inconsistent or incomplete drug-nutrition interaction information in their product summaries (Nikolić, 2020). Many electronic medical record systems may have basic built-in alerts to help prescribers and pharmacists identify drug-nutrition interactions in healthcare settings. Unfortunately, the content and capability remain quite limited and do not address much of what is described in this review. Despite their important patient counseling role, community pharmacists' knowledge about common food-drug interactions was inadequate, requiring greater awareness and education (Radwan et al., 2018). Clinicians continue to rely on summaries of the primary literature in review articles and reference texts.

Recognition of drug-nutrition interactions continues to grow as more comprehensive approaches to systematically identify and classify interactions, describe their mechanisms, and define clinical relevance evolve (Santos and Boullata, 2005; Boullata and Hudson, 2012; Boullata, 2013; Chan, 2002; Mason, 2010; Boullata and Armenti, 2010). For example, a systematic protocol aimed to identify knowledge gaps in and synthesize and evaluate data on clinical characteristics and safety issues for drug-nutrition interactions focused on food-drug interactions (Orellana-Paucar and Vintimilla-Rojas, 2020). The research community continues systematic evaluations using computer modeling, cell culture studies, and literature review to help further build the database. For example, an analysis using high-throughput screening and *in silico* modeling suggested a number of drugs that can inhibit the thiamin transporter (ThTR-2) (Vora et al., 2020). Interestingly, thiamin is also a substrate for other uptake transporters (e.g., OCT1) whose activity is further modulated by genotype. But given the yet unknown gene-protein influences on thiamin disposition itself, the clinical implications for drug interaction at the less specific transporters are not yet known (Jensen et al., 2020). The broad topic of drug-nutrition interactions has been covered in more depth and in the context of specific populations elsewhere (Boullata, 2013; Boullata and Hudson, 2012; Boullata and Armenti, 2010; Boullata, 2010, 2019a, 2021a,b).

Regardless of the object or precipitant, for a drug-nutrition interaction to be considered clinically significant, there is an expectation that nutritional status is compromised and/or therapeutic drug response is modified (Boullata and Hudson, 2012). The change may be recognized by a 20% or greater shift in biomarkers, physiologic indicators, or kinetic parameters from a baseline value or an anticipated effect. The time frame over which the change occurs will vary with the precipitating factor and object. The severity of consequences may vary, with some individuals at higher risk based on their age, genetic variants, organ function, or disease state. As a result, the clinical significance or severity of an interaction may be difficult to predict in the absence of biomarkers for individual predisposition.

The management of an interaction may only require close clinical monitoring in some instances but may require significant changes in eating patterns, nutrient intake or pharmacotherapeutic regimens in others. Clinicians with sharp clinical awareness can recognize or predict drug-nutrition interactions, assess their clinical relevance, and manage them, with a basic understanding of pharmacology, and an appreciation for the mechanisms of interactions. When known, their cellular mechanisms can be intricate.

Mechanisms

Drug-nutrition interactions essentially occur as a result of two different constructs: (1) the inherent physicochemical properties of drugs, nutrients, and the food matrix, and (2) similarities in the physiologic disposition[1] and physiologic effect of drugs and nutrients.

[1]Disposition refers to the absorption, distribution, metabolism and excretion of a drug or nutrient.

The physical chemistry that is operational for each substance governs their interactions when combined. This occurs in the gut lumen, but also may occur in mealtime dishware or in a nutrition support container prior to consumption/administration. Important properties include the solubility and permeability of a substance, and the viscosity, osmolality, and macronutrient content of a meal (Boullata, 2021b). Physiologically, the processes involved in substrate transport (absorption, distribution, excretion) and metabolism are shared by both drugs and nutrients, at a systemic and cellular level, so an interaction risk always exists (Semen et al., 2020).

Substrate absorption, distribution, and excretion require transporters; metabolism requires one or more enzyme systems; and therapeutic effect requires cellular or molecular targets. These transporters, enzymes, and most targets are proteins coded for by genes. Transporters and enzymes represent products of some of the most highly expressed genes in key tissues (e.g., intestine, liver, kidney). The proteins nearly always require micronutrients for optimal function. The expression of the genes themselves comes under the influence of compounds including nutrients and drugs. Genetic variability and the influence of disease states also need to be considered. Therefore, systems biology models and approaches that incorporate nutrient roles may more efficiently predict potential for interaction than traditional biomarker-based studies. This concept is beginning to evolve within the framework of quantitative systems biology-pharmacology to more closely analyze "-omics" data reflecting broader physiologic roles of these transporters and enzymes including nutrient interaction (Nigam et al., 2020). This will eventually be adapted to better predict drug-nutrition interactions.

The current model of drug-nutrition interactions links an interaction with its physiologic effect and clinical outcome, through differentiating mechanisms (**Table 2**) (Boullata and Hudson, 2012). Some interactions are based on physicochemical reactions that take place in the lumen of the gastrointestinal tract or ex vivo in a mealtime or nutrition support delivery device. These interactions, referred to as *pharmaceutical* interactions, have the distinct potential to alter the bioavailability of one or more substances (nutrient or drug). An interaction's relevance may be interpreted based on treatment goal. For example, the antiparasitic albendazole must be taken with food in order to create adequate dissolution to support drug absorption and achieve therapeutic blood concentrations (Lange et al., 1988). However, if the therapy with albendazole is only intended for local gut action on helminths, then administration on an empty stomach may enhance that local effectiveness (Humphries et al., 2017). In an example using neurologic agents, the addition of a dose of carbamazepine or escitalopram to an enteral nutrition formula would result in a physicochemical interaction that clogs a feeding tube, thereby limiting delivery and absorption of both drug and nutrients (Klang et al., 2013).

Interactions exist beyond the immediate physicochemical reactions of temporal combinations. Many interactions are the result of encounters at the level of cell membrane transporters and/or metabolizing enzymes, referred to as *pharmacokinetic* interactions. An interaction is based on direct competition for a transporter or enzyme, or an indirect influence often through nuclear transcription factors (e.g., PXR) on protein expression or activity. This is mechanistically similar whether transporters or enzymes are involved. With regard to transporters, short-term/acute regulation involves direct interaction of an agent with the transporter via a ligand-binding site. It may be competitive (via the orthosteric binding site) decreasing substrate transport, or non-competitive (via an allosteric non-overlapping, topographically distinct binding site) that may increase or decrease substrate transport (Domínguez et al., 2021). However long-term regulation involves indirect interactions through a change in transporter expression or modulation through a post-transcription, translation or post-translation level (Rigalli et al., 2016). For example, oleic acid can influence several efflux transporters to decrease P-glycoprotein (P-gp) activity (short-term) and P-gp and breast cancer-resistance protein (BCRP) expression (longer) but increase multidrug resistance-associated protein (MRP2) activity through translocation from intracellular compartments (Domínguez et al., 2021). Potential interactions can occur with drugs that are substrates for any of those transporters.

Table 2 Mechanisms of drug-nutrition interaction.

Mechanism type	*Description*	*Examples (as cited within the article)*
Pharmaceutical	Physicochemical reaction with temporal combination in the gut lumen or nutrition support delivery device	Meal increases (e.g., carbamazepine) or decreases (e.g., indinavir) drug bioavailability; Drug (e.g., ciprofloxacin) chelates nutrient (e.g., calcium) from meal or enteral nutrition formula, reducing bioavailability of both
Pharmacokinetic	Direct competition for transporter or metabolizing enzyme;	Beverage (e.g., grapefruit juice) may increase (e.g., atorvastatin) or decrease (e.g., levothyroxine) drug bioavailability;
	Indirect influence on expression or activity of transporter or enzyme;	Drug (e.g., carbamazepine) may decrease nutrient absorption (e.g., biotin) or increase nutrient metabolism (e.g., vitamin D);
	Obesity and malnutrition alter drug distribution and clearance	Nutritional status (e.g., protein-energy malnutrition) increases bioavailability and toxicity risk of a drug (e.g., clarithromycin)
Pharmacodynamic	Therapeutic interference or synergistic augmentation directly at physiologic target site or through indirect targets (e.g., signaling molecules) to alter physiology	Excessive nutrient will interfere with (e.g., vitamin K) or exacerbate (e.g., ubiquinone) the anticoagulant action of warfarin; Drug use (e.g., azole antifungals) may reduce food intake by several mechanisms (gastrointestinal complaints, fatigue, arthralgia, myalgia, tremor)

Although less commonly recognized, pharmacokinetic interactions can also occur through the influence of nutritional status (i.e., body composition variability) on kinetic parameters. For example, systemic drug exposure and toxicity may be increased in protein-energy malnutrition (PEM) as the result of altered drug distribution or clearance of the antimicrobial clarithromycin (Ahn et al., 2003). The potential consequences of all pharmacokinetic interactions are altered bioavailability, distribution, and clearance. The latter can occur through metabolism and/or excretion. This parallels pharmacokinetic drug-drug interactions whose mechanism involves drug transporters alone or along with metabolizing enzymes (Yu et al., 2019).

Still other interactions may take place at target receptors that influence cell signaling, referred to as *pharmacodynamic* interactions, which yield an effect (direct or indirect) on physiologic function, perhaps including nutrient sensing pathways. The potential consequences are altered cellular responses that may translate to a biomarker change or clinical manifestation. Many medications can affect physiologic function to influence food intake. For example, azole antifungal agents may be associated with fatigue, arthralgia, myalgia, tremor, and gastrointestinal complaints (e.g., dry mouth, anorexia, nausea, vomiting, diarrhea, abdominal pain), with potential to decrease food intake and body weight with longer courses of therapy (Boullata, 2021a). Whenever an interaction is sufficient to alter food intake, drug or nutrient disposition, or their clinical effects, it may be considered clinically significant.

Food–meal

Food is well established for influencing oral drug absorption and bioavailability (Welling, 1977; Boullata and Armenti, 2010; Deng et al., 2017). The potential impact of food on the absorption of a medication is an interaction that all new drug applications to the US Food and Drug Administration (FDA) are required to describe. The potential interaction can occur for a variety of physicochemical and physiologic reasons. This "meal-effect" occurs because the presence of food changes the conditions within the gut lumen into which a drug is administered. Following a meal there are changes in gut pH, viscosity, volume, gastric emptying rates, bile flow and pancreatic secretion that influence drug dissolution, as well as changes to enterocyte permeability, transport, and metabolism. The influences of a meal will be different depending on individual drug properties (e.g., pKa, solubility). There is additionally physiologic variability in drug dissolution and absorption based on the consistency and composition of the meal (Rubbens et al., 2019). Whether the addition of a drug diluted in water to the fed stomach can evade the chyme bolus and be emptied more rapidly is of interest, as this has been noted to differ between orally administered drugs (Schick et al., 2019).

Depending on the drug, altered conditions in the gut may influence the *rate* of a drug's absorption or the *extent* of drug absorption (i.e., bioavailability). Bioavailability is more clinically relevant than absorption rate and is evaluated by examining the area under the serum drug concentration-time curve (AUC) in the fed state compared to the fasted state. For example, although there is a significant decrease in the rate of absorption for pregabalin when taken with a meal, there is no clinically significant difference in total absorption (i.e., bioavailability), so the drug may be taken without regard to food (Boullata, 2019b).

Well-designed meal-effect studies are useful in recognizing food-drug interactions and designing management strategies. Although not always used in trials, the FDA recommends a "test meal" of ~800–1000 kcal containing ~50% of energy from fat. Clinical significance is noted if the AUC_{fed}-to-AUC_{fasted} ratio is <0.8 or >1.25. In the absence of clinical data, a drug's known physicochemical properties (and susceptibility to removal) can be used to predict drug disposition with a meal. Although complex physicochemical mechanisms are in play, generally, drugs with low solubility but high permeability and significant gut mucosal clearance (e.g., carbamazepine), are expected to have higher bioavailability in the presence of food than in the fasted state, whereas the absorption of drugs with low permeability (e.g., indinavir) is often impaired by food, and those with high solubility and permeability are unaffected (Benet, 2013; Varma et al., 2015; Boullata, 2019b, 2021b). Thus, knowing a drug's solubility and permeability characteristics may help predict its interaction with food.

Clinical studies can then provide data on the extent to which the change in bioavailability is clinically significant (e.g., AUC change of 20%–25%). Although the bioavailability of one drug may be increased (positive effect) or another drug may be decreased (negative effect) in the presence of food, some drugs exhibit no significant meal-effect and can therefore be administered without regard to a meal. To make an informed evidence-based clinical recommendation, it is important to note what test meal conditions have been used in a study as results may vary with the meal. Varied meal types may have different influences on gut function and therefore drug absorption (Deng et al., 2017). This potential complexity of the meal-effect can be seen with ziprasidone. The drug's bioavailability is enhanced using the standard high-calorie/high-fat meal, but the effect is seen to a greater extent when a high-calorie/low-fat meal is used but to a lower extent with a low-calorie/high-fat meal (Lincoln et al., 2010; Sutton et al., 2017). Studies that examine the effects of fed vs fasted state on drug bioavailability can also expose significant interpatient variability (Kim et al., 2014).

The alteration in drug bioavailability is especially important for antimicrobials or drugs with a narrow therapeutic index (NTI).[2] Considering antimicrobials, a meal may increase (e.g., cefpodoxime) or decrease (e.g., ampicillin, penicillin, norfloxacin) the extent of drug absorption (Hughes et al., 1989; Eshelman and Spyker, 1978; Minami et al., 1993; Bolme et al., 1995). For the latter agents, administering the drug at least 1 h before or 2 h after food is recommended to improve bioavailability. Sometimes the formulation of a drug may make a difference. For example, azithromycin capsules and tablets are considered bioequivalent in the fasted state;

[2]NTI—drugs for which slight decreases (or increases) in dose or blood concentrations can lead to dose/concentration-dependent therapeutic failure (or toxicity); includes medications whose blood concentrations or biomarker effects are routinely monitored (e.g., carbamazepine, cyclosporine, digoxin, levothyroxine, lithium, phenytoin, tacrolimus, warfarin).

however, a significant meal-effect is noted only with the capsules resulting in decreased bioavailability compared with the tablets (Curatolo et al., 2011). Food may reduce the bioavailability of the antiepileptic phenytoin—at least for some brands of extended-release products (Wilder et al., 2001). Of note, not taking into account the reduced drug bioavailability with a meal may lead to subtherapeutic drug concentrations, risking therapeutic failure and drug resistance in the case of antimicrobials.

The meal-effect may have different outcomes even on agents within the same drug class. For example, the antifungal posaconazole should be administered with food, based on a 3-fold increase in bioavailability regardless of caloric density compared with the fasted state (Lin et al., 2013). While a similar antifungal, voriconazole, needs to be administered on an empty stomach to optimize oral bioavailability (Courtney et al., 2004; Purkins et al., 2003). Mixed findings are also reported for agents to treat tuberculosis where food may reduce the oral bioavailability of isoniazid and pyrazinamide, but not likely of rifampin, so that administration in the fasted state is suggested (Kumar et al., 2017). The administration of a four-component (elvitegravir, cobicistat, emtricitabine, tenofovir) fixed-dose antiviral tablet to healthy subjects in crossover studies under both fed and fasted states revealed drug-specific effects (Shiomi et al., 2014; Yamada et al., 2018). Although bioavailability of cobicistat, emtricitabine, and tenofovir from this product were similar in the fasted, standard breakfast fed and enteral nutrition formula fed states, the bioavailability of elvitegravir was significantly greater in the fed state regardless of nutrient content. As a result, this combination antiviral product is best administered with a meal.

The macronutrient content of a meal may make a difference on the effect. For example, the fat component (percent energy) may have significant influence on oral bioavailability of a drug, as with the antifungal drug griseofulvin. The oral bioavailability increases between 35% and 120% with a meal depending specifically on fat content, compared to its administration in the fasted state (Ogunbona et al., 1985). Bioavailability of ivermectin increased ~2.5-fold when administered with a meal, especially with a high-fat meal (Guzzo et al., 2002; Raman and Polli, 2016) which may be related to altered post-absorptive distribution as well as improved gut dissolution following a high-fat meal (Miyajima et al., 2015). However, administration of ivermectin with a grain-based meal did not alter bioavailability from that in the fasted state (Homeida et al., 2013). Even the content of pectin in a meal may alter drug absorption usually through an influence on permeability in part by bile salt modification (Elshahed et al., 2021).

Beyond test-meal conditions, therapeutic diets prescribed in the clinical setting may influence drug disposition. For example, a modified Atkins diet (78% of energy from fat) has been reported to significantly reduce serum concentrations over time of several antiepileptic drugs (carbamazepine, lamotrigine, topiramate, valproic acid) requiring close therapeutic monitoring (Kvernelend et al., 2015). Interestingly, in the case of irinotecan, dietary protein and energy restriction for just a few days may significantly increase the systemic exposure to the drug's active metabolite (de Man et al., 2021).

There is occasionally an interest in mixing powdered medication directly in with food to facilitate administration in patients unable to swallow intact pills. This needs to be carefully considered. For example, crushing and mixing carbamazepine in yogurt slowed drug dissolution, compared with crushing and mixing in water, which then directly influences bioavailability (Manrique et al., 2014). When ciprofloxacin is crushed and mixed in water, it is much more chemically stable compared with mixing it with any of several common foods and beverages or with enteral nutrition formula (Sadrieh et al., 2005; Wright et al., 2000). The influence of dairy products in reducing the bioavailability of some fluoroquinolones (e.g., ciprofloxacin) by at least 30% is the result of chelation with calcium in the intestinal lumen (Neuvonen et al., 1991).

Food—diet component

Distinct from the meal-effect, specific dietary components, including those available in dietary supplement products, may also influence drug disposition. The oral bioavailability of medications that are substrates for transporters (uptake or efflux) and/or metabolizing enzymes at the intestinal epithelium, lend themselves to interactions with dietary components that may directly or indirectly influence expression or activity of those proteins (Nakanishi and Tamai, 2015). Uptake transporters include the peptide transporters (e.g., PEPT1) and organic anion transporting polypeptides (e.g., OATP2B1) which both have broad substrate specificity at luminal brush-border membranes, with organic cation transporters (e.g., OCT1) at the basolateral membrane, and efflux transporters P-gp and BCRP, while common metabolizing enzymes include phase 1 cytochrome P450 (e.g., CYP3A) and phase 2 conjugation enzymes (e.g., UDP-glucuronosyl-transferases [UGTs], glutathione-S-transferases [GSTs]) (Wu and Lin, 2019). An individual's susceptibility may in part also depend on the transporter and enzyme genotype.

Administration of oseltamivir, the influenza treatment, with milk reduces initial drug bioavailability by 35% compared to administration with water, possibly the result of competition at the PEPT1 transporter for absorption (Morimoto et al., 2011). Dietary garlic intake at up to 15 cloves per week may reduce protease inhibitor (e.g., darunavir) concentrations enough that HIV plasma viral load increases (Cloarec et al., 2017). This is most likely associated with an increase in garlic-induced expression of gut P-gp efflux reducing drug bioavailability. Cruciferous vegetable intake containing isothiocyanates can within a week significantly induce the activity of CYP1A2, UGT1A1 and GSTα responsible for metabolizing a number of drugs (Eagles et al., 2020). Other compounds such as genistein found in soy-based foods or in dietary supplement products, interact with several efflux transporters to modulate their expression and activity (Rigalli et al., 2016). Consistent with the interesting co-localization and apparent coordination between gut metabolizing enzymes and efflux pumps, genistein may activate PXR to induce CYP3A4 as well as efflux transport (e.g., P-gp, MRP2) by increasing expression and activity at typical concentrations.

Box 1 The grapefruit interaction

Juice components can influence drug-metabolizing enzymes and drug transporters to increase bioavailability of some drugs and reduce bioavailability of others (Ando et al., 2005; Dolton et al., 2012; Bailey et al., 2013; Hyewon et al., 2013; Fleisher et al., 2015). This seeming dichotomy is explained by differential effects through metabolizing enzymes (e.g., CYP3A) and transporters (e.g., OATPs).

Grapefruit and its juice contain furanocoumarins (including bergamottins) which are metabolized by CYP3A4 in gut enterocytes. The metabolites then bind to and irreversibly inactivate the enzyme making it unavailable to metabolize medication, allowing more of the enzyme-susceptible drug to be absorbed. Given that many drugs are substrates for CYP3A4, there can be widespread increased bioavailability/serum concentrations for drugs as disparate as atorvastatin, budesonide, buspirone, carbamazepine, efavirenz, estradiol, fenofibrate, ibrutinib, lomitapide, lovastatin, saquinavir, sertraline, and simvastatin (Bailey et al., 2013; Boullata, 2019b; De Vries et al., 2015; Lee et al., 1999). But grapefruit juice has no influence on drugs like abacavir, glyburide, phenytoin or prednisone because these are not substrates for CYP3A4 (Kupferschmidt et al., 1998; Kumar et al., 1999; Boullata, 2019b). The effect of the interaction can last up to several days later until sufficient quantity of new enzyme is produced. In other words, the interaction is unrelated to temporal proximity of consumption. Generally, the interaction is more likely to be clinically significant with potential for adverse effects for drugs with a bioavailability <25–50% and significant clearance through CYP3A4. This explains why felodipine bioavailability increases with grapefruit juice but the related drug amlodipine does not (Vincent et al., 2000).

Grapefruit juice also contains flavonoids (e.g., naringin, hesperidin) which can inhibit uptake transporter function (OATP1A2, OATP2B1) at the gut mucosa (Nakanishi and Tamai, 2015; Johnson et al., 2017). Several drugs are substrates for OATP absorptive transport including aliskiren, celiprolol, etoposide, fexofenadine, ivermectin, levothyroxine, sulfasalazine, and talinolol (Kashihara et al., 2017; Glaeser et al., 2007; Lilja et al., 2005; Boullata, 2019b; Dolton et al., 2012). Because the interaction seems to occur through ligand competition at the transporter, the effect is most pronounced when drug is consumed with the juice or up to about 4 h afterward (Dolton et al., 2012). This interaction can generally be avoided by separating drug and juice by at least 4 h.

An individual's susceptibility to grapefruit juice-drug interactions may depend in part on the transporter and enzyme genotype, and concentrations of bioactive components in the juice. Despite much research on juice interactions involving drug-metabolizing enzymes, the effect on transporters continues to develop and both will allow better recognition and prediction of interactions (Dolton et al., 2012).

A number of beverages, containing a wide variety of polyphenols, are known to influence drug bioavailability; apart from any interactions in the gut lumen. For example, fruit juices can serve as the precipitating factor in interactions (Boullata and Armenti, 2010). Grapefruit juice is a commonly recognized example that with typical use can influence a number of drug-metabolizing enzymes and drug transporters to increase bioavailability of some drugs and reduce the bioavailability of others (Boullata and Armenti, 2010) (see **Box 1**). Similar interactions also occur with other fruit juices to varying degrees, depending on bioactive contents, by altering the bioavailability of vulnerable drugs (Chen et al., 2018a). For example, clementine and mandarin juices inhibit transporters (i.e., BCRP, OATP) to a similar extent as grapefruit juice, but grapefruit is more potent an inhibitor of CYP3A4 and CYP1A2 than are the other two citrus juices (Theile et al., 2017). Aside from interactions at transporters/enzymes caused by specific diet components, the fluid's characteristics may alter drug bioavailability. Beverage (e.g., apple juice) osmolality influences gut lumen fluid volume which then alters drug concentration and impacts absorption by decreasing the concentration gradient, especially for low permeability compounds (e.g., atenolol) (Funai et al., 2019). Given apple juice's ability to inhibit a specific transporter involved in atenolol absorption, multiple mechanisms may be in play (Hyewon et al., 2013).

The widespread use of dietary supplement products is associated with adverse effects including drug interactions (Ronis et al., 2018). Up to 45% of patients using dietary supplements with prescription drugs are at risk of interaction, with as many as 29% considered clinically serious (Peng et al., 2004; Lee et al., 2006; Sood et al., 2008; Loya et al., 2009). Individual nutrients and other isolated bioactive food substances found in dietary supplements can influence drug disposition, due to the effect on transporters and metabolizing enzymes, as well as their clinical effects. The ingredients found at higher concentrations in dietary supplement products are more concerning in some circumstances than food sources with regard to interactions (Egert and Rimbach, 2011).

Supplement products may include compounds isolated from a variety of foods including fruits (e.g., berries) and spices (e.g., turmeric) (Dreiseitel et al., 2009; Sand et al., 2010; Bahramsoltani et al., 2017; Hyrsova et al., 2019). Based on experimental data, several dietary flavonoids can decrease P-gp activity (non-competitively), and the polyphenols gallic acid and ellagic acid decrease P-gp activity (competitive), and curcumin may decrease BCRP activity through an unclear mechanism (Domínguez et al., 2021). A classic interaction with garlic supplements resulted in significant reduction of saquinavir bioavailability (Piscitelli et al., 2002).

Traditional supplements containing macro- and micronutrients are also important. Protein supplements, possibly with source-specific effects, can increase drug metabolism (Boullata and Armenti, 2010; Ronis et al., 2011). Interactions with micronutrients may improve drug outcomes in some cases but be detrimental in others. The potential benefit of vitamin B complex as an adjuvant intervention to fluconazole therapy for Candida has been described (Sun et al., 2017). A number of randomized controlled trials have administered vitamin D concurrent with pulmonary tuberculosis treatment accelerating sputum culture conversion in those with multidrug-resistant disease (Jolliffe et al., 2019). Including ascorbic acid, with or without vitamin E, in a triple-antibiotic therapy eradication regimen for *H. pylori* infection improved success rates in randomized trials (Kaboli et al., 2009; Sezikli et al., 2012).

The administration of folic acid supplementation in a patient receiving phenytoin can increase the drug's metabolic clearance over several weeks possibly requiring an increased drug dose to reach a new therapeutic state to avoid breakthrough seizures (Lewis et al., 1995; Berg et al., 1995). These two substances may also compete for similar transporters in the gut and brain. Vitamin

E can increase P-gp and CYP3A expression and activity in vitro (Podszun and Frank, 2014). Supplemental vitamin E could reduce tamoxifen concentrations (pharmacokinetic) in patients with breast cancer, or the drug's inhibitory effect on cancer cell proliferation (pharmacodynamics) (Peralta et al., 2009, 2006). Coenzyme Q_{10} (ubiquinone) supplementation increases the bleeding risk in patients using warfarin, whereas vitamin K inhibits the drug's anticoagulant activity (Shalansky et al., 2007). Although vitamin C enhanced the killing of *Candida albicans* in vitro, when combined with fluconazole the activity of the antifungal was significantly reduced (Avci et al., 2016; Wang et al., 2009).

At therapeutic doses, multivalent mineral supplements can interfere with antimicrobial (e.g., ciprofloxacin, levofloxacin, minocycline) absorption due to chelation (Lomaestro and Bailie, 1991; Pai et al., 2006). Co-administration of mineral supplements (e.g., calcium, iron, magnesium, or zinc) also significantly reduces the bioavailability of the antiretroviral drug dolutegravir, unless taken concurrently with a fat-containing meal (Song et al., 2015).

Nutritional status on drug disposition

The nutritional status of an individual, much like renal or hepatic function, is also a determining factor in drug disposition and effect. Drug distribution and clearance are the elements most influenced by poor nutritional status (e.g., obesity, PEM, micronutrient deficits) (Walter-Sack and Klotz, 1996; Boullata, 2010). Despite the prevalence of obesity and PEM, it is rare for drug product labeling to provide dosing regimen guidance for these patients as there is currently no regulatory requirement to do so (Boullata, 2010; Jacques and Erstad, 2010; Boyd et al., 2016). So clinicians rely on reviews summarizing the limited data available or apply rational principles to their decision-making (Boullata and Armenti, 2010; Boullata, 2010). The factors to appreciating the implications of nutritional status on drug dosing regimens include dose format and body composition.

Dose formats are most often a fixed dose (mg) but may also be weight-based (mg/kg). The fixed dose is established on the clinical effects in patients of otherwise healthy weight. Weight-based dosing most often uses total body weight. For drugs prescribed using weight-based dosing, it can be particularly challenging in a patient with elevated body mass index (Boullata, 2010; Pai, 2012). Neither body surface area nor an "ideal" body weight relative to height is appropriate for drug dosing in these circumstances (Boullata, 2010). The dosing weight to use depends on drug-specific characteristics in obese patients. The total body weight may be appropriate for some drugs; the lean body weight for others. A validated predictive equation for lean weight, which accounted for body composition, may be suitable for weight-based dosing to account for the altered distribution of a drug (Janmahasatian et al., 2005; Beckman et al., 2017). For other drugs an "adjusted" body weight may be more fitting. This value falls between total and lean weight (= lean body weight + [cf] [total body weight−lean body weight]), where the correction factor (cf), when known, represents the fraction of excess weight that normalizes the volume of distribution for that drug in the obese to that in a non-obese patient (Boullata, 2010). This is where body composition comes into play.

Two important kinetic parameters to examine for each drug are the volume of distribution (Vd, in L/kg) and the clearance (Cl, in L/h); the former governs a drug's initial or loading dose, while the latter shapes the maintenance dose and/or dosing interval and both parameters together determine serum and tissue drug concentrations. Using the two-compartment framework, relative to healthy body habitus, absolute content and proportions of lean mass and fat mass are altered as lower and upper extremes of body mass index are approached. A drug's Vd correlates with both anatomic compartments, while its Cl correlates predominantly with lean body mass (Boullata, 2010). So any influence of altered nutritional status on these two parameters can yield unexpected drug concentrations following a "usual" dose. Optimal drug dosing, especially for NTI drugs and antimicrobials, is required to achieve sufficient therapeutic serum concentrations while limiting adverse effects associated with supra-therapeutic concentrations. Generally, the initial (loading) dose is adjusted to account for drug distribution into existing lean and fat mass, and dosing interval is adjusted to account for altered clearance. Best practice, especially for NTI drugs and antimicrobials, is to closely follow clinical effects and serum concentrations in obesity or PEM.

In obesity, the weight-normalized Vd (L/kg) of carbamazepine is lower, and clearance is slightly reduced, suggesting use of an appropriately adjusted (between actual and lean) body weight for initial dosing followed by close monitoring (Caraco et al., 1992, 1995). Unrelated to the lipophilicity of lorazepam, there is no significant difference in Vd for the drug in obesity; however, drug clearance via glucuronidation is increased (Abernathy et al., 1983).

In malnourished children, gentamicin exhibited increased Vd and/or reduced Cl suggesting that larger doses administered less frequently might be appropriate as confirmed in a prospective study (Khan et al., 2006). Conversely, drug concentrations of sulfadiazine increase in PEM because of reduced Vd (Nehru et al., 1988), and Vd is also reduced for quinine and chloroquine (Pussard et al., 1999; Kadam et al., 2016). For antibiotics cleared primarily by renal excretion, some (e.g., cefoxitin, penicillin) are significantly reduced by malnutrition, whereas others (e.g., the aminoglycosides, cotrimoxazole) are insignificantly affected (Bolme et al., 1995; Bravo et al., 1984, 1982; Buchanan et al., 1979, 1980). PEM is associated with increased chloramphenicol concentrations despite increases in renal Cl because of a concurrent decrease in hepatic metabolism (Kohli et al., 1981; Smith et al., 1973). Hepatically cleared drugs will be influenced by any malnutrition-associated alterations in the enzymes of phase 1 (oxidation) or phase 2 (conjugation) metabolism. Rarely there may be a reported increase in drug metabolic Cl (e.g., chloroquine, quinine), but more often overall metabolic Cl is reduced (e.g., chloramphenicol, isoniazid, metronidazole, quinine) (Eriksson et al., 1983; Lares-Asseff et al., 1992; Ashton et al., 1993; Tulupule and Krishnaswamy, 1984; Roy et al., 2010; Treluyer et al., 1996; Salako et al., 1989). Modeling drug disposition based on data from malnourished patients better predicts serum concentrations than

a model derived in non-malnourished individuals (Lares-Asseff et al., 1999). Subsequently, a rational dosing approach can be prospectively evaluated and validated (Lares-Asseff et al., 2016). Following weight loss in obesity or nutritional rehabilitation in PEM, the pharmacokinetics may return toward expected values.

Micronutrient deficits can also influence drug disposition based in part on their roles in the function or stability of enzyme systems involved in drug metabolism (Walter-Sack and Klotz, 1996). For example, inadequate folate status decreases metabolic clearance of phenytoin, and subsequent repletion increases the metabolism (Berg et al., 1995). Carnitine deficiency may also increase phenytoin toxicity (Ling et al., 2012). Selenium deficiency in an animal model decreased the activity of several CYP-isoenzymes relative to controls (Jiang et al., 2020). Magnesium and zinc deficits compound the ototoxicity associated with aminoglycoside use making it potentially irreversible (Gunther et al., 1988; Lautermann and Schacht, 1995).

Drug-induced alteration in nutrition

In addition to any therapeutic benefit, medication use is associated with potential adverse effects. Among the adverse effects are those that can influence metabolic biomarkers, nutritional status in general, or the status of specific nutrients, all of which are important to consider during patient care (Piccolo and Boullata, 2016). The influence of medication on metabolic profile (e.g., hyperglycemia, hypertriglyceridemia), on overall nutritional status (e.g., weight gain, weight loss), or on the status of specific nutrients (e.g., hyperkalemia, zinc deficiency) can be mechanistically multifactorial (White, 2010; Lombardi et al., 2010). These include long lists of drugs including specific neurologic and antimicrobial agents described elsewhere (Lombardi et al., 2010; Piccolo and Boullata, 2016; Boullata, 2019a, 2021a). Even close analysis of a single drug can identify multiple influences on nutrition—including positive effects (Luciano-Mateo et al., 2017). From a clinician's perspective, drug-induced change in metabolic status is reflected in available biomarkers, alteration of overall nutritional status evaluated by physical exam, and in the case of specific nutrients is assessed through nutrition-focused physical exam and biomarkers.

Common general effects of a drug on metabolic profile or broad nutritional parameters are often available from clinical trials and found in the FDA's product labeling and therefore not specifically cited further in this section (US-FDA, 2021). Specific nutrient biomarkers of interest are rarely available until post-marketing reports surface from case studies. These represent "drug-nutrient" interactions in the strict sense of the term in which a medication alters the disposition of a nutrient. The number of drugs used by an individual has been correlated with deficits/excesses of specific nutrients (Heuberger and Caudell, 2011).

Metabolic status

Metabolic changes can include hyperglycemia as seen with several antiretrovirals (e.g., abacavir, atazanavir, dolutegravir, lamivudine, saquinavir), other antimicrobials (e.g., posaconazole), and neurologic agents (e.g., aripiprazole, bupropion, escitalopram, lamotrigine, olanzapine, quetiapine). Hypoglycemia may be expected with some drugs (e.g., insulin, glyburide) but also occur with others as an "off-target" effect (e.g., fluoroquinolones). Hypertriglyceridemia and hypercholesterolemia have been reported during treatment with some antimicrobials (e.g., clindamycin, dolutegravir, efavirenz), and neurologic drugs (e.g., mirtazapine, olanzapine, quetiapine).

Overall nutritional status

The effect on food intake may occur through the central nervous system, more local gut mechanisms, or both. Some medications may indirectly influence intake by altering a patient's capacity to gather, prepare or ingest food. An individual drug may cause several adverse effects that together influence nutritional status—fatigue, loss of appetite, change in taste, and disturbances of gastrointestinal function combined can lead to reduced dietary intake and decreased body weight. The impairment of the ability to gather, prepare, and ingest food may even occur following drug-induced cognitive, visual, movement, or gait disturbances. Many of these are at play in the elderly in particular given their generally high use of medication (Zadak et al., 2013). An influence on appetite or body weight is important to recognize especially for medications used chronically. The effects can be additive if multiple drugs are being used.

Central nervous system effects that may contribute to reductions in food gathering, preparation, and intake can include severe dizziness, fatigue, ataxia, and/or tremor. Visual disturbances (e.g., rifampin) and optic neuritis (e.g., chloramphenicol) may also influence food gathering, preparation, and intake. Peripheral neuropathy and even myositis, myalgia, and/or arthralgia may influence food gathering, preparation, and intake and negatively impact nutritional status over time. The influence on appetite varies, but rarely, a drug may alter impulse control that includes binge eating (e.g., aripiprazole). Reduced appetite has been reported for several unrelated antimicrobials (e.g., aminoglycosides, fluconazole, lamivudine, tenofovir), which with other factors may contribute to weight loss. The influence on weight gain or weight loss may even be indirect through a change in the gut microbiota (Angelakis et al., 2013; Raoult, 2017). This likely involves the gut-brain axis.

Through an effect on the gastrointestinal tract itself (i.e., taste disorders, stomatitis, nausea, vomiting, diarrhea and malabsorption), many medications can impair a patient's overall nutritional status. An influence on oral health which interferes with food intake is possible with several agents. Dry mouth and altered taste as well as glossitis or stomatitis occur with a number of drugs (Lombardi et al., 2010; Piccolo and Boullata, 2016; Boullata, 2019a, 2021a). Furthermore, common gastrointestinal disturbances

(i.e., anorexia, nausea, vomiting, abdominal pain, and diarrhea or constipation) are reported with many medications. Less commonly occurring are dysphagia, liver function abnormalities or hepatitis, and pancreatitis.

Changes in body weight, body composition, or extracellular fluid volume over time are easier to recognize. Weight gain including body fat accumulation (central, dorsocervical) or redistribution with peripheral wasting may occur with several antiretroviral agents (e.g., abacavir, atazanavir, dolutegravir, efavirenz, lamivudine, nevirapine, saquinavir, tenofovir). Weight gain is also associated with several neurologic drugs (e.g., aripiprazole, mirtazapine, olanzapine, quetiapine, and valproic acid), while weight loss is associated with others (e.g., phenytoin, sertraline, topiramate, and venlafaxine). Some drug-associated weight gain (e.g., lamotrigine, trazodone, valproic acid) may be attributed to edema.

Specific nutrients

The classic term "drug-nutrient" interaction is appropriately applied here for medication that alters nutrient disposition. The absorption, distribution, metabolism and excretion of specific nutrients can be influenced by drugs. When the detrimental effect on a specific nutrient's status is significant, it may be addressed by supplementation (e.g., pyridoxine administered with isoniazid, folic acid with phenytoin). With the common findings of micronutrient deficits, an undefined proportion of them may result from a drug-nutrient interaction requiring concerted effort to recognize and prevent (Karadima et al., 2016).

Although overt classic nutrient deficiency syndromes are rarely seen, lesser degrees of deficit may still be associated with clinical manifestations noted as adverse drug effects. For example, drug-induced osteomalacia can be caused by influence (e.g., antiepileptic drugs) on vitamin D metabolism (Pascussi et al., 2005; Oscarson et al., 2006; Xu et al., 2006). Carbamazepine increases the catabolism of vitamin D and subsequent association of long-term use with poor bone mineral density (Farhat et al., 2002; Fitzpatrick, 2004; Souverein et al., 2006; Kim et al., 2007). Intervention with vitamin D therapy may improve serum 25(OH) vitamin D concentrations in these patients (Pedrera et al., 2000; Elliott et al., 2007). Treatment for tuberculosis (e.g., isoniazid, rifampin) may decrease 25(OH) vitamin D concentrations by inhibiting 25-hydroxylation and accelerating 24,25-hydroxylation (Brodie et al., 1982). Despite low 25(OH) vitamin D concentrations often found at baseline in HIV infection, efavirenz as part of an antiretroviral regimen is associated with significant vitamin D deficits, increased bone turnover, and risk for osteomalacia (Welz et al., 2010; Orkin et al., 2014; Nylen et al., 2016). This may occur as a result of drug-induced interference with vitamin D metabolism and regulation in osteoblasts (Wegler et al., 2016). Vitamin D supplementation increased serum 25(OH) vitamin D concentrations in these patients and improved biomarkers of bone metabolism (Etminani-Esfahani et al., 2012). Knowing a patient's vitamin D status may help determine whether supplementation is required as part of these drug regimens.

Drug-induced hepatotoxicity with hyperammonemia (e.g., valproic acid) may result from carnitine deficiency (Van Wouwe, 1995; Werner et al., 2007; Melegh et al., 1987; Opala et al., 1991; Moreno et al., 2005). Reductions in serum carnitine concentrations occur in the face of diminished tissue distribution (Tein et al., 1993). In addition to the inhibition of intracellular transport, deficits are also thought to be related to decreased endogenous synthesis, and decreased renal reabsorption (Farkas et al., 1996; Lheureux et al., 2005; Segura-Bruna et al., 2006).

Sometimes the effect may be through more than one mechanism. For example, carbamazepine impairs biotin status through decreased intestinal absorption and accelerated metabolism to inactive metabolites (Said et al., 1989; Mock and Dyken, 1997). This influences the activity of the biotin-dependent carboxylases that may in turn alter neurologic function noted as an adverse drug effect. Phenytoin has an effect on folate status by inhibiting intestinal conjugases and limiting folate absorption. The drug may also interfere with folate transport, and consume folate during metabolism leading to folate deficits in at-risk individuals (Lewis et al., 1995; Schwaninger et al., 1999; Apeland et al., 2002; Sener et al., 2006). Serum folate decreases significantly in a few weeks, but adding folic acid to the regimen improves folate status and allows for steadier control of drug concentrations (Berg et al., 1995). For other drugs the mechanism of altered nutrient status is less clear.

The classic influence that isoniazid has on pyridoxine metabolism has been well described for decades in the treatment of tuberculosis (Biehl and Vilter, 1954; Clark, 1976). To prevent associated peripheral neuropathy or seizures, or to manage the presenting adverse effect, pyridoxine is administered along with the isoniazid regimen (Minns et al., 2010; Aiwale et al., 2015). The coadministration of high-dose pyridoxine with isoniazid is not expected to significantly affect overall drug bioavailability (Zhou et al., 2013). The use of highly active antiretroviral therapy was associated with lower folate and vitamin B_{12} concentrations in circulation and in breast milk in postpartum HIV-infected women (Allen et al., 2015; Flax et al., 2015). Highly active antiretroviral therapy is also associated with reduced α-tocopherol concentrations in nearly 20% of patients with long-standing HIV (Kaio et al., 2014). Carnitine deficits have also been suspected as an adverse effect of antiretroviral therapy (Mintz, 1995). Elevated partial thromboplastin time (PTT) and international normalized ratio (INR) with or without bleeding has been associated with azithromycin and with cotrimoxazole despite no history of malabsorption, that responded in each case to vitamin K administration (Stork et al., 2011; Fotouhie et al., 2016). Aminoglycosides may interfere with host selenium (Se-cys) incorporation into proteins (e.g., selenoprotein P) but requires further examination (Renko et al., 2017). Plasma selenium and selenoprotein P can serve as valuable biomarkers in such studies (Wiehe et al., 2016).

Electrolyte abnormalities are associated with a number of medications mostly at the level of excretion. This includes hyponatremia (e.g., carbamazepine, cotrimoxazole, escitalopram, mirtazapine, sertraline, venlafaxine), hypokalemia (e.g., aminoglycosides, amphotericin, caspofungin, flucytosine, posaconazole), hypomagnesemia (e.g., aminoglycosides, amphotericin, posaconazole, caspofungin), and hypocalcemia (e.g., aminoglycosides, amphotericin, posaconazole), and hypophosphatemia (topiramate). Chronic adefovir treatment for viral hepatitis is also associated with renal tubular dysfunction, hypophosphatemia,

and osteomalacia (Chen et al., 2018b). Hyperkalemia (e.g., cotrimoxazole, macrolide antibiotics) and hyperphosphatemia (e.g., macrolides) may also occur. Cotrimoxazole's association with reversible hyperkalemia is likely due to inhibition of renal potassium secretion attributed to the trimethoprim fraction of the combination product (Marinella, 1995).

Summary

It is a given that each person has a relationship with nutrition that supports their growth and development, and with varying degrees of success maintains or improves their health. However, many persons use medication acutely or on a more routine basis, setting up the potential for drug-nutrition interactions. A variety of potential interactions between drug and nutrition may influence a specific outcome and need to be accounted for. Such interactions include the influence of meals, dietary components, dietary supplements, or nutritional status on the disposition and clinical effect of the drug, as well as the potential effect of each drug on metabolic profile, overall nutritional status, or the status of individual nutrients. Depending on the drug and the patient, the pharmaceutical, pharmacokinetic, or pharmacodynamics interactions may have clinically significant effect. Drug-nutrition interactions are becoming better recognized and understood in clinical practice. In cases where enough data are available, interactions may be identified or predicted, and then appropriately managed. With ongoing advances, the expectation is that patient care can be further improved in the context of drug-nutrition interactions.

References

Abernathy, D.R., Greenblatt, D.J., Divoll, M., et al., 1983. Enhanced glucuronide conjugation of drugs in obesity: studies of lorazepam, oxazepam, and acetaminophen. J. Lab. Clin. Med. 101, 873–880.

Ahn, C.Y., Kim, E.J., Kwon, J.W., et al., 2003. Effects of cysteine on the pharmacokinetics of intravenous clarithromycin in rats with protein-calorie malnutrition. Life Sci. 73, 1783–1794.

Aiwale, A.S., Patel, U.A., Barvaliya, M.J., Jha, P.R., Tripathi, C., 2015. Isoniazid induced convulsions at therapeutic dose in an alcoholic and smoker patient. Curr. Drug Saf. 10, 94–95.

Allen, L.H., Hampel, D., Shahab-Ferdows, S., et al., 2015. Antiretroviral therapy provided to HIV-infected Malawian women in a randomized trial diminishes the positive effects of lipid-based nutrient supplements on breast-milk B vitamins. Am. J. Clin. Nutr. 102, 1468–1474.

Ando, H., Tsuruoka, S., Yanagihara, H., et al., 2005. Effects of grapefruit juice on the pharmacokinetics of pitavastatin and atorvastatin. Br. J. Clin. Pharmacol. 60, 494–497.

Angelakis, E., Merhej, V., Raoult, D., 2013. Related actions of probiotics and antibiotics on gut microbiota and weight modification. Lancet Infect. Dis. 13, 889–899.

Apeland, T., Mansoor, M.A., Pentieva, K., McNulty, H., Seljeflot, I., Strandjord, R.E., 2002. The effect of B-vitamins on hyperhomocysteinemia in patients on antiepileptic drugs. Epilepsy Res. 51, 237–247.

Ashton, M., Bolme, P., Alemayehu, E., Eriksson, M., Paalzow, L., 1993. Decreased chloramphenicol clearance in malnourished Ethiopian children. Eur. J. Clin. Pharmacol. 45, 181–186.

Avci, P., Freire, F., Banvolgyi, A., et al., 2016. Sodium ascorbate kills Candida albicans in vitro via iron-catalyzed Fenton reaction: importance of oxygenation and metabolism. Future Microbiol. 11, 1535–1547.

Bahramsoltani, R., Rahimi, R., Farzaei, M.H., 2017. Pharmacokinetic interactions of curcuminoids with conventional drugs: a review. J. Ethnopharmacol. 209, 1–12.

Bailey, D.G., Dresser, G., Arnold, M.O., 2013. Grapefruit-medication interactions: forbidden fruit or avoidable consequences? Can. Med. Assoc. J. 185, 309–316.

Beckman, L.M., Boullata, J.I., Fisher, P.L., Compher, C.W., Earthman, C.P., 2017. Evaluation of lean body weight equation by dual-energy X-ray absorptiometry measures. J. Parenter. Enter. Nutr. 41, 392–397.

Benet, L.Z., 2013. The role of BCS (biopharmaceutics classification system) and BDDCS (biopharmaceutics drug disposition classification system) in drug development. J. Pharm. Sci. 102, 34–42.

Berg, M.J., Stumbo, P.J., Chenard, C.A., et al., 1995. Folic acid improves phenytoin pharmacokinetics. J. Am. Diet Assoc. 95, 352–356.

Biehl, J.P., Vilter, R.W., 1954. Effects of isoniazid on pyridoxine metabolism. J. Am. Med. Assoc. 156, 1549–1552.

Bolme, P., Eriksson, M., Paalzow, L., et al., 1995. Malnutrition and pharmacokinetics of penicillin in Ethiopian children. Pharmacol. Toxicol. 76, 259–262.

Boullata, J.I., Armenti, V.T., 2010. Handbook of Drug-Nutrient Interactions, second ed. Humana Press, New York, NY.

Boullata, J.I., Hudson, L.M., 2012. Drug-nutrient interactions: a broad view with implications for practice. J. Acad. Nutr. Diet. 112, 506–517.

Boullata, J.I., 2010. Drug disposition in obesity and protein-energy malnutrition. Proc. Nutr. Soc. 69, 543–550.

Boullata, J.I., 2013. Drug and nutrition interactions: not just food for thought. J. Clin. Pharm. Therapeut. 38, 269–271.

Boullata, J.I., 2019a. Drug-nutrition interactions and the brain: it's not all in your head. Curr. Nutr. Rep. 8, 92–98.

Boullata, J.I., 2019b. Guidebook on Enteral Medication Administration. American Society for Parenteral and Enteral Nutrition, Silver Spring.

Boullata, J.I., 2021a. Drug-nutrition interactions in infectious diseases. In: Humphries, D., Scott, M., Vermund, S.H. (Eds.), Nutrition and Infectious Diseases: Shifting the Clinical Paradigm. Springer Nature, Switzerland AG, pp. 367–410.

Boullata, J.I., 2021b. Enteral medication for the tube-fed patient: making this route safe and effective. Nutr. Clin. Pract. 36, 111–132.

Boyd, S.E., Charani, E., Lyons, T., Frost, G., Holmes, A.H., 2016. Information provision for antibacterial dosing in the obese patient: a sizeable absence? J. Antimicrob. Chemother. 71, 3588–3592.

Bravo, M.E., Arancibia, A., Jarpa, S., Carpentier, P.M., Jahn, A.N., 1982. Pharmacokinetics of gentamicin in malnourished infants. Eur. J. Clin. Pharmacol. 21, 499–504.

Bravo, I.G., Bravo, M.E., Plate, G., Merlez, J., Arancibia, A., 1984. The pharmacokinetics of cotrimoxazole sulphonamide in malnourished (marasmic) infants. Pediatr. Pharmacol. 4, 167–176.

Brodie, M.J., Boobis, A.R., Hillyard, C.J., et al., 1982. Effect of rifampicin and isoniazid on vitamin D metabolism. Clin. Pharmacol. Ther. 32, 525–530.

Buchanan, N., Davis, M.D., Eyberg, C., 1979. Gentamicin pharmacokinetics in kwashiorkor. Br. J. Clin. Pharmacol. 8, 451–453.

Buchanan, N., Mithal, Y., Witcomb, M., 1980. Cefoxitin: intravenous pharmacokinetics and intramuscular bioavailability in kwashiorkor. Br. J. Clin. Pharmacol. 9, 623–627.

Caraco, Y., Zylber-Katz, E., Berry, E.M., Levy, M., 1992. Significant weight reduction in obese subjects enhances carbamazepine elimination. Clin. Pharmacol. Ther. 51, 501–506.

Caraco, Y., Zylber-Katz, E., Berry, E.M., Levy, M., 1995. Carbamazepine pharmacokinetics in obese and lean subjects. Ann. Pharmacother. 29, 843–847.

Chan, L.N., 2002. Drug-nutrient interactions in clinical nutrition. Curr. Opin. Clin. Nutr. Metab. Care 5, 327–332.

Chan, L.N., 2013. Drug-nutrient interactions. J. Parenter. Enter. Nutr. 37, 450–459.

Chen, M., Zhou, S.Y., Fabriaga, E., Zhang, P.H., Zhou, Q., 2018a. Food-drug interactions precipitated by fruit juices other than grapefruit juice: an update review. J. Food Drug Anal. 26, S61–S71.
Chen, N., Zhang, J.-B., Zhang, Q., et al., 2018b. Adefovir dipivoxil induced hypophosphatemic osteomalacia in chronic hepatitis B: a comparative study of Chinese and foreign case series. BMC Pharmacol. Toxicol. 19, 23.
Clark, F., 1976. Drugs and vitamin deficiency. Adv. Drug React. Bull. 57, 196–199.
Cloarec, N., Solas, C., Ladaique, A., et al., 2017. Sub-therapeutic darunavir concentration and garlic consumption; a "Mediterranean" drug-food interaction, about 2 cases. Eur. J. Clin. Pharmacol. 73, 1331–1333.
Courtney, R., Wexler, D., Radwanski, E., Lim, J., Laughlin, M., 2004. Effect of food on the relative bioavailability of two oral formulations of posaconazole in healthy adults. Br. J. Clin. Pharmacol. 57, 218–222.
Curatolo, W., Liu, P., Johnson, B.A., et al., 2011. Effects of food on a gastrically degraded drug: azithromycin fast-dissolving gelatin capsules and HPMC capsules. Pharm. Res. 28, 1531–1539.
de Man, F.M., van Eerden, R.A.G., van Doorn, G.M., et al., 2021. Effects of protein and calorie restriction on the metabolism and toxicity profile of irinotecan in cancer patients. Clin. Pharmacol. Ther. 109, 1304–1313.
De Vries, R., Smit, J.W., Hellemans, P., et al., 2015. Stable isotope-labelled intravenous microdose for absolute bioavailability and effect of grapefruit juice on ibrutinib in healthy adults. Br. J. Clin. Pharmacol. 81, 235–245.
Deng, J., Zhu, X., Chen, Z., et al., 2017. A review of food-drug interactions on oral drug absorption. Drugs 77, 1833–1855.
Dolton, M.J., Roufogalis, B.D., McLachlan, A.J., 2012. Fruit juices as perpetrators of drug interactions: the role of organic anion-transporting polypeptides. Clin. Pharmacol. Ther. 92, 622–630.
Domínguez, C.J., Tocchetti, G.N., Rigalli, J.P., Mottino, A.D., 2021. Acute regulation of apical ABC transporters in the gut: potential influence on drug bioavailability. Pharmacol. Res. 163, 105251.
Dreiseitel, A., Oosterhuis, B., Vukman, K.V., et al., 2009. Berry anthocyanins and anthocyanidins exhibit distinct affinities for the efflux transporters BCRP and MDR1. Br. J. Pharmacol. 158, 1942–1950.
Eagles, S.K., Gross, A.S., McLachlan, A.J., 2020. The effects of cruciferous vegetable-enriched diets on drug metabolism: a systematic review and meta-analysis of dietary intervention trials in humans. Clin. Pharmacol. Ther. 108, 212–227.
Egert, S., Rimbach, G., 2011. Which sources of flavonoids: complex diets or dietary supplements? Adv. Nutr. 2, 8–14.
Elliott, J.O., Jacobson, M.P., Haneef, Z., 2007. Homocysteine and bone loss in epilepsy. Seizure 16, 22–34.
Elshahed, M.S., Miron, A., Aprotosoaie, A.C., Farag, M.A., 2021. Pectin in diet: interactions with the human microbiome, role in gut homeostasis, and nutrient-drug interactions. Carbohydr. Polym. 255, 117388.
Eriksson, M., Paalzow, L., Bolme, P., Mariam, T.W., 1983. Chloramphenicol pharmacokinetics in Ethiopian children of differing nutritional status. Eur. J. Clin. Pharmacol. 24, 819–823.
Eshelman, F.N., Spyker, D.A., 1978. Pharmacokinetics of amoxicillin and ampicillin: crossover study of the effect of food. Antimicrob. Agents Chemother. 14, 539–543.
Etminani-Esfahani, M., Khalili, H., Jafari, S., Abdollahi, A., Dashti-Khavidaki, S., 2012. Effects of vitamin D supplementation on the bone specific biomarkers in HIV infected individuals under treatment with efavirenz. BMC Res. Notes 5, 204.
Farhat, G., Yamout, B., Mikati, M.A., et al., 2002. Effect of antiepileptic drugs on bone density in ambulatory patients. Neurology 58, 1348–1353.
Farkas, V., Bock, I., Cseko, J., Sandor, A., 1996. Inhibition of carnitine biosynthesis by valproic acid in rats: the biochemical mechanism of inhibition. Biochem. Pharmacol. 52, 1429–1433.
Fitzpatrick, L.A., 2004. Pathophysiology of bone loss in patients receiving anticonvulsant therapy. Epilepsy Behav. 5 (Suppl. 2), S3–S15.
Flax, V.L., Adair, L.S., Allen, L.H., et al., 2015. Plasma micronutrient concentrations are altered by antiretroviral therapy and lipid-based nutrient supplements in lactating HIV-infected Malawian women. J. Nutr. 145, 1950–1957.
Fleisher, B., Unum, J., Shao, J., An, G., 2015. Ingredients in fruit juices interact with dasatinib through inhibition of BCRP: a new mechanism of beverage-drug interaction. J. Pharm. Sci. 104, 266–275.
Fotouhie, A., Desai, H., King, S., Parsa, N.A., 2016. Gastrointestinal bleeding secondary to trimethoprim-sulfamethoxazole-induced vitamin K deficiency. BMJ Case Rep. https://doi.org/10.1136/bcr-2016-214437.
Funai, Y., Shirasaka, Y., Ishihara, M., et al., 2019. Effect of osmolality on the pharmacokinetic interaction between apple juice and atenolol in rats. Drug Metab. Dispos. 47, 386–391.
Gezmen-Karadağ, M., Çelik, E., Kadayifçi, F.Z., et al., 2018. Role of food-drug interactions in neurological and psychological diseases. Acta Neurobiol. Exp. 78, 187–197.
Glaeser, H., Bailey, D.G., Dresser, G.K., et al., 2007. Intestinal drug transporter expression and the impact of grapefruit juice in humans. Clin. Pharmacol. Ther. 81, 362–370.
Gunther, T., Rebentisch, E., Vormann, J., Konig, M., Ising, H., 1988. Enhanced ototoxicity of gentamicin and salicylate caused by Mg deficiency and Zn deficiency. Biol. Trace Elem. Res. 16, 43–50.
Guzzo, C.A., Furtek, C.I., Porras, A.G., et al., 2002. Safety, tolerability, and pharmacokinetics of escalating high doses of ivermectin in healthy adult subjects. J. Clin. Pharmacol. 42, 1122–1133.
Heuberger, R.A., Caudell, K., 2011. Polypharmacy and nutritional status in older adults: a cross-sectional study. Drugs Aging 28, 315–323.
Homeida, M.M., Malcolm, S.B., El Tayeb, A.Z., et al., 2013. The lack of influence of food and local alcoholic brew on the blood level of Mectizan® (ivermectin). Acta Trop. 127, 97–100.
Hughes, G.S., Heald, D.L., Barker, K.B., et al., 1989. The effects of gastric pH and food on the pharmacokinetics of a new oral cephalosporin, cefpodoxime proxetil. Clin. Pharmacol. Ther. 46, 674–685.
Humphries, D., Nguyen, S., Kumar, S., et al., 2017. Effectiveness of albendazole for hookworm varies widely by community and correlates with nutritional factors: a cross-sectional study of school-age children in Ghana. Am. J. Trop. Med. Hyg. 96, 347–354.
Hyewon, J., In-Jin, J., Hwan, L.S., et al., 2013. Apple juice greatly reduces systemic exposure to atenolol. Br. J. Clin. Pharmacol. 75, 172–179.
Hyrsova, L., Vanduchova, A., Dusek, J., et al., 2019. Trans-resveratrol, but not other natural stilbenes occurring in food, carries the risk of drug food interaction via inhibition of cytochrome P450 enzymes or interaction with xenosensor receptors. Toxicol. Lett. 300, 81–91.
Jacques, K.A., Erstad, B.L., 2010. Availability of information for dosing injectable medications in underweight or obese patients. Am. J. Health Syst. Pharm. 67, 1948–1950.
Janmahasatian, S., Duffull, S.B., Ash, S., et al., 2005. Quantification of lean body-weight. Clin. Pharmacokinet. 44, 1051–1065.
Jensen, O., Matthaei, J., Blome, F., et al., 2020. Variability and heritability of thiamine pharmacokinetics with focus on OCT1 effects on membrane transport and pharmacokinetics in humans. Clin. Pharmacol. Ther. 107, 628–638.
Jiang, Z., Gu, L., Liang, X., et al., 2020. The effect of selenium on CYP450 isoform activity and expression in pigs. Biol. Trace Elem. Res. 196, 454–462.
Johnson, E.J., Won, C.S., Köck, K., Paine, M.F., 2017. Prioritizing pharmacokinetic drug interaction precipitants in natural products: application to OATP inhibitors in grapefruit juice. Biopharm. Drug Dispos. 38, 251–259.
Jolliffe, D.A., Ganmaa, D., Wejse, C., et al., 2019. Adjunctive vitamin D in tuberculosis treatment: meta-analysis of individual participant data. Eur. Respir. J. 53, 1802003.
Kaboli, S.A., Zojaji, H., Mirsattari, D., et al., 2009. Effect of addition of vitamin C to clarithromycin-amoxicillin-omeprazole triple regimen on *Helicobacter pylori* eradication. Acta Gastroenterol. Belg. 72, 222–224.
Kadam, P.P., Gogtay, N.J., Karande, S., Shah, V., Thatte, U.M., 2016. Evaluation of pharmacokinetics of single-dose chloroquine in malnourished children with malaria: a comparative study with normally nourished children. Indian J. Pharmacol. 48, 498–502.

Kaio, D.J.I., Rondó, P.H.C., Luzia, L.A., et al., 2014. Vitamin E concentrations in adults with HIV/AIDS on highly active antiretroviral therapy. Nutrients 6, 3641–3652.
Karadima, V., Kraniotou, C., Bellos, G., Tsangaris, G.T., 2016. Drug-micronutrient interactions: food for thought and thought for action. EPMA J. 7, 10.
Kashihara, Y., Ieiri, I., Yoshikado, T., et al., 2017. Small-dosing clinical study: pharmacokinetic, pharmacogenomics (SLCO2B1 and ABCG2), and interaction (atorvastatin and grapefruit juice) profiles of 5 probes for OATP2B1 and BCRP. J. Pharm. Sci. 106, 2688–2694.
Khan, A.M., Ahmed, T., Alam, N.H., Chowdhury, A.K., Fuchs, G.J., 2006. Extended-interval gentamicin administration in malnourished children. J. Trop. Pediatr. 52, 179–184.
Kim, S.H., Lee, J.W., Choi, K.G., Chung, H.W., Lee, H.W., 2007. A 6-month longitudinal study of bone mineral density with antiepileptic drug monotherapy. Epilepsy Behav. 10, 291–295.
Kim, S.H.J., Jackson, A.J., Hunt, C.A., 2014. In silico, experimental, mechanistic model for extended-release felodipine disposition exhibiting complex absorption and a highly variable food interaction. PLoS One 9, e108392.
Klang, M., McLymont, V., Ng, N., 2013. Osmolality, pH, and compatibility of selected oral liquid medications with an enteral nutrition product. J. Parenter. Enter. Nutr. 37, 689–694.
Kohli, K., Aggarwal, K.K., Bhatt, I.N., 1981. The pharmacokinetic profile of chloramphenicol in protein-malnourished rats. Indian J. Med. Res. 73, 208–217.
Krishnaswamy, K., 1978. Drug metabolism and pharmacokinetics in malnutrition. Clin. Pharmacokinet. 3, 216–240.
Kumar, N., Garg, S.K., Prabhakar, S., 1999. Lack of pharmacokinetic interaction between grapefruit juice and phenytoin in healthy male volunteers and epileptic patients. Methods Find Exp. Clin. Pharmacol. 21, 629–632.
Kumar, A.K.H., Chandrasekaran, V., Kumar, A.K., et al., 2017. Food significantly reduces plasma concentrations of first-line anti-tuberculosis drugs. Indian J. Med. Res. 145, 530–535.
Kupferschmidt, H.H., Fattinger, K.E., Ha, H.R., Follath, F., Krahenbuhl, S., 1998. Grapefruit juice enhances the bioavailability of the HI protease inhibitor saquinavir in man. Br. J. Clin. Pharmacol. 45, 355–359.
Kvernelend, M., Taubøll, E., Selmer, K.K., et al., 2015. Modified Atkins diet may reduce serum concentrations of antiepileptic drugs. Acta Neurol. Scand. 131, 187–190.
Lange, H., Eggers, R., Bircher, J., 1988. Increased systemic availability of albendazole when taken with a fatty meal. Eur. J. Clin. Pharmacol. 34, 315–317.
Lares-Asseff, I., Cravioto, J., Santiago, P., Perez-Ortiz, B., 1992. Pharmacokinetics of metronidazole in severely malnourished and nutritionally rehabilitated children. Clin. Pharmacol. Ther. 51, 42–50.
Lares-Asseff, I., Lugo-Goytia, G., Pérez-Guillé, M.G., et al., 1999. Bayesian prediction of chloramphenicol blood levels in children with sepsis and malnutrition. Rev. Investig. Clin. 51, 159–165.
Lares-Asseff, I., Pérz-Guillé, M.G., Camacho Vieyra, G.A., et al., 2016. Population pharmacokinetics of gentamicin in Mexican children with severe malnutrition. Pediatr. Infect. Dis. J. 35, 872–878.
Lautermann, J., Schacht, J., 1995. Reduced nutritional status enhances ototoxicity. Laryngoscope 74, 724–727.
Lee, A.J., Chan, W.K., Harralson, A.F., Buffum, J., Bui, B.C.C., 1999. The effects of grapefruit juice on sertraline metabolism: an in vitro and in vivo study. Clin. Therapeut. 21, 1890–1899.
Lee, A.H., Ingraham, S.E., Kopp, M., et al., 2006. The incidence of potential interactions between dietary supplements and prescription medications in cancer patients at a Veterans Administration Hospital. Am. J. Clin. Oncol. 29, 178–182.
Lewis, D.P., Van Dyke, D.C., Willhite, L.A., Stumbo, P.J., Berg, M.J., 1995. Phenytoin-folic acid interaction. Ann. Pharmacother. 29, 726–735.
Lheureux, P.E.R., Penaloza, A., Zahir, S., Gris, M., 2005. Science review: carnitine in the treatment of valproic acid-induced toxicity—what is the evidence? Crit. Care 9, 431–440.
Lilja, J.J., Laitinen, K., Neuvonen, P.J., 2005. Effects of grapefruit juice on the absorption of levothyroxine. Br. J. Clin. Pharmacol. 60, 337–341.
Lin, T.Y., Yang, M.H., Chang, F.Y., 2013. A randomized, phase I, 3-way crossover study to examine the effects of food on the pharmacokinetics of single doses of 400 mg posaconazole oral suspension in healthy male Taiwanese subjects. Ther. Drug Monit. 35, 223–227.
Lincoln, J., Stewart, M.E., Preskorn, S.H., 2010. How sequential studies inform drug development: evaluating the effect of food intake on optimal bioavailability of ziprasidone. J. Psychiatr. Pract. 16, 103–114.
Ling, P., Lee, D.J., Yoshida, E.M., Sirrs, S., 2012. Carnitine deficiency presenting with encephalopathy and hyperammonemia in a patient receiving chronic enteral tube feeding: a case report. J. Med. Case Rep. 6, 227.
Lomaestro, B.M., Bailie, G.R., 1991. Quinolone-cation interaction: a review. Ann. Pharmacother. 25, 1249–1258.
Lombardi, L.R., Kreys, E., Gerry, S., Boullata, J.I., 2010. Nutrition in the age of polypharmacy. In: Bendich, A., Deckelbaum, R.J. (Eds.), Preventive Nutrition, fourth ed. Humana Press, New York, NY, pp. 79–125.
Loya, A.M., Gonzalez-Stuart, A., Rivera, J.O., 2009. Prevalence of polypharmacy, polyherbacy, nutritional supplement use and potential product interactions among older adults living on the United States–Mexico border: a descriptive, questionnaire-based study. Drugs Aging 26, 423–436.
Luciano-Mateo, F., Hernandez-Aguilera, A., Cabre, N., et al., 2017. Nutrients in energy and one-carbon metabolism: learning from metformin users. Nutrients 9, 121.
Manrique, Y.J., Lee, D.J., Islam, F., et al., 2014. Crushed tablets: does the administration of food vehicles and thickened fluids to aid medication swallowing alter drug release? J. Pharm. Pharmaceut. Sci. 17, 207–219.
Marinella, M.A., 1995. Case report: reversible hyperkalemia associated with trimethoprim-sulfamethoxazole. Am. J. Med. Sci. 310, 115–117.
Mason, P., 2010. Important drug-nutrient interactions. Proc. Nutr. Soc. 69, 551–557.
Melegh, B., Kerner, J., Kispal, G., et al., 1987. Effect of chronic valproic acid treatment on plasma and urine carnitine levels in children. Acta Paediatr. Hung. 28, 137–142.
Minami, R., Inotsume, N., Nakano, M., et al., 1993. Effect of milk on absorption of norfloxacin in healthy volunteers. J. Clin. Pharmacol. 33, 1238–1240.
Minns, A.B., Ghafouri, N., Clark, R.F., 2010. Isoniazid-induced status epilepticus in a pediatric patient after inadequate pyridoxine therapy. Pediatr. Emerg. Care 26, 380–381.
Mintz, M., 1995. Carnitine in human immunodeficiency virus type 1 infection/acquired immune deficiency syndrome. J. Child Neurol. 10 (Suppl. 2), S40–S44.
Miyajima, A., Yamamoto, Y., Hirota, T., 2015. Effect of high fat intake on the pharmacokinetic profile of ivermectin in rabbits. Drug Metabol. Pharmacokinet. 30, 253–256.
Mock, D.M., Dyken, M.E., 1997. Biotin catabolism is accelerated in adults receiving long-term therapy with anticonvulsants. Neurology 49, 1444–1447.
Moreno, F.A., Macey, H., Schreiber, B., 2005. Carnitine levels in valproic acid-treated psychiatric patients: a cross-sectional study. J. Clin. Psychol. 66, 555–558.
Morimoto, K., Kishimura, K., Nagami, T., et al., 2011. Effect of milk on the pharmacokinetics of oseltamivir in healthy volunteers. J. Pharm. Sci. 100, 3854–3861.
Nakanishi, T., Tamai, I., 2015. Interaction of drug or food with drug transporters in intestine and liver. Curr. Drug Metabol. 16, 753–764.
Nehru, B., Mehta, S., Nain, C.K., Mathur, V.S., 1988. Disposition of sulphadiazine in young rhesus monkeys with protein calorie malnutrition. Int. J. Clin. Pharmacol. Ther. Toxicol. 26, 509–512.
Neovonen, P., Gothoni, G., Hackman, R., Bjorksten, K., 1970. Interference of iron with the absorption of tetracyclines in man. BMJ 4, 532–534.
Neuvonen, P.J., Kivistö, K.T., Lehto, P., 1991. Interference of dairy products with the absorption of ciprofloxacin. Clin. Pharmacol. Ther. 50 (5 Pt 1), 498–502.
Nigam, S.K., Bush, K.T., Bhatnagar, V., Poloyac, S.M., Momper, J.D., 2020. The systems biology of drug metabolizing enzymes and transporters: relevance to quantitative systems pharmacology. Clin. Pharmacol. Ther. 108, 40–53.
Nikolić, B., 2020. Antimicrobial drug-nutrient interactions: consistency of information for generic drugs. Vojnosanit. Pregl. 77, 1073–1079.
Nylen, H., Habtewold, A., Makonnen, E., et al., 2016. Prevalence and risk factors for efavirenz-based antiretroviral treatment-associated severe vitamin D deficiency: a prospective cohort study. Medicine 95 (34), e4631.
Ogunbona, F.A., Smith, I.F., Olawoye, O.S., 1985. Fat contents of meals and bioavailability of griseofulvin in man. J. Pharm. Pharmacol. 37, 283–284.
Opala, G., Winter, S., Vance, C., Vance, H., Hutchison, H.T., Linn, L.S., 1991. The effect of valproic acid on plasma carnitine levels. Am. J. Dis. Child. 145, 999–1001.
Orellana-Paucar, A., Vintimilla-Rojas, D., 2020. Interactions of clinical relevance associated with concurrent administration of prescription drug and food or medicinal plants: a systematic review protocol. Syst. Rev. 9, 1.
Orkin, C., Wohl, D.A., Williams, A., Deckx, H., 2014. Vitamin D deficiency in HIV: a shadow on long-term management? AIDS Rev. 16, 59–74.

Oscarson, M., Zanger, U.M., Rifki, O.F., et al., 2006. Transcriptional profiling of genes induced in the livers of patients treated with carbamazepine. Clin. Pharmacol. Ther. 80, 440–456.
Pai, M.P., Allen, S.E., Amsden, G.W., 2006. Altered steady state pharmacokinetics of levofloxacin in adult cystic fibrosis patients receiving calcium carbonate. J. Cyst. Fibros. 5, 153–157.
Pai, M.P., 2012. Drug dosing based on weight and body surface area: mathematical assumptions and limitations in obese adults. Pharmacotherapy 32, 856–868.
Pascussi, J.M., Robert, A., Nguyen, M., et al., 2005. Possible involvement of pregnane X receptor-enhanced CYP24 expression in drug induced osteomalacia. J. Clin. Invest. 115, 177–186.
Pedrera, J.D., Canal, M.L., Carvajal, J., et al., 2000. Influence of vitamin D administration on bone ultrasound measurements in patients on anticonvulsant therapy. Eur. J. Clin. Invest. 30, 895–899.
Peng, C.C., Glassman, P.A., Trilli, L.E., et al., 2004. Incidence and severity of potential drug dietary supplement interactions in primary care patients: an exploratory study of 2 outpatient practices. Arch. Intern. Med. 164, 630–636.
Peralta, E.A., Viegas, M.L., Louis, S., Engle, D.L., Dunnington, G.L., 2006. Effect of vitamin E on tamoxifen-treated breast cancer cells. Surgery 140, 607–615.
Peralta, E.A., Brewer, A.T., Louis, S., Dunnington, G.L., 2009. Vitamin E increases biomarkers of estrogen stimulation when taken with tamoxifen. J. Surg. Res. 153, 143–147.
Péter, S., Navis, G., de Borst, M.H., et al., 2017. Public health relevance of drug-nutrition interactions. Eur. J. Nutr. 56 (Suppl. 2), S23–S36.
Piccolo, K.M., Boullata, J.I., 2016. The influence of polypharmacy on nutrition. In: Bendich, A., Deckelbaum, R.J. (Eds.), Preventive Nutrition, fifth ed. Humana Press, New York, pp. 83–113.
Piscitelli, S.C., Burstein, A.H., Welden, N., Gallicano, K.D., Falloon, J., 2002. The effect of garlic supplements on the pharmacokinetics of saquinavir. Clin. Infect. Dis. 34, 234–238.
Podszun, M., Frank, J., 2014. Vitamin E-drug interactions: molecular basis and clinical relevance. Nutr. Res. Rev. 27, 215–231.
Purkins, L., Wood, N., Kleinermans, D., Greenhalgh, K., Nichols, D., 2003. Effect of food on the pharmacokinetics of multiple-dose oral voriconazole. Br. J. Clin. Pharmacol. 56 (Suppl. 1), 17–23.
Pussard, E., Barennes, H., Daouda, H., et al., 1999. Quinine disposition in globally malnourished children with cerebral malaria. Clin. Pharmacol. Ther. 65, 500–510.
Radwan, A., Sweileh, A., Shraim, W., et al., 2018. Evaluation of community pharmacists' knowledge and awareness of food-drug interactions in Palestine. Int. J. Clin. Pharm. 40, 668–675.
Raman, S., Polli, J.E., 2016. Prediction of positive food effect: bioavailability enhancement of BCS class II drugs. Int. J. Pharm. 506, 110–115.
Raoult, D., 2017. Microbiota, obesity and malnutrition. Microb. Pathog. 106, 1–2.
Renko, K., Martitz, J., Hybsier, S., et al., 2017. Aminoglycoside-driven biosynthesis of selenium-deficient selenoprotein P. Sci. Rep. 7, 4391.
Rigalli, J.P., Ciriaci, N., Mottino, A.D., Catania, V.A., Ruiz, M.L., 2016. Modulation of expression and activity of ABC transporters by the phytoestrogen genistein: impact on drug disposition. Curr. Med. Chem. 23, 1370–1389.
Roe, D.,A., 1985. Drug effects on nutrient absorption, transport, and metabolism. Drug Nutr. Interact. 4, 117–135.
Ronis, M.J.J., Chen, Y., Liu, X., et al., 2011. Enhanced expression and glucocorticoid-inducibility of hepatic cytochrome P450 3A involve recruitment of the pregnane-X-receptor promoter elements in rats fed soy protein isolate. J. Nutr. 141, 10–16.
Ronis, M.J.J., Pedersen, K.B., Watt, J., 2018. Adverse effects of nutraceuticals and dietary supplements. Annu. Rev. Pharmacol. 58, 583–601.
Roy, V., Gupta, D., Gupta, P., Sethi, G.R., Mishra, T.K., 2010. Pharmacokinetics of isoniazid in moderately malnourished children with tuberculosis. Int. J. Tubercul. Lung Dis. 14, 374–376.
Rubbens, J., Brouwers, J., Tack, J., Augustijns, P., 2019. Gastric and duodenal diclofenac concentrations in healthy volunteers after intake if the FDA standard meal: in vivo observations and in vitro explorations. Mol. Pharm. 16, 573–582.
Sadrieh, N., Brower, J., Yu, L., et al., 2005. Stability, dose uniformity, and palatability of three counterterrorism drugs: human subject and electronic tongue studies. Pharm. Res. 22, 1747–1756.
Said, H.M., Redha, R., Nylander, W., 1989. Biotin transport in the human intestine: inhibition by anticonvulsant drugs. Am. J. Clin. Nutr. 49, 127–131.
Salako, L.A., Sowunmi, A., Akinbami, F.O., 1989. Pharmacokinetics of quinine in African children suffering from kwashiorkor. Br. J. Clin. Pharmacol. 28, 197–201.
Sand, P.G., Dreiseitel, A., Stang, M., et al., 2010. Cytochrome P450 2C19 inhibitory activity of common berry constituents. Phytother Res. 24, 304–307.
Santos, C.A., Boullata, J.I., 2005. An approach to evaluating drug-nutrient interactions. Pharmacotherapy 25, 1789–1800.
Schick, P., Sager, M., Wegner, F., et al., 2019. Application of the GastroDuo as an in vitro dissolution tool to simulate gastric emptying of the postprandial stomach. Mol. Pharm. 16, 4651–4660.
Schwaninger, M., Ringleb, P., Winter, R., et al., 1999. Elevated plasma concentrations of homocysteine in antiepileptic drug treatment. Epilepsia 40, 345–350.
Segura-Bruna, N., Rodriguez-Campello, A., Puente, V., Roquer, J., 2006. Valproate-induced hyperammonemic encephalopathy. Acta Neurol. Scand. 114, 1–7.
Semen, K.O., Yeliyeyeva, O.P., Bast, A., 2020. Interaction of diet and drugs in lung disease. Curr. Opin. Pulm. Med. 26, 359–362.
Sener, U., Zorlu, Y., Karaguzel, O., Ozdamar, O., Coker, I., Topbas, M., 2006. Effects of common anti-epileptic drug monotherapy on serum levels of homocysteine, vitamin B_{12}, folic acid and vitamin B_6. Seizure 15, 79–85.
Sezikli, M., Çetinkaya, Z.A., Güzelbulut, F., et al., 2012. Supplementing vitamins C and E to standard triple therapy for the eradication of *Helicobacter pylori*. J. Clin. Pharm. Therapeut. 37, 282–285.
Shalansky, S., Lynd, L., Richardson, K., Ingaszewski, A., Kerr, C., 2007. Risk of warfarin-related bleeding events and supratherapeutic international normalized ratios associated with complementary and alternative medicine: a longitudinal analysis. Pharmacotherapy 27, 1237–1247.
Shiomi, M., Matsuki, S., Ikeda, A., et al., 2014. Effects of a protein-rich drink or a standard meal on the pharmacokinetics of elvitegravir, cobicistat, emtricitabine and tenofovir in healthy Japanese male subjects: a randomized, three-way crossover study. J. Clin. Pharmacol. 54, 640–648.
Smith, J.A., Butler, T.C., Poole, D.T., 1973. Effect of protein depletion in Guinea-pigs on glucoronate conjugation of chloramphenicol by liver microsomes. Biochem. Pharmacol. 22, 981–983.
Song, I., Borland, J., Arya, N., Wynne, B., Piscitelli, S., 2015. Pharmacokinetics of dolutegravir when administered with mineral supplements in healthy adult subjects. J. Clin. Pharmacol. 55, 490–496.
Sood, A., Sood, R., Brinker, F.J., et al., 2008. Potential for interactions between dietary supplements and prescription medications. Am. J. Med. 121, 207–211.
Souverein, P.C., Webb, D.J., Weil, J.G., Van Staa, T.,P., Egberts, A.C., 2006. Use of antiepileptic drugs and risk of fractures: case control study among patients with epilepsy. Neurology 66, 1318–1324.
Stork, C.M., Marraffa, J.M., Ragosta, K., Wojcik, S.M., Angelino, K.L., 2011. Elevated international normalized ratio associated with long-term azithromycin therapy in a child with cerebral palsy. Am. J. Health Syst. Pharm. 68, 1012–1014.
Sun, M.G., Huang, Y., Xu, Y.H., Cao, Y.X., 2017. Efficacy of vitamin B complex as an adjuvant therapy for the treatment of complicated vulvovaginal candidiasis: an in vivo and in vitro study. Biomed. Pharmacother. 88, 770–777.
Sutton, S.C., Nause, R., Gandelman, K., 2017. The impact of gastric pH, volume, and emptying on the food effect of ziprasidone oral absorption. AAPS J. 19, 1084–1090.
Tein, I., Dim Auro, S., Xie, Z.W., et al., 1993. Valproic acid impairs carnitine uptake in cultured human skin fibroblasts: an in vitro model for pathogenesis of valproic acid-associated carnitine deficiency. Pediatr. Res. 34, 281–287.
Theile, D., Hohmann, N., Kiemel, D., et al., 2017. Clementine juice has the potential for drug interactions: in vitro comparison with grapefruit and Mandarin juice. Eur. J. Pharmaceut. Sci. 97, 247–256.
Treluyer, J.M., Roux, A., Mugnier, C., Largadere, B., 1996. Metabolism of quinine in children with global malnutrition. Pediatr. Res. 40, 558–563.
Tulupule, A., Krishnaswamy, K., 1984. Chloroquine kinetics in the undernourished. Eur. J. Clin. Pharmacol. 24, 273–276.

U.S. Food and Drug Administration, 2021. Drug Databases: FDA Approved Drug Products. https://www.accessdata.fda.gov/scripts/cder/daf/. (Accessed 15 March 2021).
Van Wouwe, J.P., 1995. Carnitine deficiency during valproic acid treatment. Int. J. Vitam. Nutr. Res. 65, 211–214.
Varma, M.V., Steyn, S.J., Allerton, C., El-Kattan, A.F., 2015. Predicting clearance mechanism in drug discovery: extended clearance classification system (ECCS). Pharm. Res. 32, 3785–3802.
Vincent, J., Harris, S.I., Foulds, G., et al., 2000. Lack of effect of grapefruit juice on the pharmacokinetics and pharmacodynamics of amlodipine. Br. J. Clin. Pharmacol. 50, 455–463.
Vora, B., Green, E.A.E., Khuri, N., et al., 2020. Drug-nutrient interactions: discovering prescription drug inhibitors of the thiamine transporter ThTR-2 (SLC19A3). Am. J. Clin. Nutr. 111, 110–121.
Walter-Sack, I., Klotz, U., 1996. Influence of diet and nutritional status on drug metabolism. Clin. Pharmacokinet. 31, 47–64.
Wang, Y., Jia, X.M., Jia, J.H., et al., 2009. Ascorbic acid decreases the antifungal effect of fluconazole in the treatment of candidiasis. Clin. Exp. Pharmacol. Physiol. 36, e40–e46.
Wegler, C., Wikvall, K., Norlin, M., 2016. Effects of osteoporosis-inducing dugs on vitamin D-related gene transcription and mineralization in MG-63 and Saos-2 cells. Basic Clin. Pharmacol. Toxicol. 119, 436–442.
Welling, P.G., 1977. Influence of food and diet on gastrointestinal drug absorption: a review. J. Pharmacokinet. Biopharm. 5, 291–334.
Welz, T., Childs, K., Ibrahim, F., et al., 2010. Efavirenz is associated with severe vitamin D deficiency and increased alkaline phosphatase. AIDS 24, 1923–1928.
Werner, T., Treiss, I., Kohlmueller, D., et al., 2007. Effects of valproate on acylcarnitines in children with epilepsy using ESI-MS/MS. Epilepsia 48, 72–76.
White, R., 2010. Drugs and nutrition: how side effects can influence nutritional intake. Proc. Nutr. Soc. 69, 558–564.
Wiehe, L., Cremer, M., Wisniewska, M., et al., 2016. Selenium status in neonates with connatal infection. Br. J. Nutr. 116, 504–513.
Wilder, B.J., Leppik, I., Hietpas, T.J., Cloyd, J.C., Randinitis, E.J., Cook, J., 2001. Effect of food on absorption of Dilantin Kapseals and Mylan extended phenytoin sodium capsules. Neurology 57, 582–589.
Wright, D.H., Pietz, S.L., Konstatinides, F.N., Rotschafer, J.C., 2000. Decreased in vitro fluoroquinolone concentrations after admixture with an enteral feeding formulation. J. Parenter. Enter. Nutr. 24, 42–48.
Wu, K.C., Lin, C.J., 2019. The regulation of drug-metabolizing enzymes and membrane transporters by inflammation: evidences in inflammatory diseases and age-related disorders. J. Food Drug Anal. 27, 48–59.
Xu, Y., Hashizume, T., Shuhart, M.C., et al., 2006. Intestinal and hepatic CYP3A4 catalyze hydroxylation of 1α25-dihydroxyvitamin D3: implications for drug-induced osteomalacia. Mol. Pharmacol. 69, 56–65.
Yamada, H., Ikushima, I., Nemoto, T., et al., 2018. Effects of a nutritional protein-rich drink on the pharmacokinetics of elvitegravir, cobicistat, emtricitabine, tenofovir alafenamide, and tenofovir compared with a standard meal in healthy Japanese male subjects. Clin. Pharmacol. Drug Dev. 7, 132–142.
Yu, J., Petrie, I.D., Levy, R.H., Ragueneau-Majlessi, I., 2019. Mechanisms and clinical significance of pharmacokinetic-based drug-drug interactions with drugs approved by the U.S. Food and Drug Administration in 2017. Drug Metab Disp 47, 135–144.
Zadak, Z., Hyspler, R., Ticha, A., Vlcek, J., 2013. Polypharmacy and malnutrition. Curr. Opin. Clin. Nutr. Metab. Care 16, 50–55.
Zhou, Y., Jiao, Y., Yu-Hui, W., et al., 2013. Effects of pyridoxine on the intestinal absorption and pharmacokinetics of isoniazid in rats. Eur. J. Drug Metab. Pharmacokinet. 38, 5–13.

Eicosanoids: Prostaglandins, leukotrienes, thromboxanes and related compounds

Ángel Gil[a,b,c] **and Luis Fontana**[a,b,c], [a] Department of Biochemistry and Molecular Biology 2, School of Pharmacy, Granada, Spain; [b] Institute of Nutrition and Food Technology "José Mataix", Biomedical Research Center, Parque Tecnológico Ciencias de la Salud, Granada, Spain; and [c] Instituto de Investigación Biosanitaria ibs.GRANADA, Complejo Hospitalario Universitario de Granada, Granada, Spain

© 2023 Elsevier Ltd. All rights reserved.

This is an update of G.E. Caughey, M.J. James, L.G. Cleland, Prostaglandins and Leukotrienes, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 104–110, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00238-5.

Introduction 93
Biosynthesis 94
Mechanism of action 97
Catabolism 98
Biological effects 98
Effects on the formed elements of the blood 98
Effects on the cardiovascular system 98
Effects on the uterine musculature 98
Effects on the gastrointestinal tract 99
Effects on the kidney 99
Effects on the bronchial tree 99
Effects on the central nervous system 99
Effects on the endocrine system 99
Effects on metabolism 99
Pathophysiological implications 99
Eicosanoids, inflammation and anti-inflammatory drugs 99
Eicosanoids and bronchial asthma 100
Eicosanoids and immune response 100
Eicosanoids and bone resorption 100
Eicosanoids and reproduction 100
Eicosanoids and arteriovenous duct persistence in neonates 100
Eicosanoids and wound healing 100
Eicosanoids and diet 101
Regulation of platelet aggregation 101
Conclusions 102
References 103

Key points

- To understand the nutritional and metabolic origin of eicosanoids, as well as the main aspects of their functions.
- To understand the fundamental biochemical processes involved in platelet aggregation.
- To understand the role of eicosanoids in inflammation and immune responses.
- To know the main pathophysiological implications of eicosanoids.

Introduction

Given the complexity of multicellular organisms, good intercellular communication is necessary to regulate metabolism, growth and other functions of the various organs and tissues. This is a very complex network of signals that is gradually becoming better understood. Many chemical compounds (e. g. hormones, cytokines and eicosanoids) are involved in intercellular communication. Alterations in the signaling networks cause diseases as important as cancer and diabetes (Johnson et al., 2020).

Eicosanoids are derivatives of polyunsaturated fatty acids of 20 carbon atoms, especially arachidonic and eicosapentaenoic acids. They include many types of molecules, such as prostaglandins, prostacyclins, thromboxanes, leukotrienes, lipoxins, resolvins and

other oxidized compounds. Although they are derived from fatty acids, they are not very lipid soluble and therefore act primarily on membrane receptors (Calder, 2020).

While hormonal signals function in an endocrine fashion, acting on cells far from those that produce them, eicosanoids act in a paracrine (on neighboring cells) or autocrine (on the same producing cell) way (Fredriksson and Schioth, 2005). They are involved in the regulation of numerous physiological processes and are therefore implicated in many pathological alterations (Calder, 2020). Their pharmacological interest is also noteworthy, because there are inhibitors of their synthesis with therapeutic activity (especially anti-inflammatory agents) (Nakahata, 2008; Casati et al., 2021).

The aim of this chapter is to explain the intercellular communication system mediated by eicosanoids.

Biosynthesis

Eicosanoids are formed from three polyunsaturated fatty acids of 20 carbon atoms: eicosatrienoic or dihomo-γ-linolenic (20:3 n-6), eicosatetraenoic or arachidonic (20:4 n-6) and eicosapentaenoic (20:5 n-3) acids (**Fig. 1**).

The first two are derived from linoleic acid, while the third is derived from α-linolenic acid. The latter two are essential fatty acids of 18 carbon atoms. The formation of the long-chain polyunsaturated acids from the essential fatty acids takes place by enzymatic processes of desaturation and chain elongation in the liver. Subsequently, these acids are exported by the liver to the rest of the tissues, being incorporated into the membrane phospholipids, preferably in position two. There is a wide variety of stimuli capable of triggering the hydrolysis of the ester bond at position two, so that the corresponding fatty acids are released. From dihomo-γ-linolenic acid, series 1 of eicosanoids originate; from arachidonic acid, series 2 of eicosanoids are derived, and from eicosapentaenoic acid, series 3 of eicosanoids (Calder, 2020).

The first eicosanoids investigated were those of a cyclic nature (prostanoids), formed by the activity of cyclooxygenase: prostaglandins (PG), thromboxanes (TX) and prostacyclins (PC). Subsequently, other derivatives of a linear nature were discovered, originating from the action of various lipooxygenases (LOX), among which leukotrienes (LT) and lipoxins (LX) stand out (Biringer, 2020). **Figs. 2** and **3** show the formation pathways of the main eicosanoids derived from arachidonic acid (series 2), which are the most important in our usual dietary conditions.

Series 1 and 3 of eicosanoids are structural analogs of arachidonic-derived series 2, with one less and one additional bond, respectively.

Probably, the main mechanism of the regulation of eicosanoid biosynthesis is the availability of arachidonic acid and the other precursors of series 1 and 3, which is very small under normal conditions. Therefore, the action of enzymes that release these acids from the phospholipids that contain them is critical. In turn, these enzymes are influenced by membrane transduction systems in response to a wide variety of hormonal and other stimuli (Biringer, 2020). Most of the receptors related to eicosanoid synthesis operate in connection with G proteins, although, in some cases, these proteins are not well characterized (Fredriksson and Schioth, 2005). As a result of the activation of this system, activation of phospholipase A_2 or phospholipase C occurs.

Phospholipase A_2 (PLA_2) is an enzyme that acts on various phospholipids, including phosphatidylcholine, and requires calcium ions for its activity. The calcium ions can either come directly from outside the cell (as occurs in the case of physical stimuli on the

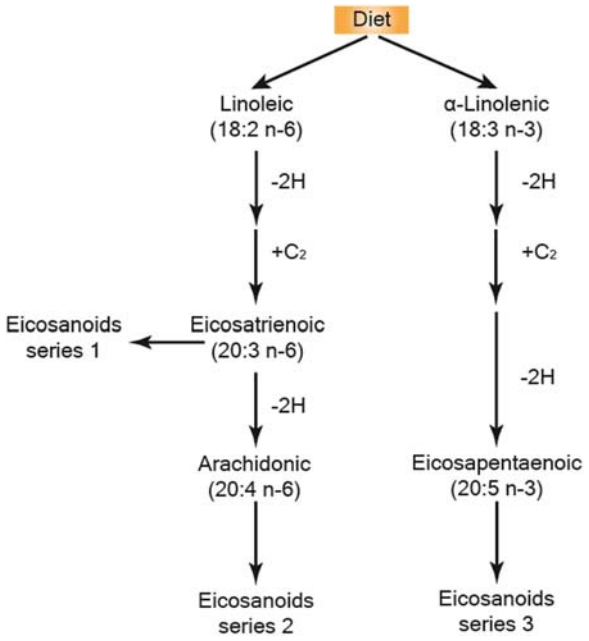


Fig. 1 Origin of series 1, 2 and 3 of eicosanoids.

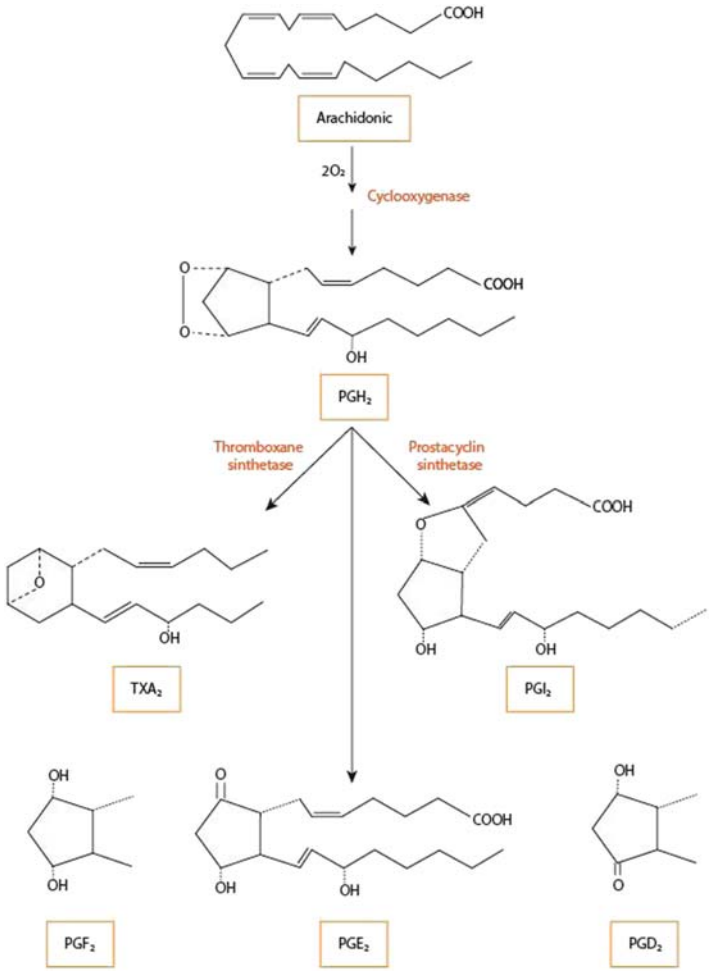


Fig. 2 Cyclic pathway of eicosanoid synthesis. PG: prostaglandin; TXA_2: thromboxane A_2.

membrane) or from the cytosol, once the activation of phospholipase C has occurred (see below). PLA_2 directly releases arachidonic acid from membrane phospholipids. Among the positive effectors is angiotensin II. In contrast, natural corticosteroids (and their pharmacological derivatives) inhibit this enzyme through the induction of a protein called lipocortin or lipomodulin.

Phospholipase C acts on phosphatidylinositol bisphosphate (PIP_2), releasing diacylglycerols and inositol triphosphate (IP_3). The latter releases calcium ions from the endoplasmic reticulum into the cytosol, which may result in further activation of PLA_2. On the other hand, there are lipases that can act on diacylglycerols, leading to the release of arachidonic acid from position two (Fig. 4).

The enzymes of the cyclic pathway of eicosanoid formation are found in the microsomal fraction of numerous tissues. In almost all tissues there is a prostaglandin synthetase or cyclooxygenase, known by the acronym COX-1, which is constitutive in nature, not inducible. Another form of cyclooxygenase, COX-2, is not normally expressed in cells, but can be induced by some growth factors or cytokines. Cyclooxygenase acts in two stages corresponding to two distinct enzymatic activities. First, with the intervention of molecular oxygen, a cyclic endoperoxide called prostaglandin G2 (PGG_2) is produced. Subsequently, the hydroperoxide group at position 15 is reduced to hydroxyl by peroxidase activity and the probable assistance of glutathione, giving rise to another cyclic endoperoxide called PGH_2 (Fig. 2). Cyclooxygenase is inhibited by aspirin and other non-steroidal anti-inflammatory drugs. COX-2 also appears to be inhibited by corticosteroids.

The remaining enzymes are not as ubiquitous. Some tissues, such as the lung, contain all of them and can synthesize the full range of prostanoids. In contrast, only D-series prostaglandins and thromboxanes can be synthesized in platelets, whereas the vascular endothelium synthesizes only prostacyclins (Biringer, 2020).

The enzymes of the linear pathway are found primarily in the cytosolic fraction of neutrophils (5-LOX and 12-LOX), eosinophils (15-LOX) and platelets (12-LOX). 5-LOX is the best known enzyme of this group. Its activation requires intracellular calcium ions and an activator protein that allows it to bind to the membrane.

There are three different lipoxygenases capable of synthesizing lipoxins from arachidonic acid: 5-LOX, 12-LOX and 15-LOX. One of the synthesis pathways takes place in platelets and involves the transformation of leukotriene A4 (LTA_4) into lipoxin A_4 (LXA_4) via 12-LOX. Another route involves 15-LOX and 5-LOX in neutrophils, erythrocytes and reticulocytes. In this case,

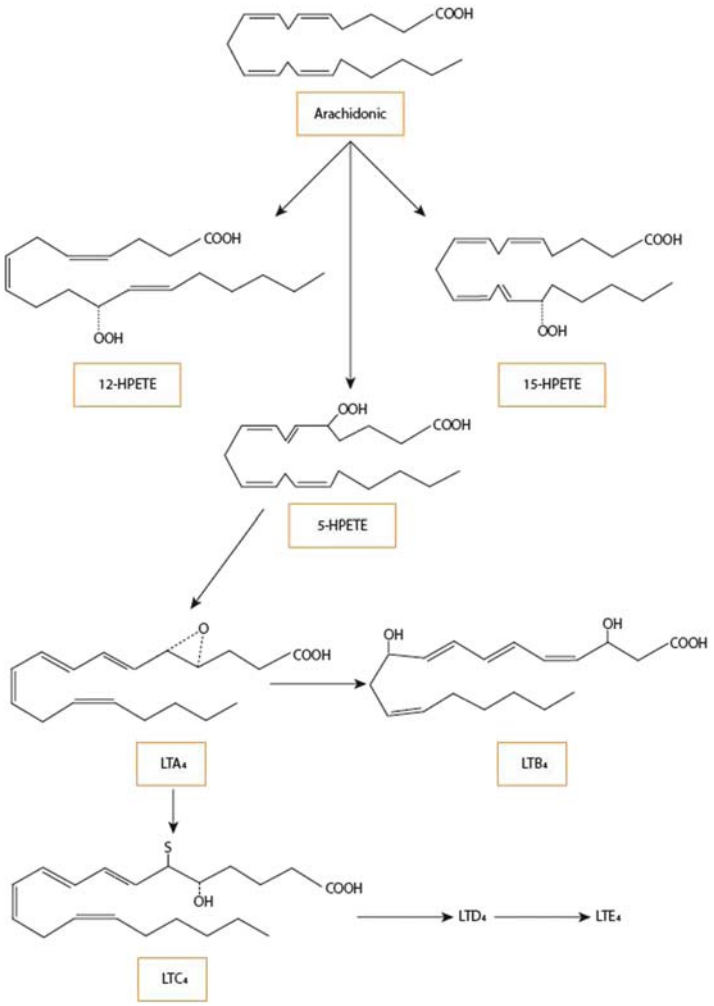


Fig. 3 Linear pathway of eicosanoid synthesis. HPETE: hydroxyperoxyeicosatetraenoic acid; LT: leukotriene.

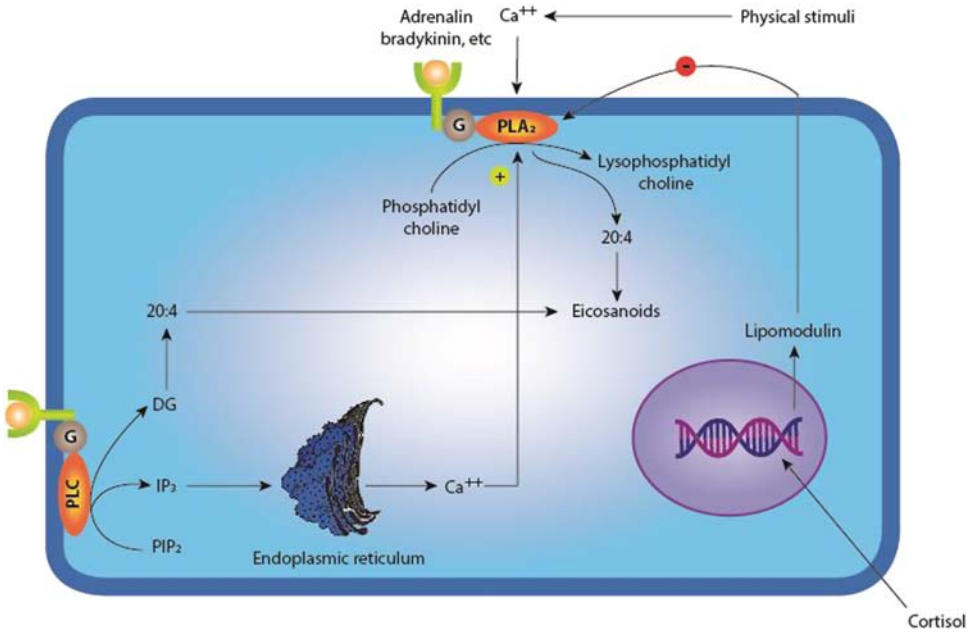


Fig. 4 Arachidonic acid release from membrane phospholipids. DAG: diacylglycerol; IP_3: inositol triphosphate; PIP_2: phosphatidylinositol bisphosphate; PLA_2: phospholipase A_2; PLC: phospholipase C.

15-hydroxyperoxyeicosatetraenoic acid (15-HPETE) is obtained from arachidonic acid, which is eventually converted to lipoxins A and B. The other route by which lipoxins can be generated is mediated by the action of aspirin, which gives rise to 15-epilipoxin A_4 (15-epi-LXA_4) and 15-epilipoxin B_4 (15-epi-LXB_4), also known as ATL (aspirin-triggered lipoxins) (Chandrasekharan and Sharma-Walia, 2015).

In addition to these compounds, there are biologically active derivatives produced by oxidation of arachidonic acid through the cytochrome P-450 system, although their physiological significance is less clear. On the other hand, there is the possibility of non-enzymatic transformation of arachidonic acid to prostaglandin-like compounds called isoprostanes. These eicosanoids also exhibit biological activity and may play a role in the inflammatory response, since they are formed by the action of free radicals released in this process (Kumar et al., 2020).

In recent years, derivatives of long-chain polyunsaturated fatty acids capable of binding to cannabinoid receptors and referred to as endocannabinoids, have gained special relevance. Anandamide or arachidonylethanolamide is an ethanolamide derivative of arachidonic acid and a member of a larger class of N-acylethanolamines (NAEs) lipids, and thus also referred to as NAE 20:4. Other endocannabinoids are N-homo-γ-linolenoylethanolamine (NAE 20:3(n-6)), N- docosatetraenoylethanolamine (NAE 22:4(n-6)) and 2-arachidonyl glycerol (Kilaru and Chapman, 2020).

Endocannabinoid signaling has been implicated in a number of physiological and cognitive processes in humans. The endocannabinoids are well-known modulators of various neurological functions such as perception of pain, mood, memory, sleep, appetite etc. Endocannabinoids also play an important role in activation of immune system, pre- and postnatal development, embryo implantation and fertility, among others. Thus, maintaining physiological homeostasis is considered as the primary function of endocannabinoid signaling. The biological consequences of endocannabinoids result from their interactions with several target proteins, including cannabinoid receptors, G-proteins and other downstream effectors. As such, synthesis, receptor activation and degradation of endocannabinoids are all highly regulated and is believed to start "on-demand" via a stimulus-dependent synthesis (Kilaru and Chapman, 2020).

Mechanism of action

There are membrane receptors for each type of eicosanoid, which have been studied especially in platelets and smooth muscle. It is noteworthy that there can be, however, several types of receptors for the same eicosanoid, even in the same cell membrane. In all cases, membrane signal transduction is carried out by G proteins (Fredriksson and Schioth, 2005). In some cases, stimulation or inhibition of adenylate cyclase occurs, increasing or decreasing cyclic AMP production. In other cases there is stimulation of phospholipase C, with a consequent increase in diacylglycerols and phosphoinositol bisphosphate. Leukotrienes appear to act through stimulation of phospholipase C, while PGI_2 (prostacyclin) and prostaglandins D and F activate adenylate cyclase.

On the other hand, PGE_2 and thromboxane A_2 (TXA_2), as well as isoprostanes, oxidized derivatives of eicosanoids, can act through both systems. The action of TXA_2 occurs through its TP receptor that interacts with different G proteins, which causes the activation of different enzymes, such as adenylate cyclase and phospholipases C of type a and d, as well as the Rho guanine nucleotide exchange protein (Rho-GEF) (Nakahata, 2008). In a very general way it can be said that stimulation of the phosphoinositol bisphosphate system and the subsequent release of calcium ions result in constriction phenomena, whereas the production of cyclic AMP results in relaxation (**Fig. 5**).

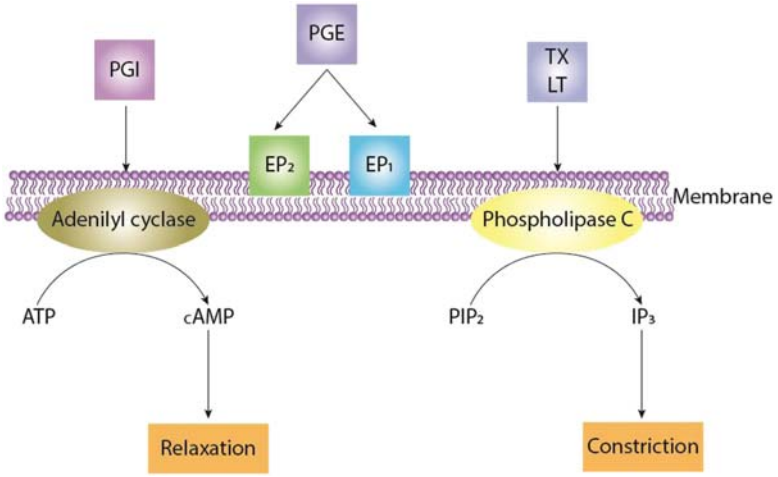


Fig. 5 Mechanism of action of eicosanoids. cAMP: cyclic adenosine monophosphate; ATP: adenosine triphosphate; PD: prostaglandin E receptor; IP_3: inositol triphosphate; LT: leukotriene; PIP_2: phosphatidylinositol bisphosphate; PGE: prostaglandin E; PGI: prostacyclin.

Recently, it has become clear that some eicosanoids can act on nuclear receptors of the peroxisome proliferator-activated receptor (PPAR) type. Calcium-activated PLA_2 can translocate to the nuclear membrane and release arachidonic acid or other polyunsaturated fatty acids, leading to the formation of eicosanoids, such as the prostacyclin PGI_2, which interact with nuclear transcription factors such as PPARs and retinol receptors (RXRs), which bind to DNA and modulate the expression of numerous genes (Katusic et al., 2012).

Lipoxins exert their action through their interaction with the high-affinity receptor ALX, also known as FPR-2, curiously the same receptor to which the serum amyloid protein binds. When the interaction between lipoxins and their receptor ALX/FPR-2 occurs, their internalization is then triggered, which induces phagocytosis of polymorphonuclear cells, thanks to the regrouping of actin filaments. On the other hand, several intracellular signaling pathways are activated with important consequences, such as inhibition of the action of B-lymphocyte nuclear factor kappa (NF-κB), a key factor in the development of inflammation. In addition, the growth factor EGR-1 (early growth response protein 1) controls the production of proinflammatory cytokines, such as interleukin-2 (IL-2) or tumor necrosis factor-α (TNF-α). This transcription factor is inhibited by another called NAB-1 (Nerve growth factor 1A-binding protein). Several studies have shown that lipoxins secreted by neutrophils are able to potentiate the synthesis of NAB-1 to the detriment of EGR-1 to control inflammation. On the other hand, lipoxins also exert their action on another transcription factor that helps to enhance inflammation: PPAR-γ (Chandrasekharan and Sharma-Walia, 2015).

Lipoxins stop polymorphonuclear leukocyte infiltration and limit dendritic cell migration by inhibiting IL-1 synthesis by antigen-presenting cells (APCs). Likewise, lipoxins cause an inhibition of proinflammatory cytokine production by Th1 cells, resulting in an inhibition of proinflammatory type 1 macrophages. In addition, lipoxins stimulate the recruitment of monocytes and the uptake of apoptotic polymorphonuclears by type 2 macrophages, a process that also involves mediators of inflammation resolution called maresins. Finally, lipoxins inhibit the production of metalloproteases by fibroblasts, limiting their tissue-destroying action (Chandrasekharan and Sharma-Walia, 2015).

Catabolism

Eicosanoids have a very short half-life, degrading very rapidly once they have been synthesized. In general, inactivation takes place in a first phase by the action of specific enzymes, which act very rapidly on the various types of eicosanoids. In a second, slower phase, the metabolites resulting from the first inactivation continue their catabolism by non-specific enzymes, some of which are those that usually act in the degradation of fatty acids (β-oxidation system). The main site of eicosanoid catabolism is the lung (Yonker et al., 2021).

Biological effects

As noted above, eicosanoids have a very short half-life, so they act only in the cells that produce them or in their immediate environment. Their biological effects are, in general, broad and intense. It is important to note, however, that these effects are highly variable. There are differences depending on the species, tissue or organ, as well as between the different types of eicosanoids. But there may even be opposing effects, depending on the amounts of a given compound used. The most prominent effects of eicosanoids are described below (Kumar et al., 2020).

Effects on the formed elements of the blood

As will be described in detail later, PGI_2 inhibits platelet aggregation, while TXA_2 stimulates it. LTB_4 is a potent chemotactic agent on polymorphonuclear leukocytes, monocytes and eosinophils, which contributes significantly to the inflammatory reaction. In addition to activating phagocytes, LTB_4 is involved in the differentiation of T lymphocytes into Th1 and Th2 in the early stages of the inflammatory process and stimulates the adhesion of neutrophils to the vascular endothelium and their transendothelial migration to the inflammatory focus. In contrast, lipoxins act in the opposite way on neutrophils. PGE_2, at high concentrations, inhibits the differentiation of B lymphocytes and the proliferation of T lymphocytes, thus behaving as an immunosuppressant (Imig, 2020).

Effects on the cardiovascular system

Prostaglandins are vasodilators in general, especially those of the E series, although exceptions occur in certain vascular areas. Generally, they also lower blood pressure and increase cardiac output. The hypotensive effect is very marked for PGI_2. In contrast, PGF_2, LTC_4 and LTD_4 and TXA_2 are vasoconstrictors and hypertensors (Piper and Garelnabi, 2020).

Effects on the uterine musculature

Prostaglandins of the E and F series, as well as TXA_2, contract the uterine musculature. In fact, prostaglandins E and F are inducers of labor and can be used in therapeutic abortion (Szczuko et al., 2020).

Effects on the gastrointestinal tract

PGI_2 and PGE_2 inhibit acid secretion in the stomach. Prostaglandins of the E series also increase mucus secretion in the stomach and small intestine and stimulate intestinal motility.

Effects on the kidney

PGE_2 and PGI_2 decrease vascular resistance and increase renal blood flow. In addition, PGE_2 inhibits the reabsorption of sodium chloride in the cortical collecting tubes, hinders the action of antidiuretic hormone and stimulates the release of renin. TXA_2, which is vasoconstrictor, is also produced in the kidney.

Effects on the bronchial tree

Prostaglandins of the E series relax the musculature of the bronchi and trachea, while those of the F series produce bronchospasm. The bronchoconstrictor action is especially important in the case of the leukotrienes LTC_4 and LTD_4 (Yonker et al., 2021).

Effects on the central nervous system

PGE_2, PGI_2 and LTB_4 act on afferent nerve endings, lowering the pain threshold at nociceptors. These compounds can, therefore, amplify the sensation of pain in inflammatory processes.

Effects on the endocrine system

The prostaglandins of the E series exert a multitude of actions on the endocrine system, being able to emphasize the stimulus on the production of adrenocorticotropic hormone (ACTH), growth hormone, gonadotropins and prolactin, as well as the inhibition of insulin secretion. On the other hand, 12-hydroxy eicosatetraenoic acid (12-HETE) seems to favor the latter secretion.

Effects on metabolism

In addition to the effects derived from the action of eicosanoids on the endocrine system, it is interesting to note the direct anti-lipolytic effect on adipose tissue of prostaglandins of the E series.

Pathophysiological implications

Given the diversity of biological actions of eicosanoids and taking into account the fact that most of the cells of the organism are capable of synthesizing them, it is not surprising that these compounds are involved in numerous pathological processes. In most cases, functional alterations seem to be due to excess production, although there can also be problems due to deficiencies. In this case, the cause may be a deficiency of precursors, but they are often the result of certain drug treatments that prevent their tissue synthesis.

Eicosanoids, inflammation and anti-inflammatory drugs

Some eicosanoids are important biochemical mediators in the inflammatory reaction. As noted above, LTB_4 has a very marked chemotactic activity on various blood cells involved in this process, especially polymorphonuclear cells and monocytes. Once they have crossed the endothelium, these cells reach the inflammatory focus and can perform their characteristic phagocytic functions. On the other hand, PGE_2 and PGI_2 are responsible for the hyperemia typical of inflammatory lesions due to their vasodilator character and contribute to pain and fever, amplifying the action of other mediators such as bradykinin.

Lipoxins intervene, above all, in the resolution of inflammation, because they impede the flow of neutrophils to the inflammatory focus. Some recently discovered polyunsaturated fatty acid derivatives of the n-3 series, including E-resolvins, derived from EPA, and the docosanoids D-resolvins, protectins and maresins, are also involved in the resolution of inflammation (Chandrasekharan and Sharma-Walia, 2015).

There is a large number of anti-inflammatory drugs, most of which inhibit the synthesis of eicosanoids. Some, such as aspirin and other non-steroidal anti-inflammatory drugs, exert their mechanism of action through the inhibition of cyclooxygenase, which prevents the formation of prostaglandins, prostacyclins and thromboxanes. Others, such as cortisol derivatives, also suppress the formation of the linear series eicosanoids, because they act by inhibiting COX-2 and phospholipase A_2 (steroidal anti-inflammatory drugs). All these drugs are widely used in the symptomatic treatment of inflammation, often on a chronic basis (Kumar et al., 2020).

Logically, the suppression of eicosanoid biosynthesis does not exclusively affect the inflamed regions, but also produces a decrease in this biosynthesis in other organs and tissues, causing the corresponding pathological alterations. The side effects of continuous or high-dose use of anti-inflammatory drugs are particularly evident in the stomach and kidney.

Indeed, prostaglandins, especially PGE_2 and PGI_2, inhibit the secretion of hydrochloric acid and pepsin and promote mucus formation. Therefore, they can be considered cytoprotective of the gastric mucosa. Inhibition of the synthesis of these prostaglandins by the use of anti-inflammatory drugs can therefore lead to ulceration and bleeding.

In the kidney, PGE_2 and PGI_2 are produced in significant amounts when perfusion decreases, causing compensatory vasodilatation. In addition, sodium chloride reabsorption is inhibited and there is an inhibitory effect on antidiuretic hormone, increasing diuresis. On the other hand, these prostaglandins stimulate the release of renin, which represents a potential feedback mechanism. Inhibition of prostaglandin synthesis causes less vasodilatation, eventually leading to renal lesions. Blood pressure does not change much because, although there is less diuresis, renin production also decreases. However, the decrease in water excretion leads to edema. And the reduced functioning of the renin-angiotensin system produces hyperkalemia.

Eicosanoids and bronchial asthma

Although the different types of prostaglandins have opposing effects on the bronchial tree (see above), the action of peptidyl-leukotrienes, LTC_4 and LTD_4, which are very potent bronchoconstrictors, strongly predominates. Their release by sensitized cells, in the anaphylactic process, strongly contributes to the clinical picture, especially during the intercritical periods. Precisely, before their isolation, peptidyl-leukotrienes were called "slow reacting substances of anaphylaxis" (Sokolowska et al., 2021).

Eicosanoids and immune response

Eicosanoids appear to be involved in the regulation of the immune response, although many aspects of this regulation are still unknown. In any case, it is well established that PGE_2 is necessary for lymphocyte proliferation, but it behaves as an immunosuppressant when found in high concentrations, for example, when polyunsaturated fatty acids of the n-6 series are used in parenteral nutrition. This prostaglandin decreases, especially, the effects of IL-1 and IL-2 on T lymphocytes. The immunosuppressive effect is counterproductive in patients undergoing parenteral nutrition, because it favors postoperative infections. However, it can be useful when it comes to slowing down transplant rejection (Calder, 2020).

Eicosanoids and bone resorption

Some tumors that do not affect the parathyroid glands, such as medullary thyroid carcinoma or mammary gland carcinoma, cause hypercalcemia due to bone resorption. It seems likely that the mediator for this effect is PGE_2, which is released in large quantities from the tumor. This prostaglandin may also be the cause of hypotension and other systemic effects that occur in these conditions. Other sources of PGE_2 that also cause bone resorption are synovial fluid in rheumatoid arthritis and certain dental cysts.

Eicosanoids and reproduction

Prostaglandins of the F series strongly contract the uterine musculature and sensitize afferent pathways for pain. Therefore, prostaglandin synthesis inhibitors are used to relieve symptoms in dysmenorrhea. Conversely, using the same effect on the uterus, structural analogs of these prostaglandins have been tested for therapeutic abortions. There seems to be a relationship between the concentration of prostaglandins in the seminal fluid and male fertility, although the data are not yet incontrovertible. On the other hand, the vasodilator effect of PGE1 has been used to achieve penile erection by intracavernous injection (Szczuko et al., 2020).

Eicosanoids and arteriovenous duct persistence in neonates

PGI_2 is capable of dilating this duct in newborns. For this reason, it is used palliatively while waiting for surgery, in cases in which pulmonary circulation is affected by the existence of heart disease, thus improving tissue oxygenation. In contrast, prostaglandin synthesis inhibitors are indicated when there is pathological persistence of the arteriovenous duct.

Eicosanoids and wound healing

Among arachidonic acid metabolites, several COX and cytochrome P-450 products have demonstrated a positive effect on skin wound healing through various activities, including modulation of clotting, macrophage polarization, keratinocyte migration, and angiogenesis, and the metabolites produced by the 5-LOX pathway such as LTs delayed the process of skin wound healing partially via the mechanisms of excessive inflammation, including neutrophil recruitment and reduced macrophage polarization to M2 phenotype (Yasukawa et al., 2020).

Eicosanoids and diet

It is important to note that the relative proportions of the different series of eicosanoids in the body depend on the type of diet. The usual diet in the Western world, based on vegetables and terrestrial animals, leads to a preponderance of series 2, derived from arachidonic acid. It should be remembered that this compound, in turn, is derived from linoleic acid, which is abundant in the vegetable world. On the other hand, since α-linolenic acid is very abundant in algae, aquatic animals are rich in this acid and its long-chain derivatives, including eicosapentaenoic acid (EPA). Therefore, the intake of fatty fish in relatively high quantities leads to a significant increase in eicosanoids of the 3 series in the organism. Series 1 of eicosanoids, i.e., derived from eicosatrienoic acid (20:3 n-6), are not very abundant, unless large amounts of γ-linolenic acid are ingested. This compound, which is found only in some terrestrial plant species, such as evening primrose (*Oenothera biennis*) or borage (*Borago officinalis*), is easily converted to eicosatrienoic acid, which greatly favors the formation of series 1 of eicosanoids.

In general, eicosanoids of series 2, the most abundant in our organism given the characteristics of our diet, are very active. On the contrary, series 3 of eicosanoids tend to have less biological activity. Particularly, TXA_3 is very little platelet proaggregant and LTB_5 is very little chemotactic on leukocytes. As has just been pointed out, these compounds originate in certain quantities when the intake is very rich in polyunsaturated acids of the n-3 series, derived from fish and other aquatic animals. Under these conditions, certain pathophysiological processes can be altered, as described below, when studying the regulation of platelet aggregation.

Regulation of platelet aggregation

Platelets are corpuscles without nuclei derived from the fragmentation of megakaryocytes, which play a fundamental role in hemostasis. Platelet activity is triggered by primary stimuli (collagen, thrombin, adrenaline) and by substances that can be released by activated platelets themselves (serotonin, ADP, calcium ions, platelet activating factor (PAF) and TXA_2). There are also inhibitory factors, such as prostacyclin (PGI_2) and nitric oxide (NO). All these substances appear to have specific receptors on the platelet membrane (Imig, 2020). **Fig. 6** summarizes the mechanisms of the platelet response following stimulation of some of these receptors.

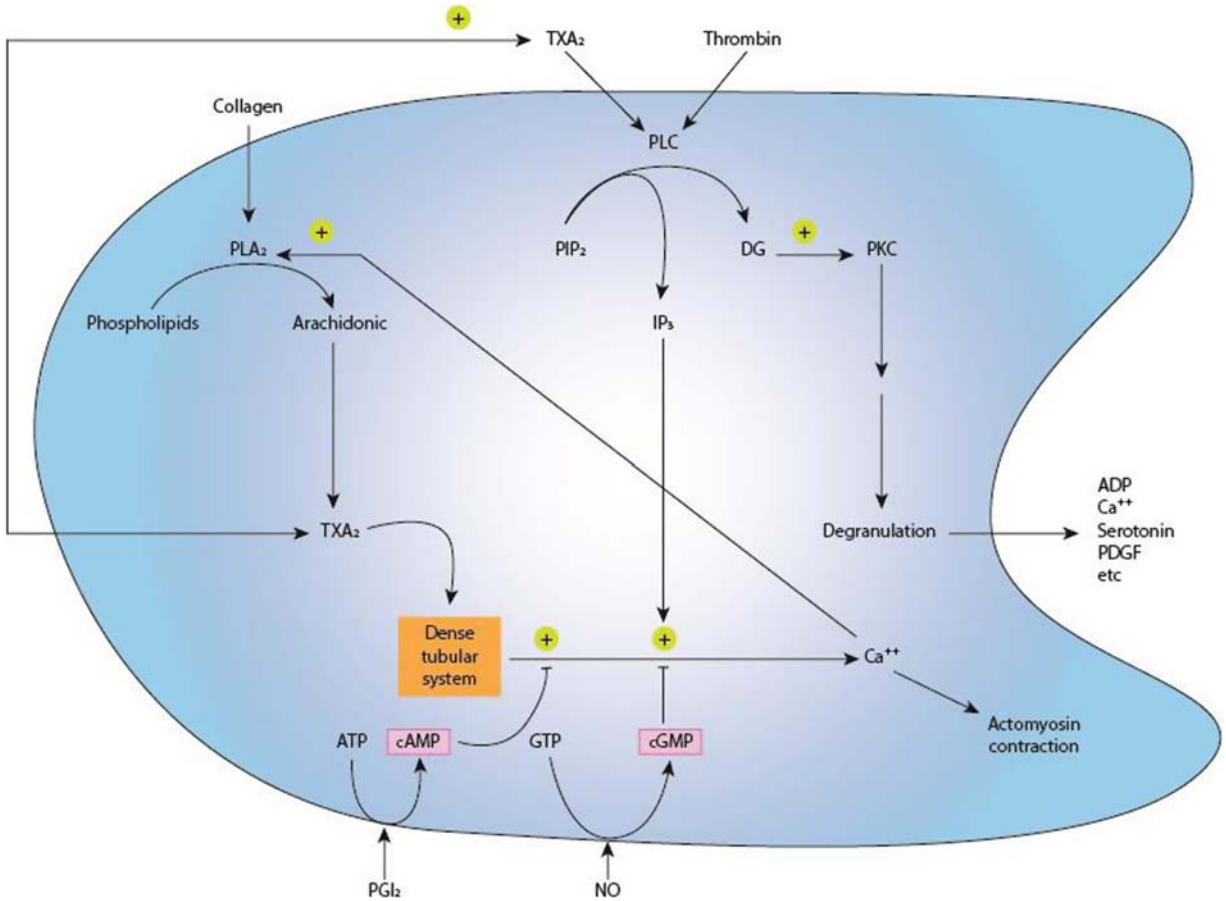


Fig. 6 Regulation of platelet aggregation. DAG: diacylglycerol; IP_3: inositoltriphosphate; NO: nitric oxide; PDGF: platelet-derived growth factor; PGI_2: prostacyclin; PIP_2: phosphatidylinositol bisphosphate; PKC: protein kinase C; PLA_2: phospholipase A_2; PLC: phospholipase C; TXA_2: thromboxane A_2.

The interaction between primary stimuli and platelet substances with receptors results in the release of calcium ions from the dense tubular system into the cytoplasmic space. In this cellular compartment, calcium ions stimulate calmodulin, leading to phosphorylation of the myosin light chains, which finally causes the actomyosinic contraction responsible for platelet shape changes and platelet movements.

The release of calcium ions into the cytosol can be considered the fundamental mechanism for triggering platelet activity. The main stimulus for this release appears to be IP_3, which is produced when phospholipase C is activated by thrombin, ADP, calcium ions themselves or TXA_2. This last compound can be considered as the main positive regulator of platelet activity, since it is capable, in turn, of directly intervening in the release of calcium ions from the dense tubular system into the cytosolic space and, in addition, once it is synthesized inside the platelet, it can exit into the plasma and stimulate other platelets. The compounds that trigger thromboxane formation are collagen, PAF and probably calcium ions and ADP. All these effectors activate phospholipase A_2, arachidonic acid is released from membrane phospholipids and thromboxane is produced as the main product, thanks to the intense activity of thromboxane synthetase in the platelet.

Activation of phospholipase C by thrombin and the other effectors mentioned above produces 2 s messengers: IP_3 and diacylglycerol. The latter activates protein kinase C, which causes phosphorylation of a platelet protein, pleckstrin, which acts on the alpha and dense granules, producing the release of their contents. The dense granules contain, among other substances, ADP, calcium ions and serotonin, all of which stimulate platelet activity. The content of the alpha granules is more related to the coagulation phenomenon (platelet factor 4, β-thromboglobulin, platelet-derived growth factor [PDGF], fibrinogen, etc.). It is worth noting that the release of PDGF plays a key role in the growth of the atheroma, when platelets adhere to it as a consequence of its rupture, since this growth factor acts by strongly stimulating the proliferation of smooth muscle cells.

In the platelet there is also a system for reuptake of calcium ions into the tubular system, a process that is stimulated by cyclic AMP. The synthesis of this nucleotide increases when the adenylate cyclase of the platelet membrane is activated by the action of prostacyclin (PGI_2). This eicosanoid is the main product of arachidonic acid in endothelial cells after activation of phospholipase A_2. Therefore, the release of prostacyclin by the endothelium constitutes a negative regulatory system of aggregation. The role of endothelial cells in the control of platelet activity is not limited to prostacyclin release (Katusic et al., 2012). These cells can also generate nitric oxide, which is capable of interacting with platelets and stimulating the reuptake of calcium ions, in this case through cyclic GMP.

Although the mechanisms are more complicated, it can be concluded that platelet aggregation is largely regulated by the relationship between the concentrations of two eicosanoids: thromboxane and prostacyclin. It is interesting to remember that TXA_2 is also a vasoconstrictor, whereas prostacyclin is a vasodilator. Therefore, any factor that alters the delicate balance between platelet production of TXA_2 and endothelial release of PGI_2 can affect the aggregation process, as well as the state of constriction or dilatation of the corresponding artery. Thus, when the endothelium is subjected to aggression, prostacyclin production decreases and the balance shifts toward aggregation. This is what occurs both in the development of the atherogenic process and in the thrombotic phenomena that trigger clinical problems (Nakahata, 2008).

As previously described, the type of diet can condition the proportion of the different series of eicosanoids in the organism. Thus, the intense consumption of fats of marine origin and, consequently, the increase in polyunsaturated fatty acids of the n-3 series, will give rise to important quantities of eicosanoids of the 3 series. Under these conditions, platelets form TXA_3, while endothelial cells generate PGI_3. The latter retains an anti-aggregating activity very similar to that of PGI_2. In contrast, TXA_3 is very poorly aggregative. As a result, it decreases platelet activity, as evidenced even by the occurrence of spontaneous nosebleeds. In fact, populations that consume many marine animals (such as Eskimos or the inhabitants of certain Japanese islands) have a very low incidence of thrombotic problems.

It is also interesting to note that platelet activity can be decreased by administration of very low doses of acetylsalicylic acid. This drug binds irreversibly and covalently to platelet cyclooxygenase, completely inhibiting its activity. This inhibition prevents the production of thromboxane during the entire platelet lifetime, because, lacking a nucleus, the disabled enzyme cannot be replaced. In contrast, the inhibition of endothelial cyclooxygenase is not permanent, since these cells have the capacity to synthesize new enzymatic proteins. The use of very low doses of the drug allows inhibition of platelet thromboxane synthesis and has little effect on endothelial prostacyclin synthesis, achieving a good antiaggregant effect (Kumar et al., 2020).

Conclusions

Prostaglandins, prostacyclins, thromboxanes and leukotrienes belong to a large, heterogeneous group of lipid mediators, collectively named eicosanoids, that exhibit a diverse array of physiological activities. Eicosanoids are synthesized by oxygenation and remodeling of their precursor 20-carbon polyunsaturated fatty acids, namely arachidonic acid. Although pivotally involved in many homeostatic processes, eicosanoids are also implicated in the pathophysiology of many chronic disorders. Since the discovery in the mid- 1930s of prostaglandins as a component of human semen that potently induced uterine contractility, the field of eicosanoid biology has expanded to include other groups such as hydroxyeicosatetraenoic acids, lipoxins, isoprostanes, and the cyclopentanone prostaglandins. Except for the latter two classes, which are generated by nonenzymatic oxidation, synthesis of these mediators is tightly regulated by a number of enzymes. Eicosanoids generally act as paracrine or autocrine agents, in that they exert their biological effects locally, either on the cell from which they were synthesized or on neighboring cells.

References

Biringer, R.G., 2020. The enzymology of human eicosanoid pathways: the lipoxygenase branches. Mol. Biol. Rep. 47, 7189–7207.
Calder, P.C., 2020. Eicosanoids. Essays Biochem. 64, 423–441.
Chandrasekharan, J.A., Sharma-Walia, N., 2015. Lipoxins: nature's way to resolve inflammation. J. Inflamm. Res. 8, 181–192.
Casati, S., Giannasi, C., Niada, S., Bergamaschi, R.F., Orioli, M., Brini, A.T., 2021. Bioactive lipids in MSCs biology: state of the art and role in inflammation. Int. J. Mol. Sci. 22, 1481.
Fredriksson, R., Schioth, H.B., 2005. The repertoire of G-protein–coupled receptors in fully sequenced genomes. Mol. Pharmacol. 67, 1414–1425.
Imig, J.D., 2020. Eicosanoid blood vessel regulation in physiological and pathological states. Clin. Sci. 134, 2707–2727.
Johnson, A.M., Kleczko, E.K., Nemenoff, R.A., 2020. Eicosanoids in cancer: new roles in immunoregulation. Front. Pharmacol. 11, 595498.
Katusic, Z.S., Santhanam, A.V., He, T., 2012. Vascular effects of prostacyclin: does activation of PPARδ play a role? Trends Pharmacol. Sci. 33, 559–564.
Kilaru, A., Chapman, K.D., 2020. The endocannabinoid system. Essays Biochem. 64, 485–499.
Kumar, V., Abbas, A., Aster, J. (Eds.), 2020. Robbins & Cotran Pathologic Basis of Disease, tenth ed. Elsevier, New York.
Nakahata, N., 2008. Thromboxane A_2: physiology/pathophysiology, cellular signal transduction and pharmacology. Pharmacol. Ther. 118, 18–35.
Piper, K., Garelnabi, M., 2020. Eicosanoids: atherosclerosis and cardiometabolic health. J. Clin. Transl. Endocrinol. 19, 100216.
Sokolowska, M., Rovati, G.E., Diamant, Z., Untersmayr, E., Schwarze, J., Lukasik, Z., Sava, F., Angelina, A., Palomares, O., Akdis, C.A., O'Mahony, L., Sanak, M., Dahlen, S.E., Woszczek, G., 2021. Current perspective on eicosanoids in asthma and allergic diseases: EAACI Task Force consensus report, part I. Allergy 76, 114–130.
Szczuko, M., Kikut, J., Komorniak, N., Bilicki, J., Celewicz, Z., Ziętek, M., 2020. The role of arachidonic and linoleic acid derivatives in pathological pregnancies and the human reproduction process. Int. J. Mol. Sci. 21, 9628.
Yasukawa, K., Okuno, T., Yokomizo, T., 2020. Eicosanoids in skin wound healing. Int. J. Mol. Sci. 21, 8435.
Yonker, L.M., Barrios, J., Mou, H., Hurley, B.P., 2021. Untapped potential: therapeutically targeting eicosanoids and endocannabinoids in the lung. Clin. Pharmacol. Ther. https://doi.org/10.1002/cpt.2165.

Electrolytes: Acid–base balance

P.B. Mark, K.K. Stevens, and A.G. Jardine, University of Glasgow, Glasgow, UK

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of P.B. Mark, K.K. Stevens, A.G. Jardine, Electrolytes: Acid-Base Balance, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 139–145, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00087-8.

Definitions, Acids, Bases, and Buffers **105**
pH 105
Acids and Bases 105
Buffers 105
Maintenance of the pH of Blood and Extracellular Fluids **106**
Acid and Alkali Load 106
Regulation **106**
Blood/Extracellular Fluid 106
Lungs 106
Kidneys 107
Liver and Bone 108
Measurement of Urinary Acid Excretion **108**
Effects of Acid–Base Disturbance **108**
Abnormalities in Acid–Base Balance **109**
Metabolic Acidosis 109
Diabetic Ketoacidosis 109
Lactic Acidosis 110
Excess Bicarbonate Loss 110
Renal Failure 111
Drugs and Other Causes 111
Compensation **111**
Treatment **111**
Metabolic Alkalosis 111
Respiratory Acidosis 111
Respiratory Alkalosis 112
Transporter Mechanisms: Physiology and Pathophysiology **112**
Further Reading **112**

Glossary

Acidemia The presence of a low blood pH.

Acidosis The clinical term for a state of excess acid in the blood, further defined by its etiology, either metabolic or respiratory in origin.

Alkalemia The presence of an elevated blood pH.

Alkalosis The clinical term for a state of excess alkali in the blood, further defined by its etiology, either metabolic or respiratory in origin.

Buffer Solution of a salt of a weak acid, which is able to bind hydrogen ions.

Renal tubular acidosis A group of conditions where metabolic acidosis results from diminished tubular secretion of hydrogen ions by the kidney.

Maintenance of cellular and extracellular pH (hydrogen ion concentration) is essential to life, in view of the exquisite pH dependence of processes such as enzyme function. Hydrogen ions (H^+) are generated by cellular metabolism and, to a lesser extent by the ingestion of acids in the diet. Acid–base homeostasis regulates pH between 7.36 and 7.44 (corresponding to a [H+] of 36–44 nmol l^{-1}) in extracellular fluids, such as blood, whereas intracellular pH is more acidic (pH 6.3–7.4) depending on individual organs and circumstances. The pH of subcellular organelles may be yet more acidic, reflecting their physiological function (e.g., lysosomes). Blood and extracellular fluid pH are tightly regulated by the presence of buffer systems, which effect change as a consequence of acid load. These buffer systems, both extracellular and intracellular, include hemoglobin, other proteins, phosphate, and bicarbonate – the latter being most important. However, the acid load must ultimately be eliminated by subsequent excretion of volatile acids by the lungs and fixed acids by the kidney.

Definitions, Acids, Bases, and Buffers

pH

The term pH is an expression of hydrogen ion (H^+) concentration (such that pH and H^+ are inversely related)(eqn [1]).

$$pH = -\log_{10}[H^+] \quad [1]$$

Acids and Bases

Acids are substances that dissociate to donate H^+ (eqn [2]); the stronger the acid, the more readily it dissociates. The dissociation constant (pK_a) is the pH at which 50% of the acid is dissociated. At pH values greater than pK_a more H^+ will dissociate; the lower the pK_a, the stronger the acid. A base is a substance that accepts hydrogen ions. In the following text, 'fixed acid' is used to describe formed acid, and 'volatile acid' is used to describe the potential acid load imposed by carbon dioxide (CO_2). Where 'A' represents an acid, the following applies:

$$AH \leftrightarrow A^- + H^+ \quad [2]$$

The importance of this relationship in physiological terms is that because the pK_a of most organic acids is much lower than the pH of extracellular fluids, most organic acids exist in a dissociated state – as acid anion salts – the free H^+ being buffered. In urine, where the minimum achievable pH is approximately 5, most strong acids (with a pK_a below this value) will be in a dissociated state, necessitating the excretion of H^+ together with urinary buffers.

Acidosis is the term used to describe conditions where pH is low and those where pH would be low were it not appropriately buffered; similarly alkalosis is the term used for a high pH and for a potentially elevated pH that has been appropriately buffered. Acidemia and alkalemia reflect low or elevated blood pH. It is common to describe an acidosis/alkalosis as respiratory or metabolic depending on causation.

Buffers

Buffering is the ability of weak acids, present in excess, to accept H^+ donated from strong acids, thus limiting changes in free H^+ concentrations and pH changes (eqn [3]):

$$AH + Buffer \leftrightarrow Buffer{-}H^+ + A^- \quad [3]$$

The principal buffer system in blood (and other extracellular fluids) is based on bicarbonate (HCO_3^-), accounting for approximately 70% of the buffering capacity of blood. In blood, CO_2 (the major product of oxidative metabolism) reacts with water, in the presence of the enzyme carbonic anhydrase (CA), to form carbonic acid (H_2CO_3). This compound is relatively unstable and tends to dissociate (eqn [4]). The rate of formation of carbonic acid is dependent on the concentration of carbon dioxide and the rate constant of reaction [i]; the dissociation of carbonic acid to generate H^+ and HCO_3^- is governed by the rate constant of reaction [ii]. In practice, these two reactions can be combined, and the relationship between pH ($[H^+]$), carbon dioxide, and bicarbonate is described by a single equation – the Henderson–Hasselbalch eqn [5]:

$$CO_2 + H_2O \overset{[i]}{\leftrightarrow} H_2CO_3 \overset{[ii]}{\leftrightarrow} H^+ + HCO_3^- \quad [4]$$

$$pH = 6.1 + \log_{10}([HCO_3^-]/K.S.PCO_2) \quad [5]$$

pH reflects − log $[H^+]$; 6.1 is the value of –log (1/K), K being the equilibrium constant describing the overall eqn [4]; PCO_2 is the partial pressure of carbon dioxide; S is the solubility constant for carbon dioxide. K.S. is constant and equal to 0.225 (when PCO_2 is measured in kP_a, 0.03 when PCO_2 is measured in mmHg). Table 1 shows the normal range for these parameters in humans.

From eqn [5] the principles of acid–base balance can be appreciated. Acidification may occur in two ways: either by production of CO_2 or by the consumption of bicarbonate (as part of buffering of fixed acid). The excretion of CO_2 is controlled by the lungs, and excretion of fixed acid takes place in the kidney.

Table 1 Normal ranges

Variable	*Normal range*
pH	7.36–7.44
Hydrogen ion (H^+)	37–44 nmol l^{-1}
Partial pressure CO_2 (PCO_2)	34–46 mm Hg 4.5–6.1 k Pa
Bicarbonate HCO_3^-	24–30 mmol l^{-1}

The Henderson–Hasselbalch equation allows basic understanding of acid–base physiology, in health and disease but has limitations. In the presence of either metabolic or respiratory derangement of acid–base homeostasis it does not allow assessment of the severity of the metabolic derangement, analogous to the respiratory component. It also does not assess the influence of other acids other than carbonic acid. For this reason some authors propose analysis of acid–base physiology using a more complex method based on the principals of physical chemistry. This method proposes that all changes in pH of plasma can be explained in terms of relative concentrations of CO_2, relative electrolyte, and weak acid. This concept allows more rigorous interrogation of acid–base disorders and may permit greater insight into their pathophysiology and management.

Maintenance of the pH of Blood and Extracellular Fluids

Acid and Alkali Load

The sources of acids (and alkalis) are the diet and metabolism. The major potential source of acid is CO_2 ('volatile acid') (eqn [4]), generated by oxidative metabolism; a total of 12–20 moles of CO_2 are produced daily. Other metabolic products include lactic acid, other organic acids, and urea, the synthesis of which produces H^+. Because of its role in the metabolism of lactic acid and in the synthesis of urea, the liver plays a major role in acid–base homeostasis that is often not appreciated.

The lungs excrete volatile acid, CO_2, whereas the breakdown of sulfur-and phosphorus-containing compounds are 'fixed' acids. For example, cysteine or methionine metabolism leads to production of sulfuric and phosphoric acid (H_2SO_4, H_3PO_4), and metabolism of other amino acids (lysine, arginine, and histidine) to hydrochloric acid (HCl). In contrast, organic acids (e.g., lactate, fatty acids) may be completely metabolized to CO_2 and H_2O and thus excreted by the lungs. In addition, absorption of dietary phosphate and fecal loss of bicarbonate represent an additional acid load. In total, the net acid load of fixed acid is approximately 1 mmol kg^{-1} day^{-1} and may be increased by a high protein intake or reduced by a strict vegetarian diet.

There is surprisingly little information on the direct contributions of individual foods to the acid burden. However, this source of dietary acid is of increasing importance in view of weight reduction diets (e.g., the Atkins diet). The major acids contained in food are citric acid (in fruit), acetic acid (as a preservative, pickles, vinegar), lactic acid (yogurt, fermented foods), malic acid (fruit), oxalic acid (vegetables, that contain smaller amounts of citric and malic acids), and tartaric acid (wine). Oxalic acid precipitates in the gut to form calcium salts, excreted in the stool, and little is absorbed. The others are absorbed but quickly metabolized and present an acid burden in the form of carbon dioxide. The largest source of fixed acid comes from the metabolism of amino acids (particularly those from animal proteins). The significance of this source of acid is readily demonstrated in patients taking a high-protein diet – particularly one rich in animal protein – who have increased urinary acid excretion. Based on studies on the relationship between diet, renal excretion of acid, and urine pH it is theoretically feasible to quantify urinary acid excretion for individual foods. However due to daily variation in diet (and therefore absence of a metabolic steady state) and inherent variation in the composition of foodstuffs, it has not been possible to estimate accurately the effects of diet on renal acid–base metabolism in circumstances reflective of normal dietary intake.

Alkalis are often prescribed to compensate metabolic acidosis and have often been used to neutralize gastric acidity. Milk and milk products are also alkaline but seldom cause any disturbance, unless consumed in great excess. Excessive consumption of milk or alkali is now rarely seen.

Regulation

Blood and extracellular fluid pH is regulated at three levels: (1) buffering within the blood and tissues; (2) excretion of volatile acids by the lungs; (3) excretion of fixed acids by the kidney. Although buffering is immediate, respiratory compensation occurs over minutes to hours and renal excretion can take from hours to days (Table 2).

Blood/Extracellular Fluid

Immediate buffering of an acid load, for example by release of lactic acid and CO_2 by anaerobic and aerobic metabolism in exercising muscle, occurs in blood and other extracellular fluids. Together they contain approximately 350 mmol of bicarbonate buffer. 60–70% of the buffering capacity of blood is accounted for by the bicarbonate buffer system; 20–30% is dependent on direct binding to hemoglobin and to other proteins, including plasma proteins. Blood is in equilibrium with extracellular fluid H^+. H^+ ions move across cell membranes depending on concentration and charge, thus H^+ ions may move into cells in exchange for K^+ (and to a lesser extent Na^+ ions) when extracellular H^+ is increased. Hence acidosis is often accompanied by increased serum K^+, and alkalosis by low K^+. Large amounts of H^+ may be 'buffered' by direct binding to proteins within cells and tissues, particularly bone where H^+ are also buffered by calcium salts, such as apatite.

Lungs

The lungs excrete volatile acid (CO_2) by changes in the rate and volume of respiration. This is regulated by respiratory centers in the brainstem, that respond to changes in the pH of cerebrospinal fluid (that is in equilibrium with extracellular fluids elsewhere in the

Table 2 Buffering and acid–base regulation

Mechanism	*Site*	*Role (time)*
Protein (e.g., Hb)	Cell	Rapid binding of H^+ (s)
Bicarbonate buffer	ECF	Buffering of H^+ (s)
Ventilation	Lungs	Excretion of CO_2, respiratory compensation (h)
Fixed acid excretion	Kidney	Excretion of H^+, reabsorption and regeneration of bicarbonate, renal compensation (h–days)

body), and signals from chemoreceptors in the carotid and aortic bodies that are responsive to changes in pH and PCO_2 of arterial blood (increased PCO_2 or reduced pH cause an increase in respiration). Thus, acidosis leads to an increase in respiratory rate and ventilatory volume (the pattern in severe acidosis being described as Kussmaul breathing) and alkalosis to the opposite effect.

Kidneys

The kidneys have two major roles in acid–base homeostasis: the recovery of filtered bicarbonate and generation of new bicarbonate; and excretion of fixed acid (**Figures 1 and 2**; eqn [5]). Blood is filtered in the glomeruli and the glomerular filtrate is subsequently modified in the renal tubules so that the final urine volume is less than 1% of the glomerular filtrate volume. Plasma bicarbonate concentration is approximately 25 mmol l^{-1} and glomerular filtration rate (GFR) 100 ml min^{-1}, thus 3600 mmol of bicarbonate must be reabsorbed daily. Bicarbonate reabsorption mainly occurs in the proximal convoluted tubule (PCT). 85% of filtered bicarbonate is reabsorbed here, 10% in the thick ascending limb of the Loop of Henle, the remainder being titrated to regulate total acid excretion, in the collecting duct (**Figures 1 and 2**). As shown, different mechanisms are involved at each tubular site. The enzyme carbonic anhydrase, on the luminal brush border of tubular cells, catalyzses the combination of filtered bicarbonate with H^+, secreted by the apical H^+-ATPase and Na^+/H^+ exchangers on tubular cells, to generate CO_2. CO_2 then diffuses into the tubular cells down its concentration gradient. Within the tubular cell, carbonic anhydrase catalyzes the reverse reaction, which generates H^+ and HCO_3^- Hydrogen ions are then recycled to the tubular lumen and bicarbonate is secreted into the extracellular fluid by basolateral anion exchangers or Na^+–HCO_3^-, cotransporters. Tubular cells are also exposed to CO_2 in the extracellular fluid and will continue to generate H^+ even in the absence of filtered bicarbonate. This H^+ is then buffered by other buffers in the glomerular filtrate including HPO_4^{2-} and, to a lesser extent, creatinine. Strong acids, (e.g., H_2SO_4) with low pK_a values will dissociate in the urine (pH range 5–8), and be buffered whereas weaker acids may be excreted intact. In the presence of alkalosis, cellular transporters

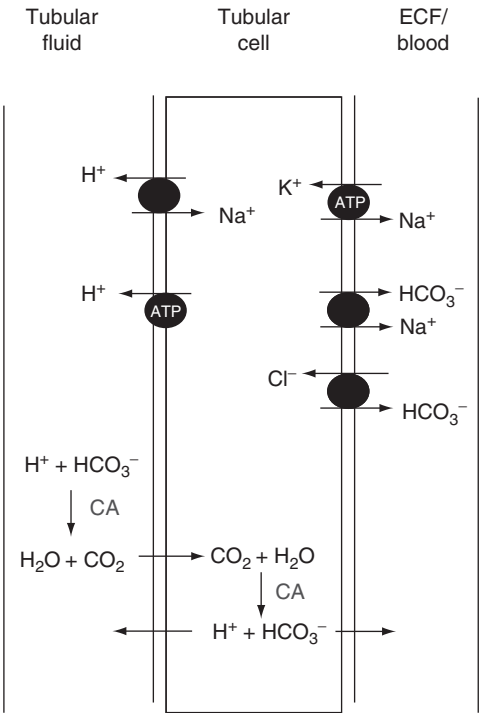


Figure 1 Recovery of filtered bicarbonate in the proximal convoluted tubule.

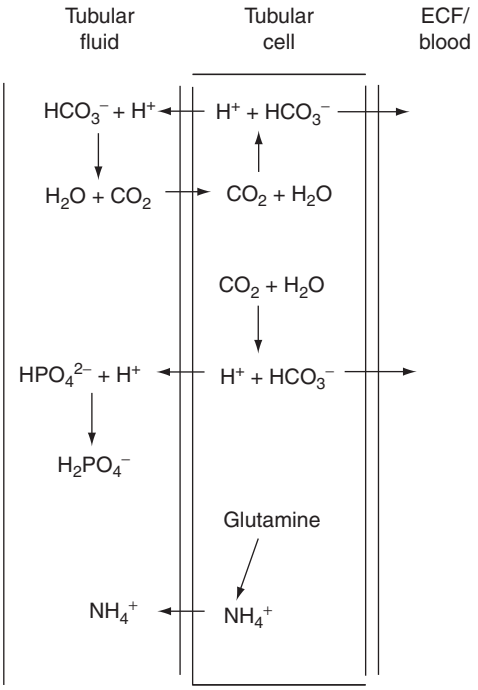


Figure 2 Excretion of acid in the collecting duct.

and their function may be reversed so that H^+ secretion occurs on the basolateral membrane and HCO_3^- on the brush border of tubular cells resulting in alkaline urine.

Classically, the final mechanism by which the kidney can excrete H^+ is by generation of ammonium (NH_4^{2+}) from metabolism of glutamine by glutaminase (Figure 2), a process that is stimulated by low pH and increased PCO_2. Excretion of H^+ as part of ammonium accounts for approximately 70 mmol day^{-1}, increasing several-fold (albeit over a period of days) in the face of an acid load. Whether this is truly a urinary buffer is the subject of some debate as ammonium (NH_4^{2+}) is generated directly from glutamine rather than accepting additional protons. There are alternative mechanisms for the role of NH_4^+ in overall acid–base homeostasis that involve the liver. After being pumped into the glomerular filtrate, NH_4^+ may be reabsorbed by the tubule and used by the liver to synthesize urea, generating free H^+ ions. Thus, there is no net loss of H^+. Thus, the overall role of NH_4^+ in acid–base balance is dependent on the balance between tubular reabsorption of NH_4^+ and the hepatic synthesis of urea. The latter function may also be directly influenced by extracellular pH.

Liver and Bone

The liver also plays additional roles in acid–base balance that may be underestimated. For example, the liver metabolizes lactate and ketoacids; the rate of metabolism is dependent on pH (e.g., ketogenesis is suppressed at low pH) and may be exceeded at higher concentrations of lactate or in liver disease. The synthesis of urea from ammonium and carbon dioxide (see Section on Kidneys) results in genesis of two protons, and is reduced in the presence of acidosis. Some buffering also occurs in bone due to slow exchange of bone calcium carbonate for extracellular phosphate.

Measurement of Urinary Acid Excretion

Urinary pH can be measured by commercially available 'dipsticks' or by using a pH meter on a fresh sample of urine. Loss of CO_2 or the production of NH_4^+ from urea-splitting organisms in infected urine will alter the pH with time. The excretion of fixed acid can be determined by chemical titration of urine to pH 7.4, and is commonly termed 'titratable' acidity. The amount of NH_4^+ is usually estimated from the difference between the most abundant cation (Na^+, K^+) and anion (Cl^-) concentrations in the urine.

Effects of Acid–Base Disturbance

In addition to the adaptive changes occurring in acidosis, a range of metabolic and pathophysiological changes occur; alkalosis tends to produce opposite but milder effects. Metabolism of carbohydrate is altered; both glycolysis and gluconeogenesis are

inhibited in the liver. Delivery of oxygen to the tissues is increased by the reduced ability of hemoglobin to retain oxygen in an acid environment (the Bohr effect). Consciousness is impaired, leading to coma in severe cases. However, the most important effects from a clinical perspective are cardiovascular: vasodilatation occurs in peripheral tissues, cardiac contractility is impaired resulting in reduced blood pressure and, when severe, in reduced tissue perfusion. These adverse effects of acidosis contribute to the high mortality of conditions like septic shock. Conversely, in critically ill patients, often after surgery, severe alkalosis may be equally or more dangerous, particularly in the event of cardiac arrest, where resuscitation is extremely difficult.

Abnormalities in Acid–Base Balance

Disturbances in acid–base balance are classified either as 'acidosis', indicating an excess of H^+ ions in the blood (reduced pH) or 'alkalosis', indicating the opposite. In practice, acidosis is often the more common, varied, and serious problem, although profound alkalosis can be life threatening in the critically ill. Disturbances in acid–base balance are usually labeled according to their origin. For example, respiratory acidosis reflects a primary problem in gas exchange with impaired excretion of CO_2, whereas metabolic acidosis reflects over-production of fixed acid or loss of bicarbonate. Compensation refers to the body's ability to offset the primary problem. Thus, the response to a primary metabolic acidosis is to increase excretion of CO_2; respiratory compensation. In a primary respiratory acidosis, increased H^+ is secreted by the kidney with increased bicarbonate generation; metabolic compensation. If the pH returns to normal the problem is said to be 'fully compensated' whereas most disturbances tend to only be partially compensated (Table 3). Additionally in some circumstances, a mixed respiratory and metabolic acidosis may be present (e.g., septic shock in a patient with underlying respiratory disease affecting gas exchange).

Metabolic Acidosis

The main causes of metabolic acidosis are excessive acid production, inappropriate urinary loss of bicarbonate or failure of the kidney to excrete fixed acid. Although the Henderson–Hasselbalch equation provides mathematical information concerning the equilibrium of bicarbonate species, in practice it provides little information regarding the nature of the underlying cause of the acid–base disorder and the concept of 'anion gap' is useful in assessing cause of metabolic acidosis. This is derived from the principle of electroneutrality and is calculated thus (eqn [6]):

$$([Na^+]+[K^+])-([Cl^-]+[HCO_3^-]) \quad [6]$$

The anion gap represents an artificial disparity between the concentrations of these cations and anions routinely measured in clinical practice, therefore signifying concentration of unmeasured anions such as proteins (the most important in healthy subjects), sulfate, phosphate, and other unmeasured anions. The normal anion gap is 10–18 mmol l^{-1} although recent calculations using more sensitive measurements estimate this as 6–12 mmol l^{-1}. This concept has limitations but is useful for dividing metabolic acidoses into those characterized by an increased anion gap as a marker of excess generation of organic acids and those with a normal anion gap due to decreased excretion of acid or external losses of bicarbonate. There are exceptions to this rule for example, the acidosis of chronic renal failure but nonetheless, it remains a useful concept in clinical practice. Classification of the causes of metabolic acidoses according to presence of an increased or normal anion gap is shown in Table 4.

Diabetic Ketoacidosis

The absence of pancreatic insulin secretion in insulin-dependent diabetes results in increased plasma glucose and reduced tissue uptake and utilization of glucose. In place of glucose, there is increased utilization of nonesterified fatty acid (NEFA) as an alternative source of energy that is metabolized to acetyl coenzyme A (acetyl-CoA). Under normal circumstances this substance is further metabolized in the liver *via* the tricarboxylic acid (TCA) cycle to carbon dioxide and water. In diabetic crises this cycle cannot accommodate the excess acetyl-CoA that is instead converted to acetoacetic acid, which can be further reduced to ß-hydroxybutyric acid or

Table 3 Changes in blood and ECF during acid–base disturbance, the mechanism and degree of compensation

Problem	*$[H^+]$*	*HCO_3^-*	*PCO_2*	*Compensation*
(a) Metabolic				
Acidosis	↑	1°↓	2°↓	Partial respiratory
Alkalosis	↓	1°↑	2°↑	Partial respiratory
(b) Respiratory				
Acidosis	↑	2°↑	1°↑	Complete renal
Alkalosis	↓	2°↓	1°↓	Complete renal

↑, increase; ↓, decrease; 1°, primary; 2°, secondary.

Table 4 Causes of metabolic acidoses according to the presence of an increased or normal anion gap

Increased anion gap	*Normal anion gap*
Ketoacidosis Diabetic Starvation Alcoholic Inborn enzyme defects of metabolism	Decreased renal acid excretion Distal renal tubular acidosis
Lactic acidosis	Loss of alkali Diarrhea Ureterosigmoidostomy (urinary conduit)
Renal failure	Increased renal bicarbonate loss Proximal renal tubular acidosis Azetazolamide Renal tubular damage
Intoxication Salicylates Methanol Ethylene glycol Paraldehyde	Increased HCl production Ammonium chloride ingestion Increased catabolism of lysine, arginine

decarboxylated to acetone. These metabolites are known as 'ketone bodies' and their accumulation results in metabolic acidosis. In diabetic ketoacidosis, homeostatic compensation is to increase ventilation and CO_2 excretion, leading to the characteristic pattern of Kussmaul breathing.

Lactic Acidosis

Reduced tissue perfusion, or perfusion that is inadequate to meet the metabolic demands of tissues (such as exercising muscle), results in an inadequate supply of oxygen and a change from oxidative metabolism (the end products of which are CO_2 and H_2O) to anaerobic metabolism. The end product of anaerobic glycolysis is lactic acid, which is normally metabolized (to CO_2 and H_2O) by the liver or used in synthesis of glucose (gluconeogenesis). The normal plasma lactate concentration is less than 1 mmol l^{-1} but may increase 10-fold in extreme exercise. When the ability to metabolize lactate is exceeded, either by increased production, or reduced delivery to the liver (in, for example, circulatory shock) or in the presence of impaired liver function, accumulation results in metabolic acidosis. Thus, lactic acidosis may occur in a variety of conditions, including both overt circulatory shock and conditions leading to circulatory shock (for example, septic shock), severe diabetic ketoacidosis, as a consequence of drugs (e.g., the oral hypoglycemic agent metformin that inhibits gluconeogenesis and lactate transport), chronic liver disease and poisoning (including ethanol and methanol).

Excess Bicarbonate Loss

The secretion of acid into the stomach is neutralized by intestinal alkaline secretions. It follows that excessive loss of pure intestinal secretions (e.g., due to an enteric fistula) may lead to acidosis. A more common circumstance is a urinary diversion procedure such as an ileal conduit where the ureters are implanted into an isolated loop of intestine, which is then externalized (a 'urinary conduit'). The delivery of urine rich in chloride to the isolated intestine leads to exchange of Cl^- for HCO_3^-, and thence to excessive loss of HCO_3^- in the conduit, resulting in metabolic acidosis. Historically ureterosigmoidostomies performed for similar clinical indications tends to provoke a more severe metabolic acidosis by the same mechanism.

There is also a group of conditions known as renal tubular acidosis (RTA). These are mostly inherited but may be acquired, for example as a consequence of recurrent infection. There are two major forms – proximal and distal – reflecting the site of the tubular defect in the nephron. In distal tubular RTA (Type I) H^+ secretion is impaired resulting in impaired H^+ excretion, whereas in proximal RTA (Type 2) HCO_3^- reabsorption is impaired (usually as part of multiple tubular abnormalities) leading to net loss of bicarbonate. Both cause acidosis, with low pH and hypokalemia as a result of increased distal tubular H^+/K^+ exchange. The precise causes of these conditions are not known but is likely to reflect genetic defects on individual transporter subtypes, for example those of the Na^+/H^+ exchanger (**Figure 1**).

A third form of RTA (Type 4), also called hyperkalemic RTA, is the result of generalized transport abnormalities of the distal tubule. Electrolyte transport of sodium, chloride, and potassium are impaired. This form differs from classical distal and proximal RTA because there are high (not low) levels of potassium in the blood. Type 4 RTA occurs with deficiency of the hormone aldosterone or when the kidneys fail to respond to it. Aldosterone acts *via* the kidneys to control levels of sodium, potassium, and chloride

in the blood. Type 4 RTA also occurs when tubular transport of electrolytes such as sodium, chloride, and potassium is impaired due to an inherited disorder or use of certain drugs (e.g., spironolactone, nonsteroidal anti-inflammatory drugs or angiotensin-converting enzyme inhibitors) and in diseases which alter kidney structure and function (e.g., diabetic nephropathy, sickle cell disease, urinary tract obstruction, lupus, or amyloidosis). Treatment is directed at the underlying cause; in aldosterone deficiency, fludrocortisone, a mineralocorticoid is used.

Renal Failure

In progressive renal failure, renal clearance of all substances is progressively impaired, reflecting progressive loss of individual nephron function. Reduced excretion of fixed acid leads to bicarbonate consumption in extracellular fluids and to acidosis. Tubular recovery of HCO_3^- may also be impaired (see *Excess bicarbonate loss*), as may production of tubular NH_4^+, and be associated with over-production of urea in the liver, although this is unlikely to be a major concern to excess urea production in clinical practice.

Drugs and Other Causes

Many drugs can cause metabolic acidosis, generally in overdose. A classic example is aspirin (acetylacetic acid). Lactic acidosis is also associated with oral hypoglycemic agents (specifically metformin, used in treatment of noninsulin dependent diabetes), paracetamol, alcohol, and ethylene glycol (antifreeze) poisoning.

Compensation

The body's response to metabolic acidosis is a compensatory increase in ventilation to excrete excessive CO_2, restoring equilibrium in the Henderson–Hasselbalch equation (**eqn [5]**). This respiratory compensation is usually incomplete, resulting in pH values or H^+ concentrations at, or marginally outside, the limits of 'normal' (**Table 3**). Complete compensation depends on renal excretion of excess H^+, or resolution of the underlying condition.

Treatment

Treatment of metabolic acidosis is essentially that of the underlying condition: correction of tissue hypoxia in lactic acidosis; correction of fluid depletion and insulin therapy in diabetic ketoacidosis; dialysis in renal failure. Rapid correction of pH can be achieved by administration of intravenous sodium bicarbonate if necessary with the caveat that in a mixed respiratory and metabolic acidosis caution should be exerted as sodium bicarbonate can increase CO_2 levels. The treatment of chronic metabolic acidosis (e.g., in chronic renal failure or RTA) may be achieved by the administration of oral sodium bicarbonate. In uremia the prescription of a low protein diet will reduce acid load and may prevent uremic symptoms.

Metabolic Alkalosis

Metabolic alkalosis may be caused either by excessive loss of acid or intake of alkali. The latter may be iatrogenic or factitious, with the excessive intake of prescribed antacids (such as sodium bicarbonate for heartburn or peptic ulcer disease) – the 'milk-alkali' syndrome. The loss of acid-rich gastric secretions in severe vomiting for example, in cases of gastric outlet obstruction (due to pyloric stenosis, or a consequence of peptic ulcer disease), also leads to alkalosis. Administration of alkali only causes significant metabolic alkalosis when the absorption is considerable and sustained such as transfusions of citrated blood or sodium bicarbonate therapy for metabolic acidosis due to circulatory arrest in the presence of oliguric renal failure. Compensation is by reducing ventilation to promote retention of CO_2 and thus balance the Henderson–Hasselbalch equation. Treatment is usually of the underlying condition. In the case of severe metabolic alkalosis in critically ill patients with associated coma or arrhythmia where alternative treatment (hemodialysis or filtration) has failed or is contraindicated, administration of intravenous hydrochloric acid *via* a central vein can be beneficial.

Respiratory Acidosis

Impaired ventilation, reduces CO_2 excretion, increases $PaCO_2$ and thus lowers pH. This may occur acutely or chronically. Causes of respiratory acidosis include factors that interfere with the neurological 'drive' for respiration (e.g., head injury, cardiac arrest, opiate, and anesthetic drugs), diseases of the respiratory muscles (e.g., poliomyelitis, Guillain–Barré syndrome), or primary lung diseases (acute-pulmonary edema or pneumonia; chronic bronchitis, emphysema). In acute conditions, pH may fall dramatically, whereas in chronic conditions, such as chronic lung disease, pH is generally nearer normal. In chronic conditions complete compensation occurs in the kidney where elevated $PaCO_2$ levels are offset by the increased generation of bicarbonate and excretion of fixed acid by the kidney, to balance the Henderson–Hasselbalch equation.

Respiratory Alkalosis

Respiratory alkalosis occurs as a result of inappropriately increased ventilation and increased excretion of CO_2. This may occur as a transient response to pain or hysteria. Such stimuli tend to be short-lived and can be offset by analgesia, sedation or short-term rebreathing of expired air. Additional causes include the early phases of aspirin poisoning (where the respiratory centers are activated), hypoxia, sepsis, hepatic failure, stroke, and other conditions affecting brainstem respiratory control centers. Most causes of respiratory alkalosis are short-term and, although adaptive responses would be expected to require excretion of bicarbonate to balance the Henderson–Hasselbalch equation, resolution usually occurs with treatment of the underlying condition.

Transporter Mechanisms: Physiology and Pathophysiology

Developments in molecular biology have lead to major improvements in our understanding of the physiology, and pathophysiology of renal tubular function. It is now possible to subdivide the various types of renal tubular acidosis for example, by the precise biochemical defect rather than simply the tubular location. Thus, distal (or type 1) RTA may be a consequence of impaired distal tubular H^+ excretion, either due to increased permeability to H^+ or to impaired secretion, the latter – in turn – being a consequence of a variety of defects that include carbonic anhydrase type 2 deficiency, mutations in anion transport protein AE1, or deficiency of collecting-duct proton transport ATPase. Although specific knowledge of the molecular defect is not necessary to either diagnose or manage these disorders, it is likely that future classification of acid–base disorders will change to recognize the underlying defect.

Further Reading

Adrogué, H.J., Gennari, F.J., Galla, J.H., Madias, N.E., 2009. Assessing acid-base disorders. Kidney International 76 (12), 1239–1247.
Adrogué, H.J., Madias, N.E., 1998. Management of life-threatening acid-base disorders. First of two parts. New England Journal of Medicine 338 (1), 26–34.
Adrogué, H.J., Madias, N.E., 1998. Management of life-threatening acid-base disorders. Second of two parts. New England Journal of Medicine 338 (2), 107–111.
Corey, H.E., 2003. Stewart and beyond: New models of acid-base balance. Kidney International 64, 777–787.
Gennari, F.J., Weise, W.J., 2008. Acid-base disturbances in gastrointestinal disease. Clinical Journal of the American Society of Nephrology 3 (6), 1861–1868.
Gluck, S.L., 1998. Acid-base. The Lancet 352, 474–479.
Halperin, M.L., Goldstein, M.B., 1999. Fluid, Electrolyte and Acid-Base Physiology: A Problem-Based Approach, 3rd edn. W.B.Saunders, London.
Halperin, M.L., Kamel, K.S., 2010. Some observations on the clinical approach to metabolic acidosis. Journal of the American Society of Nephrology 21 (6), 894–897.
Kellum, J.A., 2007. Disorders of acid-base balance. Critical Care Medicine 35 (11), 2630–2636.
Kellum, J.A., 2000. Determinants of blood pH in health and disease. Critical Care 4 (1), 6–14.
Palmer, B.F., Alpern, R.J., 2007. Normal acid-base balance. In: Feehally, J, Floege, J, Johnson, R.J. (Eds.), Comprehensive Clinical Nephrology, 3rd edn. Elsevier, Philadephia, pp. 141–146.
Remer, T., 2000. Influence of diet on acid-base balance. Seminars in Dialysis 13 (4), 221–226.
Sirker, A.A., Rhodes, A., Grounds, R.M., Bennett, E.D., 2002. Acid-base physiology: The 'traditional' and the 'modern' approaches. Anaesthesia 57, 348–356.
Stewart, P.A., 1983. Modern quantitative acid-base chemistry. Canadian Journal of Physiology and Pharmacology 61, 1444–1461.
Williams, A.J., 1998. ABC of oxygen: Assessing and interpreting arterial blood gases and acid-base balance. British Medical Journal 317, 1213–1216.

Energy metabolism

Malcolm Watford, Department of Nutritional Sciences, School of Environmental and Biological Sciences, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

© 2023 Published by Elsevier Ltd. All rights reserved.

This is an update of S.E. Cox, Energy Metabolism, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 177–185, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00091-X.

Introduction 115
Carbohydrates 117
Glycolysis 117
Pyruvate metabolism 118
Fat catabolism 118
Protein and amino acid metabolism 120
Ethanol 121
The tricarboxylic acid cycle and oxidative phosphorylation 121
Uncoupling and brown adipose tissue 122
Energy storage and fuel use in the post-absorptive state 123
Triglycerides 123
Glycogen 124
Glycogen breakdown and gluconeogenesis provide glucose to the body 125
Summary 125
References 125

Key points

- The major energy supplying fuels in the body are glucose, lactate, fatty acids, ketone bodies, some amino acids and ethanol, with minor contributions from fructose, galactose, glycerol and limited amounts from short chain fatty acids produced by the colonic microbiota.
- During metabolism the energy in the fuels is converted by specific pathways to a number of end products but the useable energy is conserved in reducing equivalents (NADH and $FADH_2$) and primarily acetyl CoA.
- The only anaerobic pathway of energy metabolism is glycolysis, the breakdown of glucose (and other sugars) to lactate which only yields 2 ATP per glucose metabolized. When this occurs in the presence of oxygen the process is termed aerobic glycolysis. Most cells however can continue the breakdown of glucose by oxidation in the mitochondria a process yielding 27–29 ATP per glucose.
- Acetyl CoA is itself oxidized in the Tricarboxylic Acid Cycle (a central pathway for the final stages of catabolism of all fuels) releasing the carbon as carbon dioxide and producing NADH and $FADH_2$
- NADH and $FADH_2$ are oxidized in the process of oxidative phosphorylation in the inner membrane of mitochondria where electrons flow along a transport chain to the final electron acceptor, molecular oxygen, to produce water. At key stages along the electron transport chain protons are pumped out of the mitochondrial matrix to concentrate in the inter-membrane space. These protons flow back into the matrix driving the ATP synthase complex to produce ATP.
- During the absorptive phase the major fuel utilized is glucose and extra glucose and dietary fatty acids are stored as glycogen and triglycerides respectively. During the post-absorptive period fatty acids are mobilized and used as fuel by most tissues, but not by the brain. Within the liver fatty acids undergo partial oxidation to produce ketone bodies which can be used as fuel by most tissues including the brain. During this time blood glucose concentrations are maintained by the release of glucose from liver glycogen stores and then from gluconeogenesis (the synthesis of new glucose).
- In the post-absorptive period and starvation the utilization of fatty acids and ketone bodies as fuels by different tissues replaces much of the need for glucose utilization and is known as "the glucose sparing effect of fat derived fuels". But since most of the glucose in starvation is made from amino acids derived from the breakdown of proteins, this could also be referred to as "the protein sparing effect".

Glossary

Acidosis In this text referring to metabolic acidosis, a disturbance of plasma acidity with a drop in pH and bicarbonate levels. Arising from an accumulation of acids such as acetoacetic acid and β-hydroxybutyric acid (the ketone bodies)

Acyl Referring to the fatty acid group in a compound for example, the acetyl and palmitoyl groups in acetyl CoA and palmitoyl CoA may simply be referred to as acyl CoA

Acyl CoA A fatty acid linked by a thio-ester bond to Coenzyme A (CoASH)

Anabolism The synthetic part of metabolism, the building up of cellular components and storage of fuels

ATP, ADP, AMP Adenosine Triphosphate, Adenosine Diphosphate, Adenosine Monophosphate. ATP is the final energy containing compound derived from the breakdown of macromolecules in cells. The third phosphate bond is used to provide energy for other reactions with the production of ADP. Further breakdown to AMP yields more useable energy

β-oxidation The breakdown of fatty acids by removal of 2 carbon units, ultimately yielding acetyl CoA or ketone bodies

BAT Brown adipose tissue that contains Uncoupling protein 1 (UCP 1) which generates heat by allowing protons to re-enter the mitochondrial matrix without the synthesis of ATP

Cellular Respiration The biochemical process by which the energy in fuels is released as ATP by the process of oxidative phosphorylation (also known as oxidative respiration)

CoASH Coenzyme A (derived from vitamin B5, pantothenic acid). The SH indicates a thiol group that is used to link to acyl groups by a thio-ester bond

Chylomicrons Large triglyceride rich lipoprotein complexes formed in the enterocytes and released to carry dietary lipids around the body via the lymph and peripheral blood circulation

Catabolism The breaking down of fuels and cellular components

Carnitine Palmitoyl Transferase The enzyme that replaces the CoA moiety of Acyl CoA with carnitine enabling the acyl (fatty acid) group to enter the mitochondria for oxidation

Coupled respiration The coupling of electron transfer to ATP synthesis

Dehydrogenase Enzymes that carry out reduction (and oxidation) reactions usually involving NAD^+/NADH, $NADP^+$/NADPH, or FAD/$FADH_2$

De novo Synthesis of a compound from basic components, for example de novo fatty acid synthesis is the synthesis of fatty acids from glucose

Electron Transport Chain A series of electron acceptors embedded in the inner mitochondrial membrane that passes electrons from NADH and $FADH_2$ to the final electron acceptor oxygen

FAD, $FADH_2$ Flavin adenine dinucleotide (oxidized and reduced) redox related coenzymes. In this text referring to the production of $FADH_2$ in the TCA cycle and its oxidation to FAD in the electron transport chain

Free fatty acid Otherwise known as non-esterified fatty acids, meaning the fatty acid is not bound to other compounds and is present as the free acid

Fructose A six carbon sugar that comprises 50% of sucrose (table sugar). Almost exclusively catabolized in the liver

Galactose A six carbon sugar comprising 50% of lactose and found exclusively in mammalian milk. Almost exclusively catabolized in the liver

Gluconeogenesis The synthesis of glucose from non-glucose precursors including, lactate, pyruvate, amino acids, glycerol and propionate. Gluconeogenesis only occurs in the liver and kidneys

Glucose A six carbon sugar that is the principal sugar in the body and is the preferred fuel of most tissues in the fed/absorptive state. In healthy individuals blood glucose concentrations are held within very tight limits

Glycerol A three carbon alcohol that is the backbone of triglycerides. It is released during lipolysis and can be used for gluconeogenesis

Glycogen A polymer of glucose mainly stored in liver (for release as glucose into the circulation) and skeletal muscle (for use as a local fuel for muscle contraction)

Glycolysis Glycolysis occurs in the cytosol and is the breakdown of glucose to two lactates, an anaerobic process. When this occurs in the presence of oxygen it is called aerobic glycolysis. In most cells however, the pyruvate formed from glucose is further oxidized in mitochondria and the process of glucose to pyruvate in the cytosol is simply referred to as glycolysis

Glyceroneogenesis The formation of glycerol 3-phosphate for esterification with 3 fatty acids to form triglycerides from non-glucose precursors such as lactate

Hexokinase The enzyme that phosphorylates glucose to glucose 6-phosphate. In liver a unique form, glucokinase, is employed

Isomerase An enzyme that rearranges a the structure of a molecule from one isomer (form) to another. For example glucose 6-phosphate to fructose 6-phosphate

Ketone bodies The ketone bodies, often referred to simply as ketones, are acetoacetate, β-hydroxybutyrate and acetone. They are water soluble end products of fatty acid oxidation in the liver. Acetone is highly volatile and is lost from the body in expired air. Ketones rise in the circulation in the post-absorptive period and continue to rise as starvation develops. They are used as an alternative fuel by most tissues including the brain, but not the liver. Very high levels of ketones may occur in Type 1 diabetes resulting in keto acidosis. Lower levels of ketones arise on consuming high fat, very low carbohydrate diets, otherwise known as ketogenic diets

Kinase An enzyme that adds a phosphate group from ATP (or GTP) to a substrate
Lactate Also known as L-lactic acid. A three carbon hydroxy acid that is the end product of the anaerobic pathway of glycolysis. It is readily converted to and from pyruvate in most tissues, and also to alanine (via pyruvate) in a number of tissues. Lactate can also be used directly a fuel in some tissues
Lipolysis The hydrolysis of neutral lipids such as triglycerides to yield free fatty acids
NAD^+, NADH Nicotinamide adenine dinucleotide (oxidized, reduced) the major redox related coenzyme. Since it donates H^+ in many reactions, including the synthesis of ATP, it is often referred to as "reducing equivalents"
$NADP^+$, NADPH Nicotinamide adenine dinucleotide phosphate (oxidized, reduced) a redox related coenzyme mainly involved in donating H^+ in cytosolic synthetic reactions
Oxidative phosphorylation The process by which ATP is produced in mitochondria by the passage of electrons from NADH and $FADH_2$ along the electron transport chain
Pentose Phosphate Pathway/Hexose Monophosphate Shunt A cytosolic pathway that is, in part an alternative to glycolysis. The purpose of the PPP is to produce NADPH for synthetic reactions and the ribose phosphate required for the synthesis of nucleic acids
Pi Inorganic phosphate, released together with ADP when ATP is hydrolyzed
PPi Pyrophosphate released together with AMP when ATP is used in some reactions
Propionate A three carbon fatty acid, produced from the oxidation of odd carbon chain fatty acids or as a fermentation product of the colonic microbiota. It can be used as a gluconeogenic substrate
Proton In chemistry proton refers to the hydrogen ion H^+ and a reaction where a proton is donated to form a product is termed a reduction. Conversely the removal of a proton is an oxidation
Proton Motive Force The build up of protons in the inter-membrane space of the mitochondria that can be used to drive the ATP synthase to synthesize ATP from ADP and Pi
Substrate level phosphorylation The transfer of a phosphate directly from a substrate to form ATP. An example is the transfer of phosphate from phosphoenolpyruvate to form ATP and pyruvate in glycolysis. Such a process does not require oxygen
Thio-ester bond An ester bond where sulfur replaces oxygen. The best known example is acetyl CoA where acetate is linked by a thioester bond to Coenzyme A (CoASH)
Thiolase An enzyme that hydrolyzes thioester bonds as in the terminal reaction of both de novo fatty acid synthesis and fatty acid β-oxidation where a thiolase releases palmitoyl CoA and acetyl CoA respectively
Tricarboxylic Acid Cycle, Citric Acid Cycle, Krebs Cycle The central pathway whereby the final products of catabolism are metabolized releasing CO_2 and NADH, $FADH_2$, and GTP. Parts of the TCA cycle are also involved in a number of other pathways such as fatty acid synthesis and gluconeogenesis
Triglyceride (triacylglycerol) A neutral lipid consisting of three fatty acids bound to a glycerol. It is the major component of dietary lipids and is the major form of storage lipid in the body
Transamination A reaction whereby an amino group (NH_2) is transferred to a keto group (C=O) forming a new amino acid and a new keto acid catalyzed by a transaminase (aminotransferase). An example is alanine transaminase whereby pyruvate (keto acid) and glutamate (amino acid) transaminate to yield alanine and α-ketoglutarate. Such reactions involve pyridoxal phosphate (vitamin B6) and are usually freely reversible
Uncoupler A compound that uncouples electron transport and oxygen consumption from ATP production
VLDL Very low density lipoprotein. A triglyceride rich lipoprotein that is the major transporter to move triglycerides, cholesterol and other lipids throughout the body

Introduction

The first law of thermodynamics states that "energy can neither be created nor destroyed". Thus, when we refer to energy metabolism in the body we mean the conversion of the potential energy contained in food into the kinetic energy required to maintain life. The major fuels in the body are glucose, lactate, fatty acids, ketone bodies, some amino acids, and ethanol, with minor contributions from fructose, galactose, glycerol, and limited amounts from short chain fatty acids produced by the colonic microbiota. This section will consider how these fuels are used to release energy through the process of intermediary metabolism (Berdanier and Berdanier, 2015; Bronk, 1999; Frayn and Evans, 2019; Salway, 2004). Most fuels undergo a series of reactions often resulting in the production of reducing equivalents (H^+) in the form of NADH that can be used to reduce other metabolites or undergo further oxidation (Fig. 1). Ultimately, most of the energy released from macronutrient metabolism will yield ATP, the universal energy containing intermediate within the cell. The synthesis of ATP can occur by two mechanisms, substrate level phosphorylation and oxidative phosphorylation, the latter requires mitochondria and yields considerably more ATP than substrate level phosphorylation. The body turns over ATP at an amazing rate of the order of 40 kg per day (much more if you are doing hard physical labor). This translates to 10^{21} ATP per second per body, or 3×10^4 ATP per second per mitochondria.

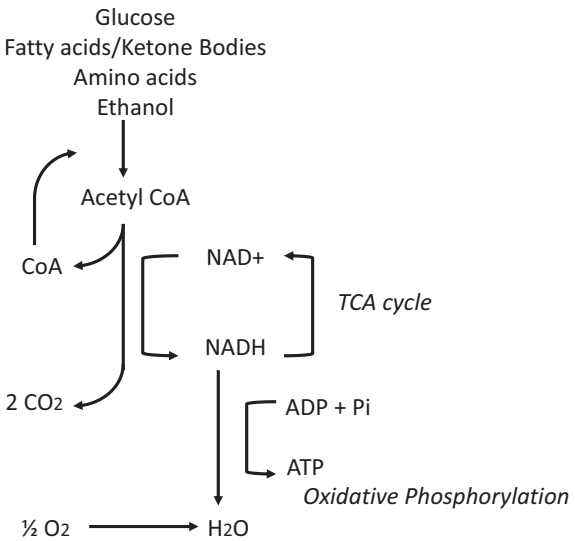


Fig. 1 The energy in the major fuels is broken down to acetyl CoA and NADH and used to produce ATP. Most fuels undergo catabolism and ultimately form Acetyl CoA plus NADH. The Acetyl CoA is then further metabolized in the TCA cycle releasing the carbons as carbon dioxide and forming addition NADH. The NADH then passes to the electron transport chain to be oxidized with oxygen to form water which is coupled to ATP formation (oxidative phosphorylation).

It is important to realize that different tissues and cells show differences in their metabolism (Table 1). Some tissues can use most fuels but a few are very specific in their fuel usage. Of particular importance is the brain which usually just oxidizes glucose but can use ketone bodies if they are available, and possibly lactate under certain conditions. Notably the brain does not utilize fatty acids as fuel. Since the brain uses about 20–25% of the body's energy per day and it completely oxidizes glucose it must be supplied with about 120 g of glucose per day. Furthermore, the brain is not able to obtain sufficient glucose once the circulating

Table 1 Tissue specific fuel metabolism.

Brain	Normally completely oxidizes only glucose. If ketone body concentrations are elevated they can be oxidized decreasing the use of glucose. Under some conditions the brain also can oxidize lactate in place of glucose
Liver	The liver is important in the utilization (glucose, fatty acids, amino acids, glycerol and alcohol), interconversion (synthesis of fatty acids, triglycerides, ketone bodies and glucose) and storage of fuels (glycogen). Ethanol is almost totally metabolized in the liver. The liver is not able to use ketone bodies or to carry out the initial reactions of branched chain amino acid metabolism.
Skeletal muscle	Different types of muscle will use different fuels, mostly fatty acids and glucose, with some metabolism of branched chain amino acids. In starvation most muscles can use ketone bodies.
Cardiac muscle	Mainly fatty acids, but during the absorptive period can use lactate or glucose. During starvation ketone bodies also may be used.
Red blood cells	Since these cells lack mitochondria they are only able to use glucose which is fermented to lactate via glycolysis.
Kidney	Can use most fuels but also is able to carry out gluconeogenesis.
Proliferating cells (e.g., Tumors)	Very high rates of glucose metabolized to lactate (Warburg effect) and high rates of glutamine utilization.

glucose concentration drops below about 3 mg/dL (60 mg/dL) and therefore blood glucose is maintained at around 4.4–5.0 mM (80–90 mg/dL). The careful balance of the utilization of the different fuels enables the body to provide different tissues with sufficient energy from the absorptive state after a meal through the postabsorptive state and into starvation (Cahill, 2006; Watford, 2015).

Carbohydrates

The end products of the digestible carbohydrates (starches and simple sugars) are glucose, fructose and galactose. Some may be metabolized in the cells of the small intestine but most will enter the portal vein for direct delivery to the liver where, under normal dietary loads, all the fructose and galactose will undergo metabolism. The rise in blood glucose after a meal will trigger the release of insulin from the beta cells of the pancreas and this will have a number of important effects to maintain glucose homeostasis. It will increase glucose uptake into skeletal muscle and adipose tissue via the stimulation of the glucose transporter GLUT 4. Equally importantly insulin also shuts down glucose output by the liver and the release of fatty acids from lipolysis in adipose tissue (Frayn and Evans, 2019). This means that glucose is now the major fuel available and, not surprisingly, insulin signals to many cell types to increase glucose utilization. There are essentially three separate pathways for glucose utilization, glycolysis, the pentose phosphate pathway (hexose monophosphate shunt), or storage as glycogen (**Fig. 2**) (Berdanier and Berdanier, 2015; Bronk, 1999; Frayn and Evans, 2019; Salway, 2004).

Glycolysis

Strictly speaking glycolysis is a fermentation, it does not require oxygen, and the penultimate product, pyruvate, is converted to lactate (which has the same oxidation state as glucose). Most cells however, are capable of taking the pyruvate into the mitochondria where is oxidized to CO_2 and water. There is some confusion in biochemistry texts regarding the use of the terms aerobic and anaerobic glycolysis. In this text the metabolism of glucose to pyruvate with further oxidation will be referred to simply as glycolysis. Many cell types will convert glucose to lactate regardless of oxygen availability and this will be referred to as aerobic glycolysis. The classic example of such a cell is the mammalian red blood cell which, lacking mitochondria, is unable to carry out oxidative metabolism. A number of other tissues such as some muscles, the renal cortex, and a number of other cell types, including tumors, also exhibit high rates of aerobic glycolysis, widely known as the Warburg effect.

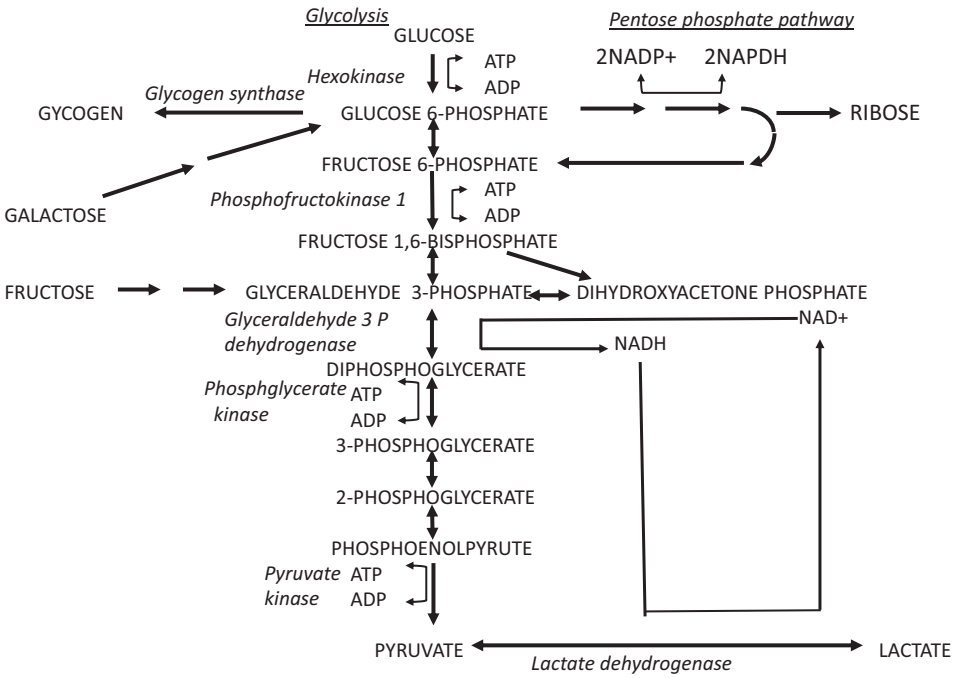


Fig. 2 Fate of glucose in the cell. Glucose is phosphorylated by hexokinase to glucose 6-phosphate which can then be used for glycogen synthesis, the pentose phosphate pathway or glycolysis. Glycolysis results in the formation of pyruvate which can be transported into the mitochondria for oxidation or simply reduced to lactate. Fructose and galactose are metabolized to feed into the glycolytic pathway. The pentose phosphate pathway is used to make ribose for nucleic acid synthesis or it can be used to produce NADPH for use in synthetic and anti-oxidant reactions with the carbon skeleton feeding back into the glycolytic pathway. Hexokinase, Phosphofructokinase 1 and Pyruvate Kinase are irreversible steps in glycolysis.

Within the cytosol glucose (a six carbon sugar) undergoes phosphorylation by the enzyme hexokinase (glucokinase in liver) to yield glucose 6-phosphate (G6P) with the phosphate donated from ATP (Fig. 2). An isomerase will convert G6P to fructose 6-phosphate, then to fructose 1,6-bisphosphate (F1,6P2) in another reaction requiring ATP catalyzed by phosphofructokinase 1 (PFK 1). Thus, two ATPs have been used in these initial reactions. The F1,6P2 is split by an aldolase to yield 2 three carbon compounds which progress along the pathway via a series of intermediates to pyruvate. Glyceraldehyde 3-phosphate dehydrogenase will yield NADH and two substrate phosphorylation steps, phosphoglycerate kinase and pyruvate kinase, are linked to the synthesis of ATP. Since each three carbon unit yields 2 ATPs the total yield is 4 ATPs but because two ATPs were utilized in the first part of the pathway the net yield is two ATP per glucose plus two pyruvates and 2 NADH. To maintain flux in the pathway, and to keep using glucose, this NADH must be oxidized back to NAD^+. In oxidative cells the reducing power of the NADH can be transferred to the mitochondria but if this is not possible the NADH will be used by lactate dehydrogenase to reduce the pyruvate to lactate limiting the ATP yield to 2 per glucose metabolized. Sugars such as fructose and galactose can enter the glycolytic pathway after being phosphorylated by specific isozymes of hexokinase and metabolized to intermediates, triose phosphates for fructose and glucose 6 phosphate for galactose, that feed into the glycolytic pathway.

If however, the pyruvate is destined to be fully oxidized in the mitochondria then the NADH formed in the cytosol also must be transferred to the mitochondria. The mitochondrial membrane, however, is impermeable to NAD and NADH and shuttles exist to transfer the reducing power (H^+) in NADH into the mitochondria for oxidative phosphorylation. There are two possible shuttles one that yields NADH within the mitochondria and one that yields $FADH_2$.

An additional pathway of glucose metabolism is the pentose-phosphate pathway (hexose-monophosphate shunt) (Fig. 2). This takes G6P and produces the ribose required for RNA and DNA synthesis. Alternatively the pathway can produce NADPH which is the reducing power required by a number of cytosolic synthetic pathways, such as de novo fatty acid synthesis. NADPH is also used to reduce oxidized glutathione the major cellular antioxidant. When used for NADPH production the carbon re-enters the glycolytic pathway as fructose 6-phosphate.

Pyruvate metabolism

Within the mitochondrial pyruvate is metabolized by pyruvate dehydrogenase, a multi-subunit complex, to yield acetyl CoA (Fig. 3) (Jeoung and Harris, 2010). The first reaction, a thiamine dependent decarboxylase yields a 2 carbon (acetyl) group that is transferred to CoASH forming acetyl CoA. In so doing a lipoate group is reduced which must be re-oxidized to allow continued flux through the complex. This is achieved by transferring the reducing power from the lipoate to NAD to form NADH via a FADH2 intermediate. The reaction is highly regulated mainly by phosphorylation (inhibition) and dephosphorylation (activation) of the decarboxylation step. Regulation of PDHC is extremely important at times of limited glucose availability (when glycogen stores are becoming depleted) since it is possible to synthesize glucose (gluconeogenesis) from pyruvate (3 carbons) but not from acetate (2 carbons).

The acetyl CoA formed can now enter the TCA cycle (Fig. 3). Alternatively, the acetyl CoA can enter the pathway of fatty acid synthesis (Fig. 4) which is important in producing cell membranes and other lipids. In humans however, although theoretically possible, there is very little de novo fatty acid synthesis for storage (Aarsland et al., 1997; Acheson et al., 1988; Hellerstein et al., 1996). Note, this does not mean if you eat carbohydrates you will not get fat. As with any overeating of calories you will become obese, it just means that most of the fat you store will be derived from the diet.

Fat catabolism

Fat in the diet is predominantly in the form of triglycerides (three fatty acids esterified to a glycerol). The end products of digestion are free fatty acids and monoglycerides. These are absorbed in the enterocytes where they are resynthesized to triglyceride and then packaged with a number of specialized proteins into large lipoprotein complexes, chylomicrons. The chylomicrons pass into the lymph system to be transported to the thoracic vein where they are released into the blood circulation. This process is much slower than the digestion of carbohydrates and dietary lipids do not appear in the blood circulation until a number of hours after a meal. On reaching the target tissues the triglycerides in the chylomicrons are hydrolyzed to fatty acids and glycerol by lipoprotein lipase present on the exterior of the cells. Similarly, fatty acids released by the liver in the form of VLDL (very low density lipoproteins) are also hydrolyzed by lipoprotein lipase. The fatty acids can then enter the cells where they can be oxidized to liberate the energy or re-esterified to triglyceride for storage.

As indicated above lipids are transported between tissues as triglycerides (neutral lipids) and a limited amount of free fatty acids. The latter are transported in the circulation bound to the protein albumin since a free acid group is potentially toxic. Usually most dietary lipids are stored for at least some time but when glucose levels fall and insulin levels drop there is an increase in lipolysis (the release of the fatty acids from triglycerides). In adipocytes the triglycerides are stored in a one large lipid droplet, comprising about 80% of the cell volume. This droplet is surrounded by a number of proteins with perilipin being the most important in terms of lipolysis. When lipolysis is stimulated, for example by norepinephrine, perilipin can be phosphorylated and this them allows access of a number of lipases to release the fatty acids from triglycerides. Both free fatty acids and glycerol are released from the cell to be taken up by other tissues.

On being taken up by cells the fatty acids are esterified to CoASH to form a fatty acyl-CoA involving the breakdown of ATP to AMP plus PPi (Fig. 4). It is at this stage that the fate of that acyl CoA may be re-esterification with glycerol-3-phosphate to

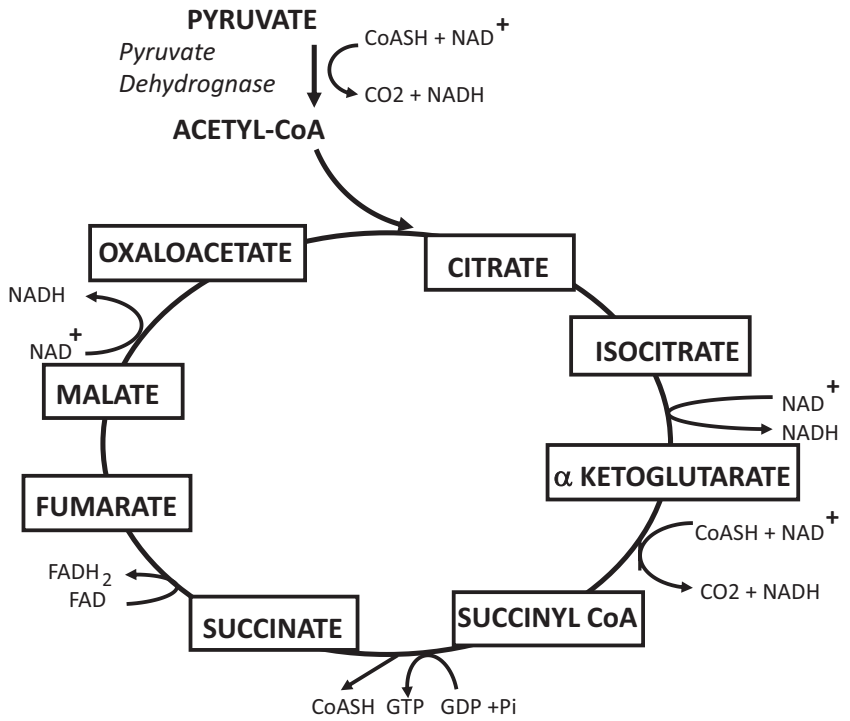


Fig. 3 Pyruvate dehydrogenase and the TCA cycle. Within the mitochondria pyruvate is decarboxylated to form Acetyl CoA which, together with Acetyl CoA from the catabolism of other fuels, enters the TCA cycle. Acetyl CoA (2 C) combines with oxaloacetate (4 C) to form citrate (6 C). In a series of reactions two carbons are lost as CO2 with the production of 2 NADH, 1 $FADH_2$ and a GTP to regenerate the oxaloacetate.

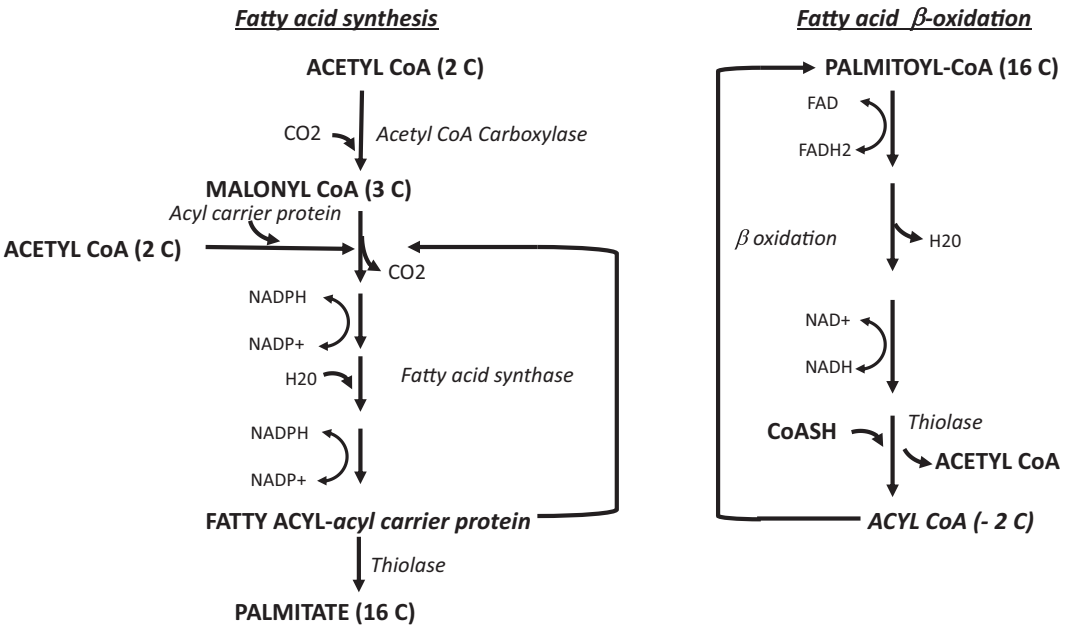


Fig. 4 Fatty acid synthesis and fatty acid β oxidation. Acetyl CoA carboxylase forms malonyl CoA which binds to an acetyl CoA linked to the acyl carrier protein of fatty acid synthase. Within the cytosol, after a decarboxylation, a series of reductions by NADPH, and a hydration, the resulting four carbon acyl group is bound to the acyl carrier protein (ACP) in fatty acid synthase. The process repeats utilizing a new malonyl CoA until the final product, palmitate (16 carbons) is formed and this is released by hydrolysis from the ACP by a thiolase. The palmitate can be further modified by elongation and/or desaturation. Within the mitochondria fatty acyl CoA (e.g., palmitoyl CoA) will be catabolized releasing a FADH2 and a NADH in a process equivalent to the reverse of fatty acid synthesis. The final step is the release of acetyl CoA by a thiolase leaving a fatty acyl CoA that is two carbons shorter than the original. The process then repeats until the final acetyl CoA is released. In liver the end product is usually ketone bodies. See **Fig. 6**.

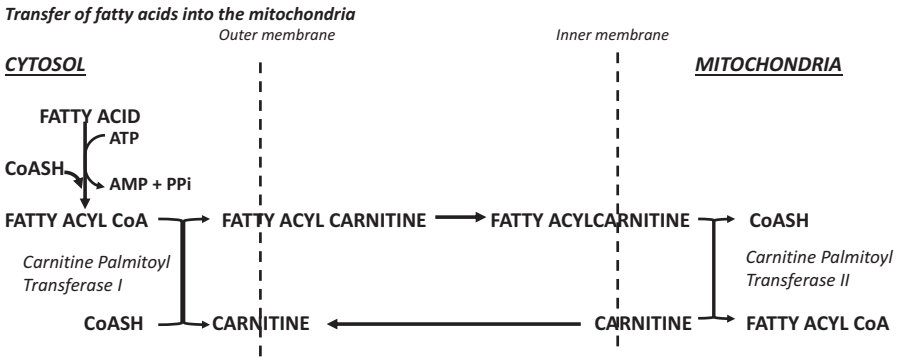


Fig. 5 The Carnitine shuttle. Fatty acids are activated by the addition of CoASH in the cytosol. At this stage the fatty acyl CoA could be used to synthesize triglyceride. If the fatty acid is to be oxidized it must enter the mitochondria but fatty acyl CoA cannot cross the mitochondrial membrane. Carnitine Palmitoyl Transferase I (CPT 1) on the outside of the outer mitochondrial membrane transfers carnitine to the acyl group to replace CoASH. The fatty acyl carnitine now enters the mitochondria where CPT II will reverse the reaction transferring CoASH to the acyl group and liberate carnitine that can return back to CPT 1.

triglyceride for storage in the cytosol or oxidation in the mitochondria. If the fatty acid is to be oxidize the fatty acyl group from the fatty acyl CoA is transferred to carnitine to form a fatty acyl-carnitine by the action of palmitoyl carnitine transferase 1 (CPT 1) (Fig. 5). This occurs on the outer mitochondrial membrane and the fatty acyl-carnitine is transported into the mitochondrial matrix in exchange for free carnitine. Once within the matrix the fatty acyl group is transferred back to CoASH to yield the fatty acyl-CoA that can enter the pathway of β oxidation. CPT1 is strongly inhibited by malonyl CoA which is abundant in the cell at times of high rates of glucose metabolism and therefore inhibits fatty acid oxidation. The β oxidation of fatty acids (Fig. 4) involves a dehydrogenase producing $FADH_2$, followed by a hydratase (addition of water), a further dehydrogenase producing NADH and finally a thiolase to release the acetyl CoA leaving a fatty acyl CoA that is two carbons shorter than the original substrate. The complete process then repeats removing two carbons at a time until the final acetyl CoA is produced. The oxidation of unsaturated fatty acids requires two additional steps to remove the double bond (Berdanier and Berdanier, 2015; Bronk, 1999; Frayn and Evans, 2019; Salway, 2004).

In the liver fatty acids do not undergo complete oxidation to acetyl CoA rather the end products are acetoacetate, β-hydroxybutyrate and acetone (Fig. 6). Collectively these are known as the ketone bodies, although only acetone is actually a ketone. Within the liver two acetyl CoA molecules are linked to yield acetoacetyl CoA that is elongated by the addition of another acetyl CoA group and then lysis to yield acetyl CoA and the free acetoacetate. Acetone, produced by the spontaneous decarboxylation of acetoacetate, is relatively unstable and most is lost in the urine and in expired air. Acetoacetate however can be reduced to β-hydroxybutyrate and these two products circulate throughout the body to be used as a fuel by a number of tissues, including the brain. Ketone bodies are metabolized in non-hepatic tissues by reacting with succinyl CoA to form acetoacetyl-CoA that is cleaved to yield two acetyl CoA (Fig. 6). Essentially all ketone bodies are made in the liver but the liver cannot used them since it lacks the succinyl CoA 3-ketoacid CoA transferase activity required to initiate their breakdown (Berdanier and Berdanier, 2015; Bronk, 1999; Frayn and Evans, 2019; Salway, 2004).

Protein and amino acid metabolism

The end products of protein digestion are free amino acids and a mixture of di- and tri-peptides. Once these enter the enterocyte however, the peptides are immediately hydrolyzed to yield the amino acids. Some amino acids (notably glutamine and glutamate) are metabolized within the enterocyte but most others simply enter the circulation via the portal vein. After a meal there is a rise in the circulating concentrations of amino acids which is accompanied by a general increase in protein synthesis in most tissues. Any amino acids in excess of immediate needs for protein synthesis and non-protein functions, are predominantly metabolized by the liver. Notable exceptions are the branched chain amino acids (BCAA), leucine, valine and isoleucine (usually 25% of the amino acids in a meal). They are not metabolized to any great extent by the intestine or the liver but are degraded in tissues such as skeletal muscle, adipose tissue and kidney. The first step in degradation of the BCAA is a transamination whereby the amino group is transferred to α-ketoglutarate to form glutamate yielding a branched chain keto acid. Some of these keto acids can undergo further degradation within the tissues but large amounts are released into the circulation to be further broken down in the liver. The end products of BCAA metabolism are acetyl CoA and propionate (isoleucine) propionate (valine) and acetoacetyl CoA and acetyl CoA (leucine). Since leucine does not yield a three carbon product leucine cannot be used as a gluconeogenic substrate. Similarly, lysine cannot be used for gluconeogenesis since the end product of lysine metabolism is also acetyl CoA (Berdanier and Berdanier, 2015; Bronk, 1999; Frayn and Evans, 2019; Salway, 2004).

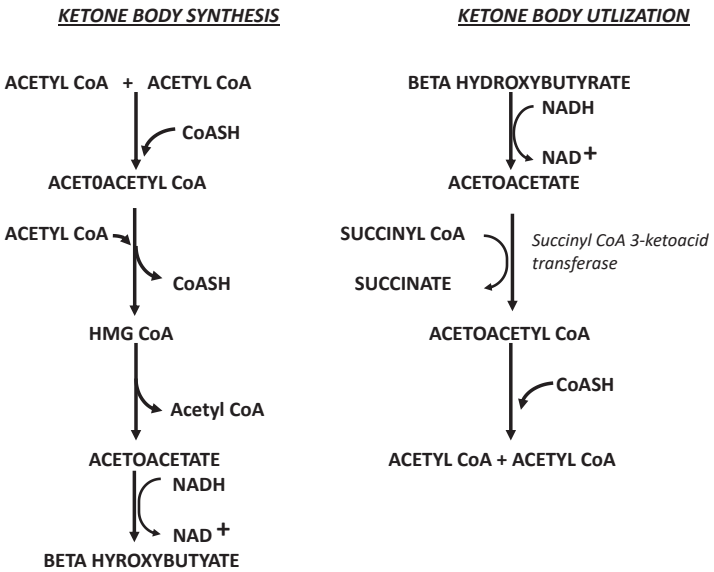


Fig. 6 Ketone body metabolism. In the liver the end products of fatty acid β-oxidation, acetyl CoA, are joined to make Acetoacetyl CoA. An additional Acetyl CoA is added to form HMG CoA when then cleaved to release acetyl CoA and form acetoacetate. Acetoacetate can be reduced using NADH to form β-hydroxybutyrate. Acetoacetate and β-hydroxybutyrate are released into the circulation as fuel for other tissues. Some amino acids can also give rise to acetoacetate. When ketone bodies are oxidized the b-hydroxybutyrate is first oxidized to aceto acetate followed by the addition of CoASH from succinyl CoA releasing succinate. The acetoacetyl CoA is then split to form 2 acetyl CoA. The liver is unable to use ketones as a fuel since it does not express the succinyl CoA 3-oxoacid transferase.

Although the pathways of catabolism are different for every amino acid they do follow a common sequence. Usually the nitrogen is removed, often by transamination, although a few do undergo direct hydrolysis with release of the nitrogen as ammonia. The carbon skeleton will then be further metabolized to acetyl CoA, ketone bodies, propionate and a number of other products that can feed directly into the TCA cycle (**Fig. 7**). In the liver the nitrogen is converted to urea to be excreted and the bulk of the carbon skeletons will be used for gluconeogenesis (replenishing hepatic glycogen stores in the absorptive state) (Watford, 2020a). Similarly, the degradation of glutamine in the kidney is linked to the provision of urinary ammonia to correct acidosis with the carbon skeleton being recovered as glucose through renal gluconeogenesis (Watford, 2020b). Certain cell types, notably those that are replicating and proliferating rapidly, or are very short lived, for example, cells of the immune system, enterocytes, and cancer cells, utilize large amounts of glutamine as their major respiratory fuel through metabolism of the carbon skeleton through parts of the TCA cycle (Watford, 2020b). Such cells also show limited glucose oxidation with high rates of lactate formation (aerobic glycolysis).

Ethanol

Humans express considerable alcohol dehydrogenase the main enzyme of ethanol metabolism. Most is in the liver with some evidence of limited expression in the stomach. On entering the cell alcohol undergoes reduction to yield acetaldehyde and NADH in the cytosol. The acetaldehyde enters the mitochondria where a further reduction occurs to form acetate and NADH. The acetate may be further metabolized within the liver or released to be used by other tissues.

The tricarboxylic acid cycle and oxidative phosphorylation

As can be seen above the major product arising from macronutrient metabolism is acetyl CoA (even if the initial pathway yielded ketone bodies or glucose). Thus the fate of acetyl CoA is important in all cells that possess mitochondria and acetyl CoA will enter the TCA cycle, also known as the citric acid cycle or the Krebs cycle (**Fig. 3**). As implied from the name the pathway is cyclic in nature with a four carbon compound, oxaloacetate (OAA), being the starting point. Acetyl CoA condenses with OAA to form citrate (6 carbons). This then converts to α-ketoglutarate with the release of CO_2 and production of NADH. The α-ketoglutarate (5 carbons) undergoes a reaction similar to PDHC, with the release of another CO_2 and NADH and yields succinyl CoA. The CoASH from succinyl CoA is released involving a substrate level phosphorylation of GDP to GTP (an ATP equivalent) to produce succinate (4 carbons). Further metabolism to fumarate yields $FADH_2$, and then to malate followed by another dehydrogenase yielding another NADH and regeneration of the OAA. The TCA rotates twice for each glucose molecule (Berdanier and Berdanier, 2015; Bronk, 1999; Frayn and Evans, 2019; Salway, 2004).

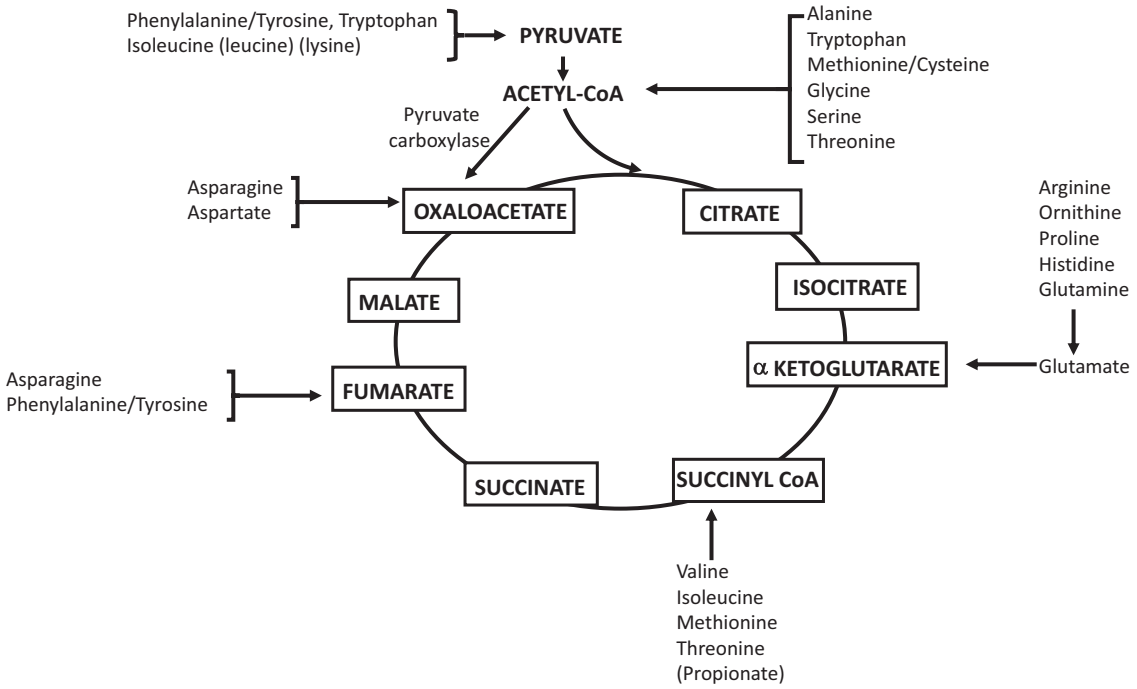


Fig. 7 Entry of carbon skeletons into the TCA cycle. In addition to acetyl CoA for oxidation the TCA can use acetyl and other substrates for oxidation or as part of other pathways such as gluconeogenesis. Leucine and lysine are shown in parentheses since these amino acids only give rise to acetoacetate (a 4 carbon molecule equivalent to two acetyl CoA) and therefore cannot be used for gluconeogenesis. The entry of propionate (3 carbons) is also shown since this is the only fatty acid that can be used for gluconeogenesis.

Thus at the end of glucose oxidation the mitochondrial matrix accumulates 8 NADH, 2 FADH2 ad a GTP from the TCA cycle, 2 NADH from PDHC, plus 2 NADH (shuttled in from cytosolic glycolysis) and the release of 6 CO_2. The NADH/$FADH_2$ is now used to generate ATP via the process of oxidative phosphorylation.

As stated earlier the hydrogen of NADH is generally referred to as "reducing power" or "reducing equivalents" since it is extensively used to reduce other metabolites. Within the mitochondria these hydrogens dissociate into protons (H^+) and electrons (e^-). The electrons pass along what is known as the electron transfer chain within the inner membrane of the mitochondria (**Fig. 8**). NADH enters at complex 1 where it is oxidized to NAD^+ with the H^+ transferring to FADH to yield $FADH_2$. At this step a number of protons are pumped from the mitochondrial matrix into the inter-membrane space. The $FADH_2$, both from NADH oxidation and pathways such as β-oxidation and the TCA cycle, is oxidized at Complex II by reduction of ubiquinone (coenzyme Q), and then on to the cytochromes each one being reduced and then re-oxidized as it reduces the next one. At complexes III and IV more protons are pumped into the inter-membrane space. The cytochromes, named because of their coloration, contain a reactive ferric ion that becomes reduced to ferrous and then finally is re-oxidized at Complex IV by cytochrome oxidase (a copper containing enzyme) with the electrons passed along to the final acceptor oxygen to yield water. Thus the oxidative metabolism of acetyl CoA yields CO_2 and water as the end products. The result of this process is a buildup of protons in the intermembrane space (crista space) forming a what is known as the proton-motive force. These protons then re-enter the matrix via complex V, the ATP synthase. In so doing they drive a "motor" subunit in the matrix that powers the formation of ATP from ADP and Pi (Berdanier and Berdanier, 2015; Bronk, 1999; Frayn and Evans, 2019; Salway, 2004). The number of protons pumped per NADH oxidized is commonly accepted as 10. From this textbooks usually state that total oxidation of glucose will yield 36–38 ATP. It is now realized that NADH oxidation does not yield 3 ATPs and the actual ATP yield from glucose oxidation is 27–29 ATPs (Bronk, 1999; Rich, 2003). The process is generally known as oxidative phosphorylation due to the coupling of the electron transport chain with ATP synthesis and the mechanism as the chemiosmotic theory.

Uncoupling and brown adipose tissue

A number of metabolic poisons work by interfering with oxidative phosphorylation. Cyanide will bind to the copper in cytochrome oxidase, while rotenone inhibits electron transfer. Another group of poisons are uncouplers, of which dinitrophenol (DNP) is perhaps the best known example. Uncouplers simply punch holes in the membrane and allow large amounts of protons to leak back into the matrix without ATP synthesis in so doing they do generate considerable amounts of heat. In the 1920s DNP was sold as a weight loss supplement but was withdrawn from the market due to severe side effects, often as severe as death.

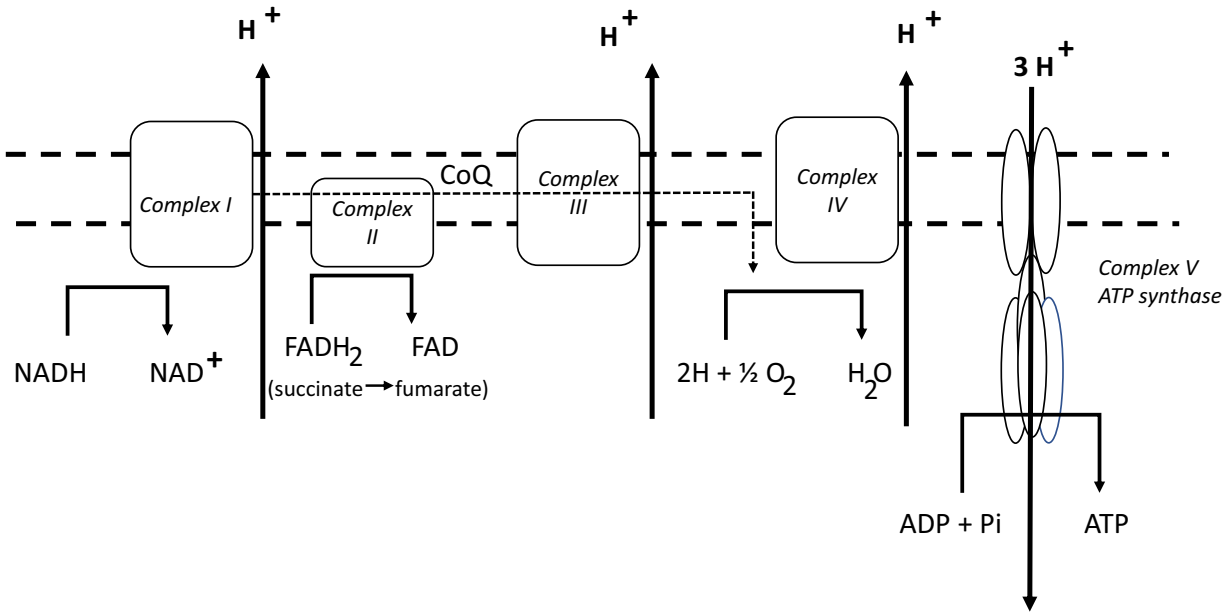


Fig. 8 Oxidative phosphorylation. Within the mitochondria the protons and electrons from NADH and FADH2 are split and the electrons pass along the electron transport chain (shown as the dotted line) embedded in the inner membrane to the final electron acceptor molecular oxygen to produce water. At complexes I, III and IV protons are pumped out of the matrix into the inter-membrane space. This establishes the proton motive force (a build up of protons in the inter-membrane space). The protons then re-enter the matrix at Complex V driving the ATP synthase reaction to phosphorylate ADP to ATP. The process is also referred to as coupled respiration but some compounds, and the Uncoupling protein 1, can allow the protons to reenter the matrix without passing through Complex 5 with the energy released being dissipated as heat.

Unfortunately, with the rise of the internet, DNP has recently re-appeared as a weight loss supplement and is resulting in a rising number of deaths (Grundlingh et al., 2011). Some uncoupling also occurs naturally in mitochondria in most cells which again decreases the ATP yield from electron transfer.

Most adipose tissue is classified as white, due to its color, and is the major site of energy storage in the body. A different form of adipose tissue, known as brown adipose tissue (BAT), also exists which functions to product heat (Siddosis and Kajimura, 2015). Unlike white adipose tissue BAT contains numerous mitochondria (hence the brown color) and many lipid droplets. But BAT also expresses a unique protein, uncoupling protein 1 (UCP 1). As its name implies, it is an uncoupler and allows protons to re-enter the matrix and bypass the ATP synthase dissipating the energy as heat. For many years it was assumed that BAT was only present in humans as babies and was not found in adults. By using sophisticated imaging techniques however, the presence of BAT was detected in adults but it is only detectable when active, i.e., in the cold. The everyday response to cold is to shiver but activation of BAT allows non-shivering thermogenesis (heat production). BAT is more common in northern climates, more common in winter, higher in women than men, and lower in obese subjects. Once cold is detected lipolysis within BAT increases the concentration of fatty acyl CoA that activates UCP1 and heat is produced. White adipose cells and brown adipose have very different cell origins and recent work has shown that some of the UCP1 containing cells in adult humans are derived from the same lineage as white adipose cells. Such cells are not as brown as pure BAT and are termed beige.

Energy storage and fuel use in the post-absorptive state

The previous sections have dealt with how fuels are catabolized releasing their energy to do useful work. Two of the fuels, carbohydrates and fats are also stored in the body to be used between meals and longer periods of starvation, or during exercise (Cahill, 2006; Watford, 2015). There is no purely storage form of protein since all proteins have unique functions other than as a fuel. Similarly, there is no storage of ethanol and the metabolism of ethanol is actually prioritized over all other fuels by the liver.

Triglycerides

In humans white adipose tissue is the most important site of triglyceride storage. Most fatty acids stored as fuel are derived from the diet. A minor contribution will arise from de novo fatty acid synthesis (Fig. 4) that occurs in the cytosol from acetyl CoA which is converted to malonyl CoA followed by a series of reactions in the fatty acid synthase enzyme complex (Berdanier and Berdanier, 2015; Bronk, 1999; Frayn and Evans, 2019; Salway, 2004). In a reversal of β-oxidation acetyl CoA (2 carbons)

are sequentially added as malonyl CoA until the chain length is 16 carbons which is then released as palmityl CoA. This can then undergo elongation and desaturation to form the variety of long chain fatty acids. Within the cytosol the high levels of malonyl CoA will block CPT1 preventing the fatty acid from entering the mitochondria to be oxidized. Therefore any long chain fatty acids, from de novo synthesis or taken up by the cell, will be esterified to glycerol 3 phosphate to form triglyceride (a neutral lipid) and then stored in the lipid droplet. The glycerol 3-phosphate can arise from glucose via glycolysis or from lactate by the pathway of glyceroneogenesis. In addition, the liver expresses glycerol kinase that enables the glycerol released from lipolysis in other tissues to be phosphorylated and either used for triglyceride synthesis or as a substrate for gluconeogenesis. In the post-absorptive state insulin levels drop which activates lipolysis in adipose tissue and the free fatty acids become available as fuel for other tissues.

Glycogen

During the absorptive phase the blood glucose concentration rises and any in excess of immediate needs will be stored as glycogen mainly in liver and skeletal muscle through the action of glycogen synthase (**Fig. 2**). Glycogen is a polysaccharide (a polymer of glucose with extensive branching), sometimes called animal starch. Glycogen is synthesized from glucose 6-phosphate derived from glucose in muscle and both from glucose and gluconeogenesis in liver. Glycogen is mobilized by release of glucose 6-phosphate by the enzyme (glycogen) phosphorylase (**Fig. 9**) (Berdanier and Berdanier, 2015; Bronk, 1999; Frayn and Evans, 2019; Salway, 2004).

Muscle glycogen stores are of the order of 300 g and muscle glycogen functions as a local, intracellular, fuel. Stimulation of the muscle to contract releases calcium within the cell and this stimulates glycogen breakdown and glycolysis and so provides the ATP for contraction. White (Type 2b) muscles possess high levels of glycogen which fuels glycolysis to lactate (anaerobic metabolism) yielding very high power (ATP formation) but limited endurance. Glycogen levels in muscle are also important in aerobic exercise, where the oxidation of glucose yields twice the power (rate of ATP production) compared with the oxidation of fatty acids. Thus, the more glycogen present in muscle longer the endurance at maximal aerobic metabolism and endurance athletes develop

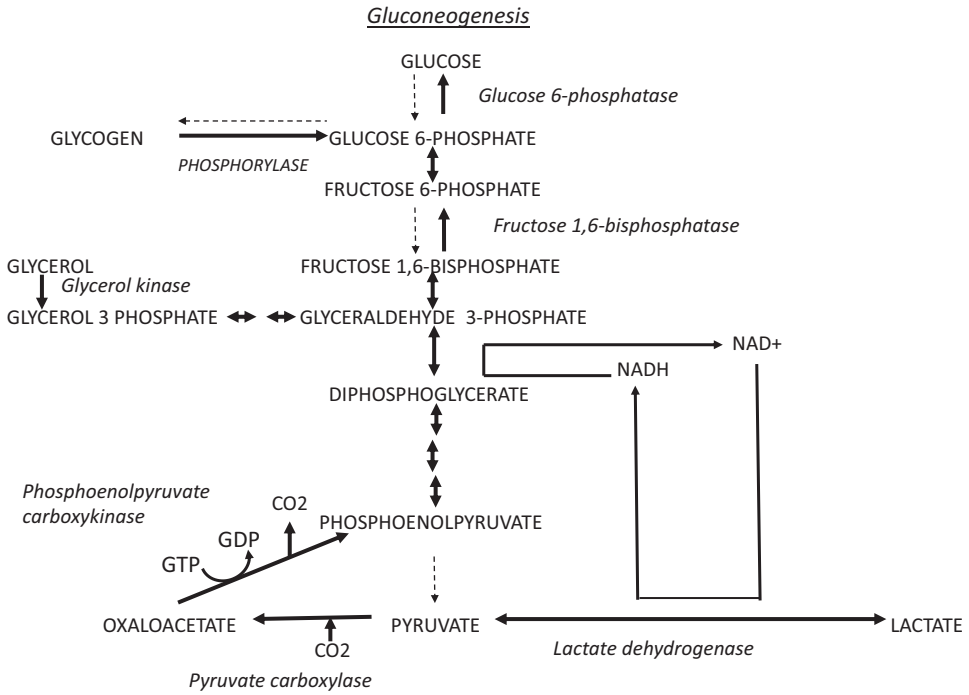


Fig. 9 Glycogen breakdown and gluconeogenesis. In the liver excess glucose can be stored as glycogen (see **Fig. 2**). In the post absorptive state this glycogen is broken down by phosphorylase to yield G6P that can be hydrolyzed by glucose 6-phosphatase with the resulting glucose being released into the circulation to maintain blood glucose levels. As liver glycogen becomes depleted blood glucose levels are maintained by the process of gluconeogenesis. The main substrates are lactate (together with pyruvate and alanine), glycerol and amino acids (see **Fig. 7**). Glycerol is phosphorylated by glycerol kinase and enters the pathway at glyceraldehyde 3-phosphate. The other substrates all enter the pathway within the mitochondria where they give rise to oxaloacetate. The oxaloacetate then transfers to the cytosol via shuttles and is converted back to phosphoenolpyruvate by phosphoenolpyruvate carboxykinase (PEPCK). The PEP then essentially reverses the glycolytic pathway to yield free glucose. Three irreversible steps (shown as dotted arrows) in glycolysis are reversed by gluconeogenic specific enzymes: Pyruvate Kinase, reversed by Pyruvate carboxylase and PEPCK, Phosphofructokinase 1 reversed by fructose1,6-bisphosphatase, Glucokinase (Hexokinase) reversed by glucose 6-phosphate. Gluconeogenesis also occurs in the kidney where it is predominantly linked to the production of ammonia from glutamine at times of metabolic acidosis.

specific dietary regimes to maximize their muscle glycogen stores (Frayn and Evans, 2019). The phenomenon of "hitting the wall" in a marathon, when the runner can no longer maintain high intensity, is believed to occur when the muscle glycogen reserves have been depleted and the muscles must rely of fat oxidation.

Glycogen breakdown and gluconeogenesis provide glucose to the body

Liver glycogen is an important source of circulating glucose in the post-absorptive state (**Fig. 9**). This is possible since liver expresses glucose 6-phosphatase which enables the release of free glucose to be released into the circulation on stimulation by glucagon. Liver glycogen stores are of the order of 100 g in an adult but the turnover of glucose is approximately 10 g per hour which would mean that liver glycogen stores would be rapidly deleted. But as insulin levels fall in the post-absorptive state lipolysis in adipose increases and the availability of free fatty acids in the circulation can replace some of the glucose utilization in most tissues. Nevertheless, given that the brain continues to fully oxidize 120 g of glucose per day liver glycogen stores can only maintain the brain for less than a day. In addition to stimulating glycogen breakdown glucagon also stimulates gluconeogenesis in liver (Cahill, 2006; Watford, 2015).

Gluconeogenesis (**Fig. 9**) is the synthesis of glucose from non-carbohydrate precursors and will increase during the post-absorptive period and early starvation to keep blood glucose concentrations stable. Glucose can only be formed from 3 carbon substrates and not from 2 carbon units such as those in fatty acids and ketone bodies. The substrates for gluconeogenesis are lactate (pyruvate and alanine), glycerol from triglyceride lipolysis, amino acids, and propionate (see **Fig. 7**). The last is the product of fermentation of dietary fiber by the colonic microbiota and is of little importance during prolonged starvation. Lactate, pyruvate and most alanine are derived from pyruvate that is itself derived from glucose in peripheral tissues. Therefore, the use of these three substrates for gluconeogenesis simply recycles the carbon skeletons and does not contribute "new" glucose for other tissues. This means that glycerol and amino acids are the only gluconeogenic substrates that can provide glucose for the brain to oxidize. From day one of starvation glycerol, derived from lipolysis of triglycerides, contributes approximately 20 g of new glucose. Therefore, once hepatic glycogen stores have been depleted gluconeogenesis from amino acids must provide 100 g of glucose. It has been estimated that it takes 180 g of amino acids to make 100 g of glucose which would require the breakdown of this much protein (approximately 800 g of muscle tissue) each day. Clearly this cannot be maintained for long and during starvation ketone bodies and fatty acids become the dominant oxidative fuels for most tissues. This is known as "the glucose sparing effect of fat derived fuels" or as the protein sparing effect. As ketone concentrations rise the brain will begin to oxidize them instead of glucose thereby further reducing the requirement for glucose. In very long starvation however, the brain still utilizes some glucose but even the metabolism of most of that is restricted to glycolysis to lactate, so conserving the glucose carbon skeletons. During prolonged starvation the high levels of ketone bodies produces an acidosis which is titrated by as ammonium salts with the ammonia being synthesized in the kidney from glutamine. The carbon skeleton of glutamine is further metabolized within the kidney as a substrate for renal gluconeogenesis (Watford, 2015, 2020a).

Summary

Within the body different tissues show distinct patterns of fuel use to provide the ATP that is required for optimal function. In the absorptive state glucose is plentiful which increases insulin secretion and thereby increases glucose uptake into cells, suppresses hepatic glucose output, and inhibits the release of fatty acids from triglyceride stores. Insulin also increases the utilization of glucose both for energy production in most cells and storage as glycogen in liver. Similarly, dietary fats are stored, as triglycerides, in adipose tissue. Excess proteins will be degraded with the carbon skeletons used for gluconeogenesis (glycogen storage). During the post-absorptive period blood glucose levels decrease slightly and insulin levels fall. This removes the inhibition of lipolysis which then releases the fatty acids to be oxidized in place of glucose. Glucagon rises to stimulate the mobilization of hepatic glycogen to provide glucose to those tissues, such as the brain, that continue to use glucose. As hepatic glycogen stores become depleted blood glucose levels are maintained by gluconeogenesis, mainly in the liver using glycerol (from triglycerides) and amino acids (from protein degradation) as substrates. The partial oxidation of fatty acids in the liver results in increasing ketone body levels that will be utilized by the brain in place of glucose. Together the utilization of fatty acids and ketones as fuels is known as the "glucose sparing effect of fat derived fuels" (Randle et al., 1963). An alternative description is, the protein sparing effect, since the glucose would have been made from amino acids released from protein breakdown. This well-orchestrated pattern of fuel metabolism maintains fuels (glucose and ketones) to the brain and fatty acids and ketones to other tissues from the absorptive period into starvation (Cahill, 2006; Watford, 2015).

References

Aarsland, A., Chinkes, D., Wolfe, R.R., 1997. Hepatic and whole-body fat synthesis in humans during carbohydrate feeding. Am. J. Clin. Nutr. 65, 1774–1782.

Acheson, J.R., Schutz, Y., Bessard, T., Anantharaman, K., Flatt, J.P., Jequier, E., 1988. Glycogen storage capacity and de novo lipogenesis during massive carbohydrate overfeeding in man. Am. J. Clin. Nutr. 48, 240–247.

Berdanier, C.D., Berdanier, L.A., 2015. Advanced Nutrition: Macronutrients, Micronutrients and Metabolism. CRC Press, Boca Raton.

Bronk, J.R., 1999. Human Metabolism. Addison Welsley Longman Limited, Harlow.
Cahill Jr., G.F., 2006. Fuel utilization in starvation. Annu. Rev. Nutr. 26, 1–22.
Frayn, K., Evans, R., 2019. Human Metabolism, fourth ed. Wiley Blackwell, Hoboken, Chichester.
Grundlingh, J., Dargan, P.I., El-Zanfaly, M., Wood, D.M., 2011. 2,4-Dinitrophenol (DNP): a weight loss agent with significant acute toxicology and risk of death. J. Med. Toxicol. 7, 205–212.
Hellerstein, M.K., Schwarz, J.M., Neese, R.A., 1996. Regulation of hepatic de novo lipogenesis in humans. Annu. Rev. Nutr. 16, 523–557.
Jeoung, N.H., Harris, R.A., 2010. Role of pyruvate dehydrogenase kinase 4 in the regulation of blood glucose levels. Korean Diabetes J. 34, 274–283.
Randle, P.J., Garland, P.B., Hales, C.N., Newsholme, E.A., 1963. The glucose fatty acid cycle: its role in insulin sensitivity and the metabolic disturbances of diabetes mellitus. Lancet 785–789.
Rich, P., 2003. A re-evaluation of yields of ATP synthesis from glucose oxidation. Biochem. Soc. Trans. 31, 1095–1105.
Salway, J.G., 2004. Metabolism at a Glance, third ed. Blackwell Scientific Publications, London, Boston.
Siddosis, L., Kajimura, S., 2015. Brown and beige fat in human: thermogenic adipocytes that control energy and glucose homeostasis. J. Clin. Invest. 125, 478–486.
Watford, M., 2015. Starvation: Metabolic Changes. Encyclopedia of Life Sciences. John Wiley and Sons, Ltd, Chichester. https://doi.org/10.1002/9780470015902.a0000642.pub2.
Watford, M., 2020a. Ornithine Cycle. Encyclopedia of Biochemistry. Elsevier. https://doi.org/10.1016/B978-0-12-819460-7.00062-1.
Watford, M., 2020b. Glutamine. Encyclopedia of Biochemistry. Elsevier. https://doi.org/10.1016/B978-0-12-819460-7.00028-1.

Energy: Adaptation

Abdul G. Dulloo, Department of Medicine, Physiology Unit, University of Fribourg, Fribourg, Switzerland

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A.G. Dulloo, Energy: Adaptation, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 146–153, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00088-X.

Introduction 128
Energy adaptation: beyond mass action 128
Energy adaptation: inter-individual variability 129
What constitutes adaptive thermogenesis? 129
Adaptive thermogenesis in the resting state 129
Adaptive thermogenesis in the non-resting state 130
Variability in spontaneous physical activity 131
Efficiency of muscle work 131
Interactions and compensations across components of EE 131
Autoregulation of body weight and body composition 131
A compartmental model 132
Biological significance of a dual control system for adaptive thermogenesis 133
Energy adaptation during a longitudinal human study of large weight fluctuations 134
Energy adaptations and susceptibility to leanness and fatness 135
Concluding remarks 136
Further reading 136

Key points

- Humans can adapt energetically to changes in body weight through autoregulatory control systems that operate through an intrinsic (internal) control of body energy partitioning between lean and fat tissues (lean-fat partitioning) and through adaptive thermogenesis (regulated heat production); these contribute to energy conservation during energy deficit or increased energy dissipation during energy surplus.
- Even under standardized conditions of experimental overnutrition (weight gain) or undernutrition (weight loss), there is a large inter-individual variability in lean-fat partitioning and in adaptive thermogenesis, both of which contribute to the large inter-individual variability in weight gain during overfeeding or weight loss during underfeeding.
- The available evidence suggest the existence of two distinct control systems for adaptive thermogenesis: a relatively rapid reaction system that respond to energy imbalance and a relatively slow reaction system responding to changes in the body's fat reserves *per se*.
- Adaptive thermogenesis may occur in any of the components of daily energy expenditure, and could hence result in altered efficiency of energy utilization in the resting and/or non-resting states, whether during weight gain, weight loss or after the dynamic phase of weight change.
- There may hence be considerable inter-individual differences among humans in metabolic strategies to conserve in response to energy deficit or dissipate some of the excess energy in response to energy surplus.
- Subtle variations between individuals in lean-fat partitioning and in adaptive thermogenesis, cumulated over the long-term amid a fluctuating body weight, can be important in determining long-term weight maintenance over months and years in some individuals and in provoking the drift toward obesity in others.
- These energy adaptations are relevant as much to the etiology of obesity and in undermining therapeutic slimming as to the difficulties in regaining a healthy body composition (and functional capacity) during nutritional rehabilitation after disease cachexia or perturbed growth.

List of abbreviations

BMR Basal metabolic rate
DIT Diet-induced thermogenesis
EE Energy expenditure
FFM Fat-free mass
Pc Partitioning characteristic

REE Resting energy expenditure
SNS Sympathetic nervous system
SPA Spontaneous physical activity

Introduction

Current notions about how humans adapt energetically to food deficit or surplus have origins in seminal works that were conducted in the late 19th and early 20th centuries.

First, it has long been advocated that the body's fat reserves play a critical role in ensuring the energy needs of the individual during long periods of food scarcity, while protecting the functional integrity of the lean mass (and vital organs). Indeed, an inverse association between the contribution of the body's protein as fuel during starvation and the initial (pre-starvation) degree of fatness was observed from inter-species studies more than a century ago. It was only several decades later (in the 1980s) that this lean-fat interrelationship in response to energy deficit became firmly established in humans. This was notably through (i) the work of Gilbert Forbes on the "companionship" of the lean and fat compartments of the body in response to dietary-induced alterations in body weight, and (ii) that of Marinos Elia in his remarkable analysis of human body composition data pertaining to the effect of starvation and very low calorie diets on protein–energy interrelationships. These publications have formed the basis of mechanistic models to help explain human variability in maximum percentage of body weight that can be lost and in predicting survival duration during prolonged starvation.

Second, in early human investigations into physiological adaptations pertaining to protein-energy metabolism during starvation, Francis Benedict's documentation of the day-to-day changes in basal oxygen consumption during the course of prolonged fasting revealed a greater-than-expected fall in basal metabolic rate (BMR) relative to the loss in body weight, thereby suggesting the possible existence of an adaptive energy conservation phenomenon in response to food deficit. This notion was strongly supported a few decades later in men subjected to 6 months of semistarvation in the classic Minnesota Starvation Experiment directed by Ancel Keys, and subsequently extended to people with obesity undergoing therapeutic dieting. This phenomenon of energy conservation during weight loss, nowadays often referred to as "metabolic adaptation", thus accounts for some of the less-than-expected weight loss from the prescribed dietary regimen and/or exercise therapies, and can hence influence the short- and long-term outcomes of weight loss interventions.

Third, the notion that certain individuals can offer considerable resistance to weight gain during overeating through the ability to burn the excess calories also has a long history. Carl Voit, in his 1881 theory of plethora, suggested that excess calories may be converted directly to heat, a phenomenon which Rudolf Neumann later referred to as "Luxus Konsumption". This term was replaced by "thermogenesis" several decades later in mid-1960 when Miller and Mumford discussed the potential importance of variations in heat production in the etiology of obesity. In subsequent discussion of their "Gluttony experiments" in university students, they attributed the large inter-individual variability in weight gain to varying capacity in disposing of excess calories as heat, and introduced the term "dietary-induced thermogenesis". This was later shortened to "diet-induced thermogenesis" and abbreviated as "DIT". To-day, the variations in heat production that serve the purpose of buffering body weight against energy imbalance—whether when referring to DIT in response to energy surplus or to "metabolic adaptation" in response to energy deficit—are embodied in the concept of regulated heat production or adaptive thermogenesis.

Given the famine-and-feast lifestyle that has characterized human evolutionary history, it is conceivable that these phenomena of energy adaptations—through adaptive thermogenesis and the management of the body's main energy-containing (protein and fat) compartments—evolved as key control systems in the regulation of body weight and body composition. This article provides an overview of current concepts about the nature of these control systems, and their modes of operation in enabling the human body to adapt to nutritional stresses and to achieve weight homeostasis.

Energy adaptation: beyond mass action

Understanding how long-term weight homeostasis is achieved is still challenging for human research today. It is nonetheless recognized that the regulation of body weight results from a complex integration of genetic, epigenetic, physiological and behavioral factors which are mediated by numerous hormonal, metabolic and neural signals. In individuals in whom body weight is tightly regulated in the face of widely varying levels of food intake and energy expenditure (EE), this is achieved through autoregulatory feedback control systems that enable the precise matching of energy intake and EE in order to achieve long-term energy balance.

In fact, it has long been realized that there is a built-in stabilizing mechanism into the overall regulatory system that enables weight homeostasis. Any imbalance between energy intake and energy requirements will eventually lead to a change in body weight which, in turn, will alter the maintenance energy requirements in a direction that will tend to counteract the original imbalance and restore body weight to its original level. In this context, the system exhibits "dynamic equilibrium". For example, an increase in body weight will be predicted to lead to an increase in EE due to the extra energy needed for synthesis and maintenance of additional lean

and fat tissues. This will tend to produce a negative energy balance and thus a subsequent decrease in body weight toward its "set" or "preferred" value. Conversely, a reduction in body weight will lead to a reduction in EE due to the loss in lean and fat tissues, which will tend to produce a positive energy balance and a subsequent return toward the "set" or "preferred" weight. In reality, however, the homeostatic system is more complex than this simple effect of mass action as the efficiency of metabolism—through the extent of lean-fat partition and adaptive thermogenesis—can also change following alterations in body weight.

The most compelling evidence in support of a role for adaptive thermogenesis in the adjustments of daily EE following altered body weight derives from the "weight-clamping" experiment in which the subjects were made to lose or gain 10% of their habitual body weight, and which they had to maintain for several weeks prior to the determination of their daily EE. Subjects who were deliberately maintaining body weight at 10% above their habitual level showed an increase in daily EE of 15% on average even after adjusting for changes in body weight and body composition. Conversely, in subjects who were deliberately made to maintain body weight at 10% below their habitual level, daily EE, after adjusting for the losses in weight and lean tissues, was lower by about 15% on average. These compensatory changes in daily EE reflect changes in metabolic efficiency (i.e., adaptive thermogenesis) that oppose the maintenance of a body weight that is above or below the "set" or "preferred" body weight, and were observed in both men and women, and in subjects with or without obesity.

Energy adaptation: inter-individual variability

Another important outcome of the above-mentioned "weight clamping" experiment is that it also underscored a wide inter-individual variability in the ability to readjust daily EE following weight gain or weight loss. Some individuals showed little or no evidence for adaptive thermogenesis while some others exhibited marked ability to decrease or increase EE through adaptive thermogenesis that was estimated to correspond to as much as 2 MJ day^{-1}.

Indeed, a large inter-individual variability in weight gain under standardized overfeeding conditions or in the efficiency of weight gain (i.e., the amount of weight gain per unit of excess energy consumed) has been observed in virtually all experiments of human overfeeding that ranged in duration from a few weeks to a few months. These differences could be attributed in part to inter-individual variability in the gain of lean tissue relative to fat tissue (thus reflecting variability in lean-fat partitioning), and in part to variability in the capacity for DIT. A subsequent detailed reanalysis of data from some 150 men and women who participated in overeating experiments conducted in the second half of the 20th century suggested that more than 40% of these overfed subjects must have exhibited an increase in DIT, albeit to varying degrees.

An important role for genetic factors in accounting for some of the human variability in metabolic susceptibility to weight gain has in fact been established from an overfeeding study in identical twins. The demonstration of intra-pair resemblance in their gain in body weight, body fat and lean mass have strongly suggested that genes play an important role in lean-fat partitioning of the excess energy consumed and in EE responses. Conversely, a role for genes in determining the large inter-individual variability in the amount and composition of weight lost, and in variability in the estimated increase in metabolic efficiency during therapeutic slimming has been suggested by intrapair resemblance during weight loss on a very low calorie diet in female identical twins with obesity.

Taken together, these studies suggest that in addition to the control of food intake, changes in lean-fat partitioning and in metabolic efficiency (adaptive thermogenesis) play an important role in the regulation of body weight and body composition, and that the magnitude of these adaptive changes is strongly influenced by genetic factors.

What constitutes adaptive thermogenesis?

The quantification of adaptive thermogenesis during different states of weight changes remain challenging, amid several lines of evidence indicating that it could be occurring in any of the various components and subcomponents of daily EE, i.e., within the resting and/or non-resting state (Fig. 1).

Adaptive thermogenesis in the resting state

The majority of studies investigating the existence of adaptive thermogenesis pertaining to weight regulation have assessed EE only as BMR or resting EE (REE), which constitutes the largest component of daily EE in most sedentary humans typical of modern societies. In response to underfeeding or overfeeding, the changes in mass-adjusted BMR or REE, i.e., changes in EE that are greater than those predicted by changes in fat-free mass (FFM) and fat mass, can be quantified with the use of regression models to calculate the residuals (measured-predicted) values. The latter "residuals" (hence deviations from the predicted values) are considered to reflect changes in metabolic "efficiency" due to adaptive thermogenesis.

Reports of increases in mass-adjusted BMR, as well as in the thermic effect of food (expressed as % of meal energy) during experimental overfeeding is equivocal, and may be influenced by protein intake or the inclusion of exercise in the overfeeding protocol. By contrast, most studies of experimental starvation or hypocaloric dieting have reported reductions in mass-adjusted BMR or REE, with average values in the range of 0.4–1 MJ day^{-1}. Based upon the application of imaging techniques to assess the contribution of

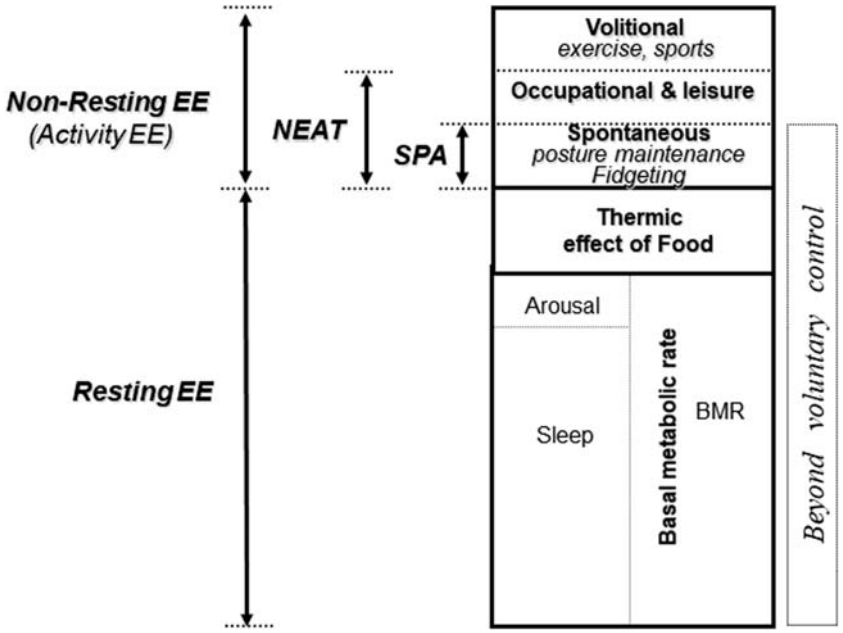


Fig. 1 Schematic representation of daily energy expenditure components and sub-components. Daily energy expenditure (EE) can be divided into resting and non-resting components from which resting EE is the larger component in sedentary, light-activity individuals; these various components or sub-components can be assessed by ventilated canopy or facemask indirect calorimetry or in a respiratory chamber. The compartment of resting EE comprises all measurements of EE generally made at rest—the basal metabolic rate (BMR), sleeping EE and the thermic effect of food (TEF)—and which are essentially beyond voluntary control. Non-resting EE (often referred to as activity EE) comprises both involuntary and voluntary physical activities; it is divided into volitional and non-exercise activity thermogenesis (NEAT), which in turn is subdivided into occupational/leisure activity and spontaneous physical activity (SPA), the latter being essentially involuntary and subsconscious. Adapted from Dulloo, A.G., Jacquet, J., Montani, J.P., Schutz, Y., 2012. Adaptive thermogenesis in human body weight regulation: more of a concept than a measurable entity? Obes. Rev. 13 (Suppl. 2), 105–121.

the loss of individual organ mass to the decline in REE following weight loss on a low-calorie diet in women who were overweight or with obesity, more than a-third of the fall in REE could be attributed to adaptive thermogenesis.

A reduction in mass-adjusted BMR or REE has also been shown in subjects with obesity following weight reduction programs that involved dieting in combination with intense exercise and in which the loss of FFM is prevented or minimal. The most spectacular reductions in mass-adjusted REE were observed in "The Biggest Loser" nationally televised weight loss competition in which the participants who showed severe obesity (lost more than a-third of their body weight during a weight loss program consisting of diet restriction and vigorous exercise for 30 weeks. Despite a relative preservation of FFM, exercise did not prevent the dramatic reduction in mass-adjusted REE (~2 MJ day^{-1} on average) observed several months later.

A lower thermic effect of food (expressed as a % of meal energy) measured under resting conditions has also been reported after weight loss. Post-prandial thermogenesis, assessed during a test of oral glucose tolerance, has been reported to be reduced by more than two-thirds following weight loss in patients with obesity, while the thermic response to a mixed meal has been shown to be markedly lower after weight loss in subjects who were no longer obese compared to lean controls who never developed obesity.

Overall, these studies suggest that energy conservation through adaptive thermogenesis can occur both in the post-absorptive (BMR) and postprandial (thermic effect of food) components of EE; these being consistent with the reduced circulating levels of thyroid hormones and diminished sympathetic nervous system (SNS) activity that characterize energy sparing following weight loss.

Adaptive thermogenesis in the non-resting state

Adaptive thermogenesis can also occur in various sub-components of non-resting EE (Fig. 1). While it is well established that the efficiency of muscular contraction during dynamic exercise is low (~25%), that of spontaneous physical activity (SPA)—including fidgeting, muscle tone and posture maintenance, and other low-level physical activities of everyday life—is even lower since these essentially involuntary activities comprise a larger proportion of isometric work which is simply thermogenic. Since the actual work done on the environment during SPA is very small compared to the total energy spent on such activities, the energy cost associated with SPA has been referred to as movement-associated or SPA-associated thermogenesis. Because SPA is primarily a biologically-driven, involuntary (subconscious) behavior, a change in the level or amount of SPA in a direction that defends body weight may also constitute autoregulatory changes in EE. In this context, a decrease in SPA occurring following weight loss or an increase in SPA during overeating, also constitute adaptive changes in thermogenesis in the non-resting component of daily EE.

Variability in spontaneous physical activity

The potential importance of SPA-associated thermogenesis in human weight regulation has in fact been underscored by the findings that even under conditions where subjects are confined to a respiratory chamber, the 24 h EE attributed to SPA (as assessed by radar) was found to vary widely (between 0.4 and 3.3 MJ day^{-1}), and to be predictive of subsequent weight gain. This notion of a role of SPA in the etiology of obesity received considerable support by the findings that about two-thirds of the increase in total daily EE in experimentally overfed subjects could be attributed to non-exercise activity thermogenesis (NEAT), in which SPA is an important subcomponent (Fig. 1), and that NEAT was the most significant predictor of resistance to body fat gain. The more recent findings that overfeeding resulted in an increase in physical activity after adjusting for changes in weight and body composition, is consistent with the notion that an elevation in low-level physical activities and SPA might contribute to DIT, although there seems to be considerable inter-individual variability in this behavioral response to overfeeding. Conversely, evidence that diminished SPA contributes to adaptive thermogenesis following underfeeding can be derived from data obtained from the participants of the Biosphere 2 experiment, a self-sustaining ecologic "miniworld" and prototype planetary habitat built in Arizona several decades ago. Following an unexpected shortage of food, their losses in body weight (ranging between 8% and 25%) over a 2 year period were found to be accompanied by a markedly lower SPA than controls, which along with their reduced mass-adjusted 24 h EE (assessed in a respiratory chamber), was found to persist several months later after substantial weight recovery. In more recent years, diminished physical activity EE (including SPA), after adjustment for the change in body mass, has also been found to be reduced after several months of modest caloric restriction. Thus, in addition to suppressed thermogenesis in the resting component of EE, reduced SPA may contribute importantly to adaptive reductions in thermogenesis following caloric restriction.

Efficiency of muscle work

Although changes in SPA can constitute an adaptive response to under- or over-nutrition, significant alterations in SPA have not always been found to be a main component of adaptive changes in non-resting EE. Indeed, in the "weight-clamping" experiment in which the participants maintained their body weight at 10% above or below the habitual level, the reported adaptive thermogenesis which was found to occur primarily in the non-resting component of EE (estimated by the difference between total daily EE and REE), could not be explained by changes in the amount of time spent in physical activity assessed by accelerometry. Instead, changes in the mechanical contraction efficiency of skeletal muscle following weight loss or weight gain have been shown to account for about a-third of the change in daily EE associated with physical activity. The notion that weight loss leads to increased muscle contraction efficiency is consistent with other reports of a lower (net) energy cost of walking or cycling exercise after weight reduction in subjects with or without obesity, in female dieters exercising at different workloads or in chronically undernourished subjects—all of which would contribute to further reduce energy needs after weight loss. They are also consistent with diminished circulating leptin, thyroid hormones and SNS activity that characterize energy sparing following weight loss.

Interactions and compensations across components of EE

It should be underlined that the separation of adaptive thermogenesis into resting and non-resting states is arbitrary, and that potential interactions across these two components of EE cannot be disregarded (Fig. 1). For example, sleep EE, which is generally classified as EE at rest, also comprises a non-resting component due to spontaneous movements (i.e., SPA) that occur during sleep, the frequency of which appears to be highly variable between individuals. In addition, non-resting EE and NEAT could also include EE resulting from the impact of physical activity (exercise or SPA) on postabsorptive or postprandial EE. There is indeed some evidence that relatively low-intensity exercise can potentiate the thermic effect of food and that the effect of physical activity on EE can persist well after the period of the physical activity. There is also considerable interest in the phenomenon of exercise-induced "energy compensation" whereby the increases in EE that occur during moderate-to-intense exercise activities may not necessarily all translate into an increase in daily EE because of at least partial compensatory reductions of EE in NEAT or when at rest over subsequent day(s). Diminished post-exercise stimulation of EE has also been proposed as a mechanism for energy conservation in individuals who are considered to be chronically-energy deficient since childhood.

Overall, any change in metabolic efficiency in resting or non-resting state that would tend to *attenuate energy imbalance* or to *restore body weight and body composition* toward its "set" or "preferred" value constitute adaptive changes in thermogenesis.

Autoregulation of body weight and body composition

Based upon a series of re-analyses of classic longitudinal studies of human starvation, refeeding and overfeeding, the available evidence suggests that the mechanisms of energy adaptation for optimal survival in an environment of famine-and-feast are embodied in three distinct autoregulatory control systems. These comprise an intrinsic (internal) control of partitioning between protein and fat (i.e., the control of lean-fat partitioning) and two distinct control systems underlying adaptive thermogenesis—the latter being depicted in the conceptual diagram shown in Fig. 2. One control system is a direct function of changes in dietary energy intake and responds relatively rapidly to energy deficit. Its effector mechanisms are suppressed early during starvation, restored rapidly as a function of energy re-availability, and are further activated if hyperphagia occurs during refeeding. Since the efferent limb of this control system—which is primarily under the control of the SNS—is dictated not only by the dietary energy intake but also by a variety of other environmental factors such as diet composition, specific nutrient deficiencies, ambient temperature,

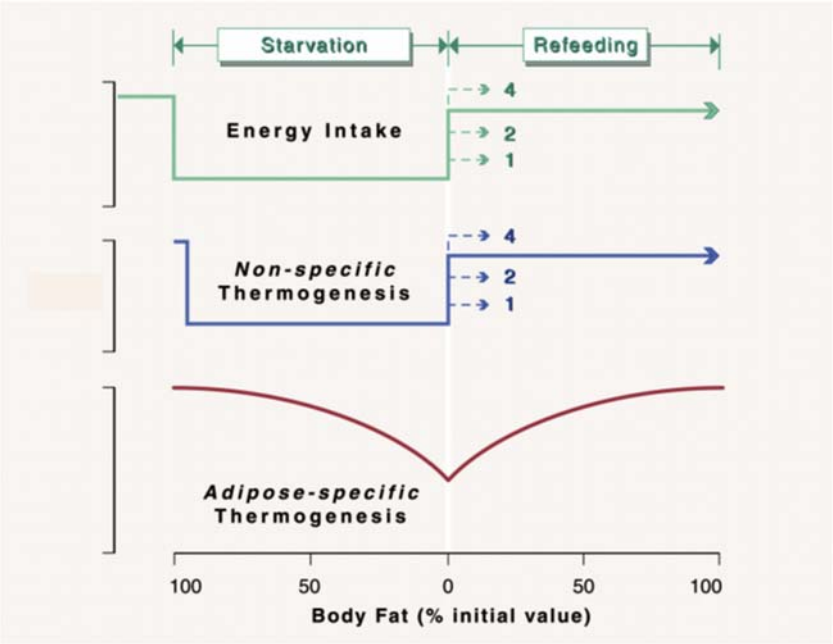


Fig. 2 Schematic representation of the concept of two distinct control systems underlying adaptive thermogenesis during prolonged starvation and subsequent refeeding. One control system, which is a direct function of changes in the food energy supply, responds relatively rapidly to the energy deficit. Its effector mechanisms are suppressed early during the course of starvation, and upon refeeding they are restored relatively rapidly as a function of energy re-availability (levels 1–4), and are activated further if hyperphagia occurs during refeeding (level 4). Because, (as depicted in **Fig. 3**) the efferent limb of this control system is primarily under the control of the sympathetic nervous system (SNS) whose functional state is dictated by overlapping or interacting signals arising from a variety of environmental stresses including food deprivation, deficiency of essential nutrients, excess energy intake and exposure to cold or to infections, it is referred to as the *non-specific* control of thermogenesis; it is likely to occur primarily in organs/tissues with a high specific metabolic rate (e.g., liver, kidneys, brown adipose tissue). The other control system, by contrast, is independent of the functional state of the SNS, has a much slower time-constant by virtue of its response *only* to putative signals (denoted by symbol "?" in **Fig. 3**) arising from the state of depletion/repletion of the fat reserves; it is therefore referred to as the control system operating through an *adipose-specific* control of thermogenesis. While suppression of this *adipose-specific* thermogenesis during starvation and during refeeding leads to energy conservation, the energy thus spared during refeeding is directed specifically at the replenishment of the fat reserves, resulting in an accelerated fat recovery—a phenomenon that could contribute to the disproportionately rapid rate of fat relative to lean tissue recovery during refeeding after substantial fat reserves depletion. Adapted from Dulloo, A.G., Jacquet, J., 2001. An adipose-specific control of thermogenesis in body weight regulation. Int. J. Obes. 25 (Suppl. 5), S22–S29.

psychological stress, etc., it has been referred to as the *non-specific* control of thermogenesis. By contrast, the other control system has a much slower time-constant as it responds to signals arising from the state of depletion/repletion of body fat reserves; it has therefore been referred to as the control system operating through an *adipose-specific* control of thermogenesis. Thus, these control systems are distinct in their responses to signals reflecting changes in energy balance and the fat reserves.

A compartmental model

An integration of these control systems in the autoregulation of body weight and body composition during a cycle of weight loss and weight recovery is discussed using a schematic diagram presented in **Fig. 3**. This diagram embodies the findings that the intrinsic control of body energy-partitioning between protein and fat (i.e., lean-fat partitioning) is an individual characteristic and that the intrinsic lean-fat partitioning characteristic (*Pc*) of the individual during weight loss is relatively conserved during weight recovery. It also takes into account the two distinct control systems of adaptive thermogenesis which can function independently of each other.

During starvation, the intrinsic control of partitioning determines the relative proportion of protein and fat to be mobilized from the body as fuel—the individual's *Pc* being, to a large extent, dictated by the initial body composition (i.e., initial body fat%). The functional role of the control of lean-fat partitioning is to meet the individual's fuel needs in such a way that the energy-reserve component in both the fat and protein compartments (i.e., the part that can be lost without causing death or irreversible damage) would reach complete depletion simultaneously. Such a strategy would ensure the maximum duration of survival for a given individual during long-term food shortage. Furthermore, the energy conservation resulting from suppressed thermogenesis is directed at reducing the energy imbalance, with the net result that there is a slowing in the rate of protein and fat mobilization in the same proportion as defined by the *Pc* of the individual. Indeed, the demonstration in lean individuals that the proportion of fuel energy derived from protein (i.e., the P ratio) remains relatively constant during the course of prolonged fasting implies that suppressed thermogenesis is not directed at sparing specifically protein or specifically fat, but at sparing both protein and fat compartments

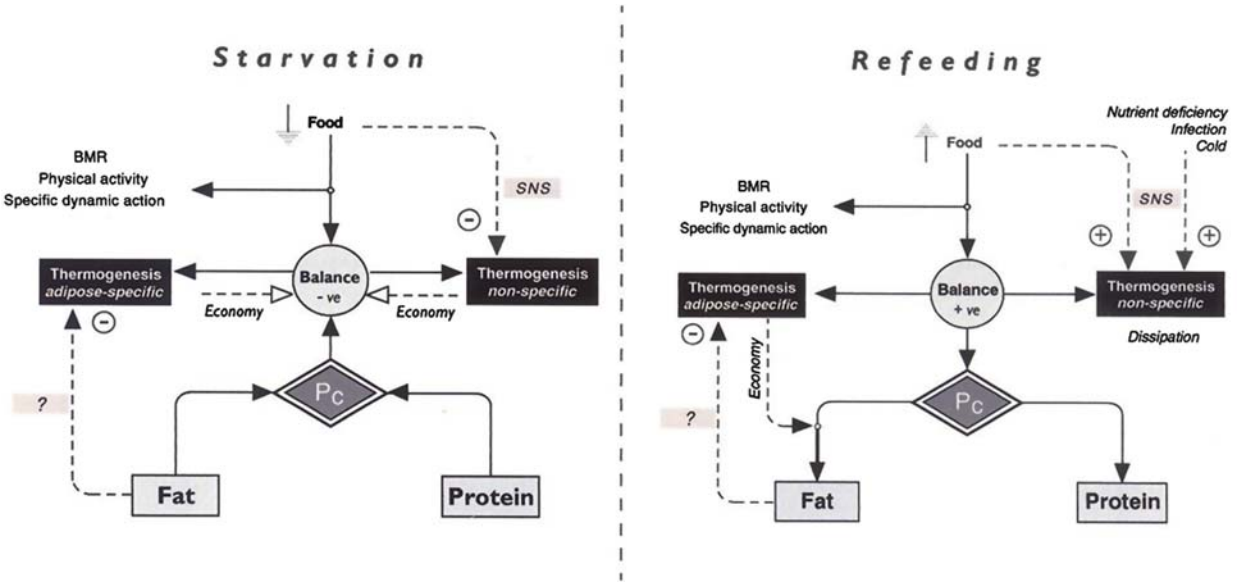


Fig. 3 Schematic representation of a compartmental model for the regulation of body weight and body composition during a cycle of weight loss (prolonged starvation) and weight recovery (refeeding). In this model, the two distinct control systems underlying adaptive thermogenesis—the non-specific control and the adipose-specific control—are integrated with the more "basal" control of lean-fat partitioning between the body fat and protein compartments as determined by the partitioning characteristic (*Pc*) of the individual. During weight loss (starvation), the energy economy resulting from a reduction in thermogenesis is directed in such a way as to spare (non-specifically) both the lean and fat compartments as defined by the *Pc* of the individual. In contrast, during weight recovery, the energy economy resulting from a reduction in thermogenesis is specifically directed at replenishing the fat compartment, with the net result that fat is deposited in excess of that determined by the *Pc* of the individual during weight recovery. The existence of these two distinct forms of adaptive thermogenesis provides an explanation for the co-existence of a hypermetabolic state (e.g., in response to infections, nutrient deficiencies, cold exposure) and the suppression of thermogenesis specific for accelerating fat replenishment. SNS = sympathetic nervous system; ? = unknown signal(s) (see legend of **Fig. 2**). Adapted from Dulloo, A.G., Jacquet, J., 2001. An adipose-specific control of thermogenesis in body weight regulation. Int. J. Obes. 25 (Suppl. 5), S22–S29.

simultaneously. During the course of prolonged starvation therefore, the functional role of the two control systems underlying suppressed thermogenesis is to reduce the overall rate of fuel utilization, i.e., to conserve energy that would reduce the rate at which both lean and fat energy reserves are being depleted.

During refeeding, the control of lean-fat partitioning operates in such a way that protein and fat are deposited in the same relative proportion as determined by the individual's *Pc* during starvation, thereby restoring the individual's pre-starvation surviving capacity. In addition, the increased availability of food energy would lead to rapid withdrawal of the suppressed *non-specific* (SNS-mediated) control of thermogenesis. In contrast, the suppression of thermogenesis under *adipose-specific* control is slowly withdrawn as a function of fat recovery, so that the energy that continues to be spared is directed specifically toward replenishing the body's fat reserves. The net effect is that fat is deposited in excess of that determined by the individual's *Pc*, thereby contributing to the disproportionate rate of fat recovery compared to that of lean tissue. This phenomenon of preferential catch-up fat has often been observed during nutritional rehabilitation in adults as well as during catch-up growth in infants or children. A role for thrifty (energy conservation) mechanisms underlying this catch-up fat phenotype has been proposed as a key link between fetal or neonatal growth retardation, catch-up growth and risks for developing type 2 diabetes and cardiovascular diseases later in life.

Biological significance of a dual control system for adaptive thermogenesis

Such an adaptive phenomenon that accelerates the restitution of the fat reserves rather than diverting the saved energy toward a compensatory increase in body protein synthesis (an energetically costly process) would have survival value in an ancestral lifestyle of periodic food scarcity. This is because, by virtue of the fact that body fat has a greater energy density and a lower energy cost of synthesis/maintenance than protein, it would provide the organism with a greater capacity to rapidly rebuild an efficient energy reserve, and hence to cope with recurrent food shortages.

Thus, the functional role of the *adipose-specific* control of thermogenesis during weight recovery is to accelerate specifically the replenishment of the fat stores when food availability improves following severe food deficit and marked depletion of body fat reserves. It provides an alternative mechanism to accelerate the recovery of survival capacity in the absence of food abundance and hyperphagia. But equally important for mammalian survival during weight loss and weight recovery is the need to retain the ability to increase heat production (i.e., to activate thermogenesis) in response to a number of other environmental stresses namely: (i) for increased thermoregulatory needs in cold environments, (ii) for generating fever upon exposure to infections or (iii) for increased heat production as an adaptation to nutrient-deficient diets.

Indeed, a re-analysis of human data on experimental overfeeding studies conducted between 1965 and 1999 demonstrated that subjects who were overfed diets almost deficient in protein (<3% of energy intake) showed the highest energy cost of weight gain, and were hence most likely exhibiting a substantial level of DIT. From an evolutionary perspective, the capacity to increase DIT in the face of nutrient-deficient diets probably had survival advantage of "homeostatic waste" because it allows individuals to overeat relatively large amounts of poor quality food in order to obtain essential nutrients without depositing excess, non-essential energy in the form of fat. Excessive weight gain is thought to interfere with optimal locomotion, hunting capabilities and the ability to perform the "fight or flight" reaction. It has in fact been proposed that DIT may have evolved as a means of regulating the metabolic supply of essential nutrients (protein, minerals, vitamins) with only a secondary role in regulating energy balance and body weight.

Whatever the exact functional significance of DIT, however, it is clear that in a context of weight recovery, an increased efficiency for catch-up fat can be shown to persist even under conditions of hypermetabolism (a net increase in thermogenesis) induced by hyperphagia or nutrient-deficient diets. To explain this apparent paradox, the model presented in **Fig. 3** provides a structural framework that illustrates how suppressed *adipose-specific* thermogenesis which drives the acceleration of fat deposition during refeeding (with the energy sparing postulated to occur in the skeletal muscle) persists under conditions when the *non-specific* control of thermogenesis is activated in organs/tissues recruited by the SNS (liver, kidneys, heart, brown adipose tissue). Such differentially-regulated control systems for adaptive thermogenesis may thus have arisen during the course of mammalian evolution as dual-adaptive processes that can satisfy the need for energy conservation during weight loss, or for catch-up fat during weight recovery, even under conditions of environmental stresses when activation of heat production by the SNS has equally important survival values. This concept of dual-control system for adaptive thermogenesis has received considerable support from studies in rodent models of refeeding after caloric restriction, and is also supported by some studies in humans (see section below). Furthermore, using animal models of semistarvation and controlled refeeding, considerable progress has been made in elucidating the postulated role of diminished skeletal muscle thermogenesis in the effector mechanisms of the "adipose-specific" control of thermogenesis. These include reductions in the rate of muscle protein turnover and other energetically costly substrate cycles, reductions in subsarcolemmal mitochondria, diminished ratio of white to red muscle fibers and slowed muscle contraction/relaxation cycle—all of which can be related to a state of skeletal muscle hypothyroidism resulting from altered activity of muscle deiodinase enzymes. Still to be unraveled, however, are the sensors and signals of the feedback loop between adipose tissue fat depletion and suppressed skeletal muscle thermogenesis.

Energy adaptation during a longitudinal human study of large weight fluctuations

The existence and functions of this dual-control systems for adaptive thermogenesis in humans are consistent with the temporal changes in BMR and body composition during the longitudinal study of semistarvation, refeeding and subsequent overeating in men participating in the Minnesota Starvation Experiment—the most tightly controlled domiciled feeding study ever conducted in humans. The pattern of changes in food intake and body weight, together with kinetics of altered thermogenesis (BMR adjusted for FFM and fat mass, and expressed as a percentage of baseline BMR) are presented in **Fig. 4**; the data on altered thermogenesis being derived from a re-analysis of data from this experiment.

- First, during the phase of weight loss, the operation of the two control systems for adaptive thermogenesis is suggested by the fact that reduction in thermogenesis is *biphasic* in nature, with an initial rapid reduction in adjusted BMR at week 4, corresponding to 10% of baseline BMR, followed by a slower reduction in adjusted BMR, corresponding to 20% and 25% of baseline BMR at week 12 and 24, respectively. At these latter time-points during semistarvation (at S12 and S24), the magnitude of reductions in adjusted BMR was found to be associated with the reduction in fat mass—i.e., the greater the degree of depletion of the fat reserves, the greater the suppression of thermogenesis.
- Second, during the phase of weight recovery, the functioning of the two control systems for thermogenesis is also suggested by the following:
  (a) the relationship between the degree of depletion of fat reserves and suppressed (*adipose-specific*) thermogenesis persisting at week 12 of restricted refeeding, at which time-point (R12) the mean adjusted BMR was still about 10% below baseline BMR level, the body fat was 80% recovered, while the recoveries of body weight and FFM were less than 50%, and
  (b) during the subsequent period of *ad libitum* refeeding, the development of hyperphagia was accompanied by a rapid (*non-specific*) increase in thermogenesis, as evidenced by increases in adjusted BMR corresponding to about 20% of baseline BMR at week 14 of refeeding.
- Third, during the phase of weight overshot, it is noticeable that by week 20 after the onset of refeeding (at R20), when FFM has been recovered to almost 100% and body fat has exceeded the baseline (pre-starvation) level by more than 50%, the adjusted BMR remained significantly higher (by about 10% on average) above baseline (pre-starvation) values despite the fact that hyperphagia was no longer present and FFM was normalized. This sustained rise in thermogenesis after post-starvation overfeeding is consistent with a feedback mechanism existing between thermogenesis and body fat, and may well have contributed to the subsequent slow return of the "overshot" body weight toward the baseline (pre-starvation) level.

It should be pointed out that the data available for the re-analysis of the Minnesota Starvation Experiment only allowed the analysis of adaptive thermogenesis in the BMR component of daily EE, and that no measurements were performed pertaining to the thermic effect of food or to the energy cost of physical activity. Nonetheless, the authors observed that there was a profound decrease in

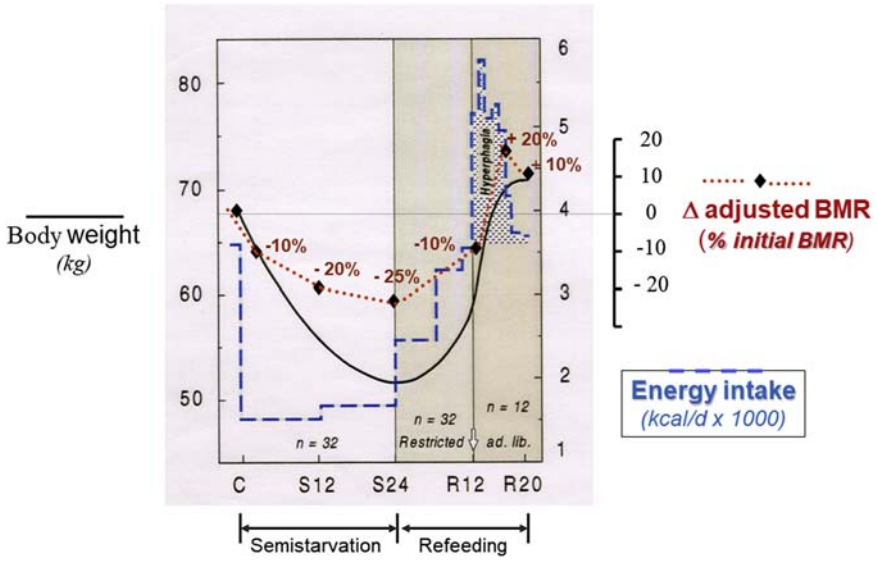


Fig. 4 Pattern of changes in body weight, food intake, and adaptive thermogenesis during the various phases of the longitudinal Minnesota Experiment of human semistarvation and refeeding. The changes in adaptive thermogenesis at the various time-points are assessed as changes in basal metabolic rate (BMR) after adjusting for changes in fat-free mass (FFM) and fat mass, and expressed as a percentage of the baseline (control, C) level. C = end of control (baseline) period; S12 and S24 = week 12 and week 24 of semistarvation, respectively; R12 and R20 = week 12 and week 20 after onset of refeeding. Drawn from data of: (i) Keys, A., Brozek, J., Henschel, A., Mickelson, O., Taylor, H.L., 1950. The Biology of Human Starvation. Minneapolis: University of Minnesota Press, 1950; (ii) Dulloo, A.G., Jacquet, J., 1998. Adaptive reduction in basal metabolic rate in response to food deprivation in humans: a role for feedback signals from fat stores. Am. J. Clin. Nutr. 68, 599–606.

physical activity of the subjects, especially during second 12 week period (S12-S24) of semistarvation, thereby suggesting that adjustments in EE occurred importantly in both resting and non-resting EE. Indeed, further evidence in humans in support of the adipose-specific suppression of thermogenesis driving catch-up fat can be derived from:

(i) the demonstration of an association between mass-adjusted REE and the recovery of fat mass during nutritional rehabilitation of patients recovering from non-neoplastic gastrointestinal disease, and
(ii) the findings of persistently low mass-adjusted 24 h EE (including in sleeping EE and SPA) during weight recovery that was primarily as body fat after 2 years of sustained caloric restriction in the above-mentioned Biosphere 2 Experiment.

Energy adaptations and susceptibility to leanness and fatness

In addressing the issue of energy adaptation in human susceptibility to leanness and fatness, it must be emphasized that even in individuals who appear to maintain a relatively stable body weight over decades, there is no absolute constancy of body weight over days, weeks and months. Instead, body weight tends to fluctuate or oscillate around a mean constant value, with deviations from a "set" or "preferred" value being triggered by events that are cultural (week-end parties, holiday seasons), psychological (stress, anxiety or emotions), intentional (multiple attempts to lose weight but resulting in repeated weight regain, i.e., weight cycling), and pathophysiological (ranging from minor health disturbances to more serious disease states). Within such a dynamic state in which weight homeostasis occurs, it is likely that long-term constancy of body weight is achieved through a network of regulatory systems and subsystems through which autoregulatory changes in food intake, body composition and EE are interlinked.

The above-described autoregulatory control systems—operating through the intrinsic control of lean-fat partitioning and through the two distinct control systems underlying thermogenesis—can play a crucial role in attenuating and correcting deviations of body weight from its "set" or "preferred" value. The extent to which these adjustments are made depends on the environment (e.g., diet composition), and varies greatly between individuals, largely due to the individual's previous nutritional status and genetic variations. They probably conferred different abilities to defend the body's lean mass and fat reserves in an ancestral hunter-gatherer lifestyle of famine-and-feast, but now underlie our different metabolic susceptibilities to fatness in societies where palpable foods are abundant throughout the year. The resultant subtle variations between individuals in lean-fat partitioning and in adaptive thermogenesis can, over the long-term, be important in determining constancy of body weight in some and in provoking drift toward obesity in others. This may have particular relevance in explaining the findings of numerous prospective studies which suggest that repeated dieting to lose weight and subsequent weight regain (weight cycling) is a risk factor for later obesity. Indeed, using a mathematical model of weight cycling that integrates these autoregulatory control systems regulating body composition—including those underlying the preferential catch-up fat phenomenon leading to fat overshooting—it could be predicted that in lean individuals, the cumulative amount of fat overshoot over several weight cycles would ultimately result in obesity.

Furthermore, with particular relevance for obesity management, the adaptive responses to energy deficits, discussed so far, primarily in the context of weight recovery after experimental starvation in normal-weight (lean) individuals, also occur in individuals in whom obesity has developed spontaneously, and contribute the defense of the obese state once acquired. Indeed, several studies conducted in subjects with obesity have shown that mass-adjusted reductions in REE and/or in non-resting EE persist in the stable weight loss maintenance phase (i.e., in the weight reduced state) for months to years after weight loss. For example, in a follow-up study of the "Biggest loser" competition, the reduced mass-adjusted REE (2 MJ day^{-1} on average) after losing about 40% of their weight during the competition, persisted some 6 years later despite having regained more than two-thirds of their lost weight. Taken together, these studies suggest that people with obesity respond to weight loss therapy by conserving energy through adaptive thermogenesis (in the resting and/or non-resting components of EE) well beyond the dynamic phase of weight loss and into the weight-reduced state and during weight regain.

These results support the notion that suppressed thermogenesis in response to food deprivation is a factor that reduces the efficacy of therapeutic regimens, and contributes to the relapse of obesity. In addition, since the initial body composition (% body fat) is the most important determinant of energy partitioning between lean and fat tissue (i.e., *Pc* of the individual) during weight loss and weight recovery, the higher %body fat (i.e., the more obese the individual), the lower the fraction of energy mobilized as protein, and hence the greater the propensity to mobilize fat during weight loss and to subsequently deposit fat during recovery. The low *Pc* of the individual with obesity, coupled with sustained (adipose-specific) suppression of thermogenesis in response to their relative state of body fat depletion, will contribute to the relapse of obesity.

Concluding remarks

Given the famine-and-feast lifestyle that has characterized much of human evolutionary history, it is conceivable that what constitutes normal physiology of weight regulation evolved within the constraints of large fluctuations in body weight and body composition. In this context, the autoregulatory control systems that regulate body composition must have evolved to optimize not only the survival capacity of the individual during prolonged periods of food scarcity, but also the recovery of this survival capacity whenever food availability improves (in readiness for the next period of food scarcity), and in some situations to exhibit DIT in the face of abundance of nutrient-deficient diets to avoid excessive weight gain. These multiple control systems therefore constitute the "energy adaptations" which nowadays manifest themselves to varying degrees in proneness to obesity and in undermining obesity therapy and disease-cachexia rehabilitation.

See Also: Body composition; Energy balance; Energy metabolism; Starvation and fasting: Biochemical aspects

Further reading

Bray, G.A., Bouchard, C., 2020. The biology of human overfeeding: a systematic review. Obes. Rev. 21, e13040. https://doi.org/10.1111/obr.13040.

Dulloo, A.G., Jacquet, J., Seydoux, J., Montani, J.P., 2006. The thrifty "catch-up fat" phenotype: its impact on insulin sensitivity during growth trajectories to obesity and metabolic syndrome. Int. J. Obes. 30 (Suppl. 4), S23–S35. https://doi.org/10.1038/sj.ijo.0803516.

Dulloo, A.G., Jacquet, J., Montani, J.P., Schutz, Y., 2012. Adaptive thermogenesis in human body weight regulation: more of a concept than a measurable entity? Obes. Rev. 13 (Suppl. 2), 105–121. https://doi.org/10.1111/j.1467-789X.2012.01041.x.

Dulloo, A.G., 2021. Physiology of weight regain: lessons from the classic Minnesota Starvation Experiment on human body composition regulation. Obes. Rev. 22 (Suppl. 2), e13189. https://doi.org/10.1111/obr.13189. PMID: 33543573.

Ravussin, E., Smith, S.R., Ferrante Jr., A.W., 2021. Physiology of energy expenditure in the weight-reduced state. Obesity 29 (Suppl. 1), S31–S38. https://doi.org/10.1002/oby.23095.

Stock, M.J., 1999. Gluttony and thermogenesis revisited. Int. J. Obes. 23, 1105–1117. https://doi.org/10.1038/sj.ijo.0801108.

Stubbs, R.J., Turicchi, J., 2021. From famine to therapeutic weight loss: hunger, psychological responses, and energy balance-related behaviors. Obes. Rev. 22 (Suppl. 2), e13191. https://doi.org/10.1111/obr.13191.

Wells, J.C., 2006. The evolution of human fatness and susceptibility to obesity: an ethological approach. Biol. Rev. 81, 183–205. https://doi.org/10.1017/S1464793105006974.

Hunger

K. Beaulieu and J.C.G. Halford, School of Psychology, University of Leeds, Leeds, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of J.C.G. Halford, E.J. Boyland, Hunger, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 431–435, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00145-8.

Introduction 137
The measurement of hunger 138
Hunger and satiety 138
Hunger is driven by energetic needs 139
Other physiological determinants of hunger 139
Learning and hunger 140
Hunger and eating behavior 141
Disorders of hunger 141
Conclusion 142
References 142
Further reading 142

Key points

- While perceived hunger is the drive to consume in response to a biological need, several other influences can impact how much and when a person eats.
- Within the homeostatic appetite system, excitatory and inhibitory signals shape the overall expression of hunger and satiety.
- Hunger is a key component of the relationship between energy expenditure and energy intake.

Glossary

Food reward or food hedonics Thoughts about food and sensory appreciation of certain food attributes that determine food preference and choice. Comprises of two distinct components: liking and wanting for food. Liking is the sensory pleasure elicited by contact with food and wanting is the motivation to consume a specific food, manifesting explicitly (craving/desire) or implicitly (food cue responsiveness)

Hunger The drive to consume, a subjective experience that stimulates and sustains a behavioral event, eating. It has a strong learned component and can be triggered by physiological need or environmental stimuli. Hunger usually predicts subsequent eating behavior, and is suppressed by that ingestion, an effect that depends on the sensory, cognitive, gastric, gastrointestinal, and metabolic consequences of ingestion and learned associations with specific foods

Satiation The within-meal processes that bring eating episodes to an end, predominantly driven by cognitive factors and learned associations, the physical and sensory characteristics of the meal, the bulk of that meal in the stomach, and the rate of gastric emptying

Satiety The suppression of hunger and eating behavior after a meal. Often referred to as a state sustained by numerous post-ingestive and post-absorptive mechanisms, triggered as food is digested and nutrients are absorbed and metabolized

Introduction

The term "hunger" is used in more than one sense by both scientists and the public. One traditional use refers to insufficient food availability, as it may occur in poor regions of the world. Most experts, however, prefer to use the term "food insecurity" to refer to situations of food scarcity. In the context of this article, hunger describes the drive or the motivational force that urges us to seek and consume food. It is the expression of a biological need to sustain growth and life. Hunger is therefore a purposeful experience that possesses a clear biological function (Blundell, 1979). Therefore, in traditional motivation theory, hunger is a biological drive to satisfy a physical need. In true homeostatic fashion, the drive to consume reduces and that need is satisfied (sated).

There are two components to hunger within nutritional science. Hunger is inferred from directly observable and measurable events. In this way, by inferring increased or high levels of hunger from a long period of food deprivation or an increased willingness to expend effort in order to obtain food, hunger becomes a mediating concept or an intervening variable. However, linked to these

directly observable behavioral events are a collection of conscious feelings or sensations that are linked to a desire to obtain and eat food. This is the sense in which lay people understand the term hunger. It is through these objectively observable behavioral events and the vigorous measurement of these subjective feelings, by means of rating scales and other measurement devices, that researchers attempt to capture changes in hunger motivation.

Early investigators used questionnaires which asked people to note the presence of physical sensations in a number of bodily areas, together with moods, urges to eat, and preoccupation with thoughts of food and of the everyday experience of hunger (Monello and Mayer, 1967; Friedman et al., 1999). It was found that the observation, "I feel hungry," was typically based on the perception of bodily feelings, which at times are very strong. Gastric sensations, a hollow feeling or stomach rumbling, are frequent indicators of hunger, although people also report sensations in the mouth, throat, and head. These accompany more diffuse feelings of restlessness and excitability as well as an urge to eat. The consumption of food changes both the pattern of physical sensations and the accompanying emotional feelings, with unpleasant and aversive sensations becoming replaced by more pleasant ones. Thus, as classic drive reduction theory would suggest, hunger is associated with unpleasant feelings, and food consumption eases these. For example, an aching stomach becomes relaxed, and the feeling of excitement and irritability is replaced by one of contentment.

Despite this commonality, subsequent study has shown a great deal of variability both within and between individuals. In other words, hunger demands neither the consistent presence of sensations before every act of eating nor that these sensations occur in every person sitting down to eat. However, even with this variability, people are able to, and frequently do, make judgments regarding their state of hunger, partly through reference to these sensations.

The measurement of hunger

The process of measuring hunger is not straightforward. There is an inherent mistrust of subjective reports of appetite. Critics point to the variability in response between individuals and the absence of any objective "standard" by which internal experience can be calibrated. This has driven many to use biological proxies (the release or inhibition of particular hormones or activation of certain brain areas) as "biomarkers" for appetite in the hope that these may be more reliable indicators of behavior. However, hunger must be seen in a wider individual and situation context and cannot rest in the fluctuations of any single biological system or biomarker. The issues of "validity" and "reliability" are more complex than criticism suggests.

The two most common methods for quantifying hunger are fixed-point rating scales and visual analog scales. Fixed-point scales are quick and simple to use, and the data they provide are easy to analyze. Past examples of these scales show that they vary considerably in complexity. In considering the appropriate number of points to be included in this type of scale, the freedom to make a range of possible responses must be balanced against the precision and reliability of the device. Research seems to indicate that scales with an insufficient number of fixed points can be insensitive to subtle changes in subjective experience. In addition, the fixed points are important determinants of the way in which people use the scales and distribute their ratings. One way of overcoming some of these failings is to abolish such points. Thus, visual analog scales are horizontal lines; these lines should be at least 100 mm long (or longer to allow the freedom to make a range of responses), unbroken and unmarked, except for word anchors at each end (Flint et al., 2000). The user of the scale is instructed to mark the line at the point that most accurately reflects the intensity of the subjective feeling at that time. The researcher measures the distance to that mark in millimeters from the negative end (no hunger), thus yielding a score of 0–100 (in the case of 100 mm lines). This is done either by hand or automatically if presented by a computer screen (the latter method ensures that they are completed at designated times during a study day). By doing away with all the verbal labels except the end definitions, visual analog scales retain the advantages of fixed-point scales, while avoiding many of the problems with uneven response distributions.

An important aspect of these methods concerns the interpretation of differences between the fixed points or intervals on a visual analog scale. For example, it cannot be assumed that the difference between 20 and 30 mm on a hunger scale is perceptually the same as the difference between 80 and 90 mm. Nor can a hunger rating of 80 mm be said to represent a feeling of hunger that is twice the intensity of that rated at 40 mm. The variation of scores on these scales represents the fluctuations in the perception of feeling and not absolute change of some physical commodity. Individuals will use these scales idiosyncratically (each individual may interpret the scale and gauge their response differently) but generally consistently on repeated occasions (they will tend to scale any changes in response within their own parameters, producing a reliable record of alterations in an individual's appetite). One example of this is the problem of "end effects." This refers to the reluctance of a minority of subjects to make ratings away from the upper or the lower end points of the scale, despite clear instructions. Despite these limitations, data from such scales are often analyzed using parametric statistical procedures, such as analysis of variance, and in general, this appears to be a satisfactory approach.

Hunger and satiety

If hunger is that feeling that drives us to seek food and to consume, then eating eventually relieves hunger, albeit until the next snack or meal. The capacity of a food to reduce the experience of hunger can be termed "satiating power" or "satiating efficiency" (Kissileff, 1984) and can be quantified by the satiety quotient (Green et al., 1997). This power is the product of the body's handling of the

nutritional composition and structure of the food eaten. It follows that some foods will have a greater capacity to maintain suppression over hunger than other foods. This will depend on their sensory impact, physical characteristics, energy density, and macronutrient composition (Holt et al., 1995; Kirkmeyer and Mattes, 2000).

The distinction between hunger and satiety is both conceptual and technical. As hunger diminishes, satiety rises. However, it is useful to further separate those events that occur across the course of a meal (prandial) from those between meals (pre- and postprandial). In this way, the prandial process of satiation can be clearly distinguished from the postprandial state of satiety. Satiation can be regarded as the process that develops during eating and that eventually brings a period of eating to an end. Accordingly, satiation can be defined in terms of the measured size of an eating episode (such as its energy, weight, or volume). Hunger declines as satiation develops, and usually reaches its lowest point at the end of a meal. Satiety is defined as the state of inhibition over further eating that follows at the end of a meal and that arises from the consequences of food ingestion. The intensity of satiety can be measured by the duration of time until eating starts once more or by the amount consumed at the next meal. The strength of satiety is also measured by the time that hunger is suppressed, but as satiety subsides, hunger is restored.

In examining the mechanisms responsible for suppressing hunger, it is clear that they range from those that occur when food is initially sensed to the effects of metabolites on body tissues following the digestion and absorption of food (across the wall of the intestine and into the bloodstream). By definition, satiety is not an instantaneous event but occurs over a considerable time period. Sensory effects are generated through the smell, taste, temperature, and texture of food, and it is likely that these factors have effects on eating in the very short term. Cognitive influences represent the beliefs held about the properties of foods, and these factors may also help inhibit hunger in the short term. These will start to affect hunger even before consumption has commenced and continue to suppress it over the course of the meal.

Post-ingestive processes such as gastric distension and rate of emptying, the release of hormones such as cholecystokinin, peptide-YY, glucagon-like peptide-1, and the stimulation of certain receptors along the gastrointestinal tract increasingly suppress hunger and, as the meal terminates, further inhibit its rebound during early post-meal satiety. The postabsorptive phase of satiety comprises those mechanisms arising from the action of metabolites after absorption into the bloodstream. These include the action of glucose and amino acids, which act directly on the brain after crossing the blood–brain barrier and which influence the brain indirectly via neural inputs following stimulation of peripheral chemoreceptors. These will continue to suppress hunger late into the postprandial period.

Hunger is driven by energetic needs

As that feeling that drives us to seek food, it makes sense from an evolutionary perspective that hunger should arise from the need to replace the energy used by the body. Our body's basic energy needs are reflected by the resting metabolic rate, largely influenced by lean or fat-free mass, which includes all high-metabolic-rate organs of the body such as the heart, liver, kidneys, brain, muscles, etc. Studies have shown that fat-free mass and resting metabolic rate are associated with hunger ratings and energy intake (Blundell et al., 2020). Being physically active and performing essential daily behavioral tasks also contribute to a person's energy needs and constitute another drive for energy from food (albeit to a smaller degree than resting metabolic rate). Thus, both metabolism and behavior are involved in generating hunger as a functional response to energy needs. It is therefore of interest to examine how habitual physical activity affects hunger and satiety. This is different from the effect of acute exercise on hunger: high-intensity exercise has been shown to suppress hunger, but this effect is transient (<1 h) and does not appear to subsequently affect food intake. This will not be discussed further. Rather, habitual physical activity level, i.e., the degree to which someone performs physical activity over a long period of time, has been shown to be non-linearly related to energy intake in a J-shaped relationship (Beaulieu et al., 2018). This means that in people with higher levels of physical activity, daily energy expenditure and energy intake are closely matched, but in those with lower levels of physical activity, this coupling is lost, such that daily energy intake exceeds expenditure. Additionally, lower levels of physical activity appear to be accompanied by higher levels of adiposity. This J-shaped relationship between physical activity level (daily energy expenditure) and energy intake suggests that higher levels of physical activity are associated with more sensitive appetite control. This better matching between intake and expenditure could be due to enhanced satiety signaling. In contrast, lower levels of physical activity (being inactive/highly sedentary) are associated with dysregulated appetite and a reduction of energy expenditure which does not downregulate hunger and food intake. Consequently, this results in overconsumption relative to energy requirements and weight gain. This overconsumption may be related to greater hedonic and psychocognitive inputs increasing the susceptibility to overconsumption at lower levels of physical activity. The effect of physical activity on hunger and satiety can be conceptualized by a "dual-process action" on appetite whereby an increase in fasting hunger with greater levels of physical activity occurs in conjunction with an enhancement in satiety signaling (King et al., 2009).

Other physiological determinants of hunger

According to Rogers (1999), even before food reaches the mouth, potent physiological signals are generated by the mere sight or smell of food, or even learned contextual cues such as location and time of day. These signals, produced in response to exposure to external stimuli, comprise the "cephalic phase" of appetite. This cephalic response is expressed in several parts of the gastrointestinal tract and acts to anticipate food ingestion. Later, stomach distension and the detection of macronutrients such as fat or protein

within the gut are all powerful satiety cues. They bring a meal to an end and, for a time, inhibit further consumption. Eventually, hunger again prevails and food intake follows. The flux between hunger and satiety is episodic and underpins the expression of our eating behavior throughout the day. However, it is not just the absence of episodic satiety cues (e.g., stomach distension and intestinal or absorbed nutrients) that influences the expression of hunger. Reduction in blood glucose levels or in the levels of the circulating adipose tissue hormone leptin indicates a deficit in available energy and in energy reserves. Such events are linked with feelings of hunger. Fluctuation of these factors indicates the metabolism and storage of the body's energy reserves. These are a tonic class of physiological signals that also influence the expression of appetite. Like episodic satiety signals, these tonic signals normally act on inhibitory mechanisms with the hypothalamus (anorexogenic circuits). Their absence elicits an active feeding response. Other tonic factors that indicate the body's energy status, such as adiponectin, cytokines, and gonadal hormones, also appear to act on energy regulator centers within the brain, particularly the hypothalamus, mainly to suppress hunger. Again, their absence serves to unleash appetite.

However, not all physiological signals, episodic or tonic, inhibit hunger. For instance, blood levels of the gut hormone ghrelin have been shown to increase before a meal. Subsequent intake has been shown to suppress ghrelin release. Further studies have shown that ghrelin infusions increase food intake. Thus, this is a hormone that acts to promote food intake. Interestingly, ghrelin receptors are found in various hypothalamic locations that form part of the orexigenic circuits promoting food intake. These circuits contain many neuropeptides, such as neuropeptide Y, orexins, melanocortin concentrating hormone, and galanin, which all stimulate food intake. The precise nature of the physiological and neurobiological regulation of appetite is discussed elsewhere in this encyclopedia.

Finally, it should be noted that the biological mechanisms critical to the expression of hunger are not independent of psychological ones. Indeed, the sensory and cognitive cues that stimulate hunger produce physiological changes that anticipate the ingestion and metabolism of energy and subsequently aid these processes. This gives rise to the psychological factors critical in the expression of appetite. Furthermore, the taste and thoughts of foods can generate sensations resembling hunger, sometimes called "hedonic hunger" (Lowe and Butryn, 2007). This other concept of hunger arises from the hedonic appetite system, which interacts with the homeostatic appetite system in the expression of our motivation toward and desire to eat certain foods. It has been argued that the concept of hedonic hunger actually represents a form of implicit wanting and is simply a reflection of food reward (Beaulieu and Blundell, 2020). The food reward components of liking and wanting can override homeostatic satiety signals and lead to overconsumption by increasing the size of a meal (overriding satiation) and by "eating in the absence of hunger" during the postingestive period (overriding satiety). It is important to make the distinction between hunger as a biological drive and implicit wanting as a component of the hedonic system as separate but interacting entities.

Learning and hunger

As the classical work of Pavlov demonstrates, hunger responses can be easily entrained to specific stimuli. Omnivores, like carnivores, are meal eaters (in contrast to grazers); however, their dietary decisions are more complex. One of the essentials for an omnivore faced with a variety of new and different foods is the capacity to learn. It is not possible for an inborn preference or aversion to guide the choice of every possible food. Therefore, we learn which foods are beneficial (and which are not) by eating them. If the consequence of consumption is normal satiety, the pleasant postingestive signals this generates serve as a positive experience that can lead to learned preferences, as studied by Birch (1999). Alternatively, if consumption leads to an association between that food and negative gastrointestinal consequences such as nausea, then a conditioned aversion to that food (or even other similar tastes and flavors) is likely to develop. The learning process involves the association between the sensory and the postabsorptive characteristics of foods. In this way, the sensory characteristics of foods act as cues and come to predict the impact that foods will later have. Consequently, these cues should suppress hunger according to their relationship with subsequent physiological events. It is possible to demonstrate experimentally how human beings adapt their eating to a food's energy content. A distinctively flavored food that contains "extra" hidden energy, presented on several occasions, will result in a change in eating and in preference. When deprived of food, subjects' preference for the taste increases with gained experience. If presented when satiated, preference for the taste decreases. This process is also observable in young children, who eat smaller meals following a taste previously associated with a high-energy snack, and larger meals following a taste previously associated with a low-energy snack. Our hunger can thereby be potently reduced by the learned associations of satiety.

In environments where food sources remain limited and dietary variety is controlled, learned responses may effectively control subjective feelings of appetite. However, when there is distortion, variation, or extreme complexity in the relationship between sensory characteristics and nutritional properties, then the conditioned control of hunger is weakened or lost. In such a scenario, hunger may become less controllable. In many respects, the variety of foods available to us represents a cacophony of different sensory characteristics and has the added complication of ingredients that preserve the sensory qualities while altering their nutritive value. Learned hunger therefore is a relatively less important factor when the food supply contains many food items with identical tastes but differing metabolic properties.

Hunger and eating behavior

If hunger is biologically useful and a subjective experience that indicates a depleted nutritional state, then a close correspondence between hunger and eating would be expected. Hunger should be either a necessary or a sufficient condition for eating to occur. However, this is not invariably the case. Instances of people deliberately refraining from eating in spite of hunger (fasting for moral or political conviction) show hunger not to be a sufficient condition. People can, and regularly do, resist the drive to consume. As aforementioned in relation to food reward, examples, from both research and daily experience, of eating a tempting food when otherwise satiated show hunger not to be necessary for eating to take place. In the laboratory, merely changing the food on offer, or increasing its palatability, can encourage eating beyond the limits of normal satiety. Many of us have experienced festive overconsumption or had our interest tweaked by the dessert trolly. However, although the relationship between hunger and eating is not based on biological inevitability, they are not entirely uncoupled and under many circumstances they remain closely linked.

The lack of a one-to-one correspondence between hunger and eating is repeatedly used to question the validity of hunger ratings. But should a high correlation between hunger ratings and subsequent food intake be expected in all circumstances? The previous examples show that in certain circumstances the two can be disengaged. Thus, for example, eating can occur when hunger is low (such as when highly palatable food is offered unexpectedly) and not at other times when hunger is high (when food is unavailable or other activities have priority). The proximity to food cues can have dramatic, instantaneous effects on appetite expression. For some individuals, this responsiveness to food cues overwhelms normal appetite control and the normal rhythm of hunger. For others, eating in the absence of hunger is a component of their natural eating repertoire. It is clear that hunger ratings cannot be used simply as a proxy measure for food intake. Equally, when such factors are taken into account, there is good evidence that self-report ratings of hunger correlate statistically and meaningfully with eating.

In questioning the relationship between hunger and eating, we are also forced to place the action of hunger within a broader context of social and psychological variables that moderate food choice and eating behavior. Eating patterns are maintained by enduring habits, attitudes, and opinions about the value and suitability of foods, and an overall liking for them. These factors, derived from the cultural ethos, largely determine the range of foods that will be consumed and sometimes the timing of consumption. The intensity of hunger experienced may also be determined, in part, by the culturally approved appropriateness of this feeling and by the host of preconceptions brought to the dining table. Fluxes in hunger over time are therefore only one portion of the range of determinants of eating in any given situation.

Disorders of hunger

The clinical eating disorders anorexia nervosa and bulimia nervosa are commonly believed to encompass major disturbances of hunger. Yet the role that hunger may play is not entirely clear. Contrary to the literal meaning of the term, "anorexia" is not experienced as a loss of appetite. Rather, clinicians recognize that people with anorexia may endure intense periods of hunger during their self-restricted eating. For some, their strength in resisting intense episodes of hunger provides a feeling of self-mastery and control that is absent in other areas of their lives. Research suggests that people with anorexia restricting food intake (compared with those who binge) have the greatest blunting of hunger response, and that this disturbance in hunger is not a product of other areas of perceptual confusion (Halmi and Sunday, 1991).

There is evidence that under conditions of total starvation, hunger may become temporarily diminished. This circumstance is extremely rare and obviously relatively brief. Once eating is recommenced, hunger returns rapidly and with extreme intensity. The accounts of the male volunteers who submitted to a 6-month period of semistarvation during World War II (the "Minnesota Experiment") are a testament to the extreme power of hunger (Keys et al., 1950). Referred to as semistarvation neurosis, these men's activities were shaped by their need for food, and their hunger experience was extreme. Nearly two-thirds reported feeling hungry all the time and a similar proportion experienced physical discomfort due to hunger. Participants described a marked increase in what was referred to as "hunger pain." For some, this was mildly discomforting and vaguely localized in the abdomen. For others, it was extremely painful. This account is especially useful in reminding us why energy-reduced diets aimed at achieving weight loss are often difficult to maintain and easy to abandon.

Like anorexia, bulimia finds its literal meaning in changed hunger–"ox hunger." Again, however, the term is imprecize. Close analysis of the precursors of binge episodes shows hunger to be lower than it is before a normal meal. In addition, although the urge to eat may be strong during a binge, the large amount of food consumed implies some defect in satiation rather than in hunger. Moreover, bingeing is often a well-practiced behavior that develops and changes with time. As with people with anorexia, it is likely that a stable eating pattern is necessary in order to normalize the experience of hunger, a process that may take a long time to establish.

The question of whether obesity reflects a disorder of hunger is now regarded as largely redundant. There is hardly any evidence of heightened levels of hunger contributing to excessive energy intake in people living with obesity. However, an exception to this is the rare disorder Prader–Willi syndrome. Genetically determined and characterized mainly by intellectual disability, obesity is a well-recognized feature of the syndrome. Research suggests that the excessive levels of food intake are associated with both a delayed reduction in hunger while eating and a more rapid return to premeal states when eating has finished. It has been postulated that tonic inhibitory signals such as leptin and insulin are weakened at higher levels of adiposity due to leptin and insulin resistance, also affecting episodic satiety signaling, which may promote overconsumption and further weight gain in people living

with obesity. Clearly, a better understanding of the biological events that accompany such aberrant eating patterns will strengthen understanding of the psychobiological framework that supports hunger.

Conclusion

Hunger is a universal human experience and sensation. The sensation of hunger as a motivational drive to eat can be measured using validated tools and be quantified in its intensity. This intensity oscillates on a daily basis in response to food intake and inversely with satiety. Hunger is related to a fundamental process in human biology in order to replenish energy used up in metabolic processes and during activity. However, other intervening variables such as food reward, learned responses, culture and the environment can influence its intensity and periodicity as well as shape its expression.

References

Beaulieu, K., Blundell, J., 2020. The psychobiology of hunger—a scientific perspective. Topoi 40, 565–574.
Beaulieu, K., Hopkins, M., Blundell, J., Finlayson, G., 2018. Homeostatic and non-homeostatic appetite control along the spectrum of physical activity levels: an updated perspective. Physiol. Behav. 192, 23–29.
Birch, L.L., 1999. Development of food preferences. Annu. Rev. Nutr. 19, 41–62.
Blundell, J.E., Gibbons, C., Beaulieu, K., et al., 2020. The drive to eat in homo sapiens: energy expenditure drives energy intake. Physiol. Behav. 219, 112846.
Blundell, J.E., 1979. Hunger, appetite and satiety-constructs in search of identities. In: Nutrition and Lifestyles. Applied Science, London, pp. 21–42.
Flint, A., Raben, A., Blundell, J.E., Astrup, A., 2000. Reproducibility, power and validity of visual analogue scales in assessment of appetite sensations in single test meal studies. Int. J. Obes. Relat. Metab. Disord. 24 (1), 38–48.
Friedman, M.I., Ulrich, P., Mattes, R.D., 1999. A figurative measure of subjective hunger sensations. Appetite 32 (3), 395–404.
Green, S.M., Delargy, H.J., Joanes, D., Blundell, J.E., 1997. A satiety quotient: a formulation to assess the satiating effect of food. Appetite 29 (3), 291–304.
Halmi, K.A., Sunday, S.R., 1991. Temporal patterns of hunger and fullness ratings and related cognitions in anorexia and bulimia. Appetite 16 (3), 219–237.
Holt, S.H., Miller, J.C., Petocz, P., Farmakalidis, E., 1995. A satiety index of common foods. Eur. J. Clin. Nutr. 49 (9), 675–690.
Keys, A., Brožek, J., Henschel, A., et al., 1950. The Biology of Human Starvation. University of Minnesota Press, Minneapolis.
King, N.A., Caudwell, P.P., Hopkins, M., et al., 2009. Dual-process action of exercise on appetite control: increase in orexigenic drive but improvement in meal-induced satiety. Am. J. Clin. Nutr. 90 (4), 921–927.
Kirkmeyer, S.V., Mattes, R.D., 2000. Effects of food attributes on hunger and food intake. Int. J. Obes. 24 (9), 1167–1175.
Kissileff, H.R., 1984. Satiating efficiency and a strategy for conducting food loading experiments. Neurosci. Biobehav. Rev. 8 (1), 129–135.
Lowe, M.R., Butryn, M.L., 2007. Hedonic hunger: a new dimension of appetite? Physiol. Behav. 91 (4), 432–439.
Monello, L.F., Mayer, J., 1967. Hunger and satiety sensations in men, women, boys, and girls. Am. J. Clin. Nutr. 20 (3), 253–261.
Rogers, P.J., 1999. Eating habits and appetite control: a psychobiological perspective. Proc. Nutr. Soc. 58 (1), 59–67.

Further reading

Beaulieu, K., Hopkins, M., Blundell, J.E., Finlayson, G., 2016. Does habitual physical activity increase the sensitivity of the appetite control system? A systematic review. Sports Med. 46, 1897–1919.
Blundell, J.E., De Graaf, C., Finlayson, G., et al., 2009. The measuring of food intake, hunger and satiation in the laboratory. In: Allison, D.B., Baskin, M.L. (Eds.), 2nd Edition of the Handbook of Assessment Methods for Obesity and Eating Behaviours. Sage, Newbury Park, California, pp. 283–326.
Halford, J.C.G., Harrold, J.A., 2008. Neuropharmacology of human appetite expression. Dev. Disabil. Res. Rev. 14, 158–166.
Hill, A.J., Rogers, P.J., Blundell, J.E., 1995. Techniques for the experimental measurement of human eating behavior and food intake: a practical guide. Int. J. Obes. 19, 361–375.
Ogden, J., 2002. The Psychology of Eating. From Healthy to Disordered Behavior. Blackwell, Oxford.
Womble, L.G., Wadden, T.A., Chandler, J.M., Martin, A.R., 2003. Agreement between weekly vs. daily assessment of appetite. Appetite 40, 131–135.

Hydration and thirst

J. Leiper, Loughborough University, Loughborough, UK

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of J. Leiper, Thirst Physiology, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 280–287, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00267-1.

Role of Thirst in Water Balance 143
Perception of Thirst 143
Assessment of Thirst 143
The Physiological Regulation of Thirst 144
Mechanisms of Thirst Regulation 146
Osmotic Regulation of Thirst 147
Volemic Regulation of Thirst 148
Sensory Regulation of Thirst 149
Mechanisms for Terminating the Sensation of Thirst 149
Fluid Requirements 150
Further Reading 150

Role of Thirst in Water Balance

Approximately 70% of the lean body mass of an individual is composed of water, with approximately two-thirds of the total body water (TBW) volume being held within the cells of the body (intracellular pool), with the remaining one-third (extracellular pool) divided between the circulating blood plasma (intravascular pool) and the fluid-filled spaces between the cells (interstitial pool). The volume and distribution of the body fluids are mainly determined by the amounts of body water and sodium. In humans, the TBW content is regulated daily to within approximately 0.2% of the lean body mass under normal, temperate conditions by factors that control input and output. The kidneys regulate water excretion in excess of the evaporative loss and the fecal and obligatory urine losses. Water intake occurs in the form of food and drink, with the sensation of thirst underpinning drinking behavior.

The mechanisms that monitor the body's hydration status also interact with the thirst control centers in the brain to regulate the desire to drink. This article focuses on the physiological factors that govern the perception of thirst and how this is altered by drinking.

Perception of Thirst

Thirst is a sensation that is best described as the desire to drink. The reason for drinking may not be directly involved with a physiological need for water intake, but can be prompted by habit, ritual, taste, nutrients, craving for alcohol, caffeine, or other drugs in a beverage, or a desire to consume a fluid, which will provide a warming or a cooling sensation. Much of the perception of thirst is a learned or a conditioned process, with signals such as dryness of the mouth or throat initiating drinking, whereas a feeling of fullness of the stomach can stop ingestion before a fluid deficit has been restored.

Currently, the thirst response is thought to be regulated by neural modulators that operate as a reward mechanism, integrating the effective requirement for water intake with the sensations of taste and pleasantness of the fluid ingested. Thus, when the individual is hypohydrated multiple areas of the brain are activated, promoting the intensity of the thirst sensation. As the water deficit is restored, the feeling of thirst diminishes and this subjective sensation correlates well with a reduction in neural activation. However, areas of the brain associated with taste, which are activated by water when thirsty, remain active following drinking to satiety when water is ingested.

Although it is true that thirst in humans is a poor indicator of acute hydration status and that daily fluid intake is normally in excess of obligatory water loss, the preservation of the TBW volume under a variety of environmental and nutritional stresses is remarkably robust and is mainly due to the drive to drink that the sensation of thirst chronically induces.

Assessment of Thirst

In humans, two main techniques have been used to identify the perception of thirst and its alleviation by drinking. The first method is to monitor the volume of drink voluntarily ingested by an individual within an allotted time period and to compare the amount consumed with the volume of fluid required to restore a given water deficit or other imbalance of the body water pools. The other method is to assess the individual's perceived rating of thirst by asking them to record on a visual analog scale their responses to

a series of questions that are thought to relate to the sensation of thirst (Figure 1). The questionnaire technique has the advantages that it allows a series of measurements to be made before, during, and following the period of drinking, and appears to provide an indication of the relative strength of a given stimulus. Although in many studies both methods are used to gauge the sensation of thirst and the responses have been correlated to a number of physiological parameters that are known to influence the drive to drink, it is widely recognized that at present there is no consistently reliable measure of the thirst sensation.

A more recently introduced technique has been the use of noninvasive methods of imaging to identify the specific regions of the brain that are activated during the genesis and satiation of the thirst response. Different techniques are used to image brain activation. Both Positron Emission Tomography (PET) and functional Magnetic Imaging (fMRI) are used to visualize brain activation by detecting either temporal changes in blood flow or in the chemical composition of regions in the brain that occur when individuals are exposed to specific stimuli. The number of brain regions, their specificity, and the intensity of activation have also been correlated with the subjective perception of thirst.

The Physiological Regulation of Thirst

As thirst is the major factor controlling water intake, the physiological regulation of thirst is associated with the need to maintain a relatively stable volume of TBW. Although water is lost from the body continually, albeit usually in relatively small amounts, and hence the body is almost always developing a water deficit, water intake is intermittent. The amount of fluid usually ingested is in excess of that required to replace the losses incurred since the last water intake. The factors that initiate, maintain, and end the drinking response are various and are not fully understood. However, as the regulation of the volume and composition of the various water pools of the body play an essential role in controlling the perception of thirst, an understanding of the homeostatic mechanisms involved has provided us with the best insight we have into the complexities of the perception of thirst.

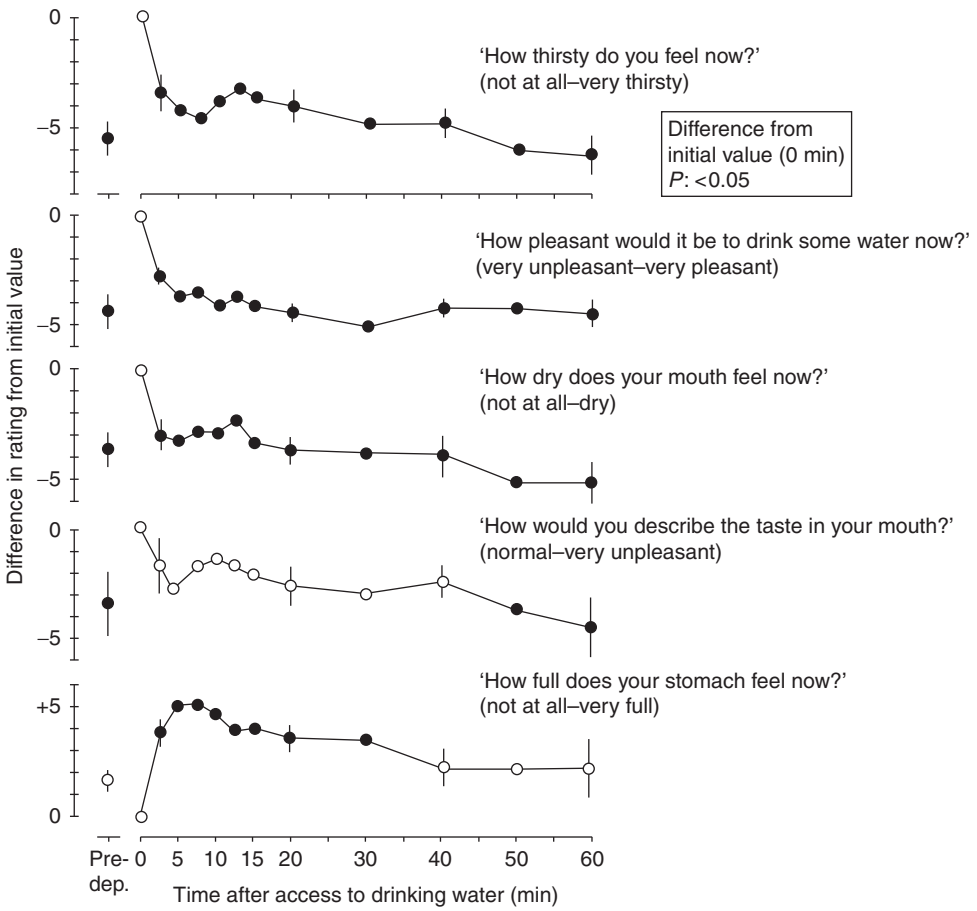


Figure 1 Subjective responses to a series of five questions assessed by visual analog scale ratings from 24-h water-deprived individuals before and during a 60-min rehydration period. The data are shown as differences (in centimeters) from the initial values and significant differences are indicated by filled symbols. Reproduced from Rolls BJ, Wood RJ, and Rolls ET (1980) Thirst: The initiation, maintenance, and termination of drinking. *Progress in Psychobiology and Physiological Psychology* 9: 263–321.

The total volume, distribution, and composition of body fluids must be regulated within narrow limits for normal cellular function to be maintained. Body water is passively distributed between the extracellular and intracellular pools according to osmotic, oncotic, and hydrostatic forces as shown in **Figure 2**. The sodium and chloride contents of the extracellular fluid constitute the two greatest osmotically active components of this fluid and are therefore important in maintaining its volume. Potassium, phosphate, and protein play a similar role in regulating the intracellular fluid volume. The distribution of water between the intravascular and the extravascular pools is dependent on the balance of hydrostatic and oncotic pressures across the capillaries and post capillary venules.

Variation in the water to solute ratio of a body fluid pool results in changes in the tonicity and hence effective osmolality of the fluid. As the various body water pools are in dynamic equilibrium with each other (**Figure 2**), there is a tendency for adjustments to occur throughout the body as water moves from regions of low solute concentration to those of higher solute concentration. Changes in plasma osmolality are relatively easy to monitor; therefore, there is a tendency to consider changes in the circulation as the effector of fluid balance control. It is, however, important to remember that any alteration in one body pool will affect the others and that receptors that initiate responses affecting water balance may reside at sites far removed from the circulation.

Loss of water from the body or an increase in the circulating solute concentration causes an increase in the osmolality of primarily the extracellular fluid; water then moves into the extracellular space from the cells, producing a reduction in cell volume. Changes in plasma osmolality are therefore thought to be signaled to the effector mechanisms by changes in the cell volume of specific specialized cells, collectively termed osmoreceptors. As the main solute determining the tonicity of the extracellular fluid is sodium, there has been some debate as to whether the receptor cells detect changes in osmolality or changes in the sodium ion content. The evidence to date suggests that at least the majority of the receptors respond to osmolality rather than to sodium concentration. These osmoreceptors play a regulatory role not only in the perception of thirst but also in the maintenance of the circulating levels of hormones that regulate the excretion of water and solute by the kidneys (**Figure 3**). As increases in the extracellular osmolality effectively decrease the volume of the cells in the body, this form of dehydration is termed cellular dehydration.

Alteration in the volume of the extracellular fluid pool without change in its osmolality also affects the fluid balance hormone concentrations and the sensation of thirst. Changes in blood volume affect the blood and capillary pressures and atrial filling pressure. The effect on capillary pressure will tend to redistribute body water and help to adjust the circulating fluid volume, and the change in venous return to the heart will alter the cardiopulmonary and arterial stretch receptor (baroreceptor) activity. The level of afferent activity from these baroreceptors directly affects both the sensation of thirst and the secretion of some fluid balance hormones. Additionally, modifications to the arterial blood pressure can directly affect renal perfusion, which, together with baroreceptor activity to the kidneys, regulates the renin–angiotensin system (**Figure 4**). Although the effect on the kidneys can influence the perception of thirst, the main renal response is to regulate urine water and solute excretion. A decrease in the volume of the extracellular pool with no concomitant change in plasma osmolality is termed extracellular dehydration.

When humans are given access to fluids after the development of a water deficit, their drinking response usually follows a pattern of rapid ingestion of more than 50% of the total intake, followed by intermittent consumption of relatively small volumes of drink over a longer period. Although initiation of the response to drink is due mainly to osmotic or blood volume (volemic) changes, other mechanisms appear to be involved in the control of the continuation and satiety responses. Receptors in the mouth, esophagus, and gastrointestinal tract appear to be major factors in the acute regulation of thirst satiation, with the effects that the volume and solute content of the ingested drink have on restoring the fluid deficits controlling the chronic regulation of thirst (**Figure 5**). There is a close relationship between eating and drinking, with approximately 70% of daily fluid intake normally being associated with meals (**Figure 6**). The desire to drink while eating is probably produced by a series of responses including the mechanochemical composition of the food before absorption, the neuroendocrine response to digestion, the movement of water into the intestine during digestion, and the osmotic solute load that occurs following absorption. The intake of minerals is essential to replace that lost from the body and for growth. The majority of mineral intake is supplied by the food ingested, and indications of a desire or appetite for ingesting specific minerals have been shown in animals and humans. Although sodium appetite has been linked to the sensation of thirst, anatomically and functionally, the controlling mechanisms are distinct and separate.

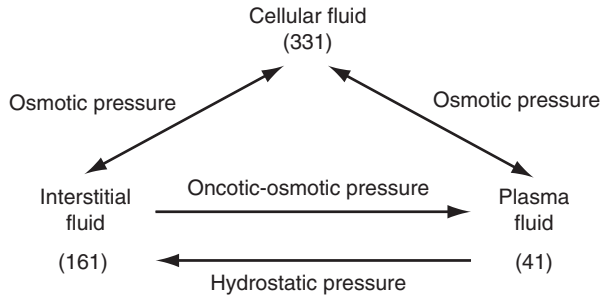


Figure 2 Diagrammatical representation of the forces that regulate the distribution of the body water pools. The volumes given are those determined in a single male subject with a lean body mass of 75.8 kg.

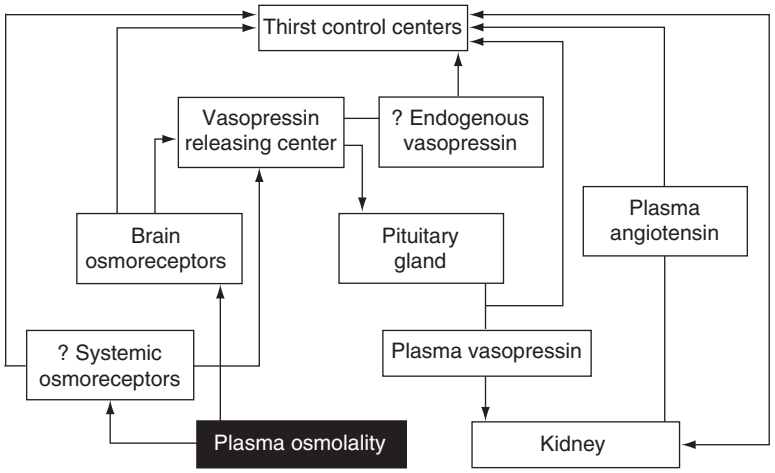


Figure 3 Schematic representation of the main factors proposed in osmotically induced regulation of the sensation of thirst and their interaction with the control of diuresis. An increase in plasma osmolality will tend to stimulate greater excitatory activity, whereas a decrease in osmolality will activate more inhibitory inputs. Neural pathways are indicated by →? and hormonal input by →.

Mechanisms of Thirst Regulation

The sensation of thirst is regulated separately both by the osmotic pressure and by the volume of the body fluids and as such is closely related to the control mechanisms that are responsible for the secretion of the fluid balance hormones, which affect water and solute reabsorption in the kidneys and play a role in blood pressure control. These hormones, arginine vasopressin, atrial natriuretic peptide, oxytocin, and the renin–angiotensin–aldosterone system, are central to the regulation of thirst. Regions of the brain, in the hypothalamus and the forebrain, appear to be the main areas involved in the control of thirst and antidiuresis, and collectively, these parts of the brain have been termed the thirst control centers. Neurons that are responsive to changes in osmolality, intravascular volume (volemia), and blood pressure are found within these areas of the brain, as are other receptors that are responsive to many of the fluid balance hormones. Neural pathways from the thirst control centers and the kidneys may allow some direct integration between the control of thirst and excretion, whereas within the brain, all of the major fluid balance hormones are present as neurohormones. Afferent input from systemic receptors monitoring osmolality, circulating sodium concentration, and changes in intravascular volume and pressure also play roles in controlling the feeling of thirst. Therefore, there appears to be a complex integrated system for both monitoring the status of the body water pools and controlling intake and excretion (**Figures 3 and**

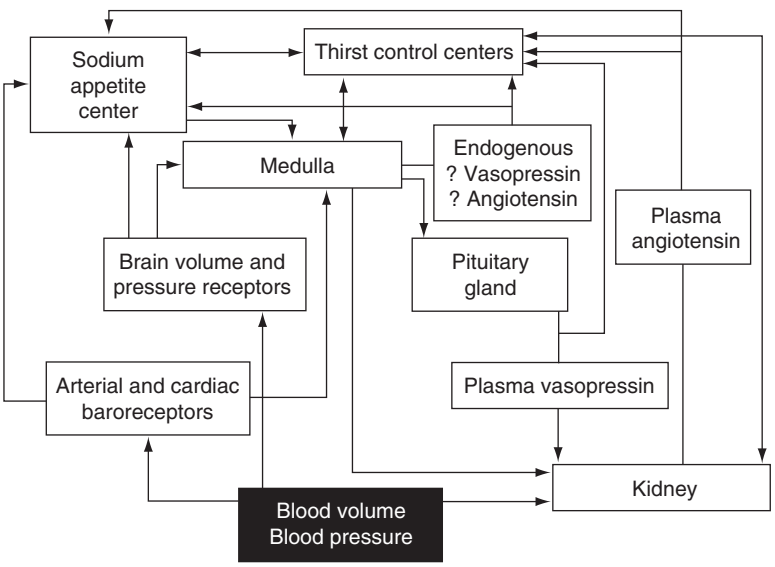


Figure 4 Schematic representation of the main factors proposed in volemic-induced regulation of the sensation of thirst and their interaction with the control of diuresis and sodium appetite. A decline in circulating blood volume will decrease baroreceptor activity, which will increase excitatory activity, whereas an increase in volume will have the opposite effect. Reduction in blood pressure will decrease renal perfusion, which will activate the renal renin–angiotensin system. Neural pathways are indicated by → and hormonal input by →.

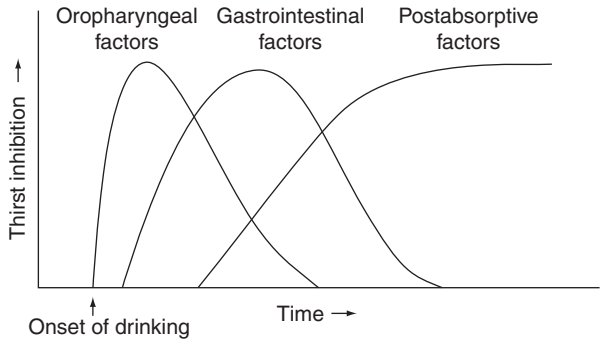


Figure 5 Schematic diagram depicting the proposed onset, duration, and overlap of various inhibitory signals to continue fluid ingestion following initiation of drinking in response to a fluid deficit. Reproduced from Verbalis JG (1990) Inhibitory controls of drinking: Satiation of thirst. In: Ramsay DJ and Booth DA (eds.) *Thirst: Physiological and Psychological Aspects. ILSI Human Nutrition Reviews*, pp. 313–330. London: Springer-Verlag.

4). Much of the regulatory mechanisms controlling water balance appear to overlap, with several stimuli appearing to subserve the same response; however, it is assumed that this effect is required in order to ensure that the blocking of one type of stimulus will not prevent homeostatic control.

Osmotic Regulation of Thirst

The osmolality of circulating plasma is normally maintained within a very narrow limit of between 270 and 295 mosmol kg^{-1}, with the circulating levels of the antidiuretic hormone arginine vasopressin playing a major role in its homeostatic regulation. An increase of as little as 2–3% in plasma osmolality is sufficient to produce a strong sensation of thirst and a significant increase in circulating arginine vasopressin concentration (Figure 7). The osmoreceptors that monitor the tonicity of the body pools appear to reside mainly in the subfornical organ (SFO) and organum vasculosum of the lamina terminalis (OVLT), an area of the brain that lacks a blood–brain barrier; therefore, they appear to respond mainly to the changes that occur in the osmolality of the blood rather than in the cerebral interstitium. Although the changes in the circulating levels of arginine vasopressin and the perception of thirst appear to parallel one another, it is unlikely that the same receptors are responsible for both responses; it is likely that there are different neurons that react to the same stimulus. However, there may be some neuro-hormonal interaction between the osmotically

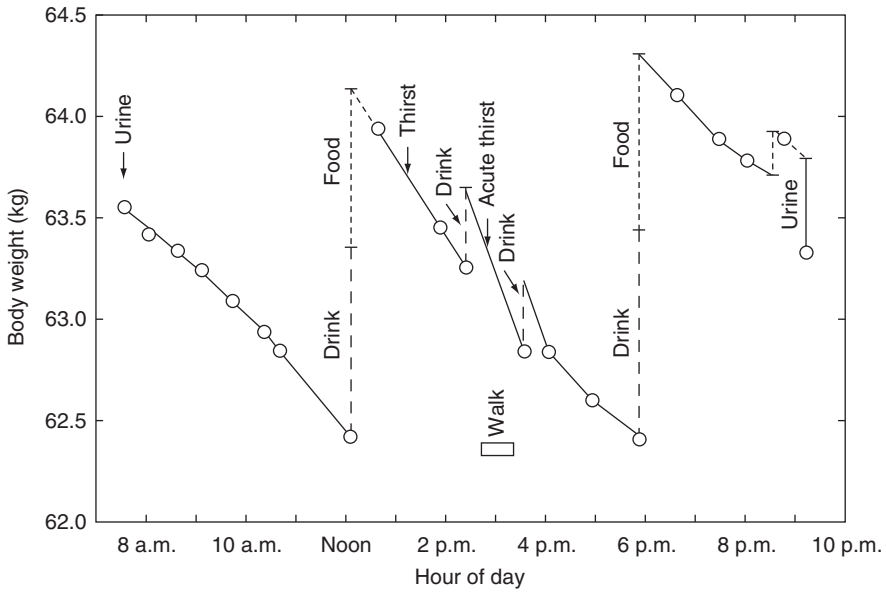


Figure 6 Changes in body weight during 13 h in the desert. The majority of the volume of drink ingested was associated with food intake. Sweat loss varied from 150 to 700 ml h^{-1}; the total fluid intake was 3.05 l. At the end of the period, body mass was essentially the same at the beginning and end of the day; therefore, water intake and output were equal. Reproduced from Adolph ED (1947) *Physiology of Man in the Desert*. New York: Interscience.

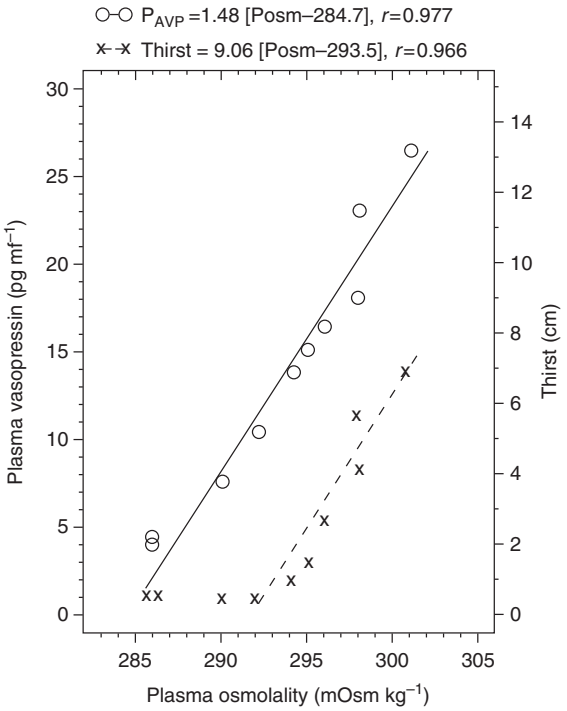


Figure 7 The relationship of plasma vasopressin (○) and thirst (x) with plasma osmolality in a volunteer during an infusion of 5% saline. Reproduced from Robertson GL (1984) Abnormalities of thirst regulation (Nephrology forum). *Kidney International* 25: 460–469.

activated thirst centers and the 'vasopressin releasing center' in the brain, and arginine vasopressin-responsive neurons have been detected within the thirst centers (Figure 3).

The current theory of the osmotic control of thirst suggests that there is a constant output of both inhibitory and excitatory neural activity from the respective osmoreceptors to the thirst centers and the arginine vasopressin-releasing centers. Within the normal range for plasma osmolality, the inhibitory and excitatory activities in the thirst centers effectively cancel out one another and there is neither a sensation of thirst nor satiety. However, at this level of activity, there is release of a basal level of arginine vasopressin, which is sufficient to maintain a state of half-maximum antidiuresis. An increase in plasma osmolality above the normal level stimulates greater excitatory output, causing an increase in the feeling of thirst and higher levels of circulating arginine vasopressin. Elevated levels of arginine vasopressin increase the concentrating ability of the kidneys. A decrease in plasma osmolality below the normal range increases the inhibitory output, producing a feeling of satiety, and arginine vasopressin release is suppressed, allowing urinary dilution (Figure 3).

Cells and fibers within the brain have been shown to contain several hormones, including angiotensin and vasopressin, within the same cell. Although neurons associated with the thirst centers can be activated *in vitro* by vasopressin, it is not clear at present whether either peripheral- or neural-generated arginine vasopressin levels do influence the perception of thirst.

Volemic Regulation of Thirst

The receptors that initiate hypovolemic thirst are generally thought to be the cardiovascular baroreceptors, which respond to underfilling of the circulation by reducing their inhibitory nerve impulse activity to the thirst centers. However, in areas of the brain associated with the thirst centers, there are neurons that are separately responsive to volemic, pressure, and osmotic changes. This suggests that at least part of the response to changes in blood volume originates in the brain. It is thought that changes in blood pressure and osmolality are monitored mainly within the brain, whereas variations in blood volume are principally sensed by the peripheral baroreceptors, with a degree of overlap between the different receptors. The mechanisms that respond to changes in intravascular volume and pressure appear not to be as sensitive as those responsive to osmotic changes, for example, a decrease of approximately 10% of the plasma volume is required to initiate hypovolemic thirst. As fairly large variations in blood volume and pressure occur during normal daily activity, such as postural changes and physical activity, this apparent lack of sensitivity presumably prevents overactivity of the volemic control mechanisms. As with osmotic thirst, the control of volemic thirst is thought to be a balance between continuous inhibitory and excitatory neural activity, although in this system, the basal level appears to be essentially inhibitory. Another difference in the basic control mechanism between the two systems occurs due to the requirement for

both solute, mainly sodium, and water to restore the extracellular volume. Therefore, extracellular dehydration causes an initial thirst and a delayed increase in sodium appetite.

Reduction in the intravascular volume sufficient to lower cardiac output and arterial blood pressure decreases afferent activity from the low- and high-pressure cardiovascular baroreceptors to the thirst centers and increases sympathetic activity to the kidneys. As afferent input from the baroreceptors to the thirst centers is inhibitory, a decrease in activity produces a reflex increase in the perception of thirst and also appears to directly stimulate arginine vasopressin release. The increase in sympathetic activity to the kidneys directly promotes greater renal renin release. In addition, reduction in blood pressure lowers the renal perfusion pressure, which stimulates renin release both as a direct pressure effect and by decreasing the delivery of sodium to the kidneys.

Increased activation of the renin–angiotensin–aldosterone system also helps regulate hypovolemic thirst. Although circulating levels of both vasopressin and aldosterone affect water and sodium reabsorption in the kidneys and thereby control water and solute loss, angiotensin acts directly on the thirst and sodium appetite centers to stimulate their respective responses. Neurons that are stimulated by angiotensin are found in areas of the brain that lack a blood–brain barrier; therefore, circulating angiotensin has direct access to both centers. In addition, the release of neurally generated angiotensin is promoted by suitable neuron activity responding to sensory stimuli (**Figure 4**).

There are therefore a variety of neural and hormonal responses that interact to modulate and control both thirst and urine excretion. A number of other hormones including oxytocin, atrial natriuretic peptide, tachykinins, neuropeptide Y, thyroid hormones, corticotrophin releasing factor, and steroid hormones have also been shown experimentally to affect the drinking response. Under normal conditions of water and solute loss, both osmotic and volemic dehydration occur; therefore, stimuli from receptors for both systems are usually involved in the sensation of thirst. Increases in extracellular osmolality appear to be more effective than hypovolemia in promoting the thirst and hence the drinking response. Greater than 70% of the stimulus to drink appears to be generated by increased osmolality.

Sensory Regulation of Thirst

The sensations of a dry mouth or the desire for a specific taste or effect also generate the desire to drink while there may be no physiological requirement to drink. A dry mouth promotes changes in neural activity in the parahippocampus, amygdala, thalamus, cingulate, insula, allocortex, and transitional cortex of the brain. This finding has strengthened the hypothesis that the perception of thirst is a primitive vegetative function that appeared long before vertebrates evolved.

Drinking water activates areas of the anterior insular and frontal opercular cortex that are also involved in the perception of taste. Areas of the orbitofrontal cortex are also activated by the ingestion of water or sweet or salty tastes, but here, activation is greatest when the subjects are thirsty and it diminishes when the subjects have drunk water to satiety. This has been interpreted as functionally separated areas of the brain, one of which responds to taste stimuli that are not diminished following drinking to satiety, whereas the other is highly active during drinking when water is physiologically required but that reduces as the need for water is met.

Mechanisms for Terminating the Sensation of Thirst

Although undoubtedly decreasing osmolality and increasing extracellular volume promote a reduction in the perception of thirst by reactivating inhibitory neuron activity, usually, there is a decrease in the perception of thirst and termination of drinking before circulating osmolality, volume, and hormonal levels have returned to pre-dehydration levels. Although it could be argued that receptors in the brain may be responsible for the early cessation of the perception of thirst, the majority of the evidence suggests that it is receptors in the upper gastrointestinal tract that promote the early termination of drinking. Although the nature and neural connections of these proposed receptors have not been fully characterized, most appear to have an inhibitory response. It has been suggested that as much of the thirst and drinking response is behavioral, an individual learns what volume of drink is required to restore a given water deficit. Termination of drinking therefore could be a learned response, which anticipates a future fluid deficit or matches a known current level of dehydration. The stimuli for gauging the current level of dehydration may be the same as that which initiates the sensation of thirst.

The mere presence of a liquid, particularly cold liquid, in the mouth reduces the perception of thirst. Receptors in the mouth and esophagus responding to different chemical, tactile, pressure, and temperature stimuli are thought to be part of the mechanism that influences the relative intensity of the perception of thirst. Although the neural activity involved in swallowing, and perhaps oropharyngeal and gastric receptors are thought to be effective in sensing or metering the volume of liquid ingested, distension of the stomach tends to inhibit drinking due to increased gastric stretch receptor activity, although this response does not always reduce the perception of thirst. Taste and other psychological factors can have a stimulatory effect on consumption of a drink that is considered to be palatable.

The continuation and termination of the acute sensation of thirst is regulated by a series of stimuli that operate before all of the drink consumed has been absorbed and before disturbances in the body water pools have been corrected. A variety of receptors located from the mouth to the upper part of the small intestine and probably neural control from the higher centers of the brain appear to monitor and regulate the initial volume consumed. After absorption, if restoration of body water pools does not occur, the

sensation of thirst is once again initiated, presumably by the same homeostatic stimuli that initially induced the feeling of thirst, and drinking restarts. The integration of the pre- and post-absorptive stimuli modulates the sensation of thirst and finally the volume of drink consumed.

Fluid Requirements

Renal reabsorption can reduce the volume of water and solute loss, and hence slow the rate of progress of a fluid deficit, but it cannot stop its development. Intake of fluid either as food or drink is required to restore a fluid deficit. Daily fluid intake is highly variable between individuals and the rate of loss is dependent on factors such as environmental temperature, physical activity, sweating rate, antidiuretic function, and dietary solute load. A representative normal daily water turnover in a sedentary individual living in a temperate climate and eating a typical western diet is approximately 2–3 l, and a minimum daily fluid intake of approximately 1.7 l is necessary to conserve fluid balance. The water content of the typical western diet approximates to approximately 1 l and metabolically derived water produces in the order of approximately 300 ml, which together almost offsets the daily obligatory water loss. Therefore, in many situations, the requirement for fluid intake can actually be very low.

There are conditions in which water loss is greater than that indicated above and replacement obviously requires a compensatory increase in the daily fluid intake (**Figure 6**). Urine volume is related to the solute content of the diet, with a minimum volume of approximately 500 ml being necessary to eliminate the usual daily solute load. Diets rich in protein or foods with a high sodium content require a greater obligatory urine output for excretion. The renal concentrating ability at maximum antidiuresis determines the minimum urinary water loss for a given dietary solute load. Normally, there is a wide range of urinary osmolality such that the same solute load can be excreted in 500 ml of urine with an osmolality of 1400 mosmol kg^{-1} or in 23 l of urine with an osmolality of 30 mosmol kg^{-1}. This feature of renal excretion allows body water balance to be maintained while fluid intake volume is varied.

Prolonged relatively intense muscular activity, elevated ambient temperature, and fever all increase the rate of evaporative sweat loss. Individual sweat rates are highly variable, with average exercise-induced losses usually in the order of 1 l h^{-1}, but daily losses of between 10 and 15 l have been recorded. Daily fecal losses associated with a western diet are usually between 100 and 200 ml; however diarrhea, particularly infectious diarrhea, can produce prodigious fecal water losses that are potentially fatal.

Inappropriate fluid intake can be produced following lesions or development of tumors in regions of the brain associated with the thirst centers. Diabetes insipidus promotes an increase in the volume of fluid ingested, which is caused by a lowering of the basal threshold set point for osmotic thirst. A similar, although less pronounced, lowering of the osmotic thirst threshold occurs during pregnancy. In both the young and the elderly, the thirst response can be blunted and inappropriate drinking habits may occur. Psychogenic disturbances in the sensation of thirst and hence fluid intake have also been reported for a variety of clinical conditions.

Further Reading

Adolph, ED, 1947. Physiology of Man in the Desert. Interscience, New York.

de Araujo, IET, Kringelbach, ML, Rolls, ET, McGlone, F, 2003. Human cortical responses to water in the mouth, and the effects of thirst. Journal of Neurophysiology 90, 1865–1876.

Denton, D, Shade, R, Zamarippa, F, et al., 1999. Correlation of regional cerebral blood flow and change of plasma sodium concentration during genesis and satiation of thirst. Proceedings of the National Academy of the United States of America 96, 2532–2537.

Engell, DB, Maller, O, Sawka, MN, Francesconi, RN, Drolet, L, Young, AJ, 1987. Thirst and fluid intake following graded hypohydration levels in humans. Physiology and Behaviour 40, 229–236.

Johnson, AK, Edwards, GL, 1990. Central projections of osmotic and hypovolaemic signals in homeostatic thirst. In: Ramsay, DJ, Booth, DA (Eds.), Thirst: Physiological and Psychological Aspects. ILSI Human Nutrition Reviews. Springer-Verlag, London, pp. 149–175.

Maughan, RJ, 1994. Fluid and electrolyte loss and replacement in exercise. In: Harries, BJ, Williams, C, Stanish, CW, Micheli, A (Eds.), Oxford Textbook of sports medicine. Oxford University Press, New York, pp. 82–93.

McKinley, MJ, Johnson, AK, 2004. The physiological regulation of thirst and fluid intake. News in Physiological Sciences 19, 1–6.

Phillips, PA, Rolls, BJ, Ledingham, JGG, Morton, JJ, 1984. Body fluid changes, thirst and drinking in man during free access to water. Physiology and Behaviour 33, 357–363.

Ramsay, DJ, 1989. The importance of thirst in the maintenance of fluid balance. In: Bayliss, PH (Ed.), Water and Salt Homeostasis in Health and Disease. Bailliere Tindall, London, pp. 371–391 (Clinical Endocrinology and Metabolism, vol 3, no 2).

Ramsay, DJ, 1990. Water: Distribution between compartments and its relationship to thirst. In: Ramsay, DJ, Booth, DA (Eds.), Thirst: Physiological and Psychological Aspects. ILSI Human Nutrition Reviews. Springer-Verlag, London, pp. 23–34.

Ramsay, DJ, Thrasher, TN, 1986. Hyperosmotic and hypovolemic thirst. In: de Caro, G, Epstein, AN, Massi, M (Eds.), The Physiology of Thirst and Sodium Appetite. Plenum Press, New York, pp. 83–96.

Robertson, GL, 1984. Abnormalities of thirst regulation (Nephrology forum). Kidney International 25, 460–469.

Rolls, BJ, Rolls, ET, 1982. Thirst. Cambridge University Press, Cambridge.

Rolls, BJ, Wood, RJ, Rolls, ET, 1980. Thirst: The initiation, maintenance, and termination of drinking. Progress in Psychobiology and Physiological Psychology 9, 263–321.

Rolls, BJ, Wood, RJ, Rolls, ET, Lind, H, Lind, W, Ledingham, JGG, 1980. Thirst following water deprivation in humans. American Journal of Physiology 239, R476–482.

Verbalis, JG, 1990. Inhibitory controls of drinking: Satiation of thirst. In: Ramsay, DJ, Booth, DA (Eds.), Thirst: Physiological and Psychological Aspects. ILSI Human Nutrition Reviews. Springer-Verlag, London, pp. 313–330.

Indirect calorimetry

Michele N. Ravelli[a,b], **Timothy C. Shriver**[b], **Dale A. Schoeller**[b,c], **and Chad Cook**[d], [a] Department of Neurology, University of Wisconsin, Madison, WI, United States; [b] Biotechnology Center, University of Wisconsin, Madison, WI, United States; [c] Nutrition Science, University of Wisconsin, Madison, WI, United States; and [d] Body Science Center, LLC, Gurnee, IL, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of D.A. Schoeller, C.M. Cook, A. Raman, Energy Expenditure: Indirect Calorimetry, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 170–176, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00092-1.

Introduction 152
Laboratory methods 153
Whole body indirect calorimetry 153
Metabolic carts 155
Field methods 157
Douglas bag 157
Max Plank/Kofranyi–Michaels respirometer 157
Tracer methods of indirect calorimetry 158
Labeled bicarbonate 158
Doubly labeled water (DLW) 159
Interpretation of indirect calorimetry results 159
Clinical cases 160
Conclusion 160
References 160

Key points

- Indirect calorimetry assesses the rate of metabolic heat production based on the rates of consumption of oxygen and production of carbon dioxide and nitrogen waste products.
- In nutrition, indirect calorimetry typically referring to a measurement of respiratory gas exchange across the lungs.
- Indirect calorimetry is used to optimize nutrition therapy of hospitalized patients and determine the cardiovascular fitness of otherwise healthy athletes during training or cardiac patients before and during rehabilitation.
- Indirect calorimeters are attached in small carts for transport to individual patients and can be designed to be small enough to hand carry to testing facilities.
- When used under unstable clinical conditions, the interpretation of results from indirect calorimetry should be done carefully to avoid misinterpretation.

Glossary

Bicarbonate The chemical form of carbon dioxide when dissolved in body water

Total body water The sum of all water within and between the tissues of the body

Doubly labeled water A field method of indirect calorimetry to measure energy expenditure over one to three weeks based on the elimination of stable isotope-labeled water from body water

Resting metabolic rate The energy expended while completely at rest. It is the energy expended in breathing, circulating blood, sustaining organ function, and supporting neurological function

Indirect calorimetry A method to measure the metabolic heat produced by an organism calculated from the rates of oxygen consumption and/or carbon dioxide production usually by monitoring respiratory gas exchange as gases enter and exit the lungs

Metabolic cart A common term for a portable, table-mounted indirect calorimeter

Respiratory Exchange Ratio (RER) or Respiratory Quotient (RQ) The ratio of the amount of carbon dioxide produced to the amount of oxygen consumed, used to calculate rates of carbohydrate versus fat catabolized to support energy metabolism

VCO_2 The rate of CO_2 production, often expressed in ml (mL) min^{-1}

VO_2 The rate of O_2 consumption, often expressed in ml (mL) min^{-1}

Introduction

All living organisms require a source of energy for survival. Among aerobic animals, this energy is provided in the form of chemical energy in the macronutrients they consume, which are converted to carbon dioxide and water through oxidative phosphorylation. This conversion is subject to the same laws of thermodynamics that govern all energy systems. The first law of thermodynamics states that energy can neither be created nor destroyed; it can only be exchanged from one form to another. Hence, the chemical energy consumed in the form of macronutrients in food is converted into mechanical energy for work performed by the body, heat energy lost to the environment or retained by raising body temperature, or stored as chemical energy in tissues such as fat, protein, or carbohydrates (glycogen). This conservation of energy can be stated mathematically as:

$$\text{Energy}_{\text{in}} = \text{Energy}_{\text{work}} + \text{Energy}_{\text{heat}} + \text{Energy}_{\text{stored}}$$

The sum of energy converted to work and heat is defined as metabolism, which includes synthesis of biochemicals, smooth muscle contractions involved in breathing and blood circulation, generation and maintenance of ion gradients the synthesis and catabolism of biochemicals. Although life-sustaining metabolic functions constitute thousands of chemical reactions occurring at the same time throughout the body that cannot readily be individually measured, their sum can be measured as the sum of work and heat energy or, in the absence of any measurable work, just the rate of heat production by the body. This assumes that all cellular events ultimately result in heat production.

Calorimetry, therefore, is defined as the measurement of heat production by the body during oxidative combustion of the macronutrients from food. At the cellular level, each macronutrient is ultimately combined with oxygen to produce carbon dioxide, water, and heat as described by the following general equation:

$$\text{Substrate} + O_2 \rightarrow CO_2 + H_2O + \text{heat}$$

These chemical reactions are virtually the same as those that would be observed if the nutrient were combusted in a flame, except the reaction in the body is an enzymatic process that does not produce a flame. Still, the chemical reaction is the same; O_2 is a necessary reactant, and CO_2 is an end product of reactions that constitute metabolism. The term "indirect" calorimetry is used when heat production is calculated by measuring rates of oxygen consumption (VO_2) and carbon dioxide production (VCO_2) over time. After VO_2 and VCO_2 are measured, the Weir equation (**Table 1**) or similar equations can be used to calculate the rate of heat production, which is commonly called energy expenditure. In addition, the ratio of VCO_2/VO_2, which is termed the respiratory exchange ratio (RER) when measured on a whole-body basis. The RER, in addition to its use in calculating heat production using the Weir equation, provides useful information about which macronutrient (carbohydrate, lipid, or protein) has been oxidized. The metabolism of these macronutrients consumes oxygen and produces carbon dioxide. When used to calculate nutrient utilization, the whole-body RER is referred to as the global respiratory exchange ratio (RER) as it is the oxygen consumed and the carbon dioxide produced by the total of macronutrients metabolized. The first step in calculating the three individual rates of macronutrient oxidation is to calculate protein oxidation. Protein is the most difficult to describe on a chemical basis because a protein is made from a mixture of amino acids and, for each dietary protein, the number and composition of amino acids differ, and this is also influenced by

Table 1 Respiratory gas exchange constants and calculation of energy expenditure.

	VO_2	*VCO_2*	*Heat*	*RER*
Substrate	*L g^{-1}*	*L g^{-1}*	*kcal g^{-1}*	*VCO_2/VO_2*
Carbohydrates[a]	0.829	829	4.2	1.00
Fat	2.019	1.427	9.5	0.71
Protein	0.966	0.794	4.4	0.82
Nitrogen	6.037	4.963	–	0.82

Variable	*Formula*
VO_2 (L min^{-1})	$=V_i\ (FiO_2) - V_e\ (FeO_2)$
VCO_2 (L min^{-1})	$=V_e\ (FeCO_2) - V_i\ (FiO_2)$
$NPVO_2$ (L min^{-1})	$=VO_2 - 6.037Nu$
$NPVCO_2$ (L min^{-1})	$=VCO_2 - 4.963Nu$
Full Weir equation (kcal min^{-1})	$=3.941\ VO_2 + 1.106\ VCO_2 - 2.17Nu$
Modified Weir equation (kcal min^{-1})	$=(3.94 + 1.11\ RER)\ VO_2$ (L min^{-1})

VO_2 = rate of O_2 consumption L min^{-1}; VCO_2 = rate of CO_2 production L min^{-1}, Vi = volume of inspired air L min^{-1}, Ve = volume of expired air L min^{-1}, Fi = fraction of inspired O_2 or CO_2, Fe = fraction of expired air that is O_2 or CO_2, NPV = Volume of non-protein gas L min^{-1}, Nu = urinary nitrogen g min^{-1}.

[a]Starch or glycogen.

Adapted from Jéquier et al. (1987).

variation in the fractions of nitrogen excreted as urea, ammonia, and creatinine. The breakdown of the standard dietary protein, however, can be described by the following chemical reaction:

$$C_{100}H_{159}N_{26}O_{32}S_{0.7} + 104O_2 = 86.6CO_2 + 50.6H_2O + 11.7N_2H_4CO + 1.3NH_4OH + 0.43N_3C_4H_7O + 0.7H_2SO_4 + \text{heat}$$

In this example, 86.6 molecules of CO_2 are produced whereas 104 molecules of O_2 are consumed when one molecule of protein is oxidized (RER = 0.83). Although this RER value is intermediate between carbohydrate and fat, protein is unique among the three energy substrates because it is the only one to contain the element nitrogen. As such, urinary N (g/min) can be assayed to obtain an estimate of protein oxidized by an individual. Because protein is the only macronutrient containing nitrogen, the oxidation of protein can be calculated from the rate of urinary nitrogen excretion. When greater accuracy has desired the sum of urinary nitrogen production plus and change in the amount of urea circulating in the blood. Assuming circulating urea does not change, protein oxidation is calculated as follows:

$$\text{Protein oxidation}_{g/min} = 6.25N_{\text{urine}g/min}$$

After calculating protein oxidation, the VO_2 and VCO_2 due to protein oxidation can be calculated in units of mL/min for each g/min of urinary nitrogen, and these subtracted from the global VO_2 and VCO_2 to yield the nonprotein VO_2 and VCO_2 (Table 1).

To calculate the percentage of carbohydrate versus fat being used to support energy expenditure. For example, one molecule of sugar (glucose) breaks down as follows:

$$C_6H_{12}O_6 + 6O_2 = 6CO_2 + 6H_2O + \text{heat}$$

In this reaction, six molecules of CO_2 are produced and six molecules of O_2 are consumed. Thus, the ratio of CO_2 produced to O_2 consumed (RER) has a value of 1.0. When one molecule of fat (tripalmitin) is broken down completely, the chemical reaction is as follows:

$$C_{57}H_{104}O_6 + 80O_2 = 57CO_2 + 52H_2O + \text{heat}$$

When this molecule of fat is completely oxidized, 57 molecules of CO_2 are produced whereas 80 molecules of O_2 are consumed (RER = 0.71).

Physiologic nonprotein RER values generally range between 0.71 (100% fat oxidation) and 1.00 (100% carbohydrate oxidation). Thus, an RER of 0.85 reflects approximately 50% carbohydrate and 50% fat oxidation supporting energy expenditure.

Urinary N_2 is often not measured and therefore results from indirect calorimetry often use the abbreviated Weir equation to calculate rates of energy expenditure (Table 1), which assumes protein oxidation supports 12% of total energy expenditure.

A healthy individual consuming an average mixed diet of carbohydrates, fat, and protein will usually have an RER between 0.80 and 0.85. For example, assume a healthy male consumes 350 mL O_2 min^{-1} and produces 298 mL CO_2 min^{-1} on average throughout 24 h. That individual will be expending 1.7 kcal min^{-1} (using the abbreviated Weir equation in Table 1). This equals a 24 h energy expenditure of approximately 2448 kcal day^{-1} (1.7 kcal min^{-1} × 1440 min). An RER ≥ 1.00 usually indicates whole body de novo lipogenesis (endogenous synthesis of fatty acids) but can also result from hyperventilation during respiratory gas analysis. High-intensity exercise will also result in an RER above 1.00 as the accumulation of lactic acid in the blood is buffered by the body bicarbonate pool, resulting in additional CO_2 output.

Laboratory methods

The first calorimeters developed in the early 1800s directly measured the rate of body heat production of an animal in a thermally isolated chamber (direct calorimetry). Antoine Lavoisier was one of the first to perform such direct calorimetry. He placed a guinea pig inside a chamber surrounded by ice and then measured the rate of melting of the ice to obtain a measure of the animal's rate of heat production. By burning a small amount of coal equal in mass to the mass of carbon in the CO_2 produced by the guinea pig, he demonstrated that animal respiration was, in fact, combustion of the chemicals provided by food in a manner that is chemically identical to what occurs when macronutrients were burned in a flame. Modern direct calorimeters are complex to construct but are both an accurate and precise method for measuring energy expenditure. The disadvantage is that RER is not measured and thus no information on which substrates are being oxidized. As such most clinic and calorimeter research is performed using indirect calorimeters. A detailed account of the history and theory of calorimetry can be found in the further reading section at the end of the article. This material on laboratory and field indirect calorimetry techniques commonly used in most clinical and research settings is, however, briefly described below.

Whole body indirect calorimetry

The advent of indirect calorimetry was a significant event in the history of animal and human nutrition. In whole-body indirect calorimetry, a person is housed in a small, sealed room with controlled ventilation, often referred to as a metabolic chamber. It is a small chamber, but similar to a studio (one room) living environment and hence a little more representative measurement

of free-living energy expenditure. The chambers in use today are typically furnished with a bed, a chair, desk or table, toilet, sink, TV/computer, telephone, and an airlock used for passing food and physiologic specimens between subjects and study personnel with minimal leakage of air into or out of the chamber.

Two common types of whole-body indirect calorimetry systems have been constructed. One is closed-circuit indirect calorimetry involves the recirculation of the same air through the sealed chamber. The recirculated air is kept breathable by removing and measuring CO_2 produced by the subject and replacing it with O_2 to be consumed by the subject. As the subject consumes O_2, a sensor detects the decrease in volume, and a signal is sent to an external source to release calibrated pulses of O_2 back into the system to restore the O_2 concentration to its original level. The rate of O_2 consumption by the subject is measured by recording the amount of O_2 that is added to the air during recirculation. The CO_2 produced by the subject is removed from the recirculated air by an absorber and is measured from the increased weight of the absorber (**Fig. 1**).

The more common type is the open-circuit indirect calorimetry involves a system in which both ends of the breathing system are open to the atmosphere (**Fig. 2**). Outside air is drawn into the chamber at a constant flow rate while maintaining a slight negative pressure inside the chamber. At the same time, the composition of this outside air is measured by CO_2 and O_2 gas analyzers. Well-mixed expired chamber room air is drawn out at a constant flow rate and also analyzed for O_2 and CO_2 composition. The difference in O_2 and CO_2 composition of the outside air and chamber room air is used to calculate the energy expenditure and macronutrient oxidation of the subject using the equations in **Table 1**.

A limitation of room calorimeters is that the volume of the room is large, often over 10,000 L, and that the flow rate of air through the chamber is usually between 50 and 100 L/min. Thus, the room requires many hours to respond to changes in the rate of the subject's gas exchange rates. One means to obviate this limitation is to measure not only the difference in O_2 and CO_2 as it enters and leaves the chamber but to add to this the changes in the volumes of O_2 and CO_2 in the chamber itself—e.g., VO_2 = flow rate $(Ve(feO_2) - Vi(fiO_2) + (\text{chamber volume})*((fO_2)_{t=i} - (fO_2)_{t=0})$. During the 1990s, the precision of the O_2 and CO_2 analyzers was modest, and minute to minute variation in the measured gas concentrations only allowed investigators to measure energy expenditure over a period of one to 4 h. Improvements in the performance of the gas analyzers since then have improved the time resolution of modern room indirect calorimeters is such that energy expenditure can be measured over intervals as short as minutes and thus it is whole room indirect calorimeters to measure individual bouts of light to moderate physical activity, the energy costs of daily life (dressing, food preparation, playing video games), or the changes in energy expenditure when nutrients are ingested or infused.

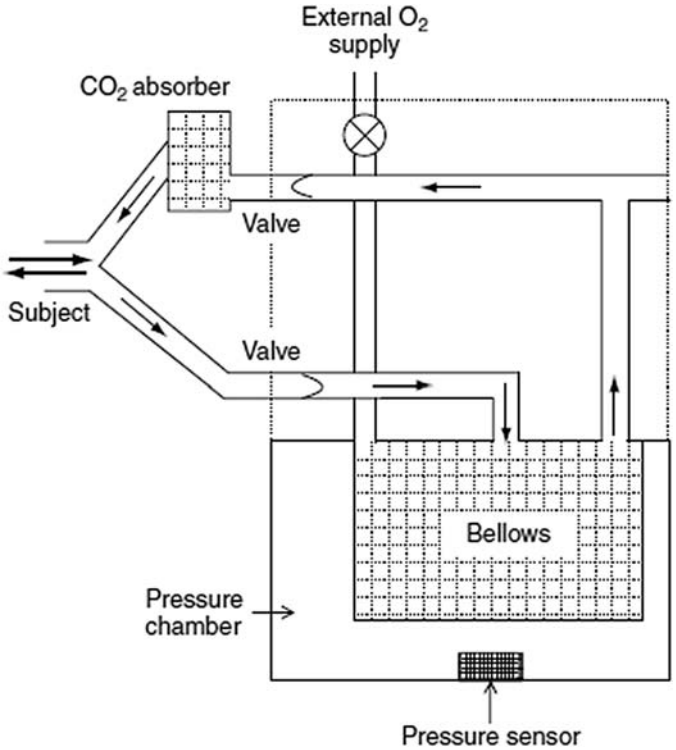


Fig. 1 Closed-circuit metabolic chamber in which the subject's oxygen consumption is measured to calculate the corresponding energy expenditure. The change in volume of air in the system is constantly monitored by the sensors and a measured quantity of oxygen is added back to the system. Carbon dioxide is taken out of the recirculated air by a CO_2 absorber.

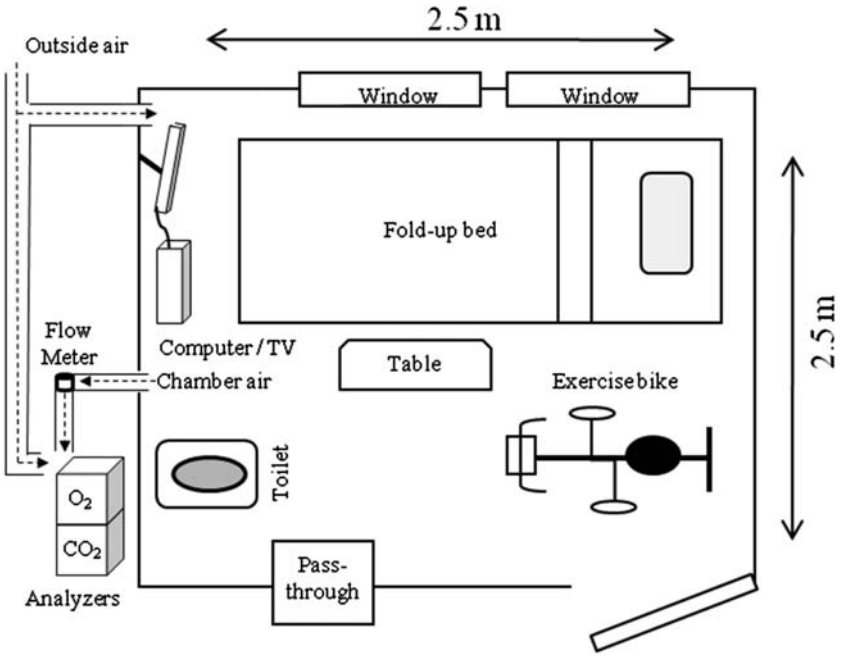


Fig. 2 An open system respiration chamber is illustrated above. The 2.5 m^2 area ids large enough to contain a bed, small table, wall mounted computer/TV station, toilet, and stationary exercise bike (or treadmill). A door is provided for access, and windows to provide contact with the outside world. Air is continuously drawn into the sealed chamber. The volume of air is measured along with the concentrations of oxygen and carbon dioxide entering and exiting the chamber.

Metabolic carts

The concept of indirect calorimetry is often applied on a scale smaller than a whole room in many research and clinic settings. Many smaller indirect calorimeters have been developed over the years. The instrumentation used for each varies in complexity and the degree to which they restrict the subject's movement.

Metabolic cart is a common name for a respiratory gas analyzer that has been made small enough to be placed on a cart with wheels so that it can be rolled to different locations within a building. Two designs are generally available: the ventilated hood (canopy) and the mouthpiece systems. The ventilator hood system is an open-circuit indirect calorimeter that usually consists of a pliable plastic or rigid Perspex hood placed over the subject's head with latex or thin plastic apron providing a rough seal around the neck or chest. These allow air to be drawn across a subject's face while in a reclining or lying position. For longer-term measurements, ventilated plastic tents are available that cover all or part of the patient's bed. Because these hoods operate on a suction principle, a tight seal of the hood is not required. For field measurements, whole body transparent plastic ventilated boxes have been used successfully in infants. Many of the ventilatory hoods are constructed by researchers from the components according to the requirements of their study. The components include a pump, a flowmeter, and a means of regulating the airflow. Samples of the air drawn from the hood can be directed to gas analyzers, which are usually connected in series to the hood. Respiratory gas exchange is calculated from the difference in O_2 and CO_2 concentration between the air entering and exiting the hood times the known controlled airflow (Figs. 3A, B).

Instruments have been developed to operate in adult and pediatric modes and differ with respect to flow rates and internal volume because metabolic rates, and corresponding gas exchange rates, of children, are up to five smaller than that of an adult. With adult and pediatric systems, the expired air enters a mixing chamber within the instrument to eliminate concentration variation resulting from the normal inspiration and expiration phases associated with breathing. The mixed sample enters O_2 and CO_2 analyzers and the concentration differences between the air entering and exiting the hood or canopy. For state-of-the-art instruments, the data are input into a microprocessor providing a minute-by-minute calculation of the O_2 consumption, CO_2 production, RER, and energy expenditure. These instruments are generally used for measurements of subjects at rest as part of nutritional studies of energy expenditure and macronutrient utilization. These units can also be connected to mechanical ventilators for use in hospitalized patients.

When measured in the absence of movement or skeletal muscle contraction, the rate of energy expenditure is termed the resting metabolic rate (RMR). Time since the last consumption of nutrients, ambient temperature, and arousal are controlled and recorded. The measurement of RMR under near basal conditions requires an overnight fast of about 15 h, absence of vigorous exercise in previous 24 h, minimal physical activity of any intensity in the previous 30 min, thermal neutrality (20–27 °C for individuals without excess body fat), absence of noise or talking and dim lighting and referred to near basal, fasting or in older literature basal RMR.

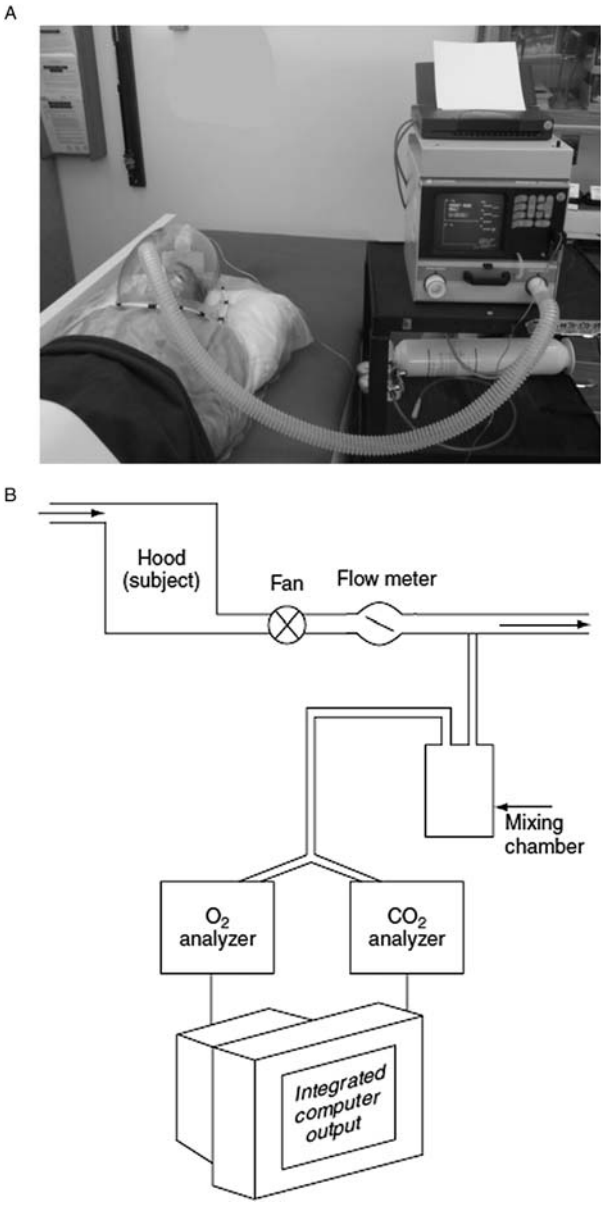


Fig. 3 (A, B) Ventilatory hood system showing a hood that is placed on the subject's head, a mixing chamber, and O_2 and CO_2 analyzers. A fan maintains a slight negative pressure in the hood to pull room air into the chamber and prevent the escape of expired air from the system. The expired air is mixed in the mixing chamber and is analyzed for oxygen and carbon dioxide by the respective analyzers. Results are calculated by the computer.

Mouthpiece systems are similar to ventilated hood systems in principle, but instead of placing a hood over the subject's head, the subject wears a mouthpiece connected to the analyzer and nose clips to prevent breathing through the nose. The mouthpiece is connected to a one-way valve system that allows the subject to breathe in atmospheric air while directing the exhaled air to the gas analysis system. The expired breath is again subjected to analysis of O_2 and CO_2 concentration, but rather than passing the breath through a mixing chamber to smooth out the changes in a concentration gradient of these gases from the start to the end of an exhalation, the concentration profile is measured in real-time along with the rate of gas flow from the exhalation. Again, the data are logged into a microprocessor for calculation of O_2 consumption and CO_2 production, but in this case, the calculation is performed on a breath-by-breath basis. Results are averaged over time, usually provided as minute-by-minute averages of O_2 consumption, CO_2 output, and the rate of energy expenditure. The mouthpiece systems are generally used for studies of gas exchange and energy metabolism during exercise. The mouthpiece and nose clip used with some of the instruments make long-time measurements highly cumbersome. Also, breathing through the mouthpiece often causes untrained subjects to involuntarily hyperventilate leading to inappropriate O_2 and CO_2 rates. It is also often difficult with mask systems to obtain an airtight seal without excessive pressure at the site of contact with the mask and face. The use of mouthpiece systems is often employed for

measures made when the subject is exercising as part of an exercise tolerance test to diagnose exercise limitations in patients with chronic disease or to assess safe limits of exercise in patients before they begin rehabilitation after an adverse metabolic event. A measure of maximal O_2 consumption during exercise is used to measure cardiovascular fitness and training in athletes.

Different types of metabolic carts or monitors are available that are designed for various applications ranging from nutrition to exercise science. Most have built-in gas analyzers and data processing computers, making them highly user-friendly, easy-to-use tools for the measurement of energy metabolism. They generally provide accurate and reliable data but do require periodic calibration. Ventilated hood systems often use a combination of gases with known concentrations and weighed ethanol or methanol burns for such calibration, whereas breath-by-breath systems use a combination of large volumetric syringes and gases of known O_2 and CO_2 concentration.

Field methods

As for whole-body indirect calorimetry, ambulatory and portable systems measure the respiratory gas exchange with the VO_2 and VCO_2 measurements. Ambulatory methods and less refined laboratory methods often dispense with the measurement of CO_2 to avoid the need for two gas analyzers. The error incurred by assuming a CO_2 production rate is several percentage points, which many researchers and clinicians are prepared to compromise on. When only O_2 consumption is measured, however, it is not possible to compute macronutrient-specific oxidation rates. The accuracy of ambulatory and portable methods is generally between 4% and −2%. Field methods involve the collection of expired air over a fixed period of time as in the Douglas bag or small online analysis systems that sample inspired and expired air through a mouthpiece.

Field methods discussed below are used to measure the energy expenditure of individuals performing a specific task such as agricultural jobs or other manual labor for the estimation of the energy requirements or costs of the jobs being performed so that appropriate supply of food is provided, and the costs of the labor can be calculated. This may be part of an ergometric study, or for planning the logic requirements for an exploration or military project.

Douglas bag

The Douglas bag method is a classic example of how expired air can be collected to measure energy expenditure in the field during both rest and physical activity. It consists of a gas-impermeable bag with a capacity of 100 L. A classic Douglas bag is made of either a rubber sheeting cemented between two layers of canvas or plastic material lined by PVC or aluminum with welded seams. The rubber bags are subject to slow leakage of CO_2 by diffusion, which is unavoidable, but PVC and metalized bags reduce this loss. If the bags are filled to their capacity and analyzed with 20 min of collection, the effects of diffusion are minor. The subject wears a nose clip and mouthpiece or a face mask. Outside air or its equivalent is inhaled through a one-way valve and exhaled into a Douglas bag for a precise period of time. It is important that the mouthpiece and connecting tubing provide minimal resistance to airflow, or the cost of breathing will increase the energy expenditure. The volume of air collected in the bag or tank is measured and a sample of exhaled air is obtained to measure the O_2 and CO_2 concentrations using gas analyzers. The volume of oxygen consumed, and carbon dioxide produced are calculated by analyzing the gas from the Douglas bag for the precise time period during which it was collected. This method is relatively simple and inexpensive yet gives reliable results. It is suitable only for short durations of field measurement, because wearing the mask and nose clip for the whole duration of the study may be cumbersome, interfere with daily activities, and is socially undesirable to the subject. Spirometers were used in the past for the measurement of the volume of the respired air. With the advent of continuous flow electronic analyzers and superior gas flowmeters, spirometers are now rarely used. Ambulatory methods also consist of a mouthpiece incorporating light action-sensitive but robust one-way gas valves, corrugated tubes, and three-way taps. The volume of air and the concentrations of O_2 and CO_2 in the expired air are measured using O_2 and CO_2 gas analyzers. These small analyzers have replaced the systems used in the past such as the Haldane system or micro-Scholander chemical gas analyzers, which used reagents to absorb CO_2 and O_2, with the weight of absorbents measured before and after use.

Max Plank/Kofranyi–Michaels respirometer

A Max Plank respiration gas meter is a small, compact, and lightweight backpack-mounted respirometer (Fig. 4). It combines a gas volume meter and a sampling device for continuous sampling of each breath of expired air. The Max Plank respirometer consists of a dry, bellow-type gas meter for measuring the total volume of expired air during activity. The subject breathes through a low-resistance valve and the expired volume is monitored. A measured quantity of expired air is removed continuously (0.3% or 0.6%) by an aliquoting device to be sent to a small butyl rubber bag. This rubber sampling bag can be connected directly to the oxygen analyzer, eliminating the need for the transfer of samples to gastight syringes for analysis. The respirometer is suitable for flow rates between 15 and 50 L min^{-1} or for a period of 110 min on a slow flow rate and 55 min on a faster rate. It is smaller, more compact, and lighter than the Douglas bag apparatus.

In addition, several miniaturized systems have been developed that have been made possible by miniaturization or the development of novel new chemical sensors. Such devices help to extend the use of indirect calorimetry to a field site requires because they are lighter and easily carried to remote locations for measurement of exercise load during mountain climbing, endurance athletes, and laborious tasks in non-industrialized countries. Such devices require validation against established indirect calorimeters to ensure that the gas collection attachments do not restrict breathing and cause hyper- or hypoventilation, are not subject to

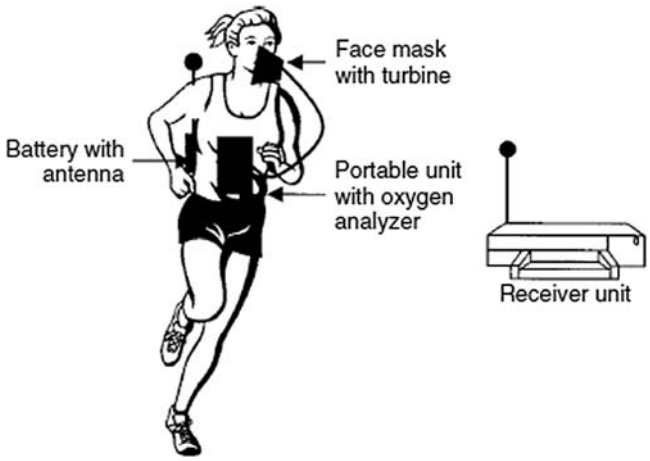


Fig. 4 A Max Plank type respirometer is illustrated above. The subject is fitted with a mask covering the mouth and nose. A pair of one-way values on the mask allows the subject to inhale and directs the exhaled breath into the flowmeter and oxygen analyzer. The flow rate and oxygen concentration data are transmitted to a data logger for storage.

leakage during breath collection, meet the requirements for accuracy and precision of the application, and are rugged enough to survive transport and use in remote locations.

Calibration of metabolic carts and portable indirect calorimeters remains important to obtaining accurate measurements of RMR and RER. Both the flow and gas sensors that comprise the components of the O_2 and CO_2 must be periodically confirmed or adjusted. With some instruments, the two can be done at the same time by measurement of the area under the curve on gas concentration and time by measuring the amount of O_2 consumed or CO_2 recovered from the utilization of a known quantity of material. An example of this the combustion of a known quantity of alcohol, which will consume a known volume of O_2 and produce a known volume of CO_2 based on the stoichiometry of the chemical reaction. This does require that the analyzers are impervious to damage from incomplete combustion and that combustion can be performed without issues arising from the open flame. The frequency of calibration depends on the specific instrument being calibrated and its susceptibility to the drift of its response to the gas being analyzed. Calibration can also be performed using a combination of the infusion of gas of known O_2 and CO_2 concentration and calibrated flow generator such as a 1 or 3 L calibrated syringe or a calibrated flow sensor. Gases of known O_2 and CO_2 concentration and flow calibrators are available from many vendors. It should also be remembered that gases must be dried because when water vapor is added to expire air, it will dilute the O_2 and CO_2 and create a measurement artifact. Lastly, the volumes of O_2 and CO_2 gas need to be corrected to standard temperature and pressure for accurate calculation of RMR and RER.

Tracer methods of indirect calorimetry

A third category of techniques has gained popularity among investigators during the past two decades. These techniques provide a measure of CO_2 production through the use of dilution techniques using isotopic tracers. These are the labeled bicarbonate method and doubly labeled water. These methods do not measure O_2 consumption, and the RER must be assumed or estimated in the calculation of energy expenditure. When used for periods of days, the error in estimating RER is small as the RER is usually close to the food quotient (ratio of CO_2 produced to O_2 consumed if all of the food is completely metabolized). The error incurred by any difference between true long-term RER and assumed or estimated RER is 0.9% for each 0.01 RER unit and such differences have a standard deviation of 0.03 units except under conditions of large (>500 kcal/d) energy imbalance or low (<10% of energy) carbohydrate or fat diets. These techniques have been used to study the causes and treatment of obesity, the establishment of energy requirements, or the prevention of malnutrition.

Labeled bicarbonate

A constant infusion-labeled bicarbonate method is useful in estimating the net CO_2 production and hence energy expenditure in animals and humans. This method is based on an isotopic dilution technique whereby the administered label is diluted by the CO_2 produced endogenously by the body. The extent of this isotope dilution is used to measure the rate of CO_2 production and then, to estimate the energy expenditure of the individual. A micro-infusion of ^{13}C or ^{14}C labeled bicarbonate is given to an individual and the specific activity or enrichment of his or her physiological fluids, especially breath or urine, are measured to estimate the rate of label elimination and hence the rate of endogenous CO_2. Thus, variation in the endogenous CO_2 production rate will be reflected in the dilution of the body pool and consequently in the breath samples. These measurements are subject to the effects of label

sequestration over shorter periods. Sequestration refers to trapping, or fixation, of the label in tissues that utilize bicarbonate/CO_2 for their metabolic functions. A shorter duration of collection of breath samples requires a correction for the fraction of the label that is sequestered. This is based on the assumption that similar amounts of the label are sequestered in various individuals. When breath samples are collected over longer durations, the sequestration is often assumed to be negligible.

Some investigators have used a bolus bicarbonate administration rather than the continuous infusion. These investigators measured the rate at which the label concentration decreases with time as a measure of CO_2 turnover and the initial concentration as a measure of the body's bicarbonate pool size. Taken together, these provided a measure of energy expenditure during a short period of constant physical activity.

Doubly labeled water (DLW)

The doubly labeled water (DLW) technique, discussed in detail in the next article, uses isotope dilution wherein deuterium and heavy oxygen-labeled water (DLW) are given to individuals and timed urine samples are collected to measure the elimination rates of ^{2}H and ^{18}O in the urine. The ^{2}H label from DLW mixes with the body water and is eliminated as water in the urine. Similarly, the ^{18}O label from DLW is eliminated as water, but it exchanges with the oxygen in circulating bicarbonate and is also eliminated in the breath as CO_2. The difference in turnover rates of isotopic ^{2}H and ^{18}O-labeled water is proportional to CO_2 production. DLW is technically a method of indirect calorimetry even though respiratory gas exchange measurements across the lung are not made. Energy expenditure, oxygen consumption, water intake, and metabolic water production can be calculated using standard indirect calorimetry equations with an estimated RER. Unlike the majority of the other methods, the DLW method provides a measure of average energy expended over a period of 3–21 days, which provides a measure of habitual, free-living energy expenditure.

Since the predictive equations of energy requirement are subject to inaccuracies, more reliable methodologies are fundamental for investigating metabolism in conditions of health and disease. Due to advances in technology, portability, handling, and affordability, IC has been applied to optimize nutrition therapy through the individualization of nutrition support, maximizing its benefits, and thereby, improving the treatment outcome (Haugen et al., 2007; Oshima et al., 2017). This is readily achievable because IC can provide real-time energy expenditure data, and when combined with measures of urinary nitrogen excretion, also assesses substrate utilization and when combined with energy intake an assessment of the patients' energy balance (Lam and Ravussin, 2017).

One of the most important applications of the IC is to assess the near basal RMR, which comprises between 50% of total daily energy expenditure (TEE) in highly active individuals to 85% of TEE in hospitalized individuals who are restricted to bed. Among healthy individuals, measurement of near basal RMR is used to describe an individual's energy budget in terms of the fraction of daily energy expenditure used to support life functions, process foods, support physical activity, calculate physical activity energy expenditure by difference from TEE, with or without adjustment for the energy expended in the processing of macronutrients from meals (thermic effect of meals).

The time course of the changes in RMR among hospitalized individuals is such that energy requirements can change daily and normalize with a week or less after hospitalization. Therefore, monitoring the metabolism, in subjects in the Intensive Care Unit (ICU), has been shown crucial for the recovery of the patient (McClave et al., 2013). Conditions that significantly modifying RMR, the difficulty of maintaining or restoring body weight due to inadequate nutrition support based on predicted energy requirements, and acute disease associated with intense changes of metabolic stress level are strongly benefited with the RMR measurement (Delsoglio et al., 2019). Other chronic conditions as obesity, cancer, chronic kidney diseases, chronic obstructive pulmonary disease, diabetes, neuromuscular degenerative diseases, and anorexia, also significantly affect the RMR, therefore, an individual nutrition therapy benefits from measuring RMR (Delsoglio et al., 2019). Benefits include the prevention of malnutrition or overnutrition which can worsen an individual's health during treatment.

Interpretation of indirect calorimetry results

Even being an easy handling device to use in health or critical illness conditions, misinterpretation of IC results can lead to erroneous prescriptions. Less reliable IC measurements are observed under unstable clinical conditions, therefore, the RMR should be assessed within the first 24–48 h to reflect the metabolic evolution caused by the illness. Standard values of VO_2, VCO_2 are reported to be 120 mL min^{-1} m^{-2} and 100 mL min^{-1} m^{-2}, respectively. The RMR varies between individuals and can be between 25 and 40 kg^{-1} day^{-1}. During an individual 30 min measurement, RMR measurements can vary from minute to minute and the coefficient of variation of up to 10% is acceptable. The measurement of RER also has diagnostic value. A normal range is ~0.7–1, while a value over 1.0 indicates overnutrition which increases the work needed for breathing and can lead to fat accumulation in the liver, and a value below 0.8 indicates undernutrition and can result in reductions in immune function. Other illness conditions as hypermetabolism, hyperventilation, metabolic acidosis, excessive carbohydrate intake, sepsis, and hemodialysis can cause high values of VCO_2 and/or RER. On the other hand, hypometabolism, hypoventilation, starvation, fasting, ketosis, hypothermia, and paralysis can induce low values of VCO_2 and/or RER (Mtaweh et al., 2018). Respiratory gas exchange constants and equations to estimate the energy expenditure are summarized in **Table 1**.

Clinical cases

Analyzing clinical cases provides insight into the understanding of the proper use of an IC device, the interpretation of IC results, and the decision-making process to apply its results for a proper prescription of nutrition support. Here we discuss two clinical cases adapted from Achamrah et al. (2021), where more details about IC use in clinical care can be found.

Case 1: A 25-years-old man weighing 75 kg was admitted to Intensive Care Unit due to a traumatic brain injury. He was on mechanical ventilation, with a FiO_2 of 60% and positive end expiratory pressure (PEEP) of 8 cm H_2O, with intracranial pressure monitored by ventricular drain and cerebrospinal fluid drainage. He was under sedation with propofol and pain controlled with morphine, without vasoactive agents. In the unit, continuous enteral nutrition of 1500 kcal day^{-1} was initiated. Measurements of IC were taken on the second day after admission for 30 min, on which he was showing normal body temperature, but was in an agitated state. The IC results were: $VO_2 = 490$ mL min^{-1}; $VCO_2 = 350$ mL min^{-1}; RER = 0.75; RMR = 3560 kcal day^{-1}; The minute-to-minute coefficient of variation (first 5 min) was 14% for VO_2 and VCO_2. Interpretation: The coefficient of variation is out of acceptable limits for VO_2 and VCO_2 (>10%), which could be caused by the physical agitation, therefore, this IC measurement is not reliable. Additionally, the measurement was performed in the early phase of the acute illness, on which the energy needs are covered by endogenous energy stores, therefore, to avoid overfeeding, the use of a weight-based equation (20–25 kcal kg^{-1} day^{-1}, with an energy target of 1875 kcal day^{-1}) for the first 5 days of ICU admission would be prescribed and adjusted after the 3–4th day of IUC admission, based on a new IC measurement.

Case 2: A 21 years-old woman with a BMI of 13.3 kg m^{-2}, FM (2 kg, 6.3%), and FFM (30 kg, 93.7%) measured by DEXA, was assisted at the Nutrition Outpatients Clinic. The RMR measurement was conducted using a canopy mode, in a quiet room, with a stable temperature and humidity, in the morning and after an overnight fast. The IC results were: $VO_2 = 119$ mL min^{-1}; $VCO_2 = 95$ mL min^{-1}; RER = 0.85; RMR = 900 kcal day^{-1}; Coefficient variation (first 5 min) = 5% for VO_2 and VCO_2. Interpretation: the IC measurements were collected in an appropriate environment and the coefficient variation was within acceptable limits for VO_2 and VCO_2 (<10%). A blood urea nitrogen (BUN − 4 mg/dL) and blood creatinine (0.65 mg/dL) were measured to assess the BUN to creatinine ratio (BCR = 6:1) and support an overview of the protein metabolism. Interpretations: the IC measurements were collected in an appropriate environment and the coefficient variation was within acceptable limits for VO_2 and VCO_2 (<10%). The RER value and a value of BCR below the normal range (10:1 and 20:1) indicates predominant use of protein as a fuel source, thus a need for increased energy intake from carbohydrates and lipids sources to reduce protein catabolism. Protein turnover and patient needs can be assessed by measuring 24 h urinary nitrogen excretion or estimated by measuring urinary urea excretion, which represents 85% of total nitrogen in humans. Considering the extreme malnutrition, more than 900 kcal day^{-1} should be prescribed to achieve a positive energy balance and weight gain, however, to avoid the refeeding syndrome the energy prescription should start at 10 kcal kg^{-1} day^{-1}.

Conclusion

Max Kleiber's 20th century, seminal textbook is entitled *The Fire of Life* in recognition of the discoveries made by Antoine Lavoisier. The measurement of the intensity of that metabolic fire using indirect calorimetry provided and continues to provide great insight into the metabolic processes that form human and animal metabolism. The metabolic rate assessed by measuring oxygen and carbon dioxide exchange is the largest component of daily energy expenditure and hence the major (50–85%) contributor to an individual's energy requirement. The ratio of carbon dioxide production to oxygen consumption or RER indicates the relative contributions of macronutrients being combusted and further illuminate information on metabolism and is valuable in understanding physical performance as well as alterations in metabolism that underlie various acute disease states.

References

Achamrah, N., Delsoglio, M., De Waele, E., et al., 2021. Indirect calorimetry: the 6 main issues. Clin. Nutr. 40, 4–14.
Delsoglio, M., Achamrah, N., Berger, M.M., Pichard, C., 2019. Indirect calorimetry in clinical practice. J. Clin. Med. 8, 1387. https://doi.org/10.3390/jcm8091387.
Haugen, H.A., Chan, L.-N., Li, F., 2007. Indirect calorimetry: a practical guide for clinicians. Nutr. Clin. Pract. 22 (4), 377–388.
Jéquier, E., Achesson, K., Schuttz, Y., 1987. Assessment of energy expenditure and fuel utilization in man. Ann. Rev. Nutr. 7, 187–208.
Lam, Y.Y., Ravussin, E., 2017. Indirect calorimetry: an indispensable tool to understand and predict obesity. Eur. J. Clin. Nutr. 71, 318–322.
McClave, S.A., Martindale, R.G., Kiraly, L., 2013. The use of indirect calorimetry in the intensive care unit. Curr. Opin. Clin. Nutr. Metab. Care 16, 202–208.
Mtaweh, H., Tuira, L., Floh, A.A., Parshuram, C.S., 2018. Indirect calorimetry: history, technology, and application. Front Pediatr. 6, 257. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6157446/.
Oshima, T., Ragusa, M., Graf, S., et al., 2017. Methods to validate the accuracy of an indirect calorimeter in the in-vitro setting. Clin. Nutr. ESPEN 22, 71–75.

Microbiota of the intestine: Biology and physiological functions

Julio Plaza-Diaz[a,b,c], **Francisco Javier Ruiz-Ojeda**[b,c,d,e], **and Angel Gil**[b,c,d,f], [a] Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada; [b] Department of Biochemistry and Molecular Biology II, School of Pharmacy, University of Granada, Granada, Spain; [c] Instituto de Investigación Biosanitaria IBS.GRANADA, Complejo Hospitalario Universitario de Granada, Granada, Spain; [d] Institute of Nutrition and Food Technology "José Mataix", Center of Biomedical Research, University of Granada, Granada, Spain; [e] RG Adipocytes and Metabolism, Institute for Diabetes and Obesity, Helmholtz Diabetes Center at Helmholtz Center Munich, Munich, Germany; and [f] CIBEROBN (CIBER Physiopathology of Obesity and Nutrition), Instituto de Salud Carlos III, Madrid, Spain

© 2023 Elsevier Ltd. All rights reserved.

Introduction 161
Human microbiota of the gastrointestinal tract: biology and biogeography 162
Microbiota of the intestine: physiological functions 162
Gut microbiome and development of disease 162
Gut microbiota and obesity 163
Gut microbiota and type 2 diabetes 165
Gut microbiota and cardiovascular disease 165
Dysbiosis of gut microbiota in NAFLD 165
Diet and gut microbiota 166
Personalized nutrition and gut microbiota changes 166
Conclusion 167
References 167

Key points

- To describe the presented microbiota in the intestine
- To understand the biology of the microbiota
- To know the physiological functions of the microbiota
- To mention some aspects in personalized nutrition that are related to the gut microbiota

Introduction

The microbiome is defined as the collective genomes of the microorganisms in a particular environment, and microbiota is the community of microorganisms themselves. Approximately 100 trillion microorganisms (most of the bacteria, viruses, fungi, and protozoa) exist in the human gastrointestinal tract; the gut microbiome is now best thought of as a virtual organ of the body. The human genome consists of about 23,000 genes. In contrast, the microbiome encodes over three million genes producing thousands of metabolites, which replace many of the functions of the host, consequently influencing the host's fitness, phenotype, and health (Valdes et al., 2018).

The gut microbiota acts as a barrier against harmful microbes using competition for nutrients and ecological binding site occupancies, and the production of antimicrobial substances. Numerous antimicrobial compounds, such as defensins, cathelicidins, and C-type lectins, are produced by eukaryotic cells in the gastrointestinal tract. The presence of commensal bacteria or their structural components, as well as the presence of products of bacterial metabolism, can induce the expression and activation of these antimicrobial substances, contributing to host protection against invading pathogens and preventing the overgrowth of the commensals themselves. Induction can be mediated through various signaling pathways (Kelly et al., 2017).

Gram-positive anaerobic fecal isolates have been reported to exhibit a greater inhibitory effect on the growth of enteric pathogens in vitro than Gram-negative anaerobic isolates. The competitive activity between commensal microbes has been shown to vary between individuals, as well as from different time points in the same individuals, highlighting the interindividual variations of the gut microbiome and its dynamic fluctuations over time (Kelly et al., 2017).

Most human pathogens enter the body across a mucosal surface, e.g., in the intestine. Indeed, the intestinal immune system is the main and most complex part of the immune system (Gomez-Llorente et al., 2010). In addition to its continuous contact with dietary and environmental antigens, the adult human intestine is home to a huge number of commensal bacteria. How the host distinguishes between commensal and non-pathogenic bacteria is still not well understood (Gomez-Llorente et al., 2010).

The immune system can be classified into innate and adaptive systems. The adaptive immune response depends on B- and T-lymphocytes, which are specific for particular antigens. In contrast, the innate immune system responds to common structures, called pathogen-associated molecular patterns (PAMP), shared by the vast majority of pathogens. The primary response to

pathogens is triggered by pattern recognition receptors (PRR), which bind PAMP. PRR comprises Toll-like receptors (TLR), nucleotide-binding oligomerization domains, adhesion molecules, and lectins (Gomez-Llorente et al., 2010; Valdes et al., 2018). The present article aims to describe: (1) The biology and biogeography of the gut microbiota with detailed information about physiological functions. (2) The gut microbiota metabolites and their effects on the host; and (3) Gut microbiota influences the development of several diseases and personalized nutrition.

Human microbiota of the gastrointestinal tract: biology and biogeography

Microbiota composition in the gastrointestinal tract reflects the physiological properties in a given region and is stratified on both a transverse and longitudinal axis. The density and composition of the microbiota are affected by chemical, nutritional and immunological gradients along the gut. There are typically high levels of acids, oxygen and antimicrobials, and a short transit time in the small intestine. These properties limit bacterial growth, such that only rapidly growing, facultative anaerobes with the ability to adhere to epithelial/mucus are thought to survive. In mice, the small-intestine microbial community is largely dominated by *Lactobacillaceae*. In contrast, colonic conditions support a dense and diverse community of bacteria, main anaerobes with the ability to utilize complex carbohydrates, which are undigested in the small intestine. In the colon *Prevotellaceae*, *Lachnospiraceae* and *Rikenellaceae* have been reported to dominate (Thursby and Juge, 2017).

In contrast with the differing microbiota composition between varying gastrointestinal organs, the microbiota of different colorectal mucosal regions within the same individual is spatially conserved in terms of composition and diversity. This feature is apparent even during periods of localized inflammation. On the other hand, the fecal/luminal and mucosal compositions are significantly different. For example, the abundance of the phylum *Bacteroidetes* appears to be higher in fecal/luminal samples than in the mucosa. In contrast, the phylum *Firmicutes*, specifically *Clostridium* cluster XIVa, are enriched in the mucus layer compared with the lumen. Interestingly, experiments in mice colonized with a diverse specific pathogen-free microbiota show that the outer mucus of the large intestine forms a unique microbial niche and that bacterial species present in the mucus exhibit differential proliferation and resource utilization compared with the same species in the intestinal lumen. These observations highlight the need for careful consideration in choosing a sampling method when analyzing the microbiota composition (Thursby and Juge, 2017). Inter-individual differences in the species and subspecies arrangement are proposed to outweigh differences in the community arrangement within an individual. Suggestions have been made of the presence of a "core microbiota," proposed to be a set of the same abundant organisms present in all individuals. However, more similarity can be observed in the repertoire of microbial genes present between individuals than the taxonomic profile, suggesting that the core microbiota may be better defined at a functional rather than organism level (Thursby and Juge, 2017).

Microbiota of the intestine: physiological functions

Alteration of the microbial structure and function (dysbiosis) is associated with the pathogenesis of various disorders. Fermentation is the process by which anaerobic bacteria (*Firmicutes* and *Bacteroidetes*) break down indigestible carbohydrates to lactate, short-chain fatty acids (SCFAs; acetate, propionate, and butyrate), and other metabolites collaborating with species specialized in oligosaccharide fermentation (e.g., bifidobacteria). Butyrate and propionate can regulate intestinal physiology and immune function, while acetate acts as a substrate for lipogenesis and gluconeogenesis. The gut microbiota regulates immune homeostasis via the induction of regulatory T cells and Th17 cells. In addition, butyrate has strong anti-inflammatory effects, possibly through the inhibition of histone deacetylase activity. Metabolic products generated by the gut microbiota, such as SCFAs, gamma-aminobutyric acid (GABA), tryptophan, serotonin, and catecholamine, transmit signals to resident cells in the gut (Andoh, 2016).

Furthermore, gut microbiota might interfere in nutrient sensing and signaling from the gut to the brain, where the information is processed to control energy homeostasis. This gut-microbiota-brain crosstalk or axis is mediated by different metabolites such as SCFAs, secondary bile acids, or amino acid-derived metabolites. Besides, some bacterial components participate in this gut-brain crosstalk influencing metabolic homeostasis (Romani-Perez et al., 2021).

An overwhelming number of reports demonstrate that disruptions in gut microbiome taxonomy and functional potential relate to numerous pathological phenotypes. The majority of studies in humans and animals are observational and lack experimental mechanistic data. However, as revealed by the genetic repertoire of the human gut microbiome, the trillions of commensal or mutualistic bacteria and *Archaea* are an immense chemical factory that can synthesize a multitude of compounds needed for their existence and survival with their host.

Gut microbiome and development of disease

The microbiome is the collection of all intestinal microbial genes from bacteria, archaea, bacteriophages, eukaryotic viruses, and fungi, which are living on human surfaces and in all body cavities (**Fig. 1**). Most of these microorganisms are influenced by the mode of birth, lifestyle, infant feeding, medication, and host genetics. The gut microbiome has significant roles in training host immunity, regulating gut endocrine function, digesting food, and neurological signaling, modifying drug action and metabolism,

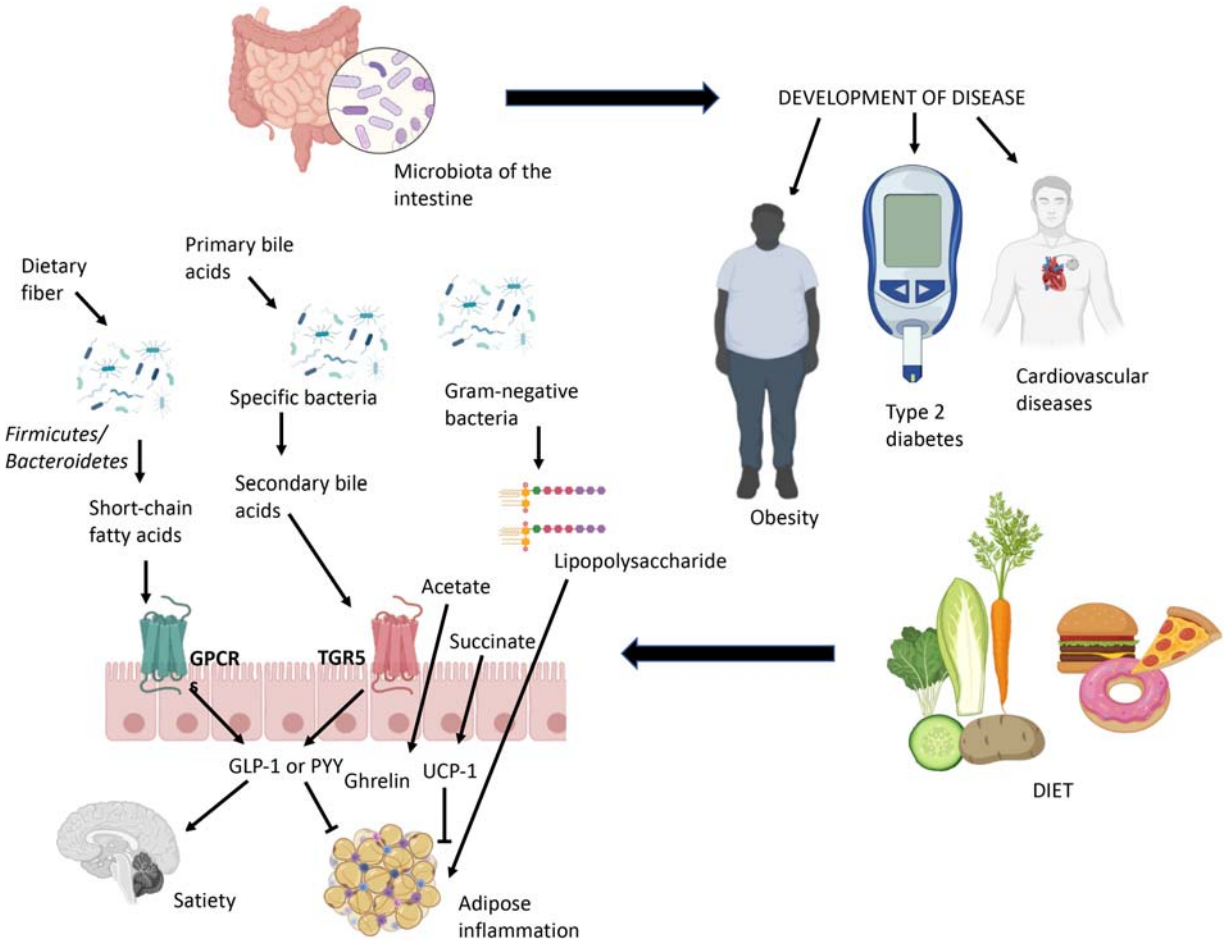


Fig. 1 Microbiota of the intestine: physiological functions. Abbreviations: GLP-1, Glucagon-like peptide 1; GPCRs, G-protein-coupled receptors; PYY, Peptide YY; UCP-1, Uncoupling protein 1; TGR5, G-protein-coupled bile acid receptor.

eliminating toxins, and producing numerous compounds that affect the host. Hence, host-microbe interactions are determined by the complexity of the human body and the diversity of the microbiome. In this context, the gut microbiota has a deep role in the development of intestinal lymphoid tissue, affecting the balance between Th1 and Th2 responses and inducing Th17 responses. Furthermore, the gut microbiota affects the enteroendocrine system and modulates gut hormones production, such as glucagon-like peptide 1 (GLP-1), which promotes insulin sensitivity in peripheral tissues and modulates the intraepithelial lymphocytes and gut transit. Likewise, work in mice and humans reveals that the gut microbiota composition oscillates with the circadian clock, implying that it may be important to standardize sampling procedures. In general, alterations of the composition of the human gut microbiome is linked to several diseases, including human metabolic diseases such as obesity; type 2 diabetes (T2D), cardiovascular diseases (CVD), cancer, liver disease, allergy, brain disorders, and irritable bowel disease (IBD), establishing causality with some experiments in animals. Nevertheless, the underlying mechanisms remain unclear (Wu et al., 2015; Fan and Pedersen, 2021) (Fig. 2).

Gut microbiota and obesity

Obesity affects 650 million adults worldwide, and 39% of the world population has overweight, increasing the risk of insulin resistance, glucose intolerance, and T2D, non-alcoholic fatty liver disease (NAFLD), several types of cancer, including breast, endometrial, and colorectal cancer, obstructive sleep apnea, and dyslipidemia with subsequent CVD. Research has increasingly focused on the role of gut microbiota and its metabolites in the pathogenesis of obesity and associated comorbidities, with important microbial produced metabolites including the microbial formed trimethylamine (TMA) and its oxidized form TMAO, lipopolysaccharide (LPS), secondary bile acids, and the discovered association of phenylacetylglutamine (PAG) and its role in thrombosis, among others.

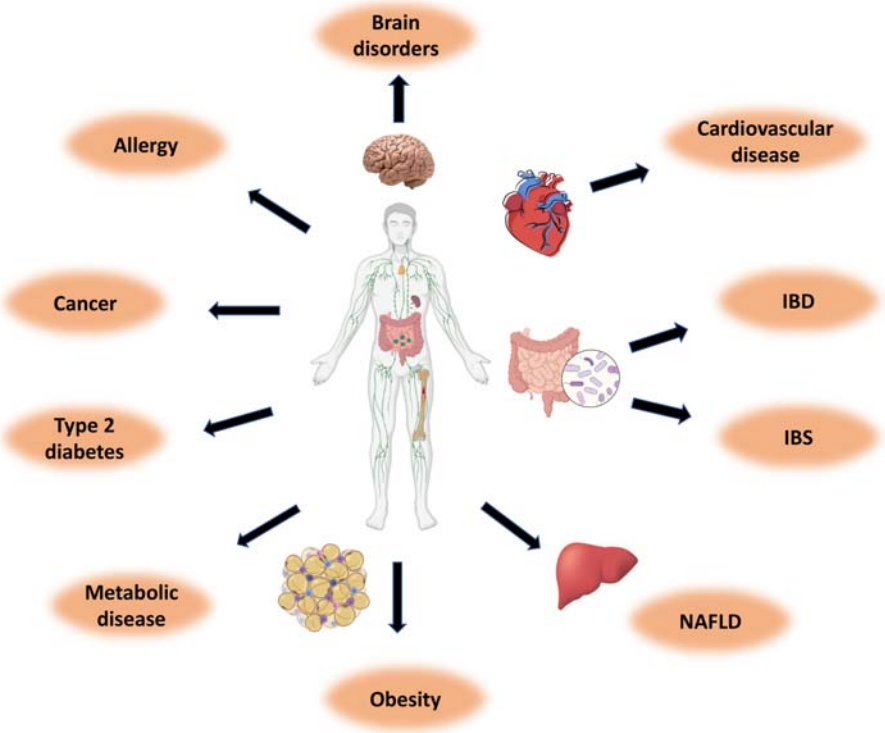


Fig. 2 Microbiota and its effects on different organs. Abbreviations: IBD, inflammatory bowel disease; IBS, irritable bowel syndrome; NAFLD, non-alcoholic fatty liver disease.

Gut microbiota composition seems to change in time, varies among individuals despite a similar diet and depends on age, genetics, and environment. The most commonly described change in gut microbiota ecology is the fluctuations in the relative abundance of the bacterial phyla of *Bacteroidetes* as compared to *Firmicutes* in individuals with obesity, and the increased *Firmicutes*-to-*Bacteroidetes* ratio as a predictor of obesity and cardiometabolic disease, observed in both murine models as well as in human studies. Evidence is accumulating for the role of the gut microbiota in mediating some of the environmental effects in obesity pathogenesis. After the discovery in 2006 that a transferrable obesity-associated microbiota could induce weight gain in lean mice, epidemiological studies showed differences between the gut microbiota of individuals with obesity and lean individuals. Thus, at the species level, twin studies revealed that the abundance of SCFA producers such as *Eubacterium ventriosum* and *Roseburia intestinalis* is associated with obesity. In contrast, butyrate producers such as *Oscillospira* spp. and the methanogenic archaeon *Methanobrevibacter smithii* may be associated with leanness. Besides, a metagenome-wide association study of lean individuals and individuals with obesity displayed that the abundance of *Bacteroides thetaiotaomicron*, a glutamate-fermenting commensal, is markedly decreased in individuals with obesity and is inversely correlated with serum glutamate concentration. Furthermore, gavage of mice with *B. thetaiotaomicron* protected against adiposity, pointing to possible future modalities for obesity intervention targeting the gut microbiota with potential probiotic or microbial compounds. Analyses of gut microbial pathways and gene families suggest that obesity is associated with a decreased capacity for unidirectional conjugation. Observational studies of fecal microbiota transplantation (FMT) demonstrate that transfer of stool from twins discordant for obesity into germ-free mice, in a diet-dependent manner, shifts the phenotype of the human donor to the recipient animal, establishing a link between gut microbiota and whole-body metabolism. Interestingly, a case report has revealed a significant weight gain in a woman receiving FMT from a healthy donor with overweight as a treatment for *C. difficile* infection, suggesting that obesity may also be transmissible in humans.

Obesity promotes a reduced fecal microbiota richness, which is linked to increased inflammation and decreased insulin sensitivity. A large-scale twin study showed that the abundance of *Christensenellaceae* was highly correlated in monozygotic twin pairs and associated with a lean phenotype. FMT of an obese-associated microbiota supplemented with a cultivable member of the *Christensenellaceae* caused reduced weight gain and altered gut microbiota composition in the recipient mice, demonstrating a strong relationship between heritable microbial taxa and host metabolic regulation. Therefore, the development of bacterial taxa as novel probiotics may be useful to target obesity and associated metabolic disease.

Concerning bariatric surgery as the most successful long-term treatment for obesity, it is worth highlighting that this process alters the human gut microbiota, promoting an increase in *Enterobacteriaceae*, and microbiota richness in the short term up to 6 months after surgery. Moreover, by colonizing germ-free mice with stools from the patients, it has been demonstrated that the

surgically altered microbiota promoted reduced adiposity in recipient mice. In addition, the mice exhibited a lower respiratory quotient, indicating decreased utilization of carbohydrates as a fuel. These findings suggest that gut microbiota may play a direct role in reducing adiposity observed after bariatric surgery (Zwartjes et al., 2021; Wu et al., 2015).

Gut microbiota and type 2 diabetes

The gut microbiota is altered in T2D, and the production of butyrate and secondary bile acids may be important for insulin sensitivity. Mechanistic studies in rodents have reported that hyperglycemia might increase intestinal barrier permeability through a GLUT2-dependent transcriptional reprogramming of intestinal epithelial cells and alteration of tight junction integrity, causing a leaky mucosa. Indeed, a leaky mucosa promotes translocation of whole bacteria, bacterial products such as metabolites, and bacterial wall components into the circulation and distant tissues, contributing to remote organ injury in T2D and other metabolic diseases. Besides hyperglycemia, the T2D phenotype often exhibits dyslipidemia with high circulating levels of triacylglycerols and low-density lipoprotein cholesterol, hypertension, and an increased tendency for platelet aggregation. Recent epidemiological studies have focused on elucidating the link between the gut microbiota and T2D by the drug-naive early stages of T2D, called prediabetes. In fact, in individuals with prediabetes, the gut microbiota shows a loss of butyrate-producing taxa, a decrease in abundance of *Akkermansia muciniphila*, and an increase in the abundance of bacteria with pro-inflammatory potentials.

Among the drugs that are frequently prescribed for T2D, metformin, an anti-hyperglycemic compound, seems to promote shifts of multiple genera and species in the gut microbiota such as *Escherichia* and *Intestinibacter*, and enhancement of several microbiome functional potentials, such as propionate and butyrate production that induces intestinal gluconeogenesis.

On the other hand, T2D has been associated with high levels of *Lactobacillus* and low levels of *Roseburia* in both European and Chinese populations, showing that some consistent features may be linked to T2D independent of geographical and dietary differences. Additionally, FMT from lean individuals into patients with metabolic syndrome improved insulin sensitivity over a 6-week period. This has also been linked to higher levels of *Roseburia* in the fecal microbiota and of another butyrate producer, *Eubacterium hallii*, in the microbiota from the small intestinal mucosa (Zhang et al., 2021; Wu et al., 2015).

Gut microbiota and cardiovascular disease

CVD, in particular arteriosclerosis, is one of the leading causes of morbidity and mortality in the western world. People who have atherosclerosis frequently had prior silent metabolic dysfunction such as elevated circulating concentrations of glucose, insulin, and lipids, insulin resistance, and low-grade inflammation. Hence, an aberrant gut microbiota on early-stage metabolic perturbations from an imbalanced gut microbiota is escalating the pathogenesis of ischemic heart disease. On top of this situation, people with the cardiometabolic disease usually are, like in T2D, heavily medicated making it challenging to tease apart authentic arteriosclerosis signals in the gut microbiota from signatures induced by complex medication and pre-morbidities and co-morbidities. Individuals with cardiometabolic disease exhibit in their gut microbiota enriched abundances of *Enterobacteriaceae*, including *Escherichia coli*, *Klebsiella* spp., and *Enterobacter aerogenes*, as well as an increased abundance of oral cavity species when compared with healthy subjects. However, individuals with the cardiometabolic disease show a lower abundance of *Bacteroides* spp. and anti-inflammatory *Faecalibacterium prausnitzii*. Furthermore, ischemic heart failure is associated with a dysbiotic gut microbiota with a higher abundance of *Ruminococcus*, *Acinetobacter*, and *Veillonella* spp., and a decreased abundance of *Alistipes*, *Faecalibacterium*, and *Oscillibacter* spp. Overall, cardiometabolic disease patients show a more inflammatory and less fermentative microbiome. Other studies have focused on showing how microbial metabolism of dietary such as phosphatidylcholine and L-carnitine to TMA may be involved in atherosclerosis and CVD, through a mechanism involving oxidation of TMA to TMA-N-oxide (TMAO) in the liver, establishing a link between TMAO and CVD. Hence, microbial metabolism in combination with diet interactions might be involved in the development of CVD (Wu et al., 2015).

Dysbiosis of gut microbiota in NAFLD

Non-alcoholic fatty liver disease (NAFLD) is a hepatic manifestation of the metabolic syndrome that generally occurs in many countries with a prevalence of 20–40% of the adult population. Disrupted gut microbiota is one of the essential factors in the pathophysiology of NAFLD and non-alcoholic steatohepatitis (NASH), an aggressive inflammatory form of NAFLD. Thus, subjects with NAFLD have an augmented abundance of species assigned to *Clostridium*, *Anaerobacter*, *Streptococcus*, *Escherichia*, and *Lactobacillus*, *whereas Oscillibacter*, *Flavonifractor*, *Odoribacter*, and *Alistipes* spp. are less abundant. Furthermore, children with steatosis or NASH are depleted in *Oscillospira* spp. accompanied by a higher abundance of *Dorea* and *Ruminococcus* spp. compared with normal-weight children.

Gut microbiota in NALFD patients is more enriched in ethanol-producing bacteria such as *E. coli*; therefore, it has been hypothesized that the altered gut microbiome of individuals with NAFLD may produce more ethanol than microbiomes of healthy individuals, as evidenced by increased concentrations of intrinsically generated ethanol in the circulation and breath. Furthermore, ethanol activates nuclear factor-κB (NF- κB) signaling pathways, causing tissue damage by impairing gut barrier function and

contributing to portal endotoxemia. In the liver of individuals with NALFD, the detoxification pathway is weakened, providing a constant source of reactive oxygen species that potentially cause oxidative damage to the hepatocytes, which in turn may induce hepatic inflammation and steatohepatitis. Similarly, liver cirrhosis drives an altered microbiota, including enrichment of *Proteobacteria* and *Fusobacteria* phyla (Fan and Pedersen, 2021).

Diet and gut microbiota

Diet influences the human gut microbiota, displaying the potential of therapeutic dietary strategies to manipulate microbial diversity, composition, and stability. Energy intake and the macronutrient composition of the diet affect human health and influence the composition of the human gut microbiota. However, the gut microbiota responds rapidly to dietary interventions, both short-term and long-term. Generally, a metabolically healthy microbiota is mainly achieved by a high-fiber, low animal fat, and low animal protein diet. On the contrary, microbial dysbiosis is induced by a high-fat diet, characterized by high amounts of fat and protein, and sedentary life, smoking, alcohol intake, and relatively infrequent defecation might cause a leaky mucosa, inflammation, and reduced production of SCFAs. Of note, a dysbiotic microbiota is often associated with a prolonged colonic transit time, resulting in a shift in colonic metabolism leading to increased microbial proteolysis. Furthermore, the gut microbiota of people consuming high amounts of protein and animal fat is dominated by the *Bacteroides* genus and lower amounts of *Firmicutes* genus, while the Prevotella genus dominates the gut microbiota of people consuming more fiber and carbohydrates (Fan and Pedersen, 2021; Zhang et al., 2021). Diet-gut microbiota interactions and their impact on human physiology have been extensively reviewed in the article "Microbiota of the intestine: Dietary interactions".

Personalized nutrition and gut microbiota changes

Personalized nutrition is defined as an approach that counts on details of individual characteristics to evolve a package of nutritional counsel, products, or services, explained as a perspective "assists individuals in achieving a lasting dietary behavior change that is beneficial for health." Diet is increasingly appreciated to impact numerous aspects of the host's biology tremendously, both in health and disease. Dietary content and timing are also central in shaping the gut microbiome and contribute to its taxonomic and functional diversity. Thus, personal variations lead to consider that host and gut microbiome features might contribute to the inter-individual variability in response to environmental factors (e.g., diet). In this context, we highlight recent findings regarding the diet-microbiome interactions and the large personalized data that it represents toward integration into nutritional recommendations affecting human health.

As we abovementioned, diet is a key determinant of personalized microbiome composition and function. Accordingly, dietary components such as carbohydrates, proteins, fatty acids, and some additives influence the human gut microbiome structure and function. Furthermore, the gut microbiome dynamically changes with immigration-related dietary shifts, seasons, and timing throughout a 24 h period, related to chrononutrition.

Different studies have elucidated how intake of dietary components modify gut microbiota in a person-specific fashion. For instance, a report showed that short-term intake of dietary fiber in healthy adults was associated with high proportions of *Prevotella* and *Coprococcus* species and high levels of caproate and valerate in a person-specific manner, with fewer microbiota shifts upon fiber intake. On the other hand, non-caloric artificial sweeteners (NAS) outline the gut microbiome structure and function in both mice and humans and impact postprandial glycemic responses in some, but not in all humans, indicating an effect in an individual-specific manner. In this context, human NAS responder and non-responder subsets are suggested to feature differential glycemic responses to food, and they might be recognized based on individualized microbiome patterns. Similarly, consumption of the probiotic *Lactobacillus paracasei* DG induced individual-specific alterations in fecal *Clostridiales* and butyrate production, depending on the initial intestinal microbial ecology, establishing a predictable fate by combining baseline host and microbiome features. The stable colonization of the probiotic *Bifidobacterium longum* AH1206 happened only in 30% of individuals, which was related to low pre-treatment resident *B. longum* levels and underrepresented microbiome-related carbohydrate utilization genes.

In individuals with obesity, the composition and function of the individual's gut microbiota after weight-stabilization or weight-loss diet differed significantly from person to person, which was negatively correlated with their microbial diversity. Moreover, inter-individual microbiome responses to dietary interventions in obesity were also related to specific bacterial species at baseline, in both males and females. These personalized microbiome responses to dietary exposure may also be important in the context of disease. In obese men, the compositional and functional changes of the individual's gut microbiota under a weight-stabilization or weight-loss diet varied significantly from person to person, which was negatively correlated with their microbial diversity, and relying on an individual's baseline microbiome composition. Inter-individual microbiome responses to obese-related dietary interventions are also associated with specific bacterial species at baseline, in both males and females. Indeed, *Firmicutes* have been used as predictors of both host and microbiota responses to a weight-control diet in patients with obesity. In a large cohort containing 800 subjects with overweight/obesity in Israel, the high inter-individual variability in postprandial glycemic response was predicted by incorporating microbiome features and host characteristics, including anthropometrics and dietary habits, using machine-learning algorithms.

Overall, growing evidence suggests a contributing role of the gut microbiota in predicting the personalized host response to dietary interventions. However, still further studies are needed to decipher downstream mechanisms by which microbe's sense and react to specific food-derived compounds and their downstream impact on the host. Additionally, integration of multi-omics strategies not only limited to microbiome analyses but in combination with proteomics, metabolomics, transcriptomics, and epigenomics are required to identify potential clinical predictors of person-specific response, using advanced machine-learning algorithms and accurate validation methods (Kviatcovsky et al., 2021; Gibney et al., 2016).

Conclusion

Future sequencing-based and culture-based gut microorganism surveys combined with mechanistic exploitations of the gut microbiome will exponentially expand our knowledge about the interactions within the global intestinal microbial community. Finally, yet importantly, exciting new knowledge about the multitude of chemicals that the global gut microbiome produces affecting host physiology and numerous pathologies may foster novel efficacious paths to stabilize metabolic health of the human host and prevent or combat common human metabolic disorders.

References

Andoh, A., 2016. Physiological role of gut microbiota for maintaining human health. Digestion 93, 176–181.
Fan, Y., Pedersen, O., 2021. Gut microbiota in human metabolic health and disease. Nat. Rev. Microbiol. 19, 55–71.
Gibney, M., Walsh, M., Goosens, J., 2016. Personalized nutrition: paving the way to better population health. In: Eggersdorfer, M., Kraemer, M., Vordaro, J.B., et al. (Eds.), Good Nutrition: Perspectives for the 21st Century. Karger Publishers, pp. 235–248.
Gomez-Llorente, C., Munoz, S., Gil, A., 2010. Role of toll-like receptors in the development of immunotolerance mediated by probiotics. Proc. Nutr. Soc. 69, 381–389.
Kelly, D., Yang, L., Pei, Z., 2017. A review of the oesophageal microbiome in health and disease. Methods Microbiol. 44, 19–35.
Kviatcovsky, D., Zheng, D., Elinav, E., 2021. Gut microbiome and its potential link to personalized nutrition. Curr. Opin. Physiol. 22, 100439.
Romani-Perez, M., Bullich-Vilarrubias, C., Lopez-Almela, I., Liebana-Garcia, R., Olivares, M., Sanz, Y., 2021. The microbiota and the gut-brain axis in controlling food intake and energy homeostasis. Int. J. Mol. Sci. 22.
Thursby, E., Juge, N., 2017. Introduction to the human gut microbiota. Biochem. J. 474, 1823–1836.
Valdes, A.M., Walter, J., Segal, E., Spector, T.D., 2018. Role of the gut microbiota in nutrition and health. BMJ 361, k2179.
Wu, H., Tremaroli, V., Bäckhed, F., 2015. Linking microbiota to human diseases: a systems biology perspective. Trends Endocrinol. Metab. 26, 758–770.
Zhang, L., Chu, J., Hao, W., Zhang, J., Li, H., Yang, C., Yang, J., Chen, X., Wang, H., 2021. Gut microbiota and type 2 diabetes mellitus: association, mechanism, and translational applications. Mediat. Inflamm. 2021, 5110276.
Zwartjes, M.S.Z., Gerdes, V.E.A., Nieuwdorp, M., 2021. The role of gut microbiota and its produced metabolites in obesity, dyslipidemia, adipocyte dysfunction, and its interventions. Metabolites 11.

Microbiota of the intestine: Dietary interactions

Julio Plaza-Diaz[a,b,c] **and Angel Gil**[b,c,d,e], [a] Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada; [b] Department of Biochemistry and Molecular Biology II, School of Pharmacy, University of Granada, Granada, Spain; [c] Instituto de Investigación Biosanitaria IBS.GRANADA, Complejo Hospitalario Universitario de Granada, Granada, Spain; [d] Institute of Nutrition and Food Technology "José Mataix", Center of Biomedical Research, University of Granada, Granada, Spain; and [e] CIBEROBN (CIBER Physiopathology of Obesity and Nutrition), Instituto de Salud Carlos III, Madrid, Spain

© 2023 Elsevier Ltd. All rights reserved.

Introduction 168
Microbiota of the intestine 169
High-throughput sequencing technologies 170
Roles of intestinal microbiota 170
Protective and structural functions 170
Metabolic functions 170
Microbiota and malnutrition 171
Gut microbiota: dietary interactions 172
Gut microbiota: microbial metabolites 173
Manipulating the gut microbiota through diet 173
Personalized nutrition through the microbiota 174
Conclusion 174
References 174

Key points

- To describe the major microbes of the intestine
- To understand the main functions of the gut microbiota
- To know the new technologies for the detection of specific intestinal microbes
- To understand the dietary interactions between gut microbes and foods
- To mention some aspects in personalized nutrition using the gut microbiota

Introduction

The microbiota can be described as "the full collection of microbes, i.e., bacteria, fungi, virus, etc., that naturally exists within a particular biological niche such as an organism, soil or a body of water, among others". Actually, the "microbiota" word is also referred to as the metagenome of the microbiota. Indeed, the microbiome comprises all of the genetic material within a microbiota (the entire collection of microorganisms in a specific niche, such as the human gut). Gut microbes are key to many aspects of human health including metabolic, immune, and neurobehavioral traits (Valdes et al., 2018).

The microorganism's communities in the human digestive tract are an ecosystem that is both critical for health and also a possible pilot of several pathologies. The adult human intestine is colonized by more than 1000 microbial species including all domains of life: Archaea, Bacteria, and Eukarya (Plaza-Diaz et al., 2019; Plaza-Díaz and Gil, 2017). The microbiota properties in the human small intestine are fewer well described, primarily because of the complications in testing this segment of the digestive tract. Learning the small intestine ecosystem is predominantly important to digestive health because the duodenum and jejunum are tasked with facilitating the majority of nutrient assimilation and absorption. The intestinal epithelium is in constant and closed contact with luminal substances and the variable, dynamic enteric microbiota. The intestinal barrier is the main defense mechanism used to preserve epithelial integrity and to safeguard the organism from the environment. Defenses of the intestinal barrier consist of the antimicrobial peptides, secretory IgA, mucous layer, and the epithelial junction adhesion complex. After this barrier function is distorted, food and bacterial antigens can reach the submucosa and can provoke inflammatory responses, which may result in intestinal disorders, such as inflammatory bowel disease. In the present article, the objective is to describe the main microbes present in the intestine, to mention the new technologies for the detection of specific species, and to discuss the interactions between gut microbes and diet.

Microbiota of the intestine

Specifically, there are more than 50 phyla of bacteria on Earth, but of these, the bacterial communities associated with humans are predominantly dominated by four, *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, and *Proteobacteria*, and other phyla to a lesser extent, depending on the individual's characteristics and the distribution of microbes throughout the human body.

The relative abundance of these four phyla tends to be consistent across individuals, regardless of their place of origin, residence, or type of diet, e.g., in almost all humans studied so far, *Bacteroidetes* and *Firmicutes* predominate in the colon. In contrast, the composition of the vaginal microbiota is more variable; most women have a preponderance of *Firmicutes* with some other representatives, while a minority of women have a preponderance of *Actinobacteria* with some other representatives. It is estimated that 20–80% of the human phylum are directly associated with habitat (Plaza-Díaz and Gil, 2017, O'hara and Shanahan, 2006).

Fig. 1 shows that the greatest amount of bacterial load is found in the intestine and colon. The distribution of the microbiota in the human intestine is not homogeneous and the number of bacteria present ranges from 10 to 10^3 colony-forming units (CFU) in the stomach and duodenum, increasing from 10^{11} to 10^{12} CFU in the colon. There are also differences in longitudinal heterogeneity throughout the digestive tract and latitudinal variation because the intestinal epithelium is separated from the lumen by a thick and complex layer of mucus, generating a different type of habitat to that which may exist in the lumen or on the surface of the intestinal epithelium, causing a clear difference in the microbiota in each of these habitats.

The human intestine is the natural habitat for a large and dynamic bacterial community, but a substantial part of these populations remains as yet unknown. The dominant phyla of bacteria in the intestine are *Bacteroidetes*, *Firmicutes*, *Actinobacteria*, *Proteobacteria*, and *Verrucomicrobia*. In the small intestine, the *Enterobacteriaceae* and *Lactobacillaceae* families dominate, whereas in the colon there is a marked presence of species from the *Bacteroidaceae*, *Prevotellaceae*, *Rikenellaceae*, *Lachnospiraceae*, and *Ruminococcaceae* families (O'hara and Shanahan, 2006; Plaza-Díaz and Gil, 2017; Plaza-Diaz et al., 2019). Within the small intestine, in the duodenum, the composition of the microbiota is sparse and usually contains less than 10^5 CFU; *Haemophilus*, *Actinomyces*, and some anaerobes, and lactobacilli are found. In the jejunum and ileum, there is a continuous increase in the number ($>10^8$ CFU) and variety of the microbiota with bifidobacteria, facultative anaerobes (*Bacteroides* and *Fusobacterium*), and strict anaerobes being present in increasing numbers from the ileocecal valve. The microbiota in the distal part of the small intestine is denser ($>10^9$ CFU) and resembles that of the cecum, with large numbers of strictly anaerobic bacteria. In the large intestine,

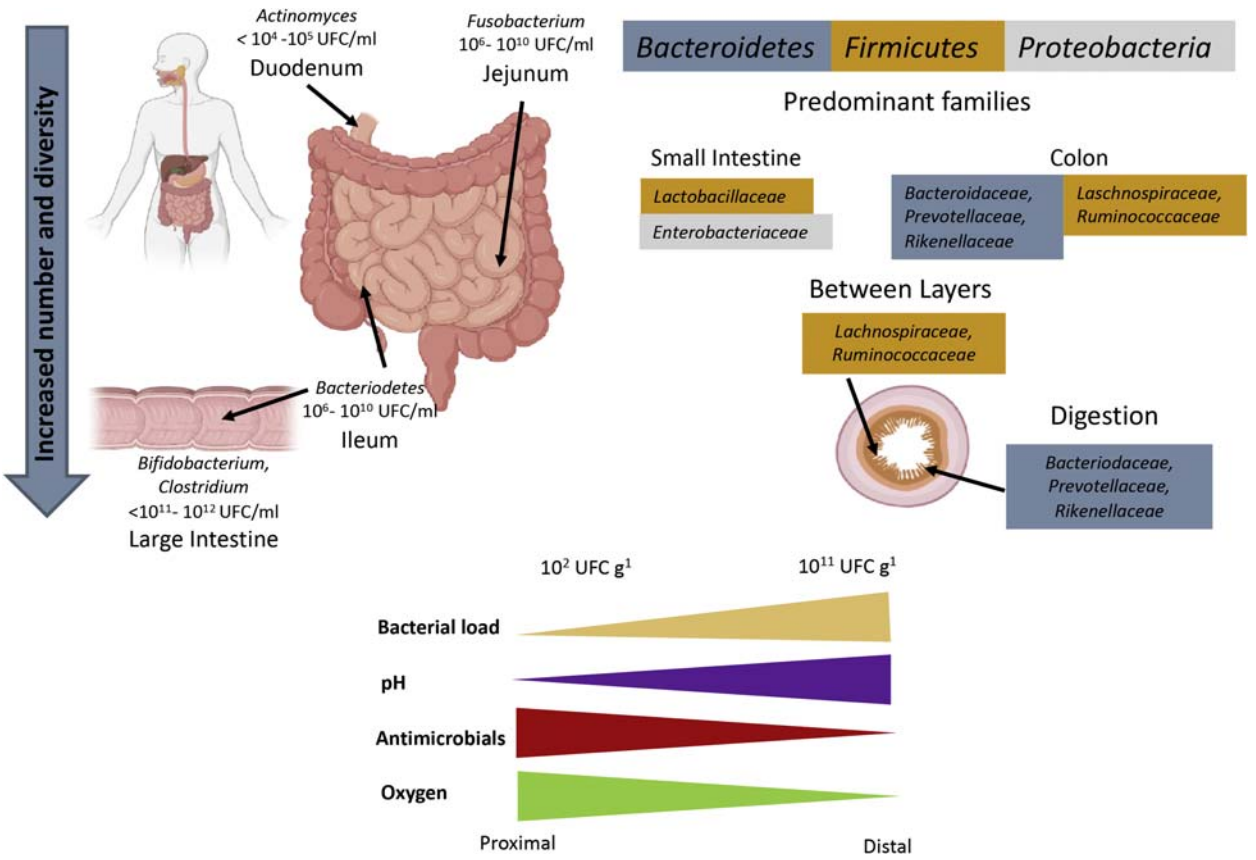


Fig. 1 Variation in composition and number of the microbiota along the human gut. Abbreviations. UFC, unit forming-colony.

facultative anaerobic microorganisms (*Streptococcus* and *Enterococcus*) are commonly found. Finally, in the cecum, eubacteria, bifidobacteria, *Clostridium*, and Gram-positive bacteria are prominent (O'hara and Shanahan, 2006; Plaza-Díaz and Gil, 2017; Plaza-Diaz et al., 2019) (**Fig. 1**).

The gut microbiota in adults is estimated to consist of between 1000 and 1150 bacterial species, with some experts suggesting that only 160 of these species constitute the core microbiota that is present in most individuals. Although many of these species are found in most people, their relative abundance may vary, thus there is great individual variability. The complex bacterial community present in the human intestine does not remain constant over time but can vary due to several factors, among which we can find the environmental conditions of the digestive tract itself, the quantity and variety of bacteria in the different regions of the digestive tract, which is determined by a great diversity of complex intrinsic and extrinsic factors. There is also a reciprocal relationship between the host and the microbiota that inhabits it since the microbiota can have an important impact on the organism that hosts it and these effects can be beneficial or detrimental to the health of the individual. In recent years, great strides have been made in characterizing the composition of the microbiota and how such microbial numbers have functional interactions with the host. Such studies on function are becoming fundamental in understanding the role of the microbiota in the maintenance of human homeostasis and the pathogenesis of certain diseases (O'hara and Shanahan, 2006; Plaza-Díaz and Gil, 2017; Plaza-Diaz et al., 2019).

High-throughput sequencing technologies

More than a decade ago, the majority of the knowledge and available information about the adult human gut microbiota came from culture-based methods. Nowadays, our faculty to understand the gut microbiota has significantly upgraded and developed due to high-throughput and low-cost sequencing methods targeting the bacterial 16S ribosomal RNA (rRNA) gene. Because this gene exists in all bacteria and Archaea with nine highly variable regions (V1–V9), which permits specific bacteria to be easily described (Poretsky et al., 2014; Thursby and Juge, 2017). More recently, the 16S rRNA sequencing focus has moved to analyze shorter sub-regions of the gene in greater depth. However, the employment of shorter read lengths could introduce some errors in the sequencing (Poretsky et al., 2014).

Roles of intestinal microbiota

The influence of bacteria on intestinal physiology has been demonstrated in studies with germ-free animals through colonization. Thus, it has been observed that reconstitution of the microbiota in germ-free mice with an intestinal microbiota from other normal mice is sufficient to recover the immune function of the intestinal mucosa. Repopulation of the gut of germ-free mice with a single species, *Bacteroides thetaiotaomicron*, has been shown to affect the expression of several genes that control certain functions associated with nutrient absorption, metabolism, angiogenesis, mucosal barrier function, and enteric nervous system development (Plaza-Díaz and Gil, 2017; O'hara and Shanahan, 2006).

Protective and structural functions

Host defense against pathogenic microorganisms mediated by the intestinal microbiota requires a fine interpretation of the microenvironment present, as it must distinguish between commensal microorganisms and occasional pathogens and, in addition, differentiate subsequent responses for both cases (Plaza-Díaz and Gil, 2017; O'hara and Shanahan, 2006).

The mucosal epithelia and, in particular, the intestinal epithelium represent the first line of defense against action by pathogens; superficial enterocytes serve as afferent sensors of the luminal microenvironment by secreting antibacterial peptides, immunoglobulin A, and chemokines that alert and direct the immune response to the site of infection. M cells lining the lymphoid follicles transport luminal antigens to underlying dendritic cells and other antigen-presenting cells. Finally, intestinal dendritic cells play a role essentially as sensors of the immune system and can directly monitor intestinal contents, either by entry or extension of dendrites between superficial enterocytes without disrupting tight junctions. In addition, they can ingest and keep commensal bacteria alive, travel to mesenteric lymph nodes, where a local immune response against commensal bacteria is induced (Plaza-Díaz and Gil, 2017; O'hara and Shanahan, 2006).

The fine distinction between pathogenic and commensal bacteria is mediated by the antigenic pattern recognition receptor (PRR) system, the toll-like receptor (TLR) family, and NOD-like receptors (NLR). On the other hand, commensal bacterial fragments influence the normal development of the mucosal immune system and profoundly affect the development of the humoral components of this system, modulating the response of T and T helper cells type 1 and 2. The human microbiota controls the proliferation and differentiation of epithelial cells and is also responsible for modulating the maturation and activity of the innate and adaptive immune response (**Fig. 2**) (Plaza-Díaz and Gil, 2017; O'hara and Shanahan, 2006).

Metabolic functions

The microbiota has an adaptive yet renewable metabolic activity. Through the production of short-chain fatty acids (SCFA), resident bacteria positively influence intestinal epithelial cell differentiation and proliferation and may mediate other metabolic effects. All

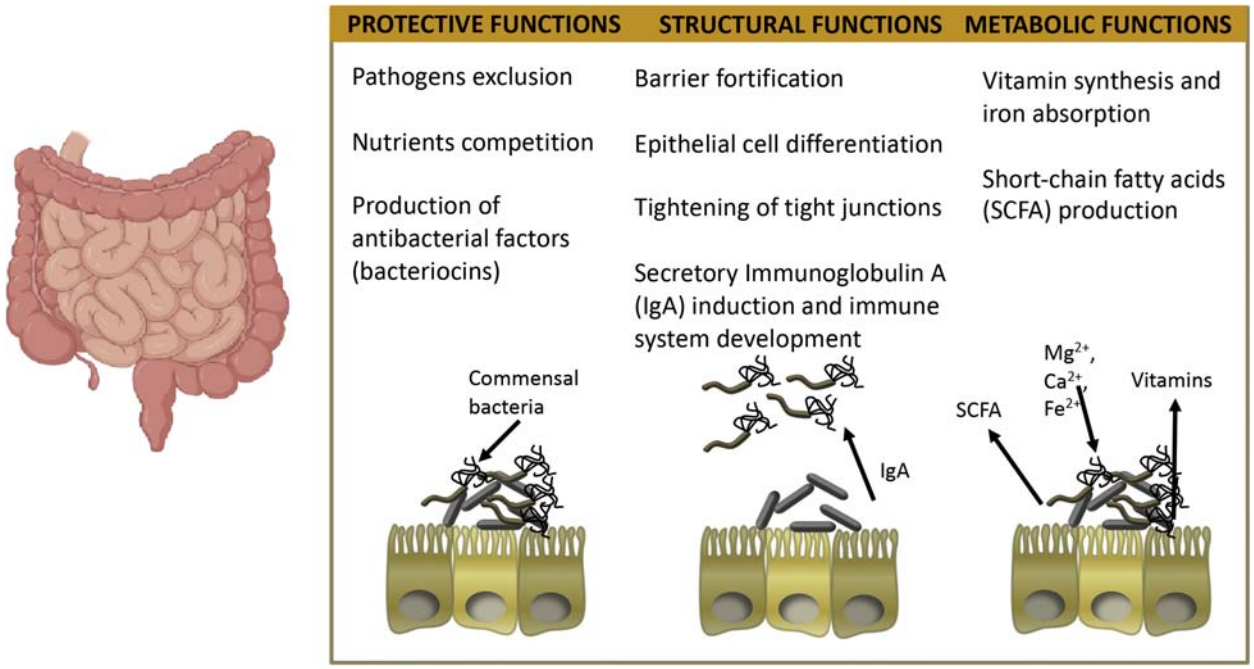


Fig. 2 Functions of the gut microbiota. Abbreviations. IgA, immunoglobulin A; SCFA, short-chain fatty acids.

this complex metabolic activity recovers valuable energy and absorbable substrates for the host and at the same time provides energy and nutrients for bacterial growth and proliferation. In addition, the commensal microbiota synthesizes some vitamins that are efficiently used by the host, such as vitamin K and several B vitamins, including biotin, folic acid, and vitamin B12 (Plaza-Díaz and Gil, 2017; O'hara and Shanahan, 2006).

It is important to note that each of the functions of the intestinal microbiota varies with age. Some factors as the diet may account for changes in host physiology associated with the development of immune system reactivity, but may also be caused by changes in the bacterial population of the large intestine. Some studies have indicated that bifidobacteria content decreases in older people, increasing those of clostridia and enterobacteria, favoring the progression of infections. Lactobacilli and bifidobacteria are some of the candidate microorganisms to be used in humans since there is evidence that some of these organisms can increase resistance to intestinal infections by inhibiting pathogens and improving the immune response of the host (**Fig. 2**) (Plaza-Díaz and Gil, 2017; O'hara and Shanahan, 2006).

Microbiota and malnutrition

Child malnutrition accounts for almost half of all deaths in children under 5 years of age worldwide. At present, health efforts to restore nutritional intake remain inconclusive and have yielded modest results (Thursby and Juge, 2017). One of the new alternatives proposed is to modify the population of bacteria inhabiting the intestine which, as mentioned above, modulates intestinal metabolic activity. Childhood malnutrition has been associated with an altered microbiota (Thursby and Juge, 2017). New studies indicate that the composition of the microbiota is altered in chronically malnourished children. The time between conception and three years of age is crucial for human growth and development. Decades of research on the onset of health and disease have shown that environmental influences during this period can contribute to the onset of disease later in life. For example, suffering nutritional deficits or excess nutrients in early life can have metabolic and cardiovascular health consequences. Because of the symbiotic role of the gut microbiota in intestinal metabolism, it is thought that changes in its composition may contribute to these health problems (Plaza-Díaz and Gil, 2017).

The foods we eat influence the type of microbes that colonize the gut. Each microbial species has an optimal metabolic environment that is preferably compatible with its growth. As such, a diet rich in plant fiber promotes a gut microbiota considerably different from that promoted by a diet rich in animal fat. The microbiota converts non-digestible dietary components, such as fiber, into useful compounds to fuel the growth of intestinal cells that, in turn, promote the development of a healthy immune system (Myhrstad et al., 2020).

To investigate the stool microbiota of Malawian children with and without varying degrees of growth impairment, results from children who grew normally were compared with the idea of deriving a model of a healthy microbiota (Thursby and Juge, 2017). When such a model was applied to the entire cohort it was observed that children with weight-associated growth impairment had an immature microbiota compared to those of a healthy weight. Subsequently, it was investigated whether the difference in the gut

microbiota contributed to the observed impairment in growth (Gehrig et al., 2019). Bacteria were isolated from the fecal samples of the children to colonize the gut of germ-free mice. The mice were then fed a nutrient-poor diet that resembled a typical Malawian diet. After several weeks the mice harboring a microbiota-derived from growth-impaired donors gained significantly less weight than the control group that received the microbiota from healthy children. However, if both groups were housed in the same cage, the microbiota from the controls quickly transferred to the low-growing mice, regaining weight similar to the healthy animals. This effect was attributed to two bacterial species, *Ruminococcus gnavus*, and *Clostridium symbiosum*, which were introduced into germ-free mice along with the microbiota of stunted donors, and sustained weight gain followed. In addition, the levels of amino acid metabolism by-products were reduced in the livers of the mice that had regained weight compared to those of the growth-stunted mice (Gehrig et al., 2019).

In an animal study, the growth of normal young mice was compared with that of germ-free mice indicating that the presence of a previous microbiota promoted growth by increased production of Insulin-like Growth Factor-Binding Protein (IGFBP). If the mice had a nutrient-poor diet, microbial stimulation of this growth factor partially ameliorated the growth deficit. Not all strains of bacteria could promote growth in the manner detailed above. Different strains of a species had different effects on growth. These data obtained highlight the fact that each microbe interacts with its host in a heterogeneous manner, indicating that the beneficial growth-promoting effects of bacteria in mice cannot be applied to humans (Plaza-Díaz and Gil, 2017).

Growth and its relationship to breast milk were analyzed in malnourished children and the content of the breast milk they consumed. The breast milk of malnourished children had a lower amount of sugars, especially oligosaccharides. These sugars are of interest because they are abundant in human milk, but not in cow's milk. Strategies aimed at supplementing the infant's diet may include milk supplemented with oligosaccharides to promote healthy growth in infants who do not receive sugars through their mother's milk. However, the large-scale manufacture of these oligosaccharides is still difficult because of the technology required to achieve their use on an industrial scale.

The gut microbiota of children living in an urban slum in Bangladesh was investigated by collecting fecal samples from birth to the first two years of life. It was concluded that growth problems were associated with immature microbiota even in the early stages of malnutrition (Plaza-Díaz and Gil, 2017).

It is becoming increasingly clear that our diet, gut microbiota, and health are highly correlated. Therefore, we need to be aware that dietary interventions can affect the growth of billions of bacteria and those have an impact on the host in multiple domains (Plaza-Díaz and Gil, 2017).

Gut microbiota: dietary interactions

Humans can only synthesize 11 of the 20 essential amino acids themselves; they trust food intake for the other nine along with all 13 essential vitamins. Most of these amino acids and vitamins came from meat, eggs, milk products, fruits, and vegetables, but a few essential compounds are produced by microbes, which are important producers of essential amino acids and vitamins themselves (Hirt, 2020).

Diet influences microbial colonization of the gastrointestinal tract, affecting health throughout the lifespan. The short-term consumption of diets composed entirely of animal or plant products alters microbial community structure and overwhelms inter-individual differences in microbial gene expression. An animal-based diet increases the abundance of bile-tolerant microorganisms (*Alistipes*, *Bilophila*, and *Bacteroides*) and decreases the levels of *Firmicutes* that metabolize dietary plant polysaccharides (*Roseburia*, *Eubacterium rectale*, and *Ruminococcus bromii*). Increases in the abundance and activity of *Bilophila* spp. on the animal-based diet support a link between dietary fat, bile acids, and the outgrowth of microorganisms capable of triggering inflammatory bowel disease (David et al., 2014).

A Western-type diet is related to metabolic disease incidence increased in developed nations. In contrast, diets high in fruits, vegetables, nuts, seeds, legumes, and whole grains and low in animal fats and sugary foods are protective against cardiovascular disease and metabolic disorders. These diets are supportive of health-promoting microbiota and metabolomic profiles, as evidenced by associations with the higher relative abundance of *Lachnospira*, *Prevotella*, and *Roseburia* and increased concentrations of fecal SCFAs (Hirt, 2020).

The gut microbiota has crucial capacities for non-digestible substrates fermentation as dietary fibers and endogenous intestinal mucus, and this fermentation generates the growth of the specific microbes that create SCFAs and gases. The main SCFAs formed are acetate, propionate, and butyrate (Valdes et al., 2018).

Butyrate is the key source of energy for human colonocytes, can stimulate colon cancer cells apoptosis, and can stimulate intestinal gluconeogenesis with positive effects on glucose and energy homeostasis. Butyrate is vital for epithelial cells to consume large amounts of oxygen through β oxidation, causing a state of hypoxia that preserves oxygen balance in the gut, avoiding dysbiosis in the gut microbiota. Propionate is transported to the liver, where it controls gluconeogenesis and signaling for satiety through interaction with the gut fatty acid receptors. Finally, acetate is the main abundant SCFA and is critical for bacterial growth, when acetate goes to peripheral tissues it is used in lipogenesis and cholesterol metabolism and could have some participation in central appetite regulation.

Some clinical evidence, especially from randomized controlled trials has shown that greater SCFAs production associates with lower diet-induced obesity and with decreased insulin resistance. Gut microbial enzymes contribute to bile acid metabolism, producing unconjugated and secondary bile acids that act as signaling molecules and metabolic regulators to impact central

host pathways. Also, some evidence has reported that higher diet quality previses lower mortality and type 2 diabetes incidence. Besides, the positive influence of diet quality on body weight, body fat distribution, and inflammatory status, food components, and dietary patterns have been associated with the gut microbiota composition (Zhao et al., 2018). In addition, evidence has accumulated for a relation between gut microbial characteristics and type 2 diabetes.

A meta-analysis showed that data-derived patterns characterized by red/processed meat, refined grains, high-fat dairy, and fried products were related to a 44% higher diabetes risk, whereas a diet rich in vegetables, legumes, fruits, poultry, and fish predicted a 16% lower risk (Jannasch et al., 2017). Similar results were observed with a reduction in diabetes risk of 30% with the Mediterranean diet (Jannasch et al., 2017). The literature has important information showing that dietary components bioavailable to microbiota, such as dietary fats, proteins, carbohydrates, and polyphenols, impact the gut microbiota composition.

Some studies have added new information about the triangular relation of dietary patterns, the presence of prediabetes as an indicator of risk to develop diabetes, and gut microbiota composition. In participants who had completed a validated quantitative dietary questionnaire before stool collection, the dietary indices were positively related to alpha diversity indices, and decreased relative abundance of *Actinobacteria*. Some genera were associated negatively and positively with the diet quality, especially from *Roseburia* and *Lachnospiraceae* families (Liu et al., 2019). *Roseburia* and *Lachnospira* both metabolize fermentable carbohydrates, producing butyrate and that affects glucose metabolism and diabetes risk. In the case of *Lachnospira*, these bacteria use pectin from fruits and vegetables for the production of acetate: Also, *Roseburia* in the presence of these fermentable carbohydrates could condense two moles of acetate to form one mole of butyrate. A systematic review performed to analyze the dietary influences on fecal microbiota composition in healthy humans 1–20 years of age, indicates that a diet rich in indigestible plant polysaccharides is related to *Prevotella* species, i.e., *Prevotella copri*. In contrast, a diet with a content of a high-fat and -sugar diet is related to *Bacteroides* (Dinsmoor et al., 2021).

Gut microbiota: microbial metabolites

Metabolomics is an important platform to distinguish the profile of metabolites by performing large-scale metabolic profiling after dietary interventions. In the nutrition field and health specifically, metabolic profiling may provide information about diet—that is, as food intake biomarkers, as well as the ongoing physiological responses triggered by the diet. Investigations of dietary responses, however, remain challenging due to the complex interplay between any bioactive compound with other nutritional components, and an immense variation between individuals, as previously reported.

Microbial metabolites do not seem to have a comparable behavior next to dietary interventions, which makes sense due to their great diversity. Secondary bile acids are conversion products of primary bile acids by the gut microbiota. Included in this group are deoxycholic and lithocholic acid, as well as their conjugates of taurine and glycine. They have been reported to increase following various interventions, including oral-glucose-tolerance test, yogurt, acidified milk, and wheat flour fortified with vitamin B and minerals, except for lower deoxycholate concentrations after the oral-glucose-tolerance test.

Propionate diminished after an oral-glucose-tolerance test. However, it augmented after the consumption of cranberry juice and green tea. Acetate, conversely, decreased following consumption of dark chocolate and black tea. Other specific gut microbiota products have been associated directly with human health outcomes. The trimethylamine production from meat and dairy differs on the gut microbiota and thus its amount in the blood varies between people. Trimethylamine is oxidized in the liver to trimethylamine N-oxide, which is positively related to an increased atherosclerosis risk and major adverse cardiovascular events. Indole propionic acid is highly related to the intake of dietary fiber and has strong radical scavenging activity in vitro, which looks to decrease the incidence of type 2 diabetes (Myhrstad et al., 2020). Indole propionic acid has been shown to grow following an optimized Nordic diet and acidified milk, but declines after an oral-glucose-tolerance test.

Manipulating the gut microbiota through diet

Gut microbiota changes can occur within days of changing diet; remarkable differences have been found after African Americans and rural Africans switched diets for only two weeks (Zhao et al., 2018).

A subset of dietary fiber sources is fermentable, which means that they serve as growth substrates for microbes in the distal bowel. The consumption of resistant starches has been shown to facilitate the specific growth of bacterial groups (e.g., *Bifidobacterium adolescentis*, *Ruminococcus bromii*, and *Eubacterium rectale*) in some populations. Microbiota accessible carbohydrates offer a potential strategy to improve useful minority members of the microbiota. These changes only last as long as the carbohydrate is consumed, and they are highly individual, which provides a basis for personalized approaches. Many short-term feeding trials with purified dietary fibers or even whole plant-based diets either do not affect microbiota diversity or reduce it (Zhao et al., 2018), but can still have clinical benefits, potentially through metabolites such as SCFAs (Zhao et al., 2018; David et al., 2014).

Low fiber intake decreases SCFA production and shifts the gastrointestinal microbiota metabolism to use less favorable nutrients, leading to the production of potentially detrimental metabolites. Several meta-analyses have found clear links between dietary fiber and health benefits in a wide range of pathologies (Zhao et al., 2018), and dietary fibers significantly reduced insulin resistance in patients with type 2 diabetes, with clear links to the shifts in the microbiota and beneficial metabolites (such as butyrate).

Probiotics are living microorganisms that confer health benefits to the host when administered in adequate amounts. However, dead bacteria and their components can also exhibit probiotic properties. *Bifidobacterium* and strains of lactic acid bacteria are the most widely used bacteria that exhibit probiotic properties and are included in many functional foods and dietary supplements (Plaza-Diaz et al., 2019). Probiotics can affect health independently of the gut microbiota through direct effects on the host; for example, through immune modulation or the production of bioactive compounds. The therapeutic effect of probiotic supplementation has been studied in a broad range of diseases (Valdes et al., 2018). *Bifidobacterium longum* has been shown to depend on individualized features of the gut microbiota, providing a rationale for the personalization of probiotic applications.

In addition, gut microbiota changes in the composition and function, for example in the case of dysbiosis could be related to driving obesity-related pathogenesis and may be one of the most important drivers of mucosal barrier dysfunction. Several microbiota effects on the host are mediated by microbiota-derived metabolites (Wei et al., 2021). Also, novel products are tested for the food industry with anti-obesity activities (Shen et al., 2021; Su et al., 2021).

Personalized nutrition through the microbiota

Blood glucose levels after a meal (postprandial response) are determined by an individual's daily food and nutrient intake. It has shown that the postprandial response to standardized meals is highly variable between individuals and depends on several factors that may include the gut microbiota. This is important because tracking individual responses to different foods can generate an integration of this information between health parameters and microbiota composition into a learning algorithm to develop individualized nutrition plans to improve the postprandial response (Plaza-Díaz and Gil, 2017). A study assessed glucose levels in a cohort of 800 people every 5 min for seven days during which participants followed their normal routine, except for the consumption of standardized meals. These data were used to calculate the postprandial response. A positive correlation between the abundance of *Proteobacteria*, *Enterobacteriaceae*, and *Actinobacteria* and elevated values in the postprandial response to some of the standardized meals was found, while the presence of *Clostridium* and *Prevotellaceae* correlated with lower responses (Zeevi et al., 2015). In that study, a computer learning algorithm based on clinical parameters and the microbiota was useful to predict individual postprandial responses. An initial analysis of individual postprandial responses from the algorithm showed that *Eubacterium rectale* growth is associated with low glycemic responses while *Parabacteroides distasonis* abundance was associated with higher values. The findings in this field indicate that the postprandial response is different and variable among individuals, even when they consume the same meal. The postprandial response has a multifactorial component with differences in the composition of the gut microbiota. Further studies should address in greater detail the possible mechanisms of action between the effect of diet and its relationship with the gut microbiota and metabolic diseases.

Conclusion

Dietary patterns broadly agree on components of a healthy diet: primarily high intakes of fruits/vegetables, nuts, and whole-grain products and low intakes of red/processed meats and products high in sugar and certain fats. The beneficial health outcomes are likely a result, in part, of higher levels of SCFA derived from gut microbial fermentation of plant-based nutrients typical for a Mediterranean diet.

Dietary interventions and targeted nutritional therapies, such as medical foods, dietary supplements, living microorganisms, and nutraceutical food, could provide a great promise for the prevention and treatment of microbiota-related diseases.

Some specific genera from the gut microbiota are related to human health, and those are principally regulated through dietary interactions: However, more clinical trials, especially randomized clinical trials with an adequate number of subjects are needed to elucidate the specific changes in the gut microbiota related with the food intake.

References

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., Biddinger, S.B., Dutton, R.J., Turnbaugh, P.J., 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505, 559–563.

Dinsmoor, A.M., Aguilar-Lopez, M., Khan, N.A., Donovan, S.M., 2021. A systematic review of dietary influences on fecal microbiota composition and function among healthy humans 1–20 Years of age. Adv. Nutr. 12 (5), 1734–1750.

Gehrig, J.L., Venkatesh, S., Chang, H.W., et al., 2019. Effects of microbiota-directed foods in gnotobiotic animals and undernourished children. Science 365.

Hirt, H., 2020. Healthy soils for healthy plants for healthy humans: how beneficial microbes in the soil, food and gut are interconnected and how agriculture can contribute to human health. EMBO Rep. 21, e51069.

Jannasch, F., Kröger, J., Schulze, M.B., 2017. Dietary patterns and type 2 diabetes: a systematic literature review and meta-analysis of prospective studies. J. Nutr. 147, 1174–1182.

Liu, Y., Ajami, N.J., EL-Serag, H.B., Hair, C., Graham, D.Y., White, D.L., Chen, L., Wang, Z., Plew, S., Kramer, J., 2019. Dietary quality and the colonic mucosa–associated gut microbiome in humans. Am. J. Clin. Nutr. 110, 701–712.

Myhrstad, M.C.W., Tunsjo, H., Charnock, C., Telle-Hansen, V.H., 2020. Dietary fiber, gut microbiota, and metabolic regulation-current status in human randomized trials. Nutrients 12.

O'hara, A.M., Shanahan, F., 2006. The gut flora as a forgotten organ. EMBO Rep. 7, 688–693.

Plaza-Díaz, J., Gil, Á., 2017. Microbioma Humano. In: Gil, A. (Ed.), Tratado de Nutrición. Editorial Panamericana, Madrid.
Plaza-Diaz, J., Ruiz-Ojeda, F.J., Gil-Campos, M., Gil, A., 2019. Mechanisms of action of probiotics. Adv. Nutr. 10, S49–S66.
Poretsky, R., Rodriguez-R, L.M., Luo, C., Tsementzi, D., Konstantinidis, K.T., 2014. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. PLoS One 9, e93827.
Shen, C.Y., Hao, Y.F., Hao, Z.X., Liu, Q., Zhang, L., Jiang, C.P., Jiang, J.G., 2021. Flavonoids from Rosa davurica Pall. fruits prevent high-fat diet-induced obesity and liver injury via modulation of the gut microbiota in mice. Food Funct. 12, 10097–10106.
Su, Y., Li, J., Wu, L., Kuang, H., 2021. Polysaccharides from TCM herbs exhibit potent anti-obesity effect by mediating the community structure of gut microbiota. Pharmazie 76, 473–479.
Thursby, E., Juge, N., 2017. Introduction to the human gut microbiota. Biochem. J. 474, 1823–1836.
Valdes, A.M., Walter, J., Segal, E., Spector, T.D., 2018. Role of the gut microbiota in nutrition and health. BMJ 361, k2179.
Wei, Y.X., Zheng, K.Y., Wang, Y.G., 2021. Gut microbiota-derived metabolites as key mucosal barrier modulators in obesity. World J. Gastroenterol. 27, 5555–5565.
Zeevi, D., Korem, T., Zmora, N., et al., 2015. Personalized nutrition by prediction of glycemic responses. Cell 163, 1079–1094.
Zhao, L., Zhang, F., Ding, X., et al., 2018. Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. Science 359, 1151–1156.

Nutriepigenetics

Concepción M. Aguilera[a,b,c], **Augusto Anguita-Ruiz**[a,b,c], **and Ángel Gil**[a,b,c], [a] Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology "José Mataix", Center of Biomedical Research, University of Granada, Armilla, Spain; [b] Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain; and [c] CIBEROBN (Physiopathology of Obesity and Nutrition Network CB12/03/30038), Institute of Health Carlos III (ISCIII), Madrid, Spain

© 2023 Elsevier Ltd. All rights reserved.

Introduction 177
Epigenetic modifications and chromatin structure 177
Nutrient metabolic regulation by epigenetic mechanisms 178
Epigenetics and dietary intervention 180
Hypercaloric and high-fat diets 180
Caloric restriction and intermittent fasting 181
Mediterranean diet 182
Nutritional factors involved in epigenetic regulation 182
Methyl donor molecules 182
Fatty acids and amino acids 183
Minerals and vitamins 183
Polyphenols and other plant compounds 184
Conclusions and future perspectives 186
References 187

Key points

- To understand epigenetic modifications, chromatin structure and their impact on gene expression.
- To analyze the relationship between metabolic diseases and nutriepigenetics.
- To consider the hereditary transmission of epigenetic marks.
- To identify the ways in which chromatin can be modified by nutrients and dietary patterns and their functional importance.

Glossary

Precision nutrition A concept that goes beyond personalized nutrition as it brings together knowledge based on the gene pool (gene sequence, epigenetics, transcriptomics), but also on phenotypic markers such as family history and previous illnesses, lifestyle, exercise, diet, microbiota, stress, culture, religion, aversions, preferences, allergies and intolerances, etc.

Epigenetics Science that studies modifications in gene expression that are not due to an alteration in the DNA sequence and are heritable

DNA methylation Epigenetic process by which a methyl group is added to the cytosine base at position 5 of the pyrimidine ring to form 5-methylcytosine. A high degree of methylation is associated with a closed conformation of chromatin and thus with gene silencing

Covalent modifications of histones Post-translational modifications of histone proteins (H1, H2A, H2B, H3 and H4), such as methylation, phosphorylation, acetylation, deamination, ubiquitination and proline isomerization, which can modify histone configuration and alter their interaction with DNA and other nuclear proteins, thus affecting chromatin compaction and gene expression

Biomarker Any molecule or chemical structure that can be used as an indicator of a biological state (normal or pathological), the response to a dietary, pharmacological or surgical intervention, the risk of suffering a certain pathology or the future evolution of a disease

Non-coding RNA RNA molecule that is not translated into a protein, such as transfer RNA (tRNA), ribosomal RNA (rRNA), and others such as snoRNAs, microRNAs (miRNAs), siRNAs, piRNAs and long non-coding RNAs (lncRNA). Many of the latter have regulatory functions on gene expression

MicroRNAs miRNAs are single-stranded non-coding RNAs, between 21 and 25 nucleotides in length, that have the ability to inhibit the expression of other genes through protein degradation during translation, inhibition of translation elongation, premature termination of translation (by ribosome misfolding), or inhibition of translation initiation

Methyl donors Molecules involved in the process of transferring a methyl group from S-adenosylmethionine to proteins or nucleic acids. For example, betaine, choline, some vitamins (folic acid, riboflavin, vitamins B6 and B12) and amino acids (methionine, cysteine, serine, glycine)

Introduction

"Epigenetics" is defined as heritable modifications in gene expression that cannot be explained by changes in DNA sequence. Mechanisms involved cause heritable changes to cells without affecting DNA sequence (e.g., DNA methylation; histone modification; DNA replication timing; nucleosome positioning; and heterochromatization, which result in selective gene expression or repression). Besides, "epigenomics" can be defined as the systematic study of the global gene expression changes due to epigenetic processes and not due to DNA base sequence changes.

The current scientific evidence shows that nutrients are capable of modulating metabolic, cellular and molecular processes by mediating transcriptional, translational and post-translational pathways and operations, including DNA organization and conservation and gene expression. Given the incontrovertible evidence of the involvement of environmental factors, such as dietary patterns, intake of specific nutrients and physical activity, in the incidence of certain diseases including metabolic syndrome, cardiovascular diseases and cancer, a new concept termed "nutriepigenomics," has been established. Nutriepigenomics joins nutrigenomics (the study of the influence of nutrients on gene expression) and nutrigenetics (the way in which the genetic sequence conditions the individual response to nutrients) and makes it possible to explain some of the effects on human health nutrients, foods and their bioactive components, e.g., phytochemicals, through changes in epigenetic marks. Epigenetic changes are modulated by environmental exposure (including nutrition and physical activity), thus epigenetics is possibly involved in the development of aging-related and other non-communicable chronic diseases.

Epigenetic marks influence gene function and expression without modifying the primary nucleotide sequence of DNA. These include methylation and hydroxymethylation of certain nucleotides, chromatin compaction, various covalent modifications of histones, and the expression of miRNAs. This "overgenetic" signaling has made possible to show that, if there are nutritional alterations and metabolic imbalances in certain critical periods of development, the epigenetic modifications generated can induce stable changes in the differential expression of certain genes with effects on the structure or function of tissues and organs, thus predisposing to various pathophysiological disorders and diseases.

Epigenetic changes can be heritable and long-lasting. In addition, epigenetic processes are one of the ways by which environmental agents can influence gene expression and metabolic regulation throughout the life cycle, providing possible etiopathological explanations for chronic diseases typical of adulthood, whose origin may be based on physiological alterations in perinatal stages or on parental feeding behaviors during that period. Besides, epigenetic phenomena are potentially reversible, which allow us to think about possible therapeutic targets based on nutritional guidelines or personalized pharmacological therapies for their control, treatment and eventual modification. Early detection of changes in epigenetic signals (biomarkers) can also be used for the early diagnosis of diseases and precise individualized preventive action based on epigenetic biomarkers.

Epigenetic modifications and chromatin structure

Among the epigenetic mechanisms that regulate gene function there are some "canonical" modifications, including histone acetylation, histone methylation and DNA methylation, and "emerging" modifications, including histone acylation, homocysteinylation, and serotonylation (Table 1). Others recently described include changes that affect chromatin folding (euchromatin vs. heterochromatin) or chromosome stability, and, in general, those processes that affect gene expression patterns without altering or being mediated by the DNA sequence (Fig. 1).

Histones are proteins responsible for DNA packaging as well as chromatin conformation and are subject to a wide variety of post-translational modifications including arginine and lysine methylation, lysine acetylation, serine and threonine phosphorylation, or lysine ubiquitination and sumoylation, among other molecular processes. These transformations occur mostly in the amino-terminal tails of histones on the nucleosome surface, as well as in the region of the nucleus. Histone alterations can affect chromosomal function by modifications in the electrostatic charge of the histone, resulting in a structural change or its binding to DNA and modifications in the binding sites for protein recognition modules such as bromodomains or chromodomains, which recognize acetylated or methylated lysine residues, respectively. These epigenetic mechanisms, which mediate chromatin rearrangement and heterochromatin formation, influence the regulation of gene expression. Acetylation is the most studied histone modification and the main enzymes involved in this process are histone acetyltransferases and histone deacetylases (HATs and HDACs, respectively).

DNA methylation occurs at cytosine bases which are converted to 5-methylcytosine by a DNA methyltransferase (DNMT). Cytosine residues with an added carbon ($-CH_3$) are usually located next to a guanine (CpG dinucleotide), resulting in two diagonally arranged, side-by-side methylated cytosine residues opposing the complementary DNA strands. Different DNMT-type enzymes are

Table 1 Types of epigenetics modifications.

"Canonical" modifications	
Reaction substrates	*Epigenetic modification*
SAM	DNA methylation
SAM	Histone methylation
Acetyl-CoA	Histone acetylation
"Emerging" modifications	
Reaction substrates	*Epigenetic modification*
Lactate	Histone lactylation
Succinyl-CoA	Histone succinylation
Benzoyl-CoA	Histone benzoylation
Crotonyl-CoA	Histone crotonylation
Butyrate	Histone butyrylation
β-Hydroxybutyrate	Histone β-hydroxybutyrylation
Homocysteine	Histone homocysteinylation
Serotonin	Histone serotonylation
Dopamine	Histone dopaminylation
UDP-GlcNAc	Histone O-GlcNacylation
NAD+	Histone ADP-ribosylation
Methylglyoxal	Histone methylglyoxal adduction
RNA modifications	
Reaction substrates	*Epigenetic modification*
SAM	RNA methylation
Acetyl-CoA	RNA acetylation

ADP, Adenine diphosphate; CoA, Coenzyme A; NAD, Nicotinamide adenine dinucleotide; SAM, S-adenosylmethionine; UDP-GlcNAc, Uridine diphosphate N-acetylglucosamine.

involved in these processes, either by de novo DNMTs, incorporating the initial pattern of methyl groups on a DNA sequence where there was no methylation before, or by maintenance DNMTs, copying the methylation pattern from a pre-existing DNA strand after cell replication. It has been proposed that the methyl group can intercept the DNA binding of certain transcription factors that control the transcriptional process. In some cases, genes that are more highly methylated (specifically in the promoter region) tend to be less accessible, often resulting in inhibition of their expression. DNA methylation in non-promoter regions may also be related to various functional tasks. For example, DNA methylation of intragenic regions may modulate gene expression by acting as alternative promoters. However, the correlation between DNA methylation of intragenic regions and gene expression is unclear, as controversial findings have been published in different cells and tissues. The evaluation of DNA methylation as a cause or biomarker of disease is consistent in some tumor types or obesity and diabetes.

The scientific literature also refers to several types of non-coding RNAs as epigenetics modifications, such as long non-coding RNAs (lncRNA), Piwi-associated RNAs (piRNA) or circular RNAs (circRNA). All these RNAs have in common that they do not code for proteins and appear to play roles in the control of gene expression, that's why they are recognized as epigenetic modification. The best known of which are microRNAs (miRNA), which are small non-coding RNA molecules (about 22 nucleotides) whose mission is usually the silencing of messenger RNA or the post-transcriptional regulation of protein synthesis. Several studies have shown that miRNAs are involved in the incidence of various diseases (from different types of tumors to obesity, diabetes and cardiovascular disease) and have high potential as biomarkers for both prognosis and diagnosis and also as possible therapeutic agents and targets, given that nutrients, hydration and physical activity can modulate the expression and actions of some of these RNAs.

Nutrient metabolic regulation by epigenetic mechanisms

Nutrients such as glucose, fatty acids, amino acids and vitamins are utilized by cellular metabolic pathways to produce metabolites that are used as substrates or activity modulators of chromatin-modifying enzymes (**Fig. 2**). These molecules are included in the regulation of a plethora of both "canonical" modifications, including histone acetylation, histone methylation and DNA methylation, and "emerging" modifications, including acylation, homocysteinylation, serotonylation and so forth (Dai et al., 2020). Pathways related to central-carbon, one-carbon and methionine metabolism, acetate metabolism, ketogenesis, and redox balance feed the pools of several of these metabolites and thus help to regulate the epigenomic landscape, in concert with chromatin modifiers, remodelers and transcription factors.

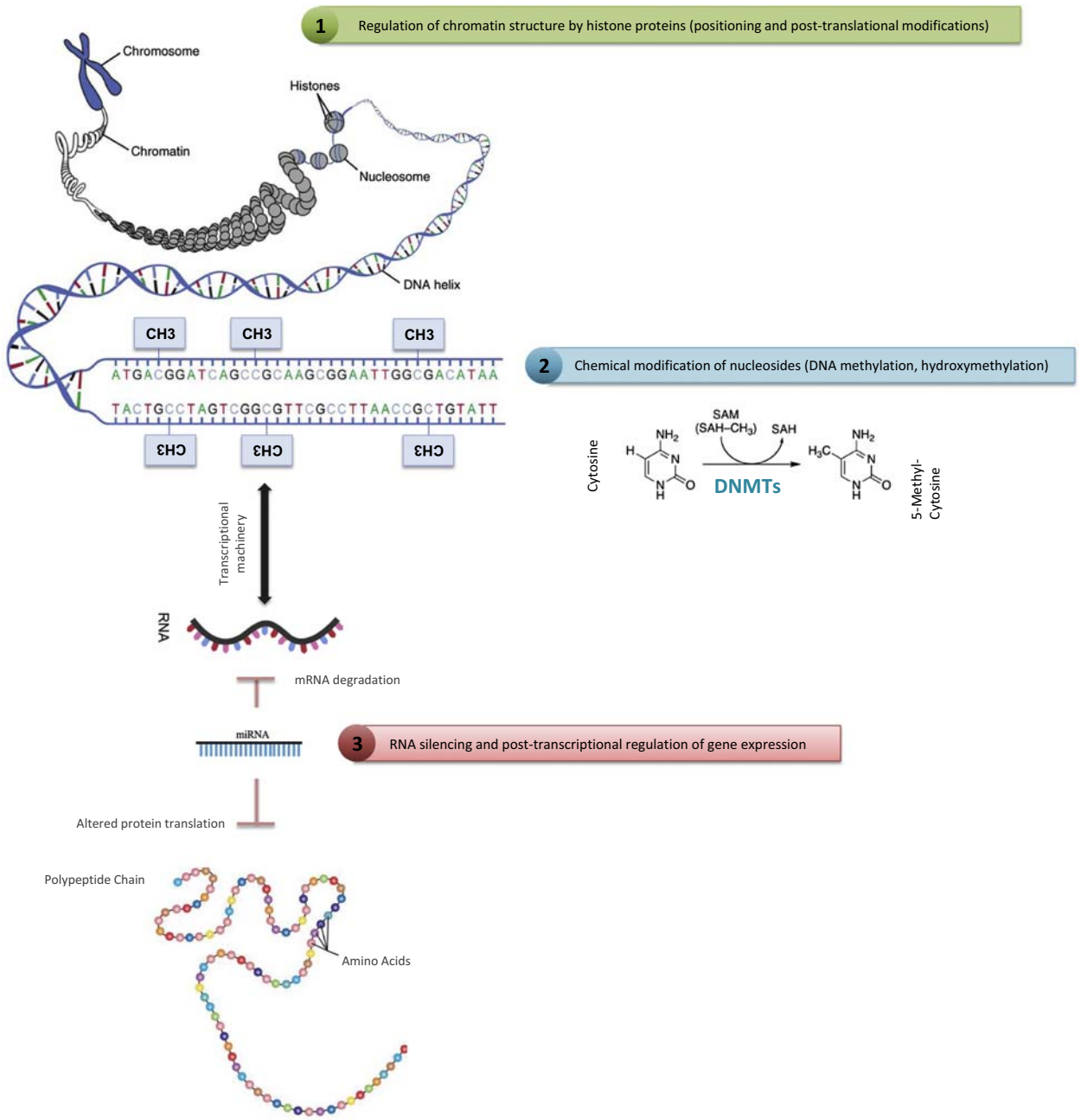


Fig. 1 The main forms of epigenetic modifications. Modification of nucleosides in DNA such as by methylation and hydroxymethylation. Post-translational modification of histone proteins by methylation, acetylation, ubiquitylation, SUMOylation, citrullination and ADP-ribosylation. Changes in small non-coding RNA expression. Nucleoside and histone modifications regulate gene transcription by modulating the conformation of the chromatin and the access of DNA binding factors. Small non-coding RNAs (such as microRNAs) regulate gene expression by prompting mRNA degradation or modulating protein translation.

The ability of epigenetic modifications to respond to fluctuations in metabolic activities is a consequence of the intrinsic thermodynamic and kinetic parameters of different chromatin-modifying enzymes. The addition and removal of most of these modifications are catalyzed by enzymes (known as, "writers" and "erasers") that utilize metabolites as substrates or cofactors (known as chromatin-modifying metabolites). The intracellular abundance of chromatin modifying metabolites is regulated by several mechanisms. Metabolites that are taken up by cells can passively or actively diffuse through the plasma and nuclear membrane to modify chromatin. Alternatively, metabolites can be processed internally by the activity of metabolic enzymes that convert them into substrates or cofactors for chromatin-remodeling enzymes. The metabolic enzymes can also translocate to the nucleus, where they can locally produce substrates for chromatin modification. The resulting consequences of metabolite abundance for the rate of chromatin modification are dependent on the kinetic and thermodynamic parameters of the particular enzyme. These

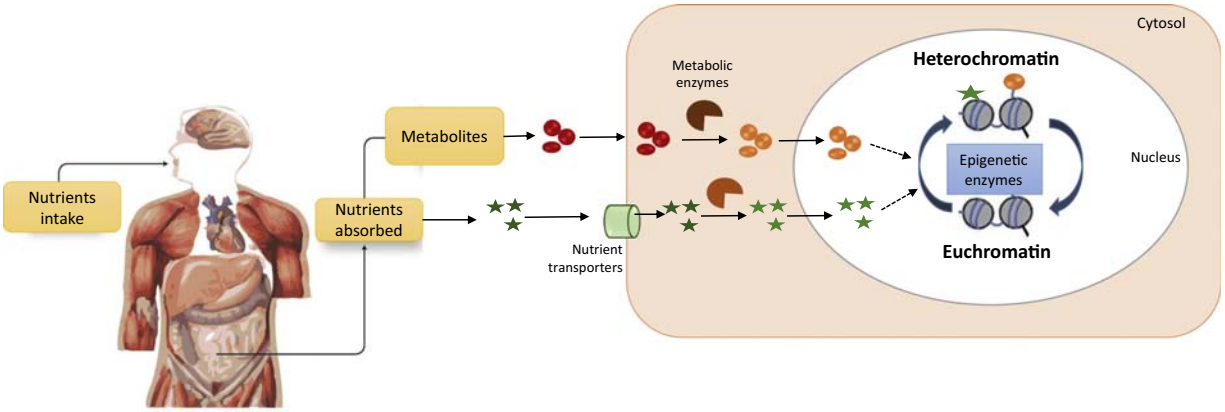


Fig. 2 Metabolic regulation by epigenetic mechanisms. Nutrients are utilized to produce metabolites that are taken up by cells and can passively or actively diffuse through the plasma and nuclear membrane in order to modify chromatin. Alternatively, metabolites can be processed internally by the activity of metabolic enzymes that convert them into the substrates or cofactors for chromatin-remodeling enzymes.

enzymes include methyltransferases and acetyltransferases, among others. Finally, once the modifications have been deposited, effector proteins can recognize and bind to them using specific protein-binding modules, through which the markings determine a variety of intracellular fates, involving processes such as development, immune-system regulation or disease development.

A good example of metabolic enzymes involved in the synthesis of chromatin modifying metabolites, is acetyl-CoA and S-adenosylmethionine (SAM), which localize to the nucleus and interact with nucleosomes and chromatin-modifying enzymes. Levels of chromatin-modifying metabolites such as SAM are controlled by multiple mechanisms, including both environmental inputs such as nutrient availability and intracellular methyl group sinks that consume SAM. Methyl group sinks are mediated by enzymes that metabolize SAM, allowing them to divert methyl groups away from enzymes such as histone or DNA methyltransferases, thus affecting the activity of these enzymes. These mechanisms provide avenues for the control of metabolite levels and thus for chromatin to sense intracellular metabolic status.

Epigenetics and dietary intervention

Due to the plasticity of epigenetic factors, environmental changes, such as dietary interventions, which alter food intake and composition or meal timing, have a significant impact on the epigenome. Dietary interventions, including caloric restriction, intermittent fasting or time-restricted feeding, have shown to induce promising improvements in patients' overall metabolic profiles (i.e., reduced body weight, improved glucose homeostasis), and an increasing number of studies have associated these beneficial effects with epigenetic alterations (Asif et al., 2020; ElGendy et al., 2018).

Hypercaloric and high-fat diets

The response to high-fat diets (HFD) in animal models is associated with changes in the expression of hypothalamic neuropeptides, which regulate energy metabolism and appetite, as well as gene expression in hepatocytes and adipocytes where different epigenetic processes could participate in this adaptation. Additionally, in adult rodents, it has been shown that long-term HFD feeding has an effect on the methylation of obesity-related genes such as leptin (*LEP*) in adipose tissue, or melanocortin 4 receptor (*MC4R*) in the brain, probably contributing to changes in gene expression and appetite regulation. Moreover, while fasting decreases the amount of acetylated histone H3 and H4-positive cells in the ventrolateral subdivision of the ventromedial hypothalamus, dietary consumption of an HFD pattern for four weeks results in overexpression of histone deacetylases HDAC5 and HDAC8. These studies demonstrate that hypercaloric diets alter some epigenetic mechanisms related to the regulation of the expression of genes involved in the control of energy homeostasis.

As gestation and lactation are the periods of life most susceptible to epigenetic modifications, and with effects that persist into adulthood, a series of studies have been performed during the perinatal period, especially in rodents, which have proven that maternal overfeeding with HFD in rats induces overweight in the offspring independently of postnatal nutrition and also nonalcoholic steatohepatitis, explained by an alteration of mitochondrial metabolism and lipogenesis in the liver. The origin of other metabolic diseases in adults, such as hypertension, hyperlipidemia or insulin resistance and diabetes, may also be related to maternal overnutrition. These metabolic alterations in the offspring appear to be attributable, at least in part, to epigenetic modifications that persist into adulthood. Thus, neonatal overfeeding alters the DNA methylation patterns of the hypothalamic promoter region of the major anorexigenic neuropeptide, POMC. Maternal consumption of HFD can modify epigenetic marks in the offspring's brain and alter the expression of genes related to the mesocorticolimbic reward circuitry (dopamine and

opioids), inducing a preference for foods richer in sucrose and fat. Interestingly, these epigenetic modifications persist for at least two generations and may contribute to understanding part of the increased prevalence of obesity observed in most countries. However, not all studies reached the same conclusion, as there are examples in which an HFD intake during gestation and lactation at least partially protects the offspring from excessive weight gain through different hypothalamic mechanisms. Therefore, further studies are needed to elucidate the intrinsic mechanisms related to these effects, as well as complementary studies in humans (Cai et al., 2021). Within this context, unbalanced nutrition can lead to changes in epigenetic marks that may contribute to the development of metabolic complications. For example, an epigenome-wide screen identified 625 significant differentially methylated regions (DMRs) associated with diet-induced obesity phenotypes, of which 232 DMRs correlated with HFD alone, and 249 regions were conserved in adipose tissue from obese subjects. Another human study showed that acute intake of a HFD induced changes in the methylation of 6508 genes in skeletal muscle. In the same nutritional intervention, the observed changes did not reverse after 6–8 weeks of a normocaloric diet, suggesting that the recovery of epigenetic marks may be slow and that the set of modifications may influence gene expression levels over time. This finding could help to explain the "yo-yo" effect, which is characterized by an increasing difficulty to lose weight after each stage of treatment with low-calorie diets. However, a study in rats has demonstrated the reversibility of some epigenetic marks: the intake of a hypercaloric diet for 20 weeks altered the methylation of several CpG sites located in the promoter of the leptin gene in visceral adipose tissue. When the hypercaloric diet was replaced by a normocaloric diet, the original methylation levels for some of the CpG sites of the leptin promoter were recovered. This experiment confirms the possible reversibility of phenotypic and epigenetic changes induced by hypercaloric diet intake. However, more meticulous and longer-term follow-up would be needed to ensure that the initial pre-hypercaloric diet methylation levels can be reached for all genes.

Caloric restriction and intermittent fasting

Numerous epidemiological and animal studies link suboptimal early nutrition and poor intrauterine growth with an increased risk of hypercholesterolemia, hypertension, type 2 diabetes and obesity in adulthood. In relation to the multiple mechanisms examined, it is clear that epigenetics plays a key role. Both energy restriction and low protein diets induce epigenetic modifications and metabolic alterations that persist into adulthood. Therefore, a moderate restriction of caloric intake during the periconceptional period is considered a stress factor and is accompanied not only by an increase in adrenal mass and an exacerbation of the cortisol stress response but also by epigenetic modifications such as decreased methylation in the *IGF2* and *H19* genes in the adrenal gland. As noted above, a present classic example of an association between periconceptional energy restriction and DNA methylation in humans comes from the analysis of people born during the Dutch Hunger Winter (between 1944 and 1945), which found persistent epigenetic differences associated with prenatal exposure to famine and increased risk of chronic disease in the 1970s by those individuals.

Furthermore, in adult mice and humans, it has been shown that weight loss, as a consequence of hypocaloric diets, may alter the DNA methylation pattern of different tissues, mainly adipose tissue, liver, pancreas or brain, among others. In adult mice, caloric restriction seems to reprogram orexigenic pathways and alter the reward circuitry in the brain by affecting epigenetic mechanisms. In these animals, caloric restriction promotes binge eating, and this result opens perspectives to study whether the yo-yo effect, which makes weight loss maintenance so complicated, maybe, at least in part, mediated by epigenetic mechanisms. However, it is difficult to determine whether these epigenetic effects are caused directly by caloric restriction or weight loss. Female obese patients subjected to bariatric surgery with significantly reduced body weight (27%) and food intake show reductions in global DNA methylation levels and differentially methylated genes associated with obesity and T2D in adipose tissues, thus providing context for weight loss and adipocyte reprogramming. These genes are associated with the regulation of body weight (*LEPR*, *FTO*), cholesterol homeostasis (*CETP*, *LCAT*), blood glucose (*IRS1*, *INSR*), adipose tissue function (*mTOR*, *RPTOR*), and epigenetic modifications (*FOXP2*, *HDAC4*, *DNMT3B*).

Modulation of lipid compartmentalization and efficient utilization of excess energy in adipose tissues are critical targets for the treatment of obesity and related metabolic dysfunctions. Dietary interventions, including intermittent fasting and caloric restrictions, markedly reduce adipocyte size and weight depots in rodent models of obesity, and confer improvements in adipose tissue inflammation and insulin sensitivity. Additionally, adipose tissue thermogenesis, via the induction of WAT "browning" (beige fat), and activation of BAT appear to be predominant pathways which elevate energy expenditure, mitochondrial biogenesis and energy dissipating capacity. However, studies investigating DNA methylation changes in adipose tissue upon dietary interventions are limited. In one study, obese women on a 6-months caloric restriction (1100–1800 kcal/day) who lost >3% of their body fat showed hypermethylation at three genomic loci in their subcutaneous adipose tissue. These loci were associated with lipid (e.g., *PLCL4*) and glucose (e.g., *ENC1*) homeostasis and epigenetic regulation (e.g., *PRDM8*). In another study, 36 h of fasting in young, healthy men increased DNA methylation at the promoter site of *LEP* in subcutaneous adipose tissue, leading to a 3-fold decrease in plasma leptin levels. In addition to DNA methylation, histone acetylation and deacetylation in adipose tissues are associated with the beneficial metabolic effects seen with dietary interventions. For example, 30% caloric restriction in HFD-fed mice led to a significant increase in histone 4 acetylation (H4ac) at the *Glut4* promoter, which is associated with increased *Glut4* mRNA expression in white adipose tissue and decreased plasma glucose levels. Additionally, several HATs and HDACs are involved in regulating adipose thermogenesis, which is one of the primary beneficial mechanisms of dietary interventions.

Mediterranean diet

There is currently interest in knowing the influence of the different components of the Mediterranean diet on the beneficial effects of this dietary pattern for metabolic syndrome and cardiovascular disease and the epigenetic mechanisms potentially involved. In fact, some experiments have observed indications that some compounds frequently found in the Mediterranean dietary pattern, such as choline, but above all betaine, could be involved in the health effects through changes in DNA and histone methylation. Besides, the expression of some miRNAs also seems to be related to the beneficial effects of the Mediterranean diet. GWAS genomic studies have described an SNP of the *LPL* gene (rs13702) in the binding site for the miR-410 associated with circulating triglyceride concentrations and HDL cholesterol. In a nutritional intervention with a Mediterranean diet and a control (PREDIMED study), subjects carrying the C allele reduced their hypertriglyceridemia and their risk of cerebral infarction more when they followed a Mediterranean pattern especially rich in unsaturated fatty acids. With this initial knowledge, new findings are expected to associate the Mediterranean diet or some of its specific components with epigenetic "signatures" that contribute to explain at least partially the mechanisms of action by which this dietary pattern mediates its beneficial effect. In fact, in a very recent PREDIMED analysis (Arpón et al., 2017), specific components of the Mediterranean diet (MedDiet), particularly nuts and extra-virgin olive oil, were able to induce methylation changes in several peripheral white blood cell genes and modulate exosomal RNA content, with the former affecting a higher number of miRNAs (Mantilla-Escalante et al., 2021). These changes may have potential benefits in health, especially those changes in genes related to intermediate metabolism, diabetes, inflammation and signal transduction, which may contribute to explain the role of MedDiet and fat quality on health outcomes.

Nutritional factors involved in epigenetic regulation

Given the responsiveness of epigenetic marks to dietary factors, research has attempted to relate the administration of compounds of plant origin with effects, i.e., phytochemicals, on epigenetic processes to apply them in the prevention and treatment of different diseases. In the future, it may be possible to speak of "epigenetic foods" as a group of functional foods containing bioactive compounds capable of modulating the expression of miRNAs and DNA methylation or inducing covalent modifications in histones. As in other epigenetics-related topics, this concept has been developed mainly about cancer prevention so that diverse plant components are being studied in relation to apoptosis, cell cycle regulation, differentiation, inflammation, angiogenesis, autophagy and metastasis, as well as in stress response (Kumari et al., 2020). Similarly, understanding the epigenetic effects of botanical compounds consumed with the diet may provide insights into prevention strategies to reduce the prevalence of metabolic diseases and prevent their comorbidities.

Among the components (nutrients and non-nutrients) of food that have been associated with epigenetic changes are methyl group donor molecules (methionine, choline, folic acid, betaine and vitamins B2, B6 and B12, which are involved in the methionine cycle), carbohydrates and proteins, short-chain fatty acids (butyrate, acetate and propionate) or omega-3/6 fatty acids, some minerals and antioxidant vitamins (vitamins A, E and C), as well as different bioactive compounds of plant origin, including various polyphenols, catechins, isoflavones and isothiocyanates.

Methyl donor molecules

Dietary methyl groups are derived from foods containing methionine, serine, folic acid, biotin and choline, among other compounds, which are molecules that allow the transfer of a methyl group to DNA and histones via S-adenosylmethionine (SAM) (Mahmoud and Ali, 2019). The synthesis of methionine from homocysteine requires zinc, selenium and vitamins B6 and B12. Thus, methionine can be synthesized in the liver from homocysteine using methyl groups of betaine (derived from choline) or methyltetrahydrofolate. Therefore, when people have low levels of choline, dietary folate requirements increase and, conversely, when there is a folic acid deficiency, dietary choline requirements increase. However, not all individuals tend to develop similar health problems with methyl donor intake because other factors may be involved. In this sense, the presence of SNPs in genes involved in the regulation of choline and folic acid metabolism may increase or decrease the risk of functional methyl group deficiency and interfere with the proper methylation of macromolecules.

The importance of maternal methyl donor intake on germline epigenetic status was first published in 2006 for the *agouti Avy* mouse model. Another subsequent paper (2008) concluded that methyl donor supplementation induces DNA hypermethylation during development and prevents transgenerational amplification of obesity in *agouti* mice. Also, a methyl donor-deficient diet has been established as a good model for the induction of nonalcoholic steatosis in rodents, which is accompanied by numerous epigenetic and transcriptomic changes. Following these findings, many studies have analyzed the epigenetic effects of methyl donor-deficient or methyl donor-supplemented diets during pregnancy on different metabolic diseases, including possible alterations in offspring behavior (e.g., anxiety) through permanent changes in hippocampal DNA methylation. However, few studies have associated perinatal methyl donor administration with long-term changes in body weight and metabolism. In animal models, perinatal dietary methyl donor deficiency has been associated with increased adult adiposity, insulin resistance and hypertension in offspring, which is accompanied by alterations in global DNA methylation. Low maternal dietary choline content at the embryonic

stage in mice results in a significant reduction in the degree of genomic DNA methylation in the fetal hippocampal neuroepithelial layer, and the methylation level of protein kinase inhibitor (Cdkn3), the vicinity of the promoter of the vascular endothelial growth factor C and angiopoietin 2 genomic DNA also significantly decreases (Cai et al., 2021). However, few such studies have been carried out in humans. Therefore, further longer-term nutritional intervention trials are needed to understand the benefits and risks of methyl donor doses in relation to the development of obesity, even though nutrient supplementation during pregnancy is widespread in the fight against anemia and other micronutrient deficiencies.

The intake of methyl donors in adulthood also seems to be a determinant for the subsequent development of metabolic disorders. Thus, in adult mice, a diet deficient in methyl groups resulted in progressive fat accumulation and morphological changes in the liver similar to human nonalcoholic steatohepatitis, accompanied by important epigenetic alterations, such as aberrant histone modifications and loss of DNA methylation, especially in satellite regions. In patients with nonalcoholic fatty liver disease, a choline deficiency is often accompanied by increased fibrosis. Interestingly, it has been shown that fetal epigenetic programming is reversible in adult life by methyl donor supplementation. Thus, in rodents, maternal programming of the stress response is reversed by central infusion of methionine, whereas methyl donor supplementation prevents HFD-induced nonalcoholic fatty liver disease. Nevertheless, further studies both in animal models and human intervention are needed to precisely define the effect of supraphysiological intake of methyl donors on chronic diseases and be cautious in recommending these compounds during pregnancy. In this regard, some complementary studies have focused on the beneficial effects of betaine administration on adipose tissue dysfunction and insulin resistance and of folic acid supplementation in obese subjects with type 2 diabetes.

Fatty acids and amino acids

In addition to methyl donors, other nutrients act as regulators of DNA methylation and histone covalent modifications, either by directly inhibiting enzymes that catalyze the processes or by altering the availability of substrates required for the corresponding reactions. The ability of different types of fatty acids to induce epigenetic modifications is less well known. Concerning the effects of *n*-3 polyunsaturated fatty acids (PUFA) on the DNA methylation status, a possible mechanism that has been proposed is that *n*-3 PUFA enhance the expression of DNMTs and consequently the DNA methylation. Furthermore, a potential interaction between *n*-3 PUFA and MeCP2 (methyl CpG binding protein 2) has been proposed, mainly in promoter regions, and consequently could be associated with the regulation of gene expression. Another possible mechanism by which *n*-3 PUFAs can affect methylation is that these FA are natural ligands of some transcriptional factors, such as PPARγ. In this context, it has been reported that interactions between PPARγ and fatty acids result in a decrease in cytokine expression, and in murine models, *Pparg* expression is modulated by DNA methylation in its promoter region. However, more studies are needed to elucidate the role of FA in the regulation of epigenetic mechanisms in the context of metabolic alterations and chronic diseases. Regarding the other types of FA, a specific mechanism in which they could alter epigenetic landmarks has not been described (González-Becerra et al., 2019).

A special case is that of butyric acid, a short-chain fatty acid produced during fermentation (mediated by gut microbiota) of dietary fiber sources. This molecule induces apoptosis in vitro and cell cycle arrest, presumably through inhibition of HDACs. It has been described how butyrate triggers histone H3K9ac to activate steroidogenesis through PPARγ and PGC1α pathways in ovarian granulosa cells (Ye et al., 2021). As obesity is associated with major changes in the composition of the gut microbiota, butyrate-producing bacteria could be one of the factors linking both phenomena through epigenetic mechanisms. Other short-chain fatty acids produced by the gastrointestinal microbiota have also been associated with epigenetic changes. Thus, acetate, which enters the peripheral circulation to be metabolized by peripheral tissues, increases histone acetylation in nervous tissue and inhibits HDAC activity and expression in the brain, whereas propionate is involved in the propionylation of lysine 23 of histone H3 in different cell lines.

Moreover, the amino acid that seems to play the most determinant role in epigenetic mechanisms is methionine, an important source of methyl groups in biomethylation reactions and is key in regulating the monocarbon metabolism pathway. However, other amino acids (particularly serine, glycine and histidine), together with some vitamins, could also play an important role in providing methyl donors for DNA and histone methylation.

Minerals and vitamins

A study of cord blood samples from the offspring of Gambian women has shown that periconceptional micronutrient supplementation, providing 14 minerals and vitamins, affects fetal methylation of some promoters in genes that, in some cases, persist in children until at least 9 months of age with clear differences between sexes. This strategy may not be adequate to prevent obesity and other chronic diseases since multivitamin supplementation in rats during gestation has been associated with increased food intake and the development of obesity in the offspring when fed HFD, which could be mediated by epigenetic mechanisms.

Specifically, some minerals have been associated with changes in epigenetic mechanisms that regulate gene expression. For example, selenium and zinc are involved in the regulation of DNMTs activity and play a key role in the one-carbon metabolism and in the activation of most HDACs. High intakes of various inorganic and organic forms of heavy metals, including chromium, arsenic, lead, cadmium, copper and nickel, have also been linked to epigenetic effects. Selenium, through selenoproteins, has been

related to some pathologies such as cancer, cardiovascular and autoimmune diseases. Although the importance of the interaction of selenium with the epigenome for human health is debated, epigenetic effects caused by selenium are not ruled out. Magnesium is another element that is capable of modifying epigenetic marks. Thus, magnesium deficiency in pregnant rats induces metabolic complications in the offspring by altering the methylation of specific cytosines in the hepatic 11β-hydroxysteroid dehydrogenase-type-2 (*HSD11B2*) promoter. Increased pre-conceptional and prenatal chromium exposure could lead to epigenetic control of endocrine imbalances and some metabolic functions.

Calcium is a mineral that appears to be related to weight regulation and other biochemical functions. However, this element has only been indirectly associated with epigenetic modifications and many more studies are needed to rule out that calcium could be a metabolic regulator through epigenetic mechanisms.

Some vitamins, including A, C and E, and carotenoids can decrease circulating concentrations of inflammatory and oxidative markers. Consequently, studies have been conducted with these molecules to assess whether they can induce epigenetic changes. On this regard, vitamin C promotes widespread DNA demethylation in human embryonic stem cells, while retinol alters histone phosphorylation levels in rat Sertoli cells. The epigenetic effects of vitamin E and lipoic acid require further analysis, although molecular modeling studies suggest that both may have a role as HDAC inhibitors. In recent years, the role of vitamin D on DNA methylation and other epigenetic mechanisms has gained special relevance (Ong et al., 2020). Observational and randomized controlled human trials have proven the vitamin D effects on DNA methylation. In this sense, different mechanisms have been proposed to account for these effects, mainly the altered expression of genes that directly or indirectly affect DNA methylation (*Bhmt1*, *DNMT1*, *DNMT3*, *TET2* and *TET3*). In summary, although further studies in animal models are still needed to understand the basic mechanisms of specific nutrients, current evidence suggests that dietary supplementation with certain vitamins and minerals could be useful as epigenetic therapy for metabolic disorders.

Polyphenols and other plant compounds

Polyphenols and other plant compounds are good candidates to show epigenetic properties for their potential applications as therapeutic agents that can be used to prevent or treat inflammation and oxidative stress, as well as health problems related to cancer and metabolic diseases such as type 2 diabetes, atherosclerosis, and hypertension. Among the plant-derived bioactive compounds that mediate epigenetic modifications are agents such as genistein (soybeans), resveratrol (grapes), curcumin (turmeric), catechins (green tea) and sulforaphane (cruciferous vegetables), which have been considered in cancer prevention and therapy because of their possible epigenetic interaction, by modulating DNMT, HAT and HDAC activities and endogenous miRNAs synthesis (Cione et al., 2019). Therefore, numerous compounds and plant extracts capable of modulating the activity of enzymes that regulate histone covalent modifications and DNA methylation have been studied in recent years. Some of these compounds about epigenetic modifications are shown in **Fig. 3**.

Tea catechins, the most abundant being epigallocatechin-3-gallate (EGCG), are a group of flavonoids considered as one of the most promising dietary treatments for metabolic syndrome. The administration of EGCG to mice significantly reduces body weight gain, blood glucose levels and insulin resistance, decreases liver damage and triglyceride levels in this organ and reduces plasma cholesterol and inflammatory cytokines. Some of these effects are undoubtedly mediated by epigenetic mechanisms such as the direct inhibition of DNMT by interacting not only with the catalytic site of the DNMT molecule but also by inhibiting acetyltransferase activity in histones.

Genistein is a flavonoid present in soybeans that can act as an endocrine disruptor. However, at certain doses and in a sex- and age-dependent manner, it may also inhibit adipogenesis in vitro in a manner similar to estrogens. A pioneering study demonstrated that supplementation of the maternal diet with genistein during early embryonic development can change the color of yellow *agouti* heterozygous (Avy/a) mice, which was significantly associated with increased methylation at a retrotransposon of the transcriptional start site of the *agouti* gene. In this same model, genistein decreased ectopic *agouti* gene expression and protected offspring from obesity in adulthood by altering the epigenome. Similarly, in non-human primates, consumption of soy protein and isoflavones improved body weight, insulin sensitivity and lipid profiles, accompanied by modifications in DNA methylation patterns in the liver and muscle.

Curcumin is a polyphenol with potent anti-inflammatory activity. It is also one of the natural compounds with the greatest potential as an anti-obesity treatment because, in vitro, it can suppress 3T3-L1 differentiation and induce apoptosis in adipocytes (Hassan et al., 2019). Curcumin is able to modulate histone deacetylases and acetyltransferases, DNA methyltransferase I activity and the expression of miRNAs, the involvement of curcumin-induced epigenetic mechanisms in animal models. It is an effective medicinal agent, including decreased HAT activity and HDAC2 activation, as it regulates several important molecular signaling pathways that modulate survival, govern anti-oxidative properties like nuclear factor E2-related factor 2 (Nrf2) and inflammation pathways, e.g., nuclear factor kappa B (NF-κB).

Resveratrol is a stilbenoid with potent free radical scavenger properties that show beneficial effects on type 2 diabetes and cardiovascular disease, sharing some healthful actions with caloric restriction. Among other consequences, this compound decreases the severity of hepatic steatosis in rats and exerts an inhibitory effect on insulin secretion. Apart from its potent antioxidant activity, resveratrol is an activator of SIRT1, an NAD(+)-dependent histone deacetylase that modulates gene expression in all tissues.

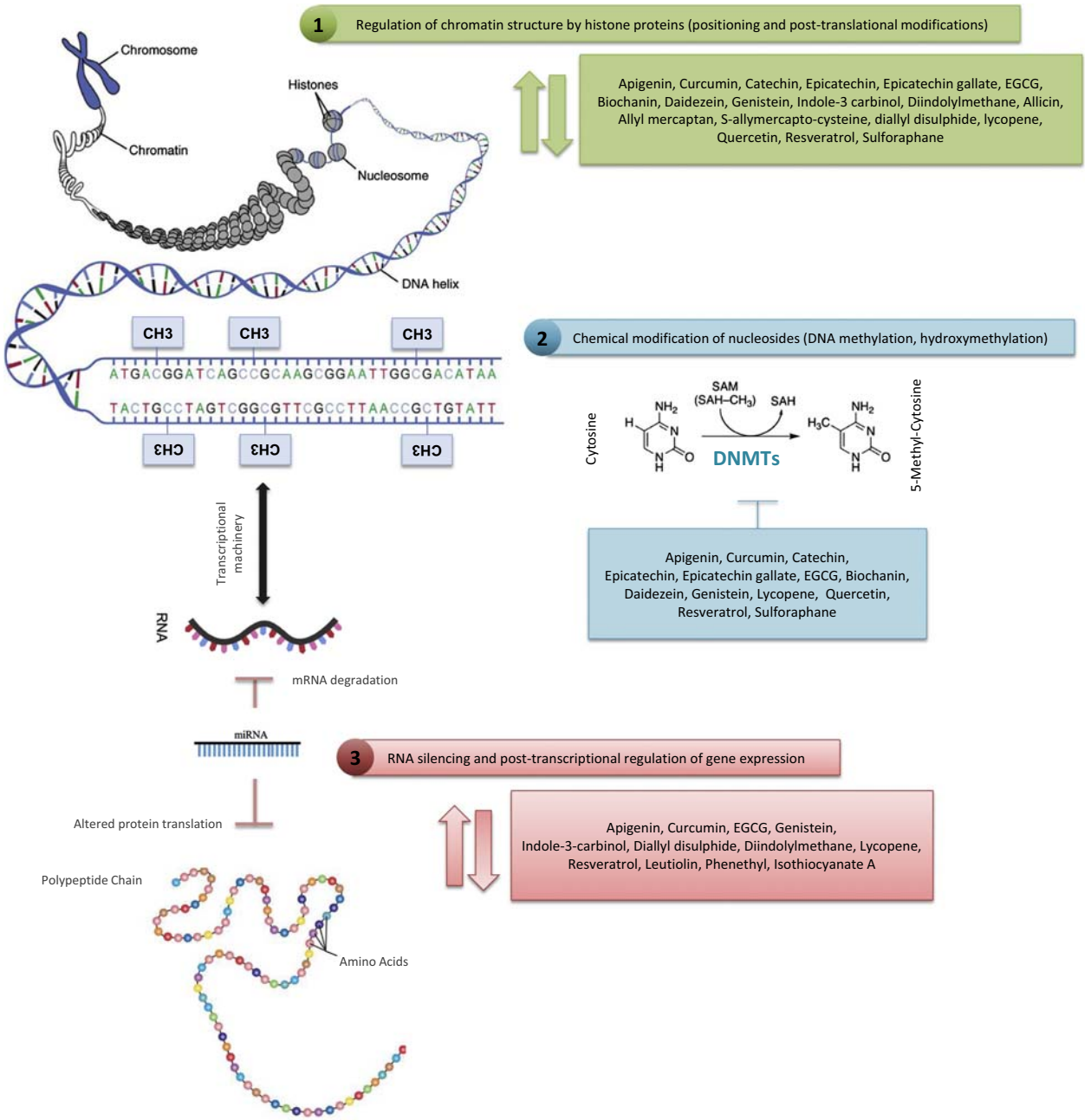


Fig. 3 Dietary phytochemicals regulate gene expression through epigenetic mechanisms by altering DNA methylation patterns, histone modification, and changes in non-coding (micro) RNAs levels. Up and down *arrows* indicate abundance levels and ⊢ refers to inhibition.

Different organic sulfur compounds have shown certain anticancer effects, in part by acting through inhibition of HDAC activity. In this regard, the most interesting natural molecules are sulforaphane, an isothiocyanate present in cruciferous vegetables that exerts beneficial properties in colon, prostate and breast cancer cells, and diallyl disulfide, which metabolizes to allyl-mercaptan and induces histone acetylation in human colon cancer cells. Although there are no studies that have reliably demonstrated the protective effects of sulforaphane on obesity in animal models, this molecule is capable of inhibiting the differentiation of adipocytes in culture through cell cycle arrest and, in relation to diabetes, reducing the production of free radicals and the consequent inflammatory damage in diabetic neuropathy and nephropathy. Therefore, further studies are needed to investigate the role of sulfur compounds in the regulation of adipogenesis and body weight, as well as the involvement of epigenetic mechanisms in the process.

Other bioactive compounds (lycopene, garcinol, luteolin, buteine, apigenin, silymarin, rosmarinic acid, anacardic acid or baicalein, among others) could inhibit the activity of DNMTs or modify histone acetylation, and consequently they could be beneficial

for the treatment or prevention of some metabolic disorders. Many studies have recently demonstrated that olive oil, especially EVOO phenolic compounds, are bioactive molecules inducing epigenetics modifications such as miRNAs expression, DNA methylation and histone modifications with anti-cancer, anti-inflammatory, anti-aging and neuroprotective activities (Fabiani et al., 2021).

The low bioavailability of most polyphenolic compounds and the impossibility of accurately determining the amounts present in food makes it difficult to establish the actual doses and effects of polyphenols on DNA methylation in humans. Therefore, new long-term studies should be designed to test the effects of different amounts of pure compounds, both in animals (periconceptionally, growing or adult) and in humans. In addition, the design of new compounds that enhance the stability and bioavailability of natural molecules is an important objective in this field. On the other hand, it seems likely that synergies may exist between different phytochemicals and nutrients, so combinations should be investigated. Finally, the combination with an epigenetic perspective of plant-derived compounds and synthetic drugs could constitute a new therapeutic approach.

Conclusions and future perspectives

To conclude, three core ideas could be highlighted: (1) epigenetic mechanisms are involved in the appearance and development of nutrition-related metabolic dysfunctions; (2) epigenetic marks could be useful for the prognosis and early diagnosis of nutritional diseases; and (3) nutritional or pharmacological agents could be used as new therapeutic strategies thanks to their capacity to induce epigenetic modifications. In the coming years, research on epigenetic mechanisms related to obesity, diabetes, and dyslipidemias could help prevent and control excessive fat deposition, glucose intolerance, and the development of cardiovascular disease through precision nutrition that takes into account personalized epigenetic aspects of nutritional homeostasis. Obviously, more in-depth studies suitably designed concerning the "epigenetic diet" are necessary to establish whether some nutrients or food components actually have a protective or adverse effect through the modulation of epigenetic processes.

In recent years, numerous omics techniques have been developed for the analysis of epigenetic modifications. With the rapid advancement of high-throughput sequencing technologies, epigenetic variations, especially variations in DNA methylation, have been extensively reported to be associated with a wide range of traits in humans, including not only phenotypes and diseases, but also behaviors and environmental exposures. Epigenome-Wide Association Studies (EWAS) have become a powerful approach to identify DNA methylation variations associated with biological traits. For genome-scale profiling of histone modifications, although chromatin immunoprecipitation followed by sequencing (ChIP-seq) is still the most widely used technique, alternatives have been developed to increase the coverage and resolution of epigenomic profiles in both bulk tissues and single cells. A new chromatin-profiling technology termed "cleavage under targets and release using nuclease" (CUT&RUN) is being used, in which DNA fragments bound to the modified histones are directly cleaved and released, instead of undergoing crosslinking, sonication and immunoprecipitation as they do in standard ChIP-seq. This new analysis has shown increased signal-to-noise ratio, efficiency and resolution and has enabled the profiling of chromatin modifications in very small numbers of cells. Based on CUT&RUN, techniques for chromatin profiling at the single-cell level recently have also been developed.

The coming years will undoubtedly offer remarkable advances in this field, although long-term experiments will probably be necessary to define new epigenetic paradigms. Selecting a multi-omics approach compared to a single-omic analysis offers some profound advantages on this line, allowing for the identification of associated factors from different biological processes, i.e., gene expression, protein synthesis and post-translational modifications, cellular metabolic processes, glycosylation, etc., maximizing the available information, and thus, increasing the possibility of identifying the root effect of the nutritional agent. For example, a single genetic variant or DNA methylation may be weakly associated with the pathophysiology of multifactorial disease. However, when a genetics finding is further supported with significant alterations in mRNA expression and in protein concentration, the possibility that the gene is an important factor in the pathogenesis of the disease increases. Similarly, individual changes in metabolites, lipids or glycans may have limited translational potential, but when combined, they may reveal important pathways associated with the etiology of a disease.

Genetic variants can introduce or delete methylation sites in CpG context, thereby induce changes in DNA methylation at the SNP site. Moreover, SNPs located in *cis* or *trans* of the CpG site can alter action of methylation enzymes. To further test for potential interactions between epigenetic profiles and genetic factors several studies have performed correlation analyses between DNA methylation and SNP genotypes. These methylation quantitative trait loci (meQTL) analyses are valuable tools to understand whether DNA methylation changes represent a mediator between genetic predisposition and clinical variables, a consequence of the interaction between genotype and phenotype or the effects genotype and methylation are independent from the genotype-phenotype associations. Blood circulating proteins as intermediate phenotypes can reveal the underlying disease-causing pathways. While changes in CpG methylation have been shown to induce changes in gene expression that can lead to changes in protein and metabolite levels, differences in transcript, protein, and metabolite levels, can in turn induce changes in CpG methylation. Therefore, mining the complex molecular networks from a multi-omics perspective, connecting genetic variants, CpG methylation, gene expression, protein, and metabolite levels to disease endpoints is necessary. In fact, the number of GWAS with deep molecular phenotypes, especially metabolomics and proteomics, is rapidly increasing. In the Website http://www.metabolomix.com it can be found tables of published GWAS with metabolomics (meQTLs), proteomics (pQTLs), and DNA methylation (mQTLs), as

well as EWAS with metabolomics (meCpGs) and with transcriptomics (eQTMs). Such networks might eventually guide a more personalized nutrition of complex disorders. In addition, machine learning can be used to analyze large data sets, as the ones that derive from multi-omics measurements, and can lead to algorithms with predictive value.

References

Arpón, A., Milagro, F.I., Razquin, C., Corella, D., et al., 2017. Impact of consuming extra-virgin olive oil or nuts within a mediterranean diet on DNA methylation in peripheral white blood cells within the PREDIMED-Navarra randomized controlled trial: a role for dietary lipids. Nutrients 10 (1), 15. https://doi.org/10.3390/nu10010015.

Asif, S., Morrow, N.M., Mulvihill, E.E., et al., 2020. Understanding dietary intervention-mediated epigenetic modifications in metabolic diseases. Front. Genet. 11, 590369. https://doi.org/10.3389/fgene.2020.590369.

Cai, S., Quan, S., Yang, G., et al., 2021. Nutritional status impacts epigenetic regulation in early embryo development: a scoping review. Adv. Nutr. 19, nmab038. https://doi.org/10.1093/advances/nmab038.

Cione, E., La Torre, C., Cannataro, R., et al., 2019. Quercetin, epigallocatechin gallate, curcumin, and resveratrol: from dietary sources to human microRNA modulation. Molecules 25 (1), 63. https://doi.org/10.3390/molecules25010063.

Dai, Z., Ramesh, V., Locasale, J.W., 2020. The evolving metabolic landscape of chromatin biology and epigenetics. Nat. Rev. Genet. 21 (12), 737–753. https://doi.org/10.1038/s41576-020-0270-8.

ElGendy, K., Malcomson, F.C., Lara, J.G., et al., 2018. Effects of dietary interventions on DNA metehylation in adult humans: systematic review and meta-analysis. Br. J. Nutr. 120 (9), 961–976. https://doi.org/10.1017/S000711451800243X.

Fabiani, R., Vella, N., Rosignoli, P., 2021. Epigenetic modifications induced by olive oil and its phenolic compounds: a systematic review. Molecules 26 (2), 273. https://doi.org/10.3390/molecules26020273.

González-Becerra, K., Ramos-Lopez, O., Barrón-Cabrera, E., et al., 2019. Fatty acids, epigenetic mechanisms and chronic diseases: a systematic review. Lipids Health Dis. 18 (1), 178. https://doi.org/10.1186/s12944-019-1120-6.

Hassan, F.U., Rehman, M.S., Khan, M.S., et al., 2019. Curcumin as an alternative epigenetic modulator: mechanism of action and potential effects. Front. Genet. 10, 514. https://doi.org/10.3389/fgene.2019.00514.

Kumari, A., Bhawal, S., Kapila, S., et al., 2020. Health-promoting role of dietary bioactive compounds through epigenetic modulations: a novel prophylactic and therapeutic approach. Crit. Rev. Food Sci. Nutr. 1–21. https://doi.org/10.1080/10408398.2020.1825286.

Mahmoud, A.M., Ali, M.M., 2019. Methyl donor micronutrients that modify DNA methylation and cancer outcome. Nutrients 11 (3), 608. https://doi.org/10.3390/nu11030608.

Mantilla-Escalante, D.C., López de Las Hazas, M.C., Crespo, M.C., et al., 2021. Mediterranean diet enriched in extra-virgin olive oil or nuts modulates circulating exosomal non-coding RNAs. Eur. J. Nutr. https://doi.org/10.1007/s00394-021-02594-0.

Ong, L.T.C., Booth, D.R., Parnell, G.P., 2020. Vitamin D and its effects on DNA methylation in development, aging, and disease. Mol. Nutr. Food Res. e2000437. https://doi.org/10.1002/mnfr.202000437.

Ye, Q., Zeng, X., Wang, S., Zeng, X., et al., 2021. Butyrate drives the acetylation of histone H3K9 to activate steroidogenesis through PPARγ and PGC1α pathways in ovarian granulosa cells. Faseb. J. 35 (2), e21316. https://doi.org/10.1096/fj.202000444R.

Nutrigenetics

Ángel Gil[a,b,c], **Concepción M. Aguilera**[a,b,c], **and Augusto Anguita-Ruiz**[a,b,c,d], [a] Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology "José Mataix," Center of Biomedical Research, University of Granada, Granada, Spain; [b] Instituto de Investigación Biosanitaria ibs.Granada, Granada, Spain; [c] CIBEROBN (Physiopathology of Obesity and Nutrition Network CB12/03/30038), Institute of Health Carlos III (ISCIII), Madrid, Spain; and [d] Barcelona Institute for Global Health, ISGlobal, Barcelona, Spain

© 2023 Elsevier Ltd. All rights reserved.

Introduction 188
Origin of the term and development of nutrigenetics 189
Basic concepts of genetics applied to nutrigenetics 191
Genetic variants and susceptibility to disease 192
Influence of gene variants on the metabolic requirements and utilization of nutrients and foods 193
References 196

Key points

- To know the concept of nutrigenetics, its conceptual and methodological origin as well as its evolution over time.
- To know the main concepts behind gene-diet interactions and how they are assessed in nutrient-centered nutrigenetics.
- Estimate and interpret a gene-diet interaction in nutrigenetics both at the level of a single genetic variant and at the level of the combination of multiple genetic variants through the so-called "genetic risk scores" (GRS).
- To describe and interpret gene-diet interactions related to the main non-communicable chronic diseases.
- To assess the level of scientific evidence of the different types of epidemiological studies that analyze gene-food interactions at the individual or dietary level.

Glossary

Copy number variation A phenomenon in which sections of the genome are repeated and the number of repeats in the genome varies between individuals

Nutritional genomics The joint study of nutrition and the genome including all other omics derived from genomics (transcriptomics, proteomics, metabolomics, epigenomics, etc.)

Human genome The complete set of nucleic acid sequences for humans, encoded as DNA and grouped by genes within the 23 chromosome pairs in cell nuclei and the small DNA molecule found within individual mitochondria

Nutrigenetics Study of gene variants within individuals and their impact on the metabolic utilization of nutrients

Single nucleotide polymorphism (SNP) A germline substitution of a single nucleotide at a specific position in the genome

Introduction

Over the last 50 years, nutrient recommendations have been established for certain segments of the population according to sex, age and physiological condition. However, even when population-level nutrient intake recommendations exist, individuals respond differently to lifestyle and dietary interventions because their gene variants influence the absorption and metabolic utilization of nutrients.

One of the most important contributions of the discovery of the human genome has been the realization that there are millions of differences in the gene sequences of different individuals (International HapMap Consortium, 2007). The phenotypic differences that distinguish individuals in the human species are largely due to differences in the sequence of their genes, mainly centered on the existence of single-nucleotide polymorphisms (SNPs), copy number variants (CNVs) of some genes, and other variants such as inversions, insertions and deletions. About 88 million genetic variants (84.7 million SNPs, 3.6 million short insertion-deletions and 60,000 structural variants) are currently known (The 1000 Genomes Project Consortium, 2015). **Fig. 1** shows a summary of the main types of genetic variants. In addition, the methylation patterns of various genes that specify functional proteins and their promoter genes are an important source of individual diversity. This influences the expression of numerous genes and thus the types and concentrations of the encoded proteins, resulting in specific metabolic changes that ultimately lead to specific phenotypes.

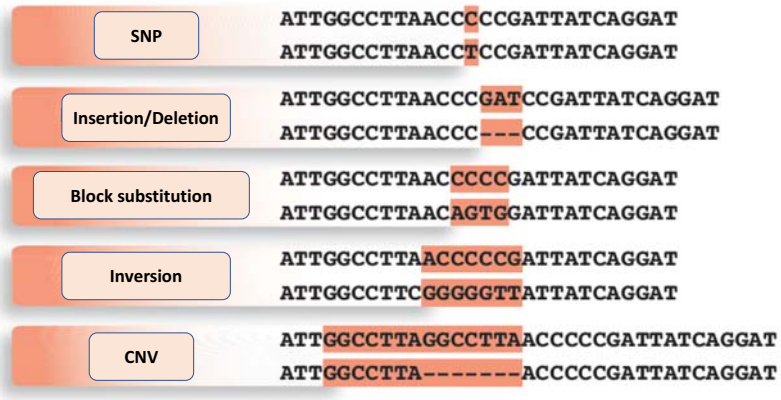


Fig. 1 Classes of human genetic variants. CNV: copy number variants; SNP: single nucleotide polymorphism.

Based on the data provided by the human genome project and knowledge of genetic polymorphisms or SNPs, as well as CNVs and other genetic variants, it has been found that there are certain variations in individuals that alter the interactions between nutrients, dietary components, foods and dietary patterns, and metabolic responses, affecting the susceptibility to the development of certain diseases. This inter-individual genetic variation questions the extent to which dietary recommendations, usually based on epidemiological studies, are valid for all racial and ethnic groups.

The study of gene variants within individuals and their impact on the metabolic utilization of nutrients is called "Nutrigenetics". Currently, there is broad consensus in considering Nutrigenetics as the discipline that studies the different phenotypic responses to diet depending on the genotype of each individual, both for the treatment of a health problem and the prevention of disease. Therefore, Nutrigenetics studies the variable response of individuals to diet as a function of SNPs and CNVs and other functional variants in the genome including the identification and characterization of these genetic variants. Ultimately, Nutrigenetics aims to generate specific recommendations on the best dietary composition for the optimal benefit of each individual, i.e., to achieve "personalized nutrition" (Kohlmeier et al., 2016). Indeed, it is thought that the human genome and its inter-individual variations in key candidate genes may be very important in helping to decipher the molecular mechanisms that determine the interindividual response to diet and thus generate a series of response biomarkers that allow us to know in advance of dietary intervention, the success of the same or its best implications in the prevention and treatment of disease (Ferguson et al., 2016).

Origin of the term and development of nutrigenetics

The term nutrigenetics as such was first used in 1975 by Dr. R.O. Brennan in his book entitled "Nutrigenetics: New Concepts for Relieving Hypoglycemia". Although this term had not yet been coined, its concept and meaning had already been explained by several authors many decades earlier. Among them, Garrod (1902–1923), in his work on inborn errors of metabolism, was the first to suggest that the effect of genes on metabolism depended on the diet. Later, Beadle, 1945 and Wagner and Mitchell, 1955 outlined the concept of gene-diet interaction through their studies on insect eye pigments and went on to show that the phenotypic correction of a genetically determined disorder could be modified by an appropriate environment. Similar to these observations, the work of Beadle & Tatum in the 1940s experimenting on *Neurospora* also contributed to the concept of gene-diet interaction by observing that *Neurospora* species with lethal genetic mutations could or could not manifest that lethal phenotype by regulating the concentration of arginine in the medium, thus concluding that the nutritional environment could strongly modulate the effects of a genetic mutation even associated with death. Compiling these and other studies by contemporary authors, J.A. Roper of the Department of Genetics at the University of Glasgow published a review in 1960 entitled "Genetic determinants of nutritional requirements", a review that was a pioneer of current Nutrigenetics and very advanced for the biotechnological limitations of the time. Interestingly, the conclusions of the review indicated that "nutritional-genetic" studies should be prioritized in future research and particularly in genetically complex diseases.

In parallel to these more basic researchers, Carl Rehnborg, best known for creating the first multivitamin in the 1930s, is also considered a pioneer of Nutrigenetics. While in China between 1917 and 1927, Carl observed that people living in agricultural areas in China were much healthier than people living in large cities. This led him to speculate on how nutrition can alter the expression of one's genetic makeup.

The beginning of the Human Genome Project in the 1980s and the discovery in the early years of important genetic variants in candidate genes for different diseases, allowed us to make great progress in the identification of inter-individual variability through the analysis of the so-called SNPs.

Preceded by several animal studies on the influence of different nutrients in the diet on gene expression, the first studies on gene-diet interactions in humans were published in the late 1980s-early 1990s. It should be noted that the first studies of nutritional

intervention with different lipids were carried out with short-term nutritional interventions, fundamentally fat loads and measuring the effects in postprandial ways after a few hours. Among these studies, the one carried out by JL Breslow et al. at the Laboratory of Biochemical Genetics and Metabolism at Rockefeller University, New York, published in 1987, stands out for its pioneering nature. This study, in which 27 individuals were analyzed, showed different postprandial responses in chylomicron concentrations after administration of a fat overload with added vitamin A depending on the genotype of the common polymorphism in the apolipoprotein E *(APOE)* gene. This apolipoprotein, produced mainly in the liver, is a major component of triacylglycerol-rich lipoproteins and high-density lipoproteins. It is involved in the binding and subsequent clearance of these particles through specific cellular receptors. Given its role in lipoprotein metabolism, it is considered to be very relevant in the pathophysiology of atherosclerosis. Indeed, the discovery of several isoforms of this gene, as well as the genetic basis of the common polymorphism that determines them, made this polymorphism the main object of study of gene-diet interactions in the first nutrigenetic studies focused on the influence of the different types of fatty acids in the diet on lipid metabolism. Those studies found that the *APOE* gene, depending on the DNA base change in a common polymorphism in exon 4, results in an amino acid change affecting three common alleles, resulting in six genotypes: e2/e2, e2/e3, e2/e4, e3/e3, e3/e4, e4/e4. In general, it is observed that carriers of the e2 allele have lower plasma concentrations of LDL-cholesterol (LDLc) than e3/e3; whereas the e4 allele is associated with higher concentrations of LDLc and subsequently with a greater cardiovascular risk. The association of the *APOE* genotype with LDLc concentrations was replicated very consistently in various populations and aroused scientific interest in investigating whether genetic determinism existed or whether this association could be modulated by diet.

In 1990, Pietinen's group from the Department of Medicine at the University of Helsinki, Finland, published a gene-diet interaction study in 110 participants, with a longer duration of the intervention (up to 12 weeks), in which they analyzed the effect of a high-fat diet versus a low-fat and cholesterol diet on plasma lipid concentrations as a function of *APOE* genotype. They observed that the effect of *APOE* genotype on lipid changes was not deterministic and was diet-dependent, with the e4 allele being associated with greater reductions for LDLc in those subjects with a low-fat diet and greater increases when changing diet.

Prompted by these results, several nutritional intervention studies analyzing the effect of genotype (*APOE* and other polymorphisms related to lipid metabolism: *APOA1*, etc.) were performed in different populations during the 1990s, but they were mainly studies performed on a small number of volunteers and with short-term interventions. Therefore, it was necessary to conduct larger studies including samples of thousands of individuals. A pioneering study in this regard was carried out by Fumeron et al. at INSERM-Faculté de Médicine Xavier Bichat in Paris, published in 1995. In this study, they analyzed 608 cases of myocardial infarction and 724 controls from the ECTIM study who were genotyped for the common polymorphism in intron 1 (named TaqIB after the restriction enzyme used to perform the genetic analysis based on restriction fragment length polymorphisms, RFLPs) in the cholesterol ester transfer protein encoded by the *CETP* gene. This polymorphism gave rise to two alleles, designated B1 and B2. It was known that people carrying the B2 allele had higher plasma concentrations of HDLc than B1B1, a finding frequently replicated in several studies. They also found that greater alcohol consumption was associated with a greater increase in HDLc concentrations in B2B2, as well as greater protection against myocardial infarction.

To consolidate a body of knowledge with solid evidence on gene-diet interactions in multiple populations, it was necessary to provide results in other cohorts. Dr. Ordovás' group at the JM-Human Nutrition Research Center on Aging in Boston pioneered the study of gene-diet interactions in the Framingham study, a cohort study focused on cardiovascular risk factors. In that study, despite observing a clear association of the CETP-TaqIB polymorphism with higher HDLc concentrations, the interaction with alcohol was not observed, and the effect of the B2 allele was to increase HDLc concentrations in both drinkers and non-drinkers. In contrast, in the Framingham study, Dr. Ordovás's group reported in 2001 an interaction between alcohol consumption and the *APOE* polymorphism in determining LDLc concentrations. These studies were replicated years later in the EPIC (European Prospective Investigation into Cancer and Nutrition Cohort)-Spain cohort that included 41,440 participants followed for up to 10 years (Ordovas and Corella, 2004).

Since the publication in 2001 of the first study showing a gene-diet interaction (*APOE**alcohol) in the Framingham cohort, the number of publications on gene-diet interactions in other relevant cohorts in nutritional epidemiology began to grow exponentially, thus contributing to the birth and consolidation of nutrigenetics as a new discipline (Corella and Ordovas, 2009).

The first nutrigenetic studies focused on polymorphisms in candidate genes because the technology for mass genome screening was not yet available. Moreover, they all started using the RFLP-based genotyping technique which was tedious for large samples of thousands of participants so that rarely more than one genetic variant was examined. Also, these early studies focused on examining the interaction with dietary nutrients (primarily macronutrients and, within these, the different types of fatty acids) as nutritional epidemiology at the time was focused on the analysis of nutrients rather than foods. The focus on dietary fatty acids was also because the first nutrigenetic studies were conducted in the field of cardiovascular disease, and, at that early time, the prevailing paradigm in the etiology of cardiovascular disease and its relationship to diet was the deleterious influence of diets high in fat, primarily saturated fatty acids.

As the cost of massive genotyping chips became cheaper and several genome-wide association studies (GWAS) were carried out, considering not only phenotypes related to cardiovascular disease but also many other pathologies, new genetic variants were discovered for which it was necessary to know whether their genetic effect was deterministic or could be modified by diet (Atanasovska et al., 2015). More diverse gene-diet interaction studies then began to be carried out, incorporating at the same time the determination of multiple genetic variants either employing chips or by allele-specific probes. To combine genetic variants, gene-diet interaction studies began to be performed using the concept of "genetic risk score" (GRS) (see below) instead of individual genetic variants to better capture the multigene influence on complex diseases.

Currently, the design of nutrigenetic studies is also being improved, integrating other omics to better understand the mechanisms and performing meta-analyses of several studies that analyze the same phenotypes, genetic variants and dietary components to determine the homogeneity or heterogeneity of the interactions found. All this contributes to increasing the level of evidence of gene-diet interactions so that they can be transferred to the recommendations of more personalized nutrition and allow better prevention or treatment of the disease. A current challenge is to perform gene-diet studies using the information obtained by direct sequencing that gives us the sequence of the whole genome of the participants, but it is still computationally complex and huge sample sizes are required.

Basic concepts of genetics applied to nutrigenetics

Some basic concepts are essential for understanding nutrigenetic studies. When we refer to the study of "candidate genes", we are referring to genes that are known or hypothesized to be related to the phenotype of interest because they are directly involved in the metabolic pathways associated with it due to the functionality of the protein they encode. Whereas when we refer to new genes identified through GWAS, we focus on genes that have been discovered to be associated with the disease or phenotype of interest through statistical association analyses that have obtained very significant P-values, but whose function or the previous relationship with the phenotype of interest was unknown until then (Atanasovska et al., 2015).

As mentioned earlier, there are many types of genetic variants in DNA depending on their size, ranging from a base change to alterations in pieces of entire chromosomes. The most studied genetic variants are SNPs. In nutrigenetic studies, it is necessary to be familiar with the different models of inheritance that can be used to study how these polymorphisms influence phenotypes: dominant, recessive, codominant, and additive, since analyzing the data with one model or another has important implications. Fig. 2 shows how the genotypes are grouped in each of these models. Also in this figure, the concepts of genotypic frequencies and allele frequencies are presented, as well as the Hardy-Weinberg equilibrium. This equilibrium must always be checked when starting studies of gene-diet interactions to avoid misclassification biases (differential or non-differential), either due to errors in the genotyping technique or other types of errors in obtaining the samples, manipulation, etc.

Although the first nutrigenetic studies focused on one genetic variant, millions of SNPs have now been described in the genome, making essential to study several of them at the same time in the case of complex diseases. Some of these variants are inherited in clusters and constitute haplotypic blocks consisting of SNPs with high linkage disequilibrium. Other times, instead of analyzing several SNPs forming a haplotype, a single SNP chosen as a "tag" SNP, or indicator SNP of the whole haplotype, is analyzed, thus reducing the costs of determination. There are different strategies for the selection of these tag SNPs. The reduction in the cost of genotyping has led to an increasing number of SNP combinations being studied in the so-called "GRSs".

The GRSs aggregates the effects of many genetic variants into a single number which predicts genetic predisposition for a phenotype. Normally the genetic variants most associated with a particular phenotype are chosen from the main GWAS that can be consulted through specific publications or also from the databases of GWAS (GWAS catalog of the National Human Genome Research Institute, NHGRI, in collaboration with the European Molecular Biology Laboratory (EMBL)). There exist several requirements for a genetic variant to be included in a GRS. These requirements will vary according to the study but basically, they are: to be strongly

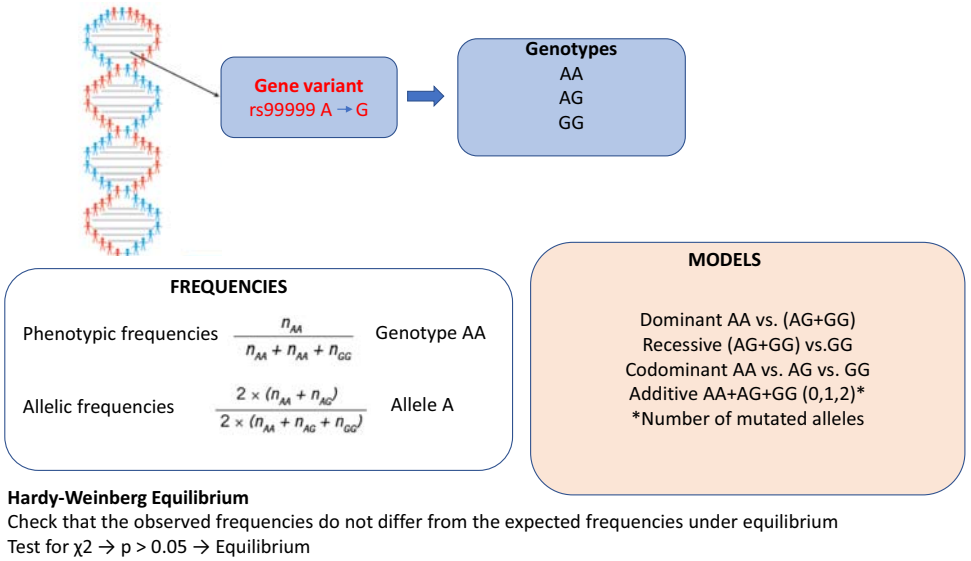


Fig. 2 Basic concepts of genetics for the interpretation of gene-nutrient interactions: genetic models, allele frequencies, genotypic frequencies and Hardy-Weinberg equilibrium.

associated with the phenotype of interest, to be in Hardy-Weinberg equilibrium, to have an allele frequency of the least frequent allele (MAF) exceeding a pre-established value, and not to have high linkage disequilibrium between them so that they are independent (Fig. 3). GRSs are typically composed of hundreds to millions of genetic variants which can be combined in several ways: unweighted and weighted. In weighted GRSs, we use a weighted sum of allele dosages multiplied by their corresponding effect sizes, as estimated from a relevant GWAS. In unweighted GRSs only the sum of the number of risk alleles contributed by each SNP included in the GRS is taken into account. However, the use of unweighted GRSs has the disadvantage that it does not take into account the magnitude of the association with the phenotype of interest for each of the SNPs included in the GRS, since the presence of each risk allele in the GRS scores the same regardless of whether it is an SNP highly associated with the risk phenotype or whether the association is 5 or 10 times smaller. Currently, there has become available a great resource from the EMBL, called PGS Catalog, that gathers all published GRSs consistently annotated with relevant metadata; including scoring files (variants, effect alleles/weights), annotations of how the PGS was developed and applied, and evaluations of their predictive performance. These GRS, both weighted and unweighted, are used as variables measuring genetic susceptibility in subsequent statistical analyses of gene-diet interactions.

Genetic variants and susceptibility to disease

Thanks to genomic research, it is now known that the vast majority of human genetic information is shared by all individuals of our species. If any two subjects are considered, they share 99.9% of the DNA sequence, i.e., less than one variation every 100 base pairs. Many of these variations are SNPs and these, individually or as groups (haplotypes) alter the regulation of the expression of the affected genes, the processing of mRNA, and the activities of synthesized proteins and enzymes. Thus, each individual can trigger unique responses to environmental factors, including dietary components, based on a combination of SNPs in their genomic DNA.

Similarly, all racial and ethnic groups share most genetic variations. The small differences that do exist are responsible for human diversity, such as skin and hair colors, potential height and weight, etc. Some of these small differences are important variations from a medical point of view because they modify the susceptibility to the development of some pathologies.

Certain minority populations have a higher incidence of certain chronic diseases than majority populations, such as obesity, diabetes, asthma, cardiovascular disease (CVD), and certain types of cancer. For example, North Americans of African descent have a 60% higher risk of developing prostate cancer than those of European descent. Likewise, epidemiological studies such as NHANES III (The third National Health and Nutrition Examination Survey) have shown that African-American and Mexican-American women over the age of 50 and African-American men have a higher risk of CVD.

Currently, GWAS and GRS studies make it possible to analyze millions of polymorphisms in hundreds of thousands of subjects and to analyze risk haplotypes for certain diseases and their interaction with nutrients, foods and food patterns (Atanasovska et al., 2015). However, although much progress has been made in the last two decades, we are still far from knowing and understanding all the relevant interrelationships between genes, their variants and nutrients.

Among the potential benefits of understanding the genetic causes of complex diseases may include the following

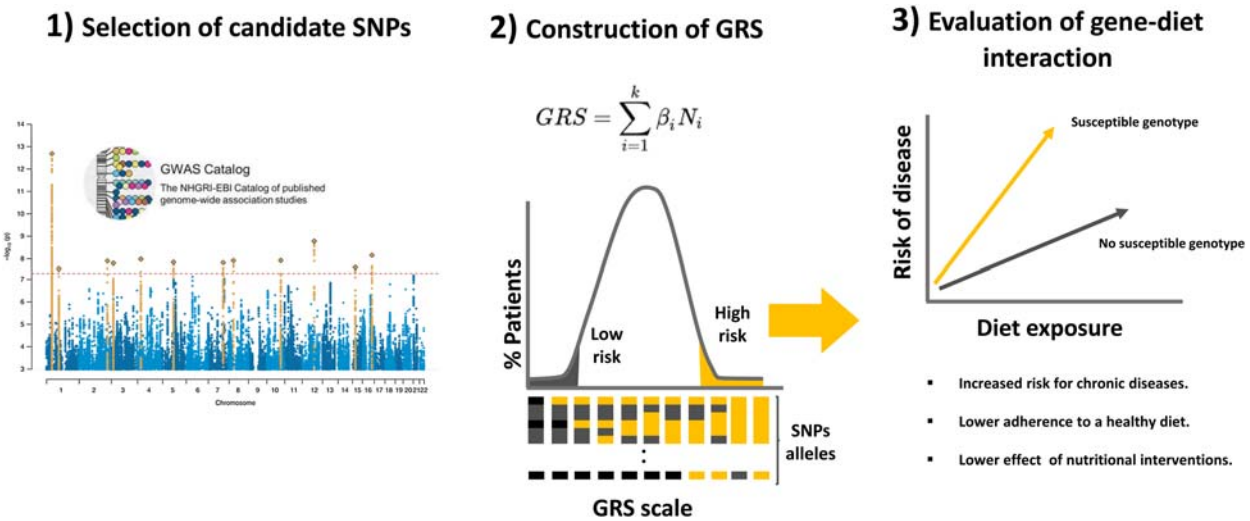


Fig. 3 The genetic risk scores (GRS) aggregates the effects of many genetic variants into a single number which predicts genetic predisposition for a phenotype.

1. The identification of individuals with a higher disease susceptibility will allow them to change their lifestyles to reduce the risk and will allow for increased medical surveillance to detect early signs of disease onset. Early diagnosis is almost always a determining factor in the success of treatment.
2. The identification of underlying genetic risk factors eliminates a variable in the study of genotype-environment interaction. Thus, the identification of environmental risk factors will be easier and more accurate.
3. Many diseases with similar clinical symptoms may be heterogeneous in their etiology. This ambiguity may affect the prognosis and result in a different response of patients to treatment.
4. If the different defects leading to the same disease can be developed, better treatments can be adopted.
5. The identification of the genes involved in the risk will allow the characterization of the proteins they encode and whose cellular functions are altered in the disease. This will facilitate a better understanding of the details of the normal and pathological state, respectively, and to develop of more rational therapies. In particular, the affected proteins can be used as a reference for designing and testing drugs that can modify their action.

Influence of gene variants on the metabolic requirements and utilization of nutrients and foods

One of the classic examples of the relationship between genetic variants, disease and nutrients is phenylketonuria (PKU), caused by mutations in the gene coding for phenylalanine hydroxylase. Individuals with PKU have to eat a diet low in phenylalanine. Lactose intolerance is another example of genome-diet interactions. In this pathology, the consumption of milk and some dairy products such as evaporated milk, powdered milk and products made with these ingredients, leads to the appearance of digestive pain accompanied by nausea, gas and diarrhea. Epidemiological data indicate that the frequency of lactose intolerance varies widely depending on geographic area, age, race and ethnicity.

In humans, the ability to digest milk lactose is conferred by a β-galactosidase enzyme called lactase-phlorizin hydrolase (LPH). While in some humans, approximately two-thirds of humankind, the levels of this enzyme decline drastically after the weaning phase (a trait known as lactase non-persistence (LNP)), some other individuals are capable of maintaining high levels of LPH lifelong (lactase persistence (LP)), thus being able to digest milk during adulthood (Anguita-Ruiz et al., 2020). The appearance of genetic polymorphisms associated with LP seems to have occurred in a relatively recent period, 10,000–12,000 years ago, coinciding with the domestication of animals. There are currently 11 known polymorphisms in the lactase gene grouped into four haplotypes called A, B, C, and U. Haplotype A, which confers lactose tolerance, has a frequency of 86% in the Northern European population, but only 36% in Southern European populations. Cultures that drink fresh milk usually have a higher frequency of this haplotype and the persistence of the polymorphism confers selective advantages such as better nutrition, prevention of dehydration, and improved calcium nutritional status. In particular, an SNP polymorphism (C/T13910) has been identified at a locus 14 kb upstream of the gene coding for lactase, at a locus on the long arm of chromosome 2 (2q21). This variant, initially identified in nine widespread Finnish families, is the most widely spread responsible for lactose tolerance among Europeans. In other words, members of these families can consume milk and dairy products without any complications. The polymorphism located in the lactase promoter gene alters the interactions of a regulatory protein that interacts with DNA. Interestingly, different polymorphisms affecting the same promoter region have been identified throughout the years in other populations (e.g., Middle East and Africa), although in a lower frequency, being responsible for milk tolerance (Anguita-Ruiz et al., 2020). The discovery of these variants makes it possible to individualize dietary interventions, specifically in infancy, by applying a genetic test for lactose intolerance.

Variants in genes required for lipid metabolism are valuable examples of Nutrigenetics. Indeed, gene variants encoding for *CETP*, lipoprotein lipase, low-density lipoprotein receptor and apolipoproteins B and E, may contribute to an increased risk of coronary heart disease. There is now sufficient evidence to indicate that variations in the *APOA1*, *APOA4*, *APOE*, and *APOB* genes contribute to the heterogeneity of lipid response after dietary intervention and that these genes are regulated directly or indirectly by PPARs and other nuclear receptors. On the other hand, in recent years the influence of different SNPs related to postprandial lipid metabolism has been studied. An example of this is the SR-BI gene, which belongs to the scavenger receptor family, because of its property to bind modified LDL particles and because it plays an important role as a receptor for high-density lipoprotein particles and in mediating the selective uptake of cholesterol esters. The presence of the minority allele 2 in exon 1 of the SR-BI gene has been associated with faster clearance of small triglyceride-rich lipoproteins (TRL), probably related to faster hepatic uptake. Also, the c.1119C > T polymorphism present in exon 8 of the gene has been shown to decrease the postprandial triacylglycerols response in LRTs in healthy males.

Although environmental factors such as diet, physical activity and alcohol play an important role in setting blood triacylglycerols concentrations, both triacylglycerols and cholesterol levels associated with LDL particles are strongly influenced by genetic variability. In the case of genes influencing LDL subclass patterns, diet-gene interactions contribute greatly to explaining the interindividual differences in the effects of low-fat and high-carbohydrate diets on heart disease risk. Studies indicate that the response of plasma lipids to diet is highly complex and variable, and involves numerous SNPs involved in multiple metabolic pathways (Lange et al., 2015). Few observational studies have examined the interaction of SNPs present in potential candidate genes for the development of CVD and the consumption of diets with different fat compositions. However, in recent years some evidence has begun to emerge on the beneficial effects of some of them, such as the Mediterranean diet, concerning some genetic variations. This is the case of the Pro12Ala polymorphism in the PPAR gamma *(PPARγ)* gene locus. This polymorphism modifies the

peripheral insulin sensitivity response following high oleic acid intake. In addition, it has been shown that individuals carrying the Ala12 allele have a lower risk of developing diabetes mellitus. In line with these findings, it has been shown that monounsaturated fatty acid consumption is not associated with an increase in body mass index (BMI) or the risk of obesity in individuals with a specific variant at the *APOA5* gene locus, which encodes for apoprotein A V. Data from that study indicate that the interaction between the -1131T > C polymorphism of the *APOA5* gene and ingested fat influences BMI. This interaction is dose-dependent (Atanasovska, 2015; Ordovas and Corella, 2004).

Another interesting work was designed to study the interaction between the C677T polymorphism in the methylenetetrahydrofolate reductase *(MTHFR)* gene and the Mediterranean diet, concerning the response to LDL oxidation. This gene expresses the corresponding enzyme involved in the formation of methyl-tetrahydrofolate from the methylene form. The values of oxidized LDL (LDL-ox) were significantly higher in TT individuals compared to carriers of the C/T and C/C genotypes. An interesting finding of this study is that greater adherence to the Mediterranean diet correlated inversely with lower LDL-ox values. This effect was observed only in carriers of the T/T and C/T genotypes, but not in the C/C genotypes. The evidence derived from this study indicates a pathophysiological explanation by which the type of fat in the Mediterranean diet could modify coronary risk by lowering LDL-ox values.

Other studies have investigated genetic variants associated with obesity or resistance to weight loss in human populations, which has contributed to understanding some of the pathophysiological mechanisms of this disease. One of the most prominent examples is the fat mass and obesity-associated *(FTO)* gene; the minority (16%) of individuals carrying two copies of the rs9939609 SNP weigh around 3 kg more and have a 1.67 times higher risk of obesity than non-carriers. Variants in numerous other genes, such as those for peroxisome proliferator-activated receptors *(PPAR)*, uncoupling proteins for oxidative phosphorylation (*UCP-1* and *UCP-3*), leptin receptor *(LEPR)*, melanocortin receptor 4 *(MCR4)*, 11β-hydroxysteroid dehydrogenase and several genes coding for antioxidant defense system proteins are also associated with obesity and may affect weight gain or loss in genetically predisposed individuals. Despite the need for further validation studies, valuable findings have been reported describing the interaction of macronutrient intake in carriers of various SNPs in genes that regulate processes such as thermogenesis, appetite and lipid or carbohydrate metabolism. Such is the case for carriers of *FTO*, *APOA5* and *PPARG* variants that generally increase their risk of obesity when they consume a certain type of fat, carbohydrate or protein (Albuquerque et al., 2015; Ruperez et al., 2014).

One of the best examples of diet-genotype interactions is type 2 diabetes, which occurs frequently in sedentary and obese individuals and certain minority groups such as Pima Indians (Bhori et al., 2022). Once diagnosed with diabetes, some individuals can control symptoms by increasing physical activity and reducing caloric intake, especially fat. In these cases, the expression of genetic information is changed by the environment. However, other individuals are refractory to such environmental changes and require treatment with medications.

Numerous chronic diseases do not show the metabolic flexibility observed in type 2 diabetes, i.e., symptoms do not reverse after the onset of the disease. Nevertheless, genotype and diet interactions contribute to the incidence and severity of many diseases such as obesity, cancer, atherosclerosis, asthma, etc. (Franzago et al., 2020).

Another example of the effects of gene polymorphisms on the metabolic utilization of nutrients is related to the perilipin (PLIN) gene, a key adipocyte protein involved in the action of hormone-sensitive lipase, and the discovery of new genetic variants associated with a lower risk of obesity in the Caucasian population. Thus, variants in this gene modulate the response to the hypocaloric diet in morbid obesity after 1 year of follow-up, so that carriers of the 11,482G > A allele have greater difficulties in losing weight (Ordovas, 2006).

Diet-genome interactions have also been studied with regard to inflammatory processes. Inflammation is an essential part of the body's response to infection, surgery and trauma, playing an important role in killing pathogens by creating a hostile tissue environment through the production of oxidative molecules and the activation of T and B lymphocytes. In addition, the body releases chemical mediators derived from the inflammatory process including three potent pro-inflammatory cytokines (interleukins 1 (IL-1) and 6 (IL-6)) and tumor necrosis factor. It has been observed that SNPs in genes responsible for producing molecules involved in inflammatory processes exert a modulatory effect on the intensity of inflammation. It has also been shown that both pro- and anti-inflammatory cytokines are influenced by differences in genotype.

Several oxidative molecules, especially oxygen free radicals, activate the production of the transcription factor nuclear factor kappa B (NFB), the most important mediator in the development of inflammation, and genetic variability influences the ability of some individuals to produce oxidative molecules that determine the activation of NFB which increases cytokine production and the expression of adhesion molecules, all of which increase the risk of host injury.

Natural Resistance Associated Macrophage Protein 1 (NRAMP1) has effects on TNF production and activation of nitric oxide-induced nitric oxide synthase (iNOS). There are four variants of the NRAMP1 gene, where alleles 1,2 and 4 are poor promoters and allele 3 causes elevated gene expression. Macrophage hyperactivity, associated with allele 3, is associated with susceptibility to autoimmune diseases and high resistance to infection, while allele 2 increases susceptibility to infection and protects against autoimmune diseases.

Some molecules suppress the production of proinflammatory cytokines and exert an anti-inflammatory influence; these include antioxidant defenses and interleukin-10 (IL-10). There are at least three polymorphic sites (−1082, −819 and −592) in the IL-10 promoter that influence its production and SNPs also occur in genes encoding enzymes involved in antioxidant defense, such as superoxide dismutase and glutathione peroxidase, which influence its activity.

Polymorphisms in cytokine genes may play a role in longevity by mediating individual responses to inflammatory stimuli. For example, possession of high IL-10-producing alleles decreases morbidity and mortality by having a protective effect against chronic inflammation.

One method to reduce inflammatory stress is the consumption of foods that suppress the production of proinflammatory cytokines (fish oil) or act as antioxidants (vitamin E). The mechanism by which fish oil decreases the inflammatory response is multiple since at the metabolic level the polyunsaturated fatty acids of the n-3 series limit the production of eicosanoids derived from arachidonic acid, which are pro-inflammatory, but at the genetic level they intervene in the suppression of the production of TNFα and in the activation of some transcription factors such as PPAR. Vitamin E is an antioxidant that also modifies cytokine production. This vitamin suppresses the production of superoxide and other oxygen-free radicals.

Deficiencies of the approximately 40 micronutrients (vitamins and minerals) necessary for human development are associated with numerous diseases. Many micronutrient deficiencies are due to poor dietary intake, but approximately 50 genetic diseases can be attributed to enzyme polymorphisms. The latter can be remedied or ameliorated by administering high levels of the relevant vitamin or coenzyme. Changes in coenzyme concentrations may contribute to increased binding of the coenzyme to the affinity apoenzyme altered by a given polymorphism. One of the now classic examples of the influence of gene variants on nutrient metabolism is the C677T polymorphism in the MTHFR gene. This gene expresses the corresponding enzyme involved in the formation of methyl-tetrahydrofolate from the methylene form. The genotypic frequency of the population is about 50% for CC, 40% for CT and 10% for TT. The activity of the enzyme expressed by individuals homozygous for CC is higher than that of CT heterozygotes and that of CT heterozygotes higher than that of TT homozygotes. The lower activity of the enzyme in TT subjects results in the accumulation of homocysteine in plasma, due to lower levels of methyl-tetrahydrofolate, and consequently, a higher cardiovascular risk. This effect reverts to normal if individuals with the TT genotype ingest amounts higher than those recommended for the general population (Fig. 4).

Other examples of genetic polymorphisms associated with genes encoding enzymes whose cofactors are derived from micronutrients are FAD oxidase, for cardiovascular disease and migraines, NAD(P) quinone oxidoreductase 1 (C609T) and FAD, for cancer, glucose-6-phosphate dehydrogenase (C131G) and NADP concerning Fabism and hemolytic anemia, and aldehyde dehydrogenase E487K, present in half of Asians, and NAD about alcohol intolerance, Alzheimer's disease and cancer.

The focus on diet as a factor in determining genomic stability is more important than previously imagined because of the impact that food has on all relevant pathways, such as exposure, carcinogen activation and deactivation, DNA synthesis and repair, and apoptosis. For example, dietary deficiency of micronutrients required for DNA maintenance can produce effects similar to genetically inherited diseases that impair the activity of enzymes required for genomic stability and can alter DNA like exposure to carcinogens and radiation. Adjusting dietary intakes of micronutrients for some individuals with similar genotypes and ages may minimize the damage to chromosomal and mitochondrial DNA observed in numerous alterations related to micronutrient deficiencies, optimizing health, prolonging quality of life, and preventing the risk of early onset of certain cancers and other age-associated degenerative diseases.

There are numerous investigations of statistically significant gene-food interactions. One example is that related to the interaction between sugar-sweetened beverage consumption and genetic predisposition to obesity according to BMI (Qi et al., 2012). Two

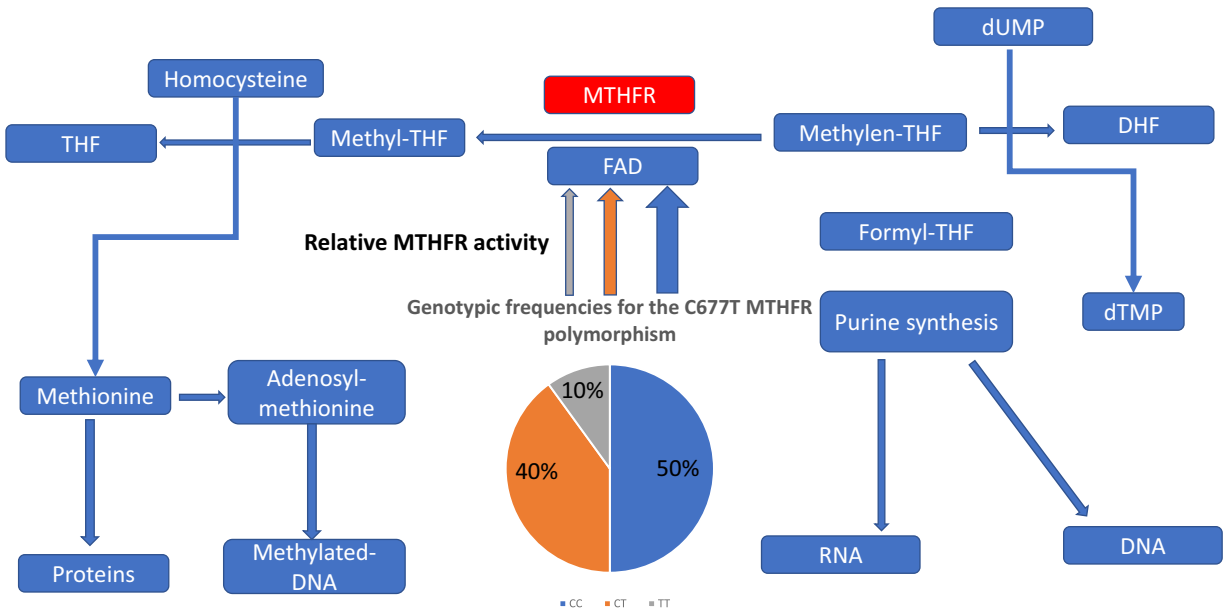


Fig. 4 Relative methylenetetrahydrofolate reductase (MTHFR) activity according to C677T polymorphism in the *MTHFR* gene. DHF: Dihydrofolate; dUMP: deoxyuridine monophosphate; dTMP: Deoxythymidine monophosphate; THF:Tetrahydrofolate.

cohorts of healthcare professionals in the United States, which included 6934 women participants in the US Nurses' Health Study (NHS, Nurses' Health Study) and 4423 men from the Health Professionals Follow-Up Study (HPFS) cohort were studied. In addition, the investigators included a replication cohort of 21,740 women participating in the Women's Genome Health Study (WGHS). For the study of genetic predisposition, they used a GRS that included 32 SNPs related to obesity. As for the diet, they focused on the consumption of sugary beverages. They found a statistically significant interaction between GRS and the consumption of these beverages by determining BMI prospectively. According to their results, the effect of consumption of sugar-sweetened beverages is associated with a higher risk of obesity in people with a greater genetic predisposition. These results have been replicated across the different cohorts analyzed, and constitute another important piece of evidence of gene-food interactions in phenotypes.

The genetic polymorphisms rs9939609 (C > A) in the FTO gene and rs17782313 (T > C) in the MC4R gene were determined in 7052 participants of the PREDIMED study. For both polymorphisms, the allele (A and C, respectively), has been associated with an increased risk of obesity. An unweighted GRS was constructed to capture in an additive manner the genetic risk of obesity associated with the combined presence of these two genetic variants so that in this GRS a score of 0 (equivalent to having no mutated allele), 1 (having one mutated allele, either FTO or MC4R), 2 (having two mutated alleles, one of each SNP or being homozygous for any of them), 3 (having three mutated alleles in any combination) or 4 (having four mutated alleles, which corresponds to the situation of being homozygous mutated for both SNPs). A higher GRS score was associated with a higher BMI.

Likewise, in more than 7000 participants of the PREDIMED study, important gene-diet modulations on the incidence of cardiovascular disease (total and separately for stroke and myocardial infarction) after a median follow-up of about 5 years, have been observed, Indeed, the polymorphism *TCF7L2-rs7903146* is significantly associated with a higher incidence of stroke in homozygotes for the TT genotype in the control group. However, TT individuals, who were intervened with a Mediterranean diet, have a lower incidence (Corella et al., 2013).

Furthermore, the PREDIMED study described the statistically significant interaction between the rs7903146-TCF7L2 (transcription factor 7 like 2) polymorphism and adherence to the Mediterranean diet by determining fasting plasma glucose concentrations at the start of the PREDIMED study (n > 7000). When adherence to the Mediterranean diet was low (score less than 9), the risk allele (T) was significantly associated with higher blood glucose, but with high adherence (equal to or greater than 9), the risk allele was not significantly associated with higher fasting blood glucose (Corella et al., 2013).

Despite the great advances achieved to date, nutrigenetics is an alive science branch in continuous evolution and progress. In the next years, thanks to the reduction in the cost of omics technologies and the availability of new and powerful analytics strategies such as GRSs and other machine learning approaches, we will experience a considerable boost in our ability to assess gene-diet interactions, not only restricting to the study of gene sequence variations but also focusing in epigenetics and other molecular alterations as responsible for the different responses of human to diet. Based on all these premises, it is fair to state that, through the elucidation of diet and gene interactions, nutrigenetics offer great opportunities to support patients with more personalized and effective dietary interventions helping in the management of chronic and serious metabolic diseases such as obesity, type 2 diabetes, CVD and several inflammatory diseases.

References

Anguita-Ruiz, A., Aguilera, C.M., Gil, Á., 2020. Genetics of lactose intolerance: an updated review and online interactive world maps of phenotype and genotype frequencies. Nutrients 12 (9), 2689.

Atanasovska, B., Kumar, V., Fu, J., Wijmenga, C., Hofker, M.H., 2015. GWAS as a driver of gene discovery in cardiometabolic diseases. Trends Endocrinol. Metab. 26, 722–732.

Albuquerque, D., Stice, E., Rodríguez-López, R., Manco, L., Nóbrega, C., 2015. Current review of genetics of human obesity: from molecular mechanisms to an evolutionary perspective. Mol. Genet. Genom. 290, 1191–1221.

Bhori, M., Rastogi, V., Tungare, K., Marar, T., January 2022. A review on interplay between obesity, lipoprotein profile and nutrigenetics with selected candidate marker genes of type 2 diabetes mellitus. Mol. Biol. Rep. 49 (1), 687–703.

Corella, D., Ordovas, J.M., 2009. Nutrigenomics in cardiovascular medicine. Circ. Cardiovasc. Genet. 2, 637–651.

Corella, D., Carrasco, P., Sorlí, J.V., Estruch, R., Rico-Sanz, J., Martínez-González, M.A., et al., 2013. Mediterranean diet reduces the adverse effect of the TCF7L2-rs7903146 polymorphism on cardiovascular risk factors and stroke incidence: a randomized controlled trial in a high-cardiovascular-risk population. Diabetes Care 36, 3803–3811.

Ferguson, L.R., De Caterina, R., Görman, U., Allayee, H., Kohlmeier, M., Prasad, C., Choi, M.S., Curi, R., de Luis, D.A., Gil, Á., Kang, J.X., Martin, R.L., Milagro, F.I., Nicoletti, C.F., Nonino, C.B., Ordovas, J.M., Parslow, V.R., Portillo, M.P., Santos, J.L., Serhan, C.N., Simopoulos, A.P., Velázquez-Arellano, A., Zulet, M.A., Martinez, J.A., 2016. Guide and position of the International Society of Nutrigenetics and Nutrigenomics on personalised nutrition: Part 1—fields of precision nutrition. J. Nutrigenet. Nutrigenomics 9, 12–27.

Franzago, M., Santurbano, D., Vitacolonna, E., Stuppia, L., 2020. Genes and diet in the prevention of chronic diseases in future generations. Int. J. Mol. Sci. 21 (7), 2633.

International HapMap Consortium, 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 851–861.

Kohlmeier, M., De Caterina, R., Ferguson, L.R., Görman, U., Allayee, H., Prasad, C., Kang, J.X., Nicoletti, C.F., Martinez, J.A., 2016. Guide and position of the International Society of Nutrigenetics and Nutrigenomics on personalized nutrition: Part 2—ethics, challenges and endeavors of precision nutrition. J. Nutrigenet. Nutrigenomics 9, 28–46.

Lange, L.A., Willer, C.J., Rich, S.S., 2015. Recent developments in genome and exome-wide analyses of plasma lipids. Curr. Opin. Lipidol. 26, 96–102.

Ordovas, J.M., 2006. Nutrigenetics, plasma lipids, and cardiovascular risk. J. Am. Diet Assoc. 106, 1074–1081.

Ordovás, J.M., Corella, D., 2004. Nutritional genomics. Annu. Rev. Genom. Hum. Genet. 5, 71–118.

Qi, L., Chu, A.Y., Kang, J.H., Jensen, M.K., Curhan, G.C., Pasqua le, L.R., et al., 2012. Sugar-sweetened beverages and genetic risk of obesity. N. Engl. J. Med. 367, 1387–1396.

Rupérez, A.I., Gil, A., Aguilera, C.M., February 20, 2014. Genetics of oxidative stress in obesity. Int. J. Mol. Sci. 15 (2), 3118–3144.

The 1000 Genomes Project Consortium, 2015. A global reference for human genetic variation. Nature 526, 68–74.

Nutrigenomics

Ángel Gil[a,b,c], **Concepción M. Aguilera**[a,b,c], **and Augusto Anguita-Ruiz**[a,b,c], [a] Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology "José Mataix", Center of Biomedical Research, University of Granada, Granada, Spain; [b] Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain; and [c] CIBEROBN (Physiopathology of Obesity and Nutrition Network CB12/03/30038), Institute of Health Carlos III (ISCIII), Madrid, Spain

© 2023 Elsevier Ltd. All rights reserved.

Introduction 197
Modulation of gene expression mediated by carbohydrates 198
Expression of genes related to the digestion and absorption of carbohydrates in the small intestine 198
Regulation of insulin synthesis in pancreatic β-cells by glucose 198
Control of the expression of metabolic genes by glucose 198
Direct regulation of the transcription factor ChREBP activity by glucose and its metabolites 199
Interaction of ChREBP with other transcription factors 201
Regulation of lipogenesis by carbohydrates in the liver 201
Modulation of gene expression mediated by lipids 202
Modulation of gene expression mediated by amino acids 204
Regulation of gene expression by vitamins A and D 205
Regulation of gene expression by trace metals 207
Conclusions 207
References 208

Key points

- To understand the role of carbohydrates in the regulation of genes involved in their digestion and absorption.
- To know the role of glucose as a regulator of the insulin gene function and expression.
- To understand the action of glucose on glycolytic and lipogenic gene expression.
- To identify ChREBP as the main transcription factor responsible for the effects of glucose on gene expression.
- To understand the function and mode of action of the transcription factors PPARs, SREBP, LXRs, HNF-4α, NF-kB, TLRs, and Nrf2 as regulators of gene expression in lipid metabolism.
- To identify the main genes regulated by amino acids and to unveil the mechanisms by which amino acids regulate gene expression.
- To understand the genomic actions of vitamins A and D.
- To know trace metals are involved in the regulation of gene expression.

Introduction

During the second half of the 20th century, nutrition has developed rapidly based on basic sciences such as biochemistry and physiology and applied sciences such as epidemiology and public health. One of the most important aspects of research in human nutrition is to try to understand the functions of nutrients and non-nutrient components of food at the molecular level. This includes the actions of nutrients and bioactive compounds in foods and their metabolites on cellular transporters and receptors and the corresponding implications in cell signaling cascades and interactions with the genome. Likewise, another great challenge of nutrition is knowing how nutrients, bioactive compounds in food, and certain dietary patterns influence the human microbiome and how modifications in the intestinal microbial ecology influence the greater or lesser prevalence of chronic non-communicable diseases. The science of assessing the functions of nutrients and other food components using molecular techniques is called "Molecular Nutrition".

"Nutritional genomics" is the science that attempts to provide a molecular-level explanation of how nutrients and other food components interact with an individual's gene set and their impact on health status. The tools of the "omics" sciences (genomics, epigenomics, transcriptomics, proteomics and metabolomics) applied to nutrition enable the development of nutritional genomics (Ferguson et al., 2016; Kohlmeier et al., 2016). The term "nutrigenomics" is often used generically to refer to functional genomics. Still, actually, this term should only be used when referring to changes in gene expression due to certain nutrients or bioactive compounds in food.

Traditionally, it has been assumed that gene expression in eukaryotes was not directly influenced by nutrients but by the action of hormones, growth factors and cytokines. However, diet represents a powerful mechanism for modifying the cellular environment

of numerous tissues, organs and, consequently, the individual. Thus, during the last few years a large body of evidence has been found that environmental changes caused by nutrients and other food components in the cellular environment modify gene expression. This new science that studies the effects of nutrients and bioactive compounds in food on gene expression is strictly speaking called "nutrigenomics", although, as mentioned above, this concept is often used in a broader context to refer to all interactions between genes and nutrients and food, i.e., functional genomics. This fact opens up the possibility of modifying gene expression in both healthy and diseased individuals through dietary manipulation.

It is now known that all nutrients (amino acids, glucose and other monosaccharides, fatty acids, vitamins, and minerals), beyond their classical metabolic actions, are also capable of modulating the gene expression of numerous genes, altering the phenotype of the individual. A common approach in nutrigenomics is to determine in an organ or tissue all the mRNAs present as a function of treatment with a particular nutrient or food. In addition, it is possible to determine all the proteins that appear or disappear in a specific organ or tissue due to the effects of nutrients or food consumption. It is also possible to perform a differential study of the metabolites present. In other words, while nutrigenetics would use gene sequencing as a fundamental tool, nutrigenomics would use transcriptomics, proteomics, and metabolomics, as fundamental elements to evaluate the actions of nutrients on gene expression and their metabolic repercussions on the phenotype of the individual.

Modulation of gene expression mediated by carbohydrates

Expression of genes related to the digestion and absorption of carbohydrates in the small intestine

In the small intestine, the expression of genes associated with digestion and absorption of nutrients is restricted to enterocytes. These come from undifferentiated crypt cells which, during their migration outward from the intestinal villi, show a sudden increase in the expression of genes related to digestion and absorption such as those of the digestive enzymes sucrase-isomaltase (SI) and lactase-floricin hydrolase (LPH) and transporters such as the sodium-dependent glucose co-transporter (SGLT1), the glucose transporter GLUT2 and the fructose transporter GLUT5.

In general, a diet rich in carbohydrates will increase the expression of the disaccharidases and transporters involved in their metabolism, which in the long term will result in better utilization of these nutrients. This increase in the expression of disaccharidases and transporters will occur mainly at the level of their transcription, although, as indicated below, post-transcriptional regulation also occurs in some cases.

The enzymatic activity of disaccharidases, sucrase and lactase in the small intestine increases in response to dietary carbohydrates due to an increased synthesis of their respective mRNAs. The targets promoters of lactase and sucrase contain binding elements for transcription factors CE-LPH1 and SIF1 with a common consensus sequence, TTTTAT/C. To these elements binds Cdx-2, a transcription factor involved in the transcriptional regulation of intestinal epithelial genes and which is of great importance in a wide variety of phenomena from early cell differentiation to the maintenance of the intestinal epithelial lining. Cdx-2 is overexpressed by a carbohydrate-rich diet.

On the other hand, fructose absorption in the small intestine is carried out by GLUT5. In the presence of fructose, the regulation of this transporter is rapid and the observed increases in transport are paralleled by an increase in mRNA and protein amount. In contrast, other sugars, sugar analogs and metabolites do not alter the transcription rates of the gene. Regarding the regulation of GLUT2 expression in the small intestine, it has been described that both sucrose and fructose increase its expression and that after 30 days of a diet rich in fructose, GLUT2 is permanently localized in the apical membrane.

Regulation of insulin synthesis in pancreatic β-cells by glucose

Glucose homeostasis is maintained in the body by the action of the pancreas through the regulated release of glucagon and insulin, respectively. The pancreatic cell function coordinately regulates insulin gene expression, insulin biosynthesis, and insulin secretion.

Glucose controls all steps in gene expression, including transcription, pre-mRNA splicing and mRNA stability. Insulin gene expression is essentially restricted to pancreatic cells in adult mammals. The promoter has a highly conserved region 340 bp upstream of the start site on which both tissue-specific expression and metabolic regulation of the promoter occurs. A variety of transcription factors bind to this region, forming a transcriptional network that ensures precise regulation of the gene. Boxes A3, C1 and E1 constitute the *cis*-elements mainly involved in transcriptional activation. Glucose participates in insulin gene expression through the action of three transcription factors: PDX-1 (Pancreatic/duodenal homeobox-1), NeuroD1 (neurogenic differentiation 1) and MafA (mammalian homolog of avian MagfA/L-Maf), which act by activating insulin gene expression in a coordinated and synergistic manner in response to increased glucose levels in a process that involves recruitment of transcription factors to regulatory elements, histone modifications and initiation of transcription (Fu et al., 2013).

Control of the expression of metabolic genes by glucose

Dietary carbohydrates constitute one of the main nutrients leading to short- and long-term metabolism adaptations. These adaptations are mostly mediated by ChREBP (carbohydrate response element binding protein), a transcription factor belonging to the Mondo family of bHLH/LZ (basic helix-loop-helix leucine zipper proteins). Among the variety of dietary carbohydrates, glucose is not only a primary energy source for all organisms but has also been reported as the main regulator of the expression of genes

Table 1 Genes regulated by carbohydrate response element binding protein (ChREBP).

Ruta	*Gene (gene symbol)*	*Activation/repression by glucose*
Glycolysis	Liver pyruvate kinase (*Pklr*)	Activation
	Fructokinase (*Fk*)	Activation
	GLUT2 (*Glut2*)	Activation
	GLUT4 (*Glut4*)	Activation
Gluconeogenesis	Catalytic subunit Glucose 6 Phosphatase (*G6pc*)	Activation
	Phosphoenol pyruvate carboxykinase (*Pepck*)	Repression
Lipogenesis	Fatty acid synthase (*Fasn*)	Activation
	Acetyl CoA carboxylase 1 (*Acc1*)	Activation
	Stearoyl CoA desaturase 1 (*Scd1*)	Activation
	ELOVL *n*-6 fatty acid elongase (*Elovl6*)	Activation
	Malic enzyme (*Me*)	Activation
	Adiponutrin/patatin-like phospholipase domain-containing protein 3 (*Pnpla3*)	Activation
Transcription	Krüppel like factor-10 (*Klf10*)	Activation
	Differentially expressed in chondrocytes 1 (*Dec1*)	Activation
	Pancreatic and duodenal homeobox 1 (*Pdx 1*)	Repression
	Peroxisome proliferator activated receptor α (*Ppara*)	Repression
	Hypoxia inducible factor (*Hif1b*)	Repression
	Sirtuin 1 (*Sirt1*)	Repression
Hormones and receptors	Insulin 1 (*Ins1*)	Repression
	Insulin 2 (*Ins2*)	Repression
	Glucagon receptor (*Gcgr*)	Activation
	Regulator 16 of G protein signaling (*Rsg16*)	Activation
Redox signals	Thioredoxin interacting protein (*Txnip*)	Activation

involved in energy metabolism at the transcriptional level. Thus, glucose can activate the transcription of mainly glycolytic and lipogenic genes in which ChREBP participates as a central regulator. In addition, ChREBP plays a broader and no less crucial role in other processes ranging from glucolipotoxicity to apoptosis or proliferation of some cell types (Filhouland et al., 2013).

There is ample evidence demonstrating a direct and key role of glucose in regulating gene expression in adipose tissue and liver and in pancreatic cells. In these tissues, glucose stimulates the transcription of different genes coding for glycolytic and lipogenic enzymes: hepatic pyruvate kinase (*L-PK*), acetyl-CoA carboxylase (*ACC*), fatty acid synthase (*FAS*) and stearoyl-CoA desaturase (*SCD1*). **Table 1** summarizes the genes regulated by ChREBP (Eissing et al., 2013).

For glucose to exert its effect on the expression of these enzymes, it must first be metabolized. Thus, in the liver, glucose metabolism by glucokinase (GK), the first enzyme of glycolysis, is necessary for the initiation of glucose-mediated signal transduction. Analysis of the promoters of these genes whose transcription is increased in response to glucose identified a carbohydrate response element (ChoRE) in their promoters and the transcription factor ChREBP that is able to recognize it. ChoRE has a conserved consensus sequence characterized by a tandem E box element separated by five nucleotides (5′-CACGTGnnnnnCACGTG-3′). Spot 14 (S14) was among the first genes in which a ChoRE element was detected in its promoter. It is expressed in the mammary gland and is responsible for lipid synthesis during lactation in response to carbohydrates.

There are several characteristics that make ChREBP compatible with a key role as a glucose-regulated transcription factor:

1. Its expression is most abundant in the liver and white and brown adipose tissue, where lipogenesis is very active and in the small intestine, kidney and pancreatic islets, whereas it is low in skeletal muscle.
2. Its expression in the liver is induced in response to a diet rich in carbohydrates but not in response to fasting or to diets rich in polyunsaturated fatty acids.
3. Gene silencing prevents glucose-dependent induction of *LPK*, *ACC* and *FAS* genes in adipocytes.

Following the identification of ChREBP, a second protein was discovered that interacts with ChREBP to form a heterodimer, Mlx (Max-like protein X). Thus, two ChREBP-Mlx heterodimers bind to the two E-boxes of ChoRE to provide a glucose-regulated transcriptional complex. Mlx is essential as a partner of ChREBP in the regulation of lipogenic enzyme genes in the liver since the use of dominant-negative forms of Mlx in hepatocytes, which are unable to bind DNA but do maintain their interaction with ChREBP, abrogates or decreases glucose-dependent transcription. MondoA also forms heterodimers with Mlx. Glucose signaling dependent on Mondo-Mlx interaction is evolutionarily conserved among worms, Drosophila and vertebrates.

Direct regulation of the transcription factor ChREBP activity by glucose and its metabolites

The amino-terminal domain of the ChREBP protein has several overlapping structural motifs. Of these, the GRACE and LID motifs stand out. While GRACE has a transcription-promoting effect in the absence of glucose or any of its metabolites, LID interacts with GRACE and blocks its activity. This interaction of LID with the GRACE domain extends to other regions of the amino-terminal end,

MCR I to MCR IV, regions in which signals that control trafficking to the nucleus of the transcription factor are also located and are therefore masked.

It is known that in the presence of high concentrations of glucose, the repression maintained by LID on the amino-terminal end of ChREBP is removed so that the import domains to the nucleus are exposed and the GRACE domain can bind to the target sequence, the ChoRE element. However, the mechanism by which glucose can induce this conformational change has not been confirmed. It has been proposed that glucose must be metabolized for its action to occur, so it is considered that it is a metabolite of glucose rather than glucose itself that can act as a ligand binding to ChREBP.

It is now considered likely that glucose-6-phosphate (G6P) is the metabolite responsible for this effect. The evidence supporting this assertion is that in the liver, glucokinase, the enzyme responsible for G6P synthesis, is essential for regulating ChREBP activity by glucose. Second, it has been shown that ChREBP-dependent transcriptional activity follows sigmoidal kinetics in response to variations in G6P concentration, which postulates this metabolite as an allosteric regulator of ChREBP. Structural models of the Mondo family members, MondoA and ChREBP, postulate a possible binding pocket to the MCR V domain, which would unlock the LID domain and allow it to pivot to expose the other domains present at the amino-terminal end. Although confirmatory work is required, the role of G6P in glucose regulation of ChREBP activity seems to be widely accepted (Fig. 1).

The response to G6P binding not only allows GRACE to bind to the target sequence but also exposes domains that control intracellular trafficking and at the same time allows other proteins to bind to domains that modulate transcriptional activity, as is the case for the 14-3-3 and CBP/p300 proteins.

Post-translational modifications, i.e., phosphorylation, O-Glucose-N-acylation, and acetylation modulate ChREBP function and regulate the state of ChREBP in response to glucose. ChREBP is the central protein of a complex network that regulates metabolic adaptations due to carbohydrate intake. In addition to the intrinsic mechanisms of ChREBP regulation, such as its transactivation by glucose metabolites, its translocation to the nucleus and its covalent modulation, ChREBP must be considered to be regulated by the signaling network of which it is a part. An example of this signaling is evident if we consider, for example, that insulin increases

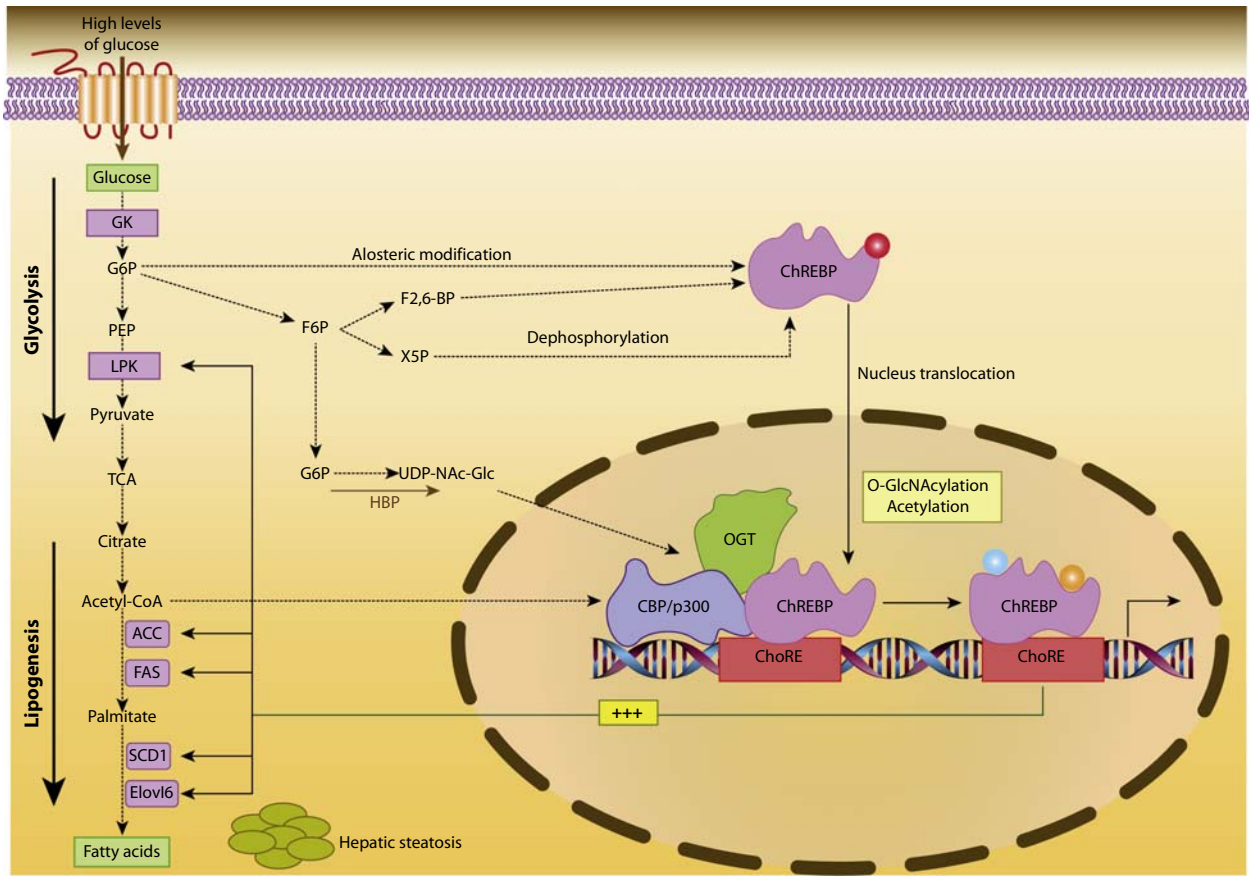


Fig. 1 Glucose regulation of glucose and lipid metabolism in the liver. Post-translational modifications by protein acetylation and O-GlcNAcylation activate ChREBP in response to glucose in liver cells. ACC, acetyl CoA carboxylase; CBP/p300, ChREBP coactivators; ChoRE, carbohydrate response element; ChREBP, carbohydrate response element-binding protein; Elovl6, fatty acid elongase; FAS, fatty acid synthase; F2,6-BP, fructose 2,6-bisphosphate; G6P, Glucose-6-phosphate; GK, Glucokinase; HAT, Histone acetyltransferase; HBP, Hepatic pyruvate kinase; LPK, Hepatic pyruvate kinase; OGT, O-GlcNAc transferase; PEP, Phosphoenol pyruvate; SCD1, Stearoyl-CoA desaturase; X-5P, Xylulose-5-phosphate.

the synthesis of ChREBPα, which in turn will bind to the specific promoter of ChREBPβ and that increased transcription of the latter is responsible for many of the effects of carbohydrates in adipose tissue (Benhamed et al., 2015).

Although ChREBP responds to circulating levels of glucose, it was established early on that glucose metabolization was necessary for regulation to occur. In most tissues, glucose enters by diffusion mediated by specific transporters called GLUT and once inside the cell, it must be metabolized in order to exert its transcriptional regulatory activity. Since in liver the first stage of glucose metabolism produces glucose-6-phosphate (G6P), this was initially proposed as the molecule responsible for glucose-mediated signaling effects, particularly in lipogenesis-related genes. Among the proposed metabolites, xylulose-5-P (X5P), an intermediate of the pentose phosphate pathway, could be considered the transducer molecule since it has been described that treatment with xylitol, the precursor of X5P, activates protein phosphatase 2A, which is involved in the regulation of the activity of several transcription factors. In addition, another glucose metabolite, fructose 2,6-bisphosphate (F-2,6-BP) has been implicated in the activation of ChREBP in response to glucose in liver cells. F-2,6-BP is synthesized from fructose-6-phosphate by a bifunctional enzyme (6-phosphofructo-2-kinase-fructose-2,6-bisphosphatase, PF2K/F2,6BPase) that has different catalytic sites for its kinase and phosphatase activities. The hexosamine biosynthesis pathway may also be involved. **Fig. 1** shows the regulation of some key genes involved in glycolysis and lipogenesis in the liver.

Interaction of ChREBP with other transcription factors

The main transcription factor that acts synergistically with ChREBP is SREBP-1c (sterol regulatory element-binding protein-1c). Although SREBP was first described as a transcription factor involved in controlling the expression of genes related to cholesterol biosynthesis, the SREBP-1c isoform was later found to participate in the insulin-mediated activation of lipogenic genes. The promoter of the latter enzyme contains two SRE (sterol regulatory elements) to which SREBP-1c can bind for activation. In turn, GK synthesizes G6P, which, as mentioned above, is one of the metabolites capable of activating ChREBP. On the other hand, the promoters of several genes coding for lipogenic enzymes present binding sites for both SREBP-1c and ChREBP, so that the highest levels of expression of these enzymes are reached when both transcription factors bind to their respective binding sites in the promoters. Thus, lipogenesis is synergistically regulated in response to carbohydrates and lipids.

Another link between carbohydrate and lipid regulation of gene expression is the USFs (upstream stimulatory factors) family of transcription factors. These factors belong to the bHLH-LZ family and bind to E-boxes of the promoters of target genes to activate their transcription. USFs bind to the proximal promoter region of the gene coding for FAS and other lipogenic genes and are crucial for the recruitment of SREBP-1c and its binding to SRE for synergistic activation of lipogenic genes during feeding or insulin treatment. Coordinated activity with ChREBP in lipogenesis has been demonstrated using USF knockout animal models showing decreased lipogenic gene transcription in response to a carbohydrate-rich diet.

Another element that adds its effects to those of ChREBP is HNF4α (hepatocyte nuclear factor 4). The promoter that regulates the expression of ChREBPα presents a binding site for this nuclear receptor, which therefore controls its expression in response to variation in glucose levels. In addition, HNF4α has the ability to bind to the FAS and L-PK promoters, where it acts synergistically with ChREBP. This synergy is enhanced by high glucose levels and is also essential for reaching maximal enzyme expression levels.

Regulation of lipogenesis by carbohydrates in the liver

The liver is the main organ responsible for *de novo* lipogenesis in response to high carbohydrate intake. The expression of enzymes involved in the synthesis of triacylglycerols in the liver is a process that is very well coordinated since a set of transcription factors activated by both hormonal and nutritional stimuli will participate in the induction/repression of genes involved in this process. Dysregulation of lipogenesis leads to the storage of triacylglycerols in the liver, steatosis, and other diseases, such as metabolic syndrome (Oosterveer and Schoonjans, 2014).

The effect of carbohydrates on gene expression in the liver occurs at several checkpoints ranging from membrane transporters to key enzymes of intermediary metabolism. Thus, one of the first checkpoints is constituted by the GLUT2 transporter. GLUT2 is the major glucose transporter in the liver. Its role is to introduce glucose during the post-absorptive state and release it into the blood during fasting. GLUT2 can be considered an early glucose sensor since part of the glucose generated from G6P in the endoplasmic reticulum by glucose-6-phosphatase will equilibrate with extracellular glucose through GLUT2. Alterations of this equilibrium lead to an increase of both glucose and G6P in the cytoplasm, which will increase the transcriptional activity of ChREBP in the nucleus and, therefore, to the induction of the gene coding for L-PK and lipogenic genes.

The second control point is constituted by the GK-ChREBP axis, which should be considered as the central glucose response system in the liver. The regulation of GK expression and activity is of vital importance in the sensitivity of the liver to glucose since it constitutes the first regulatory point of glycolysis and is the enzyme that catalyzes the conversion of glucose to G6P, a metabolite involved in the activation of ChREBP.

Insulin induces GK expression and although the molecular basis of this induction is not fully understood, several transcription factors have been described such as HFN-4, HIF-1 (hypoxia-inducible factor 1), SREBP-1c, LXR (liver X receptor), PPARγ (peroxisome proliferator activated receptor gamma), KLF-6 (Kruppel-like factor 6), TCFE3 (transcription factor E3) and LRH-1 (nuclear receptor liver receptor homolog 1) that control hepatic GK transcription.

The enzymes involved in the *de novo* lipogenesis are regulated rapidly by allosteric and phosphorylation/dephosphorylation mechanisms and in the long term at the transcriptional level in a well-coordinated manner involving different transcription factors in response to hormones (insulin) or nutrients (glucose). In the presence of insulin, which is necessary to activate glucose

metabolism in hepatocytes, high concentrations of glucose induce the expression of genes coding for glucose transporters, glycolytic and lipogenic enzymes such as L-PK, ACC and FAS and decrease that of gluconeogenic genes, such as the phosphoenolpyruvate carboxykinase gene. These effects in the liver are mediated in a coordinated manner by SREBP-1c and ChREBP as indicated above. Likewise, other transcription factors such as USFs (upstream-stimulatory factors) and LXRs have also been implicated in the coordinated regulation of lipogenesis.

In a situation of insulin resistance, there is a situation of hyperglycemia and high glucose concentrations can induce glycolytic and lipogenic genes through the activation of ChREBP. ChREBP expression is positively related to hepatosteatosis. Thus, ChREBP-dependent expression of lipogenic genes when insulin-mediated signaling is inhibited may in part explain the active lipogenesis that occurs during insulin resistance.

Modulation of gene expression mediated by lipids

Dietary lipids are macronutrients that are indispensable for mammalian growth and development. In addition to their energetic function and their role in membrane composition, fatty acids, both saturated and polyunsaturated, cholesterol and other sterols, as well as various eicosanoids or oxidized derivatives of the latter, modulate the expression of numerous genes involved in the oxidation of fatty acids themselves in the liver, muscle and adipose tissue, cholesterol and bile acid synthesis, adipocyte proliferation and differentiation, immune response, antioxidant defense and angiogenesis (Georgiadi and Kersten, 2012).

In mammalian cells, the signaling molecules are free fatty acids (FFA) and oxidized derivatives of polyunsaturated fatty, which are transported in and out of the cell with the help of a membrane protein, the fatty acid transporter (FAT). FFA is taken up by a specific fatty acid binding protein (FABP) and has various metabolic fates in the cytoplasm (Fig. 2).

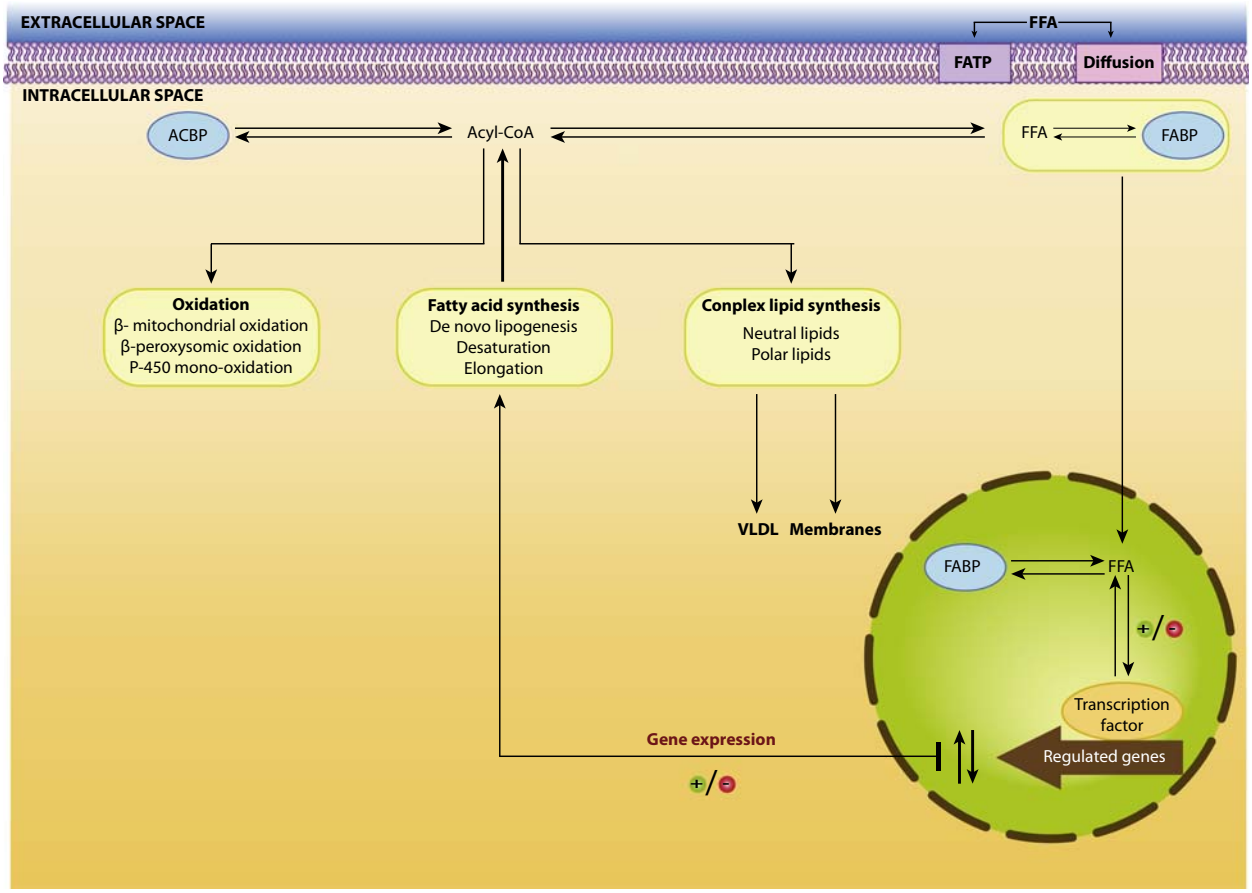


Fig. 2 Cellular transport of fatty acids and effects on the expression of genes related to lipid metabolism. ACS, acyl-CoA; ACBP, acyl-CoA-binding protein; FATP, fatty acid transport protein; FABP, fatty acid-binding protein; FATP, fatty acid transporter protein; FFA, free fatty acids; TE, thioesterase; VLDL, lipoproteins; VLDL, lipoproteins; FATP, fatty acid transport protein; VLDL, very low density lipoproteins; TE, thioesterase; FATP, fatty acid transport protein.

Free fatty acids bound to FABP, acyl-CoA or some fatty acid metabolites, such as eicosanoids, can:

- Induce a cascade of events leading to a covalent modification of a transcription factor, e.g., phosphorylation, altering its transactivation capacity.
- Directly bind and activate a transcription factor.
- Influence the transcription rate of a transcription factor and its synthesis.
- Modify the mRNA stability of both transcription factors and other target genes.

Specific transcription factors have been identified in mammals that respond to dietary lipids. These include at least seven major families of factors: PPAR (PPAR-α, PPAR-β, PPAR-δ and PPAR-γ), three variants of SREBP (SREBP-1a, SREBP-1c and SREBP-2), HNF-4, LXR-α and LXR-β, ChREBP, nuclear factor enhancer of activated B-cell kappa light chains (NF-κB) and Toll-like receptors (TLRs) (Contreras et al., 2013).

The effects of fatty acids are mediated either directly by their specific binding to various nuclear receptors (PPAR, LXR, HNF-4α), producing changes in the *trans* activation of these transcription factors, or indirectly by changes in the abundance of transcription regulatory factors (SREBP-1c, ChREBP). On the other hand, fatty acid-responsive transcription factors bind to a fatty acid recognition sequence or response element in the promoter of a target gene region, as monomers, as homodimers or as heterodimers with other transcription factors, e.g., retinoid receptors (RXR). **Fig. 3** shows the main effects on gene expression of polyunsaturated fatty acids (PUFA), indicating the transcription factors involved.

At the cellular level, the physiological response to fatty acids depends on the amount and chemical structure of the fat ingested, the specific fatty acid metabolism of particular cell types (oxidative pathways, kinetics and competitive reactions), the cellular abundance of membrane and nuclear receptors and the involvement of specific transcription factors. The mechanisms of fatty acid-mediated regulation of gene expression are involved in the control of carbohydrate and lipid metabolism, in cell growth and

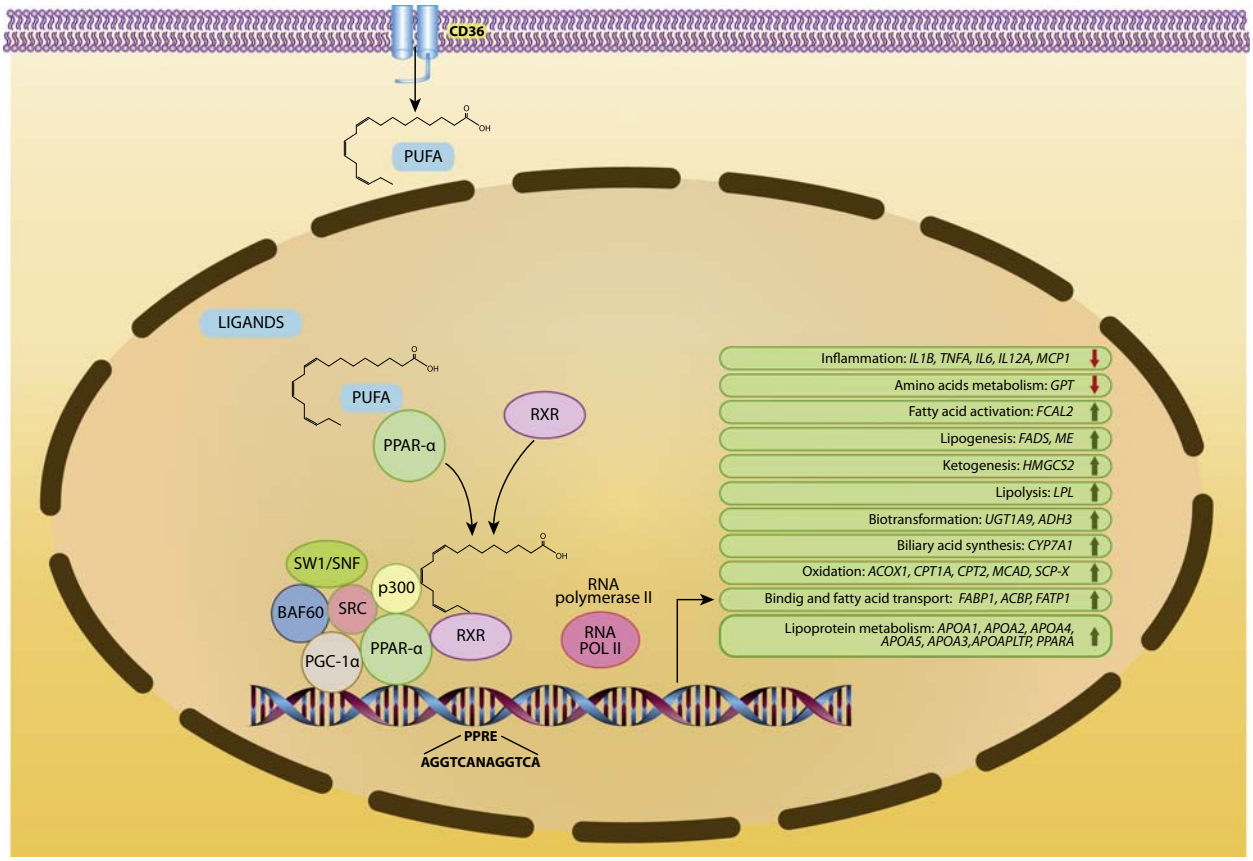


Fig. 3 Mechanism of action of polyunsaturated fatty acids (PUFA) on peroxisome proliferator-activated receptor-α (PPAR-α) and its effects on gene expression. PPAR-α can be activated by various ligands such as PUFA, fibrates and pyrinixic acid (WY14643) that modulate the transcription of certain genes. RXR is able to interact with PPARs, originating basal transcription with RNA polymerase II and transcribing certain genes such as CYP3A4, HMGCS2, SREBP1-c, APOA1, APOA2, APOA5, CPT-1, FGF21. APO: apolipoprotein; BAF60: Brahma 1 gene/Brahman-associated factor 60; CD36: differentiation cluster/cluster 36; CPT1: cRNAitin palmitoyltransferase 1; CYP: cytochrome P-450; FGF21: growth factor 21; HMGCAS2: hydroxymethylglutaryl CoA synthase; p300: histone p300 acetyltransferase; PGC1α: peroxisome proliferator-activated receptor (PPAR) co-activator 1α; PPRE: peroxisome proliferator response element; RXR: retinoid receptor; SRC: steroid receptor coactivator; SREBP1-c: sterol response element binding protein 1c.

differentiation, and in the production of cytokines, adhesion molecules, and eicosanoids that regulate numerous physiological and pathophysiological processes, including immune response and inflammation.

Modulation of gene expression mediated by amino acids

Amino acids (AAs) are the building blocks of proteins and are essential substrates for the synthesis of low molecular weight compounds such as glutathione and other active peptides, nitric oxide (NO) and polyamines, which are of great biological importance. But in addition, AAs in mammals are able to regulate gene expression by control of translation initiation and post-translational modifications, by regulation of miRNAs at the transcriptional level, by mRNA stabilization, as well as by modulation of chromatin through epigenetic processes. Although the control mechanisms of gene expression mediated by AA availability have been intensively studied in bacteria and lower eukaryotes, such as yeasts, the control of transcriptional events, including cell signaling processes, transcription factors and the corresponding cis-elements of DNA, modulated by AAs, have been studied very little in mammals and, in particular, in humans.

Mammals have evolved a wide range of adaptive mechanisms to detect and respond to nutrient fluctuations. A deprivation of AA activates the Amino Acid Response (AAR), which comprises multiple signal transduction pathways. AAR activation regulates gene expression at different levels, both at the level of chromatin structure and at the initiation of transcription. The heterogeneity of the factors involved suggests the existence of multiple AARs that depend on the particular AA, the cell type considered and the configuration of the promoter gene (Kilberg et al., 2012). The complexity of AA-mediated regulation of gene expression is increased by the fact that many of the target genes encode transcription factors that, in turn, act by directing the transcription of several subordinate genes. Studies with the asparagine synthase (ASNS) and C/EBP homologous protein (CHOP) genes, which encodes a nuclear protein related to the C/EBP family of transcription factors, -CCAAT/Enhancer Binding Protein (CCAAT/Enhancer Binding Protein), which forms dimers with other members of its family and is involved in cell apoptosis, have allowed the characterization of AA-specific response sequences in its promoters, termed AA response elements (AARE) and nutrient responsive response elements (NSRE), especially glucose. Although considerable progress has been made in understanding the mechanisms controlling AA-dependent gene expression, numerous questions remain to be resolved.

Generally, a state of malnutrition or poor AA assimilation is associated with a wide range of diseases, in which modulation of gene expression has been recognized as a component. For example, in individuals with a protein deficiency, decreased expression of insulin-like growth factor 1 (IGF-1) has been described, which together with other genes may contribute to morbidity in adults and may decrease growth in malnourished children. In healthy adults, consumption of a low-protein diet decreases IGF-1 expression and increases insulin-like growth factor binding protein 1 (IGFBP1) expression. A protein-deficient diet not only has immediate effects on the expression of certain genes, but also has long-term effects through epigenetic mechanisms. In fact, protein or AA restriction during fetal (in utero) development causes fetal hepatic DNA methylation changes and changes in gene regulation, possibly leading to the development of metabolic syndrome during adulthood. Furthermore, in rats fed wheat gluten, which is deficient in lysine and threonine, there is an induction of genes involved in cholesterol biosynthesis. Similarly, studies in animals fed a diet deficient in methionine or leucine resulted in rapid weight loss, adiposity and impaired health. These results indicate that AA are capable of regulating different metabolic pathways.

In mammary cells, at least three molecules capable of determining AA availability have been described to date:

(1) general non-derepressor control kinase 2 (GCN2), which regulates transcription initiation in cells in the absence or deprived of AA by sensing uncharged tRNAs.
(2) activation of c-JUN by JUN kinase, this pathway is important in pathologies such as cancer and in the development of diabetes.
(3) the mammalian target protein kinase of rapamycin (mTOR), which functions to confirm the existence of an adequate level of AA to support protein synthesis and cell growth.

GCN2 and mTOR kinases are highly conserved, having a major role in controlling transcription and translation, as well as regulating the adaptive response during amino acid deprivation. GCN2 has the ability to sense the levels of each of the AA, it is activated in response to protein fasting, cellular purine limitation or DNA damage. AA limitation activates this kinase through a mechanism involving the accumulation of uncharged tRNA. GCN2 phosphorylates eukaryotic translation initiation factor 2 (eIF2) α at serine 51, leading to a decrease in protein synthesis (**Fig. 4**). Paradoxically, although phosphorylation of eIF2α leads to inhibition of protein synthesis, it leads to an increase in the translation of specific mRNAs, most notably ATF4. Therefore, ATF4 protein synthesis is selectively increased in response to AA deficiency. Among the genes that are activated by ATF4 are those of the transcription factors ATF3, C/EBPβ and CHOP, and the ASNS gene. Some genes regulated by ATF4 require transcriptional upregulation of ATF2, which encodes a histone acetyltransferase. In response to a reduction of AA, and in particular leucine, ATF-2 phosphorylation is induced in human cells, leading to cellular apoptosis.

Cells use translational control to modulate gene expression throughout their life cycle in all tissues and its dysregulation can contribute to numerous diseases, including cancer. Like any biosynthetic process, translation consumes substantial amounts of energy and AA. When either is limited, protein synthesis needs to be inhibited. Therefore, mammals have evolved translation control mechanisms, most of them sensitive to the availability of nutrients, cellular energy, stress situations, hormones and growth factors.

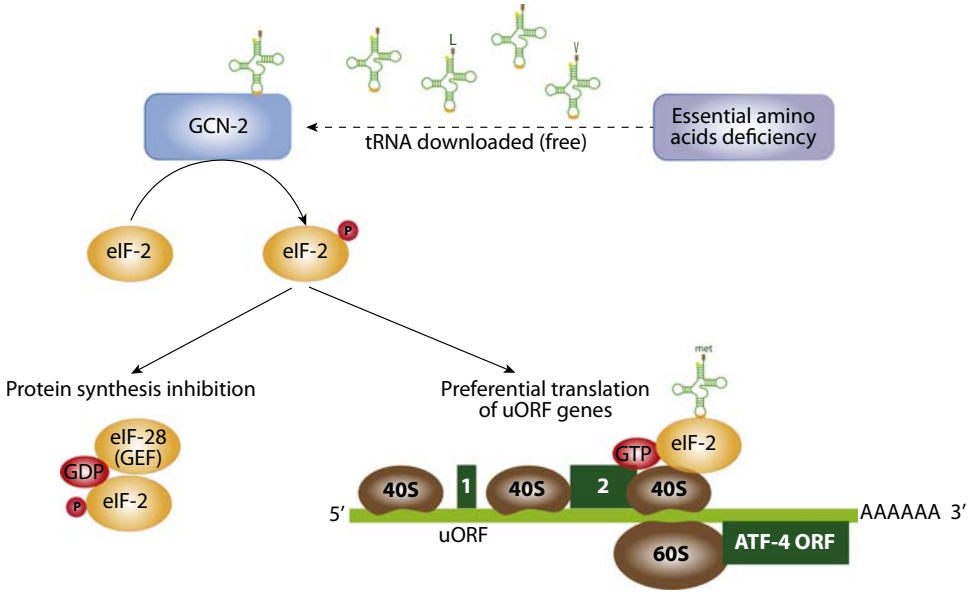


Fig. 4 Mechanism of action of general control nonderepressible 2 kinase (GCN-2) upon essential amino acid deficiency. Phosphorylation of eukaryotic translation initiation factor 2α (eIF-2α) by GCN-2 results in a coordinated response that limits the ribosomal transcription of most messenger RNAs while activating the transcription of specific genes containing upstream open reading frames of (uORF), such as activating transcription factor 4 (ATF-4). GEF: guanosine nucleotide exchange factor.

The limiting step of protein synthesis is translation initiation during which the small unit of ribosomes is recruited to the 5′ end of the mRNA to scan the initiation codon. Therefore, control mechanisms often relate to the control of the initiation factors of this process (protein synthesis, degradation and turnover). One of the mechanisms of control of translation initiation requires the assembly of translation initiation factor 4F (eIF-4F) with the cap at the 5′end mRNA. The eIF-4F consists of three factors (eIF-4E, eIF-4G and eIF-4A) and for assembly eIF-4E binds to the cap at 5′, recruiting eIF-4G and eIF-4A. The eIF-4E factor binding protein (4EBP-1) exerts an inhibitory action on the binding of eIF-4G to eIF-4E and phosphorylation of 4EBP-1 allows the recruitment of eIF-4G and eIF-4A and thereby initiation of protein synthesis.

A key pathway of translation regulation that integrates and responds to environmental causes, such as the availability of AA and energy, and the presence of insulin, is the so-called mTOR pathway, through phosphorylation and inhibition of the 4EBP-1, together with activation of ribosome kinase S6 (S6K), which are described below (Haissaguerre et al., 2014). However, this pathway controls other important metabolic and cellular functions, such as autophagy. **Fig. 5** summarizes the control of amino acid-mediated protein biosynthesis and other signals through the mTOR pathway.

The presence of amino acids, especially leucine, activates the mTOR pathway by phosphorylating 4EBP-1 and allowing the assembly of the eIF-4F complex and, therefore, the initiation of translation. Likewise, mTORC1 phosphorylates eIF-4G at several sites allowing its scaffold function for eIF-4F formation. Another important target of mTORC1 is the S6Ks, especially S6K-1, which coordinate ribosome biogenesis. It is now known that both the mTORC1 complex and S6K-1 associate with the translation initiation complex eIF-3. Thus, immediately after its activation, the mTORC1 complex is recruited by the eIF-3 complex, resulting in the phosphorylation of S6K-1, which is thus activated, as well as of factor 4EBP, which is separated from eIF-4E. On the other hand, the recruitment of active S6K-1 by newly synthesized mRNA is made possible by the scaffolding protein called SKAR, which in turn, in order to be active, needs to be phosphorylated by S6K-1 itself.

Regulation of gene expression by vitamins A and D

Vitamin A and vitamin D present a clearly transcriptional mechanism, while sharing the fact that they are both vitamins and hormones/prohormones (Al Tanoury et al., 2013; Christakos et al., 2016). Their biological actions predominantly involve interaction with specific nuclear receptors, made possible by their lipophilic structure, which allows access through the plasma membranes. Nuclear receptors are located inside the cell, in the cytoplasm or in the nucleus itself, and respond to typically lipid-soluble ligands (which otherwise would not have access) by modulating gene transcription. Class 2 of these receptors include the thyroid hormone receptor (TR), vitamin D receptor (VDR), retinoic acid receptors (RAR and RXR) and PPAR. These receptors generally act as heterodimers in which the retinoid X receptor associates with the VDR, RE, RAR or PPAR. In general, proteins belonging to the nuclear receptor superfamily are single-chain polypeptides with 3 main domains: an amino-terminal variable domain, a highly conserved DNA-binding domain, and a carboxyl-terminal ligand-binding domain. Between the DNA-binding

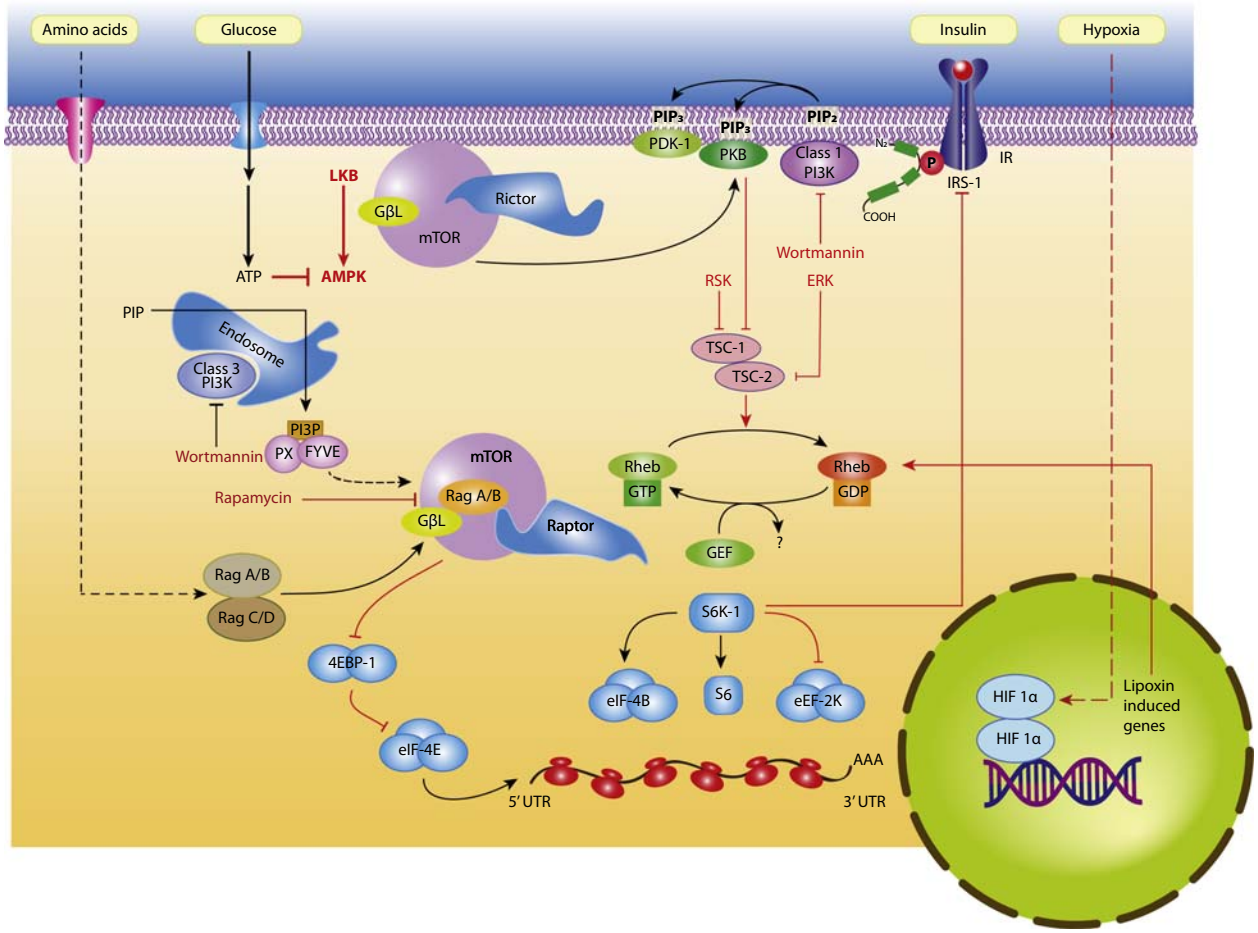


Fig. 5 Control of amino acid-mediated protein biosynthesis and other signals through the mTOR (mammalian target protein kinase of rapamycin) pathway. 4EBP-1: eIF-4E factor-binding protein; AMPK: AMP (adenosine monophosphate) activated protein kinase; ATP:adenosine triphosphate; eEF-2K: eukaryotic elongation factor 2K; eIF-4E: eukaryotic translation initiation factor 4E; ERK: extracellularly regulated kinase; FYRK: eukaryotic translation initiation factor 4E.extracellularly regulated; FYVE (Fab1p, YOTB, Vac1p, and EEA1): proteins containing Fab1p, YOTB, Vac1p, and EEA1 domains; GbL: G protein bsubunit-like, also known as LST-8 (leukocyte specific transcript); GSK-3b: glycogen synthase kinase 3b; GDP: guanosindiphosphate; GEF:guanine nucleotide exchange factors; GTP: guanosine triphosphate; HIF-1a, HIF 1b: hypoxia-inducible factors 1a and 1b; IR: insulin receptor; IRS-1: insulin receptor substrate 1; LKB-1: tumor suppressor; PDK-1: phosphoinositide-dependent protein kinase; PI3K: phosphatidylinositol-3-kinase; PIP2: phosphatidylinositol-bisphosphate; PIP3: phosphatidylinositol-3-phosphate; PKB: protein kinase B; PRAS40: proline-rich protein; PX: protein kinase B; PRAS40: proline-rich protein; PX: proline-rich protein; PX (Phox homology): proteins containing PX domains; Rag (Ras-related small GTPases): family of small GTPases; Raptor (regulatory associated protein of mTOR): regulatory protein associated with mTOR; Redd1/2 (protein regulated in development and DNA damage response): DNA damage-induced transcripts; Rheb: small GTP-bound G protein; Rictor: rapamycin-independent partner of mTOR; RSK: ribosomal S6 kinase; S6K: ribosomal S6 kinases; 5′-UTR: 5′-untranslatable region of mRNA; TSC: tuberous sclerosis complex; TSC-1: hamartin; TSC-2: tuberin.

domain and the ligand-binding domain is a hinge region. Modulation of the transcriptional activity of these receptors is critically dependent on interaction with various cofactors.

The active form of vitamin A (except for visual function) is retinoic acid (transretinoic acid), which acts as an activating ligand for RAR receptors, of which there are 3 types with different isoforms. The active form of the RAR receptors is as a heterodimer, forming a pair with RXR receptors, giving rise to multiple possible combinations. Several possible RXR receptor ligands have been identified, including 9-cisretinoic acid, but it is not clear which of these is relevant in vivo, nor indeed whether the presence of such a ligand is necessary. Other potentially relevant vitamin A receptors have been identified, most notably the PPAR/receptor. In addition, RAR receptors appear to be involved in non-transcriptional modulatory mechanisms (Al Tanoury et al., 2013).

The active form of vitamin D is calcitriol, which acts as an activating ligand of the VDR receptor. Unlike vitamin A, there is a single form of VDR, although multiple polymorphic variants have been identified, some of which appear to be clinically relevant (Christakos et al., 2016). The biologically active form of VDR is as a heterodimer, pairing with RXR receptors. As in the case of vitamin A, there are non-transductional modulation mechanisms, which depend on two types of receptors: the VDR itself, located in the

caveolae of the plasma membrane, and an independent receptor called MARRS (membrane-associated rapid steroid response protein). This type of mechanism produces faster responses than the classical transcriptional actions, but its biological significance is not clear at present.

Regulation of gene expression by trace metals

Trace metals play a fundamental role in human nutrition as most of them are considered essential for a proper cellular functioning. However, excessive intake or exposure can be highly harmful to health. This is why organisms have developed homeostatic mechanisms to control the correct concentrations of these metals (Ng et al., 2015).

Intracellular regulation of iron in man is performed by a post-transcriptional regulatory mechanism, mediated by iron regulatory proteins (IRPs) and iron responsive elements (IREs) and copper actively participates in the function of transcription factors, such as Ace-1, Mac-1 and Amt-1, which are key factors in human metabolism. Intracellular iron concentration modulates the expression of most of the proteins related to iron metabolism, including the transferrin receptor (RTf), the outbound transporter ferroportin, the iron storage protein ferritin, and the inbound iron transporter DMT1.

Iron regulatory proteins IRP1 and IRP2 are cytosolic proteins that bind to hairpin mRNA structures, called iron response elements (IRE), which are located in the 5′ (ferritin and ferroportin mRNA) or 3′ (RTf and DMT1 mRNA) untranslated regions of the mRNA (5′ or 3′ UTRs) that encodes it. In their structure, IRP proteins contain a cube-shaped core between domain 1 and domains 2–4, which allows them to detect the intracellular iron content. When the content of the labile iron pool is low, the core is in a 3Fe-4S (3-iron-4-sulfur) conformational state, i.e., open state. Under these conditions, proteins can bind to the IRE elements of mRNAs and thus regulate their expression. When IRP proteins bind to the 5′ non-coding end of the mRNA, translation is inhibited, as the entry of the major subunit of the translational complex is blocked. When they bind to the 3′ non-coding end, the mRNA is stabilized as the action of RNAases is prevented. Therefore, when the iron content in the labile iron pool is low, the IRPs bind to the IRE elements and prevent the translation of ferritin and ferroportin, resulting in decreased iron storage and reduced iron efflux from the intracellular medium, respectively, and in addition, stabilization of the RTf and DMT1 messengers occurs, resulting in increased iron uptake. When intracellular iron increases, the core IRP proteins change conformational state to 4Fe-4S, i.e., closed state, losing their ability to bind mRNAs.

Transcriptional regulation plays a key role in the control of trace metal homeostatic gene expression, in particular, cadmium, selenium and nickel directly impact transcriptional regulatory mechanisms (Ng et al., 2015). The metallothionein gene has more than one response element in its promoter, including the metal responsive element (MRE). The synthesis of metallothionein augments in response to increased levels of most of the trace metals. The transcription factor MTF-1 allows to connect the transcriptional response mediated by trace metals with the purpose of increasing the intracellular amount of metallothioneins whose function is to contain the intracellular excess of metals avoiding their toxicity. This property gives the system the ability to integrate and coordinate gene expression in order to maintain cellular homeostasis. Indeed, trace metal nutrigenomics and nutrigenetics research has enabled significant progress in understanding the genetic potential to respond to changes in trace metal availability.

Conclusions

A diet rich in carbohydrates increases the expression of the disaccharidases and transporters that are involved in their metabolism, so that in the long term there is a better utilization of the contribution of these nutrients from the diet. This increase in the expression of disaccharidases and transporters will occur mainly at the level of their transcription, although in some cases post-transcriptional regulation also occurs. Glucose participates in the coordinated regulation of insulin gene expression, insulin biosynthesis and insulin secretion. Also, glucose stimulates the transcription of genes coding for glycolytic and lipogenic enzymes: L-PK, ACC, FAS and SCD-1. ChREBP is the key element mediating glucose sensitivity in different cell types.

In addition to their energetic function and their role in membrane composition, fatty acids, both saturated and polyunsaturated, cholesterol and other sterols, as well as various eicosanoids or oxidized derivatives of the latter, modulate the expression of numerous genes involved in the oxidation of fatty acids themselves in liver, muscle and adipose tissue, cholesterol and bile acid synthesis, adipocyte proliferation and differentiation, immune response, antioxidant defense and angiogenesis. The effects of fatty acids are mediated either directly by their specific binding to various nuclear receptors (PPAR, LXR, HNF-4α) producing changes in the trans activation of these transcription factors, or indirectly by changes in the abundance or stability of transcription regulatory factors (SREBP-1c, ChREBP, Nrf2 etc.). On the other hand, fatty acid-responsive transcription factors bind to a fatty acid recognition sequence or response element in the promoter of a region of a target gene, as monomers, as homodimers or as heterodimers with other transcription factors, e.g., retinoid receptors (RXR).

Amino acids regulate gene expression through multiple mechanisms, including both transcriptional and mechanisms, including modulation of transcription, translation and epigenetic mechanisms. Three molecules capable of determining amino acid availability have been described, GCN2, which regulates the initiation of transcription in cells in the absence or deprived of amino acids; activation of c-jun by Jun kinase, an important pathway in diseases such as cancer and in the development of diabetes, and mTOR,

which is capable of determining the existence of an adequate level of amino acids and, among other mechanisms, activates protein synthesis. The amino acid leucine has been identified as the main modulator of the mTORC-1 complex.

Vitamin A and vitamin D present a clearly transcriptional mechanism, while sharing the fact that they are both vitamins and hormones/prohormones. Their biological actions predominantly involve interaction with specific nuclear receptors, made possible by their lipophilic structure, which allows access across plasma membranes. These receptors generally function as heterodimers in which RXR is associated with the VDR, RAR or PPAR.

Intracellular regulation of iron in man follows a post-transcriptional mechanism, mediated by mechanism, mediated by IRPs and IREs. Copper is actively involved in the function of transcription factors such as Ace-1, Mac-1 and Amt-1, which are essential in human metabolism. Metallothionein has more than one response element in its promoter, including the MRE. Transcriptional regulation plays a key role in controlling the expression of trace element homeostatic genes; in particular, cadmium, selenium and nickel directly impact transcriptional regulatory mechanisms.

References

Al Tanoury, Z., Piskunov, A., Rochette-Egly, C., 2013. Vitamin A and retinoid signaling: genomic and nongenomic effects. J. Lipid Res. 54, 1761–1775.

Benhamed, F., Filhoulaud, G., Caron, S., Lefebvre, P., Staels, B., Postic, C., 2015. O-GlcNAcylation links ChREBP and FXR to glucose-sensing. Front. Endocrinol. 5, 230.

Christakos, S., Dhawan, P., Verstuyf, A., Verlinden, L., Carmeliet, G., 2016. Vitamin D: metabolism, molecular mechanism of action, and pleiotropic effects. Physiol. Rev. 96 (1), 365–408.

Contreras, A.V., Torres, N., Tovar, A.R., 2013. PPAR-a as a key nutritional and environmental sensor for metabolic adaptation. Adv. Nutr. 4, 439–452.

Eissing, L., Scherer, T., Todter, K., Knippschild, U., Greve, J., Buurman, W., Pinnschmidt, H., Rensen, S., Wolf, A., Bartelt, A., Heeren, J., Buettner, C., Scheja, L., 2013. De novo lipogenesis in human fat and liver is linked to ChREBP and metabolic health. Nat. Commun. 4, 1528.

Ferguson, L.R., De Caterina, R., Görman, U., Allayee, H., Kohlmeier, M., Prasad, C., Choi, M.S., Curi, R., de Luis, D.A., Gil, Á., Kang, J.X., Martin, R.L., Milagro, F.I., Nicoletti, C.F., Nonino, C.B., Ordovas, J.M., Parslow, V.R., Portillo, M.P., Santos, J.L., Serhan, C.N., Simopoulos, A.P., Velázquez-Arellano, A., Zulet, M.A., Martinez, J.A., 2016. Guide and position of the International Society of Nutrigenetics and Nutrigenomics on personalized nutrition: Part 1—fields of precision nutrition. J. Nutrigenet. Nutrigenomics 9, 12–27.

Filhoulaud, G., Guilmeau, S., Dentin, R., Girard, J., Postic, C., 2013. Novel insights into ChREBP regulation and function. Trends Endocrinol. Metab. 24, 257–268.

Fu, Z., Gilbert, E., Liu, D., 2013. Regulation of insulin synthesis and secretion and pancreatic beta-cell dysfunction in diabetes. Curr. Diabetes Rev. 9, 25–53.

Georgiadi, A., Kersten, S., 2012. Mechanisms of gene regulation by fatty acids. Adv. Nutr. 3, 127–134.

Haissaguerre, M., Saucisse, N., Cota, D., 2014. Influence of mTOR in energy and metabolism homeostasis. Mol. Cell. Endocrinol. 397, 67–77.

Kilberg, M.S., Balasubramanian, M., Fu, L., Shan, J., 2012. The transcription factor network associated with the amino acid response in mammalian cells. Adv. Nutr. 3, 295–306.

Kohlmeier, M., De Caterina, R., Ferguson, L.R., Görman, U., Allayee, H., Prasad, C., Kang, J.X., Nicoletti, C.F., Martinez, J.A., 2016. Guide and position of the International Society of Nutrigenetics and Nutrigenomics on personalized nutrition: Part 2—ethics, challenges and endeavors of precision nutrition. J. Nutrigenet. Nutrigenomics 9, 28–46.

Ng, E., Lind, P.M., Lindgren, C., Ingelsson, E., Mahajan, A., Morris, A., Lind, L., 2015. Genome-wide association study of toxic metals and trace elements reveals novel associations. Hum. Mol. Genet. 24 (16), 4739–4745.

Oosterveer, M., Schoonjans, K., 2014. Hepatic glucose sensing and integrative pathways in the liver. Cell. Mol. Life Sci. 71, 1453–1467.

Protectins, resolvins and maresins

Ángel Gil[a,b,c] **and Luis Fontana**[a,b,c], [a] Department of Biochemistry and Molecular Biology 2, School of Pharmacy, Granada, Spain; [b] Institute of Nutrition and Food Technology "José Mataix", Biomedical Research Center, Parque Tecnológico Ciencias de la Salud, Granada, Spain; and [c] Instituto de Investigación Biosanitaria ibs.GRANADA, Complejo Hospitalario Universitario de Granada, Granada, Spain

© 2023 Elsevier Ltd. All rights reserved.

Introduction 209
Biological functions of pro-resolving mediators derived from n-3 polyunsaturated fatty acids 210
Resolvins 210
Protectins 212
Maresins 212
Conclusions 213
References 214

Key points

- To describe the biochemical mechanisms involved in the resolution of inflammation.
- To know the chemical structures of resolvins, protectins and maresins, and the major pathways involved in their synthesis.
- To understand the variety of receptors that mediate the effects of resolvins, protectins and maresins.
- To know the role of resolvins, protectins and maresins in the resolution of inflammation and other biological effects.

Introduction

The acute inflammatory response can be divided into two phases, an initial onset phase and a resolution phase (**Fig. 1**A). The onset phase is accompanied by the so-called cardinal signs or pillars of inflammation, i.e., flushing (redness), heat (fever), pain, tumor (swelling and edema) and loss of function (*functio laesa)*, and is regulated by numerous chemical mediators. Among these compounds are proteins, peptides, and lipid mediators, including eicosanoids, which give rise to chemical gradients capable of regulating leukocyte trafficking or diapedesis to the affected area. Neutrophil infiltration and accumulation of macrophages as well as monocytes result from the initiation phase of the acute inflammatory response (**Fig. 1**B). Therefore, an inappropriate elimination of these cells after their action can lead to a chronic process (Panigraphy et al., 2021).

The resolution phase is also based on five pillars: (a) elimination of microorganisms, dead cells, cellular debris and antigens; (b) restoration of vascular integrity and perfusion; (c) tissue regeneration; (d) remission of fever; and (e) elimination of pain. During the resolution phase of the inflammatory process, there is a decrease in the presence of proinflammatory mediators, as well as the synthesis of other mediators aimed at restoring homeostasis. The synthesis of these "pre-resolution" mediators has helped to demonstrate that this process is biochemically active (**Fig. 1**B) (Panigraphy et al., 2021; Serhan, 2014).

At the tissue level, restoration of homeostasis is defined as the cessation of leukocyte infiltration in response to chemotactic signals, apoptosis of polymorphonuclear cells and active clearance by macrophages of apoptotic cells. Analysis of exudates following the inflammatory process has identified several specialized pro-resolving mediators (SPMs) (Serhan, 2014; Serhan et al., 2015a). This name encompasses different chemical families of lipid mediators with different functions, including lipoxins, resolvins, protectins and maresins, which mitigate the infiltration of polymorphonuclear leukocytes, decrease the production of proinflammatory mediators and stimulate the uptake of apoptotic polymorphonuclear-dependent macrophages (**Fig. 2**) (Serhan and Petasis, 2011; Serhan et al., 2015b). These mediators are synthesized from n-3 polyunsaturated fatty acids, such as eicosapentaenoic acid (EPA, 20:5 n-3), docosapentaenoic acid (DPA, 22:5 n-3), and docosahexaenoic acid (DHA, 22:6 n-3), except for lipoxins (LX), which are synthesized from arachidonic acid (AA), 20:4 n-6. From EPA derive the E-series resolvins, and DHA is the precursor of the D-series resolvins, protectins and maresins. DHA derivatives are generally referred to as "docosanoids" (Dyall et al., 2022; Ishihara et al., 2019). Like DHA, DPA is a precursor of docosanoids whose physiological effects are however still poorly known (Drouin et al., 2019; Hansen et al., 2017).

Fig. 3 represents a general scheme of the synthesis of resolvins, protectins and maresins and their role in inflammation resolution and homeostasis.

In addition, **Fig. 4** shows, as an example, the chemical structures of arachidonic acid-derived lipoxins LXA4 and LXB4, EPA-derived resolvins E1 and E2, and DHA-derived resolvin D1, protectin D1 (also called neuroprotectin (NPD-1)) and maresin 1, all derived from DHA, as well as the interaction with currently known receptors.

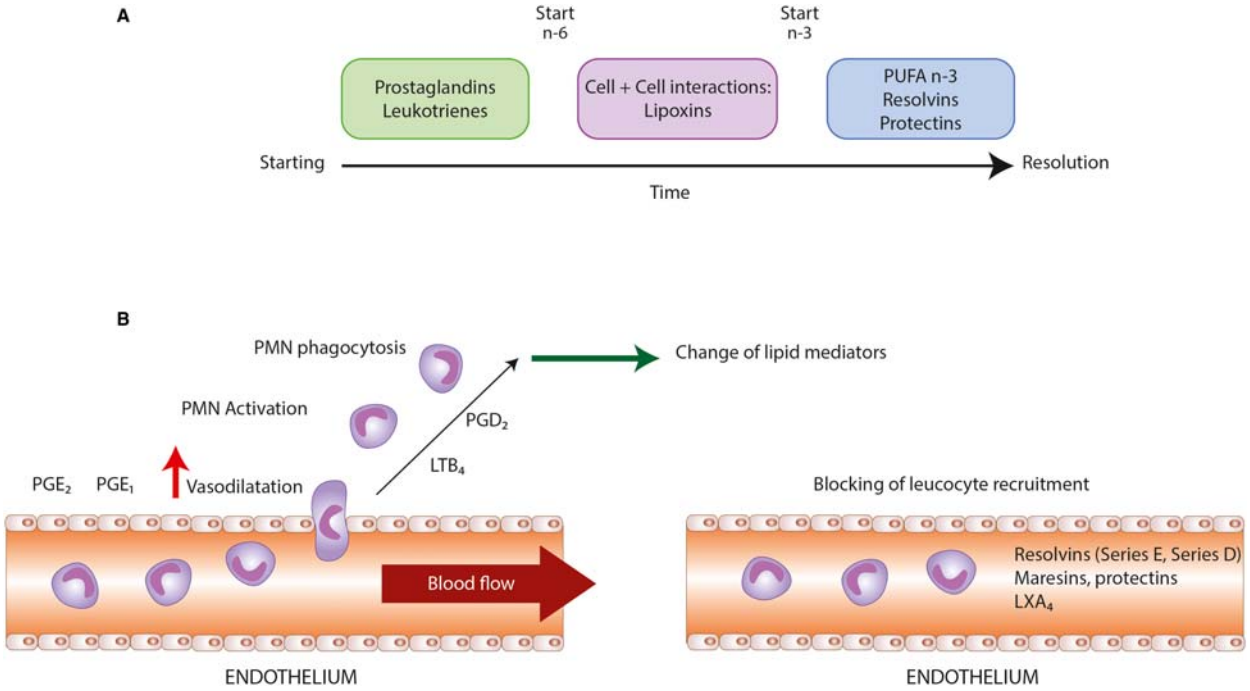


Fig. 1 Role of lipoxins, resolvins, protectins and maresins in the resolution of the inflammatory reaction. (A) Hypothetical sequence of lipid-derived mediators that moderate acute inflammation from onset to resolution and return to homeostasis. (B) Lipid mediators controlling influx to the site of inflammation, with succession from the proinflammatory leukotriene B4 (LTB4) to the resolving factor lipoxin A4 (LXA4). PUFA: polyunsaturated fatty acids; PG: Prostaglandin; PMN: Polymorphonuclears.

In addition to their participation in the resolution of inflammation, some of these compounds play a protective role on the nervous system, retina, liver and lungs (Bazan, 2009; Levy and Serhan, 2014). As their name indicates, this protection is mainly exerted by protectins. Although the mechanism is not sufficiently elucidated, these findings underline the importance of DHA in these tissues. As is schematized in Fig. 5, one of the possible mechanisms is related to cell survival, since protectins inhibit apoptosis (Serhan et al., 2015b).

Biological functions of pro-resolving mediators derived from n-3 polyunsaturated fatty acids

SPMs induce an anti-inflammatory response by inhibiting granulocyte migration and activation, as well as sensory neuronal activation, and limiting cytokine production by a variety of structural cells, including epithelial cells, endothelial cells and fibroblasts. In addition, SPMs have multiple actions by regulating the sentinel cells of the innate immune system and by decreasing the production of cytokines and amphiregulins. They also promote resolution by inducing the production of regulatory T cells and stimulating natural killer (NK) cells to trigger granulocyte apoptosis, engage macrophages to phagocytose bacteria and other noxious stimuli, and eliminate apoptotic cells via efferocytosis (Serhan, 2014; Serhan et al., 2015a,b).

Resolvins

The E series resolvins (RvE1, RvE2 and RvE3) are produced by the vascular endothelium with the intervention of aspirin-modified cyclooxygenase 2 (COX-2), which transforms EPA to 18R-hydroperoxyeicosapentanoic acid and 18S-hydroperoxyeicosapentanoic acid. These intermediates are rapidly metabolized by 5 lipoxygenase (5-LOX) to give rise to resolvins E1 and E2 (Serhan and Petasis, 2011; Spite et al., 2014). The production of resolvin E1 is increased in the plasma of individuals ingesting EPA in the diet, resulting in the amelioration of clinical signs of inflammation (Ishihara et al., 2019). Similarly, DHA-derived resolvins, the D-series resolvins, have been shown to be effective in decreasing inflammation by decreasing platelet-leukocyte adhesion and converting DHA into molecules with dual anti-inflammatory and pro-inflammatory functions (Serhan and Petasis, 2011; Spite et al., 2014). Notably, both RvE1 and RvE2 are resolvins that are antagonists for the endogenous LTB4 receptor, the BLT-1 receptor, which explains their strong ability to control leukocyte trafficking to the inflammatory focus (Fig. 6).

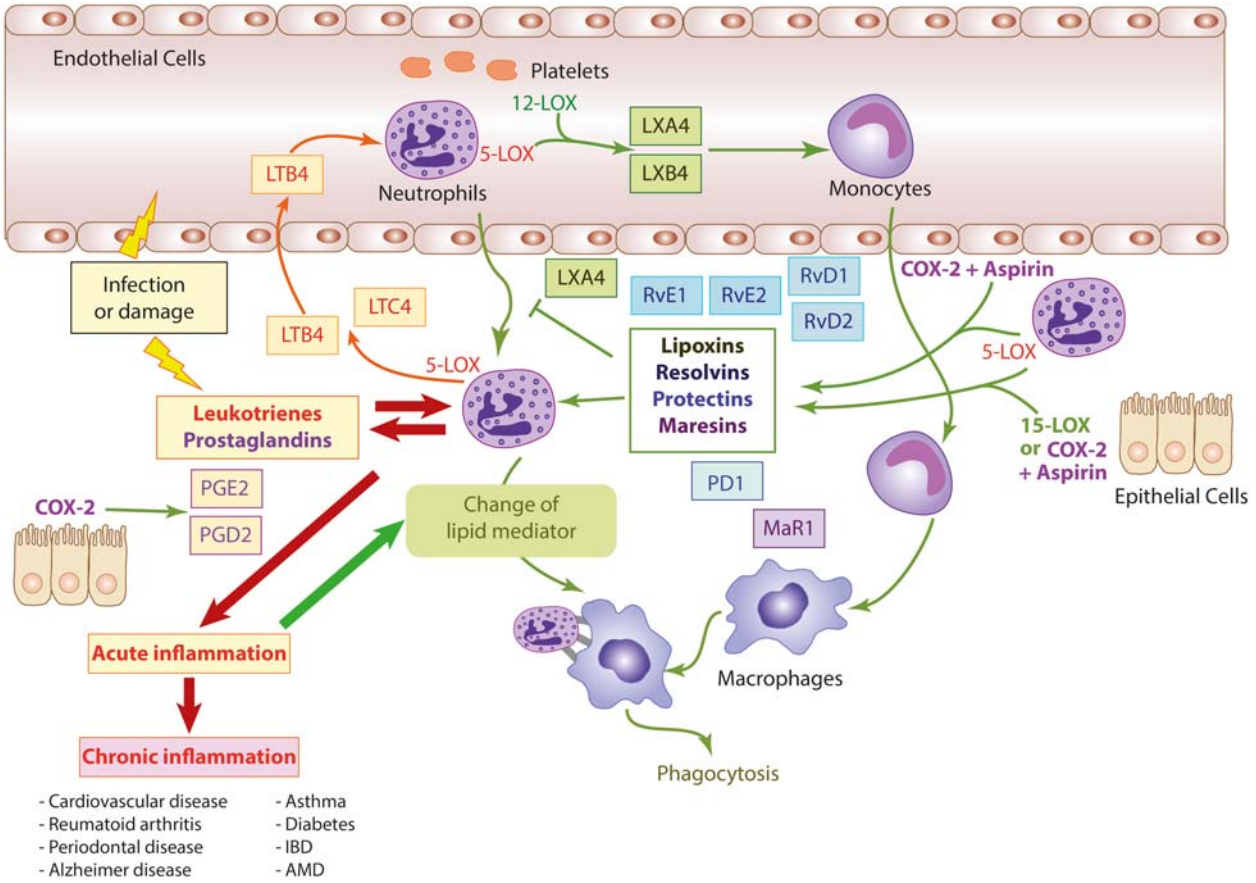


Fig. 2 Roles of lipoxins, resolvins, protectins and maresins in the resolution of acute inflammation. 5-LOX: 5-lipoxygenase; 12-LOX: 12-lipoxygenase; AMD: Age-related macular degeneration; IBD: Inflammatory bowel disease; LTA4, LTB4 and LTC4: Leukotrienes; LXA4: and LXB4: Lipoxins A4; MaR1: Maresin 1; PD1: Protectin D1; PGD2 and PGE2: Prostaglandins; RvE1 a RvE2: Resolvins derived from eicosapentaenoic acid; RvD1 and RvD2: Resolvins derived from docosahexaenoic acid.

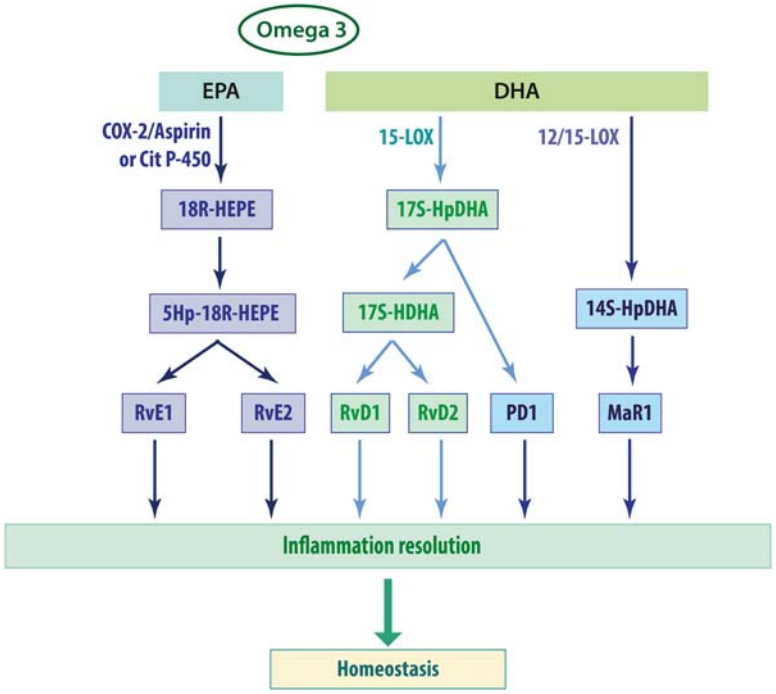


Fig. 3 General scheme of eicosanoid and docosanoid synthesis and their role in inflammation. AT-LXA4: Aspirin-mediated lipoxin A4; COX-2: Cyclooxygenase 2; DHA: Docosahexaenoic acid; EPA: Eicosapentaenoic acid; 17S-HDHA: Hydroxydocosahexaenoic acid; 14S-HpDHA and 17S-HpDHA: Hydroperoxyhydroxyeicosahexaenoic acids; 5Hp-18R-HEPE: Hydroperoxy-hydroxyeicosapentaenoic acid; 18R-HEPE: Hydroxyeicosapentaenoic acid; 15S-HETE and 15R-HETE: Hydroxyeicosatetraenoic acids; LOX: Lipoxygenases; LTA4, LTB4 and LTC4: Leukotrienes; LXA4: Lipoxin A4; MaR1: Maresin 1; PD1: Protectin D1; PGD2 and PGE2: Prostaglandins; PGH2: Prostaglandin hydroperoxide; Rv: Resolvins.

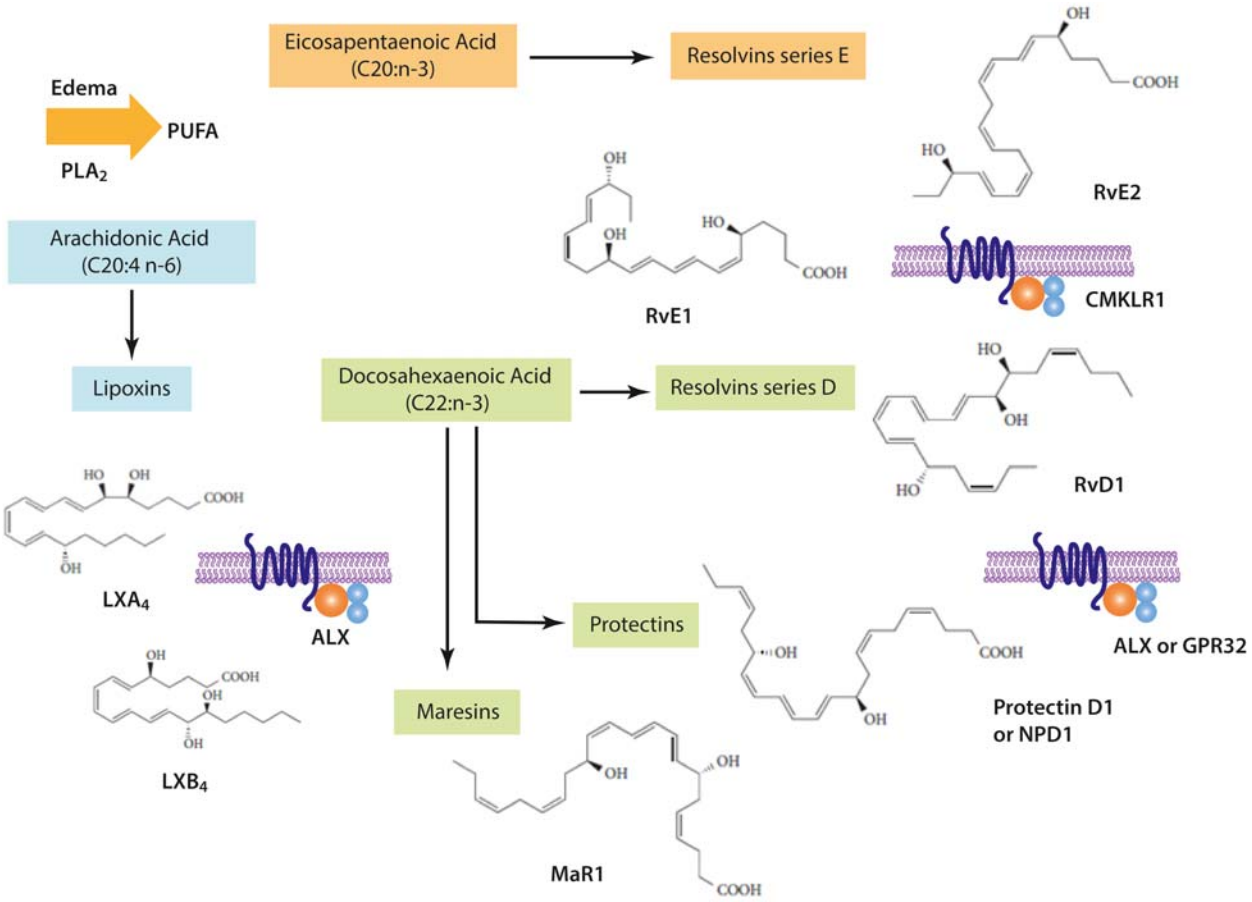


Fig. 4 Schematic of the synthesis and examples of structures of lipid mediators of inflammation resolution. PUFA: Polyunsaturated fatty acids; ALX: Lipoxin A4 receptor; CMKLR-1: Chemokine-like receptor 1; GPR-32: G protein-coupled receptor 32; LXA4 and LXB4: Lipoxins A4 and B4; MaR1: Maresin 1; PLA2: Phospholipase A2; Rv: Resolvins.

Protectins

Independently of the neuroprotective action of protectin D1 mentioned above (Fig. 5), protectins decrease the secretion of interferon-gamma (IFN-γ), as well as tumor necrosis factor-alpha (TNF-α), by blocking T-cell migration. Several studies have found a new pathway of protectin synthesis, known as protectin D1, mediated by aspirin (Serhan et al., 2015b; Hamidzadeh et al., 2022). In this new pathway, the synthesis of protectins is produced from DHA by the action of COX; this compound presents positive interactions with receptors of the peroxisome proliferative receptor-gamma (PPAR-γ) family. Both protectins, regardless of the synthesis pathway by which they are obtained, decrease neutrophil transmigration through endothelial cells and enhance the efferocytosis process (from the Latin *effere*, "to bring to the grave") performed by macrophages for the elimination of apoptotic neutrophils. Although protectin D1 can decrease leukocyte recruitment, leukocyte influx is not completely blocked. However, this fact does not compromise the host defense.

Maresins

In recent years, a new family of mediators of the resolution of acute inflammation, also derived from DHA and produced by macrophages, has been discovered and termed maresins. Members of this family have potent direct actions on phagocytes, including inhibition of neutrophil recruitment or stimulation of efferocytosis, a process very similar to macropinocytosis, whereby phagocytic cells eliminate apoptotic and necrotic cells (Serhan et al., 2015b).

Maresin biosynthesis is initiated in macrophages by the action of 14-LOX from DHA. Maresin 1 (MaR1) is the first identified member of the macrophage-derived proresolving mediator family and is known to exert direct actions on leukocytes. MaR1 is a more potent mediator than RvD1 in the case of efferocytosis stimulation and exerts a pivotal effect in regulating the resolution

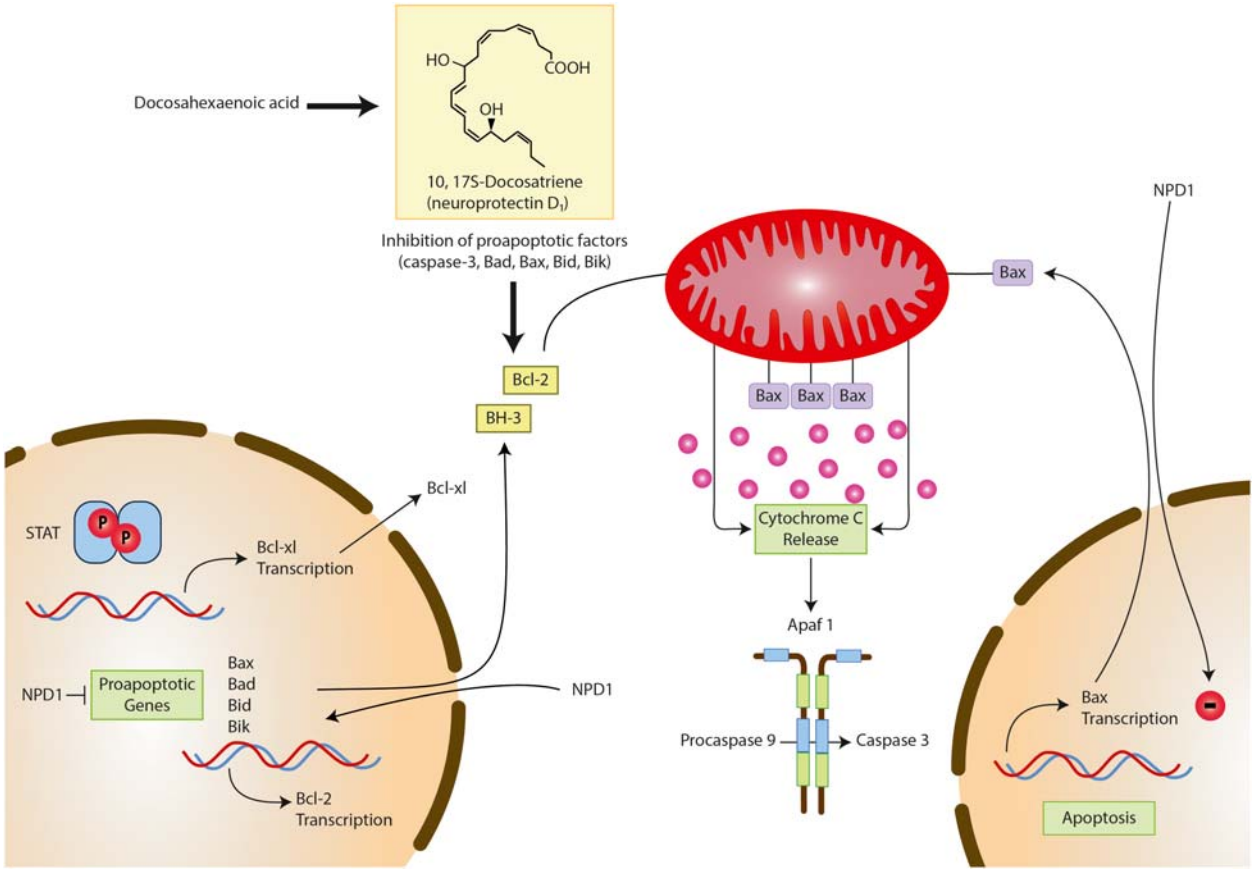


Fig. 5 Role of docosahexaenoic acid and docosanoids in neuroprotection. APAF1: Apoptosis protease-activating factor; NPD-1: Neuroprotectin D1.

phase of the inflammatory process, tissue regeneration and damage resolution. MaR2 has also been discovered, which exerts an important action *in vivo* by limiting polymorphonuclear trafficking (Dyall et al., 2022).

The presence of MaR1 in the inflammatory response occurs at late stages with the entry into play of macrophages. Macrophages are key regulators of the inflammatory response, with different subtypes linked to the spread or resolution of inflammation. Within this context, there are two broad categories of macrophages: classical or M1 macrophages, with proinflammatory properties, and M2 macrophages, linked to the resolution phase of the inflammatory response and the restoration of homeostasis. Several studies have shown that both DHA and DHA-derived mediators of resolution, including RvD1, can stimulate the macrophage phenotype from the proinflammatory profile toward the resolution phase phenotype, M2. It has also been found that macrophages with M2 phenotype are associated with high levels of MaR1, a finding that allows linking it to their regenerative and homeostatic actions. In addition to its actions on leukocytes, this mediator exerts direct actions on tissues, inducing regeneration, probably by a mechanism of stem cell regulation, promoting their differentiation, a fact similar to that recently discovered for another mediator, protectin D1. The activation of MaR1 may influence the resolution of inflammation through the selective reduction of LTB4 by direct inhibition of its synthesis.

Conclusions

SPMs include several chemical families of lipid mediators with different functions, including lipoxins, resolvins, protectins and maresins, which block the chemotaxis of polymorphonuclear leukocytes, decrease the production of cytokines and other proinflammatory mediators and stimulate the phagocytosis within inflamed tissues, leading to inflammation resolution and homeostasis.

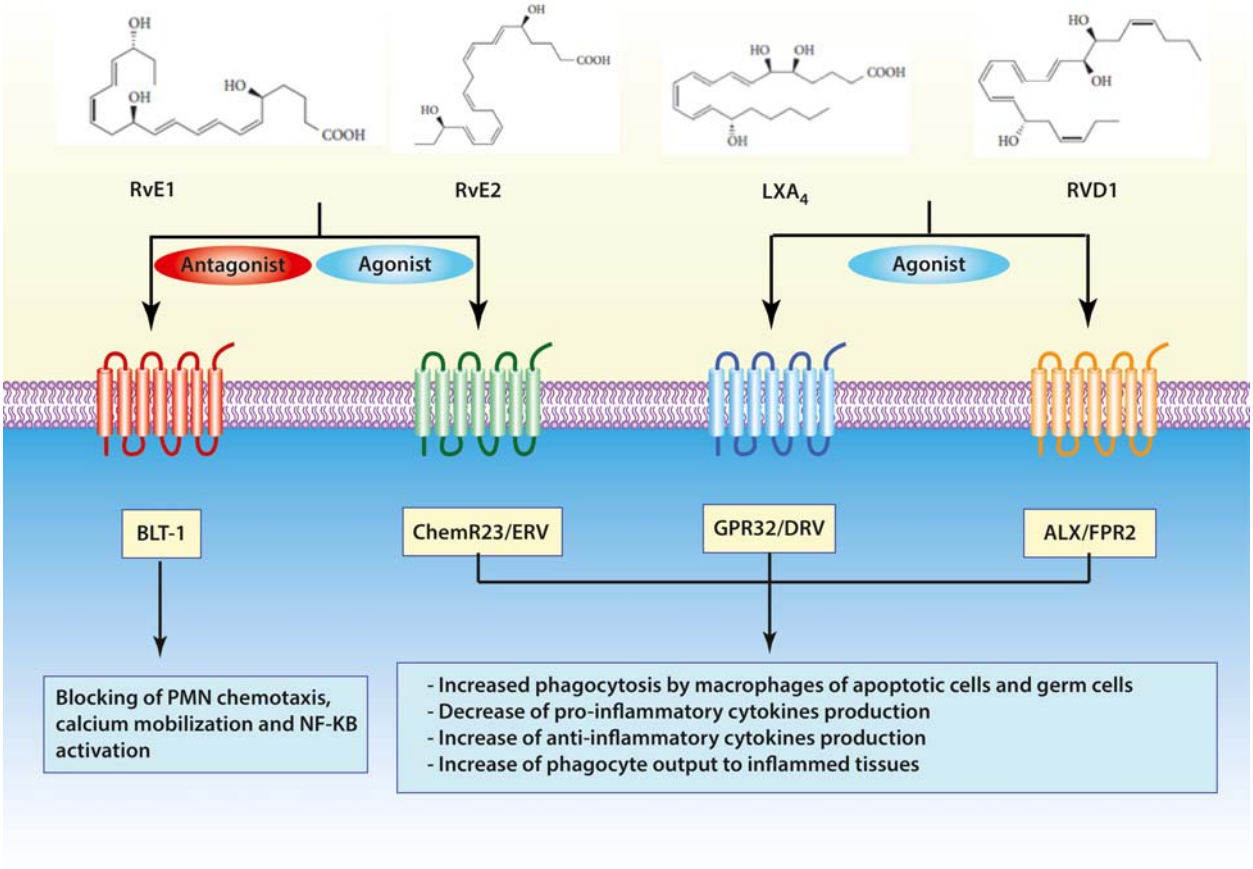


Fig. 6 Receptors for some lipid mediators specialized in the resolution of acute inflammation and their function. ALX/FPR-2: Lipoxin receptor; BLT-1: Leukotriene receptor; Chem R23/ERV: Resolvin E receptor, quemerin E23/R23; GPR-32/DRV: G protein-coupled receptor, GR32, for D-series resolvins; LXA2: Lipoxin A2; NF-kB: Nuclear factor kappa of B lymphocytes; PMN: Polymorphonuclear leukocytes; Rv: Resolvins.

References

Bazan, N.G., 2009. Neuroprotectin D1-mediated anti-inflammatory and survival signaling in stroke, retinal degenerations, and Alzheimer's disease. J. Lipid Res. 50, S400–S405. https://doi.org/10.1194/jlr.R800068-JLR200.

Drouin, G., Rioux, V., Legrand, P., 2019. The n-3 docosapentaenoic acid (DPA): a new player in the n-3 long chain polyunsaturated fatty acid family. Biochimie 159, 36–48. https://doi.org/10.1016/j.biochi.2019.01.022.

Dyall, S.C., Balas, L., Bazan, N.G., Brenna, J.T., Chiang, N., da Costa Souza, F., Dalli, J., Durand, T., Galano, J.M., Lein, P.J., Serhan, C.N., Taha, A.Y., 2022. Polyunsaturated fatty acids and fatty acid-derived lipid mediators: recent advances in the understanding of their biosynthesis, structures, and functions. Prog. Lipid Res. 86, 101165. https://doi.org/10.1016/j.plipres.2022.101165.

Hamidzadeh, K., Westcott, J., Wourms, N., Shay, A.E., Panigrahy, A., Martin, M.J., Nshimiyimana, R., Serhan, C.N., 2022. A newly synthesized 17-epi-NeuroProtectin D1/17-epi-Protectin D1: authentication and functional regulation of Inflammation-Resolution. Biochem. Pharmacol. 203, 115181. https://doi.org/10.1016/j.bcp.2022.115181.

Hansen, T.V., Dalli, J., Serhan, C.N., 2017. The novel lipid mediator PD1n-3 DPA: an overview of the structural elucidation, synthesis, biosynthesis and bioactions. Prostag. Other Lipid Mediat. 133, 103–110. https://doi.org/10.1016/j.prostaglandins.2017.06.003.

Ishihara, T., Yoshida, M., Arita, M., 2019. Omega-3 fatty acid-derived mediators that control inflammation and tissue homeostasis. Int. Immunol. 31, 559–567. https://doi.org/10.1093/intimm/dxz001.

Levy, B.D., Serhan, C.N., 2014. Resolution of acute inflammation in the lung. Annu. Rev. Physiol. 76, 467–492. https://doi.org/10.1146/annurev-physiol-021113-170408.

Panigrahy, D., Gilligan, M.M., Serhan, C.N., Kashfi, K., 2021. Resolution of inflammation: an organizing principle in biology and medicine. Pharmacol. Ther. 227, 107879. https://doi.org/10.1016/j.pharmthera.2021.107879.

Serhan, C.N., 2014. Pro-resolving lipid mediators are leads for resolution physiology. Nature 510, 92–101. https://doi.org/10.1038/nature13479.

Serhan, C.N., Petasis, N.A., 2011. Resolvins and protectins in inflammation resolution. Chem. Rev. 111, 5922–5943. https://doi.org/10.1021/cr100396c.

Serhan, C.N., Chiang, N., Dalli, J., 2015a. The resolution code of acute inflammation: novel pro-resolving lipid mediators in resolution. Semin. Immunol. 27, 200–215. https://doi.org/10.1016/j.smim.2015.03.004.

Serhan, C.N., Dalli, J., Colas, R.A., Winkler, J.W., Chiang, N., 2015b. Protectins and maresins: new pro-resolving families of mediators in acute inflammation and resolution bioactive metabolome. Biochim. Biophys. Acta 1851, 397–413. https://doi.org/10.1016/j.bbalip.2014.08.006.

Spite, M., Clària, J., Serhan, C.N., 2014. Resolvins, specialized proresolving lipid mediators, and their potential roles in metabolic diseases. Cell Metab. 19, 21–36. https://doi.org/10.1016/j.cmet.2013.10.006.

Organic: Biochemical mechanisms and regulation of vitamins and vitamin-like cofactors

Robert B. Rucker[a], **Francene M. Steinberg**[a], **and Winyoo Chowanadisai**[b], [a] University of California, Davis, Davis, CA, United States; and [b] Oklahoma State University, Stillwater, OK, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of E.D. Harris, Cofactors: Organic, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 366–377, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00058-1.

Introduction **217**
General principles **217**
Physiology: perspectives on the digestion, absorption, transport, and the metabolism of vitamins 217
Digestion and absorption 217
Whole-body, organ, and cellular transport 221
Turnover 222
Vitamin requirements **222**
Functions and mechanisms: fat-soluble vitamins **224**
Vitamin A 224
Vitamin K 227
Functions and mechanisms 227
Blood coagulation, osteocalcin, and Gla-matrix proteins 227
Vitamin K antagonists 228
Vitamin E 228
Vitamin D 229
Functions and mechanisms: water-soluble vitamins **230**
Cofactor, coenzyme, and cell signaling functions 230
Oxidation-reductions: ascorbic acid, riboflavin, and niacin 231
Ascorbic acid 231
Functions 231
Metabolism 232
Riboflavin (vitamin B2) 233
Functions 233
Metabolism 233
Niacin 233
Functions 234
Metabolism 236
Generation of leaving groups: thiamine, vitamin B6, pantothenic acid **236**
Thiamin 236
Functions 236
Metabolism 237
Vitamin B6 238
Functions 238
Metabolism 239
Pantothenic acid 239
Functions 239
Metabolism 241
Methyl transfer reactions: biotin, folic acid, vitamin B12 **241**
Biotin 241
Metabolism 242
Folic acid 243
Functions 243
Metabolism 244
Vitamin B12 244
Metabolism 245
Conditional nutrients **246**
Conclusion **249**
References **249**
Further reading **249**

Key points

- Chemical and biological features for fat- and water-soluble vitamins are described with a primary focus on their mechanisms of action. In addition, the mechanism of action for selected vitamin-like conditional nutrients are briefly reviewed.
- Descriptions for the primary functions of fat-soluble and water-soluble vitamins, including the selected conditional nutrients, are provided. Factors (e.g., dietary deficiencies, genetic polymorphisms, dietary interactions) are identified that result in perturbations of essential functions. Other perspectives include why certain vitamins are well-suited to serve as catalysts for transfer, redox, or single carbon reactions.
- Key features of the whole-body regulation of vitamins include descriptions for their digestion, absorption, transport, and cellular metabolism. It is emphasized that the dietary requirements for vitamins are allometric and most related to energy needs and metabolic rate.

Glossary

ABC transporters (ATP-binding cassette transporters) are one of the most prominent ATP-requiring transporter families

Autosomal recessive An autosomal recessive disorder has two copies of an abnormal gene (also cf., Recessive gene)

Biocatalyst A substance, especially an enzyme, that initiates or modifies the rate of a chemical reaction

Biofactors Any material that has a significant biochemical function

Carbonium and Carbanions Carbonium ion and carbanion are charged chemical species containing carbon atoms with different valences, i.e., a carbanion is a trivalent ($C^{\sigma+}$), and a carbanion is pentavalent ($C^{\sigma-}$)

Corrin ring A corrin features four pyrrole-like subunits organized into a conjugated structure

Crohn's disease A type of inflammatory bowel disease associated with abdominal pain, severe diarrhea, fatigue, weight loss, and malnutrition

Cytochromes P450 (CYPs) Are enzymes containing heme as a cofactor and function as monooxygenases oxidizing steroids, fatty acids, and xenobiotics. In addition, they are essential for the clearance and catabolism of various compounds

Energy components (Thermodynamic and Nutritional) The three principle thermodynamic energy components are entropy (system thermal energy unavailability for conversion into mechanical work), enthalpy (the total heat content of a system), and free energy (a thermodynamic quantity equivalent to the capacity of a system to do work). In nutritional terms, enthalpy most represents heat of combustion (e.g., of individual foods). Free energy broadly represents basal metabolism plus work. Finally, entropy may be viewed as a force for catalysis and heat production

Dietary Reference Intakes (DRI) Set of four reference values Estimated Average Requirements (EAR), Recommended Dietary Allowances (RDA), Adequate Intakes (AI), and Tolerable Upper Intake Levels (UL)

GPCR (G-protein-coupled receptors) are common signaling receptors in cells. Their activity is regulated by factors that control their ability to bind to and hydrolyze guanosine triphosphate

JAK-STAT signaling factors (Janus Kinase and Signal Transducer and Activator of Transcription signaling) Can interconnect with other cell-signaling pathways involved in processes such as immunity, cell division, cell death, and tumor formation

Leaving groups Moieties in compounds that have the characteristics of a weak base. The weaker the base, the better the leaving group

NF-κB Proteins are structurally-related transcription factors that control numerous cellular and organismal processes related to immune and inflammatory responses, developmental strategies, and cellular growth

Polymorphism The inheritance of a trait controlled by a single genetic locus with two alleles, in which the least common allele has a frequency of about 1% or greater. Genetic polymorphism is a difference in DNA sequence among individuals, groups, or populations

RARs (Retinoic acid receptors) and RXR (retinoid X receptors) Are ligand-controlled transcription factors that function as heterodimers to facilitate regulation of cell growth, differentiation, survival

Recessive genes Genes whose effects are masked in the presence of a dominant gene. Chromosomal DNA has two alleles one inherited from the mother and the other from the father. A recessive gene is only expressed when an organism has two recessive alleles (known as homozygous recessive). If an organism has one dominant and one recessive allele, it will show the dominant trait

REE (Resting energy expenditure) represents the energy as calories or joules required by the body at rest

ROS (Reactive oxidant species) are unstable molecules (free radicals) that contain oxygen and react with other molecules in a cell, causing a wide variety of cellular damage

TRPV6 (Transient receptor potential cation channels**)** mediate calcium movement in cells

Xenobiotics Are chemical substances that include plant constituents, drugs, pesticides, cosmetics, flavorings, fragrances, food additives, industrial chemicals, and environmental pollutants, i.e., not endogenous to animal organisms

Xerophthalmia Is an abnormal dryness and inflammation of the conjunctiva and cornea of the eye, most often associated with vitamin A deficiency

Introduction

Relationships linking food components to health were first described in texts from Greek philosophers in the fourth and fifth centuries BC, the Egyptian medical text, the *Papyrus Ebers* (written 1550–1570 BC), and the *Qianjin Fang*. The *Qianjin Fang* (Precious Prescriptions for Emergencies) was edited by the Chinese physician Sun Simiao in 652 BC and is thought to be the first nutrition-related guide used in traditional Chinese medicine. It comprises 30 volumes and addresses more than 200 topics, many of which are related to diet and nutrition. An example taken from the *Papyrus Ebers* suggests that liver consumption is a treatment to improve vision, including night blindness.

By the mid-1700s, a large number of documents focused on the treatment of specific diseases. Perhaps the best known is Dr. James Lynn's treatise (*Treatise on the Scurvy*), noting that fresh fruits and vegetables seemed effective in treating scurvy. In the late 1800s, relationships associated with the consumption of monotonous diets became appreciated, i.e., diets based on a limited selection of similar ingredients. For example, the occurrence of pellagra was associated with consuming diets based almost exclusively on corn (niacin deficiency), and the occurrence of polyneuritis was related to consumption of polished rice (Beriberi, thiamin deficiency).

By the 1920s, considerable information regarding potential micronutrient deficiencies was known, developed concurrently with the isolation of so-called underlying curative factors. However, structural elucidation methods and the ability to thoroughly assess functional mechanisms were not as available. Thus, early progress was largely inferential. For example, in an autobiographical article (Szent-Gyorgyi, 1963), Albert Szent-Gyorgyi, who in 1937 was the Nobel Prize recipient for his work in part related to vitamin C, describes attempts to publish features of a new 6-carbon sugar with strong reducing potential. The crystalline form of the new sugar was obtained from oranges, lemons, cabbage, and adrenal glands. Because he did not have good structural information, he first called the unknown carbohydrate "ignose." When the Biochemical Journal rejected this name, he humorously suggested naming it "godnose," as in "only God knows." Eventually, the "new 6-carbon sugar" became designated as factor C and eventually vitamin C. **Table 1** summarizes specific functions of the 13 established vitamins starting with the first to be recognized, fat-soluble factor A (in 1913) to the last, vitamin B12 (isolated in 1947).

From the 1950s, the search for additional vitamin and vitamin-like factors continued, along with studies focused on a better understanding of vitamin requirements and mechanisms. In the 1980s, many of the current platforms for structural and mechanistic determinations became routinely available (e.g., computing applications, high-resolution spectroscopic methods, nuclear magnetic resonance, X-ray diffraction methodologies).

General principles

When serving as biocatalysts, vitamins facilitate transition states that lower the activation energy of given reactions. Biocatalysts increase reaction rates or enable reactions at lower temperatures. They also affect the reaction environment or bind substrates to change physical and chemical properties, making them more susceptible to chemical alteration. Moreover, biocatalysts facilitate the coupling of the spontaneous processes of catabolism to the less spontaneous processes of anabolism. Another essential function is the transfer of specific functional groups from which substances are produced, maintained, or destroyed, and energy is made available to produce heat or work. Such transfer reactions range from redox reactions to chemically complex bond formation and cleavage steps that result in the transfer of functional groups and changes in cellular energy or chemical potentials. In addition, it is now clear that vitamins and vitamin-like compounds influence many processes essential to cell signaling or gene regulation.

Physiology: perspectives on the digestion, absorption, transport, and the metabolism of vitamins

Digestion and absorption

The following provides perspectives critical to understanding key features pertinent to the cellular, organ, or whole-body regulation of vitamins. The body handles the digestion of food-based vitamins and cofactors through specific and nonspecific pathways similar to other exogenous compounds (xenobiotics) undergoing absorption, digestion, metabolism, and excretion processes.

The fat-soluble vitamins (A, D, E, and K) are absorbed from the intestinal lumen by mechanisms similar to those used to absorb other dietary lipids. In the small intestine, lipid substances are first incorporated into luminal lipid-derived micelles containing bile acids. Such lipid micelles initially form as large fat droplets rich in dietary triglycerides and other lipids esters and ethers. However, lipases secreted from the pancreas and intestinal cells along with bile acids facilitate the formation of amphipathic micelles. Micelles are colloidal, i.e., heterogeneous mixtures of dispersed particles smaller than lipid enriched suspensions (e.g., 4–8 nm in diameter). Accordingly, micelles have a high probability of contact with small intestinal enterocytes capable of absorbing lipids. The lipids enter such enterocytes primarily by facilitated diffusion processes. Overall, the process is very efficient for dietary lipids; for example, over 90% of dietary triglycerides are absorbed via small intestinal enterocytes. For most fat-soluble vitamins, 40–80% are absorbed when sufficient dietary lipid is present. Next, within the enterocyte, lipid components of the micelle, including fat-soluble vitamins, are further processed, and eventually incorporated into chylomicrons and exported via exocytosis into the lymphatic system and the bloodstream. As it relates to the nutritional status of each of the fat-soluble vitamins, intestinal, biliary, and pancreatic diseases that cause decreased dietary lipid absorption may also cause a decrease in the absorption of fat-soluble vitamins and result in compromised vitamin status.

Analogous to the strategies for fat-soluble vitamins, a strategic step in enhancing water-soluble vitamin bioavailability is solubilization. Water-soluble vitamins exist in foods in their various cofactor, coenzyme, and co-substrate forms. Accordingly, the first step in their liberation by gastrointestinal enzymes (proteinases, hydrolases, phosphatases, nucleosidases, etc.) followed their association with specific carriers to increase solubility or protect from deleterious factors in the intestinal luminal environment.

Table 1 Vitamins – active forms, primary functions, and requirements.

Fat-soluble vitamins

Vitamin *(other designations) and* ***precursors***	***Adult daily maintenance requirements***	***Active forms*** *cofactor, co-substrate, co-enzyme forms*	***Physiological functions***	***Biochemical functions***	***Associated deficiency disease***
Vitamin A (retinoids, includes retinol, retinal, retinoic acid) **β-carotenoids:** β-carotene + O_2 → 2 retinals	**M**: 900 μg (3000 IU) **W**: 700 μg (2333 IU) **UL**: 3000 μg (∼10,000 IU)	Cis-trans isomers of retinol, retinal, retinoic acid	Vision, cell division, growth, reproduction, immunity; differentiation, and proliferation of epithelial cells.	Opsin to rhodopsin conversion; cell signaling (e.g., retinoic acid receptor and retinoid X receptor interactions); antioxidant defense.	Xerophthalmia (nutrition-related night blindness, nyctalopia); impaired reproductive, neonatal growth development.
Vitamin D (D_2-ergocalciferol and D_3-cholecalciferol) **7-dehydrocholesterol:** 7-DC + UV light → D_2 (plants) or D_3 (animals)	**M & W:** 15–20 μg (∼600–800 IU) **UL**: 100 μg (4000 IU)	1α,25-diOH-vitamin D_2 or 1α,25-diOH-vitamin D_3.	Bone mineralization; CA and P homeostasis; non-skeletal effects: immune, endocrine (e.g., insulin regulation); cardiovascular and neurotrophic functions.	Transcriptional activation through vitamin D receptor (VDR) and retinoid X receptor interactions; regulation of cell proliferation and differentiation.	Osteomalacia, rickets
Vitamin E (tocopherols, includes 8 isoforms: α-, β-, γ-, and δ-tocopherol and α-, β-, γ-, and δ-tocotrienol).	**M & W:** 15 mg (∼22 IU) **UL**: 100 mg (1500 IU)	α-tocopherol is considered the most active ROS scavenger and γ-tocopherol is the best scavenger of peroxyl radicals and reactive nitrogen species.	ROS scavenging; alterations in the activities of protein kinases, phospholipases, cyclooxygenases, various phosphatases; γ-tocopherol is capable of trapping electrophilic mutagens.	To a varying degree, tocopherol isomers act as antioxidants. In addition, the unsubstituted C-5 position of γ-tocopherol can also trap lipophilic electrophiles such as reactive nitrogen oxide species.	Increased potential for nerve and muscle damage; a weakened immune system; hemolytic anemias
Vitamin K (K_1-Phylloquinone and K2-menaquinone**).**	**M:** 120 μg, **W:** 90 μg **UL**: NA	Both K_1 and K_2 can act as vitamin K-dependent carboxylases.	Vitamin K serves as a cofactor for the γ-glutamyl carboxylases involved in activating six clotting factors involved in the control of coagulation. In addition, vitamin K is also a cofactor for the matrix Gla-proteins and osteocalcin essential in the regulation of matrix and bone calcification.	As a cofactor for γ-glutamyl carboxylases, the mechanism involves first the generation of a strong base by oxygenation of vitamin K. Vitamin K is converted to vitamin K epoxide, the intermediate of which is a highly basic alkoxide. This intermediate is thought to abstract hydrogen from targeted glutamyl γ-carbons to facilitate CO_2 addition to form glutamyl residues.	Vitamin K—dependent hemorrhagic disease.

Water- soluble vitamins

Vitamin (*other designations) and* ***precursors***	***Adult daily maintenance requirements***	***Active forms*** *cofactor, co-substrate, co-enzyme forms*	***Physiological functions***	***Biochemical functions***	***Associated deficiency disease***
Ascorbic acid (Vitamin C)	**F:** 75 mg **M:** 90 mg **UL:** 2000 mg	Ascorbic acid	Ascorbic acid acts as an antioxidant and as a cofactor/co-substrate for hydroxylases essential for growth, wound healing, and collagen formation.	Ascorbic acid is an electron donor for hydroxylase enzymes. Ascorbic acid functions to maintain iron in the Fe2+ state to facilitate its interaction with O_2. Peptidylglycine alpha-amidating monooxygenase (PGAMOx) also requires ascorbate. PGAMOx amidates peptide hormones to increase stability and activity. At a chemical level, ascorbate maintains vitamin E in a reduced state and contributes to glutathione reduction-related pathways.	Scurvy

Table 1 Vitamins – active forms, primary functions, and requirements.—cont'd

Water- soluble vitamins

Vitamin *(other designations) and* ***precursors***	***Adult daily maintenance requirements***	***Active forms*** *cofactor, co-substrate, co-enzyme forms*	***Physiological functions***	***Biochemical functions***	***Associated deficiency disease***
Niacin (nicotinamide, nicotinic acid, vitamin B_3, nicotinamide riboside) **Tryptophan:** Tryptophane → → →NAD→ → → *nicotinamide*	**M:** 16 mg **W:** 14 mg **UL**: 35 mg	Niacinamide is a component of the coenzymes nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP) and their derivatives.	Niacin via NAD, NADP, or their derivatives aids in converting carbohydrates into glucose and the metabolism of fats and proteins. Deficiency signs result in thick, light-sensitive scaly pigmented skin lesions, diarrhea, fatigue, depression, disorientation, memory loss.	Nicotinamide adenine dinucleotide (NAD+) is essential to ATP generation. In addition, NAD serves as a substrate for mono and poly (ADP-ribose) polymerase pathways necessary for DNA repair. NAD + serves as a substrate for sirtuin-related reactions. NAD+/NADH catalyzes two-electron transfers. First, a hydride is transferred from a substrate to NAD+, with the electrons flowing to the positively charged nitrogen of NAD+.	Pellagra
Riboflavin (Vitamin B_2)	**M:** 1.3 mg **W:** 1.1 mg **UL:** Not known	Riboflavin is a component of flavin mononucleotide (FMN, or riboflavin-5′-phosphate) and flavin adenine dinucleotide (FAD).	Important to neonatal growth and development, the function of redox-respiratory enzymes, and maintenance of epithelial tissues, reproduction	One-electron and one proton transfers in redox reactions	Ariboflavinosis
Pantothenic acid (Vitamin B_5)	**AI:** 5 mg	Pantothenic acid is a component of coenzyme A (CoASH) and acyl carrier protein (ACP).	CoASH is essential in numerous oxidative and biosynthetic reactions in intermediary metabolism. ACP is a part of the fatty acid biosynthesis cycle.	Thioesters are obligatory intermediates in key processes in which ATP is used or regenerated. Thioesters are involved in the synthesis of all esters, including those found in complex lipids. Thioesters often act as "high-energy" components for their associated reactions.	Pantothenic acid deficiency. A nutritional deficiency is rare. When observed, it is likely the result of a CoASH-related polymorphisms.
Pyridoxine (Vitamin B6)	**M:** 1.3–1.7 mg **W:** 1.3–1.5 mg **UL:** 100 mg	Pyridoxal 5′ phosphate (PLP) and pyridoxamine 5′ phosphate (PMP).	Pyridoxal 5′-phosphate is a coenzyme used in more than 140 enzyme reactions essential in amino acid, glucose, and lipid metabolism. PLP also aids in the synthesis of hemoglobin as a coenzyme for aminolevulinic acid synthase.	PLP catalyzes transaminase reactions and is a required coenzyme for glycogen phosphorylase (GP). Transamination occurs in two steps: (1) the transfer of an amino group to PLP to form pyridoxamine phosphate (PMP) and (2) oxidative deamination of PMP and amination of the keto-acid. For GP (catalyzes the sequential phosphorolysis of glycogen to release glucose-1-phosphate), the phosphate group of PLP acts as an acid catalyst for the reaction. In addition to essential roles in carbohydrate, protein, and amino acid metabolism, PLP is an essential cofactor for the biosynthesis of sphingolipids.	Vitamin B6 deficiency

(Continued)

Table 1 Vitamins – active forms, primary functions, and requirements.—cont'd

Water- soluble vitamins

Vitamin *(other designations) and* ***precursors***	***Adult daily maintenance requirements***	***Active forms*** *cofactor, co-substrate, co-enzyme forms*	***Physiological functions***	***Biochemical functions***	***Associated deficiency disease***
Biotin (Vitamin H or B7)	**AI:** ~30 μg	Peptidyl biocytin	Biotin acts as a cofactor for biotin-dependent carboxylases that play critical roles in the intermediate metabolism of gluconeogenesis, fatty acid synthesis, and amino acid catabolism.	Biotin catalyzes the transfer of CO_2 groups to various target macromolecules. Biotin comprises a ureido ring, a tetrahydrothiophene ring, and a valeric acid side chain. It is the ureido ring that functions as the CO_2 carrier. The overall reaction of biotin consists of two steps. In the first step, carboxy-phosphate is generated from bicarbonate with MgATP as an energy source. In the second step, carboxy-biotin results from the nucleophilic attack of the carboxy-phosphate intermediate's carboxyl group. The enzyme-biotin CO_2 is then transferred to an appropriate receptor.	Biotinidase deficiency, a rare inherited disorder, is the most common cause of biotin deficiency. Biotinidase. Biotinidase catalyzes the conversion of biocytin to biotin.
Folic acid, Folacin (Vitamin M. Pteroylglutamic acid; Vitamin B9)	**M&W:** 400 μg **UL:** 1000 μg	Folic acid is a collective term for pteroylglutamic acids and its oligo-glutamic acid conjugates.	Reduced folic acid (THFA) catalyzes one-carbon transfer reactions essential in amino acid metabolism and in purine and pyrimidine modifications critical to DNA and RNA production. Folate derivatives participate inosine phosphate biosynthesis. Methyl-THFA catalyzes vitamin B12 to methyl-B12 via the methionine cycle and regulates the methylation of homocysteine to methionine. THFA is also essential for S-adenosylmethionine generation.	THFA serves as a one-carbon carrier bound at positions N5, N10, or both on the folate pteridine ring. The one-carbon units come from ser, gly, his, or trp. The one-carbon units can exist in one of three oxidation states, equivalent to methanol, formaldehyde, or formate. The folate pteridine ring is also capable of redox reactions. For example, in thymidylate synthase, the folate pteridine ring undergoes a two-electron oxidation. Reduction is required to recycle dihydrofolate back to THF.	Macrocytic-megaloblastic anemia. Folate—related reproductive failure

Table 1 Vitamins – active forms, primary functions, and requirements.—cont'd

Water- soluble vitamins					
Vitamin** (other designations) and **precursors	***Adult daily maintenance requirements***	***Active forms** cofactor, co-substrate, co-enzyme forms*	***Physiological functions***	***Biochemical functions***	***Associated deficiency disease***
Cobalamin (Vitamin B12)	**M&W: 2.5** μg **UL:** Not established	Methylcobalamin and 5-deoxyadenosylcobalamin; Hydroxycobalamin and cyanocobalamin become active after conversion to methylcobalamin or 5-deoxyadenosylcobalamin.	Enzymes that require B12 are methyltransferases (use methylcobalamin as the cofactor) and isomerases (use deoxyadenosyl-cobalamin as the cofactor). Methylmalonyl coenzyme A mutase (MUT) is an isomerase that aids in the conversion of L-methylmalonyl-CoA to succinyl-CoA. Methylmalonyl-CoA results from the metabolism of odd-chain fatty acids or cholesterol sidechains. MUT is also necessary for myelin synthesis and the catabolic degradation of isoleucine, methionine, threonine, and valine.	The cobalt center in the corrin ring of cobalamin can catalyze both one- and two-electron reductive processes. The ability to shuttle between the +1, +2, and +3 oxidation states is responsible for the versatile chemistry of vitamin B12. Such versatility allows cobalamin to serve as an electron donor for the deoxyadenosyl radical form of the vitamin. Four of the six coordination sites are provided by the corrin ring and a fifth by adimethyl-benzimidazole group. The remaining coordination site allows for additions, such as a cyano group (utilized for commercial forms of the vitamin), a hydroxyl group, a methyl group, or the 5′-deoxyadenosyl group. Cyanocobalamin and hydroxocobalamin are converted to bioactive forms by enzymatically replacing the cyano or hydroxyl groups. The cobaltic ion is a strong Lewis acid. Vitamin B12 reactions resemble Grignard reactions. The reactions proceed through carbanion intermediates stabilized by the cobaltic ion.	Cobalamin deficiency; hypo-cobalaminemia, B_{12} deficiency anemia

Moreover, each water-soluble vitamin uses specific carriers associated with enterocytes for their initial absorption. Sodium-dependent multivitamin transporters have been identified (Said, 2013). Water-soluble vitamins, such as folic acid, vitamin B12, riboflavin, biotin, pantothenic acid, and thiamin, utilize specific transport systems (**Fig. 1**). From a disease or nutritional status perspective, it is important to know regions of the intestine in which a given set of vitamin carriers resides, along with biochemical and physiological features of given receptors. Regional inflammation in the intestine may promote specific vitamin deficiencies. For example, B12 receptors are found in the ileum and are compromised by Crohn's disease, which affects the ileal region. Excessive alcohol exposure is also associated with perturbations in thiamin and folic acid status. Alcohol exposure promotes lesions in the jejunum and upper ileal regions, primary sites for thiamin and folic acid absorption.

Concerning the biochemical and physiological features of given receptors, transport efficiencies may also vary due to polymorphisms or the presence of vitamin antagonists that reduce transport efficiencies. Moreover, it is now appreciated that the microbiome can be an important source of indispensable vitamins. Recent studies demonstrate that microbiota-generated water-soluble vitamins are of nutritional value and their uptake by the colon is by receptor-related mechanisms. Although the microbiota was considered insignificant because of the belief that microbiota absorptive epithelia could not carry out effective vitamin transport, it is now appreciated that human colonocytes engage in carrier-mediated transport for many, if not all, of the water-soluble vitamins.

Whole-body, organ, and cellular transport

A typical strategy for delivering fat-soluble vitamins to targeted organs and cells is transport by lymphatic chylomicrons. The lipid bilayer structure allows uncharged hydrophobic molecules, such as lipids, to pass through cell membranes by diffusion driven by differences in the concentration gradient. In addition, it is now apparent that for specific lipids, such as cholesterol and fat-soluble vitamins, various transport systems are involved. For example, cholesterol membrane transporters (e.g., scavenger receptor class B

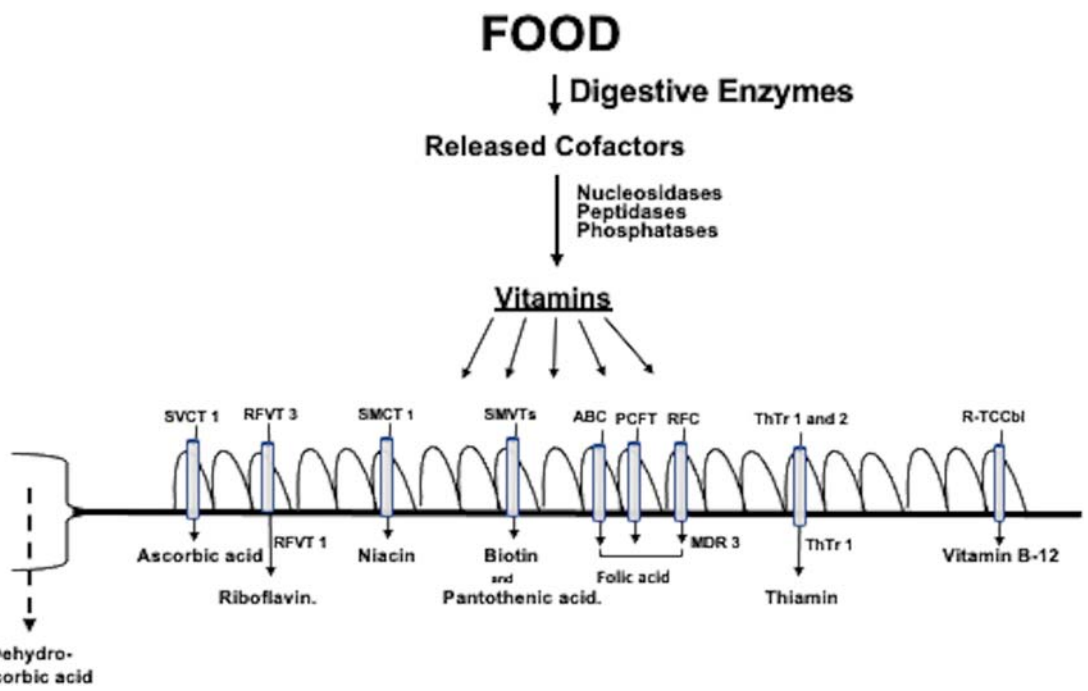


Fig. 1 Water-soluble vitamin absorption. Vitamins are first released from associated proteins, peptides, and cofactor-related moieties by the action of intestinal and pancreatic enzymes. Vitamin absorption occurs when sufficiently solubilized to interact with specific enterocyte carriers and transporters. In addition, a facilitated mechanism mediates the transport of dehydroascorbic acid via glucose transporters. Other water-soluble vitamins utilize specific receptors and transporters Abbreviations: SVCT1 (sodium-ascorbate co-transporters); RFVT3 (riboflavin transporters, e.g., RFVT/SLC52A); SMCT1 (sodium-coupled monocarboxylate transporter 1); SMVTs (sodium dependent multivitamin transporter); ABC (ATP-binding cassette transporters); PCFT (proton-coupled folate transporter); RFT (reduced-folate transporter), THTR1 and 2 (high-affinity/low-capacity thiamine transporters); R-TCCbl (Transcobalamin Receptor).

type I) and members of the ATP-binding cassette transporter family (e.g., ABCA1 and ABCG1) are involved in vitamin E and K transport. Next, cells accumulate lipid substances and package them into small lipid droplets. A current view is such lipid droplets within lipid-processing cells have multiple cellular functions, such as distributing lipids to various membrane-bound organelles within the cell for further processing.

The transport of water-soluble vitamins in blood to the liver and other organs can involve albumin or globular fractions due to nonspecific ionic or hydrophobic interactions. Entry onto cells, however, requires specific cellular membrane transporters. Without membrane transporters, polar substances do not diffuse efficiently across cell membranes. The liver is a principal organ responsible for converting vitamins to active forms. The complexity of the process is illustrated in **Fig. 2**, depicting the relationships between tryptophan and niacin and their conversion to niacin's various cofactor and co-substrate-related derivatives.

Moreover, once activated, each of the water-soluble vitamins is used directly or shuttled to highly regulated pathways that exist for transport to other cells or subsequent modifications to facilitate disposal or turnover. Much of the cellular vitamin-derived cofactor pool is also bound to transporters designed for cellular efflux and eventual delivery to other cells. For the fat-soluble vitamins, liver-derived lipoprotein particles might be used (e.g., for carotenoid pigments and retinyl esters) or specific protein carriers (e.g., α-tocopherol transfer protein). In addition, if sufficiently polar (e.g., retinoic acid derivatives), more direct transfer into blood may occur.

Turnover

All processes related to the whole body and cellular vitamin turnover occur with a high degree of metabolic control. When vitamins are in excess, they are either excreted or sequestered and compartmentalized to protect cells from toxicities associated with nonspecific reactions. Excess water-soluble vitamins are excreted usually in urine, either in their cofactor forms or as derived metabolites. Their rates of excretion often reflect intake. Although recycling via enterohepatic circulation may occur, fat-soluble vitamins and vitamin B12 are excreted in bile into the intestine.

Vitamin requirements

As catalysts and signaling molecules for energy-related processes, the dietary requirements for most vitamins are related to energy needs and metabolic rate. Consequently, vitamin requirements are similar in most animals when expressed relative to metabolic body size or per unit of food energy consumed. A common denominator for many aspects of biological research is that diet

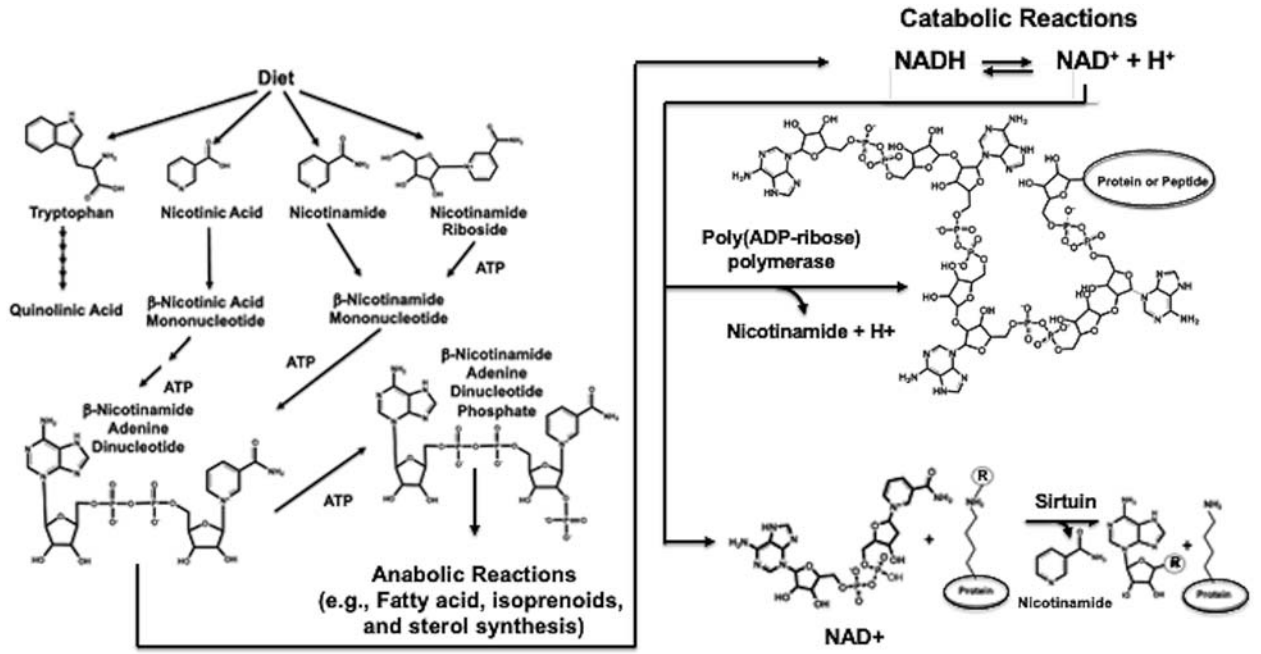


Fig. 2 Niacin metabolism. Niacin is derived from dietary sources and the cellular catabolism of tryptophan. As nicotinamide, niacin is the catalytic moiety for nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP), which serve in multiple cellular functions. In addition, oxidized NAD (NAD+) is utilized as a co-substrate in polyribosylation and sirtuin-related protein deacetylase reactions. Both types of reactions are important in regulating a range of cellular functions. Moreover, NAD + can be converted to cyclic ADP ribose (cADPR) or adenosine diphosphate ribose (ADPR). ADPR and cADPR also act as cellular signaling molecules.

composition and nutrition needs are similar even for dissimilar species. For example, animals have evolved using many of the same genetic strategies. Mice and humans each have about 30,000 genes, yet only about one in a hundred are unique to either organism. As a result, animals have similar vitamin requirements when expressed relative to the food energy consumed.

Moreover, the magnitude of such requirements may be estimated based on metabolic size or the resting or basal metabolic rate for given species using concepts that evolved from the early work of Kleiber and others (Kleiber, 1961; Rucker and Steinberg, 2002). Kleiber's law, now considered one of the few quantitative laws in biology, states that an animal's basal metabolic rate scales to the 3/4 power of the animal's mass. Simply stated, the relationship can be indicated as $q_0 \sim kM^{3/4}$, where q_0 represents the animal's resting or basal metabolic rate, M, the animal's mass in Kg, and k, a constant if needed. It is a concept that has been validated in numerous empirical and theoretical studies.

Given that vitamin requirements are energy-related and universal, an appreciation for biological quarter-power relationships opens the door to a broader range of information translatable to human health and a better understanding of nutritional risks. For example, as it relates to nutritional requirements, one may ask - how is it possible to extrapolate from a mouse of ~30 g to a human with a mass of ~70 Kg. A comparison or extrapolation based directly on body mass suggests that human requirements may be 2300-fold greater than those for mice. The use of Kleiber's law, however, indicates estimations of basal metabolic requirements only differ by ~350-fold metabolic body size (e.g., $q_{Human} \sim k(70\ Kg)^{3/4}$ compared to $q_{Mouse} \sim k(0.03\ Kg)^{3/4}$). This ratio is similar to empirical values for basal metabolism or resting energy expenditure (REE). For example, a small adult mouse's daily REE is ~ 3.5–4.5 Kcals (14–18 KJs). For a 70 Kg human, the values for REE are ~1400 Kcals (women) to ~1600 (men) Kcals or 5600–6400 KJs per day. The division of a REE_{Human} of 6000 ± 400 KJ by 16 ± 2 KJs is also ~350-fold.

The same relationships also hold for drug and toxicological assessments. For example, if a substance causes harm in a 30-g mouse at 1 mg/kg of body weight, one cannot assume that harm in a human may require 2000 times or greater that amount. Instead, concerns for humans are more likely to be encountered at dosages of 350 mg, approximately seven times less than those estimated on a per Kg body weight basis. As a final point, the Klieber relationships also have applications for nutrients that can be synthesized in some animals but are essential to others (Rucker and Steinberg, 2002). For example, in animals that produce ascorbic acid, the daily rates for its production are about the same in mice and rats as the daily dietary ascorbic acid requirements for animals that cannot produce it, such as humans and guinea pigs (Rucker and Steinberg, 2002).

The factors most important in dictating the relative need for specific vitamins include:

- Chemical stability.
- The number and types of catalytic events involving the vitamin.
- Its interactions with associated enzymes or regulatory proteins.
- The presence or absence of pathways for its synthesis.

In this regard, vitamins required daily in millimolar amounts are usually less chemically stable, function in oxidative environments, or are involved in numerous reactions. In addition, such vitamins often exist in tissues as dissociable cofactors. In contrast, vitamins required daily in micromolar amounts are usually engaged in fewer reactions. They are also more likely to be bonded covalently to the proteins they serve as cofactors or signaling molecules. Moreover, if their function is cell signaling, as few as 1000 molecules or less may initiate an event (Rucker and Steinberg, 2002).

Primary deficiencies often occur when monotonous diets or limited combinations of foods are consumed. Examples include:

- Vitamin C deficiency (Scurvy) is due to a diet absent of fruits and certain vegetables.
- Thiamin deficiency (Beriberi) is due to diets based mainly on polished or refined rice.
- Niacin deficiency (Pellagra) is due to the consumption of a corn-based diet with low levels of niacin and tryptophan, the cellular precursor for the niacin (cf., Niacin).

Vitamin deficiencies can also arise through various mechanisms, including compromised stability due to unfavorable chemical conditions (pH, heat, presence of oxidants), poor availability or bioavailability, interactions with other competing substances, and genetic influences (e.g., polymorphisms). For example, thiamin is one most unstable of the B vitamins. Baking, pasteurization, or boiling food can reduce the thiamin content by 50%. Vitamin C is also temperature-sensitive and easily degraded during cooking.

Polymorphisms of genes for proteins involved in vitamin absorption and metabolism can also influence requirements (He et al., 2017). A polymorphism is the occurrence of two or more different morphs or forms of a gene (i.e., alternative phenotypes that allow two or more possibilities of a genetic trait). Genetic polymorphisms can affect the metabolism of vitamins (e.g., transport, conversion to active forms, turnover), which in turn may affect the expression of genes or activity of proteins that require the vitamin. Knowledge of specific polymorphisms may often explain why some individuals have different vitamin requirements. Some typical examples are offered in the next section.

Lastly, there are physiological factors and competitive interactions that influence requirements. Among these are potentially higher needs during growth, pregnancy, and lactation. In addition, inflammation and fever associated with various disease states can impact global metabolism and the need for given vitamins. Vitamin absorption may also be compromised with aging. For example, Vitamin B12 deficiency is more common among the elderly because of a higher prevalence of atrophic gastritis. Atrophic gastritis is an autoimmune disorder, which, if severe, can destroy stomach parietal cells, which produce factors essential to the absorption of vitamin B12. Examples of competitive interactions include:

1. Warfarin (coumadin) interferes with vitamin K regeneration.
2. Phenothiazines (e.g., chlorpromazine) interfere with riboflavin absorption and renal reabsorption.
3. Medications (e.g., colchicine, metformin, various antibiotics) may decrease the absorption of vitamin B12.
4. Methotrexate and sulfacetamide interfere with the absorption of folic acid and its cellular transport.

In addition, food preservatives, such as sulfites, can oxidatively destroy thiamin and ascorbic acid. In addition, plant thiamin antagonists, such as caffeic acid, chlorogenic acid, and tannic acid or agents that promote intestinal inflammation (e.g., alcohol), may interfere with thiamin absorption.

Functions and mechanisms: fat-soluble vitamins

Vitamin A

Vitamin A (retinol, retinal, and retinoic acid) is essential to vision, cell division, growth, reproduction, and immunity. Relationships between vitamin A and its dietary and metabolic precursors are shown in **Fig. 3** (e.g., β-carotenoids, tetraterpenoids produced from eight isoprene moieties). Carotenoids are yellow, orange, and red pigments found in plants, algae, bacteria, and fungi. In plants and algae, carotenoids absorb light and facilitate the transformation of light into chemical energy for photosynthesis. Carotenoids also serve as primary plant antioxidants under a severe excess of radiant energy and as chloroplast antioxidants in drought and heat stress.

Many of the biochemical mechanisms that serve as underpinnings for such processes are also used in animals, particularly mechanisms essential to vision, cell signaling, and antioxidant defense. The mechanisms in part involve the isomerization of isoprene moieties in response to light energy. The resulting conformational energy is then transferred to cell-related energy or regulatory processes.

One of the earliest symptoms of vitamin A deficiency is night blindness (nyctalopia, an inability to see well in poor light). Nyctalopia can occur when the retinal form of vitamin A is insufficient. Retinal is the photosensitive cofactor for the protein opsin. When retinal and opsin are combined, rhodopsin, the active holo-form, is produced. Rhodopsin (found in retinal rod cells) mediates the ability to sense light.

In tissues, including the eye, retinal is formed from the direct oxidation of retinol or β-carotene by the irreversible oxidative cleavage (e.g., β-carotene + O_2 → 2 retinals). The latter is catalyzed by the action of ocular β-carotene 15,15′-monooxygenase or dioxygenase. In the eye, the active metabolite, 11-cis-retinal, is produced from retinol by the successive action of retinol isomerase and 11-cis-retinol dehydrogenase. 11-cis-retinal next combines with the protein opsin to form rhodopsin, the active visual pigment.

Terpenoids, such as 11-cis-retinal, interact with photons of light and isomerizes to all-trans-retinal. The resulting conformation change triggers a chemical signaling cascade linked to the eye's neural system. For example, upon absorption of a photon of light,

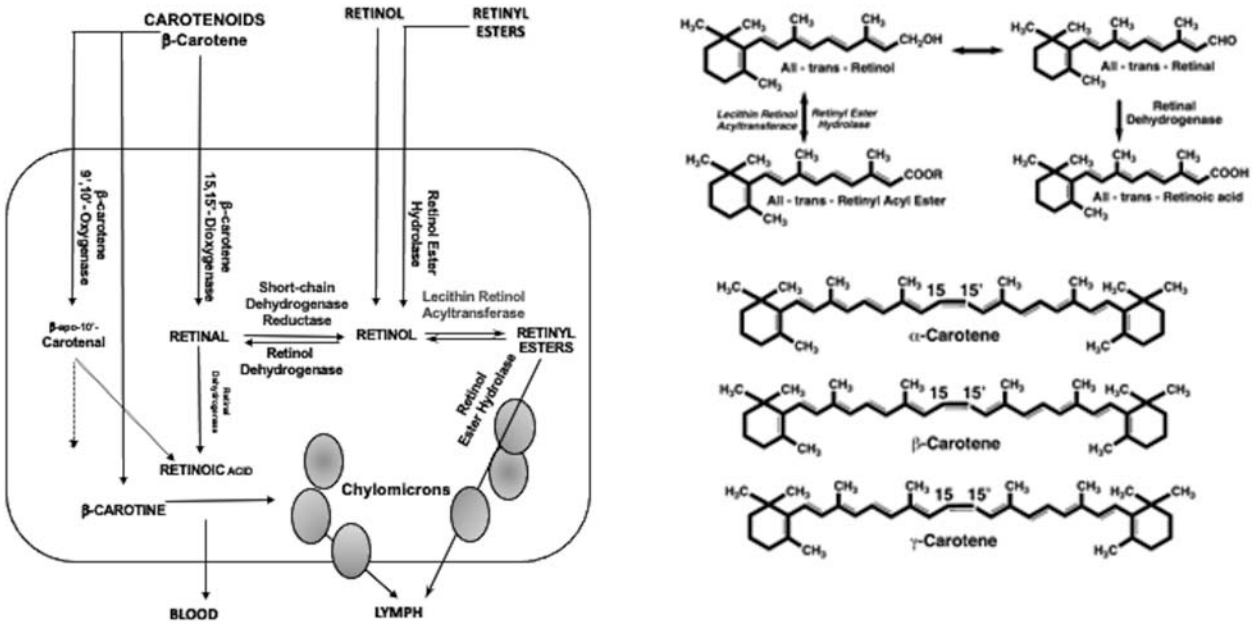


Fig. 3 Vitamin A. Relationships between vitamin A and its dietary and metabolic precursors are illustrated. The active forms of retinol (retinal and retinoic acid) are formed from the direct oxidation of retinol or β-carotene by the irreversible oxidative cleavage (e.g., β-carotene + O_2 → 2 retinals). Vitamin A is sequestered in cells as retinyl esters. Because of differences in structure, carotenoids, such as a- or b-carotene, result in one molecule of retinal following oxidative cleavage.

the covalently bound 11-cis-retinal isomerizes to the all-trans form of retinal, which results in an apo-opsin/rhodopsin-related conformational energy. This conformational energy is transferred to ocular transducin, a guanine nucleotide-binding protein or G-protein (Fig. 4).

G-proteins act as molecular switches inside cells and regulate enzymes, ion channels, and transporter proteins. For example, activation of G-protein coupled pathways leads to hydrolysis of cGMP to GMP. The decrease in intracellular cGMP induces the closure of sodium and calcium channels, leading to hyperpolarization of the rod cells and a change in the neural transmission perceived as vision. Thus, the process allows the transformation of a physical event (photo perception) into a chemical event (ionic

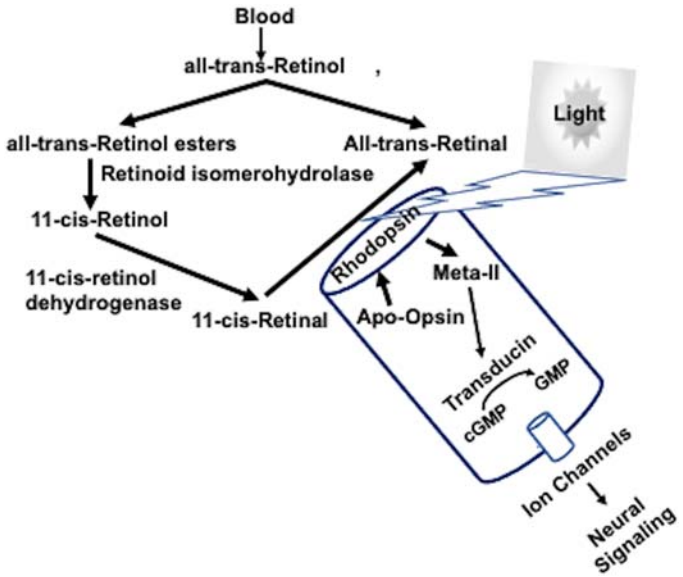


Fig. 4 Vitamin A and vision. Retinoids, such as 11-cis-retinal, interact with apo-opsins in the rod cell of the eye. 11-cis-retinal can capture photons of light, which results in its isomerization to all-trans-retinal. The resulting conformation change transforms apo-opsin into rhodopsin. Next, a chemical signaling cascade is initiated from the generated conformational energy. The cascade activates ocular transducin, a guanine nucleotide-binding protein or G-protein, which modulates ion channels to allow the signal to interact with the eye's neural network.

electrochemical transmission). Moreover, different retinal-opsin complexes, found predominately in the cones of the eye, can absorb photons of different wavelengths of light, allowing the perception of color.

Following photon activation, 11-cis retinal isomerizes to all-trans-retinal and is released from rhodopsin. The all-trans-retinal is immediately reduced to all-trans-retinol and transported back to the retinal pigment epithelium. Although complex, this process prevents inhibitions due to potentially reversible reactions; thus, it allows sensing light energy at high efficiency, even as low as a few photons of visible light.

Vitamin A deficiency also may result in dryness, conjunctivitis of the cornea, and a condition called xerophthalmia. Xerophthalmia is the primary cause of nutrition-related blindness globally, owing to the eye's sensitivity to vitamin A deficiency. For xerophthalmia, the form of vitamin A that is most important is retinoic acid and products derived from it. The following section describes the retinoic acid receptor (RAR) and retinoid X receptor (RXR) cell signaling. RAR/RXR signaling plays essential regulatory roles in inflammation, cell differentiation, and the immune system (Kiss et al., 2013).

The retinoic acid receptor (RAR) and retinoid X receptor (RXR) families belong to the large group of ligand-activated transcription factors (Fig. 5). Each has 3 isotypes designated RARα, -β, -γ or NR1B1–3 and RXRα, -β, -γ or NR2B1–3. As a cell-signaling molecule, retinoic acid can affect transcription through its interactions with retinoic acid receptors. In their native states without retinoic acid, the RARs recruit corepressor proteins to promoter regions, which leads to inhibition of selected gene transcription (Kiss et al., 2013).

The action of RXR isoforms is complex. RXR can heterodimerize with RAR and other nuclear receptors (e.g., constitutive androstane, farnesoid X, liver X, peroxisome proliferator-activated, pregnane X, thyroxine, and vitamin D receptors). When an RXR heterodimerizes with one of the nuclear receptors, nuclear response elements specific for one of the associated nuclear receptors are the target. The heterodimer is active (promotes transcription) when the associated nuclear receptor is bound to its ligand. The binding of agonist ligands results in dissociation of or less interaction with corepressors that inhibit transcription. Concerning the RAR-RXR heterodimer, the RAR ligands are all-trans retinoic acid or 9-cis retinoic acid. The 9-cis retinoic acid, however, is more potent than all-trans retinoic acid and binds with higher affinity. Without all-trans retinoic acid or 9-cis retinoic acid, transcription is inhibited. Accordingly, changes in the 9-cis retinoic acid concentration can switch genes containing response elements specific for RAR-RXR from a transcriptionally silent state to an active state. More than 500 genes are regulated by RAR/RXR signaling; however, far less is directly activated via classical RAR response element pathways. There is also the issue of retinoid derivative turnover or affinity for given RARs. For example, all-trans derivatives of retinol turnover more rapidly than derivatives of 11-cis-retinol.

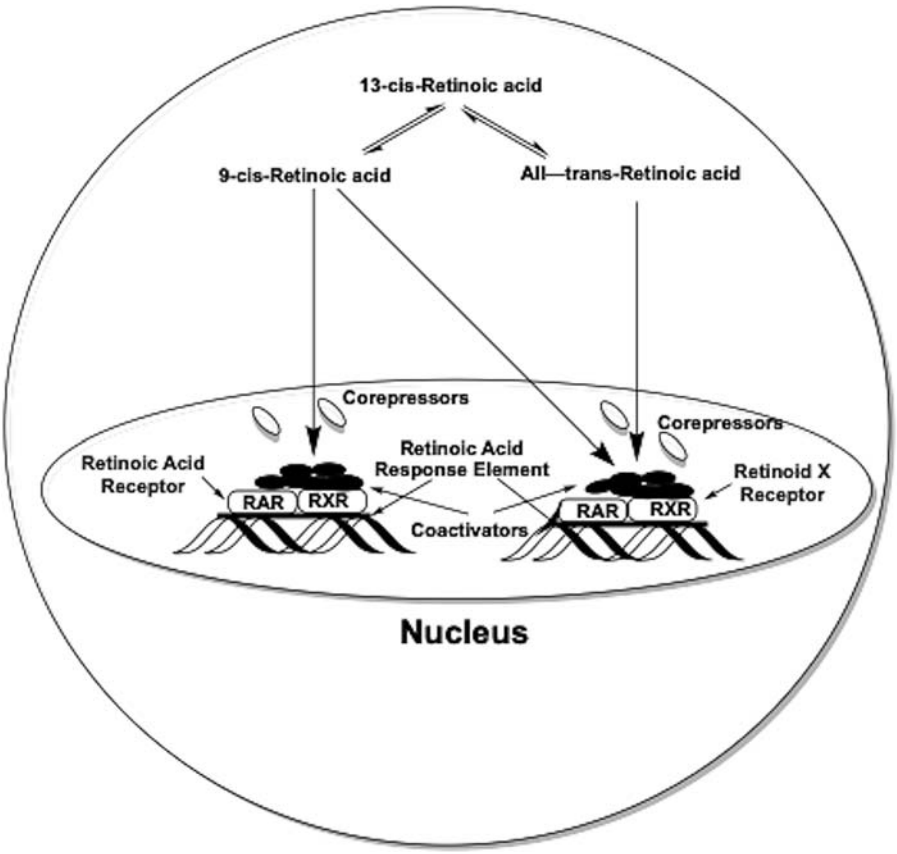


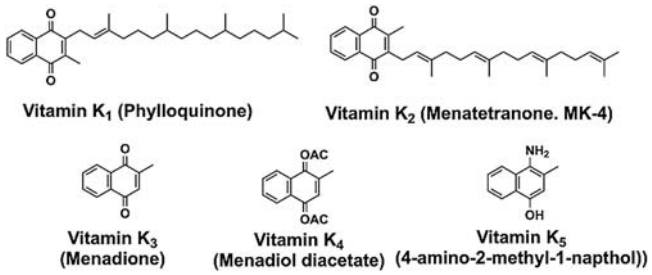
Fig. 5 The retinoic acid receptor (RAR) and retinoid X receptor (RXR) transcription factors. The RXR can heterodimerize with RAR or other nuclear receptors (e.g., the vitamin D receptor). There is transcription when the nuclear receptor (e.g., RAR) is bound to its ligand (9-cis or all-trans retinoic acid) as a heterodimer.

RAR/RXR signaling becomes increasingly important in aging due in part to vitamin-related regulatory functions related to inflammation, immunity, and protection of the neural system. Based on studies in knockout mice, vitamin A-related developmental pathologies are compounded when more than one of the RAR genes is inactivated. Although few congenital structural disabilities have been reported due to vitamin A deficiency, RAR genes are nevertheless essential for organogenesis and early mammalian development. As a final comment, in addition to transcriptional regulation, RARs and RXRs have nongenomic functions. RARs interact with several protein kinases systems whose signaling cascades are initiated by receptors at the cell surface.

Acute hypervitaminosis A has been reported to cause congenital malformations in the central nervous system, heart, thymus, and craniofacial structures. For example, congenital disabilities have been found in babies born to mothers who have been exposed to isotretinoin, a synthetic 13-cis-retinoic acid compound prescribed for severe acne and other medical conditions. Noteworthy, hypervitaminosis A is linked more to animal sources than plant carotenoid enriched sources for food-derived vitamin A sources.

Vitamin K

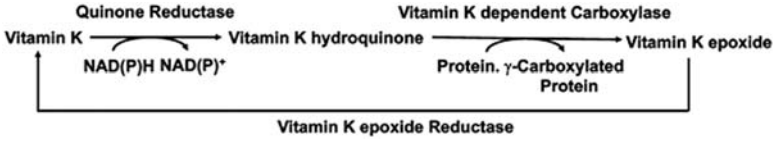
Vitamin K most often refers to one of two structures containing isoprene chains, phylloquinone (vitamin K_1) and menaquinone (vitamin K_2). Phylloquinone is abundant in green leafy vegetables, whereas menaquinone can be produced by bacteria or within animal tissues if there is a source of 1,4-naphthoquinone with a methyl group in the 2-position as a precursor. For example, menaquinone (vitamin K_2) may be formed after alkylation of menadione with isoprene units derived from the cholesterol synthesis pathway. Although not used as a supplement for humans due to safety concerns, menadione is commonly used in poultry, pig feed, and commercial pet foods as a source of vitamin K.



Scheme 1

Functions and mechanisms

Vitamin K serves as a cofactor for γ-carboxylases (designated vitamin K-dependent carboxylases) essential to coagulation, bone formation, and control of matrix calcification. Vitamin K-dependent carboxylases (4.1.1.90, peptidyl-glutamate 4-carboxylases) catalyze the addition of carboxyl groups to the Gla- or γ-glutamyl domains of given proteins. The mechanism involves the generation of a strong base by oxygenation of vitamin K. The carboxylase is a bifunctional enzyme that converts vitamin KH_2 to vitamin K epoxide. The intermediate in the process, a vitamin K alkoxide, is highly basic. This intermediate is thought to abstract hydrogen from targeted glutamyl γ-carbons to form glutamyl residues with carbanion its γ-carbon. Carbanions are carbon atoms containing a negative charge. The negative charge provides nucleophilic properties that facilitate the formation of new carbon-carbon bonds. CO_2 has a positive or carbonium ion character. Consequently, CO_2 reacts easily with the γ-carbanion, which converts the original glutamyl residues to a γ-carboxyglutamyl residue. The reaction does not require ATP but utilizes the energy of vitamin KH_2 oxidation to perform the chemical work required in Gla synthesis.



Scheme 2

Blood coagulation, osteocalcin, and Gla-matrix proteins

Vitamin K serves as a cofactor for the γ-glutamyl carboxylases involved in the activation of six clotting factors: prothrombin, also designated as factor II, factor VII, an extrinsic factor, factors IX and X in the intrinsic pathway and proteins C and S. Proteins C and S aid in preventing inadvertent blood clots and in the control of coagulation.

In each of the factors, there are N-terminal γ-glutamyl domains. Such domains engage in calcium-binding in a cooperative manner. The addition of calcium facilitates interaction with cellular phospholipid membranes at the injury site and activated platelets. For example, for calcium-bound prothrombin, the membrane-bound proteinase, prothrombinase, cleaves the calcium-enriched

N-terminal γ-glutamyl domain to yield thrombin as an active proteinase. Thrombin, in turn, serves as the proteinase catalyzing the cleavage of fibrinogen to fibrin, a near-terminal step in the coagulation cascade.

Vitamin K is also a cofactor for the so-called matrix Gla-proteins and osteocalcin, both members of another family of vitamin K_2-dependent, Gla-containing proteins. The matrix Gla-proteins have a high binding affinity for calcium and can inhibit or control the mineralization in various extracellular matrices (e.g., the matrix of blood vessels).

In bone, matrix Gla-proteins and osteocalcin work together. Moreover, it is noteworthy that although osteocalcin is secreted by osteoblasts and concentrates in bone, it is now recognized as an endocrine-like peptide hormone with multiple functions. These functions extend beyond the control of bone mineral formation in the skeletal system and include glucose homeostasis, exercise capacity, and brain development.

Vitamin K antagonists

Diet-related hemorrhagic disease was first described 100 years ago. Cattle in the Northern USA and Canada were afflicted by unusual, sometimes fatal, bleeding that resulted from minor injuries. Moldy silage made from clover was often implicated. Over the next 20 years, the active agent in clover was identified as coumadin. Coumadin (also marketed as Warfarin) inhibits the action of blood clotting factors by inhibiting the quinone and epoxide reductases essential in the regeneration of reduced vitamin K. The net result is a failure of γ-carboxylation of the Gla-domain-containing coagulation factors.

Vitamin E

Vitamin E comprises a chromanol ring with a phytyl sidechain located at the C2 position of the chromanol ring (Fig. 6). Vitamin E, as a designation, is a group of eight compounds that include four tocopherols and four tocotrienols. Vitamin E is now accepted as a free radical scavenging antioxidant found primarily in cellular and cellular organelle membranes. The α-tocopherol forms of vitamin E function within the glutathione peroxidase pathway and protect cell membranes from oxidation by reacting with lipid radicals produced in lipid peroxidation chain reactions. The oxidized tocopheroxyl products that result are recycled back to their reduced forms through reduction by other antioxidants, such as ascorbate and ubiquinol, the reduced form of CoQ10.

Vitamin E also influences cell signaling. Reactive oxygen species are known to trigger various cell signaling pathways (Lushchak, 2014). Different analogs of vitamin E can induce other cellular events, suggesting that vitamin E's impact goes beyond reactive oxygen scavenging. Some examples of enzymes identified as being modulated by Vitamin E status include the activities of protein kinase B, protein kinase C, protein tyrosine kinases, phospholipases, cyclooxygenases, and various phosphatases. Whether insufficient Vitamin E leads to imbalanced oxidative states that activate ROS-sensitive genes or alter interactions of cellular enzymes that control cell signaling and other independent strategies needs clarification.

All forms of vitamin E can act as ROS scavengers (Fig. 6). The different forms of vitamin E, however, have differing antioxidant potentials. The degree of methylation of the chromanol ring of α-tocopherol influences chromanol resonance and the location and stabilization of radical derivatives. Note that the chromanol rings of β- and γ-tocopherols contain two methyl groups; δ-tocopherol chromanol is methylated in only one position.

Tocopherols and tocotrienols quench electrons by donating a hydrogen ion from their chromanol hydroxyl group. Other reductants (e.g., vitamin C) can subsequently replace the donated hydrogen ion on Vitamin E to restore vitamin E's antioxidant capacity. Because vitamin E regeneration is dependent on a sufficient balance of hydrophilic chemical reductants within cells, the transfers occur at the interface of cell membranes.

Although the need for ROS control for the optimal health of cells is apparent, its role in specific cellular signaling events remains unclear. Moreover, the γ-tocotrienols have anti-inflammatory properties superior to those of α-tocopherol when examined in experimental models to assess inflammatory disease; also, tocotrienols have hydrophobic side chains with three carbon-carbon double

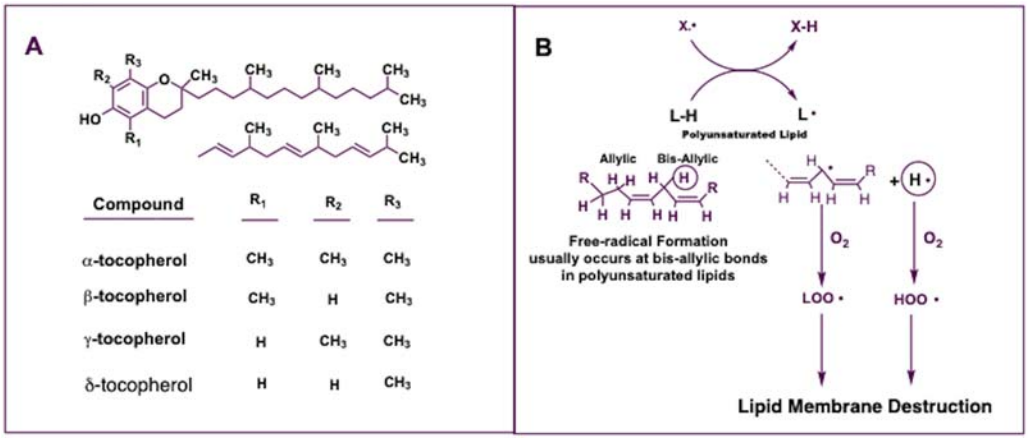
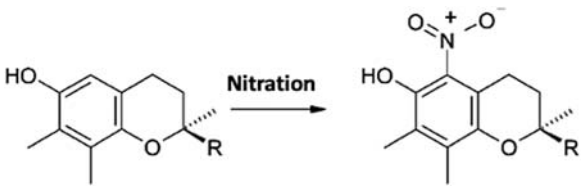


Fig. 6 Vitamin E isomers. The isomeric forms of vitamin E (tocopherol). Tocopherols and tocotrienols quench electrons by donating a hydrogen ion from the chromanol phenolic group, which reacts with and resolves peroxyl and alkoxyl radicals generated during lipid peroxidation.

bonds. This feature allows a more efficient penetration into tissue membranes than all-trans saturated phytyl chains. Furthermore, the presence of γ-tocotrienols in cells is associated with reduced cyclooxygenase- and 5-lipoxygenase catalysis of eicosanoids and suppression of NF-κB and STAT-related pro-inflammatory signaling. In addition, the γ-forms of vitamin E have functions in addition to acting as primarily H ion donors. For example, γ-tocopherol (the form not methylated at position 5) can act as a nucleophile capable of trapping electrophilic mutagens. This is important because nitric oxide (NO), an essential physiological regulator, is easily modified by rapid reaction with superoxide radical ($O\cdot^{-2}$) to yield peroxynitrite. Nitric oxide regulates the dilatation of blood vessels, raising the blood supply and lowering blood pressure. As a neurotransmitter, nitric oxide acts in the nitrergic neurons active on smooth muscle. In contrast, peroxynitrite acts as an oxidant and nitrating agent that can damage molecules, such as DNA and proteins (7).



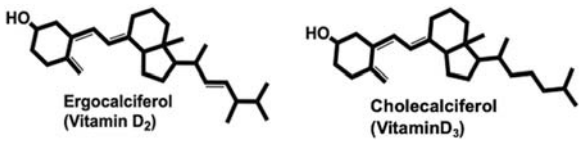
Scheme 3

The importance of vitamin E's role as an antioxidant and to aspects of cell signaling is amplified by observations that for most animals, the apparent daily requirements for vitamin E are related to the amounts of polyunsaturated fatty acids in foods. One gram of diene-fatty-acid equivalents requires the intake of ~0.5 mg *RRR*-α-TOH or its equivalent. The bioavailability of vitamin E is influenced by the amount of dietary lipid, medical conditions, such as Crohn's disease, and lipid-related genetic polymorphisms.

Regarding turnover, the initial step is β-oxidation of vitamin E's phytyl sidechain leading to the water-soluble end-product carboxyethyl-hydroxychroman (CEHC). CEHC is excreted in the urine; bile is another route of excretion. Fecal stools contain numerous vitamin E metabolites in various states of oxidation and phytyl chain-links. An essential protein in intracellular transport of Vitamin E is α-tocopherol transfer protein (α-TTP). Patients with mutations in the α-TTP gene may have low circulating vitamin E in blood and tissues due to an inability to transfer Vitamin E to lipoproteins. Moreover, many tissues (placenta, uterus, central nervous system) express α-TTP.

Vitamin D

Vitamin D is a secosteroid (a steroid with a "broken" B ring). The most important forms of vitamin D are vitamin D3 (cholecalciferol) and vitamin D2 (ergocalciferol). Each form differs only in its side chain structure. Both forms function as prohormones.



Scheme 4

Vitamin D_3 as cholecalciferol is found in foods, such as oily fish, red meat, liver, and egg yolks. It can also be synthesized in skin from 7-dehydrocholesterol, a reaction that requires exposure to UVB radiation. Vitamin D_2 is also found in fish oils and fungi such as mushrooms. Analogous to vitamin D_3, mushrooms convert the vitamin D_2 precursor, ergosterol, into vitamin D_2 when exposed to UV light. Vitamin D_3 does not occur, however, in other plants.

As prohormones, the most active forms of vitamin D_2 or D_3 are their respective 1,25-hydroxy (1,25-OH-) derivatives (1,25-OH-$D_{2\ or}$ D_3). Vitamin D_2 and D_3 are first converted to a 25-OH-derivative, mainly in the liver. Next, the kidney converts the 25-OH-derivative of cholecalciferol to 1,25-dihydroxycholrcalciferol and 25-hydroxyergocalciferol to 1,25-hydroxyergocalciferol (Fig. 7).

Vitamin D controls calcium absorption, immune and neural functions, protecting bone and muscle, apoptotic events, and insulin regulation. The interaction between activated Vitamin D (either as 1,25-OH-D_3 or 1,25-OH-D_2) and its receptors resemble steroid hormone signaling events. The vitamin D receptors (VDR) bind to specific vitamin D response elements (VDREs) to stimulate gene transcription of genes containing VDREs (cf., the previous section related to RAR/RXR signaling). Moreover, some VDREs have been located within introns or between genes, which indicates that there may be multiple ways in which vitamin D and VDR can influence transcription or regulate gene activity.

Regarding calcium absorption and transport, genes specific to these processes are activated by 1,25-OH-D to promote calcium absorption. Some examples of induced genes are transient receptor potential cation channel 6 (TRPV6) and transient receptor potential cation channel 5 (TRPV5). TRPV6 mediates calcium absorption from the intestinal lumen, whereas TRPV5 is involved in the reabsorption of calcium from the renal tubule back into the blood. For example, mutations in these receptors may cause

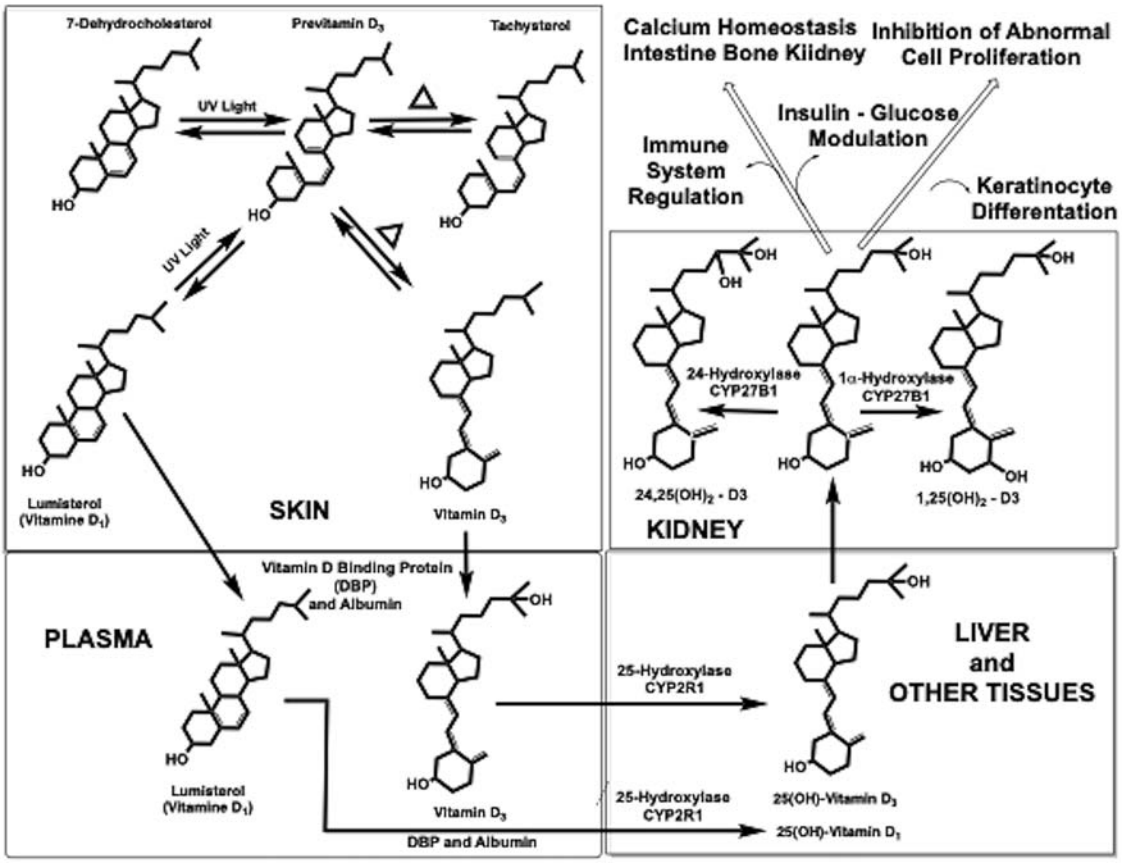


Fig. 7 Precursor and active forms of vitamin D. The active forms of vitamin D act as a steroid hormone to control functions that range from calcium absorption to insulin regulation.

hypocalcemic rickets. Mutations in TRPV6 have been found in infants with transient neonatal hyperparathyroidism. In contrast, mutations that amplify TRPV6 expression are linked to increased kidney stone formation.

Moreover, many genes with VDREs have functions beyond calcium or skeletal metabolism. Some are important in immunity and are upregulated in response to infections. In addition, several of the gene products associated with immunity, such as protein tyrosine phosphatase nonreceptor type 22 (PTPN22) and the cluster of differentiation 226 (CD226), are implicated in autoimmune diseases such as Type 1 diabetes, Crohn's disease, systemic lupus erythematosus, and rheumatoid arthritis. Other genes associated with certain cancers (e.g., colorectal cancer and chronic lymphocytic leukemia) have VDREs and respond to vitamin D exposure. Indeed, the discovery of binding sites for activated VDR has spurred significant interest in identifying roles for Vitamin D beyond bone and calcium metabolism (Fig. 7).

Following absorption, the various forms of Vitamin D are transported to targeted sites utilizing a vitamin D binding protein (VDBP). In this regard, several autoimmune diseases (e.g., multiple sclerosis, rheumatoid arthritis, and systemic lupus erythematosus) are influenced by vitamin D deficiency or polymorphisms in proteins related to vitamin D transport. Vitamin D regulates the immunomodulation of macrophages, dendritic cells, T and B lymphocytes, all of which express vitamin D receptors. Polymorphisms and defects in VDBP and vitamin D receptors account for more than one-half of the vitamin D serum variance associated with the expression of autoimmune and skeletal-related disorders.

In cells, vitamin D is converted to its various hydroxylated forms by enzymes in the cytochromes P450 superfamily of monooxygenases (hydroxylases). These proteins oxidize steroids, fatty acids, and xenobiotics (e.g., hormone synthesis and disposal). In the liver, cytochrome P450 2R1 is the vitamin D 25-hydroxylase responsible for converting vitamin D to 25-OH-D in the liver, followed by 25-OH-D hydroxylation to 1,25-OH-D in the kidney. For humans, the half-life of 25-hydroxy-Vitamin D is 15–25 days, whereas the half-life of 1,25-hydroxy-Vitamin D is only about 15 h. Accordingly, for assessing vitamin D status, the measurement of 25-hydroxy-Vitamin D is normally used.

Functions and mechanisms: water-soluble vitamins

Cofactor, coenzyme, and cell signaling functions

Fig. 8 depicts processes in which water-soluble vitamins play essential roles as co-enzymes and cofactors. For example, ascorbic acid, riboflavin, and niacin serve primarily as redox cofactors, although niacin's functions extend to serving as substrates for compounds

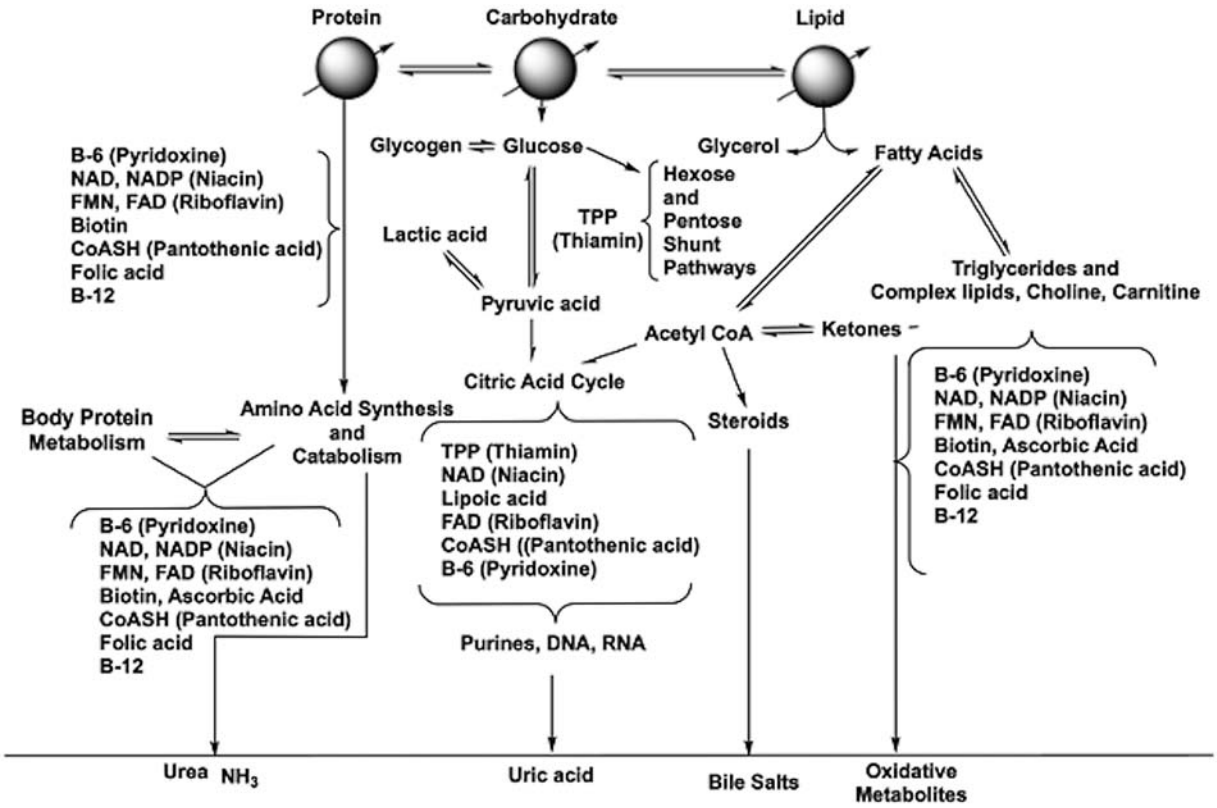


Fig. 8 Metabolic functions of water-soluble vitamins. The water-soluble vitamins in their active forms as cofactors, coenzymes, or cosubstrate are involved in all aspects of energy regulation and the metabolism and interconversions of the macronutrients protein, carbohydrates, and lipids.

important to cell-signaling and DNA repair. Thiamin, pyridoxine (vitamin B6), and pantothenic acid, as a component of coenzyme A, are distinguished because of their importance to carbohydrate and amino acid transport and acyl and acetyl transport. As cofactors, thiamin, pyridoxine, and coenzyme A facilitate the formation of good leaving groups in reactions that are often used in catabolic processes. Good leaving groups are relatively stable weak bases. As nucleophiles, in subsequent transfer steps, good leaving groups give up their pair of electrons to form bonds with molecules that have electrophilic character. Another type of transfer reaction involves those essential to the transfer of single carbons. Single or one-carbon (1C) transfers are found in a series of interlinking metabolic pathways, for which 1C units are generated for the synthetic processes (e.g., DNA, polyamines, amino acids, creatine, and phospholipid synthesis). Vitamins essential for the transfer of single carbons include biotin, folic acid, and vitamin B12 (cobalamin).

Oxidation-reductions: ascorbic acid, riboflavin, and niacin

Redox or reduction-oxidation reactions describe chemical reactions in which atoms or intermediates have their oxidation states changed, although the actual transfer of electrons may not always be obvious, particularly those involving covalent bonds. Redox reactions utilize mechanisms that range from ion hydride transfers, radical hydrogen ion transfers, or one electron plus one proton transfers. As components of coenzymes, riboflavin (as flavin mononucleotide [FMN] and flavin adenine dinucleotide [FAD]), niacin (as nicotinamide adenine dinucleotide [NAD] and nicotinamide adenine dinucleotide phosphate [NADP]), and ascorbic acid have chemical features that make them ideal as intermediate carriers of electrons. Each also acts over a wide range of chemical potentials.

Ascorbic acid

Ascorbic acid (vitamin C) is one of the most important redox cofactors in animal systems. Although most animals make sufficient ascorbic acid, for others, ascorbic acid is essential because the complete pathways for ascorbic acid synthesis are absent (e.g., higher primates, guinea pigs, bats, and some species of birds and fish).

Functions

In mono- and dioxygenases, ascorbic acid maintains metals (specifically Fe and Cu) in their reduced states (Fig. 9, Table 1). Important reactions and processes that require ascorbic acid include: (1) norepinephrine synthesis by functioning in dopamine-β-hydroxylase, (2) hormone activation by functioning in peptidyl glycine-amidating monooxygenase, (3) carnitine biosynthesis by serving as a cofactor for two of the hydroxylation steps in the pathway of carnitine biosynthesis. Ascorbic acid is also essential for

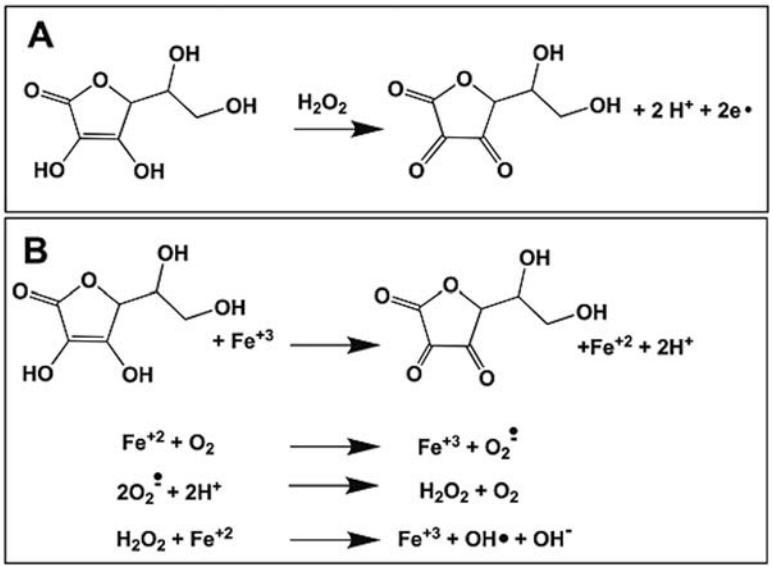
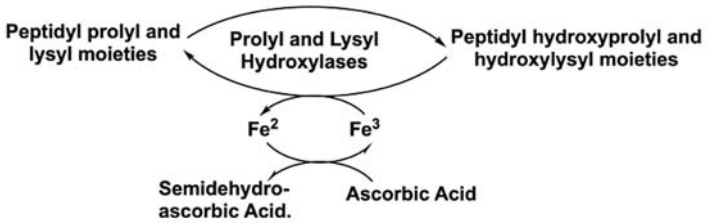


Fig. 9 Hydrogen peroxide and Fenton-related reactions catalyzed by ascorbic acid. Ascorbic acid reacts with hydrogen peroxide and its derivatives resulting in dehydroascorbic acid. Although the response eventually can lead to ROS, the initial reaction with aqueous peroxyl radicals retards their formation. Ascorbic acid also acts by restoring the antioxidant properties of fat-soluble vitamins, such as vitamin E and other antioxidants. Reduced iron and copper can serve as highly destructive Fenton catalysts. As a reductant, it can retard ROS formation; however, in the presence of iron and copper, ascorbic acid can promote their ability to act as a Fenton catalyst in the generation of ROS.

hydroxylations that optimize collagen, elastin, C1q complement, and acetylcholine esterase synthesis and function. The hydroxylations occur in two stages. First, a highly reactive high-valent iron-dioxygen complex is formed. Next, nucleophilic attack at the C2 position of proline generates an intermediate that is converted to hydroxyproline using energy supplied from the cleavage of CO_2 from α-keto-glutarate.



Scheme 5

Ascorbic acid is also associated with protecting lipid, DNA, and proteins from oxidants. For example, when peroxyl radicals are generated in plasma, ascorbic acid is consumed faster than other antioxidants, such as uric acid, bilirubin, and vitamin E. Ascorbic acid is 100 times more reactive than polyunsaturated fatty acids in reacting with peroxyl radicals. Moreover, ascorbic acid can form relatively stable free radical intermediates. This ability significantly delays or prevents free radical-initiated oxidations.

Metabolism

The bioavailability of dietary ascorbic acid is dose-dependent, but absorption (mainly in ileum and jejunum) can be as high as 70–80% at physiological intakes (10–100 mg per day). L-ascorbic acid enters cells via Na^{+-}dependent transport. Dehydroascorbic acid can also enter by facilitated diffusion utilizing the glucose transporters (GLUT 1, 3, and 4), although under physiological conditions, these transporters may play minor roles due to competition from glucose. Some tissues can accumulate as much as 100 times the level of ascorbic acid in blood (e.g., adrenal glands, pituitary, thymus, corpus luteum, and retina). The accumulation of ascorbic acid occurs because of cellular dehydroascorbate reduction systems capable of rapidly generating reduced ascorbic acid. Dehydroascorbic acid is easily destroyed by oxidation. In addition, the chemical reduction of ascorbate minimizes the potential for ascorbic acid radical generation. It is also important to appreciate that ascorbic acid acts as a pro-oxidant under aerobic conditions when metals capable of redox (Fe+2↔Fe+3; Cu+1↔Cu+2) are present. Reduced iron and copper can serve as highly destructive Fenton catalysts (Fig. 9).

During postnatal development, going from a relatively hypoxic to a hyperoxic environment results in rapid adaptive changes. From a quantitative point of view, glutathione (GSH) and ascorbic acid are the most abundant reducing agents in cells. However, glutathione levels are low in the prenatal stages of growth. Furthermore, when GSH levels are decreased, ascorbate levels are also reduced because of the degradation of dehydroascorbic acid. Therefore, in newborns, dietary sources of ascorbate are essential because they spare glutathione owing to its low rate of synthesis.

Biosynthesis of GSH occurs in the cytosol in a tightly regulated manner. Critical determinants of GSH synthesis are the availability of the sulfur amino acid precursor, cysteine, glycine, and glutamate. These amino acids, however, may be limited because of their use in growth and extracellular matrix production processes. The nutritional disease associated with ascorbic acid deficiency is scurvy. The features of scurvy (poor wound healing, osteopenic abnormalities, impaired lipid metabolism, behavioral changes) are easily linked to perturbations resulting from the inability to facilitate hydroxylations and lack of ROS protection.

When cellular levels of ascorbic acid are abnormally elevated, there is increased expression of ascorbic acid decarboxylase activity, which catalyzes the oxidation of ascorbate to CO_2 and C-4 or C-5 fragments. Moreover, ascorbic acid may be converted to 2-O-methyl or 2-sulfate derivatives and excreted. Because mechanisms are in place to homeostatically regulate ascorbic acid, evidence of toxicity, other than gastric upset, is seldom observed. However, decreased histamine production can occur when ascorbic acid is consumed in near gram quantities per 1000 kcal. High dietary amounts can also facilitate gastric nitrosamine reductions.

Riboflavin (vitamin B2)

Originally called vitamin G, riboflavin was renamed vitamin B2 when it was recognized to be part of the yeast B complex. The cofactor forms for riboflavin are flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). Riboflavin is associated with enzymes designated flavoproteins. Riboflavin's planar isoalloxazine ring facilitates the redox transfer reactions catalyzed by FMN and FAD.

Functions

FMN is usually associated with membrane proteins involved in the mitochondrial electron system, whereas FAD is found in membrane-bound and soluble enzymes. Flavins can exist in four redox states: the flavin-N(5) oxide, quinone, semi-quinone, and hydroquinone (**Fig. 10**A and B). FAD and FMN are cofactors usually used in aerobic processes, functioning as cofactors for oxidases. However, FAD also can function in anaerobic environments as cofactor dehydrogenase reactions. Flavins are particularly useful for reactions involving oxygen and sulfur, which usually proceed one electron and one proton at a time. For example, enzymes containing flavin moieties are distinguished because they can transfer hydrogen directly to molecular oxygen (**Fig. 10**C).

FAD exists as an equilibrium mixture of open and folded conformers in aqueous solutions. The planar isoalloxazine ring self-associates with the adenosine the adenine nucleotide. The planar isoalloxazine structure also promotes the resonance and localization of electrons in isoalloxazine ring nitrogens. However, the reduced form of FAD has a boat-like configuration (**Fig. 10**B).

Metabolism

Riboflavin is absorbed by active processes and is transported in blood by a riboflavin transporter. Many flavin-containing proteins are found in the smooth endoplasmic reticulum of cells associated with microsomal phase 1 and 2 enzymes and processes and xenobiotic metabolism. FMN and FAD are tightly bound in given flavoproteins. Consequently, riboflavin turnover is a function of the turnover of enzymes that it serves in cofactor form. In this regard, riboflavin deficiency is more often a function of protein inadequacy than riboflavin deprivation. Riboflavin deficiency is rarely found in isolation but may occur with other water-soluble vitamins and protein deficiencies.

Riboflavin deficiency is classically associated with the so-called oral-ocular-genital syndrome. Signs of riboflavin deficiency include lesions of the oral cavity (cheilitis), inflammation of the tongue (glossitis), and accompanying seborrhea and dermatitis in the genital area. Drugs with chemical structures similar to riboflavin (e.g., chlorpromazine, imipramine, amitriptyline, penicillin, and theophylline) can displace riboflavin from binding proteins that are important to riboflavin and FMN transport. In addition, many flavoproteins and riboflavin transporters are also targets for polymorphic alterations, which can result in the need for higher levels of systemic riboflavin for activation. Further, toxicity is seldom a concern, in part because riboflavin is not very soluble due to isoalloxazine ring self-association and its association with other planer aromatic compounds.

Niacin

Niacin (vitamin B3) is the generic name for nicotinic acid (pyridine-3-carboxylic acid), nicotinamide (pyridine-3-carboxamide), and related derivatives, such as nicotinamide riboside. As nicotinamide, niacin is the catalytic moiety for nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP), which serve in multiple functions (e.g., as cofactors, co-substrates, or signaling molecules) in numerous catabolic and anabolic reactions (cf. **Fig. 2**). For example, as enzymatic cofactors, NAD and NADP catalyze the transfer of hydrogen and electrons from one molecule to another. In addition, oxidized NAD (NAD+) is utilized as a co-substrate in polyribosylation and sirtuin-related protein deacetylase reactions. Moreover, NAD+, when converted to cyclic ADP ribose (cADPR) or adenosine diphosphate ribose (ADPR), acts as signaling molecules (**Fig. 11**).

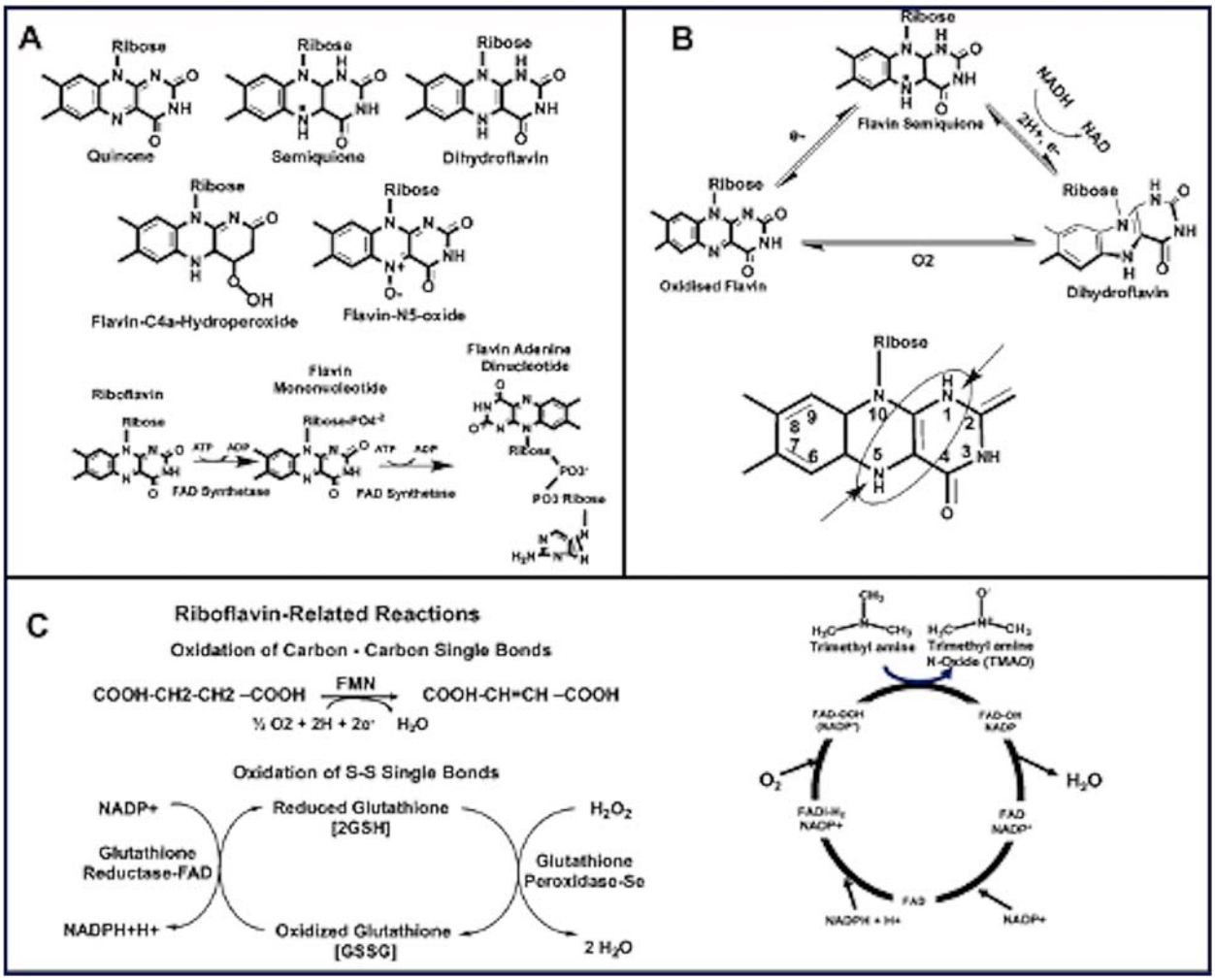


Fig. 10 Riboflavin. Riboflavin's planar isoalloxazine ring facilitates the redox transfer reactions. FMN is usually associated with membrane proteins involved in the mitochondrial electron system, whereas FAD is found in membrane-bound and soluble enzymes. Flavins can exist in various redox states, e.g., quinone, semi-quinone, hydroquinone, and flavin-N(5) oxide. Flavins are particularly useful for reactions involving oxygen and sulfur, which usually proceed one electron and one proton at a time.

Functions

With few exceptions, the redox-related reactions of NAD and NADP involve dehydrogenase (oxidoreductase) enzymes. As a family, NAD and NADP-linked dehydrogenases constitute almost 20% of all classified enzymes and utilize about half of the cellular niacin for their activation. The relative abundance of oxidized NAD(P)+ and the redox balance of NAD + to NADH and NADP + to NADPH mediate the cell's fate in a wide range of metabolic conditions. For example, decreasing the levels of NAD + can influence the rate of cellular apoptosis and, in the whole organism, the rate of malignancies, diabetes, and neurodegeneration. Thus, the ratio of NAD+/NADH and NADP+/NADPH may be viewed as redox couples that are significant determinants of the cellular redox state. NAD participates primarily in energy-yielding catabolic reactions. $NADP^+$ performs less of a catabolic role, but in its reduced form, NADPH is a reductant in anabolic reactions, especially the biosynthesis of fatty acids and other lipids.

NAD and NADP arise by phosphorylation and condensation of niacin with ATP (Fig. 2). NADP also contains a phosphate group on the 3′ position of the pentose nearest to the adenine derived from ATP (catalyzed by NAD kinase). Regarding chemical features, the functional site of NAD^+ and $NADP^+$ is the nicotinamide ring. The oxidized nicotinamide has a quaternary nitrogen with a positive charge. The oxidized ring accepts two electrons and one proton from a substrate reducing the ring and abolishing the positive charge on the nitrogen. A unique feature of the reactions catalyzed by NAD and NADP is hydrogen transfers are in the form of hydride ions, the anion form of hydrogen (H).

Moreover, there is stereospecificity associated with the reactions. The two hydrogen atoms of niacin at the 4-position are prochiral. The transfer of an H: ion to NAD or NADP can be either from the front (A-side, pro-R) or the back (B-side, pro-S) of the niacin moiety, depending on how NAD or NADP fits into its binding domain on the enzyme. There are about the same number of pro-R versus pro-S favoring dehydrogenases that require NAD and NADP in cells. The binding of NAD to dehydrogenases or

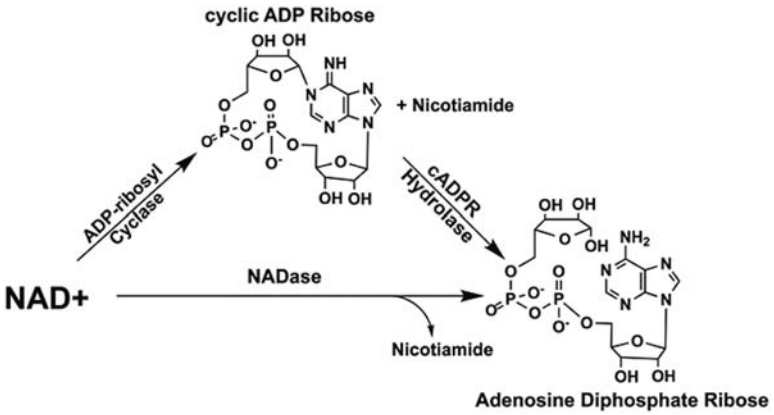
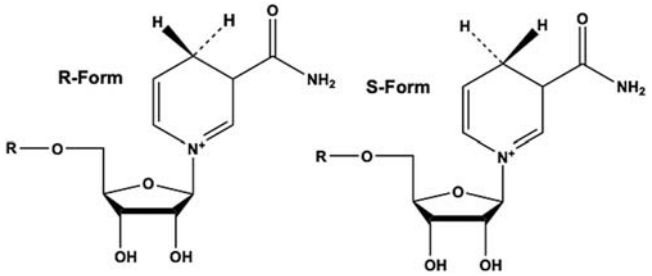


Fig. 11 Cyclic ADP ribose and Adenosine diphosphate ribose. Cyclic ADP-ribose is one of several molecules involved in the lipopolysaccharide-stimulated proliferation of human peripheral blood mononuclear cells. The adenosine diphosphate ribose activates the TRPM2 ion channel (a non-selective calcium-permeable cation channel) and part of the Transient Receptor Potential ion channel superfamily.

reductases is such that hydride transfer involves only one side of the nicotinamide moiety. Consequently, the binding of the non-preferred form is weaker, and the maximum velocity is slower. Examples of dehydrogenase and reductases that prefer the R form are alcohol, lactate, and isocitrate dehydrogenases, including dihydrofolate reductase, whereas glyceraldehyde-3-phosphate, 3–phosphoglycerate, glutamic acid, glucose-6-phosphate dehydrogenases, and glutathione reductase prefer the S form.



Scheme 6

As a substrate, the NAD + form of niacin participates in mono- and polyribosylation reactions and reactions broadly classified as deacetylations and deacetylations. Both types of reactions are important in regulating a range of cellular functions. For example, in the nuclei of cells, polyribosylation of specific histones and transcription factors precedes the normal process of DNA repair. NAD serves to generate groups of ADP-ribose moieties that are transferred to proteins to form long branched chains in reactions designated as poly-ADP-ribosylations. Mono-ADP-ribosylations also occur. As a protein modification, mono-ADP-ribosylations can switch on and off enzymatic activities ranging from glycolysis to oxidative stress.

In addition to mono- and polyribosylation reactions, NAD + also serves as a substrate in deacetylase, desuccinylase, demalonylase, demyristoylase, and depalmitoylase reactions. A family of enzymes designated sirtuins carries out these reactions. A depiction of the deacetylation mechanism is given in Fig. 2. Sirtuins are implicated in cellular processes, such as aging, responses to inflammation and stress, DNA repair, and energy efficiency. Reversible acetylation is a regulatory mechanism in that accessibility to an active site can be controlled by acetylation. Acetylation can also alter protein conformation in ways that influence catalytic parameters. For example, histone acetylation is a critical modification that changes chromatin architecture and regulates gene expression by opening or closing the chromatin structure. Mammals have seven sirtuins (SIRT, silencing information regulator transferases) in the deacyl- and deacetylation of proteins such as histones and metabolic enzymes that involve cellular energy metabolism.

As a component of cADPR, niacin signals the mobilization of $Ca2^{+}$ from the endoplasmic reticulum by activating ryanodine receptors (Fig. 11). In cardiac and skeletal muscle, ryanodine receptors release Ca2+ from intracellular compartments during excitation-contraction coupling. cADPR also activates transient receptor potential melastatin 2 (TRPM2), another calcium ion channel, which responds during inflammation and cellular temperature fluctuations. Ryanodine receptors and TRPM2 also act synergistically during muscle contraction.

Metabolism

A part of the daily need for niacin is produced from tryptophan degradation. Because of a potential internal source, niacin is often designated a "conditional" vitamin. Niacin is needed in amounts corresponding to 3–5 mg per 1000 kcal or 4.2 MJ. The niacin requirements are often expressed as equivalents, where one equivalent corresponds to 1 mg of niacin. For example, the degradation of 50–60 mg of tryptophan to niacin produces about one equivalent or 1 mg.

Factors that lead to niacin deficiency (pellagra) are diets low in tryptophan (most often corn-based) or conditions that reduce the body's ability to convert tryptophan to niacin, such as Hartnup disease or carcinoid syndrome. Pellagra is characterized by dermatitis, diarrhea, and mental disturbance. Pellagra was relatively common throughout the 1800s, mainly where corn was consumed as the primary food energy source and not nixtamalized. Nixtamalization is a process for preparing corn (maize) and other grains, in which the grain is soaked and cooked in an alkaline solution. This type of processing improves the digestion of corn and substantially increases niacin availability.

In pharmacologic doses (gram quantities), niacin-derived compounds are used therapeutically. For example, nicotinic acid is used when increased blood flow or vasodilatation is desirable. In gram quantities per day, nicotinic acid is also an effective lipid-lowering agent (increases HDL). Niacin, as a supplement, also reduces the synthesis of the low-density lipoprotein (LDL)-C precursor and very-low-density lipoprotein (VLDL). In addition, the mobilization of free fatty acids from adipose tissue is decreased. Niacin (as a supplement) can inhibit hormone-sensitive lipase in adipose tissue, reducing triglycerides catabolism to free fatty acids and the transport of free fatty acids to the liver.

There are, however, important caveats to the exposure to high levels of niacin derivatives. Niacin is excreted predominantly as methylated pyridines (e.g., methylation of niacin to 1-methylnicotinamide), requiring methionine as a methyl donor. Hyperhomocysteinemia may occur. The response is also accompanied by decreased plasma concentrations of vitamins B6 and B12.

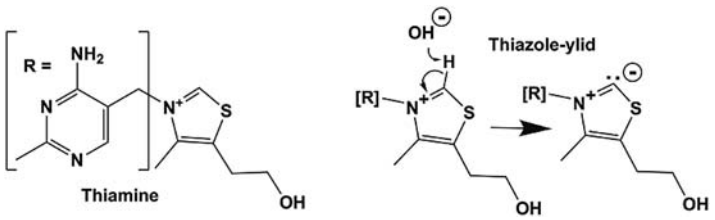
Generation of leaving groups: thiamine, vitamin B6, pantothenic acid

As noted, cofactors function in part by allowing the possibility of different transition states that lower the activation energy of given reactions. With regard to the generation of leaving groups, this usually implies enriching the nucleophilic or electrophilic character of transition state intermediates to better engage in nucleophilic or electrophilic substitutions. The vitamin-derived cofactors, thiamin pyrophosphate (TPP) and pyridoxal-5′-phosphate (PLP), and pantothenic acid as a component of coenzyme A serve as excellent examples of such functions:

Thiamin

Functions

Thiamin comprises a pyrimidine and thiazole moiety linked by a methylene bridge. There are three phosphorylated forms: thiamin monophosphate (TMP), thiamin diphosphate (TDP), and thiamin triphosphate (TTP).



Scheme 7

The thiazole moiety acts as a ylid (a neutral dipolar molecule containing a negatively charged atom attached to a heteroatom with a formal positive charge). The carbanion character of TPP's thiazole ring is retained in transition state intermediates. Examples of enzymes in which TPP is used for this purpose are pyruvate, α-ketoglutarate, or branched-chain amino acid dehydrogenase complexes and 2-hydroxyphytanoyl-CoA lyases. The relationship of the subunits essential for converting pyruvate to acetyl-CoA is shown in **Fig. 12** (Patel et al., 2014). TPP catalyzes the reversible cleavage of the carbon-carbon bond connecting a carbonyl group to an adjacent reactive group (usually a carboxylic acid or an alcohol). The TPP-substrate bond is broken in reverse reactions, reforming the thiamin ylid and substrate carbonyl.

TPP is also a cofactor for transketolases, enzymes associated with the pentose phosphate pathway and photosynthesis. In transketolases, TPP accepts a 2-carbon fragment from a 5-carbon ketose (e.g., D-xylulose-5-P) and then transfers the fragment to a 5-carbon aldose (e.g., D-ribose-5-P) to form a 7-carbon ketose (e.g., sedoheptulose-7-P). The second reaction catalyzed by transketolase in the pentose phosphate pathway involves the transfer of a 2-carbon fragment from D-xylulose-5-P to erythrose-4-phosphate, which results in fructose 6-phosphate and glyceraldehyde-3-P. This reaction connects the pentose phosphate pathway to glycolysis (**Fig. 13**).

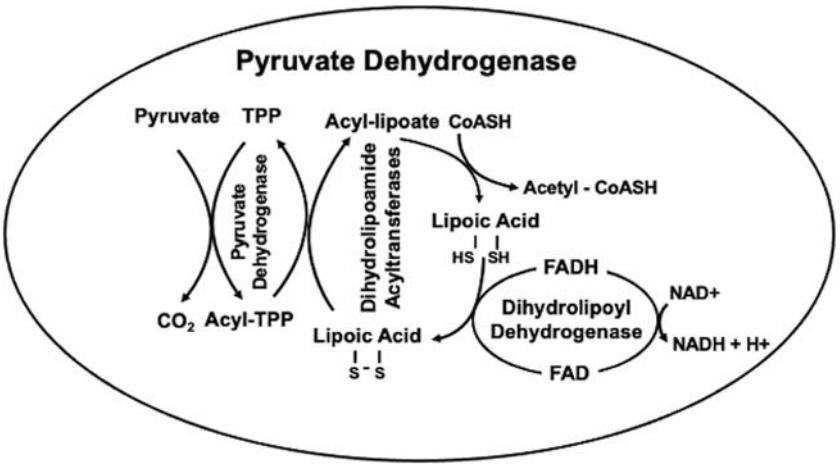


Fig. 12 Pyruvate dehydrogenase complex. The pyruvate dehydrogenase complex (PDHC) comprises three subunits essential for converting pyruvate to acetyl-CoA, CO2, and NADH. The conversion is irreversible. The control of PDHC activity is complex and involves allosteric interactions with its substrates and products, as well as covalent modification by (de)phosphorylation. For example, PDH is inhibited by its products and can be inactivated by phosphorylation by PDH kinases.

Metabolism

The majority of thiamin present in cells is TDP. It has been proposed that TTP has a specific role in nerve excitability, but this is not confirmed. Recent reports, however, suggest that TTP may play a role in cell energy metabolism. The requirement for thiamin is 0.5–1.0 mg per 1000 kcal or 4.2 MJ of diet. Factors that most influence the need for thiamin are exposure to antagonists and thiaminases (enriched in fermented foods). Antagonists used in experimental settings are pyrithiamine and oxythiamine, which inhibit the

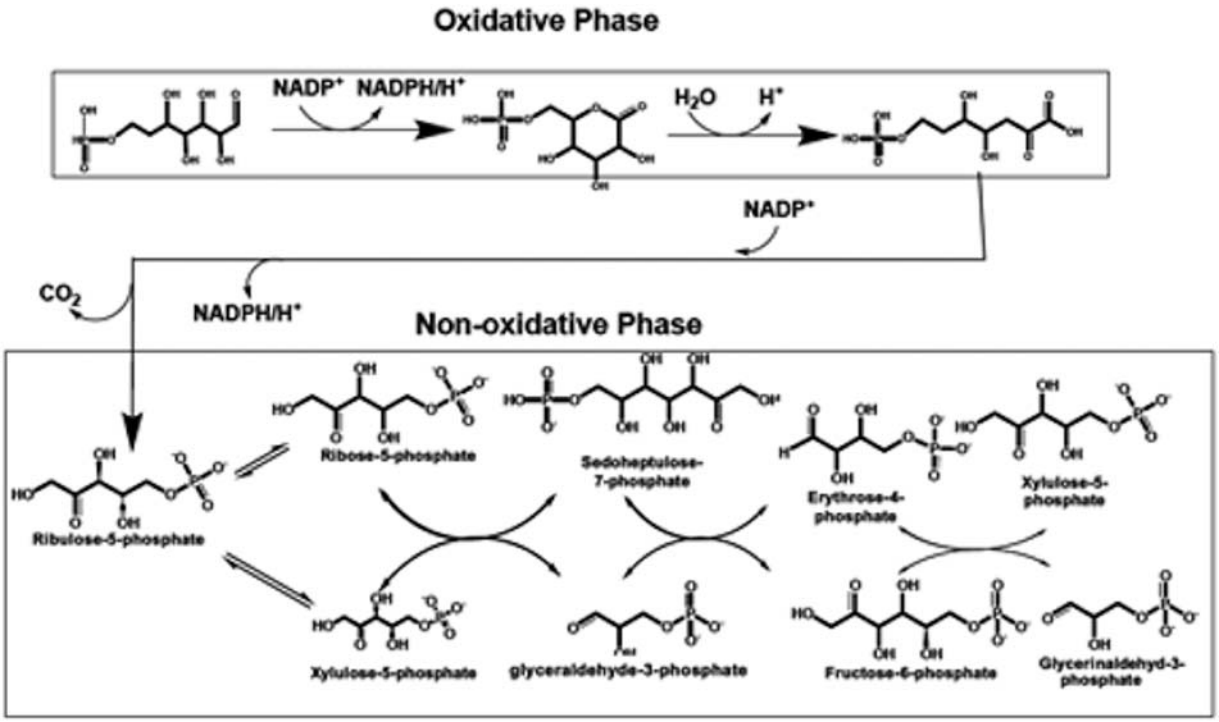


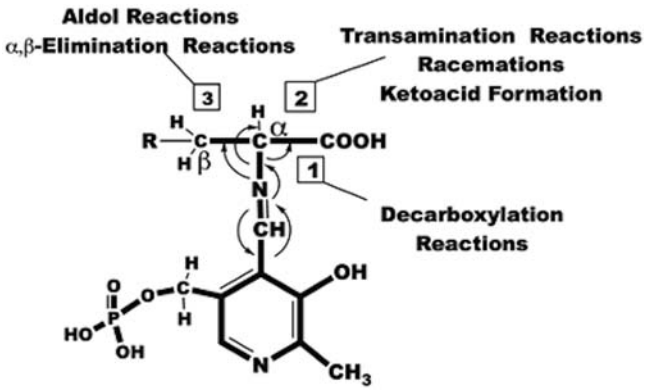
Fig. 13 The Hexose monophosphate shunt and transketolase pathway. The hexose monophosphate shunt (also known as the pentose phosphate pathway) consists of two parts. First, an oxidative pathway component converts glucose-6-phosphate to ribulose-5-phosphate and a non-oxidative component that further generates three, four, and five carbon units. The pathway plays an essential role in generating NADPH reducing equivalents, nucleic acid synthesis, and connecting the pentose phosphate pathway to glycolysis. The role of thiamin diphosphate (TPP) is to serve as a cofactor for two transketolases. In the first reaction, TPP accepts a 2-carbon fragment from D-xylulose-5-P and transfers it to ribulose-5-phosphate to generate sedoheptulose-7-P and glyceraldehyde-3-phosphate. The second reaction involves the transfer of a 2-carbon fragment from D-xylulose-5-phosphate to erythrose-4-phosphate, generating fructose 6-phosphate and glyceraldehyde-3-phosphate.

phosphorylation of thiamin. In addition, the coccidiostat, amprolium, can inhibit thiamin absorption. Alkaline conditions, heat, and oxidants can reduce or chemically modify active forms of thiamin. For example, milling whole grains to remove the bran reduces the thiamin content. Moreover, foods rich in tannins can bind and facilitate the oxidation of thiamin, which leads to reduced availability. Thiamin deficiency is characterized by peripheral neuropathy, including abnormal reflexes, diminished sensations, and cardiac failure due to impaired oxidative metabolism.

Vitamin B6

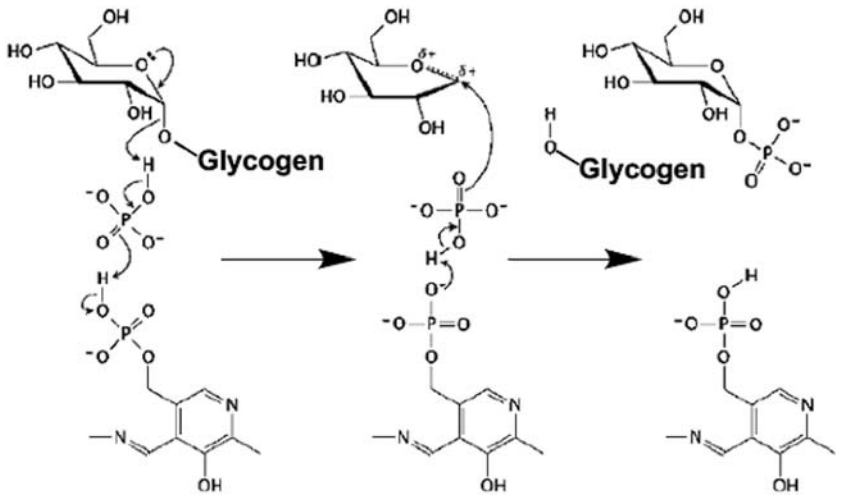
Functions

The main biological functions of vitamin B6 as a cofactor include aminotransferase, decarboxylase, and deaminations reactions. Amino transferases (i.e., transaminases) are essential in the interconversion of amino acids to corresponding α-keto acids to set the stage for eventual oxidative metabolism. Decarboxylations and aldol reactions (electron withdrawal from the α,β -carbons of amino acids) are essential in the generation of amino acid-derived neurotransmitters (e.g., histidine to histamine, tryptophan to serotonin, glutamate to γ-aminobutyric acid, or dihydroxyphenylalanine to dopamine).



Scheme 8

The cofactor form of vitamin B6 (pyridoxal-5′-phosphate, PLP) is also a cofactor for glycogen phosphorylase (Migocka-Patrzałek and Elias, 2021), which catalyzes the hydrolysis of ether bonds in glycogen to form 6-phosphoglucose. In addition, the formation of heme is vitamin B6 dependent. The synthesis of δ-aminolevulinic acid initiates the heme pathway from glycine and succinyl-CoA, the rate-limiting enzyme responsible for this reaction. Vitamin B6 is also an essential component of enzymes that catalyze sphingolipids' formation, particularly the synthesis of ceramides.



Scheme 9

In transamidase, decarboxylase, and deaminase reactions, the 4-formyl (aldehyde) substituent of PLP condenses with the α-amino group of amino acids substrates to form an azomethine (Schiff-base) linkage. The Schiff-base linkage allows the conjugated double-bond system to extend into PLP's pyridinium nitrogen, which weakens bonding around the α-carbon of the amino acid substrate. Then, depending on the enzyme, transaminations, decarboxylations, or subsequent substitutions at the carbon adjacent to the α-carbon are facilitated. Stated differently, PLP helps stabilize a carbanion at the α-carbon of the targeted amino acid. Moreover, in PLP-dependent enzymes, PLP is attached to the ε-amino group of the lysyl group at the active site within the enzyme. This can set the stage for an initial "transimidation" with an amino acid, which occurs faster than an aldehyde with an amino acid, i.e., transaminations.

In glycogen phosphorylase reactions, the organic ring of PLP functions as a reversible electron sink with the phosphate group of PLP acting as an acid-base catalyst to prime free phosphate ion for an attack on the terminal glycosidic bond of the glycogen substrate. Glycosidic bonds are ethers, which prefer cleavage by acids. For their cleavage, the 5′-phosphate moiety of PLP acts as the acid catalyst to yield as a product, glucose -6- phosphate (9).

Metabolism

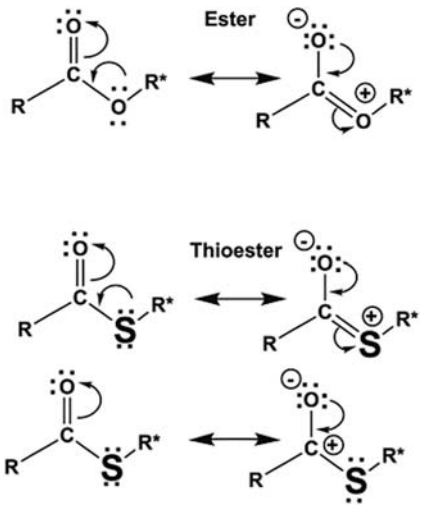
The vitamin B6 requirement is 1.6 mg–2.0 mg per day for adults, increasing during pregnancy and lactation periods. The richest sources of vitamin B6 are meats and whole grains. Vitamin B6 is absorbed in the upper gastrointestinal tract. The cofactor forms of vitamin B6 present in foods are first dephosphorylated by intestinal alkaline phosphatases and then absorbed. The B6 vitamers are transported to target cells by albumin. In cells, rephosphorylation of the B6 vitamers (pyridoxal, pyridoxamine, and pyridoxine) to corresponding 5′-phosphates is catalyzed by cellular pyridoxal kinases. The pyridoxine and pyridoxamine forms may then be oxidized, if needed, to pyridoxal phosphate. The B6 vitamers are most abundant in muscle, kidney, and liver (the primary sites for amino acid metabolism).

The products of vitamin B6 metabolism are excreted in the urine. The primary product is 4-pyridoxic acid. Other urine products may include additional pyridoxine derivatives. Drug-induced vitamin B6 deficiency can also occur following the administration of tuberculostatic drugs, such as isoniazid (isonicotinic acid hydrazide). This drug forms hydrazone derivatives with the pyridoxal forms of B6. Penicillamine (β-dimethylcysteine), a copper chelator, may also interfere with normal B6 metabolism due to the formation of thiazole derivatives. A naturally occurring antagonist to vitamin B6 is linatine (1-amino-D-proline). The signs of Vitamin B6 deficiency are energy-related and neurological due to the inability to produce α-ketoacids and vital biogenic amines from amino acid precursors. Microcytic anemia may also occur from decreased heme synthesis.

Pantothenic acid

Functions

Pantothenic acid (composed of pantoic acid linked to β-alanine) is required for coenzyme A (CoA) biosynthesis (**Fig. 14**) and as the cofactor for acyl carrier protein (ACP). CoA is an essential cofactor for the transfer of acetyl and acyl groups (Naquet et al., 2020). For CoA synthesis, pantothenic acid is phosphorylated to 4′-phosphopantothenate by pantothenate kinase, which also serves as the principal regulatory enzyme for the process. Next, cysteine is added by the enzyme phospho-pantothenoylcysteine synthetase to form 4′-phospho-N-pantothenoylcysteine, followed by decarboxylation and adenosyl-monophosphate addition to form dephospho-CoA. In a final step, dephospho-CoA is phosphorylated to coenzyme A by dephospho-coenzyme A kinase. In addition, a 4′-phosphopantetheine moiety is essential to the function of acyl carrier protein (ACP), which aids in the transfer of acyl groups in fatty acid and polyketide biosynthesis complexes. ACP is sometimes called a "macro-cofactor" because, as a small polypeptide chain (M.W. ~8500–8700 Da), it may be isolated and separated from the fatty acid synthase enzyme complex. Although few reports are specifically related to the regulation of the holo-ACP peptide or the fatty acid synthesis ACP-domain, it is reasonable to assume overall production of ACP activity is linked closely to CoA production.



Scheme 10

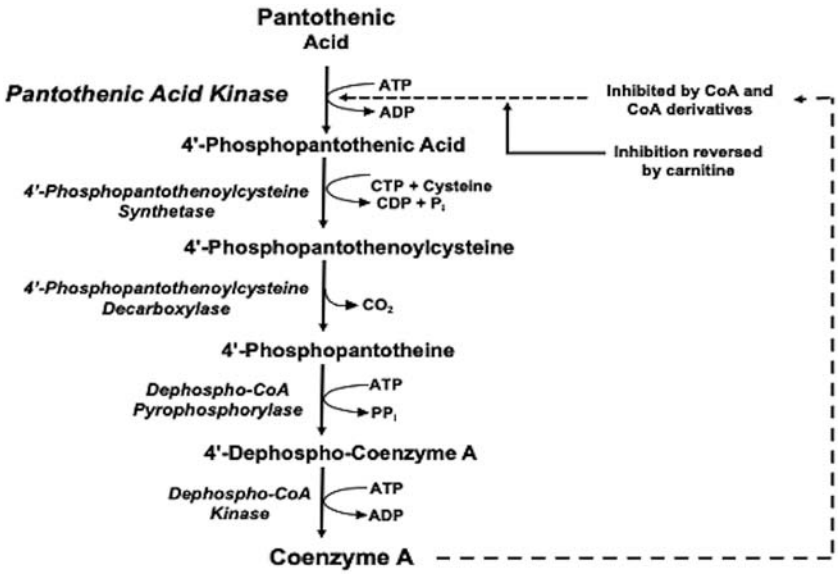
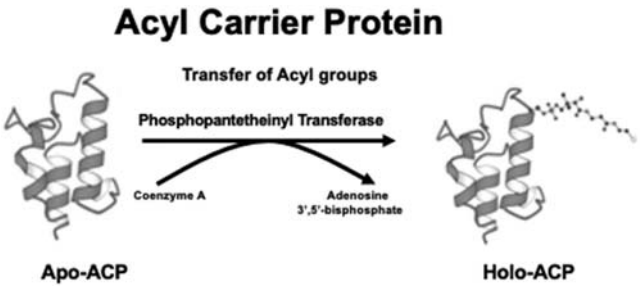


Fig. 14 Pantothenic acid and Coenzyme A (CoA) synthesis. For CoA synthesis, pantothenic acid is first phosphorylated to 4′-phosphopantothenate by pantothenate kinase (PK). PK is the central control and rate-limiting step for the CoA synthesis pathway. Feedback-inhibition by acetyl and acyl-CoA controls PK activity. The intracellular concentration of ATP is also involved in the overall regulation Of PK and CoA synthesis. In addition, carnitine is a non-competitive inhibitor for PK in part due to its role as a lipid membrane transporter for medium/long-chain fatty acids facilitating their oxidation, resulting in subsequent energy production.

An essential feature of the intermediates arising from the transfer reactions catalyzed by CoA or 4′-phosphopantetheine enzymes is they are "high energy" compounds. CoA and ACP react with acetyl or acyl groups to form thioesters. Thioesters (-S-C=O-R) are thermodynamically less stable than typical esters (-O-C=O-R) or amides (-N-C=O-R). The double bond character of the C=O bond in thioesters does not extend significantly into the C-S bond. Consequently, S-C=O-R is not stabilized by resonance. Thioester bonds have a greater negative standard free energy of hydrolysis are more reactive than oxygen esters in nucleophilic displacement reactions involving acyl groups. For most reactions involving CoA or ACP, additional energy from an energy source (e.g., ATP) is not required to transfer acetyl or acyl groups. For reference, the -ΔG of hydrolysis is ~ -7.5 kcal/mol for acetyl-CoA compared to 7–8 kcal/mol for the hydrolysis of ATP to AMP or pyrophosphate or ADP and phosphate.



Scheme 11

CoA is involved in processes that range from energy and fatty acid metabolism to protein acetylations, prenylations, and acetylations (Naquet et al., 2020). CoA is essential as a cofactor in amino acid metabolic pathways. As a β-Hydroxy β-methylglutaryl (HMG) derivative, CoA-HMG is an intermediate in mevalonate and ketogenesis pathways essential in cholesterol and steroid hormone synthesis. Moreover, CoA has also been identified as a cofactor in redox regulation, acting as a reactive oxygen species (ROS) inhibitor. The attachment of CoA to certain proteins that require –[S-S]- bonding for activity occurs during oxidative and metabolic stress. CoA addition (S-thiolation or CoA-thiolation) prevents irreversible sulfhydryl (-SH) oxidation. Proteins that require exposed cysteine residues for activity are often targets for ROS (e.g., free radicals, peroxides in various forms, or heavy metals).

The need for CoA and its derivatives is primarily controlled by the cell's ability to make and utilize ATP. A relative increase in cellular ATP has the potential to down-regulate energy-generating processes (glycolysis, citric acid cycle) and modulate energy-consuming pathways,

such as those involved in amino acid synthesis or gluconeogenesis. Conversely, a cellular decrease in CoA reduces mitochondrial energy generation due to reduced citric acid cycle activity.

As noted, pantothenic acid kinase serves as the primary control point in CoA synthesis. Accordingly, feedback inhibition of pantothenic acid kinase occurs by end products of the synthetic pathway (e.g., CoA or CoA derivatives). Other inhibitors and activators for CoA synthesis include L-carnitine, essential for transporting fatty acids into mitochondria. Carnitine does not affect pantothenic kinase activity independently but reverses the inhibition of pantothenic acid kinase by CoA. However, reversal of kinase inhibition by CoA does not occur when carnitine is acylated.

Metabolism

Pantothenic acid in food is present as CoASH or as protein- or peptide-bound 4′-phosphopantetheine. Near quantitative release of pantothenic acid or 4′-phosphopantetheine is achieved by the sequential actions of two intestinal hydrolases with pyrophosphatase and phosphatase activity. Another intestinal hydrolase, pantetheinase, further hydrolyzes 4′-phosphopantetheine to pantothenic acid. The uptake of luminal pantothenic acid occurs by a sodium-dependent multivitamin transporter (SMVT) found at the surface of intestinal and colonic enterocytes. As a multivitamin transporter, SMVT is also a transporter for biotin, lipoic acid, various sugars, myo-inositol, iodide, short-chain fatty acids, and choline.

Pantothenic acid status is reflected by both whole blood concentration and urinary excretion. Whole blood concentrations typically range from 1.6 to 2.7 μmol/L. A value of <1 μmol/L is considered low. In blood, circulating pantothenic acid is not bound tightly to a specific protein. Whole-body retention of pantothenic acid is mainly by renal tubular reabsorption. The daily excretion of pantothenic acid is closely linked to intake.

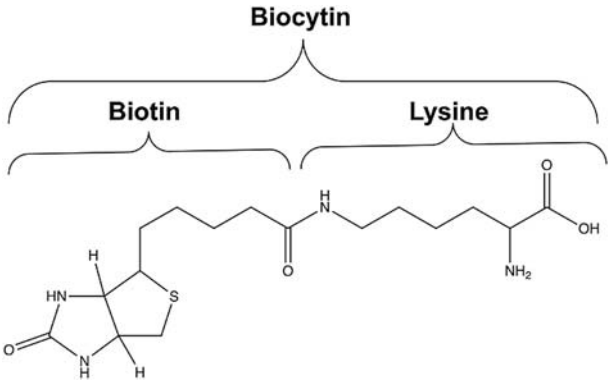
The cellular concentrations of pantothenic acid in the liver and heart are 10–15 μmol/L and ~100 μmol/L, respectively, i.e., 10–100 times that in blood. Also, the concentration of CoA derivatives is 100–400 μM, and the concentration of ACP is 10 μM or more in the cytosol of typical cells. Mitochondria account for 70–90% of the total cellular CoA content.

Adequate intakes (AIs) for adult men and women throughout the life cycle have been established based on observed mean intakes and estimates of basal excretion in urine. Five milligrams per day appears adequate, with an additional one milligram per day recommended during pregnancy and lactation (six milligrams per day). However, an Estimated Average Requirement (EAR) or Recommended Daily Allowance (RDA) has not been established. CoA and ACP are involved in so many aspects of metabolism that biochemical markers or clinical deficiency criteria specific for pantothenic acid deficiency have been challenging to define.

Methyl transfer reactions: biotin, folic acid, vitamin B12

Biotin

Biotin is a cofactor for enzymes collectively known as biotin-dependent carboxylases that catalyze the transfer of carboxyl groups between donor and acceptor molecules (cf., **Table 2**). To activate biotin-dependent apocarboxylases, biotin must be attached to a given apocarboxylase by biotin ligase (holocarboxylase synthetase, HCS). An amide linkage binds biotin to a lysyl ε-amine group at the apocarboxylase active site to form biocytin.



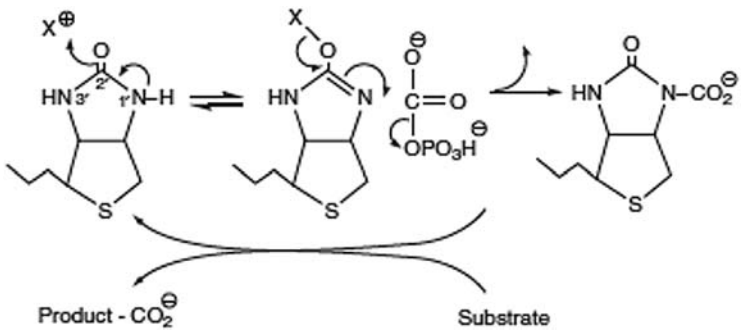
Scheme 12

The attachment reaction requires ATP and precedes via a biotinyl-AMP intermediate. The resulting structure combines the aliphatic chains of biotin and lysine to permit the ring structure of biotin to extend approximately 14 Å from the enzyme's surface. Such extensions facilitate transfers from the donor sites on one enzyme or enzyme subunit to another containing receptor sites. For perspective, a -C-C- is about 1.5 Å in length.

The catalytic site on biotin is the N in its 5-member ring. The reaction occurs in two steps. The first requires the formation of a carboxy biotinyl intermediate. The mechanism involves tautomerization of the ureido nitrogen, which enhances the intermediate's nucleophilicity. The second step is phosphorylation of bicarbonate by ATP to produce carbonyl phosphate, an electrophilic

Table 2 The biotin-dependent carboxylases.

Substrate	*Enzyme*	*Product*	*Importance*
Acetyl CoA	Acetyl CoA carboxylase	Malonyl CoA	Acetyl CoA carboxylase catalyzes the formation of malonyl-CoA for the biosynthesis of fatty acids.
Pyruvate	Pyruvate carboxylase	Oxaloacetate	Pyruvate carboxylase is essential to gluconeogenesis. Oxaloacetate is a TCA cycle intermediate,
Propionyl CoA	Propionyl CoA carboxylase	Methylmalonyl CoA	The conversion of Propionyl CoA to malonyl CoA is an essential reaction in the catabolism of odd-chain fatty acids, methylated-branched-chain fatty acids, and amino acids, such as valine, isoleucine, and methionine. In a subsequent step (see vitamin B-12), methylmalonyl CoA isomerizes to the TCA cycle intermediate succinyl-CoA.
3-Methylcrotonyl CoA	3-Methylcrotonyl CoA carboxylase	3-Methylglutaconyl CoA	3-Methylcrotonyl CoA is utilized in leucine catabolism.



Scheme 13

mixed-acid anhydride, which reacts to generate as a product, a carboxylated biotinyl-enzyme. The carbon in the CO_2 takes on carbanion character facilitating subsequent reactions, such as converting acetate to malonate or pyruvate to oxaloacetate.

Another recently identified function of biotin is the biotinylation of histones. Nuclear holocarboxylase synthetases catalyze the covalent binding of biotin to histones. Eleven biotinylation sites have been identified in histones H2A, H3, and H4. Recruitment of histone-modifying complexes enables the regulation of targeted target genes (e.g., the biotinyl carboxylases). Thus, there are two biotin-dependent roles; the metabolic one related to biotin's carboxylase cofactor functions and a distinct regulatory function mediated by nuclear holocarboxylases.

Metabolism

Biotin is present in relatively high concentrations in cereal grains; however, the bioavailability of biotin varies widely. Absorption involves the proteolysis of biotin-containing enzymes to release biocytin. Biocytinase is an enzyme that catalyzes the cleavage of the peptide linkage between biotin and lysine to release free biotin. Like the other water-soluble vitamins, biotin is taken up by specialized Na-dependent multivitamin carrier-mediated mechanisms. In adult humans, biotin uptake is significantly higher in the duodenum and jejunum compared to the ileum. When biotin-containing carboxylases are degraded in cells, biotin is also released as biocytin. Cellular biocytinases catalyze cleavage to release free biotin for reutilization.

An intake of 30 μg of biotin per day is recommended for adult humans. The dietary biotin intake in Western populations has been estimated to be 35–70 μg/d. The biochemical manifestation of biotin deficiency includes ketolactic acidosis, organic aciduria, and hyperammonemia. Symptoms of overt biotin deficiency include hair loss and dermatitis. Neurological symptoms include lethargy, numbness, and tingling of the extremities. In addition, hepatic steatosis may occur.

Nutritional biotin deficiency is rare, although problems may arise because biotin and biocytin have an affinity for certain proteins, particularly avidin in egg white. Avidin is not easily digested, and as a consequence, biotin is not released in regions of the small intestine where efficient absorption occurs. The inclusion of raw eggs in diets can cause biotin deficiency. For example, the response in commercial fur-bearing animals (alopecia, hair loss) to ingestion of significant quantities of raw egg white has been described as "egg white injury." The search for the "egg white injury" factor eventually led to the discovery of biotin. Genetic disorders involving cellular biocytinase can also precipitate a biotin deficiency. Although rare (<1 in 50,000), biotinidase deficiency is an autosomal recessive disorder in which the body cannot recycle biotin. Its signs and symptoms typically appear within the first few months of life and, if severe, can cause seizures, muscle hypotonia, ataxia, skin rashes, and alopecia. Affected children have delayed development.

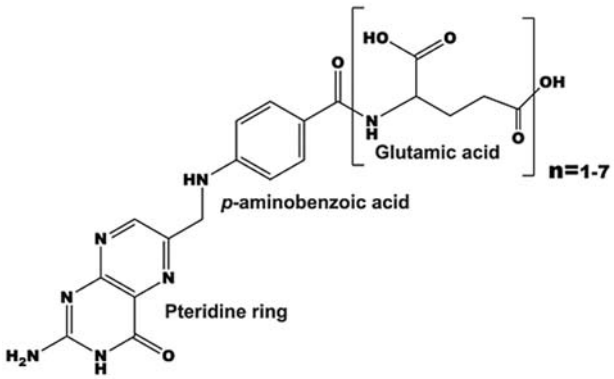
Lifelong treatment can prevent some of these complications from occurring or improve them if they have already developed. For example, symptomatic children with biotinidase deficiency improve after treatment with 5–10 mg of oral biotin per day; however, some of the features, such as developmental delay, optic atrophy, and hearing loss, are often irreversible.

Folic acid

Functions

Folic acid (folate) is essential for DNA, RNA, and amino acid metabolism. Folate-catalyzed reactions include one-carbon transfers required for purine synthesis, methylation of DNA, thymidine, choline, and *S*-adenosylmethionine [cf. 11]. Folate catalyzed one-carbon transfers are also crucial in the catabolism of histidine and tyrosine.

The structure of folic acid (N-pteroyl-L-glutamic acid) is composed of a pteridine ring attached to *p*-aminobenzoic acid (PABA) linked via the carboxyl group to the α-nitrogen of glutamic acid. The active coenzyme, tetrahydrofolate (FH_4), is formed by dihydrofolate reductase, adding four electrons and four hydrogens to the pteridine ring. Mechanistically, hydride ions are transferred from NADPH to the C6 and C8 positions of the pteridine ring. In addition, FH_4 may have up to seven glutamyl residues and exist in different chemical forms, although many are inconvertible.



Scheme 14

One-carbon group transfers are initiated by adding a one-carbon group from serine to form tetrahydrofolate (THF). The reaction is catalyzed by serine hydroxymethyltransferase with glycine and 5,10-methylenetetrahydrofolate (5,10-CH_2-THF) as products (Fig. 15A). Next, methylenetetrahydrofolate dehydrogenase oxidizes 5,10-methylenetetrahydrofolate, which is hydrolyzed to produce 5-formyl-THF or 10-formyl-THF. Alternative carbon sources include formate. Formate–tetrahydrofolate ligase catalyzes the addition of formate to yield 10-formyl-THF. Glycine, histidine, and sarcosine can also directly contribute to the pool of single

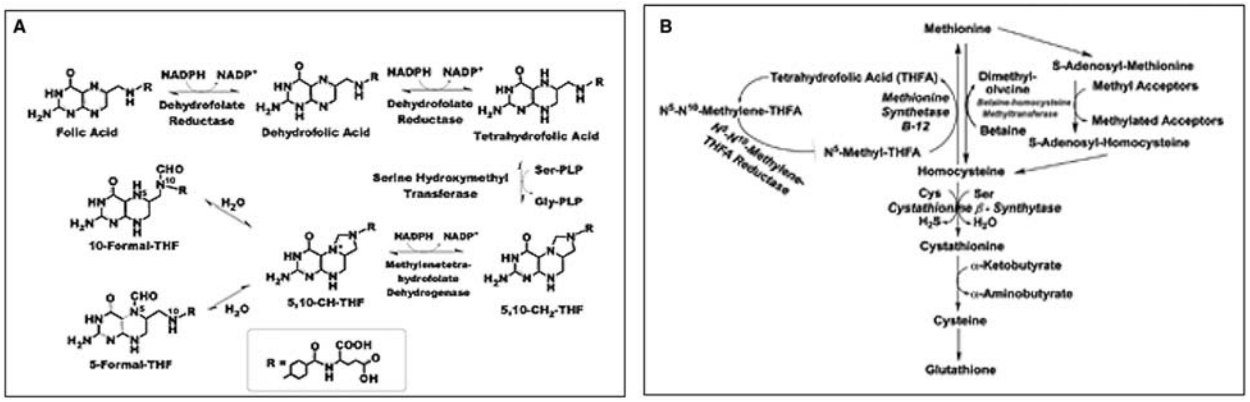


Fig. 15 Folic Acid and single carbon metabolism. (A)-carbon group transfers are initiated by adding a one-carbon group from serine to tetrahydrofolate (THF). The reaction is catalyzed by serine hydroxymethyltransferase with glycine and 5,10-methylenetetrahydrofolate (5,10-CH2-THF) as products. Next, methylenetetrahydrofolate dehydrogenase oxidizes 5,10-methylenetetrahydrofolate, which is hydrolyzed to produce 5-formyl-THF or 10-formyl-THF. Depending on the process, folic acid cofactors can facilitate the transfer of one-carbon units as methyl, formyl, formimino, or methylene groups. (B) Folic acid plays a key role in the methionine cycle. The methionine cycle generates the methyl donor S-adenosyl methionine for the transmethylation of proteins, nucleic acids, and other molecules. The cycle also yields homocysteine which can lead to cysteine and glutathione production.

carbon units. Depending on the process, folic acid cofactors can facilitate the transfer of one-carbon units as methyl (Fig. 15B), formyl, formimino, or methylene groups (Lyon et al., 2020).

Metabolism

By the late 1940s, folic acid was recognized as one of the factors associated with macrocytic anemias. Two proteins, glutamate γ-carboxypeptidase (folate conjugase) and reduced folate carrier protein are essential to the folate absorption process. Following deconjugation or hydrolysis of glutamyl residues, folic acid is taken up by enterocytes and reduced and methylated to 5-methyl-THF. It is the 5-methyl-THF form that is transported across the intestinal basolateral membrane and carried to target organs, such as the liver.

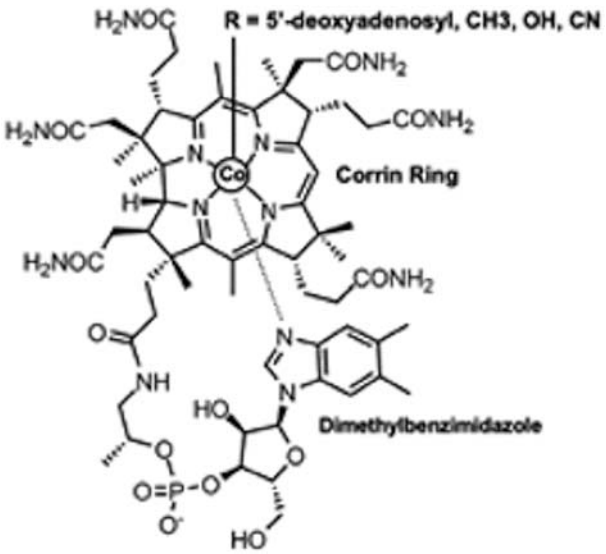
The uptake of 5-methyl-THF by the liver and other tissues involves carrier-mediated transport. 5-methyl-THF is converted to folylpolyglutamate, which keeps the coenzyme inside cells and directs its binding to specific enzymes. Regarding folate turnover, 5%–20% of liver folate undergoes biliary secretion (as 5-methyl-THF) and is subject to enterohepatic recirculation with the remaining passes into the systemic circulation. Maintenance of body pools is also dependent on renal filtration and reuptake. About 1% of the total body folate pool is excreted daily in the urine and 0.1% in the feces. Absorption of folic acid is about 85%, compared to 50% or less for the more complex dietary forms of folylpolyglutamates.

The requirements for folic acid range from 2 to 5 mg per kg of diet or 0.5–3 mg per 1000 kcal. Because folate is required for maintaining nucleotide balance during DNA synthesis, its deficiency is expressed by increased cell death and, in some cases, by a compensatory increase in cell proliferation. For example, in bone marrow, megaloblastosis and macrocytosis of enterocytes reflect defective DNA synthesis. There is a production of larger dysfunctional cells that eventually translates into macrocytic anemia, a condition in which the larger red cells are insufficient in number and hemoglobin. Folic acid deficiency can also result in developmental and neural tube defects and contribute to the development of cardiovascular disease. Hyperhomocysteinemia due to folic acid deficiency is associated with carotid artery narrowing.

As a final point, the discovery that THF is required for DNA synthesis led to the identification of antimetabolites, which function as inhibitors of folic acid reductase and DNA synthesis. The best example is methotrexate, which inhibits the proliferation and regeneration of rapidly replicating cells. Methotrexate irreversibly inhibits dihydrofolate reductase, an enzyme that participates in THF synthesis. With insufficient THF, cell division is blocked in the S phase due to impaired DNA synthesis. Drugs such as methotrexate are widely used to treat autoimmune diseases and cancer chemotherapies.

Vitamin B12

Among its unique features, B_{12} is the only vitamin with a transition metal ion, cobalt, coordinated to its structure. The presence of a cobalt(III) allows the formation of an alkyl bond, which in effect promotes catalysis comparable to a chemical Grignard reagent. The core of the vitamin consists of a corrin ring with a central cobalt atom. The corrin ring contains four pyrrole rings linked together, similar to a heme structure. The cobalt atom within the ring can have a +1, +2, or +3 oxidation state. A fifth valence (below the ring plane) allows dimethylbenzimidazole attachment. Depending on the reaction or enzyme, the remaining sixth valence can either associate with a methyl group, an –OH group, or a 5′-deoxyadenosyl group. Vitamin B12, as 5′-deoxyadenosylcobalamin, is the most common form of the coenzyme found in cells. The 5′-deoxyadenosylcobalamin arises by an attack on the -5′-carbon of ATP by the Co^{+}, which displaces the entire triphosphate group. Enzymes that require B_{12} fit one of two functional categories: those that transfer methyl groups from the coenzyme to the substrate and those that take part in the positional rearrangement of neighboring groups on the substrate or group transfer reactions.



Scheme 15

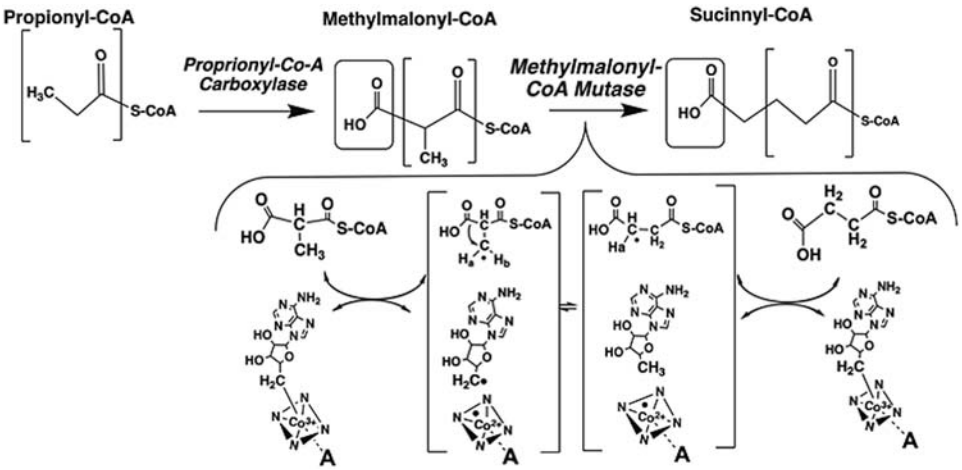


Fig. 16 Methylmalonyl-CoA mutase. Methylmalonyl-CoA mutase catalyzes positional rearrangement reactions, such as converting methylmalonic acid to succinyl-CoA. Methylmalonyl-CoA mutase uses deoxyadenosyl-cobalamin as a cofactor. The reaction mechanism begins with the homolytic cleavage deoxyadenosyl-cobalamin C—Co(III) bond with the C and Co atoms, each acquiring one of the electrons. The Co ion fluctuates between its Co(III) and Co(II) oxidation states and serves as a reversible generator of a free radical.

Methionine synthase catalyzes the methyl-B12-dependent methylation of homocysteine to methionine (Fig. 15B). The reaction is required to produce methionine and generate S-adenosylmethionine, an essential methyl group donor in cells.

In contrast, methylmalonyl-CoA mutase is an example of a positional rearrangement reaction. It is required for the catabolism of branched-chain amino acids into anaplerotic substrates, i.e., substrates essential in maintaining a high level of efficiency in metabolic cycles, such as the TCA cycle. An example is the conversion of methylmalonic acid to succinyl-CoA for ultimate use as a metabolic fuel. Moreover, methylmalonyl-CoA mutase uses the deoxyadenosyl-cobalamin derivative as a cofactor. Thus, the inability to generate deoxyadenosyl-cobalamin may also result in impaired protein catabolism and concomitant mitochondrial energy disruption (Fig. 16).

Metabolism

Vitamin B12 absorption is complex owing to vitamin B12 instability in acid and nonspecific interactions with proteins and complex carbohydrates. For example, vitamin B12 released from food sources immediately associates with so-called "R-proteins" (haptocorrins and cobalophilins). This action serves to protect vitamin B12 during the process of digestion in acidic environments. At acidic pH, these binding proteins have a very high affinity for vitamin B12.

In addition, another vitamin B12 binding protein, intrinsic factor (IF), is synthesized by the gastric parietal cells in humans. However, this binding protein functions more effectively at neutral pH. In the small intestine, pancreatic proteases digest the R-binding proteins, releasing vitamin B12 which then binds to intrinsic factor. For eventual absorption, it is the B12-IF complex that is recognized by the receptors specific for vitamin B12 located in the terminal ileum. Thus, interference with R-protein or IF production can influence the availability of vitamin B12. Although gastric acid production is highly regulated, hypochlorhydria and achlorhydria can occur. When there is low or no gastric acid in the stomach, there is a greater risk of digestive tract infections. In addition to the potential for impaired vitamin B12 absorption, bacterial overproduction increases the possibility of competition between the host and bacteria for vitamin B12. Such problems are often associated with aging.

The recommended daily amount of vitamin B12 for adults is 2.4 μg per day. Although cobalamin can be synthesized by intestinal flora in the colon because the primary absorption site is the ileum, benefits from endogenous synthesis are precluded. Fortunately, the daily requirement of cobalamin is very small compared to the body pool size (1–2 mg); therefore, it takes many weeks or months to become deficient from dietary inadequacy alone. Nevertheless, nutritional vitamin B12 deficiencies do occur. Diseases of the stomach (e.g., hypochlorhydria and achlorhydria), proximal duodenum, ileum, or pancreatic insufficiency can decrease vitamin B12 absorption. Such conditions left untreated can lead to anemia, fatigue, muscle weakness, intestinal problems, nerve damage, and mood disturbances.

In this regard, it is also important to review the relationships between folate and vitamin B12, along with sparing effects that may be ascribed to methionine. The interrelationship is explained by the so-called "methyl trap hypothesis," the physiological response to a methyl group deficiency resulting from a low supply of methionine or a decrease in the cellular S-adenosyl-methionine (SAM) concentration (Lyon et al., 2020). In such situations, 5′-methyltetrahydrofolate accumulates in a vitamin B12 deficiency due to the inability to transfer methyl groups to homocysteine, which alters the rate of folate reduction. Moreover, 5-methyl-THF is a poor substrate for folylpolyglutamate synthetase. Thus, there is a decrease in folylpolyglutamates synthesis and retention in tissues.

Restoring the methyl group on methionine primes the system to further methylation because methionine can act independently as a primary donor of methyl groups for SAM production (Lyon et al., 2020). The sparing effect of methionine can also be explained by adenosylmethionine inhibition of methylenetetrahydrofolate reductase, which prevents the buildup of 5-methyl-THF, i.e.,

a vitamin B12 deficiency does not, in itself, cause a folate deficiency if there is sufficient methionine. Nevertheless, deficiencies of any one of these conditions can provoke abnormalities that range from anemia to neurologic deficits to oxidative stress. For example, Vitamin B12 deficiency is associated with increased cellular H_2O_2 and reduced glutathione levels, because SAM is involved in the up-regulation of cystathionine-β-synthase.

Conditional nutrients

Compounds with a conditional dietary requirement are often grouped with vitamins, particularly if they act as cofactors and catalytic facilitators in metabolic processes or are involved in metabolite transport (Laurie, 2014). **Table 3** lists compounds that are often considered "conditional nutrients," i.e., there may be a dietary need under certain conditions. Those derived from amino acids or glucose include carnitine, choline, taurine, and inositol. Although not described in **Table 3**, amino acids such as glycine, arginine, cystine, and methionine are often included. For example, in late pregnancy and early development, the capacity for glycine biosynthesis does not satisfy its use related to collagen production and neonatal growth. In addition, glycine is a precursor in metabolic pathways for creatine, glutathione, heme, purines, and δ-aminolevulinic acid, a porphyrin precursor; thus, a lack of glycine may affect their relative production.

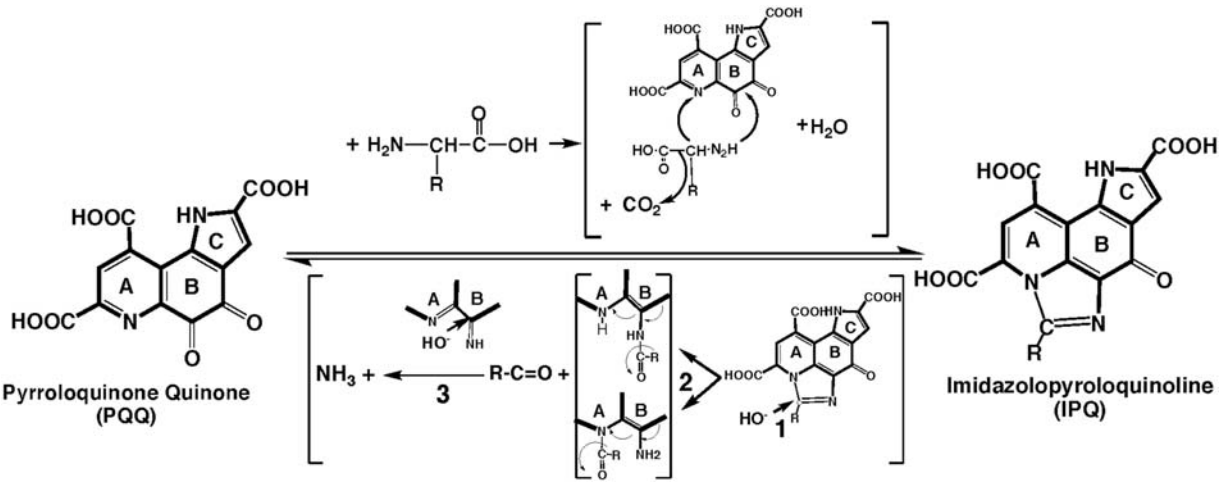
Some lipid-derived products also have conditional requirements (e.g., lipoic acid, lutein, zeaxanthin, and ubiquinone). Moreover, a case for a conditional need may be made for nucleic acid-derived products, such as queuine, a component of tRNA that is symbiotically derived from the microbiome, and biopterin, a cofactor for the aromatic amino acid hydroxylases.

Finally, natural products, such as the plant-derived natural polyphenols, are worthy of mention. Natural polyphenols have been consistently a part of the human diet throughout our entire biological evolution. More than 8000 types of polyphenols have been identified and fall into four general classes of compounds:

- Flavonoids (e.g., flavonoids, proanthocyanidins, anthocyanins) accounts for ~60% of food-derived polyphenols.
- Phenolic acids (e.g., stilbenes and lignins) accounts for ~30% of food-derived polyphenols.
- Polyphenolic amides (e.g., capsinoids and avenanthramides).
- Other polyphenols (e.g., ellagic acid, lignans, tyrosol, and pyrroloquinoline quinone).

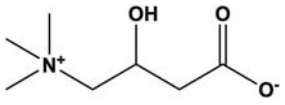
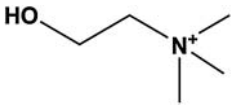
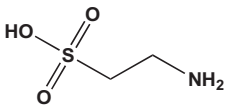
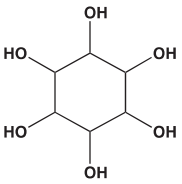
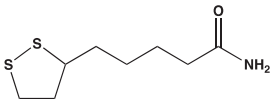
Although consistent physiological effects have been challenging to demonstrate for individual polyphenolic compounds (e.g., quercetin, rutin, curcumin, resveratrol), as a group of compounds, a case can be made for a conditional dietary need. For example, for individual polyphenolic compounds, a given therapeutic response often requires dietary intakes in the hundreds of mgs/day range; but most are present in typical diets at levels less than 10 mg/day. However, in the aggregate, Western diets provide from 0.3 g to 2 g of total polyphenolics per day. Further, many polyphenolic compounds promote physiological responses through the same signaling pathways (e.g., MAP kinase-related signaling pathways) or contribute synergistically to overall ROS defense. Collectively, polyphenolic compounds also cause modification in the expression of cytochrome p450-related xenobiotic pathways. In this regard, arguments have been made using the evolutionary principles of xenohormesis that molecules like polyphenols, which respond to stress-related signals in the plants, have the same health-conferring effects in organisms that evolved as consumers of plants.

Moreover, although consistent physiological effects have been challenging to demonstrate for many individual polyphenolic compounds, there are exceptions. For example, effective doses of [-]-epicatechin (EC) and pyrroloquinoline quinone (PQQ) may be obtained from foods without additional supplementation (Jonscher et al., 2021). EC is associated with increased nitric oxide production, improved blood flow, and physical endurance. Moreover, EC modulates cell signaling, including the MAP kinase-related pathways involved in cell proliferation. In humans, supplementation with EC-rich foods improves insulin sensitivity and glucose. Regarding pyrroloquinoline quinone, PQQ is a biofactor for which a nutritional deficiency can be defined in multiple species of animals (Jonscher et al., 2021). A clear and specific mechanism of action has also been reported for PQQ, and dietary requirements are definable and dose dependent.



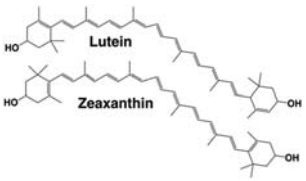
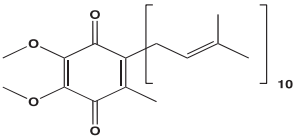
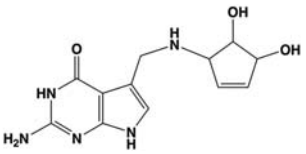
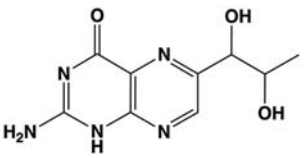
Scheme 16

Table 3 Vitamin-like conditionally nutrients.

Nutrient	*Source*	*Function*	*Structure*
Amino acid and glucose derived			
Carnitine	The L-isomer of carnitine is a derivative of the amino acid lysine. Typical diets result in 40–250 mg L-carnitine/1000 Kcal daily intakes.	Carnitine is essential for the trans-membrane transport of medium- to long-chain fatty acids into the mitochondria. Acyl-CoA is attached to the hydroxyl group of carnitine to form acylcarnitine. This transesterification is catalyzed by carnitine acyltransferase. The acylcarnitine ester is then transported through the intermembrane space facilitated by carnitine-acylcarnitine translocase.	
Choline	Choline can also be synthesized de novo from ethanolamine when the supply of methionine, dimethylcysteine, and betaine is adequate. Lecithin is the most abundant source of choline in the diet. Humans make choline de novo; however, production is generally insufficient.	Choline is a component of phosphatidylcholines and sphingomyelins found in all cell membranes. Phosphatidylcholines are also needed for the synthesis of VLDL and pulmonary surfactants. Moreover, choline is required to produce the neurotransmitter acetylcholine. In addition, choline can serve as a source of methyl groups for SAM production and oxidized in liver mitochondria to glycine betaine aldehyde, which can be converted to trimethylglycine. Trimethylglycine is utilized as an osmoregulator. Therefore, an adequate dietary intake level for choline is 400–600 mg/day for adults. However, the dietary choline requirement is reduced by increased folate, methionine, and vitamin B12.	
Taurine	Taurine is a product of the cysteine sulfinic acid pathway and the transsulfuration pathway for the conversion of homocysteine into cystathionine. About 400 mg of taurine are consumed per day.	Taurine has multiple functions and is essential to physiological processes, such as osmoregulation, immunomodulation, and bile salt formation. Furthermore, it has biological roles in the conjugation of bile acids, ROS defense, osmoregulation, membrane stabilization, and calcium signaling. In addition, it is essential for cardiovascular function and the development of skeletal muscle, the retina, and the central nervous system.	
Inositol (myo-inositol)	*Myo*-inositol is synthesized from glucose 6-phosphate	Inositol derivatives (inositol, phosphatidylinositol, and various mono- and polyphosphates) act as intracellular second messengers in regulating thyroid-stimulating hormone, follicle-stimulating hormone, and insulin. Inositol derivatives also facilitate the binding of given neurotransmitters and steroid hormones to their receptors. Such actions are essential to insulin signal transduction, neuro signaling, and cell membrane potential regulation.	
Lipid derived			
Lipoic acid	Lipoic acid is synthesized in mitochondria from octanoic acid (C8:0). There is no established dose. However, subjects with diabetic neuropathy benefit from daily doses of ~600 mg.	Lipoic acid participates in the oxidative decarboxylation of α-ketoacids in the TCA cycle. The biosynthesis of lipoic acid links mitochondrial fatty-acid synthesis and SAM metabolism with TCA cycle oxidative capacity.	

(Continued)

Table 3 Vitamin-like conditionally nutrients.—cont'd

Nutrient	*Source*	*Function*	*Structure*
Lutein and zeaxanthin	Lutein and zeaxanthin, however, can be interconverted in the body. Adults in the United States consume on average 1.5–2.5 mg/day of lutein and zeaxanthin combined.	There is some epidemiological evidence that increasing lutein and zeaxanthin intake may lower the risk of cataract development.	
Coenzyme Q10 Ubiquinone	Coenzyme Q10 or ubiquinone is derived from tyrosine following conversion to *p*-hydroxybenzoate. The benzoquinone ring is conjugated with isoprenoid units (derived from the cholesterol synthesis pathway). There is no ideal dose for ubiquinone. The average dietary intake is 3–6 mg/day.	Ubiquinone is a mobile component of the mitochondrial electron transport chain. The ability of the benzoquinone moiety of coenzyme Q10 to accept and donate electrons is a critical feature of its electron transport function. Ubiquinone exists in three oxidation states. In its reduced form, ubiquinone is an effective fat-soluble antioxidant. In addition, it may facilitate the recycling of α-tocopherol.	
Nucleic acid derived			
Queuine Queuosine	Queuine or its nucleotide form is found in the tRNA of nearly all eukaryotic organisms; however, it is produced exclusively by bacteria. Humans obtain queuine from the diet or salvage it from symbiotic microbes. The human queuine requirements are not well understood, and the prevalence of queuine deficiency in humans is unknown.	tRNAs possess numerous modified nucleosides, which are essential in maintaining specific tRNA structures, translational efficiency, and codon recognition. One of the modified nucleosides is queuosine. Queuosine is irreversibly inserted during the maturation of Q-tRNAs (queuosine-modified tRNAs) associated with cell proliferation. Humans acquire queuine from the diet and intestinal microflora. In the absence of queuosine modification, translation at Q-decoded codons slows down, and proper protein folding is affected. In addition, queuine depletion impairs the recycling of the reduced forms of biopterin, resulting in a deficit of tetrahydrobiopterin (see below), which impairs the activity of the aromatic amino acid hydroxylase enzymes.	
Biopterin	Guanosine-5′-triphosphate serves as the substrate for de novo biopterin synthesis. Under rare circumstances, a deficiency of the active form of biopterin may occur due to genetic defects in synthesis pathways. Treatment of BH4 deficiencies consists of BH4 2–20 mg/kg per day supplements or diet to control blood phenylalanine, which is toxic when abnormally elevated.	Reduced biopterin (tetrahydrobiopterin, THB) is a cofactor for aromatic amino acid hydroxylases involved in neurotransmitter synthesis, including dopamine, norepinephrine, epinephrine, and serotonin. Nitric oxide synthesis also uses biopterin derivatives as cofactors. In addition, biopterin is needed to produce skin pigments (catecholamines, melanin, serotonin, and melatonin) and cell signaling (nitric oxide).	

PQQ acts as an accessory factor for lactate acid and other dehydrogenases, promoting NADH oxidation to NAD^+. Consequently, the effects of PQQ exposure, both in vivo and in vitro, mimic cellular NAD^+ augmentation (cf., **Fig. 2**). As noted (see Niacin), NAD^+ is a substrate in deacetylase, desuccinylase, demalonylase, demyristoylase, and depalmitoylase reactions catalyzed by sirtuins. Maintaining cellular NAD^+ levels optimizes mitochondriogenesis and neural-muscular communication. PQQ also appears essential for reproduction in mice and rodent models. In humans, PQQ is associated with attenuating of clinically relevant dysfunctions such as ischemia, neurogenic losses, inflammation, and lipotoxicity. Notably, the levels of PQQ needed for given responses in tissues are in the pM to nM range and easily achieved. For example, PQQ and its derivatives, such as imidazopyrroloquinoline (IPQ) compounds, are found in human milk at levels compatible with reproduction and normal growth in animal models (Jonscher et al., 2021).

Conclusion

A mechanistic understanding of function facilitates the effective and meaningful translational of findings at all levels. Although the criterion for mechanistic insight varies among disciplines, such insight is nevertheless interdependent. Knowledge of mechanisms at a chemical and physical level aids in identifying the question to be asked related to given cellular and organ regulatory functions. At the population level, mechanistic information is essential in defining limits regarding a recommended level of intake. As it relates to fat- and water-soluble vitamins, the current focus on individualized diets requires an abundance of basic information ranging from a person's genetics to the potential of microbiome synergisms with different dietary exposures. A goal for many of the discriptions and examples was to emphasize that a better understanding of mechanisms helps to link genetics, metabolomics, and clinical observations to recommendations for nutrient provision. In this regard, much may be learned from animal models when appropriate attention is given to the allometric approaches needed for correct comparisons.

There is still much to be done. For example, the role of vitamins and related nutrients on the potential and beneficial promotion of the microbiota remains unclear. In addition, more information is needed regarding the potential synergisms between dietary biofactors and groups of biofactors, such the natural polyphenolics. Further, there are compounds, such as pyrroloquinoline quinone, for which a case may be made for essentiality.

References

He, H.Y., Liu, M.Z, Zhang, Y.L., Zhang, W., 2017. Vitamin pharmacogenomics: new insight into individual differences in diseases and drug responses. Genom. Proteom. Bioinf. 15, 94–100. https://doi.org/10.1016/j.gpb.2016.10.005.

Jonscher, K.R., Chowanadisai, W., Rucker, R.B., 2021. Pyrroloquinoline-quinone is more than an antioxidant: a vitamin-like accessory factor important in health and disease prevention. Biomolecules 11, 1441. https://doi.org/10.3390/biom11101441.

Kiss, M., Czimmerer, Z., Nagy, L., 2013. The role of lipid-activated nuclear receptors in macrophage and dendritic cell function: from physiology to pathology. J. Allergy Clin. Immunol. 132, 264–286.

Kleiber, M., 1961. The Fire of Life: An Introduction lo Animal Energetics. John Wiley & Sons, Inc., New York, pp. 1–454.

Laurie, K., 2014. Mischley conditionally essential nutrients: the state of the science. J. Food Nutr. 1, e204J.

Lushchak, V.I., 2014. Free radicals, reactive oxygen species, oxidative stress and its classification. Chem. Biol. Interact. 224, 164–175.

Lyon, P., VStrippoli, Fang, B., Cimmino, L.B., 2020. Vitamins and one-carbon metabolism: implications in human health and disease. Nutrients 12, 2867.

Migocka-Patrzałek, M., Elias, M., 2021. Muscle glycogen phosphorylase and its functional partners in health and disease. Cells 10, 883. https://doi.org/10.3390/cells10040883. PMID: 33924466; PMCID: PMC8070155.

Naquet, P., Kerr, E.W., Vickers, S.D., Leonardi, R., 2020. Regulation of coenzyme A levels by degradation: the "ins and outs". Prog. Lipid Res. 78, 101028. https://doi.org/10.1016/j.plipres.2020.101028.

Patel, M.S., Nemeria, N.S., Furey, W., Jordan, F., 2014. The pyruvate dehydrogenase complexes structure-based function and regulation. J. Biol. Chem. 289, 16615–16623.

Rucker, R.B., Steinberg, F.M., 2002. Vitamin Requirements: relationship to basal metabolic need and function. Biochem. Mol. Biol. 30, 86–89.

Said, H.N., 2013. Recent advances in transport of water-soluble vitamins in organs of the digestive system: a focus on the colon and the pancreas. Am. J. Physiol. Gastrointest. Liver Physiol. 305, G601–G610.

Szent-Gyorgyi, A., 1963. Lost in the twentieth century. Annu. Rev. Biochem. 32, 1–15.

Further reading

Ames, B.N., 2018. Prolonging healthy aging: longevity vitamins and proteins. Proc. Natl. Acad. Sci. U. S. A. 115, 10836–10844. https://doi.org/10.1073/pnas.1809045115.

Bernadette Marriott, M., Birt, Di, Stalling, V., Yates, A. (Eds.), 2020, Present Knowledge in Nutrition, eleventh ed., vol. 1. Academic Press, Elsevier, San Diego, CA 92101, pp. 1–655.

Capone, K., Sentongo, T., 2019. The ABCs of nutrient deficiencies and toxicities. Pediatr. Ann. 48, e434–e440. https://doi.org/10.3928/19382359-20191015-01. PMID: 31710362.

Rucker, R.B., Chowanadisai, W., 2016. Coenzymes and cofactors. In: Caballero, B., Finglas, P., Toldrá, F. (Eds.), Encyclopedia of Food and Health, vol. 2. Academic Press, Oxford, UK, pp. 206–224.

Rucker, R.B., Keen, C.L., Steinberg, F.M., 2014. Vitamins and food-derived biofactors. In: Van Alfen, N.K. (Ed.), Encyclopedia of Agriculture and Food Systems, vol. 5. Academic Press, San Diego, CA, pp. 356–377.

Stevens, S.L., 2021. Fat-soluble vitamins. Nurs. Clin. N. Am. 56, 33–45. https://doi.org/10.1016/j.cnur.2020.10.003. PMID: 33549284.

Alcohol: Absorption, metabolism, and physiological effects

R. Rajendram[a,d], **R.J. Hunter**[b], **and V.R. Preedy**[c], [a] College of Medicine, King Saud Bin Abdulaziz University for Health Sciences, Ministry of National Guard-Health Affairs, Riyadh, Saudi Arabia; [b] The Barts Heart Center, St Bartholomew's Hospital, Barts Health NHS Trust and Queen Mary, University of London, London, United Kingdom; [c] Department of Nutrition and Dietetics, Life Course Sciences, Faculty of Life Science and Medicine, King's College London, London, United Kingdom; and [d] Department of Medicine, King Abdulaziz Medical City, King Abdullah International Medical Research Center, Ministry of National Guard-Health Affairs, Riyadh, Saudi Arabia

© 2023 Elsevier Ltd. All rights reserved.

This is an update of R. Rajendram, R. Hunter, V. Preedy, Alcohol: Absorption, Metabolism, and Physiological Effects, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 40–49, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00005-2.

Introduction 251
Physical properties of ethanol 252
Absorption and distribution of alcohol 252
Metabolism of alcohol 253
Alcohol dehydrogenase 253
Alcohol dehydrogenase isoenzymes 254
Catalase 254
Microsomal ethanol oxidizing system 254
Acetaldehyde metabolism 255
Metabolism of acetate 255
Nonoxidative metabolism of alcohol 255
Kinetics of ethanol elimination in vivo 256
Genetics of ethanol metabolism 256
Blood ethanol concentration 256
Factors affecting blood ethanol concentration 257
Gender differences in blood ethanol concentration 257
Period over which the alcohol is consumed 257
Effects of food on blood ethanol concentration 257
Beverage alcohol content and blood ethanol concentration 258
First-pass metabolism of ethanol 258
Contribution of first-pass metabolism to overall ethanol metabolism 259
Models of ethanol pharmacokinetics 259
Physiological effects of alcohol 259
Effects of alcohol on the central nervous system 259
Neuroendocrine effects of alcohol 260
Effects of alcohol on muscle 260
Alcohol and nutrition 261
Effects of alcohol on the cardiovascular system 262
Effects of alcohol on liver function 262
Effects of acetaldehyde 262
Effects of acetate 263
Conclusion 263
References 263

Key points

- When administered orally, ethanol travels down the gastrointestinal tract until absorbed.
- Absorbed ethanol is oxidized to acetaldehyde.
- Alcohol dehydrogenase (ADH), Microsomal ethanol oxidizing system (MEOS) or Catalase are the three main enzyme systems that catalyze the conversion of ethanol to acetaldehyde.
- Acetaldehyde is highly toxic and forms harmful adducts by binding cellular constituents.
- Acetaldehyde is oxidized to acetate.
- The further metabolism of acetate and its role in alcohol-induced disease are incompletely understood.
- Up to 90% of all ethanol metabolism is hepatic.

Glossary

Alcohol dehydrogenase Alcohol dehydrogenase (ADH) is an enzyme that couples oxidation of ethanol to reduction of nicotinamide adenine dinucleotide (NAD+) to NADH. ADH has a wide range of substrates and functions, including dehydrogenation of steroids and oxidation of fatty acids

Aldehyde dehydrogenase Aldehyde dehydrogenase (ALDH) is an enzyme that couples oxidation of acetaldehyde to reduction of NAD+. The presence of ALDH in tissues may reduce the toxic effects of acetaldehyde

Blood ethanol concentration Blood ethanol concentration is commonly used as a measure of intoxication. It is commonly expressed as mg L^{-1}, g dL^{-1}, or mmol L^{-1}

Catalase is a common enzyme found in nearly all living organisms. The main action is the decomposition of hydrogen peroxide (H_2O_2) to water. It can also oxidize ethanol in a reaction that requires H_2O_2

First-pass metabolism First-pass metabolism of ethanol. Reduces the concentration of ethanol before it reaches the systemic circulation

Microsomal ethanol oxidizing system The microsomal ethanol oxidizing system (MEOS) is another pathway of ethanol metabolism. The key enzyme of the MEOS is cytochrome P4502E1 (CYP2E1). This catalyzes the oxidation of NADPH to $NADP^+$

Introduction

After caffeine, ethanol is the most commonly used recreational drug worldwide. Alcohol is synonymous with ethanol, and drinking often describes the consumption of beverages containing ethanol. In the United Kingdom (UK), a unit of alcohol (standard alcoholic drink; **Table 1**) contains 8 g (equivalent to 10 mL) of ethanol (ethyl alcohol).

In 2016, the Department of Health in the UK published updated guidelines for low risk alcohol intake based on these units of alcohol. This guidance suggests that regularly consuming any amount of alcohol is associated with long-term risks (Department of Health, 2016). However, rather than advocating universal abstinence, the Department of Health (2016) advised regular drinkers (both male and female) to drink less than 14 units per week spread over at least three days. Women were also advised not to drink any alcohol at all while pregnant (Department of Health, 2016).

However, as the amount of ethanol in one unit or a standard alcoholic drink varies throughout the world (**Table 2**), the unit system does not allow international comparisons. It is therefore important to note that the guidance from the UK Department of Health (2016) recommends consumption of less than 112 g ethanol per week (i.e., 14 UK Units of alcohol). Recommendations for sensible limits for alcohol intake also vary worldwide.

Despite these guidelines, the quantity of alcohol consumed varies widely. Many enjoy the pleasant psychopharmacological effects of alcohol. However, some experience adverse reactions due to genetic variation of enzymes that metabolize alcohol. Misuse of alcohol undoubtedly induces pathological changes in most organs of the body (Preedy and Watson, 2004). Many of the effects of alcohol correlate with the peak concentration of ethanol in the blood during a drinking session (Morgan and Ritson, 2003; Abrahao et al., 2017). It is therefore important to understand the factors that influence the blood ethanol concentration (BEC) achieved from a dose of ethanol.

Table 1 Unit system of ethanol content of alcoholic beverages.[a]

Beverage containing ethanol	*Units of ethanol*
Half pint of low-strength beer (284 mL)	1
Pint of beer (568 mL)	2
500 mL of high-strength beer	6
Pint of cider	2
One glass of wine (125 mL)	1
Bottle of wine (750 mL)	6
One measure of spirits (e.g., whisky, gin, vodka)	1
Bottle of spirits (e.g., vodka; 750 mL)	36

[a]The unit system is a convenient way of quantifying consumption of ethanol and offers a suitable means to give practical guidance. However, there are several problems with the unit system. The ethanol content of various brands of alcoholic beverages varies considerably (for example, the ethanol content of beers/ales is in the range 0.5–9.0% so a pint may contain 2–5 units) and the amounts of alcohol consumed in homes bear little in common with standard measures.

Table 2 Geographical variation in the amount of ethanol in one unit.[a]

Country	*Amount of ethanol (g)*
Sweden	20
Japan	19.75
United States	14
Australia and New Zealand	10
United Kingdom	8

[a]The unit system does not permit international comparisons.

To avoid confusion, all subsequent references to "units" of alcohol or "standard drinks" will allude to the UK definition (i.e., 8 g) unless otherwise specified. In addition, all concentrations specified as a percentage will refer to the weight volume (i.e., $w\,v^{-1}$) percentage.

Physical properties of ethanol

Ethanol is produced from the fermentation of glucose by yeast. Ethanol (**Fig. 1**) is highly soluble in water due to its polar hydroxyl (OH) group. The nonpolar (C_2H_5) group enables ethanol to dissolve lipids and thereby disrupt biological membranes. As a relatively uncharged molecule, ethanol crosses cell membranes by passive diffusion.

Absorption and distribution of alcohol

The basic principles of alcohol absorption from the gastrointestinal (GI) tract and subsequent distribution are well understood. Beverages containing ethanol pass down the esophagus into the stomach. The endogenous flora of the GI tract can also transform food into a "cocktail" containing several alcohols including ethanol (Cordell et al., 2019). This is particularly important if there are anatomical variations in the upper GI tract (e.g., Crohn disease or diverticulae; Cordell et al., 2019).

Alcohol continues down the GI tract until absorbed (Halsted et al., 1973). The ethanol concentration therefore decreases down the GI tract. There is also a concentration gradient of ethanol from the lumen to the blood. The concentration of ethanol is much higher in the lumen of the upper small intestine than in plasma (**Table 3**; Halsted et al., 1973). Alcohol diffuses passively, traversing the mucosal cell membranes to enter the submucosal area and then the submucosal capillaries (Kalant, 2004).

Absorption occurs across all of the GI mucosa but is fastest in the duodenum and jejunum (Kalant, 2004). The rate of gastric emptying is the main determinant of alcohol absorption. This is because most ethanol absorption occurs after it has passed from the stomach via the pylorus (Kalant, 2004).

Alcohol diffuses from the blood into tissues across capillary walls. Ethanol concentration equilibrates between blood and the extracellular fluid within a single pass (Kalant, 2004). However, equilibration between blood water and total tissue water may take several hours, depending on the cross-sectional area of the capillary bed and tissue perfusion (Kalant, 2004).

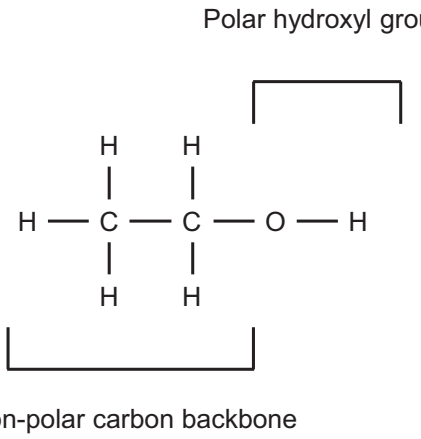


Fig. 1 Chemical structure of ethanol (ethyl alcohol). Ethanol is soluble in water due to its polar hydroxyl (OH) group. The non-polar (C2H5) group enables ethanol to dissolve lipids and thereby disrupt biological membranes.

Table 3 Approximate ethanol concentrations in the gastrointestinal tract and in the blood of a 70 kg male after consumption of 7 units of alcohol.[a]

	Ethanol concentration	
Site	*g/dL*	*mmol/L*
Stomach	8	1740
Jejunum	4	870
Ileum	0.1–0.2	22–43
Blood (15–120 min after dosage)	0.1–0.2	22–43

[a]Ethanol appears in the blood as quickly as 5 min after ingestion and is rapidly distributed around the body. A dose of 0.8 g ethanol per kg body weight (56 g ethanol (7 units) consumed by a 70 kg male) should result in a blood ethanol concentration of 100–200 mg dL^{-1} (22–43 mmol L^{-1}) between 15 and 120 min after dosage. Highest concentrations occur after 30–90 min.

Ethanol enters most tissues but its solubility in bone and fat is negligible. Therefore, in the post-absorption phase, ethanol's volume of distribution will reflect total body water composition. Thus, after a specific dose of ethanol, the BEC will correlate with lean body weight.

Metabolism of alcohol

The rate at which alcohol is eliminated from the blood by oxidization varies from 6 to 10 g h^{-1} (Kalant, 2004). This is reflected by the BEC, which falls by 9–20 mg dL^{-1} h^{-1} after consumption of ethanol (Kalant, 2004). After a dose of 0.6–0.9 g per kg body weight without food, elimination of ethanol is approximately 15 mg dL $blood^{-1}$ h^{-1} (Fisher et al., 1987). However, several factors may modulate this rate and there is considerable individual variation.

Absorbed ethanol is initially oxidized to acetaldehyde (**Fig. 2**) by one of three pathways (**Fig. 3**) (Lieber, 2005; Rajendram et al., 2016a):

1. Alcohol dehydrogenase (ADH)–cytosol
2. Microsomal ethanol oxidizing system (MEOS)–endoplasmic reticulum
3. Catalase–peroxisomes

Alcohol dehydrogenase

The oxidation of ethanol is coupled to the reduction of nicotinamide adenine dinucleotide (NAD^+) to NADH by ADH (Kwo and Crabb, 2002; Kalant, 2004; Höög and Ostberg, 2011). This enzyme has a wide range of substrates and performs several important functions, including dehydrogenation of steroids and oxidation of fatty acids (Kwo and Crabb, 2002).

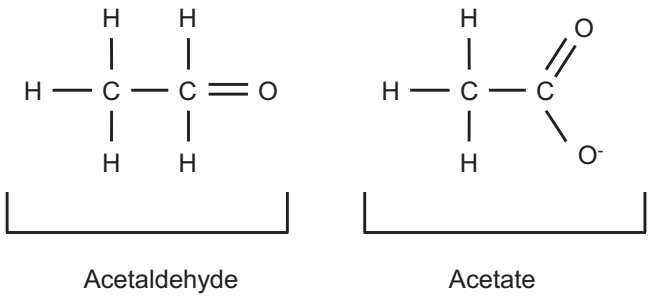


Fig. 2 Chemical structures of acetaldehyde and acetate. Chemical structures of acetaldehyde and acetate, the products of ethanol metabolism. Acetaldehyde and acetic acid/acetate are the current preferred or common names for these chemicals. However, some texts may use their systematic names, i.e., ethanal and ethanoic acid, respectively.

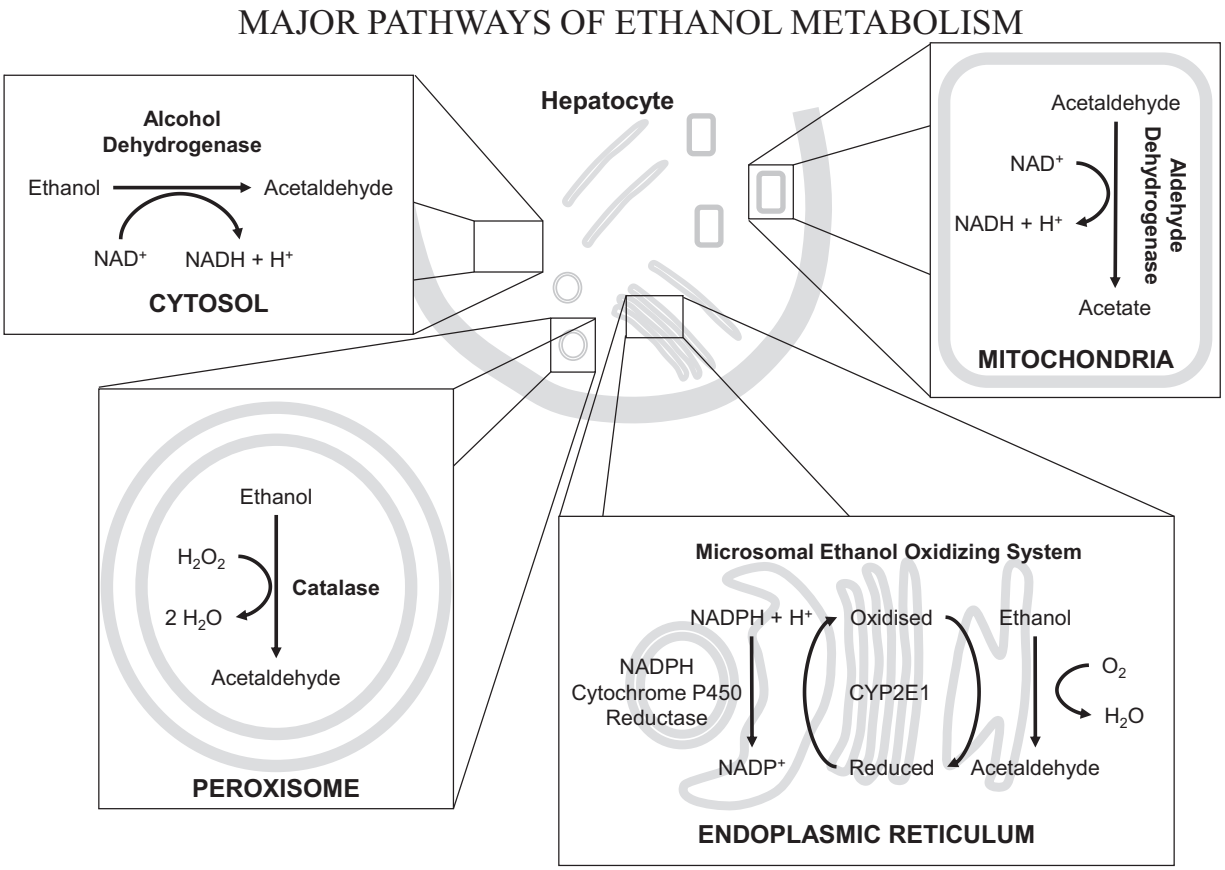


Fig. 3 Pathways of ethanol metabolism. This figure illustrates the most important enzyme pathways that oxidize ethanol to acetaldehyde and acetate. Adapted from Rajendram et al. (2013).

Alcohol dehydrogenase isoenzymes

ADH is a zinc metalloprotein with at least five different isoenzyme classes (Kwo and Crabb, 2002; Edenberg, 2007). These isoenzymes are generated by the dimerization of eight different subunits (Kwo and Crabb, 2002). There are at least seven genes which encode these isoenzymes of ADH (Wang et al., 2012). Class 1 isoenzymes generally require a low concentration of ethanol to achieve "half-maximal activity" (low Km), whereas class 2 isoenzymes have a relatively high Km (Kwo and Crabb, 2002). Class 3 ADH has a low affinity for ethanol and does not participate in the hepatic oxidation of ethanol (Lieber, 2005). Class 4 ADH is found in the human stomach (Moreno and Pares, 1991; Stone et al., 1993; Kwo and Crabb, 2002) and class 5 ADH has been is found in the liver and stomach (Yasunami et al., 1991; Kwo and Crabb, 2002). Whereas the majority of ethanol metabolism occurs in the liver, gastric ADH is responsible for a small portion of ethanol oxidation. It is therefore important to note that the gastric ADH activity in women is less than that in men (Frezza et al., 1990). This may explain some (but not all) of the differences between the effects of alcohol on men and women.

Catalase

Peroxisomal catalase, which requires the presence of hydrogen peroxide (H_2O_2), is usually of little significance in the metabolism of ethanol (Lieber, 2005). Metabolism of ethanol by ADH inhibits catalase activity because H_2O_2 production is inhibited by the reducing equivalents (NADH) produced by ADH (Lieber, 2005). However, metabolism of ethanol by catalase may be more significant if the other pathways for ethanol metabolism are inhibited, for example, by mitochondrial damage in a chronic alcoholic.

Microsomal ethanol oxidizing system

Chronic administration of ethanol with nutritionally adequate diets increases clearance of ethanol from the blood. The MEOS has a higher Km for ethanol (8–10 mmol L^{-1}) than ADH (0.2–2.0 mmol L^{-1}) so at low BEC, ADH is more important (Lieber, 2005).

However, unlike the other pathways, MEOS is highly inducible by chronic alcohol consumption (Lieber, 2005). Cytochrome P4502E1 (CYP2E1) is the most important ethanol metabolizing enzyme in the MEOS (Lieber, 2005). Chronic alcohol use is associated with a 4- to 10-fold increase of CYP2E1 due to increases in mRNA levels and rate of translation (Lieber, 2005).

When ethanol is metabolized by CYP2E1, free radicals are generated and acetaldehyde is produced (Tan et al., 2020). These highly reactive toxins deplete intracellular defenses against oxidative stress (e.g., glutathione) (Lieber, 2005; Konishi and Ishii, 2007). Hepatotoxicity occurs when these defenses are exhausted. The production of lipid hydroperoxides is accelerated by CYP2E1 activity. Lipid hydroperoxides contribute to the development of fatty liver disease and steatohepatitis (Kong et al., 2019). These diseases are commonly associated with obesity, type 2 diabetes, and hyperlipidemia as well as alcohol misuse.

Several commonly used medications are substrates for CYP2E1 (Konishi and Ishii, 2007). These include acetaminophen, diazepam and warfarin. Thus, the metabolism of these medications is increased when the expression of CYP2E1 is induced. This causes tolerance to these medications and reduces their efficacy (Konishi and Ishii, 2007). Induction of CYP2E1 also increases production of toxic metabolites such as, N-acetyl-p-benzoquinoneimine from acetaminophen (Yang and Beard, 2006). These toxins may compound ethanol-induced hepatotoxicity.

This cytochrome also has important roles in normal metabolism and homeostasis. CYP2E1 is involved in fatty acid oxidation and the diversion of ketones to gluconeogenesis (Lieber, 2005; Konishi and Ishii, 2007). The conversion of acetone to acetol is catalyzed by CYP2E1 (Bondoc et al., 1999). This acetol is further metabolized to methylglyoxal (Bondoc et al., 1999). Both methylglyoxal and acetol are involved in gluconeogenesis.

Acetaldehyde metabolism

Humans' main source of exposure to acetaldehyde is their diet. Acetaldehyde is highly toxic but is rapidly converted to acetate. This conversion is catalyzed by aldehyde dehydrogenase (ALDH) and is accompanied by reduction of NAD+ (**Fig. 3**). Like ADH, ALDH also has several isoenzymes (Kwo and Crabb, 2002). The most important of these isoenzymes are ALDH1, which is cytosolic, and ALDH2 which is present in mitochondria (Kwo and Crabb, 2002). The presence of ALDH in most tissues (for example in skeletal muscle) may reduce the toxicity of acetaldehyde.

There are several polymorphisms of ALDH2. The ALDH2*2 allele, for example, encodes an inactive form of ALDH (Edenberg, 2007; Wang et al., 2012). Furthermore, this allele's expression is so dominant that there is virtually no detectable ALDH2 activity even in heterozygotes. While Europeans and Africans rarely have this allele, the ALDH2*2 allele is often found in those of east Asian origin (for example, Japan, China) (Edenberg, 2007; Wang et al., 2012). The enzyme produced on translation of the ALDH2*2 allele does not contribute to acetaldehyde metabolism.

Those who express the ALDH2*2 allele can be subjected to the toxic effects of acetaldehyde by consuming just small amounts of alcohol (Peng et al., 2002). When individuals with this form of ALDH ingest ethanol the plasma concentration of acetaldehyde spikes sharply. This induces unpleasant sensations that probably reduce the risk of alcohol dependence (Peng et al., 2002). These sensations include nausea, facial flushing (see below for the section on facial flushing) and an increase in the heart rate. This is similar to the disulfiram reaction due to the rise of acetaldehyde after inhibition of ALDH (Edenberg, 2007). Disulfiram is used therapeutically to encourage abstinence in alcohol rehabilitation programs. The aversive effects of acetaldehyde may reduce the development of alcoholism and the incidence of cirrhosis in "flushers" (Edenberg, 2007; Wang et al., 2012).

In alcoholics, the oxidation of ethanol is increased by induction of MEOS. However, the capacity of mitochondria to oxidize acetaldehyde is reduced. Hepatic acetaldehyde therefore increases with chronic ethanol consumption (Di Padova et al., 1987). A significant rise in the acetaldehyde concentration of hepatic venous blood matches the high tissue levels of acetaldehyde.

Metabolism of acetate

Acetaldehyde is converted to acetate by ALDH. In non-fasting humans, serum acetate concentrations are usually under 0.2 mmol/L (Ballard, 1972). However, after consumption of alcohol, serum acetate may increase more than 20 times above this baseline concentration (Cornier, 2004). Serum acetate concentrations may also be increased by fasting (i.e., starvation ketosis) and insulin deficiency (i.e., type 1 diabetes). The principles of acetate metabolism are not well described. However, the current understanding has been reviewed by Cornier (2004), Rajendram et al. (2016a) and Bose et al. (2019).

The metabolism of acetate is clearly very important to several homeostatic pathways and functions in cells' cytosol, mitochondria, and nuclei (Bose et al., 2019). Dysregulation of the metabolism of acetate and acetyl-CoA is associated with several diseases including alcoholic hepatitis (Kendrick et al., 2010; Bose et al., 2019).

Nonoxidative metabolism of alcohol

Non-oxidative metabolism of alcohol, which results in formation of ethyl esters from fatty acids occurs in several organs which lack an oxidative system to metabolize alcohol (e.g., pancreas, heart, and adipose tissue) (Heier et al., 2016). These organs often develop alcohol-induced disease so fatty acid ethyl esters may play a role in the pathogenesis of the lesions induced by alcohol consumption

(Heier et al., 2016). The nonoxidative metabolism of ethanol may be more significant if the other pathways for ethanol metabolism are overwhelmed or inhibited.

Kinetics of ethanol elimination in vivo

Initially, it was believed that the process of ethanol elimination always followed zero-order kinetics. That is to say, the rate of elimination of ethanol is always constant and does not vary with the concentration of ethanol (Norberg et al., 2003). Indeed, the metabolism of ethanol does usually follow zero order kinetics because the Km of most ADH isozymes for ethanol is only about 1 mmol/L. Thus, ADH can be saturated even at relatively low ethanol concentrations. When ADH is saturated the rate of ethanol elimination then proceeds at maximal velocity and is then independent of concentration (Norberg et al., 2003).

However, concentration-dependent ethanol metabolism of has been observed in some studies of the elimination of alcohol (Ramchandani et al., 2001). There are several possible explanations for this. If ADH is not saturated the rate of elimination of ethanol will not be constant. At low ethanol concentrations its elimination follows Michaelis-Menten kinetics. In that setting, the ethanol concentration and the kinetic constants Km and Vmax determine the rate of elimination (Matsumoto and Fukui, 2002). Furthermore, because the Km of CYP2E1 and some isoenzymes of ADH for alcohol is high, the elimination of ethanol may remain concentration-dependent even at high BEC. Regardless, because of the intermittent concentration dependence of ethanol metabolism, a single specific rate of alcohol metabolism cannot be defined.

Genetics of ethanol metabolism

Single nucleotide polymorphisms (SNPs) in the genes encoding ADH influence the translated enzymes' characteristics (Wang et al., 2012). The ADH polymorphisms vary between racial subgroups. There are three known polymorphisms in the ADH1B gene and two in the ADH1C gene. The ADH1B*1 allele encodes the β1 subunit of ADH (Edenberg, 2007; Wang et al., 2012). This is the reference allele for ADH1 (Edenberg, 2007; Wang et al., 2012). The enzyme produced on translation of the ADH1B*1 allele has arginine at positions 48 and 370. The ADH1B*2 allele results in the production of an ADH enzyme subunit which has histidine in position 48 (β2). The ADH1B*3 allele produces an ADH enzyme subunit with cysteine in position 370 (β3). These seemingly insignificant differences in amino acids have profound effects on NAD^+ binding (Edenberg, 2007; Wang et al., 2012).

The ADH1B*2 allele is common in Asians and the ADH1B*3 allele is common Africans (Zintzaras et al., 2006; Edenberg, 2007; Wang et al., 2012). Yet, individual studies of the polymorphisms of ADH2, ADH3, CYP2E1 and ALDH2 failed to conclusively identify any associations with alcohol misuse or toxicity. However, a large meta-analysis suggested that the risk of alcohol misuse is increased by the ADH2*1 and ADH3*2 alleles, or the ALDH2*1 allele (Zintzaras et al., 2006). As ADH2*1 and ADH3*2 are less active forms of ADH and ALDH2*1 is very active; these polymorphisms prevent acetaldehyde from accumulating. Rapid clearance of acetaldehyde by ALDH2*1 also reduces the risk of alcoholic liver disease. On the other hand, neither ADH2 nor the ADH3 polymorphisms are implicated in alcoholic liver disease (Zintzaras et al., 2006). The risk of alcohol misuse or alcoholic liver disease is influenced by CYP2E1 polymorphism (Zintzaras et al., 2006).

Research in this field is progressing rapidly, but currently is still in its infancy. Alcohol misuse has been associated with select polymorphisms in the genes encoding enzymes that metabolize ethanol (e.g., ADH1B) has been identified by individual genome-wide association studies (GWAS; Lai et al., 2019). However, the findings of GWAS on other polymorphic loci have been inconsistent (Lai et al., 2019). This may be because the symptoms of alcohol misuse and ethanol toxicity are heterogeneous (Lai et al., 2019).

Considerable effort is still required to tease out the nature of the relationships between alcohol-induced disease and these polymorphisms. Regardless, it is already clear that the expression of these polymorphic genes has significant impact on alcohol metabolism and therefore blood ethanol concentration.

Blood ethanol concentration

The relationship between BEC and the effects of alcohol is complex and varies between individuals and with patterns of drinking. Many of the effects correlate with the peak concentration of ethanol in the blood and organs during a drinking session. Other effects are due to products of metabolism and the total dose of ethanol ingested over a period of time. These two considerations are not entirely separable because the ethanol concentration during a session may determine which of its metabolic pathways are dominant.

It is of considerable clinical interest to understand what factors increase the probability of higher maximum ethanol concentrations for any given level of consumption.

Factors affecting blood ethanol concentration

Ethanol has a low lipid: water partition coefficient. Thus, differences in body composition (i.e., fat and water content) will influence the differences between the BEC of drinkers who ingested the same dose of ethanol per kg body weight (Marshall et al., 1983; Goist and Sutker, 1985; Smith et al., 1993). This is relevant to gender differences in BEC.

Gender differences in blood ethanol concentration

Women achieve higher peak BEC than men given the same dose of ethanol per kilogram of body weight (Marshall et al., 1983; Goist and Sutker, 1985; Smith et al., 1993). Total body water (TBW) strongly influences ethanol's volume of distribution (Marshall et al., 1983). Because the bodies of women contain a greater proportion of fat, it is not surprising that the BEC is higher in women. Moreover, the differences in the BEC of men and women disappears if ethanol is dosed by TBW rather than body weight (Goist and Sutker, 1985). However, the gender differences in the gastric metabolism of ethanol described above may also be relevant. Women have a lower rate of first-pass embolism compared to men when given high concentrations of alcohol (Baraona et al., 2001).

Period over which the alcohol is consumed

Rapid intake of alcohol increases the concentration of ethanol in the stomach and small intestine. The greater the concentration gradient, the faster the absorption of ethanol and therefore peak BEC. If alcohol is consumed and absorbed faster than the rate of oxidation, then BEC increases.

Effects of food on blood ethanol concentration

Gastric emptying is delayed by food. Thus, eating flattens the increase in BEC seen when drinking alcohol on an empty stomach. The mean rate of clearance of ethanol from the blood increases by up to 50% after eating (Jones, 1993). Thus, drinking alcoholic beverages with or after food attenuates the peak BEC and reduces the area under the curve (AUC) when BEC is plotted against time (Fig. 4).

These facts can be illustrated by considering the observations of a study conducted by Jones and Jönsson (1994). In this study ten men were given a drink containing ethanol in the morning either immediately after consuming a standard meal or after fasting overnight (Jones and Jönsson, 1994). Jones and Jönsson measured the BEC in their subjects at several time points after they had consumed the alcoholic beverage.

After administration of 0.80 g kg^{-1} to fasted subjects, Jones and Jönsson (1994) found that the mean peak BEC was 104 mg dL^{-1} (standard deviation; SD 14.5; 23.7 mmol/L SD 3.1). The mean peak BEC was significantly lower if the same dose

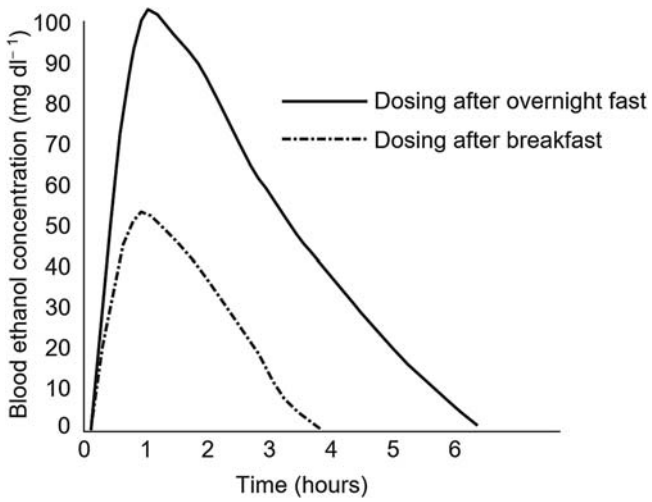


Fig. 4 Blood ethanol concentration curve after oral dosing of ethanol. Blood ethanol concentration curve after oral dosing of ethanol. This figure illustrates the possible trajectory of a subject's blood ethanol concentration after ingestion of 0.8 g kg^{-1} ethanol over 30 min either after an overnight fast (unbroken line) or after breakfast (broken line). The peak blood ethanol concentration and the area under the curve are reduced if ethanol is consumed with food (broken line).

of ethanol was taken after the overnight fast had been broken by a morning meal (67 mg dL^{-1} SD 9.5 mg dL^{-1}; 14.4 mmol/L SD 2 mmol/L; $p < 0.001$).

The mean AUC of the BEC time chart (0–6 h) was 398 mg $dL^{-1} \times$ h (SD 56 mg $dL^{-1} \times$ h) in fasting subjects compared with 241 mg $dL^{-1} \times$ h (SD 34 mg $dL^{-1} \times$ h) in those who drank after breakfast ($p < 0.001$). The ethanol was metabolized approximately 2 h earlier if subjects drank before breaking their overnight fast (Jones and Jönsson, 1994).

The contributions of various nutrients to these effects have been studied, but small, often conflicting, differences have been found. The meal's caloric load seems to more relevant than specific nutritional content (Carbonnel et al., 1994).

However, food increases splanchnic blood flow which maintains the ethanol diffusion gradient in the small intestine (Kalant, 2004). Food-induced impairment of gastric emptying may be partially offset by faster absorption of ethanol in the duodenum (Kalant, 2004).

In animal studies liquid diets are often used to administer ethanol as well as other nutrients. It is, therefore, important to note that when ethanol is given in water, the AUC is greater than when a liquid diet is used to deliver with the same dose of ethanol (de Fiebre et al., 1994). The different blood ethanol profile could affect the expression of pathology in these models (Kalant, 2004).

Beverage alcohol content and blood ethanol concentration

While the type of alcoholic drink (i.e., beer, wine or spirit) does not affect the BEC per se; the ethanol concentration of the beverage consumed (**Table 4**) does affect ethanol absorption and can, therefore, affect BEC (Roine et al., 1991, 1993). Absorption is fastest if the ethanol concentration is 10–30%. Under 10%, the low ethanol concentration in the GI tract reduces diffusion and the greater volume of liquid slows gastric emptying.

The concentration of ethanol in the GI tract also affects first-pass metabolism (FPM) (see below). Gastric ADH requires a high ethanol concentration for optimal activity. The ethanol concentration in the beverage consumed will therefore affect oxidation of ethanol to acetaldehyde (Roine et al., 1991). After ingestion of equivalent amounts of ethanol less FPM and higher blood levels occur after consumption of beer which has a low ethanol concentration than whisky which has a high ethanol concentration (Roine et al., 1993).

First-pass metabolism of ethanol

The AUC is significantly lower after oral dosing (AUCoral) of ethanol than after intravenous (AUCiv) or intraperitoneal administration. The total dose of intravenously administered ethanol is available to the systemic circulation. The difference between AUCoral and AUCiv represents the fraction of the oral dose that was either not absorbed or metabolized before entering the systemic circulation FPM. The ratio of AUCoral to AUCiv reflects the oral bioavailability of ethanol.

The investigation of ethanol metabolism has primarily focused on the liver and its relationship to liver pathology. The role of the stomach in ethanol metabolism is controversial. However, several ADH isoenzymes are present in human gastric mucosa (Kwo and Crabb, 2002). This suggests that the stomach contributes to the FPM of ethanol (Moreno and Pares, 1991; Yasunami et al., 1991; Lieber, 2005). However, gastric metabolism could account for 5% of ethanol oxidation and 2–10% of ethanol is eliminated by urination, sweating, or breathing. The remaining ethanol undergoes hepatic metabolism.

Thus, up to 90% of all ethanol metabolism may be hepatic. However, several factors including ethnicity (Dohmen et al., 1996), gender (Frezza et al., 1990) and alcoholism (Di Padova et al., 1987) alter overall FPM. The observations that their effects on FPM correlate with their effects on gastric alcohol dehydrogenase (ADH) activity (Di Padova et al., 1987; Dohmen et al., 1996; Frezza et al., 1990) is further evidence for ethanol metabolism in the stomach.

The activity of gastric ADH is high. Its Km is approximately 40 mmol/L (Kwo and Crabb, 2002). Thus, although, hepatic oxidation of ethanol cannot increase once ADH is saturated, gastric ADH can metabolize ethanol significantly at the high concentrations in the stomach. Thus, gastric ADH can protect the body somewhat if excessive amounts of ethanol are consumed. If gastric emptying

Table 4 Alcohol content of selected beverages.

	Alcohol content		
Beverage	*g/dL (%)*	*mmol/L*	*mol/L*
Low-strength beers	3–4	650–870	0.65–0.87
High-strength beers	8–9	1740–1960	1.74–1.96
Wine	7–14	1520–3040	1.52–3.04
Brandy	35–45	7610–9780	7.61–9.78
Vodka	35–50	7610–10,870	7.61–10.87
Gin	35–50	7610–10,870	7.61–10.87
Whisky	35–75	7610–16,300	7.61–16.30

is delayed, prolonged contact with gastric ADH increases FPM of ethanol. Conversely, fasting which increases the speed of gastric emptying, reduces gastric FPM (Di Padova et al., 1987).

Contribution of first-pass metabolism to overall ethanol metabolism

After absorption, ethanol is transported to the liver in the portal vein. Some is metabolized by the liver before reaching the systemic circulation. However, hepatic ADH is saturated at a BEC that may be achieved in an average-size adult after consumption of one or two UK units. If hepatic ADH is saturated by ethanol entering the liver from the systemic circulation via the hepatic artery, the ethanol in the portal blood must compete for ADH binding sites.

Thus, fractionally, hepatic FPM is greater when less alcohol is consumed. For example, the fractional FPM after consumption of 0.15 g ethanol kg bw^{-1} will be greater than the fractional FPM after consumption of 0.3 g ethanol kg bw^{-1} (Levitt, 2002). When large amounts of ethanol are consumed, proportionally, the pre-systemic metabolism is relatively insignificant in comparison to the systemic metabolism (Gentry et al., 1994; Crabb, 1997). However, when small doses of alcohol are administered, the contribution of FPM to overall ethanol metabolism is more important (Gentry et al., 1994; Crabb, 1997).

Models of ethanol pharmacokinetics

Several attempts have been made to model the complex pharmacokinetics of ethanol (see Smith et al., 1993; Levitt, 2002). These pharmacokinetic models aim to predict, define and quantify ethanol's absorption and subsequent metabolism. For example, on computerized analysis of data from human subjects, Levitt (2002) found that the ethanol dose and the rate of gastric emptying are the main determinants of ethanol absorption and FPM.

When fasting subjects drink alcohol, ethanol is absorbed rapidly and the FPM was predicted to be small (Levitt, 2002). On the other hand, when alcohol is consumed with food, gastric emptying is delayed so the absorption of ethanol is slow. The FPM was also predicted to be less when more ethanol is consumed. It was predicted that the fractional FPM would be 36% after 0.15 g kg bw^{-1} (10.5 g for a 70 kg man) but after 0.3 g kg bw^{-1} (21 g for a 70 kg man) the model predicted that the fractional FPM would only be 7% (Levitt, 2002).

However, there are several pathways of ethanol metabolism, each of which has variable kinetics. So, the accuracy of these models is poor.

Physiological effects of alcohol

Ethanol or the products of its metabolism affect nearly all cellular structures and functions.

Effects of alcohol on the central nervous system

Ethanol generally decreases the activity of the central nervous system. In relation to alcohol, the most important neurotransmitters in the brain are glutamate, gamma-aminobutyric acid (GABA), dopamine, and serotonin.

Glutamate is the major excitatory neurotransmitter in the brain. Ethanol inhibits the N-methyl-D-aspartate (NMDA) subset of glutamate receptors (Abrahao et al., 2017). Ethanol thereby reduces the excitatory effects of glutamate. GABA is the major inhibitory neurotransmitter in the brain. Alcohol facilitates the action of the GABA-a receptor, increasing inhibition (Abrahao et al., 2017). Changes to these receptors seem to be important in the development of tolerance of and dependence on alcohol (Abrahao et al., 2017).

Dopamine is involved in the rewarding aspects of alcohol consumption (Abrahao et al., 2017). Enjoyable activities such as eating or use of other recreational drugs also release dopamine in the nucleus accumbens of the brain. Serotonin is also involved in the reward processes and may be important in encouraging alcohol use.

The most obvious effects of ethanol intoxication on the central nervous system begin with behavior modification (e.g., cheerfulness, impaired judgment, and loss of inhibitions; Morgan and Ritson, 2003; Abrahao et al., 2017). These excitatory effects result from the disinhibition described previously (inhibition of cells in the brain that are usually inhibitory).

As a result of these effects, it is well recognized that operating vehicles such as cars or heavy machinery under the influence of ethanol is unsafe. However, the BEC after consumption of a specific amount of ethanol and the impairment caused by a specific BEC vary significantly between individuals (Morgan and Ritson, 2003; Abrahao et al., 2017). Despite this variation, BEC is used to define intoxication and provide a rough measure of impairment for legal purposes (Morgan and Ritson, 2003; World Health Organization, 2018) because it is an objective measurement that is difficult to contest.

Most countries have set maximum legally permissible BEC levels for drivers to reduce harm from "drink driving". Governments define these levels after reviewing the available evidence. However, the definition of what is safe or acceptable varies between countries (**Table 5**; World Health Organization, 2018). These BEC thresholds range from zero tolerance (0.0 mg mL^{-1}) to 1.0 mg mL^{-1}.

Table 5 Legal limits of blood ethanol concentrations for the general population to drive.[a]

	Blood ethanol concentration	
Legal limit[b]	*mg/dL*	*mmol/L*
The Czech Republic, Uruguay, and Hungary	0	0
Norway and Sweden	20	4.3
Japan and Montenegro	30	6.45
France, Germany, Italy, and Australia	50	11
United States, and Canada	80	17
Palau	100	22

[a]Ethanol impairs judgment and coordination. It is well recognized that driving under the influence of ethanol is unsafe. However, the definition of what is safe or acceptable varies between countries and can change as a result of social, political, or scientific influences. Moreover, many countries have lower legal blood ethanol levels for professional drivers than the general population.
[b]Legislation regarding legal limits of blood ethanol for driving may change.
Data derived from the World Health Organization (2018).

Some countries are considering the potential social benefits of lowering BEC limits. However, opponents cite factors such as the drinking culture, convenience, the unpalatability of tighter legislation and the impact on the alcohol industry.

The effects of ethanol are dose dependent (**Table 6**; Morgan and Ritson, 2003; Abrahao et al., 2017) and further intake causes agitation, slurred speech, memory loss, double vision, and loss of coordination (Morgan and Ritson, 2003; Abrahao et al., 2017). This may progress to reduce level of consciousness, and induce respiratory depression with loss of airway protective reflexes, with danger of aspiration, suffocation, and death (Morgan and Ritson, 2003; Abrahao et al., 2017).

This sequence of events is particularly relevant in the hospital setting, where patients may present intoxicated with a reduced level of consciousness. It is difficult to determine whether there is coexisting pathology such as an extradural hematoma or overdose of other drugs in addition to ethanol. Although measurement of BEC is helpful (**Table 6**), it is safest to assume that alcohol is not responsible for any disturbance in consciousness and to search for another cause.

Neuroendocrine effects of alcohol

Alcohol activates the sympathetic nervous system, increasing circulating catecholamines from the adrenal medulla. Hypothalamic–pituitary stimulation results in increased circulating cortisol from the adrenal cortex and can, rarely, cause a pseudo-Cushing's syndrome with typical moon-shaped face, truncal obesity, and muscle weakness (Bessemer et al., 2011; Rachdaoui and Sarkar, 2017). Alcoholics with pseudoCushing's show many of the biochemical features of Cushing's syndrome, including failure to suppress cortisol with a 48 h low-dose dexamethasone suppression test (Bessemer et al., 2011).

Ethanol affects hypothalamic osmoreceptors, reducing vasopressin release (Rachdaoui and Sarkar, 2017). This increases salt and water excretion from the kidney, causing polyuria (Döring et al., 2003). Significant dehydration may result particularly with consumption of spirits containing high concentrations of ethanol and little water. Loss of hypothalamic neurons (which secrete vasopressin) has also been described in chronic alcoholics, suggesting long-term consequences for fluid balance (Döring et al., 2003). Plasma atrial natriuretic peptide, increased by alcohol consumption, may also increase diuresis and resultant dehydration (Döring et al., 2003).

Alcoholism also affects the hypothalamic–pituitary–gonadal axis (Rachdaoui and Sarkar, 2017). These effects are further exacerbated by alcoholic liver disease (Rachdaoui and Sarkar, 2017). Testosterone is usually decreased in men, but it may increase in women. Estradiol is increased in men and women, and it increases as hepatic dysfunction deteriorates (Rachdaoui and Sarkar, 2017).

In women, the hormonal changes may reduce libido, disrupt menstruation, or even induce premature menopause (Rachdaoui and Sarkar, 2017). Sexual dysfunction is also common in men with reduced libido and impotence (Rachdaoui and Sarkar, 2017). Fertility may also be reduced, with decreased sperm counts and motility (Rachdaoui and Sarkar, 2017).

Effects of alcohol on muscle

Myopathy is common, affecting up to two-thirds of all alcoholics (Preedy et al., 2001; Preedy and Watson, 2004; Simon et al., 2017). It is characterized by wasting, weakness, and myalgia and improves with abstinence (Preedy et al., 2001; Simon et al., 2017). Histology correlates with symptoms and shows selective atrophy of type II muscle fibers (Preedy et al., 2001). Ethanol causes

Table 6 Relationship between amount of ethanol consumed, blood ethanol concentration (BEC) and effect of ethanol on the central nervous system.

Alcohol consumed[a] (UK units)	*Possible BEC*	*Effect*
1–5	10–50 mg/dL 2–11 mmol/L	No obvious change in behavior
2–7	30–100 mg/dL 7–22 mmol/L **Euphoria** **Sociability**	Increased self-confidence; loss of inhibitions Impaired judgment, attention, and control Mild sensorimotor impairment, delayed reaction times Legal limits for driving generally fall within this range (see Table 5)
8–15	90–250 mg/dL 20–54 mmol/L	Loss of critical judgment Impairment of perception, memory, and comprehension Reduced visual acuity Reduced coordination, impaired balance Drowsiness
11–20	180–300 mg/dL 39–65 mmol/L **Confusion**	Disorientation Exaggerated emotional states Disturbances of vision and perception of color, form, motion, and depth Increased pain threshold Further reduction of coordination, staggering gait, slurred speech
15–25	250–400 mg/dL 54–87 mmol/L **Stupor**	Loss of motor functions Markedly reduced response to stimuli Marked loss of coordination, inability to stand/walk Incontinence Impaired consciousness
22–30	350–500 mg/dL 76–108 mmol/L **Coma**	Unconsciousness Reduced or abolished reflexes Incontinence Cardiovascular and respiratory depression (death possible)
38	>600 mg/dL >130 mmol/L **Death**	Respiratory arrest

[a]Approximate amounts of alcohol required by a 70 kg male to produce the corresponding blood ethanol concentration (BEC) and intoxicating effects of ethanol. One UK unit of alcohol contains 8 g of ethanol. It should be noted that these theoretical data are provided for illustrative purposes only. The BEC and the effects after consumption of ethanol are dependent on several factors and varies significantly between individuals.
Adapted with permission from Morgan and Ritson (2003).

a reduction in muscle protein and ribonucleic acid content (Preedy et al., 2001). The underlying mechanism is unclear, but rates of muscle protein synthesis are reduced, whereas protein degradation is either unaffected or inhibited (Preedy et al., 2001). Attention has focused on the role of acetaldehyde adducts and free radicals in the pathogenesis of alcoholic myopathy (Preedy et al., 2001; Simon et al., 2017).

Alcohol and nutrition

The nutritional status of alcoholics is often impaired (Rajendram and Preedy, 2008). Some of the pathophysiological changes seen in alcoholics are direct consequences of malnutrition (Rajendram and Preedy, 2008). However, in the 1960s and 70s, Charles Lieber demonstrated that many alcohol-induced pathologies, including alcoholic hepatitis, cirrhosis, and myopathy, are reproducible in animals fed a nutritionally adequate diet (Lieber et al., 1975). Consequently, the concept that all alcohol-induced pathologies are due to nutritional deficiencies is outdated and incorrect.

Myopathy is a direct consequence of alcohol or acetaldehyde on muscle and is not necessarily associated with malnutrition (Preedy et al., 2001; Preedy and Watson, 2004). Assessment of nutritional status in chronic alcoholics using anthropometric measures (e.g., limb circumference and muscle mass) may be misleading in the presence of myopathy (Rajendram and Preedy, 2008).

Acute or chronic ethanol administration impairs the absorption of several nutrients, including glucose, amino acids, biotin, folate, and ascorbic acid (Rajendram and Preedy, 2008). There is no strong evidence that alcohol impairs absorption of magnesium, riboflavin, or pyridoxine, so these deficiencies are probably due to poor intakes. Hepatogastrointestinal damage (e.g., villous injury, bacterial overgrowth of the intestine, pancreatic damage, or cholestasis) may impair the absorption of some nutrients such as the

fat-soluble vitamins (A, D, E, and K) (Rajendram and Preedy, 2008). In contrast, iron stores may be adequate as absorption can be increased (Rajendram and Preedy, 2008; Ribot-Hernández et al., 2020).

Effects of alcohol on the cardiovascular system

Alcohol affects both the heart and the peripheral vasculature. Acutely, alcohol causes peripheral vasodilatation, giving a false sensation of warmth that can be dangerous. Heat loss is rapid in cold weather or when swimming, but reduced awareness leaves people vulnerable to hypothermia. The main adverse effect of acute alcohol on the cardiovascular system is the induction of arrhythmias i.e., "Holiday Heart" (Piano, 2017). These are often harmless and experienced as palpitations but can rarely be fatal (Piano, 2017). Chronic ethanol consumption can cause systemic hypertension and congestive cardiomyopathy (Piano, 2017). Alcoholic cardiomyopathy accounts for up to one-third of dilated cardiomyopathies but may improve with abstinence or progress to death (Piano, 2017).

The potential for beneficial, cardioprotective effects with alcohol consumption have been broadcast widely (Wood et al., 2018). This observation was based on population studies of mortality due to ischemic heart disease, case–control studies, and animal experiments (Renaud and de Lorgeril, 1992; Wood et al., 2018). However, there is no evidence from randomized controlled trials. The apparent protective effect of alcohol may therefore result from confounding factors. For example, the diets of drinkers are different from those of nondrinkers. Even the diets of beer drinkers are different from those of wine drinkers. Furthermore, on the population level, the burden of alcohol-induced morbidity and mortality far outweighs any possible cardiovascular benefit (Department of Health, 2016; Wood et al., 2018).

Effects of alcohol on liver function

Fundamental to the effects of ethanol is the liver, where the majority of ethanol metabolism occurs (Utne and Winkler, 1980). Ethanol displaces many of the substrates usually metabolized in the liver. Metabolism of ethanol by ADH in the liver generates reducing equivalents (Kwo and Crabb, 2002; Höög and Ostberg, 2011). ALDH also generates NADH with conversion of acetaldehyde to acetate (Kwo and Crabb, 2002). The NADH:NAD+ ratio is increased, with a corresponding increase in the lactate:pyruvate ratio (McGuire et al., 2006). If lactic acidosis combines with a β-hydroxybutyrate predominant ketoacidosis, the blood pH can fall to 7.1 and hypoglycemia may occur (McGuire et al., 2006). Severe ketoacidosis and hypoglycemia can cause permanent brain damage (McGuire et al., 2006). However, in general, the prognosis of alcohol-induced acidosis is good. Ketosis also reduces the renal capacity for urate excretion (Yamamoto et al., 2005). Hyperuricemia is further exacerbated by alcohol-induced ketosis and acetate-mediated purine generation (Yamamoto et al., 2005). Hyperuricemia explains, at least in part, the clinical observation that alcohol misuse can precipitate gout.

The excess NADH is reported to promote fatty acid synthesis and inhibits lipid oxidation in the mitochondria, resulting in fat accumulation (Kong et al., 2019). Fatty changes within the liver are usually asymptomatic but can be seen on ultrasound or computed tomography scanning, and they are associated with abnormal liver toxicity tests (e.g., raised activities of serum g-glutamyl transferase, aspartate aminotransferase, and alanine transaminases) (Kong et al., 2019). The supposition that most of the hepatic damage in alcoholism is due to increases in the NADH:NAD ratio per se is somewhat outdated. Now, molecular and cellular processes in response to alcohol, and acetaldehyde toxicity, have been shown to be major contributors to the disease process (Rajendram et al., 2016b; Kong et al., 2019).

Progression to alcoholic steatohepatitis involves invasion of the liver by neutrophils with hepatocyte injury (Kong et al., 2019). The injured tissue may release damage-associated molecular patterns (DAMPs). These DAMPs activate immune cells, promoting liver fibrosis and metaplasia (i.e., cancer) (Kong et al., 2019). Thus, continued alcohol consumption may lead to cirrhosis. However, not all alcoholics progress to cirrhosis. The reason for this is unclear. It has been suggested that genetic factors and differences in immune response may play a role.

In alcoholic cirrhosis there is fibrocollagenous deposition, with scarring and disruption of surrounding hepatic architecture. There is ongoing necrosis with concurrent regeneration. Alcoholic cirrhosis is classically said to be micronodular, but often a mixed pattern is present. The underlying pathological mechanisms are complex and are the subject of debate. Induction of the MEOS and oxidation of ethanol by catalase result in free radical production (Halsted and Medici, 2012). Glutathione (a free radical scavenger) is reduced in alcoholics, impairing the ability to dispose of free radicals (Halsted and Medici, 2012). Mitochondrial damage occurs, limiting their capacity to oxidize fatty acids. Peroxisomal oxidation of fatty acids further increases free radical production. These changes eventually result in hepatocyte necrosis, and inflammation and fibrosis ensue. Acetaldehyde also contributes by promoting collagen synthesis and fibrosis (Rajendram et al., 2016b; Kong et al., 2019).

Effects of acetaldehyde

Acetaldehyde is highly toxic and can bind cellular constituents (e.g., proteins including CYP2E1, lipids, and nucleic acids) to produce harmful acetaldehyde adducts (Rajendram et al., 2016b). Adduct formation changes the structure and the biochemical

properties of the affected molecules (Rajendram et al., 2016b). The new structures may be recognized as foreign antigens by the immune system and initiate a damaging response.

Adduct formation leads to retention of protein within hepatocytes, contributing to the hepatomegaly, and several toxic manifestations, including impairment of antioxidant mechanisms (e.g., decreased glutathione (GSH); Rajendram et al., 2016b). Acetaldehyde thereby promotes free radical-mediated toxicity and lipid peroxidation (Rajendram et al., 2016b). Binding of acetaldehyde with cysteine (one of the three amino acids that comprise GSH) or GSH also reduces liver GSH content (Rajendram et al., 2016b). Acute ethanol administration inhibits GSH synthesis and increases losses from the liver (Rajendram et al., 2016b). Furthermore, mitochondrial GSH is selectively depleted and this may contribute to the marked disruption of mitochondria in alcoholic cirrhosis (Rajendram et al., 2016b).

Effects of acetate

The role of acetate in alcohol-induced pathology is not well understood. The uptake and utilization of acetate by tissues depend on the activity of acetyl-CoA synthetase (Cornier, 2004; Rajendram et al., 2016a; Bose et al., 2019). Acetyl-CoA and adenosine are produced from the metabolism of acetate. Acetate crosses the blood–brain barrier easily and is actively metabolized in the brain (Cornier, 2004; Rajendram et al., 2016a; Bose et al., 2019). Many of the central nervous system depressant effects of ethanol may be blocked by adenosine receptor blockers. Thus, acetate and adenosine may be important in the intoxicating effects of ethanol.

Ethanol increases portal blood flow, mainly by increasing GI tract blood flow. This effect is reproduced by acetate. Acetate also increases coronary blood flow, myocardial contractility, and cardiac output. Acetate inhibits lipolysis in adipose tissue and promotes steatosis in the liver. The reduced circulating free fatty acids (a source of energy for many tissues) may have significant metabolic consequences. Thus, some of the effects of alcohol may be due to acetate, though this area remains under explored.

Conclusion

The second most commonly used recreational drug worldwide; ethanol, can affect the physiology of every organ. When a beverage containing ethanol is consumed the ethanol is rapidly absorbed via the upper gastrointestinal tract. Ethanol is initially oxidized to acetaldehyde via one of three major enzyme pathways. This highly toxic metabolite is further oxidized to acetate. Any ethanol remaining in the blood after FPM is rapidly distributed to the organs. Ethanol and the products of its metabolism affect nearly every cellular structure or function. The physiological effects of ethanol correlate with the blood ethanol concentration after drinking an alcoholic beverage. However, the relationship between the amount of ethanol ingested and its acute, physiological effects are complex. Several factors are relevant including gender, body composition and the rate of gastric emptying. Recent GWAS have investigated polymorphisms in the enzyme systems that metabolize ethanol. However, further studies are required to understand the interplay between an individual's alcohol intake, their blood ethanol concentration and the resulting physiological effects of ethanol.

References

Abrahao, K.P., Salinas, A.G., Lovinger, D.M., 2017. Alcohol and the brain: neuronal molecular targets, synapses, and circuits. Neuron 96 (6), 1223–1238. https://doi.org/10.1016/j.neuron.2017.10.032.

Ballard, F.J., 1972. Supply and utilization of acetate in mammals. Am. J. Clin. Nutr. 25 (8), 773–779. https://doi.org/10.1093/ajcn/25.8.773.

Baraona, E., Abittan, C.S., Dohmen, K., Moretti, M., Pozzato, G., Chayes, Z.W., Schaefer, C., Lieber, C.S., 2001. Gender differences in pharmacokinetics of alcohol Alcohol. Clin. Exp. Res. 25 (4), 502–507. https://pubmed.ncbi.nlm.nih.gov/11329488/.

Besemer, F., Pereira, A.M., Smit, J.W., 2011. Alcohol-induced cushing syndrome. Hypercortisolism caused by alcohol abuse. Neth. J. Med. 69 (7), 318–323.

Bondoc, F.Y., Bao, Z., Hu, W.Y., Gonzalez, F.J., Wang, Y., Yang, C.S., Hong, J.Y., 1999. Acetone catabolism by cytochrome P450 2E1: studies with CYP2E1-null mice. Biochem. Pharmacol. 58 (3), 461–463. https://doi.org/10.1016/s0006-2952(99)00111-2.

Bose, S., Ramesh, V., Locasale, J.W., 2019. Acetate metabolism in physiology, cancer, and beyond. Trends Cell Biol. 29 (9), 695–703. https://doi.org/10.1016/j.tcb.2019.05.005.

Carbonnel, F., Lemann, M., Rambaud, J.C., Mundler, O., Jian, R., 1994. Effect of the energy density of a solid-liquid meal on gastric emptying and satiety. Am. J. Clin. Nutr. 60, 307–311.

Cordell, B.J., Kanodia, A., Miller, G.K., 2019. Case-control research study of auto-brewery syndrome. Global Adv. Health Med. 8. https://doi.org/10.1177/2164956119837566, 2164956119837566.

Cornier, M.-A., 2004. Disposal of ethanol carbon atoms. In: Preedy, V., Watson, R. (Eds.), Comprehensive Handbook of Alcohol Related Pathologies, 1st ed., 1. Elsevier, USA, pp. 103–110.

Crabb, D.W., 1997. First pass metabolism of ethanol: gastric or hepatic, mountain or molehill? Hepatology 25, 1292–1294.

de Fiebre, N.C., de Fiebre, C.M., Booker, T.K., Nelson, S., Collins, A.C., 1994. Bioavailability of ethanol is reduced in several commonly used liquid diets. Alcohol 11, 329–335.

Department of Health, 2016. Alcohol Guidelines Review—Report From the Guidelines Development Group to the UK Chief Medical Officers.

Di Padova, C., Worner, T.M., Julkunen, R.J.K., Lieber, C.S., 1987. Effects of fasting and chronic alcohol consumption on the first pass metabolism of ethanol. Gastroenterology 92, 1169–1173.

Dohmen, K., Baraona, E., Ishibashi, K., Pozzato, G., Moretti, M., Matsunaga, C., Fujimoto, K., Lieber, C.S., 1996. Ethnic differences in gastric sigma-alcohol dehydrogenase activity and ethanol first-pass metabolism. Alcohol Clin. Exp. Res. 20, 1569–1576.

Döring, W.K., Herzenstiel, M.N., Krampe, H., Jahn, H., Pralle, L., Sieg, S., Wegerle, E., Poser, W., Ehrenreich, H., 2003. Persistent alterations of vasopressin and N-terminal proatrial natriuretic peptide plasma levels in long-term abstinent alcoholics. Alcohol Clin. Exp. Res. 27 (5), 849–861. https://doi.org/10.1097/01.ALC.0000065433.17403.DE.

Edenberg, H.J., 2007. The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. Alcohol Res. Health 30, 5–13.
Fisher, H.R., Simpson, R.I., Kapur, B.M., 1987. Calculation of blood alcohol concentration (BAC) by sex, weight, number of drinks and time. Can. J. Public Health 78, 300–304.
Frezza, M., Di Padova, C., Pozzato, G., Terpin, M., Baraona, E., Lieber, C.S., 1990. High blood alcohol levels in women. The role of decreased gastric alcohol dehydrogenase activity and first-pass metabolism. N. Engl. J. Med. 322, 95–99.
Gentry, R.T., Baraona, E., Lieber, C.S., 1994. Agonist: gastric first pass metabolism of alcohol. J. Lab. Clin. Med. 123, 32–33.
Goist Jr., K.C., Sutker, P.B., 1985. Acute alcohol intoxication and body composition in women and men pharmacol. Biochem. Behav. 22, 811–814.
Halsted, C., Medici, V., 2012. Aberrant hepatic methionine metabolism and gene methylation in the pathogenesis and treatment of alcoholic steatohepatitis. Int. J. Hepatol. 2012, 959746. https://doi.org/10.1155/2012/959746.
Halsted, C.H., Robles, E.A., Mezey, E., 1973. Distribution of ethanol in the human gastrointestinal tract. Am. J. Clin. Nutr. 26, 831–834.
Heier, C., Xie, H., Zimmermann, R., 2016. Nonoxidative ethanol metabolism in humans-from biomarkers to bioactive lipids. IUBMB Life 68 (12), 916–923. https://doi.org/10.1002/iub.1569.
Höög, J.O., Ostberg, L.J., 2011. Mammalian alcohol dehydrogenases-a comparative investigation at gene and protein levels. Chem. Biol. Interact. 191, 2–7.
Jones, A.W., 1993. Disappearance rate of ethanol from the blood of human subjects: implications in forensic toxicology. J. Forensic Sci. 38, 104–118.
Jones, A.W., Jönsson, K.-Å., 1994. Food-induced lowering of blood-ethanol profiles and increased rate of elimination immediately after a meal. J. Forensic Sci. 39, 1084–1093.
Kalant, H., 2004. Effects of food and body composition on blood alcohol levels. In: Preedy, V.R., Watson, R.R. (Eds.), Comprehensive Handbook of Alcohol Related Pathology, vol. 1. Academic Press, London, pp. 87–102.
Kendrick, S.F., O'Boyle, G., Mann, J., Zeybel, M., Palmer, J., Jones, D.E., Day, C.P., 2010. Acetate, the key modulator of inflammatory responses in acute alcoholic hepatitis. Hepatology 51 (6), 1988–1997. https://doi.org/10.1002/hep.23572.
Kong, L.Z., Chandimali, N., Han, Y.H., Lee, D.H., Kim, J.S., Kim, S.U., Kim, T.D., Jeong, D.K., Sun, H.N., Lee, D.S., Kwon, T., 2019. Pathogenesis, early diagnosis, and therapeutic management of alcoholic liver disease. Int. J. Mol. Sci. 20, 2712. https://doi.org/10.3390/ijms20112712.
Konishi, M., Ishii, H., 2007. Role of microsomal enzymes in development of alcoholic liver diseases. J. Gastroenterol. Hepatol. 22 (Suppl. 1), S7–S10.
Kwo, P.Y., Crabb, D.W., 2002. Genetics of ethanol metabolism and liver disease. In: Sherman, D.I.N., Preedy, V.R., Watson, R.R. (Eds.), Ethanol and the Liver. Mechanisms and Management. Taylor and Francis, London, pp. 95–129.
Lai, D., Wetherill, L., Bertelsen, S., Carey, C.E., Kamarajan, C., Kapoor, M., Meyers, J.L., Anokhin, A.P., Bennett, D.A., Bucholz, K.K., Chang, K.K., De Jager, P.L., Dick, D.M., Hesselbrock, V., Kramer, J., Kuperman, S., Nurnberger Jr., J.I., Raj, T., Schuckit, M., Scott, D.M., et al., 2019. Genome-wide association studies of alcohol dependence, DSM-IV criterion count and individual criteria. Gene Brain Behav. 18 (6), e12579. https://doi.org/10.1111/gbb.12579.
Levitt, D.G., 2002. PKQuest: measurement of intestinal absorption and first pass metabolism - application to human ethanol pharmacokinetics. BMC Clin. Pharmacol. 2, 4.
Lieber, C.S., 2005. Alcohol metabolism: general aspects. In: Preedy, V.R., Watson, R.R. (Eds.), Comprehensive Handbook of Alcohol Related Pathology, vol. 1. Elsevier, Amsterdam, pp. 15–26.
Lieber, C.S., DeCarli, L., Rubin, E., 1975. Sequential production of fatty liver, hepatitis, and cirrhosis in sub-human primates fed ethanol with adequate diets. Proc. Natl. Acad. Sci. U. S. A. 72 (2), 437–441. https://doi.org/10.1073/pnas.72.2.437.
Marshall, A.W., Kingstone, D., Boss, M., Morgan, M.Y., 1983. Ethanol elimination in males and females: relationship to menstrual cycle and body composition. Hepatology 3, 701–706.
Matsumoto, H., Fukui, Y., 2002. Pharmacokinetics of ethanol: a review of the methodology. Addiction Biol. 7, 5–14.
McGuire, L.C., Cruickshank, A.M., Munro, P.T., 2006. Alcoholic ketoacidosis. Emerg. Med. J. 23 (6), 417–420. https://doi.org/10.1136/emj.2004.017590.
Moreno, A., Pares, X., 1991. Purification and characterization of a new alcohol dehydrogenase from human stomach. J. Biochem. 266, 1128–1133.
Morgan, M.Y., Ritson, E.B., 2003. Alcohol and Health: A Handbook for Students and Medical Practitioners, fourth ed. Medical Council on Alcohol, London.
Norberg, A., Jones, W.A., Hahn, R.G., Gabrielsson, J.L., 2003. Role of variability in explaining ethanol pharmacokinetics. Clin. Pharmacokinet. 42, 1–31.
Peng, G.S., Yin, J.H., Wang, M.F., Lee, J.T., Hsu, Y.D., Yin, S.J., 2002. Alcohol sensitivity in Taiwanese. Association 101, 769–774.
Piano, M.R., 2017. Alcohol's effects on the cardiovascular system. Alcohol Res. Curr. Rev. 38 (2), 219–241.
Preedy, V.R., Watson, R.R. (Eds.), 2004. Handbook of Alcohol-Related Pathology: Volumes 1–3. Academic Press, London.
Preedy, V.R., Adachi, J., Ueno, Y., Ahmed, S., Mantle, D., Mullatti, N., Rajendram, R., Peters, T.J., 2001. Alcoholic skeletal muscle myopathy: definitions, features, contribution of neuropathy, impact and diagnosis. Eur. J. Neurol. 8 (6), 677–687. https://doi.org/10.1046/j.1468-1331.2001.00303.x.
Rachdaoui, N., Sarkar, D.K., 2017. Pathophysiology of the effects of alcohol abuse on the endocrine system. Alcohol Res. Curr. Rev. 38 (2), 255–276.
Rajendram, R., Hunter, R., Preedy, V., 2013. Alcohol: absorption, metabolism, and physiological effects. In: Encyclopedia of Human Nutrition, Third. Academic Press, USA, pp. 40–49.
Rajendram, R., Preedy, V.R., 2008. The effect of chronic alcoholism on nutrition. In: Martin, C.R. (Ed.), Identification and Treatment of Alcohol Dependency. MandK Update Ltd, UK.
Rajendram, R., Rajendram, R., Preedy, V.R., 2016a. Ethanol metabolism and implications for disease. In: Preedy, V.R. (Ed.), Neuropathology of Drug Addictions and Substance Misuse. Elsevier, USA.
Rajendram, R., Rajendram, R., Preedy, V.R., 2016b. Acetaldehyde: a reactive metabolite. In: Preedy, V.R. (Ed.), Neuropathology of Drug Addictions and Substance Misuse. Elsevier, USA.
Ramchandani, V.A., Bostron, W.F., Li, T.K., 2001. Research advances in ethanol metabolism. Pathol. Biol. 49, 676–682.
Renaud, S., de Lorgeril, M., 1992. Wine, alcohol, platelets, and the French paradox for coronary heart disease. Lancet 339, 1523–1526.
Ribot-Hernández, I., Martín-González, C., Vera-Delgado, V., González-Navarrete, L., de Armas-González, J.F., Viña-Rodríguez, J., Sánchez-Pérez, M.J., Rodríguez-Gaspar, M., González-Reimers, E., 2020. Prognostic value of serum iron, ferritin, and transferrin in chronic alcoholic liver disease. Biol. Trace Elem. Res. 195 (2), 427–435. https://doi.org/10.1007/s12011-019-01887-0.
Roine, R.P., Gentry, R.T., Lim Jr., R.T., Baraona, E., Lieber, C.S., 1991. Effect of concentration of ingested ethanol on blood alcohol levels. Alcohol Clin. Exp. Res. 15, 734–738.
Roine, R.P., Gentry, R.T., Lim Jr., R.T., Heikkonen, E., Salaspuro, M., Lieber, C.S., 1993. Comparison of blood alcohol concentrations after beer and whiskey. Alcohol Clin. Exp. Res. 17, 709–711.
Simon, L., Jolley, S.E., Molina, P.E., 2017. Alcoholic myopathy: pathophysiologic mechanisms and clinical implications. Alcohol Res. Curr. Rev. 38 (2), 207–217.
Smith, G.D., Shaw, L.J., Maini, P.K., Ward, R.J., Peters, T.J., Murray, J.D., 1993. Mathematical modelling of ethanol metabolism in normal subjects and chronic alcohol misusers. Alcohol. Alcohol. 28, 25–32.
Stone, C.L., Thomasson, H.R., Bosron, W.F., Li, T.K., 1993. Purification and partial amino acid sequence of a high-activity human stomach alcohol dehydrogenase. Alcohol Clin. Exp. Res. 17, 911–918.
Tan, H.K., Yates, E., Lilly, K., Dhanda, A.D., 2020. Oxidative stress in alcohol-related liver disease World. J. Hepatol. 12 (7), 332349. https://doi.org/10.4254/wjh.v12.i7.332.
Utne, H.E., Winkler, K., 1980. Hepatic and extrahepatic elimination of ethanol in cirrhosis. With estimates of intrahepatic shunts and Km for ethanol elimination. Scand. J. Gastroenterol. 15 (3), 297–304. https://doi.org/10.3109/00365528009181473.
Wang, J.-C., Kapoor, M., Goate, A.M., 2012. The genetics of substance abuse. Annu. Rev. Genom. Hum. Genet. 13, 241–261.
Wood, A.M., Kaptoge, S., Butterworth, A.S., et al., 2018. Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599 912 current drinkers in 83 prospective studies. Lancet 391 (10129), 1513–1523.
World Health Organization, 2018. Legal BAC Limits by Country. In: Global Health Observatory Data Respository. Available at. https://apps.who.int/gho/data/view.main.54600. (Accessed 10 July 2021).

Yamamoto, T., Moriwaki, Y., Takahashi, S., 2005. Effect of ethanol on metabolism of purine bases (hypoxanthine, xanthine, and uric acid). Clin. Chim. Acta 356 (1–2), 35–57. https://doi.org/10.1016/j.cccn.2005.01.024.

Yang, F., Beard, D.A., 2006. Thermodynamically based profiling of drug metabolism and drug-drug metabolic interactions: a case study of acetaminophen and ethanol toxic interaction. Biophys. Chem. 120 (2), 121–134. https://doi.org/10.1016/j.bpc.2005.10.013.

Yasunami, M., Chen, C.S., Yoshida, A., 1991. A human alcohol dehydrogenase gene (ADH6) encoding an additional class of isozyme. Proc. Natl. Acad. Sci. U. S. A. 88, 7610–7614.

Zintzaras, E., Stefanidis, I., Santos, M., Vidal, F., 2006. Do alcohol-metabolizing enzyme gene polymorphisms increase the risk of alcoholism and alcoholic liver disease. Hepatology 43, 352–361.

Alcohol: Effects of consumption on diet and nutritional status

R. Rajendram, College of Medicine, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia; and Department of Medicine, King Abdulaziz Medical City, King Abdullah International Medical Research Center, Ministry of National Guard-Health Affairs, Riyadh, Saudi Arabia

© 2023 Elsevier Ltd. All rights reserved.

This is an update of C.H. Halsted, V. Medici, Alcohol: Effects of Consumption on Diet and Nutritional Status, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 50–56, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00007-6.

Introduction 267
Effects of alcohol consumption on dietary intake 267
The complex relationship between alcohol consumption and adverse outcomes 268
The risks of alcohol misuse 269
Alcoholic liver disease 269
Pancreatitis and pancreatic insufficiency 270
Heart 270
Neurological effects 270
Cancers 270
Anemia 270
Effects of chronic alcohol misuse on nutritional status 271
Body weight and energy balance 271
Micronutrient deficiencies in chronic alcohol misuse 271
Thiamine deficiency 271
Folate deficiency 272
Vitamin B12 deficiency 273
Pyridoxine deficiency 273
Deficiencies of fat-soluble vitamins (A and D) 273
Zinc deficiency 273
Iron deficiency 273
Conclusion 273
References 274

Key points

- Metabolism of ethanol provides 7.1 kcal g^{-1}
- A significant calorific load can be obtained from alcohol intake.
- In alcohol misusers, ethanol can displace dietary macro- and micronutrients, damage several organ systems and increase the risk of several cancers
- The relationship between ethanol intake and mortality consists of many, distinct, opposing, dose–response curves rather than a single J-shaped curve.

Glossary

Alcoholic liver disease This is characterized by "fatty liver" i.e., (steatosis; increased fat in the liver), inflammation, necrosis (death of liver cells), and fibrosis (scarring that leads to cirrhosis). It is most common in chronic alcohol misusers.

Chronic alcohol misuse Long-standing alcohol misuse. Excessive ethanol intake is typically daily may but may be episodic (e.g., at weekends).

Standard Drink A unit of ethanol consumption also known as a "unit of alcohol." In the United States of America (USA) a standard drink contains 14 g ethanol but the definition varies internationally. It is usually 12–15 g, as found in 12 oz of beer, 1.2 oz of spirits, and 4 oz of wine. Note that the actual drinks people pour generally contain more ethanol than standard drinks.

Moderate drinking Up to 2 drinks per day for men but 1 drink per day for women.

Nutritional consequences of alcohol misuse Alcohol intake and the and alcohol-induced disease result in weight loss, protein malnutrition and deficiencies of several micronutrients (e.g., vitamin A, thiamine, pyridoxine, folate, vitamin B12, vitamin D, zinc and iron).

Risks of chronic alcohol misuse Liver disease, chronic pancreatitis, heart disease, myopathy, anemia, and cognitive disorders.

Introduction

Ethanol is a highly addictive nutritional toxin. Its use is so ubiquitous that "drinking" usually describes consumption of beverages containing ethanol and "alcohol" and "ethanol" are effectively synonyms. The terms "standard drink" and "unit of alcohol" are also synonymous, but their definitions vary between countries. For example, in the United States of America (USA) a standard drink contains 14 g ethanol, while in the United Kingdom a unit of alcohol contains 8 g ethanol (Rajendram and Preedy, 2008). This may be confusing to the general public and complicates international comparisons of research and guidelines for "responsible" use of ethanol. So, for simplicity all references to standard drinks in this article will be based on the definition used in the USA.

The diet of active adult men should provide circa 11.7 MJ day^{-1} (2800 kcal day^{-1}) while active adult women require circa 10 MJ day^{-1} (2400 kcal day^{-1}; USA Department of Health and Human Services and USA Department of Agriculture, 2005). While the metabolism of ethanol provides 29 kJ g^{-1} (7.1 kcal g^{-1}), only miniscule amounts of other nutrients (e.g., proteins, vitamins, and minerals) are present (Rajendram and Preedy, 2008).

Light drinkers consuming 1 standard drink per day obtain 406 kJ day^{-1} (99.4 kcal day^{-1}) from ethanol alone. Furthermore, actual alcoholic beverages generally contain more ethanol than standard drinks. Thus, a significant calorific load can be obtained from alcohol intake. These additional calories can cause truncal obesity even in those who do not misuse alcohol (Tremblay et al., 1995).

Consuming 20 units of alcohol per week provides circa 2000 kcal $week^{-1}$ from ethanol alone. As a result of the high energy content of alcoholic beverages or because of associated socioeconomic factors and medical disorders, heavy drinkers often replace a substantial proportion of nutrient-derived calories with alcohol (Suter, 2001).

Moreover, consumed in excess, ethanol can displace dietary macro- and micronutrients and damages several organ systems (Rajendram and Preedy, 2008). Alcohol misuse contributes to generalized malnutrition (Perlow et al., 1977; Suter, 2001; Morgan and Ritson, 2003). It affects the availability and metabolism of both water- and fat-soluble vitamins including folate, thiamine, pyridoxine, and vitamins B12, A and D (Wu et al., 1975; Majumdar et al., 1982; Kanazawa and Herbert, 1985; Rajendram and Preedy, 2008; Halsed and Medici, 2011; Malham et al., 2011; Clugston and Blaner, 2012). The presence of alcoholic liver disease (ALD) magnifies the effects of alcohol misuse on nutrition (Halsted, 2004; Lieber, 2004, 2005; Halsed and Medici, 2011). This article reviews the effects of alcohol consumption with a focus on the effects of ethanol intake on nutritional status in *Homo sapiens*.

Effects of alcohol consumption on dietary intake

In the USA over 85% of those over 18 years old report that they have ever ingested ethanol (Substance Abuse and Mental Health Services Administration, 2019). Alcohol misuse is more common in young adults between 18 and 24 years of age than in any other age group (Substance Abuse and Mental Health Services Administration, 2019). While this group frequently present to hospital with complications of acute alcohol misuse (Hingson et al., 2009); organ damage as a result of chronic alcohol misuse generally manifests decades later.

Of those who ingest ethanol, most partake in moderation (Substance Abuse and Mental Health Services Administration, 2019). Moderate drinking can be defined as consuming no more than 2 standard drinks (i.e., units of alcohol) per day for men or 1 drink per day for women, where 1 standard drink contains 14 g of alcohol (USA Department of Agriculture and USA Department of Health and Human Services, 2020). Heavy drinking is defined as consuming more than 4 drinks on any given day per week for men or 3 drinks on any given day per week for women (National Institute on Alcohol Abuse and Alcoholism, 2021).

Chronic alcohol misusers are addicts who typically consume excessive amounts of alcohol on a daily basis. Binge drinkers are chronic alcohol misusers who episodically escalate their alcohol intake over weeks or months. Regardless of the style of heavy drinking, ethanol often displaces essential components from the diets of chronic alcohol misusers (Suter, 2001). Alcoholic beverages differ in their ethanol content, such that, wine contains approximately 12 g 100 mL^{-1}, beer contains approximately 4.5 g 100 mL^{-1} and spirits contain approximately 40 g 100 mL^{-1}. Thus, the ethanol content of 4 oz of wine, 12 oz of beer and 1.2 oz of spirits are similar.

The effects on dietary intake depend on the amount of ethanol consumed each day and overall changes in eating habits. Although ethanol contains 7.1 kcal g^{-1}, its metabolism to acetaldehyde in the liver is rapid (rates up to 50 g h^{-1}). None is stored as energy equivalents in the body. The metabolism of ethanol also affects fat and carbohydrate metabolism.

There are three main metabolic routes for the elimination of ethanol from the body (Rajendram et al., 2016; Kong et al., 2019). Alcohol dehydrogenase (ADH) is present in the cytosol of hepatocytes and metabolizes relatively low levels of alcohol that would be expected after moderate drinking (Rajendram et al., 2016). The metabolism of alcohol by ADH causes redox changes that contribute to hepatic lipid synthesis and reduce gluconeogenesis (Klop et al., 2013; Rajendram et al., 2016). Thus, even moderate

drinking can increase serum triglycerides and cause fatty liver (Klop et al., 2013). In the absence of dietary carbohydrate, excessive alcohol intake may result in hypoglycemia (i.e., low blood glucose levels). This can impair concentration and even consciousness.

The microsomal ethanol oxidizing system is the second group of enzymes involved in ethanol metabolism. The most important enzyme in this is cytochrome P450 2E1 (CYP2E1; Rajendram et al., 2016; Kong et al., 2019). The MEOS metabolizes alcohol at levels that may be generated by heavy drinking (Lieber, 2005; Rajendram et al., 2016). When metabolizing large amounts of ethanol, CYP2E1 utilizes adenosine triphosphate (ATP) "wasting" stored calories. This metabolic pathway, therefore, has the potential to cause weight loss.

Another form of ADH is present in the stomach (Rajendram et al., 2016). This is the first ethanol-metabolizing enzyme to encounter the alcohol contained in beverages. It accounts for approximately significantly more ethanol metabolism in men than in women (Baraona et al., 2001). This difference may explain why men's tolerance of alcohol is much higher than that of women (Baraona et al., 2001).

The complex relationship between alcohol consumption and adverse outcomes

Mortality increases progressively in those drinkers who consume more than 100 g week^{-1} (Wood et al., 2018). Nevertheless, in 1992, Serge Renaud and Michel de Lorgeril reported the "French Paradox." Despite similar diet and alcohol consumption; the cardiovascular mortality of the wine-drinking residents of the southern Mediterranean provinces of France is significantly lower than that in the northern provinces where wine is less frequently ingested (**Table 1**; Renaud and de Lorgeril, 1992). Subsequent population studies also identified increased mortality in abstainers. Thus, it is often suggested that the curve for alcohol-related mortality is J-shaped (Wood et al., 2018).

However, in 2018, an international study of almost 600,000 individuals who consume alcohol demonstrated that this concept oversimplifies the relationship between alcohol intake and cardiovascular disease risk (Wood et al., 2018). Higher alcohol intake was linearly associated with a higher risk of all types of stroke, coronary disease (except myocardial infarction), cardiac failure, and several other less common subtypes of cardiac disease (Wood et al., 2018). However, higher alcohol intake was associated with

Table 1 Risks and Benefits of alcohol consumption.

	Minimal amount or duration of drinks per day	*Mechanism*
Benefits		
Coronary disease protection	1–2 (women), 2–4 (men)	Flavonoid antioxidants
Cerebrovascular disease (nonhemorrhagic) protection		Elevated HDL lipoprotein reduced platelet adhesiveness
Risks		
Cancer	42 (women), 44 (men)	Unknown; higher risk in smoking alcohol misusers
Oropharynx and esophagus		
Breast (women)	42	Increases estrogen production
Colon	42 (women), 44 (men)	Initiation risk increases with low folate, proliferation risk increases with excessive folate
Alcoholic liver disease		
Fatty liver	42	Increased liver fat synthesis, decreased oxidation and export
Alcoholic hepatitis	43 (women) × 10 years 46 (men) × 15 years	Toxicity of alcohol metabolism Increased collagen synthesis
Alcoholic cirrhosis	43 (women) × 15 years 46 (men) × 20 years	
Pancreas		
Pancreatitis	B10 years	Acute inflammation of pancreas
Pancreatic insufficiency	B10–15 years	Loss of exocrine and endocrine pancreatic cells
Cardiomyopathy	Binge drinking	Mitochondrial damage of muscle cells or thiamine deficiency
Neurological effects		
Acute trauma, e.g., motor vehicle accidents	1–2 in social setting	Legal intoxication
Coma and death	10–20 in rapid succession	Severe toxicity
Withdrawal syndrome	Follows binge drinking	Neuronal hyperexcitability
Wernicke–Korsakoff syndrome	Unknown	Thiamine deficiency
Anemia	Unknown	Combination of iron, folate, and pyridoxal deficiencies

Source: Reproduced from Halsted, C.H., 2006. Alcohol: effects of consumption on diet and nutritional status. In: Caballero. B., Allen, L., Prentice, A.M. (Eds.), Encyclopedia of Human Nutrition, second ed. Elsevier, Amsterdam, pp. 62–69.

a lower risk of myocardial infarction. Furthermore, the threshold for lowest risk of all-cause mortality was about 100 g week^{-1} (Wood et al., 2018).

Thus, this recent, large, international cohort study revealed that it is more likely that the relationship between ethanol intake and mortality consists of many, distinct, opposing, dose–response curves rather than a single J-shaped curve (Wood et al., 2018). It also called for reduction of all current recommendations for "responsible" alcohol intake to less than 100 g week^{-1} (Wood et al., 2018).

The 1992 report describing the "French Paradox" attributed specific cardioprotective benefits to wine (Renaud and de Lorgeril, 1992). However, this was soon countered by in vitro data showing that the protective effect of wine on the oxidation of low-density lipoprotein could be mimicked by constitutive antioxidant flavonoids present in many fruits and vegetables (Zenebe et al., 2001). Another epidemiological study conducted by Holahan et al. (2012) concluded that, in comparison to non-wine drinkers, the lower mortality risk in wine drinkers could be due in large part to a better lifestyle (e.g., less smoking, more exercise, and better diet).

Any potential benefits of moderate drinking are confined to reductions in incidences of myocardial infarctions (Wood et al., 2018). Although red and white wine each contain protective antioxidant flavonoids, moderate amounts of alcohol may improve the circulating lipid profile by increasing levels of high-density lipoprotein and tissue plasminogen activator while reducing platelet function.

The risks of alcohol misuse

Chronic alcohol misuse affects many different organ systems. These include the liver, pancreas, heart, skeletal muscle and brain. Alcohol misuse also increases the risk of certain cancers (**Table 1**). Although these risks are apparent among the citizens of the USA who misuse alcohol, their prevalence is generally no less in countries such as France, Italy, and Spain (GBD 2016 Alcohol Collaborators, 2018) where partaking of wine with meals is embedded into the culture. The organ damage from chronic alcohol misuse (e.g., chronic liver and pancreatic disease) may impair nutrient assimilation and metabolism. However, the organ damage is also exacerbated by nutrient deficiencies (e.g., thiamine deficiency can impair brain function). The specific effects of alcohol abuse on select organs are described below. This provides a background for consideration of specific effects of alcohol on nutrition.

Alcoholic liver disease

Alcoholic liver disease (ALD) is one of the most common causes of mortality worldwide (GBD 2016 Alcohol Collaborators, 2018; Poznyak and Rekve, 2018). The pathogenesis of ALD is complex; several mechanisms have been implicated. Alcohol-induced translocation of bacterial lipopolysaccharide from the intestinal lumen through the portal vein initiates an inflammatory process in the liver (i.e., hepatitis) by activating tumor necrosis factor alpha (Kong et al., 2019). This cytokine promotes oxidative injury with necrosis of liver cells. It also has systemic effects including fever and anorexia with weight loss.

Several factors can initiate steatosis (i.e., increased lipid in the liver). These include the effects of alcohol on liver metabolism which promote lipid synthesis in hepatocytes. Other mechanisms reduce fatty acid oxidation and the export of lipid from the liver. Altered hepatic metabolism of methionine also contributes to the reduction of the antioxidant glutathione and apoptosis (i.e., cell death).

Fibrosis results from collagen synthesis by hepatic stellate cells (Kong et al., 2019). This is, in part, initiated by their incorporation of apoptotic liver cells as well as switching to collagen production. Of the three stages of ALD, fatty liver is related to the acute effects of alcohol on hepatic lipid metabolism. It is completely reversible with abstinence from alcohol (Kong et al., 2019). In contrast, alcoholic hepatitis usually occurs after a decade or more of chronic alcohol misuse. It is associated with steatosis and hepatitis (i.e., inflammation of the liver) with necrosis. It is associated with a high risk of mortality (approximately 40% within 6 months).

Alcoholic cirrhosis represents irreversible scarring of the liver secondary to fibrosis with loss of liver cells and function (Kong et al., 2019). It is usually preceded by alcoholic hepatitis. The scarring process greatly alters the circulation of blood through the liver in the sinusoids. It is associated with portal hypertension (i.e., increased blood pressure in the visceral (portal) circulation). As a result, blood flow is shunted away from the liver to other organs such as the esophagus.

The veins in these organs become dilated and tortuous (varicose). Potentially lethal complications of portal hypertension include hemorrhage due to rupture of esophageal varices and spontaneous bacterial peritonitis. Hepatic encephalopathy can also occur due to inadequate detoxification of waste products of metabolism (e.g., ammonia) in the visceral blood shunted around the scarred liver.

There is a clear dose response relationship between alcohol intake and ALD. The risk of developing alcoholic cirrhosis is dependent on the extent of exposure to alcohol regardless of the presence or absence of malnutrition (Lieber, 2004). However, while up to 95% of all chronic alcohol misusers develop fatty infiltration of the liver, only approximately 30% of heavy drinkers develop clinically significant ALD (Wood et al., 2018). This statistic indicates that genetic susceptibility is likely to be highly relevant in the pathogenesis of ALD (Wood et al., 2018).

Pancreatitis and pancreatic insufficiency

Pancreatitis is less common in chronic alcohol misusers than ALD (Samokhvalov et al., 2015). Acute pancreatic inflammation is characterized by attacks of severe epigastric abdominal pain. Chronic pancreatitis occurs after multiple bouts of acute pancreatitis. Pancreatitis damages exocrine pancreatic cells that secrete digestive enzymes and the endocrine cells which secrete insulin and other hormones. Thus, pancreatic insufficiency is a feature of chronic pancreatitis. The pancreas is scarred and the pancreatic ducts are distorted and partially blocked. This further limits secretion and promotes recurrent episodes of acute inflammatory pancreatitis.

The pancreas produces proteases and lipases. These enzymes are required for protein and lipid digestion. Thus, destruction of the pancreas results in significant malabsorption of these major dietary constituents (Rasmussen et al., 2013). Patients with pancreatic insufficiency exhibit severe loss of muscle protein and body fat (Rasmussen et al., 2013). The increased excretion of fat with the feces causes steatorrhea (foul smelling feces that float). The absorption of fat-soluble vitamins is dependent on pancreatic lipase for solubilization of dietary fat. So, these patients are at risk of deficiencies of vitamins A, D, and E (Rasmussen et al., 2013). As insulin secretion is reduced Type 3c diabetes mellitus can also develop (Rasmussen et al., 2013).

Heart

While the effects of alcohol on the heart are complex, they have been extensively investigated (Piano, 2017). While the risk of myocardial infarction may be decreased by moderate alcohol consumption, alcohol misuse can impair cardiac function (Piano, 2017). Episodic heavy drinking can cause arrhythmias with potential for sudden death ("holiday heart" syndrome; Piano, 2017). Chronic alcohol misusers are prone to left-sided heart failure secondary to mitochondrial dysfunction. This could be mediated by abnormal fatty acid metabolism in cardiac myocytes (Piano, 2017). A specific form of high-output heart failure, or "wet beriberi," is caused by thiamine deficiency. This is described in more detail in the Section on Micronutrient Deficiencies in Chronic Alcohol misuse.

Neurological effects

Acute and chronic alcohol misuse have a broad range of adverse neurological effects. These can be broadly categorized as being directly to alcohol, secondary to ALD, or due to thiamine deficiency. The variable effects of alcohol on the brain are related to several factors including the duration and amount of drinking, the age when drinking started, malnutrition, genetic background, and family history of alcohol misuse (Sachdeva et al., 2016). The stages of acute alcohol toxicity worsen from legal intoxication with reduced reaction time and judgment (blood alcohol levels $>0.08\ g\ dL^{-1}$ that usually define legal intoxication), to coma and death ($>0.35\ g\ dL^{-1}$) (Morgan and Ritson, 2003). Mild intoxication is common with social drinking (Morgan and Ritson, 2003). Yet, coma and death have been described in young men who consume excessive amounts of alcohol within a very short period of time (Hingson et al., 2009).

More common in inebriated pedestrians than drunk drivers; automobile accidents account for a large portion of alcohol-related mortality (GBD 2016 Alcohol Collaborators, 2018). Intoxication also causes falls and head trauma. Subdural hematoma presents with delayed but progressive loss of cognition and headaches but may cause death.

Chronic alcohol misusers are prone to symptoms of alcohol withdrawal. This is characterized by stages of tremulousness, seizures, and delirium tremens with hyperexcitability and hallucinations at any time up to 6 days from the last drink (Jesse et al., 2017). Delirium tremens is associated with progressive slowing of cerebral functions with stages of confusion, loss of cognition, and eventual coma and death (Jesse et al., 2017). Progressive impairment of cognition and judgment can also result from cerebral atrophy (Sachdeva et al., 2016). This generally occurs after years of heavy drinking but may also be exacerbated by thiamine deficiency (described in greater detail in the Section on Micronutrient Deficiencies in Chronic Alcohol Misuse; Sachdeva et al., 2016). These states of cognitive dysfunction are distinct from hepatic encephalopathy in chronic ALD (described above).

Cancers

Chronic alcohol misusers are at increased risk of cancers of the oropharynx and esophagus, colon, and breast (Boffetta and Hashibe, 2006). The risk of oropharyngeal cancer is greatest when heavy smoking is combined with excessive daily alcohol intake. Increased risk of squamous cell cancer of the esophagus is also compounded by smoking and may be associated with deficiencies of vitamin A and zinc (Boffetta and Hashibe, 2006). Breast cancer in women may be mediated by increased estrogen production during heavy alcohol intake (Boffetta and Hashibe, 2006).

Anemia

The causes of anemia in chronic alcohol misusers are legion. These include iron deficiency secondary to occult bleeding (e.g., episodic gastritis or bleeding from other sites of the gastrointestinal tract), folate deficiency from inadequate diet, malabsorption, and increased renal excretion of folic acid, and deficiency of pyridoxine (vitamin B6) due to abnormal effects on its metabolism

(Savage and Lindenbaum, 1986). Consequently, the bone marrow may demonstrate absent iron and mixtures of megaloblastosis from folate deficiency and sideroblastosis from pyridoxine deficiency (Savage and Lindenbaum, 1986).

Effects of chronic alcohol misuse on nutritional status

Body weight and energy balance

The effects of alcohol misuse on body weight depend on the timing and amount of alcohol consumption in relation to meals and on the presence or absence of organ damage, in particular, ALD (**Table 2**) (Halsted, 2004; Rajendram and Preedy, 2008; Halsed and Medici, 2011). Although body weight is usually unaffected by moderate alcohol intake, chronic alcohol misusers who drink daily, substituting alcohol for other dietary constituents, lose weight as dietary alcohol is energy-neutral.

However, alcohol decreases control over eating. Thus, moderate drinkers on weight loss regimens who take alcohol with their meals are less likely to lose weight. At the same time, those who consume alcohol with high-fat meals are more likely to gain weight due to an acute effect of alcohol on reducing lipid metabolism and promoting the storage of fat (Tremblay et al., 1995; Feinman and Lieber, 1999).

Alcoholic liver disease has significant impact on body composition and energy balance (Halsted, 2004; Lieber, 2004). Large multicenter studies, demonstrate universal evidence of protein-calorie malnutrition in patients with alcoholic hepatitis (Fung and Pyrsopoulos, 2017). This is based on the physical findings of muscle wasting and edema, low levels of serum albumin and other visceral proteins, and decreased cell-mediated immunity (Fung and Pyrsopoulos, 2017). Their short- and long-term mortality is related to the severity of malnutrition (Fung and Pyrsopoulos, 2017).

Anorexia is a major cause of weight loss in ALD. Furthermore, active alcoholic hepatitis contributes to increased resting energy expenditure (Fung and Pyrsopoulos, 2017). However, resting energy expenditure is normal in stable alcoholic cirrhotic patients who are also typically underweight or malnourished in part due to preferential metabolism of endogenous fat stores. At the same time, the digestion of dietary fat and the absorption of fat-soluble vitamins A, D, E, and K are decreased in cirrhotics as secretion of bile salts and pancreatic enzymes is diminished (Rasmussen et al., 2013; Fung and Pyrsopoulos, 2017).

Micronutrient deficiencies in chronic alcohol misuse

Chronic exposure to excessive amounts of ethanol is associated with deficiencies of several micronutrients. These include thiamine, folate, pyridoxine, vitamin A, vitamin D, zinc, and iron (Wu et al., 1975; Majumdar et al., 1982; Kanazawa and Herbert, 1985; Rajendram and Preedy, 2008; Halsed and Medici, 2011; Malham et al., 2011; Clugston and Blaner, 2012; Skalny et al., 2018; Ribot-Hernández et al., 2020). The frequency of these deficiencies is increased in the presence of ALD. This decreases the numbers of hepatocytes available for storage of vitamins and metabolism. Many clinical signs of ALD are related to vitamin deficiencies.

Thiamine deficiency

Low circulating levels of thiamine have been described in the majority of patients with alcoholic cirrhosis (Majumdar et al., 1982; Halsted, 2004; Chandrakumar et al., 2018). Thiamine pyrophosphate is a coenzyme in the intermediary metabolism of carbohydrates, in particular for transketolases that play a role in cardiac and neurological functions (Chandrakumar et al., 2018). Alcoholic beverages are essentially devoid of thiamine and acute exposure to alcohol decreases the activity of intestinal transporters required for thiamine absorption.

Table 2 Effects of alcohol on body weight.

Drinking behavior	*Explanation*
Moderate drinking	
Reduce weight	Substitution of carbohydrate by alcohol; more likely in women
Increase weight	Decreased dietary restraint
Heavy drinking	
Reduce weight	Substitution of nonalcohol calories by alcohol calories, which are “wasted” during metabolism
Increase weight	Alcohol metabolism decreases lipid metabolism in the liver and promotes fat storage

Source: Reproduced from Halsted, C.H., 2006. Alcohol: effects of consumption on diet and nutritional status. In: Caballero, B., Allen, L., Prentice, A.M. (Eds.), Encyclopedia of Human Nutrition, second ed. Elsevier, Amsterdam, pp. 62–69.

The major neurological signs and symptoms of thiamine deficiency include peripheral neuropathy, partial paresis of ocular muscles, wide-based gait secondary to cerebellar lesions, cognitive defects, and severe memory loss (Chandrakumar et al., 2018). The presence of peripheral neuropathy is sometimes referred to as "dry beriberi." The other symptoms constitute the Wernicke–Korsakoff syndrome. Although abnormal eye movements can be treated acutely by thiamine injections, the other complications are often permanent and contribute to the dementia that often afflicts alcohol misusers after years of drinking (Sachdeva et al., 2016; Chandrakumar et al., 2018).

"Wet beriberi" is a high-output cardiac failure that can also occur in thiamine-deficient alcohol misusers (Piano, 2017). It improves with thiamine in addition to conventional treatment for heart failure.

As endogenous thiamine is used during carbohydrate metabolism, acute symptoms can be precipitated by the administration of intravenous glucose to malnourished and marginally thiamine-deficient patients by depletion of remaining thiamine stores (Chandrakumar et al., 2018). This can be prevented by the intravenous addition of water-soluble vitamins including thiamine to malnourished chronic alcohol misusers who receive treatment for medical emergencies (**Table 3**).

Folate deficiency

Folates, are a family of B vitamins with folic acid at its core. They are important in DNA synthesis and cell turnover and play a central role in methionine metabolism in the liver (Halsed and Medici, 2011). Originally recognized as a cause of megaloblastic anemia, the expanding known sequelae of folate deficiency are related to hyperhomocysteinemia (raised blood levels of homocysteine). These include increased risk of neural tube defects and other congenital abnormalities in newborns as well as altered cognition in the elderly people. Low serum folate levels in chronic alcohol misusers was common before grains were fortified with folate (Wu et al., 1975; Majumdar et al., 1982). There are no data on the incidence of folate deficiency in chronic alcohol misusers post-fortification.

Megaloblastic anemia, due to the negative effects of folate deficiency on DNA synthesis, has been described in approximately one-third of chronic alcohol misusers (Wu et al., 1975; Savage and Lindenbaum, 1986). Furthermore, folate deficiency may play a role in the pathogenesis of ALD by reducing hepatic levels of S-adenosylmethionine with consequent reduction in antioxidant glutathione and DNA methylation with resultant increased activation of genes relevant to alcoholic liver injury (Halsted and Medici, 2011).

There are multiple causes of folate deficiency in chronic alcohol misuse. All alcoholic beverages (except beer) are devoid of folate. The typical diet of the binge-drinking chronic alcohol misuser does not include fresh vegetable sources and fortified grains. Chronic alcohol misuse causes intestinal folate malabsorption, decreased liver folate uptake, and accelerated folate excretion in the urine (Halsted, 2004; Halsted and Medici, 2011). In addition, ALD decreases liver folate stores, so the time to development of folate deficiency with marginal diet is shortened (Halsted, 2004; Halsted and Medici, 2011).

Table 3 Common micronutrient deficiencies in chronic alcohol misusers.

Deficiency	*Cause*	*Effect*
Thiamine	Poor diet Intestinal malabsorption	Peripheral neuropathy Wernicke–Korsakoff syndrome high-output heart failure megaloblastic anemia
Folate	Poor diet Intestinal malabsorption decreased liver storage increased urine excretion	Hyperhomocysteinemia and liver disease neural tube defect Altered cognition peripheral neuropathy sideroblastic anemia
Vitamin B_6	Poor diet Displacement from circulating albumin promotes urine excretion	
Niacin	Poor diet	Pellagra with dermatitis, diarrhea, and dementia
Pantothenic acid	Poor diet	Paresthesias "burning feet" syndrome
Vitamin A	Malabsorption Increased biliary secretion	Night blindness May promote development of fibrosis in alcoholic liver disease calcium deficiency
Vitamin D	Malabsorption decreased sun exposure	Metabolic bone disease Night blindness
Zinc	Poor diet Increased urine excretion	Decreased taste Decreased immune function
Iron	Gastrointestinal bleeding	Anemia

Source: Reproduced from Halsted, C.H., 2006. Alcohol: effects of consumption on diet and nutritional status. In: Caballero, B., Allen, L., Prentice, A.M. (Eds.), Encyclopedia of Human Nutrition, second ed. Elsevier, Amsterdam, pp. 62–69.

Vitamin B12 deficiency

The incidence of vitamin B12 deficiency in chronic alcohol misuse is undefined. This is because, while B12 is decreased in the liver of patients with ALD, serum levels may be normal or even increased (Kanazawa and Herbert, 1985; Halsted and Medici, 2011). The uptake and retention of vitamin B12 by damaged hepatocytes is reduced (Kanazawa and Herbert, 1985; Halsted and Medici, 2011). As vitamin B12 is a cofactor for methionine synthase, low hepatic vitamin B12 may contribute to abnormal liver methionine metabolism with hyperhomocysteinemia (Halsted and Medici, 2011).

Pyridoxine deficiency

Pyridoxine (vitamin B6) is required for transamination. This is necessary for the elimination of homocysteine (Halsted and Medici, 2011). Pyridoxine deficiency in chronic alcohol misuse occurs because pyridoxal phosphate is displaced from hepatic protein binding sites by acetaldehyde (Halsted and Medici, 2011). This increases its urinary excretion. Low serum levels of pyridoxal phosphate are common in chronic alcohol misusers (Halsted and Medici, 2011). Pyridoxine deficiency manifests with peripheral neuropathy and sideroblastic anemia. In alcoholic hepatitis, measurements of serum alanine transaminase (ALT) activity is disproportionately low compared with aspartate transaminase (AST), as pyridoxine is required for ALT activity.

Deficiencies of fat-soluble vitamins (A and D)

Chronic alcohol misusers are at increased risk of metabolic bone disease due to vitamin D deficiency and the resulting calcium insufficiency. Alcoholic liver disease increases the risk of low circulating levels of 25-hydroxy vitamin D (Malham et al., 2011). This deficiency is multifactorial in etiology. While malabsorption of fat-soluble vitamins due to biliary dysfunction is relevant, other factors include impaired hydroxylation of vitamin D by the liver, poor diet, and impaired production in the skin (Malham et al., 2011).

Serum levels of vitamin A are usually normal in chronic alcohol misusers in the absence of organ dysfunction (Clugson and Blaner, 2012). However, liver retinoids progressively fall as ALD progresses (Clugson and Blaner, 2012). This reduction in hepatic retinoids occurs independent of intake and malabsorption is not a key player in its etiology (Clugson and Blaner, 2012). Other possible explanations include reduced hepatic uptake of retinoids from chylomicrons and reduced retinol transfer and storage (Clugson and Blaner, 2012).

Complications of vitamin A deficiency include night blindness and increased risk of several cancers (Clugson and Blaner, 2012). However, it is important to note that patients with ALD are very susceptible to hepatotoxicity from vitamin A. Thus, it has not yet been clarified whether supplementation of vitamin A is beneficial in this cohort (Clugson and Blaner, 2012).

Zinc deficiency

Zinc is a cofactor for many enzymes including retinol dehydrogenase. It is stored in the pancreas, in the circulation is mainly bound to albumin. Chronic alcohol misusers are frequently zinc deficient (Skalny et al., 2018). This is due to poor diet, pancreatic deficiency, and increased urine excretion because of hypoalbuminemia. The consequences of zinc deficiency include impaired cellular immunity. So, its deficiency may contribute to increased risk of infections in alcohol misusers (Skalny et al., 2018). Furthermore, zinc concentrations may reflect alcohol-induced metabolic dysfunction and supplementation should be considered in the treatment of alcohol-induced disease (Skalny et al., 2018).

Iron deficiency

Chronic alcohol misusers are often iron deficient but may also be iron overloaded. Iron deficiency may result from gastrointestinal bleeding, typically due to alcoholic gastritis or esophageal tears from frequent retching and vomiting, or from rupture of esophageal varices in patients with cirrhosis and portal hypertension. The major consequence of iron deficiency is anemia (Savage and Lindenbaum, 1986). This may be compounded by the concurrent effects of folate and pyridoxine deficiencies (Savage and Lindenbaum, 1986). On the other hand, hepatic iron content is increased by ethanol (Ribot-Hernández et al., 2020). The oxidant effects of iron in the liver stimulates synthesis of ferritin and activates stellate cells (Ribot-Hernández et al., 2020). This promotes inflammation and fibrosis, worsening ALD (Ribot-Hernández et al., 2020). Yet, while albumin and transferrin saturation index are independently related to mortality in this cohort, serum iron and ferritin are not (Ribot-Hernández et al., 2020).

Conclusion

The relationship between alcohol intake, nutritional status and alcohol-induced disease is complex. The calorific value of alcohol is high and can induce primary malnutrition in alcoholics by displacement of other nutrients from the diet. Secondary malnutrition

may result from the toxic effects of ethanol on several organs involved in the absorption and metabolism of micro- and macronutrients (Rajendram and Preedy, 2008). The ethanol-induced reduction of dietary intake, maldigestion and malabsorption often results in deficiency of one or more nutrients in alcohol misusers (Rajendram and Preedy, 2008). However, the effects of ethanol on nutrient metabolism affect even those alcoholics who are not malnourished, resulting, for example, in impaired protein synthesis (Rajendram and Preedy, 2008).

The investigation of the relationship between ethanol and nutrition has traditionally been hypothesis-driven. Importantly, this methodology has elucidated that some processes are not affected by either acute or chronic alcohol exposure. While hypothesis-driven research is clearly important, the application of holistic, or "omic," technologies could rapidly highlight significant pathways, processes or products (Rajendram and Preedy, 2008). In particular, genomics, proteomics and metabolomics could identify thousands of molecular and cellular targets involved in the interaction between ethanol, the nutritional status of alcohol misusers and alcohol-induced disease.

Until this complex relationship is better defined, while it is sensible to encourage alcohol misusers to consume balanced diets; the most important nutritional intervention is to encourage and maintain abstinence.

See Also: Nutrition and susceptibility to tuberculosis; Selenium; Vitamin A: Deficiency and interventions; Vitamin A: Physiology, dietary sources and requirements; Vitamin B_6; Vitamin D: Role in chronic and acute diseases

References

Baraona, E., Abittan, C.S., Dohmen, K., Moretti, M., Pozzato, G., Chayes, Z.W., Schaefer, C., Lieber, C.S., 2001. Gender differences in pharmacokinetics of alcohol. Alcohol Clin. Exp. Res. 25, 502–507.

Boffetta, P., Hashibe, M., 2006. Alcohol and cancer. Lancet Oncol. 7 (2), 149–156. https://doi.org/10.1016/S1470-2045(06)70577-0.

Chandrakumar, A., Bhardwaj, A., 't Jong, G.W., 2018. Review of thiamine deficiency disorders: Wernicke encephalopathy and Korsakoff psychosis. J. Basic Clin. Physiol. Pharmacol. 30 (2), 153–162. https://doi.org/10.1515/jbcpp-2018-0075.

Clugston, R.D., Blaner, W.S., 2012. The adverse effects of alcohol on vitamin A metabolism. Nutrients 4 (5), 356–371. https://doi.org/10.3390/nu4050356.

Feinman, L., Lieber, C.S., 1999. Ethanol and lipid metabolism. Am. J. Clin. Nutr. 70 (5), 791–792. https://doi.org/10.1093/ajcn/70.5.791.

Fung, P., Pyrsopoulos, N., 2017. Emerging concepts in alcoholic hepatitis. World J. Hepatol. 9 (12), 567–585. https://doi.org/10.4254/wjh.v9.i12.567.

GBD 2016 Alcohol Collaborators, 2018. Alcohol use and burden for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 392 (10152), 1015–1035.

Halsted, C.H., 2004. Nutrition and alcoholic liver disease. Semin. Liver Dis. 24, 289–304.

Halsted, C.H., Medici, V., 2011. Vitamin-dependent methionine metabolism and alcoholic liver disease. Adv. Nutr. 2 (5), 421–427. https://doi.org/10.3945/an.111.000661.

Hingson, R.W., Zha, W., Weitzman, E.R., 2009. Magnitude of and trends in alcohol-related mortality and morbidity among U.S. college students ages 18–24, 1998–2005. J. Stud. Alcohol Drugs Suppl. 16, 12–20.

Holahan, C.J., Schutte, K.K., Brennan, P.L., North, R.J., Holahan, C.K., Moos, B.S., Moos, R.H., 2012. Wine consumption and 20-year mortality among late-life moderate drinkers. J. Stud. Alcohol Drugs 73 (1), 80–88. https://doi.org/10.15288/jsad.2012.73.80.

Jesse, S., Bråthen, G., Ferrara, M., Keindl, M., Ben-Menachem, E., Tanasescu, R., Brodtkorb, E., Hillbom, M., Leone, M.A., Ludolph, A.C., 2017. Alcohol withdrawal syndrome: mechanisms, manifestations, and management. Acta Neurol. Scand. 135 (1), 4–16. https://doi.org/10.1111/ane.12671.

Kanazawa, S., Herbert, V., 1985. Total corrinoid, cobalamin (vitamin B12), and cobalamin analogue levels may be normal in serum despite cobalamin in liver depletion in patients with alcoholism. Lab. Invest. 53 (1), 108–110.

Klop, B., do Rego, A.T., Cabezas, M.C., 2013. Alcohol and plasma triglycerides. Curr. Opin. Lipidol. 24 (4), 321–326. https://doi.org/10.1097/MOL.0b013e3283606845. PMID: 23511381.

Kong, L.Z., Chandimali, N., Han, Y.H., Lee, D.H., Kim, J.S., Kim, S.U., Kim, T.D., Jeong, D.K., Sun, H.N., Lee, D.S., Kwon, T., 2019. Pathogenesis, early diagnosis, and therapeutic management of alcoholic liver disease. Int. J. Mol. Sci. 20 (11), 2712. https://doi.org/10.3390/ijms20112712.

Lieber, C.S., 2004. Alcoholic fatty liver: its pathogenesis and mechanism of progression to inflammation and fibrosis. Alcohol 34, 9–19.

Lieber, C.S., 2005. Alcohol metabolism: general aspects. In: Preedy, V.R., Watson, R.R. (Eds.), Comprehensive Handbook of Alcohol Related Pathology, vol. 1. Elsevier, UK, Amsterdam, pp. 15–26.

Majumdar, S.K., Shaw, G.K., O'Gorman, P., Aps, E.J., Offerman, E.L., Thomson, A.D., 1982. Blood vitamin status (B1, B2, B6, folic acid and B12) in patients with alcoholic liver disease. Int. J. Vitam. Nutr. Res. 52 (3), 266–271. Internationale Zeitschrift fur Vitamin- und Ernahrungsforschung. Journal international de vitaminologie et de nutrition.

Malham, M., Jørgensen, S.P., Ott, P., Agnholt, J., Vilstrup, H., Borre, M., Dahlerup, J.F., 2011. Vitamin D deficiency in cirrhosis relates to liver dysfunction rather than aetiology. World J. Gastroenterol. 17 (7), 922–925. https://doi.org/10.3748/wjg.v17.i7.922.

Morgan, M.Y., Ritson, E.B., 2003. Alcohol and Health: A Handbook for Students and Medical Practitioners, fourth ed. Medical Council on Alcohol, London.

National Institute on Alcohol Abuse and Alcoholism (NIAAA), 2021. Drinking Levels Defined. NIAAA Alcohol and Your Health website. https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking. (Accessed 8 June 2021).

Perlow, W., Baraona, E., Lieber, C.S., 1977. Symptomatic intestinal disaccharidase deficiency in alcoholics. Gastroenterology 72, 680–684.

Piano, M.R., 2017. Alcohol's effects on the cardiovascular system. Alcohol Res. 38 (2), 219–241.

Poznyak, V., Rekve, D., 2018. Management of Substance Abuse Team, Global Status Report on Alcohol and Health 2018. World Health Organization, Geneva, Switzerland.

Rajendram, R., Preedy, V.R., 2008. The effect of chronic alcoholism on nutrition. In: Martin, C.R. (Ed.), Identification and Treatment of Alcohol Dependency. MandK Update Ltd., UK.

Rajendram, R., Rajendram, R., Preedy, V.R., 2016. Ethanol metabolism and implications for disease. In: Preedy, V.R. (Ed.), Neuropathology of Drug Addictions and Substance Misuse. Elsevier, USA.

Rasmussen, H.H., Irtun, O., Olesen, S.S., Drewes, A.M., Holst, M., 2013. Nutrition in chronic pancreatitis. World J. Gastroenterol. 19 (42), 7267–7275. https://doi.org/10.3748/wjg.v19.i42.7267.

Renaud, S., de Lorgeril, M., 1992. Wine, alcohol, platelets, and the French paradox for coronary heart disease. Lancet 339, 1523–1526.

Ribot-Hernández, I., Martín-González, C., Vera-Delgado, V., González-Navarrete, L., de Armas-González, J.F., Viña-Rodríguez, J., Sánchez-Pérez, M.J., Rodríguez-Gaspar, M., González-Reimers, E., 2020. Prognostic value of serum iron, ferritin, and transferrin in chronic alcoholic liver disease. Biol. Trace Elem. Res. 195 (2), 427–435. https://doi.org/10.1007/s12011-019-01887-0.

Sachdeva, A., Chandra, M., Choudhary, M., Dayal, P., Anand, K.S., 2016. Alcohol-related dementia and neurocognitive impairment: a review study. Int. J. High Risk Behav. Addict. 5 (3), e27976. https://doi.org/10.5812/ijhrba.27976.

Savage, D., Lindenbaum, J., 1986. Anemia in alcoholics. Medicine 65, 322–338.
Skalny, A.V., Skalnaya, M.G., Grabeklis, A.R., Skalnaya, A.A., Tinkov, A.A., 2018. Zinc deficiency as a mediator of toxic effects of alcohol abuse. Eur. J. Nutr. 57 (7), 2313–2322. https://doi.org/10.1007/s00394-017-1584-y.
Samokhvalov, A.V., Rehm, J., Roerecke, M., 2015. Alcohol consumption as a risk factor for acute and chronic pancreatitis: a systematic review and a series of meta-analyses. eBioMedicine 2 (12), 1996–2002. https://doi.org/10.1016/j.ebiom.2015.11.023.
Substance Abuse and Mental Health Services Administration, 2019. Center for Behavioral Health Statistics and Quality. National Survey on Drug Use and Health. https://www.samhsa.gov/data/sites/default/files/reports/rpt29394/NSDUHDetailedTabs2019/NSDUHDetTabsSect2pe2019.htm. (Accessed 8 June 2021).
Suter, P.M., 2001. Alcohol. its role in health and nutrition. In: Bowman, B.A., Russell, R.M. (Eds.), Present Knowledge in Nutrition, third ed. International Life Science Institute Press, Washington, D.C., pp. 497–507
Tremblay, A., Buemann, B., Theriault, G., Bouchard, C., 1995. Body fatness in active individuals reporting low lipid and alcohol intake. Eur. J. Clin. Nutr. 49, 824–831.
U.S. Department of Agriculture and U.S. Department of Health and Human Services, December 2020. Dietary Guidelines for Americans, 2020–2025, ninth ed. DietaryGuidelines.gov. Available at. (Accessed 8 June 2021).
U.S. Department of Health and Human Services and U.S. Department of Agriculture, 2005. Dietary Guidelines for Americans 2005. U.S. Government Printing Office, Washington.
Wood, A.M., Kaptoge, S., Butterworth, A.S., et al., 2018. Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599,912 current drinkers in 83 prospective studies. Lancet 391 (10129), 1513–1523.
Wu, A., Chanarin, I., Slavin, G., Levi, A.J., 1975. Folate deficiency in the alcoholic—its relationship to clinical and haematological abnormalities, liver disease and folate stores. Br. J. Haematol. 29 (3), 469–478. https://doi.org/10.1111/j.1365-2141.1975.tb01844.x.
Zenebe, W., Pechánová, O., Bernátová, I., 2001. Protective effects of red wine polyphenolic compounds on the cardiovascular system. Exp. Clin. Cardiol. 6 (3), 153–158.

Beverages and health

Liwei Chen, Department of Epidemiology, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of L. Chen, Beverages and Health, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 142–148, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00023-4.

Introduction 277
Coffee and tea 278
Tea 278
Coffee 280
Milk 281
Animal milk 281
Soy milk 281
100% fruit or vegetable juice 282
Fruit juice 282
Vegetable juice 282
Soft drinks 282
Diet beverages 283
Sugar-sweetened beverages 283
Summary 283
References 284
Further reading 284

Key points

- Define beverages and beverages categories
- Provide nutrient composition of commonly consumed beverages
- Provide a review of consumption history, levels and patterns of various beverage
- Provide a review of nutritional values of beverages by category
- Provide a review of health benefits and risks of beverage consumption by category
- Summarize the current recommendations for beverage consumption based on available scientific evidence

List of abbreviations

CHD Coronary heart disease
CVD Cardiovascular diseases
EGCG Epigallocatechin-3-gallate
HFCS High-fructose corn syrup
IGF-1 Insulin like growth factor I
LDL Low-density lipoprotein
NAS Non-caloric artificial sweeteners
ORAC Oxygen Radical Absorbance Capacity
RCT Randomized controlled trial
US United States
UK United Kingdom
WCRF World Cancer Research Fund

Introduction

Beverages, usually excluding drinking water, are referred to as fluids which are prepared for human consumption. In this article, we focus on nonalcoholic beverages. Water and alcohol are reviewed in other articles. For clarification, definitions for different beverage categories and other terminologies that are discussed in this article are provided in **Table 1**.

In history, water and breast milk are the only beverages that were consumed by humans until 11,000 and 12,000 years ago. It is believed that other beverages appeared in the human diet no more than 11,000 years ago. **Fig. 1** provides a timeline of when major beverages entered the human food chain (Popkin, 2010). With industrialization, more and more beverages have been produced and introduced into the human diet. Today, beverages are playing very important roles in the human diet and health.

Currently, a large proportion of the world's population is consuming a significant amount of beverages on a daily basis. Globally, the consumption of packed beverages in 2019 was about 224 billion liters for carbonated soft drinks, 260 billion liters for milk and dairy drinks, and 86 billion liters for fruit and vegetable juice. In the United States (US), water accounted for 51% of total nonalcoholic beverage consumption among adults on a given day, followed by coffee (15%), sugar-sweetened beverages (10%), tea (9%), fruit beverages (6%), milk (5%), and diet beverages (4%) based a national survey. Beverages share the same feature as fluids but vary tremendously in their energy content and nutrient composition. The nutrient composition of commonly consumed beverages is provided in **Table 2**. Given current consumption levels, these beverages could contribute significant amounts of calories as well as sugars, caffeine, and several other vitamins and minerals to our diet. In 2015–2016, beverages provided about 17% of total calories, 50% of added sugar intake, 1/3 of vitamin C intake (60% from 100% juice), and 4/1 of calcium (80% from milk) intake for the general American population. Sugar-sweetened beverages are the leading source of added sugars and calories in the Americans' diet, responsible for 1/3 of the total added sugar consumed by Americans. In 2011–2014, 63% of youth and 49% of adults drank sugar-sweetened beverages on a given day. On average, US youth consume 143 calories and US adults consume 145 calories from sugar-sweetened beverages on a given day. The United Kingdom (UK) and many other developed counties are heading in the same direction.

Beverages are important sources of fluids in our diet. There is no doubt that we need fluid to provide adequate water for maintaining normal physiological function and metabolic homeostasis of the human body. It has been estimated that about 20% of the water we need is derived from food, and the other 80% comes from drinking water and beverages. One fundamental question we should ask is: In addition to fluids, do adults need beverages to provide energy and other nutrients? The answer is no. Historically, adults consumed predominantly water which provided no calories. Scientifically, no evidence suggests that a healthy diet for adults relies on beverages to provide energy or other nutrients. Drinking water can provide the fluids our body needs as long as we maintain a healthy, balanced diet.

One of the major concerns regarding such a large amount of calories from beverages is that beverages are less satiating than solid foods. Evidence suggests that compensation of energy consumed in a liquid form (liquid calories) may be less complete than energy consumed in a solid form in humans. In other words, consumption of calorie-rich beverages does not reduce the intake of solid foods proportionally to maintain the energy balance. Therefore, when a large amount of calorie-rich beverages are consumed,

Table 1 Definitions of beverages.

Liquid calories	*All calories derived from beverages*
Fruit drinks	Sugar-sweetened drink with a small percentage of fruit juice or juice flavoring
100% juices	Beverages composed exclusively of liquids extracted from fruits and vegetables with no added caloric sweeteners
Milk	A translucent white liquid produced by the mammary glands of mammals.
100% juices	Beverages composed exclusively of liquids extracted from fruits and vegetables with no added caloric sweeteners or flavorings
Soft drinks	Non-alcoholic carbonated beverages containing caloric and artificial sweeteners, including regular soft drinks and diet soft drinks
Diet beverages (diet soft drinks)	Carbonated or non-carbonated beverages sweetened with artificial sweeteners
Sugar-sweetened beverages (regular soft drinks)	Carbonated or non-carbonated beverages containing added caloric sweeteners
Added caloric sweeteners	All composite sugars added to foods and beverages during production and processing, including sucrose, high-fructose corn syrup, honey, molasses, and other syrups.
Artificial sweeteners	Synthetic food additives that have sweet taste but no or less energy content

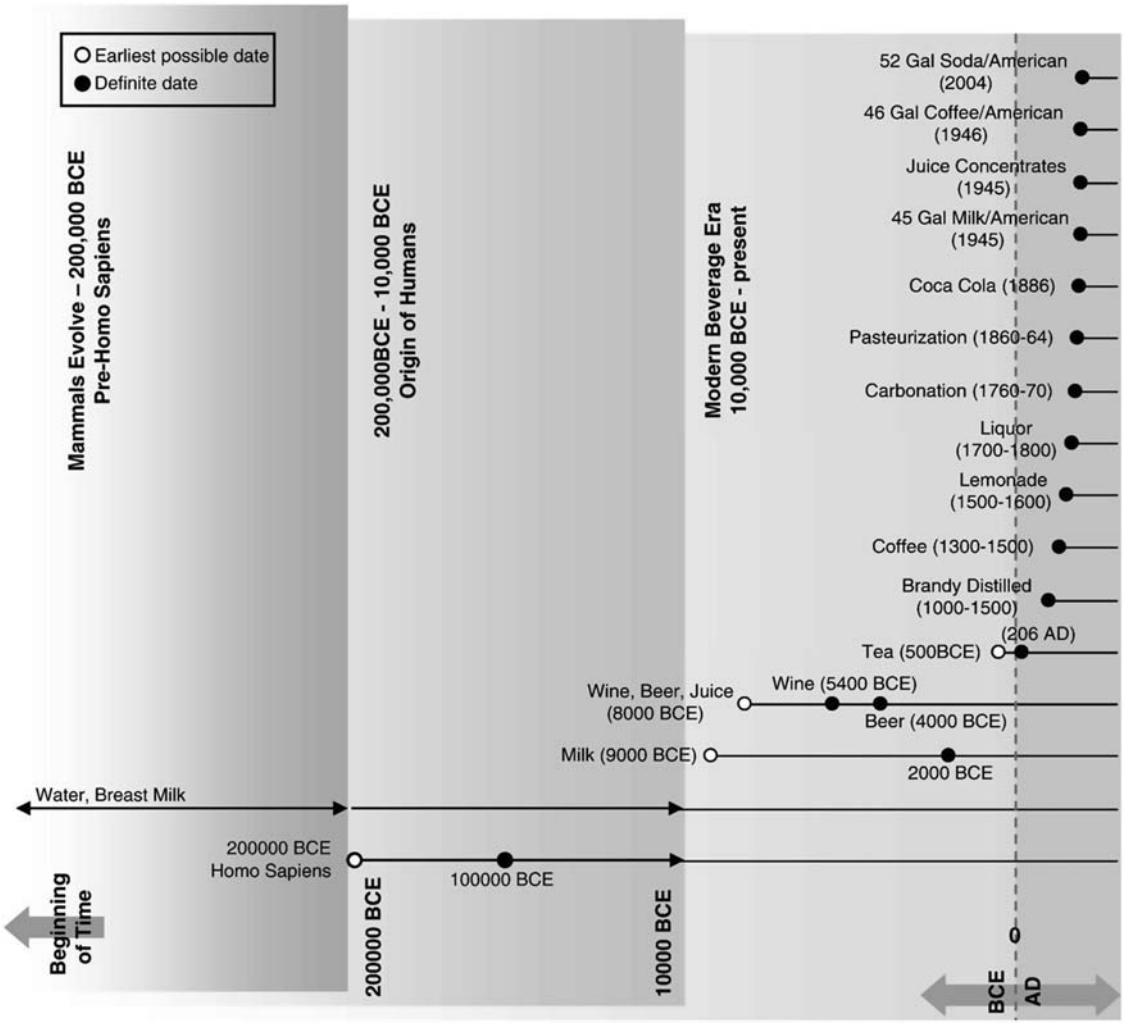


Fig. 1 Beverage history timeline. Adapted with permission from **Fig. 1** in Popkin, B.M., 2010. Patterns of Beverage Use Across the Lifecycle.

the total calorie intake will increase, which may result in weight gain and other health issues over time. In addition, high consumption of soft drinks is linked with less healthy behaviors. For example, the frequent consumers of soft drinks are more likely to be smokers, don't get enough sleep, don't exercise much, eat fast food often, and don't eat fruit regularly. Among adolescents who frequently drink soft drinks also have more screen time, including televisions, cell phones, computers, and video games.

In the following sections, we will provide a review of the relative nutritional and health benefits and risks of various beverage categories. The beverage categories reviewed in this article are ordered according to their health benefits (from the relatively healthy to unhealthy) based on current scientific evidence.

Coffee and tea

Tea

Tea has been widely consumed in many countries and regions for a long period of time. The origin of tea drinking can be traced back as early as around 500 BC in China. Tea remained a primarily Chinese beverage until the 17th century when Europeans started trading for teas and introducing them to a broader market. In 1908, teabags were invented in the US, which largely increased the convenience of tea consumption.

Tea is traditionally prepared by adding tea leaves to boiling water. The tea leaves are leaves, leaf buds, or internodes of the *Camellia sinensis* plant. According to how the tea leaves are produced and processed, tea can be classified into six major varieties: green tea (unwilted and unoxidized), yellow tea (unwilted and unoxidized, but allowed to be yellow), white tea (wilted and unoxidized), Oolong tea (wilted, bruised, and partially oxidized), black tea (wilted, crushed, and fully oxidized), and Pu-erh tea (fully

Table 2 Beverages nutrient composition.

Beverages[a]	*Black tea*[b]	*Coffee, brewed*[b]	*Whole milk (3.25% fat)*[b]	*1% fat milk, vitamin A & D-fortified*[b]	*Skim milk, vitamin A & D-fortified*[b]	*Soy milk, plain (silk)*[c]	*Orange juice, original (minute maid)*[d]	*Apple juice*[b]	*Tomato juice (V8)*[e]	*Diet coke*[f]	*Diet pepsi*[g]	*Coca-cola classic*[f]	*Pepsi cola*[g]	*Arizona green tea*[h]
Calories, kcal	2	2	149	102	83	100	110	112	50	0	0	100	105	70
Protein, g	0	0.3	7.7	8.2	8.3	7	0	0.3	2	0	0	0	0	0
Total fat, g	0	0	7.9	2.4	0.2	4	0	0.2	0	0	0	0	0	0
Saturate fat, g	0	0	4.6	1.5	0.1	0.5	0	0	0	0	0	0	0	0
Carbohydrate, g	0.7	0	11.7	12.2	12.2	8	27	28	10	0	0	26	27	17
Sugars, g	0	0	12.3	12.7	12.5	6	24	26	8	0	0	26	27	17
Total dietary fiber, g	0	0	0	0	0	1	0.2	0.2	2	0	0	0	0	0
Vitamin A, IU	0	0	395	478	500	500	0	0	2000	0	0	0	0	0
Vitamin C, mg	0	0	0	0	0	0	0	1.4	60	0	0	0	0	0
Vitamin D, IU	0	0	5	117	115	120	0	0	0	0	0	0	0	0
Folate, mcg	12	5	12	12	12	24	60	0	0			0	0	0
Calcium, mg	0	5	276	305	299	300	20	14	20	0	0	0	0	0
Potassium, mg	88	116	322	366	382	300	450	301	470	0	20	0	10	0
Sodium, mg	7	0	105	107	103	85	15	17	420	70	25	33	25	20
Magnesium, mg	7	7	24	27	27	40	24	12	NA	0	0	0	0	0
Caffeine, mg	47	95	0	0	0	0	0	0	0	31	24	23	25	10

[a]Amounts are per 8 fl oz (237 mL).
[b]Data are compiled from US Department of Agriculture, Nutritional Data Laboratory: http://www.nal.usda.gov/fnic/foodcomp/search/.
[c]Data are compiled from www.silkissoy.com.
[d]Data are compiled from www.minutemaid.com.
[e]Data are compiled from www.v8juice.com.
[f]Data are compiled from www.coca-cola.com.
[g]Data are compiled from www.pepsi.com.
[h]Data are compiled from www.arizonabev.com.

fermented and composted). Worldwide, the most popular consumed tea is black tea which accounts for 78% of all tea consumed, followed by green tea (20%). Other types of tea are only consumed in very small amounts or only by particular populations. While black tea is mostly consumed in the Western countries, green tea is consumed primarily in Southeastern countries like China and Japan. In the US, 21% of adults are tea consumers. Tea leaves contain more than 700 bioactive compounds, including flavanoids (primarily catechins), amino acids, vitamins (C, E, and K), caffeine, and polysaccharides. Many of these compounds are antioxidants. For instance, the Oxygen Radical Absorbance Capacity (ORAC) is 1253 μmolTE/100 g for green tea and 1128 μmolTE/100 g for black tea.

Because of tea's rich contents of many bioactive compounds, numerous studies have investigated the health benefits of tea consumption, particularly on its role in the prevention of cancer and cardiovascular diseases (CVD). The beneficial effects of tea consumption on cancer and CVD have been shown in animal models. Recent meta-analyses of human observational studies also suggest that higher consumption of tea is associated with a lower risk of cancer, but the benefits may be site-specific. The inverse association was observed for lung, liver, gastric, biliary tract, breast, ovarian, prostatic, and oral cancer, but not for pancreatic and bladder cancer (Filippini et al., 2020). Results from the human experimental studies on cancer prevention and treatment remain inconclusive. For example, three randomized controlled trials (RCT, a study design with the highest research rigorous) conducted in men with high-risk of prostate cancer have suggested a decreased risk but the effect size varied substantially across studies. Also, no effect was found on the prevention of non-melanoma skin cancer with each consumption. The potential therapeutic effect of epigallocatechin-3-gallate (EGCG), an active compound of green tea, has been investigated. Although promising, the levels of EGCG in some experimental studies are too high to be reached by consuming brewed tea. In addition, the EGCG has poor bioavailability and can be eliminated from the human body quickly after intake, which limits the use of this compound as a treatment option for cancer. Several observational studies also investigated the association of tea consumption with CVD risk or CVD mortality; however, the results are inconsistent across populations. Studies that reported an inverse association were mainly conducted in Japanese and Chinese populations. Possible explanations on tea's CVD protective effect could be related to its impact on reducing serum total or low-density lipoprotein (LDL)-cholesterol, triglyceride, and coronary vessel function. Several observational studies also reported that higher tea consumption was associated with a lower risk of depression and cognitive disorders. Tea consumption may also enhance the innate immunity, increase bone density, promote weight loss, improve glucose metabolism,

and reduce kidney stones. However, data on these health benefits of tea consumption are still limited. Most observed health benefits of tea are modest and at the consumption levels of 3–4 cups per day, particularly from the green tea.

Coffee

Coffee is one of the most popular beverages consumed worldwide. Coffee consumption originated in Africa around the ninth century. Early coffee was probably made using green (unroasted) coffee beans. The modern form of coffee (roasted, grounded, and brewed) likely appeared in Yemen in the late 14th century. By the 16th century, coffee had reached rest of the Middle East, and from there, spread to Italy and to the rest of Europe. Europeans began to add honey and milk to their coffee and vastly expanded the techniques for coffee brewing. In 1907, instant coffee was invented by preparing (freeze-drying or spray-drying) the soluble powder from brewed coffee beans. It rapidly gained popularity because of its convenience.

The most important bioactive compound in coffee is caffeine. A cup of coffee (7–8 oz), depending on the variety of coffee beans and brewed methods, may contain 80–175 mg of caffeine. The caffeine content in decaffeinated coffee is largely reduced to about 2–4 mg per cup. Coffee contains over 1000 bioactive compounds and is also a good resource of antioxidants. Both caffeinated and decaffeinated versions appear to have similar antioxidant levels. Vitamins and minerals that can be found in coffee include folate, vitamin K, pantothenic acid, riboflavin, calcium, magnesium, and manganese. Plain coffee contains only a trace amount of proteins and carbohydrates. The calories in coffee without any additives are only 2–3 kcal per cup. However, adding milk, cream, or sugar to coffee significantly increases the levels of calories as well as proteins, fats, and carbohydratse.

The stimulant effect of coffee is mainly due to its caffeine content. Human metabolic studies have found that caffeine ingestion acutely induces cardiac arrhythmias and increases plasma renin activity, catecholamine concentrations, and blood pressure. Numerous studies have been conducted to understand the long-term health effects of coffee consumption (van Dam et al., 2020). The relationship between coffee intake and cancer has been investigated for decades because of the potential cancer prevention effect of its main compounds such as caffeine, caffeic acid, and chlorogenic acid. Many meta-analyses, based on observational studies, have reported an inverse association between coffee consumption and the risk of liver, colorectum, colon, breast, prostate, oral, and endometrial cancer, but not of pancreatic, gastric or biliary tract cancer.

Many observational studies and several meta-analyses also examined the association between coffee consumption and CVD risk, including coronary heart disease (CHD), stroke, and heart failure. Up to date, there is no strong evidence to suggest that a higher intake of coffee or caffeine is associated with an increased risk of CVD. In a recent meta-analysis of 40 studies, a non-linear, inverse association was reported of coffee consumption with all-cause and CVD mortality, with the lowest risk at the intake of 2.5–3.5 cups per day (Kim et al., 2019). The inverse association between coffee consumption and CVD remained in subgroup analyses stratified by age, overweight status, alcohol drinking, smoking status, and caffeine content of coffee. A large prospective study of a half million people from the UK Biobank cohort also found an inverse association of coffee drinking with all-cause mortality and the results were the same when stratified by genetic polymorphisms affecting caffeine metabolism. On the other hand, coffee consumption has been related to several CVD risk factors such as serum lipids and blood pressure in the meta-analysis of RCTs. Coffee intake, especially unfiltered coffee, increases the plasma total and LDL-cholesterols and triglycerides. It is worthy to note that these RCTs were conducted only in Western countries with a mean duration of 45 days. Studies examining the long-term effects and in other populations are lacking. The cholesterol-raising effects may be related to the diterpenes cafestol and kahweol in coffee. Since diterpenes cafestol and kahweol can be trapped by paper filters, their concentrations are higher in boiled and espresso than in filtered coffee. Caffeine also has a well-known acute pressor effect. However, tolerance to the caffeine-induced pressor effect develops in habitual coffee drinkers. In addition, ingredients other than caffeine may also have blood pressure control effects. A meta-analysis of RCTs that longer than a week suggests that coffee consumption has no effect on blood pressure or risk of hypertension. Therefore, the long-term effect of habitual coffee consumption on blood pressure is still unclear.

Studies have consistently linked higher coffee consumption to a lower risk of type 2 diabetes. It has been suggested that compounds other than caffeine may contribute to such benefits. High coffee and caffeine intake were also associated with a reduced risk of Parkinson's disease in men, but not in women, which may be due to the modifying effect of estrogen. The relationships between coffee consumption and risk of other cognitive disorders (i.e., Alzheimer's disease, dementia, cognitive decline, and cognitive impairment) were reported as a "J-shaped" in a meta-analysis with the lowest risk at a daily consumption level of 1–2 cups of coffee. Similarly, a "J-shaped" relationship was also reported from a meta-analysis that examined the association between coffee consumption and risk of depression with the lowest risk at consumption of 1.5–2 cups per day.

There are concerns that coffee intake may be harmful to bone health. Animal studies have found that caffeine could increase urinary calcium excretion and decrease intestinal calcium absorption with a net result of a negative calcium balance. One meta-analysis of 10 prospective observational studies in humans found that a higher intake of coffee was associated with an increased risk of fracture, especially for women. However, a randomized cross-over study in women found that there was no difference in calcium balance between the moderate intake of caffeine (400 mg/day) and the placebo group for up to 19 days. Therefore, the current evidence on coffee consumption and bone health is inconclusive. Coffee consumption during pregnancy has raised some concerns regarding fetal health, mainly due to its caffeine content (Jahanfar and Jaafar, 2015). Several meta-analyses have reported that higher maternal caffeine intake during pregnancy was associated with a higher risk of pregnancy loss, low birth weight, and small for gestational age; however, the results are mixed for preterm birth. The adverse birth outcomes reported from observational studies are generally modest with the caffeine intake lower than 150 mg (e.g., 1–2 cups of coffee). Up to date, only one RCT was conducted among over 1000 Danish women to test whether the reduction in coffee intake during

pregnancy can decrease the risk of adverse birth outcomes. This trial found that there was no effect on birth weight and length of gestation between women who consumed decaffeinated instant coffee and caffeinated instant coffee (the mean caffeine difference was approximately 180 mg). Based on the available evidence, pregnant women are recommended to limit their caffeine intake to less than 200 mg per day. Pregnant women should be aware that caffeine can also be obtained from other beverages such as tea and soft drinks (See **Table 2**).

No safety limits of caffeine intake have been established for young children. There are some concerns about the potential adverse health effects of caffeine for this age group. Therefore, coffee and tea should be avoided for children younger than age 2.

Milk

Animal milk

Animal milk may have been consumed by humans since 2000 BC or even earlier. Later, the development of pasteurization largely enhanced the safety of milk consumption and expanded the market of milk trading. By the 1930s, most milk consumed in the UK and US was pasteurized. Milk consumption in the US peaked in the 1940s and has fallen steadily since then, probably due to the competition from other beverages, such as soft drinks.

Today, cow's milk is produced on an industrial scale and is the most commonly consumed form of milk. Milk and milk products are rich in many vitamins and minerals, including calcium, selenium, phosphorus, vitamin A, D (due to fortification), B2 (riboflavin), B5 (pantothenic acid), and B12 (cobalamin). Lactose, a disaccharide, is the dominant carbohydrate contained in plain milk (12–13 g/cup) and gives milk a slightly sweet taste. Interestingly, only certain human populations maintain the ability to digest lactose into adulthood. Some individuals cannot produce sufficient lactase, an enzyme to help absorb the lactose, and, therefore, may suffer intestinal gas, cramps, bloating and diarrhea when drinking milks. This is usually called "lactose intolerance" and it is more common among Asians, Africans, and Native Americans but less common among Caucasians. Processed cow milk contains about 8 g of proteins per cup. The fat content of milk varies by type of milk. A cup of whole milk (3.3% fat) contains about 8 g of fat, whereas reduced-fat (2% fat), low fat (1% fat), and skim milk contain 4.8, 2.4 and 0.2 g of fat, respectively. The majority of milk fats (57% for whole milk) are saturated fatty acids. For this reason, whole milk contributes significantly to the saturated fat intake for Americans and many other populations. Therefore, low-fat or skim milk are recommended instead of whole milk.

For children and adolescents, milk is an excellent source of calcium, vitamin D, and high-quality proteins. For adults, low fat and skim milks can contribute to a healthy diet, but are not essential. Both beneficial and detrimental health effects have been observed from milk consumption. Since milk is an important source for calcium and vitamin D, milk consumption can be beneficial for bone health, particularly among children and adolescents. A positive association between milk consumption and bone mineral density has been consistently reported by both human observational studies and RCTs. However, meta-analyses of observational studies have found that there is no clear pattern of association of milk intake with fractures, and findings from randomized trials are lacking. Some studies have shown that a higher intake of milk was associated with a reduced risk of metabolic syndrome, hypertension, and CVD, but the evidence is still inconclusive. Recent two meta-analyses concluded that there is no association between milk consumption and CVD risk, including CHD and stroke (both ischemic and hemorrhagic stroke) (Fontecha et al., 2019). A recent meta-analysis found an inverse association of fermented dairy consumption (including sour milk products, cheese, or yogurt) with mortality and CVD risk. Fermented milk is produced from fresh milk by the addition of bacterial cultures to initiate the fermentation process. As a result, fermented milk may impact human health through its influence on the gut microbiome.

The role of milk on weight control has been suggested by some studies but not others. It has been hypothesized that milk's effect on weight regulation is attributed to its high calcium content. High calcium intake may regulate body adiposity through plasma parathyroid hormone and $25(OH)_2\ D_3$ by repartitioning dietary energy from adipose tissues to lean body mass and increasing thermogenesis. However, it is unlikely that increasing milk consumption can be an option for weight loss without a substantial reduction in total caloric intake.

The most important health concern of milk consumption among adults comes from cancer. Several studies have reported that a higher intake of milk was associated with an elevated risk of prostate cancer in men, and breast and ovarian cancers in women. It is speculated that such detrimental effects may be related to the high level of insulin-like growth factor I (IGF-1) in cow milk. IGF-1 has been linked with increased cell proliferation and inhibition of apoptosis, and thus has been associated with cancer risk. Dairy producers have been injecting growth hormones into cows to increase milk production, which results in an increased concentration of IGF-1 in the milk. The 2018 World Cancer Research Fund (WCRF) report concluded there is no strong evidence to conclude the relationship between the consumption of dairy products (including milk) and the risk of cancer (World Cancer Research Fund International). The evidence is divergent between cancer sites: consumption of dairy products may decrease the risk of colorectal cancer, but increase the risk of prostate cancer.

Soy milk

Soy milk is an aqueous extraction of soybeans and has a milk-like appearance. Soy milk is traditionally produced by soaking the soybeans and grinding them with water. The resulting slurry is brought to heat at a boiling point for 15–20 min, followed by filtration to remove the insoluble residues (soy pulp). Soy milk was originally consumed in China and the oldest evidence can be traced to around 25-220 AD. Soy milk is widely consumed in China and other Asian countries such as Japan, Malaysia, and Singapore. The

drink is slowly getting popular in Western countries. Soy milk has a comparable amount of protein as cow milk, but very little saturated fat and no cholesterol. Unlike cow milk, soy milk does not contain lactose. Therefore, soy milk is safe for people with lactose intolerance. For the above reasons, soy milk is considered a healthy alternative to cow milk for children, adolescents, and adults. However, soy milk only provides about the 75% of calcium bioavailability from cow milk and cannot be legally fortified with vitamin D.

Both cow milk and soy milk should not be consumed by infants as a replacement of human milk or infant formula before age of 12 months. Unsweetened cow milk or soy milk can be consumed by children after age of 12 months to help meet calcium, potassium, vitamin D, and protein needs.

100% fruit or vegetable juice

Juices are the liquids extracted from fruits or vegetables. Juices can be prepared by mechanically squeezing or mashing fresh fruits or vegetables. While the early history of juice consumption is not well documented, lemonade appeared during the 16th century in Italy and was considered the pioneer of modern juice. Important technologies in the juice industry happened in the past 100 years, including the use of pasteurization and the introduction of juice concentration.

Fruit juice

Sometimes, the labels of commercial fruit juices may be misleading. In the US and several other countries, fruit juice can only be legally used to describe a product that is 100% pure fruit juice. According US Food and Drug Administration (FDA), juices reconstituted from concentrate must be labeled that the product is reconstituted from concentrate. Beverages that contain less than 100% and more than 0% fruit juice should be called fruit beverages, fruit cocktails, fruit drinks, or juice drinks. Most of such juice products should be considered as sugar-sweetened beverages because they are mainly composed of water and added sugar.

Fruit juice consumption has increased in recent years, probably due to the public perception of juices as a natural source of many nutrients. Fruit juices provide most of the nutrients of their natural source, including vitamins (A, C, folate, etc.), minerals (calcium, potassium, and magnesium), and other phytochemicals. However, they have relatively higher energy content and lower fiber levels compared to whole fruits (many commercial juices are filtered to remove the fibers or pulps). Carbohydrates, including sucrose, fructose, glucose, and sorbitol, are the most prevalent nutrient in juice. The carbohydrate concentration varies from 26 g/cup (8 oz) to more than 38 g/cup, depending on the type of juices. Juices fortified with calcium have approximately the same calcium content as milk but lack other nutrients present in milk. Some juices have high contents of vitamin A, C, and flavonoids, and certain types of juices are fortified with vitamin D, which may provide beneficial health effects. However, the high energy and sugar contents in fruit juices may contribute to excessive energy intake (American Academy of Pediatrics, 2001). Studies have linked excessive intake of fruit juice to the risk of obesity and type 2 diabetes, but results are still mixed. Nevertheless, fruit juice offers no nutritional advantage over whole fruit, particularly for healthy adults. Consumption of whole fruits should be encouraged to meet individuals' nutrient needs and energy balance.

Vegetable juice

Vegetable juices, such as tomato and multi-vegetable juices, are getting more and more popular in many countries. In general, vegetable juice has fewer calories and sugars than fruit juice and can be considered as a healthy alternative to fruit juice. However, commercial vegetable juices usually contain a considerable amount of sodium. As with fruit juices, whole vegetables, rather than vegetable juices, should be encouraged.

100% fruit or vegetable juices should not be offered to infants before age of 12 months. After 12 months, 100% fruit juice can be added but is not necessary if children can eat the whole fruit. If children drink 100% fruit juice, the consumption should be limited to 4 ounces per day. Any juices that contain added sugars should be avoided for children at any age.

Soft drinks

Soft drinks are usually referred to as nonalcoholic beverages typically containing water, sweeteners, and some flavorings. Sometimes, they are also called pops, soda, or fizzy drinks. They are called "soft" is in contrast to "hard" drinks which are generally referred to drinks with significant alcohol content. Many of these beverages are carbonated and sweetened with either sugar or high-fructose corn syrup (HFCS). Some of them may contain a small proportion of additional ingredients such as fruit juice or alcohol. If alcohol is added to a beverage, the content must be less than 0.5% of the total volume if the beverage is to be considered as a non-alcoholic beverage. Widely, soft drinks include cola, lemonade, iced tea (usually sweet), sports drinks, fruit drinks, fruit punch, and root beer (non-alcoholic form). Carbonated water (also called sparkling water, seltzer, or fizzy water) which contains carbon dioxide gas and flavorings, as well as smart water (distilled water with added electrolytes), fruit water (distilled water with

zero-calorie fruit flavorings), and vitamin water (distilled water with vitamins and flavorings) are usually classified between the soft drinks and plain drinking water.

The history of soft drinks can be traced back to the 1760s when carbonation techniques were developed. The original purpose was to produce beverages similar to naturally occurring mineral water. Back then, no sugars or artificial flavorings were added to these beverages. It is unclear when and by whom sweeteners and/or flavorings were added to the carbonated waters. In 1886, J.S Pemberton, an Atlanta pharmacist, combined kola (a caffeinated nut from Africa) with coca (a parent plant of cocaine from South America) to create Coca-Cola, which is considered a landmark event in the history of soft drinks. In the following section, we divided soft drinks into 2 major categories based on whether they are sweetened with added caloric sweeteners (sugars or HFCSs) or non-caloric artificial sweeteners (NAS).

Diet beverages

Diet beverages (diet soft drinks) are carbonated or non-carbonated beverages sweetened with NAS that provide less than 1 kcal in one serving. Currently, 5 artificial sweeteners (saccharin, acesulfame, aspartame, neotame, and sucralose) have US FDA approval. Beverages sweetened with these artificial sweeteners are considered to be safe. In the past several decades, there is an increasing trend in consumption of diet beverages worldwide, likely due to the increased concern regarding the adverse effects of added sugars. In the US, 19% of children and 31% of adults in the US consume diet beverages on a daily basis.

Diet beverages are preferable to sugar-sweetened beverages because they provide water and sweetness but no calories. Randomized interventions have shown that drinking diet beverages was more favorable for weight management compared to drinking sugar-sweetened beverages. However, evidence is emerging to suggest that long-term consumption of these sweet beverages may increase people's desire for sweetness, resulting in increased intake of other sweet foods and beverages. NAS are hundreds to thousands of times sweeter than sugars. Individuals who habitually consume NAS may find less intensely sweet or unsweet foods unpalatable. In addition, diet beverages have a high degree of sweetness but no calories, which might produce a dissociation between sweet taste and calorie intake. One concern is that the dissociation of these physiological events might disrupt the hormonal and neurobehavioral pathways regulating hunger and satiety, potentially leading to seeking and consuming more palatable foods. Recent data also suggested that consumption of NAS could alert the microbial metabolic pathways that are linked to glucose intolerance in mice. A few small RCTs have reported an increase in insulin levels following the consumption of diet beverages or NAS in humans. Results from human observational studies on the relationship of long-term consumption of diet beverages with the risk of CHD, stroke, and death are highly inconsistent (Mossavar-Rahmani et al., 2019). These new data clearly called a need for large studies with rigorous design to determine the role of diet beverages and NAS in metabolic health, particularly for habitual consumption and for different types of NAS.

Sugar-sweetened beverages

Sugar-sweetened beverages (regular soft drinks) include carbonated and noncarbonated beverages which are usually sweetened with HFCS or sucrose. These beverages have relatively high calorie and sugar contents, but no or a very small amount of other nutrients. Higher intake of sugar-sweetened beverages has been consistently associated with higher energy intake and increased risk of dental caries, obesity, type 2 diabetes, metabolic syndrome, hypertension, CVDs, gout, and cancer in several populations. Consumption of sugar-sweetened beverages was also positively associated with CVD and cancer mortality, independent of other dietary and lifestyle factors (Malik et al., 2019). Several RCTs have shown that reducing sugar-sweetened beverage consumption might result in significant weight loss and blood pressure reduction. It has been proposed that sugar-sweetened beverages contribute to weight gain because of their high sugar content and incomplete energy compensation. Fructose from sucrose or HFCS may also increase blood pressure and promote the accumulation of visceral adiposity through the increase of hepatic de novo lipogenesis. In addition, consuming a large amount of sugar-sweetened beverages can contribute to a high glycemic load diet by providing a large amount of rapidly absorbable sugars, leading to insulin resistance, chronic inflammation, and possibly impaired β-cell function. Therefore, high sugar-sweetened beverage consumption may increase type 2 diabetes and cardiovascular risk independently of obesity. Recent meta-analyses also reported that a higher intake of sugar-sweetened beverages was associated with a higher risk of fracture in children and lower bone mineral density in adults. Sugar-sweetened beverages, thus, should be consumed as less as possible.

Summary

In summary, the primary purpose of drinking beverages is to provide fluids rather than calories or other nutrients. Beverages with more nutritional values but few calories and sugars can be consumed to meet our body's requirement for fluids. In general, beverages should provide less than 10% of total energy for healthy adults. By considering the energy density, nutrient density, and health benefits/risks linked with each beverage category, unsweetened tea and coffee are mostly recommended for adults, followed by low-fat and skim milk, and 100% fruit and vegetable juices. Soft drinks, both diet and sugar-sweetened beverages, are not necessary for a healthy diet and should be avoided. Special populations such as children and pregnant women may have additional

considerations when choosing beverages. No beverages should be consumed by infants before 12 months of age. Beverages that contain high levels of caffeine should be avoided for children and pregnant women.

References

American Academy of Pediatrics, 2001. The use and misuse of fruit juice in pediatrics. Pediatrics 107, 1210–1213.

Filippini, T., Malavolti, M., Borrelli, F., Izzo, A.A., Fairweather-Tait, S.J., Horneber, M., Vinceti, M., 2020. Green tea (*Camellia sinensis*) for the prevention of cancer. Cochrane Database Syst. Rev. (3) Art. No.: CD005004.

Fontecha, J., Calvo, M.V., Juarez, M., Gil, A., Vicente Martínez-Vizcaino, V., 2019. Milk and dairy product consumption and cardiovascular diseases: an overview of systematic reviews and meta-analyses. Adv. Nutr. 1 (10), S164–S189.

Jahanfar, S., Jaafar, S., 2015. Effects of restricted caffeine intake by mother on fetal, neonatal and pregnancy outcomes. Cochrane Database Syst. Rev. (6) Art. No.: CD006965.

Kim, Y., Je, Y., Giovannucci, E., 2019. Coffee consumption and all-cause and cause-specific mortality: a meta-analysis by potential modifiers. Eur. J. Epidemiol. 34, 731–752.

Malik, V.S., Li, Y., Pan, A., Koning, L.D., Schernhammer, E., Walter, C., Willett, W.C., Hu, F.B., 2019. Long-term consumption of sugar-sweetened and artificially sweetened beverages and risk of mortality in US adults. Circulation 139 (18), 2113–2125.

Mossavar-Rahmani, Y., Kamensky, V., Manson, J.E., Silver, B., Rapp, S.R., Haring, B., Beresford, S.A., Snetselaar, L., Wassertheil-Smoller, S., 2019. Artificially sweetened beverages and stroke, coronary heart disease, and all-cause mortality in the women's health initiative. Stroke 50 (3), 555–562.

Popkin, B.M., 2010. Patterns of beverage use across the lifecycle. Physiol. Behav. 100, 4–9.

van Dam, B., Hu, F.B., Willett, W.C., 2020. Coffee, caffeine, and health. N. Engl. J. Med. 383, 369–378.

World Cancer Research Fund International World Cancer Research Fund International/American Institute for Cancer Research Continuous Update Project Report: Meet, Fish, and Dairy Products and Risk of Cancer. Available at dietandcancerreport.org.

Further reading

Popkin, B.M., Armstrong, L.E., Bray, G.M., Caballero, B., Frei, B., Willett, W.C., 2006. A new proposed guidance system for beverage consumption in the United States. Am. J. Clin. Nutr. 83, 529–542.

U.S. Department of Agriculture, U.S. Department of Health and Human Services, 2020. Dietary Guidelines for Americans, 2020–2025, ninth ed. Available at. DietaryGuidelines.gov.

Biofortification: A primer on nutrient enriched crops

Mduduzi N.N. Mbuya[a], **Valerie M. Friesen**[b], **and Lynnette M. Neufeld**[c], [a] Global Alliance for Improved Nutrition (GAIN), Washington, DC, United States; [b] Global Alliance for Improved Nutrition (GAIN), Geneva, Switzerland; and [c] Food and Agriculture Organization of the United Nations (FAO), Rome, Italy

© 2023 Elsevier Ltd. All rights reserved.

This is an update of C. Hotz, Biofortification, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 175–181, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00025-8.

Introduction 286
Overview of biofortification 286
History and progress 286
Target foods and nutrients 287
The potential of biofortification to improve micronutrient status 288
Nutrient retention 288
Minerals: zinc and iron 288
Vitamin A 289
Nutrient bioavailability 289
Minerals: zinc and iron 289
Vitamin A/provitamin A 289
Efficacy 290
Iron enriched crops 290
Provitamin A enriched crops 290
Zinc enriched crops 290
Effectiveness 290
Achieving potential for impact 291
Conclusion 291
Conflict of interest 292
References 292
Further reading 292
Relevant websites 292

Key points

- Biofortification, or nutrient enrichment, of staple food crops, is the process of increasing the density of vitamins and minerals in widely consumed food crops through plant breeding or agronomic means.
- Nutrient retention in the biofortified food following storage, processing, and cooking, nutrient bioavailability compares favorably to non-biofortified foods suggesting potential for impact. Efficacy is well demonstrated, but less is known about factors determining effectiveness, including gene and environment interactions and use and re-use of seeds.
- There is a need to generate evidence on the effectiveness of biofortification in real-world settings, explore and test innovative market delivery models, and ensure coherent policies are in place that create an enabling environment for biofortification.

Glossary

Agronomy Refers to the application of mineral enriched fertilizers to the soil or plant leaves to increase the micronutrient content of the resulting edible crop

(Nutrient) Bioaccessibility Is defined as the amount of an ingested nutrient that is available for absorption (released from the food matrix) in the gut after digestion. It is often considered a subset of bioavailability

(Nutrient) Bioavailability Refers to the proportion of an ingested nutrient that reaches the systemic circulation and the specific sites where it can exert its biological action. It is also referred to as absorbability

Biofortification Refers to the process of increasing the micronutrient content and/or bioavailability of micronutrients in crops through conventional plant breeding techniques. Several commonly consumed foods have been developed and shown to be

effective for improving micronutrient intakes, including iron-rich beans and millet, and vitamin A-rich sweet potato, maize, and cassava

Conventional plant breeding Refers to the cross-pollination of two genetically closely related plants to increase desirable characteristics. These characteristics include yields as well as micronutrient levels among others

Micronutrient Is a general term that refers to all vitamins (e.g., vitamin A, B vitamins such as folate, and vitamin D) and minerals (e.g., zinc, iron) that are required in small amounts for a range of different processes in the body, including the healthy functioning of all organ systems, growth, development, and health. An essential micronutrient is one that must be obtained from the diet because the body cannot produce it or does so in quantities insufficient to meet needs

Staple foods Refers to a food that makes up the dominant part of a population's diet. Food staples are eaten regularly—even daily—and usually in large quantities such that they contribute a major proportion of a person's energy and nutritional needs. They commonly include cereals, starchy roots, tubers, and legumes, and often extend to include vegetable oils, salt, and sugar

Introduction

Biofortification (also known as nutrient enrichment) is a public health intervention in which the density of vitamins and minerals in a widely consumed crop is increased through agricultural means as a strategy to increase regular consumption of nutrients, thus decreasing risk of micronutrient deficiencies, particularly in low- and middle-income countries. Several processes are used to achieve biofortification, and these include conventional breeding, transgenic techniques, and agronomic approaches (Bouis, 2018). These are presented in detail in an earlier edition of this article (Hotz, 2013). The dominant method of enrichment of crops is through the use of conventional plant breeding, which applies the same basic techniques as used to breed plants with improved agronomic traits, such as increased yield or pest resistance (Hotz, 2013). Furthermore, conventional plant breeding does not encounter the same regulatory and acceptability barriers as transgenic approaches. In this article we present an overview of the history and current progress of biofortification, summarize the evidence on the efficacy and effectiveness of biofortified crops to improve micronutrient status and functional outcomes, and discuss the key elements necessary to maximize potential for impact of biofortification with a focus on conventionally bred biofortified crops.

Overview of biofortification

History and progress

The main objective of plant breeding—which is the most predominant form of biofortification—has historically been to improve farm productivity, usually by developing crops with higher yields, greater resilience to drought, and resistance to pests. This strategy, which is the primary focus of this article, takes advantage of the considerable genetic variation in crop species and entails crossing varieties with various traits in a bid to develop a variety with desirable characteristics. Agricultural research stations and smallholder farmers alike have used this technique to isolate desired traits. In comparison, transgenic approaches entail the introduction of new genes derived from another species to the target plant species. An example of this technique in practice is Golden Rice, in which genes from bacteria were introduced to trigger production of β-carotene in rice grain endoplasm (Beyer et al., 2002). Agronomic methods include the application of mineral (e.g., zinc or iron) micronutrient fertilizers to soils or plant leaves (foliar) to increase micronutrient contents in edible parts of crops (de Valença et al., 2017).

Biofortification using convention plant breeding began in the 1990s when it was in essence prompted by Howarth Bouis suggesting that dietary quality was just as important as quantity and that crops could "fortify themselves" with micronutrients to address the "hidden hunger" of micronutrient deficiencies. He posited three questions (Bouis, 2003):

1. "Can commonly eaten food staple crops be developed that fortify their seeds with essential minerals and vitamins?
2. Can farmers be induced to grow such varieties?
3. If so, would this result in a marked improvement in human nutrition at a lower cost than existing nutrition interventions?"

Over the next decade scientists at the Consultative Group on International Agricultural Research (CGIAR) worked to provide proof of concept and evidence of the viability of this concept. This culminated in the establishment of HarvestPlus: the Biofortification Challenge Program in 2003 (Pfeiffer and McClafferty, 2007). Later renamed HarvestPlus, this international consortium of researchers and implementers has led, in collaboration with other CGIAR programs such as the International Potato Center (CIP), agricultural and biotechnology research, food science and nutritional research, economic and policy impact research, and direct programs delivering biofortified crops through the seed and agricultural inputs supply system to farmers. To date, more than 340 biofortified varieties of 12 crops have been released in 40 countries (Council for Agricultural Science and Technology (CAST), 2020).

The value proposition of biofortification is threefold. First, it has the potential for cost-effectiveness in the long-term by increasing regular consumption of essential nutrients (Bouis, 2003). This is predicated on initial development and dissemination having been completed and commercialization achieved, after which recurring costs are estimated to be low. Recent evidence on this is detailed in a comprehensive report of the Council for Agricultural Science and Technology (Council for Agricultural Science and Technology (CAST), 2020). Second, biofortification has the inherent ability to reach rural populations who rely heavily on locally-produced staple foods and are at greater risk of micronutrient deficiencies–and thus have the greatest potential to benefit. This population segment may also have limited access to a high variety of nutrient-rich foods and to industrially fortified foods from markets (that is, those to which micronutrients are added during processing), leaving them at additional risk and unreached by other population-based micronutrient deficiency mitigation programs. Third, it has potential for sustainability as once planting material is obtained, it can be saved, recycled, and shared with other farmers.

On the left side of the ledger are a few notable disadvantages. Mainly, biofortification can only provide modest additional amounts of micronutrients compared to supplements and, in some cases, industrially fortified foods; therefore, potential for impact may be modest. We provide more detail on target nutrient profiles in the next section. Another limitation that relates to delivery and timeliness in the time required to develop a viable, stable crop, which can take up to 6–8 years. This results in less flexibility to adjust nutrient contents and/or combinations over time. As a result, biofortification is considered a complementary strategy to other micronutrient deficiency strategies (e.g., industrial food fortification, supplementation, and promotion of a diverse diet).

Target foods and nutrients

A diet that provides an adequate and consistent supply of essential micronutrients is out of the reach for many of the world's poor due to the relatively higher cost of nutrient dense foods and their availability in some regions. This explains the high prevalence and recalcitrance of micronutrient deficiencies and their harmful effects on morbidity and mortality. The fundamental premise of biofortification is therefore that enhancing the micronutrient content of energy rich commonly consumed foods, on which resource-constrained people often rely, can contribute to intakes and concomitant reductions in micronutrient malnutrition. The first step to achieving this goal is for plant breeders to identify plant varieties with high micronutrient content that can be bred into local varieties.

Such research has been largely successful in identifying nutrient rich varieties of the world's most important staple food crops, including maize, rice, wheat, cassava, potato, sweet potato, common beans, pearl millet, sorghum, and banana/plantain. Efforts have been mainly focused on increasing the content of provitamin A carotenoids, iron, and zinc (Pfeiffer and McClafferty, 2007) likely due to the high prevalence of deficiencies associated with these micronutrients in low- and middle-income countries.

It is important to note that a wider range of nutritional components are also being considered for biofortification (Council for Agricultural Science and Technology (CAST), 2020). This includes other minerals (i.e., calcium, iodine, magnesium, selenium), vitamins (i.e., vitamins C, E, folate), non-provitamin A carotenoids (i.e., lutein, lycopene), protein and protein quality, essential amino acids, fatty acids, other secondary nutritional components such as flavonoids, and probiotic compounds (i.e., inulin and other fructans). So, the research agenda is more expansive and ongoing.

In addition to addressing the inherent nutrient content of foods, biofortification has considered the presence of antinutritive compounds. Cereal grains tend to have a high content of phytate, a phosphorus storage compound found within the hulls and kernels of these seeds that has a strong binding affinity to dietary minerals such as calcium, iron and zinc, inhibiting their absorption in the small intestine. As a secondary strategy, biofortification research has therefore sought also to improve mineral bioavailability from cereals by targeting reductions in the phytate content or increases in phytase activity in the grain to reduce the inhibitory effect of phytate. The focus on high yield crops is important to note as it makes biofortified crops appealing to farmers and value chain actors.

To date, there are several examples of biofortified staple food crops under production using conventional plant breeding (**Table 1**). Some of these traits are visible, e.g., vitamin A biofortified staple food crops are often yellow or orange because of the β-carotene. While others are invisible, e.g., high iron beans look like conventional varieties.

Minimum breeding target levels for increased micronutrient contents are estimated during the development stage (Pfeiffer and McClafferty, 2007). Breeding targets are set based on micronutrient needs and usual daily quantities consumed of the staple crop by non-pregnant, non-lactating women of reproductive age and children 1–6 years of age and account for any micronutrient losses during storage, processing and/or cooking and the fractional absorption of the micronutrient content. These population groups are selected as they are at greatest risk of micronutrient deficiencies and thus the targets of most micronutrient interventions. Considering this, most biofortified staple food crops aim to provide an additional 20–100% of the micronutrient requirements for women and children compared to conventional crops (**Table 1**).

Current varieties of biofortified staple crops developed through conventional plant breeding focus on single nutrients. This can be considered a limitation because many population segments in low- and middle-income countries suffer from a range of co-existing micronutrient deficiencies and therefore combining micronutrients in the same staple food (also known as multi-biofortification) would be advantageous to tackling this problem. Some examples of this have been developed, such as

Table 1 Target nutrient levels and estimates of contribution to the estimated average requirement (EAR)[a] of biofortified staple crops developed by conventional breeding.

			Estimated proportion of the EAR provided for non-pregnant, non-lactating women		*Estimated proportion of the EAR provided for children, 1–6 years old*	
Crop	*Nutrient*	*Target increment provided by biofortification*	*Before biofortification*	*After biofortification*	*Before biofortification*	*After biofortification*
Beans	Iron	+44 ppm	45%	90%	40%	75%
Pearl millet	Iron	+30 ppm	50%	85%	45%	75%
Sweet potato	Pro-vitamin A	+70 ppm	0%	>100%	0%	>100%
Cassava	Pro-vitamin A	+15 ppm	0%	>100%	0%	95%
Maize	Pro-vitamin A	+15 ppm	0%	55%	0%	60%
	Zinc	+12 ppm	45%	65%	55%	80%
Rice	Zinc	+12 ppm	40%	70%	40%	70%
Wheat	Zinc	+12 ppm	35%	55%	20%	25%

[a]Rounded to nearest 5%.

Source: https://www.harvestplus.org/content/estimated-average-requirements-provided-biofortification.

multi-vitamin corn in South Africa that contains both provitamin A and folate; however, it was developed using transgenic techniques. Whether or not it is feasible to target multiples nutrients simultaneously through conventional plant breeding has yet to be determined.

The potential of biofortification to improve micronutrient status

The extent to which biofortification can impact micronutrient status and health outcomes at scale is influenced by several factors, which include the presence of nutrient deficiency in the target population (i.e., the need), usual intake of the major staple food or foods (i.e., the opportunity), nutrient retention in the biofortified food following storage, processing, and cooking, nutrient bioavailability, and the efficacy of the biofortified food to improve nutrient status when consumed under controlled conditions. Taken together, the latter influences (retention, bioavailability, efficacy) relate to efficiency, and we tackle them sequentially.

Nutrient retention

A major role of the food matrix is to trap the nutrients within cells or subcellular compartments and provide constituents that interact chemically with specific nutrients to either encourage or delay their release, leading to their classification as absorption promoters and inhibitors. Because biofortified crops are subjected to periods of storage, transportation, processing, and cooking or other forms of preparation (e.g., fermentation), the extent to which nutrients are retained is worth considering as the effects differ by food processing technique and micronutrient (**Table 2**).

Minerals: zinc and iron

The lower proportions of iron and zinc retained (i.e., 20–60%) following milling of cereal grains for refinement is concerning. Grains, such as wheat, are often dehulled and milled to produce a flour, which, because a large proportion of minerals occur in the outer aleurone layers, can result in significant reductions of minerals like zinc (Díaz-Gómez et al., 2017). When it comes to cooking, the retention of minerals in plant-based staple foods following typical household cooking methods is generally quite high, with ≥90–95% of iron and zinc retained after cooking of cereal grains, roots, and tubers and ≥75% retained in boiled legumes. Minerals are heat-stable, although it is plausible that losses can occur due to leaching. However, the extent to which retention of minerals after cooking will differ in mineral-biofortified staple food crops is not yet determined. On the other hand, heating can soften the cell walls and remove inhibitors thereby enhancing the bioaccessibility of zinc in particular (Díaz-Gómez et al., 2017).

Table 2 Effects of processing on the micronutrient content of food.

	Carotenoids	*Iron and zinc*
Drying	Can reduce carotenoid levels but this depends on the drying method, the temperature/time combination.	Not applicable.
Dehulling	Not applicable.	Significant quantities of minerals can be lost, up to 50% of the iron in some grains. Reduces the level of inhibitors that prevent mineral uptake.
Milling	Increases carotenoid bioavailability because the food particle size is reduced.	Degrades the cell wall, allowing minerals to interact with other components. Iron, zinc and phytate levels are reduced by milling, but the remaining iron and zinc is more bioavailable.
Fermentation	Does not usually affect carotenoid retention.	Can degrade phytate through the action of microbial phytases.
Heating	Exposure to light and heating can result in the loss of provitamin A activity.	Minerals are heat stable. Can enhance mineral absorption by softening the cell walls and removing inhibitors.

Adapted from Díaz-Gómez et al. (2017).

Vitamin A

Nutrient retention is a greater consideration for biofortification with vitamins, particularly provitamin A carotenoids as they are more susceptible to exposure to heat, light, and oxygen. As such, retention during the postharvest storage (with respect to the duration and conditions) of the biofortified food must be considered. In orange sweet potato, the retention of β-carotene averages 84% after boiling and 77% after steaming, which are the most common cooking methods (Hotz, 2013). In biofortified orange maize, retention of β-carotene was ~75% in porridge cooked from fermented or unfermented, wet, milled flour. Retention of β-carotene after drying of sweet potato or cassava to make flour or chips is moderate, but dramatic degradation occurs during storage of these dried products over weeks or months.

Nutrient bioavailability

Not all the retained nutrient is absorbed by the body and used in the physiological functions for which they are needed due to two interrelated characteristics. One is bioavailability (also called absorbability), which refers to the proportion of an ingested nutrient that reaches the systemic circulation and specific sites where it can exert its biological action. Bioaccessibility is often considered a subset of bioavailability and is defined as the amount of an ingested nutrient that is available for absorption (i.e., released from the food matrix) in the gut after digestion. The bioaccessibility and overall bioavailability of nutrients from specific foods is determined by the food matrix, the chemical form as well as the physical location of the nutrient, and the presence or absence of other food factors that inhibit or promote the bioavailability of a nutrient. Because traits often interact, it is possible that the process of biofortification will result in concomitant changes in these or other properties that may alter the relative bioavailability of the additional nutrient.

Minerals: zinc and iron

The bioavailability of iron and zinc is influenced in general by the presence of antinutrients such as phytate. Although levels of phytate can be reduced through plant breeding, this needs to be counterbalanced against the potential implication that decreasing them too much may compromise the benefits of phytates to plants (La Frano et al., 2014). Studies have been conducted using low-phytic acid mutant maize with ~60% less phytic acid content than the wild type (Mazariegos et al., 2006). Controlled laboratory studies measured large increases in calcium, iron, and zinc bioavailability. However, when this maize was incorporated into typical diets of Guatemalan children in a community-based study, phytic acid in the total diet was reduced by only 25–30% and no significant improvement in zinc bioavailability was observed (Mazariegos et al., 2006). The difference in results highlights the importance of conducting these kinds of studies under realistic dietary conditions.

There is a general indication that iron biofortification of grains will not modify the relative bioavailability of additional iron from that of intrinsic (i.e., pre-existing) iron. Ferritin is an iron storage protein in plants, and transgenic rice and wheat with increased grain ferritin have increased grain iron content. Several studies indicate that iron from ferritin is as well absorbed as iron from exogenous ferrous sulfate and that zinc biofortified rice and wheat have similar bioavailability when compared to varieties that are industrially fortified with zinc. For example, human studies found that bioavailability for iron (2.6–9.0%) and zinc (17–20%) from biofortified crops were consistent with the results of studies in non-biofortified plants (La Frano et al., 2014).

Vitamin A/provitamin A

Food related factors affecting the efficiency of absorption and conversion of dietary provitamin A to retinol include the type of provitamin A carotenoid (i.e., α-carotene, β-carotene, β-cryptoxanthin), its isomeric form (i.e., *trans-*, *cis-*), and the food matrix. In contrast to minerals, the bioconversion of β-carotene to retinol in the provitamin A biofortified crops, which varied from 2.8:1

to 6.5:1, was more efficient than the average bioconversion of 12:1 estimated for nonbiofortified foods (La Frano et al., 2014). The amount of provitamin A ingested (μg) that is required to produce 1 μg retinol is expressed as the retinol equivalency ratio, and this varies by food type. The order of increasing efficiency of conversion to retinol, or decreasing ratios, is: green or yellow vegetables ≪ orange fruits, vegetables, or sweet potato ≪ yellow or orange biofortified cereal grains (Hotz, 2013). Studies using animal and in vitro models also indicate that β-carotene in biofortified cassava is well absorbed and as efficiently converted to retinol as β-carotene from biofortified maize and rice (La Frano et al., 2014). These results support the potential for provitamin A-biofortified staple foods to serve as important sources of vitamin A.

Efficacy

Another critical assumption of linking biofortification to its intended goal is that the nutritional efficacy of a biofortified crop must be established. The extent to which various biofortified staple crops can improve nutrient status and/or functional outcomes among women and children has been demonstrated under controlled settings. We summarize this evidence below (references to studies mentioned in this section can be found in (Council for Agricultural Science and Technology (CAST), 2020) except where noted). It is important to note however that, when moving from controlled to real-world settings, there is a need to additionally consider the different factors affecting nutrient absorption and utilization, including processing, cooking methods, and storage as described above that will vary in uncontrolled settings.

Iron enriched crops

Three of four efficacy studies completed for iron biofortified crops observed positive effects on iron status. In the Philippines, iron rice improved iron stores among non-anemic women. In Rwanda, iron biofortified beans improved iron status among women. In India, iron pearl millet improved iron status among school children. Conversely, in Mexico, iron biofortified beans had no significant effect on iron status compared to conventional beans among school children 12–16 years. Further meta-analyses of the first three studies have demonstrated that consumption of iron biofortified foods resulted in significantly increased serum ferritin concentrations and total body iron compared to controls and that the greatest effects were seen among individuals who were iron deficient at baseline and thus had the greatest potential to benefit.

Positive effects of iron biofortified crops on functional outcomes, such as cognitive and physical performance, have also been demonstrated. In Rwanda, iron biofortified beans increased physical work efficiency and improved cognitive performance among women. In India, iron biofortified pearl millet increased measures of physical activity and improved cognitive performance among school children 12–16 years.

Provitamin A enriched crops

The efficacy of the various vitamin A biofortified crops on vitamin A status and some functional outcomes has been demonstrated across different countries and populations. Provitamin A orange sweet potato improved vitamin A status in South Africa and vitamin A intakes and serum retinol concentrations in Mozambique among children. In Bangladesh, orange sweet potato increased plasma β-carotene concentrations but not vitamin A status among women with low initial vitamin A status; nor did it increase breast milk vitamin A concentrations among lactating women. Orange sweet potato has also been shown to reduce diarrhea prevalence and duration among children under 3 years in Mozambique.

Provitamin A maize has been demonstrated to significantly improve serum beta-carotene concentrations, total body stores of vitamin A compared to conventional maize as well as increased pupillary responsiveness among children in Zambia. Additionally, biofortified maize improved retinol concentration in breast milk among Zambian women. Biofortified maize has also been shown to be an effective, bioavailable source of vitamin A among men and women.

Provitamin A yellow cassava increased serum retinol and β-carotene concentration, markers of vitamin A status, among children 5–13 years in Kenya and improved serum retinol among preschool children in Nigeria (Afolami et al., 2021).

Zinc enriched crops

For zinc, only a few efficacy studies have been conducted to date. In India, high zinc biofortified wheat had no significant effect on zinc status compared to low zinc wheat among women of reproductive age and children 4–6 years, but this may have been due to small differences in zinc levels between the two wheat varieties (30 ppm vs. 20 ppm) and the limited duration of the intervention (6 months). However, a significant reduction in days with pneumonia and vomiting among children and fever among women was observed in the high zinc group. A randomized-controlled trial is currently in progress in Pakistan to examine the impact of zinc biofortified wheat on zinc status and functional outcomes among adolescent girls and children (Lowe et al., 2020). More data are needed to provide definitive evidence.

Effectiveness

Evidence on the effectiveness of biofortified staple food crops to improve population-level nutritional status in real-world settings is limited (references to studies mentioned in this section can be found in (Council for Agricultural Science and Technology (CAST), 2020) except where noted). To date, two randomized controlled effectiveness trials have been conducted for provitamin A orange sweet potato. In Uganda, farming households were randomized to an intensive or reduced intervention program over 2 years that

was aimed at increasing the production and consumption of orange sweet potato. Consumption of orange sweet potato was associated with increased vitamin A intakes among children 3–5 years and women and improved vitamin A status (serum retinol concentrations) among children compared with controls. A similar study in Mozambique compared low intensity (1 year) and high intensity (3 years) intervention groups with controls and found increased vitamin A intakes among women and children in both groups compared to controls (with no difference between intervention groups). The lack of evidence on real-world impact is largely due to the nascency of the intervention and the fact that many crops have not yet been scaled up such that it would be feasible to assess impact at a population-level.

In addition to effectiveness, it is important to understand acceptance and adoption of biofortified crops as these are factors that determine their consumption and ultimately effectiveness. Evidence from a systematic review of acceptance and adoption of biofortified crops (which included sweet potato, maize, rice, and pearl millet) in low- and middle-income countries found good overall acceptance of biofortified crops and that sensory changes (such as color) were not barriers to acceptance (Talsma et al., 2017). Additionally, the most importance determinants of acceptance and adoption were availability and information on health benefits of the biofortified crops.

Achieving potential for impact

Clearly, biofortification is a food-based approach that holds promise for cost-effective impacts on micronutrient, health, and functional outcomes. For biofortification to achieve its potential for impact, scale up, and sustainability, there is a need to identify or build on cost-effective delivery strategies. Such strategies need to be grounded in the three interrelated elements of (1) supply, (2) demand, and (3) an enabling policy environment.

Considering the goal of reaching the rural poor, current delivery models have focused with good reason on farmers. To unpack the sequence, the program pathway moves from specialized seeds bred by plant specialists, to seeds then multiplied and released by seed companies, to agricultural production by farmers, to farmers allocating some of their production for home consumption. With few published exceptions (Kangile et al., 2021), there has been less focus on getting the biofortified foods in the commercial markets. As suggested by Kangile in their study of orange sweet potato (Kangile et al., 2021), we posit that a transition from farmer to commercial market models is needed to achieve potential impact, scale up, and sustainability and additionally to reduce/eliminate the need for donor funding. This will require four conditions to be met. First, a sufficient supply of biofortified seed that is accessible and desirable to farmers. Second, the traceability of the biofortified seed through initial supply chain (as described above) and then through the rest of the food supply chain (i.e., storage and distribution, processing, packaging and distribution, retail, and markets). This is needed to signal nutritional value and consequently demand by value chain actors but is easier for crops with visible traits than those with invisible traits. Third, incentive/motivation for processors to take up the biofortified food over conventional varieties (e.g., price). A critical consideration for example is how prices for biofortified crop varieties compare to those for conventional varieties, since these will determine affordability of the biofortified staple crops and food products, and similarly influence demand from consumers - the fourth condition.

This having been said, the considerable advocacy that HarvestPlus and its partners have deployed to promote investment in biofortification and adoption by countries has contributed to the release of biofortified crops in at least 22 African countries (Covic et al., 2017). Biofortification is an intervention, but is there a need for policy processes and coordination at the level needed for more intensive, and we would suggest disruptive, interventions? There has been considerable progress in integrating biofortification into regional and national policies as evidenced by some national nutrition strategies that include biofortification (e.g., Bangladesh, Malawi, Nigeria, Pakistan, Tanzania, and Uganda) and a greater number that include biofortified crops in national agriculture and nutrition plans (Bouis, 2018).

Conclusion

By enriching the micronutrient content in staples foods that are consumed by those most at-risk of micronutrient deficiencies, biofortification has the potential to be a part of the solution to the persistent problem of micronutrient deficiencies affecting many low- and middle-income countries. Significant progress in crop development and research on nutrient retention, bioavailability, and efficacy of biofortified foods has been made over the past two decades.

At the household level, the value of biofortified crops can be realized by combining the adoption of biofortified varieties by smallholder farmers and procurement of biofortified foods and ingredients by consumers with the most appropriate food preparation and cooking methods that maximize the bioavailability of different nutrients. The fact that selection of varieties also focuses on high yield or resilience co-factors helps acceptance and adoption across the value chain. Moving forward there is a need to generate evidence on the effectiveness of biofortification in real-world settings, explore and test innovative market delivery models, and ensure coherent policies are in place that create an enabling environment for biofortification.

Conflict of interest

The views expressed in this publication are those of the author(s) and do not necessarily reflect the views of FAO.

References

Afolami, I., Mwangi, M.N., Samuel, F., Boy, E., Ilona, P., Talsma, E.F., Feskens, E., Melse-Boonstra, A., 2021. Daily consumption of pro-vitamin A biofortified (yellow) cassava improves serum retinol concentrations in preschool children in Nigeria: a randomized controlled trial. Am. J. Clin. Nutr. 113 (1), 221–231.

Beyer, P., Al-Babili, S., Ye, X., Lucca, P., Schaub, P., Welsch, R., Potrykus, I., 2002. Golden rice: introducing the β-carotene biosynthesis pathway into rice endosperm by genetic engineering to defeat vitamin A deficiency. J. Nutr. 132 (3), 506S–510S.

Bouis, H.E., 2003. Micronutrient fortification of plants through plant breeding: can it improve nutrition in man at low cost? Proc. Nutr. Soc. 62 (2), 403–411.

Bouis, H.E., 2018. In: Mannar, M.G.V., Hurrell, R.F. (Eds.), Biofortification: An Agricultural Tool to Address Mineral and Vitamin Deficiencies. Food Fortification in a Globalized World Academic Press, pp. 69–81.

Council for Agricultural Science and Technology (CAST), 2020. Food Biofortification—Reaping the Benefits of Science to Overcome Hidden Hunger—A Paper in the Series on the Need for Agricultural Innovation to Sustainably Feed the World by 2050. CAST, Ames, Iowa.

Covic, N., Low, J., MacKenzie, A., Ball, A., 2017. Advocacy for biofortification: building stakeholder support, integration into regional and national policies, and sustaining momentum. Afr. J. Food Nutr. Sci. 17 (2), 12116–12129.

de Valença, A.W., Bake, A., Brouwer, I.D., Giller, K.E., 2017. Agronomic biofortification of crops to fight hidden hunger in sub-Saharan Africa. Glob. Food Secur. 12, 8–14.

Díaz-Gómez, J., Twyman, R.M., Zhu, C., Farré, G., Serrano, J.C., Portero-Otin, M., Muñoz, P., Sandmann, G., Capell, T., Christou, P., 2017. Biofortification of crops with nutrients: factors affecting utilization and storage. Curr. Opin. Biotechnol. 44, 115–123.

Hotz, C., 2013. In: Caballero, B. (Ed.), Biofortification. Encyclopedia of Human Nutrition, third ed. Academic Press, Waltham, pp. 175–181.

Kangile, R., Bakuza, E., Kuboja, N., Tairo, F., 2021. Can orange fleshed sweet potatoes processing trigger farmers to use quality improved planting materials? Empirical evidence from selected regions in Tanzania. J. Dev. Agric. Econ. 13 (1), 34–44.

La Frano, M.R., de Moura, F.F., Boy, E., Lönnerdal, B., Burri, B.J., 2014. Bioavailability of iron, zinc, and provitamin A carotenoids in biofortified staple crops. Nutr. Rev. 72 (5), 289–307.

Lowe, N.M., Zaman, M., Moran, V.H., Ohly, H., Sinclair, J., Fatima, S., Broadley, M.R., Joy, E.J.M., Mahboob, U., Lark, R.M., Zia, M.H., Ander, E.L., Sharp, P.A., Bailey, E.H., Young, S.D., Khan, M.J., 2020. Biofortification of wheat with zinc for eliminating deficiency in Pakistan: study protocol for a cluster-randomised, double-blind, controlled effectiveness study (BIZIFED2). BMJ Open 10 (11), e039231.

Mazariegos, M., Hambidge, K.M., Krebs, N.F., Westcott, J.E., Lei, S., Grunwald, G.K., Campos, R., Barahona, B., Raboy, V., Solomons, N.W., 2006. Zinc absorption in Guatemalan schoolchildren fed normal or low-phytate maize. Am. J. Clin. Nutr. 83 (1), 59–64.

Pfeiffer, W.H., McClafferty, B., 2007. HarvestPlus: breeding crops for better nutrition. Crop Sci. 47 (S3), S-88.

Talsma, E.F., Melse-Boonstra, A., Brouwer, I.D., 2017. Acceptance and adoption of biofortified crops in low- and middle-income countries: a systematic review. Nutr. Rev. 75 (10), 798–829.

Further reading

Foley, J.K., Michaux, K.D., Mudyahoto, B., Kyazike, L., Cherian, B., Kalejaiye, O., Ifeoma, O., Ilona, P., Reinberg, C., Mavindidze, D., 2021. Scaling up delivery of biofortified staple food crops globally: paths to nourishing millions. Food Nutr. Bull. SAGE Publications Sage CA: Los Angeles, CA. 42 (1), 116–132.

Relevant websites

Consultative Group on International Agricultural Research, http://www.cgiar.org.
HarvestPlus, http://www.harvestplus.org.
International Potato Center (CIP), http://www.cipotato.org.

Cereal grains

Sumathi Swaminathan, St John's Research Institute, Bangalore, Karnataka, India

© 2023 Elsevier Ltd. All rights reserved.
This is an update of R.K. Price, R.W. Welch, Cereal Grains, In Encyclopedia of Human Nutrition (Third Edition), edited by Benjamin Caballero, Academic Press, 2013, https://doi.org/10.1016/B978-0-12-375083-9.00047-7.

Introduction **293**
Types of cereal grains **294**
Maize 295
Wheat 295
Rice 295
Millets 295
Barley 295
Sorghum 295
Oats 296
Rye 296
Triticale 296
Production and consumption of cereal grains **296**
Structure of cereal grains **297**
Post-harvest storage of cereal grains **298**
Processing of cereal grains **298**
Milling 299
Germination 299
Heat treatment 300
Fermentation and parboiling 300
Nutrient composition of cereal grains **300**
Carbohydrates 300
Proteins 301
Fats 302
Micronutrients 302
Bioactive compounds 302
Anti-nutrient compounds 303
Contaminants 303
Cereals grains and adverse health outcomes **303**
Gluten allergy 303
Association with non-communicable diseases 303
Conclusions **304**
Conflicts of interest **304**
References **304**

Key points

- Worldwide cereals grains are staple foods that provide a large proportion of energy, carbohydrates, proteins in the human diet
- Wheat, maize and rice are the three most highly produced and consumed cereal grains
- All cereal grains go through some level of processing that determines the quantity and quality of nutrients ingested
- Different varieties of cereals and cereal-based food products when routinely consumed are associated with various health outcomes

Introduction

The edible monocotyledonous seeds from grasses, both cultivated and wild, of the *Poaceae* or *Gramineae* family are termed cereal grains. The term cereal originated from the word "Ceres", the ancient Roman god of agriculture. Cereals have served as the primary staple food for human beings from ancient times ensuring food security. Its consumption contributes to approximately 50% of

caloric intake (Food and Agricultural Organization, 2016, 2020). Not only are over half of the cereals produced used for human consumption, but they are also used for consumption by livestock, poultry, and other birds. Further, they are the primary source for production of alcohol.

From an evolutionary perspective, cereal grains were ingested long before domestication of crops when humans were hunter-gatherers. The realization that cereal grains can be stored and used as a safe reliable food source may have served as a stimulus for humans to dwell close to where cereal crops were cultivated and become an agrarian society. Cereal grains were preferred as a source of food mainly because of the possibility of longer storage periods and the preservation of available nutrients in the grain compared to perishable foods (Wrigley, 2017b). Historically, domestication of maize first occurred in Mexico or Central America, wheat in the Far East and rice in China (Awika, 2011). Processing and cooking of cereals too probably began as early as 105,000 years ago when cereal-based snacks were eaten. Evidence from Africa where residues of grass seeds, mainly sorghum was found on implements were suggestive of the use of some form of processing technique to make the cereal grains conducive to consumption (Thielecke et al., 2020; Curry, 2021).

Cereals crops are the dominant agricultural produce grown over 60% of cultivable land available worldwide (Koehler and Weiser, 2013). There are several cultivars of each type of cereal grain with cultivation dependent on climate, temperature, availability of water, soil conditions prevailing in each region, apart from other cultural and economic factors (Awika, 2011).

A large proportion of food energy, protein, some of the B complex vitamins and minerals is contributed by cereals in human diets mainly because of the quantity of cereals consumed. The developing world relies more on cereal foods for their nutritional needs compared to the developed countries (Awika, 2011; Wrigley, 2017b). With the advent of processing and further advances in technology especially over the last century, changes in the form of consumption of cereals has occurred.

Types of cereal grains

Global production and consumption of cereals are largely dominated by wheat, maize and rice followed by sorghum, barley, oats, rye, millets, and other cereals. Within each species of cereal grains, diverse varieties exist due to breeding for improvement of agronomic, technological, and nutritional properties. Some cereals are also termed coarse cereals. Coarse cereals exclude rice and wheat and refers to sorghum (*Sorghum vulgare*), maize (*Zea mays*), oats (*Avena sativa*), barley (*Hordeum vulgare*), buckwheat (*Fagopyrum esculentum*) and some millets species, namely pearl millet (*Pennisetum glaucum*), finger millets (*Eleusine coracana*), foxtail millets (*Setaria italica*) and kodo millet (*Paspalum setaceam*), etc (Fu et al., 2020). Pseudo cereals are dicotyledenous grains resembling cereals but not part of the Gramineae family. Buckwheat (*Polygonaceae*), amaranth (*Amaranthaceae*), chia (*Salvia hispanica*) and quinoa (*Chenopodium quinoa*) are the commonly consumed varieties of pseudo cereals. They are used as grains or ground into flour (Békés et al., 2017). The cereal species and the region of cultivation are provided in **Table 1**.

Generally, all cereals grains are harvested only once annually. Maize, rice, sorghum and millets are tropical grasses grown in regions with temperate climates under frost free conditions, while wheat, rye, barley, and oats are grown in regions with a moderate climate during cooler seasons. Cereals such as sorghum and millets can adapt in arid conditions and are tolerant to drought (Awika, 2011; Koehler and Weiser, 2013).

Table 1 Cereal species.

Type	*Sub-family*	*Genus and species mainly used*	*Region grown*
Cereals			
Barley	Pooideae	*Hordeum vulgare*	Arctic and semi-arid zones
Rye	Pooideae	*Secale cereale*	Arctic zone, temperate zone
Wheat (common, bread)	Pooideae	*Triticum aestivum*	Temperate zones
Wheat (durum)	Pooideae	*Triticum durum*	Temperate zones
Triticale		×Triticosecale sp.	Arctic zone, temperate zone
Oats	Pooideae	*Avena sativa*	Temperate and sub-tropical zone
Maize (corn)	Andropogonoideae	*Zea mays*	Tropical, subtropical, temperate
Sorghum	Andropogonoideae	*Sorghum bicolor*	Tropical, sub-tropical regions
Millets (proso, pearl. Foxtail)	Panicoideae	Pennisetum	Tropical, subtropical, semi-arid
Millets (finger millet, teff)	Cynodonteae	Eleucine coracana	Tropical, subtropical, semi-arid
Rice	Oryzoideae	*Oryza sativa*	Tropical, subtropical
Pseudocereals			
Amaranth	Amaranthaceae	Amaranth caudatus; amaranth cruentus; amaranth hypochondriacus	Tropical, subtropical
Buckwheat	Polygonaceae	*Fagopyrum esculentum*	Tropical, high altitude
Quinoa	Chenopodiaceae	*Chenopodium quinoa*	Tropical, subtropical, temperate

Wheat, rye and barley are from the sub family Pooideae, tribus Triticeae, with oats being distantly related under the same subfamily. In evolutionary terms, rice, maize or corn, sorghum and millets are from a different species (Koehler and Weiser, 2013).

Maize

Maize (*Zea mays*) also called as corn in some countries, is largely grown and consumed as a staple in Africa and North and South America. It grows in warmer climates and regions with adequate rainfall and is generally sown in spring and harvested in summer (Awika, 2011). The use of maize as food is projected to increase primarily in Sub-Saharan Africa where population growth is expected to rise. Maize grown across the world are of different colors (white, yellow, red, black). The maize grown in Africa and Central America are mostly white, while that grown in the United States of America are both white and yellow (Ranum et al., 2014). Maize, particularly white maize, will remain an important staple, accounting for about a quarter of total caloric intake of the world.

Wheat

Wheat of the Triticum family, is the second most largely produced cereal in the world (OECD/FAO, 2021). It is mainly grown in the temperate zones and adapts itself to different agro-climatic conditions. Of the manifold species grown, the main commercial varieties grown are *T. aestivum* subspecies Vulgare and the hard wheat *T. durum*. Over 90% of wheat grown is *T. aestivum*. These are varieties that can be easily threshed. Wheat grows in a wide range of environments from low to high water availability and wide-ranging temperatures. Wheat grows in temperate regions both in winter and summer and is the predominant staple in North Africa and West and Central Asia although populations across the world ingest them (McKevith, 2004; Uthayakumaran and Wrigley, 2017).

Rice

Rice (*Oryza Sativa*) is consumed by over 65% of the population largely in Asia followed by Latin America. Most rice is produced in Asia, with over half of the world's production from China and India (OECD/FAO, 2021). Rice, the major staple food in Asia. It is cultivated in standing water (approximately 5–10 cm depth) in tropical and subtropical climates and sometimes, on dry uplands (MacEvilly, 2003; OECD/FAO, 2021). The average estimated per capita consumption per year between the years 2018–2020 was highest in Asia (77.2 kg) followed by Latin America and Caribbean (28.0 kg), Africa (27.4 kg), Europe (20.7 kg), Oceania (13.5 kg) and North America (6.3 kg) (OECD/FAO, 2021). Rice is generally grown in warmer tropical and sub-tropical regions with high rainfall and requires a lot of water for production and hence production is highest in Asia. However, some varieties also thrive in temperate regions (Fitzgerald, 2017b). Wild rice is unrelated to rice but has a higher protein content than rice. It is from a North American plant *Zizania aquatica* which is both difficult to harvest and more expensive compared to other grains (McKevith, 2004).

Millets

Millets are small, grained cereals belonging to a separate sub family *Panicoidacea* and dince ancient times have been used by mankind in Asia and Africa (Belitz and Grosch, 2009). Pearl millet (*Pennisetum glaucum*), finger millet (*Eleusine coracana*), little millet (*Panicum sumatrense*), Italian or foxtail millet (*Setaria italica*), barnyard millet (*Echinochloa crusgalli*), proso or common millet (*Panicum miliaceum*) and kodo millet (*Paspalum scrobiculatum*) are the millets most consumed. Most millets are grown in the tropical and sub-tropical regions, while some are grown in the temperate regions. Millets form the staple food of populations inhabiting the arid and semi-arid tropics of the world and serve as food in most of the Asian and African countries and parts of Europe (MacEvilly, 2003; Taylor and Duodu, 2017). The ability of millets to adapt to extreme ecological conditions particularly in dry lands and other nutritional attributes makes it a sustainable crop for the future (Bandyopadhyay et al., 2017).

Barley

Barley (*Hordeum vulgare*) grows mainly in the colder regions and is thus largely grown in northern Europe, United States of America and Canada. Nevertheless, being a versatile crop, with a short period for growth, it adapts to a wide range of environments and geographical locations. The crop is tolerant to drought and saline soil, too (Győri, 2017). Barley is used as animal and human feed and to produce alcoholic beverages, but most barley is mainly malted and consumed. It is consumed as food in a few countries in Europe, and the Middle and Far East (McKevith, 2004; Győri, 2017).

Sorghum

Sorghum (*Sorghum bicolor*) is a crop grown in tropical and subtropical regions with 90% grown in Africa and Asia. It is the staple food in Africa and some parts of Asia and the Middle East (Leff et al., 2004; McKevith, 2004). Sorghum produced in North and Central America, South America and Oceania are primarily used for animal feeds (McKevith, 2004). It grows in warm climates and cannot tolerate low temperatures. Sorghum is both a pest and disease resistant crop.

Oats

Oats (*Avena sativa*) is a grass grown annually in cold regions. A small quantity that is produced is used for human consumption while most is used as animal feed and for other purposes (McKevith, 2004; Butt et al., 2008).

Rye

Rye (*Secale cereale*) is grown mainly in the colder regions, in high altitudes and semi-arid regions and is a winter crop (MacEvilly, 2003; McKevith, 2004; Wrigley and Bushuk, 2017). The principal use of the crop is as food and as a forage crop. Most rye is produced in Russia, the European countries of Poland, Germany and the Scandinavian countries and Canada where rye bread is mainly prepared (McKevith, 2004; Wrigley and Bushuk, 2017).

Triticale

A hybrid of durum wheat and rye is Triticale (x Triticosecale). The hybrid variety was developed for the purpose of integrating the agronomic properties of growth of rye in winter and poor soil conditions and the baking qualities of wheat (Wrigley and Bushuk, 2017).

Production and consumption of cereal grains

Cereals are produced in about 50% of the world's harvested area, of which wheat constitutes 14% of the global crop area. Among cereal grains, 40% of wheat is produced in China, India and the Russian federation, 65% of maize in the United States of America, China and Brazil and about 60% of rice in China, India and Indonesia (Food and Agricultural Organization, 2020). The current figure for production of cereal grains is 2817.2 million tonnes (Food and Agricultural Organization, 2021). (http://www.fao.org/worldfoodsituation/csdb/en/). The highest produced cereal grain is maize, followed by wheat, rice and other coarse grains (**Fig. 1**). Rice and maize are predominantly produced in the developing countries (**Table 2**).

Cereal grains have evolved from being the main or sole purpose of cultivation as food to being utilized for other purposes. Of the total cereal grains produced about 50% are utilized as human food. 35–40% as animal feed, and the rest as raw material for biofuels or other uses like for production of starch and related substances, malting and for production of alcoholic beverages. These proportions vary across different cereal grains and across countries. About 90% of rice grown is consumed as human food in the regions produced, except in some countries in Europe where it is primarily used as animal feed. The primary usage of oats is as human food in Britain and Brazil, but as animal feed in other parts of the world. A large proportion of most coarse grains produced, except for millets are used as animal feeds and for industrial purposes (Batey, 2017; OECD/FAO, 2021). In countries such as Mexico, most of the maize or corn produced is consumed as food, whereas, in the United States of America, a large proportion is processed industrially for starch and oil production (Wrigley, 2017b) although it is also used for the manufacture of glucose, beer and alcohol. The consumption in the various regions indicate that the per capita food consumption of maize is the highest in Latin America and Africa, wheat in Europe, Rice in Asia and other coarse grains in Africa (**Table 3**).

Growth in feed use of cereals is expected to continue exceeding the rate of expansion for food use. Feed use of cereals is projected to grow at 1.2% p.a. over the coming decade as livestock production expands and intensifies in low and middle-income countries, compared to a projected growth of 1% per annum for food use (OECD/FAO, 2021).

Production of cereals is likely to be significantly affected by temperature and carbon-dioxide emissions. An increase in carbon dioxide emissions is likely to increase temperatures and absolute humidity which in turn affects the soil's moisture content and the duration of the growing season. The higher temperatures are likely to accelerate crop maturity and result in water stress due to rise in temperature. Therefore, there can be lesser yields of cereals such as barley, wheat and oats. However, the increase in water stress may

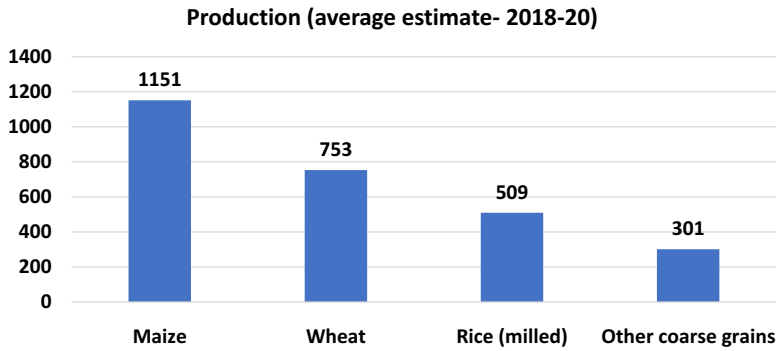


Fig. 1 Production of major cereals.

Table 2 Production of major cereal grains by region.

	Production (in million tonnes, MT)			
Region	Maize	Wheat	Rice	Other coarse grains
World	1151.4	752.7	509.3	301.0
Developed countries	515.4	384.9	17.7	181.0
Developing countries	636.4	367.8	491.6	120.0
North America	370.4	84.6	6.5	28.0
Latin America	206.7	31.4	18.6	19.8
Europe	126.5	251.6	2.8	132.7
Africa	88.1	27.1	24.3	57.8
Asia	359.2	336.2	456.8	50.0
Oceania	445.0	21.7	0.2	12.7

Source: OECD-FAO (2021).

Table 3 Consumption of major cereals by region.

	Consumption (average, 2018–20 est)							
	Maize		Wheat		Rice (milled)		Other coarse grains	
Region	Total consumption (MT)	Food (kg/ capita)	Total consumption (MT)	Food (kg/ capita)	Total consumption (MT)	Food (kg/ capita)	Total consumption (MT)	Food (kg/ capita)
World	1166.3	18.9	748.7	67.6	506.3	54.4	287.8	10.5
North America	323.9	17.6	39.5	79.6	5.0	13.5	20.1	4.4
Latin America	171.1	50.5	40.6	55.8	19.4	28.0	18.2	3.5
Europe	109.6	8.4	181.4	106.7	4.8	6.3	110.6	13.7
Africa	107.3	43.9	76.1	50.5	40.8	27.4	59.6	32.8
Asia	454.5	9.4	401.5	66.8	435.5	77.2	73.6	5.2
Oceania	0.5	2.3	9.6	69.0	0.9	20.7	5.6	6.3

Source: OECD-FAO (2021).

be beneficial for cereal crop yields of maize and sorghum. Decreased rainfall particularly in the African and Asian countries due to climate change with magnitude and timing of rainfall being affected can contribute to change in crop yields (Grace et al., 2019).

Sustainable cereal production is being encouraged in farming systems by adopting save and grow methods of conservation agriculture practices, improvement of soil health, improvement of yields and diversification of crops and efficiency in water and pest management (Food and Agricultural Organization, 2016). For example, across the world, wheat farmers grow legumes through crop rotation. This ensures nitrogen fixation of the soil and improves wheat yields. In South and Southeast Asian countries, rice farmers grow high yielding hybrid maize which require less water in the dry season (Food and Agricultural Organization, 2016).

The consumption of cereals is likely to be underestimated since direct consumption of cereals rather than indirect consumption from products such as corn-based sweeteners which provide additional food energy are not considered for per capita consumption (Awika, 2011). Maize is one of the most important commodities used as animal feed (**Fig. 2**). Since the early 2000s, demand in Asian middle-income countries have fueled the drive for production of biofuels from cereals. This increase has been due to the implementation of policies with three main objectives that is, support countries' commitments to reduce their carbon dioxide (CO2) emissions, reduce the dependency on imported fossil fuels and create additional demand for feedstock crops to support domestic producers (OECD/FAO, 2021).

Structure of cereal grains

Botanically, a cereal grain is referred to as a kernel or caryopsis. Caryopsis refers to a single seeded indehiscent fruit where the pericarp or fruit coat is fused with the seed coat. Although cereal grains may be taxonomically uniform, between and within the same species of cereals, their morphology and composition vary and this determines their utilization (Koehler and Weiser, 2013; Wrigley, 2017a). The size and weight of the grain varies from large, with maize having the highest weight of about 350 mg to the small, seeded millets which are about 9 mg in weight (Koehler and Weiser, 2013). Apart from the kind of cereal, the cultivar and the technique for crop production determines the kernel size (weight in gram per 1000 kernels) (Belitz and Grosch, 2009).

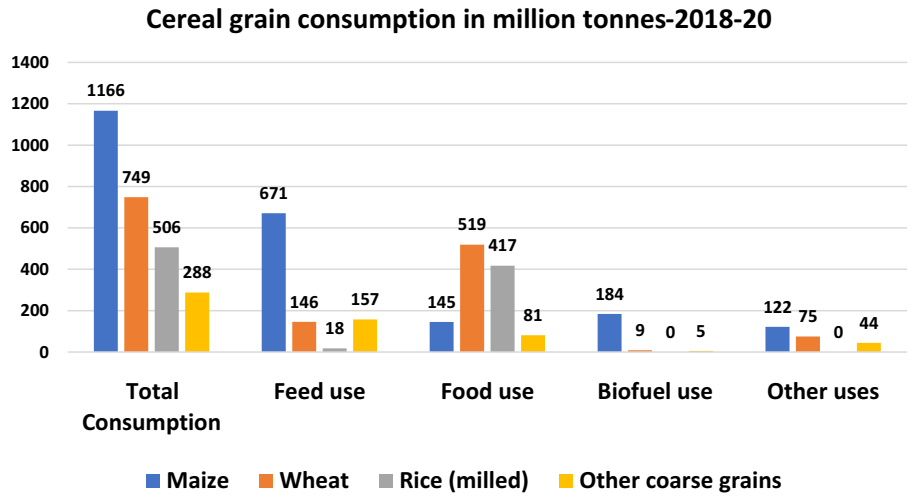


Fig. 2 Consumption of major cereals. Source: OECD-FAO (2021).

All grains have an outer protective hard inedible hull or husk or chaff (Slavin, 2004; Jones et al., 2015). Beneath the hull, all cereal grains consist of 3 parts, that is the seed coat, embryo (germ) and endosperm characteristic to each species. The embryo is separated from the endosperm by the scutellum which during germination mobilizes the reserves of food. The seed coat is the thick-walled outer structure and consists of the outer wall called the pericarp, followed by a semipermeable layer, testa and then a permeable layer of thick-walled cells, the aleurone, altogether called the bran. The endosperm is covered by the aleurone layer, either by a single layer as in wheat, maize, rye, oats and sorghum or by three layers as in rice and barley (McKevith, 2004).

Cereal grains have a moisture content of 11–14%, 56–74% carbohydrates, 8–11% of proteins, 2–4% lipids, and 1–3% minerals (Koehler and Weiser, 2013). The seedcoat mainly contains about 2–13% fiber, cellulose, hemicellulose, proteins and lignins. The embryo contains protein, minerals, fat. The endosperm contains about 70–80% starch, and some amounts of protein (Koehler and Weiser, 2013; McKevith, 2004).

Wheat grains are ovoid in shape. The kernel is classified by color, that is, red or white, soft or hard and winter or spring. Compared to other cereal kernels, the seed and the embryo of maize (10–13%) of grainare the largest (Wrigley, 2017a).

Post-harvest storage of cereal grains

Before cereal grains are processed, they are stored at the farm or at a collection center. Prior to storage, the kernels need to be dried either through natural (wind/solar) or mechanical drying mechanisms. Drying minimizes the nutrient losses. Post-harvest losses because of mold spoilage, infestation by pests and germination of the grain due to the moisture content present in the grain, affect grain quality. A higher moisture content during storage and microbial amylases results in breakdown of starch affecting quality. Oxidation of unsaturated acids can result in rancidity and flavor can be affected (McKevith, 2004).

Rice is harvested as paddy, wherein the caryopsis is within the hull which provides protection against insects, pests and adverse climatic conditions (Fitzgerald, 2017a).

Processing of cereal grains

Some form of processing is required before cereal grains can be consumed to render the indigestible carbohydrate digestible and to increase palatability. Even before domestication of cereals, processes such as cooking, sprouting, pounding, grinding or milling, and fermentation were employed (Dietrich et al., 2019; Hübner and Arendt, 2013). From the early days, cracking, flattening, grinding, and milling were done using rudimentary tools (Curry, 2021; Jones et al., 2016). By milling, cereal grains are rendered esculent by removal of the husk and sometimes the bran, and if required, ground into flour of varying particle size. The level of milling determines the level of nutrients removed (Oghbaei and Prakash, 2016).

Traditional methods of cooking used are soaking and germination of whole cereals. In some instances, kernels are soaked and then cooked in its whole form. This removes some of the toxins or bitter components and some antinutrients and increase nutrient availability, importantly for maize. These processes permit water to enter the kernel (Jones et al., 2016). Processing has evolved over the years to improve the life and retard the growth of microbes and toxins, increase nutrient availability, improve the sensory characteristics and functional properties, for convenience, and to reduce waste and any losses (Thielecke et al., 2020).

Between harvest and consumption, all cereals undergo a series of processes, starting from preparing the harvested grain for storage (threshing, winnowing, drying to reduce moisture content), primary processing to clean, sort and remove the inedible

husk by milling, pounding, grinding, tempering, parboiling, soaking, drying, sieving etc and finally secondary processing through fermentation or thermal treatment (baking, puffing, flaking, roasting, frying and extrusion) for utilization of cereal grains in the form of whole grains, flakes or flour (Tiwari and Pojić, 2020; Thielecke et al., 2020). Whole grains are consumed whole or intact, cracked or flaked but with the endosperm, germ and bran present in the same proportions as the original grain composition. With rapid advances in technology recombining the various components of grain, that is, bran, the heat-treated germ, and the milled white flour in different proportions compared to the intact grain (25%, 50%, 75%, or 100% of the germ and aleurone layer) is possible (Jones et al., 2016; Swaminathan et al., 2021). Rice grains are usually milled only after a storage period of at least 3–4 months so that the milled rice can expand more during cooking (McKevith, 2004).

Milling

There are slight variations in the way different cereals are milled. But in general, most grains undergo the process of grinding, sifting, separation of fractions and regrinding (McKevith, 2004). Both dry and wet milling techniques are used. For maize, dry milling is used for producing grits, flour, and corn germ, while wet milling is used for starch production, dextrose and syrups for use in food (Slavin et al., 2000; Gyŏri, 2017). Traditional methods use mortar and pestle or stones for grinding into flour and this removes the outer layer but retains the germ, while the modern methods strip the maize grains of the germ too and make them more refined (Slavin et al., 2000). As rye has a softer kernel than wheat, the wet milling technique is used (McKevith, 2004). Wet milling is done in cereals to toughen the pericarp and germ so that they do not break during milling.

The type and level of processing further determines the quantity and quality of nutrients consumed and varies with the type of cereal grain (Oghbaei and Prakash, 2016; Thielecke et al., 2020). The outer layers of a grain are rich in antinutrients, and these are removed by dehulling. Heat treatments not only gelatinize starch but also help improve storage stability and may aid in retaining nutrients such as vitamin E by the inactivation of some enzymes, thus preventing rancidity (Oghbaei and Prakash, 2016). A distinction exists between processing and refining. Generally, if the different parts of the kernel remain as per the original ratio, they are said to be processed. However, in refined grains, the outer husk, bran and germ are removed by milling (Jones et al., 2016). In the case of rice paddy, the nutrients lost depends on the degree of polishing it undergoes (Oghbaei and Prakash, 2016). Paddy is dehulled and separated as brown rice and bran, after which to produce white rice, whitening or pearling and polishing are done. Barley undergoes dehulling and pearling to produce pearl barley, flaked or milled into barley flour (Slavin et al., 2000; McKevith, 2004).

After the grain is crushed during milling, the milled grain is separated by particle size into different milling streams and then crushed further depending on the type of grain required (Borneo and León, 2012; Jones et al., 2016; Oghbaei and Prakash, 2016). If the whole grain is directly made into flour it will result in whole grain flour or meal which are termed as flour with high extraction rates of 90%, indicating a high proportion of the whole grain in the flour. However differential milling is also done to separate the fractions as for example in wheat where through roller milling refined wheat flour, bran, germ, semolina etc. can be obtained. Low extraction flours have less than 75% of the whole grain and are considered as refined grains mostly consisting of the endosperm and miniscule amounts of bran. Thus, refined grains have less of vitamins, minerals, fiber and bioactive compounds than whole grain flours, and this loss may be mitigated by adding the nutrients lost during milling by enrichment of the flour. The nutrients usually added are dependent on the type of cereal flour and the needs of the specific population consuming the flour. The nutrients that can be added are thiamine, riboflavin, niacin, pyridoxine and iron, but not fiber. On the other hand, flours can be fortified by either adding nutrients more than the quantity originally present in the grain or by adding a nutrient not present, as for example the addition of folic acid (Jones et al., 2016; Oghbaei and Prakash, 2016). Rye is milled into flour with different extraction rates for production of soft rye breads, coarse rye meal for preparation of hard breads and rye flakes for breakfast cereals (Slavin et al., 2000).

The embryo and aleurone comprise only 10% of the dry weight of the grain and is usually removed by milling in wheat, polishing in rice, pearling in barley or decortication in sorghum, before it is consumed by humans. The concentration of fiber, protein, vitamins and minerals decrease from the outer bran layer to the inner part of the seed and depends on the extent to which the different fractions of the grain are removed and the type of cereal grain. The milling of wheat to white flour results in a decrease of 27% thiamine, 20% niacin and 19% riboflavin. Further bleaching removes nearly 98% of vitamin E (MacEvilly, 2003). During milling mechanical changes to the starch occur that can increase the level of enzyme activity which is beneficial in bread making. The extent to which this change occurs is dependent on the quality of the grain and the type of milling done. Denaturation of protein is likely due to high heat treatment during milling resulting in lower gluten yield and decreased absorption of water in the flour (McKevith, 2004).

Germination

Germination and malting (germination under controlled conditions) are techniques that have been used since ancient times and not only change the taste, flavor and appearance of the grain, but also improve the nutrient content while decreasing the amount of antinutrients present thus improving the digestibility and availability of nutrients (Hübner and Arendt, 2013; Oghbaei and Prakash, 2016).

Heat treatment

Heat treatment such as boiling, baking, frying and extrusion of cereal-based foods can cause losses in some vitamins. Antioxidant activities increase by processes such as baking or toasting by the browning reaction due to the Maillard reaction wherein reducing sugars interact with amino acids. Some heat treatments hydolyse the phytates and thus increase the availability of iron, calcium and zinc (McKevith, 2004). Wheat is used to make a variety of food products such as biscuits, cookies, cakes, pasta, noodles, bread and many more. With globalization, wheat products such as ready to eat cereals and bread are now eaten across the world (Thielecke et al., 2020). The process of extrusion, a heat treatment where a shaped product is extruded through a screw press under heat is used for manufacturing products such as pasta, noodles and ready-to-eat breakfast cereals (McKevith, 2004). Among maize products, popcorn prepared using intense heat is an intact whole grain (Slavin et al., 2000).

Fermentation and parboiling

Fermentation of several cereal products such as bread, maize, rice etc. are used in various regions. The carbon-di-oxide produced during fermentation increases the volume of the dough. Protein digestibility is improved by bacterially produced lysine and bioavailability of minerals are increased probably by hydrolization of phytates (McKevith, 2004).

Rice is also eaten parboiled, when unhusked rice is soaked for 4–5 h in water at a temperature of 65 °C, steamed under pressure and then dried and milled. Parboiling reduces the loss of nutrients by the movement of B vitamins and oil from the bran into the endosperm (McKevith, 2004).

Nutrient composition of cereal grains

Cereal grains are eaten across the world in substantial quantities, thus providing the highest proportion of energy from food apart from proteins and micronutrients. **Tables 4–7** provide the macronutrient, essential amino-acid contents, mineral and vitamin content per 100 g of cereal or cereal-based foods.

Carbohydrates

Generally cereal grains contain about 60–80% carbohydrates, mainly as starch of which 70–75% is as amylopectin and the rest amylose. The proportion varies by the cereal grain especially in rice, maize, barley and sorghum. A higher proportion of amylase is associated with increased satiety, lower glycemic index, and insulin release. Among the cereals, those which are waxy cultivars have higher amylopectin. Apart from starch, non-starch polysaccharides (NSP), or dietary fiber, in the outer layers of the grain including arabinoxylans and beta glucans are also present (Koehler and Weiser, 2013; Wrigley, 2017a; Thielecke et al., 2020).

Both insoluble and soluble forms of NSP or dietary fiber are present; the insoluble dietary fiber contents are similar across cereals, but the soluble fraction differs (McKevith, 2004). A high content of dietary fiber is present in barley. Arabinoxylans are the principal water-soluble NSPs in wheat, rye and barley and beta glucans in oats. The beta-glucan content in oats makes it difficult to refine oats and it thus remains a whole grain (Fu et al., 2020).

Table 4 Macronutrient composition of cereals and pseudo cereals (per 100 g).

Type	*Water content (g)*	*Energy (kcal)*	*Carbohydrate (g)*	*Dietary fiber (g)*	*Fat (g)*	*Protein (g)*
Cereals						
Wheat (durum)	10.9	339	71.1	NA	2.5	13.7
Wheat flour, whole grain	10.7	340	72.0	10.7	2.5	13.2
Wheat flour, white, unenriched	11.9	364	76.3	2.7	1.0	10.3
Bread, white wheat	35.2	274	47.5	4.0	4.5	10.7
Corn grain, white (maize)	10.4	365	74.3	NA	4.7	9.4
Rice, white, medium-grain, raw, unenriched	12.9	360	79.3	NA	0.6	6.6
Sorghum, grain	12.4	329	72.1	6.7	3.5	10.6
Barley, pearled, raw	10.1	352	77.7	15.6	1.2	9.9
Rye, grain	10.6	338	75.9	15.1	1.6	10.3
Oats	8.2	389	66.3	10.6	6.9	16.9
Millet, raw	8.7	378	72.8	8.5	4.2	11.0
Pseudo cereals						
Buckwheat	9.8	343	71.5	10.0	3.4	13.2
Amaranth grain, uncooked	11.3	371	65.2	6.7	7.0	13.6
Quinoa, uncooked	13.3	368	64.2	7.0	6.1	14.1

Source: USDA Food Composition Database. https://fdc.nal.usda.gov/. Accessed September 7, 2021.

Table 5 Amino acid content of cereals and pseudo cereals (per 100 g).

Type	Histidine (g)	Isoleucine (g)	Leucine (g)	Lysine (g)	Methionine (g)	Cystine (g)	Phenylalanine (g)	Tyrosine (g)	Threonine (g)	Tryptophan (g)	Valine (g)
Cereals											
Wheat (durum)	0.322	0.533	0.934	0.303	0.221	0.286	0.681	0.357	0.366	0.176	0.594
Wheat flour, whole grain	0.357	0.443	0.898	0.359	0.228	0.275	0.682	0.275	0.367	0.174	0.564
Wheat flour, white, unenriched	0.23	0.357	0.710	0.228	0.183	0.219	0.520	0.312	0.281	0.127	0.415
Bread, white wheat	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Corn grain, white (maize)	0.287	0.337	1.160	0.265	0.197	0.170	0.463	0.383	0.354	0.067	0.477
Rice, white, medium-grain, raw, unenriched	0.155	0.285	0.546	0.239	0.155	0.135	0.353	0.221	0.236	0.077	0.403
Sorghum, grain	0.246	0.433	1.490	0.229	0.169	0.127	0.546	0.321	0.346	0.124	0.561
Barley, pearled, raw	0.223	0.362	0.675	0.369	0.190	0.219	0.556	0.284	0.337	0.165	0.486
Rye, grain	0.189	0.208	0.563	0.286	0.153	NA	0.435	0.200	0.289	0.108	0.317
Oats	0.405	0.694	1.280	0.701	0.312	0.408	0.895	0.573	0.575	0.234	0.937
Millet, raw	0.236	0.465	1.400	0.212	0.221	0.212	0.580	0.340	0.353	0.119	0.578
Pseudo cereals											
Buckwheat	0.309	0.498	0.832	0.672	0.172	0.229	0.520	0.241	0.506	0.192	0.678
Amaranth	0.389	0.582	0.879	0.747	0.226	0.191	0.542	0.329	0.558	0.181	0.679
Quinoa, uncooked	0.407	0.504	0.840	0.766	0.309	0.203	0.593	0.267	0.421	0.167	0.594

Source: USDA Food Composition Database. https://fdc.nal.usda.gov/. Accessed September 7, 2021.

Table 6 Mineral content of cereals and pseudo-cereals (per 100 g).

Type	Calcium (mg)	Iron (mg)	Magnesium (mg)	Phosphorus (mg)	Potassium (mg)	Sodium (mg)	Zinc (mg)	Copper (mg)	Manganese (mg)	Selenium (μg)
Cereals										
Wheat (durum)	34	3.5	144	508	431	2	4.16	0.55	3.01	89.4
Wheat flour, whole grain	34	3.6	137	357	363	2	2.60	0.41	4.07	61.8
Wheat flour, white, unenriched	15	1.2	22	108	107	2	0.70	0.14	0.68	33.9
Bread, white wheat	684	4.9	26	103	127	478	0.95	0.15	0.63	16.1
Corn grain, white (maize)	7	2.7	127	210	287	35	2.21	0.31	0.49	15.5
Rice, white, medium-grain, raw, unenriched	9	0.8	35	108	86	1	1.16	0.11	1.10	NA
Sorghum, grain	13	3.4	165	289	363	2	1.67	0.28	1.60	12.2
Barley, pearled, raw	29	2.5	79	221	280	9	2.13	0.42	1.32	37.7
Rye, grain	24	2.6	110	332	510	2	2.65	0.37	2.58	13.9
Oats	54	4.7	177	523	429	2	3.97	0.63	4.92	NA
Millet, raw	8	3.0	114	285	195	5	1.68	0.75	1.63	2.7
Pseudo cereals										
Buckwheat	18	2.2	231	347	460	1	2.40	1.10	1.30	8.3
Amaranth	159	7.6	248	557	508	4	2.87	0.53	3.33	18.7
Quinoa, uncooked	47	4.6	197	457	563	5	3.10	0.59	2.03	8.5

Source: USDA Food Composition Database. https://fdc.nal.usda.gov/. Accessed September 7, 2021.

Proteins

A large proportion of the protein from the diet comes from cereal grains particularly in low- and middle-income countries. Protein ranges on an average from 7 to 12% of dry weight of cereals. The protein content present is dependent on the genotype, that is, cereal, species, variety, and the conditions during cultivation, which includes climate, soil and fertilization, especially that of nitrogen. The proteins are mainly concentrated in the germ and aleurone layer, although protein is distributed across the grain with the bran containing about 7% and the endosperm about 13% of the protein (Belitz and Grosch, 2009; Koehler and Weiser, 2013). Not only does it determine the nutritional value, but it also impacts the functional properties in food processing. Prolamins and globulins are the storage proteins present in varying amounts in cereal grains. Globulins are largely present in the embryo and the outer aleurone layer of the endosperm. In oats and rice, however, globulins which constitute 70–80% of the total protein, are the main endosperm storage protein fraction. In other cereals, prolamins constitute the major protein storage fraction in the

Table 7 Vitamin content of cereals and pseudo-cereals (per 100 g).

Type	*Thiamin (mg)*	*Riboflavin (mg)*	*Niacin (mg)*	*Pantothenic (mg)*	*Vitamin B6 (mg)*	*Total folate (μg)*
Cereals						
Wheat (durum)	0.42	0.12	6.74	0.94	0.42	43
Wheat flour, whole grain	0.50	0.17	4.96	0.61	0.27	44
Wheat flour, white, unenriched	0.12	0.04	1.25	0.44	0.04	26
Bread, white wheat	0.51	0.26	4.46	0.46	0.08	127
Corn grain, white (maize)	0.39	0.20	3.63	0.42	0.62	NA
Rice, white, medium-grain, raw, unenriched	0.58	0.05	5.09	1.34	0.15	231
Sorghum, grain	0.33	0.10	3.69	0.37	0.44	20
Barley, pearled, raw	0.19	0.11	4.60	0.28	0.26	23
Rye, grain	0.32	0.25	4.27	1.46	0.29	38
Oats	0.76	0.14	0.96	1.35	0.12	56
Millet, raw	0.42	0.29	4.72	0.85	0.38	85
Pseudo cereals						
Buckwheat	0.10	0.43	7.02	1.23	0.21	30
Amaranth	0.12	0.20	0.92	1.46	0.59	82
Quinoa, uncooked	0.36	0.32	1.52	0.77	0.49	184

Source: USDA food composition database. https://fdc.nal.usda.gov/. Accessed September 7, 2021.

endosperm. Wheat flour contain the globulins gliadin and glutenin, which form the gluten fraction a factor that determines the end use quality (Wrigley, 2017b). Coarse cereals have a higher content of protein and a balanced amino acid profile (Fu et al., 2020).

The amino acid composition of the cultivar of cereal determines the quality of protein. Cereals are not only deficient in the essential amino acid lysine but also in threonine and specifically in maize, tryptophan too (Shewry and Halford, 2002; Belitz and Grosch, 2009). Lysine content is low in most cereals particularly in wheat, rye, barley, oats and corn. In oats protein content is higher than some of the other cereals like wheat but they have a higher arginine and lysine content, which improve the quality of protein (Slavin et al., 2000).

Fats

Cereal grains in general contain a low level of lipids. Most of the lipids are present in the germ. Barley, rice, rye and wheat contain 1–3% lipids, corn 5–9% and oats 5–10% of the dry matter. The essential fatty acid linoleic acid is the main lipid present in grains (McKevith, 2004).

Micronutrients

Micronutrients are mainly concentrated in the pericarp, germ and aleurone layer. About 45% of vitamins and minerals are present in the bran and germ components of the whole grain (Oghbaei and Prakash, 2016).

Cereals are a good source of potassium but contain a lower amount of sodium. Whole grains have a considerable amount of iron, magnesium and zinc and lower levels of trace elements such as selenium (rice has the highest content at 10–13 μg per 100 g), and the content is dependent on soil conditions. A high amount of calcium is found in finger millet and iron in barnyard millet (McKevith, 2004).

In general, B vitamins thiamine, riboflavin and niacin and vitamin E are present in cereal grains but do not contain vitamin B12 and C, and in most grains, vitamin A. Whole grain cereals and bran and the main dietary sources for vitamin E (Jones et al., 2016).

Bioactive compounds

Bioactive substances such as fiber, vitamins, minerals and trace elements (vitamins B, E and folate, zinc, magnesium, selenium etc), phytochemicals (polyphenols, Ɣ-oryzanol, carotenoids, phytic acid, alkyl resorcinols, phytosterols, inulin, B-glucans flavonoids etc) sulfur containing compounds and alpha linolenic acid, methyl donors and lipotrophes (for example betaine, choline, melatonin, etc) constitute 15% of the dry weight of whole grains, with about 52% and 24% concentrated in bran and embryo respectively (Fardet, 2013; Fu et al., 2020). Lignans, tocotrienols, phenolic compounds, and antinutrients including phytic acid, tannins, and enzyme inhibitors present in whole grains are associated with better health outcomes (Oghbaei and Prakash, 2016).

Among phytochemicals, small quantities of flavonoids and lignans (phytoestrogen) are present in cereals. The major contributor to the antioxidant activity of cereals are bound phytochemicals (McKevith, 2004).

Anti-nutrient compounds

Some of the anti-nutrient compounds present in cereals are phytates, tannins and polyphenols. Phytates mainly reside in the germ of maize, the aleurone layer of wheat, and is uniformly distributed in millets (Tiwari and Pojić, 2020). Polyphenols and phytates are mainly found in the pericarp, seed coat and aleurone of the grain. When outer layers of the grain are removed there is a reduction in tannins, phytates and polyphenols which increases the digestibility of proteins and carbohydrates and availability of minerals. But these nutrients also have antioxidant properties (Oghbaei and Prakash, 2016).

Phytates form insoluble complexes with iron and calcium decreasing bioavailability. Tannins, phenolic compounds from the flavonoid group reduces protein digestibility by phenol-protein complexing, reduce palatability as flavor and color are affected. As tannins are mainly present in the outer layers, removal of these layers reduces the quantity of tannins in the cereal, but this results in loss of nutrients. In sorghum, the anti-nutrient effects can be mitigated by germination and treatment with calcium oxide, potassium carbonate, ammonium bicarbonate or sodium bicarbonate (McKevith, 2004).

Goitrogens, substances that decrease bioavailability of iodine or interfere within the thyroid hormone biosynthetic pathway, are present in pearl millet and barley, while trypsin inhibitors that reduce protein digestibility anti-nutritive compounds are present in pearl millet and rye, Heat treatment of the grain reduces these antinutrient contents (McKevith, 2004).

Contaminants

Pests such as mites and weevils can affect cereal crops before and after harvesting and can produces unpleasant smells and tastes, decrease carbohydrate content while increasing the free fatty acid levels and reducing vitamin B levels of thiamine and riboflavin (McKevith, 2004). The mycotoxin, aflatoxin present in raw maize, rice, wheat and sorghum are harmful carcinogenic substances present in wheat and rice and corn (MacEvilly, 2003; Mousavi Khaneghah et al., 2018; Andrade and Caldas, 2015). Analysis of worldwide aflatoxin (s (AFB1, AFB2, AFG1, AFG2)) occurrence data has reported that rice, wheat, maize and sorghum consumption contributed to 41.6%, 35.4%, 21.2% and 1.8% of the total aflatoxin intake respectively (Andrade and Caldas, 2015). Aflatoxins are naturally present in cereals, but the presence of a substantial amount can cause harm (Andrade and Caldas, 2015). Rye and other temperate cereals can be affected by Ergot (*Claviceps purpurea*) and produces alkaloid toxins.

Cereals grains and adverse health outcomes

Gluten allergy

Some individuals are allergic to gluten, a protein present in cereals such as wheat, oats barley and rye. Glutenin and gliadin in wheat, avenin in oats, hordein in barley, and secalin in rye are the gluten compounds present leading to celiac disease (MacEvilly, 2003). It is precisely for this reason that gluten free sorghum, millets and pseudo cereals are considered safe options for individuals with gluten allergy.

Association with non-communicable diseases

With rapid changes in technology in processing of cereal grains, there is a focus now on their association with some of the non-communicable diseases mainly through evidence from epidemiological studies. Higher intakes of fiber from cereal foods or a mix of whole grains and bran are associated with a lower risk of obesity, type 2 diabetes, cardiovascular disease and some cancers (Slavin et al., 2000; Cho et al., 2013; Fardet, 2013).

Coarse cereals contain betaglucans (or lichenins, a water-soluble fiber), polyphenols, phytosterols and fagopyritols and phytochemicals, such as phenolic acids, flavonoids, lignans, tocols and phytosterol which have anti-inflammatory properties and a beneficial effect on non-communicable diseases such as cardiovascular disease, diabetes and some cancers (Fu et al., 2020). Beta glucan in oats and the arginine and lysine content have been shown to have a cholesterol lowering effect, reduce the risk of heart disease and also promote weight control by enhancing the feeling of satiety (Slavin et al., 2000; Wrigley, 2017b; Fu et al., 2020). Carbohydrates in foods can be classified based on the extent to which they can be digested and absorbed and assimilated into the human body resulting in a post-prandial glucose response. The higher the glucose response, the higher will be the glycemic index of the food. Cereals are the principal sources of carbohydrate and therefore, the glycemic response of cereal-based foods are important. The starch present in the food matters. Amylose is slowly digested by the enzyme α-amylases in the human duodenum, while amylopectin is rapidly digested mainly because its branched structure allows enzymatic hydrolysis at multiple sites. The ratio of the incremental area under the curve (AUC) on ingestion of a certain amount of carbohydrate from a test food, to the to the AUC in a reference food with similar amount of carbohydrate usually glucose or white bread is termed the glycemic index. A lower glycemic index in foods including cereals combined with high fiber diets can decrease post-prandial glucose and insulin responses and improve the lipid profile and better metabolic and health outcomes The extent of fiber present further lowers the glycemic response and therefore cereal grains consumed whole are beneficial in preventing cardiovascular disease and diabetes. The presence of resistant starch also matters. More the resistant starch present, lower the digestion of the starch. The size, shape and crystallinity, amylose content, lipid, protein and phosphate content of the starch granule are important for this resistance. Physiologically, there is an increased production of the small chain fatty acids, butyrate, acetate and propionate in the large bowel due to fermentation of

the unabsorbed resistant starch by the gut microbiota, thereby improving g gut health. This in turn increases the level of the gut hormones, glucagon-like peptide (GLP-1) and peptide YY (PYY) that regulate energy intake (Lafiandra et al., 2014).

Acrylamides formed during heat treatment such as baking and frying are quoted as potential carcinogen, but there is no convincing evidence to support this (McKevith, 2004).

Conclusions

Cereal grains are the principal foods that provide a large proportion of energy, carbohydrate, protein and some vitamins and minerals in the diet. Wheat, maize and rice and the three most highly produced and consumed cereal grains. Other coarse cereals are ingested to a lesser extent but are important due to their nutritional properties. Processing enhances palatability and alters the nutrient cont. Both traditional and technologically advanced techniques are used for processing cereal and cereal-based foods. There are potential benefits of consumption of cereal grains which is dependent on both the inherent property of the cereal and the level of processing. Apart from immunological disturbances, there are other adverse health outcomes due to consumption of cereal grains, particularly refined cereals that need to be addressed and tackled.

Conflicts of interest

The author declares no conflict of interest.

References

Andrade, P.D., Caldas, E.D., 2015. Aflatoxins in cereals: worldwide occurrence and dietary risk assessment. World Mycotoxin J. 8 (4), 415–431. https://doi.org/10.3920/WMJ2014.1847.

Awika, J.M., 2011. Major cereal grains production and use around the world. ACS Symp. Ser. 1089, 1–13. https://doi.org/10.1021/bk-2011-1089.ch001.

Bandyopadhyay, T., Muthamilarasan, M., Prasad, M., 2017. Millets for next generation climate-smart agriculture. Front. Plant Sci. 8, 1–6. https://doi.org/10.3389/fpls.2017.01266.

Batey, I., 2017. The diversity of uses for cereal grains. In: Cereal Grains: Assessing and Managing Quality: Second Edition. Elsevier Ltd. https://doi.org/10.1016/B978-0-08-100719-8.00003-6.

Békés, F., Schoenlechner, R., Tömösközi, S., 2017. Ancient wheats and pseudocereals for possible use in cereal-grain dietary intolerances. In: Cereal Grains: Assessing and Managing Quality: Second Edition. Elsevier Ltd. https://doi.org/10.1016/B978-0-08-100719-8.00014-0.

Belitz, H., Grosch, W., 2009. Cereals and cereal products. In: Food Chemistry. Springer. https://doi.org/10.1016/B978-0-12-374136-3.00011-0.

Borneo, R., León, A.E., 2012. Whole grain cereals: functional components and health benefits. Food Funct. 3 (2), 110–119. https://doi.org/10.1039/c1fo10165j.

Butt, M.S., et al., 2008. Oat: unique among the cereals. Eur. J. Nutr. 47 (2), 68–79. https://doi.org/10.1007/s00394-008-0698-7.

Cho, S.S., et al., 2013. Consumption of cereal fiber, mixtures of whole grains and bran, and whole grains and risk reduction in type 2 diabetes, obesity, and cardiovascular disease. Am. J. Clin. Nutr. 98 (2), 594–619. https://doi.org/10.3945/ajcn.113.067629.

Curry, A., 2021. The ancient carb revolution. Nature 594, 488–491. https://doi.org/10.1093/nq/s2-ix.233.471d.

Dietrich, L., et al., 2019. Cereal processing at early neolithic göbekli tepe, southeastern Turkey. PLoS One. https://doi.org/10.1371/journal.pone.0215214.

Fardet, A., 2013. Whole Grains From a Mechanistic View. CFW Plexus. https://doi.org/10.1094/cplex-2013-1001-01b.

Fitzgerald, M., 2017a. Rice: grain-quality characteristics and management of quality requirements. In: Cereal Grains: Assessing and Managing Quality: Second Edition. Elsevier Ltd. https://doi.org/10.1016/B978-0-08-100719-8.00009-7.

Fitzgerald, M., 2017b. Rice: Grain-Quality Characteristics and Management of Quality Requirements in Cereal Grains: Assessing and Managing Quality, second ed. https://doi.org/10.1186/s40066-020-00282-6

Food and Agricultural Organization, 2016. Save and Grow in Practice: Maize, Rice, Wheat, Food and Agricultural Organization. FAO, Rome.

Food and Agricultural Organization, 2020. World Food and Agriculture—Statistical Yearbook 2020, World Food and Agriculture—Statistical Yearbook 2020. https://doi.org/10.4060/cb1329en. Rome.

Food and Agricultural Organization, 2021. World Food Situation: FAO Cereal Supply and Demand Brief. Available at. http://www.fao.org/worldfoodsituation/csdb/en/. (Accessed 2 September 2021).

Fu, J., et al., 2020. Concise review: coarse cereals exert multiple beneficial effects on human health. Food Chem. 325. https://doi.org/10.1016/j.foodchem.2020.126761.

Grace, P.K., et al., 2019. Key agri-climatic and socio-economic indicators for cereal production across the world. Int. J. Environ. Agric. Biotechnol. 4 (2), 279–300. https://doi.org/10.22161/ijeab/4.2.5.

Győri, Z., 2017. Corn: grain-quality characteristics and management of quality requirements. In: Cereal Grains: Assessing and Managing Quality, second ed. https://doi.org/10.1016/B978-0-08-100719-8.00009-7

Hübner, F., Arendt, E.K., 2013. Germination of cereal grains as a way to improve the nutritional value: a review. Crit. Rev. Food Sci. Nutr. 53 (8), 853–861. https://doi.org/10.1080/10408398.2011.562060.

Jones, J.M., et al., 2015. Nutritional impacts of different whole grain milling techniques: a review of milling practices and existing data. Cereal Foods World 60 (3), 130–139. https://doi.org/10.1094/CFW-60-3-0130.

Jones, J.M., et al., 2016. CIMMYT series on carbohydrates, wheat, grains, and health: carbohydrates, grains, and whole grains and disease prevention Part II. Blood pressure, metabolic syndrome, and diabetes. Cereal Foods World 61 (3), 106–122. https://doi.org/10.1094/CFW-61-3-00106.

Koehler, P., Weiser, H., 2013. Handbook on sourdough biotechnology. In: Gobbetti, M., Gänzle, M. (Eds.), Chemistry of Cereal Grains in Handbook on Sourdough Biotechnology. Springer Science+Business Media, New York. https://doi.org/10.1007/978-1-4614-5425-0.

Lafiandra, D., Riccardi, G., Shewry, P.R., 2014. Improving cereal grain carbohydrates for diet and health. J. Cereal. Sci. 59 (3), 312–326. https://doi.org/10.1016/j.jcs.2014.01.001.

Leff, B., Ramankutty, N., Foley, J.A., 2004. Geographic distribution of major crops across the world. Global Biogeochem. Cycles 18 (1). https://doi.org/10.1029/2003gb002108.

MacEvilly, C., 2003. CEREALS | contribution to the diet. In: Encyclopedia of Food Sciences and Nutrition, pp. 1008–1014. https://doi.org/10.1016/b0-12-227055-x/00186-3, 5.

McKevith, B., 2004. Nutritional aspects of cereals. Nutr. Bull. 29 (2), 111–142. https://doi.org/10.1111/j.1467-3010.2004.00418.x.
Mousavi Khaneghah, A., et al., 2018. Aflatoxins in cereals: state of the art. J. Food Saf. 38 (6), 1–7. https://doi.org/10.1111/jfs.12532.
OECD/FAO, 2021. OECD-FAO Agricultural Outlook 2021–2030. OECD Publishing, Paris.
Oghbaei, M., Prakash, J., 2016. Effect of primary processing of cereals and legumes on its nutritional quality: a comprehensive review. Cogent Food Agric. 2 (1). https://doi.org/10.1080/23311932.2015.1136015.
Ranum, P., Peña-Rosas, J.P., Garcia-Casal, M.N., 2014. Global maize production, utilization, and consumption. Ann. N. Y. Acad. Sci. 1312 (1), 105–112. https://doi.org/10.1111/nyas.12396.
Shewry, P.R., Halford, N.G., 2002. Cereal seed storage proteins: structures, properties and role in grain utilization. J. Exp. Bot. 53 (370), 947–958. https://doi.org/10.1093/jexbot/53.370.947.
Slavin, J.L., Jacobs, D., Marquart, L., 2000. Grain processing and nutrition. Crit. Rev. Food Sci. Nutr. 40 (4), 309–326. https://doi.org/10.1080/10408690091189176.
Slavin, J., 2004. Whole grains and human health. Nutr. Res. Rev. 17 (1), 99–110. https://doi.org/10.1079/nrr200374.
Swaminathan, S., et al., 2021. Associations of cereal grains intake with cardiovascular disease and mortality across 21 countries in prospective urban and rural epidemiology study: prospective cohort study. BMJ 372. https://doi.org/10.1136/bmj.m4948.
Taylor, J.R.N., Duodu, K.G., 2017. Sorghum and millets: grain-quality characteristics and management of quality requirements. In: Cereal Grains: Assessing and Managing Quality: Second Edition. Elsevier Ltd. https://doi.org/10.1016/B978-0-08-100719-8.00013-9.
Thielecke, F., Lecerf, J.M., Nugent, A.P., 2020. Processing in the food chain: do cereals have to be processed to add value to the human diet? Nutr. Res. Rev. https://doi.org/10.1017/S0954422420000207.
Tiwari, U., Pojić, M., 2020. Introduction to cereal processing. In: Innovative Processing Technologies for Healthy Grains, pp. 9–35. https://doi.org/10.1002/9781119470182.ch2.
Uthayakumaran, S., Wrigley, C., 2017. Durum wheat: grain-quality characteristics and management of quality requirements. In: Cereal Grains: Assessing and Managing Quality: Second Edition. Elsevier Ltd. https://doi.org/10.1016/B978-0-08-100719-8.00006-1.
Wrigley, C., Bushuk, W., 2017. Rye: grain-quality characteristics and management of quality requirements. In: Cereal Grains: Assessing and Managing Quality, second ed. Elsevier Ltd. https://doi.org/10.1016/B978-0-08-100719-8.00009-7.
Wrigley, C., 2017a. Cereal-grain morphology and composition. In: Cereal Grains: Assessing and Managing Quality: Second Edition. Elsevier Ltd. https://doi.org/10.1016/B978-0-08-100719-8.00004-8.
Wrigley, C., 2017b. The cereal grains: providing our food, feed and fuel needs. In: Cereal Grains: Assessing and Managing Quality: Second Edition. Elsevier Ltd. https://doi.org/10.1016/B978-0-08-100719-8.00002-4.

Dietary fiber: Physiological effects and health outcomes

Ian T. Johnson, Quadram Institute, Norwich, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of D.L. Topping, Dietary Fiber: Physiological Effects and Health Outcomes, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 50–54, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00106-9.

Introduction 306
Definitions 307
Analytical methods 307
Physiological effects 308
Dietary fiber in the upper alimentary tract 308
Mouth and pharynx 309
Stomach 309
Small intestine 309
Physiological effects in the lower alimentary tract 309
Colon and rectum 310
Health outcomes attributable to dietary fiber 310
Overweight and obesity 310
Etiology and management of type 2 diabetes 311
Cholesterol metabolism 311
Blood pressure 311
Cardiovascular disease 312
Diseases of the alimentary tract 312
Constipation and diverticular disease 312
Irritable bowel syndrome (IBS) 312
Colorectal cancer (CRC) 313
Conclusion 313
References 314
Further reading 314

Key points

- Dietary fiber comprises a range of food-borne polysaccharides that share the single physiological property of resistance to human digestive enzymes.
- These polysaccharides are confined to the intestinal lumen during digestion of food in the upper alimentary tract. Their presence influences digestion in ways that depend on their individual physical and chemical properties.
- After passage through the small intestine, undigested polysaccharides enter the colon, where they differ in their accessibility for fermentation by the colorectal microbiota.
- As a result of these various properties and effects, fiber influences metabolic processes throughout the body.
- Epidemiological evidence demonstrates that the inclusion of adequate levels of fiber in the diet is a protective factor in relation to metabolic, cardiovascular and gastrointestinal diseases.
- Future research should focus particularly on the metabolic and physiological effects of novel polysaccharides that may become increasingly important due to global trends in agricultural production, food supply and culinary practice.

Introduction

In addition to digestible sugars, starch and protein, plant foods contain poorly digestible carbohydrate components that do not contribute to nutrition in the classical sense of providing macronutrients or other essential substances, but which are nevertheless important for human health. The existence of undigestible residues in plant foods, first described as *crude fiber*, was recognized early in the 20th century by some of the pioneers of scientific nutrition, including Atwater in the USA and Rubner in Germany. Later the British scientists McCance and Lawrence categorized the *available carbohydrates* present in plant-foods, which consist of starch and sugars, and the *unavailable carbohydrates*, which in cereal seeds consist mainly of cell wall polysaccharides that cannot be hydrolyzed by human digestive enzymes. During the middle years of the 20th century, TL Cleave and other medical scientists went on to develop the view that various chronic diseases of western society were caused by excessive consumption of highly refined

carbohydrates from which non-digestible cell wall material had been removed. The colonial medical officer Denis Burkitt became convinced from his own field observations that diets rich in unrefined carbohydrate foods protect rural African populations against gastrointestinal disorders, including diverticular disease, appendicitis and colorectal cancer. Contemporaneously, his colleague, the physician Hugh Trowell, argued that consumption of unrefined carbohydrates was protective against obesity, coronary heart disease and type 2 diabetes. Trowell proposed the umbrella term *dietary fiber* to describe "that portion of food which is derived from cellular walls of plants which are digested very poorly by human beings" (Trowell, 1972). The "dietary fiber hypothesis", as it soon came to be known, has been developed and refined over subsequent decades, and it is now firmly established as a principle of human nutrition. This article explores the physiological mechanisms by which food components that, at least in the early stages of digestion, are unavailable for absorption and metabolism by the body, but can nevertheless be deemed essential for normal function and the prevention of disease.

Definitions

Trowell's early concept of dietary fiber was logical but chemically naive. Once the hypothesis began to be explored, it became increasingly obvious that a more precise definition of fiber was needed, together with a standardized method of analysis. However the diversity of human diets and the complexity of their carbohydrate constituents delayed these goals for many years. The available carbohydrates in plant foods are mainly those that function in the plant as energy stores. They consist of alpha-linked linear or branching glucose polysaccharides known collectively as *starches*, and the *sugars*, notably the monosaccharides glucose and fructose, and the disaccharide sucrose. The presence of alpha-linkages ensures that both sucrose and starch are readily hydrolyzed to monosaccharides by human digestive enzymes, and rapidly absorbed in the small intestine. In contrast, the structural carbohydrates of plant tissues consist of non-starch polysaccharides (NSP) containing a variety of saccharide units, including glucose, galactose, arabinose, xylose and uronic acids, joined by beta linkages, which are not hydrolyzed by human digestive enzymes. Cellulose, which is one of the most abundant components of this group, consists of polymers of 1–4 beta-linked glucose molecules, often with molecular weights of a million Daltons or more. Individual cellulose polymers are combined to form fibrils which, when embedded within a matrix of branching xylose, galactose and arabinose polymers, provide structural rigidity to the cell walls of undifferentiated parenchyma cells in fruits and vegetables. The walls of more specialized structural cells, and the vascular components in stem and leaf tissue, typically require the greater rigidity and mechanical strength provided by even higher levels of cellulose and other matrix polysaccharides; these structures are often toughened further by the deposition of lignin, a branching polymer of phenyl propane units. Pectins, and another class of beta-linked glucose polysaccharides, the beta-glucans present in the outer layers (bran) of cereal seeds such as oats and barley, have the important property of solubility in water. Some important sources and constituents of dietary fiber in human foods are summarized in **Table 1**.

Analytical methods

Quantification of fiber in foods is essential for the compilation of food composition tables, for dietetic practice and for nutrition research, but the great diversity of substances classified as fiber leads to technical difficulties. The earliest methods for measuring the undigestible residues in animal feeds emerged in the 19th century. The approach was to remove the digestible components by chemical degradation, and then measure the residues by weighing. This *gravimetric* approach formed the basis of the crude-fiber procedure for food analysis, which was widely used until the 1970s. However, because the old analytical procedures typically employed strong acids and alkalis for the initial digestion step, many of the cell wall carbohydrates falling within the new concept of dietary fiber were broken down and lost. To address this problem, Southgate developed methods based on enzymatic hydrolysis of polysaccharides followed by colorimetric analysis of the constituent sugars. Using a distinctly different approach, Prosky and

Table 1 Some important sources and constituents of dietary fiber in the human diet.

Food source	*Non-starch polysaccharides*	*Associated substances*
Fruits and vegetables	Cellulose, xyloglucans, pectic polysaccharides	Lignin, suberin, glycoproteins
Cereal grains	Cellulose, branched arabinoxylans, xyloglucans, beta-glucans, glucomannans	Lignin, and other phenolic esters
Leguminous seeds	Cellulose, xyloglucans, mannans, galactomannans, pectic substances	Oligosaccharides (verbascose, stachyose, inulin)
Manufactured foods	Galactomannan gums (guar gum, locust bean gum), beta-glucans (oat gum), pectins, alginates, carrageenans, modified cellulose gums (carboxymethyl cellulose, methyl cellulose)	

Table 2 Total dietary fiber, non-starch polysaccharides (NSP) and resistant starch (RS) in some typical human foods, measured using two official AOAC methods and the Englyst method for resistant starch and NSP.

	AOAC 991.43	*AOC 2009.01*	*Englyst method*		
Food source	*Total dietary fiber*	*Total dietary fiber*	*Total RS*	*NSP*	*Total dietary fiber*
Whole wheat cereal	11.1	12.1	1.0	7.5	10.8
Puffed rice cereal	2.3	3.2	0.5	1.1	2.2
Bread	5.9	7.2	2.1	2.9	6.1
Biscuit	3.2	3.0	0.2	2.1	4.0
Canned potato	10.3	13.6	5.4	6.1	11.1
Canned chick peas	22.7	22.5	6.9	11.4	19.8

Illustrative data selected from Englyst et al. (2013).

colleagues developed methods for the analysis of *total dietary fiber* (TDF) based on stepwise enzymatic hydrolysis and accurate weighing of the residue. This gravimetric approach was adopted for official methods of fiber analysis by the Association of Official Analytical Chemists (AOAC International). Though more physiologically relevant than the crude fiber methods, these new approaches allowed varying quantities of *resistant starch* (RS) to be in the final value for TDF. It is now recognized that RS is not merely an analytical artifact, but an important component of food, which does, to varying degrees, escape digestion in the human small intestine. To address this issue, Englyst and others developed analytical refinements to enable the physical separation of starch residues from *non-starch polysaccharides* (NSP). Though analytically more precise, these methods are also more complex and time-consuming than gravimetric methods. Some typical values for TDF and RS in processed foods are given in **Table 2**. In general, the AOAC and Englyst methods do give similar results for TDF, but in some foods a substantial proportion of this value is accounted for by RS.

For regulatory purposes most countries now use definitions of dietary fiber consistent with those agreed upon in 2009 by the food regulation arm of the World Health Organization, *Codex Alimentarius* (Codex Alimentarius, 2010). The definition states that: *Dietary fiber means carbohydrate polymers with ten or more monomeric units, which are not hydrolyzed by endogenous enzymes in the small intestine of human beings and belong to the following categories:*

- *Edible carbohydrate polymers naturally occurring in food as consumed.*
- *Carbohydrate polymers, which have been obtained from raw material in food by physical, enzymatic, or chemical means and which have been shown to have physiological effect of benefit to health by generally accepted scientific evidence to competent authorities,*
- *Synthetic carbohydrate polymers, which have been shown to have physiological effect of benefit to health by generally accepted scientific evidence to competent authorities.*

Thus, both RS and NSP are included in the modern definition of dietary fiber. It is left to national authorities to decide whether to include purified or synthetic polysaccharides with established health benefits, and oligosaccharides with degrees of polymerization between 3 and 9. The Codex Alimentarius Commission recommends several AOAC methods for analysis of dietary fiber. The more elaborate NSP method is also recognized as being appropriate for research purposes, where separation of NSP and RS and a full analysis of the monosaccharide composition of fiber may be required.

Physiological effects

The modern definition of dietary fiber is convenient from a regulatory perspective but, given the diverse chemical and physical properties of the substances captured by the gravimetric approach to analysis, a single analytical value for TDF cannot predict the physiological properties or the health benefits of any particular food or diet. In this section the physiological effects of dietary fiber will be discussed in relation to the properties and behavior of its constituents at sequential locations along the digestive tract. However, for many components of dietary fiber, the metabolic consequences of their presence, and hence their effects on health, extend well beyond the gut.

Dietary fiber in the upper alimentary tract

The human alimentary tract is essentially a tube extending from mouth to anus, although it is subdivided into anatomically distinct regions that function as separate organs. After ingestion food is conveyed progressively through the gut and physically disrupted by rhythmic contractions of the muscular walls (peristalsis). Hydrolysis of polymers occurs under the influence of digestive enzymes released into the lumen at the appropriate anatomical sites. The presence of dietary fiber influences digestion at each stage.

Mouth and pharynx

The earliest steps in digestion begin in the mouth, where food particles are reduced in size by mastication, lubricated with saliva, and prepared for swallowing. Salivary amylase initiates the hydrolysis of starch at this stage. Intact cell walls confer three-dimensional structure on plant foods, they are an important determinant of texture, and they can influence the degree of mechanical breakdown that tissues undergo prior to swallowing. The rate at which starch is digested and absorbed from cubes of cooked potato has been shown to be slower when they are swallowed whole than when they are chewed. Thus, simple mechanical factors related to food texture can limit the rate at which glucose enters the circulation.

Stomach

After passage from the mouth and through the esophagus, food accumulates in the stomach, where large fragments are further degraded by the activity of gastric smooth muscle and the effects of hydrochloric acid and proteolytic enzymes. The physical structure of fiber-rich complex carbohydrates modifies gastric digestion. For example, the persistence of intact cereal fragments in the stomach can delay starch digestion significantly, and soluble non-starch polysaccharides with high viscosity, such as pectin and guar gum, can slow the delivery of glucose solutions to the small intestine by delaying gastric emptying.

Small intestine

The small intestine, and particularly the jejunum which accounts for about forty percent of its total length, is the main site of nutrient absorption. The semi-liquid products of gastric digestion are released into the duodenum and then propelled through the jejunum by peristaltic movements. Further hydrolysis of proteins, triglycerides and starch, catalyzed by pancreatic enzymes, proceeds rapidly within the jejunal lumen, but the final stages of digestion occur at the jejunal mucosa, under the influence of cell-surface enzymes. The products of hydrolysis are then absorbed into the circulation, along with water and electrolytes, through the epithelial cells of the mucosal villi. Waves of muscular activity ensure that the mixture of ingested fluids, gastric secretions and partially digested food is well stirred.

The presence of dietary fiber modifies the rate of nutrient digestion and absorption by various mechanisms. For example, the effects of intact cell walls have been investigated in cooked legume seeds and shown to delay the hydrolysis of starch by restricting the process of gelatinization, inhibiting the access of pancreatic amylase to starch granules, and perhaps by the physical binding of amylase. Soluble, viscous polysaccharides added as isolates to carbohydrate test-meals can also inhibit glucose absorption, but this effect is abolished if the viscosity is eliminated. In contrast to the barrier-effects of intact cell wall polysaccharides, which inhibit digestion, the importance of viscosity lies in the inhibition of transport, through suppression of convective stirring in the boundary layer immediately adjacent to the mucosal surface. The absorptive activity of the epithelial cells reduces the concentration of substrates in the fluid layer adjacent to the mucosal cells. To maintain rapid transport the absorbed substrates must be replaced by diffusion or, more efficiently, by physical stirring due to peristalsis; viscous forms of dietary fiber inhibit this latter process, thus slowing the delivery of glucose across the mucosa and into the circulation.

Prolonged consumption of viscous types of dietary fiber is associated with a reduction in plasma cholesterol levels in humans. Unlike the acute suppression of glucose absorption, the change in cholesterol metabolism is a chronic effect, but it probably also results from a suppression of mucosal transport, this time in the lower small intestine (ileum) where, as a crucial step in their enterohepatic circulation, bile salts are reabsorbed. If this step is inhibited, bile salts can be lost to the feces; it is probable that a compensatory increase in bile-salt synthesis from endogenous cholesterol pools leads to the observed reduction in plasma cholesterol.

Dietary fiber also influences the rate of transit through the upper alimentary tract, but the importance of its physical properties in this context is complex and still poorly understood. Thus, although insoluble fiber (coarse wheat bran) added to a carbohydrate test-meal delays gastric emptying, the same material apparently accelerates the rate of transit through the small intestine. Some of the effects on motility probably reflect the ability of dietary fiber to modify the release of gastrointestinal peptide hormones. For example, addition of viscous fiber to test-meals inhibits insulin secretion, and the release of gut peptides, including glucagon-like peptide 1 (GLP-1) and gastric inhibitory peptide (GIP), which both regulate the release of insulin and inhibit the motility of the upper gastrointestinal tract. The distal small intestinal mucosa also contains endocrine cells which, in response to the presence of nutrients in the gut lumen, secrete gastrointestinal regulatory peptides, including enteroglucagon, GLP-1 and PYY, into the circulation. Among other effects, these peptides inhibit gastric emptying, via the so-called *ileal brake*, a mechanism which may be important in appetite control. Thus, by slowing digestion and absorption in the upper small intestine, dietary fiber can displace nutrients to more distal regions of the small intestine and thereby activate this regulatory system. However, dietary fiber also exerts important effects on gut motility via endocrine mechanisms initiated in the large bowel.

Physiological effects in the lower alimentary tract

Having escaped hydrolysis in the small intestine, the various polysaccharides that comprise dietary fiber enter the large bowel, where they encounter the complex array of carbohydrase enzymes expressed by the gut microbiota.

Colon and rectum

Humans do not possess the enlarged cecum typical of herbivores, but the proximal colon is well adapted for bacterial colonization. It typically contains around 200 g of semi-liquid dietary residues, mucus and intestinal secretions, which provide ideal conditions for anaerobic fermentation, to the extent that the human large intestine is one of the richest known bacterial eco-systems. The dominant groups are gram-negative anaerobes of the genus *Bacteroides*, and gram-positive organisms including *bifidobacteria*, *eubacteria*, *lactobacilli*, and *clostridia*. The technical difficulty of culturing many of the anaerobic bacteria long delayed our understanding of the gut microbiome but, with the use of DNA sequencing technology and advanced bioinformatics, much of this complexity has been unlocked in recent years.

In contrast to the very limited enzymic repertoire of human saliva and pancreatic secretions, gut bacteria express a diverse array of enzymes capable of degrading many of the components of the plant cell wall. The microflora profit by gaining energy and substrates that are converted to bacterial mass, whereas the human host gains from the release of the short chain fatty acids (SCFA) acetate, propionate and butyrate, some of which are reabsorbed and utilized as a source of metabolic energy. Colorectal epithelial cells are adapted to utilize butyrate as their primary source of energy. Propionate is metabolized in the liver, but acetate enters the circulation and is used as an energy source by various peripheral tissues.

Despite the huge array of carbohydrase enzymes deployed by the gut microbiota, some polysaccharides, including both natural and chemically modified celluloses, and heavily lignified non-cellulose polysaccharides, remain resistant to gut bacterial fermentation, and serve only to contribute bulk and increased water-retention to the fecal mass. Wheat-bran is a familiar source of dietary fiber. It has long been recognized and recommended as *roughage* by dieticians for its mild laxative properties, and it is widely consumed in wholemeal breads and breakfast cereals. The resistance to bacterial fermentation of the lignified bran enables the material to retain much of its particulate structure in the fecal mass. The resulting matrix favors the retention of water, which adds to fecal bulk in the rectum, and softens the stool. All types of dietary fiber increase fecal bulk to some extent, but raw wheat bran is about five times more effective than more readily fermented polysaccharides such as pectin, beta-glucan and most types of RS.

Because of the varying fermentability of different components of dietary fiber, the precise contributions made by cell wall polysaccharides and RS to total human energy consumptions are difficult to quantify. For practical purposes a figure of 8.4 kJ/g (2.0 kcal/g) is considered a realistic approximation for the energy value of dietary fiber derived from mixed diets. This figure, which is half the energy value of glucose, has been endorsed by the Food and Agriculture Organization of the United Nations, but, as with all food energy factors, this is only an approximation. Indeed, the capacity to derive energy from fermentation of dietary fiber differs between individuals, presumably because of differences in the composition of the colonic microflora, and in the transit time, which determines the time available for fermentation in the large bowel.

Apart from their role as energy substrates, SCFA exert important metabolic effects on the host via their role as signaling molecules. The free fatty acid receptors FFAR2 and FFAR3 are G protein-coupled receptors expressed in gut epithelial cells, and in a variety of other tissues and organs, including those of the immune system. Thus, by stimulating the production and absorption of SCFA, fermentable components of dietary fiber may modulate important physiological pathways in distant tissues, including some related to immunity and inflammation. SCFA have also been shown to stimulate the release of the gut regulatory peptides GLP-1 and PYY from the L-cells of the colonic mucosa. It is hypothesized that through this mechanism the presence of fiber in the diet can modify energy metabolism via effects on upper bowel motility, and hence on satiety.

SCFA may also directly influence gene expression, both in the colorectal mucosa and in other tissues. Butyrate is a potent inhibitor of histone deacetylase (HDAC), a cellular enzyme that reduces acetylation of the histones of chromatin, thus modifying the interaction of gene transcription factors with their target sequences, and thereby altering gene-expression. This *epigenetic* mechanism may exert profound effects on the growth and functioning of cells, both in the gut mucosa, and in other organs including the heart. The consequences for human physiology and health are important subjects of ongoing research.

Because gut bacteria vary in their capacities to degrade different polysaccharides, certain components of dietary fiber can selectively favor the growth of bacteria specialized to utilize them. This so-called *prebiotic* effect is an important determinant of bacterial diversity in the human gut. The recognition of these effects has led some scientists recently to argue that the term *microbiota-accessible carbohydrates* (MAC) should be used to highlight the importance of fermentable components of fiber in relation to both the diversity and the metabolic activity in the gut microbiota. It is increasingly recognized that although all MAC are components of dietary fiber, not all dietary fiber provides MAC.

Health outcomes attributable to dietary fiber

Overweight and obesity

Several of the physiological effects of dietary fiber during digestion have the potential to increase satiety and cause a reduction in energy intake in humans. The average consumption of dietary fiber in the adult populations of high-income countries, including those of North America and Western Europe, is usually significantly lower than the 25 g–50 g per day widely recommended by bodies responsible for public health. Such populations tend also to have a high prevalence of overweight and obesity, but it is not clear that these associations provide evidence for a causal link. Epidemiological studies have shown that average consumption of dietary fiber is often inversely related to body-mass index, but given the complexity of self-selected human diets, and the difficulty of correcting for all the variables involved, it is very difficult to establish a causal mechanism at the population level.

Experimental evidence for a suppression of weight gain by dietary supplementation with, for example, pectin, has been obtained with rodent models and there is some support from human intervention trials. For example, Reynolds et al. (2019) conducted a meta-analysis of 27 controlled intervention trials in which the effect of a higher intake of dietary fiber on weight-gain was investigated and concluded that there was evidence for an approximately 37% reduction in weight-gain in the subjects consuming higher levels of fiber. The evidence for a similar effect of RS is, reportedly, less strong.

Etiology and management of type 2 diabetes

Trowell's early work on dietary fiber led him to propose that both type 2 diabetes and insulin-dependent diabetes were largely diseases of fiber-depletion, caused in high-income countries by mechanized flour milling. His argument was that fiber-depleted starchy foods were diabetogenic and, conversely, that lightly processed high-fiber foods were protective. In the case of type 2 diabetes, subsequent epidemiological evidence provides circumstantial support for this hypothesis. For example, The UK Scientific Advisory Committee on Nutrition conducted a meta-analysis of eleven prospective epidemiological studies and observed a statistically significant protective effect of higher fiber consumption against type 2 diabetes, estimated to provide a 6% reduction in risk for every additional 7 g of fiber consumed per day. Another more recent systematic review of observational studies also demonstrated a reduction in risk of type 2 diabetes among higher consumers of fiber compared to lower consumers, which was estimated to result in a relative risk of 0.85 for every additional 8 g of fiber consumed.

Though the long-term consumption of insoluble cereal fiber is associated with a reduced risk of type-2 diabetes the reasons remain unclear. One early proposal was that because the presence of some types of fiber in carbohydrate foods slows the delivery of glucose to the bloodstream, the reduced demand for insulin would protect against impaired glucose metabolism. The rate at which digestion of a carbohydrate food increases blood glucose is expressed as a numerical quantity, the "glycemic index" (GI), which is determined empirically in human subjects. Despite its proven importance for diabetes management, the role of GI in the causation of type 2 diabetes remains controversial. There is some epidemiological evidence that high-GI diets are diabetogenic, but the dietary fiber content of foods is not a strong predictor of GI. Thus any protective effect of fiber may not be driven primarily by the regulation of intestinal glucose uptake. An alternative explanation is that, through its effects on the gut microbiota, fermentable fiber reduces systemic low-grade inflammation, which is an established risk-factor for insulin-resistance and other metabolic abnormalities. This hypothesis remains under investigation.

Whatever the precise role of fiber in the causation of type 2 diabetes, it is well established that increased fiber consumption is beneficial as a means of managing the condition. Many studies on postprandial glycemia have been conducted using soluble fiber supplements, including guar gum, beta-glucan and pectin, added to carbohydrate test-meals. Although wheat bran and other insoluble cell wall materials have little effect on glucose absorption viscous soluble polysaccharides do (Jenkins et al., 1978). This hypoglycaemic effect is among the physiological properties deemed appropriate to enable isolated or synthetic polysaccharides to be classified as dietary fiber for commercial purposes. Beyond these acute effects on glucose uptake, prolonged supplementation with viscous polysaccharides, including guar gum and beta-glucan, has been shown to improve long-term control of blood glucose in patients with type-2 diabetes.

Cholesterol metabolism

As mentioned previously, consumption of guar gum and beta-glucan has been shown to reduce plasma cholesterol in humans. Dietary guidelines for patients with impaired glucose and cholesterol metabolism often recommend a high intake of carbohydrate foods rich in soluble fiber, although only a few common foods, such as oat products, contain significant quantities of soluble viscous polysaccharides. Most studies in which significant effects on plasma cholesterol are achieved use isolated polysaccharide supplements. One meta-analysis of randomized, controlled intervention trials of this type showed that oat beta-glucan reduced plasma cholesterol by 0.13 mmol/L total cholesterol for every 3 g soluble fiber consumed per day. Reductions of this magnitude are potentially beneficial to those with total plasma cholesterol levels in excess of the desirable upper limit of 5 mmol/L. Health claims relating to cholesterol metabolism are officially recognized for oat bran, and for isolated beta-glucan, by regulatory authorities in both Europe and the USA.

Blood pressure

Observational studies on human populations indicate that higher fiber consumption is associated with lower systolic and diastolic blood pressure. However experimental research indicates that this does necessarily apply equally to all forms of fiber. For example, Evans et al. (2015) carried out a meta-analysis of 18 human intervention trials and observed a reduction in blood pressure in response to higher consumption of all fiber types combined (0.9 mmHg), but this was not statistically significant. Further analysis of the effects of specific fiber types indicated that supplementation with soluble fiber (beta-glucans) reduced systolic blood pressure by 2.9 mmHg (95% Confidence Interval 0.9–4.9 mmHg) and diastolic blood pressure by 1.5 mmHg (95% Confidence interval 0.2–2.7 mmHg) for every extra 4 g of beta glucans consumed. Although these effects seem modest, they are potentially significant

for public health. At the population level, a reduction of 2.0 mmHg in the average systolic blood pressure has been estimated to lead to reductions in annual mortality from coronary heart disease and stroke of 4% and 6% respectively.

Cardiovascular disease

In addition to protective effects against risk factors for cardiovascular disease (CVD), including overweight, impaired glucose tolerance and hypertension, a high consumption of dietary fiber may be directly protective against stroke and coronary heart disease (CHD). Pereira et al. (2004) used pooled data from 10 prospective studies to conduct a systematic analysis of CHD and fiber intake in 91,058 men and 245,186 women. Food-frequency questionnaires were used to estimate intakes of TDF, and fiber from cereals, fruits and vegetables separately. Each 10 g increment in total dietary fiber intake was associated with a 14% reduction in risk of coronary events, and a 27% reduction in risk of death from CHD. A more recent metanalysis of prospective cohort studies on dietary fiber intake and mortality from cardiovascular disease observed a dose-dependent protective effect such that the pooled relative risk for a 10 g increment in fiber intake was 0.91 (95% confidence interval 0.88–0.94). A similar relationship was observed for CHD in particular. No distinction was observed for any particular type or source of fiber, and no conclusions could be drawn as to mechanism. Total dietary fiber has recently been shown to be protective against ischemic stroke, in an analysis of a large cohort spanning nine European countries.

One possibility for the protective effect of dietary fiber against cardiovascular disease is that MAC modifies the metabolic activity of the gut microbiota and suppresses systemic low-grade inflammation and its adverse effects via increased production of SCFA. In a recent study, Ma et al. (2021) compared habitual dietary fiber intake, levels of the inflammation marker C-reactive protein (CRP) in plasma, and the composition of the gut microbiome in 307 healthy male volunteers. They observed a statistically significant relationship such that a higher intake of fiber was associated with a lower level of CRP and with consistent changes in the colorectal microbiome. Studies of this type do provide valuable evidence for the hypothesis that MAC influences systemic physiology by modulating the metabolism of the gut microbiome, but further research is needed to establish a full causal relationship between fiber intake and disease pathology. Most intervention studies using dietary fiber for primary prevention of CVD have been short-term investigations on risk-factors such as LDL cholesterol, rather than studies with disease as the endpoint of interest. There is a clear need for long-term, randomized intervention trials to explore the full potential of dietary fiber for the primary prevention of cardiovascular disease events.

Diseases of the alimentary tract

Constipation and diverticular disease

One of the earliest hypotheses for the health benefits of dietary fiber was the idea advanced by Denis Burkitt that undigested plant material provided high fecal bulk which, he proposed, was associated with an optimal rate of movement of stool through the large intestine during peristaltic contractions, lower intraluminal pressure and minimal straining at stool (Burkitt, 1975). In contrast, he argued, the low-fiber diets of western populations, led to both chronic constipation and age-related degenerative changes in the muscular wall of the colon. Burkitt based these conjectures largely on his own observations of bowel function and disease prevalence in the rural communities of Africa in which he had worked.

Chronic constipation is a relatively common complaint in western populations. It is defined as a condition in which patients regularly experience fewer than three bowel movements per week, with hard stools and a sense of incomplete evacuation or excessive straining. Epidemiological evidence from western populations suggests that higher intakes of dietary fiber are associated with a reduced incidence of constipation in the general population. A systematic review by Suares and Ford (2011) concluded that soluble fiber was probably effective as a treatment for chronic idiopathic constipation but found the evidence inconclusive for insoluble fiber. In routine medical practice, patients with chronic constipation may be advised to increase their consumption of foods rich in fiber, or they may be prescribed pharmaceutical preparations such as refined isphagula husk, derived from the seeds of *Plantago psyllium*.

Diverticulae are small inward protrusions of the mucosa, extending deep into the muscular wall of the colon, common in older age-groups in western populations. They are often symptomless, but if they become filled with impacted fecal material, they can enlarge and become inflamed; causing a condition termed *diverticulitis*, which is both painful and potentially serious. There is epidemiological evidence to support the hypothesis that a high intake of dietary fiber protects against diverticular disease in populations in the USA and Europe, but it is not known whether this is due to physical causes as envisaged by Burkitt, or to some unidentified effect acting via the colonic microbiota. Increased consumption of fiber is often recommended for the management of established disease, but the evidence base to support this advice remains relatively weak.

Irritable bowel syndrome (IBS)

IBS is a poorly defined functional bowel disorder with a variety of symptoms, including abdominal pain, bloating, constipation or chronic diarrhea, with no established organic cause. Increased consumption of dietary fiber, either from high-fiber foods, or from supplements, is often recommended for the management of IBS but the clinical evidence for this is not strong. Various types of

dietary fiber have been used in the management of IBS but the most common have been wheat bran, which has a high fecal bulking capacity, and ispaghula, which is soluble, only partially fermentable and commonly prescribed as a mild laxative and stool-softener. The literature suggests that ispaghula is generally more effective than wheat bran in the treatment of IBS symptoms. However, given the diversity of IBS symptoms, and the multiple properties that characterize the various components of dietary fiber, a much more selective and personalized approach is probably needed for the rational use of fiber in IBS therapy.

Colorectal cancer (CRC)

Colorectal carcinoma remains one of the most prevalent and potentially fatal cancers in industrialized countries, where the incidence is around five times higher than in the rural communities of many countries of Africa and southern Asia. Burkitt observed this disparity at first hand in Africa and proposed that diets rich in dietary fiber protect against CRC due to the stool bulking effects of undigested polysaccharides. His hypothesis, simply stated, was that Western dietary patterns, typically low in dietary fiber, caused prolonged exposure of the colonic mucosa to fecal carcinogens. Consistent with this hypothesis, Cummings et al. (1992) later showed a statistically significant inverse relationship between fecal mass and risk of CRC across a range of population groups. However, as with all observational studies, these data could only demonstrate an association, with no proof of a causal mechanism. Later epidemiological studies, often conducted within populations such as that of the USA where the range of intakes of dietary fiber is relatively limited, failed to confirm the hypothesis. For example, in a systematic review of cohort studies, Park et al. pooled the results of 13 investigations and showed that among those in the highest quintile of fiber consumption the risk of CRC was 16% lower than among those in the lowest quintile. However, when the data were adjusted for confounding factors. The protective association was reduced to 6% and became statistically non-significant. In contrast, strong evidence for a protective effect of fiber consumption against CRC emerged from the European Prospective Investigation of Cancer and Nutrition (EPIC) project, which is based on a cohort of approximately half a million individuals drawn from geographically and culturally distinct European populations with widely differing diets. Analysis of these data after 11 years of follow-up indicated that total dietary fiber intake was inversely associated with colorectal cancer (Murphy et al., 2012). The hazard ratio per 10 g/day increase in fiber intake was 0.87 (95% confidence interval: 0.79–0.96). A meta-analysis performed by the Scientific Advisory Committee on Nutrition (SACN) in the UK also identified a protective dose-response relationship between consumption of dietary fiber and CRC, and this was used as a criterion for the recommendation that the adult population of the UK should consume at least 30 g of dietary fiber per day. Further support for these findings and recommendations has come from more recent meta-analyses of prospective studies. The protective effects of fiber consumption against CRC are now generally acknowledged.

In most cases CRC develops from relatively common lesions of the colorectal mucosa called adenomatous polyps, a minority of which eventually develop into carcinoma. Polyps are frequently identified during endoscopy carried out for screening or diagnostic investigations and their removal is a routine preventative measure. If a high intake of fiber is protective against CRC, then it is reasonable to expect a similar protection against the development of polyps. A recent systematic review of epidemiological studies supports this conjecture (Nucci et al., 2021). It might also be expected that an increased consumption of fiber would reduce the risk of polyp recurrence after polypectomy. A review of intervention trials with different types of dietary fiber found no evidence to support this hypothesis, but the authors questioned the overall quality of the research to date and recommended longer trials using higher levels of fiber.

Although the original proposal of Burkitt that fiber intake reduces the risk of CRC appears to be correct, the mechanisms underlying the protective effect remain uncertain. The simple hypothesis that fiber reduces both the level and duration of exposure to carcinogens has not been disproved but it is unlikely to be the whole explanation. The development of CRC begins with a sequence of somatic mutations in the crypt cells of the colorectal mucosa, which drive abnormalities of gene expression during epithelial cell-proliferation and differentiation. These changes accumulate over decades and are probably caused both by local genetic damage from fecal mutagens and adverse metabolic drivers related to overweight, insulin-sensitivity and low physical activity. Higher consumption of dietary fiber probably protects against CRC through a variety of mechanisms, ranging from local effects on the colorectal epithelia, mediated via intraluminal fermentation products, to systemic effects, including the reduced risk of obesity and improved metabolic health.

Conclusion

Although the founders of the dietary fiber hypothesis may have over-stated the role of inadequate fiber consumption as a cause of non-infectious diseases in western countries, the strength of their convictions, coupled with their pioneering observations on rural populations eating traditional diets, launched decades of experimental research and more robust epidemiology. The outcome has been a much deeper, though rather more nuanced understanding of the role of fiber in human health. There is probably no disease that can be attributed entirely to a deficiency of dietary fiber, and yet a surprising proportion of what the pioneers believed has turned out to be correct; the complex group of dietary constituents that we now call dietary fiber indisputably contributes to the maintenance of human health, though often through mechanisms that would have been beyond speculation in the mid 1960s. Despite this modern consensus, consumption of dietary fiber in most western countries continues to fall significantly below dietary recommendations. This problem may prove difficult to rectify as the human food chain evolves in response to global trends in agricultural production, food supply and culinary practice. With increasing prosperity, populations tend to rely more on manufactured,

often highly processed foods for a large fraction of their energy intake. In these circumstances a smaller proportion of fiber consumption will come from unprocessed plant foods, and more will need to be incorporated into diets in the form of isolated cell-wall polysaccharides, or as resistant starches, formed during food processing, or perhaps enriched in food products through the creation of novel varieties of starchy crops. To retain, or better still to maximize the benefits of fiber for human health, it will be essential to continue to explore the physiological effects of these substances, and their interactions with the microbiota. With a growing understanding of human molecular physiology, advances in bacterial genomic sequencing and increasingly sophisticated bioinformatics, we can be confident of continued progress.

References

Burkitt, D.P., 1975. Dietary fibre and "pressure diseases". J. Roy. Coll. Phys. Lond. 9, 138–147.

Codex Alimentarius, 2010. FAO, Rome: Joint FAO/WHO food Standards Program, Secretariat of the Codex Alimentarius Commission. Guidelines on Nutrition Labelling CAC/GL 2-1985 as last amended 2010.

Cummings, J.H., Bingham, S.A., Heaton, K.W., Eastwood, M.A., 1992. Fecal weight, colon cancer risk, and dietary intake of nonstarch polysaccharides (dietary fiber). Gastroenterology 103, 1783–1789.

Englyst, K., Quigley, M., Englyst, H., et al., 2013. Evaluation of methods of analysis for dietary fiber using real foods and model foods. Food Chem. 140, 568–573.

Evans, C.E.L., Greenwood, D.C., Threapleton, D.E., et al., 2015. Effects of dietary fiber type on blood pressure: a systematic review and meta-analysis of randomized controlled trials of healthy individuals. J. Hypertens. 33, 897–911.

Jenkins, D.J., Wolever, T.M., Leeds, A.R., et al., 1978. Dietary fibers, fiber analogues and glucose tolerance: importance of viscosity. Br. Med. J. 1, 1392–1394.

Ma, W., Nguyen, L.H., Song, M., 2021. Dietary fiber intake, the gut microbiome, and chronic systemic inflammation in a cohort of adult men. Genome Med. 13, 102. https://doi.org/10.1186/s13073-021-00921-y.

Murphy, N., Norat, T., Ferrani, P., et al., 2012. Dietary fiber intake and risks of cancers of the colon and rectum in the European Prospective Investigation into cancer and nutrition (EPIC). PLoS One 7 (6), e39361.

Nucci, D., Fatigoni, C., Salvatori, T., et al., 2021. Association between dietary fiber intake and colorectal adenoma: a systematic review and meta-analysis. Int. J. Environ. Res. Publ. Health 18 (8), 4168.

Pereira, M.A., O'Reilly, E., Augustsson, K., et al., 2004. Dietary fiber and risk of coronary heart disease: a pooled analysis of cohort studies. Arch. Intern. Med. 164, 370–376.

Reynolds, A., Mann, J., Cummings, J., et al., 2019. Carbohydrate quality and human health: a series of systematic reviews and meta-analyses. Lancet 393, 434–445.

Suares, N.C., Ford, A.C., 2011. Systematic review: the effects of fiber in the management of chronic idiopathic constipation. Aliment. Pharmacol. Ther. 33, 895–901.

Trowell, H., 1972. Ischemic heart disease and dietary fiber. Am. J. Clin. Nutr. 25, 926–932.

Further reading

Adeshirlarijaney, A., Gewirtz, A.T., 2020. Considering gut microbiota in treatment of type 2 diabetes mellitus. Gut Microb. 11, 252–264.

Askani, E., Rospleszcz, S., Rothenbacher, T., et al., 2021. Dietary habits and the presence and degree of asymptomatic diverticular disease by magnetic resonance imaging in a Western population: a population-based cohort study. Nutr. Metabol. 18 (1), 73.

Burkitt, D.P., 1971. Epidemiology of cancer of the colon and rectum. Cancer 28, 3–13.

Buyken, A.E., Mela, D.J., Dussort, P., et al., 2018. Dietary carbohydrates: a review of international recommendations and the methods used to derive them. Eur. J. Clin. Nutr. 72, 1625–1643.

Carabotti, M., Annibale, B., 2018. Treatment of diverticular disease: an update on latest evidence and clinical implications. Drugs Context 7, 212526.

Cleave, T.L., Campbell, G.D., 1966. Diabetes, Coronary Thrombosis and the Saccharine Disease. John Wright & Sons Ltd, Bristol.

Cummings, J.H., Engineer, A., 2018. Denis Burkitt and the origins of the dietary fiber hypothesis. Nutr. Res. Rev. 31, 1–15.

Dagbasi, A., Lett, A.M., Murphy, K., Frost, G., 2020. Understanding the interplay between food structure, intestinal bacterial fermentation and appetite control. Proc. Nutr. Soc. 79, 514–530.

Dukas, L., Willett, W.C., Giovannucci, E.L., 2003. Association between physical activity, fiber intake, and other lifestyle variables and constipation in a study of women. Am. J. Gastroenterol. 98, 1790–1796.

Elia, M., Cummings, J.H., 2007. Physiological aspects of energy metabolism and gastrointestinal effects of carbohydrates. Eur. J. Clin. Nutr. 61 (Suppl. 1), 40–74.

Ellis, P.R., Roberts, F.G., Low, A.G., Morgan, L.M., 1995. The effect of high-molecular-weight guar gum on net apparent glucose absorption and net apparent insulin and gastric inhibitory peptide production in the growing pig: relationship to rheological changes in jejunal digesta. Br. J. Nutr. 74, 539–556.

Englyst, H., Wiggins, H.S., Cummings, J.H., 1982. Determination of the non-starch polysaccharides in plant foods by gas-liquid chromatography of constituent sugars as alditol acetates. Analyst 107, 307–318.

Food and Agriculture Organisation, 2002. Food Energy—Methods of Analysis and Conversion Factors. Food and Agriculture Organisation of the United Nations, Rome.

Food and Agriculture Organisation, 2010. Endorsement of Methods of Analysis and Sampling. B. Methods of Analysis for Dietary Fiber: Guidelines for Use of Nutrition and Health Claims. Codex Alimentarius, Rome.

Ford, A.C., Talley, N.J., Spiegel, B.M., et al., 2008. Effect of fiber, antispasmodics, and peppermint oil in the treatment of irritable bowel syndrome: systematic review and meta-analysis. Br. Med. J. 337, a2313.

Gee, J.M., Blackburn, N.A., Johnson, I.T., 1983. The influence of guar gum on intestinal cholesterol transport in the rat. Br. J. Nutr. 50, 215–224.

Guo, J., Tan, L., Kong, L., 2021. Impact of dietary intake of resistant starch on obesity and associated metabolic profiles in human: a systematic review of the literature. Crit. Rev. Food Sci. Nutr. 61, 889–905.

Hartley, L., May, M.D., Loveman, E., Colquitt, J.L., Rees, K., 2016. Dietary fiber for the primary prevention of cardiovascular disease. Cochrane Database Syst. Rev. 1, CD011472.

Holscher, H.D., 2017. Dietary fiber and prebiotics and the gastrointestinal microbiota. Gut Microb. 8, 172–184.

Johnson, I.T., Gee, J.M., 1981. Effect of gel-forming gums on the intestinal unstirred layer and sugar transport in vitro. Gut 22, 398–403.

Jovanovski, E., Khayyat, R., Zurbau, A., 2019. Should viscous fiber supplements be considered in diabetes control? Results from a systematic review and meta-analysis of randomized controlled trials. Diabetes Care 42, 755–766.

Juvonen, K.R., Purhonen, A.-K., Salmenkallio-Martilla, M., et al., 2009. Viscosity of oat bran-enriched beverages influences gastrointestinal hormone responses in healthy humans. J. Nutr. 139, 461–466.

Kasahara, K., Krautkramer, K.A., Org, E., et al., 2018. Interactions between *Roseburia intestinalis* and diet modulate atherogenesis in a murine model. Nat. Microbiol. 3, 1461–1471.

Kim, Y., Je, Y., 2016. Dietary fiber intake and mortality from cardiovascular disease and all cancers: a meta-analysis of prospective cohort studies. Arch. Cardiovasc. Dis. 109, 39–54.
Koh, A., De Vadder, F., Kovatcheva-Datchary, P., Backhed, F., 2016. From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. Cell 165, 1332–1345.
Livesey, G., Taylor, R., Livesey, H.F., et al., 2019. Dietary glycaemic index and load and the risk of type-2 diabetes: assessment of causal relations. Nutrients 11, 1436–1470.
Mackie, A.R., Bajka, B.H., Rigby, N.M., et al., 2017. Oatmeal particle size alters glycemic index, but not as a function of gastric emptying rate. Am. J. Physiol. Gastrointest. Liver Physiol. 313. https://doi.org/10.1152/ajpgi.00005.2017.
McCance, R.A., Lawrence, R.D., 1929. The Carbohydrate Content of Foods. Medical Research Council, London.
Park, Y., Hunter, D.J., Spiegelman, D., et al., 2005. Dietary fiber intake and risk of colorectal cancer: a pooled analysis of prospective cohort studies. J. Am. Med. Assoc. 294, 2849–2857.
Pasolli, E., Asnicar, F., Manara, S., et al., 2019. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. Cell 176, 649–662.e620.
Prosky, L., Asp, N.G., Schweizer, T.F., DeVries, J.W., Furda, I., 1988. Determination of insoluble, soluble, and total dietary fiber in foods and food products: interlaboratory study. J. Assoc. Off. Anal. Chem. 71, 1017–1023.
Reynolds, A.N., Akerman, A.S., Mann, J., 2020. Dietary fiber and whole grains in diabetes management: systematic review and meta-analyses. PLoS Med. 17, e1003053.
Roediger, W.E., 1980. Role of anaerobic bacteria in the metabolic welfare of the colonic mucosa in man. Gut 21, 793–798.
SACN, 2015. Carbohydrates and Health. The Stationary Office (TSO), London.
Sanchez, H.N., Moroney, J.B., Gan, H., et al., 2020. B cell-intrinsic epigenetic responses by dietary fiber-derived short chain fatty acids. Nat. Commun. 11, 60. https://doi.org/10.1038/s41467-019-13603-6.
Sievenpiper, J.L., 2020. Low-carbohydrate diets and cardiovascular health: the importance of carbohydrate quality over quantity. Nutr. Rev. 78, 69–77.
So, D., Gibson, P.R., Muir, J.G., Yao, C.K., 2021. Dietary fibers and IBS: translating functional characteristics to clinical value in the era of personalized medicine. Gut. https://doi.org/10.1136/gutjnl-2021-324891.
Sonnenburg, E.D., Smits, S.A., Tikhonov, M., 2016. Diet-induced extinctions in the gut microbiota compound over generations. Nature 529, 212–215.
Southgate, D.A., 1969. Determination of carbohydrates in foods. II. Unavailable carbohydrates. J. Sci. Food Agric. 20, 331–335.
Tarleton, S., Dibaise, J.K., 2011. Low-residue diet in diverticular disease: putting an end to a myth. Nutr. Clin. Pract. 26, 137–142.
Tong, T.Y.N., Appleby, P.N., Key, T.J., et al., 2020. The associations of major foods and fibre with risks of ischaemic and haemorrhagic stroke: a prospective study of 418,329 participants in the EPIC cohort across nine European countries. Eur. Heart J. 41, 2632–2640.
Trowell, H.C., 1983. Causes of insulin dependent diabetes. Br. Med. J. 287, 430.
Veettil, S.K., Wong, T.Y., Loo, Y.S., et al., 2021. Role of diet in colorectal cancer incidence: umbrella review of meta-analyses of prospective observational studies. JAMA Netw. Open. https://doi.org/10.1001/jamanetworkopen.2020.37341.
Vincent, R., Roberts, A., Frier, M., Perkins, A.C., MacDonald, I.A., Spiller, R.C., 1995. Effect of bran particle size on gastric emptying and small bowel transit in humans: a scintigraphic study. Gut 37, 216–219.
Wilbrink, J., Masclee, G., Klaassen, T., et al., 2021. Review on the regional effects of gastrointestinal luminal stimulation on appetite and energy intake: (pre)clinical observations. Nutrients 13, 1601.
Wursch, P., Del Vedovo, S., Koellreutter, B., 1986. Cell structure and starch nature as key determinants of the digestion rate of starch in legume. Am. J. Clin. Nutr. 43, 25–29.
Xu, B., Fu, J., Qiauo, Y., 2021. Higher intake of microbiota-accessible carbohydrates and improved cardiometabolic risk factors: a meta-analysis and umbrella review of dietary management in patients with type 2 diabetes. Am. J. Clin. Nutr. 113, 1515–1530.
Yao, Y., Suo, T., Andersson, R., et al., 2017. Dietary fiber for the prevention of recurrent colorectal adenomas and carcinomas. Cochrane Database Syst. Rev. 1, CD003430.

Fatty acids: Metabolism

P.A. Watkins, Kennedy Krieger Institute, and Johns Hopkins University School of Medicine, Baltimore, MD, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of P.A. Watkins, Fatty acids: Metabolism, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 220–230, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00103-3.

Introduction 316
Fatty acid nomenclature conventions 317
Physical properties of fatty acids 317
Fatty acid activation 317
Mitochondrial fatty acid β-oxidation 317
Peroxisomal fatty acid β-oxidation 320
Fatty acid α-oxidation and ω-oxidation 322
Fatty acid *de novo* synthesis 323
Mitochondrial fatty acid synthesis 324
Fatty acid elongation 324
Fatty acid unsaturation and the essential FAs 325
Fatty acids as components of complex lipids 325
Oxylipin synthesis 326
Endocannabinoids 327
Fatty acylation of proteins 327
Vitamins and fatty acid metabolism 328
Regulation of fatty acid metabolism 328
Summary 328
References 328
Further reading 329

Key points

- The term "fatty acid" comprises a large group of biologically and structurally diverse molecules. The metabolism of these compounds is equally diverse.
- Fatty acids can be structural components of more complex lipids, including triacylglycerol and phospholipids. They can be degraded for energy production; conversely, fatty acids can be synthesized from carbohydrates for energy storage. Polyunsaturated fatty acids can be converted into signaling molecules such as oxylipins, and are constituents of endocannabinoids. Fatty acids can regulate protein function by reversible or irreversible covalent attachment.
- The objective of this article is to provide an overview of the major metabolic pathways in which fatty acids participate.

Glossary

Fatty acid An aliphatic hydrocarbon chain terminating in a carboxylic acid function

Fatty acid activation Formation of fatty acyl-CoA *via* the attachment of coenzyme A to the fatty acid's carboxylic acid function *via* a thioester bond

Fatty acid *de novo* synthesis Synthesis of a saturated, 16-carbon fatty acid from precursors typically derived from carbohydrate metabolism

β-Oxidation Energy-yielding cyclic process in which the hydrocarbon chain of a fatty acid is shortened, typically by two carbon atoms per cycle

Protein acylation The covalent modification of an amino acid residue in a protein by a fatty acid, most commonly a 14- or 16-carbon saturated fatty acid

Introduction

Fatty acids (FAs) are hydrophobic molecules consisting of an aliphatic hydrocarbon chain terminating in a carboxylic acid moiety. Structurally, FAs are quite diverse. Although FAs containing 16–18 carbons are the most abundant in nature, the hydrocarbon chain can vary in length from two to more than 26 carbons. The carbon chain can be fully saturated, or contain one or more double bonds.

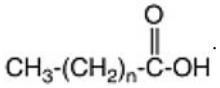
Ingestion of dietary fats and oils is a major source of FAs for humans and most other animals. In addition, many physiologically important FAs can be synthesized *de novo* from metabolites derived from the catabolism of sugars and proteins.

FAs in the human body serve two primary functions—they are an excellent source of metabolic energy, and they serve as building blocks for many diverse complex lipids. FAs are also precursors of bioactive molecules such as prostaglandins and other eicosanoids. An important covalent modification of proteins is the attachment of a 14- or 16-carbon FA. In addition, FAs and their coenzyme A derivatives have many metabolic regulatory roles.

Fatty acid nomenclature conventions

In this article, FAs will be identified by their chain length, the number of double bonds present, and the position of the first double bond from the methyl end of the molecule. Thus 14:0 denotes a 14 carbon saturated FA, 16:1ω9 denotes a 16 carbon monounsaturated FA in which one double bond occurs nine carbons from the omega (methyl) end, and 20:4ω6 denotes a 20 carbon polyunsaturated FA in which the first of four double bonds is found six carbons from the methyl end. Unless otherwise noted, all double bonds are in the *cis* configuration and double bonds in polyunsaturated FAs are always separated by a single methylene ($-CH_2-$) group. The carboxyl carbon of any FA is carbon-1. The adjacent carbon is referred to as either carbon-2 or the α-carbon; the next is carbon-3 or the β-carbon, and so on. Some examples are shown in **Fig. 1**.

Physical properties of fatty acids

FAs are aliphatic organic acids with the fundamental structure 

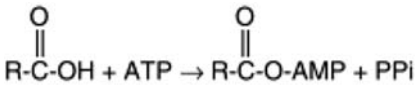
.

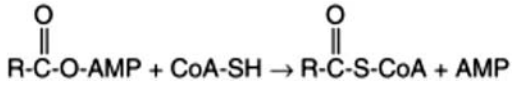
Where *n* can range from 0 to >26. Thus, FAs range from the shortest, acetic acid (2:0), to the very long-chain FAs containing 26 or more carbon atoms (e.g., 26:0). Although FAs with an odd number of carbon atoms exist in nature, most common FAs have an even number. The most abundant FAs found in human lipids, as well as in dietary lipids, are the long-chain FAs 16:0 (palmitic acid) and 18:1ω9 (oleic acid) (**Fig. 1**). The hydrophobic nature of the hydrocarbon chain of FAs containing more than eight carbons renders them quite insoluble in aqueous media. It has been estimated that for every two carbon increase in FA chain length, its solubility decreases 10 fold (Miura and Nonomura, 2019).

Owing to the poor solubility of the most abundant FAs, free (nonesterified) FAs are often found associated with binding or transport proteins (Schwenk et al., 2010; Xu et al., 2019). Serum albumin has at least six binding sites for FAs and is the primary transporter of these molecules through the bloodstream. Several low molecular weight intracellular FA binding proteins (FABPs) have also been identified. Free FAs can also associate with lipophilic cellular and organellar membranes; however, concentrations of these nonesterified compounds in membranes is typically very low.

Fatty acid activation

FAs are generally nonreactive unless first "activated" in an ATP-dependent process by thioesterification to coenzyme A (CoA) (reviewed in Watkins, 1997). Activation is catalyzed by acyl-CoA synthetases (ACS; E.C. 6.2.1.x) *via* a bi uni uni bi ping-pong mechanism. In the first half-reaction, an acyl-adenylate intermediate is formed, with the release of inorganic pyrophosphate (PPi):



Pyrophosphatases rapidly cleave PPi, effectively preventing reversal of this reaction. In the second half-reaction, CoA displaces AMP to form the acyl-CoA: 

The thioester bond between the acyl moiety and CoA is a high-energy bond that facilitates subsequent participation of the FA in metabolic pathways. Humans have more than 25 ACSs that differ in their tissue expression, subcellular location, and FA chain length preference.

Mitochondrial fatty acid β-oxidation

To recover their stored energy, FAs must be oxidized. Quantitatively, the most important energy-yielding degradation pathway is mitochondrial β-oxidation (**Fig. 2**) (reviewed in Houten et al., 2016). In the fed state, triacylglycerol in circulating lipoproteins

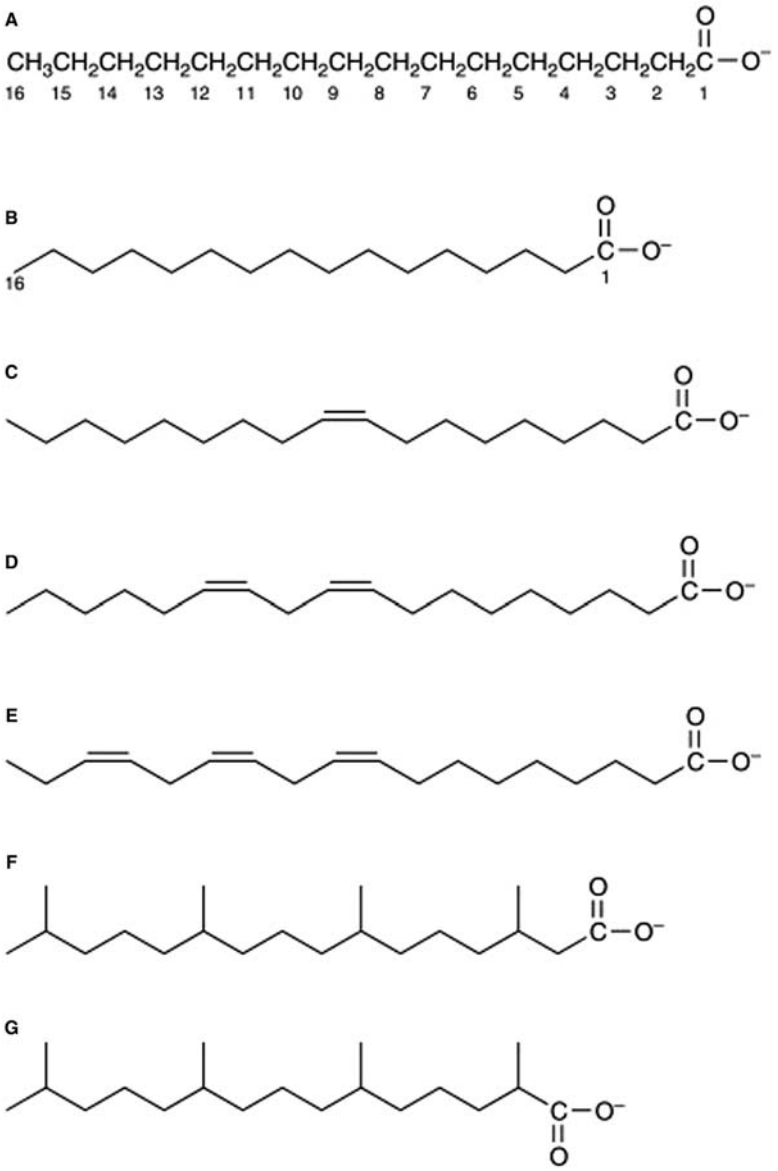


Fig. 1 Fatty acid structure and nomenclature. (A) Chemical formula and carbon numbering system for a 16-carbon saturated FA (16:0). (B) Schematic representation of 16:0. (C) A monounsaturated FA, 18:1ω9, showing the double bond nine carbons from the methyl end (carbon 18). (D) The essential ω6 FA, 18:2ω6 (linoleic acid), where the first double bond is found six carbons from the methyl end. The two double bonds are separated by a methylene (–CH_2–) group. (E) The essential ω3 FA, 18:3ω3 (α-linolenic acid), where the first double bond is found three carbons from the methyl end. (F) Phytanic acid, a dietary β-methyl-branched-chain FA (3,7,11,15-tetramethyl 16:0). The methyl group on carbon-3 prevents this FA from degradation by β-oxidation. (G) Pristanic acid (2,6,10,14-tetramethyl 15:0) is the product of phytanic acid α-oxidation, in which a single carbon (carbon 1) is lost. The methyl group on carbon-2 does not preclude subsequent degradation by β-oxidation.

delivers FAs to tissues; hydrolysis by lipoprotein lipases in the capillary endothelium releases FAs for cellular uptake. In the fasted state, FAs are released from triacylglycerol stored in adipocytes and delivered to tissues bound to serum albumin. Because of their hydrophobic nature, FAs can traverse the plasma membrane by simple diffusion; however, proteins such as CD36, **GOT2**, and **SLC27A1-6** have also been proposed to play a role in membrane FA transport (Official Symbols for genes/proteins are shown in bold typeface throughout this article; other commonly used abbreviations are in normal typeface).

ACS activity toward long-chain FA substrates is present in the outer mitochondrial membrane. However, fatty acyl-CoAs do not readily traverse biological membranes such as the inner mitochondrial membrane. A highly sophisticated transport system has evolved that allows for tight regulation of FA entry into the mitochondrion (Fig. 2). Carnitine palmitoyl transferase 1 (**CPT1**), located on the inner aspect of the outer mitochondrial membrane, catalyzes a transesterification reaction: Fatty acyl–CoA + carnitine → fatty acyl–carnitine + CoA–SH. Carnitine:acylcarnitine translocase (CACT; **SLC25A20**), located in the inner mitochondrial membrane, carries the fatty acyl-carnitine inside the mitochondrion in exchange for a free carnitine molecule.

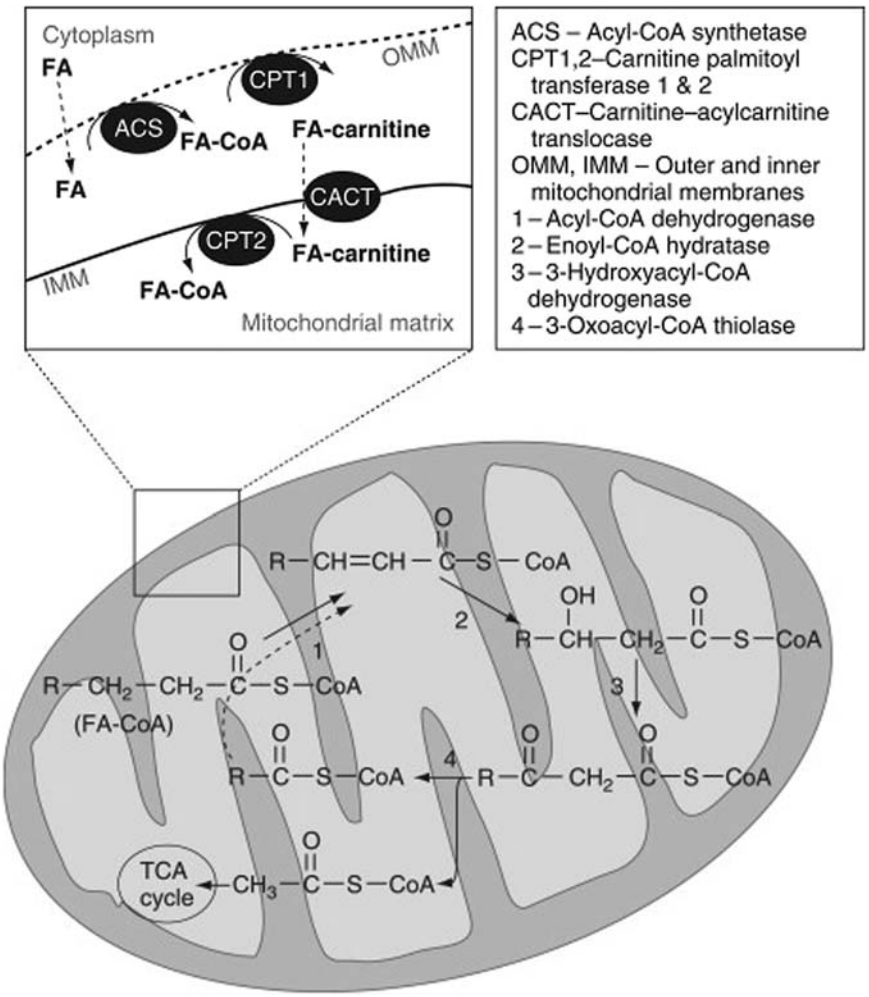


Fig. 2 Mitochondrial fatty acid β-oxidation pathway. Long-chain FAs are activated, converted to carnitine esters, transported across the inner mitochondrial membrane, and re-converted to their CoA thioester once in the mitochondrial matrix. Four sequential mitochondrial enzyme reactions shorten the fatty acyl-CoA by two carbons, which are released as acetyl-CoA. The shortened fatty acyl-CoA can undergo additional cycles of degradation until the entire carbon chain has been converted to acetyl-CoA units. $FADH_2$ and NADH, produced in reactions 1 and 3, respectively, can enter the electron transport chain for ATP production. Acetyl-CoA enters the tricarboxylic acid (TCA) cycle, yielding additional NADH and $FADH_2$ for ATP production. Mitochondrial β-oxidation is the primary pathway for recovering the energy stored as triacylglycerol, or "fat".

CPT2, located inside the mitochondrion, then catalyzes the reversal of the **CPT1** reaction. Thus the concerted actions of **CPT1**, CACT, and **CPT2** effectively translocate fatty acyl-CoA across the inner mitochondrial membrane. Entry of FAs into the mitochondrion is tightly regulated primarily at the **CPT1** step by malonyl-CoA. This intermediate in FA synthesis (*see* **Fatty acid *de novo* synthesis**, below) is a potent inhibitor of **CPT1.**

The entry of medium-chain FAs into the mitochondrial matrix does not utilize the carnitine pathway (Papamandjaris et al., 1998). These FAs can "flip-flop" across the inner mitochondrial membrane by simple diffusion, Medium-chain ACSs located in the matrix then activate them intramitochondrially. Thus, the subcellular location of the ACS responsible for activation of long-chain FAs (outer mitochondrial membrane) versus medium-chain FAs (mitochondrial matrix) dictates whether entry into the β-oxidation pathway is regulated (long-chain FA) or unregulated (medium-chain FA). The latter is of particular nutritional significance for newborns, as breast milk is rich in medium-chain FAs. Bypassing the regulated carnitine transport pathway allows for more rapid energy production.

The four primary enzyme activities of mitochondrial β-oxidation act on intramitochondrial fatty acyl-CoA by: (1) dehydrogenation (acyl-CoA dehydrogenase), (2) hydration (enoyl-CoA hydratase), (3) dehydrogenation (3-hydroxyacyl-CoA dehydrogenase), and (4) cleavage (3-oxoacyl-CoA thiolase). The products are fatty acyl-CoA that has been shortened by 2 carbons, acetyl-CoA, $FADH_2$ (from reaction 1) and NADH (from reaction 3). $FADH_2$ and NADH directly enter the electron transport chain, yielding approximately 5 ATP molecules. Acetyl-CoA can be further degraded to CO_2 and water by the tricarboxylic acid cycle, yielding additional ATP molecules. Importantly, the entire β-oxidation process can be repeated multiple times using the shortened fatty acyl-CoA as substrate until the entire carbon skeleton of the FA has been degraded to 2-carbon acetyl-CoA units. Complete oxidation of one molecule of 16:0 (β-oxidation + tricarboxylic acid cycle) yields more than 160 ATP molecules.

Depending on the acyl chain length, the first enzymatic step of mitochondrial β-oxidation is carried out by either short-, medium-, or very long-chain acyl-CoA dehydrogenase (**ACADS**, ACADM, and ACADVL, respectively). Long-chain acyl-CoA dehydrogenase (**ACADL**) functions in rodents but not in humans. Instead, VLCAD acts on long-chain acyl-CoAs in most human tissues whereas another dehydrogenase, ACAD9, performs this function in brain. For medium-to long-chain fatty acyl-CoAs, the three subsequent β-oxidation reactions are catalyzed by mitochondrial trifunctional protein (HADHA), whereas three separate enzymes (**ECHS1**, **HADH**, and **ACAA2**) are needed for oxidation of short-chain acyl-CoAs. Genetic defects in essentially all of the mitochondrial β-oxidation enzymes, as well as the carnitine transport system, are known. If unrecognized, these FA oxidation disorders are often fatal, and are a recognized cause of sudden infant death syndrome (SIDS).

FAs with an odd number of carbons in their acyl chain are degraded in the same manner as even-chain FAs by the normal β-oxidation pathway. However, the final degradation product is the CoA derivative of the 3-carbon molecule, propionic acid. Propionyl-CoA can be metabolized to succinyl-CoA, which is a tricarboxylic acid cycle intermediate, β-Oxidation of unsaturated FAs requires participation of one or more ancillary enzymes, such as enoyl-CoA isomerase and 2,4-dienoyl-CoA reductase, to change the position and/or conformation of the double bond and thus allow degradation to go to completion.

Some tissues, for example skeletal muscle, completely oxidize FAs to CO_2 and water. Others, for example liver, only partially oxidize FAs, using the acetyl-CoA product for biosynthetic needs. In particular, liver utilizes intramitochondrial acetyl-CoA for the synthesis of ketone bodies, acetoacetate, and β-hydroxybutyrate (**Fig. 3**) (Rui, 2014). Ketone bodies can be oxidized by all tissues except liver and provide an alternative fuel source during starvation. In particular, brain and nerve, which do not derive energy from FA oxidation, can oxidize ketone bodies. During prolonged starvation, increased ketone body utilization spares the brain's requirement for glucose.

Peroxisomal fatty acid β-oxidation

Like mitochondria, peroxisomes contain pathways for the β-oxidation of FAs that yield a chain-shortened FA-CoA, acetyl-CoA (or propionyl-CoA), NADH, and $FADH_2$ (**Fig. 4**) (reviewed in Wanders, 2014, and Wanders et al., 2020). Unlike mitochondria, peroxisomes do not contain an electron transport chain or tricarboxylic acid cycle and thus FA degradation is not directly coupled to energy production. Typically, peroxisomes degrade FA substrates that cannot be catabolized in mitochondria. Peroxisomes are indispensable for the degradation of saturated very long-chain FAs (VLCFA; containing more than 22 carbon atoms), which are neurotoxic if allowed to accumulate. Detoxification *via* several cycles of peroxisomal β-oxidation decreases VLCFA chain-length to 8–10 carbons, after which they are converted to carnitine derivatives, exit peroxisomes, and translocate to the mitochondrion

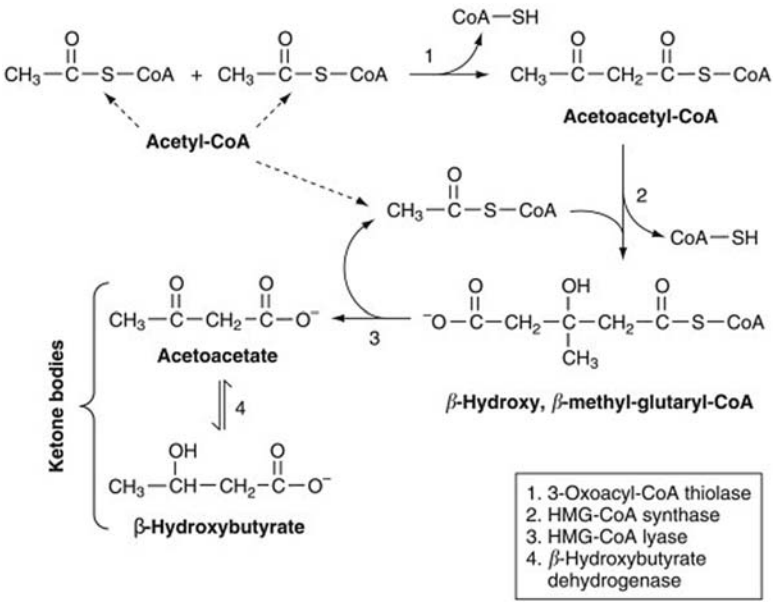


Fig. 3 Ketone body synthesis. In the mitochondrion of liver hepatocytes, acetyl-CoA derived from β-oxidation is converted to "ketone bodies", primarily acetoacetate and β-hydroxybutyrate, rather than enter the tricarboxylic acid cycle. Two molecules of acetyl-CoA condense in a reversal of the last β-oxidation reaction (3-oxoacyl-CoA thiolase). The product, acetoacetyl-CoA, condenses with another molecule of acetyl-CoA, yielding β-hydroxy, β-methyl-glutaryl-CoA (HMG-CoA), a reaction catalyzed by HMG-CoA synthase. Cleavage of HMG-CoA by HMG-CoA lyase yields acetoacetate, regenerating one molecule of acetyl-CoA. Acetoacetate is reversibly reduced to β-hydroxybutyrate *via* the NAD-dependent enzyme β-hydroxybutyrate dehydrogenase. These ketone bodies can traverse the inner mitochondrial membrane, eventually reaching the bloodstream for ultimate utilization by brain and other tissues.

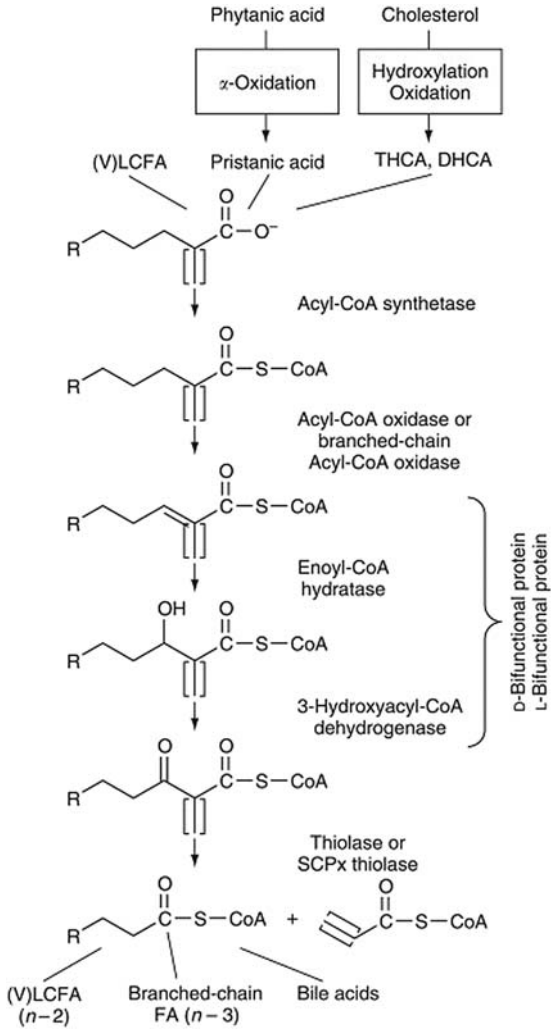


Fig. 4 Peroxisomal fatty acid β-oxidation pathways. Although saturated long-chain FAs (LCFA) are preferentially degraded in mitochondria, saturated very long-chain FAs (VLCFA) and some LCFA are shortened by peroxisomal β-oxidation. Degradation of pristanic acid, the product of phytanic acid α-oxidation, and the conversion of the cholesterol-derived 27-carbon bile acid precursors DHCA and THCA (di- and tri-hydroxycholestanoic acids) to 24-carbon bile acids also require this pathway. Four enzymatic reactions serve to shorten the substrates by either two (LCFA, VLCFA) or three (pristanic acid, DHCA, THCA) carbons. The 2-methyl group of the latter substrates is shown in brackets.

for further catabolism. Degradation of xenobiotic fatty acyl-like compounds (e.g., sulfur-substituted FAs and many nonsteroidal anti-inflammatory drugs) also takes place in peroxisomes. Oxidation of dicarboxylic acids (from the diet or from ω-oxidation) or 2-methyl-branched-chain FAs (from the diet or from α-oxidation of phytanic acid) also occurs in peroxisomes.

FAs enter peroxisomes *via* a mechanism that is distinct from the mitochondrial **CPT1**/CACT/**CPT2** pathway. Long- and very-long-chain ACSs are associated with peroxisomes, and it is currently thought that FAs are activated to their CoA derivatives before entry into the peroxisomal matrix. Homodimers of peroxisomal membrane transporter proteins ABCD1 and ABCD2 are believed to transport saturated and unsaturated VLCFA-CoAs, respectively, into the matrix, while ABCD3 homodimers are responsible for the translocation of CoA derivatives of phytanic acid, pristanic acid, dicarboxylic FAs, and bile acid precursors (Wanders et al., 2020). The basic reactions of peroxisomal β-oxidation resemble those found in mitochondria, but the peroxisomal and mitochondrial enzymes are distinct proteins (Fig. 4). In fact, peroxisomes contain two sets of β-oxidation enzymes that appear to function with distinct substrates.

The first step in the oxidation of straight-chain FAs, (e.g., VLCFAs), is catalyzed by acyl-CoA oxidase 1 (**ACOX1**) (Fig. 4). **ACOX2** and **ACOX3** catalyze this initial desaturation reaction for α-methyl branched-chain substrates—bile acid precursors and pristanoyl-CoA, respectively (*see* FA α-Oxidation). For all substrates, enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase activities are catalyzed mainly by D-bifunctional protein (**HSD17B4**), which contains both activities. For straight-chain substrates, peroxisomal 3-ketoacyl-CoA thiolase (**ACAA1**) catalyzes the final β-oxidation reaction, whereas SCPx thiolase (SCP2) catalyzes this step for pristanic acid and bile acid precursors. Peroxisomes also contain L-bifunctional protein (**EHHADH**), which is thought to contribute to straight-chain FA oxidation. Human deficiencies of **ACOX1** and **HSD17B4** are associated with significant morbidity and mortality.

The peroxisomal β-oxidation pathway also performs important biosynthetic roles. In the hepatic synthesis of bile acids from cholesterol, the aliphatic side chain, which resembles a 2-methyl-branched-chain FA, must be shortened. A single cycle of peroxisomal β-oxidation removes a three-carbon portion of the side chain (as propionyl-CoA), converting the 27-carbon bile acid precursors di- and tri-hydroxycholestanoic acids into the 24-carbon primary bile acids chenodeoxycholate and cholate, respectively. Furthermore, peroxisomal β-oxidation is required for synthesis of docosahexaenoic acid (DHA 22:6ω3) (*see* **Fatty acid unsaturation and the essential FAs**, below).

Fatty acid α-oxidation and ω-oxidation

Other important FA catabolic pathways include α-oxidation (reviewed in Wanders, 2014) and ω-oxidation (reviewed in Miura, 2013). 3-Methyl-branched FAs, for example, phytanic acid (3,7,11,15-tetramethyl-16:0; **Fig. 1**), present in ruminant meats, fats, dairy products, and certain fish consumed in the diet, cannot be degraded by β-oxidation due to the methyl group on carbon-3. Shortening the FA chain by one-carbon (α-oxidation; **Fig. 5**) effectively shifts the position of the methyl group to carbon-2, thereby allowing subsequent degradation *via* β-oxidation. Phytanic acid is first activated to phytanoyl-CoA,translocated into peroxisomes via **ABCD3**, and is then hydroxylated on the 2-carbon by phytanoyl-CoA 2-hydroxylase (**PHYH**). 2-Hydroxyphytanoyl-CoA lyase (**HACL1**) cleaves a one-carbon CoA derivative, formyl-CoA, yielding an aldehyde, pristanal. A spice variant of fatty aldehyde dehydrogenase (**ALDH3H2**) is thought to oxidize pristanal to pristanic acid (2,6,10,14-tetramethyl-15:0). Both α-oxidation and pristanic acid β-oxidation occur in peroxisomes. Deficiency of **PHYH** is the primary cause of Refsum disease, a peripheral neuropathy, which, if untreated, causes cerebellar ataxia, retinitis pigmentosa, and deafness.

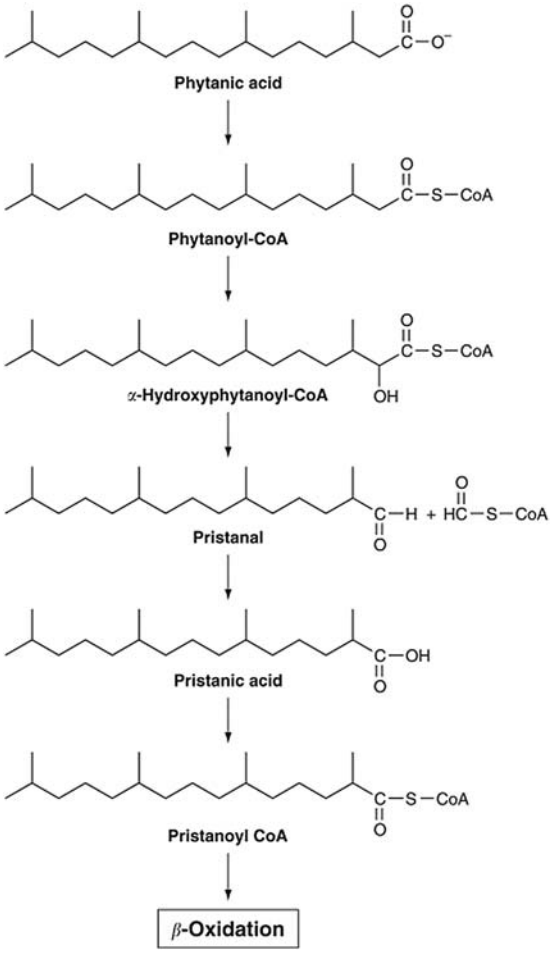


Fig. 5 Peroxisomal phytanic acid α-oxidation pathway. The dietary 3-methyl-branched FA, phytanic acid, is toxic if allowed to accumulate in tissues. Its 3-methyl group prevents degradation by β-oxidation; therefore, this FA is first shortened by one-carbon. Like the substrates for peroxisomal β-oxidation, the mechanism by which phytanic acid enters peroxisomes is not known. Activated phytanic acid is translocated into the peroxisomal matrix via **ABCD3**, after which it is hydroxylated on carbon-2. Cleavage between carbons 1 and 2 yields a one-carbon CoA compound, formyl-CoA, and an aldehyde, pristanal. After oxidation and re-activation to the CoA derivative, pristanoyl-CoA can be degraded by β-oxidation.

Like mitochondria, peroxisomes also require ancillary enzymes of FA metabolism, For unsaturated FA β-oxidation, enoyl-CoA isomerase and 2,4-dienoyl-CoA reductase are needed. In addition, the enzyme 2-methylacyl-CoA racemase (**AMACR**) is essential for proper metabolism of branched chain substrates. Methyl branches can be in either the R- or S-conformation, but **ACOX2** and **ACOX3** can oxidize only the S-conformers. **AMACR** catalyzes the interconversion of these isomers.

Another mechanism for degradation of FAs that cannot undergo β-oxidation is ω-oxidation. In this process, the terminal methyl group of a FA chain is oxidized to a carboxylic acid *via* various cytochrome P450 isozymes in the endoplasmic reticulum. The resulting dicarboxylic acids are then degraded by β-oxidation from the ω-end, primarily in peroxisomes.

Fatty acid *de novo* synthesis

Much of our need for FAs as constituents of phospholipids and other complex lipids are met by the diet. In addition, lipogenic tissues are capable of the *de novo* synthesis of FAs (**Fig. 6**) (reviewed in Nguyen et al., 2008). These tissues include liver (hepatocytes), adipose tissue, and lactating mammary gland. Much of the FAs synthesized by all three tissues are incorporated into triacylglycerol. Hepatic synthesis is primarily for export to other tissues (in very low-density lipoproteins), whereas synthesis in adipocytes and mammary gland is for local storage. In several respects, the enzymatic reactions of FA synthesis are the converse of those in FA oxidation. However, there are key differences, which are summarized in **Table 1**.

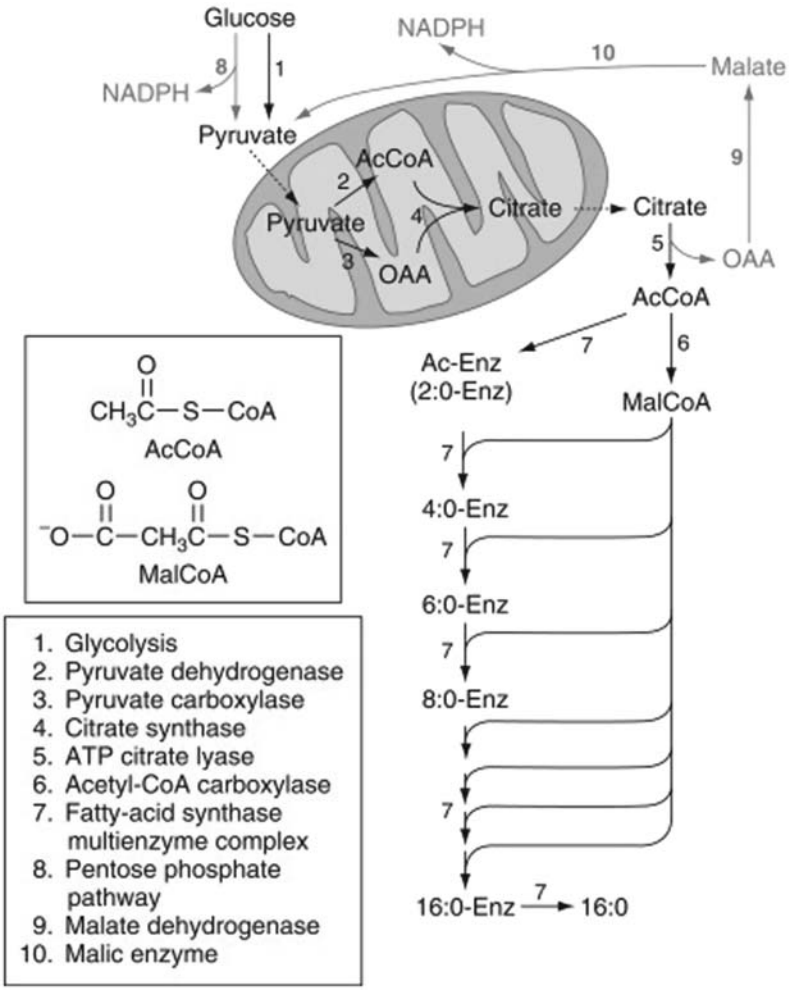


Fig. 6 Fatty acid biosynthesis. Cytoplasmic acetyl-CoA (AcCoA) is the primary substrate for *de novo* FA synthesis. This 2-carbon compound most commonly derives from the glycolytic degradation of glucose, and its formation is dependent on several reactions in mitochondria. The mitochondrial enzyme, pyruvate carboxylase, is found primarily in tissues that can synthesize FAs. AcCoA is converted to malonyl-CoA (MalCoA) by acetyl-CoA carboxylase. Using AcCoA as a primer, the FA synthase multienzyme complex carries out a series of reactions that elongate the growing FA by two carbons. In this process MalCoA condenses with AcCoA, yielding an enzyme-bound 4-carbon β-ketoacid that is reduced, dehydrated, and reduced again. The product is enzyme-bound 4:0. This process is repeated six more times, after which 16:0 is released from the complex. The reductive steps require NADPH, which is derived from enzyme reactions and pathways shown in gray.

Table 1 Distinctions between fatty acid β-oxidation and fatty acid synthesis.

	Fatty acid β-oxidation	*Fatty acid synthesis*
Tissues with active pathway	Nearly all tissues except brain, nerve, and erythrocytes	Liver, adipose, lactating mammary gland
Subcellular location	Mitochondria	Cytoplasm
Redox cofactors	NAD, FAD	NADPH
Acyl group carrier	CoA	Enzyme-bound acyl carrier protein
Stereochemistry of 3-hydroxy intermediate	L-	D-

The carbon utilized for FA synthesis typically derives from the products of glycolysis (Fig. 6). Pyruvate enters the mitochondrion and becomes the substrate for two separate reactions. In one, pyruvate is decarboxylated *via* the pyruvate dehydrogenase complex, yielding acetyl-CoA. Lipogenic tissue mitochondria also contain pyruvate carboxylase, which converts pyruvate to the 4-carbon acid, oxaloacetate. Acetyl-CoA and oxaloacetate condense to form the 6-carbon acid, citrate. As citrate accumulates within the mitochondrion, it is exported via a transporter to the cytoplasm, where it is converted back to oxaloacetate plus acetyl-CoA in a reaction catalyzed by ATP citrate lyase (**ACLY**). Cytoplasmic acetyl-CoA is the fundamental building block for *de novo* synthesis of FAs.

The first enzyme unique to FA synthesis is acetyl-CoA carboxylase (**ACC1**), which converts the 2-carbon substrate, acetyl-CoA, into the 3-carbon product, malonyl-CoA. Citrate, in addition to being the precursor of cytoplasmic acetyl-CoA, also has a regulatory role. Citrate is an allosteric activator of acetyl-CoA carboxylase and serves as a signal that there is an ample carbon supply for FA synthesis. As mentioned, malonyl-CoA is a potent inhibitor of **CPT1**. Cytoplasmic malonyl-CoA levels are high when there is significant flux through glycolysis, indicative of a high cellular energy state. Under these conditions, entry of FAs into the mitochondrion (and subsequent β-oxidation) is prevented. Interestingly, there are two isoforms of acetyl-CoA carboxylase. **ACC1** is found in the above-named lipogenic tissues. **ACC2** is found in many tissues that are not capable of synthesizing FAs, for example, heart. It is thought that the primary role of the second isozyme is to regulate mitochondrial FA β-oxidation by synthesizing malonyl-CoA when cellular energy needs are being met by carbohydrate metabolism.

The subsequent reactions of cytoplasmic FA synthesis in humans are catalyzed by a multienzyme complex, fatty acid synthase (**FASN**). After binding of one molecule each of acetyl-CoA and malonyl-CoA to unique binding sites, a condensation reaction occurs in which CO_2 is released, and an enzyme-bound 4-carbon3-ketoacid is formed. Subsequent reactions include reduction, dehydration, and a second reduction. The intermediates produced in these reactions are similar to those seen in β-oxidation (Fig. 2), in reverse order. The product (enzyme-bound) is the saturated FA 4:0, which can then condense with another molecule of malonyl-CoA to start the process anew. After seven such cycles, the ultimate product, 16:0, is released from the complex.

The reductive steps in FA synthesis require NADPH, derived from two sources. Oxaloacetate produced by the **ACLY** reaction is converted to malate (*via* cytoplasmic malate dehydrogenase). Malic enzyme (**ME1**) then catalyzes the decarboxylation of malate to pyruvate, reducing $NADP^+$ to NADPH in the process. Two reactions in the pentose phosphate pathway, glucose-6-phosphate dehydrogenase (G6PD) and 6-phosphogluconate dehydrogenase (**PGD**), also yield NADPH.

Mitochondrial fatty acid synthesis

While the function of the cytoplasmic pathway for *de novo* synthesis of FAs is well-defined, the presence of a similar pathway in mitochondria remains somewhat enigmatic. Typically, metabolic pathways for the synthesis and degradation of a compound—in this case, FA—are not found in the same subcellular compartment. Nevertheless, there is ample evidence for a mitochondrial FA synthesis pathway (reviewed in Kastaniotis et al., 2017). Acetyl-CoA is abundant in mitochondria, but no acetyl-CoA carboxylase enzyme has been found in this organelle in humans. It has been speculated that mitochondrial propionyl-CoA carboxylase (**PCC**) can also accept acetyl-CoA as substrate, yielding malonyl-CoA. Alternatively, malonate entering mitochondria can be activated to malonyl-CoA via the acyl-CoA synthetase **ACSF3**.

Unlike cytoplasmic FA synthesis, most enzymes of the mitochondrial pathway are individual proteins rather than a multienzyme complex. The growing acyl chain is tethered to acyl carrier protein (ACP). A molecule of malonyl-CoA condenses with a primer acetyl-CoA via 3-ketoacyl-ACP synthase (**OXSM**), forming the 4-carbon 3-ketoacid. Subsequent reduction, dehydration, and second reduction steps are catalyzed by oxoacyl-ACP reductase (**CBR4**), 3-oxoacyl-ACP reductase (**HDT2**), and 2-Enoyl-ACP reductase (**MECR**), respectively. The resulting 4-carbon acyl-ACP can serve as primer for additional cycles to increase chain length. Both reductase enzymes utilize NADPH.

Research over several decades has uncovered some, but likely not all, functions for the mitochondrial FA synthesis pathway. One is the synthesis of lipoic acid, derived from the 8-carbon FA octanoic acid. Lipoic acid is a critical cofactor for carboxylase enzymes such as **ACC1**, **ACC2**, and **PCC**. Other functions include RNA processing, enzyme activity regulation, and control of mitochondrial gene expression.

Fatty acid elongation

16:0 is the primary product synthesized by the *de novo* cytoplasmic FA synthesis pathway. Although 16:0 is an important FA, there is need to synthesize longer chain-length acids. Enzymes for elongation of FA have been found in endoplasmic reticulum membranes (reviewed in Jakobsson et al., 2006). Reactions involved in FA elongation are very similar to those of cytoplasmic FA synthesis. The

donor of the added carbon atoms is also malonyl-CoA, indicating that an active acetyl-CoA carboxylase is required for elongation. Whereas the four primary reactions of FA synthesis are found within the **FASN** multienzyme complex, individual proteins catalyze these reactions in FA elongation. Like FA synthesis, both reduction steps in FA elongation require NADPH.

The condensation reaction is thought to be rate-limiting, and is catalyzed by a family of very long-chain elongase enzymes (**ELOVL1-7**). **ELOVL1, 3, 6**, and **7** have been associated with elongation of saturated and monounsaturated FAs, whereas **ELOVL2, 4**, and **5** seem to prefer polyunsaturated FAs. The second reaction is catalyzed by 3-ketoacyl-CoA reductase (**HSD17B12**). No specific enzyme has been associated with the third (dehydratase) reaction. The last step is catalyzed by trans-2,3-enoyl-CoA reductase (**TECR**).

A minor pathway (in eukaryotes) for FA elongation is found in mitochondria. The donor of elongation units is acetyl-CoA, not malonyl-CoA, and the reduction reactions utilize NADH rather than NADPH. There is little knowledge of how FA elongation in either endoplasmic reticulum or mitochondria is regulated. However, the presence of different ELOVL isoforms suggests that each might direct its elongation product toward a specific metabolic fate.

Fatty acid unsaturation and the essential FAs

Monounsaturated and polyunsaturated FAs are extraordinarily important in human health and nutrition. Thus, the ability to insert double bonds into the carbon skeleton of an FA is a vital metabolic function (reviewed in Lee et al., 2016). However, humans generally cannot insert a double bond closer than nine carbons from the methyl end of FAs. Thus, we are incapable of the *de novo* synthesis of two important classes of FAs, the ω3 FAs such as docosahexaenoic acid (22:6ω3) and the ω6 FAs such as arachidonic acid (20:4ω6). The ω3 FAs have proven health benefits, for example, in the prevention of coronary artery disease. 22:6ω3 has been shown to be important for normal development of brain and retina, leading some manufacturers to include this FA into their infant formula preparations. The ω6 FAs are important constituents of membrane lipids. 20:4 is also the precursor of prostaglandins and other bioactive eicosanoids. Because humans cannot synthesize these FAs *de novo*, we are dependent on the presence of at least some ω3 and some ω6 FAs in the diet. 18:2ω6 (linoleic acid) and 18:3ω3 (α-linolenic acid) are the precursors of most biologically important ω3 and ω6 FAs; thus, they are referred to as essential FAs.

One of the most abundant FAs in humans is 18:1ω9 (oleic acid), produced by inserting a *cis*-double bond in 18:0 (stearic acid). The enzyme catalyzing this reaction, stearoyl-CoA desaturase (**SCD1**) (reviewed in Piccinin et al., 2019) is called a Δ9 desaturase because it inserts the double bond 9 carbons from the *carboxyl* carbon. Because oleic acid is so abundant, the importance of **SCD1** in metabolism was initially overlooked. Oleic acid produced by **SCD1** appears to be directed specifically toward triacylglycerol synthesis. **SCD1** knockout mice have decreased adiposity. Furthermore, genetically obese, leptin-deficient (ob-/ob-) mice in which the **SCD1** gene was also disrupted had significantly reduced body weight than did ob-/ob-mice, leading to the hypothesis that leptin regulates the synthesis of **SCD1**. Interestingly, dietary oleate seems to be more readily incorporated into lipids other than triacylglycerols, implying that the dietary pool and the **SCD1**-produced pool of this FA are metabolically distinct. Like the ω3 FAs, dietary ingestion of monounsaturated FAs like 18:1ω9 have been associated with benefits to cardiovascular health. Humans and nonhuman primates also express **SCD5**, an isozyme found primarily in brain and pancreas, whose physiological role may be distinct from that of **SCD1**. Potential functions proposed for **SCD5** include regulation of neuronal differentiation, proliferation, and survival.

Humans are also capable of inserting *cis*-double bonds either five or six carbons from the carboxyl carbon of a FA (Δ5 desaturase and Δ6 desaturase activity, respectively). These activities, when combined with FA elongation pathways, form a powerful mechanism for synthesis of highly polyunsaturated FAs such as 20:4ω6 or 22:6ω3 from dietary essential FAs. The conversion of 18:3ω3 to 22:6ω3 also theoretically requires a Δ4 desaturase to insert the 6th double bond (in 22:5ω3); however, humans lack this enzyme. It is now believed that 22:5ω3 is elongated to 24:5ω3, converted to 24:6ω3 by Δ6 desaturase, and finally chain-shortened to 22:6ω3 by one cycle of peroxisomal β-oxidation.

Fatty acids as components of complex lipids

FAs are important building blocks for various cellular complex lipids (**Fig. 7**) (reviewed in Nguyen et al., 2008). For simplicity, the pathways for incorporation of FAs into these lipids are outlined only briefly. In most cases, fatty acyl-CoA and not free FA participates in these biosynthetic reactions. Nearly all cells synthesize phospholipids, which are essential membrane constituents. Phospholipid synthesis takes place in the endoplasmic reticulum. It begins by fatty acylating the two free hydroxyl groups in α-glycerophosphate, a triose derived from glycolytic intermediates, yielding phosphatidic acid. Various head groups (e.g., choline, ethanolamine, inositol, or serine) can then be linked to the phosphate group. For synthesis of triacylglycerol, this phosphate moiety is removed, yielding diacylglycerol, and a third fatty acyl group is esterified to the free hydroxyl group. Synthesis of plasmalogens (alkyl-acyl phospholipids), which comprise approximately 20% of membrane phospholipids, requires enzymes present in both peroxisomes and the endoplasmic reticulum. Plasmalogens are thought to be part of the cellular defense mechanism against oxidative injury.

Cholesteryl esters (ChE), in which an FA is esterified to the 3-hydroxyl group of cholesterol, are a transport and storage form of cholesterol. ChE are found in high concentrations in plasma low-density lipoproteins. Intracellular lipid droplets containing ChE

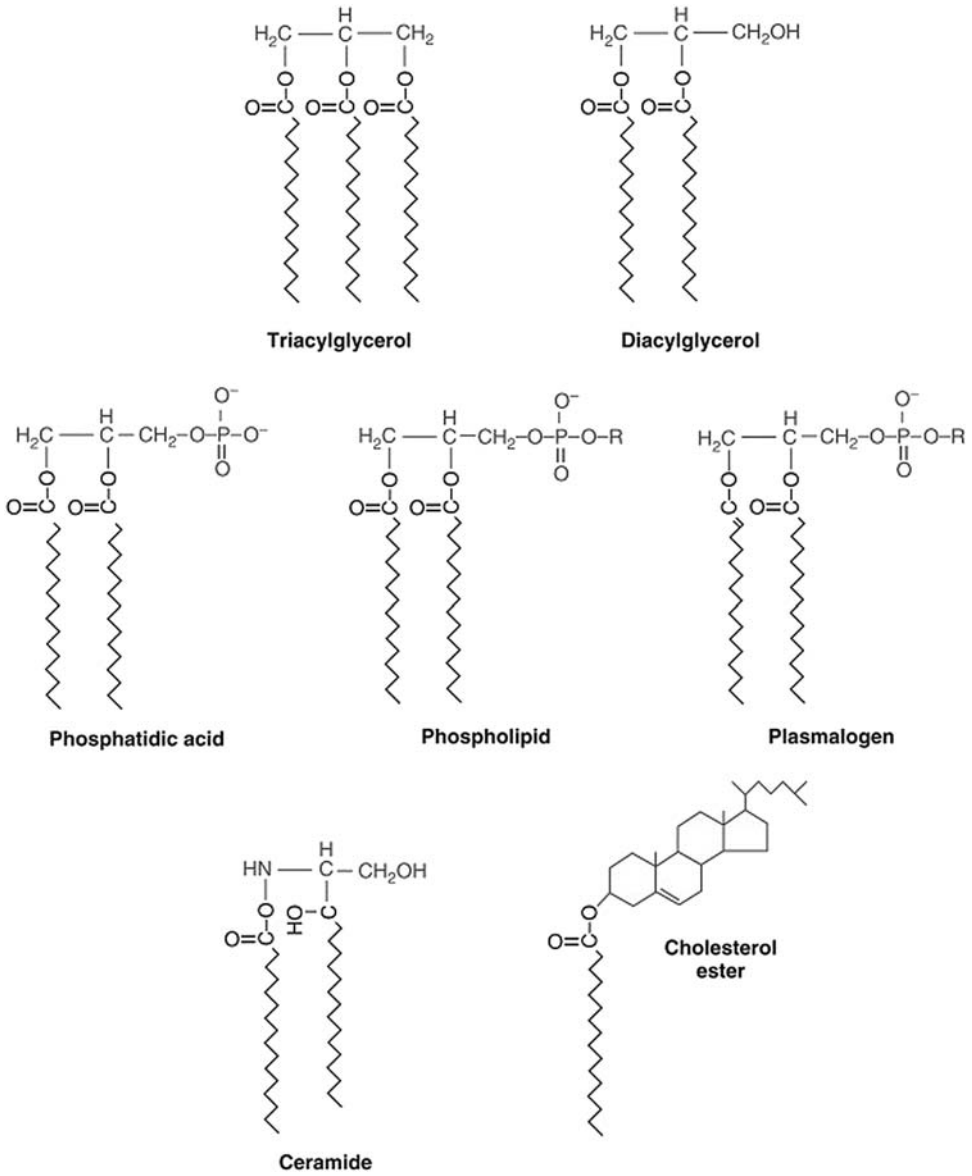


Fig. 7 Fatty acids as components of complex lipids. FAs form the basis of most complex lipids. The part of the molecule derived from FAs is black, and the part derived from other sources is gray. For phospholipids and plasmalogens, R = choline, ethanolamine, inositol, serine, or similar head group.

are found in steroidogenic tissues and are a reservoir of cholesterol for steroid hormone synthesis. ChE most commonly contain 18:1ω9, which is activated to its CoA derivative before transfer to cholesterol by the enzyme acyl-CoA:cholesterol acyltransferase (**SOAT1**). ChE are also formed within lipoproteins by the transfer of a fatty acyl chain from phosphatidylcholine to cholesterol in a reaction catalyzed by lecithin:cholesterol acyltransferase (**LCAT**).

Synthesis of sphingolipids, which include sphingomyelin, ceramides, cerebrosides, and gangliosides, begins by the condensation of palmitoyl-CoA (16:0-CoA) with serine. The amino group of serine is then acylated by a second fatty acyl-CoA to form ceramide; the chain length of the second FA can be variable. Transfer of phosphorylcholine (from phosphatidylcholine) to the hydroxyl group of ceramide yields sphingomyelin. Alternatively, sugars (from sugar nucleotide donors) are added to produce the cerebrosides, gangliosides, and related lipids.

Oxylipin synthesis

Highly unsaturated FA from both the ω6 and ω3 series can be converted to potent signaling molecules collectively referred to as oxylipins (reviewed in Christie and Harwood, 2020). As noted previously, humans are incapable of synthesizing *de novo* FAs

with double bonds in the ω6 and ω3 positions. Thus, oxylipins are derived from dietary lipids or are synthesized by elongation and unsaturation of essential FAs. The FA 20:4ω6 (arachidonic acid; AA) is the precursor of most classic eicosanoids, which include the prostaglandins, leukotrienes, and thromboxanes. Like other FAs, cellular concentrations of unesterified AA are low. Conversion of AA to eicosanoids begins with an agonist-induced release of the FA from the sn-2 position of membrane phospholipids *via* the action of phospholipase A2. Unlike most reactions of FAs, free AA rather than its CoA derivative appears to be the substrate for eicosanoid synthesis.

Cyclooxygenases (**COX1** and **COX2**) catalyze a complex, molecular O_2-requiring reaction that convert AA to prostaglandin G2. This reaction involves carbon atoms in the middle of the acyl chain, rather than at the methyl carbon (such as occurs in ω-oxidation) or the carboxyl carbon (such as occurs in nearly all other reactions of FAs). Prostaglandin G2 can subsequently be converted to other prostaglandins or to thromboxanes. As these compounds have potent biological effects, including mediation of inflammation, COX inhibitors form an important class of anti-inflammatory drugs. Free AA is also the primary substrate for the enzyme 5-lipoxygenase, which is the first step in the synthesis of leukotrienes, and for several cytochrome P450 oxidases, which lead to synthesis of hydroxyeicosatetraenoic acids (HETEs) and epoxyeicosatrienoic acids (EETs).

Important oxylipins are also derived from the highly unsaturated ω3 FAs, 20:5ω3 (eicosapentaenoic acid; EPA) and 22:6ω3 (docosapentaenoic acid; DHA). Many of these oxylipins are involved in the resolution of inflammation, and have been collectively referred to as specialized pro-resolving mediators (SPM). Enzymes that act on AA can also act on EPA and DHA, including cyclooxygenases, lipoxygenases, and cytochrome P450 oxidases. Products of these reactions and downstream pathways include bioactive lipids such as protectins, resolvins, and maresins.

Endocannabinoids

Research that identified receptors for the psychoactive ingredient in cannabis, Δ^1 -tetrahydrocannabinol, led to the discovery of endogenous ligands of these receptors, or endocannabinoids (reviewed in Joshi and Onaivi, 2019; Cristino et al. 2020). Receptors are located not only in the brain, but throughout the body. The endocannabinoid system serves important roles in human reproduction and newborn growth, among other things. The endocannabinoids are FA-derived molecules, primarily anandamide (N-arachadonoylethanolamine; AEA) and 2-arachidonoylglycerol (2-AG). The main pathway for AEA synthesis in brain begins with the action of an N-acyltransferase that links the carboxyl carbon of AA (released from membrane phospholipids) to the amino group of phosphatidylethanolamine, producing N-arachidonoyl phosphatidylethanolamine (NAPE). A NAPE-specific phospholipase D (**NAPEPLD**) cleaves this molecule, yielding AEA and phosphatidic acid. There are additional AEA synthesis pathways in the brain and other tissues. Degradation of AEA is rapid, and is catalyzed by fatty acid amide hydrolase (**FAAH**), which cleaves the carbon-nitrogen bond between AA and ethanolamine.

Phosphatidylinositol typically contain AA at the sn-2 position. Synthesis of 2-AG begins with the cleavage of phosphatidylinositol, producing diacylglycerol and phosphoinositol, by the enzyme phospholipase C (**PLCG1**). A specific diacylglycerol lipase (**DAGLA/B**) releases the FA (typically a saturated FA) from the sn-1 position, yielding 2-AG. Like AEA, 2-AG is rapidly inactivated *in vivo*. Monoacylglycerol lipase (**MGLL**) and other lipases cleave 2-AG, releasing AA from the glycerol backbone.

Fatty acylation of proteins

Covalent modification of proteins is a more recently discovered role of FAs (reviewed in Resh, 2016). Fatty acylation of proteins frequently serves as a means of targeting or anchoring a protein to a membrane. Before attachment, all FAs must be activated to their CoA derivatives. Myristoylation, the addition of 14:0 to a protein, occurs at *N*-terminal glycine residues after removal of the initiator methionine. The consensus sequence for this modification is H_2N-Met-Gly-X-X-X-Ser/Thr. The reaction, catalyzed by *N*-myristoyltransferase (**NMT1**, **NMT2**), is co-translational and irreversible. *N*-myristoyl-proteins include many signal transduction-associated proteins, for example, *src* and ADP-ribosylation factors (Arfs).

Palmitoylation, the addition of 16:0 to a protein, is also commonly observed. This modification occurs post-translationally and is reversible. Most often, palmitic acid is bound to the sulfhydryl side chain of cysteine residues. Both membrane-associated proteins and integral membrane proteins can be palmitoylated; examples include ion channels, caveolin, neurotransmitter receptors, wnt, and sonic hedgehog. Palmitoylation is catalyzed by members of the large family of palmitoyltransferases (**ZDHHC1-22**). Three protein palmitoylthioesterases (PPT1, **PPT2**, and **LYPLA1**) that catalyze depalmitoylation of proteins have been identified. Several proteins are modified with both an *N*-terminal 14:0 and an *S*-linked 16:0 elsewhere in the protein chain. Alpha-subunits of heterotrimeric G-proteins and endothelial nitric oxide synthase are examples of myristoylated/palmitoylated proteins. In addition to palmitoylation, the signaling protein wnt is modified on a serine residue by the monounsaturated FA, palmitoleic acid (C16:1ω7), which is produced by the action of **SCD1** on palmitoyl-CoA.

There are instances of acylation by FAs with chain length other than 14 or 16 carbons. Proteins modified by C8:0, C18:0, C18:1, and C20:4 have been reported. One nutritionally important example is the recently identified orexigenic peptide, ghrelin. The active form of this 28-amino acid peptide hormone has the medium-chain FA, C8:0, covalently esterified to the hydroxyl group of serine-3. Octanoylated ghrelin is believed to act at the level of the hypothalamus to stimulate appetite, perhaps *via* neuropeptide Y.

Table 2 Vitamins associated with fatty acid metabolism.

Vitamin	*Active form*	*Enzymes*	*Pathways*
Pantothenic acid	CoA	Many enzymes	Most reactions involving FAs
Niacin	NAD, NADH, NADP, NADPH	Dehydrogenases; reductases	Many pathways, particularly β-oxidation and FA synthesis and elongation
Riboflavin	FAD, $FADH_2$	Oxidases	β-Oxidation
Thiamine	Thiamine pyrophosphate	Pyruvate dehydrogenase complex; α-hydroxyphytanoyl-CoA lyase	FA synthesis from glucose; phytanic acid α-oxidation
Biotin	Biocytin	Acetyl-CoA carboxylase; pyruvate carboxylase	FA synthesis from glucose

Vitamins and fatty acid metabolism

Several of the B-vitamins are essential for normal FA metabolism (Table 2). Pantothenic acid is a constituent of CoA, and is thus required for numerous reactions of FAs. Niacin and riboflavin are necessary for the synthesis of oxidized and reduced NAD(P) and FAD, respectively. These compounds play essential roles in FA oxidation, synthesis, and elongation. Biotin is a constituent of acetyl-CoA carboxylase and pyruvate carboxylase, and thiamine is required for activity of the pyruvate dehydrogenase complex; all three enzymes are involved in the synthesis of FAs from glucose.

Regulation of fatty acid metabolism

Regulation of FA synthesis (*see* **Fatty acid *de novo* synthesis**) and β-oxidation (*see* **Mitochondrial fatty acid β-oxidation**) have already been discussed. More global regulatory mechanisms that deserve a brief mention include those mediated by insulin/glucagon, sterol regulatory element binding protein (SREBP)1c, and peroxisome proliferator-activated receptors (PPARs) (reviewed in Nguyen et al., 2008; Nakamura et al., 2014; Rui, 2014). In the fed and fasted states, control of fuel metabolism is mediated to a large extent by insulin and glucagon, respectively. Effects of glucagon are mediated *via* cAMP-dependent kinases. During fasting, increased blood glucagon levels promote release of free FAs from adipocyte triacylglycerol stores, In the liver and other tissues, glucagon promotes decreased flux through glycolysis, thereby decreasing the rate of *de novo* FA biosynthesis and increasing rates of mitochondrial β-oxidation and ketogenesis. In the fed state, insulin levels rise in response to increased blood glucose. Insulin's effects are mediated through activation of its receptor tyrosine kinase and are in general opposite to those of glucagon, stimulating glycolysis and FA synthesis while inhibiting FA degradation. Insulin and glucagon have both acute and long-term effects on FA metabolism.

SREBP1c (SREBF1) is a transcription factor thought to mediate the action of insulin in upregulating genes involved in FA synthesis such as **ACC1** and **FASN**. PPARγ (**PPARG**) is a nuclear hormone receptor that, on activation by an as yet unknown ligand, activates genes involved in adipocyte differentiation and lipid storage. It has been hypothesized that activation of SREBP1c may contribute to generation of an endogenous PPARγ ligand.

Activation of PPARα (**PPARA**) on the other hand increases rates of FA oxidation and ketogenesis. Endogenous ligands for this nuclear receptor are thought to include polyunsaturated FAs and branched-chain FAs. Both PPARs heterodimerize with the retinoid-X-receptor (RXR), and both receptors must be ligand-bound for transcriptional activation. Several mitochondrial and microsomal, as well as peroxisomal, genes associated with FA catabolism are upregulated *via* PPARα stimulation.

Summary

This article discusses the primary metabolic pathways in which fatty acids participate. The prerequisite for activation of fatty acids prior to metabolism is described. Multiple metabolic pathways for degradation and synthesis of fatty acids are explained. Fatty acid elongation and desaturation are described, as well as the incorporation of these molecules into complex lipids. The metabolic conversion of fatty acids into signaling molecules, as well as the participation of fatty acids in endocannabinoid synthesis, is discussed. Fatty acylation of proteins is mentioned. The article concludes with a brief synopsis of the regulation of fatty acid metabolism.

References

Christie, W.W., Harwood, J.L., 2020. Oxidation of polyunsaturated fatty acids to produce lipid mediators. Essays Biochem. 64, 401–421.
Cristino, L., Bisogno, T., Di Marzo, V., 2020. Cannabinoids and the expanded endocannabinoid system in neurological disorders. Nat. Rev. Neurol. 16, 9–29.

Houten, S.M., Violante, S., Ventura, F.V., Wanders, R.J., 2016. The biochemistry and physiology of mitochondrial fatty acid β-oxidation and its genetic disorders. Annu. Rev. Physiol. 78, 23–44.
Jakobsson, A., Westerberg, R., Jacobsson, A., 2006. Fatty acid elongases in mammals: their regulation and roles in metabolism. Prog. Lipid Res. 45, 237–249.
Joshi, N., Onaivi, E.S., 2019. Endocannabinoid system components: overview and tissue distribution. Adv. Exp. Med. Biol. 1162, 1–12.
Kastaniotis, A.J., Autio, K.J., Kerätär, J.M., Monteuuis, G., Mäkelä, A.M., Nair, R.R., Pietikäinen, L.P., Shvetsova, A., Chen, Z., Hiltunen, J.K., 2017. Mitochondrial fatty acid synthesis, fatty acids and mitochondrial physiology. Biochim. Biophys. Acta Mol. Cell Biol. Lipids 1862, 39–48.
Lee, J.M., Lee, H., Kang, S., Park, W.J., 2016. Fatty acid desaturases, polyunsaturated fatty acid regulation, and biotechnological advances. Nutrients 8, 23.
Miura, Y., 2013. The biological significance of ω-oxidation of fatty acids. Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. 89, 370–382.
Miura, M., Nonomura, Y., 2019. Evaluation of hydrophobicity for fatty acids using reversed-phase thin layer chromatography. J. Oleo Sci. 68, 665–670.
Nakamura, M.T., Yudell, B.E., Loor, J.J., 2014. Regulation of energy metabolism by long-chain fatty acids. Prog. Lipid Res. 53, 124–144.
Nguyen, P., Leray, V., Diez, M., Serisier, S., Le Bloc'h, J., Siliart, B., Dumon, H., 2008. Liver lipid metabolism. J. Anim. Physiol. Anim. Nutr. 92, 272–283.
Papamandjaris, A.A., MacDougall, D.E., Jones, P.J., 1998. Medium chain fatty acid metabolism and energy expenditure: obesity treatment implications. Life Sci. 62, 1203–1215.
Piccinin, E., Cariello, M., De Santis, S., Ducheix, S., Sabbà, C., Ntambi, J.M., Moschetta, A., 2019. Role of oleic acid in the gut-liver axis: from diet to the regulation of its synthesis via stearoyl-CoA desaturase 1 (SCD1). Nutrients 11, 2283.
Resh, M.D., 2016. Fatty acylation of proteins: the long and the short of it. Prog. Lipid Res. 63, 120–131.
Rui, L., 2014. Energy metabolism in the liver. Compr. Physiol. 4, 177–197.
Schwenk, R.W., Holloway, G.P., Luiken, J.J., Bonen, A., Glatz, J.F., 2010. Fatty acid transport across the cell membrane: regulation by fatty acid transporters. Prostaglandins Leukot. Essent. Fatty Acids 82, 149–154.
Wanders, R.J.A., 2014. Metabolic functions of peroxisomes in health and disease. Biochimie 98, 36–44.
Wanders, R.J.A., Vaz, F.M., Waterham, H.R., Ferdinandusse, S., 2020. Fatty acid oxidation in peroxisomes: enzymology, metabolic crosstalk with other organelles and peroxisomal disorders. Adv. Exp. Med. Biol. 1299, 55–70.
Watkins, P.A., 1997. Fatty acid activation. Prog. Lipid Res. 36, 55–83.
Xu, H., Diolintzi, A., Storch, J., 2019. Fatty acid-binding proteins: functional understanding and diagnostic implications. Curr. Opin. Clin. Nutr. Metab. Care 22, 407–412.

Further reading

Frohnert, B.I., Bernlohr, D.A., 2000. Regulation of fatty acid transporters in mammalian cells. Prog. Lipid Res. 39, 83–107.
Gibbons, G.F., 2003. Regulation of fatty acid and cholesterol synthesis: co-operation or competition? Prog. Lipid Res. 42, 479–497.
Gunstone, F.D., Harwood, J.L., Padley, F.B. (Eds.), 1994. The Lipid Handbook, second ed. Chapman & Hall, London.
Jakobsson, A., Westerberg, R., Jacobsson, A., 2006. Fatty acid elongases in mammals: their regulation and roles in metabolism. Prog. Lipid Res. 45, 237–249.
Kunau, W.H., Dommes, V., Schulz, H., 1995. Beta-oxidation of fatty acids in mitochondria, peroxisomes, and bacteria: a century of continued progress. Prog. Lipid Res. 34, 267–342.
McGarry, J.D., Foster, D.W., 1980. Regulation of hepatic fatty acid oxidation and ketone body production. Annu. Rev. Biochem. 49, 395–420.
Numa, S. (Ed.), 1984. Fatty Acid Metabolism and Its Regulation. Elsevier, NY.
Vance, D.E., Vance, J.E. (Eds.), 2002, Biochemistry of Lipids, Lipoproteins and Membranes, fourth ed., vol. 36. Elsevier, NY.
Wanders, R.J., Vreken, P., Ferdinandusse, S., et al., 2001. Peroxisomal fatty acid alpha- and beta-oxidation in humans: enzymology, peroxisomal metabolite transporters and peroxisomal diseases. Biochem. Soc. Trans. 29, 250–267.
Wanders, R.J., Ruiter, J.P., IJLst, L., Waterham, H.R., Houten, S.M., 2010. The enzymology of mitochondrial fatty acid beta-oxidation and its application to follow-up analysis of positive neonatal screening results. J. Inherit. Metab. Dis. 33, 479–494.
Watkins, P.A., 1997. Fatty acid activation. Prog. Lipid Res. 36, 55–83.

Food composition data 2021

F. Grande[a] **and U.R. Charrondiere**[b], [a] Food and Agriculture Organization of the United Nations, Sao Paulo, Brazil; and [b] (Retired) Regional Office for Latin America and the Caribbean, Food and Agriculture Organization of the United Nations, Santiago, Chile

© 2023 Elsevier Ltd. All rights reserved.

This is an update of S.P. Murphy, Food Composition Data, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00117-3.

Introduction 330
What are food composition data and where to find them? 330
Coverage of foods 331
Food description, food group and food coding 331
Component identification 331
Coverage of components 332
What are the sources of data for food composition tables and databases? 333
What is the global availability of FCTs/FCDBs? 336
Selecting high-quality food composition data and databases 336
How to assure that food composition data, FCT and FCDB are generated, compiled and used correctly 337
Conclusions and way forward 337
Conflict of Interest 338
References 338
Relevant websites 338

Key points

- Provide the key aspects of high-quality food composition tables or databases
- Discuss the different sources of food composition data
- Present the pillars to assure that food composition data are generated, compiled and used correctly

Introduction

Food composition data (FCD) play a key role in most nutrition related activities and should play an increasing role in agriculture when implementing nutrition-sensitive policies and programs. Undernutrition and micronutrient deficiencies persist in most countries, while overweight, obesity and dietary related non-communicable diseases are increasing. Many countries concentrate their efforts on supplementation or fortification, despite increasing criticism of their sustainability and effectiveness. Foods contain not only nutrients, but many other bio-active compounds with positive health effects, which supplements and fortificants cannot deliver. Thus, food-based approaches are on the rise and, for this type on intervention, FCD are key.

Over the last 40 years, progress in the availability and quality of food composition data has been realized, and much of this was achieved through INFOODS, the International Network of Food Data Systems, established in 1984 and coordinated since 1999 by the Food and Agriculture Organization of the United Nations (FAO). INFOODS and FAO have published, free of charge, many important guidelines and standards, training materials, a food composition database management system (FAO/INFOODS Compilation tool), and food composition tables and databases (FCT/FCDB), and have provided training and technical support to many countries. Other organizations, such as the European Food Information Resource Network (EuroFIR), established in 2005, complement these efforts. However, much is still to be done.

What are food composition data and where to find them?

FCD provides information about the content of nutrients and other nutrition-relevant components in foods. These data are published in national, regional or global FCT/FCDB. They may comprise a range of components, such as edible portion, macronutrients, vitamins, minerals, bioactive components (e.g., flavonoids), as well as conversion factors and contributing components, to calculate the different nutrients and their fractions (e.g., pro-vitamin A carotenes and retinol for vitamin A calculation or different fatty acids to present saturated fatty acids). Excluded are contaminants and other food components of interest for food safety, called occurrence data.

FCD can vary significantly due to the natural or artificial differences in the nutrient content of foods. Natural differences are due to the environment, processing, biodiversity, storage, cooking methods, recipe or product formulation. Additionally, artificial differences may be introduced into data due to sampling design and collection, analytical methods and quality control used, calculation methods or definitions adopted for each food component.

The quality of FCTs/FCDBs are affected mainly by the sources of FCD such as quality, coverage and description of foods and components, and their documentation.

Coverage of foods

An FCT/FCDB of sound quality should include foods as consumed by the population of interest, including raw and cooked foods, mixed dishes, manufactured and biodiverse foods.

Many FCTs/FCDBs include mainly raw foods, with a limited number of mixed dishes, biodiverse and branded or fortified foods. The lack of fortified foods in FCTs/FCDBs will result in an underestimation of the micronutrient intake for the population, particularly if these foods are highly consumed. In all countries, voluntarily fortified food products exist, and many countries have mandatory fortification of core foods. For instance, the intake for Dietary Folate Equivalents (DFE) was estimated for different regions of Brazil, where fortification with folic acid is mandatory for wheat and corn flours. The total average intake for DFE was up to 35% lower when a FCDB was used which was not adjusted according to the fortification practices in Brazil (Grande et al., 2019). Currently, 85 countries have mandatory fortification of wheat flour, meaning that the inclusion of fortified foods in the respective FCT/FCDB is of great importance (Global Fortification Data Exchange, 2021). As a rule, the more reported foods are unavailable in FCT/FCDB and would have to be matched to a similar food, the lower the quality of the nutrient intake estimations. This is not only true for fortified foods, as shown in the example, but also for biodiverse or cooked foods or recipes.

Dietary supplements also add significant amounts of nutrients and other ingredients to the diet, as in the USA, where 81% of the population over 55 years report using them regularly (Statista, 2021). There are some databases on dietary supplements using the labeling information, e.g., USA, INFOODS, but it is difficult to keep them up-to-date, to evaluate their quality and to present the data in a similar way as FCTs/FCDBs, i.e., using the same component identifiers (e.g., INFOODS tagnames) or principles. Their contributions to nutrient intake are, therefore, difficult to include in food consumption surveys.

Food description, food group and food coding

The detail of food description has significant relevance for nutrient content and data quality. Foods not well described leave room for interpretation and errors in usage. The main descriptors that should be added to the food name are the cooking or preservation method, color, maturity stage, origin (e.g., type of animal, acquisition type or harvest season), part, edible/inedible portion, fortification or enrichment and biodiversity details (e.g., variety, breed or wild food). The more complete the food description the better, as nutrient content can differ greatly with different descriptions of the same food. For instance, if we find in a FCT a poor food description such as "meat" or "cassava" it is very difficult (or not feasible) to use the data, even if they were generated by high-quality analytical methods. The meat composition will depend on the type and part of the animal and cooking method. In the 2019 FAO/INFOODS FCT for Western Africa (Vincent et al., 2020), the protein content for "Beef meat, lean, ca. 5% fat, raw" is 22.5 g per 100 g edible portion on fresh weight basis (EP) while for "Beef meat, lean, ca. 5% fat, boiled (without salt), drained" it is 37.5 g per 100 g EP. For cassava, the part of the plant, cooking method and color affect the composition. In the same FCT, when we compare different plant parts, we can find that the dietary fiber content for "Cassava, tuber, white flesh, raw" is 3.7 g per 100 g EP while for "Cassava, leaves, fresh, raw" it is 7.9 g per 100 g EP. Moreover, the vitamin A content in retinol equivalents (RE) is very different due to tuber color: white raw cassava tuber contains 4 μg RE per 100 g EP while the yellow fleshed raw cassava contains 75 μg RE per 100 g EP. These examples demonstrate the importance of detailed food description. as the nutritional contents change drastically depending on their characteristics. An ambiguous food description is likely to introduce errors into food matching and therefore into nutrient intake estimations.

Some food description systems include the food grouping and/or food coding, such as the food classification and description system (FoodEx2) (European Food Safety Authority, 2015) or the LanguaL food classification system (Møller and Ireland, 2018). Food groups help to find the foods easily and to define certain quality checks at the food group level. Common food groups used are in **Table 1**.

In an FCT/FCDB, each food entry should have their own unique food code. Often, it is a combination of the food group and food entry within the food group. For example, in the FAO/INFOODS Food Composition Table for Western Africa (WAFCT 2019) (Vincent et al., 2020), "*Millet, whole grain, raw*" has the code 01_015, where 01 indicates the food group "*01 Cereals and their products*" followed by the number of the food record.

Component identification

A clear and complete component description is key for FDC and the quality of FCT/FCDB. A complete component description should cover both the component identification (e.g., expression, definition and analytical methods), units and denominator.

For some components, there are potential problems regarding component naming, due to differences in definitions and analytical methods. These "problematic components" are important for nutrition as they, unfortunately, may have different analytical

Table 1 Common food groups used in FCT/FCDB.

Food group name
Cereals and their products
Starchy roots, tubers and their products
Legumes and their products
Vegetables and their products
Fruits and their products
Nuts, seeds and their products
Meat, poultry and their products
Eggs and their products
Fish and their products
Milk and their products
Fats and oils
Beverages
Miscellaneous
Soups and sauces

methods, expressions or calculations, yielding significantly different values. Such components include energy, protein, fat, carbohydrates, fiber, most vitamins (Vitamin A, D, E, D, C, B6, folate, niacin/niacin equivalent) and phytate. For these components, the user should carefully investigate which analytical method or definition is used in an FCT/FCDB (or another source) before utilizing their nutrient values.

To provide unambiguous and language-independent component descriptions, INFOODS developed the INFOODS food component identifiers in 1989 (Klensin et al., 1989), also known as tagnames. A list of tagnames is available at the INFOODS website (http://www.fao.org/infoods/infoods/standards-guidelines/food-component-identifiers-tagnames/en/). These have been updated and enlarged over the last decade, as new components are analyzed, or because of newly developed methods. Their use in FCT/FCDB is of upmost importance, as nutrient values will change significantly by tagname. Different tagnames exist if analytical methods, expressions or definitions generate significantly different nutrient values. The "problematic components" always have several tagnames. Luckily, most components have one tagname, meaning that existing analytical methods generate similar values and only one definition/expression exists. Tagnames are like unique codes or abbreviations according to a specific naming system that allows clear definition of components developed using the following main principles:

- Each time an analytical method, expression or calculation generates significantly different values, a new tagname should be created for the specific component. This ensures that values assigned to the same tagname are directly comparable. Only data with the same tagname are comparable and can be combined. Data with different tagnames are not comparable and should not be combined.
- Each chemical used for nutrition that is analyzed should have a specific tagname.
- The food component tagname scheme should reflect "nutrients," not just chemistry.
- For unknown methods, or multiple empirical methods, the tagname with a hyphen is used (e.g., FIB-). This is to avoid the assignment of a specific tagname without knowing if it is the most appropriate one.

The tagname system must be robust and extensible to allow new tagnames to be added in the future, as information technology and nutrition science are rapidly developing fields. Some examples for different INFOODS tagnames, due to different definition or analytical method, are shown in **Table 2**.

Vitamin A may be calculated either as Retinol Equivalents (RE with tagname VITA) or as Retinol Activity Equivalents (RAE with tagname VITA_RAE). In plant foods, they result in different values for the same carotenoid content. Thus, using RE or RAE in estimating nutrient intake would generate two significantly different figures.

FCD can also be presented with different units, such as g, mg, μg or international units, and with different denominators. In general, nutrient values in FCT/FCDB are expressed per 100 g edible portion on fresh weight basis (EP). Some use per 100 mL for beverages, per g of nitrogen for amino acids; or fatty acids may be expressed per 100 g EP or as a percentage of total fatty acids. Scientific articles often express data per 100 g edible portion on a dry matter basis (DM), or per 100 g fatty acids. These differences have great implications on the absolute value of the data. The importance of unit and denominator on the absolute values is exemplified in **Table 3**.

Users of FDC need to understand the importance of component description, unit, and denominator to calculate nutrient intake correctly. It is also key to verify that component definitions are the same between FCD and recommended daily intakes (RDI) to avoid errors in nutrient adequacy estimations.

Coverage of components

FAO/INFOODS defined the minimum number of components to be included in FCT/FCDB as: Edible portion (Refuse), energy, water, protein, fat, available carbohydrates, dietary fiber, iron, zinc, calcium, vitamin A, vitamin C, vitamin B1, vitamin B2, vitamin B12, and folate. Many FCT/FCDB contain more than these components, while some FCT/FCDB do not include all of them.

Table 2 Different INFOODS food component identifiers (tagnames) for carbohydrates, fiber and vitamin A.

Component	*Analytical/determination method or definition*	*Tagname*	*Description*
Carbohydrates	Total carbohydrates (includes fiber)	CHOCDF	Calculated by difference = 100 − (water (g) + protein (g) + fat (g) + alcohol (g) + ash (g))
		CHOCSM	By summation, this value is the sum of analytical values of sugars, starch, oligosaccharides and dietary fiber
	Available carbohydrates (excludes fiber)	CHOAVLDF	Calculated by difference = 100 − (water (g) + protein (g) + fat (g) + alcohol (g) + ash (g) + (dietary) fiber))
		CHOAVL	By weight, this value is the sum of analytical values (anhydrous form) of sugars, starch and glycogen
		CHOAVLM	Monosaccharide equivalents, this value is the sum of analytical values (hydrous form) of sugars, starch and glycogen
Fiber	Total dietary fiber	FIBTG	Analyzed by AOAC Prosky method. It is a mixture of non-starch polysaccharides, lignin, resistant starch, resistant oligosaccharides, gum and waxes (most recommended definition)
	Non-starch polysaccharide	PSACNS/NSP	Also called Englyst fiber. This includes non-starch polysaccharides but excludes lignin, resistant starch and resistant oligosaccharides
	Southgate fiber	FIBTS	Mixture of non-starch polysaccharides, lignin and some resistant starch
	Fiber, determined by acid detergent method	FIBAD	This includes lignin, cellulose, some hemicellulose and some pectin
	Fiber, determined by neutral detergent method	FIBND	This includes lignin, cellulose, and insoluble hemicellulose
	Crude fiber	FIBC	It captures only fractions of lignin, cellulose and hemicellulose (obsolete and should not be used)
	Fiber, unknown or mixed methods	FIB-	
Vitamin A	Vitamin A, expressed in retinol equivalents (RE)	VITA	VITA = retinol (μg) + 1/6 beta-carotene (μg) + 1/12 other pro-vitamin A carorenoids (μg)
	Vitamin A, expressed in retinol activity equivalents (RAE)	VITA_RAE	VITA_RAE = retinol (μg) + 1/12 beta-carotene (μg) + 1/24 other pro-vitamin A carorenoids (μg)
	Vitamin A, determined by bioassay	VITAA	No longer used in food composition tables and databases

Adapted from FAO/INFOODS e-Learning Course on Food Composition Data (FAO, 2013).

Table 3 Examples of how different units and denominators can affect the absolute fat values in cream cheese.

Fat value	*Unit*	*Denominator*
34.2	g	100 g edible portion on fresh weight basis (EP)
34.2	%	100 g EP
342	g	1 kg EP
34,200	mg	100 g EP
75	g	100 g edible portion on a dry matter basis (water content = 54.4 g/100 g EP)
75	%	Of dry matter (34.2 g fat/45.6 g DM)
75	%	Fat in dry matter

Adapted from FAO/INFOODS e-Learning Course on Food Composition Data (FAO, 2013).

An example of how FCT/FCDB should identify components was taken from the FAO/INFOODS Food Composition Table for Western Africa (WAFCT 2019) (**Table 4**). In addition to such a table with the components, the FCT/FCDB should include detailed description of the nutrients, their calculation methods and details on the analytical methods.

What are the sources of data for food composition tables and databases?

Analytical data should be the basis for FCT/FCDB. They should be generated using appropriate methods and quality standards published in scientific articles or laboratory reports. FCD generated through direct analysis, in which the values are a result of analysis carried out specifically for a certain FCT/FCDB are considered the highest quality data. In this approach, it is possible to collect representative samples across a country or region so that they accurately reflect the food supply. As this is costly, it is rarely done for most foods.

In general, FCD published in the scientific literature contain limited analytical data, with a short description of the analytical methods and quality control schemes and, sometimes, with sampling information. Many analyzed FCD, however, never get

Table 4 Components, units, analytical/determination method/definition and corresponding INFOODS component identifier listed in the FAO/INFOODS Food Composition Table for Western Africa (WAFCT 2019), expressed per 100 g edible portion.

Component	*Unit*	*Analytical/determination method/definition*	*INFOODS component identifier (tagname)*
Edible portion 1	–	Edible portion coefficient: from as purchased to as described	EDIBLE1
Edible portion 2	–	Edible portion coefficient: from as described to as eaten	EDIBLE2
Sum of proximates (SOP)	G	=Water + protein + fat + available carbohydrates + ash + dietary fiber + alcohol	SOP
Energy	kJ, kcal	Calculated (see formula below)	ENERC
Water	g	Drying	WATER
Protein, total	g	Calculated by multiplying the nitrogen conversion factor (XN) times the analyzed total nitrogen (mainly Kjeldahl method)	PROTCNT
Fat, total or [fat by Soxhlet]	g	Mixed solvent extraction or [Soxhlet method with continuous extraction]	FAT of [FATCE]
Carbohydrates, available by difference	g	=100 − (water + protein + fat + ash + dietary fiber + alcohol)	CHOAVLDF
Fiber, total dietary or [crude fiber]	g	AOAC Prosky method (AOAC 991.43) or [Weende method]	FIBTG or [FIBC]
Alcohol	g	GLC or distillation	ALC
Ash	g	Gravimetric methods	ASH
Calcium	mg	AAS, ICP-MS	CA
Iron	mg	AAS, ICP-MS	FE
Magnesium	mg	AAS, ICP-MS	MG
Phosphorous	mg	AAS, ICP-MS	P
Potassium	mg	AAS, ICP-MS	K
Sodium	mg	AAS, ICP-MS	NA
Zinc	mg	AAS, ICP-MS	ZN
Copper	mg	AAS, ICP-MS	CU
Vitamin A (expressed in retinol equivalents)	µg	=Retinol + 1/6 beta-carotene equivalent	VITA
Vitamin A (expressed in retinol activity equivalents)	µg	=Retinol + 1/12 beta-carotene equivalent	VITA_RAE
Retinol	µg	HPLC	RETOL
Beta-carotene equivalents or [beta-carotene]	µg	=Beta-carotene + 1/2 alpha-carotene + 1/2 beta-cryptoxanthin	CARTBEQ or [CARTB]
Alpha-carotene	µg	HPLC	CARTA
Beta-carotene	µg	HPLC	CARTB
Beta-cryptoxanthin	µg	HPLC	CRYPXB
Vitamin D	µg	=Vitamin D_2 + Vitamin D_3; analyzed by HPLC	VITD
Vitamin E (expressed in alpha-tocopherol equivalent) or [alpha-tocopherol]	mg	=Alpha-tocopherol + 0.4 × beta-tocopherol + 0.1 × gamma-tocopherol + 0.01 × delta-tocotrienol; analyzed by HPLC	VITE or [TOCPHA]
Alpha-tocopherol	mg	HPLC	TOCPHA
Beta-tocopherol	mg	HPLC	TOCPHB
Gamma-tocopherol	mg	HPLC	TOCPHG
Delta-tocopherol	mg	HPLC	TOCPHD
Thiamin (vitamin B1)	mg	HPLC, microbiological	THIA
Riboflavin (vitamin B2)	mg	HPLC, microbiological	RIBF
Niacin equivalents or [niacin] (vitamin B3)	mg	=Niacin + 1/60 tryptophan	NIAEQ or [NIA]
Niacin	mg	HPLC, microbiological	NIA
Tryptophan	mg	Ion exchange chromatography	TRP
Vitamin B6	mg	HPLC, microbiological	VITB6C
Folic acid, fortificant	µg	HPLC	FOLAC
Folate, naturally occurring (food folate)	µg	Microbiological	FOLFD
Dietary folate equivalents	µg	=Food folate + 1.7 × folic acid	FOLDFE
Folate, total or [sum of folates] (vitamin B9)	µg	Microbiological determination or [HPLC]	FOL or [FOLSUM]
Vitamin B12	µg	HPLC, microbiological	VITB12
Vitamin C	mg	=Ascorbic acid + dehydro-acorbic acid	VITC
Cholesterol	mg	HPLC	CHOLE
Fatty acid conversion factor	–	From literature sources	XFA
Fatty acids, total saturated	g	Calculated from FA profile	FASAT
Fatty acid, total monounsaturated	g	Calculated from FA profile	FAMS

Table 4 Components, units, analytical/determination method/definition and corresponding INFOODS component identifier listed in the FAO/INFOODS Food Composition Table for Western Africa (WAFCT 2019), expressed per 100 g edible portion.—cont'd

Component	*Unit*	*Analytical/determination method/definition*	*INFOODS component identifier (tagname)*
Fatty acids, total polyunsaturated	g	Calculated from FA profile	FAPU
Linoleic acid	g	Calculated from FA profile	F18D2CN6
Alpha-linolenic acid	g	Calculated from FA profile	F18D3CN3
Phytate, total	mg	Phytate, total, calculated from phytate phosphorus by anion exchange method (AOAC 986.11) or [phytate, determined by direct precipitation] or [phytate, determined by indirect precipitation]	PHYTCPP or [PHYTCPPD] or [PHYTCPPI]
Inositol triphosphate (IP3)	mg	HPLC	IP3
Inositol tetraphosphate (IP4)	mg	HPLC	IP4
Inositol pentaphosphate (IP5)	mg	HPLC	IP5
Inositol hexaphosphate (IP6)	mg	HPLC	IP6
Nitrogen conversion factor	–	From literature sources	XN

AAS = Atomic absorption spectrophotometry, GLC = Gas liquid chromatography, HPLC = High performance liquid chromatography, ICP-MS = Inductively coupled plasma-mass spectrometry. Components were assigned to a secondary tagname if the preferred one was not available for a given food and the values was presented between square brackets ([ ]).

published outside the laboratory reports, as either the data owners consider them propriety data (mainly by food industry), do not have time to write a scientific article, or the data quality was too low for publication. FAO/INFOODS has published several global databases as repositories of analytical data (see http://www.fao.org/infoods/infoods/tables-and-databases/faoinfoods-databases/en/), such as the FAO/INFOODS Food Composition Database for Biodiversity (BioFoodComp), FAO/INFOODS Analytical Food Composition Database (AnFood), or the FAO/INFOODS/IZiNCG Global Food Composition Database for Phytate (PhyFoodComp), while EuroFIR published eBASIS (Bioactive Substances in Food Information Systems). When compiling analytical data from scientific articles or laboratory reports into the FAO/INFOODS databases, most could not be included because of missing data or imprecize documentation. A common error is an imprecize denominator, e.g., presenting data per 100 g—is this 100 g edible portion on fresh weight basis (EP), 100 g edible portion on dry matter basis (DM), 100 g total food (edible and inedible part) on fresh weight basis, 100 g of fatty acids or something else? Other main reasons for not including the analytical data are missing or imprecize units, poor data quality and/or description of foods, components, method—or missing water value (if data is expressed in dry matter). This is unfortunate and a waste of scare funds for analysis. In many countries, high-quality analytical food composition data are still missing on commonly consumed foods, on underutilized and wild foods, and on the different varieties. They often include raw foods, but rarely different varieties (i.e., biodiversity) or indigenous foods, prepared or processed foods, recipes or branded foods. This represents a major challenge in most countries, as these latter foods represent the majority of foods consumed. Most of the published analytical compositional data are on macronutrients, some on minerals or bioactive components, but few include vitamins. Fiber is unfortunately often analyzed using the crude fiber method, which is unsuitable for human nutrition. Carbohydrates are rarely analyzed, but usually calculated by difference. Fortunately, total carbohydrates by difference (i.e., carbohydrates including dietary fibers) has been phased-out in most of the recently published food composition tables/databases (except e.g., the food composition database of the US Department of Agriculture).

Another source of FCD are nutrition label data. However, they are rarely analyzed, which means that nutrient values of manufactured products are often of lower quality in FCT/FCDB. Further, it seems that food companies calculate the nutrient content of complex foods (e.g., cornflakes, pizza, cakes) for the label by summing the nutrient values of the raw ingredients, without taking into account the weight loss during processing, which will result in a concentration of nutrients in the final product. This means that, in general, the nutrient contents on labels of complex foods are underestimated. These lower values could mean that the nutritional label holds lower than true values.

Because analytical data are incomplete, and analyses are costly and time-consuming, data in FCT/FCDB are largely imputed, calculated, presumed or borrowed from FCT/FCDB from other countries. Imputed data estimated based on similar foods (e.g., values for red beans used fill missing values for white beans). The accuracy of these values relies on good food composition knowledge and an understanding of the variation in the composition of foods by the data compiler. Calculated data are mainly used to estimate the composition of cooked foods, either single foods or mixed dishes. FCD for cooked foods and mixed dishes are commonly derived from recipe calculations. Appropriate calculations require complete recipe information (including the amount of each ingredient and the weight of the cooked recipe), application of nutrient retention factors to estimate the loss of nutrients, and yield factors to account for changes in water and/or fat during cooking, Detailed instruction on these calculations can be seen in lesson 5.2 Recipe calculation in the FAO/INFOODS e-Learning Course on Food Composition Data (FAO, 2013). Presumed zeros may be used in some cases, when a certain component is known to not exist in a food category; for example, alcohol in nonalcoholic beverages or cholesterol in plant-based products. Borrowed data from other countries are often the alternative used where local data are not available. These data need to be scrutinized to ensure that they match foods consumed locally, e.g., for fortification, cooking practices, color of fruit and vegetables, etc. Adapting data from another country can be challenging and there will always be uncertainty around borrowed data, thus its use should be limited.

FCTs/FCDBs should include as few missing values as possible for nutrients of interest. If values are missing, the user should estimate or borrow the missing data—this task demands a higher level of expertize. If this is not done and missing values are treated as zero, nutrient intakes will be underestimated.

In recent years, documentation at value and food levels, also called meta data, has allowed users to appreciate the origin and quality of compositional data in FCTs/FCDBs. This documentation is key for quality assessment and every published FCD and FCT/FCDB should take great care to adequately document their data, including the source of data.

What is the global availability of FCTs/FCDBs?

Country-specific or regional FCT/FCDB should be developed, due to the different factors that may affect the composition of foods (e.g., different food varieties/breeds, soil composition, food fortification and cooking practices). Published FCT/FCDB vary considerably in terms of data quality, documentation, food and nutrient coverage, analytical methods used and accessibility. Using FCD from another country to estimate nutrient intakes may lead to wrong conclusions when evaluating micronutrient adequacy.

Published FCTs/FCDBs can be found at the INFOODS website (http://www.fao.org/infoods/infoods/tables-and-databases/en/). For more than 190 existing countries in 2021, 105 FCT/FCDB were available (**Fig. 1**), of which only 25 were published or updated less than five years ago, and only 27 were available online in user-friendly format (e.g., Excel or Access). Others were available online, but only as PDF, as food-by-food search, or as printed versions, which limits their usability, especially for dietary assessment. Although many FCT/FCDB have been published, most of the newer FCT/FCDB have been updated only in Europe and Africa (**Fig. 1**). A successful example is the Indian Food Composition Table published in 2017, with solely analytical data for about 200 food components in over 500 raw foods collected using a national and appropriate sampling plan (Longvah et al., 2017). FCTs/FCDBs must be regularly updated to reflect changes both in the food supply and components of interest due to advances in the nutrition science.

Few countries have a fully funded and staffed national food composition program, which regularly publishes updated FCT/FCDB with new analytical data, new foods and/or components. In most developing countries and many developed countries, political and institutional support is lacking and thus FCT/FCDB are often produced within projects or as a private initiative, i.e., not within an institutional framework. This is one of the reasons why many FCT/FCDB are outdated, cover only few foods or nutrients and are not produced according to the international standards of INFOODS. Hence many countries rely on data from high-income countries, such as the Food Data Central published by the United States Department of (https://fdc.nal.usda.gov).

Selecting high-quality food composition data and databases

To assist users in selecting the most adequate FCT/FCDB, experts have developed the FAO/INFOODS Evaluation Framework to Assess the Quality of Food Composition Tables and Databases (FCT/FCDB)—(Evaluation Framework) (Charrondiere et al., 2021). This is done in two steps: a screening with eight questions and, if passed, a full evaluation from user or compiler perspective with up to 190 questions in seven categories: Documentation in the FCT/FCDB to inform users, food coverage, food identification, component and value coverage and component expression, quality of data—analysis-the compilation process and checking; public access of FCT/FCDB, and year of publication and frequency of updates.

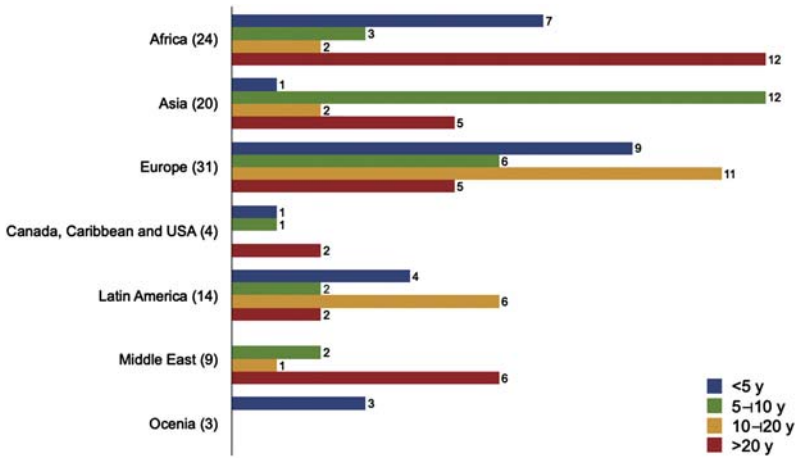


Fig. 1 Food composition tables and databases (FCT/FCDB) listed in the INFOODS inventory in March 2021 by region and according to the number of years since last update. Values in parentheses correspond to the total number of FCT/FCDB available in the region.

How to assure that food composition data, FCT and FCDB are generated, compiled and used correctly

Publishing and using high-quality FCTs/FCDBs are of great importance to ensure the accuracy of nutrient intake estimations, as previously discussed. However, incorrect use of even a high-quality FCT/FCDB can lead to poor results and wrong conclusions. The appropriateness of FCD use requires trained professionals on the basic knowledge to generate, compile, evaluate and use FCTs/FCDBs.

Three pillars are needed to ensure that high-quality food composition data are correctly generated, compiled, disseminated and used:

1. International standards, guidelines and tools on the generation and compilation of food composition data must be developed and used, e.g., by INFOODS
2. National and/or regional food composition programs must exist, which update FCT/FCDB regularly, and
3. Professionals must be trained in all aspects related to food composition.

If guidelines and programs exist, but professionals do not know how to use them, they are of little interest. Therefore, there is a great need for training professionals and students in the different aspects of food composition. Despite being the most effective means to spread food composition knowledge and recognition, food composition is still rarely included in the formal training of nutritionists and other professionals. Increasingly, there are international courses, and distance and e-learning tools.

The basic food composition knowledge that any professional using, generating and compiling FCD should have (e.g., nutritionist, dietitian, health professional giving dietary advice, chemist), are included in the FAO/INFOODS e-Learning Course on Food Composition Data (FAO, 2013). This e-learning course takes about 14 h to complete, is available in English and French and can be downloaded free-of-charge (or used online) from FAO e-learning platform (https://elearning.fao.org/course/view.php?id=191). The course explains the importance of appropriate food and component descriptions, units, denominators, biodiversity, recipe calculation and compilation methods, and data quality, and relates all these issues to their influence on the resulting nutrient values. It also describes briefly the impact good or bad FCD can have on programs, policies, laws and regulations. This e-learning course was designed to be used in universities, even in undergraduate courses.

Those who wish to acquire more knowledge on food composition may also explore the FAO/INFOODS Food Composition Study Guide (volume 1 and 2), available in English, Spanish and French and downloadable free-of-charge from the INFOODS website (http://www.fao.org/infoods/infoods/training/en/). It takes about 2 weeks to complete, as it is more comprehensive and asks learners to complete questions and exercises.

Additionally, there are classroom based Food Composition Courses, such as the International Postgraduate Courses on the Production and Use of Food Composition Data in Nutrition, offered in Wageningen, Netherlands, or there are regional courses (see http://www.fao.org/infoods/infoods/training/en/- past and upcoming courses). In addition, FAO/INFOODS have organized many one-day training courses around nutrition conferences and held webinars (http://www.fao.org/infoods/infoods/webinars/en/).

Also, there are websites with important information such as the INFOODS website (http://www.fao.org/infoods/infoods/en/) with guidelines, training material, tools and inventories of FCT/FCDB and much more. Other websites include Compilers' Toolbox™ (http://toolbox.foodcomp.info/), Food Composition Explained website (https://www.foodcomponline.info/) or the EuroFIR website (https://www.eurofir.org/).

Those who wish to become a real food composition expert, however, must work with food composition data and follow international guidelines such as the ones from INFOODS, ideally under the initial supervision of a food composition expert. Experience and expert knowledge come with time.

For dietary assessment, food matching is key for the quality of nutrient intake assessment. Therefore, users should choose quality FCT/FCDB and match correctly reported foods to the foods in the FCT/FCDB, as outlined in the FAO/INFOODS Guidelines for Food Matching (Stadlmayr et al., 2012). In general, first check if the FCT includes foods and components of interest, and then check food description and components according to **Table 5**. The guidelines also include a quality scheme for food matching which could be used to estimate the quality of nutrient intake estimations.

Conclusions and way forward

Great progress has been made in the availability and quality of FCD, and food composition standards. However, more analytical data are badly needed, especially on vitamins, and more people should have knowledge on food composition to assure that FDC are generated, managed and used correctly. Therefore, food composition should be recognized as science, and be taught in nutrition or dietetics programs, among others. In this case, future decision makers could be more willing to invest in food composition programs including chemical analysis, compilation and dissemination of FCD. With better compositional data, many policies and programs could be better designed and targeted.

Table 5 Food matching criteria from the FAO/INFOODS Guidelines for Food Matching.

Food description	
Food name and descriptors	Food description should be complete and unambiguous, including taxonomic/scientific name and forms of preparation when appropriate.
Water and fat content (and additional components of interest)	Water and fat (for some foods) should always be checked against the food description to assess whether the values correspond to the food of interest.
Food components	
Mode of expression	Important for components that can expressed in different forms (e.g., total and available carbohydrate).
Definition	Some components may have the same common name but different definitions (e.g., vitamin A defined as retinol equivalents or retinol activity equivalents).
Conversion factors	For some components not only mode of expression and definition are required, but also the conversion factors applied (e.g., nitrogen conversion factors to calculate protein values).
Analytical methods	Analytical methods may result in significantly different values (e.g., dietary fiber analyzed by AOAC-Prosky and crude fiber method).
Units and denominator	Units are used to quantify the amount of a component (e.g., g, mg, μg, kcal) while the denominator define in how much of a food or beverage the component is found. Normally, in a FCT/FCDB present data "per 100 g edible portion on fresh weight basis," but care should be taken, especially for liquids.

Conflict of Interest

The views expressed in this information product are those of the author and do not necessarily reflect the views of the Food and Agriculture Organization of the United Nations (FAO).

References

Charrondiere, U.R., Stadlmayr, B., Grande, F., Vincent, A., Oseredczuk, M., Sivakumaran, S., Puwastien, P., Judprasong, K., Haytowitz, D., Gnagnarella, P., 2021. FAO/INFOODS Evaluation Framework to Assess the Quality of Food Composition Tables and Databases (FCT/FCDB)—(Evaluation Framework). URL. http://www.fao.org/infoods/infoods/standards-guidelines/en/. (Accessed 17 May 2021).

European Food Safety Authority, 2015. The food classification and description system FoodEx 2 (revision 2). EFSA Support. Publ. 12. https://doi.org/10.2903/sp.efsa.2015.EN-804.

FAO, 2013. FAO/INFOODS e-Learning Course on Food Composition Data. Food and Agriculture Organization of the United Nations. URL. https://elearning.fao.org/course/view.php?id=191. (Accessed 24 March 2021).

Global Fortification Data Exchange, 2021. Fortification Overview: Mandatory Fortification of Wheat Flour. URL. https://fortificationdata.org/interactive-map-fortification-legislation/. (Accessed 17 April 2021).

Grande, F., Giuntini, E.B., Coelho, K.S., Menezes, E.W., 2019. Elaboration of a standardized dataset for foods fortified with iron and folic acid in Brazil. J. Food Compos. Anal. 83. https://doi.org/10.1016/j.jfca.2019.103285.

Klensin, J.C., Feskanich, D., Lin, V., Truswell, A.S., Southgate, D.A.T., 1989. Identification of Food Components for INFOODS Data Interchange. United Nations University Press, Tokyo.

Longvah, T., Ananthan, R., Bhaskarachary, K., Venkaiak, K., 2017. Indian Food Composition Tables. National Institute of Nutrition, Hyderabad.

Møller, A., Ireland, J., 2018. The LanguaL Thesaurus Technical Report 2017. Danish Food Informatics.

Stadlmayr, B., Wijesinha-Bettoni, R., Haytowitz, D., Rittenschober, D., Cunningham, J., Sobolewski, R., Eisenwagen, S., Baines, J., Probst, Y., Fitt, E., Charrondière, U.R., 2012. FAO/INFOODS Guidelines for Food Matching. FAO, Rome.

Statista, 2021. Usage of Dietary Supplements Among U.S. Adults in 2020, by Age Group. URL. https://www.statista.com/statistics/307917/dietary-supplement-usage-us-adults-by-age-group. (Accessed 17 May 2021).

Vincent, A., Grande, F., Compaoré, E., Amponsah Annor, G., Addy, P.S., Aburime, L.C., Ahmed, D., Achu, M.B.L., Dahdouh Cabia, S., Deflache, N., Dembélé, F.M., Dieudonné, B., Edwige, O.B., Ene-Obong, H.N., Fanou Fogny, N., Ferreira, M., Jemide, J.O., Kouebou, P.C., Muller, C., Nájera Espinosa, S., Ouattara, F., Rittenschober, D., Schönfelft, H., Stadlmayer, B., van Deventer, M., Razikou Yiagnigni, A., Charrondière, U.R., 2020. FAO/INFOODS Food Composition Table for Western Africa (2019). FAO, Rome.

Relevant websites

Compilers' Toolbox™, http://toolbox.foodcomp.info/.
European Food Information Resource, https://www.eurofir.org/.
FAO/INFOODS e-Learning Course on Food Composition Data, https://elearning.fao.org/course/view.php?id=191.
Food Composition Explained, https://www.foodcomponline.info/.
International Network of Food Data Systems (INFOODS), http://www.fao.org/infoods/infoods/en/.

Food fortification: Technological aspects

O. Dary[a], **M. Guamuch-Castañeda**[b], **and J.O. Mora**[c], [a] United States Agency for International Development (USAID), Washington, DC, United States; [b] Institute of Nutrition of Central America and Panama (INCAP), Guatemala, Guatemala; and [c] Private Consultant, Washington, DC, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of O. Dary, J.O. Mora, Food Fortification: Technological Aspects, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 306–314, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00120-3.

Introduction 340
What is food fortification? 340
How does food fortification work? 341
Which factors limit the amounts of micronutrients in fortified foods? 341
Safety 341
Technological compatibility 342
Cost 342
Which other technical factors are important in food fortification? 343
Physical segregation 343
Dilution factor 343
The special case of rice fortification 343
Types of food fortification 343
Mass-fortification 344
Targeted-fortification 344
Market-driven fortification 344
Home fortification or fortification at the point of consumption 344
What else is needed to implement a fortification program? 345
Compliance and enforcement 345
Dietary monitoring 347
Conclusions 349
References 349
Further reading 350
Relevant websites 350

Key points

- Understand that the biological impact of food fortification is related to the nature, quality, and quantity of the incorporated micronutrients and not to the fortified foods; however, population coverage depends on the fortification vehicle.
- Describe the use of the Estimated Average Requirement (EAR), instead of the Recommended Nutrient Intake (RNI), as the reference values for determine the need and the success of micronutrient programs for populations, and the use of the Tolerable Upper Intake Level (UL) to ensure safety and prevent excessive micronutrient intakes.
- Explain that adequate intake matters, but that bioavailability (efficient absorption) and bioconversion (transformation into the metabolically active forms) of the micronutrients to ensure their bioefficacy, i.e., their biological impact is also important.
- Illustrate that programmatic viability of a food fortification program is associated with participation of formal and centralized food industries, the presence of low cost, practical, and efficient industry quality control/assurance systems and governmental inspection actions, complemented with periodical epidemiological surveillance to demonstrate population coverage and benefit.
- Promote the recognition that more than one fortified food or more than one micronutrient intervention may be needed to correct dietary micronutrient gaps but, at the same time, their design should be coordinated to prevent excessive intake of micronutrients.
- Clarify that the distinction between mass-, targeted-, and market-driven fortification is related to the objectives and conditions of the program, not to the type of fortification vehicles, and that home-fortification is not a food fortification program.

Glossary

Codex Alimentarius Inter-country organization of the United Nations system, led by FAO and WHO, with the purpose of proposing general principles, standard models, practices, and guidelines, aimed to protect the health of consumers and ensure fair trade associated with foods. The Codex recommendations are recognized by the World Trade Organization as international reference criteria in those matters

Estimated Average Requirement (EAR) The average daily nutrient intake level estimated to meet the needs of half of the healthy individuals in a particular physiological stage-, gender-, and age-specific population group. It is used as the reference intake value for population diets

Fortificants The chemical forms that are the sources of micronutrients for food fortification

Home Fortification Term used to identify micronutrient supplements in the form of a powder that are added into foods before consumption. It is not food fortification as the micronutrient vehicle is a supplement

Micronutrients These are vitamins and minerals that are required by humans in very small amounts; most of them cannot be synthesized by the human body and therefore they should be obtained from the diet.

Premixes Blends of fortificants with antioxidants, anticaking, excipient, and other substances to provide the adequate physical and chemical properties that are required to produce homogeneous and reasonably stable fortified products

Recommended Nutrient Intake (RNI) The daily average nutrient intake that meets the nutrient requirements of almost all apparently healthy individuals in a particular physiological stage-, gender-, and age-specific population group. The RNIs are set as the EAR plus two standard deviations (i.e., for covering 97.5% of the population). In some countries, the term Recommended Daily Allowance is used instead of RNI. It is the used reference value for diets aimed to individuals; and it is also used as reference for expressing the nutrient content of foods in the nutrition panels of the food labels

Tolerable Upper Intake Level (UL) The highest average daily nutrient intake level unlikely to pose a risk of adverse health effects to almost all (97.5%) apparently healthy individuals in a particular physiological stage-, gender-, and age-specific population group. It is used as the safe reference value of intake for diets of both individuals and populations

Introduction

Food fortification is a strategy to increase the content of vitamins and minerals (micronutrients) in the diet by adding them to commonly edible products, mainly staple foods (cereal flours, oil, rice, and dairy products) and condiments (salt, sugar, bouillon cubes), when processed by the food industry. Biological impact is related to the correction of micronutrient gaps. Several fortification vehicles can be alternatives for the same micronutrient, and their selection depends on the industrial structure and dietary habits of each country. Although fortifying foods by small and underdeveloped factories may be feasible and even efficacious, the viability and sustainability of fortification programs depends on the involvement of formal and centralized food factories to take advantage of economy of scale, efficient governmental supervision, and population coverage. This article describes factors that influence food fortification program viability and effectiveness; reviews the biological, technical, economic, and programmatic factors that limit the quantities of micronutrients that can be added to fortified foods; and justifies the need for considering several micronutrient interventions - used in a coordinated manner—to reduce dietary inadequacies of at-risk populations.

What is food fortification?

The Codex Alimentarius defines food fortification or enrichment as the addition of micronutrients to foods, whether or not they are normally contained in the food, for the purposes of preventing or correcting a demonstrated micronutrient deficiency. The Codex also notes that the quantity of micronutrients to add should be sufficient to correct or prevent deficiency when the food is consumed in normal amounts by the population at risk, but not likely to result in excessive intakes by individuals with high intake of the fortified food (Codex Alimentarius, 2015). These recommendations are applicable to single foods.

The WHO/FAO proposed a more appropriate definition in the Guidelines on Food Fortification with Micronutrients, which focuses on the diet rather than on single foods. A single food may contribute toward improving the nutritional quality of the diet but may not necessarily be sufficient as the only solution to prevent a micronutrient inadequacy (i.e., insufficient intake to satisfy the nutrient requirements) (Allen et al., 2006). This is the concept adopted in this article.

In food technology, fortification and enrichment have different meanings: fortification is reserved for the addition of micronutrients to a food that does not contain those nutrients naturally, whereas enrichment is applicable when the natural contents of some micronutrients normally present in the food are intentionally increased. Thus, addition of iodine to salt is an example of fortification whereas addition of iron to cereal flours is an example of enrichment. Two other related terms are frequently used: restoration, when micronutrients are added to recover the original micronutrient contents in a food that has partially or totally lost them during processing, for example, adding vitamins A and D to defatted milk to reproduce the content of those vitamins in whole milk; and nutritional equivalence or standardization, when the content of micronutrients of a manufactured food is modified to imitate

the micronutrient content of a natural food that is intended to be replaced, as, for example, adding vitamin A and D to margarine to achieve their natural micronutrient contents in butter (Richardson, 1990).

The chemical sources of micronutrients used in food fortification are called fortificants. Thus, for example, ferrous sulfate, ferrous fumarate, and NaFeEDTA are fortificants used to increase the content of iron in foods (Hurrell, 2021). Fortificants are generally added to foods as part of premixes, which also contain substances to increase the stability and the technical and sensorial properties of the fortificants.

How does food fortification work?

Many articles in the scientific literature note that consumption of fortified foods has positive health outcomes (Keats et al., 2019; Field et al., 2020; Pachon, 2021), but rarely emphasize that those outcomes are associated with the quality and quantity of the micronutrients that are added to the foods rather than to only the consumption of the fortified foods. For example, vitamin A can be added to several food vehicles such as oil, sugar, cereal flour, milk, and even bouillon cubes, but the biological impact does not depend on the fortification vehicle but to the amount of vitamin A that is being delivered to the consumer, which is independent to the type of fortified food (Dary and Mora, 2002). The purpose of using a food vehicle is to reach a large number of individuals in vulnerable populations in an efficient and low-cost manner. Food fortification takes advantage of the manufacturing structure of the food industry and the already existent delivery system of their products. In other words, fortified foods aim to cover as many individuals as possible, but the biological impact depends on the proportion of individuals who meet their nutritional needs through the additional intake of micronutrients supplied by the fortified foods. The magnitude of impact depends on the extent to which the nutritional gap is corrected. Therefore, the benefit of fortified foods varies from population to population, as the same fortification formula does not necessarily replicate results from one community to another because the amount of fortified foods consumed and the magnitude of the nutrient intake gap may differ. The goal is that most individuals of the population reach the Estimated Average Requirement (EAR) of intake, combining diets and fortified foods (Allen et al., 2020). It is important to point out that, although the nutritional panel of food labels are expressed as percentage of Recommended Nutrient Intakes (RNI, which in some countries is similar to the Recommended Dietary Allowance, RDA), for population programs, the parameter of reference is the EAR. The RNI or RDA values are applied to diets for individuals and not for populations (Institute of Medicine, 1998).

On the other hand, sufficient intakes of micronutrients do not necessarily lead to noticeable biological impacts if the diet and/or metabolic condition of the individuals are unfavorable. The bioefficacy (i.e., the metabolic function of a nutrient) of the added micronutrients depends on three factors: intake, bioavailability (i.e., transference of the micronutrient from the intestinal track to the interior of the body), and bioconversion (i.e., transformation of the micronutrient to its metabolically active form). For example, bioavailability of synthetic β-carotene may vary, depending on the content of fat in the meal where the fortified food is consumed, because fat is needed for its absorption. Moreover, in individuals with zinc deficiency, the bioconversion of β-carotene to vitamin A may be impaired (Van Lieshout et al., 2003). Another example is folic acid (a precursor of the vitamin folate or B_9, which is oxidized and stable) that is easily added to foods and highly bioavailable but does not act alone but in coordination with vitamins B_2, B_6, and B_{12} (Selhub, 2008).

In summary, food fortification increases the population supply of micronutrients in the diet by using products manufactured by the food industry, but biological impact depends on the extent to which nutritional gaps are corrected. Nevertheless, the programmatic viability of food fortification is related to the involvement of large and centralized food industries, which is a characteristic that originated the terms Large-Scale Food Fortification (LSFF) or Industrial Food Fortification. Although small industries can fortify foods and those foods can have biological impact, the cost-efficiency (i.e., the relative absolute cost of increasing the supply of micronutrients) and the cost-effectivity (i.e., the relative beneficial cost to attain a biological impact) are unfavorable. Neither the saving of production at large scale nor practical tracing and inspection of the products are possible, and therefore fortification under this condition is unscalable and unsustainable. Food fortification is an opportunistic strategy that takes advantage of existent formal industries and their efficient distribution networks.

Which factors limit the amounts of micronutrients in fortified foods?

Although the micronutrient content in the fortified food is not the only condition to attain biological impact, it continues being an important element for this objective. Micronutrient content can be adjusted to the consumption pattern of the fortification vehicle; lower food intakes require larger micronutrient content to supply the same quantity of micronutrients. However, it is important to realize that the incorporation of micronutrients to foods may be limited because of safety, technological compatibility, or cost constrains.

Safety

Both the Codex Alimentarius (2015) and the WHO/FAO (Allen et al., 2006) recommend that fortification should be designed to ensure that almost all individuals in the population do not have excessive micronutrient intake. The total micronutrient intake for most individuals should not exceed the tolerable upper intake level (UL) (Allen et al., 2020). Young children and adult males have

the lowest and the highest UL values, respectively, and those coincide with smaller and larger intakes of staples and condiments, which are the most common fortification vehicles used in large-scale food fortification. Therefore, in practical terms, the highest micronutrient content in these fortification vehicles should not cause that some of the adult or adolescent males to reach the UL values of intake. However, for packaged foods with fixed serving sizes regardless of the population group (e.g., breakfast cereals and granola bars), the micronutrient content should be checked against the UL values for young children. It is important to point out that the calculation of total intake should be carried out for the overall supply, i.e., the combined amount provided by the diet plus all the fortified foods or any other micronutrient supplying intervention.

Technological compatibility

In general, fortificants should not change the stability, color, odor, or flavor of the food vehicle. Negative interactions between fortificants and vehicles impede the adoption of high micronutrient content. For example, vitamin B_2 and β-carotene, when added to food matrices that are white (refined flours and rice, e.g.), cause yellowish-orange colors. Folic acid and vitamin B_{12} may also change color of the matrix. Some iron salts with good water solubility and bioavailability, such as ferrous sulfate and ferrous bisglycinate, favor rancidity, whereas others such as NaFeEDTA may produce discoloration. Technological incompatibility reduces the impact of iron fortification, because fortificants with higher bioavailabity cannot be added in high amounts. In this case, the selection of the fortificant rests on a decision that combines estimation of bioavailability and the highest possible fortification content. For example, for maize flour, it is possible to use 40 mg iron/kg from ferrous fumarate, but only 15 mg iron/kg from NaFeEDTA. If one assumes that the bioavailabilities are 5% and 15%, respectively, any of these two iron fortificants would have a similar biological impact, and the compound with the lowest cost would be preferred (Dary, 2007). Another example is the iron fortification of bouillon cubes with ferric pyrophosphate, which has low bioavailability, compared to water-soluble iron compounds; however, the latter cannot be used because it causes rancidity to the bouillon cube (Gonzalez de Mejia et al., 2015).

Fortificants in the form of encapsulates have been developed for oily vitamins (A and D, e.g.), those required in very small amounts (B_{12}) or that are highly reactive (iron). In this form, micronutrients can be incorporated into dry foods. Moreover, micronutrients in encapsulate forms are more stable and negative interactions between the fortificant and food matrix are minimized.

Cost

The cost of adding fortificants to foods depends on the price of the premix and the selected fortification contents. To design programs congruent with usual production and trade practices, the increment in price due to fortification should not be too high, because it might distort the market and so promote incompliance and unfair competition. Most Large-Scale Food Fortification programs are based on staples and condiments, and the increment in price should be lower than 3%. Although this small increment in the price seems low, some governments do not allow it, because the fortification vehicles are basic components of the family food basket and, therefore, the price might be politically sensitive.

Fortificants with high cost can only be added in certain conditions. **Table 1** illustrates why the addition of iodine to salt is much easier than the addition of vitamin A or iron. For example, supplying 100% of the EAR to an adult male consuming 10 g d^{-1} of salt, fortification cost per metric ton (100 kg) of salt for iodine, vitamin A, and iron are US $1.27, US $42.90, and US $90.00, respectively. These costs represent 0.42%, 14.3%, and 30.0% of the retail salt price, respectively (assuming US $0.30/kg). Consequently, the quantity of vitamin A and iron to salt must be low for a program to be feasible, and the additional contribution of these micronutrients by fortified salt would be proportionally lower or negligible. Therefore, addition of these two micronutrients to salt may not be programmatically viable or biologically effective.

Table 1 Conditions for fortification of salt with different micronutrients.

Nutrient	*Fortificant*	*EAR*[a] of adult males (mg)	*Nutrient content*[b] (mg/kg)	*US$/metric ton*[c]	*% Price*[d]	*Dilution factor*[e]
Iodine	Potassium iodate	0.100	10	1.27	0.42	1:60,000
Vitamin A	Dry 250,000 IU vitamin A	0.429	42.9	42.90	14.3	1:1750
Iron	Micronized ferric pyrophosphate	30.0	3000	90.00	30.0	1:85

[a]Estimated average requirement (EAR), and assuming a diet with low iron bioavailabity.
[b]Estimated for supplying 100% of the EAR of adult males with a salt intake of 10 g d^{-1}, and without losses of the micronutrients.
[c]Based on estimated global prices of the fortificants.
[d]Assuming that the salt price is US $0.30/kg.
[e]Considers that the fortificants are not 100% micronutrient. In this case, iodine in potassium iodate is 59%; vitamin A in a dry beadlet form is 7.5%; and iron in micronized ferric pyrophosphate is 25%.

Which other technical factors are important in food fortification?

Physical segregation

Dry food matrices may also suffer segregation (separation) of the fortificant particle from food particles, due to differences in sizes and densities. This mainly affects food matrices with large particle sizes, such as sugar and some types of salt. To avoid segregation, the particle size of the fortificant and the food matrix should be similar, or the fortificant should be adhered to the surface of the food particles. For example, in the case of sugar fortified with vitamin A, the vitamin-containing beadlet is "glued" to the surface of the sugar crystal by means of a layer of vegetable oil. However, during storage and transportation the friction among sugar crystals may separate some of the vitamin A beadlets and reduce the content of vitamin A (Mora et al., 2000). **Fig. 1** illustrates how the vitamin A beadlets adhere to the surface of the sugar crystals.

Dilution factor

The amount of fortificant required to reach the target micronutrient content in a given food is rarely examined when food fortification programs are being considered. **Table 1** shows dilution factors in salt for iodine, vitamin A, and iron at fortification contents required to provide 100% of the EAR; these are 1:60,000, 1:1750, and 1:85 (dilution factors are reduced further when the fortificants are incorporated into premixes), respectively. This dilution explains why iodine can be added as a liquid spray, while vitamin A and iron must be incorporated into dry forms. It also explains how the low dilution factor for iron requires large amounts of the fortificant and, therefore, large amounts of premix to be transported, managed and stored. Further, special devices for addition and mixing are required. Other products derived from salt, such as bouillon cubes, face the same type of limitations for fortification with iron if biologically important quantities of this mineral are desired.

The special case of rice fortification

In developing countries, milled rice is cleaned and washed before consumption and, therefore, the methodology of adding the premix as a powder over the polished rice -as it is the practice in some industrialized countries-is not possible, because the micronutrients are going to be washed out before consumption. Therefore, rice fortification requires the incorporation of the micronutrients either through artificial fortified kernels (where micronutrients are embedded inside the kernel) or micronutrient-coated grains. In both cases, these products are manufactured by highly specialized industries. Furthermore, the dilution factor is between 1:200 and 1:100 which in the case of rice may be compatible as many rice mills mix several types of rice grains, offering different qualities of commercial rice presentations. However, the cost of the enriched kernels may be restrictive, as more than half the cost is associated with the process rather than the fortificant (for other staples, 90% or more of the total cost is due to the fortificant). The cost of rice fortification is between $10 and $20 per metric ton, which represents 10–20% of the rice price (assuming $1/kg) (Alavi et al., 2008). The special manufacturing requirements and cost limit of scaling-up rice fortification, make it viable for only a few countries or social program.

Types of food fortification

Three main types of food fortification were identified by the WHO/FAO (Allen et al., 2006) based on application and scope: mass-, targeted-, and market-driven. The same food can support any of the three types of fortification. For example, rice could be

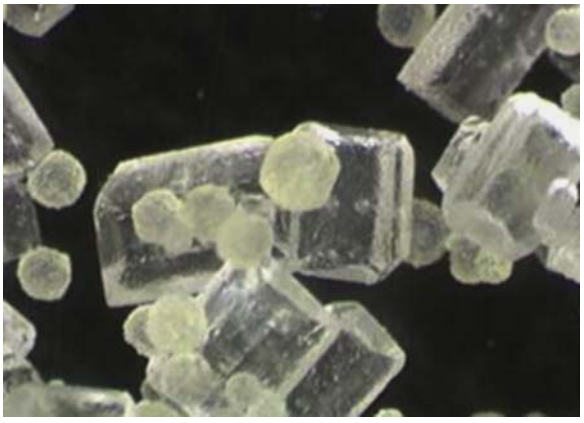


Fig. 1 Microphotography of vitamin A beadlets (fortificant) attached to sugar crystals (food matrix) by means of a thin layer of vegetable oil. Reproduced with permission from DSM.

mass-fortified if the country decides that all rice distributed to the population should be fortified, as in Costa Rica (Martorell et al., 2015). Fortified rice may also be used for targeted fortification if it is produced and distributed through specific social programs (Rai et al., 2019), such as in humanitarian food aid programs supported by USAID or the World Food Program. Finally, rice could also be used for market-driven fortification, as in China, where industries launched fortified rice aimed at high-income groups (Alavi et al., 2008).

Mass-fortification

Mass-fortification is defined as that aimed to the general population using an already widely consumed product, without requiring changes in consumer behavior. In mass-fortification, patterns of consumption of the fortified food are determined and vary considerably among countries, regions, age groups, and socio-economic strata. This variation in food intakes limits the quantity of micronutrients that can be added because of the need to avoid excessive intakes by individuals with high intake of the fortification vehicles. Mass-fortification uses staples and condiments, which are products extensively and widely consumed. Voluntary fortification of these products is difficult because the consumer usually does not select the product by brand, and because these foods are ingredients of other processed foods and therefore "hidden" to the consumer. Moreover, small differences in price may distort the market for staples and condiments or be politically sensitive. Therefore, this type of fortification is usually mandatory and instigated by governments, and the standards are usually made compulsory through specific regulations.

Mass-fortification requires strong governmental enforcement, both for ensuring compliance of the standards as well as for creating a leveled playing-field among producers. Food industry is hesitant to adopt fortification if they do not see that the market is protected or that governmental enforcement is not guaranteed.

It is important to note that mass-fortification should not increase the intake of the food vehicles, as high consumption of most (salt, sugar, oil, refined cereal flours, polished rice) are discouraged with the purpose of preventing non-communicable diseases. Therefore, promotional campaigns on mass-fortification are not needed, because the consumers should keep their usual dietary habits. The strategy is to use fortified foods as practical delivering vehicles of micronutrients, but not to raise their intake. In this sense, mass-fortification may not have an economic incentive for the involved industries, but it is more in the area of reputational prestige and social responsibility toward the societies that they serve.

Targeted-fortification

Targeted fortification describes the strategy designed for specific population groups, such as infants and young children (complementary foods), school-feeding and public health or social programs, which are implemented for benefiting special vulnerable groups. Fortified foods are generally given in specific serving sizes and, because other fortified products are rarely available, the micronutrient content is high to meet daily nutrient requirements using a few products. Furthermore, because the products are frequently subsidized, higher micronutrient content is feasible, because the cost or retaining the characteristics of the unfortified foods are not too restrictive. Fortification standards of targeted-fortification are different from those of mass-fortification, although the same food vehicle may be used, as the conditions and the intended target population for the two types of food fortification differ.

Market-driven fortification

Market-driven fortification occurs when the food industry decides to introduce additional perceived value to their product to attract consumers toward a food with enhanced nutritional value. These products are usually packaged, labeled, and branded. An increasing number of countries are enacting standards that apply to these products, including nutritional claims and advertising, with the purpose of preventing unnecessary high micronutrient exposure, or to protect the consumer against misleading advertising of products that may not be healthy. The Codex Alimentarius encourages competent national or regional authorities to establish specific provisions for the addition of nutrients in food standards according to the national or regional nutritional context, including population dietary habits and specific nutritional guidelines. Thus far, few countries have introduced standards for market-driven fortification, including minimum and/or maximum contents within which essential nutrients should be present, or specifications of foods that should not have added vitamins or minerals, based on their nutrient profile. Although market-driven fortification is applicable voluntarily, governments have the responsibility of enacting standards and regulations.

Home fortification or fortification at the point of consumption

The terms home fortification or fortification at the point of consumption have been used to describe use of micronutrient supplements in the form of powders that are mixed with foods just before consumption (WHO, 2011). The fact that the product is mixed with foods before being consumed does not represent an example of food fortification. Indeed, this is clear example of micronutrient supplementation, because the vehicle that is used for carrying the micronutrients is a supplement and not a fortified food.

What else is needed to implement a fortification program?

The usefulness and success of any food fortification program requires that it is programmatically controllable and that its dietary contribution is assessed periodically. This implies the existence of practical, efficient, and reliable mechanisms for compliance and enforcement, and the introduction of dietary monitoring systems to document the epidemiological significance of the programs.

Compliance and enforcement

Compliance and enforcement are based on the usual practices of food control, which have two components: internal checking by the industry and external confirmation by governmental authorities (Rowe, 2020). Internal checking refers to quality assurance (actions during production that are needed for complying with standards and technical specifications) and quality control (chemical and physical analyses of samples of the industry-manufactured products to check compliance with the standard and specifications, including fortification) by the food industry.

External confirmation (also known as regulatory monitoring) includes auditing (review of the industry quality assurance procedures and their documentation) and inspection (confirmation that the product complies with standards and technical specification through sampling at the production places, distribution centers and retail outlets, and importation sites) by governmental authorities responsible for public food safety.

The most important components are quality assurance and auditing, as they establish good manufacturing practices along the chain of production with the purpose of implementing timely correcting actions (Guamuch et al., 2007a).

Quality control and inspection demonstrate that the final product complies with the criteria of presentation; labeling and packaging; toxicological and microbiological safety; expected technical properties and, in the case of fortification, the specified micronutrient content. Fortification is only one element measured in these systems. Quality control and inspection require testing a few samples of foods per day, week, or month, as compliance and enforcement mostly depend on operational quality assurance and auditing actions. Fig. 2 illustrates quality control data for fortified wheat flour by formal industry testing of two daily composite samples (i.e., combination of single samples taken systematically during the day) per week.

The Central American technical standard (RTCA 67.01.15:06) specifies a minimum content of 55 mg iron/kg. Historically, standards for micronutrient content usually specify "minimum" values. However, this has created problems, because the micronutrients low concentration with variation around the mean depending on several factors: the nature of the food (liquid or dry), quantity of the fortificant (high or low), quantity of the food used for the analysis (large or small), mixing performance (good or inefficient), variation of the analytical method (small or large), and others. Fig. 3 illustrates the hypothetical example of using single or composite food samples for the determination of micronutrient content. Although the mean is similar in the two cases, the proportion of samples below the "minimum" value is much larger using single samples, even in the case where the fortification process might be appropriate and acceptable. Therefore, it is preferable to specify in the standards the mean content (i.e., the target content) as the most important value to check. For promoting homogeneity, an allowable range around the mean should also be added, and the analytical conditions should be described in detail to interpret the results against those reference values. For example, in the case

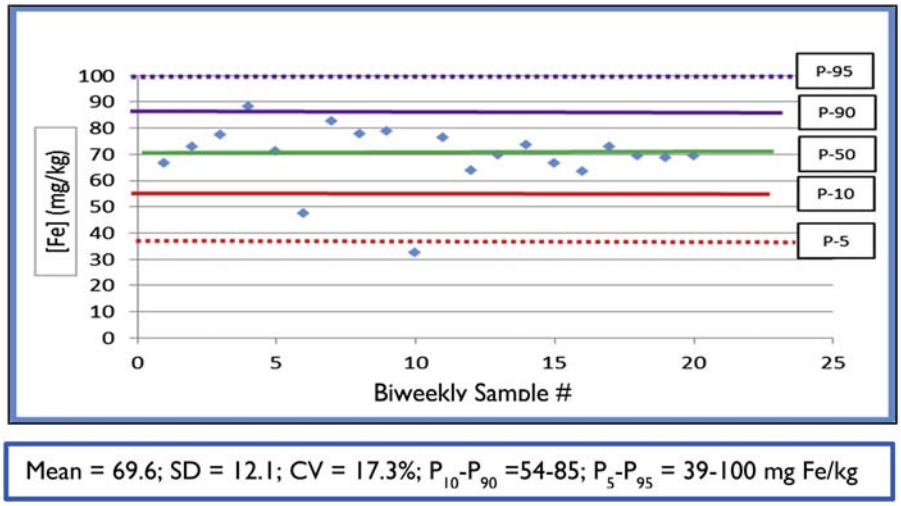


Fig. 2 Quality control data of fortified wheat flour (iron content) in the CENTIA flour mill of Guatemala, using daily composite samples (i.e., combination of several single samples taken systematically during the day). SD = Standard Deviation; CV = Coefficient of Variation; P_{10}—P_{90} = Range between percentiles 10–90 for a program with 80% compliance; P_5—P_{95} = Range between percentiles 5–95 for a program with 90% compliance.

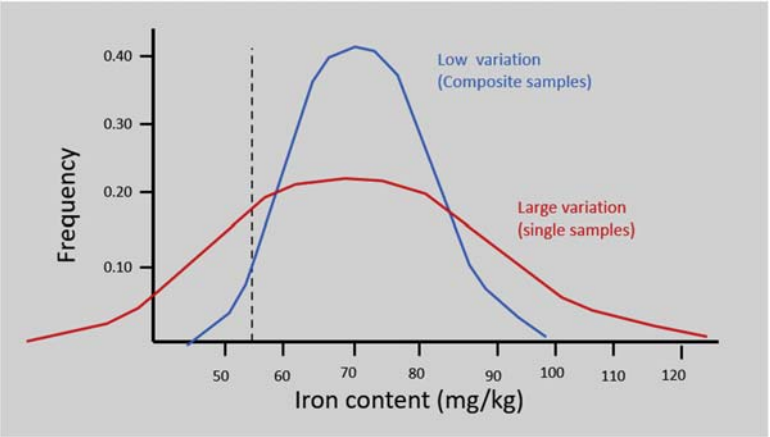


Fig. 3 Hypothetical example of the distribution of results of the iron content in fortified wheat flour analyzing single (27% below minimum reference value when the standard deviation is 25 mg iron/kg) or composite samples (11% below minimum reference value when the standard deviation is 12 mg iron/kg), when the average content is 70 mg iron/kg and the specified "minimum" content is 55 mg iron/kg.

of wheat flour fortification using the Central American technical standard, around ten percent of results may be below the "minimum" value of 55 mg iron/kg, when the average iron content is 70 mg iron/kg, if composite samples are analyzed. However, in single samples, around 27% of samples may be found with values below 55 mg iron/kg. In both cases, the fortification process is fine (Fig. 3).

The specification of "minimum" values instead of average contents in the standards of fortified foods has also caused that, in many countries the micronutrient contents are much lower than expected. Some industries have interpreted the "minimum" values as the target value and, therefore, they add lower amounts of the fortificant.

If the value of compliance is a "minimum" value, large amounts of fortificants are needed in those foods with a large micronutrient content variation (Fig. 4). If the criterion of the "minimum" is enforced, it will increase the cost and negative sensorial reactions in fortified foods, due to high micronutrient content. Further, there is risk that these products may supply excessive quantities of micronutrients to the population.

If the coefficient of variation of the micronutrient content is larger than 30%, based on the analysis of single samples, it would be necessary to prepare composite samples to reduce the variation in the results. If this is not done, the values of the allowable range that responds to the reality of the process will be too large to be useful. Table 2 shows the situation for iodized salt in Mexico. The allowable ranges, either using 80% (i.e., percentiles 10–90) or 90% compliance (i.e., percentiles 5–95), would increase from refined ("vacuum-crystalized") to fine (washed) to coarse types of salt, although all three types of salt are providing sufficient iodine to the population. Therefore, using the same standard for the three types of salt would be inappropriate. In this case, the use of composite samples reduces the variation of the results but is insufficient for "correcting" the situation for the coarse salt samples, whose results may even produce negative values (contents lower than 5 mg iodine/kg) despite that iodization is taking place. Using composite

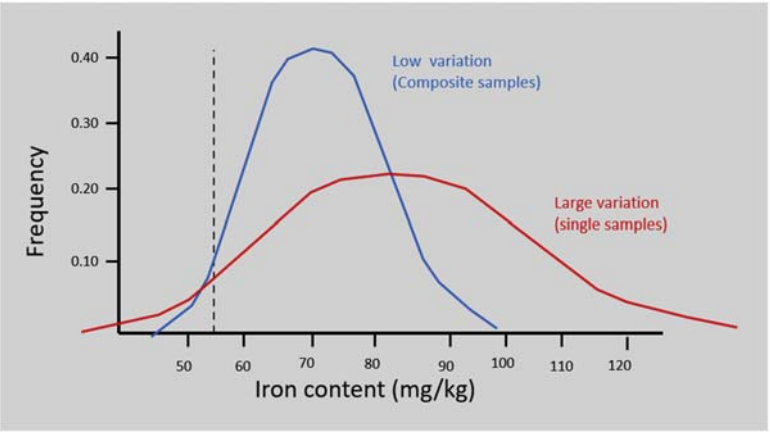


Fig. 4 Hypothetical example of the distribution of results of the iron content in fortified wheat flour if a "minimum" content of 55 mg iron/kg is enforced under two different scenarios of variation. For a compliance that less than 10% samples are below the required "minimum," if single samples (high variation, standard deviation = 25 mg iron/kg) are used, the average content should be 87 mg iron/kg. On the contrary, if composite samples (low variation, standard deviation = 12 mg iron/kg) are used, the average content should be 70 mg iron/kg.

Table 2 Iodine content in three different types of salt in Mexico.[a]

Salt type	*Refined ("vacuum crystalized")*		*Fine ("washed")*		*Coarse*	
Sample type	*Single*	*Compos.*[b]	*Single*	*Compos.*	*Single*	*Compos.*
n	20	4	42	8	32	6
Median (mg I/kg)	33.7	34.0	29.7	29.7	23.8	40.4
Mean (mg I/kg)	33.9	33.9	30.9	30.8	39.9	41.3
S.D. (mg I/kg)	2.9	0.6	10.0	6.5	46.5	31.0
C.V. (%)	7.6%	1.7%	32.4%	20.7%	116.4%	75.0%
% < 20 mg I/kg	0%	0%	16%	6%	42%	37%
% < 5 mg I/kg	0%	0%	0%	0%	23%	12%
Range P_{10}—P_{90}[c]	30.2–37.6	33.1–34.7	18.1–43.7	22.5–39.5	0.0–99.5	1.6–81.0
Range P_5—P_{95}[d]	29.1–38.7	32.9–34.9	14.5–47.3	20.1–41.5	0.0–116.4	0.0–92.3

[a]The Mexican standard for iodized salt specifies a content range of 20-40 mg iodine/kg.
[b]Composite samples prepared by mixing the same amounts of five single samples of salt.
[c]Percentiles 10–90 for determining a compliance range of 80%.
[d]Percentiles 5–95 for determining a compliance range of 90%.
Data provided by the Food Control Institution (COFEPRIS) of the Mexican Government (2013).

samples has the additional advantage of requiring a lower number of analytical tests; in this example one-fifth of the total number as using single samples.

Dietary monitoring

Finally, a food fortification program is not complete if it lacks periodic evaluation of its performance and effectiveness at the consumer level (A2Z-Project, 2008). This means estimating the contribution of consumption of fortified foods to reduce the micronutrient inadequacies in different strata of the population (Engle-Stone et al., 2014; Friesen et al., 2020; Williams et al., 2021). For this purpose, micronutrient intakes should be estimated in the absence and presence of fortification programs. Three types of data are needed: (a) the basal micronutrient intakes from the usual diet, for estimating the magnitude and extension of the micronutrient inadequacies; (b) the customary intake values of the fortification vehicles; and (c) the average of the additional content of micronutrients supplied through the fortified foods. The multiplication of the last two values permits estimating the additional micronutrient intakes associated to the food fortification programs (**Table 3**). Food intakes can be obtained through food intake

Table 3 Current contribution of vitamin A for food fortification in Malawi: %-Harmonized Average Requirement for adult women.[a]

Population strata/socio-economic group	*Oil*[b]		*Sugar*[c]		*Wheat flour products*[d]	
	% Users	*%-HAR median (IQR)*[e]	*% Users*	*%-HAR median (IQR)*	*% Users*	*%-HAR median (IQR)*
Rural (total)	**72**	**20 (8, 39)**	**48**	**36 (24, 53)**	**44**	**1 (0, 3)**
Lowest	44	6 (2, 16)	17	16 (9, 27)	20	0 (0, 1)
Lower middle	66	12 (4, 24)	34	26 (14, 34)	33	0 (0, 1)
Middle	75	18 (6, 31)	47	30 (20,41)	42	1 (0, 1)
Upper middle	83	22 (10, 41)	62	37 (27, 53)	53	1 (0, 2)
Highest	91	39 (22, 67)	80	51 (37, 73)	73	3 (1, 6)
Urban (total)	**96**	**43 (24, 69)**	**92**	**49 (36, 69)**	**86**	**5 (2, 9)**
Lowest	87	18 (8, 29)	76	33 (23, 43)	64	1 (0, 3)
Lower middle	97	31 (18, 47)	92	40 (31,56)	84	3 (1, 5)
Middle	98	45 (31, 63)	97	51 (39, 69)	92	5 (2, 7)
Upper middle	98	53 (37,76)	98	57 (43, 73)	95	7 (4, 10)
Highest	99	81 (55, 127)	97	64 (44, 96)	94	9 (6, 14)
National (total)	**76**	**24 (10, 47)**	**56**	**40 (27, 57)**	**52**	**1 (1, 5)**

[a]Using the Harmonized-Average Requirement (H-AR) vitamin A value (490 μg/day) as recommended by Allen et al. (2020); and intake data estimated as adult female equivalent by Tang et al. (2021).
[b]Vitamin A content in oil at local level: (current: 10 mg/kg).
[c]Vitamin A content in sugar at local level: (current: 7 mg/kg).
[d]Vitamin A content in wheat flour at local level: (current: 0.8 mg/kg).
[e]Median and inter-quartile range in % H-AR/day.
Courtesy of Tang et al. (2021).

Table 4 Reduction of vitamin A inadequacy through fortified foods in Malawi.

	Percent population inadequacy estimated for adult female equivalents		
Population strata/socio-economic group	*Basal*	*Currently*[a]	*Possible*[b]
Rural (total)	**77**	**58**	**46**
Lowest	91	90	86
Lower middle	84	77	65
Middle	78	62	46
Upper middle	70	42	26
Highest	60	18	18
Urban (total)	**67**	**19**	**10**
Lowest	82	56	36
Lower middle	70	24	9
Middle	72	9	2
Upper middle	61	3	1
Highest	48	1	0
National (total)	**75**	**50**	**39**

[a]It was estimated that the current contents of vitamin A in cooking oil, sugar, and wheat flour are 10 mg/kg, 7.0 mg/kg, and 0.8 mg/kg, respectively.
[b]Possible situation if compliance of these fortification programs improves. The corresponding contents of vitamin A would be 21 mg/kg, 10.5 mg/kg, and 1.8 mg/kg, for oil, sugar, and wheat flour, respectively.
Courtesy of Tang et al. (2021).

surveys, or approximating them using secondary analysis of food acquisition data collected in household economic surveys, which are identified under the generic name of "household consumption and expenditure surveys" (Fiedler et al., 2013). Information on the micronutrient content in the fortification vehicles is obtained by analyzing samples of fortified foods collected at local levels (households or local retail stores) and determining if they are fortified or not using qualitative tests and, importantly, measuring the average micronutrient content in composite samples per cluster or subregion, using quantitative tests (Guamuch et al., 2007b).

Table 4 shows that the combination of oil and sugar fortification is needed to provide sufficient vitamin A to certain populations, such as Malawi, as the inadequacy of this vitamin fails to be corrected for all population groups even with these two fortified foods. As expected, the dietary impact of food fortification programs is better in urban than in rural areas. Currently, in Malawi, the role of fortified wheat flour is negligible, as the intake of products made with wheat flour is very low in most population strata of the country. Maize flour, although an important staple in the country, was not included in the analysis because it is processed by hundreds of small mills and, therefore, is not suitable for large scale food fortification.

Often, the performance of food fortification programs has been defined as the percent of compliance with national fortification standards at the household level. This is an erroneous practice as fortification standards are not applicable to the home level because some losses of the micronutrient are expected (Table 5). Further, measuring the micronutrient contents in single samples (generally without following the recommendations of the analytical tests) unreliable results are produced. At the consumer level, it is important to estimate the average additional supply of micronutrients through the fortified foods and not the percent of samples that comply with standards. Table 4 shows that the fortification programs can have epidemiological impact even though the fortified foods do not have the micronutrient contents specified in the standards.

The final proof of impact of food fortification programs is improvement in biomarkers associated with micronutrient status. These assessments are not specific to food fortification programs and are preferable to be done in combination with the evaluation of other nutritional and health interventions, through well designed and executed nutritional surveillance programs (Williams et al., 2021).

Table 5 Prediction of vitamin A contents at households based on the added contents at factories and the estimated stability of the nutrient to homes—example of Malawi.

Food	*Added content (mg/kg)*	*Stability from factory to homes (%)*	*Predicted micronutrient content at homes (mg/kg)*
Vegetable oil	30.0	70	21.0
Sugar	15.0	70	10.5
Wheat flour	2.0	90	1.8
Maize flour	0.5	80	0.4

Conclusions

Food fortification, especially large-scale food fortification, may be an effective and low-cost intervention to reduce micronutrient inadequacies, therefore contributing to prevent their deficiencies, if appropriately designed and implemented. The design of the programs depends on the selection of fortification vehicles that are routinely consumed by the vulnerable populations and manufactured by formal and centralized food industries. As micronutrient content may be limited because of safety, technological or costs constraints, fortification vehicles should be consumed in reasonable amounts by the target population, and micronutrient content set as low as possible to fill dietary nutrient gaps. As most fortification vehicles are foods whose consumption should be reduced to prevent non-communicable diseases, fortification should not be used for promoting the consumption of those foods. Correct implementation of food fortification programs requires the introduction of systems to check processing (quality assurance by the industry and auditing by the government), complemented with practical, simple, and low-cost measures of compliance (quality control by the industry and inspection by the government). These should be based on the use of qualitative and quantitative methods, the latter applied to composite samples following detailed procedurs and instructions for interpreting results. Finally, these programs must include an epidemiological surveillance component to periodically assess population coverage and dietary contributions which, in turn, will help for scaling-up, sustainability and recognition of food fortification as an important intervention in public health nutrition.

References

A2Z-Project, 2008. Monitoring and Evaluating Food Fortification Programs: General Overview. A2Z-Project, Washington, DC, p. 19. http://www.a2zproject.org/pdf/Food-Fortification-M&E-July-2006-Consultation-AED-FINAL.pdf. (Accessed 5 July 2021).

Alavi, S., Bugusu, B., Cramer, G., et al., 2008. Rice Fortification in Developing Countries: A Critical Review of the Technical and Economic Feasibility. Project A2Z/USAID/AED/IFT, p. 74. http://www.a2zproject.org/pdf/Food-Rice-Fortification-Report-with-Annexes-FINAL.pdf. (Accessed 4 July 2021).

Allen, A., de Benoist, B., Dary, O., Hurrell, R. (Eds.), 2006. Guidelines on Food Fortification With Micronutrients. World Health Organization, Geneva, p. 341. http://www.who.int/nutrition/publications/micronutrients/9241594012/en/index.html. (Accessed 4 July 2021).

Allen, L.H., Carriquiry, A.L., Murphy, S.P., 2020. Perspective: proposed harmonized nutrient reference values for populations. Adv. Nutr. 11 (3), 469–483, 1. https://pubmed.ncbi.nlm.nih.gov/31701998/.

Codex Alimentarius, 2015. General Principles for the Addition of Essential Nutrients to Foods. CAC/GL 09-1987 (amended 1989, 1991. Revision: 2015). http://www.fao.org/fao-who-codexalimentarius/sh-proxy/en/?lnk=1&url=https%253A%252F%252Fworkspace.fao.org%252Fsites%252Fcodex%252FStandards%252FCXG%2B9-1987%252FCXG_009e_2015.pdf. (Accessed 4 July 2021).

Dary, O., Mora, J.O., 2002. Food fortification to reduce vitamin A deficiency: International Vitamin A Consultative Group recommendations. J. Nutr. 132 (9 Suppl. I), 2927S–2933S. https://doi.org/10.1093/jn/132.9.2927S.

Dary, O., 2007. The importance and limitations of food fortification for the management of nutritional anemias. In: Kraemer, K., Zimmermann, M.B. (Eds.), Nutritional Anemia. Sigth & Life Press, Basel, pp. 335–336. http://www.ernaehrungsdenkwerkstatt.de/fileadmin/user_upload/EDWText/TextElemente/Ernaehrungswissenschaft/Naehrstoffe/nutritional_anemia_book.pdf. (Accessed 18 July 2021).

Engle-Stone, R., Nankap, M., Ndjebayi, A.O., et al., 2014. Simulations based on representative 24-h recall data predict region-specific differences in adequacy of vitamin A intake among Cameroonian women and young children following large-scale fortification of vegetable oil and other potential food vehicles. J. Nutr. 144, 1826–1834. https://doi.org/10.3945/jn.114.195354. (Accessed 4 July 2021).

Fiedler, J.L., Martin-Prével, Y., Moursi, M., 2013. Relative costs of 24-hour recall and household consumption and expenditures surveys for nutrition analysis. Food Nutr. Bull. 34, 318–330. https://doi.org/10.1177/156482651303400304.

Field, M.S., Mithra, P., Estevez, D., Peña-Rosas, J.P., 2020. Wheat flour fortification with iron for reducing anaemia and improving iron status in populations. Cochrane Database Syst. Rev. (7). https://doi.org/10.1002/14651858.CD011302.pub2. Art. No.: CD011302.

Friesen, V.M., Mbuya, M.N.N., Aaron, G.J., et al., 2020. Fortified foods are major contributors to apparent intakes of vitamin A and iodine, but not iron, in diets of women of reproductive age in 4 African countries. J. Nutr. 150, 1283–2190. https://doi.org/10.1093/jn/nxaa167.

Gonzalez de Mejia, E.G., Aguilera-Gutierrez, Y., Martin-Cabrejas, M.A., Mejia, L.A., 2015. Industrial processing of condiments and seasonings and its implications for micronutrient fortification. Ann. N. Y. Acad. Sci. 1357, 8–28. https://doi.org/10.1111/nyas.12869.

Guamuch, M., Makhumula, P., Dary, O., 2007a. Manuals for Internal, External, and Commercial Monitoring of Fortified Foods. (Several Numbers Per Fortification Vehicle). http://www.a2zproject.org/node/74. (Accessed 5 July 2021).

Guamuch, M., Makhumula, P., Dary, O., 2007b. Manual of Laboratory Methods for Fortified Foods. East, Central, and Southern Health Community. (Several numbers). http://www.a2zproject.org/node/74. (Accessed 5 July 2021).

Hurrell, R.F., 2021. Iron fortification practices and implications for iron addition to salt. J. Nutr. 151 (Suppl. 1), 3S–14S. https://doi.org/10.1093/jn/nxaa175.

Institute of Medicine, 1998. Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline. National Academy Press, Washington, DC, p. 464. https://pubmed.ncbi.nlm.nih.gov/23193625/.

Keats, E.C., Neufeld, L.M., Garrett, G.S., Mbuya, M.N.N., Bhutta, Z.A., 2019. Improved micronutrient status and health outcomes in low- and middle-income countries following large-scale fortification: evidence from a systematic review and meta-analysis. Am. J. Clin. Nutr. 109, 1696–1708. https://doi.org/10.1093/ajcn/nqz023.

Martorell, R., Ascencio, M., Tacsan, L., et al., 2015. Effectiveness evaluation of the food fortification program of Costa Rica: impact on anemia prevalence and hemoglobin concentrations in women and children. Am. J. Clin. Nutr. 101, 210–217. https://doi.org/10.3945/ajcn.114.097709.

Mora, J.O., Dary, O., Chinchilla, D., Arroyave, G., 2000. Vitamin A: Sugar Fortification in Central America. Experience and Lessons Learned. MOST/The USAID Micronutrient Project/INCAP, p. 49. http://www.a2zproject.org/~a2zorg/pdf/sugarlessonsEnglish.pdf. (Accessed 4 July 2021).

Pachon, H., 2021. Wheat Flour Fortification and Human Health. Burleigh Dodds, Science Publishing, p. 67. https://doi.org/10.19103/AS.2021.0087.23.

Rai, A., Maharjan, M.R., Fry, H.A.H., et al., 2019. Consumption of rice, acceptability and sensory qualities of fortified rice amongst consumers of social safety net rice in Nepal. PLoS One 14 (1–17), e0222903. https://doi.org/10.1371/journal.pone.0222903.

Richardson, D.P., 1990. Food fortification. Proc. Nutr. Soc. 49, 39–50. https://doi.org/10.1079/pns19900007.

Rowe, L.A., 2020. Addressing the fortification quality gap: a proposed way forward. Nutrients 12, 3899. https://doi.org/10.3390/nu12123899.

Selhub, J., 2008. Public health significance of elevated homocysteine. Food Nutr. Bull. 29 (Suppl. 2), S116–S125. https://doi.org/10.1177/15648265080292S116.

Tang, K., Adams, K.P., Ferguson, E.L., Woldt, M., Kalimbira, A.A., et al., 2021. Modeling Food Fortification Contributions to Micronutrient Requirements in Malawi Using Household Consumption and Expenditure Surveys. Annales of the New York Academy of Science. https://doi.org/10.1111/nyas.14697.

Van Lieshout, M., West, C.E., van Breemen, R.B., 2003. Isotopic tracer techniques for studying the bioavailability and bioefficacy of dietary carotenoids, particularly beta-carotene, in humans: a review. Am. J. Clin. Nutr. 77, 12–28. https://doi.org/10.1093/ajcn/77.1.12.

WHO, 2011. Guideline: Use of Multiple Micronutrient Powders for Home Fortification of Foods Consumed by Infants and Children 6–23 Months of Age. World Health Organization, Geneva, p. 25. http://apps.who.int/iris/bitstream/handle/10665/44651/9789241502047_eng.pdf. (Accessed 18 July 2021).

Williams, A.M., Tanumihardjo, S.A., Rhodes, E.C., et al., 2021. Vitamin A deficiency has declined in Malawi, but with evidence of elevated vitamin A in children. Am. J. Clin. Nutr. 113, 854–864.

Further reading

Dary, O., Hainsworth, M., 2008. The Food Fortification Formulator: Technical Determination of Fortification Levels and Standards for Mass Fortification. Project A2Z/USAID/AED. p. 29. Plus spreadsheet of calculations. http://www.a2zproject.org/pdf/Food-Fortification-Formulator.pdf. (Accessed 4 July 2021).

Pan American Health Organization, Nutrition Unit, 2005. Code of Practice: Food Premix Operations. FCH/NU/66-16/04. Pan American Health Organization/World Health Organization, Washington, DC, p. 31. https://iris.paho.org/bitstream/handle/10665.2/37201/COPPremixOperations.pdf?sequence=1&isAllowed=y. (Accessed 5 July 2021).

USAID/OMNI/MOST, DSM, 1998–2005. Fortification Basics [A Collection of 9 Briefs That Describe Technologies for Food Fortification of 6 Staple Foods and Three Cross-Cutting Subjects on Food Fortification: Choosing a Vehicle, Assays, and Stability]. http://www.a2zproject.org/~a2zorg/node/103. (Accessed 4 July 2021).

Relevant websites

A2Z Project, http://www.a2zproject.org/—A2Z Project: A2Z was an USAID Project in micronutrients (2005–2011), and the website contains publications, reports, manuals, presentations, and similar material about micronutrients, including food fortification. The site also presents similar documents from the predecessor project MOST (1999–2005).

CDC/IMMPaCt Project, http://www.cdc.gov/immpact/index.html—CDC/IMMPaCt: The site of the International Micronutrient Malnutrition Prevention and Control (IMMPaCt) Program of the CDC (USA) covers a wide range of information about micronutrient biology and interventions, including food fortification.

Alimentarius, http://www.fao.org/fao-who-codexalimentarius/en/—Codex Alimentarius: This site contains guidelines, general principles, manuals, standard models, and similar materials for food safety and trade, including recommendations for food fortification, health and nutrient claims, nutritional panel, and labeling.

Fortification Initiative, https://www.ffinetwork.org/—The Food Fortification Initiative (FFI) is a public–private network of organizations dedicated to promote the use of fortified foods, especially rice and cereal flours, worldwide to improve the status of several micronutrients. The website contains documents, information, and tools associated with this area.

GAIN, http://www.gainhealth.org/—Global Alliance for Improved Nutrition: This site contains news, reports, resources, and announcements about nutrition interventions, mainly implemented under public–private partnerships, including food fortification.

Fortification Data Exchange, https://fortificationdata.org/—An inter-institutional project aimed to collect information about food fortification programs and to offer tools for analysis and visualization of the data.

Global Network, https://www.ign.org/index.cfm—International NGO with hundreds of members in most countries specifically dedicated to maintain iodine deficiency under control mainly through iodization of the salt.

Forum, https://micronutrientforum.org/—International NGO supported by several institutions with the aim to advance in the science, policies and programs to reduce micronutrient deficiencies. Among the strategies is food fortification.

International, https://www.nutritionintl.org/—Nutrition International (prior the Micronutrient Initiative, MI) is an international technical NGO based in Ottawa, Canada, with offices in Asia and Africa, and programs in several developing countries; its focus is to improve micronutrient status in vulnerable populations, and food fortification has been one of the main interventions.

Food safety heavy metals and metalloids

Fernando Gil and Pablo Olmedo, Facultad de Medicina (PTS), Dpto. Medicina Legal y Toxicología, Universidad de Granada, Granada, Spain

© 2023 Elsevier Ltd. All rights reserved.

This is an update of L.H. Allen, Food Safety: Heavy Metals, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 331–336, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00126-4.

Introduction 352
General aspects 352
Lead 353
Dietary exposure 353
Permissible intakes 354
Toxicokinetics 354
Toxicity 355
Mercury 356
Dietary exposure 356
Permissible intakes 356
Toxicokinetics 356
Toxicity 356
Cadmium 357
Dietary exposure 357
Permissible intakes 357
Toxicokinetics 357
Toxicity 357
Nickel 358
Dietary exposure 358
Permissible intakes 358
Toxicokinetics 358
Toxicicty 358
Tin 358
Dietary exposure 358
Permissible intakes 359
Toxicokinetics 359
Toxicity 359
Copper 359
Dietary exposure 359
Permissible intakes 359
Toxicokinetics 359
Toxicity 359
Arsenic 359
Dietary exposure 359
Permissible intakes 360
Toxicokinetics 360
Toxicicty 360
Conclusion 360
References 361
Relevant websites 361

Key points

- To present various aspects of interest in relation to contamination by metallic elements in foodstuffs.
- To highlight the importance of the mechanism of toxic action of certain metallic elements and the consequences of interaction with target organs.

- To know the toxicity on humans of the main heavy metals and metalloids of interest in Food Toxicology, especially in chronic intoxication.
- To know that the International Agency for Research on Cancer (IARC) has classified certain metals and metalloids (Cd, Cr, Ni and As) as carcinogenic to humans.

Glossary

Acceptable Daily Intake (ADI)
Arsenic
BMD (Benchmark Dose)
Cadmium
Copper
Delta-aminolevulinic acid dehydratase (ALA-D)
Heavy metals
Lead
Mercury
Metal
Metalloid
Nickel
Provisional Tolerable Weekly Intake (PTWI)
Tin

Introduction

Food toxicology is the branch of toxicology that specifically deals with toxicity (acute and long-term) problems arising from the presence of toxic substances in food. The great tragedies of toxic origin over the last 100 years have been linked to food contamination. In fact, food contamination by toxic agents is one of the problems of greatest concern to public health authorities. Environmental pollution represents one of the current major problems in food contamination. Pollution of air, soil, inland and marine waters is mainly a consequence of human activity (industrial, agricultural, domestic) which, indirectly, allows the incorporation of pollutants into plants and animals. The food products coming from these polluted organisms can cause harmful effects to the health of the consumers. The main xenobiotics from environmental pollution include metals and metalloids such as lead, mercury, arsenic, cadmium, tin and copper.

General aspects

The release of heavy metals from emissions, toxic discharges, etc., is one of the main consequences of industrial pollution. Although the concentration in the environment is usually not very high, the enormous power of accumulation and environmental persistence, together with the high toxicity of these compounds, determines the possibility of generating important toxicological problems. Therefore, their study in food toxicology is essential.

Environmental contamination by metals is very diverse and comes mainly from industrial emissions, agricultural effluents, fungicides and biocides, etc. As a result of the wide variety of anthropogenic activities, the concentration of these pollutants in ecosystems has increased dramatically in recent decades. Due to the intrinsic toxicity of metals, their presence in the environment constitutes a potential threat to terrestrial and aquatic biota, negatively affecting the different species of biotic communities, including the human species (Gómez et al., 2007).

A key aspect in the toxicity of metal ions is the dose. In some cases, they act as essential constituents of living organisms, and their deficiency may lead to health problems. For example, a decrease in fluoride promotes dental cariogenic processes. Also, when selenium is deficient, the enzymatic defense systems against oxidative stress are less efficient. This sometimes determines their use as dietary supplements. However, some of them (fluorine, copper, zinc, selenium, etc.) are also capable of causing alterations when they are present in excess. In fact, there is a narrow margin between what would be a toxic dose and nutritional requirements. On the other hand, other elements (e.g., arsenic, lead, mercury or cadmium) are typically toxic as they have not been attributed an essential role in the organism so far. In this regard, dietary habits can significantly influence exposure to any food contaminant, including heavy metals (Gil et al., 2017).

Another interesting aspect is the chemical form. The metallic (elemental) form is usually non-toxic, with a few exceptions (i.e., mercury vapor), whereas the ionic forms (organic and inorganic salts) are quite toxic depending on the metal. Due to the variation in toxicity of the different metal forms, the determination of the total metal concentration (i.e., all ionic forms) is not always useful. Therefore, speciation analyses are important as they are able to separate and detect each of the metal forms (Gil et al., 2017).

Moreover, it is possible to establish a number of general characteristics in relation to the toxic effects of heavy metals and metalloids. These include the following (Pla et al., 2019):

- They tend to accumulate in the body, with a high affinity for -SH (thiol or sulfhydryl) groups of enzymes (e.g., arsenic blocks the pyruvate dehydrogenase complex). This selective binding by chemical affinity may mean that even at relatively low levels of exposure, they accumulate in the body to produce the delayed toxic effect.
- Their toxicity increases with the degree of oxidation.
- Numerous factors can modify the toxicity of metals. These include possible pathological alterations affecting the kidney (essential in the excretion of metals), increased or decreased permeability of the intestinal mucosa, interaction with other nutrients in the diet (e.g., calcium competes with lead; vitamin D promotes intestinal absorption of lead, etc.), fasting situations, age, among others.

Due to their cumulative properties and toxicity, heavy metal concentrations could reach levels that are potentially harmful to human health. Cadmium (Cd), mercury (Hg), lead (Pb) and arsenic (As) are non-essential toxic elements of particular concern because of their toxicity even at low concentrations (Hsu et al., 2006; Rao et al., 2011).

On the other hand, the level of metal accumulation in plants is influenced by numerous physicochemical properties of the soil (sediment characteristics, pH, dispersion range and presence or absence of other elements), which are determinants in the bioavailability of these xenobiotics (Sarma et al., 2011; Hernández and Gil, 2016). Furthermore, although plants readily assimilate metals through the roots, they can also do so through the leaves (Łozak et al., 2002). Cooper et al. (2007) found that certain Chinese medicinal preparations were highly contaminated with As, Pb and Hg, which theoretically would imply a serious risk to consumers. However, the risk assessment of the exposure to metals in food should take into account not only the level of the metal but also its bioavailability in the body.

The International Agency for Research on Cancer (IARC) has classified certain metals and metalloids (As, Cd, Cr, and Ni) as carcinogenic to humans due to their ability to induce DNA damage. Thus, it is important to analyze their presence in food in order to avoid excessive human impregnation. There is growing concern about the health risks of the exposure to metallic and metalloid elements, especially arsenic, because of its carcinogenic and neurotoxic potential, and its possible effects on the developing brain of children (Sarma et al., 2011; Hernández and Gil, 2016).

In 2020, the US Food Drug Administration (US FDA) released Closer to Zero, the agency's action plan for reducing toxic elements (metals) in the foods eaten by babies and young children (Woodcock and Mayne, 2021).

Lead

Dietary exposure

Lead is a ubiquitous metal and its use in kitchen utensils dates back to Roman times. The Romans treated these utensils (usually of copper) with alloys containing lead. It was also used in the preservation of wines. Lead poisoning was even described in the United States as a result of drinking whisky distilled in automobile radiators that contained lead.

Probably one of the main problems with lead is that it is very easily dissolved by weak acids (carbonic), organic acids (citric, acetic, tartaric, etc.) and fatty acids released from rancidity processes.

Numerous sources of exposure, which until recently caused food contamination problems, have now been eliminated. The main sources (Gil et al., 2017; Villanueva and Gil, 2018a) include (**Table 1**):

- Canning tins, especially where pickled foods or fruit juices were stored, which had a lead-containing solder, which could solubilize on contact with acetic acid. At present, the cans have an internal coating, usually chrome, or a special lacquering, and the soldering has been eliminated.
- Earthenware and ceramic utensils that are colored and/or vitrified with lead glazes (based on lead sulfide).
- Leaded glass vessels.
- Water pipes, which were formerly made of lead. They have now been replaced by Polichloride vinyl (PVC), which is an inert material. Thus, hard water, coming from granitic soils, could drag lead from the pipes.
- Tetraethyl lead has been used until recently as an antiknock agent in "leaded" gasoline. This has been responsible for the pollution caused by deposition of lead on leaves and fruits around motorways and busy roads. As a result of the harmful consequences of environmental lead emissions, the US Environmental Protection Agency (US EPA) banned this fuel in 1996. Another consequence of the ban on the use of leaded gasoline was the drastic decrease in lead levels in general and in foods in particular, especially vegetables and, to a lesser extent, cereals, tubers and fruits (Falcó et al., 2006).

Table 1 Heavy metal and metalloid dietary exposure and toxicity.

Heavy metal	*Dietary exposure*	*Toxicity*
Lead	Canned food, leaded glass containers or ceramic food utensils, water, wine treated with pesticides. Migration is higher in the case of foods in contact with weak acids	Anemia (blockage of the hemoglobin synthesis pathway), neurocognitive impairment, intellectual deficit, encephalopathy, irritability
Mercury	Seafood products (biotic transformations to methylmercury- e.g., tuna fish-), food treated with organomercury fungicides, nuts (walnuts, almonds), some fresh fruits (raspberries)	Organic: cerebellar ataxia, paresthesias in hands and mouth, dysarthria, personality disorders, mental confusion, neurodevelopmental impairment, teratogenic effects. Inorganic: renal alterations (tubular necrosis and renal failure) and local irritant action on gastrointestinal tract (acute gastroenteritis and colitis)
Cadmium	Food contaminated with residues derived from pigment, battery industries or phosphate fertilizers, offal (liver, kidneys, lung), crabs and bivalves (oysters and scallops), vegetables (spinach, lettuce, soybeans, cereal -rice- and peanuts)	Renal dysfunction (tubular interstitial nephritis with hypercalciuria and proteinuria), osteomalacia, carcinogenic (prostate, testicular and lung)
Nickel	Vegetables (including mushrooms, spices and condiments, legumes, spinach, lettuce, nuts and oilseeds), baking powder and cocoa powder. Migration is higher in the case of foods in contact with weak acids	Carcinogenic in humans. Oral exposure does not usually cause dermatological reactions
Tin	Canned foods, cereal grains, dairy, meat and vegetables. To interfere with the metabolism of certain trace elements.	There is no evidence that cause cancer in humans
Copper	Seafood products (oysters), sheep, fruits, cereals, nuts and green vegetables, wine treated with fungicides	Irritant action on gastrointestinal tract (nausea, epigastric burning and diarrhea) and hepatotoxicity (hepatomegaly and jaundice)
Metalloid	*Dietary exposure*	*Toxicity*
Arsenic	Seafood products (crustaceans and shellfish), rice, food treated with pesticides (arsenate)	Carcinogenic in humans (skin), sensorimotor polyneuropathy and dermatological alterations (palmoplantar hyperkeratosis, mees bands on the nails)

- Wine, especially rosé and white wine, with a more acidity. The origin is multiple: insecticides treated with lead arsenate or from the leaded foil that covers the corks of bottles; when the cork dries out and retracts in bottles that have been stored horizontally for a certain period of time, the alcohol oxidizes to acetic acid, which can solubilize the foil, forming lead acetate.

Permissible intakes

The daily intake of lead is around 20–25 μg/g and the maximum contribution through the diet comes in general from cereals and in particular from bread (Falcó et al., 2006).

The average lead in food is estimated to be 0.05–0.2 μg/g and in water 10–20 μg/L. FAO/WHO expert committee recommended that the weekly intake per person (provisional tolerable weekly intake, PTWI) of lead should not exceed 25 μg/kg body weight (1500 μg/person weighting 60 kg). Although EPA established an action level of 0.015 mg/L (15 ppb or μg/L) for Pb, more recently EPA has set the maximum contaminant level goal for lead in drinking water at zero because lead can be harmful even at low exposure levels (Gil et al., 2017).

With regard to lead, which may be released from certain ceramic utensils, the International Organization for Standardization has established maximum limits of 1.7 mg/dm^2 for flat wares and 2.5–5.0 mg/L in solution for the concave ones.

The limits in foodstuffs are summarize in Regulation (EU) No 420/2011 of April 29, 2011, which amends points 3.1.6, 3.1.9, 3.1.10 and 3.1.11 of Regulation (EC) No 1881/2006.

Toxicokinetics

Absorption of lead from the digestive tract is highly variable and relatively low (10–20%). Pregnant women and children (who are more susceptible to the toxic effects of lead) may absorb a higher amount and this depends on the nutritional status (levels of calcium, phosphate or iron can influence their uptake). Once absorbed, it is distributed in blood bound to erythrocytes (>95%), most likely due to its high affinity for delta-amino levulinic acid dehydratase, leaving a small fraction in plasma. It accumulates in the epiphysis of long bones (e.g., femur, tibia and humerus), being eliminated mainly through urine (75%), to a lesser extent through feces and also through nails and hair.

Toxicity

Lead in food is mainly responsible for chronic intoxications, given its cumulative character. It follows the metabolism of calcium, hence it can be stored in the form of deposits, especially in bone tissue, in epiphyseal areas, and can be mobilized in certain situations of metabolic calcium turnover.

A limit concentration without biological effect is established at 35 μg/100 mL in blood. However, in children, due to a higher permeability of the blood-brain barrier, it has been shown that with much lower levels (even below 10 μg/100 mL), neurocognitive alterations together with intellectual deficit may appear (**Table 1**). Classically, four main effects can be distinguished (Villanueva and Gil, 2018a):

(a) Hematological effects. There are two types: metabolic and morphological.

Metabolic effects are due to the inhibition of various enzymes involved in the biosynthesis of the heme group (see **Fig. 1**):

- Delta-aminolevulinic acid dehydratase (catalyzes the conversion of delta-aminolevulinic acid to porphobilinogen). It is the most specific and selective, and can be used as a biomarker of response or effect to lead. Under normal conditions it should be higher than 100 U/10^6 erythrocytes (Gil and Hernández, 2009).
- Coproporphyrinogen oxidase (catalyzes the passage from coproporphyrinogen III to protoporphyrinogen).
- Heme synthetase or ferrochelatase (leads to the formation of the heme group, by incorporating a Fe^{2+} atom into the tetrapyrrole nucleus of the protoporphyrin IX or zinc-protoporphyrin).

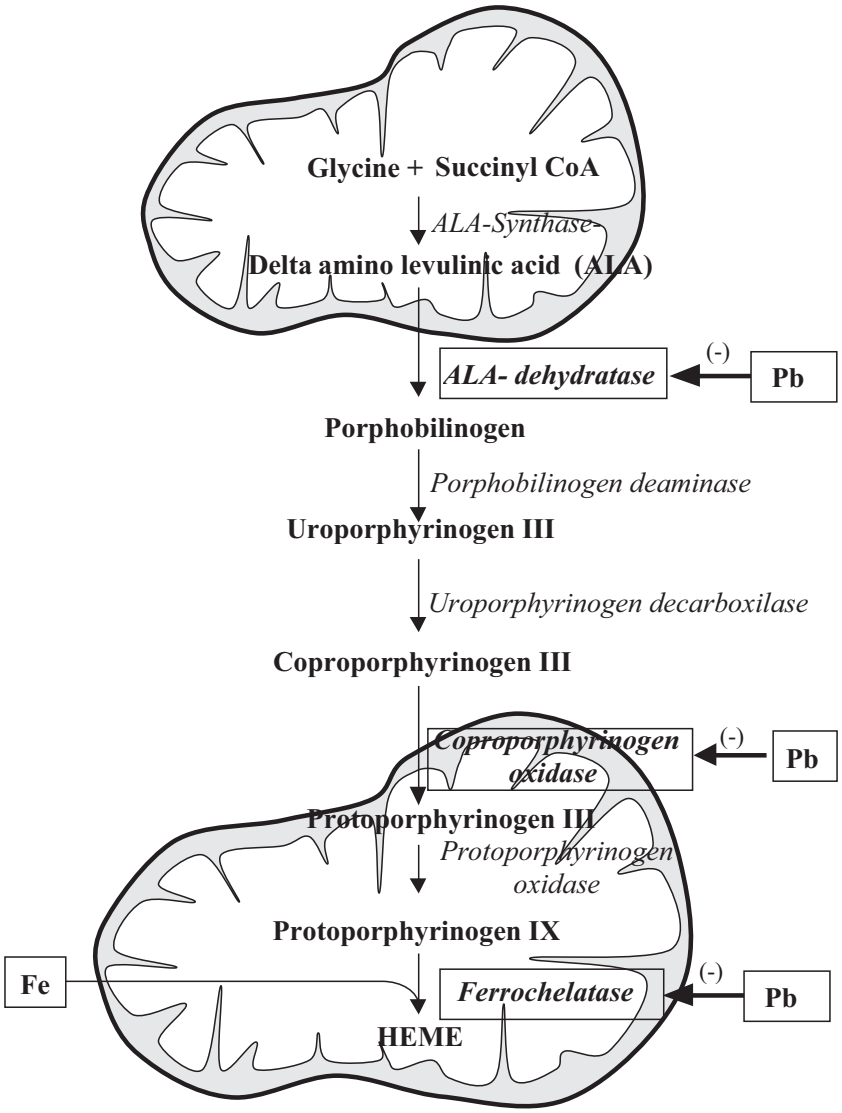


Fig. 1 Heme group synthesis in bone marrow erythroblasts and enzymatic interference (inhibition) by lead.

The morphological effects result in the appearance of red blood cells with basophilic granulations (consisting of non-heme iron, ribosomes and RNA).

The most important hematological consequence is a significant decrease in hemoglobin and anemia (an earthy skin tone).

(b) Effects on the central nervous system. The occurrence of encephalopathy is possible, especially in children (Pb can cross the placental barrier and interfere with placental transport systems, thereby increasing prenatal exposure to these toxic compounds), at levels of 45–50 μg/100 mL in whole blood and above. It is characterized by headaches, irritability, memory loss, etc. The most important consequences are epilepsy and hydrocephalus. Reduced cognitive development and learning problems have been described at relatively low doses.

(c) Effects on the peripheral nervous system. The most important alterations are revealed in the electromyogram. Motor conduction slowing, fibrillations, decrease in the number of motor units involved in the maximal contraction, etc. may appear. Sensory disturbances may also be observed. In severe cases, there is a neuritis consisting of a radial paralysis affecting the extension of the fingers excepting the supinator longus, which means that the middle and ring finger cannot extend.

(d) Other effects. At present, it is very difficult to develop chronic end-stage saturnism with glomerular involvement and gout. During the saturnism phase, the so-called saturnine colic is accompanied by intense and diffuse pain, vomiting, diarrhea that progresses to persistent constipation and significant general involvement. However, such clinical syndrome, which was relatively common in the workplace, would be an exception in food toxicology, since exposure levels are relatively low. Lead has also been described as having the capacity to cause chromosomal alterations, especially at blood levels above 65 μg/100 mL. Pb exposure has also been associated with renal tumors and increased blood pressure, and cardiovascular disease probably due to an alteration in natriuretic peptide (Goyer and Clarkson, 2001; Garcia-Rico et al., 2007).

Mercury

Dietary exposure

Mercury has probably caused the greatest food-borne problems. Several poisonings have occurred throughout history, most notably in the Minamata Bay (1953) in Japan, due to the consumption of fish contaminated with methylmercury. Also, there was a massive poisoning in Iraq (1972) due to the consumption of bread made from wheat treated with organomercury fungicides. Mercury contamination has also been described in poultry and eggs due to contamination of the fishmeal used in their feed (Villanueva and Gil, 2018b).

The most important sources of food poisoning are the treatment of seeds with organomercurial fungicides and industrial discharges into inland or marine waters (**Table 1**). In general, discharges are in inorganic form; however, fish (and particularly tuna, sharks, swordfish, etc.) are capable of methylating it (biotic transformations) thanks to the micro-organisms present in their gills or those in the marine plankton (Gil and Gil, 2015). As a result of this process, it is obtained mainly monomethylmercury and dimethylmercury (10 and 1500 μg/kg have been found in some fish). On the other hand, it should not be forgotten that the organic forms are much more toxic to humans. Therefore, the main source of contamination of foodstuffs (>90%) is organic mercury and, commonly derives from fishery products (Olmedo et al., 2013).

Plant contamination through roots is usually minimal. In general, foods of vegetal origin are not contaminated. However, significant contributions have also been described in nuts (walnuts, almonds, etc.) and in some fresh fruits (raspberries, etc.). Some mushroom species can have considerable levels of mercury.

Permissible intakes

The FAO/WHO Expert Committee on Food Additives sets a provisional tolerable weekly intake level (PTWI) of 0.3 mg of total mercury per person, of which, as a maximum, no more than 0.2 mg should be present in the form of methylmercury. These amounts are equivalent to 0.005 and 0.0033 mg respectively per kilogram of body weight (300 and 198 μg/person weighting 60 kg, respectively). When the total mercury intake through food exceeds 0.3 mg per week, the concentration of methylmercury compounds should also be investigated (Gil et al., 2017).

Mercury in hair can be used as a biomarker of chronic exposure due to the metal's affinity for keratin-rich structures.

Toxicokinetics

Organic mercury compounds are very well absorbed by the digestive tract (95%), but inorganic and metallic mercury (7%) are not. Organic forms and metallic mercury are distributed to the central nervous system (it cross the blood-brain barrier) thanks to their high liposolubility. They are selectively accumulate in the hippocampus, cerebellar gray matter and calcarine cortex. On the contrary, inorganic mercury, which is not liposoluble, is fixed to the kidney (Villanueva and Gil, 2018b).

Toxicity

Intoxication by organic mercury compounds is characterized by a progressive neurological and encephalic syndrome. Clinical syndrome is essentially cerebellar (methylmercury is distributed in the Purkinje and Golgi cells of the cerebellum as well as in three

different layers of cerebral cortical cells) with paresthesia in the hands and around the mouth, cerebellar ataxia, dysarthria, personality disorders and mental confusion (Table 1). Cortical signs also appear, leading to a progressive loss of visual field and hearing acuity. Finally, it should not be overlooked the teratogenic power of organic mercury, since it is capable of crossing the placental barrier, as it was shown in the Minamata poisoning. Therefore, swordfish and tuna are fishery products that should be avoided during pregnancy, to minimize the risk of high mercury intake. In fact, alterations observed following in utero exposure to organic mercury compounds during the Minamata and Iraq poisonings included increased percentage of spontaneous abortions, low weight at birth, brain and cerebellum malformations, abnormal neuronal migration pattern and neurodevelopmental impairment.

In the later stages of chronic poisoning, there is a possibility of renal damage.

On the other hand, poisoning by inorganic compounds is characterized by renal tubular necrosis with tubular obstruction and preglomerular vasoconstriction, leading to renal failure. It is usually accompanied by acute gastroenteritis, stomatitis and ulcerative hemorrhagic colitis as a consequence of the local irritant action of these compounds (Goyer and Clarkson, 2001; Garcia-Rico et al., 2007; Gil, 2018).

Cadmium

Dietary exposure

The presence of cadmium in the environment is due to anthropogenic pollution resulting from its industrial uses. In industry, this element is used as a pigment in the manufacture of paints, in foundries, in battery and semiconductor factories, as a waste product of the electrolysis process leading to the cadmium plating of metals, as a covering of other metals, in various fertilizers and in stabilizing additives for plastics. Another possible source of contamination comes from the use of colored pottery; some glazes, especially yellow glazes, can release cadmium when in contact with acid foods (Gil et al., 2017).

One of the most important characteristics of cadmium is its easy transfer from soil to plants, in which it differs markedly from lead. This fact determines that plants which have been fertilized with phosphate contain significant amounts of cadmium, as it is found as an impurity in these fertilizers. As a consequence, this makes exposure to Cd through the food-chain contamination a public health concern.

With regard to food of animal origin, it does not usually contain very high levels of cadmium, with the exception of offal (liver, kidneys, lung, etc.) (Table 1). Crabs tend to have the highest concentrations, and bivalves (oysters and scallops) have the capacity to bioaccumulate this element thanks to their filtration power. Seafood, such as mussels and oysters, can be a major source of cadmium in the diet, containing between 100 and 1000 μg/kg. Cadmium accumulates in certain vegetables (e.g., spinach, Romaine lettuce, soybeans, cereal crops and peanuts) due to the use of fertilizers and irrigation water containing cadmium. Fruits contain approximately 1–50 μg/kg and cereals 10–150 μg/kg. Drinks contain relatively low levels, as well as drinking water (usually less than 4 μg/L) (Satarug et al., 2017).

Permissible intakes

The Joint FAO/FAO Expert Committee established 7 μg/kg body weight per week (420 μg/person weighing 60 kg) as provisional tolerable weekly intake (PTWI). At the present moment, tolerable intake is 25 mg per kg body weight per month (0.83 mg/kg body weight per day or 58 mg/day for a 70 kg person) (Satarug et al., 2017). The maximum concentration for water proposed from US FDA is 5 μg/L. Regulation No 488/2014 of May 12, 2014 amends Regulation (EC) No 1881/2006 of subsection 3.2 of Regulation (EC) No 1881/2006 and includes the maximum permissible levels (expressed as mg/kg fresh weight) in many foodstuffs, both vegetal and animal.

Toxicokinetics

Oral cadmium absorption is high and can be increased by some trace elements such as zinc, iron or calcium. It is transported by blood cells and the main transport protein in plasma to the kidneys is the metallothionein, a low molecular weight protein cysteine-rich with numerous thiol binding sites. Metallothionein is induced by cadmium exposure and can be filtered through the kidney glomerular membrane (when its binding capacity is exceeded, renal toxicity occurs) (Gil, 2018).

Cadmium is distributed to various organs and tissues, mainly (more than 70%) to liver, muscles and kidneys, the latter being the target organs. Excretion is through the urine and at a very low rate, which justifies its extensive accumulation in the organism (the half-life in the kidney cortex and liver can be as long as 20–25 years).

Toxicity

As indicated above, the most important target organ is the kidney, which justifies that most of the manifestations that characterize cadmium poisoning are a consequence of renal damage. In chronic poisoning, there is a significant renal dysfunction (Table 1) consisting of tubular interstitial nephritis with distal progression and without glomerular impairment; this leads to a "cadmic

proteinuria" with the appearance in the urine of mainly low molecular weight proteins (β_2-microglobulin, retinol binding protein, etc.). In the long term hypercalciuria sets in, leading to bone alteration with osteomalacia of renal origin, associated with severe bone pain and pathological fractures. These clinical aspects were evidenced in "Itai Itai disease" as a result of consuming rice contaminated with cadmium in the Kakehashi River Basin (Ishikawa, Japan) (Gil, 2018).

The Agency for Research on Cancer (IARC) has determined that cadmium and cadmium compounds are carcinogenic in humans. The EPA determined that cadmium is probably carcinogenic in humans (Group B1) and is capable of inducing prostate, testicular and lung cancers (Agency for Toxic Substances and Disease Registry, ATSDR).

Nickel

Dietary exposure

Nickel is widely distributed in nature and is present in water, soil, plants and animals. It is used in alloys because of certain properties such as heat and corrosion resistance as well as ductility. In fact, it is commonly used in the metallurgical and electroplating industry and also as a catalyst in the chemical industry. The main source of nickel exposure for the general population is food and water. Plants (vegetables and vegetal products, including spices and condiments, legumes, spinach, lettuce, nuts and oilseeds, etc.) are the main sources of nickel exposure. The concentration of this metal in these products depends to a large extent on the nickel content of the soil and also varies according to climate and season (**Table 1**). Baking powder and cocoa powder may also contain an excess nickel (Klein and Costa, 2007; Mania et al., 2019).

Nickel can also be released from materials made of nickel in contact with foods (migration is higher in the case of acidic foods or foods with citric or acetic acid) (Gil and Gisbert-Calabuig, 2018).

Permissible intakes

The recommended dietary allowances for nickel have not been established but, estimated maximum guideline set by US FDA for Ni is 70–80 μg/g. The European Food Safety Authority (EFSA) estimated the tolerable daily intake of nickel from any food source at 2.8 μg/kg body weight, which is 196 μg for an adult and 56 μg for a child. It also determined the Benchmark Dose Lower Confidence Limit ($BMDL_{10}$) associated with dermatitis at 1.1 μg/kg body weight. There are currently no maximum levels in EU legislation for Ni in food nor in the Codex Alimentarius. For natural mineral water a limit of 0.02 mg/L has been established in the Codex (Gil et al., 2017).

Toxicokinetics

Nickel present in food is poorly absorbed (less than 5–10%) through the gastrointestinal tract probably due to the competition with various food components (phosphates, phytates, fibers, etc.) that would reduce such absorption. However, absorption is higher through drinking water (30–40%). Once absorbed, it is distributed in the blood and stored mainly in the kidneys, and also in the lung and brain. Finally, nickel is primarily excreted through urine (Klein and Costa, 2007).

Toxicicty

Nickel compounds are classified as Group 1 carcinogen by the IARC, meaning they are carcinogenic to humans. Metallic Nickel as Group 2B, is possibly carcinogenic to humans. EFSA considered that dietary exposure to Ni is unlikely to cause cancer in humans and that only non-cancer health effects from oral exposure to Ni may occur, such as gastrointestinal, hematological, neurological and immunological effects. Oral exposure to nickel does not cause dermatological reactions (eczema) in individuals without previous contact with the metal (**Table 1**) (Gil and Gisbert-Calabuig, 2018).

Tin

Dietary exposure

This element is found in very small amounts in vegetal and animal foodstuffs. The main sources of tin are canned foods, cereal grains, dairy, meat and vegetables (**Table 1**). Until recently, the main route of dietary exposure was canned food (tin cans). However, their internal lacquering has significantly reduced this route of contamination. Fortunately, tin usually has a very low toxicity, partly due to its minimal intestinal absorption (Gil et al., 2017). Like the other metals, it is a thiol-binding toxicant, affecting many enzymatic activities. However, it is important to note its ability to interfere with the metabolism of certain trace elements (Zn, Fe, Mg, Cu and Se). It has also been shown to interfere with calcium, reducing its bone fixation. Fish and fishery products (especially shellfish) are considered as the main source of organotin compounds (OTC). The main species are likely to be tri-substituted compounds (tributyltin TBT and triphenyltin TPT), which have been used extensively as biocides in wood preservatives, in antifouling paints for boats and as pesticides (ATSDR). Nevertheless, because these compounds are persistent in the environment they tend to accumulate through the food chain (Guérin et al., 2007).

Permissible intakes

The FAO/WHO Expert Committee of food additives suggests weekly maximum intake of 14 mg/kg of body weight (120 mg/day for an adult) (Gil et al., 2017).

Toxicokinetics

Inorganic tin compounds are usually insoluble, poorly absorbed at the gastrointestinal tract (<5%) and show low tissue accumulation. Tin is widely distributed, especially in the liver and spleen, and most of it is excreted in the feces, with slow elimination through the urine.

Toxicity

There is no evidence that tin or tin compounds cause cancer in humans. Studies of inorganic tin compounds in animals have also been negative. Several reports of food poisoning from tin contamination in canned food indicated gastrointestinal symptoms (diarrhea, abdominal pain). Organic tin can cause severe irritation of skin and can lead to systemic toxicity (renal and hepatic damage) (Ostrakhovitch and Cherian, 2007).

Copper

Dietary exposure

Sources of contamination arise from its use in the electrical and electronic industries, in the manufacture of pesticides (fungicides) used against vineyard pests (in particular copper sulfate), as well as in alloys with other metals (mainly zinc and nickel) (Table 1).

Oysters are the seafood that capture and accumulate the most copper. In general, fish is more susceptible of accumulating copper. Among the terrestrial animals, sheep have the highest levels. Fruits, cereals, nuts and green vegetables are important sources of cooper.

Permissible intakes

Taking into account the dietary requirement for copper, which is estimated to be around 0.05 mg/kg body weight (approximately 3 mg for an adult), the FAO/WHO Committee set a provisional value for the maximum daily intake of copper of 0.5 mg/kg body weight (Gil et al., 2017).

Toxicokinetics

Copper present in food is well absorbed (50–75%) in the gastrointestinal tract. It is distributed in blood bound to albumin and is also incorporated into ceruloplasmin, with the highest concentrations of copper found in bone, muscle, liver and brain. The liver is an essential organ in copper metabolism, and not only acts as a storage organ but it is also essential in the regulation of biliary excretion. The majority of copper is eliminated through biliary excretion (>95%) and a small percentage through urine (<5%).

Toxicity

The main acute toxic effects are related to the ingestion of food or drinking water heavily contaminated with copper. This metals has an irritant effect on the mucosa of the digestive tract and it mainly causes nausea, vomiting, epigastric burning and diarrhea (Table 1). Other toxic effects observed in cases of fungicide (copper sulfate) poisoning were hepatotoxicity (with typical symptoms of hepatomegaly, jaundice and alteration of the centrolobulillar area), and hemolysis with possible renal tubular pathology (Ellingsen et al., 2007).

Arsenic

Dietary exposure

The risk of food contamination by arsenic was illustrated in the early 20th century in England as a result of the consumption of beer contaminated with arsenic, which acted as an impurity in the sulfuric acid used in the manufacture of glucose (by hydrolysis of starch). Arsenic levels in the environment are usually relatively low and are present in the geological composition, mining areas, etc. … (US FDA). Arsenic occurs naturally in the soil and in certain minerals (arsenopyrites). Arsenic can therefore pass to the air, water and soils and can even be incorporated into rainwater and leach through the soil. Moreover, arsenic is widely distributed

in nature due to the use of arsenical compounds as pesticides (rodenticides) and xyloprotectants (especially sodium arsenate, arsenic anhydride and arsenious anhydride). Of all the chemical forms, those with trivalent arsenic, and more specifically arsenious anhydride, are the most toxic compounds.

Seafood is the main contributor of this metalloid in the human diet, especially if they are in contact with areas where arsenical waste is dumped. In certain marine animals, especially crustaceans and shellfish (**Table 1**), arsenic tends to concentrate as arsenobetaine. It is also concentrated in algae as arsenophospholipids. Fish contains between 2 and 80 μg/g, oysters 3–10 μg/g and mussels 10–120 μg/g. Several research studies have shown that dimethylarsinic acid (DMA) and monomethylarsonic acid (MMA), two forms of organic arsenic found in certain foods. The FDA is monitoring emerging research on possible health risks from these forms of organic arsenic and will continue to monitor both inorganic and organic forms of arsenic in foods, as needed.

The arsenic compounds present in marine sources are methylated to organic forms, which are less toxic than inorganic ones. Hence, analytical speciation is very important because inorganic arsenic comes mostly from occupational exposure, whereas the organic forms are usually incorporated into the body from a seafood-rich diet. If only total arsenic levels are determined, we will not know what percentage is inorganic arsenic and what percentage is organic (Fowler et al., 2007; Gil et al., 2017).

Permissible intakes

Based on available data, the Joint FAO/WHO Committee estimated as a provisional maximum daily intake for inorganic arsenic 0.0015 mg/kg body weight, with no data assessed for organic arsenicals in food. In any case, the total daily intake of arsenic by humans not subject to industrial exposure is usually less than 0.3 mg/day. FAO/WHO expert committee recommended that the weekly intake per person (provisional tolerable weekly intake, PTWI) of inorganic arsenic should not exceed 15 μg/kg body weight (900 μg/person).

The US EPA has established a limit of 0.01 part per million (ppm) or 10 μg/L for arsenic in drinking water (ATSDR y US FDA).

Regulation (EU) No. 2015/1006 of June 25, 2015, has found it necessary to fix the maximum level of inorganic arsenic in rice and rice-based products. A scientific opinion has established that large consumers of rice in Europe, such as certain ethnic groups and children under the age of 3 years, were most exposed to inorganic arsenic from food. The estimated exposure in children would be two to three times higher than adults. The US FDA issued guidelines for industry not to exceed inorganic arsenic levels of 100 ppb in infant rice cereals. The FDA has also assessed inorganic arsenic levels in apple juice and found it to be generally low, with total arsenic levels below 10 μg/L (limit established for arsenic in drinking water).

Toxicokinetics

More than 90% of inorganic arsenic (trivalent or pentavalent forms) is absorbed from the gastrointestinal tract. Organic arsenic compounds are also well absorbed across the gastrointestinal tract. The degree of particle fragmentation as well as the pH influences the digestive absorption of arsenic trioxide. Two processes are involved in the biotransformation of inorganic arsenic compounds: redox reactions (As III and V) and methylation reactions which convert arsenite to monomethylarsonic acid (MMA) and dimethylarsinic acid (DMA). Arsenobetaine, cacodylic acid and arsanilic acid are apparently not biotransformed. The major route of excretion of inorganic arsenic is the urine.

Toxicicty

Among the chronic toxic effects due to arsenic there is the possibility of carcinogenesis (**Table 1**). Both the US EPA and the International Agency for Research on Cancer (IARC) have determined that inorganic arsenic is carcinogenic in humans and may increase the risk of skin cancer as well as liver, bladder and lung cancers. It is also capable of causing painful sensorimotor polyneuropathy and certain dermatological alterations (palmoplantar hyperkeratosis, Mees bands on the nails, among others). Inorganic arsenic exposure intra utero (arsenic passes through the placenta) and in children is associated with impaired intellectual development. In fact, certain levels of arsenic in children's urine have been associated with impaired attention and cognitive functions, providing clear evidence that postnatal arsenic exposure impairs neurological function in children, even at levels that could theoretically be considered safe (Rodríguez-Barranco et al., 2016).

Conclusion

Therefore, as a conclusion, we must emphasize that metals represent one of the main contaminants in food as a consequence of human activity (industrial, agricultural, domestic). Metals can cause harmful effects to the health of the consumer including cancer. It will therefore be essential to establish permissible limits for these xenobiotics in foodstuffs as well as an adequate toxic risk assessment to the consumer.

References

Cooper, K., Noller, B., Connell, D., et al., 2007. Public health risks from heavy metals and metalloids present in traditional Chinese medicines. J. Toxicol. Environ. Health 70 (19), 1694–1699.

Ellingsen, D.G., Horn, N., Aaseth, J., 2007. Cooper. In: Nordberg, G.F., Fowler, B.A., Nordberg, M., Friberg, L.T. (Eds.), Handbook of the Toxicology of Metals, third ed. Academic Press, California, pp. 529–546.

Falcó, G., Nadal, M., Llobet, J.M., Domingo, J.L., 2006. Riesgo tóxico por metales presentes en alimentos. In: Camean, A., Repetto, M. (Eds.), Toxicología Alimentaria, first ed. Díaz de Santos, Madrid, pp. 309–326.

Fowler, B.A., Chou, C.-H.S.J., Jones, R.L., Chen, C.J., 2007. Arsenic. In: Nordberg, G.F., Fowler, B.A., Nordberg, M., Friberg, L.T. (Eds.), Handbook of the Toxicology of Metals, third ed. Academic Press, California, pp. 367–406.

Garcia-Rico, L., Leyva-Perez, J., Jara-Marini, M.E., 2007. Content and daily intake of copper, zinc, lead, cadmium and mercury from dietary supplements in Mexico. Food Chem. Toxicol. 45, 1599–1605.

Gil, A., Gil, F., 2015. Fish, a Mediterranean source of omega-3 fatty acids: benefits do not justify limiting consumption. Br. J. Nutr. 113, S58–S67.

Gil, F., Gisbert-Calabuig, J.A., 2018. Intoxicaciones por otros metales. Capítulo 71. In: Villanueva, E. (Ed.), Medicina Legal y Toxicología, seventh ed. Elsevier, Barcelona, pp. 1055–1071.

Gil, F., Hernández, A.F., 2009. Significance of biomarkers in applied toxicology. In: Ballantyne, B., Marrs, T.C., Syversen, T. (Eds.), General and Applied Toxicology, third ed., vol. 2. John Wiley & Sons, New Jersey, pp. 847–858.

Gil, F., Hernández, A.F., Pla, A., 2017. Toxicología de los alimentos. Capítulo 3.28. In: Gil, A. (Ed.), Tratado de Nutrición, third ed., vol. III. Médica Panamericana, Madrid, pp. 707–734.

Gil, F., 2018. Toxicología industrial (III). Toxicidad de los metales pesados. Capítulo 50. In: third ed.Gil, F. (Ed.), Tratado de Medicina del Trabajo. Elsevier, Barcelona, pp. 717–729.

Gómez, M.R., Cerruti, S., Sombra, L.L., et al., 2007. Determination of heavy metals for the quality control in Argentinean herbal medicines by ETAAS and ICP-OES. Food Chem. Toxicol. 45, 1060–1064.

Goyer, R.A., Clarkson, T.W., 2001. Toxic effects of metals. In: Amdur, M.O., Doull, J., Klaassen, C.D. (Eds.), Toxicology the Basic Science of Poisons, sixth ed. McGraw-Hill Press, New York, pp. 623–680.

Guérin, T., Sirot, V., Volatier, J.L., Leblanc, J.C., 2007. Organotin levels in seafood and its implications for health risk in high-seafood consumers. Sci. Total Environ. 388, 66–77.

Hernández, A.F., Gil, F., 2016. Nutraceuticals and adverse outcome pathways (Chapter 63). In: Gupta, R. (Ed.), Nutraceuticals: Efficacy, Safety and Toxicity, first ed. Elsevier, Amsterdan, pp. 839–854.

Hsu, M.J., Selvaraj, K., Agoramoorthy, G., 2006. Taiwan's industrial heavy metal pollution threatens terrestrial biota. Environ. Pollut. 143, 327–334.

Klein, C., Costa, M., 2007. Nickel. In: Nordberg, G.F., Fowler, B.A., Nordberg, M., Friberg, L.T. (Eds.), Handbook of the Toxicology of Metals, third ed. Academic Press, California, pp. 743–758.

Łozak, A., Sołtyk, K., Ostapczuk, P., et al., 2002. Determination of selected trace elements in herbs and their infusions. Sci. Total Environ. 289, 33–40.

Mania, M., Rebeniak, M., Postupolski, J., 2019. Food as a source of exposure to nickel. Rocz. Panstw. Zakl. Hig. 70 (4), 393–399.

Olmedo, P., Pla, A., Hernández, A.F., Barbier, F., Ayouni, L., Gil, F., 2013. Determination of toxic elements (mercury, cadmium, lead, tin and arsenic) in fish and shellfish samples. Risk assessment for the consumers. Environ. Int. 59, 63–72.

Ostrakhovitch, E.A., Cherian, M.G., 2007. Tin. In: Nordberg, G.F., Fowler, B.A., Nordberg, M., Friberg, L.T. (Eds.), Handbook of the Toxicology of Metals, third ed. Academic Press, California, pp. 839–859.

Pla, A., Hernández, A.F., Gil, F., 2019. Intoxicación por metales. In: Pla, A., Hernández, A.F., Gil, F. (Eds.), Manual de Toxicología, second ed. Avican, Granada, pp. 245–258.

Rao, M.M., Kumarmeena, A., Galib, 2011. Detection of toxic heavy metals and pesticide residue in herbal plants which are commonly used in the herbal formulations. Environ. Monit. Assess. 181 (1–4), 267–271.

Rodríguez-Barranco, M., Gil, F., Hernández, A., Alguacil, J., Lorca, A., et al., 2016. Postnatal arsenic exposure and attention impairment in Spanish school children. Cortex 74, 370–382.

Sarma, H., Deka, S., Deka, H., et al., 2011. Accumulation of heavy metals in selected medicinal plants. Rev. Environ. Contam. Toxicol. 214, 63–86.

Satarug, S., Vesey, D.A., Gobe, G.C., 2017. Current health risk assessment practice for dietary cadmium: data from different countries. Food Chem. Toxicol. 106, 430–445.

Villanueva, E., Gil, F., 2018a. Intoxicación por plomo. Capítulo 70. In: seventh ed.Villanueva, E. (Ed.), Medicina Legal y Toxicología. Elsevier, Barcelona, pp. 1037–1054.

Villanueva, E., Gil, F., 2018b. Intoxicación por mercurio. Capítulo 69. In: Villanueva, E. (Ed.), Medicina Legal y Toxicología, seventh ed. Elsevier, Barcelona, pp. 1029–1036.

Woodcock, J., Mayne, S.T., 2021. FDA Releases Action Plan for Reducing Exposure to Toxic Elements From Foods for Babies, Young Children, 2021. Food and Drug Administration. Available from. https://www.fda.gov/news-events/press-announcements/fda-releases-action-plan-reducing-exposure-toxic-elements-foods-babies-young-children.

Relevant websites

Agency for Toxic Substances and Disease Registry (ATSDR), https://www.atsdr.cdc.gov/.
European Food Safety Authority (EFSA), https://www.efsa.europa.eu/en.
FoodSafetygov, https://www.foodsafety.gov/.
U.S Food and Drug Administration (US FDA), https://www.fda.gov/.
U.S. Environmental Protection Agency, https://www.epa.gov/.
World Health Organization (Food Safety), https://www.who.int/foodsafety/areas_work/chemical-risks/jecfa/en/.

Food safety: Bacterial contamination

A.H. Simonne, Food Safety and Food Quality, University of Florida, Family, Youth and Community Science, Gainesville, FL, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of M.P. Doyle, Food Safety: Bacterial Contamination, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 322–330, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00124-0.

Introduction 363
Bacterial toxins 364
Staphylococcal food poisoning 364
Background 364
Survival and growth 364
Characteristic sequence of events 364
Clinical features 365
Diagnosis 365
Bacillus cereus 365
Background 365
Survival and growth 365
Clinical features and characteristic sequence of events 365
Diagnosis 365
Clostridium botulinum 366
Background 366
Survival and growth 366
Clinical features 366
Infant botulism 366
Diagnosis 366
Clostridium perfringens 366
Background 366
Survival and growth 367
Clinical features and characteristic sequence of events 367
Diagnosis 367
Vibrio cholerae 367
Background 367
Survival and growth 367
Clinical features 367
Characteristic sequence of events 367
Diagnosis 367
Invasive bacteria 368
Salmonella infections 368
Background 368
Survival and growth 368
Characteristic sequence of events 368
Clinical features 368
Salmonellae and hens' eggs 369
Diagnosis 369
Campylobacter infections 369
Background 369
Survival and growth 369
Characteristic sequence of events and clinical features 369
Diagnosis 370
Escherichia coli 370
Background 370
Survival and growth 370
Characteristic sequence of events 370
Clinical features 370
Diagnosis 371
Other organisms 371
Shigella species 371
Listeria monocytogenes 371

Yersinia enterocolitica 371
Vibrio parahaemolyticus 371
Vibrio vulnificus 372
Brucellosis 372
Streptococcal pharyngitis 372
Prevention of bacterial foodborne illness 372
Conclusion 373
References 373
List of relevant websites 374
Further reading 374

Glossary

Anaerobic Without oxygen
Antigens Markers on the surface of microbes such as bacteria that allow antibodies to recognize them
Cholinergic nerves Nerve cells that employ *acetylcholine* as their *neurotransmitter*
Farinaceous Consisting or made of starch
Hazard analysis critical control point system A management system in which food safety is addressed through the analysis and control of biological, chemical, and physical hazards from raw material production, procurement, and handling, to manufacturing, distribution, and consumption of finished product
Hemolytic uremic syndrome (HUS) Formation of clots in small blood vessels of kidneys which can lead to renal failure
Meningitis A serious inflammation of the thin membranous covering of the brain and spinal cord
Multilocus sequencing typing (MLST) A molecular technique for DNA sequence of different bacteria
Oviduct A passage through which ova leave the maternal body or pass to an organ communicating with the exterior of the body
Polymerase chain reaction (PCR) A rapid technique to synthesize large quantities of a given DNA segment
Septicemia Systematic disease associated with the presence of pathogenic microbes or their toxins in the blood
Serotype A group of closely related microorganisms distinguished by a common characteristic set of antigens
Temperature danger zone Temperatures between 4.4 °C (40 °F) and 60 °C (140 °F) where bacteria can grow rapidly

Introduction

The burden of *gastroenteritis* (GE) in the world, in terms of both morbidity and mortality, is extensive. The more advanced *foodborne disease* surveillance systems become, the more food-associated GE illnesses they uncover.

Not all GE is directly caused by contaminated food. Some GE is caused by poor hygiene resulting in direct or indirect transmission of infectious microbes without a food vehicle. Nevertheless, contaminated food is a major vehicle of infectious GE throughout the world. The general definition of *foodborne illness* (FI) is acute GE caused by consumption of food contaminated with *foodborne pathogens* or their preformed toxin. *Typhoid fever* and *brucellosis* are not usually considered FI, whereas *botulism* is, even though it causes paralysis and not GE, as is *listeriosis*, which principally causes *septicemia* and meningitis.

Norovirus is the most commonly known cause of FI, followed by bacterial FI. Many types of bacterial FI, such as *salmonellosis* and enterohemorrhagic *Escherichia coli* (*E. coli*) O157:H7 infections, are more common in the summer than winter months (CDC, 2021a,b,c,d; WHO, 2021a).

Bacteria produce their effects on the intestinal tract either by direct invasion of the *mucosa* or by production of toxins. Some of the toxins are produced in food outside the intestinal tract; others are formed or released in the intestine. Some invasive bacteria also produce toxins in the intestine. In the developing world, diarrhea, along with *acute respiratory tract infection*, are the leading causes of childhood death. In the more developed world, GE caused by infectious microbes is a significant cause of illness and time lost from work. Death does occur, especially among highly susceptible populations with weak immune response. This article provides an overview of key bacterial causes of FI.

Objectives: Overview of FI.

- Bacterial toxins related FI
  - Staphylococcal food poisoning
  - *Bacillus cereus* food poisoning
  - *Clostridium botulinum*

- *Clostridium perfringens*
- *Vibrio cholerae*
- Invasive bacterial FI
  - *Salmonella*
  - *Campylobacter*
  - *Escherichia coli*
- Other bacterial FI
  - *Shigella* species
  - *Listeria monocytogenes*
  - *Yersinia enterocolitica*
  - *Vibrio parahaemolyticus*
  - *Vibrio vulnificus*
  - Brucellosis
  - *Streptococcal pharyngitis*
- Prevention of bacterial food poisoning
- Conclusion/Summary/Outlook

Bacterial toxins

There are three main forms of *bacterial toxins* that cause FI: (1) *enterotoxin*, which produces excess fluid secretion into the gut (e.g., cholera and staphylococcal toxins); (2) *cytotoxin*, which causes inflammation and mucosal damage (e.g., *Shigella* and enterohemorrhagic *E. coli*); and (3) *neurotoxin*, which affects the nervous system (e.g., botulinum toxin).

Some *E. coli* strains invade and produce toxins, and are addressed under Invasive Bacteria. *Red kidney bean* toxin, scombrotoxin and other *fish toxins*, and *heavy metal poisoning* are addressed elsewhere in this encyclopedia.

Staphylococcal food poisoning

Background

Staphylococcal *food poisoning* (SFP) is one of the causes of bacterial FI commonly attributed to food handlers. Humans frequently carry *Staphylococcus aureus* either in an infected site or asymptomatically. Infected sites include wounds and abscesses, which may be the source of large numbers of *staphylococci*. Asymptomatic sites include the throat, nostrils, fingernails, or hair. Several staphylococcal species (coagulase-negative and coagulase positive) produce highly heat stable enterotoxins. Some coagulase-negative *S. aureus* strains may also produce toxins, but rarely. Because the organism is also carried by many animals, outbreaks attributable to inadequate food handling, which often involves temperature abuse of a pre-contaminated food, can also occur (FDA, 2017; Park and Seo, 2019).

Survival and growth

Staphylococci are killed by normal cooking temperatures. Any staphylococci that survive because of inadequate heating or, more frequently, by post-cooking contamination from a food handler will, if it is an enterotoxigenic strain and given the right conditions of warmth, moisture, pH, and time, produce toxin. Growth of staphylococci and production of toxins are greatest at 35 °C, but growth can occur between 7 and 47.8 °C (FDA, 2017). This toxin is heat stable, unless exposed to high temperature for a long period (i.e., autoclave at 121 °C at 15 PSI for 60 min The toxin is also resistant to *gamma irradiation*.

Many foods have been associated with SFP. *S. aureus* can grow in foods with high salt or sugar content, and salted meat products such as ham have been a common vehicle of SFP, as have desserts, especially those containing cream. Other foods implicated in SFP include *high-protein foods*, salads, canned mushrooms, cream, cheese, *salami*, and eggs. *S. aureus* can also contaminate milk by mastitic cows infected with *S. aureus*.

Characteristic sequence of events

For example, a whole leg of ham (or other foods) is prepared for consumption and cooked. It is then sliced warm (or handled) by a person with no skin lesions but who is a nasal-carrier of *S. aureus*. The slices are stacked on a deep tray. The tray is covered and left to cool for several hours before being refrigerated. Staphylococci from the nose of the food handler are conveyed to the warm ham slices. Because of the large surface area, the large bulk of covered overlapping slices of meat require several hours to cool, during which staphylococci grow and produce toxin. Refrigerating, freezing, or reheating the meat will not destroy the toxin. Holding cooked food, subsequently contaminated with *S. aureus* by a food handler, at ambient temperature for several hours is the major contributing factor to outbreaks of SFP.

Clinical features

Staphylococcal toxin is an enterotoxin and a receptor in the gut appears to be necessary. There may also be a neurotoxic effect that acts on the vomiting center in the brain. With SFP, onset of symptoms is often dramatic. Vomiting is the most prominent symptom, generally occurring within 6 h, but may range from 30 min to 8 h, after eating. Nausea, *abdominal cramps*, and diarrhea are also common. Generally, as with most toxins, the higher the toxin concentration (or the greater the amount ingested), the shorter the *incubation period* and the more severe the symptoms. Individual susceptibility is also a determining factor in severity. The illness usually completes its course within a day or two, but deaths have occurred, sometimes as a result of acute hypotension (another well-known but rare effect of the toxin) (Park and Seo, 2019).

Diagnosis

Many people carry staphylococci asymptomatically in their throat or nostrils. Hence, to identify a food handler as the source of an SFP outbreak, it is important to confirm that the type of *S. aureus* causing an outbreak of FI is the same in the carrier and in those affected; merely showing that a food handler carries staphylococci is insufficient. The organism can be grown from, or enterotoxin can be detected in, implicated foods. Generally, *S. aureus* must grow to populations of $>10^5$ *colony-forming units* per gram to produce sufficient amounts of enterotoxin to produce the illness (FDA, 2017; Park and Seo, 2019). *S. aureus* can also be isolated from vomit or stool of patients and from the hands, nose, abscess, or an infected wound of the food handler. *Phage typing* of strains, with detection and typing of enterotoxin, can also be performed. Enterotoxins A–K (but not F) are recognized, although A is the most common. With the advent of *polymerase chain reaction* (PCR) based (molecular biological detection technology), enterotoxin types other than A are being associated with illness more frequently, because they are reliable, with conclusive results. As with all FI, the absence of laboratory-supporting evidence does not necessarily mean that the diagnosis is wrong or the implicated food innocent (FDA, 2017; Park and Seo, 2019).

Bacillus cereus

Background

Bacillus cereus is widely distributed in the environment and can occur in food. It is found in rice and other *natural foods*, such as herbs and spices, cream, and dry foods.

Survival and growth

Unlike *S. aureus*, *B. cereus* is a spore-forming bacterium that can survive prolonged boiling. It causes two fairly distinct types of food poisoning, *emetic* and diarrheic. The diarrheal toxin is heat labile and, like *Clostridium perfringens* food poisoning, this toxin is released in the gut. Foods commonly associated with diarrheagenic *B. cereus* FI are "proteinaceous" and, like *C. perfringens* FI, associated with meats, stews, desserts, and sauces. The emetic type is "farinaceous," associated mainly with cooked rice, and produces an illness similar to SFP. Different *serotypes* of *B. cereus* cause these two different forms of FI, and the toxins are also different. Some strains can grow at refrigeration temperature in milk and other foods (FDA, 2017; Lindbäck and Granum, 2019). Prevalence of emetic or diarrheal attributed illness varied among different regions of the world (Lindbäck and Granum, 2019).

Clinical features and characteristic sequence of events

The emetic type of *B. cereus* FI is caused by preformed toxin (cereulide) in foods, usually rice (or starchy foods) that have cooled slowly. This typically happens when a large amount of rice is not properly cooled (i.e., left in a temperature danger zone for many hours, often overnight). The center of the mass will remain warm for a long enough period for the *spores* to germinate and form toxins. The toxin is heat stable and will not be inactivated by quick frying rice, as typically done in a restaurant. The incubation period is usually short (1–6 h), and the symptoms, predominantly vomiting, tend to be milder than those of SFP, which it otherwise resembles.

The diarrheal form of *B. cereus* FI is similar to that caused by *C. perfringens*. The toxin, unlike the emetic type, is an enterotoxin released in the intestine and is heat labile. The predominant symptoms are watery diarrhea and abdominal cramps. The incubation period, is also longer (6–15 h), and persists for approximately 24 h. A wide variety of foods, including meat, vegetables, and *dairy products*, have been associated with this type of *B. cereus* FI (Lindbäck and Granum, 2019).

Diagnosis

The mere presence of *B. cereus* in a food is insufficient evidence to confirm the food as the vehicle of FI because *B. cereus* is a normal contaminant of many natural foods. The diagnosis is confirmed by finding *B. cereus* in high concentration [10^6–10^8 g^{-1}, minimum 10^5] in cooked rice, or other foods for the diarrheic type, and obtaining it from the stool or vomit of those who are ill. Alternatively, the same serotype of *B. cereus* should be present in the implicated food and patient specimens. Detection of the emetic toxin in the food may also be sufficient.

Clostridium botulinum

Background

Clostridium botulinum is an anaerobic spore-forming bacterium widely distributed in soil and mud. Botulinum toxin is the most lethal substance known to man, with an LD_{50} of 1–2 ng kg^{-1} body weight. For comparison, *tetanus toxin* from *Clostridium tetani* and *ricin* from the *castor bean* (the next most toxic substances) have LD_{50} values of 0.0001 and 0.02 μg kg^{-1}, respectively. The seven toxin types, A–G, affect the nervous systems of *vertebrate* animals, birds, and man. Birds in *aquatic environments* are especially susceptible to mass die-offs caused by *botulism*. Invertebrates are not susceptible but can harbor the bacteria and toxins in their bodies. Botulinum toxin types A, B, and E most frequently affect humans. Type E is typically acquired from fish. Type C is the primary toxin causing botulism in birds, although types D and E are also important (Johnson, 2019; FDA, 2017).

Survival and growth

C. botulism is an *anaerobe*, hence special conditions (no oxygen) must be present for it to produce toxin. Although the organism is widespread in the environment, botulism in humans is uncommon. First, the spores must be present, which, depending on the type of food, occurs because they are widely distributed in soil and aquatic environments. Second, the spores must survive cooking, which often occurs because they can survive heating at 100 °C for 2 h. Third, they must germinate and grow in *anaerobic conditions*. Although accidents have occurred, and occasionally still occur, especially with home *canned-vegetables, fermented fish products,* and, occasionally, with preserved meat (the term botulism is derived from *botulus*, the Latin term for sausage) and preserved rotting or fermenting food, botulism is rare among humans. Commercial canning, except for the occasional process deviation, destroys spores by the heating processes used. The vegetative forms of *C. botulinum* are as susceptible to heat as most other vegetative bacteria, and the toxin can be destroyed by boiling for 5 min. The pH of food is also important: the lower the pH, the less resistant the spores are to heat, and a low pH (<4.6) prevents *vegetative cells* from growing and producing toxin. Hence, bottled *vegetables pickled* in *vinegar* tend to be safe. High concentrations of salt also affect the viability and toxin-forming properties of *C. botulinum* (Johnson, 2019; FDA, 2017).

Clinical features

The incubation period of botulism is 18–36 h (range, 4 h–8 days). The toxin destroys *cholinergic nerves* in the *motor end plates* (MEPs) of cells. These are the junctions of the nerves within muscle, preventing the release of *acetylcholine* from the cholinergic nerves in the MEP and paralyzing the muscle. Once this has happened, *antitoxin*, which is used to treat patients with botulism, is ineffective. The combination of nausea, vomiting, or diarrhea followed by symmetrical descending paralysis of cranial and *autonomic nerves* is almost diagnostic. The characteristic *neurological symptoms* are blurred vision, dry mouth, difficulty in swallowing, *dysarthria*, *diplopia*, and descending paralysis. Recovery occurs when new MEPs form. The fatality rate was once high but, with respirators, patients are often kept alive artificially until new nerve terminals have formed new MEPs, which may take several months (Johnson, 2019; FDA, 2017).

Infant botulism

Serious illness has occurred in babies (usually younger than 6 months of age) who ingested enough *C. botulinum* spores to cause colonization in the intestine, subsequently forming toxins. The initial symptom is typically constipation, leading to poor appetite, irritability, neck paralysis, and generalized weakness. Honey is a primary vehicle of infant botulism (CDC, 2021b; FDA, 2017).

Diagnosis

The diagnosis is made by demonstration of *botulinum toxin* in food, stool, or serum by mouse bioassay. Clinical symptoms alone can be used for diagnosing botulism, but it can be difficult distinguish it from other disease. Growing the organism from food is suggestive but not diagnostic, whereas fecal isolates are uncommon, except in affected individuals (Johnson, 2019; FDA, 2017; CDC, 2021b,d).

Clostridium perfringens

Background

Food poisoning caused by *C. perfringens* is also toxin mediated. It is similar to the diarrheal form of *B. cereus* FI in that toxin is formed in the intestine after *ingestion* of the bacteria. Like other *clostridia*, it is anaerobic (but can grow in the presence of low concentration of oxygen), Gram-positive, and sporeforming. There are five types (old classification: A through–E according to the enterotoxin formed; type A is the one that causes FI. Some strains, but not generally those that cause FI, can cause *gas gangrene (Pig-bel form or enteritis necroticans)* (FDA, 2017), *which is uncommon in the US* Updated classification include two additional toxin types (F and G) (García et al., 2019). *Clostridium perfringens* is primarily found in soil and is transmitted to animals and man by ingestion of vegetables and other plants. It is commonly found in the intestine of man and animals. When animals are eviscerated, the organism contaminates the inside of the *carcass*. Flies can transmit the organism to food.

Clostridium perfringens FI is frequently reported in countries such as the UK and the US (second most common FI as reported by the CDC); however, it is rarely fatal, except for those who are debilitated or immunocompromized (García et al., 2019; FDA, 2017).

Survival and growth

Clostridium perfringens does not typically multiply on the surface of raw meat. It grows optimally at warm temperatures of approximately 43 °C (range, 10–54 °C) and where there is little to no oxygen in the interior of a cooked dish. The cooking process can remove oxygen and thereby facilitate germination and subsequent growth of the organism. Vegetative cells are not resistant to heat, but spores of the FI strains of *C. perfringens* can survive boiling conditions. If cooling is slow, vegetative cells form and can grow rapidly. After ingestion, toxin is formed from multiplying cells in the intestine, although both toxin and vegetative cells appear to be necessary to produce symptoms.

Clinical features and characteristic sequence of events

As an example, a *casserole* is prepared containing, among other ingredients, meat pieces. It is cooked for 1 or 2 h until ready. However, it may not be consumed immediately and, because of its bulk and the lack of refrigeration facilities, is left unrefrigerated overnight in a warm kitchen. It is warmed the next day before serving. Symptoms of diarrhea with intense abdominal pain typically begin 6–24 h after exposure. The illness may last for up to 24 h, and there are no *sequelae*, except in those who are already debilitated, in which less severe symptoms can last for 1–2 weeks.

Diagnosis

The organism can be cultured from the stools of affected people and should be compared by molecular subtyping and for toxin production with isolates from food. Enterotoxin detection in stools is important confirmatory evidence. The organism must be detected in high numbers ($>10^5$ g^{-1}) in food to be significant (García et al., 2019; FDA, 2017).

Vibrio cholerae

Background

The first *Cholera pandemic* originated in India (Ganges delta) and spread to Asia in 1817–1823. The second pandemic reached Europe in 1826–1837 and, subsequent to this, there were five additional pandemics. *Vibrio cholerae* serogroups O1 and O139 are responsible for epidemics and pandemic cholera outbreaks. It began inexplicably in 1961 with a mild strain, the El Tor *biotype*, which had been endemic in Indonesia since 1937. *Vibrio cholerae* O139 is a relatively new strain that emerged in the Indian subcontinent in 1992 (Ceccarelli et al., 2019). More recently, cholera has become endemic in areas of South America, and other continents (WHO, 2021b). In 2010, after an earthquake in Haiti, without sanitary water for bathing and drinking, a large cholera outbreak killed more than 7000 people (CDC, 2021c).

Cholera was responsible for the introduction of sanitation and the development of "public health." Although not a common cause of FI or GE in developed countries, the *vibrios*, especially *V. cholerae*, still cause large, mainly waterborne outbreaks in the developing world (WHO, 2021b). Rare cases in the US are often associated with international travel. It is the only *gastrointestinal infection* that is internationally notifiable. Because large numbers of organisms are required for infection, person-to-person transmission is uncommon (Ceccarelli et al., 2019; CDC, 2021c; WHO, 2021b; FDA, 2017).

Survival and growth

The bacteria are aquatic and prefer briny waters. They can be found in many warm plankton-rich coastal waters, including the Mediterranean, Gulf of Mexico, and those of Southeast Asia and South America. Bivalved *mollusks* concentrate them, and other fish and shellfish can also be contaminated. Inadequate cooking and unrefrigerated storage will allow *V. cholerae* to survive and grow to sufficient cell numbers to cause FI. Vibrios are generally associated with moist, slightly salty foods. The El Tor strain is more likely to produce *asymptomatic infections*, persist longer in the environment, multiply more rapidly in food, and produce less immunity than the classical type. The organism causes illness by producing an enterotoxin in the intestine (Ceccarelli et al., 2019; FDA, 2017).

Clinical features

Cholera, in its most severe form, is characterized by an acute outpouring of watery diarrhea (rice water stools) and vomiting resulting in death within 24 h by acute loss of fluid and electrolytes. However, the clinical syndrome can also be mild. *V. cholerae* is not invasive, and if the loss of fluid and salts can be counterbalanced by infusion of equal amounts of fluid supplemented by electrolytes, the patient will survive. Patients with an absence of acid in the stomach, and those with *blood group O*, are especially prone to severe symptoms. The incubation period is typically 1–3 days (range, 12 h–5 days) (Ceccarelli et al., 2019; FDA, 2017).

Characteristic sequence of events

Seafood or water contaminated with human sewage is, by far, the most common vehicle of infection. Vibrios can grow prolifically in cooked rice and other grains contaminated by food handlers, and salad vegetables can be contaminated by water.

Diagnosis

V. cholerae is usually isolated from the stool using special media. It can also be distinguished by *light microscopy*, and specific antisera will halt motility of the organisms. *Agglutination tests* with antiserum will distinguish O1 from O139 and other serovars. The bacterium can also be isolated from the environment using enrichment media. Toxin production or the presence of the toxin gene can, and should, also be demonstrated (Ceccarelli et al., 2019; FDA, 2017).

Invasive bacteria

Salmonella infections

Background

Salmonellae are among the most commonly known causes of bacterial FI in developed (and possibly less well-developed) countries of the world (Lewis et al., 2019). *Salmonellae* cause two types of FI: Gastroenteritis, and Typhoidal fever. Annually in the US, it is estimated that *Salmonella* bacteria cause about 1.3 million infection, 26,500 hospitalizations and 420 deaths, and majority of those are non-typhoidal type (CDC, 2021a,b,d; FDA, 2017; Lewis et al., 2019; WHO, 2021a).

There are more than 2500 *serotypes* of *Salmonella*. They are typed according to their somatic [O] or flagellar [H phases 1 and 2] antigens according to the Kauffman–White scheme, and are sometimes named after a geographical location. Further typing or molecular subtyping, such as by *pulsed-field gel electrophoresis* (PFGE), can be done to distinguish strains of the more common serotypes. All serotypes are considered pathogenic for humans, although most have not been detected in infected humans. Salmonellae are gram-negative *bacilli* that do not form *spores*, but can survive for remarkably long periods (months or years) in *dried foods* such as *nonfat dried milk* (FDA, 2017; Lewis et al., 2019).

Salmonellae are frequently carried in the intestinal tract and excreted in the feces of a variety of animals, hence environmental contamination often occurs through contact with manure. *Salmonella*-contaminated protein-based feeds processed in bulk for livestock and poultry have caused widespread infection in animals and, subsequently, in humans. In the UK, for example, fishmeal imported from Peru and fed to poultry caused a large outbreak of *Salmonella* Agona infection in humans that lasted for many years, through the late 1960s and early 1970s. Since then, there have been outbreaks of *S.* Hadar infection in turkeys. More recently infections in poultry and hens' eggs by *S.* Enteritidis of various *phage types* have caused *salmonellosis* outbreaks in many countries. In eggs, transmission of *S.* Enteritidis is mainly "vertical" (i.e., through *oviducts* to the interior of eggs laid by infected hens). Before this, salmonellae largely gained entry to the interior of eggs through the shell. If egg shells were removed in bulk, contamination of just one or two shells would be enough to contaminate the entire batch of eggs and then salmonellae would grow under favorable temperature conditions (FDA, 2017; Lewis et al., 2019).

Other important sources of *Salmonella* contamination include sewage, manure, polluted water, and direct fecal contamination of foodstuffs. Hence, many fresh, unprocessed foods are bought already contaminated. Examples include raw meat and poultry, unpasteurized milk and eggs, legume and vegetable sprouts, and fresh produce. Many multistate outbreaks of salmonellosis in the United States and around the world have been traced to multiple produce items, and the list of implicated foods expanded to include nuts and other low moisture foods (i.e., peanut butter, tahini, and coconut). Cross-contamination in a kitchen or restaurant from raw meat or poultry has also been responsible for numerous outbreaks. Direct contamination of a food by a food handler can also occur. Although cases of human carriers with prolonged fecal shedding of *Salmonella* spp. occur, they are infrequent (FDA, 2017; Lewis et al., 2019).

Survival and growth

Although salmonellae do not form spores, and are fairly easily destroyed by heat (71 °C in a moist food), they can survive for a remarkably long period of time (years) in a dry environment. An outbreak of *S.* Virchow and *S.* Saintpaul infection associated with green *lentils* (mung beans) imported from Queensland occurred in several countries of Europe. The lentils were used to produce *bean sprouts*, which were grown overnight in a warm waterbath. Drying of salmonellae makes them more resistant to heat. The presence of moisture is important when using heat to kill salmonellae.

Salmonellae grow best at 37 °C, with the danger zone at 30–45 °C. Generally, growth does not occur below 7 °C and above 46 °C. Antibiotic-resistant strains are becoming an increasing concern, especially those strains that multiply are resistant to antibiotics used for human therapy (FDA, 2017; Lewis et al., 2019).

The infective dose of salmonellae in humans can be quite low, e.g., 1 cell g^{-1}, depending on a variety of factors, including the strain of *Salmonella*, type of food vehicle, and immune status of the person. Fatty foods such as chocolate, cheese, *salami*, peanut butter, and *mayonnaize* generally can confer illness with much smaller doses, and patients with *immunosuppression*, low acid levels in their stomach (achlorhydria), as well as the elderly and debilitated may also be particularly vulnerable. Salmonellae can be transmitted nosocomially, especially in geriatric or *psychogeriatric* wards of hospitals, and in such outbreaks food may not be the source (FDA, 2017; Lewis et al., 2019).

Characteristic sequence of events

Examples of scenarios of outbreaks of salmonellosis are the following.

A chicken dish is undercooked and then left in a warm environment for many hours before consumption. Alternatively, the chicken may be thoroughly cooked but is then placed in an unwashed container or on a plate with juices from the uncooked chicken, or cut with a knife that was used for raw chicken, and then held at room temperature for a few hours. The contaminated utensil may also be used on another food such as lettuce, thus contaminating it with salmonellae.

Clinical features

The *incubation period* of salmonellosis is generally 12–36 h, but can range from 6 to 72 h. Clinical features of salmonellosis can range from mild to severe. *Enteric fever* (typhoid) is usually caused by *S.* Typhi or *S.* Paratyphi A. Salmonellae can cause severe diarrhea with fever and abdominal pain. Additional, less common symptoms include chills, headache, vomiting, and nausea. Symptoms

usually resolve within 4–7 days. Some *Salmonella* spp. such as *S.* Cholerae-suis, may cause multiple abscesses, and people with *sickle cell disease* may develop bone abscesses caused by a variety of *Salmonella* spp. *Septicemia* (blood poisoning), meningitis, *Reiter's Syndrome*, and some localized infections are also occasional complications of salmonellosis. Patients with AIDS and other *immunosuppressive* conditions are particularly vulnerable to severe complications, which can lead to death (CDC, 2021a; FDA, 2017; Lewis et al., 2019).

Salmonellae and hens' eggs

In the late 1980s, *S.* Enteritidis rapidly became the most common cause of human salmonellosis in the United Kingdom. Previously, *S.* Typhimurium had been the most frequently reported *Salmonella* species associated with human illnesses. Between 1984 and 1987, the number of human *S.* Enteritidis infections increased by approximately 50% per year. In 1988, the number more than doubled and, by 1993, it was virtually 10 times that diagnosed in 1984. By 1993, *S.* Enteritidis accounted for approximately five times the number of *S.* Typhimurium infections. Most of this was due to consumption of undercooked contaminated eggs, although some cases were also attributed to chicken. Many European countries and the USA experienced similar trends (CDC, 2021a; FDA, 2017; Lewis et al., 2019).

Diagnosis

The diagnosis of a salmonellosis is usually made by isolating salmonellae from stool or food and serologically identified. Some salmonellae, such as *S.* Typhimurium and *S.* Enteritidis, are so common that further differentiation is necessary for epidemiologic purposes. Further characterization of the *Salmonella* serovar can be determined by molecular subtyping (PFGE) or phage typing (FDA, 2017; Lewis et al., 2019).

Campylobacter infections

Background

Campylobacter jejuni was determined to be a cause of human GE in the mid-1970s, and is now recognized as the most common bacterial cause of GE and FI in many developed countries. In less developed countries, asymptomatic *Campylobacter* infection is more common (Habib et al., 2019; FDA, 2017). *Campylobacter jejuni* is the most common species causing diarrhea, but *C. coli* is also common in some areas.

Campylobacter spp. are carried in the intestinal tract of many animals and birds, including cattle and horses, household pets, and chickens. Rates of contamination of chicken *carcasses* vary from >75% in the United Kingdom to <30% in Sweden and Norway. Some of these differences may be due to the isolation or detection method used. The estimated annual incidence of human *Campylobacter enteritis* in the USA is *ca.* 1.5 million cases (CDC, 2021a; FDA, 2017; Habib et al., 2019).

Survival and growth

The reason for the late recognition of campylobacters as causes of human GE is the fastidious growth conditions required to culture and isolate them. They grow best in an O_2 concentration of 5% (but not well, if at all, anaerobically or in the presence of 21% O_2) in a special medium and at a temperature of 42 °C. They are sensitive to heat, being destroyed readily by cooking, and do not survive for long (probably a few hours only) when dried on the surfaces of foods or kitchen utensils. They, nevertheless are highly successful in causing infection, probably because of their ubiquity in food-producing animals and birds and the relatively small dose needed for infection (possibly no more than a few thousand organisms may be enough) (FDA, 2017).

Characteristic sequence of events and clinical features

Although campylobacters undoubtedly cause FI, the vehicle of infection in most instances is unknown. It is probable that many cases are caused by direct contact with animals, birds, the environment in which fecal contamination occurs (both domestic and outside), meat carcasses, and possibly other people. Foodborne outbreaks have been traced to untreated water and unpasteurized milk, and also milk from bottles whose tops have been pecked by birds. Undercooked poultry is a major risk factor, and meat prepared at barbecues, which includes pork, *veal*, and beef, as well as chicken, has also been implicated as a vehicle of infection. Eating grapes was determined in one study to be a risk factor, and salads and fresh vegetables have also been implicated as vehicles, but it is possible that some of these foods were contaminated by another source or directly by a food handler.

Other risk factors include travel to foreign countries; handling and cooking of food, especially raw meat; contact with animals and pets (especially those with diarrhea); and visiting an animal farm.

The incubation period is generally 2–5 days. Symptoms are mostly associated with the lower GI tract hence vomiting is uncommon, with abdominal pain and diarrhea being the main symptoms. An accompanying fever, abdominal pain, and headache are usual, and the diarrhea can be bloody. The illness may last a few days, and the antibiotic *ciprofloxacin* is the treatment of choice for severe or prolonged illnesses.

Septicemia or other localized infections are rare complications. One of the well-known complications of *Campylobacter* infection is *Guillain–Barré syndrome*, in which a symmetrical paralysis affects the body some weeks after the infection. Recovery is usually spontaneous but may take several months. In the acute phases of the illness, *respiratory support* may be needed. Reactive arthritis is also a complication, although infrequent, of *Campylobacter* infections.

Diagnosis

The organism can be cultured from stools, rectal swabs, and food. Special media and an atmosphere of 3–5% O_2 are needed for culturing *Campylobacter*. Additionally molecular subtyping assays (i.e., PFGE, MLST) can improve accuracy of the results (FDA, 2017; Habib et al., 2019).

Escherichia coli

Background

Escherichia coli are a remarkable group of bacteria causing a wide range of infections, including *gastroenteritis*, meningitis, septicemia, and *urinary tract infections*. Many strains are nonpathogenic. Those that cause GE have a wide range of pathogenic mechanisms and are divided into various fairly distinct groups: enteropathogenic (EPEC), enterohemorrhagic (EHEC), enteroinvasive (EIEC), and enterotoxigenic (ETEC) are the main ones, although some groups—diffusely adherent (DAEC) and enteroaggregative (EAEC)—have been described. EHEC strains, which include *E. coli* O157:H7, produce one or two primary *Shiga toxins* and possess specific adherence factors. However, because the O157 strains are much more common than the other EHEC strains that cause FI, EHEC strains are classified as O157 and non-O157. Shiga toxin is produced by other bacteria also, including *S. dysenteriae* type 1. Only these EHEC Shiga toxin-producing strains are considered in detail here because they are commonly foodborne (about 75% worldwide) and can cause serious illness and death.

ETEC is not identified routinely by *stool culture* methods commonly used. Infections with ETEC strains are common in developing countries at all ages. These strains are most commonly known to cause *travelers' diarrhea*, but can also be spread by food. This toxigenic group includes strains that produce heat-labile and heat-stable *enterotoxins*. Heat-labile enterotoxin is closely related to *cholera toxin* and causes profuse watery diarrhea. EPEC strains largely cause infections in neonates and infants, which tend to spread from person-to-person, and are not commonly associated with FI. EIEC and the two newer strains are rare. EIEC outbreaks related to food have been occasionally described, including one associated with French cheese exported to the United States.

In the United States, *E. coli* O157:H7 is estimated to cause 96,000 cases and 31 deaths annually, with 68% of cases being foodborne. Swimming in contaminated water can also transmit the infection. *Escherichia coli* O157:H7 was first recognized as a cause of FI in 1982. Some strains with the Shiga toxin gene produce a toxin that causes the *hemolytic uremic syndrome* (HUS). Many serovars of *E. coli* that can produce Shiga toxin have not been associated with human illness.

Escherichia coli, including *E. coli* O157:H7, is an asymptomatic inhabitant of the intestines of many *ruminants*, including cattle, sheep, and goats. Contamination can occur directly from the intestinal content to carcass to meat or via the feces of these animals to *raw vegetables* and other foods (CDC, 2021a; Gonzalez-Escalona et al., 2019; FDA, 2017).

Survival and growth

E. coli O157:H7 can survive for days to weeks on contaminated meat and vegetables. The organism is more resistant to acid than *Salmonella*. *E. coli* O157:H7 can survive for 21 days in cider at a pH of 3.7–3.9 at 4 °C, with only approximately a 5% kill-off. It can grow successfully over several weeks in manure *slurries*. The infectious dose is thought to be very small (less than 100 cells), so person-to-person transmission may occur. Like most vegetative bacteria, it is destroyed by heat (71 °C in moist foods) (Gonzalez-Escalona et al., 2019; FDA, 2017).

Characteristic sequence of events

As an example, 61 patients in a town in North Cumbria, England had diarrhea, many with blood, over 3 weeks. A total of 114 people were infected, ranging in age from 3 months to 85 years. Investigations implicated a farm supplying *pasteurized milk*. Nine days before the first case, a problem had occurred in the *heat-exchanger* plates of the *pasteurization* unit. No tests were undertaken after new plates were fitted, and temperature monitoring was inadequate. The unit was one that a few months before had been the subject of a *food hazard* warning. *Escherichia coli* O157 was isolated from 66 environmental and animal feces samples on the farm but not from the milk or the pasteurization plant.

In an outbreak in the United States, more than 700 became ill after eating inadequately heated *hamburgers* from a restaurant chain. More than 30 cases developed HUS and four died. The largest O157:H7 outbreak (radish sprouts) in Japan affected about 10,000 people.

Undercooked hamburgers and ground beef are a common vehicle of *E. coli* O157:H7 infection. The process of grinding beef can spread the bacteria from the surface of the meat to the inside. Other vehicles of infection include raw milk, unchlorinated water, apple juice, unwashed fruits and vegetables including alfalfa sprouts and bagged spinach, lettuce, and frozen raw cookie dough or swimming in unchlorinated pools (CDC, 2021a; Gonzalez-Escalona et al., 2019; FDA, 2017).

Clinical features

The infectious dose for *E. coli* O157:H7 is thought to be fewer than 100 bacteria. The typical incubation period is 2–5 days (ranging from 2 to 8 days). Symptoms are mostly associated with the lower GI tract. Severe bloody diarrheal and *abdominal cramps* are the most common symptoms, but nonbloody diarrhea also occurs. Fever is unusual. The illness may last a few days, and may progress to HUS which is characterized by *hemolytic anemia* and renal failure, and occurs in approximately 5% of reported *E. coli* O157:H7 cases, most frequently in young children and the elderly (FDA, 2017; Gonzalez-Escalona et al., 2019).

Diagnosis

The usual method of diagnosis is to isolate the bacteria from stools or food. However, because most of the *E. coli* in the intestine is part of the normal flora and nonpathogenic, it is necessary to demonstrate virulence by further tests or assigning it to a serotype, which normally requires more sophisticated techniques in specialized laboratories. Serotyping is performed on the *somatic cell* wall antigens (O antigen) and the flagellar antigen (H). On the basis of serotyping the O and *H antigens* and detecting known *virulence factors* or virulence genes such as toxins, the organisms can be classified as EHEC, EPEC, ETEC, etc. Molecular-based tests are increasingly being used. Toxins can be tested for using *immunoassays* or molecular tests based on gene sequences. *Serology* tests of blood from infected humans are also used, but they are not reliable indicators of recent infection (FDA, 2017; Gonzalez-Escalona et al., 2019).

Other organisms

Shigella species

Humans and primates are the known reservoirs of *Shigella*, hence this bacterium is often spread by person-to-person, especially among kindergarten and primary school children. Shigellosis occurred in both economically developed and developing countries. Affected patients may excrete Shigellae for weeks. Many large outbreaks of *shigellosis* are associated with consumption of food contaminated by sewage-polluted water or food handlers. In 1995, an extensive *S. sonnei* outbreak associated with lettuce imported from Spain affected people in many countries in northern Europe. In another outbreak associated with shrimp consumption, infection was transmitted by a food handler who mixed the shrimp by hand with *mayonnaize* and tomato sauce. The *incubation period* is typically 1–4 days (ranging from 8 h to 7 days) and although large-volume bloody or mucoid diarrhea and high-grade fever are the usual symptoms, it is characteristically accompanied by *tenesmus*—a feeling of wanting to defecate without being able to do so (CDC, 2021a; Faherty and Lampel, 2019; FDA, 2017).

Listeria monocytogenes

Listeriosis, caused by the bacterium *Listeria monocytogenes*, is an unpleasant and rare infection that typically affects the more vulnerable, such as fetuses, infants, pregnant women, the elderly, and the immunocompromized. It causes *septicemia* and meningitis, which is unusual from an FI bacterium. Fatality rates for invasive disease are high, as many as one in four. *GI symptoms* may be absent or mild. Also unusual, this pathogen can grow (albeit slowly) at normal refrigeration temperatures (0–4 °C). It is also very resistant in the environment, both to cold and to heat, so that it can survive for long periods of time (months to years), especially at refrigeration temperatures. The incubation period for invasive listeriosis tends to be long (up to 3 weeks) but, for GI symptoms, very short periods of 1 day have been recorded. *Deli meats* (especially poultry-based), certain *soft cheeses*, and patés have been associated with large outbreaks of listeriosis. In the US, an estimated 1600 people get listeriosis each year and 260 die (CDC, 2021a; Ryser et al., 2019; FDA, 2017). Outbreaks occur sporadically worldwide (WHO, 2021a). South Africa experienced largest outbreak of Listeriosis, with 1060 laboratory-confirmed cases and 216 deaths (ready-to-eat process meat products from 2017 to 2018) (Kaptchouang Tchatchouang et al., 2020).

Yersinia enterocolitica

Like *L. monocytogenes*, *Y. enterocolitica* can grow at refrigeration temperatures (4 °C) or below. It is often missed in the laboratory because it requires special media and generally grows best at 25–30 °C. Many strains are nonpathogenic, although those of *serotype* O 3, 8, or 9 are most commonly associated with illness. The bacterium is largely associated with swine, especially the oral–nasal cavities. Outbreaks have been associated with raw milk and *dairy products*, as well as undercooked meats (i.e., pork, beef, lamb, chitlins) and *tofu*. In one incident, a caregiver who handled swine intestines passed the infection on to some infants. While these infections appear worldwide, according to the current PCR detection, Africa and Western Pacific had the most, and the Americas the least, prevalence of *Y. enterocolitica*. The incubation period is typically 4–7 days, and symptoms of diarrhea often bloody, abdominal pain, and fever, may last from 1 to 3 weeks or longer. In older children and adults, severe right-sided abdominal pain may be confused with *appendicitis*. Infants and young children are most often affected. Clinical features are characteristically fever and profuse watery diarrhea, but may mimic *acute appendicitis*, resulting in an unnecessary operation. Occasionally, in vulnerable patients, septicemia may occur (FDA, 2017; Ukuku and Bari, 2019).

Vibrio parahaemolyticus

Like *V. cholerae*, *V. parahaemolyticus* is an aquatic bacterium that thrives in shallow coastal waters. Deep-sea fish do not tend to harbor the organism and usually become contaminated in fish markets. Precooked frozen shrimp may be contaminated and transmit FI if served without further cooking, as in a seafood cocktail. *Vibrio parahaemolyticus* FI is associated with raw, undercooked, or cross-contaminated seafood and is especially common in Japan and other countries in which seafood is a staple of the diet. Cross-contamination from raw to cooked seafood is a common mode of transmission. Recent data revealed that the incidence of *V. parahaemolyticus* FI has become the leading cause of seafood-related gastroenteritis in Japan, the United States and several other

parts of the world, caused by a pandemic clone. Diarrhea, abdominal pain, and nausea are the predominant symptoms. The diarrhea can be severe, with blood or mucus in the stool. Vomiting is a less common feature, but fever can occur. The incubation period ranges from 4 h to 4 days, but most cases occur between 12 and 24 h. Death is uncommon.

Diagnosis is made by culture of the bacterium from feces or food. *Vibrio parahaemolyticus* can be easily isolated from most *aquatic environments*, but such strains are predominantly Kanagawa negative. Only the Kanagawa-positive strains (i.e., those producing a thermostable *hemolysin* that can be confirmed in a laboratory) cause GE, and it is thought that they multiply selectively in the human intestine. The infectious dose is reportedly 10^5–10^7 cells (Ceccarelli et al., 2019; FDA, 2017).

Vibrio vulnificus

Like other vibrios, infection with this bacterium is acquired from seafood, largely from consumption of contaminated raw oysters. It can cause a fulminating septicemia by ingress through a skin lesion in the food handler and in people with *chronic liver disease*, through consumption of *raw seafood*. *Gastroenteritis* can also occur (Ceccarelli et al., 2019; FDA, 2017).

Brucellosis

Although not usually considered as a FI, *brucellosis* deserves mention because it is associated with food consumption. *Brucella melitensis*, in particular, is largely foodborne, and cheese, milk, and other dairy products made from unpasteurized milk are primary vehicles. Occasionally, contaminated meats may be responsible. *Brucella abortus* is associated with cattle and bovine products, *B. melitensis* with goats, and *B. suis* with swine. Brucellosis is a serious and prolonged systemic illness, with fever, *night sweats*, headache, aches and pains, and, sometimes, profound depression. Many developed countries have eradicated *Brucella* from livestock (FDA, 2017; WHO, 2021a).

Streptococcal pharyngitis

Notwithstanding the definition of FI as causing GE, streptococcal sore throat with fever has been well documented to spread via foods. Usually, a food handler has a *Streptococcus* group A infection in his or her throat, which may be asymptomatic, and transfers this to a food that is then left in a warm environment for several hours before consumption. Foods that have been implicated in illnesses include cheese, milk, eggs, and meat. The incubation period is 24–48 h. To confirm the vehicle, typing of strains is important, as is sound epidemiologic evidence, because many people carry these streptococci in their throats (FDA, 2017).

Prevention of bacterial foodborne illness

With the increasing trend toward manufacturing of foods in large quantities for distribution and globalization, the potential for large outbreaks of *foodborne disease* is considerable. Outbreaks of *salmonellosis* and *E. coli* O157:H7 FI associated with cheese, *salami*, chocolate, peanut butter, beef jerky, dried infant formula, ground beef, and even raw cookie dough have all been documented. In one outbreak of *E. coli* O157:H7 infection, 34 lots of 281,000 lb of beef patties were manufactured in one plant, and 7 of 21 lots tested were contaminated. The introduction of *hazard analysis and critical control point* (HACCP) systems in food manufacturing processes has been a significant advance in the production of safer food and the prevention of FI. *Microbiological criteria* now exist including *ready-to-eat foods*. The establishment of various outbreak-identification surveillance systems (i.e., PulseNet) in the US and elsewhere is an important tool in the early detection of foodborne outbreaks and the curtailment of their effects. These various surveillance systems help link specific ("fingerprinting") information on microbes causing FI to the source and to specific steps where problems occurred. For example, most outbreaks of FI recently detected in the USA are widespread in scope, with cases occurring in many states. The most commonly identified problems are inadequate cooking, improper temperature control (i.e., leaving *prepared food* too long at too high a temperature), allowing cross-contamination from raw to cooked food, improper cleaning and sanitation of equipment, and food from unsafe sources. Some of these foods were contaminated at the source by polluted water or sewage, others during harvesting or processing by infected food handlers and others by food handlers during preparation. It is difficult to avoid or prevent such infections in the kitchen short of cooking everything, and more stringent codes for hygiene at the growing farms and processing plants are required. In the kitchen, it is important to keep raw foods, such as beef and ready-to-eat foods, entirely separate. Salads and fruit are in the ready-to-eat category. Raw meats, especially, should be handled with separate utensils, surfaces, and cutting boards than for cooked food, unless the utensils and cutting boards are washed thoroughly in very hot water and detergent or a *dishwasher*, and then left to dry. Otherwise, many FI microbes can transfer from raw to cooked food and grow.

Cooking food, especially meat, thoroughly will kill vegetative microbes, including salmonellae, although *bacterial spores* can survive. If the cooling down period is too long—normally approximately 2 h is considered the limit before refrigeration is necessary—*C. perfringens* or *B. cereus* that have survived as spores will grow. So will salmonellae and many other FI bacteria if the food was inadequately cooked. Cooking will not normally destroy preformed toxins of *S. aureus*. Infected food handlers may also cause outbreaks of FI by contaminating food during preparation. Generally, *B. cereus* and *C. perfringens* originate in the

food. Infected food handlers whose hands have been contaminated with their feces are the usual source of *Shigella*, hepatitis A, or *norovirus* in outbreaks of FI.

Moist food should be held either hot (above approximately 60 °C) or cold (below 8 °C, preferably 4 °C). Cooling food, even freezing it, will not destroy foodborne bacterial pathogens. Undercooked chicken that has been refrigerated will still need thorough cooking before it is safe to eat. When in large amounts, frozen meat or poultry should be thawed before cooking. Large frozen turkeys may need several days in a refrigerator to thaw fully. The inside of the meat is the last to thaw and the last to cook. Grinding meat will disperse organisms through it. Hence, *hamburgers* and sausages need thorough cooking.

Drinking raw milk is hazardous: A large variety of bacterial pathogens, from *E. coli* O157:H7 to *Salmonella*, can be spread in this way. Hens' eggs have been the vehicle of many cases of salmonellosis FI (mainly *S.* Enteritidis), especially since the 1980s throughout much of Europe and the United States. The rate of contamination is typically low, such as 1 per 20,000 eggs, but the number of cases has been large because of the popularity of eggs as a food and because it is common practice to eat them less than fully cooked, not only on their own but also in other dishes such as sauces and mousse. Screening and *vaccination* of flocks in recent years have reduced the risk of contamination. Irradiation of food is effective in mitigating *foodborne pathogen* contamination, but is not popular with the public. From a food service and retail food standpoint, as >50% of reported FI outbreaks in recent years are from restaurants, food protection manager certification (i.e., ServSafe®, National Registry of Food Safety Professionals, Prometric) and food handler trainings are required in many states.

Education of food handlers may be straightforward but, especially in countries in which food handlers have low status and pay, compliance is more difficult. Education of the general public has been slow but has progressed, e.g., most people now realize the importance of thawing poultry and meat thoroughly before cooking, and the large outbreaks of salmonellosis. In the US, Partnership for Food Safety Education (PFSE) was formed among federal and state government, industry, and consumer leaders to sign a Memorandum of Understanding to develop science-base, consumer oriented message focusing on four core messages (Clean, Separate, Cook and Chill), and The FightBac! Keep Food Safe from Bacteria was released in 1998.

Conclusion

Bacterial *foodborne illness* is one of the major global concerns (after norovirus), with millions of cases occurring in the US and globally. This article provides an overview of important causes of bacterial foodborne illness such Salmonellosis, *Campylobacter enteritis*, *Escherichia coli* (*E. coli*) O157:H7 infection, and *botulism*. The more advanced *foodborne disease* surveillance systems become, the more food-associated GE illnesses they uncover. Preventing bacterial FI in modern time requires a system approach with multidisciplinary filed of sciences, multifaceted approaches, multi-agency, and the latest sound science. Government body can improve sanitary infrastructure and enforce food safety regulation informed by surveillance GE data and good sciences. Changes in global environment, along with the changes in global population, and advances in science (i.e., detection, information technology), among other factors, will likely affect FI data and future intervention strategies.

References

Ceccarelli, D., Amaro, C., Romalde, J.L., Suffredini, E., Vezzulli, L., 2019. Vibrio species. In: Doyle, M.P., Diez-Gonzalez, F., Hill, C. (Eds.), Food Microbiology: Fundamentals and Frontiers, fifth ed. ASM Press, Washington, D.C., pp. 347–388

Faherty, C.S., Lampel, K.A., 2019. Shigella. In: Doyle, M.P., Diez-Gonzalez, F., Hill, C. (Eds.), Food Microbiology: Fundamentals and Frontiers, fifth ed. ASM Press, Washington, D.C., pp. 289–315

FDA, 2017. Foodborne Pathogenic Microorganisms and Natural Toxins Handbook, second ed. Bad Bug Book https://www.fda.gov/food/foodborne-pathogens/bad-bug-book-second-edition. (Accessed 10 October 2021).

García, S., Vidal, J.E., Heredia, N., Juneja, V.K., 2019. Clostridium perfringens. In: Doyle, M.P., Diez-Gonzalez, F., Hill, C. (Eds.), Food Microbiology: Fundamentals and Frontiers, fifth ed. ASM Press, Washington, D.C., pp. 487–512

Gonzalez-Escalona, N., Meng, J., Doyle, M.P., 2019. Shiga toxin-producing *Escherichia coli*. In: Doyle, M.P., Diez-Gonzalez, F., Hill, C. (Eds.), Food Microbiology: Fundamentals and Frontiers, fifth ed. ASM Press, Washington, D.C., pp. 289–315

Habib, I., De Zutter, L., Uyttendaele, M., 2019. Eleven *Campylobacter* species. In: Doyle, M.P., Diez-Gonzalez, F., Hill, C. (Eds.), Food Microbiology: Fundamentals and Frontiers, fifth ed. ASM Press, Washington, D.C., pp. 263–287

Johnson, E.A., 2019. In: Doyle, M.P., Diez-Gonzalez, F., Hill, C. (Eds.), Food Microbiology: Fundamentals and Frontiers, fifth ed. ASM Press, Washington, D.C., pp. 487–512

Kaptchouang Tchatchouang, C.D., Fri, J., De Santi, M., Brandi, G., Schiavano, G.F., Amagliani, G., Ateba, C.N., 2020. Listeriosis outbreak in South Africa: a comparative analysis with previously reported cases worldwide. Microorganisms 8 (1), 135. https://doi.org/10.3390/microorganisms8010135.

Lewis, A.M., Melendrez, M.C., Fink, R.C., 2019. Salmonella. In: Doyle, M.P., Diez-Gonzalez, F., Hill, C. (Eds.), Food Microbiology: Fundamentals and Frontiers, fifth ed. ASM Press, Washington, D.C., pp. 225–262

Lindbäck, T., Granum, P.E., 2019. Bacillus cereus. In: Doyle, M.P., Diez-Gonzalez, F., Hill, C. (Eds.), Food Microbiology: Fundamentals and Frontiers, fifth ed. ASM Press, Washington, D.C, pp. 541–554.

Park, J.Y., Seo, K.E., 2019. Staphylococcus aureus. In: Doyle, M.P., Diez-Gonzalez, F., Hill, C. (Eds.), Food Microbiology: Fundamentals and Frontiers, fifth ed. ASM Press, Washington, D.C., pp. 555–584

Ryser, E.T., Buchanan, R.L., den Bakker, H.C., 2019. Listeria monocytogenes. In: Doyle, M.P., Diez-Gonzalez, F., Hill, C. (Eds.), Food Microbiology: Fundamentals and Frontiers, fifth ed. ASM Press, Washington, D.C., pp. 451–486

Ukuku, D.O., Bari, M.L., 2019. Yersinia enterocolitica. In: Doyle, M.P., Diez-Gonzalez, F., Hill, C. (Eds.), Food Microbiology: Fundamentals and Frontiers, fifth ed. ASM Press, Washington, D.C., pp. 289–315

List of relevant websites

CDC, 2021a. Foodborne Outbreaks. https://www.cdc.gov/foodsafety/outbreaks/index.html. (Accessed 12 October 2021).
CDC, 2021b. Infant Botulism. https://www.cdc.gov/botulism/infant-botulism.html. (Accessed 15 October 2021).
CDC, 2021c. Cholera in Haiti. https://www.cdc.gov/cholera/haiti/index.html. (Accessed 15 October 2021).
CDC, 2021d. Foodborne Germs and Illness. https://www.cdc.gov/foodsafety/foodborne-germs.html. (Accessed 12 October 2021).
FDA, https://www.fda.gov/food/retail-food-protection/fda-food-code.
FDA, 2021. https://www.fda.gov/food/outbreaks-foodborne-illness/investigations-foodborne-illness-outbreaks.
FDA/FSMA, https://www.fda.gov/food/food-safety-modernization-act-fsma/fsma-final-rule-produce-safety.
FSIS.USDA, National Advisory Committee on Microbiological Criteria For Foods (NACMCF), https://www.fsis.usda.gov/policy/advisory-committees/national-advisory-committee-microbiological-criteria-foods-nacmcf.
Partnership for Food Safety Education, https://www.fightbac.org/.
WHO, 2021a. Food Safety. https://www.who.int/health-topics/food-safety. (Accessed 12 October 2021).
WHO, 2021b. Cholera. https://www.who.int/health-topics/cholera#tab=tab_1. (Accessed 15 October 2021).
WHO, 2021c. Foodborne Illness Burden. https://www.who.int/activities/estimating-the-burden-of-foodborne-diseases. (Accessed 15 October 2021).

Further reading

Doyle, M.P., Diez-Gonzalez, F., Hill, C. (Eds.), 2019. Food Microbiology: Fundamentals and Frontiers, fifth ed. ASM Press, Washington, D.C.
Morris, J.G., Vugia, D.J. (Eds.), 2021. Foodborne Infections and Intoxications, fifth ed. Academic Press, Cambridge, MA.

Food safety: Pesticides

Antonio F. Hernández[a,b,c], [a] Department of Legal Medicine and Toxicology, University of Granada School of Medicine, Granada, Spain; [b] Biomedical Research Institute of Granada (Instituto de Investigación Biosanitaria) ibs.GRANADA, Granada, Spain; and [c] Consortium for Biomedical Research in Epidemiology & Public Health (CIBER en Epidemiología y Salud Pública), CIBERESP, Madrid, Spain

© 2023 Published by Elsevier Ltd. All rights reserved.

This is an update of M. Saltmarsh, Food Safety: Pesticides, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 347–352, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00123-9.

Introduction 375
Terminology 376
Regulation 376
Classification 377
Magnitude of pesticide use worldwide 378
Toxicology of pesticides 378
Insecticides 378
Organochlorines 378
Organophosphates 380
N-methylcarbamates 381
Pyrethrins and pyrethroids 382
Neonicotinoids 382
Fungicides 383
Dithiocarbamates 383
Triazoles 383
Benzimidazoles 384
Herbicides 384
Bipyridyl (or bipyridinium) 384
Chloro-S-triazines 385
Chlorophenoxy acids 385
N-phosphonomethyl amino acids 386
Dietary risk assessment of pesticide residues in food 386
Conclusion 387
References 387

Key points

- Pesticides are largely used in agriculture to protect crops against the harmful effects insects, weeds, fungi and other pests.
- As chemical pesticides are biologically active substances, they can cause acute and chronic effects on humans and the environment because of their poor selectivity.
- Adverse effects of pesticides on human health depend on the magnitude of exposure and their intrinsic hazards.
- Although developed countries have banned the agricultural use of pesticides that are most toxic to humans and those that persist for a long time in soil and water, they are still used in many developing or underdeveloped countries.
- Agriculture workers who apply pesticides face the greatest health risks from exposure to these chemicals; however, the presence of pesticide residues in food and drinking water represent a consumer's risk.
- International regulatory agencies review evidence and set maximum residue limits (or tolerances) of pesticides in food and water to protect public health.

Introduction

The Food and Agriculture Organization (FAO) defines pesticides as any substance or mixture of substances intended for preventing, destroying or controlling any pest, including vectors of human and animal disease, and unwanted species of plants or animals causing harm (FAO, 2002). "Pesticides" is, thus, a broad term that includes insecticides, herbicides, fungicides, rodenticides, and plant and insect growth regulators, among others. Commercial formulations of pesticides used in agriculture are referred to as

"plant protection products" (PPPs) and contain the following components which can potentially be harmful to humans and the environment:

- Active ingredient or active substance (i.e., the pesticide itself)
- Safeners, which are added to PPPs to eliminate or reduce their phytotoxic effects on certain plants
- Synergists which, despite having no or weak plant protection activity, enhance the activity of active substances
- Adjuvants, which are co-formulants marketed to be mixed with a PPP to enhance its effectiveness or other pesticidal properties.
- Co-formulants which, although used in PPPs or adjuvants, are not active, safeners or synergistic substances

Most pesticides are organic synthetic derivatives, a few are of mineral origin, and others are biological (also called biopesticides, which can be of plant origin or microbials). As pesticides are biologically active compounds, they must be assessed before being authorized for use, to ensure that they will not pose unacceptable risk to humans, animals or to the environment. While older pesticides had broad targets (they could harm insects, plants, birds, and mammals), the market is continually changing toward newer pesticides intended to be largely effective only against some species (e.g., neonicotinoids), with less toxicity for non-target organisms (Gaforio et al., 2019). However, full selectivity has not been achieved, as non-target species can still be affected (e.g., neonicotinoids are harmful to bees).

Terminology

"Pesticides" is a broad term often confused with biocides and PPPs when, in fact, it encompasses both. PPPs are substances containing at least one active substance (chemical compounds or microorganisms) aimed at increasing crop yield by protecting plants and their products from harmful organisms. According to the European Food Safety Authority (EFSA), PPPs include insecticides, fungicides, herbicides, acaricides, plant growth regulators and repellents. In contrast, biocides are active substances, including microorganisms intended to destroy, counteract, neutralize, prevent the action of, or exert any other type of control over, any harmful organism to human or animal health and natural or manufactured materials. Biocides are used to control pests and vectors of communicable diseases (insects, rodents, etc.), so their purpose is not agricultural. In the European Union (EU), Regulation 528/2012 classifies biocides into four groups: disinfectants and general biocides, preservatives, pesticides and other biocides.

Regulation

In the United States (US), pesticides are regulated under two major statutes, the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) of 1996, and the Federal Food, Drug, and Cosmetic Act (FFDCA) of 2002. These laws were amended by the Food Quality Protection Act (FQPA) and the Pesticide Registration Improvement Act (PRIA) of 2003. FIFRA requires all pesticides sold or distributed in the US (including imported pesticides) to be registered by the Environmental Protection Agency (EPA). FFDCA requires EPA to set pesticide tolerances, i.e., maximum permissible level for pesticide residues allowed in or on human food and animal feed. Under FQPA, EPA must find that a pesticide poses a "reasonable certainty of no harm" before it can be registered for use on food or feed. Each pesticide registration must be reviewed at least once every 15 years (EPA, 2020).

The EPA is, thus, primarily responsible for regulating pesticides in the US. EPA's mission is to protect human health and the environment. The EPA Office of Pesticide Programs (OPP) handles most of the regulatory issues pertaining to pesticides. The FIFRA gives the EPA authority to determine which pesticides can be used in the US, and how they can be used. The EPA (together with other federal and state agencies) evaluates new pesticides and proposed uses, determines if emergency situations warrant temporary approvals of certain pesticides, and periodically reviews current research related to the safety of older pesticides. They also enforce pesticide regulations and provide support to state and regional EPA programs designed to protect, certify and train pesticide applicators (NPIC, 2019).

The EU has a comprehensive regulatory system for pesticides. While active substances are approved at the EU level after an exhaustive evaluation carried out by EFSA in collaboration with Member States, PPPs are authorized by each Member State. The approval of active substances requires the assessment of potential acute, subchronic and long-term effects, as well as genotoxic, reproductive and developmental effects, and neurotoxic or immunotoxic effects, if applicable. The marketing and use of PPPs requires an authorization process that entails the conduct of a strict risk assessment, as established by Regulation (EC) No. 1107/2009. All EU Member States apply the same evaluation and authorization procedures to place PPPs on the market. Regulation 1107/2009 is intended to guarantee a high level of protection of human, animal and environmental health, as well as to safeguard the competitiveness of community agriculture. Vulnerable population groups, such as pregnant women, infants and children, are specifically protected. According to this Regulation, the industry must demonstrate that substances produced or marketed in the EU do not have harmful effects on human or animal health or unacceptable effects on the environment. Regulation 1107/2009 also applies to safeners, synergists, coformulants, and adjuvants.

On the other hand, Regulation (EC) No. 396/2005 establishes harmonized provisions throughout the EU, regarding the maximum residue limits (MRLs) of pesticides in food and feeds of plant and animal origin to guarantee a high degree of protection. MRLs should be set at the lowest possible level in accordance with good agricultural practice (GAP) to protect the most vulnerable groups such as children, embryos and fetuses. This Regulation is aimed to ensure that pesticide residues are not present at levels that

pose an unacceptable risk to humans or animals. Any pesticide lacking an established MRL will not be authorized, and all combinations of pesticides and basic food products that do not have established a MRL are assigned a default value of 0.01 mg kg^{-1}. When setting MRLs, the cumulative and synergistic effects of pesticides sharing a common toxicity mechanism must be taken into account. Member States are obliged to carry out controls to ensure that food marketed in the EU complies with the regulations on MRLs. The results of these controls are submitted to EFSA which provides an annual report on pesticide residue levels in foods of the European market.

Classification

There are many classifications of pesticides, depending on the criteria used. Some are listed below:

1. Chemical nature of the pesticide. The most widespread method is based on the chemical composition and description of the pesticide active substance:
   - Organochlorines
   - Organophosphates
   - Carbamates (*N*-methylcarbamates, thiocarbamates and dithiocarbamates)
   - Pyrethrins and pyrethroids
   - Neonicotinoids
   - Triazoles
   - Bipyridyls
   - Chlorophenoxy acids
   - Triazines
   - Others
2. Pesticide toxicity. Since 2009, the WHO uses the Acute Toxicity Hazard Categories from the Globally Harmonized System of Classification and Labeling of Chemicals (GHS) as the starting point for pesticide classification (WHO, 2020). The GHS was introduced in 2002 with the intent to provide a globally harmonized system to address classification of chemicals, labels, and safety data sheets, and is now widely used for the classification and labeling of chemicals worldwide. Classification for acute toxicity is based on the acute oral median Lethal Dose (LD_{50}) value for rats. Nonetheless, dermal toxicity must be considered, as under most conditions of handling pesticides, a high proportion of the total exposure is dermal (**Table 1**):

In addition to acute toxicity, the GHS also provides classification of chemicals according to their chronic health and environmental hazards. The Joint FAO/WHO Meeting on Pesticide Management (FAO and WHO, 2016) recommended that highly hazardous pesticides should be defined as having one or more of the characteristics depicted in **Table 2**:

3. According to the entry mode for target organisms, pesticides are classified based on:
   - Contact: action through contact with the body of the insect causing death by paralysis of the nerve centers or by suffocation.
   - Ingestion: action when ingested by the animal, causing death by poisoning.
   - Ingestion and contact: both types of action.
   - Systemic: once absorbed by plants, pesticides travel through the plant's vascular system and reach untreated leaves, flowers and fruits.
   - Fumigants: highly volatile pesticides that act or kill target pests by producing vapor.
4. According to target pests, pesticides are classified as insecticides (act on insects), acaricides (on mites), fungicides (on fungi), nematicides (on worms), herbicides (on "weeds"), phytoregulators (regulate plant growth through plant hormones), molluscicides (act on mollusks), rodenticides (control rodent populations), avicides (control bird populations) or bactericides (kill bacteria).
5. Mixed classification: is widely used and includes information on the specific target pest and the chemical family (**Table 3**).

Table 1 Classification of pesticides according to their acute toxicity hazard categories (WHO, 2020).

		LD_{50} for the rat (mg kg body weight^{-1})	
		Oral	*Dermal*
Ia	Extremely hazardous	<5	<50
Ib	Highly hazardous	5–50	50–200
II	Moderately hazardous	50–2000	200–2000
III	Slightly hazardous	≥2000	≥2000
U	Unlikely to present acute hazard	≥5000	≥5000

Table 2 Characteristics that define highly hazardous pesticides according to the Joint FAO/WHO Meeting on Pesticide Management (FAO and WHO, 2016).

Criterion	*Characteristic*
Criterion 1	Pesticide formulations that meet the criteria of classes Ia or Ib of the GHS of classification and labeling of chemicals; or
Criterion 2	Pesticide active ingredients and their formulations that meet the criteria of carcinogenicity categories 1A and 1B of the GHS of classification and labeling of chemicals; or
Criterion 3	Pesticide active ingredients and their formulations that meet the criteria of mutagenicity categories 1A and 1B of the GHS of classification and labeling of chemicals; or
Criterion 4	Pesticide active ingredients and their formulations that meet the criteria of reproductive toxicity categories 1A and 1B of the GHS of classification and labeling of chemicals; or
Criterion 5	Pesticide active ingredients listed by the Stockholm Convention in its Annexes A and B, and those meeting all the criteria in paragraph 1 of Annex D of the convention; or
Criterion 6	Pesticide active ingredients and formulations listed by the Rotterdam Convention in its Annex III; or
Criterion 7	Pesticides listed under the Montreal Protocol; or
Criterion 8	Pesticide active ingredients and formulations that have shown a high incidence of severe or irreversible adverse effects on human health or the environment

Magnitude of pesticide use worldwide

Total pesticide use worldwide has shown a rapid increase, with more than 4 million tons utilized by the year 2017. The major contributing country was China, followed by the US, Brazil and Argentina (Fig. 1). These figures give an idea of the magnitude of pesticides consumption which, together with the pesticides intrinsic toxicity, highlights the potential health risk of their use.

Statistics on pesticide sales are used as an indicator of pesticide consumption in agriculture. According to the US-EPA, the amount of pesticides used worldwide increased from 2.20 to 2.64 million tons between 2008 and 2012 (Atwood and Paisley-Jones, 2017). Approximately half of these figures (48.9%) corresponded to herbicides and growth regulators, followed by fumigants (19.1%), insecticides (18.3%) and fungicides (13.7%). In the US, agriculture is the activity in which the greatest use of these compounds is made, as around 90% of pesticides are used for the chemical control of pests that act on vegetable and food crops. The remainder is used for environmental (infectious disease vector control) or domestic purposes.

In the EU, pesticide sales remained stable over the period 2011–2018, at around 360,000 tons per year, reaching in 2019 the lowest volume of pesticides sold since the start of data collection (~333,500 tons), which represents a decrease of ~10%. The major groups of pesticides with the highest sales volumes in 2011 and 2019 were "Fungicides and bactericides" and "Herbicides, haulm destructors and moss killers". Spain, Italy, France and Germany were the four EU countries that recorded the highest volumes sold for most major groups, and in total (Eurostat, 2021). When Regulation (EC) No 1107/2009 concerning the placing of plant protection products on the market entered into force, many of the more hazardous substances were removed from the market, as they had their authorization withdrawn.

Toxicology of pesticides

Insecticides

Organochlorines

Organochlorine pesticides (OCP) are organic compounds of low molecular weight, with a cyclic structure and chlorine atoms in their molecule. They have been widely used worldwide as insecticides and fungicides until the 1960s, when they were banned in most developed countries. OCP are ubiquitous environmental contaminants as they are lipophilic, chemically stable and non-biodegradable; hence, they persist in the environment for long periods, reach the food chain and end up accumulating in the fatty tissue of living organisms. Due to the high persistence and bioaccumulation potential of these chemicals, the Stockholm Convention classified most of them as environmental hazards and banned their use. OCP are structurally divided into 4 groups (Fig. 2): chlorinated derivatives of ethane (dichlorodiphenyltrichloroethane —DDT—, methoxychlor), chlorinated derivatives of benzene and cyclohexane (hexachlorobenzene –HCB–, α-, β- and γ-hexachlorocyclohexane –HCH– isomers, the latter also known as lindane), cyclodienes (endosulfan, aldrin, dieldrin, endrin) and chlorinated camphenes (toxaphene and chlordecone). None of these are currently authorized for use in agriculture in the EU.

OCP are efficiently absorbed by respiratory and digestive routes and can easily penetrate through intact skin due to their high lipophilicity (except for DDT). The biotransformation of DDT is extremely slow and consists of a reductive dehalogenation with formation of DDE (dichlorodiphenyldichloroethylene), which has a longer half-life. Therefore, the presence of DDE in the blood indicates long-term exposure, while that of DDT, recent exposure. The high lipophilicity of OCP determines that, once they enter the body, they are eliminated slowly because of their accumulation in the adipose tissue, where they can persist for a long time without

Table 3 Classification of pesticides according to their functional class (i.e., target pests) and chemical group.

Functional class	*Chemical class*	*Individual pesticide/chemical subgroup*	*Individual pesticide*
Insecticides	Organochlorines	Endosulfan and other compounds banned in most countries (DDT, aldrin, dieldrin, lindane, etc.)	
	Organophosphates	Chlorpyrifos, chlorpyrifos-methyl, dimethoate, ethoprophos, fenamiphos, fostiazate, malathion, phosmet, pirimiphos-methyl	
	N-methylcarbamates	Methomyl, oxamyl, carbofuran	
	Pyrethrins and pyrethroids	Acrinthrin, bifenthrin, cypermethrin, deltamethrin, fluvalinate, cyhalothrin, tralomethrin	
	Neonicotinoids	Imidacloprid, acetamiprid	
	Formamidin	Amitraz, clordimeform, formetanate	
	Natural and biological pesticides	Macrocyclic lactones	Abamectin, spinosad
		Bacterial insecticides	*Bacillus thuringiensis*
		Botanical insecticides (tetranortriterpenoid)	Azadirachtin
	Growth regulators	Benzoylureas	Flufenoxuron, lufenuron, teflubenzuron
		Juvenile hormone analogs	Fenoxycarb, pyriproxyfen
	Others	Bisacylhydrazines	Tebufenozide, methoxyfenozide
		Oxadiazine	Indoxacarb
		Phenylpyrazole	Fipronil
		Pyrazoles	Fenpyroximate, tebufenpyrad
		Pyridazinone derivative	Pyridaben
		Quinazoline	Fenazaquine
		Sulfoximines	Sulfoxaflor
		Tetronic and tetramic acid derivatives (cyclic ketoenol)	Spirodiclofen, spiromesifen, spirotetramat
Fungicides	Dithiocarbamates	Dimethyl-dithiocarbamates	Ferbam, thiram, ziram
		Ethylene-bis-dithiocarbamates	Maneb, mancozeb, metiram, nabam, zineb
		Propylen-bis-dithiocarbamates	Propineb
	Triazoles	Fenbuconazole, fluquinconazole, flutriafol, myclobutanil, propiconazole, tebuconazole, tetraconazole, triadimenol	
	Inorganic	Copper-based compounds	Copper sulfate, cuprous oxide
		Sulfur and its compounds	
	Benzimidazoles	Thiabendazole, carbendazim, benomyl, thiophanate-methyl	
	Dicarboximide	Vinclozoline, iprodione, procimidone, clozolinate	
	Halobenzonitrile	Chlorothalonil	
	Imidazole	Prochloraz	
	Morpholine group	Dimetomorf	
	Phthalimides	Captan, folpet, captafol	
	Chloro- nitrophenol	Pentachlorophenol, dinitro-*o*-cresol, dinoseb	
	Others	(Benzothiazol-2-ylthio)methyl thiocyanate (TCMTB), benzothiazole	
Herbicides	Bipyridyl (bipyridinium)	Paraquat, diquat	
	Triazines	Atrazine, simazine, cianazine, propazine	
	Chlorophenoxy acids	2,4-Dichlorophenoxyacetic acid (2,4-D)	
		2,4,5-Trichlorophenoxyacetic acid (2,4,5-T)	
	N-(phosphonomethyl) glycine derivatives	Glyphosate, glufosinate	
	Urea and substituted urea (phenylureas, sulfonylureas)	Linuron, monuron, isoproturon, diuron, fluometuron	
Others	Fumigants	Methyl bromide, 1,3-dichloropropene, metam-sodium, aluminum phosphide, zinc phosphide	
	Rodenticides	Thallium sulfate, sodium fluoroacetate, coumarin anticoagulants (warfarin), stricnine	

causing apparent harmful effects. However, OCP can be mobilized by lipolysis in cases of fasting, infection, cancer, pregnancy, etc., eventually causing detrimental health effects (Costa, 2015; Jayaraj et al., 2016).

The main target organ of OCP is the nervous system. Specifically, they inhibit $GABA_A$-gated chloride channels, where the inhibitory neurotransmitter gamma-aminobutyric acid (GABA) binds, thus producing neuronal hyper-excitability. Furthermore, DDT, like pyrethroids, binds to voltage-gated sodium channels, keeping them in open state and leading to persistent depolarization and hyperactivity of the nervous system. On the other hand, cyclodienes inhibit calcium ion influx and Ca- and Mg-ATPase, causing

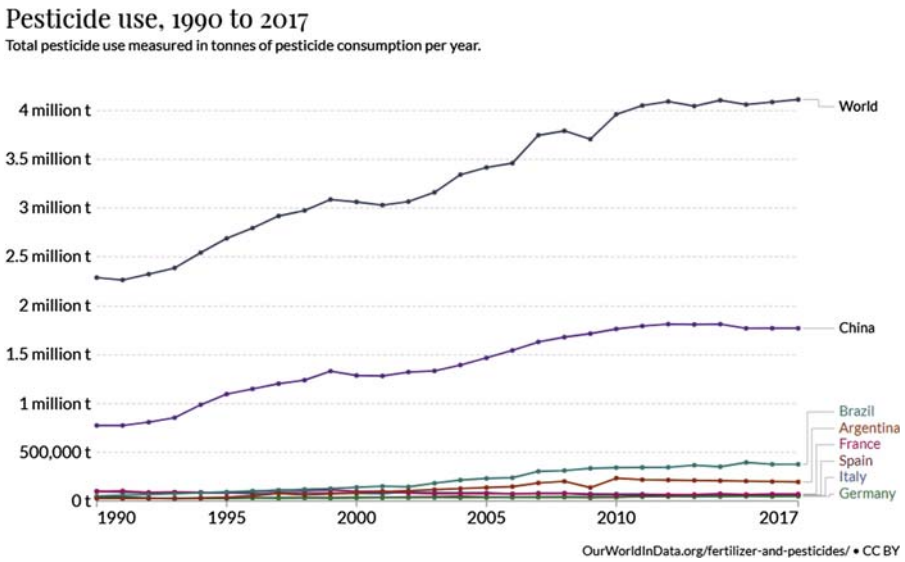


Fig. 1 Pesticides utilized worldwide from 1990 to 2017. From Roser (2019).

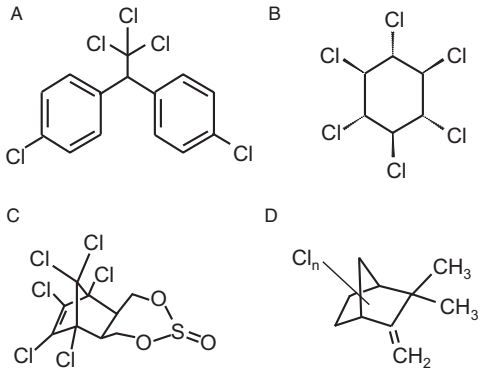


Fig. 2 Chemical structures of dichlorodiphenyltrichloroethane, DDT (A); endosulfan (B); gamma-hexachlorocyclohexane, γ-HCH (C); and toxaphene (D); representative compounds of the four organochlorine pesticide classes.

release of neurotransmitters (Costa, 2015; Jayaraj et al., 2016). In addition, long-term exposure to OCP may damage the liver and kidney, owing to the presence of chlorine atoms in their molecule, and they may cause skin and mucous membrane irritation. Some are carcinogenic, for example lindane has been classified by the International Agency for Research on Cancer (IARC) in group 1 (carcinogenic to humans) and DDT in group 2A (probably carcinogenic) (IARC, 2018).

An interesting property of OCP is their ability to induce liver enzymes involved in the biotransformation of xenobiotics, such as cytochromes P450 (CYPs), particularly subfamily CYP2B and CYP3A enzymes. This results in accelerated metabolism of both endogenous compounds (steroid hormones) and xenobiotics.

Organophosphates

Organophosphate (OP) pesticides are mostly phosphoric acid esters (Fig. 3), which differ depending on the atoms or groups attached to phosphorus. Most are insecticides, while a few (fosetyl and tolclofos-methyl) are fungicides. Currently, fosthiazate,

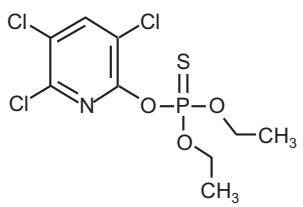


Fig. 3 Chemical structure of chlorpyrifos, an organophosphate insecticide.

malathion and pirimiphos-methyl are the only OP insecticides authorized for agriculture use in the EU. The authorization status of chlorpyrifos, chlorpyrifos-methyl, dimethoate, ethoprophos, fenamiphos and phosmet was shifted to non-authorized in the past few years.

OP compounds have partly replaced organochlorine insecticides, due to their low persistence in the environment where they undergo (bio)degradation. They do not accumulate in the body, as they are metabolized yielding non-toxic metabolites. Nevertheless, OPs have higher acute toxicity.

OPs easily enter the body by all routes. The skin is responsible for a high proportion of poisonings in occupational settings. Application of OPs generates aerosol droplets that can be inhaled by sprayers or by-standers which, like other respiratory irritants, can lead to reactive dysfunction of the airways. Irritation is generally restricted to the upper airways, although dyspnea (shortness of breath) and chest tightness may occasionally occur. When droplets are small enough (<10 μm in aerodynamic diameter) they can reach the alveoli and be absorbed through the respiratory route, eventually leading to systemic effects.

The half-life of OPs and their degradation products is relatively short (hours or days). OPs are biotransformed in the liver by oxidation, hydrolysis and conjugation. Plasma esterases (paraoxonase-1, –PON1–) also play a prominent role, particularly in chronic exposure to low doses. The major breakdown products of OPs are dialkylphosphates (DAPs), which are non-specific metabolites common to many OPs. Urinary DAP concentrations have been used to estimate human exposure to OP pesticides (Hernández et al., 2019).

Many OP pesticides are not directly active, but need to be activated by oxidative desulfuration to oxons, which are potent acetylcholinesterase (AChE) inhibitors. These oxidations are almost always catalyzed by the CYP450-dependent monooxygenase system which can be induced by some agents, including OCP, hence the potential risk from combined use of these compounds (Hernández et al., 2013).

OPs exert their main toxicological effects through non-reversible phosphorylation of AChE, an enzyme responsible for degrading the neurotransmitter acetylcholine in the central nervous system, neuromuscular junction and autonomous nervous system. The irreversible inhibition of AChE leads to an excess of acetylcholine in cholinergic synapses, causing overstimulation of muscarinic and nicotinic receptors in postsynaptic neurons, resulting in muscarinic, nicotinic and central syndromes. The severity of poisoning depends not only on the extent of AChE inhibition, but also on the speed of inhibition (Costa, 2008; Hernández and Pla, 2007).

Clinically, exposure to OPs may produce four toxic syndromes. The first is the cholinergic syndrome, typical of acute poisoning with various degree of severity. The second is the so-called intermediate syndrome, which appears 24–96 h after the acute cholinergic crisis, consisting of weakness and even paralysis of respiratory muscles, neck flexor muscles, proximal limb muscles, and muscles innervated by cranial motor nerves. The third is known as organophosphate-induced delayed polyneuropathy (OPIDP), which appears 10–21 days after exposure to some OPs (e.g., methamidophos) and is characterized by impairment of distal parts of sensory and motor axons in peripheral nerves and ascending and descending tracts of spinal cord. Neuropathy target esterase (NTE) is thought to be the molecular target of OPIDP. The fourth clinical syndrome is considered as long-term effects that persist for several months after exposure to high doses of OPs, or after long-term low-dose exposure, leading to neurobehavioral changes (e.g., anxiety, mood swings, emotional lability, depression, fatigue, irritability, drowsiness, confusion, and lethargy). These symptoms, collectively known as chronic OP-induced neuropsychiatric disorders (COPIND), are thought to involve non-cholinesterase targets, due to the highly reactive nature of OPs (Abou-Donia, 2003).

N-methylcarbamates

The term "carbamates" covers a wide family of pesticides, including derivatives of *N*-methylcarbamic, thiocarbamic and dithiocarbamic acids, which are used as insecticides, herbicides and fungicides, respectively. However, *N*-methylcarbamates are the only carbamates showing anticholinesterase effects and, thus, cause similar acute toxic effects as OPs. Examples of these chemicals include: aldicarb, carbaryl, carbofuran, methomyl (**Fig. 4**), oxamyl, pirimicarb and propoxur. *N*-methylcarbamates are not persistent in the environment and do not accumulate in the body.

N-methylcarbamates are absorbed by all routes, including the skin. They undergo liver metabolism through oxidation, hydrolysis mediated by carboxylesterases, and conjugation, yielding phenols as major metabolites that are eliminated in the urine. At sufficient doses, *N*-methylcarbamates cause similar acute toxicity as the OP insecticides, as both have the same mechanism of toxic action. However, *N*-methylcarbamates inhibit AChE in a reversible way (the carbamate–AChE binding reverts spontaneously in a few hours) such that their effects are less severe and of shorter duration. Hence, both the morbidity and mortality of *N*-methylcarbamates is limited. Another important difference is that these compounds do not easily cross the blood-brain barrier and barely enter into the brain, such that their effects on the central nervous system are limited; nonetheless, brain dysfunction may still occur in massive intoxication (Colović et al., 2013; Costa, 2008).

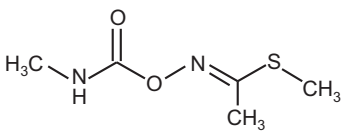


Fig. 4 Chemical structure of methomyl, an *N*-methylcarbamate insecticide.

Pyrethrins and pyrethroids

Pyrethrins are a mixture of six chemical insecticides found naturally in some chrysanthemum flowers. They do not accumulate in the body, their excretion is rapid, and they do not persist in environmental media, as they degrade rapidly. However, their cost, high biodegradability and lack of photostability limit the use of pyrethrins, which require piperonyl butoxide (with low oral toxicity) as a synergistic agent to enhance their insecticidal effectiveness. For this reason, analogous products were synthesized with structure and properties similar to pyrethrins, the so-called pyrethroids. These are esters of chrysanthemic acid with a broad activity spectrum, excellent selectivity and improved field stability. They are classified into two groups: type I pyrethroids (i.e., esters lacking the α-cyano substituent: allethrin, bifenthrin, permethrin, phenothrin, tetramethrin, resmethrin, tefluthrin) and type II pyrethroids (esters containing the α-cyan substituent: cyfluthrin, cyhalothrin, cypermethrin, deltamethrin (**Fig. 5**), phenpropathrin, fenvalerate, flucythrinate, flumethrin, fluvalinate and tralomethrin).

The acute mammalian toxicity of pyrethrins is generally low. The most common adverse effect is their great sensitizing power, related to the sesquiterpene lactones contained in the plant extract (pyrethrum). These substances cause allergic rhinitis and contact dermatitis (O'Malley, 1997). Natural pyrethrins are more toxic by skin contact than by ingestion, while synthetic pyrethroids are more potent when ingested. However, the latter are not skin sensitizers or irritants, although may cause inflammation and paresthesia after coming into contact with the skin. Pyrethroids can be absorbed by ingestion or inhalation and, to a lower extent, through intact skin, although significant neurotoxic effects have been observed in animals through this route. In the body, pyrethroids undergo rapid hydrolysis mediated by carboxylesterases, followed by oxidation of the resulting alcohol to aldehyde and acid metabolites, which are rapidly excreted from the body. This rapid metabolism leads to low mammalian toxicity. Urine concentration of 3-phenoxybenzoic acid, a common metabolite of many pyrethroids, can be used for biological monitoring of recent exposure to these chemicals (Costa, 2008, 2015).

Pyrethroids bind to voltage-gated sodium channels in nerves. Analogously to DDT, type I pyrethroids prolong channel opening only long enough to cause repetitive firing of action potential. Conversely, type II pyrethroids hold the channels open for such long periods that the membrane potential ultimately becomes depolarized and the generation of action potential is not possible (depolarization-dependent block) (Costa, 2015). Two distinct acute poisoning syndromes have been described in experimental animals: the T-syndrome, produced by type I pyretroids and pyrethrins, featured by tremor, ataxia and sensitivity to sensory stimuli; and CS-syndrome, produced by type II pyrethroids, which is characterized by choreoathetosis and salivation. Both syndromes may also induce paralysis (WHO, 2005).

Neonicotinoids

Neonicotinoids are a class of systemic insecticides that have become the most widely used group of insecticides globally. Currently there are 4 generations of neonicotinoids. Imidacloprid (**Fig. 6**), acetamiprid, thiacloprid, and nitenpyram belong to the first generation, all of which contain the chloropyridine ring and yield 6-chloronicotinic acid as a common metabolite, often used as a urinary biomarker of exposure in biomonitoring studies. The main second-generation neonicotinoid is thiamethoxam, which originates a metabolite that also exhibits insecticidal activity (clothianidin). Dinotefuran is the main third-generation neonicotinoid, and sulfoxaflor and cycloxaprid are members of the fourth generation (Simon-Delso et al., 2015).

Neonicotinoids have rapid oral absorption and are completely eliminated from the body by 48 h, so they do not accumulate. They undergo CYP450-mediated oxidation, which differs depending on animal species and each neonicotinoid. Certain metabolites are as or more toxic than the parent compound, such as *N*-denitro-imidacloprid, which has a high affinity for mammalian nicotinic acetylcholine receptors (nAChRs). The mode of action of neonicotinoid insecticides is similar to nicotine, a natural insecticide. They are agonists of postsynaptic nAChR in the nervous system, allowing for an influx of cations (Ca^{2+}, Na^{+}, or K^{+}, depending on the nAChR subtype) through the central channel pore. Differences in the properties and structure of nAChR subunits between insects and mammals partly explain the high selectivity of neonicotinoids over insects and the relative low toxicity to vertebrates.

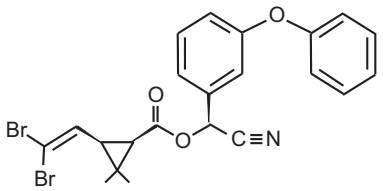


Fig. 5 Chemical structure of deltamethrin, a type II pyrethroid insecticide.

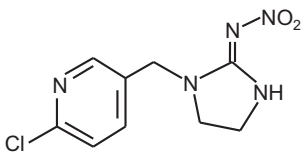


Fig. 6 Chemical structure of imidacloprid, a neonicotinoid insecticide.

While neonicotinoids are selective for receptor subtypes in the vertebrate brain (α4β2 and α7), they have little or no effect on nAChRs of the peripheral nervous system. These insecticides were originally believed to exert low human and mammalian toxicity because of their lower chemical affinity for mammalian nAChRs and their inability to cross the mammalian blood-brain barrier. However, in vitro and in vivo studies, as well as ecotoxicological studies, indicated that these insecticides have adverse effects on mammals, even at sublethal doses, as they can affect mammalian nAChRs in a similar way to nicotine (Cimino et al., 2017; Houchat et al., 2020).

Following high dose exposure, neonicotinoids cause gastrointestinal symptoms consisting of nausea, vomiting, abdominal pain, and corrosive oral mucosa damage. Skin and eye irritation can also occur. Stimulation of nAChR in the autonomic nervous system leads to tachycardia, hypertension, diaphoresis, and mydriasis. Neurodevelopmental and reproductive effects have also been reported (Cimino et al., 2017; Selvam and Srinivasan, 2019).

Fungicides

Dithiocarbamates

The dithiocarbamic acid class of fungicides includes dimethyl-dithiocarbamates (DMDTC, e.g., ferbam, thiram and ziram), ethylene-bis-dithiocarbamates (EBDC, e.g., maneb, mancozeb, metiram, nabam and zineb) and propylene-bis-dithiocarbamates (propineb).

Dithiocarbamates (DTC) (Fig. 7) are organosulfur compounds that bind to various transition metals, forming more lipophilic complexes capable of entering the central nervous system. The nomenclature of DMDTC and EDBC compounds is derived from the cationic metals to which they are associated (e.g., iron to ferbam, zinc to ziram and zineb, sodium to nabam, manganese to maneb, etc.). Some DTC metabolites are highly reactive electrophilic compounds that can covalently bind to thiol groups in cysteine residues. Although their acute toxicity in experimental animals is low or moderate, some worrying adverse effects have been reported, such as teratogenesis and reproductive disorders (embryotoxicity). Prolonged or repeated exposure to these compounds can cause dermatitis or conjunctivitis and, if inhaled, bronchospasm. Although inhibition of AChE has been reported for some DTC, it is considered weak or null (Hernández and Pla, 2007; Janz, 2014).

EBDC are metabolized in mammals, and degraded in the environment, to ethylene thiourea (ETU), a mutagenic, carcinogenic, teratogenic and immunotoxic agent in animals, which has raised concerns about the human toxicity of EBDC fungicides. In addition, ETU affect thyroid function by decreasing the synthesis of thyroid hormones, leading to decreased plasma concentration of T4 and a compensatory increase in TSH. On the other hand, DTC fungicides yield a common metabolite, carbon disulfide, which is responsible for the fungicidal and disinfectant properties posed by many DTCs. This metabolite can also cause peripheral neuropathy that persists for a period of time (Lushchak et al., 2018).

Chronic exposure (4–5 years) to maneb can lead to parkinsonism, a syndrome that has been attributed to the extrapyramidal effects caused by manganese (although its blood concentration is not usually elevated), and also to oxidative stress as a result of dopamine autoxidation, generation of reactive oxygen species, reduced glutathione (GSH) and decreased activities of antioxidant enzymes (Lushchak et al., 2018).

Triazoles

Azoles are antifungal agents widely used in agriculture, in clinical practice and in cosmetic products. They are classified into imidazole or triazole derivatives, depending on the number of nitrogen atoms in the 5-atom ring and which is responsible for the fungicidal activity (Fig. 8).

While imidazole compounds are used mostly as pharmaceuticals, triazoles can be used as drugs (fluconazole, voriconazole, itraconazole) or plant protection agents (tebuconazole, propiconazole, tetraconazole, fluquinconazole, hexaconazole, myclobutanil, triadimefon, triadimenol, etc.). The antifungal mechanism of action of azoles consists of inhibiting C14α-demethylase (CYP51),

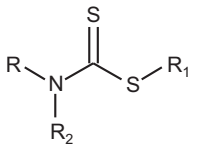


Fig. 7 General chemical structure of dithiocarbamate fungicides

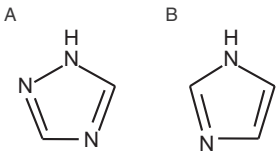


Fig. 8 General chemical structure of azole fungicides: triazole (A) and imidazole (B).

an enzyme that catalyzes the transformation of lanosterol into ergosterol, a component of the fungal cell wall. These compounds can cause adverse effects in humans due to nonspecific binding to various CYP450 isoenzymes, although triazoles show less affinity for human CYP450 than for fungal. CYP19 (aromatase), the enzyme that converts androgens into estrogens in steroidogenesis, is one of the CYP450 inhibited by azoles in humans. They also inhibit other CYP450s involved in steroid biosynthesis and may interact with nuclear receptors, especially androgen and estrogen receptors with different affinities. In rats, maternal exposure to the imidazole fungicide prochloraz produces malformation of the male reproductive tract and diminishes testicular steroidogenesis, indicating antiandrogen effects (Ye et al., 2011).

Contact with triazoles may cause irritation and redness of the skin and mucous membranes. Orally, they have low toxicity. In experimental animals, myclobutanil produces toxicity in the testes, adrenal gland, kidney and thyroid. Most triazoles have hepatotoxic effects, and only some are toxic to reproduction and development (e.g., triadimefon), causing teratogenic effects such as craniofacial and axial skeletal malformations (Marrs and Ballantyne, 2004).

Exposure to triazoles, in general, causes rodent neurotoxicity. Acute exposures to triadimefon affect catecholamines in the central nervous system and induce a transient syndrome in rats consisting of hyperactivity and stereotyped behaviors (Crofton, 1996). Triadimefon and its metabolite triadimenol can inhibit dopamine reuptake in mammals by blocking the dopamine transporter (DAT). While tebuconazole may not produce neurotoxic effects in adults, it may lead to developmental neurotoxicity when the exposure occurs during development (Filipov and Lawrence, 2001).

Benzimidazoles

The benzimidazole family of fungicides includes carbendazim, benomyl, thiabendazole and fuberidazole. Although imazalil and thiophanate-methyl are also included in this fungicide class, they lack the typical benzimidazole moiety consisting of a fusion of benzene and imidazole ring system (**Fig. 9**). These compounds inhibit microtubule formation by binding to free β-tubulin monomers. Thiabendazole is a slightly toxic compound also used to treat some helminthiases in veterinary medicine and in humans. Carbendazim is the main metabolite of benomyl and thiophanate-methyl, and produces toxic effects on the seminiferous epithelium by inhibiting polymerization of β-tubulin, thus interfering with microtubule formation. Carbendazim also exerts androgenic effects which are not related to the androgen receptor but to a lower expression of estrogen receptors (ERα and β). Carbendazim specifically acts on the blood-testicular barrier, affecting germ cell migration (Durand et al., 2017; Pisani et al., 2016).

Herbicides

Bipyridyl (or bipyridinium)

These are solid herbicides that contain a bipyridyl ring structure with quaternary ammonium that form salts with bromide and chloride anions very soluble in water. The most widely used are paraquat and diquat (**Fig. 10**), while cyperquat, diethamquat, difenzoquat, and morfamquat are little used or no longer marketed.

Paraquat is an excellent herbicide, widely used in agriculture, but withdrawn from the European market in 2007. Paraquat is poorly absorbed by any route but, given its high toxicity, it can cause severe and life-threatening poisonings with a high mortality rate. As paraquat is highly corrosive, it damages the skin and the digestive mucosa, thus improving its absorption through these routes. Accidental poisonings have occurred after drinking water from discarded paraquat bottles as well as from sustained dermal

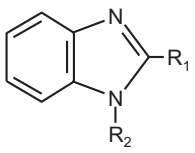


Fig. 9 General chemical structure of benzimidazole fungicides.

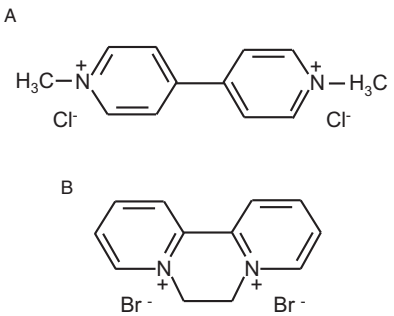


Fig. 10 Chemical structure of the bipyridyl herbicides paraquat (A) and diquat (B).

exposure in agriculture settings due to leaking backpack sprayers. Once absorbed, paraquat is rapidly distributed through all well vascularized organs, and accumulates in the lungs, the main target organ, by selective uptake via the polyamine transport system abundantly expressed in the membrane of type I and type II alveolar cells, and Clara cells. In contrast, diquat does not bind to these transporters, which explains its different toxicokinetic and toxicity. Paraquat metabolism generates oxygenated free radicals responsible, in part, for mucous membrane lesions and secondary necrosis of the gastrointestinal tract, liver, kidney and adrenal glands by lipid peroxidation. Redox cycling and intracellular oxidative stress generation are, thus, the main molecular mechanisms of paraquat toxicity. After an irritant action on the mucous membranes (which causes acute gastrointestinal distress), and following a latency period of 7–14 days in which hepatorenal toxicity and myocardial damage occurs, paraquat produces pulmonary fibrosis. Epidemiological and experimental studies have considered paraquat as a risk factor for Parkinson's disease. Despite its cationic nature, paraquat is able to cross the blood-brain barrier using membrane transporters (Marrs and Ballantyne, 2004; Bronstein, 2004; Dinis-Oliveira et al., 2008).

Chloro-S-triazines

The 2-chloro-s-triazine family of herbicides is widely used as broad-spectrum herbicides that inhibit photosynthesis in plants. They have a triazine ring and the best known are: atrazine (**Fig. 11**), cyanazine, propazine, simazine and terbuthylazine, although the latter is the only one currently approved in the EU for agricultural use.

These compounds persist in the environment and are mobile in soil and groundwater. Groundwater contamination resulted in the cancellation of atrazine and simazine registrations in Europe. Triazine herbicides have low oral toxicity and are unlikely to pose acute hazards with normal use, although some of them are moderately irritating to the eyes, skin and respiratory tract. Atrazine and other triazines have raised concerns for human health based on limited evidence of carcinogenicity, reproductive/developmental effects, organ toxicity, immune system effects, and genotoxicity to humans. Atrazine, simazine and propazine, and their major chlorinated metabolites, have been reported to induce aromatase (CYP19), the rate-limiting enzyme in the conversion of androgens to estrogens, and to disrupt the hypothalamic–pituitary–gonadal (HPG) axis. Observed changes in progesterone and testosterone concentrations can be partially attributed to triazine-induced alterations in steroidogenic gene expression (Forgacs et al., 2013; Semren et al., 2018).

Chlorophenoxy acids

These herbicides are structural analogs of auxin, a plant growth hormone not found in mammals, that causes uncontrolled and lethal growth in target plants. They are derived from phenoxyacetic acid and include 2,4-dichlorophenoxyacetic acid (2,4-D, **Fig. 12**), 4-chloromethylphenoxyacetic acid (MCPA), 2,4,5-trichlorophenoxyacetic acid (2,4,5-T), diclofop-methyl, mecoprop and dichlorprop. These herbicides can be easily transferred to surface and ground waters due to their polar nature and relatively good solubility.

Chlorophenoxy esters are absorbed from the gastrointestinal tract and excreted in urine, mostly unchanged, within 24 h. They are believed to elicit toxicity by inducing cell membrane damage and uncoupling of oxidative phosphorylation. Though these compounds generally have low mammalian toxicity, oral ingestion of overdoses causes nausea, vomiting, abdominal pain, diarrhea and myotonia. These symptoms are followed by muscle twitching, metabolic acidosis, and a hypermetabolic state with hyperthermia, tachycardia, hypertension, sweating, seizures, and coma. Cases of peripheral neuropathy have been described after exposure to significant amounts of 2,4-D via the skin for several days. 2,4-D has also been associated with adverse changes in lipid concentrations, glucose metabolism, and TSH concentration, which may predispose to heart disease and diabetes (Schreinemachers, 2010).

2,4,5-T has teratogenic properties in experimental animals, as it is inherently contaminated during its synthesis with 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), which is ultimately responsible for neurological and behavioral changes (neurasthenia gravis and depressive syndrome), and hepatic and cutaneous alterations (e.g., porphyria cutanea tarda and chloracne) (Bronstein, 2004).

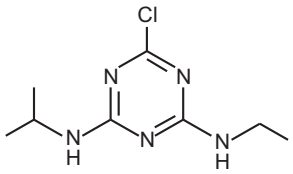


Fig. 11 General chemical structure of Chloro-S-Triazine herbicides.

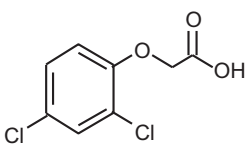


Fig. 12 Chemical structure of 2,4-dichlorophenoxyacetic acid (2,4-D), a phenoxyacetic acid derivative herbicide.

Chronic exposure to chlorophenoxy acids has been associated with non-Hodgkin's lymphoma, soft tissue sarcomas, chronic lymphocytic leukemia and bladder cancer; however the evidence is not entirely consistent (Jayakody et al., 2015).

N-*phosphonomethyl amino acids*

The two compounds in this class are glyphosate (*N*-phosphonomethyl glycine) and glufosinate (*N*-phosphonomethyl homoalanine). Both are broad-spectrum, non-selective systemic herbicides marketed primarily as isopropylamine salt (glyphosate) or ammonium salt (glufosinate). Despite containing a P = O moiety in their molecule (**Fig. 13**), they are not organophosphate esters and do not significantly inhibit AChE.

Glyphosate exerts its herbicidal activity by inhibiting 5-enolpyruvil-shikimate-3-phosphate synthase, a key enzyme in the shikimate pathway essential for the synthesis of aromatic amino acids in plants. However, this metabolic route is not found in mammals. The use of glyphosate spread due to the development of transgenic crops capable of tolerating this herbicide to the extent that it is the most widely used herbicide worldwide. As acute toxicity is low (oral and cutaneous LD_{50} for the isopropylamine salt are >5000 mg/kg), high doses are necessary for intoxication to occur, and mortality is low. Glyphosate is poorly absorbed from the digestive tract, undergoes little biotransformation yielding aminomethylphosphonic acid (AMPA) as the major metabolite, and both are excreted by the feces and urine. Thereby, glyphosate does not accumulate in the body. While mild intoxication leads to transient gastrointestinal distress, moderate or severe intoxication causes upper gastrointestinal bleeding. In severe cases, hypotension, metabolic acidosis and pulmonary, renal, cardiovascular and neurological involvement are observed, with a decrease in the level of consciousness (Marrs and Ballantyne, 2004; Costa, 2008).

Commercial formulations of glyphosate contain other substances that have particular toxicities, such as isopropylamine and the non-ionic surfactant polyoxyethyleneamine (POEA, also known as polyethoxylated tallow amine, which increases the herbicidal effect by promoting the penetration of glyphosate in plants). While the first formulations caused acute and ocular toxicity (eye irritation) after occupational exposure to the concentrated formulation, the change in co-formulants reduced this risk. Studies in animals suggest that the acute toxicity of the glyphosate formulation was due to the surfactant POEA, with an oral LD_{50} of 1200 mg kg^{-1} in rats. In 2016 POEA was banned in all commercial glyphosate formulations in the EU and replaced by propoxylated quaternary ammonium surfactants, which are less toxic to aquatic ecosystems and humans (Mesnage et al., 2019).

Glyphosate does not show teratogenic, developmental or reproductive effects and chronic toxicity studies have shown negligible effects. Genotoxicity and carcinogenicity studies in animals have yielded controversial results and, according to the US-EPA and EFSA, the available evidence does not allow the conclusion that glyphosate poses carcinogenic risk, whereas IARC classified it as probably carcinogenic in humans (Group 2A) (Tarazona et al., 2017). The largest prospective cohort study of licensed pesticide applicators found no association with solid tumors or lymphoid malignancies (Andreotti et al., 2018).

Dietary risk assessment of pesticide residues in food

The application of PPPs in agricultural crops represents a risk to consumer health due to the presence of pesticide residues in fruits and vegetables. The aim of dietary risk assessment is to determine whether the levels of pesticide residues in food commodities and drinking water pose an unacceptable risk for consumers. This assessment combines available pesticide residue data with data from food consumption surveys to estimate potential residue intake by consumers. The latter information is then compared with health-based guidance values (HBGVs) for each pesticide, which are based on referent points (also known as points of departure) derived from regulatory toxicology studies. The HBGVs used in dietary risk assessment of pesticide residues are the Acceptable Daily Intake (ADI, expressed in mg of residue per kg body weight per day) for long-term intake, and the Acute Reference Dose (ARfD, expressed in mg of residue per kg body weight), for short-term dietary intake (<24 h). Based on current scientific knowledge, when the dietary exposure to a substance is found to be lower than, or equal to, its HBGV, the risk to consumer health is low or negligible. Conversely, when the HBGV is exceeded, possible negative health outcomes cannot be excluded (EFSA et al., 2021; FAO, 2017).

In the EU, national and international monitoring programs have been implemented to determine the presence of pesticide residues in food to identify when MRL[1] are exceeded, and to provide representative data to assess the risk to consumer health derived from the presence of pesticide residues in food. According to the results of the last thirteen EU annual reports on pesticide residues in food (from 2007 to 2019), the proportion of cases where MRLs were exceeded remained stable, between 2.5% and 4.5%

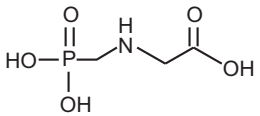


Fig. 13 Chemical structure of glyphosate, an *N*-phosphonomethyl amino acid herbicide.

[1]Highest level of a pesticide residue that is legally tolerated in or on food or feed when pesticides are applied correctly (i.e., according to Good Agricultural Practice) (https://ec.europa.eu/food/plants/pesticides/maximum-residue-levels_en).

Table 4 Pesticide residues in food according to the annual monitoring reports published by EFSA.[a]

Program	*Absence of residues*[b]	*>LOD and <MLR*[c]	*Residues < LRM*	*Residues > LRM*	*Multiple residues*[d]	*No*[e]
2019	56.6%	39.5%	96.1%	3.9%	27.0%	96,302
2018	52.2%	43.3%	95.5%	4.5%	29.1%	91,015
2017	54.1%	41.8%	95.9%	4.1%	27.5%	88,247
2016	50.7%	45.5%	96.2%	3.8%	30.1%	84,657
2015	53.3%	43.9%	97.2%	2.8%	28.0%	84,341
2014	53.6%	43.4%	97.1%	2.9%	28.3%	82,649
2013	54.6%	42.8%	97.4%	2.6%	27.3%	80,967
2012	54.9%	42.2%	97.1%	2.9%	26.1%	78,390
2011	53.4%	44.7%	97.5%	2.5%	26.5%	79,035
2010	NR	NR	97.2%	2.8%	26.6%	77,075
2009	NR	NR	97.4%	2.6%	25.1%	67,978
2008	NR	NR	96.5%	3.5%	27.0%	70,143
2007	NR	NR	96.0%	4.0%	26.2%	74,203

[a]https://efsa.onlinelibrary.wiley.com/doi/toc/10.1002/(ISSN)1831-4732.chemicalresidues-data
[b]Samples free of residues (<LOD).
[c]Residue concentrations above the LOD but below the MRL.
[d]Presence of various pesticide residues in the same food commodity.
[e]Total number of food samples analyzed. NR: not reported.

(Table 4). Nevertheless, if the uncertainty of the analytical determinations were accounted for, the percentage of exceedances of the MRL would be lower.

The analysis of the health-risk to European consumers is usually performed using a deterministic model that bases its calculations on conservative assumptions. A more realistic methodology, based on probabilistic modeling to exposure to multiple chemicals, has recently been developed and implemented by EFSA.

Overall, the risk assessment for the consumption of foods of animal or plant origin in the European population indicates that it is unlikely that short- and long-term exposure to pesticide residues could affect the consumers' health. This underlines the high standards of food safety in the EU.

Conclusion

Pesticides are biologically active substances that undergo strict safety assessment before being authorized for use, to ensure that they will not pose harmful effects to human, animal or environmental health. The limited selectivity of pesticides has increased over time to maximize their specific effect against pests while reducing their impact on non-target beneficial organisms. The necessity of developing and using more selective pesticides has led to advances in their sustainable use. The most at-risk population are agricultural workers who apply pesticides, as they are directly exposed to these chemicals, whereas the population in the vicinity area during and shortly after spraying pesticides show indirect exposure. The general population is also exposed, although to a lesser extent, to significantly lower levels of pesticides through the presence of their residues in food and water. Important achievements in relation to pesticides have been the banning of those that are most toxic to humans, and those that remain for the longest time in the environment. The protection of public health by setting MRLs in food commodities and drinking water, as well as the regular monitoring of pesticide residues in food, represent other important achievements. Nonetheless, the implementation of environmental monitoring programs is also needed.

References

Abou-Donia, M.B., 2003. Organophosphorus ester-induced chronic neurotoxicity. Arch. Environ. Health 58 (8), 484–497. https://doi.org/10.3200/AEOH.58.8.484-497.

Andreotti, G., Koutros, S., Hofmann, J.N., Sandler, D.P., Lubin, J.H., Lynch, C.F., Lerro, C.C., De Roos, A.J., Parks, C.G., Alavanja, M.C., Silverman, D.T., Beane Freeman, L.E., 2018. Glyphosate use and cancer incidence in the agricultural health study. J. Natl. Cancer Inst. 110 (5), 509–516. https://doi.org/10.1093/jnci/djx233.

Atwood, D., Paisley-Jones, C., 2017. Pesticides Industry Sales and Usage. 2008–2012 Market Estimates. U.S. Environmental Protection Agency, Office of Pesticide Programs, Washington. https://www.epa.gov/sites/production/files/2017-01/documents/pesticides-industry-sales-usage-2016_0.pdf.

Bronstein, A.C., 2004. Herbicides. In: Dart, R.C. (Ed.), Medical Toxicology, third ed. Lippincott Williams & Wilkins, Philadelphia, pp. 1515–1529.

Cimino, A.M., Boyles, A.L., Thayer, K.A., Perry, M.J., 2017. Effects of neonicotinoid pesticide exposure on human health: a systematic review. Environ. Health Perspect. 125 (2), 155–162. https://doi.org/10.1289/EHP515.

Colović, M.B., Krstić, D.Z., Lazarević-Pašti, T.D., Bondžić, A.M., Vasić, V.M., 2013. Acetylcholinesterase inhibitors: pharmacology and toxicology. Curr. Neuropharmacol. 11 (3), 315–335. https://doi.org/10.2174/1570159X11311030006.

Costa, L.G., 2008. Toxic effects of pesticides. In: Klaasen, C.D. (Ed.), Casarett and Doull's Toxicology: The Basic Science of Poisons, seventh ed. McGraw-Hill Publishers, New York, pp. 883–930. Chapter 22.

Costa, L.G., 2015. The neurotoxicity of organochlorine and pyrethroid pesticides. Handb. Clin. Neurol. 131, 135–148. https://doi.org/10.1016/B978-0-444-62627-1.00009-3.

Crofton, K.M., 1996. A structure-activity relationship for the neurotoxicity of triazole fungicides. Toxicol. Lett. 84 (3), 155–159. https://doi.org/10.1016/0378-4274(95)03618-0.
Dinis-Oliveira, R.J., Duarte, J.A., Sánchez-Navarro, A., Remião, F., Bastos, M.L., Carvalho, F., 2008. Paraquat poisonings: mechanisms of lung toxicity, clinical features, and treatment. Crit. Rev. Toxicol. 38 (1), 13–71. https://doi.org/10.1080/10408440701669959.
Durand, P., Martin, G., Blondet, A., Gilleron, J., Carette, D., Janczarski, S., Christin, E., Pointis, G., Perrard, M.H., 2017. Effects of low doses of carbendazim or iprodione either separately or in mixture on the pubertal rat seminiferous epithelium: an ex vivo study. Toxicol. In Vitro 45 (Pt 3), 366–373. https://doi.org/10.1016/j.tiv.2017.05.022.
EFSA—European Food Safety Authority, Carrasco Cabrera, L., Medina Pastor, P., 2021. The 2019 European Union Report on pesticide residues in food. Eur. Food Saf. Auth. 19 (4), e06491. https://doi.org/10.2903/j.efsa.2021.6491.
EPA—Environmental Protection Agency, 2020. Pesticide Registration. https://www.epa.gov/pesticide-registration/about-pesticide-registration.
Eurostat—European Statistic, 2021. Agriculture and the Environment. Pesticide sales. https://ec.europa.eu/eurostat/web/agriculture/data/database.
FAO & WHO—Food and Agriculture Organization of the United Nations and World Health Organization, 2016. International Code of Conduct on Pesticide Management Guidelines on Highly Hazardous Pesticides. http://apps.who.int/iris/bitstream/handle/10665/205561/9789241510417_eng.pdf?sequence=1.
FAO—Food and Agriculture Organization of the United Nations, 2002. International Code of Conduct on the Distribution and Use of Pesticides (Revised Version) (Adopted by the Hundred and Twenty-Third Session of the FAO Council in November 2002), p. 6. Rome. https://web.archive.org/web/20130404190746/http:/www.fao.org/WAICENT/FAOINFO/AGRICULT/AGP/AGPP/Pesticid/Code/Download/code.pdf.
FAO—Food and Agriculture Organization of the United Nations, 2017. Submission and Evaluation of Pesticide Residues Data for the Estimation of Maximum Residue Levels in Food and Feed, third ed. Pesticide Residues, p. 286. FAO Plant Production and Protection Paper 225. http://www.fao.org/documents/card/es/c/3829b88a-c721-434f-92c1-08e1675e8141/.
Filipov, N.M., Lawrence, D.A., 2001. Developmental toxicity of a triazole fungicide: consideration of interorgan communication. Toxicol. Sci. 62 (2), 185–186. https://doi.org/10.1093/toxsci/62.2.185.
Forgacs, A.L., D'Souza, M.L., Huhtaniemi, I.T., Rahman, N.A., Zacharewski, T.R., 2013. Triazine herbicides and their chlorometabolites alter steroidogenesis in BLTK1 murine Leydig cells. Toxicol. Sci. 134 (1), 155–167. https://doi.org/10.1093/toxsci/kft096.
Gaforio, J.J., Visioli, F., Alarcón-de-la-Lastra, C., Castañer, O., Delgado-Rodríguez, M., Fitó, M., Hernández, A.F., Huertas, J.R., Martínez-González, M.A., Menendez, J.A., Osada, J., Papadaki, A., Parrón, T., Pereira, J.E., Rosillo, M.A., Sánchez-Quesada, C., Schwingshackl, L., Toledo, E., Tsatsakis, A.M., 2019. Virgin olive oil and health: summary of the III International Conference on Virgin Olive Oil and Health Consensus Report, JAEN (Spain) 2018. Nutrients 11 (9), 2039. https://doi.org/10.3390/nu11092039.
Hernández, A.F., Pla, A., 2007. Toxicidad de los plaguicidas. In: Guillen Enríquez, J., Serrano-Ramírez, J.L. (Eds.), Respuesta ante las intoxicaciones agudas por plaguicidas. Manual para el sanitario. Vigilancia Epidemiológica. Consejería de Salud (Junta de Andalucía), Spain, ISBN 84-8486-089-2, pp. 7–27. Available online at. https://www.juntadeandalucia.es/export/drupaljda/salud_5af0653873762_intoxicaciones_agudas_plaguicidas.pdf.
Hernández, A.F., Parrón, T., Tsatsakis, A.M., Requena, M., Alarcón, R., López-Guarnido, O., 2013. Toxic effects of pesticide mixtures at a molecular level: their relevance to human health. Toxicology 307, 136–145. https://doi.org/10.1016/j.tox.2012.06.009.
Hernández, A.F., Lozano-Paniagua, D., González-Alzaga, B., Kavvalakis, M.P., Tzatzarakis, M.N., López-Flores, I., Aguilar-Garduño, C., Caparros-Gonzalez, R.A., Tsatsakis, A.M., Lacasaña, M., 2019. Biomonitoring of common organophosphate metabolites in hair and urine of children from an agricultural community. Environ. Int. 131, 104997. https://doi.org/10.1016/j.envint.2019.104997.
Houchat, J.N., Cartereau, A., Le Mauff, A., Taillebois, E., Thany, S.H., 2020. An overview on the effect of neonicotinoid insecticides on mammalian cholinergic functions through the activation of neuronal nicotinic acetylcholine receptors. Int. J. Environ. Res. Publ. Health 17 (9), 3222. https://doi.org/10.3390/ijerph17093222.
IARC—International Agency for Research on Cancer, 2018. Working Group on the Evaluation of Carcinogenic Risks to Humans. DDT, Lindane, and 2,4-D. International Agency for Research on Cancer, Lyon (FR). PMID: 29912510. Bookshelf ID: NBK507424.
Janz, D.M., 2014. Dithiocarbamates. In: Wexler, P. (Ed.), Encyclopedia of Toxicology, third ed. Academic Press, pp. 212–214. https://doi.org/10.1016/B978-0-12-386454-3.00139-1.
Jayakody, N., Harris, E.C., Coggon, D., 2015. Phenoxy herbicides, soft-tissue sarcoma and non-hodgkin lymphoma: a systematic review of evidence from cohort and case-control studies. Br. Med. Bull. 114 (1), 75–94. https://doi.org/10.1093/bmb/ldv008.
Jayaraj, R., Megha, P., Sreedev, P., 2016. Organochlorine pesticides, their toxic effects on living organisms and their fate in the environment. Interdiscip. Toxicol. 9 (3–4), 90–100. https://doi.org/10.1515/intox-2016-0012.
Lushchak, V.I., Matviishyn, T.M., Husak, V.V., Storey, J.M., Storey, K.B., 2018. Pesticide toxicity: a mechanistic approach. EXCLI J. 17, 1101–1136. https://doi.org/10.17179/excli2018-1710.
Marrs, T.T., Ballantyne, B., 2004. Pesticide Toxicology and International Regulation. John Wiley & Sons, Chichester. https://doi.org/10.1002/0470091673.
Mesnage, R., Benbrook, C., Antoniou, M.N., 2019. Insight into the confusion over surfactant co-formulants in glyphosate-based herbicides. Food Chem. Toxicol. 128, 137–145. https://doi.org/10.1016/j.fct.2019.03.053.
NPIC—National Pesticide Information Center—EPA Pesticide Regulation, (2019). http://npic.orst.edu/reg/epareg.html.
O'Malley, M., 1997. Clinical evaluation of pesticide exposure and poisonings. Lancet 349 (9059), 1161–1166. https://doi.org/10.1016/S0140-6736(96)07222-4.
Pisani, C., Voisin, S., Arafah, K., Durand, P., Perrard, M.H., Guichaoua, M.R., Bulet, P., Prat, O., 2016. Ex vivo assessment of testicular toxicity induced by carbendazim and iprodione, alone or in a mixture. ALTEX 33 (4), 393–413. https://doi.org/10.14573/altex.1601253.
Roser, M., 2019. Pesticides. OurWorldInData.org. Retrieved from. https://ourworldindata.org/pesticides. Licence CCBY.
Schreinemachers, D.M., 2010. Perturbation of lipids and glucose metabolism associated with previous 2,4-D exposure: a cross-sectional study of NHANES III data, 1988–1994. Environ. Health 9, 11. https://doi.org/10.1186/1476-069X-9-11.
Selvam, V., Srinivasan, S., 2019. Neonicotinoid poisoning and management. Indian J. Crit. Care Med. 23 (Suppl. 4), S260–S262. https://doi.org/10.5005/jp-journals-10071-23308.
Semren, T.Ž., Žunec, S., Pizent, A., 2018. Oxidative stress in triazine pesticide toxicity: a review of the main biomarker findings. Arh. Hig. Rad. Toksikol. 69 (2), 109–125. https://doi.org/10.2478/aiht-2018-69-3118.
Simon-Delso, N., Amaral-Rogers, V., Belzunces, L.P., Bonmatin, J.M., Chagnon, M., Downs, C., Furlan, L., Gibbons, D.W., Giorio, C., Girolami, V., Goulson, D., Kreutzweiser, D.P., Krupke, C.H., Liess, M., Long, E., McField, M., Mineau, P., Mitchell, E.A., Morrissey, C.A., Noome, D.A., et al., 2015. Systemic insecticides (neonicotinoids and fipronil): trends, uses, mode of action and metabolites. Environ. Sci. Pollut. Res. Int. 22 (1), 5–34. https://doi.org/10.1007/s11356-014-3470-y.
Tarazona, J.V., Court-Marques, D., Tiramani, M., Reich, H., Pfeil, R., Istace, F., Crivellente, F., 2017. Glyphosate toxicity and carcinogenicity: a review of the scientific basis of the European Union assessment and its differences with IARC. Arch. Toxicol. 91 (8), 2723–2743. https://doi.org/10.1007/s00204-017-1962-5.
WHO—World Health Organization, 2005. Safety of Pyrethroids for Public Health Use. WHO Pesticide Evaluation Scheme (WHOPES) & Protection of the Human Environment Program on Chemical Safety (PCS). http://libdoc.who.int/hq/2005/WHO_CDS_WHOPES_GCDPP_2005.10.pdf.
WHO—World Health Organization, 2020. WHO Recommended Classification of Pesticides by Hazard and Guidelines to Classification, 2019 ed. WHO, Geneva. Licence: CC BY-NC-SA 3.0 IGO. https://apps.who.int/iris/rest/bitstreams/1278712/retrieve.
Ye, L., Su, Z.J., Ge, R.S., 2011. Inhibitors of testosterone biosynthetic and metabolic activation enzymes. Molecules 16 (12), 9983–10001. https://doi.org/10.3390/molecules16129983.

Fructose: Metabolism and health effects

Chrysanthi Stylianopoulou, Nutrition Services, Veria, Greece

© 2023 Elsevier Ltd. All rights reserved.

This is an update of N.L. Keim, P.J. Havel, Fructose: Absorption and Metabolism, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 361–365, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00128-8.

Introduction 390
Properties and sources of fructose 390
High fructose corn syrup (HFCS) 390
Fructose consumption 391
Physical properties 391
Chemical properties 392
Fructose absorption 392
Fructose metabolism 392
Hepatic fructose metabolism 392
Fructose vs. glucose metabolism 392
Extrahepatic fructose metabolism 393
Free vs bound fructose metabolism 393
Health implications 394
Fructose consumption and dyslipidemia 394
Fructose consumption and obesity 394
Fructose consumption and insulin resistance and type 2 diabetes 394
Fructose consumption and coronary heart disease risk 394
Fructose consumption and caries 394
Inborn errors of fructose metabolism 395
Essential or benign fructosuria 395
Hereditary fructose intolerance or hereditary fructosuria 395
Fructose-1,6-bisphosphatease (FBPase) deficiency 396
Summary 396
References 396
Relevant websites 396

Key points

- Outline the sources of fructose in the food supply
- Describe the production and uses of high fructose corn syrup in the food industry
- Outline the estimated intake of fructose and sugars in the Western diets
- Briefly outline the physical and chemical properties of fructose
- Describe the absorption and metabolism of fructose
- Discuss the health implications of increased fructose consumption, namely the effects on dyslipidemia and lipid accumulation, obesity, insulin resistance and type 2 diabetes, coronary heart disease risk and dental caries
- Describe inborn errors of fructose metabolism

Glossary

Glycosidic bond A covalent bond that joins the hemiacetal group of a saccharide molecule and the hydroxyl group of some organic compound (e.g., an alcohol)

Nomenclature

mM Millimolar, a molar concentration measurement unit.

Introduction

The most frequently encountered naturally occurring sugars are sucrose, fructose and glucose (dextrose). Among these, fructose (also known as levulose) and glucose (also known as dextrose) are simple monosaccharides, while sucrose (also known as table sugar) is a disaccharide formed by a fructose and a glucose molecule linked together through a glycosidic bond. Fructose is the sweetest of all monosaccharides and it is naturally present alongside glucose in ripening fruits and honey. It can be used as fuel for the human body and it is a major constituent of industrial food sweeteners (Tappy and Lê, 2010; Scientific Advisory Committee on Nutrition, 2015).

Sugar consumption became increasingly popular after the 18th century, when sugar cane became abundant and extraction methods of refined sugars from sugar cane became available (Tappy and Lê, 2010). Sugar consumption increased dramatically during the 20th century and became a major component of the human diet.

Properties and sources of fructose

Natural sources of dietary fructose are fruits, fruit juices, honey and some root vegetables. In these foods, fructose is found unbound and also as a component of sucrose (**Table 1**). However, the primary source of fructose in Western diets is sugars added to baked goods, candies, soft drinks, and other beverages sweetened with sucrose and high fructose corn syrup (HFCS), also known as added sugars. Both sucrose and fructose are used extensively in the food industry as additives to provide sweetness, texture, and palatability. These sugars also contribute to the appearance, preservation, and energy content of the food product.

High fructose corn syrup (HFCS)

Until the 1960s sucrose was the sole sweetener ingested, with only small amounts of glucose and fructose consumed. Technological advances in the food industry in the 1960s made possible the extraction of starch from corn, its hydrolysis to glucose and subsequent conversion of part of glucose into fructose. Consequently, sweeteners from corn were produced, among which was HFCS. HFCS has useful physical and functional characteristics, among which are high sweetening power, flavor enhancement, long-lasting moisture preservation, long self-life maintenance and improved texture. These characteristics contributed to the gradual substitution of HFCS at the expense of sucrose as an industrial nutritive sweetener. HFCS is produced with varying fructose-to-glucose ratios. The predominant forms of HFCS used are HFCS-55, which contains 55% fructose, and HFCS-42, which contains 42% fructose. HFCS and sucrose contain comparable amounts of glucose and fructose. The difference of the two sweeteners is that in HFCS glucose and fructose exist in solution in their free form, while in sucrose the two sugars exist as a disaccharide.

Table 1 Sugar content of selected common plant foods (g sugar per 100 g food).

Food item	*Total carbohydrate*	*Total sugars*	*Free fructose*	*Free glucose*	*Sucrose*	*Fructose to glucose ratio*	*Sucrose as a % of total sugars*
Apple	13.8	10.4	5.9	2.4	2.1	2	19.9
Apricot	11.1	9.2	0.9	2.4	5.9	0.7	63.5
Banana	22.8	12.2	4.9	5	2.4	1	20
Beet, red	9.6	6.8	0.1	0.1	6.5	1	96.2
Carrot	9.6	4.7	0.6	0.6	3.6	1	70
Corn, sweet	19	3.2	0.5	0.5	2.1	1	64
Fig, dried	63.9	47.9	22.9	24.8	0.07	0.93	0.001
Grapes	18.1	15.5	8.1	7.2	0.2	1.1	1
Peach	9.5	8.4	1.5	2	4.8	0.9	56.7
Pear	15.5	9.8	6.2	2.8	0.8	2.1	8
Pineapple	13.1	9.9	2.1	1.7	6	1.1	60.8
Plum	11.4	9.9	3.1	5.1	1.6	0.66	0.16
Red pepper, sweet	6	4.2	2.3	2.3	0	1.2	0
Onion, sweet	7.6	5	2	1	0.7	0.9	14.3
Sweet potato	20.1	4.2	0.7	2.3	2.5	0.9	60.3
Yam	27.9	0.5	Trace	Trace	Trace	Not applicable	Trace
Sugar cane		13–18	0.2–1.0	0.2–1.0	11–16	1	100
Sugar beet		17–18	0.1–0.5	0.1–0.5	16–17	1	100

Data obtained from https://nal.usda.gov/legacy/fnic/food-composition, Nutritive and Nonnutritive Sweetener Resources accessed on February 9, 2022.

Table 2 Relative sweetness of nutritive sweeteners.

Type of sweetener	*Relative sweetness to sucrose*[a]
HFCS[b] (42% fructose)	100
HFCS[b] (55% fructose)	100–110
HFCS[b] (90% fructose)	120–160
Lactose	40
Dextrose	70–80
Fructose	150–170
Sucrose	100

[a]Compared to sucrose in a 15% solution.
[b]HFCS high fructose corn syrup.
Data from Pancoast and Junk (1980).

HFCS is produced by hydrolyzing corn starch to glucose, using α-amylase and glucoamylase. This process is followed by treatment with glucose isomerase, which isomerizes glucose to a mixture of 90% fructose and 10% glucose (HFCS-90). Subsequently, HFCS-90 is mixed with glucose syrup to produce HFCS-55 and HFCS-42. Both HFCS-55 and HFCS-42 have distinct physical and chemical properties, which enable their use in specific food manufacturing areas. HFCS-55 has high fructose content, is sweeter than sucrose and it is utilized as a sweetener in sugar-sweetened beverages. HFCS-42 is mildly sweet and does not mask natural food flavors. Therefore, it is utilized in canned and baked goods, and in dairy processed foods. Relative sweetness of various nutritive sweeteners is outlined in **Table 2**.

Fructose consumption

The major sources of fructose in the United States diet are HFCS and sucrose, which are used as caloric sweeteners in the food industry, followed by other foods sources, namely fruits, vegetables, honey and other syrups. Fructose consumption for the United States can be calculated from the United Stated Department of Agriculture (USDA) Loss-Adjusted Food Availability Database. According to this database, HFCS consumption in 1970 was close to zero and gradually increased with the increasing use of HFCS in the food industry (usda). The increase in HFCS intake was coupled with a decrease in the consumption of other nutritive sweeteners, mainly sucrose, peaked in 1999 and has been declining ever since. Overall, total sugar consumption did not change significantly from 1970 to 2020 and has remained less than 90 g per day in the United States, with HFCS consumption being approximately 26 g per day (usda).

Fructose consumption in the United States can be also calculated from survey data, using the third National Health and Examination Survey (NHANES) which provided data from single 24 h dietary recalls administered in 21,483 children and adults (Vos et al., 2008). According to this survey, the mean consumption of fructose was 55 g per day and 10% of total energy intake. It was observed that adolescents had higher fructose consumption at 73 g per day, which was 12% of total energy intake. Furthermore, this study showed that the largest source of fructose was sugar-sweetened beverages (30% of total energy intake), grains (22% of total energy intake) and fruit or fruit juice (19% of total energy intake).

A recent review of sugar consumption from nationally representative dietary surveys (24 h dietary recalls or dietary records) of developed countries suggests that total sugar consumption as a percentage of energy was the highest in infants and young children, with a mean value of 28% and gradually decreased in adults to 20% (Newens and Walton, 2016). Added sugars intakes in preschool children were reported as higher than 10% in Australia, the United Kingdom, and the United States. In school-aged children and adolescents, mean added sugars intakes were higher than 19%, and gradually decreased during adulthood to lower than 10% of total energy intake (Newens and Walton, 2016). Therefore, it was concluded that added sugar intakes were the largest among school-aged children and adolescents.

Another review of worldwide trends in intake of dietary sugars of developed countries concluded that dietary sugar consumption had been either decreasing or stable in most of the countries studied (Wittekind and Walton, 2014). Multiple national surveys revealed that dietary sugar intake varied by subpopulation, and an increase of dietary sugar intake was observed only in a few countries and specific subpopulations.

Physical properties

Fructose is a white solid, and it is the most water soluble of all sugars. It dissolves rapidly in small amounts of water, a characteristic applicable in candy making. The high solubility of fructose produces softer sweets than those containing other sugars. This property prevents fructose from forming ice crystals and is responsible for its hydroscopicity and humectancy. Therefore, fructose may improve the quality and texture, as well as enhance the self life of food products.

Fructose is the sweetest of all natural sugars and this is the main reason for its utilization in the food and beverage industry as a sweetener. Sweetness ratings of fructose are between 150% and 170%, compared to the standard, sucrose, rated at 100%. Heating reduces the sweetness of fructose. Fructose can be used as a flavor enhancer, since it displays synergistic sweetness enhancing effects when blended with other sweeteners and amplifies sweetness.

Chemical properties

Fructose is a 6 carbon polyhydroxy ketone, shares the same molecular formula with glucose, but has distinct structural characteristics, which means that it is an isomer of glucose. The structure of fructose, similarly to all monosaccharides, can be presented as a 6-carbon linear molecule (open chain). Fructose also forms cyclic structures called hemiketals due to internal hydrogen bonding. When it is in solution, fructose exists in equilibrium mixture of its cyclic structures and open chain conformation. Fructose undergoes the chemical reactions of ketones.

Fructose absorption

Dietary fructose is ingested either as a simple monosaccharide (pure fructose or HFCS) or as part of the disaccharide sucrose, which is digested at the intestinal brush border. Subsequently, it is transported by a fructose-specific hexose transporter, GLUT-5, located at the apical pole of the enterocyte and it is mainly expressed in intestinal epithelial cells. Fructose transport, contrary to glucose, does not require energy in the form of adenosine triphosphate (ATP) and does not depend on sodium. When fructose is inside the enterocyte, it is absorbed by facilitated diffusion using a basolateral transporter, GLUT-2, which also transports glucose and galactose. GLUT-2 is highly expressed in intestinal epithelial cells, hepatocytes, pancreatic β-cells, and kidney epithelial cells. When fructose is absorbed by the intestine, it arrives at the liver through the hepatic portal vein and then it is metabolized in hepatocytes.

However, the human body appears to have a limited capacity to absorb fructose, especially when ingested without concomitant glucose ingestion. Consumption of a large amount of fructose may result in abdominal bloating, diarrhea, and flatulence. The intestinal absorptive capacity for fructose increases when glucose is consumed along with fructose. Thus, co-ingesting glucose to roughly balance fructose, as occurs when most fruits or sucrose are consumed, largely alleviates problems of fructose malabsorption.

Fructose metabolism

Hepatic fructose metabolism

Following absorption, fructose is quickly transported into hepatic cells through the GLUT-2 transporter (Tappy and Lê, 2010). Fructose is readily extracted by the liver because of the presence of an active hepatic enzyme system for metabolizing fructose, and the majority of ingested fructose is cleared in a single pass through the liver (Tappy and Lê, 2010). Thus, the concentration of fructose circulating in the blood is low after consumption of moderate amounts of fructose. Fructose enters the glycolytic pathway through conversion to intermediate compounds in the liver (**Fig. 1**), and as a result, fructose is not generally available for uptake by other tissues.

In the liver, fructose is phosphorylated and forms fructose-1-phosphate. This reaction requires ATP and is catalyzed by fructokinase, an enzyme with high affinity and specificity for fructose. Fructose-1-phosphate is then cleaved by hepatic aldolase (aldolase B) to form dihydroxyacetone phosphate and glyceraldehyde. Dihydroxyacetone phosphate is an intermediate metabolite in both the gluconeogenic and glycolytic pathways. Thus, a portion of the original fructose carbon structure forms glucose, and, in fact, a small increase in circulating glucose occurs after ingestion of fructose. The glyceraldehyde intermediate is phosphorylated by triokinase to form glyceraldehyde-3-phosphate, another intermediate in the glycolytic pathway. The triose phosphate compounds provide substrates for glycolysis and oxidative metabolism. Nonetheless, most fructose gets converted into glucose. With the formation of the triose phosphates, the metabolism of fructose and glucose converges.

Increased fructose consumption may saturate the glycolytic pathway with intermediates, stimulating de novo lipogenesis and triacylglycerol synthesis. Dihydroxyacetone phosphate may be converted to glycerol, and acetyl CoA may enter the lipogenic pathway to form fatty acids that are then esterified to the glycerol molecule to form triacylglycerols.

The metabolic fate of ingested fructose is the following: in hepatic tissue, the largest part of fructose (∼67%) is converted to glucose (Tappy and Lê, 2010). The majority (∼50%) is released as plasma glucose and a small part (∼17%) is stored as glycogen in the liver. A proportion of the fructose (∼25%) is converted into lactate in the enterocytes and in the liver, and a minor portion of fructose is converted into fatty acids, through the process of de novo lipogenesis.

Fructose vs. glucose metabolism

There are important differences in fructose and glucose metabolism that impact both carbohydrate and lipid metabolism. Initially, when glucose enters the glycolytic pathway, it is controlled by glucokinase, an enzyme with high affinity for glucose. The rate of glucose phosphorylation, which is catalyzed by glucokinase, changes with respect to fluctuations in glucose concentration. Glucose 6-phosphate is then converted to fructose 6-phosphate, a reaction catalyzed by the allosterically regulated enzyme

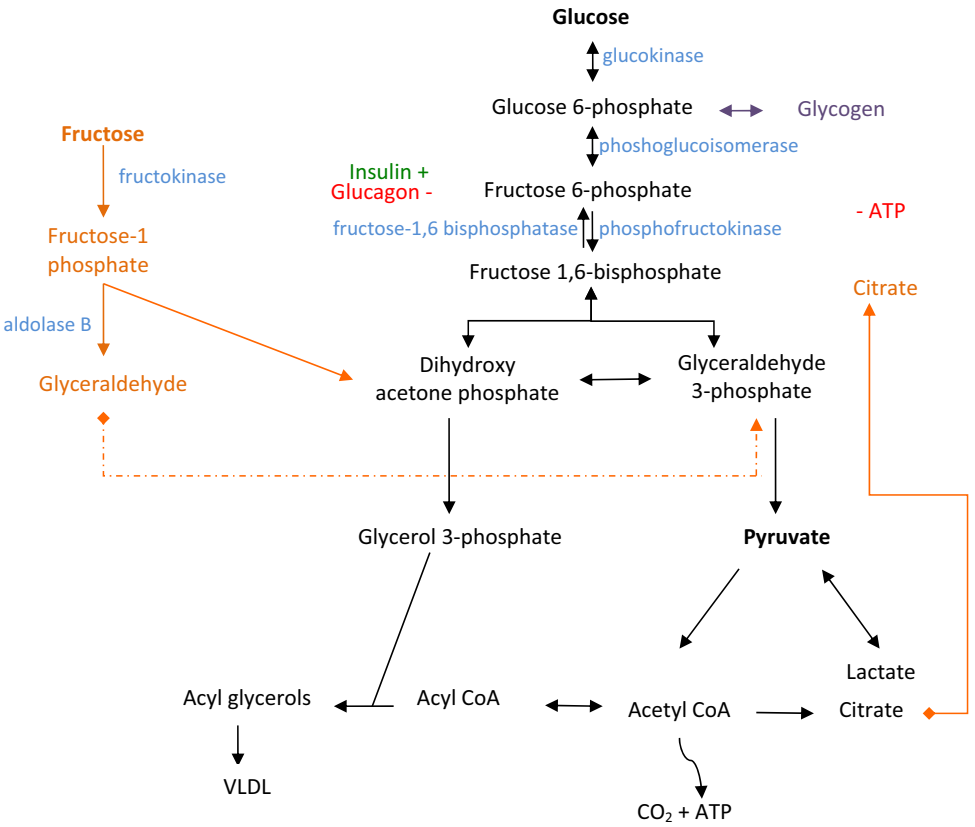


Fig. 1 Utilization of fructose and glucose in the liver. Adapted from Elliott et al. (2002).

phosphofructokinase (PFK), which is the most important control point in the glycolytic sequence. Among the multiple effectors of PFK are ATP and citrate; these products of glucose oxidation exert an inhibitory effect on the enzyme. The allosteric inhibition of PFK effectively reduces the rate of glycolysis and decreases hepatic glucose uptake overall. Furthermore, conversion of glucose to pyruvate is controlled by insulin and energy requirements of the cell. On the contrary, fructose metabolism does not depend on insulin, and it occurs rapidly due to the lack of the inhibitory action of ATP and citrate. Consequently, there is a short-lived decrease of free phosphate and ATP in liver cells after fructose ingestion. Furthermore, large amounts of lactate are produced after ingestion of fructose molecules in comparison to equal amounts of glucose molecules.

Extrahepatic fructose metabolism

Extrahepatic fructose metabolism does not occur under normal conditions, since extrahepatic cells do not express fructokinase for the phosphorylation of fructose (Tappy and Lê, 2010). After parenteral fructose administration, fructose concentration in the plasma raises 1–2 mM. Even high-dose fructose infusions that raise plasma fructose to 3 mM have shown to raise fructose uptake by the kidney not more than 20% of total fructose turnover (Tappy and Lê, 2010).

Free vs bound fructose metabolism

It has been proposed that free fructose, in the form of HFCS, has different effects compared to bound fructose, in the form of sucrose. Current data from short-term studies in healthy and type 2 diabetic volunteers do not support the hypothesis that free fructose and bound fructose are metabolically different. Short-term ingestion of fructose in the form of HFCS has been shown to have similar metabolic responses (fasting plasma glucose, insulin, leptin and ghrelin concentrations) to fructose in the form of sucrose in healthy volunteers and type 2 diabetes (Tappy and Lê, 2010). Furthermore, no differences were observed between consumption of preload drinks containing sucrose, HFCS or milk on insulin, glucose, glucagon-like peptide-1 (GLP-1), ghrelin and appetite score in lean volunteers (Tappy and Lê, 2010). There is no evidence to date to support the hypothesis that free fructose in the form of HFCS is metabolically different from bound fructose in the form of sucrose.

Health implications

As fructose is metabolized in a different manner than glucose and other monosaccharides, it has been proposed that it promotes lipid accumulation, insulin resistance, obesity and other metabolic abnormalities. Most of the accusing evidence originates from consumption of fructose in the form of sweetened beverages.

Fructose consumption and dyslipidemia

Some studies have shown that de novo lipogenesis is increased after acute fructose intake (Tappy and Lê, 2010). Several randomized controlled trials have investigated the effects of short-term fructose consumption on fasting blood lipids and other measures of adiposity and lipid dysregulation at various levels of dietary intake. Some short-term studies showed increased plasma triacylglycerols after consumption of fructose sweetened beverages, in comparison to non-fructose sweetened beverages, at fructose intake 8–30% of total energy intake (Tappy and Lê, 2010). Randomized trials have reported no adverse effects of fructose vs non-fructose intake on total cholesterol, low density lipoprotein (LDL) or high density lipoprotein (HDL) at fructose intake of up to 30% total energy intake (Scientific Advisory Committee on Nutrition, 2015).

Most of these studies have used extremely high fructose intakes (25–30% of total dietary intake), which far exceed the normal quantities of fructose in the human diet. Furthermore, most studies compare pure fructose with non-fructose carbohydrate, which is different than the sucrose or HFCS consumed in the typical diet (Rippe and Angelopoulos, 2013). Dietary sources of fructose contain comparable amounts of fructose and glucose and a diet which is fructose dominant is rarely encountered.

It can be concluded that some short-term studies have reported unfavorable effects of fructose consumption on plasma triacylglycerols. Nonetheless, there is lack of concrete evidence that fructose consumption at typical population levels may contribute to lipid dysregulation. Further studies are needed to address this hypothesis.

Fructose consumption and obesity

A large proportion of fructose in the diet originates from sugar-sweetened beverages. Foods high in sugars are highly palatable, and can create a potential risk for energy overconsumption and weight gain. Furthermore, consumption of sugar-sweetened drinks may contribute to weight gain because of the low satiety of liquid foods.

Most of the evidence on fructose-sweetened foods and beverages is derived from cohort studies. Systematic reviews of large cross-sectional studies investigating the effect of consumption of fructose-sweetened beverages and the risk of weight gain have shown conflicting results (Tappy and Lê, 2010; Rippe and Angelopoulos, 2013; Scientific Advisory Committee on Nutrition, 2015). There is a lack of experimental studies to examine the impact of individual sugars, such as glucose, fructose or sucrose on body weight.

Fructose consumption and insulin resistance and type 2 diabetes

Historically, in the nutritional management of diabetes mellitus, the ingestion of fructose was recommended as a sweetener for patients because it causes smaller increases in blood glucose following ingestion, compared to similar amounts of glucose, sucrose, or starches. In fact, fructose, in small quantities, increases the hepatic uptake of glucose and promotes glycogen storage, probably by stimulating the activity of hepatic glucokinase. Also, in individuals with type 2 diabetes, the addition of a small amount of fructose to an oral glucose tolerance load improves the glycemic response, indicating improved glycemic control (Scientific Advisory Committee on Nutrition, 2015).

Prolonged feeding of fructose or sucrose to animals impairs insulin signaling and induces insulin resistance (Baena et al., 2016). Less is known about the effect of fructose ingestion on glucose tolerance and insulin resistance in humans, as the scientific literature contains conflicting results (Tappy and Lê, 2010; Scientific Advisory Committee on Nutrition, 2015). Studies of long-term beverage consumption using more direct measures of insulin sensitivity are clearly warranted.

Fructose consumption and coronary heart disease risk

Prospective cohort studies have been analyzed to investigate the relationships of sugar-sweetened beverages consumption with biomarkers of coronary heart disease. Sugar-sweetened beverage consumption has been positively associated with increased risk of coronary heart disease and adverse changes in inflammatory factors (Scientific Advisory Committee on Nutrition, 2015).

Meta-analyses of prospective cohort studies have reported an increased risk of hypertension with higher intake of sugar-sweetened beverages (Malik and Hu, 2019). Further experimental studies are needed to confirm these findings and examine the impact of individual sugars, such as glucose, fructose, or sucrose on coronary heart disease.

Fructose consumption and caries

Sucrose and fructose are considered the most cariogenic sugars. The dental plaque formed while sucrose is present has been demonstrated to have lower mineral content of calcium, inorganic phosphate, and fluoride, all of which re-mineralize enamel and dentin (Fidler Mis et al., 2017). Fructose and glucose are more cariogenic than starches because they are metabolized quickly and produce

a higher decrease in pH (Fidler Mis et al., 2017). Saliva can neutralize and repair the loss of minerals resulting from this process (Scientific Advisory Committee on Nutrition, 2015). When this loss of minerals exceeds the capacity of saliva to re-mineralize the tissues, tooth decay occurs (Scientific Advisory Committee on Nutrition, 2015). The amount of dietary sugar intake has been shown to affect caries formation, along with oral hygiene, fluoride supplementation and use. The incidence of dental caries has decreased worldwide, because of the increased use of fluoride and improvement of oral hygiene (World Health Organization, 2015).

Prospective cohort studies show that the quantity and frequency of dietary sugar and sugar-containing beverage intake play a significant role in tooth decay and are positively associated with high risk of caries development (World Health Organization, 2015; Scientific Advisory Committee on Nutrition, 2015; Fidler Mis et al., 2017). Data indicate that dental caries are less common in countries where sugar intake is below 15–20 kg per person per year (World Health Organization, 2003). This figure translates to a daily intake of 40–55 g per person and to a total energy intake of 6–10% (World Health Organization, 2003). Worldwide ecological studies and randomized controlled trials have consistently shown a relationship between high sugar consumption, mainly sucrose, and tooth demineralization and tooth decay (Scientific Advisory Committee on Nutrition, 2015; World Health Organization, 2003). When annual sugar consumption exceeds 15 kg per person per year, dental caries also increase (World Health Organization, 2003).

Inborn errors of fructose metabolism

Several genetically based abnormalities in fructose metabolism have been described in humans. They are summarized in **Table 3**.

Essential or benign fructosuria

Essential fructosuria is an autosomal recessive disorder of fructose metabolism which occurs due to a deficiency of fructokinase, the enzyme which catalyzes the conversion of fructose to fructose 1-phosphate. Consequently, fructose cannot be contained in the liver in the form of fructose 1-phosphate and accumulates in the blood. Fructose is then partly excreted in the urine and the remaining is converted to fructose-6-phosphate in adipose tissue and muscle, a reaction catalyzed by hexokinase (McGrane, 2012). Therefore, blood fructose values decline. This disease is not harmful and does not have any symptoms.

Hereditary fructose intolerance or hereditary fructosuria

Fructose intolerance is an autosomal recessive disease, caused by a genetic defect of fructose 1-phosphate aldolase (aldolase B) in the liver (McGrane, 2012). The symptoms of aldolase B deficiency usually start when the infant is exposed to fructose or sucrose through their diet at the time of weaning. Fructose and sucrose consumption results in hepatic and renal accumulation of fructose 1-phosphate and subsequent toxicity. Furthermore, fructose 1-phosphate causes fructokinase inhibition and decreased fructose uptake in the liver, resulting in fructosemia (Malik and Hu, 2015). If left untreated, depletion of the cellular phosphate pool and severe organ damage can occur.

Fructose intolerance in characterized by abdominal pain, nausea, vomiting, and metabolic abnormalities, such as hypoglycemia (Malik and Hu, 2015). Early diagnosis of fructose intolerance is essential as the toxic effects of fructose can be fatal. The primary therapy is elimination of all fructose sources, intravenous glucose to alleviate hypoglycemia and supplementation with folate and vitamin C (Tran, 2017).

Table 3 Inborn errors of fructose metabolism.

Name of disease	*Enzyme defect*	*Main clinical symptoms*	*Treatment*
Essential fructosuria	Fructokinase	Asymptomatic	No treatment
Hereditary fructose intolerance	Aldolase B	Abdominal pain	Withdrawal of all sources of fructose
		Nausea	Intravenous glucose for hypoglycemia
		Hypoglycemia symptoms	Supplementation with folate and vitamin C
		Shock-like syndrome after fructose ingestion	Avoiding catabolic triggers
Fructose-1,6-bisphosphatase deficiency	Fructose-1,6-bisphosphatase	Life-threatening episodes of hypoglycemia	Frequent feeding (avoid overfeeding)
		Coma triggered by a febrile episode, fasting orLarge amount of fructose ingestion	Hypoglycemia correction with oral and/or IV glucose
			In the absence of triggers no carbohydrate supplements needed
			Restriction of fructose, sucrose and sorbitol

Adapted from Tran (2017).

Fructose-1,6-bisphosphatease (FBPase) deficiency

FBPase deficiency is an autosomal recessive disease, caused by a genetic defect of FBPase (Tran, 2017). FBPase has a critical role in the enzyme complex regulating glycolysis and gluconeogenesis. FBPase deficiency inhibits glucose production from lactate, glycerol and gluconeogenic amino acids. It is diagnosed in infants within the first weeks of life, when they become dependent on carbohydrate metabolism. When infants are fasting, normoglycemia is maintained by glycogenolysis and depends on the availability of hepatic glycogen. FBPase deficiency results in hypoglycemia and inability to synthesize glucose from gluconeogenic substrates.

Deficient individuals exhibit hypoglycemia, acidosis, ketonuria, hyperventilation, and often hypotonia and hepatomegaly. Furthermore, urinary excretion of many organic acids is altered. FBPase deficiency results in elevation of urinary glycerol, which is used in the diagnosis of this disease. It may result in life-threatening hypoglycemic episodes and coma (Tran, 2017). The treatment includes avoidance of dietary fructose, sorbitol, and prolonged fasting.

Summary

Dietary fructose is a natural component of fruits, fruit juices, honey, and some root vegetables. Fructose is the sweetest of all natural sugars and it extensively utilized in the food and beverage industry as a nutritive sweetener in the form of sucrose or HFCS. Worldwide surveys of fructose consumption have revealed either steady or declining trends in dietary intake of fructose. It is mainly metabolized in the liver, in a different manner than glucose. Although fructose in the form of sweetened beverages has been implicated in accumulation of lipids, in the development of obesity and other metabolic abnormalities, further studies are necessary to confirm this hypothesis. Fructose appears to play a role in the development of dental caries. Several genetic abnormalities of fructose metabolism have been described in humans, causing benign or serious disease.

References

Baena, M., Sangüesa, G., Dávalos, A., et al., 2016. Fructose, but not glucose, impairs insulin signaling in the three major insulin-sensitive tissues. Sci. Rep. 6, 26149.
Elliott, S.S., Keim, N.L., Stern, J.S., Teff, K., Havel, P.J., 2002. Fructose, weight gain, and the insulin resistance syndrome. Am. J. Clin. Nutr. 76 (5), 911–922.
Fidler Mis, N., Braegger, C., Bronsky, J., et al., 2017. Sugar in infants, children and adolescents: a position Paper of the European Society for Paediatric Gastroenterology, Hepatology and Nutrition Committee on Nutrition. J. Pediatr. Gastroenterol. Nutr. 65 (6), 681–696.
Malik, V.S., Hu, F.B., 2015. Fructose and cardiometabolic health: what the evidence from sugar-sweetened beverages tells us. J. Am. Coll. Cardiol. 66 (14), 1615–1624.
Malik, V.S., Hu, F.B., 2019. Sugar-sweetened beverages and cardiometabolic health: an update of the evidence. Nutrients 11 (8), 1840.
McGrane, M.M., 2012. Carbohydrate metabolism—synthesis and oxidation. In: Stipanuk, M., Caudill, M. (Eds.), Biochemical, Physiological, and Molecular Aspects of Human Nutrition, third ed. WB Saunders Company, Philadelphia, pp. 209–255.
Newens, K.J., Walton, J., 2016. A review of sugar consumption from nationally representative dietary surveys across the world. J. Hum. Nutr. Diet. 29 (2), 225–240.
Pancoast, H.M., Junk, W.R., 1980. Fructose. In: Pancoast, H.M., Junk, W.R. (Eds.), Handbook of Sugars, second ed. AVI Publishing Co., Westport, CT, pp. 387–388.
Rippe, J.M., Angelopoulos, T.J., 2013. Sucrose, high-fructose corn syrup, and fructose, their metabolism and potential health effects: what do we really know? Adv. Nutr. 4 (2), 236–245.
Scientific Advisory Committee on Nutrition, 2015. SACN Carbohydrates and Health Report. In: The Scientific Advisory Committee on Nutrition Recommendations on Carbohydrates, Including Sugars and Fibre. Public Health England.
Tappy, L., Lê, K.A., 2010. Metabolic effects of fructose and the worldwide increase in obesity. Physiol. Rev. 90 (1), 23–46.
Tran, C., 2017. Inborn errors of fructose metabolism. What can we learn from them? Nutrients 9 (4), 356.
Vos, M.B., Kimmons, J.E., Gillespie, C., Welsh, J., Blanck, H.M., 2008. Dietary fructose consumption among US children and adults: the Third National Health and Nutrition Examination Survey. Medscape J. Med. 10 (7), 160.
Wittekind, A., Walton, J., 2014. Worldwide trends in dietary sugars intake. Nutr. Res. Rev. 27 (2), 330–345.
World Health Organization, 2003. Diet, Nutrition and the Prevention of Chronic Diseases: Report of a Joint WHO/FAO Expert Consultation. WHO Technical Report Series No. 916. Geneva.
World Health Organization, 2015. Guideline: Sugars Intake for Adults and Children. Geneva.

Relevant websites

https://www.ers.usda.gov/topics/crops/sugar-sweeteners/data/, (Accessed on January 20, 2022).
https://www.acs.org/, (Accessed on February 9, 2022).

Fruits and vegetables

María Dolores Ruiz-López and Belén García-Villanova Ruiz, Department of Nutrition and Bromatology, Universidad de Granada, Granada, Spain

© 2023 Elsevier Ltd. All rights reserved.

Introduction 398
Definition and classification of fruits 398
Nutritional components of fruits 399
Energy 399
Water 399
Carbohydrates 399
Sugars 399
Starch 399
Fiber 399
Polyalcohols 400
Nitrogenous compounds 400
Lipids 400
Vitamins 400
Minerals 401
Definition and classification of vegetables 401
Nutritional components of vegetables 401
Energy 401
Water 401
Carbohydrates 401
Nitrogenous compounds 403
Lipids 403
Vitamins 403
Minerals 404
Other components in fruits and vegetables 404
Organic acids 404
Phenolic compounds 405
Aromatic compounds 405
Pigments 405
Bioactive compounds 406
Vitamin C 406
Vitamin E 406
Sulfur compounds 406
Glucosinolates (GSLs) 406
Organosulfur compounds (OSCs) 406
Carotenoids 407
Phytosterols 407
Phenolic compounds 407
Health effects of fruits and vegetables 408
Gastrointestinal health 408
Cardiovascular diseases (CVD) 408
Obesity 409
Glucose metabolism 410
Cancer 410
Vision 411
Conclusions 411
References 411

Key points

- To describe the chemical composition and nutritional compounds found in fruits and vegetables.
- To describe the bioactive compounds and their health effects.
- To establish the health benefits of a diet rich in fruits and vegetables.
- To present evidence on the consumption of fruits and vegetables and the risk of certain chronic diseases.

Introduction

Daily consumption of fruits and TO PGN: check the copyright year as in nimblevegetables is essential to achieve a balanced, nutritious, and adequate diet to reach the feeling of satiety without exceeding the energy value appropriate for a correct diet. From this point of view, fruits and vegetables are essential foods for achieving "adequate nutrition" because of their relatively low caloric density, providing a wide variety of nutrients. In recent years, there has been a series of advances in nutrition on the scientific evidence showing that some food components of the diet can offer beneficial physiological effects beyond the nutritional effects traditionally considered.

According to the World Health Organization (WHO), insufficient intake of fruits and vegetables is one of the ten major risk factors contributing to mortality. It is estimated that up to 1.7 million lives could be saved each year if the global consumption of fruits and vegetables was adequate. This intake helps ensure adequate intake of vitamins, minerals, dietary fiber, and bioactive compounds, and their increased intake can help displace foods rich in saturated fat, sugars, or salt.

Fruits and vegetables have a wide variety of bioactive compounds that, even in low concentrations, have shown effects on the reduction of chronic noncommunicable diseases. Plants synthesize these substances, called phytochemicals, as a defense system against aggression by external agents, and some of them are responsible for the color, taste, and odor of these foods.

The nutritional composition of fruits and vegetables varies significantly according to type, variety, and origin, and other factors related to handling, conservation, and storage conditions. These food groups are characterized by high water, fiber, and some minerals and vitamins, and low protein and fat content. Their high capacity of hydration, low energy density, providing feeling of satiety, and hedonic character (color and attractive taste) are remarkable. They also possess small amounts of other chemical compounds such as organic acids, phenolic compounds, aromatic compounds, and pigments.

Currently, modern conservation methods, genetic improvement, the great varietal dynamism existing within crops, and the wide range of presentations available on the market (fresh, dried, canned, frozen) are facilitating, without doubt, the consumption of these foods throughout the year.

Definition and classification of fruits

In Botany, fruits are defined as the edible parts of plants that develop from the flowers and contain seeds inside. Within this context, edible fruits of a fleshy nature that can be consumed without preparation are called fruits. Although tomato or cucumber, for example, meets the requirements of this definition, they are considered vegetables for cultural reasons (Ruiz-López and Garcia-Villanova Ruiz, 2017).

From a botanical point of view, the fruits can be classified into:

1. Fleshy fruits (derived from a single flower):
   - Drupe: Fleshy pericarp that surrounds a seed with a woody shell.
     - Apricot, plum, peach.
     - Avocado, mango, sour cherry, cherry.
     - Aggregates (strawberry, raspberry, blackberry).
     - Rowanberries: Part of the flower is the floral receptacle (peduncle).
     - Apple, pear, quince.
   - Berries: Fleshy fruits with tiny seeds freely arranged in the pulp.
     - Grape, blueberry, bananas, papaya, cherimoya.
     - Date, watermelon, melon, currant, kiwi.
     - Hesperidium or citrus fruits (orange, lemon, tangerine, lime).
2. Compound fleshy fruits (derived from an inflorescence):
   - Sorosis: Pineapple.
   - Syconium: Fig.

Nutritional components of fruits

Energy

The fruits are characterized by low in calories, as shown in Table 1, except avocado, which in many Mediterranean countries is consumed with the salad and not as dessert fruit itself, and banana, whose carbohydrate content (starch and simple sugars) provides a higher energy. The other fruits provide an energy of approximately 30 kcal/100 g of edible portion.

Water

The mean water content is between 81% and 93%. Banana and avocado have a lower content. In general, the mean water content of the most consumed fruits is shown in Table 1.

Carbohydrates

Sugars

The main sugars are sucrose, glucose, and fructose. The type of predominant sugar depends on the type of fruit. Thus, the drupes (plum, apricot, cherry, peach, etc.) have mainly sucrose, except for cherries. For reducing sugars (glucose and fructose), glucose is usually in a larger proportion. Seed fruits (apples, quinces, pears) also have glucose and fructose, but in this case, the proportion of fructose is higher and continues to increase, even after harvesting.

Starch

It is usually found only in unripe fruit, decreasing its concentration throughout ripening. Exceptions are cherimoya and banana with 1.5% and more than 3%, respectively.

Fiber

It is composed of cellulose, hemicellulose, pectin, and lignin. At harvest time, it reaches values of up to 3%. Hemicelluloses contribute to the firmness of the fruits and are hydrolyzed when maturing, producing mainly pentoses, mannoses, and uronic acids. Pectins are polymeric compounds derived from galacturonic acid that highly influence the texture and consistency of fruits. The texture variations experienced by fruits during the development in the tree are closely related to the modifications that the pectins experience. Insoluble pectic substances (protopectins) are degraded by enzymatic action to soluble forms and these, in turn, to simpler compounds (galacturonic acid); the latter usually occurs in overripened fruits. Pectins are included within the so-called soluble fiber and are therefore associated with the effect of satiety, slowing glucose absorption, decrease in cholesterol absorption, and regulation of gastrointestinal motility. The total fiber content ranges from 0.3% of watermelon to 2.5% of banana.

Table 1 Energy and major components of some fruits of habitual consumption (expressed in 100 g edible portion).

Fruit	*Energy (kcal)*	*Water (g)*	*Carbohydrate*[a]	*Total sugar (g)*	*Fiber (g)*
Apple	52	85.6	11.4	10.4	2.4
Apricot	48	86.3	9.1	9.2	2.0
Banana	89	74.9	19.8	12.2	2.6
Blueberries	57	84.2	12.1	9.9	2.4
Cantaloupe melon	34	90.2	7.3	7.8	0.8
Cherries	63	82.2	14	12.8	2.1
Grapes	69	80.5	17.2	15.5	0.9
Kiwi	58	84.0	11	8.9	3.0
Mango	60	83.5	13.4	13.6	1.6
Orange	47	86.7	9.3	9.3	2.4
Peach	42	88.3	8.6	8.3	1.5
Pear	57	83.9	12.1	9.7	3.1
Pineapple	50	86.0	11.7	9.8	1.4
Plum	46	87.2	10	9.9	1.4
Strawberries	32	91	5.7	4.9	2.0
Tangerine	53	85.2	11.5	10.6	1.8
Watermelon	30	91.4	7.1	6.2	0.4

[a]Carbohydrates do not include dietary fiber.

Source: U.S. Department of Agriculture, Agricultural Research Service (2019).

Polyalcohols

Significant amounts of sorbitol, which may have a concentration-dependent laxative effect, are found in drupes—such as apples, plums, or pears. However, this polyalcohol is not found in bananas, strawberries, or pineapple, which can even be used to identify its presence in fruit preparations.

Nitrogenous compounds

Fruits are characterized by the low content of nitrogenous compounds with values ranging from 0.1 to 1.5% and consist of free amino acids, amines, and proteins. Free amino acids are widely distributed and constitute 50% of soluble nitrogen compounds.

Lipids

The lipid fraction of fruits comprises triglycerides, glycolipids, phospholipids, carotenoids, triterpenoids, and waxes and are found in very low concentration, with values between 0.1% and 0.5% of fresh weight. There is an exception in the avocado pulp that can reach up to 30%. The seeds of the drupes are rich in lipids (50%–37%), while those pomes have a content of 20%. In pomes, waxes covering their epidermis have also been studied because of their interest in controlling fruit transpiration and the protection they provide against the atmospheric agents and the attack of insects and parasites.

Vitamins

Fruits supply a significant proportion of vitamins C and A (the latter, in the form of provitamin). They contain other vitamins, such as vitamin E and B-vitamins, but their contribution to the diet is small, with some exceptions (**Table 2**). Vitamin C is formed in plants from hexoses, and its distribution in them is uneven. For example, a gradient of the vitamin C content was found in peach from the skin, the richest part, to the fleshy portion near the stone, the poorest part. In apples, during all stages of development, the vitamin C content of the skin is three to five times higher than that of the pulp. The richest apples in this nutrient are the colored varieties and those exposed to the sun in the tree. Most of the ascorbic acid of oranges is in the peel and only a quarter in the orange juice. Strawberries ($\approx$60 mg/100 g), which are consumed seasonally, or kiwi ($\approx$94 mg/100 g), consumed throughout the year, are also notable for their contents in vitamin C. Fruits that provide vitamin C include acerola (*Malpighia emarginata*), also called cherry of the Antilles: It is an acid fruit of red color (there are other varieties of yellow and purple colors), which is characterized by containing a quantity of vitamin C of 690–4800 mg/100 g. It can be eaten fresh, although its main consumption is as juice. It is also used in the food industry for vitamin C enrichment of fruit juices. The folic acid content of fruits is low, although in oranges it is 40 μg/100 g; it is 45–62 μg/100 g in strawberries and raspberries; and it is 30 μg/100 g in kiwi (**Table 2**). The vitamin A intake from fruits comes from carotenoids (α-carotene, β-carotene, γ-carotene, and β-cryptoxanthin), which are the ones that can become retinol inside the organism.

Table 2 Vitamin content of some commonly consumed fruits (100 g of edible portion).

Fruit	*Thiamin mg*	*Riboflavin (mg)*	*Niacin (mg Eq)*	*Folate (μg)*	*Vit. C (mg)*	*Vit. A (μg RAE)*	*Vit E (mg)*
Apple	0.02	0.03	0.1	3	4.6	3	0
Apricot	0.03	0.04	0.6	9	10	96	0.9
Banana	0.1	0.1	0.7	20	9	3	0.1
Blueberries	0.04	0.04	0.41	6	9.7	3	0.57
Cantaloupe melon	0.05	0.03	0.7	14	10.9	232	0.05
Cherries	0.03	0.03	0.1	4	7	3	0.1
Grapes	0.1	0.1	0.2	2	3	3	0.2
Kiwi	0.02	0.02	0.3	25	93	4	1.5
Mango	0.03	0.04	0.67	43	36.4	54	0.9
Orange	0.09	0.04	0.28	30	53.2	11	0.18
Peach	0.02	0.03	0.8	4	6.6	16	0.7
Pear	0.01	0.03	0.2	7	4	1	0.1
Pineapple	0.1	0.1	0.5	18	48	3	0.1
Plum	0.03	0.03	0.42	5	9.5	17	0.3
Strawberries	0.02	0.02	0.4	24	59	1	0.3
Tangerine	0.06	0.04	0.4	16	27	34	0.2
Watermelon	0.03	0.02	0.2	3	8	28	0.1

Source: U.S. Department of Agriculture, Agricultural Research Service (2019).

Table 3 Mineral content of some commonly consumed fruits (mg/100 g edible portion).

Fruit	*Sodium*	*Potassium*	*Calcium*	*Magnesium*	*Phosphorus*	*Iron*	*Zinc*
Apple	2	100	6	5	10	0.1	0.1
Apricot	1	259	13	10	23	0.4	0.2
Banana	1	358	5	27	28	0.3	0.1
Blueberries	1	77	6	6	12	0.3	0.2
Cantaloupe melon	16	267	9	12	15	0.2	0.2
Cherries	0	222	13	11	21	0.4	0.1
Grapes	2	191	10	7	20	0.4	0.1
Kiwi	3	312	34	17	34	0.3	0.1
Mango	1	168	11	10	14	0.2	0.1
Orange	0	181	40	10	14	0.1	0.1
Peach	0	190	6	9	20	0.4	0.2
Pear	1	116	9	7	12	0.2	0.1
Pineapple	1	109	13	12	8	0.3	0.1
Plum	0	157	6	7	16	0.2	0.1
Strawberries	2	150	30	13	26	0.7	0.1
Tangerine	2	166	37	12	20	0.1	0.1
Watermelon	1	112	7	10	11	0.2	0.1

Source: U.S. Department of Agriculture, Agricultural Research Service (2019).

Minerals

Potassium is the most representative mineral element of the fruit; in some fruits, potassium can account for 50% of the ashes. Avocado, kiwi, melon, black grapes, and banana stand out as fruits with potassium content of up to 300 mg/100 g of edible portion. The low sodium content should also be highlighted, making these foods very suitable for people with hypertension. **Table 3** shows the composition in the most representative minerals.

Definition and classification of vegetables

The definition of vegetables considers the edible parts of a plant and is usually grouped according to the consumed plant portion. This definition excludes the fruits defined above (Garcia-Villanova Ruiz, 2017).

Thus, vegetables can be classified into:

- Tubers: Sweet potato, potato
- Bulbs: Garlic, onion, leek
- Roots: Carrot, beetroot, turnip
- Fruits and inflorescence: Tomato, pepper, zucchini, pumpkin, eggplant, cucumber, artichoke
- Leaves and stems: Spinach, lettuce, chard, endive and asparagus
- Flowers: Cabbage, savoy cabbage, red cabbage, broccoli, Brussels sprouts, cauliflower.

Nutritional components of vegetables

Energy

Vegetables are low energy density foods, with values less than 50 kcal/100 g, except starchy vegetables (potato, sweet potato, and cassava) and garlic (**Table 4**).

Water

Water is the major component of vegetables. Water content in vegetables ranges from 80% to 95%. A high percentage of vegetables show values above 90%. Starchy vegetables (sweet potato and cassava) and garlic (59%) have lower water content than other vegetables (**Table 4**).

Carbohydrates

They are the main macronutrient of vegetables. Carbohydrate content can be very variable, and in most vegetables the values are less than 7%, except tubers, potato, sweet potato, cassava (12–38%), and some bulbs such as garlic (33%). They are found as free sugars,

Table 4 Energy and major components of some vegetables of habitual consumption (expressed in 100 g edible portion).

Vegetables	*Energy (kcal)*	*Water (g)*	*Carbohydrate*[a] (g)	*Total sugar (g)*	*Fiber (g)*
Artichoke	53	84.08	11.95	0.99	5.7
Asparagus	20	93.22	3.88	1.88	2.1
Broccoli	34	89.3	6.64	1.7	2.6
Brussels sprouts	43	86.0	8.95	2.2	3.8
Cabbage	25	92.18	5.8	3.2	2.5
Cabbage (red)	31	90.39	7.37	3.83	2.1
Carrot	41	88.29	9.58	4.74	2.8
Cassava	160	59.68	38.06	1.7	1.8
Cauliflower	25	92.07	4.97	1.91	2
Chard	19	92.66	3.74	1.1	1.6
Eggplant	25	92.3	5.88	3.53	3
Endive	17	93.79	3.35	0.25	3.1
Garlic	149	58.58	33.06	1	2.1
Green beans	31	90.32	6.97	3.26	2.7
Green peas	81	78.86	14.45	5.67	5.7
Kale	35	89.6	4.42	0.8	4.1
Leek	61	83.0	14.15	3.9	1.8
Lettuce	14	95.31	2.92	1.38	1.2
Onion	40	89.11	9.34	4.24	1.7
Pepper green	20	93.89	4.64	2.4	1.7
Pepper red	26	92.21	6.03	4.2	2.1
Potato[b]	58	83.29	12.44	0.82	2.5
Pumpkin	26	91.6	6.5	2.76	0.5
Spinach	23	91.4	3.63	0.42	2.2
Sweet potato	86	77.28	20.12	4.18	3
Tomato	18	94.52	3.89	2.63	1.2
Zucchini[b]	17	94.79	3.11	2.5	1

[a]Carbohydrates do not include dietary fiber.

[b]Potato and zucchini with skin.

Source: U.S. Department of Agriculture, Agricultural Research Service (2019).

oligosaccharides, and polysaccharides. Starch is present in most vegetables, with values of less than 1% and stands out in starchy vegetables. It is the main polysaccharide of the tubers (potato, sweet potato, tapioca). In some countries, tubers form the basis of energy food consumed by the population. Starch is also found in garlic, pumpkin, chard, and green legumes. The most common sugars are glucose, fructose, and sucrose, and the content ranges from 0.2% to 5%. Some, such as sweet potato and garlic, have larger amounts (**Table 4**).

Mannitol, a polyalcohol from the hydrogenation of mannose, has been found in varieties of the Cucurbitaceae and Cruciferae families. Oligosaccharides are uncommon for rare group in foods. They are found in some types of vegetables. The galactosyl-sucrose family includes raffinose, stachyose, and verbascose, consisting of a sucrose molecule together with one, two, or three galactose molecules, respectively. They are found mainly in fresh legumes, peas, and beans. Fructans, consisting of fructooligosaccharides (FOS) (Fru_n) and inulin polymer (Gal-Fru_n), are found in artichoke, chicory, asparagus, onion, beet, garlic, and tomato. Raffinose, stachyose, verbascose, and fructooligosaccharides cannot be hydrolyzed by pancreatic enzymes or by those of the brush border of enterocytes, so they pass into the large intestine where they are fermented by the action of the intestinal microbiota and produce short-chain acids and gases (CO_2).

Dietary fiber comprises polysaccharides, cellulose, hemicellulose, pectin, inulin, and the polymer called lignin. These compounds are not hydrolyzed by endogenous digestive enzymes. Cellulose, hemicellulose, and lignin are the main components of the cell walls, and cellulose is the most abundant. In vegetables, their values range from 0.2 to 4 g/100 g of product. Green legumes, beans (4.1 g/100 g), peas (5.7 g/100 g), artichoke (5.7 g/100 g), and cabbage (kale and Brussels sprouts) show the highest values; the rest of the vegetables have values ranging from (0.5 to 2.7 g/100 g) (**Table 4**). Hemicelluloses are associated with cellulose as constituents of cell walls. They can be soluble and insoluble, consisting of linear and branched polysaccharides containing hexoses and pentoses. They are characterized by their high capacity to bond with the water and serve as agents that increase the volume. The presence of acidic components in some gives them the ability to bind to cations. In the colon, hemicelluloses ferment to a higher degree than celluloses. Pectins are associated with cellulose and hemicellulose. In the immature vegetables, they are found in the form of insoluble protopectins, which ensures rigidity to the tissues, but during the maturation, they are degraded to pectins, sugars, and acids, with the consequent softening of the tissues. Some raw starch granules, such as potato granules, resist the action of β-amylases enzymes, probably due to the crystalline nature of starch (crystalline regions of starch are less susceptible to attack by acids and enzymes than amorphous zones). The presence of phosphate groups in potato starch also makes digestion

difficult. Cooking results in starch gelatinization and facilitates digestion, which could explain the difficulty of digesting the raw potato. Lignins are insoluble macromolecules; they are not digested or absorbed, nor are they attacked by the colon microbiota. Lignin is the most complex natural polymer due to its structure and heterogeneity. The lignification of vegetables increases the lignin content in the cell wall due to maturation, which gives these plants resistance to bacterial degradation (Garcia-Villanova Ruiz, 2017).

Nitrogenous compounds

Vegetables are deficient in nitrogen compounds (0.6–5%), consisting of proteins, amino acids, peptides, amines, and other compounds. Vegetable proteins are of low biological value, although it is higher in potato. Most of the protein fraction is constituted of enzymes producing positive or negative effects in handling and preparing the harvested vegetables, modifying the organoleptic characteristics, such as aroma, color, taste, and texture. Oxidoreductases (such as lipoxygenases, polyphenol oxidases, and peroxidases) and hydrolases (such as glycosidases, esterases, proteases, and transferases) are responsible for many of the organoleptic changes occurring after harvesting and processing. They have protein and non-protein amino acids, including homoserine, homomethionine, and aminoadipic acid. Choline and betaine are found in cabbage; histamine and its derivatives in spinach; and serotonin, tryptamine, tyramine, and melatonin in tomato and eggplant.

Lipids

The lipid content of vegetables is very low (0.1–1% of fresh weight). It consists of triacylglycerols, glucolipids, and phospholipids. They have saturated, monounsaturated, and polyunsaturated fatty acids.

Vitamins

Vegetables are characterized by the supply of fat-soluble and water-soluble vitamins. The type, variety, and conditions before and after harvest, storage, and technological or culinary treatments influence the vitamin content. Some are an important source of vitamin A (in the form of its precursors, carotenoids), vitamin C, and folic acid (**Table 5**). They also contain other fat-soluble

Table 5 Vitamin content of some commonly consumed vegetables (100 g of edible portion).

Vegetables	*Thiamin (mg)*	*Riboflavin (mg)*	*Niacin (mg Eq)*	*Folate (μg)*	*Vit. C (mg)*	*Vit. A (μg RAE)*	*Vit E (mg)*
Artichoke	0.05	0.089	1.11	89	7.4	1	0.19
Asparagus	0.143	0.141	0.978	52	5.6	38	1.13
Broccoli	0.071	0.117	0.639	63	89.2	31	0.78
Brussels sprouts	0.139	0.09	0.745	61	85	38	0.88
Cabbage	0.061	0.04	0.234	43	36.6	5	0.15
Carrot	0.066	0.058	0.983	19	5.9	835	0.66
Cassava	0.087	0.048	0.854	27	20.6	1	0.19
Cauliflower	0.05	0.06	0.507	57	48.2	0	0.08
Chard	0.04	0.09	0.4	14	30	306	1.89
Eggplant	0.039	0.037	0.649	22	2.2	1	0.3
Endive	0.08	0.075	0.4	142	6.5	108	0.44
Garlic	0.2	0.11	0.7	3	31.2	0	0.08
Green beans	0.082	0.104	0.734	33	12.2	35	0.41
Green peas	0.266	0.132	2.09	65	40	38	0.13
Kale	0.113	0.347	1.18	62	93.4	241	0.66
Leek	0.06	0.03	0.4	64	12	83	0.92
Lettuce	0.056	0.053	0.249	34	6	198	0.2
Onion	0.046	0.027	0.116	19	7.4	0	0.02
Pepper green	0.057	0.028	0.48	10	80.4	18	0.37
Pepper red	0.054	0.085	0.979	46	127.7	157	1.58
Potato [a]	0.021	0.038	1.033	17	11.4	0	–
Pumpkin	0.05	0.11	0.6	16	9	426	1.06
Red cabbage	0.064	0.069	0.418	18	57	56	0.11
Spinach	0.078	0.189	0.724	194	28.1	469	2.03
Sweet potato	0.078	0.061	0.557	11	2.4	709	0.26
Tomato	0.037	0.019	0.594	15	13.7	42	0.54
Zucchini[a]	0.045	0.094	0.451	24	17.9	10	0.12

[a]Potato and zucchini with skin.

Source: U.S. Department of Agriculture, Agricultural Research Service (2019).

vitamins, such as E and K, and water-soluble vitamins, such as those in group B. The vitamin content can be very variable, according to the large number of foods constituting this group. Vitamin C content can range from 2.4 to 128 mg/100 g. The vegetables with the highest content are red and green pepper and the cabbage family (kale, broccoli, brussels sprouts, red cabbage, cauliflower, cabbage). The vegetables with the highest vitamin A content are carrot, sweet potato, pumpkin, spinach, chard, lettuce, kale, and red pepper and folic acid stand out in spinach and endives.

Minerals

Minerals account for 1–2% of fresh food. Vegetables are characterized by low sodium and high potassium content. The latter is the most representative mineral element, ranging from 150 to 560 mg/100 g (Table 6). They also contribute to the daily intake of calcium, phosphorus, magnesium, iron, and zinc. The bioavailability of iron in this food group is low, but vitamin C can help increase its absorption. The most representative anions are phosphate, chloride, and carbonate.

Other components in fruits and vegetables

Organic acids

The fruits are characterized by their richness in organic acids, among which are the hydroxy acids, such as citric, malic, succinic, tartaric, and tannic acids. These compounds are responsible for the acidity in the immature fruit and decrease during ripening by transforming partly into simple sugars. The main acids in the stone fruits are malic, citric, and quinic acids. Malic acid is the one that predominates, and in some fruits, such as plum and cherry, it constitutes 90% of the acids. In grapes, tartaric acid is the majority, and in tropical fruits and berries, it is citric acid.

The content of these organic acids is lower in vegetables than in fruits. Most are in the form of salts, which makes the pH value relatively high (5.5–6.5), except in tomatoes. The free acid content is 0.2–0.4 g/100 g of the fresh product, and citric acid and malic acid contents are outstanding. Oxalic acid is found in some vegetables in high quantities (e.g., chard, spinach, and lettuce). This can lead to corrosion of tinplate containers and in the body and may contribute to the formation of urinary stones. There is controversy among researchers about the influence of oxalate in the formation of urinary stones. However, certain groups of researchers have

Table 6 Mineral content of some commonly consumed vegetables (mg/100 g edible portion).

Vegetables	*Sodium*	*Potassium*	*Calcium*	*Magnesium*	*Phosphorus*	*Iron*	*Zinc*
Artichoke	60	286	21	42	73	0.61	0.4
Asparagus	2	202	24	14	52	2.14	0.54
Broccoli	33	316	47	21	66	0.73	0.41
Brussels sprouts	25	389	42	23	69	1.4	0.42
Cabbage green	18	170	40	12	26	0.47	0.18
Carrot	69	320	33	12	35	0.3	0.24
Cassava	14	271	16	21	27	0.27	0.34
Cauliflower	30	299	22	15	44	0.42	0.27
Chard	213	379	51	81	46	1.8	0.36
Eggplant	2	229	9	14	24	0.23	0.16
Endive	22	314	52	15	28	0.83	0.79
Garlic	17	401	181	25	153	1.7	1.16
Green beans	6	211	37	25	38	1.03	0.24
Green peas	5	244	25	33	108	1.47	1.24
Kale	53	348	254	33	55	1.6	0.39
Leek	20	180	59	28	35	2.1	0.12
Lettuce	19	168	27	10	24	0.64	0.16
Onion	4	146	23	10	29	0.21	0.17
Pepper green	3	175	10	10	20	0.34	0.13
Pepper red	4	211	7	12	26	0.43	0.25
Potato[a]	10	413	30	23	38	3.24	0.35
Pumpkin	1	340	21	12	44	0.8	0.32
Red cabbage	27	243	45	16	30	0.8	0.22
Spinach	79	558	99	79	49	2.71	0.53
Sweet potato	55	337	30	25	47	0.61	0.3
Tomato	5	237	10	11	24	0.27	0.17
Zucchini[a]	8	261	16	18	38	0.37	0.32

[a]Potato and zucchini with skin.

Source: U.S. Department of Agriculture, Agricultural Research Service (2019).

agreed that spinach, rhubarb, beet, nuts, cocoa, wheat bran, strawberries, peanuts, almonds, and tea cause a significant increase in urinary oxalate.

Currently, a new consumer habit is identified intake vegetables raw which are typically cooked. This may bear several risks: an increase in kidney and ureter stones by increased intake of oxalic acid; an increasing intake of oxalic acid might lead to demineralization of food and a deficiency of calcium and iron intake; an increased intake of nitrate by frequently consuming an excess of leafy greens and microbial contamination by preparing a large stock of green smoothies with insufficient cooling of this stock.

Phenolic compounds

Fruits are very rich foods in phenolic compounds, with flavonoids being the largest group. Phenolic acids are widely distributed in fruits; they contribute to the color and flavor of many fruits. Phenolic acids are responsible for the astringency of many fruits, although they disappear with maturation in most of them. Polyphenol oxidases oxidize diphenols to quinones. The latter are the starting point of the browning reactions of the fruits, generating brown colorations when the fruit is peeled or cut. Chlorogenic acid is the key substrate for the enzymatic browning of apples and pears. Flavonoids are the largest group of polyphenolic compounds and are usually linked to sugars.

In vegetables, phenolic compounds are widely distributed, although their content compared to fruits is lower. Phenolic acids are made up of hydroxycinnamic and hydroxybenzoic acids and are generally found as esters and to a lesser extent as glycosides. Hydroxycinnamic acids are characteristic of some vegetables, such as eggplant, artichoke, potato, and chicory, with values ranging from 100 to 650 mg/100 g of fresh vegetables.

Flavonoids are the most representative group of vegetable phenolic compounds, and flavonols and flavones predominate. In some red vegetables (e.g., red cabbage and red leaf lettuce), anthocyanidins are found, and flavanones in artichokes and tomatoes.

Aromatic compounds

The aroma is one of the most attractive organoleptic characteristics of fruits, which is due to the presence of volatile compounds. More than 500 aromatic compounds of small molecular weight, such as esters, ketones, and alcohols, are available. Only those whose concentration in the food is higher than the olfactory threshold are considered aromatic compounds. Those compounds providing the characteristic aroma of the food are called "impact compounds." In some fruits, such as lemon, pear or banana, the aroma is due to a single compound; whereas in other fruits, the aroma is due to a group of compounds with a well-characterized main compound, or to a large group of compounds difficult to characterize, such as is the case of strawberries aroma.

In some vegetables, the aroma is produced by sulfur compounds and provides less attractive odors, such as in the *Cruciferae* family (e.g., cabbage, cauliflower, and Brussels sprouts). In fresh products, the compounds are bound to sugars (glucosinolates) and are odorless, but when the tissue structure is broken by cutting, bruising, or chewing, the glucosinolates are released, and myrosinase transforms them into their hydrolysis products. The smell of those of the genus *Allium* (e.g., onion, garlic, spring onion, leek) are due to organosulfur compounds (allyl sulfur compounds): When the tissue structure is broken by cutting, bruising, or chewing, the compounds are released, and the alliinase enzyme transforms them into volatile compounds, some with penetrating odor and others with tear effect.

Pigments

Color is one of the most attractive organoleptic characters of fruits and vegetables. It is due to the presence of pigments such as chlorophyll, carotenoids, flavonoids, and betalains.

Chlorophyll is the pigment responsible for the green color of some fruits and a variety of vegetables. It is the only pigment present in immature fruits, responsible for its green color. It has a lipophilic character; that is, as the fruit matures, a change of color occurs due to the disappearance of chlorophyll and the formation of carotenoids and flavonoids that are specific to each of them. This occurs in peaches, apricots, cherries, and strawberries, but not in some varieties of apples, pears, or plums, in which chlorophyll provides a characteristic green color that masks the presence of other pigments. In some vegetables, such as red pepper and tomato, chlorophyll disappears during maturation to allow carotenoids to show. Chlorophylls change during thermal treatments, resulting in different green tones after blanching, cooking, and sterilization.

Carotenoids are responsible for yellow, orange, purple, and red tones in fruits and vegetables. They are found in a much higher proportion in vegetables, especially β-carotene, with values ranging from 270 μg/100 g in celery to 8285 μg/100 g in carrot. The liposoluble pigments are of nutritional importance, as they are some precursors of vitamin A.

Other carotenoids, such as lutein, zeaxanthin, and lycopene, are potent antioxidants. Lycopene is not found in most fruits, except watermelon and tomato in vegetables. Lutein and zeaxanthin stand out in vegetables with values above 10,000 μg/100 g in chards, watercress, and spinach. Fruits also have these compounds, being more abundant on the skin than on the fleshy portion. Peaches, mandarins, watermelon, and apricots are particularly rich in these compounds.

Anthocyanidins are water-soluble and dissolved in cell juice. The colors they provide range from red to blue or purple. They are the pigment of many fruits, such as strawberries, raspberries, currants, blueberries, pomegranates. These anthocyanins are found in the skin, such as plums, apples, pears, or grapes; however, they may also be in the fleshy portion, as seen in some varieties of cherries or plums. They are very unstable due to changes in pH and the action of oxygen and SO_2. The red color of red cabbage and red leaf

lettuce are also due to these compounds, while betalains are pigments responsible for the red color of beet and some types of mushrooms.

Bioactive compounds

Vitamin C

Vitamin C is a hydrosoluble antioxidant with high reducing power. It has the ability to scavenge a wide variety of reactive oxygen and nitrogen species in aqueous media. It is found mainly in foods of plant origin, such as ascorbic acid (reduced form) and dehydroascorbic acid (oxidized form), which is rapidly regenerated *in vivo*. Fruits and vegetables are a good source of vitamin C.

In most cell functions, ascorbate neutralizes reactive oxygen species (ROS) generated by cellular metabolism. Because of the role of ascorbate in protecting cells from oxidative stress and the involvement of ROS in neurodegenerative disorders (Alzheimer's and Parkinson's disease) or inflammatory responses (atherosclerosis), it is suggested that vitamin C could prevent heart, chronic inflammatory, and neurodegenerative diseases (Shashirekha et al., 2015).

Vitamin E

Vitamin E (tocopherols and tocotrienols) is present in all cell membranes and plasma lipoproteins. α-tocopherol is a potent lipophilic antioxidant, a suppressor of oxidative damage in biological membranes, lipoproteins, and tissues, by removing free radicals such as singlet oxygen and superoxide and hydroxyl radicals. Vitamin C regenerates oxidized vitamin E. Vitamin E protects DNA, low-density lipoproteins (LDL), and polyunsaturated fatty acids from oxidative damage (Shashirekha et al., 2015). It also plays a role in hemoglobin biosynthesis, modulation of the immune response, and stabilization of the membrane structure. Vitamin E is present in most fruits and vegetables with values between 0.1 and 2 mg/100 g.

Sulfur compounds

Glucosinolates (GSLs)

The GSLs are a large group of secondary metabolites of plants belonging to the genus Brassica (families Brassicaceae and Cruciferae); this genus includes vegetables such as broccoli, cauliflower, kale, brussels sprout, cress, radish, cabbage, and mustard. These vegetables represent a good source of phytochemicals, including sulfur glycosides (GSLs), phenolic compounds, and carotenoids. However, their anticarcinogenic and antioxidant potential has been attributed mainly to their high GSLs content (Mandrich and Caputo, 2020). The breakdown products of GSLs play a vital role in plant defense mechanisms against pathogens and insects.

GSLs are thioglucosides with one cyano group and one sulfate group, composed of more than 200 different types, and 30 of them are identified in Brassica crops (Quirante-Moya et al., 2020). When the tissues of these plants are processed by cutting, crushing, or chewing and other destruction of tissues, the GSLs are exposed to enzymes named myrosinases, which hydrolyze them to isothiocyanates. A likely source of isothiocyanates is the microbial degradation of GSL by the intestinal microflora. However, hydrolysis by microflora has been reported to be not very efficient, and in humans, it is very diverse and variable. These enzymes are inactivated by thermic treatment (cooking, steam cooking, microwave), and in this case, the GSL could be partially absorbed in their intact form.

The most studied compound is the sulforaphane (SFN), an isothiocyanate occurring in stored form as glucoraphanin (GRA) located in the cruciferous such as cabbage, cauliflower, and kale, and at high levels in broccoli especially in broccoli sprouts which is recognized as the best source of SFN, and GRA corresponds to 90% of the GSL content in some of its species. Sulforaphane affects oxidative stress, and antioxidant capacity, and neuroinflammation. The SFN is considered a very promising compound because it has properties that prevent, delay or reverse the development of preneoplastic lesions and improve survival rates, acting on cancer cells as a therapeutic agent. The sulforaphane may also decrease the risk of cardiovascular diseases (CVD) and help to prevent osteoporosis.

Organosulfur compounds (OSCs)

Vegetables of the Alliaceae family are considered a rich source of a large variety of beneficial bioactive compounds, such as oligosaccharides, arginine, flavonols, and OSCs. Vegetables such as onions, garlic, shallots, leeks, scallions, and chives belong to the genus *Allium*. Allium plants are consumed worldwide as vegetables, seasonings, and spices.

The sulfur compounds are mainly obtained when these vegetables have been damaged by processed by cutting, crushing or chewing, facilitating a reaction between the enzyme alliinase and the substrate alliin (diallyl thiosulfinate) is converted into intermediate compounds (sulfenic acid, pyruvic acid, and ammonia). Two molecules of sulfenic acid condense and form allicin, intermediate compound of diallyl sulfide, diallyl disulfide, or diallyl trisulfide.

The alliin is the organosulfur compound extracted mainly from garlic (*Allium sativum* L.). Experimental studies have revealed beneficial health effects including, antioxidant, cardioprotective, antihepatotoxic, immunomodulatory, and antineoplastic effects (Rizwani and Shareef, 2011).

These intermediate products in alliaceous vegetables are considered health-promoting bioactive components. Several types of Allium products, such as pastes, oils, salts, powders, and flakes, have been used in industries and homes as food seasonings and

flavorings. These products can be prepared by different methods, including boiling, frying, baking, and drying. Commercial garlic products usually are standardized on the content of sulfur compounds.

The temperature, processing type, pH, time, and food matrix can affect the alliinase activity. Heating results in the inactivation of the alliinase enzyme, thus decreasing the amount of allicin metabolites in products. A large amount of allicin is lost during the processing of alliaceous vegetables. This reduction in OSCs is linked to decreased flavor and reduced healthy effects of Allium products. The OSCs in Allium plants are thermally unstable and tend to be lost during pasteurization, sterilization, cooking, and drying (Barba et al., 2017).

Garlic organosulfur compounds decrease total cholesterol, LDL-C, and oxidized LDL due to a decrease in cholesterol biosynthesis by inhibition of hydroxymethyl-glutaryl-coenzyme A reductase (HMG-CoA reductase); they also show a moderate hypotensive effect due to their vasodilation property and are attributed inhibitory effect on platelet aggregation and decrease of fibrinolytic activity. Regular consumption of garlic significantly increases the antioxidant activity of cells, inhibiting the formation of free radicals, reinforcing the mechanism of endogenous radical uptake, and increasing antioxidant enzymes. Garlic is one of the most consumed natural supplements because it promotes a lowering of cholesterol and blood pressure, delays the progression of atherosclerosis, prevents heart disease, improves circulation, and prevents cancer.

Carotenoids

Carotenoids have shown beneficial health effects that include improving the immune system and reducing the risk of chronic degenerative diseases such as macular degeneration, type 2 diabetes, obesity, certain types of cancer, and CVD.

Recent studies have shown a positive correlation between consuming a diet rich in fruits and vegetables with high carotenoid content and preventing mortality and morbidity related to CVD. High concentrations of lutein and β-cryptoxanthin have been associated with a decreased risk of myocardial infarction, and lycopene, the carotenoid with larger antioxidant capacity and higher plasma concentration, is associated with a decrease in inflammatory processes (Rodriguez-Concepcion et al., 2018).

Epidemiological studies relate the high dietary intake of fruits and vegetables rich in carotenoids to health benefits. However, supplementation of doses higher than 20 mg/day of purified or synthetic β-carotene may cause adverse effects, including lung cancer risk in smokers and workers exposed to asbestos and mortality from CVD.

The xanthophylls, lutein and zeaxanthin are among the carotenoids with effect on macular degeneration. These xanthophylls are found in green leafy vegetables, kale, spinach, chard, watercress, eye tissue, serum, and plasma. Their function is to protect the macula and the lens from the oxidizing action of blue light and, in addition, protect the eye from photochemical reactions due to the antioxidant capacity that both possess. Observational studies indicate that a diet high in lutein and zeaxanthin could reduce both the risk of cataracts and age-related macular degeneration.

Lycopene is the carotenoid with greater antioxidant capacity and higher plasma concentration, and it tends to accumulate in the tissues because it is lipophilic. It is found mainly in some fruits and especially in tomatoes. Epidemiological studies suggest that diets rich in fruits and vegetables such as tomato are associated with a low risk of stomach and prostate cancer. The antioxidant effect of lycopene could be the main mechanism in preventing lipid peroxidation, DNA damage and protein oxidation, processes involved in carcinogenesis, aging, and CVD.

The antioxidant and *anti*-inflammatory activity of carotenoids or their derivatives may also be relevant for obesity because obesity leads to oxidative stress and inflammation in adipose tissue, which play an essential role in the development of insulin resistance. Likewise, the mechanisms of action of lutein and zeaxanthin in relation to cognitive function may be related to their antioxidant and *anti*-inflammatory abilities, among others.

The bioavailability of carotenes depends on various factors, such as the type and amount, the food matrix, and culinary preparation. The heat treatment (cooking) improves their absorption. Dietary fat also seems to stimulate absorption because mucosal cells absorb β-carotene from lipid micelles, and the amount and type of fat condition the formation and type of micelles; unsaturated fats are more efficient than saturated fats. Some studies show that the bioavailability of β-carotene in dark green leafy vegetables is lower than that of fruits (Barba et al., 2017).

Culinary treatments have very few carotenoid losses. Treatment losses range from 5 to 10%. Vegetables have higher values of carotenoids than fruits. The high content of α-carotene in pumpkin and carrot and the high levels of zeaxanthin and lutein in chard, watercress, and spinach are worth noting.

Phytosterols

Phytosterols are found in some fruits and vegetables and have cholesterol-lowering properties reducing the risk of CVD. Phytosterols can compete with cholesterol because of their structural similarity, limiting cholesterol absorption from fat matrices in the intestinal tract.

Phenolic compounds

Phenolic compounds are compounds from the secondary metabolism of plants and contribute to the color, flavor, and odor of plant foods. These compounds play an important role in color stability, aroma profile, and antioxidant activity.

Recently, dietary intake of phenolic compounds has been associated with a lower incidence of chronic diseases such as CVD, diabetes, cancer, and neurological diseases. Vegetables and especially fruits are sources rich in phenolic compounds.

Oxidative stress plays an essential role in carcinogenesis. The development of tumors may be due to oxidative damage, and phenolic compounds are powerful antioxidants. Several *in vitro* and *in vivo* studies have shown that flavonoids can inhibit different

stages of carcinogenesis due to their antioxidant activity and other mechanisms. Polyphenols can affect molecular events in the initiation, promotion, and progression stages of carcinogenesis, and isoflavones and lignans can affect activities related to hormone-dependent tumors. Flavonoids modulate enzymes and receptors involved in the signal transduction pathways of cell proliferation, differentiation, apoptosis, inflammation, angiogenesis, metastases, and reversal of resistance to many drugs. Apoptosis or programmed cell death is necessary to maintain a balance between cell proliferation and loss. Deregulation of this balance may lead to malignancy, while induction of apoptosis suppresses the development of cancer.

Several epidemiological studies and interventional trials suggest that polyphenols and mainly flavonoids present in fruits and vegetables are associated with a decrease in the risk of CVD. Oxidative stress can play a role in the pathogenesis of CVD, such as atherosclerosis. Polyphenols have recently been shown to have immunomodulatory and vasodilator properties, which can also help reduce the risk of CVD. Phenolic compounds may inhibit *in vitro* oxidation of LDL lipoproteins; oxidized LDL is related *in vivo* to the formation of atheromatous plaques, which may allow the development of coronary heart disease.

Anthocyanins are flavonoids responsible for the red and purple color of many fruits and some vegetables; they inhibit LDL oxidation and platelet aggregation, hence their cardioprotective effect.

Pancreatic α-amylase and α-glucosidase enzymes play an essential role in starch hydrolysis. *In vitro* studies have shown that phenolic compounds can act by inhibiting these enzymes, providing a strategy for postprandial management of hyperglycemia in patients with type 2 diabetes. Several *in vivo* studies have shown that phenolic compounds can reduce the blood glucose level of hyperglycemic rats. Subacute, chronic inflammation is considered an important factor in developing insulin resistance and diabetes in animals and humans. Several non-flavonoid polyphenols have been shown to reduce the production of inflammatory mediators in animal models with diabetes.

Health effects of fruits and vegetables

Gastrointestinal health

Constipation is a very common condition; fruits and vegetables have always been considered an adjuvant to prevent or relieve constipation. Higher dietary fiber intake and exercise are associated with a lower prevalence of this condition. Most clinical practice guidelines indicate that increased dietary fiber intake should be the first therapeutic measure in patients with chronic functional constipation (without symptoms of irritable bowel syndrome). Soluble or insoluble fiber content influences intestinal transit through different mechanisms. Constipation is associated with an increased risk of diverticulosis, colon cancer, appendicitis, and various anal conditions.

The amount of fiber from fruit (considering a daily consumption of 450 g) is estimated at 9 g. In addition, other components in the fruit can have a laxative effect, such as sorbitol or phenolic compounds. The fruit recommended due to its laxative effect is dried plum. The physiological reasons explaining the laxative effect of dried plum—which has a high fiber content (17%)—is not fully understood. Plum juice also affects intestinal motility and has small fiber content, so the substances responsible for this effect are believed to be hydrosoluble. Both dried plum and plum juice is rich in sorbitol. Studies have shown that doses of sorbitol of 0.4 g/kg weight for men and 1 g/kg weight for women have laxative effects. The average amount of sorbitol is 15% in dried plums and 6.1% in plum juice; therefore, this compound could not be solely responsible for the laxative effect. This fact suggests that there should be other compounds explaining the laxative effect of plums. In this sense, the phenolic compounds present in this fruit can also contribute to the laxative effect. Diphenyl xanthine, considered in the 1950s as one of the compounds responsible for the laxative effect of dried plums, has been discarded today, as this compound has not been isolated from any natural plant.

In recent decades, the increased consumption of kiwi has made it possible to study the effect of this fruit as a laxative. Thus, the fiber content is 3 g/100 g, and its laxative effect has been demonstrated in the elder population, provided that two pieces of fruit are consumed a day. In a 2018 review, they are indicated as a characteristic of this fruit: water retention, increases the size of the fecal bolus, softens its consistency and reduces transit time. They point to the enzyme actinidine together with soluble and insoluble fiber and polyphenols as responsible for its effect on intestinal malaise.

The fiber content of vegetables ranges from 1% to 4.3%, with values of 9.4% for artichokes. The inulin content of artichokes appears to be responsible for increasing the size of the fecal bolus, softening its consistency, and decreasing intestinal transit time. The cabbage family is the one with the highest inulin content. The ratio between soluble and insoluble fiber in vegetables is 1/1.5. The dietary fiber of wheat bran has a high insoluble fiber content.

Cardiovascular diseases (CVD)

Coronary heart disease and infarction are the two main manifestations of CVD. The evidence accumulated so far seems to support the idea that increased intake of fruits and vegetables can reduce this type of condition.

Protective mechanisms are associated with bioactive nutrients with antioxidant, *anti*-inflammatory, and electrolytic properties, as well as functional properties such as low glycemic load and energy density. Initially, the effects were associated with the presence of antioxidant substances, folates, fiber, potassium, flavonoids, and other phytochemicals. However, currently, the evidence suggests that it is the set of these components that can interact with genetic factors influencing cardiovascular risk. Therefore, it is more interesting to look at whole foods and dietary patterns rather than at a single food component (Eman and Gordon, 2017).

According to WHO data, insufficient intake of fruits and vegetables is estimated to be responsible globally for approximately 31% of ischemic heart disease and 11% of stroke.

In this regard, the effects on blood pressure of increased consumption of fruits and vegetables, both alone and together with a restrictive diet in total fat and saturated fatty acids, were evaluated in the DASH (Dietary Approaches to Stop Hypertension) trial. Although the combined diet reduced blood pressure more effectively, the fruit and vegetable diet also reduced it (systolic blood pressure: 2.8 mm Hg; diastolic blood pressure: 1.1 mm Hg) compared to the control diet. These reductions, apparently discrete on an individual scale, would result in a significant reduction in CVD risk at the population level due to the displacement of blood pressure distribution. Furthermore, diets with a high proportion of fruits and vegetables have been shown to reduce plasma lipid levels, mainly because of dietary fiber. Soluble fiber appears to have a larger effect on lowering low-density lipoprotein (LDL-C) cholesterol levels, and total fiber levels are inversely related to CVD. Diets rich in fiber help reduce triacylglycerols and blood pressure. Dietary recommendations encouraging increasing dietary fiber intake are relevant in people at higher risk for strokes, such as smokers, overweight, and those with hypertension.

In 2014 a meta-analysis of clinical trials and observational studies found that adherence to a vegetarian diet was associated with lower blood pressure (Wang et al., 2014).

Recent studies have related the intake of folates from fruits and vegetables to the decrease in blood homocysteine levels, which are positively correlated with the development of coronary diseases through various mechanisms. However, high homocysteine levels are not a specific indicator of inadequate folate intake because they may be due to insufficient intake of other vitamins, such as B_{12} or B_6, or other factors.

Concerning the role of potassium in the etiology of these diseases, the WHO suggests that supplementation is not necessary; it is sufficient with the recommended consumption portions of fruits and vegetables, keeping in mind that these food groups also have the advantage of providing only little sodium amount and therefore are beneficial in hypertensive states too.

The high intake of fruits and vegetables determines increased antioxidant capacity in plasma shortly after intake. Therefore, these antioxidant compounds could act against the oxidative processes, which are responsible for the development of CVD. Quercetin is the most frequently isolated flavonoid in vegetables and is a powerful antioxidant capable of inhibiting LDL oxidation and exerting a hypotensive action. Sulfur compounds of the genus *Allium* have shown the most prominent effects on the cardiovascular system, so garlic sulfur compounds are commercial products of high consumption in the United States as hypotensive (Suman and Shukla, 2016).

Increased potassium intake in adults reduces blood pressure, both systolic and diastolic. The WHO recommends increasing potassium intake from food to reduce blood pressure and the risk of CVD, stroke, and coronary heart disease. Despite no significant association has been found between potassium intake and the incidence of CVD or coronary heart disease, the positive relationship between blood pressure and these diseases is a piece of indirect evidence that an increase in potassium intake may reduce their incidence due to its effect on the decrease of blood pressure.

The WHO urges adults to consume at least 3510 mg/day of potassium and reduce this amount proportionally to the energy requirements of the child. Increasing potassium concentrations decreases systolic blood pressure in children, but this variation is not significant. The potassium recommendations are complementary to the WHO guidelines on sodium intake. If the recommended amount of potassium is consumed, the approximate ratio would be one to one, which is considered beneficial to health (WHO, 2012).

In summary, strong evidence about the beneficial effect of consuming appropriate amounts of fruit and vegetables on CVD prevention is available (Hartley et al., 2013).

Obesity

The intake of fruits and vegetables, which have a high-water content, decreases the average energy density of the diet. The caloric reduction in a diet high in fruits and vegetables could reduce the energy supply by up to 30%. Dietary fiber can slow down gastric emptying and create a feeling of satiety, especially soluble fiber, with higher water retention capacity, preventing excessive food intake.

Evidence-based nutritional recommendations for the prevention and treatment of overweight, developed by various scientific societies, conclude the following:

- Weight gain can be prevented with diets consisting of low energy-density foods.
- Strategies to enable the availability of healthy foods, such as fruits and vegetables, should be sought.
- Increasing fiber consumption from plant foods can prevent weight gain in healthy adults.
- Dietary prevention of weight gain can be modulated through diets with high fruit and vegetable consumption.
- A better control of diabetes mellitus could prevent overweight and obesity.

The European Guidelines for adult obesity management, published in 2019 (Durrer Schutz et al., 2019) specifically continues to recommend, decrease the energy density of food; increase vegetables and eat two portions of fruit per day.

Glucose metabolism

The role of different dietary components in developing type 2 diabetes mellitus has not been clearly established. However, several studies suggest a negative correlation between a relatively high intake of fruits and vegetables and the risk of developing this condition.

The fiber contained in vegetables could slow down the rate of nutrient absorption, resulting in savings in insulin secretion and reduced blood glucose levels.

Epidemiological cohort studies show that these compounds have a protective effect, regardless of age, against type 2 diabetes. Experimental studies have shown that high fiber intake reduces blood glucose and insulin levels in people with diabetes and results in low glucose tolerance.

Experimental studies suggest that soluble forms of fiber are beneficial, whereas prospective cohort studies seem to attribute this protective effect to insoluble forms; therefore, WHO has chosen to classify the relationship as "probable" rather than "convincing."

Foods with a low glycemic index, such as many vegetables, regardless of fiber content, are associated with a lower postprandial glucose response and an overall improvement in blood glucose control. However, a low glycemic index in itself is not a guarantee of overall health benefit because a high fat or fructose content of a food can also result in a lower glycemic index, and these foods can be of high energy density.

There is no universal dietary strategy to prevent or delay diabetes onset. Maintaining the ideal body weight and consuming a prudent diet or a Mediterranean diet pattern is considered the best strategy for preventing type 2 diabetes, especially if dietary recommendations consider individual preferences, which allows higher adherence to diet in the long term.

Cancer

The relationship between fruit and vegetable consumption and cancer was evidenced by the Public Health Service (1988) and the National Research Council of United States of America (1989) studies in 1990. In its 2002 World Health Report, WHO estimated that low fruit and vegetable intake was the cause of 19% of gastrointestinal cancer cases. The consistency of this evidence is reinforced by the large number of studies showing a statistically significant relationship.

Fruits and vegetables are sources of fiber, vitamins and minerals, carotenoids, and other antioxidants, as well as a large number of phytochemicals, such as organosulfur compounds, flavonoids, and related compounds; that could be the mechanism explaining the effects of these foods on cancer prevention. Most of these studies conclude with the need for further study of these mechanisms.

The World Cancer Research Fund (WCRF/AICR, 1997) has published a report on nutrition and cancer based on case-control studies conducted to that date and concluded that there was "convincing" evidence that vegetables reduced the risk of mouth

Table 7 Vegetables and fruit and the risk of cancer.

WCRF/AICR Grading		*Decreases risk*		*Increases risk*	
		Exposure	*Cancer site*	*Exposure*	*Cancer site*
Strong evidence	Probable	Foods containing dietary fiber	Colorectum, 2017		
		Non-starchy vegetables and fruit aggregated	Aerodigestive cancer and some other cancers (aggregated)		
Limited evidence	Limited suggestive	Non-starchy vegetables	Mouth, pharynx and larynx, 2018	Non-starchy vegetables	Colorectum, 2017
			Nasopharynx, 2017		
			Esophagus (adenocarcinoma and squamous cell carcinoma) 2016	Preserved non-starchy Vegetables	Nasopharynx, 2017
			Lung (people who smoke or used to smoke tobacco) 2017		
			Breast (estrogen receptor-negative) 2017		
		Fruit	Esophagus (squamous cell carcinoma) 2016		
			Lung (people who smoke or used to smoke tobacco) 2017		
		Citrus fruit	Stomach (cardia) 2016		
		Non-starchy Vegetables And fruit	Bladder, 2015	Fruit (low intake)	Stomach, 2016 Colorectum, 2017
		Foods containing carotenoids	Lung 2017		
			Breast 2017		
		Foods containing beta-carotene	Lung 2017		
		Foods containing vitamin C	Lung (people who smoke tobacco) 2017		
			Colorectum (colon) 2017		

Source: Work cancer research fund international (https://www.wcrf.org/dietandcancer/exposures/wholegrains-veg-fruit).

and pharynx cancer, esophagus, stomach, lung, colon and rectal cancer. Based on these findings, they established the following recommendation: «eat between 400 and 800 g, or five or more portions a day, of different vegetables and fruits, throughout the year». Later, a study comparing the effects of a diet poor in fruits and vegetables with the effects of another diet with a high content of these foods showed a significant decrease in DNA damage and lipid oxidation in individuals fed with the second diet; this was attributed to the high intake of fruit and vegetables.

The latest updates on this report published by WCRF International in 2018 are shown in **Table 7**. In conclusion, they indicate that higher consumption of fiber-containing foods is likely to protect against colorectal cancer and generally that higher consumption of starch-free fruits and starch-free vegetables is likely to protect against aerodigestive cancer. In addition, this intake has been consistently related to a dietary pattern associated with a lower risk of cancer and other noncommunicable diseases, as well as obesity (WCRF, 2018.)

Vision

Some fruits and vegetables have remarkable amounts of lutein and zeaxanthin, excellent fat-soluble antioxidants with an affinity for the eye system. Cataracts and macular degeneration are common vision-related aging diseases, affecting a high percentage of the population over 65 years old. Consumption of some varieties of these food groups, in particular those containing lutein and zeaxanthin, appears to reduce the risk of cataracts.

Conclusions

Fruits and vegetables are essential foods in a daily diet because of their content in nutrients and bioactive compounds. Their low energy density and high hydration capacity provide a feeling of satiety and contribute to adequate nutrition. Recent studies have shown that fruit and vegetable consumption prevents the risk of chronic diseases such as obesity, cardiovascular disease, type II diabetes, and cancer. Therefore, the WHO recommends a daily intake of fruits and vegetables of at least 400 g.

References

Barba, F.J., Mariutti, L.R.B., Bragagnolo, N., Mercadante, A.Z., Barbosa-Cánovas, G.V., Orlien, V., 2017. Bioaccessibility of bioactive compounds from fruits and vegetables after thermal and nonthermal processing. Trends Food Sci. Technol. 67, 195–206.

Durrer Schutz, D., Busetto, L., Dicker, D., et al., 2019. European practical and patient-centred guidelines for adult obesity management in primary care. Obes. Facts 12 (1), 40–66.

Eman, M.A., Gordon, A.F., 2017. Dietary fruits and vegetables and cardiovascular diseases risk. Crit. Rev. Food Sci. Nutr. 57 (9), 1950–1962.

Garcia-Villanova Ruiz, B., 2017. In: de Nutrición, G.A.T. (Ed.), Hortalizas y productos hortícolas, 3ª ed., vol. III Madrid: Panamericana.

Hartley, L., Igbinedion, E., Holmes, J., et al., 2013. Increased consumption of fruit and vegetables for the primary prevention of cardiovascular diseases. Cochrane Database Syst. Rev. 6. Art. No.: CD009874.

Mandrich, L., Caputo, E., 2020. Review. Brassicaceae-derived anticancer agents: towards a green approach to beat cáncer. Nutrients 12, 868.

Quirante-Moya, S., García-Ibañez, P., Quirante-Moya, F., Villaño, D., Moreno, D.A., 2020. The role of brassica bioactives on human health: are we studying it the rightway? Molecules 25, 1591.

Rizwani, G.H., Shareef, H., 2011. Genus *Allium*: the potential nutritive and therapeutic source. J. Peripher. Nerv. Syst. 1, 158–165.

Rodriguez-Concepcion, M., Avalos, J., Bonet, M.L., et al., 2018. A global perspective on carotenoids: metabolism, biotechnology, and benefits for nutrition and health. Prog. Lipid Res. 70, 62–93.

Ruiz-López, M.D., Garcia-Villanova Ruiz, B., 2017. Frutas y productos derivados. In: Gil, A. (Ed.), Tratado de Nutrición, 3ª ed., vol. III Madrid: Panamericana.

Shashirekha, M.N., Mallikarjuna, S.E., Rajarathnam, S., 2015. Status of bioactive compounds in foods, with focus on fruits and vegetables. Crit. Rev. Food Sci. Nutr. 55, 1324–1339.

Suman, S., Shukla, Y., 2016. Diallyl sulfide and its role in chronic diseases prevention. In: Chandra Gupta, S., Prasad, S., Aggarwal, B.B. (Eds.), Drug Discovery From Mother Nature, Advances in Experimental Medicine and Biology, vol. 929. Springer International Publishing, Switzerland.

U.S. Department of Agriculture, Agricultural Research Service, 2019. FoodData Central. Available in. https://fdc.nal.usda.gov/.

Wang, X., Ouyang, Y., Liu, J., et al., 2014. Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. Br. Med. J. Open 349, g4490.

WCRF. World Cancer Research Fund/American Institute for Cancer Research, 2018. Food, Nutrition, Physical Activity, and the Prevention of Cancer: A Global Perspective. Continuous Update Project Expert Report 2018.

WCRF/AIRC, World Cancer Research Fund/American Institute for Cancer Research, 1997. Food, Nutrition, Physical Activity, and the Prevention of Cancer: A Global Perspective. AIRC, Washington, DC.

WHO, 2012. Guideline: Potassium Intake for Adults and Children. WHO, Genève.

Glucose and its polymers: Chemistry, sources, digestion and metabolism

Julie Miller Jones, St. Catherine University, St. Paul, MN, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of D.J.A. Jenkins, L.S.A. Augustin, A. Malick, A. Esfahani, C.W.C. Kendall, Glucose: Chemistry and Dietary Sources, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 372–380, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00133-1.

Glucose: an introduction 413
Chemistry of glucose and glucose polymers 413
Monosaccharides 413
Glucose in disaccharides and oligosaccharides 414
Glucose from carbohydrate staple foods 416
Resistant starch 416
Nondigestible glucose polysaccharides 417
Digestion of glucose and its polymers 418
Digestion of sugars 418
Digestion of starch and related polymers 418
Slowly digested and resistant starch 419
Blood glucose response and glycemic index 419
Conclusion 420
References 421
List of Relevant Websites 421

Key points

- The structure, chemistry, digestion and dietary sources of glucose monosaccharides, disaccharides, oligosaccharides and polysaccharides—both available (energy yielding) and unavailable (dietary fiber)—will be delineated.
- The utilization and function of these carbohydrates including starch moieties amylose and amylopectin, and dietary fibers such as resistant starches, celluloses and hemicelluloses will be discussed.
- The digestion and absorption of various glucose entities, their rate of entry into the blood stream, and potential health impacts will be described
- The glycemic index (GI) will be introduced and the issues and controversies around its use and variability will be explored.

Glossary

Amylase Salivary and pancreatic enzymes that hydrolyze α-glycosidic bonds of starch

Amylose A moiety of starch with glucose units joined linearly by α(1→4) glycosidic bonds, which enables them to form tight crystalline arrays

Amylopectin A moiety of starch with glucose units joined by α(1→4) glycosidic bonds but has α(1→6) bonds creating a more open, branched structure

Beta-glucan Unavailable glucose polymers from certain plants that are soluble, viscous fibers with documented health benefits

Brush-border Small intestinal, epithelial cells covered by finger-like microvilli, that are the key to splitting of disaccharides and glucose transfer from the intestinal lumen into the bloodstream

Cellulose A linear, strong, unavailable glucose polymer forming structural components of plant tissues and in the diet becomes insoluble dietary fiber and the main form of fecal bulk

Degree of Polymerization (DP) The number of sugar molecules in an oligosaccharide or polysaccharide.

Dietary fiber A carbohydrate polymer that is unavailable, meaning that it is not digested by enzymes in the upper gastrointestinal tract of humans

Disaccharide Two basic carbohydrate units (e.g., sucrose, lactose, maltose)

Glycemic Index (GI) The physiologic classification of carbohydrate-containing foods based on postprandial blood glucose responses

GLUT carriers One of 5 carriers of glucose that facilitate transport of sugars across cells and membranes

Hemicelluloses Unavailable glucose polymers are often associated with cellulose in the plant and function as dietary fiber and tend to be readily or partially fermented by colonic bacteria

Insoluble fiber Dietary fiber resistant to digestion and insoluble in cold and hot water. Some may be partially degraded by colonic bacteria
Monosaccharides A single sugar molecule (e.g., glucose, galactose, fructose)
Oligosaccharides Small polymers with from 3 to 20 two sugar units
Polysaccharides Large polymers of many sugar units such as starch or cellulose
Resistant starch Starches that escape digestion in the upper gastrointestinal tract. Five types have been identified
Short-chain fatty acids (SCFA) Metabolically active fatty acids with few carbons that are byproducts of bacterial fermentation of food components in the colon (e.g., acetic, propionic, and butyric acids)
Sodium-linked glucose transporter (SGLT) Carriers using active transport (ATP-requiring) to carry glucose (and galactose) across the mucosa and through the membrane
Soluble fiber Glucose polymer or dietary fiber resistant to digestion but soluble in cold and hot water
Starch Glucose polymers formed by at least 1000 glucose units joined by $\alpha(1\rightarrow4)$ glycosidic bonds in the linear amylose and $\alpha(1\rightarrow4)$, $\alpha(1\rightarrow6)$ in the branched amylopectin found in grains, roots, tubers and legumes
Viscous fiber A type of soluble fiber which becomes viscous on contact with water

Glucose: an introduction

Glucose is the most abundant of the simple sugars and the preferred energy source of most living cells. Plant foods are major dietary sources of sugars and available and unavailable carbohydrates. Plantchloroplasts capture energy in photons of light during photosynthesis and join carbon dioxide from air to water, usually from the roots, and form glucose and oxygen. The glucose stores energy in its bonds. Thus energy-yielding carbohydrates contain 4 kcal g^{-1}. Glucose and other sugars are found naturally in fruits, vegetables and other foods such as honey, but sugar refined from cane and sugar beets is also a significant dietary component in many diets. The primary source of blood glucose should be from starchy carbohydrate foods -grains, roots and tubers and legumes, where it is recommended that ~50% of calories be derived.

While glucose in the diet is ready for absorption, digestible sugars and carbohydrates must be split to their monosaccharide form(s), usually glucose, for absorption. When absorbed glucose becomes circulating form of carbohydrate in the bloodstream. It is also a key component in plant cell walls and fibrous plant structures that provide our diets with important unavailable carbohydrates in the form of dietary fibers. This entry will look at the chemistry, structures, functions, digestion of monosaccharides, disaccharides and polysaccharide components. Further this entry will discuss their impacts of plant and starch chemistry on blood glucose response, glycemic index and other aspects of human nutrition.

Chemistry of glucose and glucose polymers

Plant-based carbohydrate foods formed the basis of dietary patterns worldwide for millennia and usually provide more than half the energy in the human diet. This comes from foods that contain "available" carbohydrate, which means that it can be broken down to glucose or other monosaccharides for absorption in the human small intestine and for delivery in the bloodstream to provide energy throughout the body. Energy stored in bonds of available carbohydrate is released via glycolysis and other pathways as ATP and other high-energy metabolites. Glucose polymers, not split by enzymes in human upper gut, move into the large intestine to function as dietary fiber.

Monosaccharides

Glucose and other monosaccharides in plant foods exist either free or bound. When bound to nutrients or phytochemicals, they may impact nutrient availability and stability, impart flavor or mask bitterness.

Chemically, the monosaccharide D-glucose ($C_6H_{12}O_6$, MW 180.16 g mol^{-1}), from the Greek (*glukus*, sweet) or dextrose, is 2,3,4,5,6-pentahydroxyhexaldehyde. Translated, glucose is a six-carbon aldohexose. Glucose differs from other aldohexoses in positioning and conformation of hydroxyl groups on the hexose ring and from the ketohexose, fructose with its 5-member furanose ring.

Positioning of hydroxyl groups influences sugar solubility and ability to bind to the sweet taste receptor on the tongue for eliciting a sweet taste by. Glucose ranks second, after fructose, in sweetness compared and sugars and syrups (**Table 1**). Sweet taste is hard-wired, and its pleasurable sensation have caused humans to seek sources of sugar and sweetness, such as honey or maple sap, for preservative and sensory properties. Despite their being major sources of glucose, sweetness is not a characteristic of most starchy foods, although roasting or toasting may increase sweetness because starch dextrinizes or react with protein in the Maillard reaction. Slight sweetness is characteristic of some glucose oligosaccharides such as inulin and chicory root, which also function as dietary fiber.

Table 1 Relative sweetness of sugars and syrups.

Sugars	*Sweetness rating*
Fructose	120–180[a]
Sucrose—glucose bonded to fructose	100
Invert sugar—hydrolyzed sucrose to glucose + fructose	85–100
Glucose (dextrose)	70–80
Trehalose (3 glucoses)	50
Galactose	35
Corn syrups (20–42 DE[b])	23–48
Maltose = glucose bonded to glucose	30–50
Lactose = glucose bonded to galactose	16–20
Maltodextrin	6–21
Syrups	
Date sugar/syrup -glucose, fructose, trace sucrose	120
Birch syrup—fructose, glucose, trace sucrose	110
Honey- fructose, glucose, maltose and some sucrose	110
Cane juice (sucrose, glucose, fructose)	100
High fructose corn syrup[c]	100–120
Palm sugar or syrup—sucrose, glucose, fructose	100

[a]Depends on concentration and pH.
[b]DE = Dextrose equivalents.
[c]Sweetness depends fructose: glucose ratio, often 55:45.

Free glucose is found often with other sugars usually fructose or sucrose in fruits, vegetables and syrups. **Table 2** gives an average content of sugars of common fruits and vegetables, but the actual amounts vary by variety, agronomic conditions, and ripeness. For example, an unripe banana is not sweet and is quite astringent, but as the banana ripens acids decrease and some starch is metabolized to release glucose, changing both the taste and texture and decreasing the amount of resistant starch in the banana.

Glucose in disaccharides and oligosaccharides

Common **disaccharides** -sucrose, lactose, galactose, and maltoses -often contain glucose as one of the sugars. In sucrose (table sugar), glucose bonds to fructose by an $\alpha(1\rightarrow2)$ link, making it a non-reducing sugar. This property differentiates it from most common saccharides, which are all reducing sugars, so they readily attach to lysine in the Maillard browning reaction. Sucrose must by be split by heat or sucrase (invertase). Baking properties such as color and flavor and lysine availability by amount of Maillard reaction.

Lactose, the sugar in nearly all mammalian milks, has glucose joined to galactose by an $\alpha(1\rightarrow4)$ glycosidic bond. Compared to other sugars, lactose dissolves slowly and is the least sweet of the sugars (**Table 1**). Maltose, isomaltose and trehalose have two glucose molecules, but their linkages differ, $\alpha(1\rightarrow4)$, $\alpha(1\rightarrow6)$ and $\alpha(1\rightarrow1)$, respectively. They are less sweet than many sugars. Digestion of starch and related dextrins supply the body's major source of maltose and isomaltose, but trehalose is found in honey, some plants, fungi, and fermented foods. Glucose polymers in carbohydrate foods subjected to dry heat (e.g., toasting) or moist heat, especially with acid, can be hydrolyzed to produce maltose. In addition, sprouting, malting, and fermentation of grains or other carbohydrates, as in mashes or yeast-leavened doughs, also produce maltose.

Oligosaccharides (Greek *oligo*, few) contain 3–20 sugar units, making their degree of polymerization (DP) 3–20. Plant oligosaccharides, function both as energy storage as chicory root inulin or structural or gummy components such as β-glucan in oats, barley and fungi. All oligosaccharides, including small ones such as the trimer (maltotriose) and tetramer (maltotetrose) even with $\alpha(1–4)$ glycosidic linkages, cannot be hydrolyzed by human amylases. They function as prebiotics and other fibers fermented by colonic microorganisms. Human milk is a source of animal oligosaccharides, but glucose is not the dominant sugar. Galactooligosaccharides contain glucose, galactose, sialic acid, fucose and N-acetylglucosamine with the precise combination varying by stage of lactation in human milk and can also be found in some plant foods. In human milk the oligosaccharides serve as probiotics that help foster a healthy microbiome and fight infection. Most jurisdictions label all oligosaccharides as dietary fiber, but a few disallow those with DPs 3–9.

Polysaccharides have 1000 or more sugar units and high molecular weights. In plants, many are glucose polymers that either provide stored energy, usually as starch, or materials for structural plant components (e.g., celluloses and hemicelluloses). Starch chains are formed when glucose, not needed immediately by the plant, is shuttled to amyloplasts for polymerization. Starch chains align into starch granules whose shape, size, organization, crystallinity, number, and location (e.g., roots, rhizomes, seeds, stems, tubers and corms) and amylose: amylopectin ratio are determined by plant type and variety.

Table 2 Sugars of fruits and vegetables (g/100 g of fruit).

Fruit	*Carbohydrate*	*Total sugar*	*Glucose*	*Fructose*	*Sucrose*
Apple	13.8	10.4	2.3	5.9	3.1
Apricot	11.1	9.3	1.6	0.7	5.2
Avocado	8.5	0.9	0.5	0.2	0.1
Banana	22.8	12.2	5.0	4.9	2.4
Blackberry	9.6	8.1	3.1	4.1	0.4
Blueberry	8.2	7.3	3.5	3.6	0.2
Cantaloupe	7.9	8.7	1.2	1.8	5.4
Cherry	16	14.6	4.2	3.3	0.5
Grapefruit	8.1	6.2	1.3	1.2	3.4
Grapes	18.1	18.1	6.5	7.6	0.2
Kiwi	14.7	10.5	5.0	4.3	1.1
Lemon	9.3	2.5	1.0	0.8	0.6
Limes	10.5	0.4	0.2	0.2	0
Mango	17	14.8	0.7	2.9	9.9
Nectarine	10.6	8.5	1.2	0	6.2
Orange	11.7	9.2	2.2	2.5	4.2
Peach	9.5	8.4	1.2	1.3	5.6
Pear	15.5	9.8	2.8	6.2	0.8
Pineapple	13.1	11.9	2.9	2.1	3.1
Plum	11.4	7.5	2.7	1.8	3.0
Pomegranate	16	10.1	5.0	4.7	0.4
Raspberry	11.9	9.5	3.5	3.2	2.8
Strawberry	7.7	5.8	2.2	2.5	1.0
Tomatoes	3.9	2.3	1.0	1.2	0
Watermelon	7.6	9.0	1.6	3.3	3.6
Beet	9.6	6.8	0.1	0.1	6.5
Carrot	9.6	4.7	0.6	0.6	3.6
Corn	19.0	3.2	0.5	0.5	2.1
Red pepper	6.0	4.2	2.3	1.9	0
Onion	7.6	5.0	2.0	2.3	0.7
Sweet potato	20.1	4.2	0.7	1.0	2.5

Data from USDA Nutrient Database, Nutritionist 5. FSANZ Australian Food Composition.

Starch, $(C_6H_{12}O_6)n$, is comprised of two moieties—the mostly linear amylose or the branched amylopectin (Fig. 1). While both have $\alpha(1–4)$ bonds, amylopectin has $\alpha(1–6)$ bonds at the branch points, approximately every 25 glucose units. Amylose (DP1000–10,000; MW 10^5–10^6 g mol^{-1}) is smaller than amylopectin (DP $> 10^6$; MW $\sim 10^8$ g mol^{-1}).

Linear amylose chains form helical structures because hydroxyl molecules of one glucose bonds to the aldehyde portion of another glucose forming an internal hemiacetal bond causing the chain to coil. Adjacent helices hydrogen-bond together into compact, crystalline arrays, making amylose require more energy for hydration and gelatinization, absorb less water during cooking, form stronger gels, and retrograde more readily than amylopectin.

Amylopectin has areas of crystallinity, but its branched, less compact structure is far less crystalline than amylose. Amylopectin hydrates easily forming soft pastes (sols) characterized by clarity, stability, and resistance to retrogradation and gelling on aging.

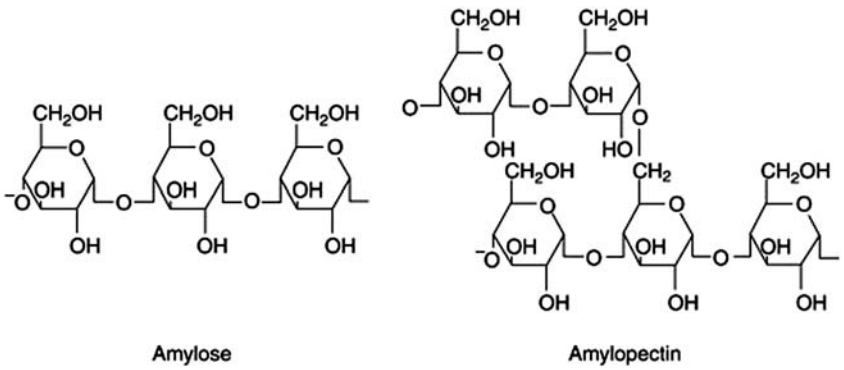


Fig. 1 Partial structures of amylose (linear) and amylopectin (branched) starches.

Further, its many branches give multiple sites for amylase attachment. All these aspects make amylase hydrolyze hydo amylopectin faster than amylose and speed glucose digestion and delivery to the bloodstream.

Starch granule structure also matters. Small granules with their larger surface area enable greater water absorption, greater aggregation, and higher glycemic responses than larger granules. Large intact granules can slow glucose release, but the resulting gelatinized starch may gel can be fragile making the gel easily penetrated by amylose and hasten glucose release into the bloodstream.

Dextrins and smaller maltodextrins (MW $>$ 1000 g mol^{-1}) are glucose polysaccharides formed from starch, either by dry heat dextrinization or hydrolysis by enzymes or moist heat and acid. The latter process is used to produce traditional corn syrup, a mildly sweet food (Table 1). Manufacturers add glucose isomerase to corn syrup, which converts some glucose to fructose, to yield a sweetening power equivalent to sucrose (Table 1). Powdered dextrins, used as food additives, add viscosity, promote water holding, or enhance browning.

Like starch, most dextrins are digestible in the upper gut and yield 4 kcal g^{-1}. However, resistant dextrins with their α-1,2 or α-1,3 linkages cannot be hydrolyzed by amylases in the upper gut rendering them soluble dietary fibers.

Glucose from carbohydrate staple foods

Glucose needs have traditionally been met by starchy staples, which can be prepared in countless ways, or their starches isolated as ingredients. Roots and tubers such as potato, yuca, and cassava or seeds of cereal and pseudocereals, and legumes are the major sources. Their starch content averages 50–80% of dry matter, except in legumes, which range from 1 to 58% dry matter with soy and peanut being very low. The actual amounts for all staples are affected by variety, cultivar, and agronomic conditions.

Most native starches are a mix of amylose and amylopectin with root starches having more amylopectin than many grains (Table 3). Some natural and hybrid grain varieties are high in amylose, which lowers their glycemic responses. Others are entirely amylopectin (sometimes called "waxy" or "glutinous" grains) and have high glycemic responses. Some pseudocereals such as amaranth have starches characterized by small granules and very little amylose and crystallinity, so they elicit high glycemic responses.

Legume starches are higher in amylose than other staples, averaging 30–50% amylose, but ranging from 14% in chickpea and 88% in peas (Table 3). Most tend to have low glycemic responses due not just to amylose content, but also to starch structures that favor creation of resistant starch (RS).

Resistant starch

Until the 1980s, all starch was believed to be completely digested in the upper gut. However, research studies detected undigested starch in the large bowel. These starches were dubbed "resistant," and function as dietary fibers. RSs are fermented to short-chain fatty acids (SCFA), which have beneficial effects locally in the colon but can be absorbed to impact metabolism throughout the body. Studies show substitution of RS for available starch can lower the blood glucose, improve insulin sensitivity, and reduce hunger.

Table 3 Amylose, amylopectin, and resistant starch in staple foods.

Source	*Amylose (%)*	*Amylopectin (%)*	*Resistant starch*
Cereal grains			
Maize	28	72	1.7
Wheat	26	74	
Hi amylose maize			47.4
Barley, pearled	14–38	42–86	
Waxy maize	–	100	
White rice	20–24	76–80	1.3
Legumes	13–88	12–87	1.2–3.4
Navy bean	29	71	2.0
Cow pea	43	57	
Pea	88	12	2.6
Root tuber/other			
Cassava	17	83	
Potato	21	79	
Tapioca	17		
		83	
Buckwheat			1.8

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6836628/table/T1/?report=objectonly USDA Nutrient Data Database.

Table 4 Resistant starch (RS) classifications and examples.

Designation	*Description*	*Example*
RS1	Physically inaccessible starch	Whole kernel and coarsely ground seeds, grains or meals
RS2	Granular starch with the B- or C-polymorph	Starches high in amylose—basmati rice, high amylose grain varieties/cultivars, raw potato, raw banana starch
RS3	Retrograded starch	Cooked and cooled starchy foods, heat and moisture treated grains or starches
RS4	Chemically modified starches	Cross-linked starches, used as additives
RS5	Starch-lipid complex	Lipids -complexed amylose starch and longer branch chains of amylopectin

Five RS classes exist (**Table 4**) (Patterson et al., 2020). RS1 escapes small intestinal digestion because of large particle size, compact structure, or cellular compartments that sequester starch from amylase attack (e.g., whole-grain pumpernickel or grainy breads, bulgur wheat, or corn kernels). RS2 is high in amylose (e.g., green banana, high-amylose grains, and legumes). RS3 occurs in starchy foods that have been cooked and cooled (e.g., potato salad, fried rice, and "stale" bread) causing retrogradation. Starches higher in amylose are more likely to form RS3. RS4 is chemically cross-linking to make them resist digestion. RS5 entraps lipids within helical amylose rendering both fat and starch resistant to digestion.

Amounts of RS in foods are highly variable. Most cereals contain <3% RS, but high amylose varieties have 45–60% and green bananas contain ~75% (**Table 3**). In legumes RS content varies from 3% for pea and chickpea to >26% in raw beans but increases to over 50% in some processed products. For all foods the amount of RS depends on their processing, storage conditions, and accompanying ingredients. RS comprises only a small part of Western diets. For example, the daily average per capita intake in the USA is 3–8 g.

Glycogen is the only form of carbohydrate storage in the human body. Storage capacity is limited ~100–120 g in liver and 400 g in muscle. Glycogen's branched structure is akin to amylopectin but contains many fewer glucose molecules (2000–60,000) and is smaller (MW ~ 10^6 g mol^{-1}). Glycogenolysis can generate glucose needs when none is readily available. Because glycogen stores are small, glycogen cannot be depended as way to meet long term glucose needs. During energy surfeit, signaled by an excess of phosphorylated glucose from glycolysis, insulin activates enzymes that start glycogen synthesis.

Nondigestible glucose polysaccharides

Cellulose $(C_6H_{10}O_5)n$, a linear, β-(1–4)- linked, glucose polymer (DP ~15,000), forms strong intrachain bonds that result in fibrous structures critical for plant cell walls and structures. Human enzymes cannot break β-(1–4)- links. Thus, cellulose reaches the colon, where it undergoes limited fermentation by colonic bacteria, thereby, contributing a significant amount of insoluble dietary fiber. The proportion fermented depends on the source, with vegetable cellulose degraded to a greater extent than cereal cellulose. Chemically modified celluloses (e.g., carboxy-methyl cellulose) undergo limited fermentation. The proportion of cellulose and other fibers that escape colonic bacterial fermentation contributes directly to fecal bulk and water holding (Older food composition tables report "crude fiber" derived fiber content by using strong reagents capable of destroying material that functions biologically as dietary fiber. Thus, these tables underreport the actual fiber content compared to tables using methods approved to emulate physiology).

Hemicelluloses got their name because they are often found with cellulose in plant structures. They are heteropolysaccharides that contain glucose, other sugars, and a variety of linkages. While some contain saccharide backbones with β(1–4)-linkages, they are unlike cellulose as they much smaller (DP 500–3000), highly branched, amorphous structures that form weak intrachain bonds. Rather than fibrous, they are soluble and often viscous making them readily fermented by colonic microbes. Their four classes are as follows: (1) mixed linkage β-glucans, (2) xyloglucans, (3) xylans, and (4) mannans. β-glucans and xyloglucans have glucose backbones, so they will be discussed here.

***β*-Glucans** are water-soluble glucose polymers found in cell walls of certain plants. Linear bacterial and algal glucans have β-(1→3) glycosidic bonds; whereas branched yeast and mushroom glucans have β-(1→3) and -(1→6), or oat and barley glucans have β-(1→4) and β-(1→3) glycosidic bonds. All are soluble in water and capable of forming viscous sols, but the viscosity is dependent on the number of β(1→3) linkages and the molecular weight. β-glucans of barley and oats have been shown to lower serum cholesterol and reduce blood glucose, and heavier and more viscous molecules show greater physiological impact. Oat β-glucan has an average MW 206 -230 × 10^4 g m^{-1}. However, processes such as fine grinding, long dough fermentation, or treatments such as acid, heat, or enzymatic hydrolysis that decrease molecular weight or viscosity, lower physiological impact.

Extracted β -glucan is used to increase viscosity, cohesiveness, and dietary fiber content of various food products. Extraction from oat is easier because the glucan is concentrated in outer bran layers but barley, has β-glucan dispersed throughout the endosperm. High β-glucan cultivars of both barley and oats are being sought. Small amounts of β-glucans are found in wheat, rye, millets, yeasts, bacteria, and some mushroom and seaweed varieties. However, the health impacts are not well studies and may be limited by their presence in small amounts.

Xyloglucans are cell wall polymers that interact with cellulose microfibrils. Structurally, it is like cellulose with α-D-xylose, a 5-carbon sugar, replacing β-1,4-D-glucose in position 6. They help control plant functions such as ripening and impart pest resistance to plants. In humans they may have an immune modulating role.

Arabinoxylans are the predominant hemicellulose polysaccharide in cereals. Glucose is a minor constituent because the pentoses arabinose and xylose form their backbone. Despite low concentrations, they may have important physiological impact because the quantity of grains such as wheat eaten in many cultures.

Digestion of glucose and its polymers

Digestion of sugars

Most carbohydrates must be broken down to monosaccharides during digestion, but monosaccharides are ready for absorption upon ingestion. All sugars are digested quickly as they travel to the stomach where they mix with gastric acid and stomach contents. They become part of chyme, which leaves the stomach through the pylorus into the duodenum. There, monosaccharides can be absorbed quickly, especially if in liquid or little fat or protein is present.

Disaccharides must be split to their respective monosaccharides by brush border enzymes—sucrase–isomaltase complex, maltase, and lactase-found in the villi of the jejunum and upper ileum prior to absorption. Several specific transporters are required to move the monosaccharides from the lumen and through the enterocytes of the intestine (Koepsell, 2020). These include sodium-linked glucose transporter (SGLT1) that carries glucose (and galactose) across the mucosa and through the membrane using active (ATP-requiring) transport. Sugars transport via facilitative diffusion involves GLUT carriers, which are sugar and tissue specific. After glucose enters the intestinal epithelium, it is carried to the basal cell side by the facilitative transporter GLUT2 for delivery to the portal vein. Once in the bloodstream glucose travels to the liver, where liver GLUT2 extracts 30–40% of the glucose to phosphorylate it in preparation for glycolysis or synthesis of glycogen. The glucose remaining in circulation is delivered to other tissues where most express GLUT1 that transfers glucose into cells for glycolysis. The largest amount of glucose goes to muscle and adipose where GLUT4 is expressed to speed glucose uptake. While neither brain nor red blood cells require insulin for efficient glucose for uptake, however, insulin helps regulate oxygen and glucose metabolism and prevent hemolysis.

Digestion of starch and related polymers

Starch must go through a series of steps to release maltoses and glucose ready to be metabolized and absorbed as described (Lovegrove et al., 2017). Salivary amylase begins starch digestion in the mouth but acts only on gelatinized starch and for the brief time while the food bolus remains pH neutral. Stomach acid stops amylases. As stomach's acid chyme moves into the duodenum, it triggers bicarbonate release and returns pH to neutral. This allows pancreatic amylase to hydrolyze starch to maltose and maltotriose. In addition to amylase, amylopectin requires α-glucosidases to breaks down the α-limit dextrin to glucose ready for absorption (Fig. 2).

Most starch digestion is completed upon leaving the jejunum; however, both rate and completeness depend on numerous factors such as the following.

(1) Gelatinized starches are 85% hydrolyzed, but under 20% of ungelatinized may be.
(2) Amylopectin with its multiple reducing ends for enzyme attachment and open structure is rapidly and completely digested.
(3) Starches high in amylose with its linear chains and many fewer reducing ends for amylase attachment digest slowly. Further, the compact, crystalline structure impedes amylase penetration and enzymatic activity.

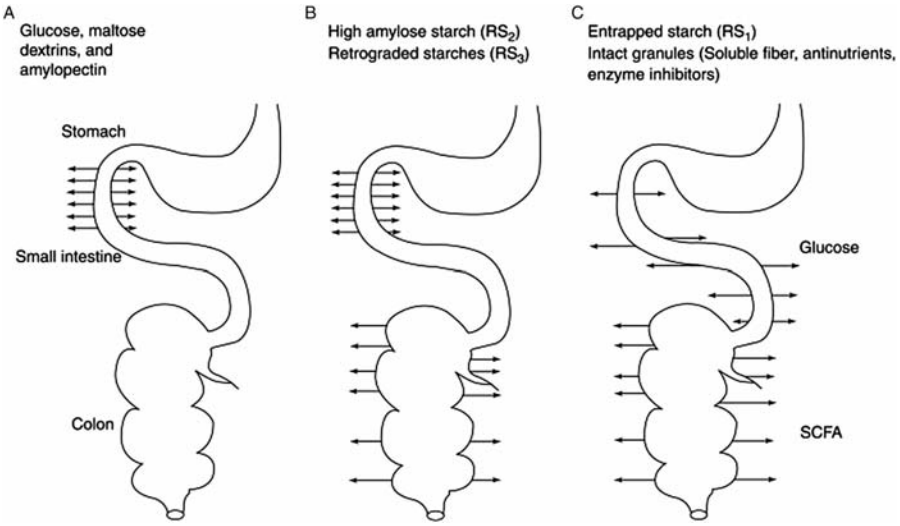


Fig. 2 Effect of different forms of glucose on glucose absorption and short-chain fatty acid (SCFA) production and uptake from the gut.

(4) Starches that are surrounded by seed coats, fibers or fat impair enzyme penetration and reduce both the speed and completeness of starch digestion.
(5) Viscous soluble fibers can delay gastric emptying and act as physical barriers to intestinal absorption.
(6) Starch–protein and starch–fat interactions both reduce digestion rate, and the fat slows gastric emptying.
(7) Antinutrients (e.g., amylase inhibitors, lectins, tannins and phytates) and acids can inhibit starch breakdown in various ways. For example, phytate binds calcium needed by amylase, so it reduces the rate of small-intestinal starch digestion.

Food processing can change the rate of digestibility. It can increase the rate by reducing the concentration or activity of inhibitory food components or by modifying the structure of the food or the starch change its available to amylase. Examples include heating to gelatinize starch or milling crushing, flaking, or extruding crack seed kernels to increase surface area and decrease particle size. Rates of starch digestion can decrease when formulations or processes bind with or coat the starch or add or generate acid to lower the pH as in sourdoughs. Starch retrogradation generated by repeated heating and cooling, or extrusion can produce compact physical structures, as occurs during production of slow-release breakfast bars, pasta and some cereals, can decrease enzyme penetration.

Slowly digested and resistant starch

Digestion is slowed if undigested starch reaches the ileum because it triggers the ileal brake; in the large intestine, the colonic brake. Its presence also stimulates ileal-activated glucagon-like peptide-1(GLP-1) and peptide YY (PYY). These can suppress food intake, increase satiety, and decrease gastric emptying.

Foods that naturally contain so-called "slowly digestible starch" (SDS) are legumes or intact wholegrains or manufactured foods produced using formulations and processes usually repeated heating and cooling. Studies confirm that increasing SDS in food products slows the rate of appearance of dietary (exogenous) glucose in the bloodstream and lowers the rate of plasma glucose disappearance (Breyton et al., 2021). Thus, swings in blood glucose and insulin are blunted resulting in a steady fuel supply. SDS consumption is associated with a lower blood glucose response and beneficial changes in the microbiome and its fermentation. Such changes are factors associated with reducing disease risks related to impaired glucose tolerance.

Undigested starch in the colon can also cause marked changes in the microbiome and in fermentation products and is deemed resistant starch (**Fig. 2**).

Any starch not digested in the small intestine and moves into the colon becomes RS. Bacteria ferment it to hydrogen, methane and SCFAs. These provide the colon with about 10% of its energy and acidify the colonic lumen. The SCFA mix averages 60% acetate, 20% propionate, and 20% butyrate. However, the proportion depends on the size and composition of the meal, speed of transit, where in the colon fermentation is occurring and the species inhabiting the microbiome.

Different fatty acids have different roles. Butyrate specifically promotes healthy colonocytes by maintaining the mucous lining, combating inflammation, and reducing pathogenic and promoting beneficial microorganisms. Butyrate and SCFAs not utilized in the colon are absorbed and travel to the liver, where they enter the tissues with G-protein-coupled receptors. They act as cytokines to control energy metabolism by activating kinases that regulate hormone release, impact appetite, and communicate with the central nervous system (Hu et al., 2018). They also appear to mitigate negative effects of high-fat diets by inhibiting lipid synthesis and hepatic cholesterol synthesis and promoting lipolysis.

Blood glucose response and glycemic index

Different carbohydrate foods vary dramatically in their postprandial blood sugar response, so the role of specific carbohydrate foods in health and disease is a subject of intense interest and debate. The glycemic index (GI) measure was developed to compare the rates at which different starchy foods deliver glucose to the bloodstream and raise postprandial glycemia. The GI theory suggests that choosing foods that slow blood glucose entry might aid in treating conditions with abnormal glucose tolerance and decrease the risk of diabetes and other chronic diseases.

Specifically, the GI ratio compares the 2 h blood glucose response curve for one subject ingesting 50 g of available carbohydrate from the test food with his/her own glucose response curve generated using a standard, usually 50 g of glucose. The integrated area-under-the curve (AUC) for the test food is divided by the subject's AUC for the glucose standard. The quotient is then multiplied by 100. A food's glycemic load (GL) is calculated by multiplying the grams of carbohydrate in the portion eaten multiplied by the GI (It is important to note that 50 g of available carbohydrate is rarely the same as 50 g of food and may not reflect amounts of food normally ingested. For example, USDA nutrient database shows that 28 g of beets contain 2.7 g of carbohydrate and 1.9 g of available carbohydrate, so 50 g of available carbohydrate from beets, the test amount eaten, is 0.72 kg or 5–6 medium beets).

Some argue that GI is similar to values in tables for nutrients. While similarities exist, there are important differences. First, the nutrients are determined by precise chemical analyses, but GI is a physiological measure that averages data from 10 human subjects, who have followed rigid diet and exercise protocols for 3 days prior to the test and ingest 50 g of available carbohydrate food after fasting. The blood glucose response is monitored for 2 h, and this is repeated on 3 separate occasions. The resulting AUCs are compared to each subject's own standard. Despite the strict protocols, variability within and among subjects is high (DeVries, 2008; Matthan et al., 2016). The application of data, collected on a single food after a strict protocol, to situations where foods

are eaten in combination in free-living conditions raises questions. Further, unlike other nutrients, the values in the tables are nutrient is used as touchstones for choosing or avoiding a food.

Since Jenkins et al. (1981) introduced GI, it has been used in over 10,000 scientific articles. The International Carbohydrate Quality Consortium, a group of aligned scientists, assessed the data and issued a consensus in 2013 stating the importance of postprandial glycemia in overall health and suggesting that GI is a valid and reproducible method of classifying carbohydrate foods. This group stated that low GI and GL diets were relevant for preventing and managing chronic diseases with the strongest agreement around use for diabetes to improve hemoglobin A1c (HbA1c) levels in moderately controlled type 1 and 2 diabetics (Augustin et al., 2015).

This consensus reflected findings of many epidemiological studies that showed low, compared to high, GI diets were associated with reduced risk impaired glucose tolerance and chronic diseases (Shahdadian et al., 2019; Ojo et al., 2018; Zafar et al., 2019). However, there are many studies that show high variability and mixed results (Afandi et al., 2021; Barclay et al., 2021; Gaesser et al., 2021; Vega-López et al., 2007, 2018). Reasons for variability include: (1) Imprecision with respect to assigning GI values data from food frequencies and similar instruments occurs because critical factors affecting GI are often unavailable. These include determinants of GI such as processes used, liquidity, grain structure, particle size, phytonutrients, dietary fiber, RS, and amylose: amylopectin ratios, which are important (Jones, 2013; Kanter et al., 2021). Even notes accompanying the most recent International Tables of Glycemic Index and Glycemic Load Values (2021) point out high GI variability in some food categories such as cereals (Atkinson et al., 2021).

Rice exemplifies the problem. There are >115 entries for boiled rice in the Sydney University Glycemic Index Database (Kaur et al., 2016; University of Sydney Database, 2021). Values range from very low GI (27) to extremely high (112). Knowledge of grain size, brand, variety, amount of water and cooking times, or amylose content helps, but values may still not reflect the GI of the food eaten.

(2) Epidemiological studies that compare high and low GI diets show significant confounding (Gaesser et al., 2021). Among studies the strongest evidence supporting GI is for use by diabetics. The quality of evidence for the findings for association with HbA1C is rated in meta-anlyses as "moderate" and for most other endpoints is rated as "weak" or of "low certainty" (Clar et al., 2017; Reynolds et al., 2019; Jayedi et al., 2020). Eaters of diets deemed high GI and low GI often have significantly different nutrient, fiber and calorie intakes and different lifestyle characteristics. This begs the question of whether the observed health impacts are due to differences in GI or confounding factors.

Further if meal or daily diet GI is used, the calculation may introduce error. These values are calculated by multiplying grams of each carbohydrate component, (expressed as a percentage of total carbohydrate in the meal or day) by the relevant GIs and summing these. Multiplying may amplify any error, and there are questions about measures from a single food and using it in meals or daily patterns. While some defend its use, others provide data suggesting the calculations overestimate GI (Afandi et al., 2021; Balance et al., 2019; DeVries, 2008; Kim et al., 2019; Matthan et al., 2016; Meng et al., 2017a,b, 2018; Wolever and Bhaskaran, 2012).

(3) Intervention studies show less variability than epidemiological ones. A meta-analysis of 27 controlled trials showed that low GI diets provided small benefits including better hemoglobin A1C and other cardiometabolic endpoints for persons with diabetes (Chiavaroli et al., 2021). Certainty was graded 'moderate' for HbA1c and several secondary outcomes.

Conclusion

Glucose, the major energy source, is supplied by foods that deliver available carbohydrates. These foods may also be important sources of unavailable carbohydrates—dietary fiber and resistant starch. To supply energy, digestive processes hydrolyze available carbohydrates to monosaccharides for absorption. Digestion begins in the mouth but takes place primarily in the small intestine. Various glucose carriers facilitate sugars movement from the villi and the gut lumen through the mucosa and into portal blood for delivery to the liver and body tissues. While sugars supply some energy, most comes from starch and related polymers with thousands glucose units. These are broken down by amylases in the upper gut. Starch exists in two forms—the rapidly digested, branched amylopectin and the slowly digested, linear amylose. α-1-4 glycosidic bonds and amylopectin has α-1-6 bonds glycosidic bonds. Starch that escapes digestion is deemed "resistant" and exists in five forms. These function physiologically with other nondigestible polymers and are fermented to SCFA in the colon.

The speed of glucose entry from carbohydrate foods into the bloodstream is dependent on the rate of digestion, which is impacted by many factors including gut motility, meal size and fluidity. Factors that specifically affect the rate of starch digestion include: the amylose: amylopectin ratio, fiber content, starch granule characteristics, food texture and macronutrient composition and other meal constituents, and how the food is processed, stored, and eaten. The GI was developed to compare the blood glucose excursions for available carbohydrate from various foods. Hundreds of studies link low GIdiets with lower risks of various diseases. Thus, a group of scientists issued a statement suggesting GI defined carbohydrate quality. However, there is much controversy surrounding this as the findings show great variability and strength of the conclusion for most endpoints associating GI with health endpoints except HbA1c for diabetics is rated 'as "weak" or of "low certainty". Sources of variability and confounding in studies comparing low and high GI diets are numerous, so the controversy continues.

Glucose, its sources, digestion, and bloodstream delivery do impact health. While GI is one measure of carbohydrate quality, the role of diets, that contain the recommended carbohydrate staples with the right mix and quantity of wholegrain, dietary fiber, fruits,

vegetables, some legumes, nuts and seeds in the recommended pattern such as DASH or MyPLATE, is important for supplying the glucose that the body needs for energy.

References

Afandi, F.A., Wijaya, C.H., Faridah, D.N., Suyatma, N.E., Jayanegara, A., 2021. Evaluation of various starchy foods: a systematic review and meta-analysis on chemical properties affecting the glycemic index values based on in vitro and in vivo experiments. Foods 10, 364.

Atkinson, F.S., Brand-Miller, J.C., Foster-Powell, K., Buyken, A.E., Goletzke, J., 2021. International tables of glycemic index and glycemic load values 2021: a systematic review. Am. J. Clin. Nutr. 13:nqab233.

Augustin, L.S., Gallus, S., Bosetti, C., et al., 2015. Glycemic index, glycemic load and glycemic response: an International Scientific Consensus Summit from the International Carbohydrate Quality Consortium (ICQC). Nutr. Metabol. Cardiovasc. Dis. 25, 795–815.

Balance, S., Knutsen, S.H., Fosvold, Ø.W., Fernandez, A.S., Monro, J., 2019. Predicting mixed-meal measured glycaemic index in healthy subjects. Eur. J. Clin. Nutr. 58, 2657–2667.

Barclay, A.W., Augustin, L.S.A., Brighenti, F., et al., 2021. Dietary glycaemic index labelling: a global perspective. Nutrients 12, 3244.

Breyton, A.E., Goux, A., Lambert-Porcheron, S., et al., 2021. Starch digestibility modulation significantly improves glycemic variability in type 2 diabetic subjects: a pilot study. Nutr. Metabol. Cardiovasc. Dis. 31, 237–246.

Chiavaroli, L., Lee, D., Ahmed, A., et al., 2021. Effect of low glycaemic index or load dietary patterns on glycaemic control and cardiometabolic risk factors in diabetes: systematic review and meta-analysis of randomised controlled trials. BMJ 374, n1651.

Clar, C., Al-Khudairy, L., Loveman, E., et al., 2017. Low glycaemic index diets for the prevention of cardiovascular disease. Cochrane Database Syst. Rev. 7, CD004467.

DeVries, J., 2008. Glycemic Index: the analytical perspective. Cereal Foods World 52, 45–49.

Gaesser, G.A., Miller Jones, J., Angadi, S.S., 2021. Perspective: does glycemic index matter for weight loss and obesity prevention? Examination of the evidence on “fast” compared with “slow” carbs. Adv. Nutr. nmab093.

Hu, J., Lin, S., Zheng, B., Cheung, P.C.K., 2018. Short-chain fatty acids in control of energy metabolism. Crit. Rev. Food Sci. Nutr. 58, 215–236.

Jayedi, A., Soltani, S., Jenkins, D.J., Sievenpiper, J., Shab-Bidar, S., 2020. Dietary glycemic index, glycemic load, and chronic disease: an umbrella review of meta-analyses of prospective cohort studies. Crit. Rev. Food Sci. Nutr. 60, 1–10.

Jenkins, D.J., Wolever, T.M.S., Taylor, R.H., et al., 1981. Glycemic index of foods: a physiological basis for carbohydrate exchange. Am. J. Clin. Nutr. 34, 14–17.

Jones, J.M., 2013. Concerns about adding glycemic index values for use in the NHANES database. J. Acad. Nutr. Diet. 113, 767–768.

Kanter, M., Angadi, S., Jones, J.M., Beals, K.A., 2021. Limitations of the glycemic index and the need for nuance when determining carbohydrate quality. Cardiovasc. Res. cvab312. https://doi.org/10.1093/cvr/cvab312.

Kaur, B., Ranawana, V., Henry, J., 2016. The glycemic index of rice and rice products: a review, and Table of GI Values. Crit. Rev. Food Sci. Nutr. 56, 215–236.

Kim, J.S., Nam, K., Chung, S.J., 2019. Effect of nutrient composition in a mixed meal on the postprandial glycemic response in healthy people: a preliminary study. Nutr. Res. Pract. 13, 126–133.

Koepsell, H., 2020. Glucose transporters in the small intestine in health and disease. Pflügers Arch. 472, 1207–1248.

Lovegrove, A., Edwards, C.H., De Noni, I., et al., 2017. Role of polysaccharides in food, digestion, and health. Crit. Rev. Food Sci. Nutr. 57, 237–253.

Matthan, N.R., Ausman, L.M., Meng, H., Tighiouart, H., Lichtenstein, A.H., 2016. Estimating the reliability of glycemic index values and potential sources of methodological and biological variability. Am. J. Clin. Nutr. 104, 1004–1013.

Meng, H., Matthan, N.R., Ausman, L.M., Lichtenstein, A.H., 2017a. Effect of macronutrients and fiber on postprandial glycemic responses and meal glycemic index and glycemic load value determinations. Am. J. Clin. Nutr. 105, 842–853.

Meng, H., Matthan, N.R., Ausman, L.M., Lichtenstein, A.H., 2017b. Effect of prior meal macronutrient composition on postprandial glycemic responses and glycemic index and glycemic load value determinations. Am. J. Clin. Nutr. 106, 1246–1256.

Meng, H., Matthan, N.R., Lichtenstein, A.H., 2018. Reply to Brighenti F et al. Am. J. Clin. Nutr. 107, 846–847.

Ojo, O., Ojo, O.O., Adebowale, F., Wang, X.H., 2018. The effect of dietary glycaemic index on glycaemia in patients with type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials. Nutrients 10, 373.

Patterson, M.A., Maiya, M., Stewart, M.L., 2020. Resistant starch content in foods commonly consumed in the United States: a narrative review. J. Acad. Nutr. Diet. 120, 230–244.

Reynolds, A., Mann, J., Cummings, J., et al., 2019. Carbohydrate quality and human health: a series of systematic reviews and meta-analyses. Lancet 393 (10170), 434–445.

Shahdadian, F., Saneei, P., Milajerdi, A., Esmaillzadeh, A., 2019. Dietary glycemic index, glycemic load, and risk of mortality from all causes and cardiovascular diseases: a systematic review and dose-response meta-analysis of prospective cohort studies. Am. J. Clin. Nutr. 110, 921–937.

University of Sydney, 2021. The International Glycemic Index (GI) Database and Website. https://researchdata.edu.au/international-glycemic-index-gi-database/11115.

Vega-López, S., Ausman, L.M., Griffith, J.L., Lichtenstein, A.H., 2007. Interindividual variability and intra-individual reproducibility of glycemic index values for commercial white bread. Diabetes Care 30, 1412–1417.

Vega-López, S., Venn, B.J., Slavin, J.L., 2018. Relevance of the glycemic index and glycemic load for body weight, diabetes, and cardiovascular disease. Nutrients 10, 1361.

Wolever, T.M., Bhaskaran, K., 2012. Use of glycemic index to estimate mixed-meal glycemic response. Am. J. Clin. Nutr. 95, 256–257.

Zafar, M.I., Mills, K.E., Zheng, J., et al., 2019. Low-glycemic index diets as an intervention for diabetes: a systematic review and meta-analysis. Am. J. Clin. Nutr. 110, 891–902.

List of Relevant Websites

FoodData Central, https://fdc.nal.usda.gov/. (Accessed 10 October 2021).

Glycemic index

J. Brand-Miller[a], **G. Frost**[b], **and A. Dornhorst**[b], [a] University of Sydney, Sydney, NSW, Australia; and [b] Imperial College London, London, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of G. Frost, A. Dornhorst, Glycemic Index, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 393–398, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00136-7.

Background and definition 422
Type of dietary carbohydrate and the glycemic index 424
Chain length and composition 424
Amylose and amylopectin 424
Relationship of insoluble and soluble non-starch polysaccharides (NSPs) (fiber) to glycemic index 424
Cell structure, food preparation, and processing 424
Concerns related to the glycemic index 425
Reproducibility 425
Within-subject variation 425
Between-individual variation 425
Reproducibility of the glycemic index 425
Problems arising from different methodologies used to calculate the glycemic index 425
Mixed meals and other nutrients 425
The second meal effect 426
Clinical significance of postprandial hyperglycemia 426
Benefits of low glycemic index carbohydrates on diabetes control 426
Benefits of low glycemic index carbohydrates on CVD risk factors 427
Glycemic index and the prevention of type 2 diabetes 427
Obesity and glycemic index 427
Pregnancy and glycemic index 427
Proposed mechanism by which dietary carbohydrates/glycemic index influence insulin resistance 427
Conclusion 428
Conflict of interest statement 428
References 428

Key points

- The glycemic index of foods - a marker of postprandial physiological effect.
- Fibre content does not predict the glycemic index.
- The glycemic impact of mixed meals can be predicted from the glycemic index of component carbohydyrate foods.
- Diets based on low glycemic index foods improve glucose metabolism in people with diabetes and cardiovascular disease.

In the past 10 years, meta-analyses of epidemiological and experimental studies have linked the glycemic index to postprandial glucose metabolism, insulin resistance, diabetes and cardiovascular risk factors. A joint committee of the WHO and the Food and Agriculture Organization (FAO) have recommended that the physiological effects of dietary carbohydrates be classified according to their glycemic index (FAO/WHO Joint Expert Consultation, 1998). This review examines the historical and scientific background of the glycemic index.

Background and definition

In 1939, Conn and Newburgh noted how different carbohydrate-containing foods could have the same macronutrient composition, but produce different glycemic responses. These observations were extended by Jenkins and colleagues in 1981 (Jenkins et al., 1981), resulting in systematic classification of carbohydrate foods according to postprandial glucose levels elicited by standardized carbohydrate portions. In the beginning, the "glycemic index" (GI) was introduced as a means of comparing carbohydrate "exchanges" for people with diabetes, but later broadened to have a potential role in prevention of type 2 diabetes and cardiovascular disease in the general population.

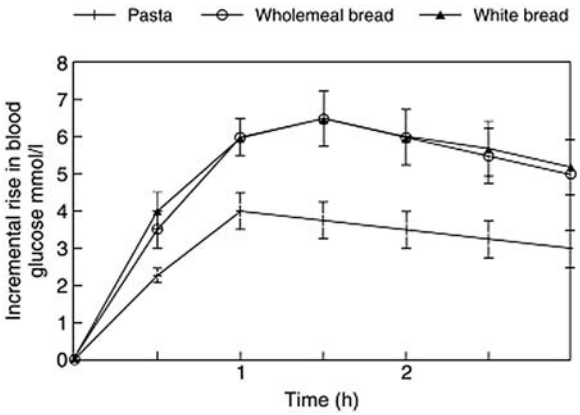


Fig. 1 Mean blood glucose increment after equi-available carbohydrate meals. Data with permission from Jenkins, D.J., Wolever, T.M., Jenkins, A.L., 1988. Starchy foods and glycemic index. Diabetes Care 11: 149–159.

Glycemic indexes of thousands of foods have been published in international nutritional tables, the most recent in 2020 (Atkinson et al., 2021). Standardized methodology for glycemic index testing was published by the International Standards Organization in 2010. Specifically, the glycemic index of a food is calculated as the average postprandial glucose response after feeding a standardized available carbohydrate portion of the food (**Fig. 1**), relative to that of a reference food. (**Table 1**). The reference food (by definition, GI = 100) is usually a 50 g glucose load in solution, but a 25 g carbohydrate portion may be used for foods with lower carbohydrate content. The 2 h incremental area under the curve is summed to reflect the period of maximum fluctuation. Low GI carbohydrates have lower areas under the glucose curve, whereas high GI foods have higher areas.

Although the insulin response is not used to define GI, it can be considered a surrogate marker of the insulin demand in response to different carbohydrates. Indeed, the insulin response in individuals without diabetes to a wide range of foods (GI between 32 and 100) is highly correlated. Like glycemia, the pattern of insulin secretion caused by different carbohydrates reflects their rate of digestion and intestinal transit times. The exception to this rule is dairy products, which have a higher insulin response than predicted by their glycemic index.

It has been argued that it is the insulin response to foods and not the glycemic response that is important in the pathogenesis of insulin resistance and related metabolic disturbances and disease risk. However, high glucose levels, per se, increase the glycation of proteins and inflammatory markers such as nuclear factor-kappa B. With the advent of continuous glucose monitoring, postprandial glycemia has been linked to many cardiovascular risk markers, including body mass index (BMI), blood lipids and glycated hemoglobin (Zeevi et al., 2015).

Dietary carbohydrates stimulate insulin secretion both directly, by stimulating the pancreatic β cell, and indirectly, through their effect on other hormones such as GIP and GLP-1. It has been postulated that the stimulation of high GIP secretion by K cells in the upper small intestine promotes lipogenesis, fatty liver, insulin resistance, and postprandial inflammation, and reduces fat oxidation in skeletal muscle (Pfeiffer and Keyhani-Nejad, 2018).

Table 1 The glycemic index model.

Incremental area under blood glucose response curve for the test food containing 50 corresponding areas after equi-carbohydrate portion of glucose × 100
Calculation of the GI of a mixed meal containing three separate carbohydrate-containing foods
GI/mixed meal = $(GI_1)(PCF_1)+(GI_2)(PCF_2)+(GI_3)(PCF_3)$
Where
The three carbohydrate-containing foods are 1, 2, and 3
The GI for each carbohydrate-containing food is GI_1, GI_2, and GI_3
The carbohydrate content is C1, C2, and C3 (g)
The total meal carbohydrate (TMC) is [C1+C2+C3] g
The proportion of carbohydrate from each food (PCF) is PCF1=C1/TMCg, PCF2=C2/TMCg, and PCF3=C3/TMCg

Type of dietary carbohydrate and the glycemic index

The GI of a carbohydrate food is influenced by the rate of digestion and intestinal absorption which, in turn, is influenced by the presence of other nutrients, the tertiary structure, type of starch, and susceptibility to enzymatic digestion. A recent meta-analysis showed significant reduction in postprandial glucose and insulin levels caused by starch with high amylose content, less gelatinized starch, retrograded starch and intact and large grain particle size (Cai et al., 2021).

Chain length and composition

In the past, carbohydrates were classified as simple or complex in chemical structure, but this has no relation to their GI. Complex carbohydrates are long, polymeric chains of repeating monosaccharide units. Starches comprise repeating glucose units. Surprisingly, the GIs of different starches are determined by their susceptibility to enzymatic digestion, and not chain length. The starch in white bread and pasta have similar chain length, but bread has a higher GI, due to its tertiary structure and solubility, which creates faster digestion by increasing the access of salivary and pancreatic amylases. Short-chain carbohydrates, such as glucose and sucrose are rapidly absorbed when consumed in solution, but when part of a solid mixture such as breakfast cereals, cakes and biscuits, their digestion and absorption is often slowed. Fructose, one of the sugars in fruit and sugar sweetened beverages, is converted to glucose in the intestinal wall and preferentially oxidized in the liver (Jang et al., 2018). The GI of fructose is, therefore, low. When a mixture of sugars is consumed (e.g., glucose, fructose and sucrose in fruit), the GI is lowered proportionally. The disaccharides sucrose and lactose consist of 50% glucose and 50% fructose or galactose, respectively, and both have a lower GI than maltose, the disaccharide formed from two molecules of glucose. Maltodextrins are short chain glucose polymers that have a high GI.

Amylose and amylopectin

The starches in cereal grains, potatoes, and other plants are composed of repeating glucose units arranged in straight (amylose) and branched-chained (amylopectin) polysaccharides. The rates of digestion and absorption and, hence, the GI of these starches is influenced by the ratio of amylose to amylopectin. The more compact structure of amylose than amylopectin results in less hydration during cooking and slower amylase digestion. Amylose-enriched starches therefore have lower GI than those enriched in amylopectin.

Relationship of insoluble and soluble non-starch polysaccharides (NSPs) (fiber) to glycemic index

Some complex carbohydrates are not digestible by human enzymes and are commonly known as dietary fibers. They can be divided into soluble and insoluble forms. Clinical studies have shown that foods and supplements rich in viscous forms of soluble fiber, such as guar gum, pectin, and psyllium, are able to lower postprandial blood glucose and insulin levels to a meal containing rapidly digestible carbohydrate. Guar gum, a β-galactomannan from the Indian locust bean, is the most viscous and also reduces postprandial lipemia. Soluble fibers, such as those in pulse vegetables, whole fruits, oats, and barley, form gelatinous gels within the stomach that delay gastric emptying and enzymatic digestion, the latter by forming a physical barrier around the carbohydrate. Viscous fiber also increases the thickness of the unstirred water layer in the small intestine, thereby slowing the rate of absorption.

On the other hand, insoluble dietary fibers such as cellulose found in wheat and rice bran, have no viscosity and usually no effect on the GI. White bread and whole-meal bread, for example, both have high glycemic indices. High fiber diets and whole-grains are, therefore, not synonymous with low glycemic foods. The exception is when the bran is largely intact around the kernel forming a physical barrier to hydration of starch during cooking. Large particles of cracked wheat are digested and absorption slowly, giving products such as All-Bran a low GI. The lack of effect of insoluble fiber on glucose and insulin should not detract from important effects on laxation, bowel function and bowel pathology.

Cell structure, food preparation, and processing

Cooking and food preparation can modify the GI. Highly gelatinized, processed convenience foods tend to have high GI. When cooking and processing disrupt the cell wall, starch granules absorb water (gelatinize), increasing the rate of amylase digestion and, therefore, the GI. The starch in bread, scones and breakfast cereals is usually fully gelatinized, while the starch in pasta, cakes and biscuits is less gelatinized, slowing digestion. Cooked pulse vegetables have low GI because their cell walls are resistant to water entry, even after canning. The largely intact cereal grains of pumpernickel rye bread, granary bread, and bulgur wheat all have low GI. However, when wheat grains are finely milled to make soft whole-meal bread, the bran is no longer a barrier to hydration and the GI rises. Cooling can lead to the formation of some starch that is resistant to human enzymes, notably in potatoes. However, the majority of the starch (>90%) is still highly gelatinized, and a cold potato salad has a high GI.

Concerns related to the glycemic index

Over the years, scientists and health professionals have expressed concerns about the GI. These include apparent lack of reproducibility within and between individuals, the practical application to mixed meals, and inconsistent findings from clinical studies. In the intervening period, these issues have been addressed by systematic reviews and meta-analyses that have resolved much of the early controversy (Augustin et al., 2015). However, the difference between GI, glycemic load (GL) and glycemic response is often misunderstood.

Reproducibility

Within-subject variation

The variability of the glycemic response for a given food for any individual is similar to that seen for the oral glucose tolerance test. In one study, a 25% coefficient of variation (CV) within individuals was seen when 11 healthy subjects had their glycemic response to different carbohydrates tested on eight separate occasions. In another study, the CV of the glycemic response in 22 healthy subjects given 50 g of white bread was 22%. This day-to-day variability is the rationale for testing the reference food at least twice, and preferably three times, in each subject. The average area under the curve is then used as the denominator in calculating the GI of a test food.

Between-individual variation

The variability of the glycemic response between individuals is larger than that within individuals. In other words, glucose tolerance varies. In a study that included 11 individuals without diabetes, 10 non-insulin-treated subjects with type 2 diabetes, 12 insulin-treated subjects with type 2 diabetes, and 14 subjects with type 1 diabetes; the CV between individuals within each group was 26%, 34%, 23%, and 34%, respectively. Thus, it can be seen that simple comparison of the absolute glycemic responses incorporates two sources of variability—both person-to-person and day-to-day variation. In GI testing of foods, this variability is managed by using subjects as their own control, i.e., expressing the glycemic response to any given food as a percentage of that individual's average glycemic response to a reference food, usually 50 g of glucose. By expressing the glycemic response of a test food against an equal amount of a standard carbohydrate, variations that occur with age, sex, BMI, sleep duration and ethnicity, as well as medical conditions such as diabetes, are minimized. By using the ratio and calculating the average of at least 10 subjects, the standard deviation around the mean is reduced. What remains reflects variation that is unavoidable and is also seen in diagnostic glucose tolerance testing for diabetes and carbohydrate counting.

Reproducibility of the glycemic index

The GI of a food is a characteristic of the specific food and not a single individual. Nonetheless, it is often assumed that the GI of certain foods varies between individuals. For example, a single study reported that the GI of lentils ranged between 23 and 70 for different subjects. However, this large variability reflects day-to-day variation in glucose tolerance within individuals, rather than between individuals. When the same food is tested again and again, the mean per individual moves closer to the mean of the group, and the standard error narrows. This statistical phenomenon is known as regression to the mean. When using standardized GI testing, the standard error around the mean is approximately 10% of the mean. Even in groups of individuals with different AMY1 gene copy number (and therefore higher or lower salivary amylase activity), the GI of foods shows the same rank order. A recent inter-laboratory study reported no significant differences in mean GI values for 6 identical foods determined by the ISO method in 3 different laboratories. Although the standard error around the mean varied between laboratories, the ISO method was sufficiently precise to distinguish a low GI food from a high GI food with 97–99% probability.

Problems arising from different methodologies used to calculate the glycemic index

Before 1998, different groups used different techniques to calculate the area under the glucose curve and to assess the postprandial glycemic response. Two reference foods, white bread and glucose solution, were also in common use, causing confusion and claims of poor reproducibility. This issue was resolved with the publication of the ISO Standard for GI methodology, that specified that when reference foods other than glucose were used, the final result should be converted to the glucose = 100 scale (International Standards Organisation, 2010). In the most recent edition of the International Tables of Glycemic Index and Glycemic load, all values are reported on the glucose scale (Atkinson et al., 2021). The GL of a food (or meal or diet) is defined as the product of its GI and the amount of carbohydrate in a serving. In the case of meals and whole diets, GL should be expressed per 1000 kJ, or otherwise adjusted for overall energy intake.

Mixed meals and other nutrients

Carbohydrate foods are frequently taken as part of a mixed meal, and the addition of fat and protein to a high carbohydrate meal lowers the glycemic response by slowing gastric emptying and stimulating more insulin secretion. However, the relative response of

one carbohydrate food to another remains, such that lentils will always have a lower response than white bread, when part of a mixed meal.

The GI of a mixed meal can be calculated if the proportion of each of the carbohydrate-containing foods and their individual GI values are known. For example, when bread and beans are mixed in equal portions, the resulting glycemic response is midway between that of bread alone and that of beans alone. A formula for calculating the GI of mixed meals has been derived by Wolever and Jenkins (Table 1). For accuracy, this method requires all individual carbohydrate components of a mixed meal to be pre-tested. However, in practice, the GI of a mixed meal should not be tested, but rather calculated. When the average of all meals in a food record is determined in this way, it is possible to calculate the overall dietary GI and glycemic load of the person's diet. When expressed per 1000 kJ or per 2000 Kcal, it allows researchers to predict the relative glycemic impact or insulin demand of one diet versus another, but not the absolute glycemic response.

The second meal effect

Dietary carbohydrates can influence the glycemic response of a second meal consumed during the postprandial period. The blood glucose response to a lunchtime meal is lower when taken after a low GI breakfast than after a high GI breakfast. Similarly, the glycemic response of a second meal taken during the postprandial period after lunch or dinner is influenced by the GI of the preceding meal (Wolever et al., 1988).

Wolever attributed the differences in the glycemic response to a second meal during the postprandial period to differences in intermediary metabolism and insulin action associated with rapidly and slowly absorbed carbohydrates. Rapidly absorbed carbohydrates produce large increases in insulin, which results in blood glucose decreasing sufficiently quickly to stimulate several counter-regulatory hormones that inhibit insulin action and glucose disposal. High GI meals are associated with significantly higher serum concentrations of glucagon, catecholamines, growth hormone, and non-esterified fatty acid (NEFA) levels in the post-absorptive period (3–5 h after consumption). The addition of guar to a meal, which slows glucose absorption and lowers the glycemic response, reduces postprandial NEFA and β-hydroxybutyrate and improves insulin action. In this way, chronic consumption of a low GI diet improves glucose and lipid metabolism.

Clinical significance of postprandial hyperglycemia

Postprandial hyperglycemia in populations without diabetes is linked to insulin resistance, diabetes and cardiovascular disease (CVD). A glucose tolerance test containing 75 g glucose in solution is used to detect a person's ability to reestablish euglycemia after a glucose load. If glucose concentration remains above $\geq$11.1 mmol L^{-1} at 120 min, this is diagnostic of diabetes; concentrations between 7.8 and 11.0 reveal a predisposition to diabetes, or pre-diabetes. The 120 min glucose concentration is also a predictor of usual response after a high GI meal. In the DECODE study, the combined 10-year mortality data from 14 studies showed that the highest versus lowest quintile for the 2 h post-load blood glucose was associated with a 2- to 3-fold increased risk of CVD mortality. Fasting glucose values were less predictive of CVD than post-load values. During a 7-year period, elderly women with isolated postprandial hyperglycemia and a 2 h value more than 11.1 mmol L^{-1}, and fasting value less than 7.0 mmol L^{-1} on a 75 g oral glucose tolerance test, had an approximately threefold increased risk of heart disease compared with women whose 2 h values were less than 11.1 mmol L^{-1}.

In healthy populations, average postprandial blood glucose concentration during the preceding months is reflected in glycosylated hemoglobin A_{1c} (HbA_{1c}). Higher HbA_{1c} is now used to diagnose both diabetes and pre-diabetes. Importantly, HbA_{1c} concentration, even within the normal range, has been associated with the development over time of type 2 diabetes and cardiovascular disease (Selvin et al., 2010).

In established diabetes, postprandial glycemia appears to have a stronger relationship with microvascular and macrovascular disease than fasting blood glucose. Similarly, in gestational diabetes, adverse pregnancy outcome is more closely related to postprandial glycemia than fasting and premeal glycemic values.

Benefits of low glycemic index carbohydrates on diabetes control

Diabetes control is the area in which there is most evidence of clinical efficacy. Several independent systematic reviews of global evidence have demonstrated the efficacy of low GI diets on glycemic control in management of both type 1 and type 2 diabetes. In the most recent meta-analysis (Zafar et al., 2019), 54 randomized controlled trials in adults and children with impaired glucose tolerance, type or type 2 diabetes, showed that low GI diets were useful for improving glycemic control and lipid metabolism, as judged by HbA_{1c}, fasting glucose, BMI, total cholesterol, high density lipoprotein cholesterol, triglycerides and insulin requirements. The greatest reduction in fasting blood glucose was seen in those with the longest duration. The comparison diets included carbohydrate exchange, high-fat and low-fat diets. However, many studies failed to achieve meaningful difference in GI between a low GI arm and the control arm, and only individuals with obesity showed significant reduction in body weight. Good glycemic control and favorable lipid and fibrinolytic profiles have also been reported in individuals with either type 1 or 2 diabetes who habitually consume low GI dietary carbohydrates. It remains to be shown whether these diets bestow long-term benefits on micro- or macrovascular complications.

Benefits of low glycemic index carbohydrates on CVD risk factors

High GI foods induce postprandial hyperinsulinemia, which is a powerful predictor for metabolic risk factors and CVD in epidemiological studies. However, it has been difficult to separate the effects of high fiber and high carbohydrate intake from those of a low GI diet. The predominant concern in heart disease prevention has been saturated fat, leading to widespread adoption of low fat, high carbohydrate diets. However, recent cohort studies suggest an even greater increase in coronary heart disease progression when highly refined and rapidly absorbed carbohydrates have replaced saturated fat. In the most recent meta-analysis, 10 studies with a mean follow-up of 11 years yielded a relative risk of 1.24 per 10 point increase in GI (Livesey and Livesey, 2019). Across the global range in GI (47–82 units), the relative risk was 2.7. Similarly, 11 studies yielded a relative risk of 1.66 per 65 g increase in GL of the diet. Across the global range for GL (55–290 g day^{-1}, the relative risk was 5.5. Based on the Bradford-Hill criteria, the findings are strong and imply causality. Some evidence suggests that the associations are more evident in females and in individuals with obesity.

Glycemic index and the prevention of type 2 diabetes

Changes in diet and physical activity levels, both alone and in combination, reduce the progression of impaired glucose tolerance to type 2 diabetes. In a recent meta-analysis of 10 studies, the relative risk of type 2 diabetes was almost 1.3 fold higher per 10-point increase in GI (Livesey et al., 2019). The relative risk was similar per 80 g day^{-1} increase in dietary GL. Two large US prospective population studies demonstrated a doubling of the relative risk of developing type 2 diabetes for both men and women when the habitual diet was characterized by high GI and high fat content. A similar protective effect against diabetes has been reported in populations consuming high-fiber foods and high quantities of fruit, and one would predict that these diets would also have a low GI.

Obesity and glycemic index

Obesity contributes to the pathogenesis and morbidity of type 2 diabetes and cardiovascular disease and is associated with changes in carbohydrate and fat metabolism central to the development of insulin resistance. Although low GI diets enhance insulin sensitivity and improve metabolic cardiovascular risk factors, they are unlikely to reduce weight unless part of a lower energy diet. In the DioGenes Study (Larsen et al., 2010), a 5-arm, 26-week multi-center study in overweight adults from 8 European countries, only the low protein-high GI diet was associated with significant weight regain after weight loss. Both high protein and low GI diets were independently associated with reduced weight regain, but the difference was no longer present at 12 months. In the PREVIEW diabetes prevention study, there was no difference in diabetes incidence or weight outcomes between the high protein-low GI diet and the control moderate protein-moderate GI diet. However, a secondary longitudinal analysis of the PREVIEW data found that the highest tertiles of GI and GL were associated with greater weight regain and increase in glycated hemoglobin (HbA1c) than the lowest tertiles (Zhu et al., 2021). Hunger and satiety ratings were correlated with changes body weight. Taken together, the findings in PREVIEW imply that greater reduction in GI is needed to impact clinical outcomes than was achieved in the PREVIEW randomized controlled trial. Evidence from animal studies, in which diet can be strictly controlled, also demonstrates that the GI of the diet affects both weight and body fat outcomes (Campbell et al., 2017).

Pregnancy and glycemic index

Throughout pregnancy, insulin resistance increases as pregnancy progresses, even in well-nourished women. As a result, many women develop hyperglycemia and gestational diabetes. Untreated hyperglycemia has serious adverse pregnancy outcomes, including macrosomia and offspring obesity. Randomized controlled trials in the West show that women with gestational diabetes assigned to a low GI diet had better pregnancy outcomes, including less need for insulin, than those consuming high GI diets (Moses et al., 2006). A meta-analysis of 5 studies showed no significant differences in risk of cesarian section, large-for-gestational age infants or small-for-gestational age infants. When the proportion of dietary carbohydrate increases above 50% in women with gestational diabetes, low GI carbohydrates improve glucose tolerance (Walsh et al., 2012).

Proposed mechanism by which dietary carbohydrates/glycemic index influence insulin resistance

Adipocyte metabolism is central to the pathogenesis of insulin resistance, and dietary carbohydrates influence adipocyte function. The previous simplistic view that insulin resistance resulted from the downregulation of insulin receptors in response to hyperinsulinemia is being replaced by the hypothesis that high circulating NEFA concentrations both impair insulin action and reduce pancreatic β cell secretion. It is plausible that low GI carbohydrates reduce insulin resistance by their ability to reduce adipocyte NEFA release. There is evidence of a loss of suppression of hormone-sensitive lipase (HSL), an enzyme that breaks down triglyceride to free fatty acids and glycerol, to small physiological increases in insulin and, to a lesser extent, insulin insensitivity of lipoprotein

lipase. HSL is normally sensitive to small increases in insulin, and is totally suppressed at much lower concentrations than those required for glucose uptake. In insulin-resistant subjects, HSL is less sensitive and adipocyte NEFA release is increased. A relationship between increased adipocyte NEFA release and insulin resistance has been shown in subjects with CHD. The metabolic consequences of increased circulating NEFA are multiple, including adverse lipoprotein and coagulation changes and a lipotoxic effect on the β cell. Accumulation of triglycerides within the β cell also impairs insulin secretion.

Many of the metabolic benefits associated with low GI carbohydrates can be attributed to their ability to reduce adipocyte NEFA release. Low GI foods have been consistently shown to reduce insulin resistance, and animal studies have shown that improvements in fat and muscle insulin sensitivity are accompanied by decreases in fatty acid synthetase activity, adipocyte size, and lipid storage.

Low GI diets attenuate the insulin response for approximately 4 h postprandially. High postprandial insulin in the later stage is insufficient to affect glucose transport, but does suppress the insulin-sensitive enzyme, HSL and, thus, ensures prolonged suppression of postprandial NEFA output. The ability of low glycemic carbohydrates to do this is in stark contrast with high glycemic diets, which can cause an elevation of NEFA release postprandially by stimulating counter-regulatory hormones. Low glycemic meals taken in the evening can effectively suppress circulating NEFA concentrations and hepatic glucose output throughout the night. These metabolic effects are predicted to promote insulin sensitivity.

Our own work has shown that insulin-resistant adults with a history of, or who are at risk of, CHD improve adipocyte insulin sensitivity after consuming a low GI diet for 3 weeks. and that their circulating NEFA decline. These human studies complement animal work showing that low GI diets improve insulin sensitivity by modulating adipocyte metabolism.

Conclusion

A low GI diet improves postprandial metabolism, reduces inflammatory markers and is associated with reduced risk of developing type 2 diabetes and cardiovascular disease. Dietary carbohydrates are absorbed and metabolized at different rates and, therefore, influence postprandial glucose, insulin, and NEFA concentrations differently. In Western society, the proportion of the day that we spend in the postprandial state is increasing as the tendency to snack throughout the day replaces the traditional three meals each day. The known detrimental consequences of high glycemic foods and snacks on postprandial metabolism should encourage us to advocate for low glycemic diets to counter the current epidemic of insulin resistance-related and inflammatory diseases. The relevance of the GI to preventable diseases of the Western world argues strongly for its greater acceptance in current nutritional guidelines.

Conflict of interest statement

JBM is the co-author of books about the GI of foods and oversees a GI testing service at the University of Sydney. She is President of a global, not-for-profit GI-based food endorsement program, a member of the International Carbohydrate Quality Consortium and a scientific advisor to the Novo Foundation and Zoe Ltd.

References

Atkinson, F., Brand-Miller, J., Foster-Powell, K., Buyken, A., Goletzke, J., 2021. International tables of glycemic index and glycemic load values 2020: a systematic review. Am. J. Clin. Nutr. 114 (5), 1625–1632.

Augustin, L.S.A., Kendall, C.W.C., Jenkins, D.J.A., Willett, W.C., Astrup, A., Barclay, A.W., Björck, I., Brand-Miller, J.C., Brighenti, F., Buyken, A.E., Ceriello, A., La Vecchia, C., Livesey, G., Liu, S., Riccardi, G., Rizkalla, S.W., Sievenpiper, J.L., Trichopoulou, A., Wolever, T.M.S., Baer-Sinnott, S., Poli, A., 2015. Glycemic index, glycemic load and glycemic response: an International Scientific Consensus Summit from the International Carbohydrate Quality Consortium (ICQC). Nutr. Metabol. Cardiovasc. Dis. 25 (9), 795–815.

Cai, M., Dou, B., Pugh, J.E., Lett, A.M., Frost, G.S., 2021. The impact of starchy food structure on postprandial glycemic response and appetite: a systematic review with meta-analysis of randomized crossover trials. Am. J. Clin. Nutr. 114 (2), 472–487.

Campbell, J.G., Senior, M.A., Bell-Anderson, S.K., 2017. Metabolic effects of high glycaemic index diets: a systematic review and meta-analysis of feeding studies in mice and rats. Nutrients 9 (7), 646.

FAO/WHO Joint Expert Consultation, 1998. Carbohydrates in Human Nutrition. FAO Food and Nutrition Paper 66. Food and Agriculture Organisation, Food and Nutrition, Geneva.

International Standards Organisation, 2010. Food Products—Determination of the Glycaemic Index (GI) and Recommendation for Food Classification. ISO 26642-2010.

Jang, C., Hui, S., Lu, W., Cowan, A.J., Morscher, R.J., Lee, G., Liu, W., Tesz, G.J., Birnbaum, M.J., Rabinowitz, J.D., 2018. The small intestine converts dietary fructose into glucose and organic acids. Cell Metabol. 27 (2), 351–361.e3.

Jenkins, D.J., Wolever, T.M., Taylor, R.H., Barker, H., Fielden, H., Baldwin, J.M., Bowling, A.C., Newman, H.C., Jenkins, A.L., Goff, D.V., 1981. Glycemic index of foods: a physiological basis for carbohydrate exchange. Am. J. Clin. Nutr. 34 (3), 362–366.

Larsen, T.M., Dalskov, S.-M., van Baak, M., Jebb, S.A., Papadaki, A., Pfeiffer, A.F.H., Martinez, J.A., Handjieva-Darlenska, T., Kunesova, M., Pihlsgard, M., Stender, S., Holst, C., Saris, W.H.M., Astrup, A., 2010. Diets with high or low protein content and glycemic index for weight-loss maintenance. N. Engl. J. Med. 363 (22), 2102–2113.

Livesey, G., Livesey, H., 2019. Coronary heart disease and dietary carbohydrate, glycemic index, and glycemic load: dose-response meta-analyses of prospective cohort studies. Mayo Clin. Proc. 3 (1), 52–69.

Livesey, G., Taylor, R., Livesey, F.H., Buyken, E.A., Jenkins, J.A.D., Augustin, S.A.L., Sievenpiper, L.J., Barclay, W.A., Liu, S., Wolever, M.S.T., Willett, C.W., Brighenti, F., Salas-Salvadó, J., Björck, I., Rizkalla, W.S., Riccardi, G., Vecchia, L.C., Ceriello, A., Trichopoulou, A., Poli, A., Astrup, A., Kendall, W.C.C., Ha, M.-A., Baer-Sinnott, S., Brand-Miller, C.J., 2019. Dietary glycemic index and load and the risk of type 2 diabetes: a systematic review and updated meta-analyses of prospective cohort studies. Nutrients 11 (6).

Moses, R., Luebcke, M., Davis, W., Coleman, K., Tapsell, L., Petocz, P., Brand-Miller, J., 2006. Effect of a low-glycemic-index diet during pregnancy on obstetric outcomes. Am. J. Clin. Nutr. 84, 807–812.

Pfeiffer, A.F.H., Keyhani-Nejad, F., 2018. High glycemic index metabolic damage—a pivotal role of GIP and GLP-1. Trends Endocrinol. Metabol. 29 (5), 289–299.

Selvin, E., Steffes, M., Zhu, H., 2010. Glycated hemoglobin, diabetes and cardiovascular disease risk in nondiabetic adults. N. Engl. J. Med. 362, 800–811.

Walsh, J.M., McGowan, C.A., Mahony, R., Foley, M.E., McAuliffe, F.M., 2012. Low glycaemic index diet in pregnancy to prevent macrosomia (ROLO study): randomised control trial. BMJ 345.

Wolever, T., Jenkins, D., Ocana, A., Rao, V., Collier, G., 1988. Second-meal effect: low-glycemic-index foods eaten at dinner improve subsequent breakfast glycemic response. Am. J. Clin. Nutr. 48 (4), 1041–1047.

Zafar, M.I., Mills, K.E., Zheng, J., Regmi, A., Hu, S.Q., Gou, L., Chen, L.-L., 2019. Low-glycemic index diets as an intervention for diabetes: a systematic review and meta-analysis. Am. J. Clin. Nutr. 110 (4), 891–902.

Zeevi, D., Korem, T., Zmora, N., Israeli, D., Rothschild, D., Weinberger, A., Ben-Yacov, O., Lador, D., Avnit-Sagi, T., Lotan-Pompan, M., Suez, J., Mahdi, J.A., Matot, E., Malka, G., Kosower, N., Rein, M., Zilberman-Schapira, G., Dohnalová, L., Pevsner-Fischer, M., Bikovsky, R., Halpern, Z., Elinav, E., Segal, E., 2015. Personalized nutrition by prediction of glycemic responses. Cell 163 (5), 1079–1094.

Zhu, R., Larsen, T.M., Fogelholm, M., Poppitt, S.D., Vestentoft, P.S., Silvestre, M.P., Jalo, E., Navas-Carretero, S., Huttunen-Lenz, M., Taylor, M.A., et al., 2021. Dose-dependent associations of dietary glycemic index, glycemic load, and fiber with 3-year weight loss maintenance and glycemic status in a high-risk population: a secondary analysis of the diabetes prevention study preview. Diabetes Care 44 (7), 1672–1681.

Legumes in human health and nutrition

Sayra N. Serrano-Sandoval[a], Daniela Guardado-Felix[b], and Janet A. Gutiérrez-Uribe[a,c], [a] Tecnologico de Monterrey, The Institute for Obesity Research, Monterrey, Nuevo León, Mexico; [b] Programa Regional de Posgrado en Biotecnología, Facultad de Ciencias Químico Biológicas, Universidad Autónoma de Sinaloa, FCQB-UAS, Culiacán, Sinaloa, Mexico; and [c] Tecnologico de Monterrey, Escuela de Ingeniería y Ciencias, Campus Puebla, Reserva Territorial Atlixcáyotl, Puebla, Pue, Mexico

© 2023 Elsevier Ltd. All rights reserved.

This is an update of L.H. Allen, Legumes, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 74–79, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00170-7.

Introduction 430
Classification and common food uses 431
Dry beans 431
Lupines 431
Bambara beans 431
Lentils 432
Dry peas 432
Chickpeas 432
Dry pigeon peas 432
Vetches 432
Dried cowpeas 432
Winged beans and sword beans 432
Nutritional value 432
Global trends of production and economic importance 433
Legume proteins/flours in emerging markets 435
Legume processing and antinutritional factors 436
Conclusions 436
References 436
Further reading 436

Key points

- After soybean and peanuts, pulses are the most consumed legumes
- Pulses are the main protein source in many countries
- Emerging technologies have been used besides thermal treatment to inactivate trypsin inhibitor activity of legumes
- Pulses are a great alternative for plant-based protein diets

Introduction

Fabaceae or *Leguminosae* family represents the third largest group of plant life in the world. Soybean and peanuts are the legumes with highest economic relevance followed by dry bean, chickpeas and peas. Pulses are a subcategory of legumes that include the seeds that are consumed as foods. They represent an important source of protein for many countries in Africa and Latin America. Among other beneficial effects on health, they are useful for moderating blood sugar levels after meals and improving insulin sensitivity (Singh, 2017). They also play a positive role in weight reduction since their consumption induces satiety. The resistant starch and fiber act as prebiotics for the probiotic or beneficial bacteria. The content of micronutrients is also an important characteristic of legumes, and their consumption can reduce the risk of anemia. Many traditional dishes in different regions of the planet include combinations of legumes and cereals, complementing the amino acids profile. The Food and Agriculture Organization of the United Nations (FAO) promoted the nutritional and environmental benefits of the diverse crops included in this category and 2016 was proclaimed as the International Year of Pulses (IYP). Since 2019, the World Pulses Day has been celebrated on February 10th. Global market trends related to more sustainable and healthy diets have increased the consumption of pulses as ingredients for foods beyond the traditional dishes that included them boiled, stewed, germinated or fermented. Trypsin inhibitor activity has been successfully reduced with thermal treatments, but emerging technologies have been developed to improve their functionality and retain phytochemicals and nutrients with diverse beneficial effects on health.

Classification and common food uses

Legumes belonging to the *Fabaceae* or *Leguminosae* family represent the third largest group of plant life in the world. Legumes that produce edible seeds for human and animal nutrition are named pulses. According to FAO (2016), the major groups are dry beans, lupines, bambara beans, broad beans, lentils, dry peas, chickpeas, dry pigeon peas, vetches, dried cowpeas, winged beans, and sword beans (Table 1).

Dry beans

Phaseolus is the genera of the dry bean pulses of America and *Vigna* is found in some regions of Asia. These pulses include common beans such pinto, black, or in various colors, all belonging to the *Phaseolus vulgaris* that is one of the most widespread crops. *Vigna* includes some dry beans such as adzuki, mung, and urd beans. Dry beans are considered one of the most important edible legumes worldwide and can be found almost in every country on earth. In fact, it is a staple food in the Middle East, South America, India, and the Mediterranean.

Lupines

Lupines include *Lupinus albus*, a native edible legume from the Mediterranean and *Lupinus mutabilis* originating in South America. They grow in Australia, Europe, Russia, and South America. Lupines have been consumed as a popular snack usually in brine or pickled, and they are also used as an additive to enrich cereal flours.

Bambara beans

Vigna subterranea legumes grow exclusively in Africa and their color depends on the variety. The seeds are very similar to peanuts and they have to be cooked a long time because of their hardness. Bambara beans represent a staple food in small African communities.

Table 1 Legumes that are considered pulses according to FAO (2016) classification.

Pulses	*Types*	*Scientific name*
Dry beans	Borlotti beans	*Phaseolus vulgaris* L.
	Black beans	
	Cannellini beans	
	Red kidney beans	
	Haricot beans	
	Flageolet beans	
	Pinto beans	
	Adzuki beans	*Vigna angularis* L.
	Mung beans	*Vigna radiata* L.
	Urd beans	*Vigna mungo* L.
	Tepary beans	*Phaseolus acutifolius* L.
Lupines	Lupines	*Lupinus* L.
Bambara beans	Bambara beans	*Vigna subterranea* L.
Broad beans	Broad beans	*Vicia faba* L.
Lentils	Red lentils	*Lens culinaris* L.
	Yellow lentils	
	Green or brown lentils	
	Puy lentils	
	Umbrian lentils	
Dry peas	Dried green peas	*Pisum sativum* L.
Chickpeas	Bambai chickpeas	*Cicer arietinum* L.
	Desi chickpeas	
	Kabuli chickpeas	
Dry pigeon peas	Dry pigeon peas	*Cajanus cajan* L.
Vetches	Vetches	*Vicia* L.
Dried cowpeas	Dried cowpeas	*Vigna unguiculata* L.
Winged beans	Winged beans	*Psophocarpus tetragonolobus* L.
Sword beans	Sword beans	*Canavalia gladiata* L.

Lentils

Lens culinaris originated from the Middle East, Asia, and North America. Lentils are one of the oldest foods and grow even in arid land and extreme temperatures. There are red, yellow, green or brown lentils that grow more in Asia and North America. Besides, there are puy and umbrian lentils.

Dry peas

Pisum originates from the Mediterranean or Middle East. The leaders in peas production are Canada, Russia, and Ukraine. Unlike lentils, peas are soft and with a spherical form. Dry peas are used to make soups, flours, and snacks.

Chickpeas

Cicer arietium or chickpeas originated in Turkey. They are very appreciated because of their potential use in the culinary sector, nutritional value, and storing potential. They grow around the globe and are a staple food in European, Arabic, Mexican, North American and western Asian cuisines. There are three types of chickpeas: Bambai, Desi, and Kabuli type.

Dry pigeon peas

Cajanus cajan are native to India region and Africa, specifically from the Congo Republic. Now, they grow easily in some parts of the Caribbean, India, and Africa, but the greater consumers are found in Asia. Pigeon peas have become an important food source in some eastern countries.

Vetches

Vicia sativa was consumed since the Neolithic Era by humans and its cultivation was standardized and documented during the Roman Empire. Now, it is used primarily for livestock fodder.

Dried cowpeas

Vigna unguiculata seeds originated from West Africa. Nowadays, they grow and are consumed all over Asia, Africa, Southern Europe, and Central and South America. Besides, it is a very important ingredient in Creole and Indian curries.

Winged beans and sword beans

Winged beans (*Psophocarpus tetragonolobus*) originating in New Guinea and sword beans (*Canavalia gladiata*) from certain tropical regions of Asia.

Pulses are mainly used for foods, but soybeans and peanuts are legumes of high economic impact. Each year, more than 342 million tons of soybeans are produced worldwide but only 6% is consumed as food, mainly in Asia (Semba et al., 2021). For peanuts, only 50% of the more than 62 Mt produced yearly are consumed as food and the other half is mainly used for oil production. Green peas and beans are also legumes but not considered pulses since they are classified as vegetables. Clover and alfalfa are also leguminous crops, but they have no economic relevance related to their consumption as food. Therefore, pulses are the most important leguminous foods, and their value has been recognized by the UN. To create a worldwide awareness of their importance for sustainable food production, the UN General Assembly proclaimed 2016 as the International Year of Pulses (IYP) and the Food and Agriculture Organization of the United Nations (FAO) promoted the nutritional and environmental benefits of the diverse crops included in this category. Since 2019, the World Pulses Day has been celebrated on February 10th.

Nutritional value

The proximate contents, protein, fats, minerals, crude fiber and carbohydrates are given in **Table 2** based on the published data of Kamboj and Nanda (2018). Legumes have high protein content, more than 20% superior quantities of protein in comparison to other plant foods and have twice the dietary protein content of cereal grains. Pulse proteins were classified into two major fractions, albumin and globulin. Globulins are the major storage proteins in pulses seeds constituting 35–72% of total protein, they have higher amounts of glutamine aspartic acid, arginine and lysine. Albumins constitute only up to 15–25% of total seed protein. The albumins have higher amounts of cystine, methionine and lysine contents as compared to globulin fractions. In general all legumes have high lysine content and are low in sulfur-containing amino acids. In the legumes methionine, cysteine and lysine are in the range of 30–90, 30–130 and 400–520 mg per g of nitrogen (N), respectively.

Table 2 Proximate composition of legumes (as per 100 g of seeds) based on the published data of Kamboj and Nanda (2018).

Legumes	*Protein (g)*	*Fat (g)*	*Minerals (g)*	*Crude fiber (g)*	*Carbohydrates (g)*
Chickpea	17.1–22.5	5.2–5.6	2.5–3.0	1.0–3.9	58.1–60.9
Field bean	24.9	0.8	3.2	1.4	60.1
Moth beans	23.6	1.1	3.5	4.5	56.5
Kidney bean	22.9	1.3	3.2	4.8	60.6
Horse	22	0.5	3.2	5.3	57.2
Cowpea	24.1	1	3.2	3.8	54.5
Pigeonpea	9.8–22.3	1.0–1.7	1.0–3.5	1.5–6.2	16.9–57.6
Khesari	28.2	0.6	2.3	2.3	56.6
Lentil	25.1	0.7	2.1	0.7	59
Black gram	24	1.4	3.2	0.9	59.6
Peas	7.2–22.9	0.1–14	0.8–2.4	4.0–4.4	15.9–58.8
Green gram	24.5	1.2	3.5	0.8	59.9

Legumes contain relatively low quantities of the essential sulfur containing amino acids methionine and cysteine which are found in higher quantities in cereal. In different regions worldwide there are traditional dishes prepared with the combination of cereals and legumes. For example, in Asia and Latin America many traditional recipes have variations of rice and beans. Traditional African cuisine makes great use of a wide variety of beans, chickpeas and grains. In Mexico and Venezuela, corn tortillas and arepas are commonly accompanied with beans. In Morocco and Algeria chickpeas are paired with whole wheat couscous and chicken for a delicious exotic stew. Kochari is a typical dish made from lentils and rice in Egypt. Congee is prepared from rice as well such as millet and served with a variety of cooked pulses.

Commonly consumed legumes have carbohydrate content in the range of 15.9–60.9%. Carbohydrate comprises monosaccharides, oligosaccharides, other polysaccharides and starch. In the legume seeds, starch is the main source of accessible carbohydrate. The range of dietary fiber in legumes is from 0.7 to 6.2%. Legumes are mainly rich in resistant starch (RS) and have low glycemic index. In addition, they can help in moderating blood sugar levels after meals and improving insulin sensitivity. They also play a positive role in weight reduction since their consumption induces satiety. The resistant starch and fiber act as prebiotics for the probiotic or beneficial bacteria. Bacterial fermentation leads to the development of short-chain fatty acids, such as butyrate, which improve colon health through promoting a healthier gut microbiome and reducing colon cancer risk.

Other health promoting components of legumes are their phenolic compounds such as phenolic acids, tannins and flavonoids, which are related with the antioxidant, anti-tumoral, antiplatelet, anti-inflammatory and anti-allergic properties of different varieties of legumes. Dark colored and pigmented pulses have more phenolic content compared to light colored varieties. In addition, legumes are an important source of saponins, particularly in chickpeas, lentils, dry beans and peas. Saponins differ from phenolic compounds in their chemical structure since they are glycosylated forms of triterpenoids, mainly oleanolic acid, or steroids. In recent years, saponins have been proven to have anticarcinogenic, antimutagenic, hypoglycemic, hypocholesterolemic, hepatoprotective, immunomodulatory, neuroprotective, anticoagulant, antiinflammatory and antioxidant activities.

Legumes are normally low in fat and have no cholesterol but contain phytosterols that contribute to their cholesterol lowering effects. Additionally, they are a superior source of linoleic and -linolenic acid in the range of 21–53% and 4–22%, respectively. Chickpeas have the highest monounsaturated fatty acid content (34 g/100 g), kidney beans the highest polyunsaturated fat content (71.1 g/100 g) and butter beans have the highest saturated fat content (28.7 g/100 g). Lupines contain a higher monounsaturated fat and lower saturated fat content. Phytosterols, mono and polyunsaturated fatty acids decrease the possibility of coronary heart diseases.

As sources of micronutrients legumes are much superior to cereals because legumes have higher initial minerals. Legumes have up to 3.5% minerals, which corresponds mainly to phosphorus, calcium, magnesium and potassium (**Table 3**). They are an important source of other minerals such as iron and zinc, essential in human nutrition (Kamboj and Nanda, 2018). The minerals are found in the seed coat (bran) in major proportions which are discarded during processing. Most legumes, including common beans are consumed whole, resulting in conserving their mineral contents, they are an important source of iron and their consumption can reduce the risk of anemia. Dry beans and chickpeas have the highest content of essential minerals.

Legumes are a good source of B complex vitamins (thiamine, niacin, riboflavin and folic acid). Chickpea, cowpea, green gram, pigeonpea, lentil, black gram, peas and red gram contain significant amounts of free folic acid, a range of 4.6–69 μg/100 g seeds. Cowpea is an excellent source of B complex vitamins.

Global trends of production and economic importance

According to FAOSTAT, the global production of pulses is about 88.7 million tons. Dry beans, including pigmented beans, represent about 34% of the production of the main legumes cultivated in the worldwide, followed by dry peas, chickpea, cowpea, fava bean, lentil, pigeonpea, miscellaneous beans, lupin, and Bambara bean (**Fig. 1**). Dry beans, often called the

Table 3 Micronutrients contents of legumes (as per 100 g of seeds) based on the published data of Kamboj and Nanda (2018).

Legumes	*Minerals (g)*	*Iron (mg)*	*Calcium (mg)*	*Magnesium (mg)*	*Potassium (mg)*	*Zinc (mg)*	*Thiamine (mg)*	*Riboflavin (mg)*	*Niacina (mg)*	*Folic acid (μg)*
Chickpea	2.5–3.0	4.6–9.5	58–202	119–169	720–808	1.7–6.1	0.20–0.48	0.15–0.18	1.3–2.9	24–34
Field bean	3.2	2.7	60	–	–	–	0.52	0.16	1.8	–
Moth beans	3.5	9.5	202	225	1096	–	0.45	0.09	1.5	–
Kidney bean	3.2	5.1	260	184	–	4.5	–	–	–	–
Horse	3.2	6.77	287	156	762	2.8	0.42	0.2	1.5	–
Cowpea	3.2	8.6	77	210	1131	4.6	0.51	0.2	1.3	69
Pigeonpea	1.0–3.5	1.1–2.7	57–73	58–90	463–1104	0.9–3.1	0.32–0.45	0.19–0.33	3	19
Khesari	2.3	6.3	90	92	644	–	0.39	0.17	2.9	–
Lentil	2.1	7.58	69	74–80	629	2.8–3.1	0.45	0.2	2.6	14.5
Black gram	3.2	3.8	154	130–154	800	1.7–6.1	0.42	0.2	2	24
Peas	0.8–2.4	1.5–7.05	20–81	34–122	79–750	2.3	0.25–0.47	0.01–0.21	0.8–3.5	4.6
Green gram	3.5	3.9	75	127	843–1150	2.8–3.0	0.47	0.21–0.27	2.1–2.4	24.5

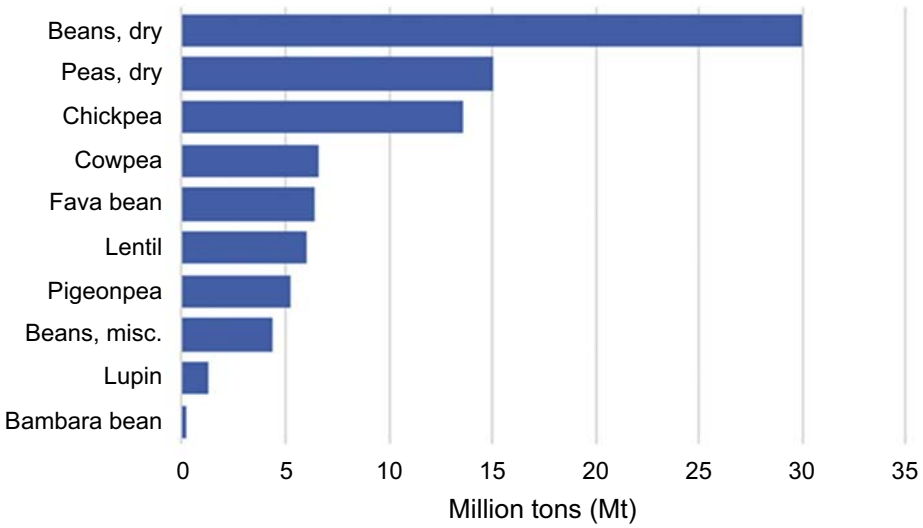


Fig. 1 Global production of different legumes, 2014–2019, with amounts shown in million tons (Mt) (FAOSTAT, 2020).

"meat of the poor", provide micronutrients to more than 300 million people in the tropics and, in many areas, are the second most important source of calories following maize. In South Africa, dry beans accounted for about 80% of total caloric supply provided by pulses. Their production is key to ensuring food security in many poor regions of the world (Semba et al., 2021). In contrast, miscellaneous beans that include winged bean (*Psophocarpus tetragonolobus*), lablab (*Lablab purpureus*), jack bean (*Canavalia* spp.), guar (*Cyamopsis tetragonoloba*), velvet bean (*Mucuna pruriens*), and yam bean (*Pachyrhizus erosus*) are produced in less quantity compared to dry beans, they only represent the 5% of the production. Dry peas and chickpea represent 17% and 15% of world production of legumes for human consumption.

Pulses are diverse in their traditional food uses in Asia, Africa, and America, where they are mainly consumed in soups, spreads, meal components, snacks, and breakfast items (**Table 4**). As an example, boiled sword beans called "avarakai" in India are considered an essential food for pregnant women. As well, fried or boiled are the base of the meals of Mesoamerican population, they are eaten up to three times a day.

Since the common food uses of pulses include them boiled or in soups or stews, one of the most popular industrialized products is canned cooked beans. For instance, the global market of canned legumes was valued at 2.56 billion of USD in 2019 and it is expected to grow at 4.0% of annual growth rate (CAGR) from 2020 to 2027. The consumer's demand for canned legumes has been increasing as they are a rich source of plant-based proteins (GVR, 2020).

Market trends such as the rise in awareness to increase protein consumption, vegan diets, and the relevance of ethnic food are stimulating the demand for pulses. The increase of legume consumption is also related to the growth of people looking for alternatives to animal protein, besides, the key motivations could be the plant-based food benefits, and awareness of climate change. The Planetary Health Diet highlights an increased consumption of legumes as one way to move toward sustainable and healthy diets. These trends have changed the past paradigms when the people thought that legumes were for the low-income population ("meat of the poor") and nowadays legumes have become the main protein source even for eccentric diets.

Table 4 Traditional dishes and their principal countries or regions.

Scientific name	*Type*	*Common food uses*	*Country*
Phaseolus vulgaris L.	Pinto beans	"Frijoles refritos" (refried beans), boiled beans, soups, in salads, stews, spreads, chilli, cowboy beans	Mesoamerica
	Black beans		Southern of United States of America
		Feijoada	Brazil
			Mozambique
			Angola
			Portugal
	Red kidney beans	Boiled beans with rice	Southern of United States of America
		Minestrone	Italy
	Haricot beans	Boston baked beans	New England
			United States of America
Vigna angularis L.	Adzuki beans	Desserts, confections, soups	Asia
Lupinus L.	Lupines	Raw, pickled, in brine, mashed, cooked in soups or stews; flour used for pasta, pastry, dairy product substitutes; salty snack	Greece
			Jordan
			Egypt
		Tremocos	Portugal
		Altramuces	Spain
		Chocho	Ecuador
		Traditional dish with olives	Italy
		Traditional fermented foods such as tempe, miso, and soy sauce	Asia
Vigna subterranea L.	Bambara beans	Cooked in soups, stews, cassoulet	Ghana
			Zimbabwe
Vicia faba L.	Broad beans	Stews, salads, burrata bruschetta	South-west Asia
			Africa
			Greece
Lens culinaris L.	Lentils	Soups, boiled or fried, stews, sprouted grain in salads, flour (breads and cakes), food for infants	Mesoamerica
			Africa
			India
		Gravies, snacks, sweets	India
		Dhal	Asia
		Koshari	Egypt
Pisum sativum L.	Dried green peas	Dhal, soups, stews	Asia
		Meat analogs, in vegetarian pakoras, pea flour (baked goods, pasta, and noodles, extruded snacks, breakfast cereal)	India
		Matzoh balls	Israel
		Falafel	Middle East
Cicer arietinum L.	Chickpeas	Boiled, fried, cooked, crushed snacks, dhal, sweets, snacks, bread, hummus	South Asia
			Mediterranean region
			Pakistán
		Falafel	Middle East
		Fermented food (dhokla), curry, chhole	India
Cajanus cajan L.	Dry pigeon peas	Stew	Caribbean
Vigna unguiculata L.	Cowpea	Boiled, steamed, or fried; fritters or akara	Africa
			Caribbean
			Brazil
		Steamed bean pudding, moimoi	Nigeria
		In salads	Mediterranean region
		Traditional dish, "Hoppin' John"	United States of America
Psophocarpus tetragonolobus L.	Winged beans	Boiled or fried seeds, stews	Asia
Canavalia gladiata L.	Sword beans	Boiled, stews, soups	India
			Africa
			Asia

Legume proteins/flours in emerging markets

The market of plant-based proteins has increased dramatically due to the massive investment, revenue increases, and the rise of new product development. In fact, in the first part of 2020 the international investment in plant-based proteins hit $1.1 billion, more than double for 2019 according to FAIRR Initiative. Likewise, the global plant protein market value of $29.4 billion in 2020 could surpass $162 billion by 2030 according to Bloomberg Intelligence.

The global inclination of sustainable, regenerative, plant-based foods, and gluten free products are some of the drivers for the demand of these products. Besides, the growing demand for a protein-rich diet, the health and wellness trend, the focus on meat alternatives, the organic trend, and the allergies and intolerances of animal proteins constitutes the core of the demand increase of the legume protein and/or flour. The main sources of legume protein are peas, chickpeas, lentils, and lupin. The main applications are food and beverages, animal feed, supplements and others. The North America region accounted for the largest market share in the global market.

There are some formulation challenges in taste, texture, nutrition, and functionality when pea, chickpea, lentils, fava beans and other pulses were used. Different products play an important role in the market such as the native flours, precooked flours, extruded products, modified vegetable proteins, legume-based foods, plant-based meats, and others.

Legume processing and antinutritional factors

Legume seeds enclose numerous rather minor proteins, which includes protease and amylase inhibitors, lipoxygenase, lectins and others, which are related to the nutritional or functional quality of the seed. The trypsin inhibitor activity (U/mg) of legumes varies, going from 1 for *Vigna mungo* to more than 90 for *Glycine max* (Avilés-Gaxiola et al., 2018). This has been a great concern to improve the digestibility of legume proteins and, traditionally, thermal treatments have been the alternative to inactive trypsin inhibitors like Kunitz (KTIs) and Bowman-Birk (BBTIs).

Thermal treatments such as boiling, roasting, cooking, are common ways to inactivate trypsin inhibitors but they affect functional properties of the resulting products as well as destroy certain phytochemicals. Biological processes like fermentation and germination have also been used for centuries and proven to reduce the trypsin inhibitor activity of legumes. In contrast to thermal treatments, germination and fermentation improve the phytochemical profile and the beneficial effects on health related to legume consumption. Chemical treatments, mainly using alkali, are also used to inactivate protease inhibitors in different soybean industrialization processes but remaining chemicals or side reaction products are an important disadvantage. Extrusion, ultrasound, irradiation and high hydrostatic pressure treatments are emerging technologies that have also been proved to reduce the trypsin inhibitory activity of legumes (Avilés-Gaxiola et al., 2018). With the growing demand for new vegetable proteins, emerging technologies are an excellent alternative for legume processing.

Conclusions

Legumes are an excellent alternative to increase the protein consumption in more sustainable diets. Pulses are in a transition from being considered the "meat of the poor" thanks to their incorporation as food ingredients. Resistant starch, micronutrients and phytochemicals are legumes components that have beneficial effects beyond nutrition. Pulses commonly consumed around the globe are very diverse and this contributes with healthy diets maintaining ethnic foods and promoting innovative uses thanks to emerging technologies.

References

Avilés-Gaxiola, S., Chuck-Hernández, C., Serna Saldívar, S.O., 2018. Inactivation methods of trypsin inhibitors in legumes: a review. J. Food Sci. 83, 17–29.

Food and Agriculture Organization of the United Nations (FAO), 2016. Pulses: Nutritious Seeds for a Sustainable Future. Available at. https://www.fao.org/3/i5528e/i5528e.pdf. (Accessed 30 December 2021).

Food and Agricultural Organization of the United Nations Statistics (FAOSTAT), 2020. Production: Crops and livestock products. Available at. https://www.fao.org/faostat/en/#data/QCL. Accessed 30 December 2021.

Grand View Research (GVR), 2020. Canned Legumes Market Size, Share & Trends Analysis Report by Product (Beans, Peas, Chickpeas), by Distribution Channel (Offline, Online), by Region, and Segment Forecasts, 2020–2027. Available at. https://www.grandviewresearch.com/industry-analysis/canned-legumes-market. (Accessed 30 December 2021).

Kamboj, R., Nanda, V., 2018. Proximate composition, nutritional profile and health benefits of legumes-a review. Legume Res. 41 (3), 325–332.

Semba, R.D., Ramsing, R., Rahman, N., Kraemer, K., Bloem, M.W., 2021. Legumes as a sustainable source of protein in human diets. Glob. Food Secur. 28, 100520.

Singh, N., 2017. Pulses: an overview. J. Food Sci. Technol. 54 (4), 853–857.

Further reading

Brewster, E., 2021. Plant Proteins Come of Age. Available at. https://www.ift.org/news-and-publications/food-technology-magazine/issues/2021/march/columns/ingredients-plant-proteins-come-of-age. (Accessed 5 January 2022).

FAIRR Initiative, 2021. Sustainable Proteins Hub. Available at. https://www.fairr.org/sustainable-proteins/. (Accessed 5 January 2022).

Food and Agriculture Organization of the United Nations (FAO). What are Pulses. Available at. https://www.un.org/en/observances/world-pulses-day. (Accessed 4 January 2022).

Institute of Food Technologists (IFT), 2021. 10 Food Trend Predictions for 2022. Available at. https://www.ift.org/news-and-publications/digital-exclusives/10-food-trend-predictions-for-2022. (Accessed 5 January 2022).

Ojiewo, C.O., Rubyogo, J.C., Wesonga, J.M., Bishaw, Z., Gelalcha, S.W., Abang, M.M., 2018. Mainstreaming Efficient Legume Seed Systems in Eastern Africa: Challenges, Opportunities and Contributions Towards Improved Livelihoods. Food and Agriculture Organization, United Nations, pp. 70–72.

Päivärinta, E., Itkonen, S.T., Pellinen, T., Lehtovirta, M., Erkkola, M., Pajari, A.M., 2020. Replacing animal-based proteins with plant-based proteins changes the composition of a whole nordic diet-a randomized clinical trial in healthy Finnish adults. Nutrients 12 (4), 943.

Lycopenes and related compounds

G. Tang, Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University, Boston, MA, USA

Published by Elsevier Ltd.
This is a reproduction of G. Tang, Lycopenes and Related Compounds, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 124–130, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00176-8.

Introduction 438
Chemical and Physical Properties of Lycopene; its Food Sources and Enteric Absorption 438
Tissue Contents and Kinetics of Lycopene Turnover 441
Functional Properties and Tissue Health 442
Health, Research Models, and Epidemiological Evidence 442
Evidence for Possible Anticancer Protection by Lycopene 443
Lycopene and Cardiovascular Disease 443
Other Disease-Related Investigations 444
Conclusions 444
Further Reading 444

Introduction

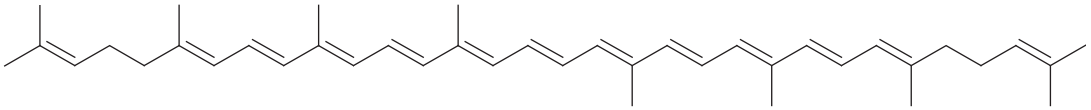
Lycopene, the most abundant pigment in ripe red tomatoes and in a few other fruits, is one of the major carotenoid pigments that is widely present in the diet of the human population. Figure 1 illustrates the chemical formula of selected carotenoids that occur widely both in human diets and in the noncellular fraction of human blood in most regions of the world. Carotenoids are yellow-to-red in color, with lycopene being nearer the red end of the carotenoid series. However, unlike the other carotenes and cryptoxanthins, it does not possess a beta-ionone ring structure at either end of the molecule, and this precludes it from becoming a precursor of vitamin A in humans and animals. Nevertheless, it is readily transformed from the all-*trans* form that is characteristic of most plants and plant foods for animals and humans, to a range of mono- and di-*cis* forms within the animal's body. In addition, oxidation to epoxides and hydroxylated derivatives occurs, although the control of these oxidation pathways and the nature of their products are not yet well understood or characterized.

In plant tissues, where it is synthesized, lycopene is thought to help protect vulnerable photosynthetic tissues from light- and oxygen-catalyzed damage. Its role in humans and other animals, which can only obtain the pigment from their diet, is less well understood. Indeed it remains unproven that there is an essential role for lycopene in animal tissues. Nevertheless, considerable research effort is currently being undertaken to test hypotheses that are attempting to link human dietary and tissue lycopene levels to the risk of degenerative diseases, such as vascular diseases, cancers, etc., especially in older people. As discussed in more detail below, this research is being performed in a wide range of tissue culture and animal model systems and human epidemiological studies.

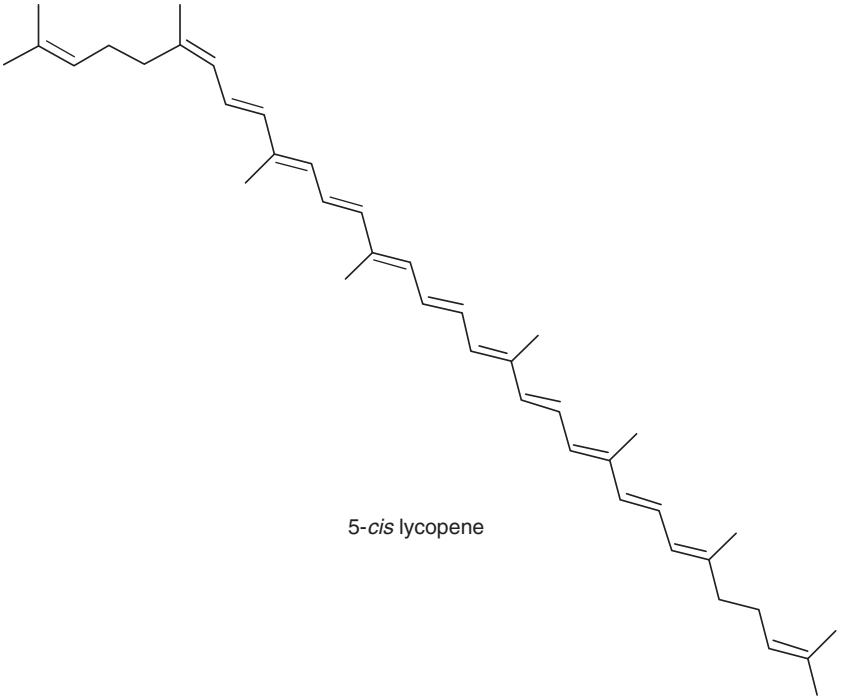
In this article, some key aspects of the chemical and physical properties, the dietary sources, biochemical status indices, and biological significance of lycopene will be described.

Chemical and Physical Properties of Lycopene; its Food Sources and Enteric Absorption

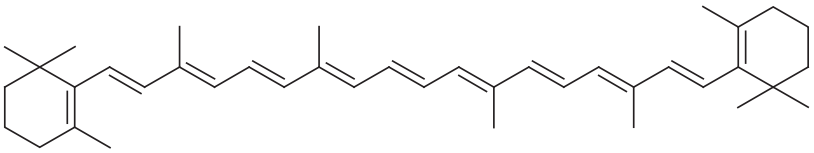
Lycopene (molecule weight 536.9) is the most commonly encountered of that subgroup of the naturally occurring carotenoids that have a straight-chain poly-isoprenoid molecule without any terminal β-ionone ring structures (Figure 1). The chain length and number of conjugated double bonds determine the absorption spectrum, which peaks at 472 nm with a molar extinction coefficient, $\varepsilon^{1\%}$ of 3450. It is one of the most nonpolar members of the carotenoids, and in organic solution it is also one of the most easily oxidized and thus is easily destroyed, which necessitates the use of rigorous precautions against its oxidative destruction during its extraction and analysis from plants, foods, animal tissues, and body fluids. Currently, such analytical determination is usually based on high-performance liquid chromatography (HPLC), using either its characteristic light absorption property, or its natural fluorescence, or its redox character, for detection and quantitation by absorbance or fluorometric or electrochemical detection. The lycopene content of selected commonly consumed foods is listed in Table 1. The original information can be found in the US Department of Agriculture (USDA) ARS recent publication (2010), Nutrition Data Laboratory Home Page. It is interesting to find out that in south Asia, the fruit Gac (*Momordica cochinchinensis*) contains more lycopene, almost three-fold more, than does the tomato fruit. Another characteristic that greatly affects lycopene stability and the problems of its storage and analysis is the phenomenon of *cis–trans* isomerization. Naturally occurring lycopene in tomatoes, the major human food source of this carotenoid, is nearly 100% all-*trans* (Figure 1), but during the processing of food, and then during the processes of absorption and accumulation in animal tissues, there is a progressive increase in the proportion of a variety of *cis*-forms. Most of these *cis*-forms contain



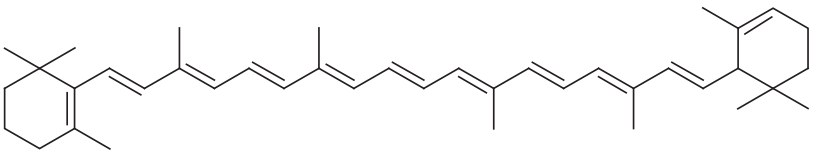
All-trans lycopene



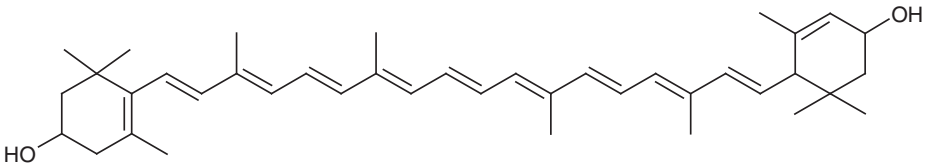
5-*cis* lycopene



All-*trans* β-carotene



All-*trans* α-carotene



Lutein

Figure 1 Structures of lycopene and certain other carotenoids found in human blood and tissues.

a single *cis*-bond (mono-*cis*-lycopene), and the 5-, 9-, 13-, and 15- mono-*cis*-lycopenes account for more than 50% of the total lycopene in human serum. Smaller quantities of di-*cis*-lycopenes are normally also present. Curiously, another food source of lycopene, red palm oil, has a much higher natural proportion of the *cis*-forms of the pigment. Isomerization is catalyzed by low pH; therefore,

stomach acid is believed to be a major factor in the conversion of the all-*trans*-lycopene ingested from tomatoes and their products to a mixture of *cis*-forms in the digestive tract. There is also evidence that further isomerization occurs between the digestive tract and the portal lymphatic lipid micelles. The *cis*-isomers differ from the all-*trans* form in their absorption and inter tissue transportation properties, and also in their functional characteristics; for instance, they are more soluble in lipophilic solvents and structures are less likely to aggregate into crystalline forms. More research data suggested that the greater bioaccessibility of *cis*-forms compared with all-*trans* form of lycopene was the reason for higher proportion of the *cis* form in tissues. A recent report showed that cholesterol membrane transporter SR-B1 was involved in lycopene intestinal absorption. However, lycopene and its various forms in association of physicochemical differences and their related biological consequences have yet to be adequately explored.

Of all the most common naturally occurring carotenoids, lycopene is by far the most efficient in reacting with and quenching singlet oxygen, $^1O_2$, which is a non-free-radical excited and reactive form of oxygen. This form of oxygen reacts rapidly with lycopene to yield nonexcited triplet oxygen and excited triplet lycopene. The latter then dissipates its extra energy by solvent interactions, thus regenerating nonexcited lycopene and preserving its original structure by recycling. However, another of its chemical interactions with molecular oxygen appears to result in irreversible oxidation to yield one or more cyclic epoxides, which then probably undergo ring-opening. Nevertheless, there are many unresolved questions about the nature and importance of the many degradation pathways that are believed to result in the irreversible destruction of lycopene both *in vitro* and *in vivo*.

As a food component, consumption of lycopene in tomato is well tolerated and generally safe. The possible side effect of high consumption of tomato products would be mild digestive upset. Lycopene from tomatoes is permitted as a food color.

Lycopene is an essential intermediate in the pathway for synthesis of the β-ionone ring-containing carotenoids such as β-carotene in plant tissues, and in most plant tissues it is present in only minor amounts. However, in a few, including tomato fruit, watermelon, and red grapefruit, this conversion to the β-ionone ring products by the enzyme lycopene cyclase is hindered, so that the intermediate carotenoid forms, lycopene, phytoene, and phytofluene, accumulate instead.

Table 1 Lycopene (mg) content of selected foods, sorted by nutrient content collected from the USDA Nutrient Database for Standard Reference, Release 23

Food description	*Lycopene content (mg per 100 g)*
Tomatoes, sun-dried	45.9
Tomato products, canned, paste, without salt added	14–28.8
Sauce, pasta, spaghetti/marinara, barbecue, ready-to-serve	4.4–12.7
Catsup (Ketchup)	16.7
Sauce, salsa, ready-to-serve	10.5
Vegetable juice cocktail, canned	9.7
Tomato juice, canned, with salt added	9.0
Pasta with meatballs in tomato sauce, canned entree	7.2
Soup, minestrone, canned, reduced sodium, ready-to-serve or tomatoes, canned, stewed, prepared with equal volume water	4.1–6.4
Watermelon, raw	4.5
Tomatoes, red, ripe, canned, stewed	4.1
Spaghetti with meat sauce, frozen entree	3.2
Soup, clam chowder, manhattan, canned, prepared with equal volume water with bean, pork, beef noodle or chunky vegetable, etc.	2.9–4.3
Papayas, raw	1.8
Tomatoes, red, ripe, raw, year round average	2.6
Pizza, pepperoni, cheese, meat and vegetable topping, regular crust	1.8–2.0
Grapefruit, raw, pink and red, all areas	1.4
Fast foods, cheeseburger; single, regular patty, with condiments	1.0
Fast foods, hamburger; single, regular patty; with condiments	
Beans, baked, canned, with franks	0.4
Salad dressing, Russian dressing, Thousand Island dressing, French dressing, reduced fat, etc.	0.6–3.6

Source: http://www.ars.usda.gov/ba/bhnrc/ndl

In the US, tomato products provide more than 85% of the total quantity of lycopene consumed by the human population. Mean lycopene intakes in the US are considerably greater than they are in the UK, where the mean daily intake is thought to be less than one-third that in the US, while lycopene intakes in Far Eastern countries such as China and Thailand appear to be much lower still. Wild tomatoes originated in Central America and were introduced into Europe following the opening up of the New World, and were later introduced back into North America from Europe. Because tomatoes are the major source of dietary lycopene in many human populations, some epidemiological studies have been designed on the simplistic assumption that tomato consumption can be used as a general proxy for lycopene consumption, and that any disease associations with tomato consumption can be attributed to the biological effects of lycopene. However, tomatoes also contain significant amounts of other carotenoids, vitamin C, bioflavonoids such as naringenin, and phenolic acids such as chlorogenic acid. Much of the existing epidemiological evidence for possible beneficial effects of lycopene (see the following sections.) cannot distinguish unequivocally between the biological effects of lycopene and those of the many other bioactive constituents present in tomatoes.

The bioavailability of lycopene from raw tomatoes is low, but it is greatly increased by cooking or by commercial processing such as conversion to soup, sauce, ketchup (catsup), etc., and its availability is also increased by increasing the fat content of the food.

A survey conducted by National Health and Nutrition Examination Survey (NHANES) in years 2007–2008 on "what we eat in America", showed that the average intake of lycopene from food was 5.5 mg per day per person, see **Table 2**. More information can be obtained from the USDAs web page.

Bioavailability of synthetic lycopene in an oil capsule was reported as better than from cooked and pureed tomatoes when taken by human subjects. Interactions with other carotenoids are complex and have only partly been studied, for instance β-carotene in the same dish seems to increase the absorption of lycopene, but large doses of β-carotene given separately seem to decrease the lycopene content in serum. The strength of the correlation between dietary lycopene intake and blood (serum or plasma) lycopene concentration varies greatly among studies and clearly depends on many factors, one of which is the degree of sophistication of the food table values, since subtle differences in food sources and meal composition affect its bioavailability very considerably.

Tissue Contents and Kinetics of Lycopene Turnover

Once absorbed, passively from lipid micelles by the enterocyte, lycopene enters the portal lymphatics and thence the liver, from which it enters the peripheral bloodstream, mainly in association with the β-lipoproteins, in which it is transported to the peripheral tissues. Its half-life in plasma is of the order of 12–33 days; longer than that of β-carotene, which is less than 12 days. Clearly, many of these factors are interdependent, and there is a need for further clarification of the key independent determinants of lycopene status, and whether plasma levels can provide an adequate picture of tissue and whole body status.

Patients with alcoholic cirrhosis of the liver have greatly reduced hepatic lycopene concentrations; indeed, hepatic lycopene seems to offer a sensitive index of hepatic health. Studies of organ concentrations (**Table 3**), suggest a gradient from circulating levels in plasma to different ones in specific tissues. The different carotenoid ratios between organs (not shown) also indicate selective transport and accumulation. However, the mechanisms involved are poorly understood. No lycopene is detectable in the retina or lens of the eye, where lutein and zeaxanthin are found; however, lycopene is present in the ciliary body.

Table 2 Daily lycopene intake from food consumed by the US population – mean amount consumed per individual and percentages of lycopene consumed with meals and snacks by age and gender

Gender and age	*Lycopene (SE) mg*	*Percentage of lycopene intake in meals and snacks*
Males		
≥20 years Sample size, *n*=2662	6.8 (0.3)	Breakfast, 7 Lunch, 31 Dinner, 51 Snacks, 11
Females		
≥20 years Sample size, *n*=2758	4.6 (0.3)	Breakfast, 4 Lunch, 27 Dinner, 57 Snacks, 13
Males and females		
≥2 years Sample size, *n*=8529	5.5 (0.2)	Breakfast, 5 Lunch, 30 Dinner, 53 Snacks, 12

Abbreviation: SE, standard error.
Source: http://www.ars.usda.gov/services/docs.htm?docid=18349.

Table 3 Concentrations of lycopene reported in human tissues

Tissue	*Range of mean or median (in Italics) lycopene concentrations (nmol per g weight)*
Adrenal	1.9–21.6
Testis	4.3–21.4
Liver	0.6–5.7
Brain	2.5
Lung	0.2–0.6
Kidney	0.1–0.6
Stomach, colon	0.2–0.3
Breast, cervix	0.2–0.8
Skin	0.4
Adipose tissue	0.2–1.3
Prostate	0.1–0.6
Plasma	0.2–1.1

Note: Values were gathered from 12 publications, all based on HPLC analysis.

Functional Properties and Tissue Health

The capacity for quenching of singlet oxygen has been mentioned above; the exceptionally high rate constant, $K=3.1\times10^{10}$ mol^{-1} s^{-1}, renders it one of the most efficient of known quenchers of this powerful oxidant. In the plant, it probably protects chlorophyll, which produces singlet oxygen as a by-product of photosynthesis. In experiments with lymphoid cells, lycopene provided better protection against singlet oxygen damage than several other carotenoids tested. In skin exposed to UV light, lycopene disappears much more rapidly than β-carotene. Lycopene is also able, in model systems, to inhibit the peroxidation of polyunsaturated lipids and the oxidation of DNA bases to products such as 8-hydroxydeoxyguanosine (8-OHdG). It can react directly with hydrogen peroxide and nitrogen dioxide. In addition, lycopene supplementation may modify DNA through DNA repair mechanisms.

Several studies in tissue culture have shown a reduction in the formation of oxidation damage products such as malondialdehyde, and have found less injury to cells exposed to oxidants such as carbon tetrachloride, if lycopene (or other carotenoids) is present.

Another characteristic of lycopene and other carotenoids that may be relevant to inhibition of cancer cell growth is the modulation of gap junction cell–cell communication processes. In particular, carotenoids including lycopene have been shown to enhance the efficacy of the protein, connexin43, which helps to ensure the maintenance of the differentiated state of cells and to reduce the probability of unregulated cell division, which is deficient in many tumors. They may also interact with and enhance the synthesis of binding proteins that down-regulate the receptor for the growth-promoting hormone insulin-like growth factor-1 (IGF-1). In one small clinical intervention with tomato drinks, it was reported that changes in circulating lycopene were inversely and significantly correlated with those of IGF-1.

In certain circumstances, lycopene can reduce low-density lipoprotein (LDL) -cholesterol levels, possibly by inhibiting hydroxymethylglutaryl CoA reductase (HMGCoA reductase), the rate-limiting enzyme for cholesterol synthesis (see the Section on Lycopene and Cardiovascular Disease). Lycopene was shown to have modest hypocholesterolemic properties in one small clinical trial.

Lycopene has shown significant antimigration and anti-invasion activity in association of its induction of nm23-H1 expression, a metastasis suppressor gene.

Health, Research Models, and Epidemiological Evidence

Table 4 summarizes the various types of evidence that have been used to test the hypothesis that lycopene may have health-promoting or protective properties in man. The ultimate proof of efficacy, which would be long-term controlled intervention studies with clinical diseases or mortality as the end points, are extremely difficult, expensive, and time-consuming to obtain, and cannot address all possible benefits in a single intervention trial.

The two disease categories that have so far received most attention for possible long-term benefits of lycopene have been the amelioration of cancers and of heart disease. Both benefits are plausible in view of the physicochemical and biological properties of lycopene outlined above, because both categories of disease are characterized by tissue damage, which is thought to be induced or exacerbated by reactive oxygen species in the environment or those generated within the body.

Table 4 Types of evidence being sought to show that a nutrient such as lycopene may protect against oxidation-induced or other disease processes

1. Model *in vitro* systems, for example, oxygen-derived free-radical trapping in cell free chemical mixtures.
2. Tissue (cell and organ) cultures, for example, reduction of optical opacity development in cultured eye lenses; reduced growth rates or apoptosis in tumor cell cultures; and protection of key macromolecules, especially nucleic acids.
3. Animal studies demonstrating a reduction of oxidation-induced damage or disease with lycopene supplements or with lycopene-rich foods such as tomatoes or tomato products.
4. Human observation studies using intermediate biochemical markers: for example, inverse relationships between lycopene intakes or blood levels and biochemical markers, such as lipid or DNA oxidation products.
5. Studies using pathology-related intermediate markers, for example, arterial thickening or reduced arterial elasticity; precancerous polyposis, etc.
6. Relationships (without intervention) between tomato intakes or estimated lycopene intakes or lycopene contents of serum, plasma, or tissues (e.g., fat biopsies) and actual disease prevalence or incidence in human cross-sectional, case–control, or prospective epidemiological studies.
7. Intervention studies: lycopene supplements producing a reduction in biochemical markers of oxidation damage or in functional markers, or, eventually, in actual human disease incidence or progression.

Evidence for Possible Anticancer Protection by Lycopene

Most of the indications with respect to cancer come from human studies linking tomato intake, total estimated lycopene intake, and serum or plasma lycopene concentrations to the subsequent development of cancers (Table 5). There is a small amount of evidence from experimental animal studies, for instance, rat and mouse dimethylbenzanthracene-induced mammary tumor studies have supported the hypothesis, as has a model of spontaneous mammary tumor formation in one strain of mice, but many of the animal models of tumor promotion have been criticized as being too dissimilar from the likely processes of spontaneous tumor genesis in humans.

Partly for historical reasons, there has been a particular interest in prostate cancer (Table 5). A large and early trial in the US (US Health Professionals Follow-up Study) reported an impressive difference between groups with high and low intakes of tomatoes and hence of lycopene for subsequent prostate cancer development, which was not shared with other carotenoids. Plausibility was enhanced by the fact that although human prostate lycopene concentrations are not especially high on an absolute basis (Table 3), they are higher than those of other carotenoids in this tissue. Subsequent studies have had variable outcomes. A small pilot study reported that tomato oleoresin supplements given for a short period to prostate cancer sufferers who were due for radical prostatectomy resulted in smaller tumor size and other apparent benefits, but this trial now needs to be repeated on a larger scale. Although lycopene has shown protective effects against prostate cancer and inhibits the progression in patients with benign prostate hyperplasia, there is insufficient evidence to recommend the use of lycopene supplements for cancer patients.

Several studies have provided evidence for protection of certain regions of the digestive tract against tumor occurrence or growth. Two studies, one in Iran and another in Italy, found an inverse relationship between esophageal cancer and tomato consumption. Two Italian and one Japanese studies reported evidence for protection against gastric cancer, and two studies claimed a reduction in pancreatic cancer. A case–control study on histologically confirmed pancreatic cancer cases and population-based controls in eight Canadian provinces reported that dietary intake of lycopene, provided mainly by tomatoes, was associated with 31% reduction in pancreatic cancer risk among men when comparing the highest and lowest quartiles of intake.

Results with other cancers have been mixed and inconclusive.

Lycopene and Cardiovascular Disease

Table 6 summarizes the evidence. The European Multicentre Euramic Study reported that risk of developing myocardial infarct was inversely related to lycopene intake, after appropriate adjustment for other cardiovascular risk factors. Some Scandinavian studies

Table 5 Summary of evidence for association of dietary intake of tomato, or serum or plasma concentrations of lycopene, with possible protection against prostate cancer

Number of studies	*Locations*	*Total number of participants*	*Types of trial*	*Outcome conclusion*
2	Greece, Canada	937	Case–control (intake of tomato or lycopene, or blood level)	Significant association
7	USA, UK, Canada, New Zealand	3824	As above	NS
3	USA	954	Prospective studies based on dietary estimates	Significant association
1	Netherlands		As above	NS
3	USA	723	Prospective studies based on serum or plasma lycopene	Inconclusive; one study found a marginal (p=.05) benefit versus aggressive cancer

Note: Significant association=p<.05 and NS=no significance (p>.05).

Table 6 Summary of evidence for association of relatively high serum or plasma lycopene with reduced risks of cardiovascular disease (CVD)

Study	*Location*	*Sex (total participants)*	*Types of trial and outcome measures*	*Outcome conclusion*
Euramic	Europe, multicenter	M (1379)	C-C, MI	Significant association with protection[a]
ARIC	USA	M+F (462)	C-C, IMT	NS
Street	USA	M+F (369)	NC-C, MI in smokers	NS
Rotterdam	Netherlands	M+F (216)	C-C, PC	Significant association with protection
Bruneck	Italy	M+F (392)	CS+PFU, PC	NS
Linkoping–Vinus	Sweden and Lithuania	M (210)	CS, mortality from heart disease	NS
Kuopio (KHID)	Finland	M (725)	PFU, acute coronary event or stroke	Significant association with protection
Kuopio (ASP)	Finland	M+F (520)	IMT	Significant protection for males; not significant for Females
Physicians' Health Study	USA	M (case 499+matched free CVD,499)	C-C, CVD	NS
Women's Health Study	USA	F (case 483+matched free CVD 483)	C-C, CVD	Significant association with protection

Abbreviations: C-C, case–control study; NC-C, nested case–control study; CS, cross-sectional study; PFU, prospective follow-up study; MI, myocardial infarct; IMT, intima-media thickness estimate; PC, plaque count; NS, no significant evidence for protection.

Note: Significant association with protection is $p<.05$, generally after appropriate adjustment for other known CVD risk factors.

[a]No association with plasma b-carotene in this study.

have subsequently supported this claim; moreover, lycopene is capable of reducing LDL-cholesterol levels, possibly by inhibiting hydroxymethylglutaryl CoA reductase (HMGCoA reductase), the rate-limiting enzyme for cholesterol synthesis.

Other Disease-Related Investigations

In an organ culture model, some evidence for protection of rat lenses against induction of cataractogenesis has been reported. There is good reason to believe that carotenoids in general may play a role in the protection of ocular tissues against the damaging effects of UV light and of reactive oxygen substances, whose exposure to light carries some analogy with the known functions of carotenoids in plant tissues. A possible protective role in the ciliary body and iris has been proposed, but not yet tested.

Conclusions

Clearly lycopene possesses chemical and biological properties which make it a very attractive candidate for tissue protection and reduction of disease, especially degenerative diseases. Lycopene probably interacts more efficiently with one particular reactive oxygen species, singlet oxygen, than any other commonly occurring nutrient. It appears to share with several other carotenoids the capacity to reduce lipid peroxidation and DNA oxidative damage, and to enhance cell–cell gap junction communication and to protect normal IGF-1 function. It may reduce cholesterol formation and its tissue accumulation in some circumstances. Studies related to cancers and cardiovascular disease are ongoing and are attracting increased research interest.

Further Reading

Basu, A, Imrhan, V., 2007. Tomatoes versus lycopene in oxidative stress and carcinogenesis: Conclusions from Clinical trials. European Journal of Clinical Nutrition 61, 295–303.
Britton, G, 1995. Structure and properties of carotenoids in relation to function. The FASEB Journal 9, 1551–1558.
Clinton, SK, 1998. Lycopene: Chemistry, biology, and implications for human health and disease. Nutrition Reviews 56, 35–51.
Haseen, F, Cantwell, MM, O'Sullivan, JM, Murray, LJ., 2009. Is there a benefit from lycopene supplementation in men with prostate cancer? A systematic review. Prostate Cancer and Prostatic Disease 12, 325–332.
International Symposium on the Role of Tomato Products and Carotenoids in Disease Prevention, 2002. 14 review articles by different authors, plus 17 symposium abstracts. Experimental Biology and Medicine 227, 843–937.
Kavanaugh, CJ, Trumbo, PR, Ellwood, KC, 2007. The U.S. Food and Drug Administration's evidence-based review for qualified health claims: Tomatoes, lycopene, and cancer. Journal of the National Cancer Institute 99, 1074–1085.
Mossine, VV, Chopra, P, Mawwhinney, TP., 2008. Interaction of tomato lycopene and ketosamine against rat prostate tumorigenesis. Cancer Research 68 (11), 4384–4391.
Nguyen, ML, Schwartz, SJ, 1999. Lycopene: Chemical and biological properties. Food Technology 53, 38–45.
Rao, AV, Agarwal, S, 1999. Role of lycopene as antioxidant carotenoid in the prevention of chronic diseases: A review. Nutrition Research 19, 305–323.
Weisburger, J, 1998. International symposium on lycopene and tomato products in disease prevention. Proceedings of the Society for Experimental Biology and Medicine 218, 93–143.

Meal size and frequency: Effect on absorption and metabolism

F.E. Lithander, Trinity College Dublin, Dublin, Ireland

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of F.E. Lithander, Meal Size and Frequency: Effect on Absorption and Metabolism, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 155–159, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00183-5.

Introduction **445**
The Effect of Meal Size on Absorption **445**
The Regulation of Meal Size by Gut-Derived Satiety Peptides and Adiposity Signals 446
The Effect of Meal Size on Metabolism **446**
The Effect of Meal Size on Energy Intake 446
The Effect of Meal Size on Energy Expenditure 447
The Effect of Meal Frequency on Absorption **447**
Mechanisms Underlying the Metabolic Effect of Meal Frequency 447
The Effect of Meal Frequency on Metabolism **448**
The Effect of Meal Frequency on Energy Intake 448
Mechanisms by which Meal Frequency may Influence Energy Intake 449
The Effect of Meal Frequency on Energy Expenditure 449
Further reading **449**

Abbreviations

AMP adenosine monophosphate
AMPK adenosine monophosphate-activated protein kinase
BMR basal metabolic rate
CCK cholecystokinin
CCK-8 octapeptide of cholecystokinin
CoA coenzyme A
DIT dietary induced thermogenesis
GIP glucose dependent insulotropic polypeptide
GLP glucagon-like peptide
HmG-CoA reductase hydroxymethylglutaryl-coenzyme A reductase
NPY neuropeptide-Y
PYY peptide YY

Introduction

Eating habits are changing. Terms such as "super-sizing," "portion distortion," "grazing," and "snacking" have appeared in the contemporary vernacular as a result. Therefore, a better understanding of meal size and frequency is particularly topical, not least when considering the potential role which these eating patterns may be playing in the dramatic rise of noncommunicable diseases such as obesity, diabetes, and cardiovascular disease.

The principal consequences that variations in meal size and frequency have on the body relate to the absorption and metabolism of that meal. There are several factors in addition to meal size and frequency which influence absorption and metabolism, such as macronutrient composition, the energy density of the diet, and the physical volume of the meal. However, the particular contribution which variations in meal size and frequency have made to the dramatic change in society's eating patterns makes them worthy of special attention.

The Effect of Meal Size on Absorption

When a meal of mixed macronutrient composition is consumed, the rate at which the carbohydrate, protein, and fat in that meal is absorbed differs. Carbohydrate in the form of glucose, and protein in the form of amino acids enter the portal vein within 30 min of meal ingestion and later appear in the general circulation. As the glucose concentration in the portal vein rises there is an incre`ase in

the uptake of glucose into the hepatocytes. Pancreatic islet cells react to the increase in blood glucose and secrete insulin, among other hormones, into the circulation. As a result of this increase in circulating glucose and insulin there is a drop in the release of nonesterified fatty acids/free fatty acids from the adipose tissue. Fatty acid oxidation in the skeletal muscle tissue decreases and, as glucose uptake takes place, the muscle cells increase the rate at which glucose is oxidized. Glycogen synthesis in the muscle and liver cells is increased and the uptake of amino acids by muscle tissue may also occur. Up to 4 h after the ingestion of the meal, fat in the form of chylomicron-triacylglycerol enters the circulation via the lymphatic system. The action of the hormone lipoprotein lipase in the adipose tissue has by now increased, which promotes the storage of fatty acids as triacylglycerol in adipocytes. This synopsis indicates that following the ingestion of a meal, there is a marked increase in glucose oxidation with a corresponding decrease in fat utilization resulting in the storage of fat.

It is therefore fair to say that the larger a meal that is consumed, the more pronounced the responses described above will be. If a large meal is eaten, for example, the plasma glucose concentration may remain elevated for up to 4 h following ingestion. Conversely, the smaller a meal, the more subtle the effect. This indicates that meal size does indeed influence absorption. However, in order for the relationship between meal size and absorption to be fully understood, the role which absorption itself plays in determining meal size needs to be considered. The following section will therefore focus on the process of absorption and the systems which are in place to control the amount of food eaten. These systems appear not to be working optimally in many cases leading to a state of positive energy balance, which is the cause of the worldwide increase in the incidence of obesity.

The Regulation of Meal Size by Gut-Derived Satiety Peptides and Adiposity Signals

The signaling systems that underlie the regulation of appetite, dietary intake, and therefore meal size are complex. These include signals from the gastrointestinal (GI) tract such as ghrelin, CCK, and GLP-1, and from adipose tissue such as leptin and adiponectin. This information is routed to the hypothalamus and brain stem where neuronal networks are activated which signal the commencement or conclusion of a meal.

Ghrelin is an orexigenic hormone whereas CCK and GLP-1 both promote satiety. Serum levels of ghrelin fall on commencement of eating whereas CCK is secreted from duodenal mucosal epithelial cells and stimulates the delivery of digestive enzymes from the pancreas, as well as bile from the gallbladder, into the small intestine. In addition, CCK is produced by neurons in the enteric nervous system and is widely distributed in the brain. Although CCK (and other GI satiety signals) acts to limit meal size, it is important to note that CCK concentration post-meal is not affected by body-fat stores, meaning that it does not take into consideration the existing energy depot of the individual when signaling the onset of satiety. Adiposity signals must therefore be considered in parallel, as they also play a part in determining the size of a meal consumed.

Until quite recently, it was thought that adipose tissue was an inert storage depot but it is now widely recognized as an active endocrine organ. Leptin and adiponectin are two adipose-derived hormones which play a role in the regulation of appetite and therefore meal size. The treatment of obese, hyperphagic, leptin-deficient individuals with exogenous leptin resulting in consumption of more appropriately sized meals has demonstrated that leptin is intimately involved in the regulation of food intake. Leptin however seems to be an overall 'caretaker' of energy intake rather than responding acutely to individual meals. When adipose mass increases, circulating leptin concentrations increase in turn suppressing appetite until which time adipose mass is lost. The mechanism by which leptin can help to reduce food intake is not fully understood although it appears to activate pathways in the brain that reduce food intake while inhibiting pathways that activate food intake.

Adiponectin is an adipokine with cardio-protective and antidiabetic properties and it is hypothesized that adiponectin regulates food intake in conjunction with leptin. When fasting, there is an increase in the adiponectin signal in the hypothalamus leading to an increase in the activity of AMP-activated protein kinase and a subsequent stimulation of food intake. The leptin signal in the hypothalamus is regulated inversely in relation to the adiponectin signal in the brain; therefore adiponectin enhances hypothalamic AMPK activity and food intake, as opposed to the action of leptin.

To conclude, the size of a meal that an individual consumes is determined by a number of physiological, behavioral, and societal factors that interact and play a critical role in the regulation of dietary intake.

The Effect of Meal Size on Metabolism

Energy homeostasis, achieved by matching energy intake with energy expenditure, is partially dependent on the regulation of meal size consumed. In order for meal size to have an effect on energy metabolism, it must affect either or both components that are involved in the regulation of energy balance, namely energy intake and energy expenditure. Energy balance is the difference between energy ingested and energy expended over a given period of time. Consequently, energy storage is equal to intake minus expenditure. The following sections examine the effect of meal size on the two components of the energy balance equation.

The Effect of Meal Size on Energy Intake

Meal portion sizes have been growing steadily since the 1970s, and have been doing so in parallel with the increasing prevalence of obesity. It is known that portion and meal sizes vary depending on the food source, location of consumption, and number of people eating a meal together. Not surprisingly, the largest portions consumed in terms of energy are generally those obtained at fast food

restaurants, although the portion sizes of home-cooked meals have been growing steadily as well. Meal size may thus be contributing to the problem of obesity by leading to a daily total energy intake which is greater than the daily total energy expenditure resulting in a positive energy balance.

The Effect of Meal Size on Energy Expenditure

Total energy expenditure can, as shown in Table 1, generally be divided into three major components: BMR, thermogenesis, and physical activity. In order for meal size to have an effect on the energy expenditure side of the energy balance equation, it must have an effect on one or more of these three components. There is no evidence to suggest that meal size has an effect on BMR, which refers to the energy expended to run the body on a day-to-day basis. Thermogenesis broadly refers to the body's production of heat, which is divided into three categories: dietary, thermoregulatory, and adaptive. It is the dietary category, commonly known as DIT, which is of greatest relevance to the current discussion of the effect of meal size on energy expenditure. It refers to the heat lost by the body as a result of the absorption and metabolism of a recently ingested meal. DIT represents approximately 10% of energy intake, and therefore the energy expended on DIT increases and decreases in relation to the size of the meal and, more importantly, the energy value of the meal consumed. The larger a meal, the more energy will be expended to absorb, transport, and metabolize the nutrients consumed in that meal. For example, in the case of a meal containing 2000 kJ of energy, approximately 200 kJ will be expended on DIT alone. It is in the physical activity component of energy expenditure that the greatest variation between individuals can be observed, as physical activity levels (and therefore the energy expended on activity) in the population are contingent on lifestyle choices such as employment and leisure time activities. The effect which meal size could have on physical activity is somewhat difficult to quantify. Meal size is perhaps more important to elite athletes whose energy expenditure is two to three times greater than untrained weight-matched athletes with up to 40% of their energy expenditure being the cost of training.

The Effect of Meal Frequency on Absorption

Humans are eating more often than before with an average eating frequency among US adults of five eating occasions per day. The perceived health advantages of increased meal frequency (as opposed to eating larger, infrequent meals) have been interesting researchers since the 1930s. It has been suggested that frequent eating increases metabolism, reduces food cravings, improves insulin and glucose control, and reduces body weight. However, eating frequently may actually cause an increased exposure to energy-dense foods, leading to increased energy intake and weight gain as opposed to weight loss.

The benefits of a frequent eating approach were originally made apparent by the discovery that insulin requirements in diabetics could be decreased in a frequent meal regime. In a series of case reports on patients taking high insulin doses, it was demonstrated that improved glycemic control and decreased insulin requirements can be achieved when glucose is sipped at hourly intervals throughout the day. Similarly, in healthy individuals a diet composed of many small meals compared with an isoenergetic one composed of larger meals results in decreased insulin and glucose fluctuations.

Meal frequency not only affects insulin and glucose levels but also influences an individual's circulating lipids. An inverse relationship exists between meal frequency and lipid levels, suggesting that infrequent feeding leads to an increased risk of cardiovascular disease due to large fluctuations in circulating lipids. Increased meal frequency, however, is associated with several benefits such as decreased serum cholesterol levels, decreased total : high density lipoprotein cholesterol (HDL-C) ratio, decreased esterified fatty acids and decreased enzyme levels in adipose tissue associated with fatty acid storage. Paradoxically, individuals who report that they eat more frequently not only have lower total and low density lipoprotein cholesterol (LDL-C), but also have a greater intake of energy, total fat, and saturated fatty acids. Considering that some of these results were found in a free-living population, it is possible that dietary mis-reporting, a common occurrence in an overweight population, could be to blame for this inconsistency.

Mechanisms Underlying the Metabolic Effect of Meal Frequency

The mechanisms underlying beneficial responses to frequent feeding as opposed to an infrequent meal pattern are not fully understood. Frequent feeding has been shown to elicit lower plasma glucose fluctuations than does a more infrequent eating pattern. The absolute amount of carbohydrate eaten at each episode of ingestion in a frequent feeding pattern is simply not great enough to

Table 1 Major components of energy expenditure

Component	*Total energy expenditure (%)*	*Represents*
Basal metabolic rate (BMR)	60–75	Day-to-day running costs of an individual, for example, circulation
Thermogenesis	10–20	Heat produced by the body through dietary, adaptive, and thermoregulatory processes
Physical activity	100 minus (BMR+thermogenesis)	The sum of work carried out by an individual

elevate glucose to the same extent as more infrequent eating. Small elevations in plasma insulin seen with frequent feeding are most likely to have been in response to minimal fluctuations in glucose. The mechanisms responsible for the effect of an increased frequency of meal eating on lipid metabolism are not as clear cut. The lower serum cholesterol levels observed during frequent feeding may be related to lower serum insulin levels. Insulin appears to have a key role in enhancing the hepatic synthesis of cholesterol through its ability to stimulate HmG-CoA reductase, the rate limiting enzyme in hepatic cholesterologenesis. Exogenous insulin quickly increases HMG-CoA reductase activity in rats with diabetes and raises levels of the enzyme in animals without the disorder. It is possible that the reduction of serum cholesterol during a diet of habitual frequent feeding in normal healthy individuals may result from a reduction in hepatic cholesterol synthesis, secondary to the maintenance of euglycemia at lower serum insulin levels. A reduction in cholesterol synthesis would result in an increase in LDL receptors, further lowering total and LDL-cholesterol levels.

Alternatively, or in addition, the benefits associated with an increased feeding regimen may reflect unintentional or uncontrolled changes in dietary energy and fat intake that may occur when an individual's meal frequency is altered. It is not clear whether a diet of frequent eating results in any adaptational responses of enzymes or hormones which in turn may be providing additional benefit to the individual.

Much of the research in which these benefits were uncovered is difficult to interpret due to the variety of methods used, the lack of information available regarding the foods consumed and the exact nature of the dietary intervention. The majority of measurements are made on fasted blood samples, when in fact most individuals are in a postprandial state for the greater part of every 24-h period. The results of such research must be interpreted with a degree of caution for a number of reasons such as the small sample size used and the interactions with other factors which may prolong absorption time (e.g., soluble fiber, low-glycemic index foods, and the administration of alpha-glycosides).

As discussed, frequent feeding has been demonstrated to lower circulating plasma glucose, insulin, and lipids in both healthy and diabetic subjects in the short term. In addition to the lack of clarity on the mechanisms involved, further research is needed to investigate any medium- and long-term benefits of frequent feeding. It is important that, if deemed desirable in terms of metabolic control, increasing the number of periods of feeding encourages the desired dietary pattern and mix of macronutrients and micronutrients and is not offset by the failure to decrease the size of the meals.

The Effect of Meal Frequency on Metabolism

The same maxim which was earlier applied to the study of meal size, namely that it can only influence energy metabolism if it affects energy intake and energy expenditure, is applicable to meal frequency. The following sections will focus on energy intake and energy expenditure, respectively.

The Effect of Meal Frequency on Energy Intake

It has long been argued that the frequency of meal intake may have an effect on body weight regulation. It has been suggested that there is an inverse association between meal frequency and body weight in individuals. However, there are a number of flaws in the design of many of the research studies from which these data have been derived, and great caution is required in the interpretation of the results. Design flaws include (1) dietary under-reporting of the number of eating occasions and of energy intake, (2) reverse causality, which refers to the possibility that people abstain from eating meals when they become overweight in an attempt to lose weight or to prevent further weight gain, (3) lack of measurement of physical activity or energy expenditure, and (4) inclusion of people in a diseased state. These important confounding factors may help to resolve the contradictory results of many research trials. Erroneous conclusions have been drawn from the misinterpretation of such results because these studies are extremely vulnerable to methodological errors that may generate spurious relationships which may not actually exist. A recently published review article examined almost 200 abstracts and articles pertaining to eating frequency and body weight which ranged from one to nine eating occasions per day over intervention periods of 2–8 weeks in duration. No association between eating frequency and health was observed and the authors concluded that a number of methodological shortcomings made it difficult to fully confirm this finding. Studies of longer duration with sufficient sample size are warranted which could include the use of dietary biomarkers to validate both reported eating frequency and energy intake.

There appears to be very little direct empirical evidence in humans to suggest that frequent feeding *per se* affects appetite and energy intake. Individuals who eat frequently seem to exhibit a greater capacity to compensate for changes in the energy content of specific meals, relative to those individuals who derive most of their energy intake from fewer larger meals. Over very short periods, and under highly controlled experimental conditions, frequent feeding can decrease energy intake at a subsequent meal, which may in turn have an effect on appetite regulation. It remains to be seen however, whether the same would occur in free-living conditions. One final consideration is that of reduced eating frequency, that is less than three meals per day, and data indicate that this pattern negatively affects appetite control. It is important to note however that many of the controlled studies that were designed to investigate the effect of meal frequency on energy intake were conducted whereby meal patterns were enforced onto individuals irrespective of the habitual eating frequency of those individuals. There may also be differences in the short-term compensation of energy intake depending on an individual's habitual eating pattern.

Mechanisms by which Meal Frequency may Influence Energy Intake

Although the evidence is inconclusive, there is a suggestion that feeding frequency may have an impact on appetite and hence affect energy intake. The control of appetite is a very complex issue which is determined by a number of factors as discussed previously. However, the question remains as to whether the frequency of feeding elicits effects on any of these factors, in turn affecting appetite and possibly body weight. It is noteworthy that rapid declines in blood glucose concentration are associated with hunger in addition to the initiation of feeding in humans.

Although difficult to measure, frequency of feeding may affect the release of appetite regulatory hormones including neuroendocrine hormones such as NPY, galanin, orexin, and melanocortins from the hypothalamus. The release of such hormones may either be stimulated or suppressed during frequent feeding, leading to either higher or lower than normal hormone levels, which may in turn have knock-on effects on energy intake and/or expenditure. When subjective appetite sensations were measured using visual analog scales, increased eating frequency tended to lead to lower peaks in perceived appetite and PYY responses compared with decreased eating frequency although this effect disappeared when investigated over a 24-h period. The release of other gut hormones such as CCK, GLP, and GIP may be altered in relation to feeding frequency. In rats, the infusion of the sulfated octapeptide of CCK (CCK-8) causes a significant reduction in meal size as previously mentioned, whereas meal frequency is increased to compensate for the small meals. However, little is known about the effects of meal pattern on CCK in animals or humans. It is possible that frequent feeding may affect CCK-release in one of two ways: (1) Frequent feeding may cause the regular release of the hormone in response to each feed, persistently alerting the brain that the individual is satiated; or (2) CCK may be released into the circulation in such small amounts in response to frequent feeding that it is not recognized by the brain and the individual continues to eat. Similar effects may occur with GLP and GIP. It is notable that the rate of gastric emptying is also unaffected by antecedent eating frequency.

The Effect of Meal Frequency on Energy Expenditure

As detailed in **Table 1**, the three components of energy expenditure are BMR, thermogenesis, and physical activity. For meal frequency to have an influence on energy expenditure, it must affect one or more of these components. The first component, BMR (which represents 60–75% of energy expenditure in sedentary individuals), is not known to be influenced by meal frequency. Much the same can be said for thermogenesis, an area where extensive research has failed to demonstrate a link between feeding frequency and DIT. It is reasonable and logical to expect that any difference between frequent and infrequent meal-eating patterns would be seen most clearly during the postprandial period when food has just been eaten, where the rate of ingestion of nutrients may alter EE and fuel storage.

Although much research has been carried out on the effects of meal frequency on total energy expenditure, little of it has isolated the physical activity component *per se*. The one area where greater attention has been paid to the relationship between meal frequency and physical activity is that of the performance of elite athletes, as the manipulation of the meal pattern can potentially be used as a tool to achieve optimal performance. Because carbohydrate requirements of elite athletes are high and endogenous glycogen reserves are limited, athletes undertaking prolonged strenuous exercise seek to maximize carbohydrate availability at all times.

Irrespective of the above, the key determinant of feeding frequency's overall effect on energy balance is whether it has an impact on 24-h energy expenditure, where energy intake is fixed in content and composition, and where physical activity is kept constant. Although many of these studies have demonstrated differences in carbohydrate, fat, and protein oxidation rates in response to a gorging versus a nibbling eating pattern, each study has reached the same conclusion, namely that no relationship exists between frequency of feeding and overall 24-h energy expenditure. The majority of these studies used either direct or indirect calorimetry or doubly labeled water in their measurements, each of which are highly reliable energy expenditure measurement techniques.

In conclusion, the contemporary terminology referring to the tendency to increase the amount of food eaten at a meal and the greater frequency at which food is eaten which has appeared in our vocabulary recently demonstrates the importance of a clear understanding of the consequences of meal size and frequency on health. We know that satiety peptides and adiposity hormones attempt to control the size of a meal eaten, and that increased meal frequency, within the constraints of energy balance, has been found to have beneficial effects attenuating circulating substrates. However, in order to elucidate the influence that meal size and frequency has on absorption and metabolism, and to clarify whether this increase in the volume of food eaten at a meal and the greater frequency at which food is eaten have a direct effect on health, further research within the free-living population is required.

Further reading

Bellisle, F., McDevitt, R., Prentice, A.M., 1997. Meal frequency and energy balance. British Journal of Nutrition 77, S57–S70.

Blevins, J.E., Schwartz, M.W., Baskin, D.G., 2002. Peptide signals regulating food intake and energy homeostasis. Canadian Journal of Physiology & Pharmacology 80, 396–406.

Blundell, J.E., King, N.A., 1996. Overconsumption as a cause of weight gain: Behavioral–physiological interactions in the control of food intake (appetite). Ciba Foundation Symposium 201, 138–154 discussion 154–158, 188–193.

Drummond, S., Crombie, N., Kirk, T., 1996. A critique of the effects of snacking on body weight status. European Journal of Clinical Nutrition 50, 779–783.

Jenkins, D.J.A., 1997. Carbohydrate tolerance and food frequency. British Journal of Nutrition 77, S71–S81.

Leidy, H.J., Campbell, W.W., 2011. The effect of eating frequency on appetite control and food intake: Brief synopsis of controlled feeding studies. Journal of Nutrition 141 (1), 154–157.

Mann, J., 1997. Meal frequency and plasma lipids and lipoproteins. British Journal of Nutrition 77, S83–S90.
McCrory, M.A., Campbell, W.W., 2011. Effects of eating frequency, snacking, and breakfast skipping on energy regulation: Symposium overview. Journal of Nutrition 141 (1), 144–147.
McCrory, M.A., Howarth, N.C., Roberts, S.B., Huang, T.T., 2011. Eating frequency and energy regulation in free-living adults consuming self-selected diets. Journal of Nutrition 141 (1), 148–153.
Melanson, K.J., Westerterp-Plantenga, M.S., Saris, W.H., Smith, F.J., Campfield, L.A., 1999. Blood glucose patterns and appetite in time-blinded humans: Carbohydrate versus fat. American Journal of Physiology 277, R337–R345.
Moran, T.H., Ladenheim, E.E., Schwartz, G.J., 2001. Within-meal gut feedback signalling. International Journal of Obesity and Related Metabolic Disorders 25, S39–S41.
Palmer, M.A., Capra, S., Baines, S.K., 2009. Association between eating frequency, weight, and health. Nutrition Reviews 67 (7), 379–390.
Verboeket-van de Venne, W.P., Westerterp, K.R., 1991. Influence of the feeding frequency on nutrient utilization in man: Consequences for energy metabolism. European Journal of Clinical Nutrition 45 (3), 161–169.
Westerterp-Plantenga, M.S., Kovacs, E.M., Melanson, K.J., 2002. Habitual meal frequency and energy intake regulation in partially temporally isolated men. International Journal of Obesity 26 (1), 102–110.
Wilding, J.P., 2002. Neuropeptides and appetite control. Diabetic Medicine 19, 619–627.

Meat, poultry, and meat products: Nutritional value

P.A. Lofgren, Oak Park, IL, USA

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of P.A. Lofgren, Meat, Poultry, and Meat Products: Nutritional Value, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 160–167, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00184-7.

Introduction **451**
Nutritional Value **452**
Classes of Nutrients and Meat Products **452**
Protein 452
Lipids 452
Carbohydrates 452
Vitamins 452
Minerals 453
Bioavailability of Nutrients and Efficiency for Child Development **453**
Nutrient Density of Meat and Meat Products 453
Meat Sources and Nutritional Values **453**
Beef 453
Pork 455
Lamb 455
Veal 457
Poultry 457
Processed Meats 457
Summary **457**
Further Reading **457**

Glossary

Glycerides Glycerol with fatty acids attached.
Lipids Fats, glycerides, fatty acids, cholesterol, sterols.
Muscle foods Meat and meat products.

Introduction

Animal source foods are major contributors to the nutrients in the food supply in many countries. Of these foods, animal muscle (or meat) foods and products are excellent examples of nutrient-dense, naturally nutrient-rich foods, which provide a relatively large amount of many nutrients per the amount of calories provided in a typical serving. For the purposes of this review, discussion will be limited to the muscle foods: beef, pork, lamb, veal, poultry, and some of the processed products made from these muscle species. Fish/seafood will be covered in a separate article. Milk and dairy products, eggs, and major nonmuscle animal source foods, will also be covered separately.

For meat and meat products there are extensive and comprehensive nutrient databases available for reference for particular products of interest. Thus, this review will provide a sampling of the data available for representative meats and meat products.

One of the best and most comprehensive listings of the nutrient values of all meat, poultry, and other meat products is the extensive nutrient database developed and maintained by the United States Department of Agriculture. In this database complete nutrient profiles are listed for more than 700 beef, 200 pork, 195 lamb, 85 veal, 140 poultry, and 130 turkey products. This database can be accessed and searched on-line from the link/website: http://www.ars.usda.gov/ba/bhnrc/ndl and continues to be updated as new data become available for various food products. The most recent version of this database is USDA National Nutrient Database for Standard Reference, Release 23, published 2010.

For another extensive listing of the nutrient values of many meat and meat products, including some by brand name, the reader is referred to Bowes & Church's Food Values of Portions Commonly Used, 19th edn. This reference, although not as extensive in terms of products listed, does provide data directly on common serving sizes and on some additional nutrient and nutrient-related components of meat products (e.g., omega-3 and *trans* fatty acids, glutathione, vitamin D activity, and other vitamin-like compounds).

Nutritional Value

The nutritional value of foods, including meat and meat products, can be defined in a number of different ways, from simply listing the quantities of various nutrients contained in the foods, to considering biological factors that affect the utilization of these nutrients by the body. Some foods may contain nutrients in forms that the body cannot readily utilize. Thus, nutrient bioavailablilty, or availability becomes important. Other articles and sections of this book will discuss some of these bioavailability issues and subsequent metabolism of food and nutrient sources.

The nutritional value of meat and meat products is related to the quantity and utilization of nutrients and the potential for these products to either enhance or restrict nutrient utilization by the body. There are five major classes of nutrients: proteins, lipids, carbohydrates, vitamins, and minerals.

The nutrient content of meat (muscle foods) is fairly similar among the various mammals, birds, and fish. However, differences in the levels of the various nutrients may result from differences in the carcass composition among species and within species due to differences in the ratio of fat to muscle in the edible portion. As fat percentage increases, nutrient concentration of the muscle portion decreases. Also, to a certain extent, the fat profile/composition and other nutrient content levels may be modified/affected by the animal's diet and genetic makeup.

In general, cooking or heat processing has only minimal effects on the nutritional value of muscle foods. In most cases, cooking usually decreases moisture content and concentrates other nutrients, including fat content, especially in lower fat products. This is due to moisture loss. However, in some intensely heated meat products fat content may also be reduced significantly with negligible loss of other nutrients.

Classes of Nutrients and Meat Products

Protein

Proteins comprise the structural unit of all muscle cells and connective tissues. As such, meat and meat products (muscle foods) are major protein sources. Further, muscle foods, as a group, are excellent sources of high-quality protein that supplies all the essential amino acids in desirable proportions for human consumption. Amino acids are the building blocks of protein and those provided by meat match or exceed the profile required by humans.

The protein content of most muscle foods, on a wet basis, is between 15% and 35%. This figure will change due to the moisture and lipid content of the specific product. On a raw weight basis as purchased at the store, the protein content is generally less than 20%. However, people do not eat muscle foods raw and visible fat in red meat products and skin in poultry products is usually trimmed away. Therefore muscle foods, as consumed, have a much higher protein content, in the range of 30%.

Lipids

The lipid component of meat and meat products includes a diverse group of substances such as glycerides (glycerol with fatty acids attached), phospholipids, and sterols. The basic component of most meat lipids is the fatty acids, which can be saturated, monounsaturated, or polyunsaturated.

The relative amount of lipid in muscle foods is probably the most variable aspect of the nutritional profile. Within the lipid components, the relative amount of the different forms of fatty acids present is another variable among meat products. Despite the common reference to animal fats (and especially meat and meat products) as 'saturated', less than half of all the fatty acids of meats are saturated. The largest proportions of fatty acids in meats are monounsaturated, followed by saturated and then polyunsaturated fatty acids. Among meat products, poultry has a higher proportion of polyunsaturated fatty acids and slightly less saturated fatty acids.

The fat in meat products provides much of the flavor associated with these foods and also contributes to the palatability and overall acceptability by consumers. In addition, the fats in meat and meat products also contain several essential fatty acids (linoleic and linolenic acids) and the fat-soluble vitamins A, D, E, and K.

Carbohydrates

Meat and meat products are not significant sources of dietary carbohydrates. Almost all dietary carbohydrates come from plant sources. The only naturally occurring carbohydrate in muscle foods is glycogen. In some processed meat products, such as those that are 'sugar-cured', there may be additional sucrose or glucose added.

Vitamins

Meat and meat products are especially good sources of most of the water-soluble vitamins. In general, meat is the major dietary source of vitamin B_{12} and is an excellent source of many of the other B-vitamins, such as pyridoxine (B_6), biotin, niacin, pantothenic acid, riboflavin, and thiamin. For vitamin B_{12}, red meat products such as beef and lamb are especially good sources. Pork products

are one of the very best sources of thiamin. Although present in muscle foods, the fat-soluble vitamins are less abundant than in plant foods. Vitamins E and K are present, but at lower levels.

Vitamin D activity may be present in some meat products, but at low levels. This is reflected in the latest update to the USDA Nutrient database (http://www.ars.usda.gov/ba/bhnrc/ndl), where vitamin D activity is listed for some beef, pork, lamb, veal, chicken/turkey, and processed meat products; however, such data is not consistently available. In recent years there has been production research with beef, pork, and lamb to determine if added vitamin D_3 or its metabolites, fed to the animal for a brief period of time before slaughter, can result in improved meat tenderness. Although the results are inconsistent, and commercial application is premature, there is some indication that tenderness may be improved with relatively low levels of vitamin D supplementation, but these seem to leave very little residual vitamin D_3 or its metabolites in the muscle. Research in Denmark found that the more biologically active 25-OH D is present at low levels in meat; however there is as yet no consensus on the conversion factor for 25-OH D to calculate vitamin D activity. Also, there is currently very little data on the vitamin D and 25-OH D levels in many meat products. This represents a potential future area of research regarding the nutrient composition of meat. Meat products are a very good source of choline, second only to whole eggs. In recent years a database for the choline content of common foods has been updated and expanded, http://www.ars.usda.gov/Services/docs.htm?docid=6232.

Minerals

Meat and meat products are good to excellent sources of most minerals. Among the macrominerals, calcium is not high in muscle foods although phosphorus and potassium are prominent. In natural meat products, sodium is present, but not a significant contributor to the diet. However, processed meat products may contain significantly higher levels of sodium (added as part of curing, preserving, or flavor-enhancing ingredients). Some of the microminerals (trace elements) are especially abundant in meat and meat products. Iron is of the greatest significance from meat sources because it is present in the heme form, which is more bioavailable than the nonheme form. Of meat products, beef is an especially rich source of iron in this bioavailable form.

Muscle tissue is an especially rich source of minerals such as phosphorus, potassium, magnesium, iron, copper, zinc, and selenium. For instance, pork, poultry, and beef are especially good sources of selenium.

Bioavailability of Nutrients and Efficiency for Child Development

Muscle foods have been shown to contain "intrinsic" factors that improve the bioavailability of a variety of nutrients. Moreover, the bioavailability of these nutrients from muscle foods is high; often exceeding the availability of the same nutrients in foods derived from plants. Heme iron is one example. Zinc and copper have been shown to be more available from meat sources than from plant sources. Several of the B-vitamins may also be more bioavailable from meat sources than from plant sources.

Another interesting aspect of meat products is the ability to promote the bioavailability of nutrients in nonmuscle foods when the two are eaten together. This has been referred to as the 'meat factor'. Perhaps the best example of this is the positive effect of meat in the diet on nonheme iron sources, also in the diet.

The efficacy of meat in the diet for its benefits for child growth and development and micronutrient status continues to be demonstrated. Studies with school children in Kenya have shown that animal-source foods, including meat, improve dietary quality, micronutrient status, growth, and cognitive function. In the US, studies with breastfed infants have shown that feeding meat as an early complementary food is feasible and associated with improved zinc status. On a larger scale, a multi-site, multinational study is underway to further determine the impact of a daily intake of 1/2 oz of meat in 6–18-month-old infants on linear growth, zinc and iron status, brain growth and neurocognitive development, and infectious disease morbidity. This is being done in populations traditionally dependent on nonmicronutrient fortified plant foods for complementary feeding.

Nutrient Density of Meat and Meat Products

The nutrient density of meat is high. Muscle foods have high levels of essential nutrients per unit of weight and per amount of calories provided. Meat and meat products (muscle foods) provide significant amounts of essential nutrients at levels/concentrations higher than from most other foods relative to the calorie content also provided. The United State Food and Drug Administration (US FDA) food labeling guidelines allow a food to be designated a 'good' source of a nutrient if it contributes 10% or more of the Daily Value (DV), and an 'excellent' source, if it contributes 20% or more of the DV, for that nutrient, per 3 oz serving. Most meat products are good or excellent sources of many nutrients. It is generally recognized that in diets that lack muscle foods, greater care is required in diet/menu selection to ensure that adequate levels of essential nutrients are present and bioavailable.

Meat Sources and Nutritional Values

Beef

Beef is an excellent source of high-quality protein, along with significant contributions of many B-vitamins and minerals. In macronutrient terms, the lean to fat ratio of the particular beef product influences the calorie and nutrient composition. In general, as the

fat content decreases the concentration of other nutrients (especially protein, B-vitamins, and minerals) in beef tend to increase. Most beef products available to the consumer are much leaner than they were 20 or 30 years ago. This is a result of changes due to feeding and genetics to produce leaner animals, and also due to closer trim levels on the products that consumers see in the meat case. Whereas in the past, beef cuts with 1/4″ of fat trim were common, now the same products have only 1/8″ fat trim, or in some cases even 0″ fat trim. In the case of ground beef products, 10 or 20 years ago, 17% fat ground beef was considered as 'extra lean'. Now ground beef is commonly available at fat levels as low as 5% or 10%. Other common fat levels for ground beef are 15%, 20%, and 25%; however, a large proportion of current ground beef sales are now in the 5–15% fat level range.

The fat content of beef contains a varied fatty acid profile, with the largest proportion being contributed by monounsaturated fat, followed by saturated fat and polyunsaturated fatty acids. In addition, being a ruminant product, beef is an excellent source of the naturally occurring fatty acid, conjugated linoleic acid (CLA), which has been demonstrated to provide anticarcinogenic properties among other health benefits.

Table 1 provides the energy, protein, and lipid profile of beef along with other meat sources. For a comparison of the mineral composition of beef products versus that of other common meat sources, see Table 2. For a comparison of the vitamin composition of beef products versus that of other common meat sources, see Table 3.

Table 1 Energy, protein, and lipid profile of meats and meat products[a] (amount per 3 oz/85 g, lean only, cooked, except as noted)

Meat species/cut[b]	*Serving size (g)*	*Energy (kcal kj^{-1})*	*Total protein (g)*	*Total fat (g)*	*Total SFA (g)*	*Total MUFA (g)*	*Total PUFA (g)*	*Total cholesterol (mg)*
Beef								
Composite, Ln 0″, ckd, all grades	85	179/751	25.4	7.9	3.01	3.32	0.27	73
Top Round, Ln 0″, brld, all grades	85	158/662	27.0	4.8	1.67	2.02	0.18	65
Top Loin, Ln 0″, brld, all grades	85	155/649	24.9	5.4	2.06	2.16	0.20	54
Shoulder Pot Roast, Ln 0″, brsd, all grades	85	167/697	26.8	6.6	2.21	2.84	0.40	83
95% Ln Ground Beef, brld	85	145/609	22.4	5.6	2.53	2.31	0.28	65
Pork								
Composite, fresh, Ln, ckd	85	171/713	23.4	7.8	2.63	3.32	0.73	71
Tenderloin, fresh, Ln, rstd	85	122/510	22.2	3.0	1.02	1.13	0.43	62
Center Loin Chop, fresh, Ln, pan-fried	85	190/796	23.5	10.0	3.66	4.51	1.27	60
Shoulder, blade steak, fresh, Ln, brld	85	193/808	22.7	10.7	3.78	4.79	0.92	80
Ham, fresh, Ln, rstd	85	179/751	25.0	8.0	2.80	3.78	0.72	80
Lamb								
Composite, Australian, Ln 1/8″, ckd	85	171/715	22.7	8.2	3.44	3.28	0.36	74
Loin, Australian, Ln 1/8″, brld	85	163/683	22.6	7.4	3.13	2.97	0.31	69
Leg, Australian, Ln 1/8″, rstd	85	162/676	23.2	6.9	2.80	2.81	0.32	76
Foreshank, Australian, Ln 1/8″, brsd	85	140/586	23.4	4.4	1.60	1.99	0.25	78
Composite, New Zealand, Ln, ckd	85	175/733	25.2	7.5	3.28	2.96	0.44	93
Composite, US Domestic, Ln 1/4″, ckd	85	175/733	24.0	8.1	2.89	3.54	0.53	78
Veal								
Composite, Ln, ckd	85	167/697	27.1	5.6	1.56	2.00	0.50	100
Cutlet, leg top round, Ln, pan-fried	85	156/651	28.2	3.9	1.10	1.40	0.35	91
Loin chops, Ln, rstd	85	149/622	22.4	5.9	2.19	2.12	0.48	90
Shoulder, blade, Ln, brsd	85	168/704	27.8	5.5	1.54	1.96	0.49	134
Chicken/turkey								
Broilers, meat only, rstd	85	162/676	24.6	6.3	1.73	2.26	1.44	76
Broilers, Lt meat only, rstd	85	147/615	26.3	3.8	1.08	1.31	0.83	72
Broilers, Dk meat only, rstd	85	174/729	23.3	8.3	2.26	3.03	1.92	79
Turkey, all classes, meat only, rstd	85	145/604	24.9	4.2	1.39	0.88	1.22	65
Turkey, all classes, Lt meat only, rstd	85	133/558	25.4	2.7	0.88	0.48	0.73	59
Turkey, all classes, Dk meat only, rstd	85	159/665	24.3	6.1	2.06	1.39	1.84	72
Processed meats								
Bacon, pork, cured, pan-fried, 1 slice	7.9	42/176	3.0	3.2	1.05	1.42	0.35	9
Sausage, pork, fresh, ckd, 2 links	48	163/680	9.3	13.6	4.38	5.94	1.79	40
Bologna, beef & pork, low fat, 1 slice	28	64/269	3.2	5.4	2.05	2.56	0.46	11
Salami, beef, ckd, 1 slice	26	68/284	3.3	5.8	2.56	2.77	0.27	18

[a]USDA, ARS (2010) USDA National Nutrient Database for Standard Reference, Release 23. Nutrient Data Laboratory Home Page, http://www.ars.usda.gov/ba/bhnrc/ndl
[b]brld, broiled; brsd, braised; ckd, cooked; Dk, Dark; Ln, lean and trim level; Lt, Light; rstd, roasted; SFA, saturated fatty acids; MUFA, monounsaturated fatty acids; PUFA, polyunsaturated fatty acids.

Table 2 Mineral composition of meats and meat products[a] (amount per 3 oz/85 g, lean only, cooked, except as noted)

Meat species/cut[b]	*Serving size (g)*	*Ca (mg)*	*Fe (mg)*	*Mg (mg)*	*P (mg)*	*K (mg)*	*Na (mg)*	*Zn (mg)*	*Cu (mg)*	*Mn (mg)*	*Se (mcg)*
Beef											
Composite, Ln 0″, ckd, all grades	85	7	2.54	22	196	302	56	5.76	0.11	0.02	18.1
Top Round, Ln 0″, brld, all grades	85	6	2.28	18	172	223	36	4.67	0.07	0.01	30.8
Top Loin, Ln 0″, brld, all grades	85	16	1.56	21	195	314	51	4.56	0.07	0.01	28.6
Shoulder Pot Roast, Ln 0″, brsd, all grades	85	11	3.06	21	199	305	52	8.02	0.11	0.01	32.4
95% Ln Ground Beef, brld	85	6	2.41	19	175	296	55	5.47	0.08	0.01	18.4
Pork											
Composite, fresh, Ln, ckd	85	15	0.85	21	196	303	47	2.46	0.07	0.01	37.6
Tenderloin, fresh, Ln, rstd	85	5	0.98	25	227	358	48	2.06	0.09	0.01	32.5
Center Loin Chop, fresh, Ln, pan-fried	85	4	0.66	23	201	386	44	1.82	0.06	0.01	38.6
Shoulder, blade steak, fresh, Ln, brld	85	28	1.33	20	187	292	63	4.27	0.05	0.01	33.4
Ham, fresh, Ln, rstd	85	6	0.95	21	239	317	54	2.77	0.09	0.03	42.4
Lamb											
Composite, Australian, Ln 1/8″, ckd	85	14	1.74	20	176	270	68	4.37	0.13	0.01	9.3
Loin, Australian, Ln 1/8″, brld	85	18	1.85	22	187	289	68	2.96	0.13	0.01	8.8
Leg, Australian, Ln 1/8″, rstd	85	8	1.83	21	182	277	61	4.11	0.13	0.01	5.0
Foreshank, Australian, Ln 1/8″, brsd	85	12	1.62	19	150	217	85	6.74	0.11	0.01	7.7
Composite, New Zealand, Ln, ckd	85	11	2.00	19	209	160	42	3.65	0.10	0.03	1.7
Composite, US Domestic, Ln 1/4″, ckd	85	13	1.74	22	178	292	65	4.48	0.11	0.02	22.2
Veal											
Composite, Ln, ckd	85	20	0.99	24	212	287	76	4.33	0.10	0.03	11.1
Cutlet, leg top round, Ln, pan-fried	85	6	0.74	27	246	376	65	2.87	0.05	0.03	8.8
Loin chops, Ln, rstd	85	18	0.72	22	189	289	82	2.75	0.10	0.03	9.9
Shoulder, blade, Ln, brsd	85	34	1.25	24	214	259	86	6.28	0.15	0.03	12.3
Chicken/Turkey											
Broilers, meat only, rstd	85	13	1.03	21	166	207	73	1.78	0.06	0.02	18.7
Broilers, Lt meat only, rstd	85	13	0.90	23	184	210	65	1.05	0.04	0.01	20.7
Broilers, Dk meat only, rstd	85	13	1.13	20	152	204	79	2.38	0.07	0.02	15.3
Turkey, all classes, meat only, rstd	85	21	1.51	22	181	253	60	2.63	0.08	0.02	31.3
Turkey, all classes, Lt meat only, rstd	85	16	1.15	24	186	259	54	1.73	0.04	0.02	27.3
Turkey, all classes, Dk meat only, rstd	85	27	1.98	20	173	247	67	3.79	0.14	0.02	34.8
Processed meats											
Bacon, pork, cured, pan-fried, 1 slice	7.9	1	0.11	3	44	47	192	0.29	0.01	0.00	5.1
Sausage, pork, fresh, ckd, 2 links	48	6	0.65	8	78	141	360	1.00	0.04	0.00	0.0
Bologna, beef & pork, low fat, 1 slice	28	3	0.18	3	51	44	310	0.42	0.02	0.00	3.1
Salami, beef, ckd, 1 slice	26	2	0.57	3	53	49	296	0.46	0.05	0.01	3.8

[a]USDA, ARS (2010) USDA National Nutrient Database for Standard Reference, Release 23. Nutrient Data Laboratory Home Page, http://www.ars.usda.gov/ba/bhnrc/ndl
[b]brld, broiled; brsd, braised; ckd, cooked; Dk, Dark; Ln, lean andtrim level; Lt, Light; rstd, roasted.

Pork

Pork, like beef, is an excellent source of high-quality protein and contributes significant amounts of many B-vitamins and minerals. As for other muscle foods, pork's nutrient composition is greatly affected by its fat and water content. As fat percentage decreases, the concentration of other nutrients increases. In addition, as pork is cooked, and moisture is removed, the concentration of nutrients also increases. Pork is an excellent source of minerals, such as selenium, iron, zinc, phosphorus, and potassium. Compared to other muscle foods, the contribution of pork to selenium in the food supply is especially significant.

In terms of the B-vitamins, pork is an excellent source. Pork is an especially good source of thiamin (vitamin B_1), being the single best source of this vitamin among commonly eaten foods. The fat profile of pork can be influenced by feeding regimes such that it is more, or less, saturated or firm. However, overall the fatty acid profile of pork is largely monounsaturated, followed by saturated and then polyunsaturated fatty acids.

Table 1 provides the energy, protein, and lipid profile of pork along with other meat sources. For a comparison of the mineral composition of pork products versus that of other common meat sources, see Table 2. For a comparison of the vitamin composition of pork products versus that of other common meat sources, see Table 3.

Lamb

Although representing a smaller portion of overall muscle food consumption, lamb still provides a nutrient profile with significant benefits for the human diet. As a source of high-quality protein, lamb is also a good source of many minerals and B-vitamins. Vitamin B_{12} is especially abundant in lamb. It is also a good source of the minerals iron and zinc.

Table 3 Vitamin composition of meats and meat products[a] (amount per 3 oz/85 g, lean only, cooked, except as noted)

Meat species/cut[b]	*Serving size (g)*	*Thiamin (mg)*	*Riboflavin (mg)*	*Niacin (mg)*	*Pantothenic acid (mg)*	*Vit. B_6 (mg)*	*Folate (mg)*	*Vit. B_{12} (mg)*	*Vit. E (mg)*	*Vit. K (mcg)*
Beef										
Composite, Ln 0″, ckd, all grades	85	0.08	0.20	3.40	0.33	0.30	7	2.64	0.14	1.50
Top Round, Ln 0″, brld, all grades	85	0.06	0.15	4.84	0.53	0.36	9	1.49	0.34	1.30
Top Loin, Ln 0″, brld, all grades	85	0.07	0.13	7.12	0.49	0.53	8	1.39	0.32	1.20
Shoulder Pot Roast, Ln 0″, brsd, all grades	85	0.07	0.23	4.01	0.70	0.46	7	2.88	0.09	1.36
95% Ln Ground Beef, brld	85	0.04	0.15	5.05	0.55	0.35	6	2.10	0.31	1.10
Pork										
Composite, fresh, Ln, ckd	85	0.57	0.26	5.53	0.62	0.49	3	0.58	0.09	0.00
Tenderloin, fresh, Ln, rstd	85	0.81	0.33	6.32	0.86	0.63	0	0.48	0.07	0.00
Center Loin Chop, fresh, Ln, pan-fried	85	0.65	0.29	4.35	0.66	0.33	7	0.52	0.19	0.00
Shoulder, blade steak, fresh, Ln, brld	85	0.64	0.37	3.66	0.69	0.26	4	0.96	0.23	0.00
Ham, fresh, Ln, rstd	85	0.59	0.30	4.20	0.57	0.38	10	0.61	0.22	0.00
Lamb										
Composite, Australian, Ln 1/8″, ckd	85	0.11	0.31	4.94	0.75	0.34	[c]	2.56	[c]	[c]
Loin, Australian, Ln 1/8″, brld	85	0.15	0.28	6.93	0.71	0.44	[c]	1.71	[c]	[c]
Leg, Australian, Ln 1/8″, rstd	85	0.12	0.36	4.87	0.84	0.39	[c]	2.71	[c]	[c]
Foreshank, Australian, Ln 1/8″, brsd	85	0.08	0.24	4.58	0.56	0.22	[c]	2.72	[c]	[c]
Composite, New Zealand, Ln, ckd	85	0.11	0.43	6.53	0.49	0.12	0	2.51	0.16	[c]
Composite, US Domestic, Ln 1/4″, ckd	85	0.09	0.24	5.37	0.59	0.14	20	2.22	0.16	[c]
Veal										
Composite, Ln, ckd	85	0.05	0.29	7.16	1.13	0.28	14	1.40	0.36	5.60
Cutlet, leg top round, Ln, pan-fried	85	0.06	0.32	10.74	1.04	0.43	14	1.28	0.36	4.20
Loin chops, Ln, rstd	85	0.05	0.26	8.04	1.08	0.32	14	1.11	0.42	4.70
Shoulder, blade, Ln, brsd	85	0.05	0.31	4.83	1.35	0.21	13	1.71	0.38	5.80
Chicken/turkey										
Broilers, meat only, rstd	85	0.06	0.15	7.80	0.94	0.40	5	0.28	0.23	2.00
Broilers, Lt meat only, rstd	85	0.06	0.10	10.56	0.83	0.51	3	0.29	0.23	0.30
Broilers, Dk meat only, rstd	85	0.06	0.19	5.57	1.03	0.31	7	0.27	0.23	3.30
Turkey, all classes, meat only, rstd	85	0.05	0.16	4.63	0.80	0.39	6	0.31	0.28	3.10
Turkey, all classes, Lt meat only, rstd	85	0.05	0.11	5.81	0.58	0.46	5	0.31	0.08	0.00
Turkey, all classes, Dk meat only, rstd	85	0.05	0.21	3.10	1.09	0.31	8	0.31	0.54	3.30
Processed meats										
Bacon, pork, cured, pan-fried, 1 slice	7.9	0.04	0.02	0.91	0.10	0.03	0	0.10	0.02	0.00
Sausage, pork, fresh, ckd, 2 links	48	0.14	0.10	3.00	0.35	0.16	1	0.57	0.26	0.20
Bologna, beef & pork, low fat, 1 slice	28	0.05	0.04	0.71	[c]	0.05	1	0.37	0.06	0.10
Salami, beef, ckd, 1 slice	26	0.03	0.05	0.84	0.25	0.05	1	0.80	0.05	0.30

[a]USDA, ARS (2010) USDA National Nutrient Database for Standard Reference, Release 23. Nutrient Data Laboratory Home Page, http://www.ars.usda.gov/ba/bhnrc/ndl

[b]brld, broiled; brsd, braised; ckd, cooked; Dk, Dark Ln, lean and trim level; Lt, Light; rstd, roasted.

[c]Comparable data not available.

In addition, as a ruminant, lamb is another naturally occurring dietary source of CL, A a unique fatty acid with anticarcinogenic and other health benefits (from animal model studies).

Table 1 provides the energy, protein, and lipid profile of lamb along with other meat sources. For a comparison of the mineral composition of lamb products versus that of other common meat sources, see Table 2. For a comparison of the vitamin composition of lamb products versus that of other common meat sources, see Table 3.

Veal

Although representing a smaller proportion of overall meat consumption, veal still provides a nutrient profile that is very beneficial. As with all meat sources, veal provides high-quality protein in a product that may be slightly leaner (in terms of fat) than other red meat sources. Compared to other meat sources, veal would have a lower iron content.

Table 1 provides the energy, protein, and lipid profile of veal along with other meat sources. For a comparison of the mineral composition of veal products versus that of other common meat sources, see Table 2. For a comparison of the vitamin composition of veal products versus that of other common meat sources, see Table 3.

Poultry

The nutrient composition of poultry (chicken and turkey) is similar to that of red meat animals (beef, pork, lamb, veal) with a few exceptions. Poultry is lower in iron content, and thus heme iron, than beef. Turkey is slightly higher in several minerals (Ca, Fe, P, K, Zn, and Cu) than chicken. As in red meats, there are significant amounts of several B-vitamins (e.g., niacin, B_6, pantothenic acid) compared to other meat sources, and these are not significantly reduced during cooking.

The fat content of poultry is predominantly monounsaturated fat, followed by saturated fat and polyunsaturated fat. Poultry fat, like pork fat, is somewhat more unsaturated than beef fat. Poultry is significantly higher in polyunsaturated fat compared to beef, pork, lamb, and veal.

Table 1 provides the energy, protein, and lipid profile of chicken and turkey along with other meat sources. For a comparison of the mineral composition of chicken and turkey products versus that of other common meat sources, see Table 2, and of the vitamin composition of chicken and turkey products versus that of other common meat sources, see Table 3.

Processed Meats

Processed meats represent a diverse array of products that have undergone additional treatment from the fresh meat form to the point of consumption. Some of these processing/treatments might include curing with other ingredients added, addition of salt or other flavor or preservative mixtures, etc. Also, these products often represent combined meat sources.

Table 1 provides the energy, protein, and lipid profile of several processed meat products along with other meat sources. For a comparison of the mineral composition of these processed products versus that of other common meat sources, see Table 2, and for the vitamin composition of these processed products versus that of other common meat sources, see Table 3.

Summary

Muscle foods (meat and meat products) provide significant amounts of essential nutrients at levels/concentrations higher than from most other foods relative to the calories provided. Most of all the essential nutrients are present in muscle foods at some level. Furthermore, muscle foods provide nutrients in a form that enhances the bioavailability of nutrients from both the meat itself and from other dietary sources. It is generally recognized that in diets that lack muscle foods, greater care is required in diet/menu selection to ensure that adequate levels of essential nutrients are present and bioavailable.

Further Reading

Council for Agricultural Science and Technology (1997) *Contribution of Animal Products to Healthful Diets*. Task Force Report No. 131. Ames: CAST.

Council for Agricultural Science and Technology (1999) *Animal Agriculture and Global Food Supply*. Task Force Report No. 135. Ames: CAST.

Foote, M.R., Horst, R.L., Huff-Lonergan, E.J., Trenkle, A.H., Parrish Jr., F.C., Beitz, D.C., 2004. The use of vitamin D3 and its metabolites to improve beef tenderness. Journal of Animal Science 82 (1), 242–249.

Godber, J.S., 1994. Nutritional value of muscle foods. In: Kinsman, D.M., Kotula, A.W., Breidenstein, B.C. (Eds.), Muscle Foods – Meat, Poultry, and Seafood Technology. Chapman & Hall, New York, pp. 430–455.

Krebs, N.F., Westcott, J.E., Butler, N, Robinson, C., Bell, M., Hambidge, K.M., 2006. Meat as a first complementary food for breastfed infants: feasibility and impact on zinc intake and status. Journal of Pediatric Gastroenterology and Nutrition 42, 207–214.

Neumann, C.G., Bwibo, N.O., Murphy, S.P., et al., 2003. Animal source foods improve dietary quality, micronutrient status, growth and cognitive function in Kenyan school children: background, study design and baseline findings. Journal of Nutrition 133 (11 Suppl), 3941S–3949S.

Neumann, C.G., Murphy, S.P., Gewa, C., Grillenberger, M., Bwibo, N.O., 2007. Meat supplementation improves growth, cognitive, and behavioral outcomes in Kenyan children. Journal of Nutrition 137 (4), 1119–1123.

Ovesen, L, Brot, C, Jakobsen, J, 2003. Food contents and biological activity of 25-hydroxyvitamin D: a vitamin D metabolite to be reckoned with? Annals of Nutrition and Metabolism 47 (3–4), 107–113.

Pennington, J.A.T., Douglass, J.S., 2009. Bowes & Church's Food Values of Portions Commonly Used, 19th edn. Lippincott Williams & Wilkins, Baltimore.

US Department of Agriculture, Agricultural Research Service (2008) USDA Database for the Choline Content of Common Foods – Release Two. http://www.ars.usda.gov/Services/docs.htm?docid=6232

US Department of Agriculture, Agricultural Research Service (2010) USDA National Nutrient Database for Standard Reference, Release 23. Nutrient Data Laboratory Home Page. http://www.ars.usda.gov/ba/bhnrc/ndl

Wright L (2010) *First Bites: Complementary Feeding – A Global Network Cluster Randomized Controlled Trial*. Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). ClinicalTrials.gov Identifier: NCT01084109.

Mediterranean diet

Emilio Martinez de Victoria Muñoz[a] **and Angel Gil**[b,c], [a] Department of Physiology, Institute of Nutrition and Food Technology "José Mataix", Biomedical Research Center, University of Granada, Health Technological Park, Granada, Spain; [b] Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology "José Mataix", Center of Biomedical Research, University of Granada, Instituto de Investigación Biosanitaria IBS, Granada, Spain; and [c] CIBEROBN (Physiopathology of Obesity and Nutrition), Institute of Health Carlos III (ISCIII), Madrid, Spain

© 2023 Elsevier Ltd. All rights reserved.

Introduction 458
Origin and historical development of Mediterranean food 459
Origin of the term Mediterranean diet and recognition of the Mediterranean diet as an Intangible Cultural Heritage of Humanity by UNESCO 459
Structure and characteristics of the Mediterranean diet 460
Adherence to the Mediterranean diet 462
The Mediterranean diet as a healthy food pattern 462
The Mediterranean diet as a sustainable food pattern 465
The future of the Mediterranean diet 465
Summary page 465
References 466

Key points

- The Mediterranean diet is more than a food pattern. It is a lifestyle of the population living in the countries bordering the Mediterranean Sea.
- Historically it is a diet resulting from the culture of numerous civilizations that inhabited this geographical area.
- It is a food pattern in which food of vegetable origin and moderate consumption of fermented dairy and fish predominate in addition to low or occasional consumption of meat.
- MD is a healthy dietary pattern due to the presence of a large number of bioactive compounds (antioxidants, anti-inflammatory, etc.) in the food included. These bioactive compounds have a positive impact on the prevention of chronic noncommunicable diseases (cardiovascular, cancer, type 2 diabetes, obesity)
- MD is a sustainable food pattern that respects biodiversity and the environment.

Introduction

It is important to analyze and define the two terms that we will use in this article. "Diet" and "Mediterranean". Although we all agree that the Mediterranean Diet (MD) in the last century, in which the term was coined, was identified more by its nutritional and health aspects, subsequently, factors related to the culture and heritage of the peoples surrounding the Mediterranean Sea have been included. It is now considered a sustainable food pattern, as we will see below. The term diet is defined as: "food and drink regularly provided or consumed" or "habitual nourishing," other meanings related to medical and health aspects are also included: "the kind and amount of food prescribed for a person or animal for a special reason" or "a regimen of eating and drinking sparingly so as to reduce one's weight" (https://www.merriam-webster.com). However, etymologically, the word Diet comes from the Latin "diaeta" which in turn comes from the Greek "diaita" (δίαιτα), "way of living," "life regime," "type of life," we now know it as a lifestyle (CIHEAM, 2012).

The UNESCO, in its declaration of the MD as an Intangible Heritage of Humanity, includes the following: "… *The Mediterranean diet involves a set of skills, knowledge, rituals, symbols, and traditions concerning crops, harvesting, fishing, animal husbandry, conservation, processing, cooking, and particularly the sharing and consumption of food. Eating together is the foundation of the cultural identity and continuity of communities throughout the Mediterranean basin. It is a moment of social exchange and communication, affirmation, and renewal of family, group, or community identity. The Mediterranean diet emphasizes values of hospitality, neighborliness, intercultural dialogue and creativity, and a way of life guided by respect for diversity. It plays a vital role in cultural spaces, festivals, and celebrations, bringing together people of all ages, conditions and social classes …*" (UNESCO, 2010).

Taking into account the above, we will include in this article all aspects of the MD such as its history and development, the foods constituting the MD, the food consumption pattern of the population, as we can know the adherence to MD, what are the health effects of this type of food and what is its scientific basis and what are the characteristics that make it a sustainable diet.

Origin and historical development of Mediterranean food

The type of food of the peoples bordering the Mediterranean Sea (etymologically, "sea in the center of the earth") since the beginnings of civilization in the eastern Mediterranean is considered the origin of the MD in its broadest sense. The strategic position of the Mediterranean Sea, situated between three continents, Europe, Africa, and Asia, the peoples that inhabit its shores, and the different cultures that sailed it constitute the seed of this style of feeding. In addition, from the 16th century, America entered this Mediterranean crucible through food that arrived from that continent.

Although we speak of an MD, it would be more appropriate to talk about Mediterranean diets because if we take a tour of the history of Mediterranean food, we will see that these patterns are not homogeneous throughout the Mediterranean geographical territory. Different cultures, traditions, and religious feelings exist in this area that model food patterns from diverse landscapes and sensitivities. Throughout history, various peoples and civilizations, such as Egyptians, Phoenicians, and Carthaginians, Mesopotamians, Greeks, Romans, Arabic (Islamic), Ottomans, etc., have contributed to this pattern of Mediterranean food and Mediterranean cuisine. If we travel through the Mediterranean, we can find a mixture of cultures in different regions, such as the presence of the Roman Empire in the Middle East, Israel and Lebanon, remnants of prehistoric cultures in Sardinia, Greek cities in Sicily, Arab presence in Spain or Ottoman Islam in Yugoslavia. This crossbreeding of cultures explains the richness of food and cuisine in what we call MD (CIHEAM, 2012).

The so-called Mediterranean triad (vine, olive, and wheat) already appears in Hosea's writings in the Bible, in the Old Testament: "*She did not acknowledge that it was I who gave her the wheat, the new wine, and the fresh oil; who gave her the silver and the gold which they used for Baal*" (Hosea 2:8). These three foods are present throughout the Mediterranean, especially the olive tree, as its area of cultivation coincides with the Mediterranean basin from the Middle East (Israel, Lebanon) to Spain, from northern Italy to Tunisia or Egypt. Wheat appeared as a crop in the Neolithic with ancestral varieties and is started to be processed to facilitate its consumption in the Palestinian area and later in Mesopotamia and Egypt. It thereafter constitutes the leading food in Greek and Roman culture. The origins of the olive tree are located in Asia Minor, Crete, and Cyprus, and its cultivation spread throughout the Mediterranean thanks to Phoenicians, Greeks, and Romans. Vine and its product (wine) were already cultivated and consumed in Egypt, also appeared in Greek celebrations. Concerning beer, there is also evidence of its consumption in Egyptian civilization.

From this traditional triad as the backbone of this Mediterranean pattern, other foods have formed its structure that we will analyze later. Nuts are a fundamental part of the MD, consumed since prehistoric times (almonds and pistachios) and have continued to be consumed by different Mediterranean civilizations. Other characteristic and widely consumed foods in the Mediterranean area are the fruits and vegetables that have flooded the Mediterranean cuisine, such as artichokes, eggplants, asparagus, figs. We must not forget legumes (lentils, chickpeas): Also, the inclusion in the Mediterranean pattern of the foods that arrived from America in the 16th century and that today are characteristic of the food and cuisine of the Mediterranean basin; potatoes to a lesser extent, and especially tomatoes and peppers: The latter two, in conjunction with onions, are the base of the essential "sofrito" (sautéed or braized in cooking oil) in the Mediterranean cuisine.

It is important to mention the role of herbs and spices in feeding in the Mediterranean peoples. Garlic, saffron, aniseed, rosemary, sesame, myrrh, dill, coriander, thyme, etc., are some of the most representative and incorporated in the early moments of Mediterranean history. Herbs and spices have always been used generously in Mediterranean cuisine, and the Greeks also used them for medicinal purposes (CIHEAM, 2012).

Origin of the term Mediterranean diet and recognition of the Mediterranean diet as an Intangible Cultural Heritage of Humanity by UNESCO

The term MD began to be used in the middle of the 20th century, referring to the food pattern of some areas of the Mediterranean basin such as Crete and other Mediterranean islands, southern Italy and Spain. Ancel Keys, an American physiologist from the University of Minnesota, described it for the first time during the development of The Seven Countries Study (SCS), a project that sought to understand socio-cultural influences, including dietary patterns, on the prevalence of cardiovascular disease in the population. SCS was the first study to systematically compare the rates of cardiovascular disease in different countries with very different cultures. This ecological study established correlations between lifestyles in these countries and cardiovascular risk factors and between cardiovascular risk factors and disease over time. The countries chosen were the United States, Finland, Holland, Italy, Greece, Yugoslavia, and Japan (Toshima et al., 1995).

The study results showed significant differences between countries in the incidence of cardiovascular diseases, including ischemic coronary disease, especially between the United States or Finland, with high rates of this type of disease, and Greece and Italy, which together with Japan, had a significantly lower incidence.

Comparing the diet of countries with a low and high incidence of cardiovascular diseases, for example, Greece, and particularly the island of Crete, a Mediterranean country, and Finland, a northern European country, respectively, they found significant differences in the consumption of some food groups. In Crete, the population consumed more fruit than Finnish (462–464 g/person/day vs. 34–40 g/person/day), more vegetables (191 g/person/day vs. 104–108 g/person/day) and legumes (1–8 g/person/day vs. 30 g/person/day). Another important difference was the higher milk consumption in the Nordic country (1090–11,192 g/person/day) than in Crete (70 g/person/day), where more fermented dairy products such as cheese or yogurt were consumed. The consumption of sugary products followed the same pattern (Crete:13–20 g/person/day vs. Finland: 91–101 g/person/day). Perhaps the most

striking finding is that the consumption of edible fat (added and for cooking) in both countries was similar (75–95 g/person/day in Crete vs. 72–96 g/person/day in Finland).

In the middle of the 20th century, dietary fat began to be related to blood cholesterol levels, which are associated with the development of cardiovascular diseases. In this case, the dietary fat consumption was quantitatively similar between a Mediterranean country and a Nordic country. However, there was a large difference between them from a qualitative point of view because the added fat and cooking fat in Finland was mostly of dairy origin. In contrast, the fat used in the Mediterranean country was extra-virgin olive oil, extracted from the olive, the fruit of the olive tree, by using only physical means. The differences between the two types of fat lie in their fatty acid profile. While milk fat is rich in saturated fatty acids, extra virgin olive oil is rich in monounsaturated fatty acids, specifically, oleic acid (C18:1 *n*-9). According to the Keys studies, there was a relationship between dietary saturated fatty acid consumption and higher blood cholesterol levels (Keys et al., 1965).

The terms Mediterranean diet and Mediterranean lifestyle did not appear in a scientific journal but in a cookbook published by Ancel Keys and his wife Margaret entitled "How to Eat Well and Stay Well the Mediterranean way" (Keys and Keys, 1975) in which they collected the recipes and customs of their neighbors during the period they lived in Italy, in the Cilento region where this Mediterranean pattern was maintained. In this book, the Keys (a husband-and-wife team) gathered the main characteristics of MD, its frugality, seasonality, local biodiversity, and foods common in its diet, as we detail below.

Unfortunately, in the last two decades of the last century and the first decades of the present century, Mediterranean food habits have been lost, especially in the younger population due to globalization and Western lifestyles, the loss of culinary culture, and working hours. Distancing from MD has led the population of the Mediterranean countries to an increase in body weight, involving a high prevalence of overweight and obesity with the highest prevalence (above 40%) observed in countries of the Mediterranean area, such as Cyprus, Greece, Malta, Italy and Spain (WHO, 2021). In addition, it has led to a higher prevalence of cardiovascular diseases than those described in the 1960s in these Mediterranean countries (particularly Greece and Italy). These changes have been the result of industrialization processes, migration to urban areas from rural areas, sedentary behaviors at work and in leisure (TV, computers, smartphones, etc.) and reduced physical activity, increased availability of cheap and processed foods in supermarkets and grocery stores with high total and saturated fat content, salt, and added sugars, thus increasing energy intake and lower consumption of vegetable proteins and fats. These fats have been replaced by animal fats, providing less fiber and micronutrients (minerals and vitamins) because of their low nutritional density (Russo et al., 2021).

The current situation in adherence to the Mediterranean food pattern, and more globally, the Mediterranean lifestyle, is not very encouraging. In this context, the UNESCO, thinking more in the aspects most related to the culture and heritage of the MD than just about the nutritional and health benefits and the typical foods and dishes of Mediterranean cuisine, initiated an inscription to declare the MD as an Intangible Cultural Heritage of Humanity. This process began in 2004 as a civil society initiative driven by the Mediterranean Diet Foundation with the endorsement of the Ministry of Agriculture, Fisheries and Food of Spain, the trustee of this foundation that acted as a link between the civil society initiative and the institutions. At this first moment, Italy was involved, and subsequently Greece and Morocco. The confirmation of the DM as Intangible Heritage of Humanity took place in November 2010. In 2012, it was recognized by FAO as one of the most sustainable diets on the planet. Other Mediterranean countries such as Portugal, Croatia, and Cyprus joined this UNESCO declaration in 2013, and it is open to the future incorporation of other Mediterranean countries (Preedy and Watson, 2020; Russo et al., 2021).

Today, several initiatives have been proposed to ensure the future of MD in Mediterranean countries, including new technologies that make Mediterranean food, recipes, and menus more affordable than processed foods. Additionally, to initiate campaigns that bring together young scientists and allow synergistic projects on the characteristics of MD and its dissemination. A third aspect is a commitment of nutritionists, public health experts, and the food industry to promote traditional MD. Finally, it is important to mention that the MD is sustainable when communicating its positive aspects (Russo et al., 2021).

Structure and characteristics of the Mediterranean diet

The MD is based on a food pattern that has been graphically represented in the MD pyramid (Serra-Majem et al., 2020) (**Fig. 1**). It has the following characteristics:

(1) It is a diet rich in plant products, including the following food groups:
   a. Cereals, mainly whole grain wheat and its derivatives (bread, pasta, couscous, etc.), are the predominant cereal in the Mediterranean basin. Although oats, barley, and rice are also consumed, they are unprocessed or poorly processed. Wheat and other cereals are consumed in very different culinary forms and prepared together with different foods from other groups in the traditional Mediterranean dishes.
   b. Seasonal and locally grown fruits and vegetables in salads, gazpachos, vegetable stews, fruit salads, etc. Garlic, olives, onions, herbs, and spices (oregano, thyme, basil, rosemary, etc.).
   c. Legumes (lentils, chickpeas, beans, etc.). They are the base of stews and salads.
   d. Nuts (almonds, nuts, hazelnuts, chestnuts, etc.).
(2) Moderate-high consumption of white and oily fish and seafood, especially in the coastal areas of the Mediterranean basin.
(3) Milk and dairy consumption are moderate, with fermented derivatives such as cheese and yogurt always predominant.

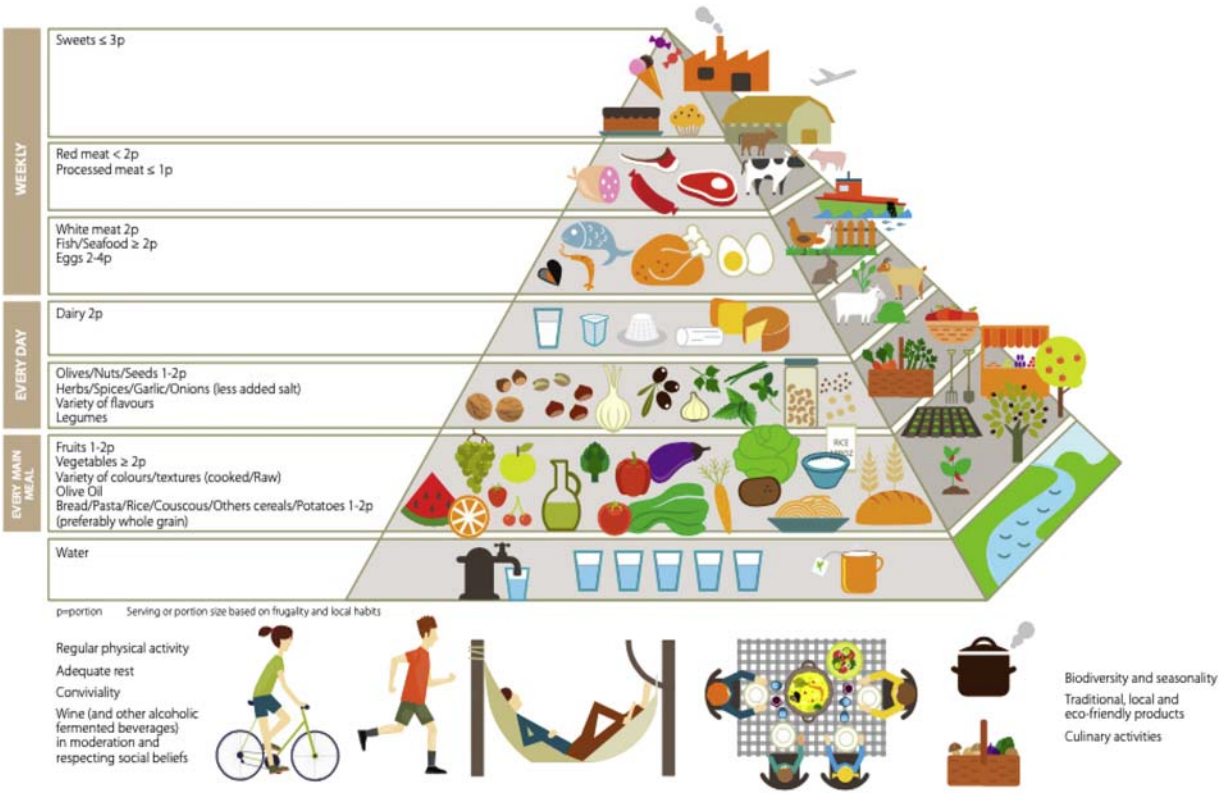


Fig. 1 Mediterranean diet pyramid. Reproduced from Serra-Majem et al. (2020).

(4) Consumption of meat and products of animal origin is occasional (low), with white meat (poultry, rabbit) predominating compared to red and processed meat (beef, lamb, pork, cold cuts/meats).
(5) Consumption of added simple sugars is also limited (pastries, candies, sugary drinks).
(6) Virgin olive oil is the predominant added fat used as a dressing for green salads and cooking (sofrito, stews, soups).

A fundamental aspect, as it appears in the pyramid of the MD (Fig. 1), is correct hydration by consuming adequate amounts of liquids, especially water and non-caloric infusions (tea, coffee, chamomile tea, pennyroyal mint, etc.), avoiding other liquids with high energy value because of added sugars.

Although not included in the pyramid, this food pattern is accompanied by other features that are also part of its overall structure, such as maintaining biodiversity by supporting the production of traditional and environmentally friendly local varieties and seasonality in their consumption. The use of traditional culinary techniques is also included in this food pattern and reduces the consumption of industrially processed foods. This makes MD a varied diet with a considerable wealth of recipes and dishes that present variations in each of the geographical regions surrounding the Mediterranean Sea.

Within the MD, it is also envisaged the consumption by adults of fermented low-alcohol beverages (wine, especially red wine, and beer), always with moderation and accompanying meals, always respecting social beliefs and religious precepts. We should not forget the hedonic aspect of food, the palatability of food associated with the response of each individual to the orosensory properties (taste, smell, flavor, texture) of food. MD is highly palatable.

In the social aspect, we must mention commensality and conviviality, eating together, sharing the table with others and enjoying during this period, consider food as a moment of social and cultural relationships. This must be complemented by avoiding sedentarism, performing regular physical activity, and adequate rest (8 h of sleep and a nap).

An important characteristic of MD is frugality, moderate in the quantitative aspect so that the caloric consumption is low and always adequate to the physical activity performed. Furthermore, the predominance of plant foods helps lowering energy consumption by having a low caloric density. All this is compatible with the maintenance of adequate body weight. This food profile of MD gives it particular nutritional characteristics due to the nutrient content and other components of the food consumed.

In general, MD is a balanced diet providing protein mainly of plant origin (legumes, whole grain cereals) (15–20% of total caloric intake), complex carbohydrates (35–40% of total caloric intake), and primarily unsaturated fat (35–45% of total caloric intake).

The high consumption of vegetables and fruits, nuts, legumes, and whole-grain cereals involves a high intake of dietary fiber and complex carbohydrates, implying a low glycemic index. Additionally, this high consumption of foods of plant origin provides a large number of micronutrients, minerals, and vitamins besides other food compounds with antioxidant and anti-inflammatory activity

such as polyphenols and other phytochemicals (allicin, isothiocyanates, etc.) with preventive properties against chronic noncommunicable diseases (cardiovascular diseases, cancer, diabetes, etc.) as discussed below.

Regarding the fatty profile of the MD, consumption of virgin olive oil as a predominant added fat involves a high intake of monounsaturated fatty acids (oleic acid; more than 50% of total fat). Moreover, fish consumption, especially oily fish, provides very long-chain polyunsaturated fatty acids of the omega 3 series such as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA). Nuts also provide monounsaturated fatty acids (oleic acid) and omega 3 polyunsaturated fatty acids (α-linolenic acid). In contrast, the low consumption of red meat and moderate consumption of fermented dairy products decreases the intake of saturated fatty acids and "trans" fatty acids, which are provided especially by low-consumption of processed foods in the Mediterranean food pattern. The low consumption of processed foods also reduces the intake of salt and added sugars.

Adherence to the Mediterranean diet

In order to know the adherence to the MD, it is necessary to have tools to inform us about the adherence of the population to this food pattern. This adherence can help us relate this type of diet to its health properties.

Several indexes and scores have been developed to establish adherence to MD, taking into account the frequency of consumption of certain food groups characteristic of MD and the intake of certain nutrients that identify MD. Thus, we have the Mediterranean Diet Score (MDS) and the modified MDS in which the adherence to the MD is classified according to the fact that the consumption of legumes, cereals, fruits, nuts, and vegetables is above the median (score 1) and the consumption of meat and derivatives, milk and dairy products and alcohol is below the median (score 1); otherwise, the value assigned to the MDS is "0". The intake of monounsaturated fatty acids is also included (score 1 if above the median and score "0" if below the median); this indicates the consumption of virgin olive oil. The difference between MDS and the modified MDS is the inclusion of fish in the latter and the ratio between the intake of monounsaturated and saturated fatty acids rather than monounsaturated fatty acids. Consumption scores are specified for alcohol according to sex.

Subsequently, new indexes have been developed, such as the Mediterranean Diet Quality Index (Med-DQI), which includes the consumption of some food groups (meat, olive oil, fish, vegetables, and fruits) and nutrients such as saturated fatty acids and cholesterol, setting consumption intervals in g/day or percentages of total dietary energy (**Table 1**). The Mediterranean Diet Adherence Screener (MEDAS) is an index developed within the PREDIMED study that assesses the number of servings consumed daily from different food groups and some typical recipes from Mediterranean cuisine.

A Mediterranean lifestyle index (MEDLIFE) has also been developed, which not only assesses adherence to the Mediterranean food pattern through the consumption of different food groups but also considers aspects related to dietary habits, physical activity and rest, and social aspects (**Table 1**).

A MD adherence questionnaire has also been developed specifically for children and adolescents, the KIDMED, which assesses the responses to questions about the consumption and acceptance of different foods and food groups and scores them positively or negatively (See **Table 1** for details) (Gil et al., 2015).

The Mediterranean diet as a healthy food pattern

The results of the Seven Countries Study provided scientific evidence on the role of MD in the prevention of cardiovascular diseases. Numerous epidemiological studies and clinical trials have shown that adherence to a traditional Mediterranean food pattern is associated with the prevention of different chronic noncommunicable diseases (cancer, type 2 diabetes, metabolic syndrome) and is also associated with a lower incidence of obesity. It has also been stated that adherence to MD is inversely related to some neurodegenerative diseases related to the aging process, such as Alzheimer's disease and vascular dementia, delaying its appearance. Other healthy properties of the MD have been published in recent years, such as its anti-inflammatory activity, stimulation of immune function, beneficial for mood disorders, such as depression. In general, adherence to the Mediterranean food pattern is consistent with lower mortality from any cause and cardiovascular diseases and is related to a better quality of life (Preedy and Watson, 2020).

The leading cause of death globally is cardiovascular disease (ischemic coronary disease and stroke). The PREDIMED project (Estruch et al., 2018) was conducted in a cohort of more than 7000 adult individuals with high cardiovascular risk divided into three groups: (1) MD + extra virgin olive oil; (2) MD + nuts; and (3) control group with a low-fat diet. At the end of approximately 5 years of follow-up and analyzing the occurrence of cardiovascular events, including death from these causes, the study concluded that the incidence of these events was lower in those adhering to the MD, supplemented with extra virgin olive oil or with nuts, compared to those who had consumed a low-fat diet. This lower incidence was observed for ischemic coronary disease, stroke, death from cardiovascular causes, and total mortality in the group supplemented with extra virgin olive oil especially. Today, the PREDIMED study is continued in the PREDIMEDplus, which sets objectives related to Mediterranean lifestyles with an intensive intervention that includes a Mediterranean hypocaloric diet, the performance of regular physical activity, and behavioral intervention. The results to be monitored are cardiovascular events and the decrease in body weight and its maintenance over time.

Studies regarding the relationship between cancer and the degree of adherence to MD show a lower risk of death for higher adherence to MD. The risk is low for different types of cancer, including colorectal, head and neck, respiratory, gastric, liver, and bladder (Morze et al., 2021).

Table 1 Scores and indexes of adherence to the Mediterranean diet.

Index	*Components*	*Criteria*	*Scoring*
Mediterranean diet score (MDS)	Monounsaturated: saturated fatty acid ratio	<Median	0
		>Median	1
	Legumes	<Median	0
		>Median	1
	Cereals	<Median	0
		>Median	1
	Fruit and nuts	<Median	0
		>Median	1
	Vegetables	<Median	0
		>Median	1
	Meat and meat products	>Median	0
		<Median	1
	Milk and dairy products	>Median	0
		<Median	1
	Alcohol	>Median	0
Mediterranean diet score modified (MED mod)	Vegetables	<Median	0
		≥Median	1
	Legumes	<Median	0
		≥Median	1
	Fruits and nuts	<Median	0
		≥Median	1
	Dairy products	<Median	0
		≥Median	1
	Cereals	<Median	0
		≥Median	1
	Meat	≥Median	0
		<Median	1
	Fish	<Median	0
		≥Median	1
	Ratio monounsaturated:saturated	<Median	0
		≥Median	1
	Alcohol	Men (10–50 gd)	1
		Women (5–25 g/d)	1
Mediterranean diet quality index (med-DQI)	Saturated fatty acids	<10 energy %	0
		10–13 energy %	1
		>13 energy %	2
	Cholesterol	<300 mg	0
		300–400 mg	1
		>400 mg	2
	Meat	<25 g	0
		25–125 g	1
		>125 g	2
	Olive oil	>15 mL	0
		15–5 mL	1
		<5 mL	2
	Fish	>60 g	0
		60–30 g	1
		<30 g	2
	Cereals	>300 g	0
		300–100 g	1
		<100 g	2
	Vegetables + fruits	>700 g	0
		700–400 g	1
		<400 g	2

(Continued)

Table 1 Scores and indexes of adherence to the Mediterranean diet.—cont'd

Index	*Components*	*Criteria*	*Scoring*
Mediterranean diet adherence screener (MEDAS)	4 or more tablespoons of olive oil/d		1
	2 or more servings of vegetables/d		1
	3 or more pieces of fruit/d		1
	<1 serving of red meat or sausages/d		1
	<1 serving of animal fat/d		1
	<100 mL of sugar-sweetened beverages/d		1
	7 or more servings of red wine/wk		1
	3 or more servings of pulses/wk		1
	3 or more servings of fish/wk		1
	Fewer than 2 commercial pastries/wk		1
	3 or more servings of nuts/wk		1
	2 or more servings/wk of a dish with a traditional sauce of tomatoes, garlic, onion, or leeks sautéed in olive oil		1
	Use of olive oil as principal source of fat		1
	Kind of meat preferable consumed		0–1
Mediterranean lifestyle index (MEDLIFE)			
Block 1: mediterranean food consumption	Sweets	≤2 servings/week	1
	Red meat	<2 servings/week	1
	Processed meat	≤1 servings/week	1
	Eggs	2–4 servings/week	1
	Legumes	≥2 servings/week	1
	White meat	2 servings/week	1
	Fish/seafood	≥2 servings/week	1
	Potatoes	≤3 servings/week	1
	Low-fat dairy products	2 servings/week	1
	Nuts and olives	1–2 servings/d	1
	Herbs, spices and garnish	≥1 servings/d	1
	Fruits	3–6 servings/d	1
	Vegetables	≥2 servings/d	1
	Olive oil	≥3 servings/d	1
	Cereals	3–6 servings/d	1
Block 2: mediterranean dietary habits	Water or infusions	6–8 servings/d or ≥3 servings/week	1
	Wine	1–2 servings/d	1
	Limit salt in meals	Yes	1
	Preference for whole grain products	Yes/fiber >25 g/d	1
	Snacks	≤2 servings/week	1
	Limits nibbling between meals	Yes	1
	Limits sugar in beverages (including sugar-sweetened beverages)	Yes	1
Block 3: physical activity, rest, social habits and conviviality	Physical activity (>150 min/week or 30 min/d)	Yes	1
	Siesta/nap	Yes	1
	Hours of sleep	6–8 h/d	1
	Watching television	<1 h/d	1
	Socializing with friends	≥2 h/weekend	1
	Collective sports	≥2 h/week	1

Modified from Gil et al. (2015).

The prevalence of type 2 diabetes in the world is increasing. Studies associating the adherence to a Mediterranean food pattern and the risk of developing type 2 diabetes show an inverse relationship. This lower risk of disease can be related to the different components of the MD, especially the anti-inflammatory and antioxidant bioactive compounds in fruits, vegetables, virgin olive oil, and other foods typical of MD. It is important to note that it is necessary to consider the role of MD and type 2 diabetes within the so-called healthy Mediterranean lifestyle mentioned above (Martin-Pelaez et al., 2020).

Another critical aspect of the MD-health relationship is the role of MD in aging and all associated pathological disorders that are among the leading causes of mortality in today's world. We have already mentioned some diseases such as cardiovascular disease, cancer, or type 2 diabetes. Neurodegenerative diseases such as Alzheimer's disease and other senile dementias also appear in this aging process. Recent systematic reviews have shown that adherence to the MD positively affects different cognitive aspects of the aging (elderly) population. In addition, it also has benefits in mood disorders such as depression). Regarding Alzheimer's disease, the MD may reduce different markers of the disease, such as the deposition of the beta-amyloid peptide and the relationship

between this peptide, the tau protein, and brain atrophy; this would lead to slowing and improving neurodegeneration and associated cognitive impairment (Limongi et al., 2020).

Some studies have been published that relate the MD pattern to the coronavirus disease 2019 (COVID-19). This relationship is established through the positive effects on risk conditions for this viral infection, such as cardiovascular diseases, obesity, and type 2 diabetes, that is related to the severity of the disease. Other authors postulate that adherence to MD may be related to the immune system, inflammation, and thrombus formation (Suardi et al., 2021).

There are numerous systematic review and meta-analysis studies on the positive effects of the MD, the Mediterranean food pattern, and the prevention of many chronic diseases (Dinu et al., 2018). All these findings can be explained based on different mechanisms involved and related to the dietary pattern of MD and the anti-inflammatory, antioxidant, antithrombotic, antineoplastic, immunomodulatory, or lipid-lowering properties of the bioactive compounds of the MD (EPA, DHA, phytosterols, dietary fiber, polyphenols, etc.). All of them have a positive role in lipid metabolism, oxidative stress, inflammation, endothelial function, peripheral tissue insulin resistance, hemostasis (platelet aggregation and coagulation), and less established processes such as autophagia, intestinal microbiome, cellular telomerase activity or nutrigenomic and epigenomic aspects (Schröder et al., 2020).

All of these studies performed to understand the relationship of the MD pattern to health prevention and maintenance have led to the inclusion of MD in Food Guidelines of many countries, including the USA since 2010, as a healthy dietary pattern along with others such as the vegetarian diet or the DASH diet (USDA, 2020).

The Mediterranean diet as a sustainable food pattern

FAO in its 2010 International Scientific Symposium on Sustainable Diets and Biodiversity. Directions and solutions for policy, research, and action define: "*Sustainable Diets are those diets with low environmental impacts which contribute to food and nutrition security and to healthy life for present and future generations. Sustainable diets are protective and respectful of biodiversity and ecosystems, culturally acceptable, accessible, economically fair and affordable; nutritionally adequate, safe and healthy; while optimizing natural and human resources*" (FAO, 2012).

One of the sections in this publication is dedicated to MD as an example of a sustainable diet. Among the arguments mentioned are its immense diversity, which guarantees its nutritional quality, as discussed above, and the maintenance of biodiversity with a commitment to local population development. The Mediterranean gastronomy has an extraordinary richness that makes the resulting dish with the same ingredients have its own personality in the different regions of the Mediterranean basin to contribute to the diversity of its landscapes. The role of cultures and traditions in the Mediterranean basin concerning eating habits (conviviality, commensality) is also an asset in its sustainable character. As a consequence of the above, the Mediterranean food system respects nature and practices the seasonality of the usually consumed products. Finally, its healthy character also makes it sustainable (Truzzi et al., 2020).

Another study states that the indicators that can evaluate MD as a sustainable food pattern can be framed in four areas: nutrition and health, environmental (carbon and water footprint and biodiversity, etc.), economic through the accessibility and development of local economies and the cultural and social area, especially the consumption of traditional products with designation of origin and protected geographical indication (Dernini et al., 2017).

The future of the Mediterranean diet

As mentioned above, the current situation of adherence to the MD in the Mediterranean countries is not at its best, especially among the youngest population. Food globalization, the pace of life, lack of time for cooking, convenience in access to ready-to-eat processed foods, and high palatability have led to a progressive abandonment of this food pattern. This has led to a negative change in the prevalence of food-related noncommunicable diseases and an increase in overweight and obesity in the Mediterranean population, particularly in the countries of the Northern shore.

This scenario raises what the future of this Mediterranean pattern, of the MD, maybe. Some authors suggest implementing strategies that act on the price and affordability of these Mediterranean foods by using new culinary and industrial technologies that provide easier access to traditional Mediterranean foods. It is also important to continue researching in this field to understand better the positive effects of the MD on both individual and planet health. Another aspect to consider is how to disseminate the MD to other non-Mediterranean parts of the world. In this sense, MD should not be exported but should try to reproduce in each of the territories by using the philosophy of this pattern, not only food but cultural and related to heritage. Use local food and the traditional way of preparing it, maintain native varieties, and therefore biodiversity, consider food a collective and social event within the family and social environment (Russo et al., 2021).

Summary page

In conclusion, MD is a food pattern that includes aspects related to the lifestyle of the population. These aspects are related to agricultural practices that allow for environmentally-friendly cultivation and maintenance of local species and biodiversity, traditional

culinary practices, and commensality and conviviality at the table. The MD is a healthy food pattern that includes foods typical of Mediterranean landscapes, with a high content in foods of plant origin (whole grain cereals, vegetables, fruits and nuts, and legumes), providing large amounts of micronutrients and other food compounds with high biological activity. Moreover, consumption of animal foods is low, especially red and processed meats, although it includes moderate consumption of fish and fermented dairy products. A backbone characteristic of the MD is extra virgin olive oil as added fat for dressing and cooking. It is an unsaturated fat with many bioactive compounds beneficial for health that, moreover, it provides high palatability to Mediterranean cuisine with worldwide fame. The maintenance of the Mediterranean lifestyle pattern, the MD, in its broadest sense, today compromised, is a commitment undertaken by the countries surrounding this sea after its recognition as the intangible heritage of humanity by the UNESCO. This will affect the health of Mediterranean populations and the sustainability of the planet.

References

CIHEAM, 2012. Mediterra. The Mediterranean Diet for Sustainable Regional Development/International Centre for Advanced Mediterranean Agronomic Studies (CIHEAM). Presses de Sciences Po, Paris.

Dernini, S., Berry, E.M., Serra-Majem, L., La Vecchia, C., Capone, R., Medina, F.X., et al., 2017. Med Diet 4.0: the Mediterranean diet with four sustainable benefits. Publ. Health Nutr. 20 (7), 1322–1330.

Dinu, M., Pagliai, G., Casini, A., Sofi, F., 2018. Mediterranean diet and multiple health outcomes: an umbrella review of meta-analyses of observational studies and randomized trials. Eur. J. Clin. Nutr. 72 (1), 30–43.

Estruch, R., Ros, E., Salas-Salvadó, J., et al., 2018. Primary prevention of cardiovascular disease with a Mediterranean diet supplemented with extra-virgin olive oil or nuts. N. Engl. J. Med. 378 (25), e34.

FAO (Food and Agriculture Organization of the United Nations), 2012. Bioversity international. In: Burlingame, B., Dernini, S. (Eds.), Sustainable Diets and Biodiversity, vol. 309.

Gil, A., Martinez de Victoria, E., Olza, J., 2015. Indicadores de evaluación de la calidad de la dieta. Nutr. Hosp. 31, 128–144.

Keys, A., Anderson, J.T., Grande, F., 1965. Serum cholesterol response to changes in the diet: IV. Particular saturated fatty acids in the diet. Metabolism 14, 776–787.

Keys, A.B., Keys, M., 1975. How to Eat Well and Stay Well. The Mediterranean Way. Doubleday, Garden City, N.Y.

Limongi, F., Siviero, P., Bozanic, A., Noale, M., Veronese, N., Maggi, S., 2020. The effect of adherence to the Mediterranean diet on late-life cognitive disorders: a systematic review. J. Am. Med. Dir. Assoc. 21 (10), 1402–1409.

Martín-Peláez, S., Fito, M., Castaner, O., 2020. Mediterranean diet effects on type 2 diabetes prevention, disease progression, and related mechanisms. A review. Nutrients 12, 2236.

Morze, J., Danielewicz, A., Przybyłowicz, K., et al., 2021. An updated systematic review and meta-analysis on adherence to Mediterranean diet and risk of cancer. Eur. J. Nutr. 60, 1561–1586.

Preedy, V.R., Watson, R.R. (Eds.), 2020. The Mediterranean Diet, second ed. Academic Press.

Russo, G.L., Siani, A., Fogliano, V., Geleijnse, J.M., Giacco, R., et al., 2021. The Mediterranean diet from past to future: key concepts from the second "Ancel Keys" international seminar. Nutr. Metabol. Cardiovasc. Dis. 31, 717–732.

Schröder, L., Morze, J., Hoffmann, G., 2020. Mediterranean diet and health status: active ingredients and pharmacological mechanisms. Br. J. Pharmacol. 177 (6), 1241–1257.

Serra-Majem, L., Tomaino, L., Dernini, S., Berry, E.M., Lairon, D., Ngo de la Cruz, J., Bach-Faig, A., Donini, L.M., Medina, F.X., Belahsen, R., Piscopo, S., Capone, R., Aranceta-Bartrina, J., La Vecchia, C., Trichopoulou, A., 2020. Updating the Mediterranean diet pyramid towards sustainability: focus on environmental concerns. Int. J. Environ. Res. Publ. Health 25, 8758.

Suardi, C., Cazzaniga, E., Graci, S., Dongo, D., Palestini, P., 2021. Link between viral infections, immune system, inflammation and diet. Int. J. Environ. Res. Publ. Health 18 (5), 2455.

Toshima, H., Koga, Y., Blackburn, H. (Eds.), 1995. Lessons for Science From the Seven Countries Study. Springer Verlag, Tokyo.

Truzzi, M.L., Ballerini Puviani, M., Tripodi, A., Toni, S., Farinetti, A., et al., 2020. Mediterranean diet as a model of sustainable, resilient and healthy diet. Prog. Nutr. 22 (2), 388–394.

U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2020. Dietary Guidelines for Americans, 2020–2025, ninth ed. Available at. DietaryGuidelines.gov.

UNESCO-United Nations Educational, Scientific and Cultural Organization, 2010. Decision of the Intergovernmental Committee: 8.COM 8.10. https://ich.unesco.org/en/decisions/8.COM/8.10. (Accessed 2 June 2021).

WHO, 2021. WHO European Childhood Obesity Surveillance Initiative (COSI): Report on the Fourth Round of Data Collection, 2015–2017. WHO Regional Office for Europe, Copenhagen. Licence: CC BY-NC-SA 3.0 IGO.

Micronutrient supplementation: Programmatic issues

Melissa F. Young[a], **Lynnette M. Neufeld**[b], **Sara Hendrix**[a], **and Usha Ramakrishnan**[a], [a] Hubert Department of Global Health, Emory University, Atlanta, GA, United States; and [b] Food and Nutrition Division, Food and Agricultural Organization of the United Nations (FAO), Rome, Italy

© 2023 Elsevier Ltd. All rights reserved.

This is an update of R.D.W. Klemm, Supplementation: Programmatic Issues, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 251–259, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00259-2.

Introduction 467
The purpose of micronutrient supplementation 468
Cost of micronutrient supplements 468
Summary of micronutrient supplementation recommendations 468
Supplementation program cycle 468
Key challenges and success factors for program implementation 473
Case studies 474
Research gaps and future directions 475
Summary 476
Conflict of Interest 477
References 477

Key points

- To review current micronutrient supplementation recommendations
- To discuss key challenges and recommendations for successful program implementation
- To identify research gaps and opportunities

Abbreviations

Hb Hemoglobin
IU International units
LBW Low birth weight
RDA Recommended daily allowance
WHO World Health Organization

Introduction

Globally almost over half of preschool-aged children and two-thirds of women are affected by deficiency of one or more micronutrient that put them at an increased risk of poor growth, morbidity, intellectual impairment, and mortality (Keats et al., 2021; Stevens et al., 2022b; Victora et al., 2021). High rates of anemia remain as a persistent global health problem, impacting over 269 million children and 571 women of reproductive age and 32 million pregnant women (Stevens et al., 2022a). Iron deficiency is one of the most prevalent micronutrient deficiencies globally and is considered the leading cause of nutritional anemia, although these estimates may vary widely by setting (Stevens et al., 2022a).

In comparison to children, less is known globally about the burden of maternal micronutrient deficiencies especially in Sub-Saharan Africa (Victora et al., 2021). However, it is well established that high nutrient demands place pregnant and lactating women at elevated risk of deficiency. Unfortunately, there is little or no data for most countries for school-age children, adolescents, adult men, and older adults (Victora et al., 2021).

Since the mid-1980s, micronutrient supplementation has been a public health strategy in many low and middle-income countries (LMIC) to prevent and control micronutrient deficiencies for infants and young children, pregnant women, and in some contexts, non-pregnant women of reproductive age. The following discussion will provide an overview of the purpose of

micronutrient supplementation, its role in the prevention and control of micronutrient deficiencies. We will briefly examine the evidence for supplementation with individual nutrients including vitamin A, iron, iodine, zinc, calcium, vitamin D, and with multiple micronutrients, and various nutrient preparations (tablets, powders, lipid-based nutrition supplements) with respect to efficacy, recommended dose, and frequency of administration, safety, and program effectiveness. Program effectiveness at scale remains a challenge in many settings and the majority of this article is dedicated to the review of key challenges and recommendations for successful program implementation.

The purpose of micronutrient supplementation

Supplementation may serve one of two objectives: (1) to provide physiological amounts (i.e., a portion of or up to daily recommended doses) of nutrients that otherwise are lacking the diets of the targeted population group as a preventive approach; or (2) to provide therapeutic doses of nutrients in response to diagnosed medical conditions (e.g., treatment of severe anemia) or to achieve proven nutrition/health outcome in specific contexts (e.g., reduced mortality in populations with high prevalence of vitamin A deficiency among children).

Generally, micronutrient supplementation is intended as a short-term means of rapidly preventing and controlling nutrient deficiencies in high-risk individuals and populations until adequate and evidence that food supplies meet those needs. However, in some population sub-groups with high micronutrient needs, supplementation may be required through the duration of that life stage (e.g., pregnancy, infancy).

Cost of micronutrient supplements

The 2008 Copenhagen Consensus a group of world–renowned economists, ranked micronutrient supplements (specifically high–dose vitamin A, and therapeutic zinc supplements for children with diarrhea) as among the top international development priorities. The criteria included the cost:benefit ratio, as well as feasibility and sustainability of the interventions. While promising, data to inform such analyses remain scarce. The cost of supplementation programs will depend on the scope and scale of the program, existing delivery mechanisms, the supplement product, among other considerations. Few studies have quantified the cost of supplementation programs, partly as result of their difficulty to estimate (Fiedler and Puett, 2015). More importantly, the cost and cost:-benefit of supplementation programs should be estimated in light of other alternatives to provide the needed nutrients to the targeted populations. We found only one such approach using optimization modeling to determine the best combination of interventions, including supplementation among others to meet nutrient needs in the short and longer term (Vosti et al., 2015).

Summary of micronutrient supplementation recommendations

There is an enormous body of evidence in the published literature on the impact of a variety of types and formulations of supplements on diverse outcomes in various population groups. To inform programs such evidence should be rigorously reviewed and consolidated to formulate guidelines that can then be adapted to context. A summary of the current World Health Organization and Lancet recommendations for micronutrient supplementation is provided in **Table 1** for vitamin A, iron, iodine, zinc, calcium, vitamin D, multiple micronutrient tablets, multiple micronutrient powders and lipid-based nutrition supplements. These recommendations vary by target group and setting depending on the country level prevalence in some circumstances.

As noted above, WHO recommendations are intended as global guidance on efficacious actions to address specific deficiency or health outcomes. It requires adaptation to context to translate this guidance into effective programs. There are several challenges to doing so. The WHO guideline development process relies on systematic reviews and meta-analyses of randomized controlled trials in the published literature. Recommendations can be made therefore, only within the bounds of the evidence that has been studied. This inevitably leads to some gaps in evidence, such as impact in specific subgroups, under varying levels of compliance, or other contextual considerations, in the extent to which such information may not have been included or data cannot be extracted to that level in the original trials. There has been a concerted efforts in the guideline development process to address this by bringing other evidence to bear (e.g., through the values and preferences, and additional considerations) (Pena-Rosas et al., 2012).

Supplementation program cycle

Well-designed programs (**Fig. 1**) should be informed by evidence of the existence of micronutrient deficiency in the population, gaps in existing approaches to address them, and other evidence that will ensure that there is potential to benefit. Ideally, such evidence would include the magnitude and distribution of deficiency, including data disaggregated by sub-groups in the population most at risk of deficiency. Program design should similarly be informed by evidence of gaps in existing micronutrient and complementary programs and resources available for procurement and delivery of supplements and all related activities. The choice of supplements should also be guided by acceptance and potential for high compliance among the target population.

Table 1 Summary of micronutrient supplementation recommendations.

Micronutrient	*Summary of benefit*	*Target group*	*Dosage*	*Frequency & duration*	*Strength of evidence*	*WHO level of recommendation/Details*	*Special considerations*
Vitamin A	Prevention of night blindness; lowered severity and mortality risk of infections; lowered risk of diarrhea and severe measles	Infants and children aged 6–59 months	6–11 months: 100,000 IU 12–59 months: 200,000 IU	6–11 mo: Once 12–59 mo: Every 4–6 months	Strong	Context-specific	Only recommended where the prevalence of night-blindness is 1% or more in children of 24–59 months, or the prevalence of vitamin A deficiency is 20% or higher in infants and children aged 6–59 months
	Prevention night blindness	Pregnant women	Daily: 10,000 IU Weekly: 25,000 IU	Daily or once weekly	Weak	Context-specific	Only recommended where vitamin A deficiency is a severe public health problem
Iron	Lowered risk of anemia, iron deficiency and iron deficiency anemia. Intermittent regimens specifically may lead to fewer side-effects than the daily regimen, increase adherence, and increase hemoglobin concentrations	Infants and children aged 6–23 months	10–12.5 mg of elemental iron	Daily for three consecutive months in a year	Strong	Context-specific	Only recommended where the prevalence of anemia in infants and young children under 2 years of age is 40% or more
Children aged 2–12 years		24–59 months: 30 mg of elemental iron 5–12 years: 30–60 mg of elemental iron	Daily for three consecutive months in a year	Strong	Only recommended the prevalence of anemia in children over 2 years of age is 40% or more		
		24–59 months: 25 mg of elemental iron 5–12 years: 45 mg of elemental iron	Once weekly	Moderate	Only recommended where the prevalence of anemia in children over 2 years of age is 20% or more		
Menstruating non-pregnant girls		30–60 mg of elemental iron	Daily for three consecutive months in a year	Strong	Only recommended where the prevalence of anemia in non-pregnant women is 40% or higher		
		60 mg of elemental iron	Once weekly	Moderate	Only recommended where the prevalence of anemia in non-pregnant women is 20% or higher		

(Continued)

Table 1 Summary of micronutrient supplementation recommendations.—cont'd

Micronutrient	*Summary of benefit*	*Target group*	*Dosage*	*Frequency & duration*	*Strength of evidence*	*WHO level of recommendation/Details*	*Special considerations*
Lowered risk of maternal anemia, puerperal sepsis, low birth weight, and preterm birth	Pregnant women	30–60 mg of elemental iron	Daily	Strong	Recommended Details: Should be a component of a combined Iron and Folic Acid supplement		
		120 mg of elemental iron	Once weekly	Moderate	Context-specific Details: Should be a component of a combined Iron and Folic Acid supplement	If daily iron is not acceptable due to side-effects, and in populations with an anemia prevalence among pregnant women of less than 20%	
Lowered risk of iron deficiency and anemia, as well as improved hemoglobin concentrations	Non-pregnant women (aged 15–59 years)	30–60 mg of elemental iron	Daily for three consecutive months in a year	Moderate	Context-specific Details: For weekly supplementation, 3 months of supplementation should be followed by 3 months of no supplementation, after which the provision of supplements should restart	Only recommended where the prevalence of anemia in non-pregnant women is 40% or higher	
		60 mg of elemental iron	Once weekly	Moderate		Only recommended where the prevalence of anemia in non-pregnant women is 20% or higher	
Iodine	Healthy brain development in the fetus and young child	Infants and children aged 6–23 months	Daily: 90 mcg Annual Iodized Oil: 200 mg	Daily or yearly (dependent on administration method)	Moderate	Context-specific	Only recommended where 20% or fewer households have access to iodized salt and complementary food fortified with iodine is not available
		Pregnant women	Daily: 250 mcg Annual Iodized Oil: 400 mg		Moderate		
		Non-pregnant women (aged 15–59 years)	Daily: 150 mcg Annual Iodized Oil: 400 mg		Moderate		

Zinc	Preventative: Lowered risk of diarrhea, lowered risk of upper respiratory tract infection, Increase of mean height	Children aged 6–59 months	Dosage to be determined through further research	Frequency and duration to be determined through further research	Moderate	N/A There are no WHO guidelines or recommendations provided	
	Therapeutic: Shortened duration of diarrhea, persistent diarrhea, and vomiting		20 mg	Daily for 10–14 days	Strong	Recommended Details: Only recommends zinc supplementation for treatment of children who already have diarrhea	Supplementation should be combined with increased fluids and continued feeding
	–	Pregnant women	–	–	Weak	Context-specific Details: Does not recommend zinc supplementation as part of routine antenatal care	Only recommended in the context of rigorous research
Calcium	Lower risk of pre-eclampsia, eclampsia, preterm births, and maternal mortality due to pre-eclampsia and eclampsia	Pregnant women	1.5–2.0 g elemental calcium	Daily	Moderate	Context-specific Details: To be administered orally	Only recommended in populations with low dietary calcium intake
Vitamin D	Oral vitamin D supplementation is not recommended for all pregnant women to improve maternal and perinatal outcomes.	Pregnant women	–	–	–	Not recommended Details: Not recommended generally, sunlight is the most important source of Vitamin D	When vitamin D deficiency is suspected, supplements of the current recommended nutrient intake of 200 IU may be given daily
Multiple Micronutrient Supplementation	May reduce risk of anemia, low birthweight, preterm birth, small for gestational age (SGA)	Pregnant women	Supplements that include 60 mg of elemental iron	–	–	Context-specific Details: Does not recommend in high-income countries or to populations not at risk of micronutrient deficiencies.	Only recommended in populations where anemia is a severe public health problem in the context of rigorous research Lancet classifies MMS as strong evidence for implementation.

(Continued)

Table 1 Summary of micronutrient supplementation recommendations.—cont'd

Micronutrient	*Summary of benefit*	*Target group*	*Dosage*	*Frequency & duration*	*Strength of evidence*	*WHO level of recommendation/Details*	*Special considerations*
Multiple Micronutrient Powders (MMP)	Reduces anemia, iron-deficiency anemia, iron-deficiency, and diarrhea. Includes logistical advantages over supplements (e.g. sachets easy to transport and store) and biological advantages (e.g. reduced severity and frequency of side-effects, a more "physiological" provision of nutrients with food)	Infants and children aged 6–23 months Children aged 2–12 years	1 Sachet *Sachet Composition: Iron*: 10–12.5 mg of elemental iron for children 6mo to 4 years, 12.5–30 mg of elemental iron for children aged 5–12 years *Vitamin A*: 6000 IU *Zinc*: 5 mg of elemental zinc with or without other micronutrients to achieve 100% of the RNI	90 sachets/doses over a 6-month period	Moderate	Context-specific Details: To be used as point-of-use fortification of foods, must be iron-containing MMPs	Only recommended in settings in which the prevalence of anemia in children under 2 years of age is 20% or more Only recommended in settings in which the prevalence of anemia in children aged 2–12 years of age is 20% or more
Lipid-based Nutrient Supplement (LNS)	Reduces risk of severe and moderate stunting, moderate wasting, and moderate underweight Preconception period: May reduce risk of small for gestational age (SGA) and stunting at birth Antenatal period: No benefits when compared to MMN, when compared to IFA there is a slight reduction in risk of SGA	Children aged 2–12 years Pregnant and lactating women	Per *The Lancet* Recommendation: Small Quantity: <120 kcal (Medium and Large Quantity are not a part of *The Lancet* recommendation)	Daily, duration to be determined through further research	Strong Weak	N/A Details: There are no WHO guidelines or recommendations provided	N/A

Summary of recommendations is adapted from the following references: Keats et al, 2021; WHO, 2016; WHO, 2019; WHO 2020a; WHO 2020b; WHO 2021.

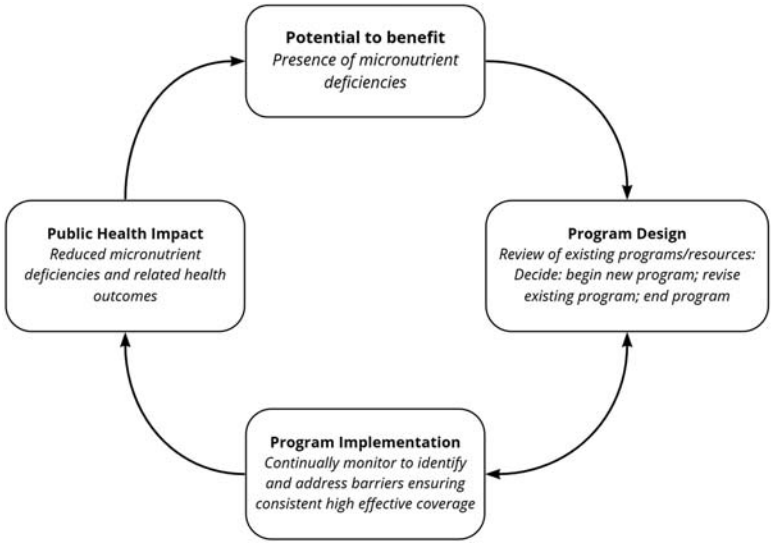


Fig. 1 Program cycle illustrating key considerations and data needs to enhance potential for impact.

Effective program implementation must consider all programmatic components from supplement specifications and procurement, to supply chain and stocks management, effective behavior change communication, among many other considerations. To ensure potential bottlenecks and barriers to implementation can be identified and resolved, a detailed impact pathway should be developed and used to identify the types of timely data needed and points for review and continual improvement. A schematic illustration of the use of such data to identify bottlenecks is shown in Fig. 2. The continual improvement of programs should be an iterative process reviewing data that may inform the need for modifications to design and/or implementation that maximize its potential for public health impact. For example, research to inform changes in supplements distributed as part of a social protection programs in Mexico provide a good example of program modifications to address design and implementation challenges (Neufeld et al., 2019). Evidence from context should also be used to identify the specific circumstances under which supplementation programs should be discontinued. Some progress has been made to identify such considerations, including biological potential for impact change (i.e., reduced prevalence of deficiency) notably for vitamin A supplementation programs (De Pee et al., 2011).

Key challenges and success factors for program implementation

Despite evidence of efficacy, rigorous guidelines and political support for supplementation programs, successful implementation and impact remains elusive in many regions across the globe (Global Nutrition Report, 2021; Victora et al., 2021). For many

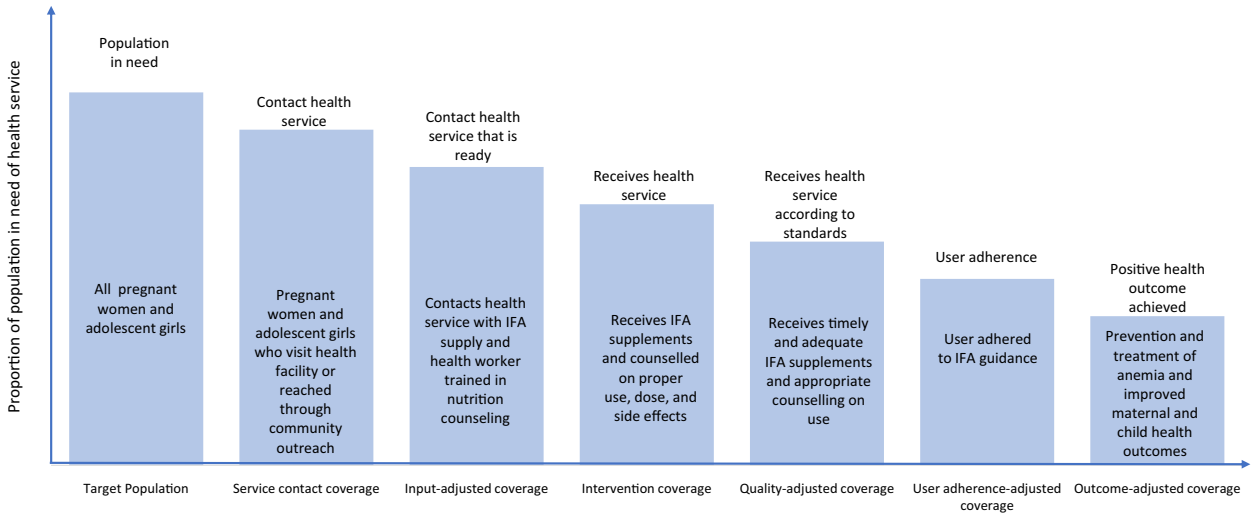


Fig. 2 Stylized example of health service coverage cascade for iron and folic acid (IFA) supplementation during pregnancy. Adapted from Marsh et al. (2020).

supplementations programs the evidence is clear on what to do but we are failing on how to effectively implement programs. There is growing recognition of the role of implementation research, guided by clear frameworks to improve understanding of the gaps and challenges to inform continual program improvement (Tumilowicz et al., 2015). There are several factors that impact the implementation of supplementation programs (Gomes et al., 2021). We consider here three essential considerations: (1) high coverage among intended groups; (2) continual supply of supplement and (3) effective behavior change.

Ensuring high coverage among intended groups is a key challenge across programs. Often programs suffer from inadequate targeting or coverage where deficient individuals are missed or reached irregularly. Inability to sustain high coverage over long periods of time as financial, political, or other health priorities change is an additional challenge for supplementation programs. For example, among low- and lower-middle-income countries the median coverage rates in 2020 were very low for childhood iron supplementation (12%), childhood zinc supplementation (15%) and for iron and folic acid supplementation in pregnancy (32%) (Global Nutrition Report, 2020). There are also large disparities by country and socioeconomic status. Marsh et al., argue that simply looking at intervention coverage does not account for intervention quality and may actually overestimate the health benefit, reiterating the need to measure effective coverage (Marsh et al., 2020). Effective coverage is defined as the extent and quality of coverage that results in the expected positive health outcome. It can be best assessed using a care cascade starting with the proportion of a population in need of the service, and mapping coverage rates through several steps that are needed to reach effective coverage. **Fig. 2** demonstrates such a cascade for iron and folic acid supplementation during pregnancy. It highlights several key steps for effective coverage to be realized, including contact with a potential delivery platform (health facility or community health worker), adequate training and quality of care from health workers, product supply where and when needed, and consumption of the supplement as per recommendations. As described in the Bihar, India case study there is often a stark contrast between the population in need (over 60% anemia among pregnant women) and receipt and consumption of recommended IFA during pregnancy. While 80% of pregnant women reported registering their pregnancies (International Institute for Population Sciences (IIPS), 2016) and attended at least one antenatal care visit, only 10% of women reported adhering to IFA guidance and consuming the supplement for at least 100 days during their pregnancy (Wendt et al., 2018).

The coverage cascade is helpful as it combines typical intervention coverage with data on continual supply and effective behavior change communication required to impact health outcomes. Issues with supplement supply are well documented across the literature (Larson et al., 2012) and a common constraint to program impact. Unfortunately, less has been published on experiences that identify and apply innovative solutions to address these challenges. Effective behavior change is likewise a challenge and often overlooked and underinvested in. For example, individuals may receive the supplement but be provided limited information on how or why to take them and what side effects may be expected, resulting in poor compliance (Fite et al., 2021). Tailored communication approaches with messages that address key barriers or common misperceptions, and are delivered through channels that targeted individuals access and trust are needed to motivate supplement use.

Case studies

To illustrate these considerations in practice, we present three case-studies including lessons learned from IFA supplementation among women in Bihar, India (**Box 1**), oral zinc supplementation among children in Bangladesh (**Box 2**), and Vitamin A supplementation in children in Nicaragua (**Box 3**). The common lessons learned from the three case-studies are summarized in **Table 2**. For example, the zinc supplementation program in Bangladesh showed that scaling-up required a shift in communication methodology. Although mass media campaigns to promote awareness were largely successful, there was little impact of the individual behavior change communication component. This emphasizes the importance of formative research to understand the local context and what might motivate the population and to use these findings in communication design. Experience from India and Nigeria demonstrate the importance of a continual supply of supplements. Not only was implementation negatively affected by the lack of a functional management systems and documentation of delivery, stock, and usage, but inconsistent and unreliable supply fostered

Box 1

Case Study 1: Distribution of antenatal iron and folic acid (IFA) supplements to pregnant women using existing health care channels in Bihar State, India

India has one of the highest prevalence of anemia, particularly the state of Bihar where over 60% of pregnant women are anemic (International Institute for Population Sciences (IIPS), 2016). In 2016, however, only 10% of pregnant women received the recommended dose of 100 IFA tablets through antenatal care services (Wendt et al., 2018). CARE-India in collaboration with Emory University launched a mixed-methods study to carefully examine the gaps and challenges that need to be addressed to improve effective coverage, i.e., the provision and consumption of IFA supplements daily by pregnant women for at least 100 days during pregnancy (Wendt et al., 2018). This case study assessed the availability of IFA supplements at various levels in 8 selected districts in Bihar combined with key interviews with various stakeholders including field level workers and community members to evaluate awareness of the program and processes related to procurement and distribution. The lack of compliance was attributed to a high rate of reported side-effects. Lack of demand and awareness of the importance of consuming supplements during pregnancy and supply issues were also identified. Specific programmatic adaptations were suggested to address these (**Table 4**). Several of these issues (e.g., health sub-center infrastructure; access to early and quality antenatal care) reflect broader health systems investment and quality assurance issues with implications well beyond supplementation.

Box 2

Case Study 2: Oral zinc supplementation for treatment of diarrhea in Bangladesh

In 2004, the World Health Organization WHO/and UNICEF revised their clinical management of childhood diarrhea guidelines to include provision of zinc supplementation based on evidence from several large randomized clinical trials that demonstrated reductions in disease duration and severity as well as the risk of a repeat episode following the receipt of 10–20 mg of zinc daily for at least two weeks (WHO/UNICEF, 2004). The recommendation was to provide oral zinc supplements that would be consumed daily for 10–14 days along with standard treatment of diarrhea in young children (under age 5 y) that included oral rehydration salts (ORS).

The Scaling up of Zinc for Young Children (SUZY) project that was established in 2003 in Bangladesh is one of the few well documented case-study that describes the various steps and challenges in the national scale-up of oral zinc supplementation to young children with diarrhea (Larson et al., 2012). Key factors that distinguish this case include: (1) the formation of a partnership with public, private, non-governmental organizations and multinational sector agencies; (2) formative and implementation research that informed product development, acceptability, awareness building and preparation of mass media messages and (3) continuous monitoring and evaluation of coverage, utilization and potential unintended consequences (i.e., reduced ORS use). The project involved the scaling up of a dispersible, 20-mg zinc sulfate tablet, referred to as "Baby Zinc" (Larson et al., 2012). The tablet is placed in a spoon, or a small cup, and water is added which leads it to disperse into a sweet, vanilla flavor that masks the taste of zinc. The treatment is packaged in a 10-tablet blister pack, and caretakers are instructed to give one tablet per day for 10 days. Studies of acceptability were promising but formative research that included caretakers of children with diarrhea, health care providers, drug vendors and medical representatives who sold drugs revealed disparities in care favoring urban settings and wealthier households (Larson et al., 2009). Nevertheless, the mass media promotion campaigns that included TV were successful in increasing awareness dramatically from less than 10% in all communities prelaunch to 90%, 74%, 66% and 50% in urban non-slum, municipal, urban slum, and rural sites by 10 months (Larson et al., 2009). However, after ~2 years, utilization was low and only 25%, 20%, 20% and 10% in the same sites, respectively. Many children also did not receive the correct ten-day course with ~50% receiving seven or fewer tablets (Larson et al., 2009, 2012). The implementation research identified several factors that may have contributed to these results (Larson et al., 2009, 2012). For example, drug vendors commonly sell medications to cover only a few days with the expectation that caregivers will purchase additional or alternative medications if child remains sick. Caregivers had limited experience continuing medications once their child appears to be cured, resulting in failure to complete the recommend ten-day course. While the project succeeded in high early awareness with mass media campaign a critical challenge was the transition from awareness to practice. There was a need to strengthen interpersonal communication and address potential barriers to use related to household level decision making for behavior change. These findings clearly showed the gaps and challenges to move from increasing awareness to actual program coverage and fidelity.

Box 3

Case Study 3: Vitamin A supplementation using an integrated approach to anemia reduction in women and children in Nicaragua

In the early 1990s, anemia among women and children in Nicaragua was identified as a problem of public health significance. The estimated prevalence of anemia at that time was about 28.5% and 33.6% in children ages 1–4 and non-pregnant women of reproductive age, respectively (Mora, 2007). While the identified causes of anemia in these populations were varied and include iron, thiamine, riboflavin, and folate deficiencies, reducing Vitamin A deficiency (VAD) was chosen to be the goal of the ensuing programmatic interventions address the issue of inadequate vitamin A intake (Mora, 2007). The program was a part of the National Micronutrient Plan created by the Ministry of Health (MOH) in Nicaragua and included a variety of multisector interventions (Mora, 2007). Vitamin A supplements were distributed for children younger than 5 years of age and post-partum women. Additionally, sugar fortification with Vitamin A was made mandatory, as sugar is regularly consumed by nearly all Nicaraguan families. Vitamin A supplements were distributed at biannual National Health Rallies (NHRs) that were largely successful, due to high community mobilization from media engagement, district-level MOH staff, the church, community groups, and NGOs. Additionally, the MOH implemented an Integrated System for Surveillance of Nutritional Interventions, an information system that monitors implementation to inform decision-making in nutrition programs (Mora, 2007). In 2000, the results of a national micronutrient survey revealed a striking reduction in VAD (Mora, 2007). In children under 5 years old, VAD prevalence was 31.3% in 1993 and dropped to 8.8%. In women, VAD was found not to be a problem of public health significance. The success of the program can be in part attributed to a series of concrete steps taken by the MOH: high-level advocacy; in-depth formative research surrounding the causes of VAD, current policies and programs in place, and resources available; and the creation of a taskforce charged with coordinating program development and implementation (Mora, 2007).

distrust among the target population. All three case studies emphasize obtaining high coverage as a key component of successful implementation. They show that this can be achieved in part by multi-intervention approaches that incorporate local ownership of the program and activities. Formative research concerning constraints and acceptability of activities, and a well-developed monitoring and evaluation plan are also referenced in all three cases as essential implementation actions.

Research gaps and future directions

There remain critical gaps in knowledge to improve the targeting, coverage and impact of supplementation programs globally (**Table 3**). First, better reporting from supplementation efficacy and effectiveness trials could lead to better formulation of recommendations with concrete suggestions for adaptation to context. For example, data are lacking to guide when programs can shift from daily to weekly IFA supplementation during pregnancy, or when population-based supplementation programs can be stopped

Table 2 Lessons learned from program implementation case studies.

Criteria	*Lessons learned*
Effective behavior change	• Mass media can be useful for drawing attention to supplementation programs, but clear messages through effective channels are needed to go beyond awareness to directly influence practices and lead to lasting behavior change • Formative research should always be carried out to design behavior change campaigns (messages and means of delivery) that respond to the local situation
Continual supply	• A functional management logistic system is critical to maintain continued availability of supplements and buffer stocks at delivery posts • Effective management and documentation is needed at lower administrative levels (village, sub-district) to more accurately track existing inventories and needs • Effective international cooperation and coordination facilitates the establishment and maintenance of a reliable supply chain when supplements are not manufactured in-country
High coverage	• The most effective programs simultaneous address effective behavior change, continual supply and high coverage • Building capacity for local production and dispersion of supplements is pivotal, as local program ownership facilitates policy decisions and program sustainability • Creating constraint frameworks is extremely useful, constraints can be initially identified via a pilot program or other methods • Effective monitoring is essential and should be designed at the beginning of a program, drawing on an explicit mapping of the program implementation process (the impact pathway) to identify and prioritize data that will identify bottlenecks and solutions to address them • Process and impact evaluation should be included wherever feasible to generate objective evidence of program delivery and effectiveness, which together with monitoring data should be used to adapt programs as needed

Table 3 Key implementation research priorities to enhance supplementation program delivery.

- Up-to-date evidence of the prevalence of micronutrient deficiencies disaggregated to potentially high-risk population subgroups is lacking in several regions of the world for most population groups. Data are particularly scarce for among pregnant women, children >5 years of age, adolescents, and older adults
- Program coverage data is lacking and requires improvements to monitor the equity in which supplementation programs are delivered. Robust coverage data should be collected to understand the coverage cascade and factors that are limiting it
- Innovations in implementation science are needed in how to best deliver supplementation programs to maximize coverage among target population and achieve impact at scale
- Further development of simple costing and cost optimization tools that can be applied to inform decisions related to program continued relevance and need for revision or cessation

given decrease in the burden of deficiency or the existence of other effective actions. Trials should always report baseline nutrient status of participants, along with detailed description of other population characteristics. In addition, trails are recommended to include details of supplement content, and the coverage cascade identifying key points that inform progress toward effective coverage. Second, up-to-date evidence of the prevalence of micronutrient deficiency is woefully lacking for most populations groups across most countries in the world (Brown et al., 2021). Similarly, program coverage data is lacking and requires improvements to monitor the equity in which supplementation programs are delivered. Finally, innovations in implementation science are needed in how to best deliver supplementation programs to maximize coverage among target population and achieve impact at scale.

Summary

Global guidance for the potential of many different types of supplementation exist, and programs are implemented in many countries globally, including many with high burden of micronutrient deficiency. Such guidelines have been informed by ample evidence of efficacy of supplements, but rigorous implementation research to inform the design, implementation and continual improvement of such programs is lacking. Further research and investment on effective strategies to achieve high coverage among intended groups, continual supply of supplement and effective behavior is required to reach impact at scale.

Table 4 Identified bottlenecks of Bihar's iron and folic acid (IFA) supply chain and proposed actions.

Stage	*Identified bottleneck/issue*	*Proposed action*
Demand	Lack of appropriate IFA need forecasting	• Standardized demand forecasting based on accurate estimates of district needs and previous consumption • Computerization and clear documentation of inventory, stock requests, and expiry dates • Estimates to include lactating women population; IFA distribution and counseling standard
Supply acquision & utilization	Late supplier deliveries resulting in inconsistent supply	• Implement centralized supply system whose policy is to deduct payment upon late delivery, damaged stock, etc.
	Perceived or actual inability to procure IFA when needed through local purchasing	• Set aside and explore use of untied funds to purchase IFA locally in times of shortage or emergency
	Lack of buffer stock use at all levels	• Implementation, monitoring, and evaluation of existing buffer stock requirements • Ensure adequate storage facilities so stock can be stored safely
	No safe disposal plan for expired medicines and pushing of expiring drugs to patients and frontline workers	• Transparent plan to prevent expired medicines through appropriate purchasing practices and safe disposal of expired medicines
	Inconsistent training on IFA counseling/distribution across FLW types	• IFA counseling/distribution training for all frontline workers who work with pregnant women • Training for all frontline workers together at health sub-center level to improve coordination and communication
Storage	Storeroom transiency and disorder	• Construct, purchase, or long-term rental of adequate storerooms, including resources for racks, labels, and shelves in resource development proposals • Training for storekeepers including storeroom order and inventory protocols

See Also: Iodine: Deficiency disorders and prevention programs; Iodine: Physiology, dietary sources, and requirements; Pregnancy: Nutrient requirements; Vitamin A: Deficiency and interventions

Conflict of Interest

The views expressed in this publication are those of the author(s) and do not necessarily reflect the views of FAO.

References

Brown, K.H., Moore, S.E., Hess, S.Y., McDonald, C.M., Jones, K.S., Meadows, S.R., Manger, M.S., Coates, J., Alayon, S., Osendarp, S.J.M., 2021. Increasing the availability and utilization of reliable data on population micronutrient (MN) status globally: the MN data generation initiative. Am. J. Clin. Nutr. 114, 862–870.

De Pee, S., Spiegel, P., Kraemer, K., Wilkinson, C., Bilukha, O., Seal, A., Macias, K., Oman, A., Fall, A.B., Yip, R., Peña-Rosas, J.P., West, K., Zlotkin, S., Bloem, M.W., 2011. Assessing the impact of micronutrient intervention programs implemented under special circumstances—meeting report. Food Nutr. Bull. 32 (3), 256–263. https://doi.org/10.1177/156482651103200311.

Fiedler, J.L., Puett, C., 2015. Micronutrient program costs: sources of variations and noncomparabilities. Food Nutr. Bull. 36 (1), 43–56. https://doi.org/10.1177/156482651503600105.

Fite, M.B., Roba, K.T., Oljira, L., Tura, A.K., Yadeta, T.A., 2021. Compliance with Iron and Folic Acid Supplementation (IFAS) and associated factors among pregnant women in Sub-Saharan Africa: a systematic review and meta-analysis. PLoS One 16 (4), e0249789. https://doi.org/10.1371/journal.pone.0249789.

Global Nutrition Report, 2020. Action on Equity to End Malnutrition. Development Initiatives, Bristol, UK.

Global Nutrition Report, 2021. The State of Global Nutrition. Development Initiatives, Bristol, UK.

Gomes, F., King, S.E., Dallmann, D., Golan, J., Da Silva, A.C.F., Hurley, K.M., Bergeron, G., Bourassa, M.W., Mehta, S., 2021. Interventions to increase adherence to micronutrient supplementation during pregnancy: a systematic review. Ann. N. Y. Acad. Sci. 1493 (1), 41–58. https://doi.org/10.1111/nyas.14545.

International Institute for Population Sciences (IIPS), 2016. India Fact Sheet: National Family Health Survey (NFHS)-4. IIPS, Mumbai.

Keats, E.C., Das, J.K., Salam, R.A., Lassi, Z.S., Imdad, A., Black, R.E., 2021. Effective interventions to address maternal and child malnutrition: an update of the evidence. Lancet 5 (5), 367–384.

Larson, C.P., Koehlmoos, T.P., Sack, D.A., 2012. Scaling up zinc treatment of childhood diarrhoea in Bangladesh: theoretical and practical considerations guiding the SUZY Project. Health Pol. Plann. 27 (2), 102–114. https://doi.org/10.1093/heapol/czr015.

Larson, C.P., Saha, U.R., Nazrul, H., 2009. Impact monitoring of the national scale up of zinc treatment for childhood diarrhea in Bangladesh: repeat ecologic surveys. PLoS Med. 6 (11), e1000175.

Marsh, A.D., Muzigaba, M., Diaz, T., Requejo, J., Jackson, D., Chou, D., Cresswell, J.A., Guthold, R., Moran, A.C., Strong, K.L., Banerjee, A., Soucat, A., 2020. Effective coverage measurement in maternal, newborn, child, and adolescent health and nutrition: progress, future prospects, and implications for quality health systems. Lancet Global Health 8 (5), e730–e736. https://doi.org/10.1016/s2214-109x(20)30104-2.

Mora, J.O., 2007. Integrated Anemia Control Strategy has Significantly Reduced Anemia in Women and Children in Nicaragua. Micronutrient Initiative, Ottawa, Canada.

Neufeld, L.M., García-Guerra, A., Quezada, A.D., Théodore, F., Bonvecchio Arenas, A., Islas, C.D., Garcia-Feregrino, R., Hernandez, A., Colchero, A., Habicht, J.P., 2019. A fortified food can be replaced by micronutrient supplements for distribution in a Mexican social protection program based on results of a cluster-randomized trial and costing analysis. J. Nutr. 149, 2302S–2309S.

Pena-Rosas, J.P., De-Regil, L.M., Rogers, L.M., Bopardikar, A., Panisset, U., 2012. Translating research into action: WHO evidence-informed guidelines for safe and effective micronutrient interventions. J. Nutr. 142, 197S–204S.

Stevens, G.A., Paciorek, C.J., Flores-Urrutia, M.C., Borghi, E., Namaste, S., Wirth, J.P., Suchdev, P.S., Ezzati, M., Rohner, F., Flaxman, S.R., Rogers, L.M., 2022a. National, regional, and global estimates of anaemia by severity in women and children for 2000–19: a pooled analysis of population-representative data. Lancet Global Health 10, e627–e639.

Stevens, G.A., Beal, T., Mbuya, M.N.N., Luo, H., Neufeld, L.M., Global Micronutrient Deficiencies Research, G., 2022b. Micronutrient deficiencies among preschool-aged children and women of reproductive age worldwide: a pooled analysis of individual-level data from population-representative surveys. Lancet Glob Health 10 (11), e1590–e1599. https://doi.org/10.1016/S2214-109X(22)00367-9.

Tumilowicz, A., Neufeld, L.M., Pelto, G.H., 2015. Using ethnography in implementation research to improve nutrition interventions in populations. Matern. Child Nutr. 11, 55–72. https://doi.org/10.1111/mcn.12246.

Victora, C.G., Christian, P., Vidaletti, L.P., Gatica-Domínguez, G., Menon, P., Black, R.E., 2021. Revisiting maternal and child undernutrition in low-income and middle-income countries: variable progress towards an unfinished agenda. Lancet 397 (10282), 1388–1399. https://doi.org/10.1016/s0140-6736(21)00394-9.

Vosti, S.A., Kagin, J., Engle-Stone, R., Brown, K.H., 2015. An economic optimization model for improving the efficiency of vitamin A interventions: an application to young children in Cameroon. Food Nutr. Bull. 36 (3), S193–S207.

Wendt, A.S., Stephenson, R., Young, M.F., Verma, P., Srikantiah, S., Webb-Girard, A., Hogue, C.J., Ramakrishnan, U., Martorell, R., 2018. Identifying bottlenecks in the iron and folic acid supply chain in Bihar, India: a mixed-methods study. BMC Health Serv. Res. 18 (1). https://doi.org/10.1186/s12913-018-3017-x.

WHO, 2016. Recommendations on antenatal care for a positive pregnancy experience. World Health Organization, Geneva. License: CC BY-NC-SA 3.0 IGO.

WHO, 2019. Essential nutrition actions: mainstreaming nutrition through the life-course. World Health Organization, Geneva. Licence: CC BY-NC-SA 3.0 IGO.

WHO, 2020a. WHO antenatal care recommendations for a positive pregnancy experience. In: Nutritional interventions update: Multiple micronutrient supplements during pregnancy. World Health Organization, Geneva. Licence: CC BY-NC-SA 3.0 IGO.

WHO, 2020b. WHO antenatal care recommendations for a positive pregnancy experience. In: Nutritional interventions update: Vitamin D supplements during pregnancy. World Health Organization; 2020, Geneva. Licence: CC BY-NC-SA 3.0 IGO.

WHO, 2021. WHO antenatal care recommendations for a positive pregnancy experience. In: Nutritional interventions update: zinc supplements during pregnancy. World Health Organization, Geneva. Licence: CC BY-NC-SA 3.0 IGO.

WHO/UNICEF, 2004. WHO/UNICEF Joint Statement on Clinical Management of Acute Diarrhoea. World Health Organization, Geneva.

Milk and dairy products

Javier Fontecha[a], **Manuela Juárez**[a], **and Ángel Gil**[b,c,d], [a] Department of Bioactivity and Food Analysis, Food Lipid Biomarkers and Health Group, Institute of Food Science Research, CIAL (CSIC-UAM), Superior Council of Scientific Investigations and Autonomous University of Madrid, Madrid, Spain; [b] Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology "José Mataix", Center of Biomedical Research, University of Granada, Granada, Spain; [c] Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain; and [d] CIBEROBN (Physiopathology of Obesity and Nutrition Network CB12/03/30038), Institute of Health Carlos III (ISCIII), Madrid, Spain

© 2023 Elsevier Ltd. All rights reserved.

Introduction 480
Milk composition 480
Proteins 480
Lipids 481
Carbohydrates 482
Mineral elements 483
Vitamins 483
Drinking milk 484
Dairy products 484
Fermented milks 485
Cheese 485
Cream and butter 485
Role of milk and dairy products in disease prevention and health promotion 486
Overall mortality 486
Pregnancy, maternal health, and newborns 486
Growth and development 486
Weight gain 486
Osteoporosis 487
Metabolic syndrome 487
Diabetes and insulin resistance 487
Cardiovascular disease (CVD) 487
Dairy proteins 487
Milk fat 488
Minerals 488
Cancer 488
Breast cancer 488
Colorectal cancer 488
Ovarian cancer 489
Liver cancer 489
Prostate cancer 489
Bladder cancer 489
Cognitive function 489
Conclusions 490
References 490
Further reading 491

Key points

- To know the overall composition of milk from different species
- To know the main proteins, lipid components and carbohydrates of milk
- To know the main mineral elements and vitamins of milk
- To know the overall composition of fluid milks and dairy products
- To understand the role of milk and dairy products in disease prevention and health promotion

Introduction

Milk is the secretion of female mammals to nourish newborns in their first months of life. Due to its high nutrient density, milk also has a recognized nutritional role in the diet of adolescents and adults. From a nutritional point of view, there are several important aspects of milk consumption that can be highlighted: (1) In milk there is a good balance between the main components, fats, proteins, and carbohydrates, which are present in comparable proportions. (2) It provides high levels of nutrients in relation to caloric content. This is especially interesting for people following a weight control diet. (3) Taking into account the consumption figures for dairy products as a whole, it can be concluded that they provide more than one-fifth of the recommended daily allowance (RDA) of protein, zinc, riboflavin (vitamin B2), and vitamin A, as well as smaller proportions of the recommended amounts of many other vitamins (B12 and D). On the other hand, its calcium content stands out, close to 60% of the RDA. In addition, recent studies consider that milk is the most important source of natural bioactive components and that its effects on health are the result of the interaction of all nutrients and go beyond the simple sum of individual effects (Thorning et al., 2017).

At the physicochemical level, milk is a balanced mixture of proteins, fat, carbohydrates, salts, and other minor components dispersed in water as emulsions, colloidal suspensions, and true solutions. In the dissolved state, lactose, water-soluble vitamins, and salts are found, totally or partially ionized, and constitute a stable phase. The micellar complex, a calcium phosphate associated with caseins, is a macromolecular aggregate of varied shape and structure, which has a net negative charge, ensuring stability by electrostatic repulsion. In the whey fraction, specific whey proteins, serum proteins, and low proportions of soluble caseins are also found. The fat globules are surrounded by a membrane of lipoprotein nature, which keeps the fat in an emulsion state.

Although the composition of milk can vary due to the animal (breed, lactation stage) or other factors (climate, diet, type of milking, etc.), variations are limited by certain physical properties such as osmotic pressure, since milk is isotonic with blood; in both fluids, the sum of molar concentrations of solutes is approximately 0.3. The main contributors are lactose and sodium, potassium, and chloride ions. Another limitation on the composition of milk is the solubility of calcium phosphate and citrate, as well as the melting point of fat, which should not be higher than body temperature (37–40 °C). Of the milk components, fat is the one with the greatest variability due to genetic factors, with notable differences not only between different breeds but also between individuals of the same breed. However, the composition of milk fat varies mainly with cattle feeding and less markedly because of other factors (Table 1).

Milk composition

Proteins

Cow's milk contains between 3 and 3.5% proteins, distributed in fractions whose importance varies according to dairy technology and nutrition: caseins, soluble proteins or whey proteins, and non-protein nitrogenous substances (Table 2). The colloidal dispersion in milk is very stable even at extreme temperatures and concentrations. In general, the dispersion recovers after drying or concentration, a circumstance of great technological importance. Milk proteins are differentiated by the size of their molecules, with molecular masses from 12,000 to 380,000 Da. They occur in two different phases: the caseins, an unstable micellar phase formed by suspended micelles that diffuse light and give the milk an opaque white appearance. The other phase is soluble, formed by different hydrophilic protein polymers named whey proteins.

Caseins are the major proteins in milk (75%–85%). They are a group of phosphorous proteins, which are present in the form of an organic complex, the micelle, of αs, β y κ-caseins, bound by colloidal calcium phosphate and to a lesser degree, citrates and magnesium. The micelle is hydrated (65% water) and its size varies from 30 to 300 nm. The most phosphorylated caseins (αs_2, αs_1, and β), are unstable in the presence of Ca, and are located in the interior of the micelle; toward the exterior is located most of the κ-casein, hydrophilic, stable in the presence of Ca. If the integrity of the casein is lost, the milk coagulates. A fraction of caseins, mostly β-casein, is soluble (1% at 20 °C), a fraction that increases when milk is kept refrigerated (10% at 5 °C), which can affect cheese yield. At pH 4.6 the micelles have zero overall charge, aggregate, and coagulate, separating from the aqueous phase. At this pH, whey proteins remain in solution. Whey proteins represent 19% of total proteins and during thermal processes; they denature and lose their solubility approximately to isoelectric point, although the degree of denaturation depends on the intensity of

Table 1 Average composition of basic nutrients in milk from different species.

Species	*Fat (g)*	*Protein (g)*	*Lactose (g)*	*Total solids (g)*	*Ash (g)*	*Energy (kcal)*
Cow	3.7	3.2	4.7	12.8	0.7	69
Sheep	7.9	5.7	4.8	19.3	0.9	105
Goat	4.0	3.5	4.3	13.0	0.9	70
Buffalo	7.6	3.8	4.0	17.2	0.8	102
Human	3.8	1.4	7.0	12.6	0.3	68

Values are expressed per 100 g.

Table 2 Average values of cow's milk protein content.

Component	*g/kg*	*%*
Total protein	33.0	100.0
Total caseins	26.0	79.5
αs1-casein	10.0	30.6
αs2-casein	2.6	8.0
β-casein	9.3	28.4
γ-casein	0.8	2.4
k-casein	3.3	10.1
Total whey proteins	6.3	19.3
α-Lactoalbumin	1.2	3.7
β-lactoglobulina	3.2	9.8
Seroalbumin	0.4	1.2
Inmunoglobulins	0.7	2.1
Membrane proteins	0.4	1.2
Others	0.8	2.4

heating and other conditions such as acidification. Whey proteins include α-lactoalbumin, β-lactoglobulin, immunoglobulins, proteoses-peptones, and other non-specific minority components such as lactoferrin and lysozyme, with bacteriostatic properties. α-lactoglobulin is the quantitatively most important whey protein in cow's milk. At normal milk pH, it exists as a dimer and due to the cysteine radicals that it contains, it can create disulfide bridges, which stabilize dimerization. The binding of casein and β-lactoglobulin that takes place in heated milk also results from an S—S bridge between the two molecules. α-Lactoalbumin stands out from the biological point of view since it participates in the synthesis of lactose. Non-protein nitrogen (NNP) varies from 3% to 7% of total nitrogen, and in sheep's milk, it represents 5% of total nitrogen.

Milk also contains a certain number of enzymes that are important from the technological point of view in cheese making and ripening, such as lipases that give rise to hydrolytic rancidity, proteases that hydrolyze casein, etc. On the other hand, the heat lability of certain enzymes provides the basis for analytical tests that determine whether milk has been subjected to heat treatment; in this sense, the analysis of alkaline phosphatase deactivation is used as a marker of adequate pasteurization. Therefore, the major milk proteins, caseins, and whey proteins contain all the essential amino acids required to meet our nutritional needs and they present high digestibility and biological value, so they are defined as high-quality proteins. Because of their high lysine content, milk proteins can complement proteins from other foods, such as cereals. Moreover, gastrointestinal digestion of milk proteins generates bioactive peptides that are reported to have numerous beneficial effects on health or on reducing the risk of disease. These peptides are also ending products of enzymatic hydrolysis of the milk proteins that are liberated during fermentation and ripening of fermented milks and cheeses. Apart from the positive data on milk proteins based on amino acid composition, milk has minor proteins with documented specific functions, such as lactoferrin (0.2 g/L) and lysozyme whose confer protection against microbial infections. Milk also contains immunoglobulins (0.7 g/L) that modulate aspects of immunity. Other minor milk proteins are bioactive molecules of great interest especially in infancy and adolescence, with the same content and molecularly identical to those present in human milk, which have been described as growth factors and modulators. On the other hand, the protein fraction includes in its structure two-thirds of the calcium and half of the phosphorus in milk, as well as practically all of the iron, manganese, and zinc, which facilitates their bioavailability. Besides, milk whey proteins alone or together with bioactive peptides, amino acids produced after digestion or combined action with other milk constituents, such as calcium, have been shown to contribute to the regulation of body weight, by providing satiety signals that affect the regulation of food intake in preference to proteins from other sources. Faster digestion of peptides and increased thermogenesis have also been reported.

Lipids

They are among the most important constituents of milk in terms of economic and nutritive aspects and because of the physical and organoleptic characteristics that they impart to dairy products. They are present in the form of globules (MFG) with a hydrophobic nucleus that consists mainly of triacylglycerols (TAG) surrounded by a membrane (MFGM) composed mostly of phospholipids (PL) and glycoproteins. These polar lipid compounds are indispensable elements that form part of cell membranes and transport liposoluble vitamins (A, D, E, and K). In the composition of milk lipids, together with the major TAG components, there are also simple lipids (diacylglycerols, monoacylglycerols, and waxes), complex lipids (mostly PL), cholesterol, cholesterol esters, antioxidants (especially tocopherols), and squalene (**Table 3**).

Milk fat is very complex, with about 400 different fatty acids (FA) which have from 4 to 26 carbon atoms, although only just under 30 are present in a proportion greater than 0.1% and the rest are only present as traces. About 60%–70% of all the FA are saturated (SFA), 20%–25% are monounsaturated (MUFA) mainly oleic acid (*cis*-9 C18:1), and 1%–4% of *trans* fatty acids (TFA) and 3%–5% are polyunsaturated (PUFA), which include essential FA such as linoleic acid (LA, *cis*-9 *cis*-12C18:2, or 18:2 n-6) and α-linolenic acid (ALA, *cis*-9 *cis*-12 *cis*-15C18:3 or 18:3 n-3), with concentrations of 1%–3% and 0.3%–1%, respectively, which have a recognized beneficial effect for cardiovascular health (Fontecha and Juárez, 2017).

Table 3 Range composition of lipid classes in cow's milk fat.

Component	*%*
Triacylglycerols	97–98
Diacylglycerolss	0.4–0.6
Monoacylglycerolss	0.02–0.03
Free fatty acids	0.03–0.1
Phospholipids	0.3–0.8
Cholesterol	0.25–0.35
Hydrocarbons	Traces
Cholesterol esters	Traces

Milk fat differs from other animal fats because of the exclusive presence in it of short-chain fatty acids (SCFA), as butyric acid (C4:0) and caproic acid (C6:0), and medium-chain FA, namely caprylic acid (C8:0) and capric acid (C10:0), which have been shown not to have an effect on increasing the levels of cholesterol in the blood. These SCFA make up 8%–12% of the total in cow's milk, while in sheep and goat milk the levels of the C6 to C10 acids are two to three times higher than in cow's milk (Park et al., 2007) (**Table 4**). In addition, milk fat has 1–4% of naturally occurring TFA. The major isomer of naturally occurring TFA is *trans*-11C18:1, a physiological precursor of conjugated linoleic acid (CLA), for which a multitude of health benefits have been described. As previously stated, whole milk has a high content of SFA and therefore there have been recommendations advising against indiscriminate consumption of it. Nevertheless, new research has been conducted in recent years which has shown the nonexistence of an association between intake of balanced dairy foods (containing all components) and cardiovascular risk in healthy individuals (Fontecha and Juárez, 2017).

Carbohydrates

Lactose is the only sugar found in milk in significant quantities. They constitute the largest fraction of the dry milk powder, and the most labile to microbial action, being transformed into lactic acid and other products responsible for the sour taste of fermented milk. Lactose is a disaccharide of glucose and galactose, found in aqueous solution as two anomers named α and β-lactose. The crystallization of lactose is of great practical importance in the manufacture of various dairy products: condensed milk, powdered milk, whey powder, ice cream, etc., and the relative insolubility of lactose is of special significance in obtaining concentrated milk. An improving effect of lactose on the absorption of calcium and other mineral elements from milk has been reported. However,

Table 4 Average composition of the main fatty acids of milk from different species.

		Species (%)		
Fatty acids	*Fatty acids*	*Cow*	*Sheep*	*Goat*
Butyric	C4:0	3.1	3.5	2.2
Caproic	C6:0	1.9	2.9	2.4
Caprylic	C8:0	1.2	2.6	2.7
Capric	C10:0	2.5	7.8	10.0
Lauric	C12:0	3.0	4.4	5.0
Myristic	C14:0	10.4	10.4	9.8
Myristoleic	*cis*-9 C14:1	1.1	0.3	0.2
iso Pentadecanoic	*iso* C15:0	0.3	0.3	0.1
Anteiso Pentadecanoic	*anteiso* C15:0	0.5	0.5	0.2
Pentadecanoic	C15:0	1.1	1.0	0.7
iso Palmitic	*iso* C16:0	0.2	0.2	0.2
Palmitic	C16:0	28.5	25.9	28.2
Palmitoleic	*cis*-9 C16:1	1.7	1.0	1.6
iso Heptadecanoic	*iso* C17:0	0.6	0.5	0.4
Anteiso Heptadecanoic	*anteiso* C17:0	0.5	0.3	0.4
Heptadecanoic	C17:0	0.7	0.6	0.7
Stearic	C18:0	10.5	9.6	8.9
Oleic	*cis*-9 C18:1	20.5	18.2	19.3
Vaccenic	*cis-11* C18:1	4.3	2.9	2.1
Linoleic	*cis*-9 *cis*-12C18:2	3.1	2.3	3.2
Conjugated linoleic	CLA	1.0	0.7	0.7
α-Linolenic	*cis*-9 *cis*-12 *cis*-15C18:3	0.6	0.6	0.4

lactose does not cross the intestinal membranes and in lactase deficient people, it is utilized by the microorganisms present in the intestine producing a series of intestinal disorders known as lactose intolerance. Individuals with difficulties to tolerate milk can consume fermented products such as cheese -since after the first days of ripening it has very low lactose content- or even yogurt. Yogurt has lower lactose levels than milk, but it has also been demonstrated that yogurt lactose is better digested.

Mineral elements

The most abundant mineral elements in milk are Ca, P, K, Na, Cl, and Mg whilst Zn, Fe, Cu, and Mn are the trace element more remarkable. Representative values for the average mineral content of milk from different species are presented in **Table 5**. The levels of Ca, P, Mg, Zn, Fe, and Cu are higher in sheep than in cows milk, while the opposite appears to be the case for K and Na.

Cows milk contains about 1% salts. Sheep milk has around 0.9% as compared to 0.7% in cow's or goat's milk (De la Fuente and Juárez M. 2015). Of the major constituents, Ca and P are especially important, both in the nutritional aspect and for their role in the physical state and stability of the caseins. These two elements are distributed in the soluble and colloidal phases of milk as follows Ca: 1/3 soluble and 2/3 micellar; P: 1/2 soluble and 1/2 micellar. The calcium/phosphorus ratio is optimal for absorption. Mg is an important element since it intervenes as Ca in the stability of the micelle and has also been related to blood pressure regulation through their effects on cardiovascular diseases (CVDs). Approximately 1/3 of the Mg is micellar and the rest is in solution. Citric acid raises the level of soluble Ca in milk, in the form of calcium citrate. Concerning the elements K, Na and Cl, together with lactose, they contribute to the osmotic pressure of milk. About 95% of Mn and Zn, and 50%–75% of Fe and Cu respectively, are distributed in the colloidal phase of milk. Soluble proteins have affinity only for Fe, which bind from 18% to 33%. Lactose, ascorbate, citrate, phosphopeptides, and lactoferrin may have a significant impact on the absorption of mineral elements. Furthermore, milk does not contain substances such as phytates which could inhibit the absorption of minerals. Therefore, the bioavailability of minerals for dairy producs is very high.

Vitamins

Milk is an important source of water-soluble vitamins for the child and adult, which come partly from forage, but are mostly synthesized by rumen bacteria and their quantity shows little variation. The recommended intake of group B vitamins (B1, B2, and B12),

Table 5 Average composition per 100 g of mineral and vitamin content of milk from different species.

	Species		
Minerals	*Cow*	*Sheep*	*Goat*
Ca (mg)	122.00	193.00	134.00
Mg (mg)	12.00	18.00	16.00
P (mg)	119.00	121.00	121.00
Na (mg)	58.00	44.00	41.00
K (mg)	152.00	136.00	181.00
Cl (mg)	100.00	160.00	150.00
S (mg)	32.00	29.00	1.89
Fe (mg)	0.08	0.08	0.07
Zn (mg)	0.53	0.57	0.56
Cu (mg)	0.06	0.04	0.05
Mn (mg)	0.02	0.01	0.03
I (mg)	0.02	0.02	0.02
Se (μg)	0.96	1.00	1.33
Vitamins			
Retinol(A) (IU)	126.00	146.00	185.00
Vitamin D (IU)	2.00	0.18	2.30
Tocopherol (E) (mg)	0.11	0.11	0.04
Vitamin K (μg)	1.50	1.80	1.40
Thiamin (B1) (μg)	45.00	68.00	68.00
Riboflavin (B2) (mg)	0.16	0.38	0.21
Niacin (B3) (mg)	0.08	0.42	0.27
Pantotenic acid (B5) (mg)	0.32	0.41	0.31
Pyridoxin (B6) (mg)	0.04	0.08	0.05
Biotin (μg)	2.00	0.92	1.50
Vitamin B12 (μg)	0.36	0.71	0.07
Vitamin C (mg)	0.94	4.16	1.29
Folic acid (μg)	5.00	5.00	1.00

and an important percentage of A, is covered with the consumption of one liter of milk. As for fat-soluble vitamins, the vitamin A content depends on the diet, vitamin D on solar radiation, and E and K are in trace form; as they are associated with fat, are concentrated in fat-rich products (Table 5).

Drinking milk

According to the U.S. Department of Agriculture (USDA) drinking or fluid milks are products pasteurized (or heat treated), homogenized and packaged. They shall be clean, sound, wholesome, and free from foreign material such as, but not limited to, dirt, insect parts, hair, wood, glass, or metal, and comply with State, local laws, regulations, or requirements. The USDA has authorized the use of the following fluid milks:

Whole milk: Not less than 3.25% total milkfat.
Reduced fat milk: At least 25% less total fat.
Lowfat milk: Maximum of 3 g or less total fat in 240 mL (1 cup or 8 fluid ounces).
Nonfat/skim/fat free milk: Less than 0.5 g of total fat in 240 mL (1 cup or 8 fluid ounces).

The European Union (EU) Regulation defines the following sales descriptions, which must be used for drinking milk, which is intended for delivery to the consumer:

- *Whole milk* such as heat-treated milk that, with respect to fat content, meets one of the following requirements: i) standardized whole milk with a fat content of at least 3.50%. However, EU member states may provide for an additional category of whole milk with a fat content of 4% or above. ii) non-standardized whole milk as milk with a fat content that has not been altered since the milking stage either by the addition or removal of milk fats or by mixture with milk whose natural fat content has been altered. However, the fat content may not be less than 3.50%.
- *Semi-skimmed milk* such as heat-treated milk whose fat content has been reduced to at least 1.50% and at most 1.80%
- *Skimmed-milk* such as heat-treated milk whose fat content has been reduced to not more than 0.50%.

Heat-treated milk not complying with the fat requirements for the above products, i.e., whole milk/semi-skimmed milk/skimmed milk can be considered drinking milk provided that the fat content is clearly indicated with one decimal and easily readable on the packaging in the form of "X% fat". The legislation states that this milk must not be described as whole milk, semi-skimmed milk or skimmed milk.

The average composition of the main nutrients in drinking milk, as well as its energy intake, can be modified by its fat content as shown in Table 6.

Dairy products

The US FDA defines milk products as "Food products made exclusively or principally from the lacteal secretion obtained from one or more healthy milk-producing animals, e.g., cows, goats, sheep, and water buffalo, including, but not limited to, the following: lowfat milk, skim milk, cream, half and half, dry milk, nonfat dry milk, dry cream, condensed or concentrated milk products, cultured or acidified milk or milk products, kefir, eggnog, yogurt, butter, cheese (where not specifically exempted by regulation), whey, condensed or dry whey or whey products, ice cream, ice milk, other frozen dairy desserts and products obtained by modifying the chemical or physical characteristics of milk, cream, or whey by using enzymes, solvents, heat, pressure, cooling, vacuum, genetic engineering, fractionation, or other similar processes, and any such product made by the addition or subtraction of milkfat or the addition of safe and suitable optional ingredients for the protein, vitamin, or mineral fortification of the product"

The different processed dairy products have wide ranges of composition so that they can cover both different consumption habits and very different uses of nutritional interest. Apart from cream and butter, which are products enriched in milk fat with low levels of other nutrients, other milk products are marketed with different fat contents and with different levels of minerals and vitamins (enriched milk), concentrated milk, especially fermented milks (yogurts) that maintain all the nutrients in milk. Different types of cheese include concentrated fat, part or all of the milk proteins (depending on the manufacturing process), and important proportions of the minerals calcium and phosphorus.

Table 6 Average composition (%) of basic nutrients in milk by its fat content.

Milk type	*Fat (%)*	*Protein (%)*	*Lactose (%)*	*Ca (mg)*	*Vit. A (μg)*	*Energy (kcal)*
Whole	>3.5	3.2	4.7	124	46	63
Semi-skimmed	1.5–1.8	3.5	4.8	125	18.9	46
Skimmed	<0.5	3.9	4.9	121	0.1	35

Fermented milks

They belong to the group of probiotic foods, i.e. those that contain live microorganisms that, when ingested in sufficient quantity, can exert beneficial effects on health by favoring the balance and maintenance of the intestinal microbiota. The bacterial groups most commonly used as probiotics are lactobacilli and bifidobacteria.

Yogurt is, in its different presentations, the best-known fermented milk. It is made from pasteurized milk and milk fractions inoculated with a mixture of microorganisms, *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Streptococcus thermophilus*, which are responsible for the metabolic transformations of carbohydrates, proteins, and lipids, leading to the development of its characteristic flavor, aroma, and texture.

During lactic fermentation, the microorganisms hydrolyze lactose, using the enzyme lactase (β-galactosidase), into glucose and galactose. Glucose is mainly transformed to lactic acid - with a consequent decrease in pH, which hinders the development of undesirable microorganisms - and to small amounts of compounds that contribute to flavor and bioactivity. On the other hand, the proteolytic enzymes of the microorganisms act on the casein giving rise to polypeptides and peptides, and although hydrolysis is slow and in low proportion, the same amino acid composition is maintained as in milk. Besides, during fermentation and as a consequence of the decrease in pH, minerals such as calcium and phosphorus pass to the soluble form, and caseins precipitate in the form of a fine coagulum, which facilitates their proteolysis and consequently favors their absorption and digestibility.

Regarding the vitamins present in yogurt, although they can be metabolized by lactic acid bacteria, during the period of exponential growth they are also synthesized by the bacteria. The levels of fat-soluble vitamins depend on the fat content, and those of thiamine, riboflavin, vitamin B6, pantothenic and folic acid may be comparable or slightly higher, and vitamin B12 content lower, than those present in milk.

The most recent scientific evidence allows us to conclude that the consumption of fermented milks, in addition to favoring digestion in people who are partially lactose intolerant, enhancing immunomodulation, and maintaining body composition, are associated with a lower risk of diabetes, hypertension, CVD, or loss of bone mineral density (especially when supplemented with vitamin D). A reduction in the risk of breast and bladder cancer has also been observed, and it has not been possible to conclude that there is an association between the intake of fermented dairy products and the risk of prostate cancer, colon cancer, or all-cause mortality.

Cheese

Cheese contains in concentrated form many of the nutrients of milk, caseins, fat and fat-soluble vitamins, colloidal minerals, and some whey constituents. Some of the water-soluble vitamins (up to 90%) may be gone in the whey, but the final levels in the cheese depend on their synthesis and utilization by microorganisms. During processing and in the first stage of the ripening process, the transformation of lactose into lactic acid takes place, which favors the action of rennet, facilitates the separation of whey, prevents the development of undesirable microorganisms, and regulates proteolysis and subsequent lipolysis.

Hydrolysis of proteins into peptides and amino acids, by residual rennet and microbial proteases, is perhaps the most important stage in the ripening of many types of cheese, as it affects both flavor and texture. The nutritive value of cheese proteins is not affected by processing and the bioavailability of lysine is comparable to that of total milk if the cheese incorporates the whey proteins, which is the case when used in the production of fresh cheese by membrane processes. During the ripening process, part of the caseins are hydrolyzed to peptides and amino acids. The decarboxylation of amino acids during cheese ripening can give rise to amines, but in general, the potential health risk is negligible, except in individuals with amine metabolism difficulties.

The fatty compounds formed during ripening by the action of microbial lipases, the free fatty acids in the first stage, contribute significantly to the aroma of the cheese. The Ca and P contents of cheese -except for those of acid coagulation-are much higher than those of milk, 4–5 times in fresh or soft cheeses, 7–8 times in a semi-hard cheese, and up to 10 times in hard cheese. The transformations that take place during cheese processing and ripening do not usually affect calcium bioavailability. On the other hand, scientific evidence collected in recent publications, on the relationship of cheese consumption with different fat contents on health, shows that there is no association or even an inverse association with cardiovascular risk and all-cause mortality. Cheese has a high content of proteins of high nutritional quality and calcium and contributes with significant health benefits, within the framework of a healthy diet.

Cream and butter

Cream has been defined as milk rich in fat (fat emulsion in skimmed milk), with low levels of protein and lactose. There are different types of cream, depending on the fat content (from 12 to 55%) and the type of heat treatment. The concentration of vitamin A increases with the fat content; e.g. in 40% cream, it is 8–12 times higher than in milk.

Butter is essentially milk fat (82% fat) with 16% water and 2% non-fat solids. It is made from cream, sweet or acidified, and can be marketed with or without salt. In its preservation, the favorable effect of vitamin E and the unfavorable effect of copper levels, the type of packaging, and storage conditions have a decisive influence on the degree of protection of butter against oxidation. The vitamin A content can be very markedly reduced by storage in light-permeable containers.

Role of milk and dairy products in disease prevention and health promotion

The relationship between diet and health has been known since ancient times, and there is increasing scientific evidence on the relationship between diet and the prevalence of chronic noncommunicable diseases such as obesity, metabolic syndrome (MS), CVD, type 2 diabetes mellitus (DM2), osteoporosis and even some types of cancer, but also for its impact on health promotion. Among the foods available in our diet, the role of milk and dairy products stands out for the wide range of nutrients present in their composition, such as fat, proteins, carbohydrates, minerals, and vitamins - with a good balance between their majority constituents -, which play a fundamental role in the diet of children, adolescents, and adults. Recent studies, moreover, consider milk as one of the most important sources of natural bioactive components and its health-promoting effects are the result of the interaction of all nutrients, which go beyond the simple sum of individual effects, and by the effect of its matrix that increases the bioavailability of many of its nutrients and bioactive compounds.

The most recent scientific evidence allows concluding that dairy products are vital in the stages of growth and development. In addition, they are not related to all-cause mortality, enhance the maintenance of body composition, promote the reduction of certain types of fractures and loss of bone mineral density, and their consumption is associated with a lower risk of MS, DM2, and CVD, as well as an anti-inflammatory effect, favor a reduction in the risk of breast, colorectal, bladder and liver cancer and no association with other cancers, and may play a beneficial role in improving cognitive function.

Overall mortality

Considerable controversy exists regarding the association between milk and dairy consumption and mortality risk. In the available meta-analysis, relative risk ratios (RR) ranged from 0.96 to 1.01 per 200 g per day of dairy consumption, from 0.99 to 1.01 per 200–244 g of milk, and from 0.99 to 1.03 per 10–50 g of cheese (Cavero-Redondo et al., 2019). Yogurt intake equal to or greater than 200 g per day was significantly associated with a decrease in all-cause mortality. However, high-fat milk consumption was significantly associated with a higher risk of overall mortality, increased by 15%, CVD by 9%, and cancer by 17%. However, total dairy consumption was associated with a lower CVD mortality risk of 7%. Dose-response analysis has revealed a significant non-linear association of total dairy consumption with all-cause and CVD mortality (Naghshi et al., 2022)

Pregnancy, maternal health, and newborns

The maternal diet should enable the mother to provide the nutrient reserves necessary for adequate fetal development, as well as good health and quality of life in childhood and later adulthood. Among the food groups, dairy plays a very important role in meeting these goals due to its high nutrient density and bioavailability, as well as its widespread consumption. There is evidence regarding maternal milk intake during pregnancy being positively associated with infant birth weight and length.

Regarding maternal health and during pregnancy, the benefits of yogurt consumption have been studied. It has been observed that the consumption of this product improves metabolic, inflammatory and infectious parameters associated with pregnancy, being remarkable the decrease in preterm births. The use of yogurt is promising as a nutritional supplement in pregnancy. In addition, the consumption of a greater amount of dairy products during pregnancy has been associated with higher birth weight and length of the baby at birth and with a lower risk of having small-for-gestational-age and low birth weight babies.

Growth and development

The relationship between milk consumption during the pediatric age and increased linear growth and bone mineralization has been extensively studied. The consumption of dairy products during that period is justified to promote growth and development because they are sources of energy, macronutrients, and micronutrients, being remarkable in their contribution of protein, calcium, phosphorus, magnesium, and several vitamins.

The consumption of dairy products increases the secretion of insulin-like growth factor type I (IGF-1) which benefits skeletal development. Specific dairy compounds, such as some peptides derived from the hydrolysis of caseins, are believed to help calcium absorption from dairy foods to be more consistent throughout the day. In developed countries, a daily intake of 500 mL is recommended for children under 9 years of age, and for adolescents, an intake greater than 600 mL per day.

Dietary supplementation with dairy has been described to significantly increase bone mineral content during childhood. However, results regarding a possible association between dairy consumption and linear growth have been inconclusive.

Weight gain

Several studies and metanalysis have documented that dairy is not associated with weight gain, even when their intake was increased, including low-fat milk, whole milk, and cheese. On the opposite, it was reported that dairy consumption had little effect on body mass index (BMI), and led to a reduction in fat mass, an increase in lean body mass, and a decrease in waist circumference. When comparing the type of dairy products to prevent long-term weight gain, the best was yogurt followed by cheese and then milk. On the other hand, increasing total dairy intake without energy restriction in adults does not affect body composition. In the context

of an energy-restricted diet, however, increased dairy intake results in lower fat mass and body weight but has no conclusive effects on waist circumference or lean mass (Lopez-Sobaler et al., 2020).

Effects of dairy consumption on body weight and body composition have been inconsistently observed in randomized control trials (RCTs). In a meta-analysis that included 37 RCTs, dairy consumption decreased body weight, body fat, and waist circumference, among adults with energy restriction. Nonetheless, high dairy consumption in the absence of caloric restriction might increase body weight (Geng et al., 2018)

Osteoporosis

Having adequate peak bone mass and maintaining it as long as possible is important to prevent osteoporosis. One of the most important sources of calcium is dairy. Milk should be consumed regularly at different stages of life to try to prevent bone loss.

Dairy product consumption is related to hip fracture risk, but numerous studies have found inconsistent results. Higher levels of dairy intake, compared to lower levels, are associated with a 5% decrease for osteoporotic fractures at any location, 13% for hip fractures and 16% for vertebral fractures. Regarding bone mineral density, studies described between 1.7 and 3% lower bone mineral density in young and postmenopausal women with low milk intake in their youth and a positive correlation between milk consumption and change in bone mineral density at the radius in women over 65 years of age.

Hip fracture risk reduction has also been studied concerning yogurt consumption, which is often associated with healthy lifestyles and dietary patterns that contribute to improved bone health. Accumulating evidence from studies following numerous individuals over a long period suggests that the risk of hip fracture may be maintained among people who consume dairy and that higher milk or yogurt consumption may be associated with a lower risk of hip fracture (Cuesta-Triana et al., 2019).

Metabolic syndrome

MS is a set of clinical and biochemical alterations that result in altered carbohydrate and lipid metabolism with negative health consequences. It is defined as the presence of three or more CV risk factors: abdominal obesity, elevated blood pressure, elevated glucose and serum TAG concentrations and decreased HDL-cholesterol levels relative to normal ranges MS is associated with two of the most prevalent diseases worldwide, DM2 and CVD.

Numerous studies have associated the consumption of dairy products with a lower risk of MS. In a systematic review (SR) and meta-analysis, it could be observed that subjects who consumed higher amounts of total dairy products had a 27% lower risk of developing MS than people classified in the lowest consumption categories. Intake of low-fat dairy products and total yogurt consumption has been associated with a decreased risk of developing over time SM of 23% and 26%, respectively. It could be observed that, for one serving of yogurt consumed per day, there was a 23% decrease in risk. In the case of low-fat yogurt, a decrease of 28% was observed, and for the whole yogurt, 19%. Total milk consumption was associated with a 21% decrease in the risk of this disease.

Diabetes and insulin resistance

Total dairy intake has been associated with a lower risk of DM2, especially for yogurt and low-fat dairy products, the association with cheese being more moderate. In a SR and meta-analysis, the scientific evidence indicated that higher total dairy consumption is associated with a decreased incidence of DM2 (between 9% and 14%). In addition, a decrease of 17–19% was observed for low-fat dairy products, 18% specifically for low-fat milk, and 14%–26% for yogurt. (Alvarez-Bueno et al., 2019). Analyses, according to the amount of consumption, showed a decrease in the risk of DM2 of 3%–7% for total dairy products (consumption of 200–400 g per day) and for low-fat dairy products, a decrease of 9%–12% (intake of 200 g per day).

Numerous investigations suggest that dairy intake, especially low-fat dairy products, presents a beneficial effect on an indicator of insulin resistance called HOMA-IR (Homeostasis Model Assessment-Insulin Resistance), waist circumference, and body weight.

Cardiovascular disease (CVD)

Among the dairy components with the greatest relevance in CVD, the following stand out:

Dairy proteins

The major dairy proteins, caseins, and whey proteins, contain all the essential amino acids and have high digestibility and biological value, and are therefore defined as high-quality proteins. Due to their high lysine content, milk proteins can complement proteins from other foods, such as cereals. In addition, during the process of gastrointestinal digestion or by enzymatic hydrolysis during fermentation and ripening of fermented milks and cheeses, "bioactive peptides" with important biological activities are generated, among which is their antihypertensive capacity, by inhibition of the angiotensin-converting enzyme (ACE), which is associated with a reduction of blood pressure (BP) and lower risk of hypertension and CVD. Therefore, milk proteins in isolation, or together with other milk components, such as mineral elements, are associated with a lower risk of hypertension. These functionalities of peptides in human health and physiology also include antimicrobial, antioxidative, antithrombotic, opioid, anti appetite,

immunomodulatory, and mineral-binding activities. The results of several clinical studies and meta-analyses have shown that milk proteins alone, or together with other milk components such as mineral elements, are associated with a lower risk of hypertension.

Milk fat

As stated before, unlike other animal fats, the exclusive presence in milk fat of short- and medium-chain fatty acids (from C4 to C10), has been shown to not affect LDL-cholesterol levels in the blood. In addition, TAG from these FA in the diet are hydrolyzed during digestion and absorbed from the intestine into the circulatory system without TAG resynthesis. Hence, they are used as a fast source of energy, have a low tendency to accumulate in adipose tissue, and are therefore recommended in weight control programs, and increasing their intake increases muscle mass and reduces the percentage of body fat. The stearic acid (C18) in milk fat (10%–12%) is considered neutral from the viewpoint of human health, although it is undoubtedly effective for reducing plasma cholesterol as well as oleic acid, also present in milk fat in high concentrations, as has been mentioned. Therefore, only a third of the concentration of FAs present in milk, corresponding to the SFA C12, C14, and C16 (lauric acid, myristic acid, and palmitic acid), could be considered detrimental if consumed in excess. Recent research has shown that milk matrix components, mainly calcium, peptides, phosphorus, and fat globule membrane modify the blood lipid response to SFA intake (Thorning et al., 2017). As a result, these SFA considered hypercholesterolemic would have no impact on CV health parameters when supplied in the dairy matrix. Milk fat, and especially goat milk fat, also has methyl-branched saturated acids, the importance of which is due to their anticancer properties, reported in tumor cell cultures and maintenance of microbiota and the inhibition of growth of some tumors in animal models. C15 and C17 FAs with an odd-numbered chain of carbon atoms are considered biomarkers of consumption of dairy fat and have been used to document non-association or even a protective effect against type 2 diabetes and CVD. The presence in milk fat of naturally occurring TFA has also been described, although scientific evidence concludes that consumption of moderate amounts - less than 1% of energy - does not contribute to increased CV risks.

In the last decade, a large number of clinical studies have been conducted on the effect of dairy consumption and CVD. Dairy consumption, regardless of fat content, is not associated with an increased risk of CVD. In a SR and meta-analysis study regarding the effect of dairy consumption with CVD, it is concluded that there is no association of whole or skimmed dairy consumption, including dose-response studies, with CVD and, in addition, an inverse association with stroke was found (Fontecha et al., 2019). This study also reports that dairy consumption has a moderate effect on risk factors associated with CVD, especially blood pressure and plasma lipids, decreasing systolic and diastolic blood pressure and LDL-cholesterol (Fontecha et al., 2019). Along the same lines, according to data obtained in other studies to estimate associations between specific dairy product substitutions and risk of myocardial infarction (MI), it was concluded that intake of whole yogurt or cheese, regardless of fat content, was associated with a lower risk of developing MI. Indeed, it is clear that the health effects of saturated fats seem to depend on the food source, and whole or low-fat dairy products have no impact on or protection against CVD when consumed in moderation and as part of a varied and balanced diet. Therefore, current dietary recommendations recognize the contribution of dairy products to a healthy diet, since their consumption involves raising the levels of multiple nutrients, mineral elements, vitamins and proteins of high nutritional quality, without risk of CVD, hypertension and general mortality.

Minerals

In recent years, many publications have reported scientific evidence of the benefits of consuming calcium-rich dairy products, with health claims approved by the EFSA-EU, not only in dental health and osteoporosis prevention, in weight control and body mass index (especially in obese individuals), but also protection against hypertension and cardiovascular risks. In this context, it has been reported that diets with high levels of calcium gave rise to greater fecal excretion of fat and they correlated with benefits in cardiovascular markers such as lower levels of total cholesterol and LDL-cholesterol. In addition, other elements are present in milk in very small quantities; some of them are indispensable in nutrition and their content can be increased by external factors.

Cancer

Breast cancer

Breast cancer is the most common cancer in women with almost 1.7 million new cases diagnosed in 2012 and is the second most common cancer overall. This represents about 12% of all new cancer cases and 25% of all cancers in women.

Regarding breast cancer, in an SR it was observed that a higher intake of milk and dairy products was associated with a lower risk of breast cancer in most studies. To better understand whether dairy consumption may influence the risk of developing this type of cancer, more research is needed on the biological mechanisms, including the relationship of different dairy products with hormones, as well as the study of other bioactive components contained in these foods.

Colorectal cancer

Colorectal cancer (CRC) is the third most commonly diagnosed malignancy and the fourth leading cause of cancer-related death worldwide. In 2017, 1.8 million incident cases of colon and rectal cancer were diagnosed, and 896,000 deaths occurred. The global CRC burden is estimated to increase by 60% to over 2.2 million new cases and 1.1 million deaths by 2030.

High dairy consumption has been associated with a lower risk of developing CRC in any anatomic location. In an SR and meta-analysis with 15 cohort studies and 14 case-control studies (>22,000 cases), a significant and consistent decrease in CRC risk associated with higher dairy consumption, approximately 20%, compared with lower consumption was observed. These studies also showed a protective association between low-fat milk consumption and colon cancer, 27%. Cheese consumption was associated with a 15% lower risk of CRC and 26% lower risk of proximal colon cancer. Low-fat milk intake was associated with a lower risk of CRC. Cheese consumption was associated with CRC prevention, specifically proximal colon cancer (Barrubés et al., 2019).

In another meta-analysis of 31 prospective cohort studies, which included 24,964 and 2302 cases for CRC incidence and mortality, respectively, a 29% lower risk of death from CRC in subjects with high dairy consumption compared with those with low intakes of dairy products was found, but each type of dairy consumption did not show a significant association (Jin et al., 2020).

Moreover, numerous studies have suggested that calcium-rich foods, such as dairy, may have a more effective role than calcium in the form of supplements concerning the risk of colorectal adenomas (CA).

Ovarian cancer

Ovarian cancer is one of the most common cancers in women, accounting for 2.5% of all malignancies. In 2018, 295,414 new cases and 184,799 deaths from ovarian cancer were reported worldwide. Of all gynecological malignancies, this cancer is the one associated with the worst prognosis.

Results on the role of dairy, calcium, and vitamin D on ovarian cancer risk remain controversial. Some evidence suggests that whole milk intake may contribute to an increased risk of ovarian cancer, whereas low-fat milk, dietary calcium and vitamin D, may reduce it. In a SR, in comparisons of higher versus lower intakes, it was observed that a high intake of whole milk was associated with an increased risk of ovarian cancer, a 35% increase. However, a decreased risk was observed at the highest intakes of low-fat milk, 16% decrease and dietary calcium and vitamin D, 29% and 20%, respectively. For every 100 g per day increase in consumption, higher risks of ovarian cancer were found for total dairy products, 3% increase and for whole milk 7% increase. However, lower risks were observed for 100 g per day increase in consumption of low-fat milk (5%), cheese (13%), dietary calcium (4%), total calcium (2%), and dietary vitamin D (8%).

Liver cancer

Primary liver cancer is ranked as the sixth most common form of cancer worldwide and the third leading cause of cancer death, with approximately 800,000 cases in 2012. Analysis according to amounts consumed showed that the risk of liver cancer decreased by 5.4% with an increase of 40 g per day of yogurt consumption.

Prostate cancer

According to the latest IARC report, prostate cancer is the second most common cancer in men worldwide, with more than 1.1 million new cases diagnosed in 2012 and 307,000 deaths recorded, accounting for 15% of diagnosed cancers. In more developed regions, prostate cancer is the most frequent cancer in men (759,000 cases).

The literature on the association of dairy consumption and the risk of developing prostate cancer is controversial and limited. The overall positive association between milk consumption and the risk of prostate cancer development and prostate cancer mortality has been documented in multiple epidemiological studies. However, there is limited literature on the association between types of milk, as classified by fat content (skim, low fat, and whole), and the risk of developing prostate cancer. In an SR it was observed that high consumption of total dairy products compared to low consumption accounted for a 7% increased risk of developing this cancer (400 g per day). No significant associations were found for all types of milk and cheese. Further research is needed to examine the relationship between dairy consumption and the increased risk of prostate cancer that has been observed in some studies due to numerous controversies.

Bladder cancer

According to the International Agency for Research on Cancer (IARC), bladder cancer is the ninth most common cancer worldwide, with 430,000 new cases diagnosed in 2012 (3% of all new cancer cases).

An SR and meta-analysis observed that average total dairy intake compared to low intake was associated with a lower risk of bladder cancer (10%). Milk intake also reduced the risk by 10% and fermented dairy products by 13%. High consumption, compared to low, was significantly associated with a lower risk of this cancer: for milk, a decrease of 11% and for fermented dairy products, a decrease of 12%. However, high compared to low consumption of whole milk was significantly associated with a higher risk (21% increase). Another recently published meta-analysis has described that the intake of milk and dairy products decreases the risk of bladder cancer (Wu et al., 2020).

Cognitive function

The activity, functioning, and maintenance of the nervous system depend substantially on the nutrients that are incorporated in the daily intake. Although the importance of breastfeeding for the cognitive development of the infant is well known, recent studies

indicate that dairy consumption also supports proper brain function during adulthood and later life. Milk components with action on brain function mechanisms include calcium, proteins, especially α-lactalbumin, bioactive peptides, vitamin B12, and the phospho- and sphingolipids of the MFGM, which affect different aspects of the nervous system, such as learning and memory abilities, mood maintenance, as well as their potential activity in the prevention of age-related cognitive decline (Bermejo-Pareja et al., 2020). Thus, clinical studies have shown that dairy consumption in adults, in addition to increasing nutrient intake, also has a positive impact on the performance of cognitive functioning. Dairy proteins provide all the essential amino acids essential in the synthesis of neurotransmitters and neuromodulators that contribute to proper brain function. Alpha-lactoalbumin (α-La) is a good source of tryptophan and cysteine, precursors of the neurotransmitter serotonin and glutathione respectively.

Longitudinal nutritional studies included in a SR, describe that, both in the adult population and in the elderly, those individuals who consumed dairy products had a better global cognitive function which they explain as a consequence of the nutritional contribution of dairy products and especially by components such as vitamin B12, calcium and vitamin D that have been associated with neuroprotective, antioxidant and anti-inflammatory effects. The lipids in MFGM have been shown to provide important benefits to nervous system functions such as memory by reducing the risk of senile dementia and cognitive dysfunction in old age. These lipids also appear to play an important role in immune-related diseases and anti-inflammatory responses. Among the different PL, phosphatidylserine and sphingomyelin present in MFGM have been linked to positive effects on diseases such as Alzheimer's disease, depression, and stress. Clinical trials have been conducted using MFGM concentrates and the results are encouraging and would allow MFGM to be considered as a potential bioactive ingredient in functional foods. In conclusion, these studies suggest that dairy products may play a beneficial role in improving cognitive function and thus aid thought processing and mental capacity.

Conclusions

Milk and dairy products provide practically all the essential nutrients for the different stages of life, especially during childhood and adolescence, but also in adulthood. Milk is still considered the most complete food from a nutritional point of view, providing proteins with all the essential amino acids and high bioavailability, fat with a great variety of short and medium-chain fatty acids, and carbohydrates such as lactose, as well as a great contribution of minerals and vitamins. All this in a balanced and low caloric value.

Excessive consumption of milk and dairy products is associated with several chronic diseases, mainly CVD and some cancers. However, that moderate dairy consumption has many benefits on health outcomes. Indeed, milk and dairy contribute to the maintenance of body composition, favor the reduction of some types of fractures and loss of bone mineral density, and their consumption is associated with a lower risk of MS, DM2, and CVD, as well as an anti-inflammatory effect, they favor a reduction in the risk of breast, colorectal, bladder, and liver cancer and there is no association with other cancers or all-cause mortality. Finally, they may play a beneficial role in improving cognitive function. Therefore, the most recent scientific evidence allows us to conclude that dairy products are vital in all stages of growth and development.

References

Álvarez-Bueno, C., Cavero-Redondo, I., Martínez-Vizcaino, V., Sotos-Prieto, M., Ruiz, J.R., Gil, A., 2019. Effects of milk and dairy product consumption on type 2 diabetes: overview of systematic reviews and meta-analyses. Adv. Nutr. 10, S154–S163.

Barrubés, L., Babio, N., Becerra-Tomás, N., Rosique-Esteban, N., Salas-Salvadó, J., 2019. Association between dairy product consumption and colorectal cancer risk in adults: a systematic review and meta-analysis of epidemiologic studies. Adv. Nutr. 10, S190–S211.

Bermejo-Pareja, F., Ciudad-Cabañas, M.J., Llamas-Velasco, S., Tapias-Merino, E., Hernandez Gallego, J., Hernandez-Cabria, M., Collado-Yurrita, L., Lopez-Arrieta, J.M., 2020. Is milk and dairy intake a preventive factor for elderly cognition (dementia and Alzheimer's)? A quality review of cohort surveys. Nutr. Rev. 1–15.

Cavero-Redondo, I., Alvarez-Bueno, C., Sotos-Prieto, M., Gil, A., Martinez-Vizcaino, V., Ruiz, J.R., 2019. Milk and dairy product consumption and risk of mortality: an overview of systematic reviews and meta-analyses. Adv. Nutr. 10, S97–S104.

Cuesta-Triana, F., Verdejo-Bravo, C., Fernández-Pérez, C., Martín-Sánchez, F.J., 2019. Effect of milk and other dairy products on the risk of frailty, sarcopenia, and cognitive performance decline in the elderly: a systematic review. Adv. Nutr. 10, S105–S119.

De la Fuente, M.A., Juárez, M., 2015. Chapter 28: milk and dairy products. In: Garrigues, S., de la Guardia, M. (Eds.), Handbook of Mineral Element in Foods: Chemistry, Analysis and Food Composition. Wiley-Blackwell, Chichester; UK, pp. 645–668.

Fontecha, J., Calvo, M.V., Juarez, M., Gil, A., Martínez-Vizcaino, V., 2019. Milk and dairy product consumption and cardiovascular diseases: an overview of systematic reviews and meta-analyses. Adv. Nutr. 10, S164–S189.

Fontecha, J., Juárez, M., 2017. Recent advances in dairy ingredients and cardiovascular diseases with special interest in milk fat components. In: Watson, R.R., Collier, R.J., Preedy, V.R. (Eds.), Milk in Human Health and Disease across the Lifespan. Academic Press, London, pp. 251–261.

Geng, T., Qi, L., Huang, T., 2018. Effects of dairy products consumption on body weight and body composition among adults: an updated meta-analysis of 37 randomized control trials. Mol. Nutr. Food Res. 62 (1).

Jin, S., Kim, Y., Je, Y., 2020. Dairy consumption and risks of colorectal cancer incidence and mortality: a meta-analysis of prospective cohort studies. Cancer Epidemiol. Biomarkers Prev. 29 (11), 2309–2322.

López-Sobaler, A.M., Aparicio, A., López Díaz-Ufano, M.L., Ortega, R.M., Álvarez-Bueno, C., 2020. Effect of dairy intake with or without energy restriction on body composition of adults: overview of systematic reviews and meta-analyses of randomized controlled trials. Nutr. Rev. 78 (11), 901–913.

Naghshi, S., Sadeghi, O., Larijani, B., Esmaillzadeh, A., 2022. High vs. low-fat dairy and milk differently affects the risk of all-cause, CVD, and cancer death: a systematic review and dose-response meta-analysis of prospective cohort studies. Crit. Rev. Food Sci. Nutr. 62 (13), 3598–3612.

Park, Y.W., Juárez, M., Ramos, M., Haenlein, G.F.W., 2007. Physico-chemical characteristics of goat and sheep milk. Small Rumin. Res. 68, 88–113.

Thorning, T.K., Bertram, H.C., Bonjour, J.F., de Groot, L., Dupont, D., Feeney, E., Ipsen, R., Lecerf, J.M., Mackie, A.S., McKinley, M.C., Michalski, M.-C., Remond, D., Riserus, U., Soedamah-Muthu, S.S., Tholstrup, T., Weave, C., Astrup, A., Givens, I., 2017. Whole dairy matrix or single nutrients in assessment of health effects: current evidence and knowledge gaps. Am. J. Clin. Nutr. 105, 1033–1045.
Wu, J., Yu, Y., Huang, L., Li, Z., Guo, P., Xu, Y.W., 2020. Dairy product consumption and bladder cancer risk: a meta-analysis. Nutr. Cancer 72 (3), 377–385. https://doi.org/10.1080/01635581.2019.1637909. PMID: 31295044.

Further reading

Amigo, L., Fontecha, J., 2022. Goat milk. In: McSweeney, P.L.H., McNamara, J.P. (Eds.), Encyclopedia of Dairy Sciences, third ed. Elsevier, Academic Press Inc., San Diego, USA, pp. 797–808.
Fox, P.F., 1992. Advances in dairy chemistry. In: Proteins, vol. 1. Blackie Academic & Professional, Londres.
Fox, P.F., 1995. Advances in dairy chemistry. In: Lipids, , 2ª ed.vol. 2. Chapman & Hall, Londres.
Gómez-Cortés, P., Amigo, L., 2022. Sheep milk. In: McSweeney, P.L.H., McNamara, J.P. (Eds.), Encyclopedia of Dairy Sciences, third ed. Elsevier, Academic Press Inc., San Diego, USA, pp. 676–687.
Jensen, R.G., 2002. The composition of bovine milk lipids: January 1995 to December 2000. J. Dairy Sci. 85, 295–350.
López-Plaza, B., Bermejo, L.M., Santurino, C., Cavero-Redondo, I., Álvarez-Bueno, C., Gómez-Candela, C., 2019. Milk and dairy product consumption and prostate cancer risk and mortality: an overview of systematic reviews and meta-analyses. Adv. Nutr. (10), S212–S223.
Malmir, H., Larijani, B., Esmaillzadeh, A., 2020. Consumption of milk and dairy products and risk of osteoporosis and hip fracture: a systematic review and meta-analysis. Crit. Rev. Food Sci. Nutr. 60, 1722–1737.
Mena-Sánchez, G., Becerra-Tomás, N., Babio, N., Salas-Salvadó, J., 2019. Dairy product consumption in the prevention of metabolic syndrome: a systematic review and meta-analysis of prospective cohort studies. Adv. Nutr. 10, S144–S153.
Walstra, P., Wouters, J.T.M., Geurts, T.J., 2005. Dairy Science and Technology, second ed. CRC Press.

Mycotoxins: Toxicity, occurrence, risk assessment and prevention

L. Manyes and G. Font, Department of Preventive Medicine and Public Health, Food Science, Toxicology and Forensic Medicine, Faculty of Pharmacy, University of Valencia, Spain

© 2023 Elsevier Ltd. All rights reserved.

This is an update of J.D. Groopman, T.W. Kensler, F. Wu, Food Safety: Mycotoxins – Occurrence and Toxic Effects, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 337–341, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00122-7.

Introduction 492
Toxicity 493
Occurrence and human exposure of mycotoxins 495
Exposure and risk assessment of mycotoxins 496
Exposure assessment 496
Risk assessment 496
Aflatoxins 496
Ochratoxin A 497
Fumonisins 497
Deoxynivalenol 497
T-2 and HT-2 498
Zearalenone 498
Patulin 498
Mycotoxin regulations 498
Mycotoxin mitigation 499
Climate change 499
Conclusions 500
References 500
Further reading 500

Key points

- To obtain an overview of the toxicity of major mycotoxins
- To have information about the occurrence and human exposure to mycotoxins
- To understand the process of mycotoxin risk assessment
- To have information about mycotoxin regulations
- To know the mitigation processes for reducing the presence of mycotoxins

Introduction

Mycotoxins are secondary metabolites produced by microfungi, capable of causing diseases and death in humans and animals. The chemical structures of mycotoxins vary considerably, but they are all relatively low molecular mass organic compounds (Fig. 1).

The untoward effect of fungi was known in ancient times. In the seventh and eighth centuries BC the festival "Robigalia" was established to honor the god Robigus, who had to be propitiated to protect grain and trees. It was celebrated on April 25th because that was the most likely time for crops to be attacked by plant rust diseases or mildew. In the Middle Ages, outbreaks of ergotism caused by ergot alkaloids from *Claviceps purpurea* reached epidemic proportions, mutilating and killing thousands of people in Europe. Ergotism was also known as *ignis sacer* (holy fire) or St Anthony's fire because, at the time, it was thought that a pilgrimage to the shrine of St Anthony would bring relief from the intense burning sensation experienced. Renewed interest in mycotoxins as causing diseases resulted from the death of thousands of turkeys and ducks, named Turkey X disease in England in the early 1960s. The animals were fed diets containing peanut meal contaminated by aflatoxins (AFs), produced by the fungus *Aspergillus flavus* (Janik et al., 2020).

The use of certain molds in the production of cheese and salami and in the fermentation of beer and wine has a long history. Molds are also used in the production of drugs. The classification of mold metabolites as antibiotics or mycotoxins is based on their toxicity or beneficial effect in treating diseases.

Mycotoxin-producing molds may grow on many foods before or after harvest, during storage, especially in hot and humid environments. Most mycotoxins are chemically stable and persist after food processing. Several hundreds of mycotoxins have been

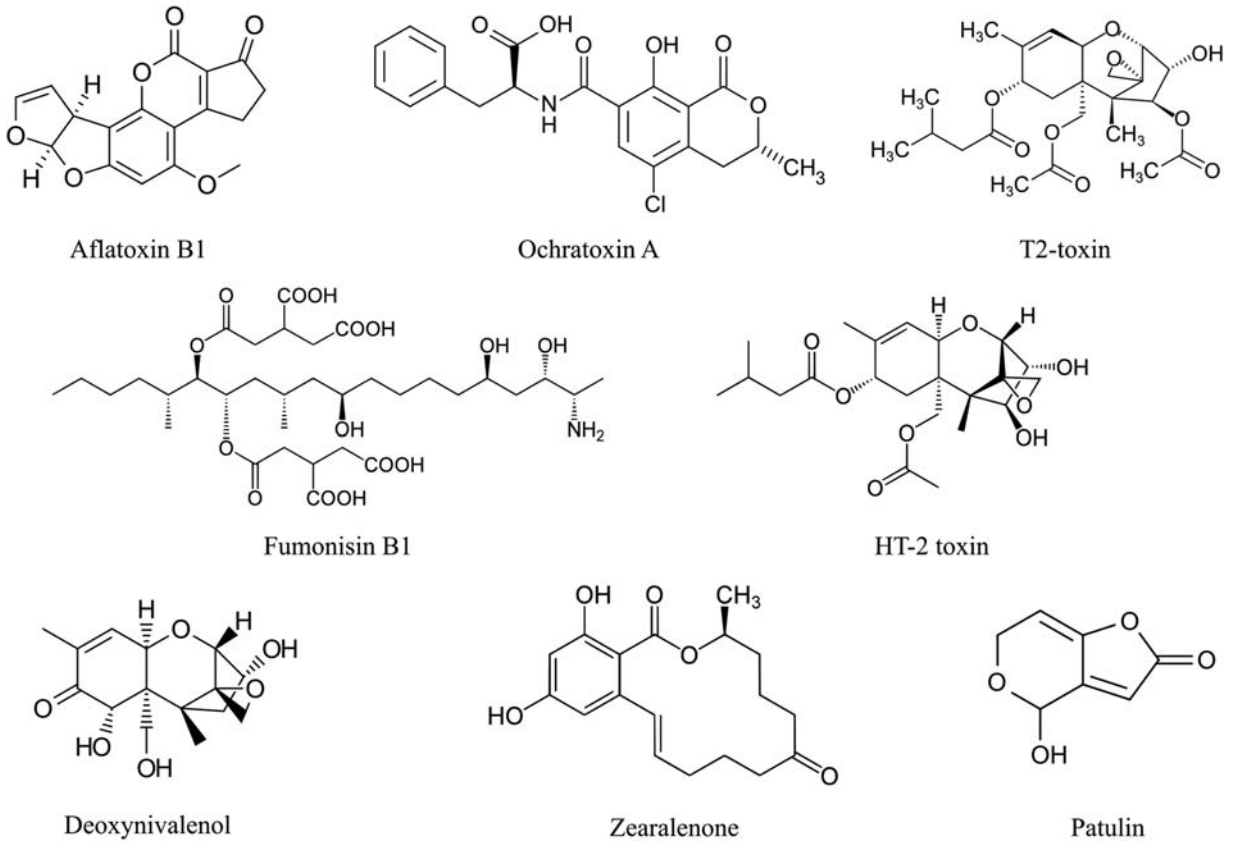


Fig. 1 Structures of major mycotoxins.

identified, but the most common concerning human and livestock health are AFs, ochratoxin A (OTA), fumonisins (FBs), zearalenone (ZEA). patulin, (PAT) and trichothecenes T-2 and HT-2, and deoxynivalenol (DON).

Mycotoxins appear throughout the entire food chain, from field crops to raw or processed food and feed. Some unprocessed foods susceptible to mycotoxin contamination and contributing to exposure include cereals, oilseeds, fruits, vegetables, nuts, dried fruits, coffee beans, cocoa beans, and spices. As mycotoxins are not destroyed during processing, cereal-based products (bread, pasta, breakfast cereals, etc.), beverages (wine, beer, coffee, cocoa, juices, etc.) are important sources of exposure to mycotoxins, as are foods of animal origin (milk, cheese) and baby foods (**Table 1**).

Despite many years of research and the introduction of improved agricultural and manufacturing practices in the food chain, mycotoxin occurrence remains a global problem. Significant economic losses are associated with their impact on human health, animal welfare and productivity, and both domestic and international trade. Regulation of mycotoxins varies in the different regions of the world. At the global level, the Codex Alimentarius Commission of International Food Standards, the joint intergovernmental body of the Food and Agriculture Organization (FAO) and World Health Organization (WHO), is the main operator in standard setting. Because compliance with mycotoxin concentrations in food and feed maximum limits (MLs) allows food and feed business operators to place their commodities on the market, a large amount of analytical occurrence data is generated for this purpose.

Toxicity

Chemical structures and characteristics of major mycotoxins are shown in **Fig. 1** and **Table 1**. A broad range of adverse health effects have been identified for mycotoxins in animals and humans (Marin et al., 2013)

Aflatoxins are acutely and chronically toxic compounds. AFB1 is a hepatocarcinogen, one of the most potent known, and long-term chronic exposure to very small amounts of this toxin through the diet has important consequences for human health. AFB1 is a potent chromosome-damaging mutagen in a variety of plant, animal, and human cells. The carcinogenic effect of AFB1 has been studied in at least 12 different species and AFs are classified by IARC as a group 1 carcinogen. Slightly less potent as liver carcinogens than AFB1, and acting more as renal carcinogens, AFG1 and M1 appear to be toxicologically similar to AFB1, although they have not been studied extensively. Among the subacute and chronic effects of AFs are liver cancer, chronic hepatitis, jaundice, hepatomegaly

Table 1 Main characteristic of major mycotoxins and US-EU limits.

Mycotoxin	*Fungal species*	*Food commodity*	*Toxic effects*	*US FDA (μg/kg)*	*EU ([a]μg/kg)*
Aflatoxins B1, B2, G1, G2	*Aspergillus flavus*, *A. parasiticus*, *A. nomius*	Maize, wheat, rice, peanut, sorghum, pistachio, almond, ground nuts, tree nuts, figs, cottonseed, spices	Liver cancer, chronic hepatitis, jaundice, hepatomegaly, cirrhosis	20 for total	2-12 for B1 4-15 for total
AFM1	Metabolite of AFB1	Milk, milk products	Possible carcinogen	0.5	0.05 milk 0.025 in infant formulas and infant milk
Ochratoxin A	*Penicillium verrucosum*, *P. nordicum*, *A. ochraceus*, *A. carbonarius*	Cereals, dried vine fruit, wine, beer, grapes, coffee, cocoa, cheese	Carcinogenic, nephrotoxic, teratogenic, immunotoxic, and possibly neurotoxic properties	Not set	2–10
Fumonisins B1, B2, B3	*Fusarium verticillioides*, *F. proliferatum*	Maize, maize products, sorghum, asparagus	Neurotoxic effect Possible carcinogens	2000–4000	200–4000
Deoxynivalenol	*F. graminearum*, *F. culmorum*, *F. cerealis*	Cereals, cereal products	Decrease in growth and weight gain, immunotoxicity	1000	50–200
T-2, HT-2	*F. langsethiae*, *F. sporotrichioides*	Maize, wheat, barley, oat, rye	Immunotoxicity	15	25–1000
Zearalenone	*F. graminearum*, *F. culmorum*, *F. equiseti*, *F. verticillioides,*	Cereals, cereal products, maize, wheat, barley	Hyperestrogenic effects, reproductive toxicity	Not set	20–100
Patulin	*P. expansum*, *A. clavatus*, *P. patulum*	Apples, apple juice, and concentrate, pears, peaches, grapes, apricots, olives low acid fruit juices	Harms liver, kidney, intestinal tissues and immune system	50	10–50

[a](EC, 2006)

and cirrhosis, caused by the frequent ingestion of small amounts present in food. AFs can also affect the immune system (Cimbalo et al., 2020).

Ochratoxin A is a powerful toxin mainly affecting the kidneys, where it can cause acute and chronic effects, depending on the dose exposure and duration, as OTA accumulates in kidney tissues. Its nephrotoxic capacity has been demonstrated in all mammalian species in which it has been evaluated. Studies on its acute toxicity show different degrees of affect, depending on the animal species exposed. Dogs and pigs are especially susceptible. OTA has been shown to be a potent teratogen in mice, rats, hamsters, and chickens but, apparently, not in pigs. Its teratogenic and reproductive effects have been demonstrated. It is also known to affect the immune system in some mammals. It is genotoxic both *in vivo* and *in vitro*, but the mechanism by which it is genotoxic has not been established. Ultimately, OTA is a mycotoxin with carcinogenic, nephrotoxic, teratogenic, immunotoxic, and possibly neurotoxic properties. OTA has a long half-life in humans.

Effects of Fumonisins have been observed for years in the form of a fatal disease in horses and related species, called equine leukoencephalomalacia. Compared with AFs based solely on the dose necessary to produce acute poisoning, fumonisins have lower toxic capacity, but FBs in corn can reach very high concentrations, even higher than 300 mg/kg. The toxicity of FBs is mainly due to their effects on the synthesis of sphingolipids in neurons. Alterations in the synthesis of these sphingolipids occur almost immediately after oral FB exposure through contaminated product. The range of effects that FBs can produce in mammals is dependent on the species. For example, affected equines start losing their appetite, then appear lethargic and develop neurotoxic effects after consumption of contaminated feed. Generally, the liver is also affected and, in severe cases, significant liver lesions with fibrosis can be seen in areas of the central lobe. Porcines induce pulmonary edema and hydrothorax. In rats, the development of hepatocellular carcinomas has been observed. Further, FB1 has shown effects on fetuses in pregnant rats, decreasing litter weight and affecting bone development. In humans, there appears to be a relationship between consumption of highly contaminated corn in some parts of the world (China and South Africa) and development of esophageal cancer, but more studies are needed for confirmation.

Acute Deoxynivalenol intoxication in pigs is characterized by the appearance of vomiting (hence the common name as vomitoxin), rejection of feed, weight loss and diarrhea. Acute poisoning can produce necrosis in various tissues, such as the gastrointestinal tract and the lymphatic system. Subchronic oral exposure to DON in experimental animals (pigs, mice, rats) can also cause rejection of food and decreases in growth. Studies also suggest that DON may have effects on the immune system. However, there is no evidence that this toxin has carcinogenic, mutagenic or teratogenic effects. Symptoms of food poisoning in humans caused by consumption of DON contaminated wheat (0.34–8.4 ppm) in India in 1987, included abdominal pain, dizziness, headache, throat irritation, nausea, vomiting, diarrhea and presence of blood in the stool.

The T-2 toxin inhibits protein, RNA and DNA synthesis, and other biochemical mechanisms that can alter the integrity of cell membranes, synthesis of immunoglobulins and, therefore, humoral immunity. The immune system is most affected by the action of these toxins, which can increase susceptibility to infection. Alterations of cell membrane functions and lipid peroxidation are

responsible for many of the acute effects of T-2 and HT-2 toxins, including the necrotic lesions seen in contact areas. Apoptosis in tissues that are renewed with a high growth rate (such as bone marrow - inhibition of hematopoiesis) and in the immune system is responsible for the systemic toxicity that follows dietary exposure. T-2 toxin is rapidly metabolized into other products, with HT-2 toxin as the main metabolite. The metabolic processes are varied (hydrolysis, hydroxylation, etc.) and the distribution and excretion of the T-2 toxin and its metabolites is rapid. There is no significant information on the toxicity of most of its metabolites. T-2 and HT-2 are toxic to all animal species and to humans. Historical cases of intoxication in humans associated with the consumption of moldy cereal stored throughout the winter are described as alimentary toxic aleukia, characterized by sepsis, hemorrhages and general pancytopenia. The significant differences in sensitivity to T-2 and HT-2 toxins between monogastric and ruminant species are attributable to the effective presystemic elimination of the toxins exerted by the microbial flora of the rumen.

The most important effects of Zearalenone are on the reproductive system. The ability of ZEA to cause hyperestrogenic effects, particularly in sows, has been known for many years. Feeding breeding sows with a diet containing 50 mg/kg of pure ZEA caused abortions and stillbirths, while 10 mg/kg reduced litter size and weight of the piglets. Trials with less contaminated diets (0.25 mg/kg or less) showed different effects on the reproductive system of the sows. The pig has been shown to be the most sensitive domestic animal to ZEA, but effects have also been reported in calves, cows, and sheep. Studies to determine the subacute and sub-chronic toxicity of ZEA on various species for exposure times greater than 14 weeks show that most effects are due to its estrogenic capacity. The estrogenic capacity and potency of ZEA has been compared to other estrogenic derivatives of plants, suggesting that ZEA is one of the most potent natural xenoestrogens. Studies to determine the genotoxic and carcinogenic capacity of ZEA show that results appear to be highly dependent on the species. Although ZEA was considered as the possible causal agent of the outbreak of precocious pubertal development in thousands of girls registered in Puerto Rico, more studies are needed to confirm or not possible genotoxic and carcinogenic effects in humans. It has also been suggested that it may increase risk of cervical cancer.

Patulin harms liver, kidney, intestinal tissues, and the immune system by increasing ROS production and oxidative stress in mitochondria, and activating the unfolded protein response in the endoplasmic reticulum. Several studies have demonstrated its cytotoxicity, embryotoxicity, chromosomal aberration, DNA damage, and micronucleus formation. Patulin acute and sub-acute toxicities have been analyzed in mice and fish. Its exposure in humans is associated with immunological, neurological and gastrointestinal effects.

Occurrence and human exposure of mycotoxins

The presence of mycotoxins in food for both human and animal consumption is an issue of concern for health authorities. Factors that health authorities must consider include the large number of existing mycotoxins, which belong to different chemical families, and the limited knowledge regarding the interactions that occur among them and with other toxic substances present in food. Generally, mycotoxin prevalence in feed samples is higher than in food. Cereals have been the most analyzed type of food for mycotoxin detection around the world (Alshannaq and Yu, 2017).

In the last century, mycotoxin allusion was exclusively made to the large families (AFs, OTA, FBs, trichothecenes, ZEA and some of their metabolites) and in relatively high quantities while using liquid chromatography coupled to ultraviolet and fluorometric detection systems. Today, advances in extraction, purification and concentration methods, coupled with widespread use of detection systems based on tandem mass-mass spectrometry, enable detection in the ng/Kg, including the classic mycotoxins, as well as emerging ones (enniatins, beauvericin, fusaproliferin and alternaria), together with the derivatives that are present in food (glucurunoconjugated, sulfated, hydroxylated, acetylated) known as modified mycotoxins. Therefore, the spectrum of compounds classified as mycotoxins is becoming broader. Analytical reports point to occurrence percentages in crops, ranging from 50 to 90% containing at least one mycotoxin, with about half displaying multiple contamination at any detectable concentration.

Mycotoxin values have been detected in raw and ready-to-eat foods in different quantities, resulting from agricultural handling, factory treatment, preservation efforts and packaging. Values provided in reports from scientific institutes and enterprises are the product of analyses carried out at a precise moment and may vary from one region and year to another for the same crop or product. For example, AF prevalence in unprocessed food-grade cereals has been estimated at around 55%, oscillating from 15% in America to 63% in Asia, with 10% in cashew nuts to 40% in pistachio and Brazil nuts, 45% in dried figs, and 53% in coffee. Prevalence varied for OTA, depending on the type of food crop, but was approximately 70% for beer, 50% for cereals (from 20% in Europe to 42% in America), and 36% in coffee. FBs have been found in approximately 70% of maize samples worldwide. T-2 and HT-2 toxins have been globally detected in cereals, ranging from 71% to 10% in barley, 100%–4% in rice and 22%–1% in maize. DON occurrence was 50% in oats, 70% in barley, wheat and rice and 90% in maize, 60% in beer and 43% in coffee. Worldwide incidence of 46% ZEA was calculated in cereals, being the lowest in Asia (15%) and the highest in Africa (59%).

Studies to assess the presence of mycotoxins are usually based on the analysis of one mycotoxin in one type of food although, in recent years, multi-mycotoxin analyses have become widespread. Foods may simultaneously contain several mycotoxins produced by fungi of different species. Co-occurrences found varied from 2 to 8 mycotoxins in the same food. For example, co-occurrence was found in durum wheat from Italy in 70% of the samples, of which 35% contained 3 enniatins. In wheat-based foods from Spain, 10% of the samples contained DON plus HT-2 or NIV. Livestock feedstuffs showed co-occurrence in at least 85% of the samples analyzed and, of these, 22% contained 5 mycotoxins and 4% eight mycotoxins. As one of the food safety agencies' concerns is to know the total intake of a certain mycotoxin, analytical approaches are carried out on a certain set of foods consumed during 24 h by a specific subset of the population. Therefore, all these studies lead to a partial knowledge of reality.

Food undergoes various treatments, mainly thermal, before being consumed and heat generally affects mycotoxins by hydrolyzing them. Therefore, concentrations in ingested food are lower than before preparation. Another aspect that deserves attention is the bioaccessibility and bioavailability of mycotoxins in humans and animals. Foods that generate a higher fecal content, due to the presence of indigestible substances, mainly dietary fiber, have a greater capacity to retain mycotoxins and, as they do not solubilize in the intestinal lumen, they are not accessible to the intestinal barrier cells. In addition, not all mycotoxins soluble in intestinal fluids are absorbable and only a percentage reaches the bloodstream, usually less than 50% and sometimes not even 10%, due to the different compound polarities presented, even within the same chemical family.

As occurrences in food are uncertain and variable, human biological monitoring for mycotoxin exposure through biological fluids is a useful tool. The most studied mycotoxin biomarkers are: AFB1adducts with albumin or lysine (AFB1-alb and AFB1-lys) in human plasma, AFB1-N7-guanine in urine, OTA and its metabolite OTα in urine, plasma, serum and blood and DON-glucuronides in urine. It should also be highlighted that human breast milk samples in Spain contained ZEA, HT-2 and NEO and AFM1. Importantly, in countries such as Tanzania, Mexico, Iran and Jordan, AFM1 has been detected in 100% of the samples analyzed. OTA has also been found in Chile, Turkey and Italy.

Exposure and risk assessment of mycotoxins

Risk assessment is the process of quantifying the magnitude and exposure, or probability, of a harmful effect to individuals or populations from certain agents or activities. There are four steps: hazard identification, dose-response or hazard assessment, exposure assessment, and risk characterization. Risk assessments using these principles have been conducted on the major mycotoxins as AFs, FBs, OTA, DON, and ZEA by various regulatory agencies for the purpose of setting food safety guidelines (**Table 1**). Apart from the well-established risk posed by aflatoxins, many uncertainties still exist about risk assessment for other major mycotoxins, often reflecting a lack of epidemiological data. Differences exist in risk management strategies, and in the ways different governments impose regulations and technologies to reduce mycotoxins in the food-chain. Regulatory measures have little impact on remote rural and subsistence farming communities in developing countries, in contrast to developed countries, where regulations are strictly enforced to reduce and/or remove mycotoxin contamination (IARC, 2012).

Exposure assessment

Exposure to mycotoxins depends on the level of these substances in different foods and on the intake of those foods. Large national and regional differences in the intakes of foods mean that exposure assessments are country specific. Monitoring data for food commodities of concern over several years provide data on levels. Estimates may then be refined by considering further modifications to the levels of mycotoxins in foods as consumed, including information on manufacturing and processing both by industry and at home. These may be calculated for different age or target groups.

The risk of dietary exposure to mycotoxins for consumers has been the objective of numerous recent studies. In general, two approaches are used to estimate human-mycotoxins exposure. The first and the most widely used method involves integrating mycotoxin levels in food samples with food consumption data. These results are further standardized by dividing by average human body weight (60 kg for adults) and expressed as ng/kg body weight/day (ng/kg bw/d). Dietary data are typically obtained through dietary intake surveys using 24 h recalls or food frequency questionnaires. The European Food Safety Authority (EFSA) has developed detailed guidelines for dietary data collection and processing for risk and exposure assessment purposes (EFSA, 2009). An alternative approach, considered a more accurate way of measuring human-mycotoxin exposure, is by measuring mycotoxin biomarkers in human biological fluids.

Risk assessment

Parameters of risk characterization of major mycotoxins are shown in **Table 2**.

Aflatoxins

It has not been possible to establish a toxicological reference point for Afs, making approximations that led to underestimation of the toxicological reference point and, therefore, to overestimation of risk. Therefore, both the Joint FAO/WHO Committee on Experts in Food Additives (JECFA) and EFSA have adopted the margin of exposure (MOE) approach, which is the ratio between the dose (BMD) at which no adverse effects are observed (NOAEL) divided by the actual exposure to that substance through diet. The MOE approach provides an indication of the level of health hazard on the presence of a substance in food, without quantifying the risk. Using the MOE can help risk managers define possible actions necessary to keep exposure to these substances as low as possible.

As there is no real safe level of intake of AFs, it is not appropriate to establish a tolerable daily intake, therefore, the "as low as reasonably achievable" (ALARA) principle is applied for the legal limit. EFSA selects one BMDL10 of 0.4 μg/kg body weight for AFB1. For the AFM1, use a power factor of 0.1 in relation with AFB1. For AFG1, AFB2 and AFG2, the *in vivo* data are not sufficient to derive power factors. For aflatoxins, the same power as for AFB1 has been assumed, leading to an overestimation of risk for AFTs.

Table 2 Major mycotoxins and risk characterization.MOE, margins of exposure. TDI, tolerable daily intake. PMTDI, provisional maximum TDI.

Aflatoxins B1, B2, G1, G2	The calculated MOE is below 10,000, particularly for the younger age groups, so EFSA concludes that a concern for the health of European consumers cannot be ruled out in light of these results.
Ochratoxin A	The estimated MOE for OTA is below 10,000 across most consumer groups, suggesting a possible health concern.
Fumonisins B1, B2, B3	2 μg/kg bw/day TDI based on increased incidence of megalocytic hepatocytes found in a chronic study in mice
Zearalenone and modified forms	0.25 μg/kg bw/day TDI based on estrogenic activity in pigs
Deoxynivalenol DON, 3-Ac-DON, 15-Ac-DON and DON-3-glucoside	1 μg/kg bw/day TDI based on reduced body weight gain in mice
H2 y HT2 and modified forms	0.02 μg/kg bw/day TDI based a reduction of total leukocyte count in rats
Patulin	0.4 μg/kg bw/day PMTDI based on a long-term toxicity/carcinogenicity study in rodent species

In risk characterization, MOEs were below 10,000, particularly for younger age groups, so EFSA concluded that there was no concern for the health of European consumers in view of these results. The MLs in food and feed has been established by Regulation (EC) No. 1881/2006, being 0.1–2 μg/kg.

Contamination of AFs in human food is strictly regulated by the US. Food and Drug Administration (FDA) and AFs in cereals and cereal-based foods should be less than 20 μg/kg. Notably, AFs are the only group of mycotoxins with FDA established advisory and action levels. Some other countries have also set maximum level or limits of AFs in various foodstuffs.

In addition to the establishment of maximum limits in legislation, there are several Codes of Hygienic Practice at the international level in the Codex Alimentarius that help to reduce the presence of aflatoxin-producing fungi in certain foods.

Ochratoxin A

EFSA experts have estimated a margin of MOE for OTA in food, considering possible human safety problems, including genotoxic and carcinogenic. EFSA concluded that current exposure levels posed a potential risk for the health of the majority of studied consumer groups. The calculated MOEs for non-neoplastic effects were above 200 in most of the dietary surveys for average and high consumers and, therefore, of low health concern. However, in high consumers, MOEs were below 200 in infants, toddlers and other children, indicating a possible health concern for these age groups.

Based on current risk calculations from foodstuffs in American marketplaces, there is negligible risk to the US population from OTA exposure. Current OTA concentrations are not high enough to elicit toxic effects, even at high intakes of the foods that may contain OTA. Exposure to OTA in the United States is highest among individuals less than 5 years of age, because of lower body weight and relatively higher consumption of oat-based cereals, however their lifetime cancer risk is considered negligible (Kuiper-Goodman et al., 2010).

Fumonisins

JECFA has identified kidney and liver effects as crucial effects of fumonisin mediated toxicity, and established a provisional maximum tolerable daily intake (PMTDI) of 2 μg/kg bw/day. It must be noted that the mandate from the Commission for the present assessment concerns FB1 and FB2. However, occurrence data available covered the sum of FB1, FB2 and FB3. Considering that most toxicological assessments of PMTDI are derived from data with FB1, and the structural similarities of the different fumonisin derivatives, this is considered acceptable. The calculated exposure to in European countries was lower than the PMTDI, whereas exposure in toddlers and other children was above the PMTDI in several surveys. This indicates a concern for fumonisins exposure in toddlers and other children with high intakes of fumonisins in food, without the additional exposure from modified forms of fumonisins. However, the unexpectedly high contribution from wheat-based products to the total exposure is associated with high uncertainty, due to the low number of quantified results. This uncertainty also affects exposure estimates for their modified forms.

Deoxynivalenol

Deoxynivalenol (DON) is a mycotoxin primarily produced by *Fusarium* fungi, occurring predominantly in cereal grains. The TDI of 1 μg/kg bw/day was established for DON by JECFA and EFSA, based on reduced body weight gain in mice with exposure to the sum of DON, 3-Ac-DON, 15-Ac-DON and DON-3-glucoside. A group-acute reference dose (ARfD) of 8 μg/kg bw/eating occasion was

calculated, based on epidemiological data from mycotoxicosis. Available occurrence data suggest that acute dietary exposures to the sum of DON, 3-Ac-DON, 15-Ac-DON and DON-3-glucoside were below the group ARfD of 8 μg/kg bw/eating occasion for all age groups of humans and, therefore, not considered an acute health concern. However, estimates of chronic dietary exposure to the sum of DON, 3-Ac-DON, 15-Ac-DON and DON-3-glucoside, based on available occurrence data, were above the group TDI of 1 μg/kg bw/day for infants, toddlers and other children and, to some extent, for adolescents and adults. The limited data on dietary habits of vegetarians, with data available for only five European countries, with few subjects in four of them, did not indicate notable differences in acute and chronic dietary exposure between the vegetarians and the general population.

T-2 and HT-2

T-2 toxin (T2) and HT-2 toxin (HT2) belong to the group of compounds known as trichothecenes, which are part of the largest group of Fusarium mycotoxins, comprising more than 150 compounds. In 2017, EFSA published a scientific opinion on the appropriateness of setting a group health-based guidance value (HBGV) for T-2 and HT-2 and its modified forms. In an *in vivo* acute toxicity study, emetic events in mink exposed to T-2 and HT-2 were identified as critical effects for setting an ARfD of 0.3 μg/kg bw. Another *in vivo* subchronic toxicity study with rats identified a reduction in total leukocyte count as the critical effect for setting a TDI of 0.02 μg/kg bw. In food, the highest amounts of T-2 and HT-2 have been reported in grains for human consumption, particularly in oat-containing breakfast cereals. Very high levels have been reported for specific plant extract formula dietary supplements. In humans, the mean chronic dietary exposure to the sum of T-2 and HT-2 was highest in toddlers and infants. Among processed foods, the main contributors were cereal flakes, bakery products including, for acute exposure, breads and rolls. Among elderly adults, dietary supplements made an important contribution.

Zearalenone

The main biological activity of zearalenone (ZEA) is its oestrogenicity, i.e. the ability to act like the endogenous steroidal sex hormone 17-b-estradiol. ZEA was evaluated by JECFA in 2000, establishing a PMTDI of 0.5 μg/kg bw. In that same year, the European Scientific Committee on Human Nutrition (SCF) established a TDI 0.2 μg/kg bw. In 2011, the European Commission requested information from the European Authority to Food Safety (EFSA) on the safety of ZEA and possible risk to consumers in the event of a possible increase in the maximum level for this mycotoxin in breakfast cereals. In the human 2011 ZEA risk assessment from EFSA, (EFSA CONTAM Panel, 2011), a TDI of 0.25 μg/kg bw/day was derived from a non-observed effect level (NOAEL) of 10 μg/kg bw for estrogenic effects in immature gilts. The highest concentrations of ZEA were found in wheat bran, corn and products derived from these grains, such as corn flour or corn flakes. EFSA established a TDI for ZEA of 0.25 μg/kg bw and has considered extending this TDI to the group of ZEA and its modified forms (metabolites in Phase I and II). Dietary exposure estimates suggest that current exposure to ZEA may be close to the TDI.

Patulin

At the international level, JECFA established a NOEL of 43 μg/kg bw/day and a PMTDI of 0.4 μg/kg bw/day for patulin, based on a long-term toxicity/carcinogenicity study in rodent species. In 2000, these values were adopted by the EU Scientific Committee for Human Nutrition (SCF) at the EU level. Apple containing products are the major source for patulin exposure, with major intake differences according to age and living region. Young children are most at risk of patulin intoxication, because they consume more apple products than adults.

Mycotoxin regulations

Mycotoxin presence and dispersal across markets due to trade make it impossible to ban mycotoxins completely from food. Many countries have set levels that ensure consumer protection and food safety and keep the exposure to mycotoxins as low as possible. Tolerances, guidelines, and residue levels have been set in several countries and maximum admissible levels have been established for mycotoxins that occur in several food commodities, especially those traded and consumed extensively. Current regulations encompass 13 mycotoxins, and they are increasingly based on scientific opinions of authoritative bodies as JECFA, EFSA and FDA. Although efforts have been made, mycotoxin regulation is not global and many countries still lack of appropriate guidelines, particularly in Africa and Latin America. Owing to their significant toxicological impacts on both human and animal health, the focus on mycotoxins has been a high priority by the FAO and WHO.

Food legislation must protect both the health of consumers and economic interests of food producers and traders. Therefore, it is preferable to harmonize regulations, especially in countries with trade agreements such as the North American Free Trade Agreement (NAFTA) between the US, Canada, and Mexico, the European Union or MERCOSUR between Argentina, Brazil, Uruguay, Paraguay, and Venezuela, and to adapt uniform international food safety standards regarding mycotoxins. However, current mycotoxin regulations vary significantly, and the absence of a unified and transparent approach has led to a wide range of disparities in guidelines, with several differences in maximum admissible levels set in different countries. Further, in many developing countries, especially those facing food availability issues, regulations may not be present or, if present, not enforced (Barajas-Ramirez et al., 2021).

Regulations serve as a safeguard for food markets from contaminated imported commodities. In many cases, commodities are rejected due to food safety threats. For example, according to the European Commission Rapid Alert System for Food and Feed (RASFF) annual report, mycotoxins were the primary hazard in border rejection from non-EU countries in 2019, with 440 notifications. Strict regulations may create an economic imbalance and affect exporting countries, which may face difficulties finding new markets or maintaining their usual ones. This may also lead to an abundance of foods contaminated with mycotoxins in local markets in developing countries, particularly those with hot and humid climates.

Mycotoxin mitigation

Prevention is the main tool to reduce the presence of mycotoxins in food and feed. Therefore, the development of toxigenic fungi in agricultural crops should be avoided through the use of Good Agricultural Practices. In 2003, the Codex Alimentarius established recommendations for the prevention and/or reduction of mycotoxins, including crop rotation, use of varieties resistant to toxigenic fungi and insects, control of water supply, use of suitable fungicides and insecticides, use of collection and transport equipment that avoid damage and free of fungal contamination, and control of humidity and temperature in the silo. However, when mycotoxin values still exceed the maximum limits established for human or animal consumption, reducing procedures based on the use of physical, chemical, and biological methods have been proposed. The application of these procedures should not alter the nutritional and organoleptic qualities of the food, while not leaving residues of any kind that could represent a hazard (Agriopoulou et al., 2020).

Physical procedures include ionizing radiation such as γ and X rays, which have been widely used to reduce aflatoxins in seeds, due to their penetration capacity. However, a large part of society rejects the use of ionizing radiation, although it is authorized in various countries. Non-ionizing radiation such as UV rays are effective only on the surface. Thermal inactivation generates partial results against the various chemical families of mycotoxins due to its varied heat resistance.

Chemical mitigation procedures are prohibited in the European Union due to the possibility of generating artifacts of unknown toxicity, although the use of oxidants such as chlorine, ozone, hydrogen peroxide and ammonization has been proposed. More recently, the use of natural substances present in food such as isothiocyanates from *Brassicaceae*, have been shown to form higher molecular weight complexes that present less toxicity *in vitro* by joining the amino and aldehyde groups of fumonisins and zearalenones. This may be an alternative in the future, provided that research releases more data about safety.

Biological methods are based on the addition to food of microorganisms capable of stopping fungal growth or transforming mycotoxins. Lactic acid bacteria (LAB) naturally present in food and in the intestinal flora are an important bet for the future. LABs affect the development of fungi by competing for space and nutrients, and also produce metabolites with antifungal activity, such as aliphatic and phenolic organic acids (lactic, phenyl-lactic, acetic, propionic, benzoic, 4-hydroxybenzoic, ferulic, vanillic, hydroxycinnamic) as well as ethanol, hydrogen peroxide, and diacetyl, which act by altering the plasma membrane of fungi to modify the electrochemical potential and increase its permeability. LABs also produce antimicrobial peptide compounds (AMP), whose mechanism of action is linked to their amphipathic character, which results in the ability to destabilize the phospholipids of the fungal membrane. The fermentation of the lacto-albumin and caseins of milk by *Lactobacillus spp*. produce AMP, which can contain from 3 to 20 amino acids with antimicrobial activity; their degradation by the proteases of the digestive tract would limit their action on the intestinal flora when they are used as food additives. The production of mycotoxins generally begins at the end of the fungal growth phase, and the direct influence of LAB and other fungal growth inhibitors lowers the presence of mycotoxins. In addition, LABs can reduce mycotoxins by adsorption, as polysaccharides and peptidoglycans surrounding the cytoplasmic membrane. Finally, LABs can also reduce the presence of mycotoxins through hydrolysis, specifically OTA is transformed into a non-toxic metabolite, α-OTA, and an amino acid, phenylalanine.

For the detoxification of feed, additives capable of adsorbing mycotoxins and preventing their intestinal absorption are being widely used. Good results are seen with aluminosilicates, such as bentonites and zeolites, diatomaceous earth and activated carbon. However, they have the disadvantage of also adsorbing vitamins and trace elements and, therefore, cannot be used in high quantities. Naturally, not all adsorbents are capable of binding all mycotoxins, the chemical structure of both determine the possibility of interacting.

Climate change

Fungal growth in crops depends on the plant species and various environmental factors, including temperature, humidity and the presence of insects. All of these are closely related to climate and, therefore, dependent on climate change. A temperature increase of 2–3 °C may change the presence of different species and genera of fungi. *Aspegillus flavus*, a predominant fungus in tropical and subtropical areas, is spreading to temperate areas as temperatures progressively increase, including in southern Europe and the Midwestern US. This is of concern, as it is generating production of aflatoxins in cereals and nuts, especially in corn. When temperatures of up to 37–40 °C, are reached, the synthesis of aflatoxins is inhibited. In contrast, a decrease in aflatoxins in maize is predicted in the Philippines, due to increased rainfall. In northern Europe, the DON-producing *Fusarium culmorum* was predominant in wheat and is being replaced by *F. graminearum*, a DON producer more typical of southern Europe, together with other *Fusarium* species that produce T-2 and HT-2 toxins, both more toxic than DON. Insects and other vectors influence the spread of fungal infections by

altering the outer layer of grain pods and ears. High temperatures may favor insect reproduction and their activity. In some temperate zones where winters are milder and summers are hotter, winter survival and the number of annual generations will increase. The change in climatic conditions at a global level will produce a migration of plant pests and diseases in the direction of the cooler latitudes. This may lead to the substitution of some species of fungi for others and the presence of different mycotoxins. Some researchers point to the possibility that a change from toxigenic species to others that do not produce mycotoxins may occur, especially in hot countries. Although the new fungus/host plant/geographical area combinations are of great interest regarding changes in mycotoxin production, there is currently insufficient knowledge for diagnosis. Omics and other new tools will provide keys for prediction and the fight in the dissipation of mycotoxins (Daou et al., 2021).

Conclusions

Mycotoxins are the most abundant and abiotic toxins present in food, a significant risk to the health and well-being of humans and animals, and a food safety issue with potential for important economic losses. Although efforts to prevent mycotoxin formation have been undertaken, contamination of those secondary fungal metabolites still occur The presence of mycotoxins in food is inevitable, and the complexity of these contaminants will possibly increase due to the influence of climate change. Furthermore, our ability to detect existing complex contaminants will develop with the increase in knowledge regarding the diversity of molecules and modified forms, their co-presence, and toxicological interactions. The entire population is exposed to mycotoxins but, given the high proportion of cereals in the diet with mycotoxins, children and high consumers are of greatest concern.

The application of Good Practices of Hygiene throughout the food production chain and MLs in food management are the most effective measures in reducing exposure to mycotoxins.

It is essential for toxicologists together with other health professionals to continue working on their knowledge, risk assessment and prevention with the collaboration of regulatory bodies.

References

Agriopoulou, S., Stamatelopoulou, E., Varzakas, T., 2020. Advances in occurrence, importance, and mycotoxin control strategies: prevention and detoxification in foods. Foods 9, 137. https://doi.org/10.3390/foods9020137.

Alshannaq, A., Yu, J.-H., 2017. Occurrence, toxicity, and analysis of major mycotoxins in food. Int. J. Environ. Res. Publ. Health 14 (6), 632. https://doi.org/10.3390/ijerph14060632.

Barajas-Ramirez, J.A., Moncada-Abaunza, D.A., Gomez-Espinoza, M.G., 2021. Mycotoxins in foods, from the field to the plate: a review. Int. Food Res. J. 28 (2), 230.

Cimbalo, A., Alonso-Garrido, M., Font, G., Manyes, L., 2020. Toxicity of mycotoxins in vivo on vertebrate organisms: a review. Food Chem. Toxicol. 137, 111161. https://doi.org/10.1016/j.fct.2020.111161.

Daou, R., Joubrane, K., Maroun, R.G., El Khoury, A., 2021. Mycotoxins: factors influencing production and control strategies. AIMS Agric. Food 6 (1), 416–447. https://doi.org/10.3934/agrfood.2021025.

EFSA, 2009. Guidance of the Scientific Committee on a request from EFSA on the use of the benchmark dose approach in risk assessment. EFSA J 1150, 1–72. https://doi.org/10.2903/j.efsa.2009.1150.

EFSA CONTAM Panel (EFSA Panel on Contaminants in the Food Chain), 2006. Scientific Opinion on the risks for public health related to the presence of zearalenone in food. EFSA J 9 (6), 2197. https://doi.org/10.2903/j.efsa.2011.2197.

European Commission, 2006. Commission Regulation No. 1881/2006 of December 19th setting maximum levels of certain contaminants in foodstuffs. Off. J. Eur. Union L 364, 5–24, 20.12.2006.

Janik, E., Niemcewicz, M., Ceremuga, M., Stela, M., Saluk-Bijak, J., Adrian Siadkowski, A., Bijak, M., 2020. Molecular aspects of mycotoxins—a serious problem for human health. Int. J. Mol. Sci. 21, 8187. https://doi.org/10.3390/ijms21218187.

Kuiper-Goodman, T., Hilts, C., Billiard, S.M., Kiparissis, Y., Richard, I.D.K., Hayward, S., 2010. Health risk assessment of ochratoxin A for all age-sex strata in a market economy. Food Addit. Contam. 27 (2), 212–240. https://doi.org/10.1080/02652030903013278.

Marin, S., Ramos, A.J., Cano-Sancho, G., Sanchis, V., 2013. Mycotoxins: occurrence, toxicology, and exposure assessment. Food Chem. Toxicol. 60, 218–237. https://doi.org/10.1016/j.fct.2013.07.047.

IARC, 2012. Risk assessment and risk management of mycotoxins. IARC Sci. Publ. 158, 105–117. PMID: 23477199.

Further reading

EFSA, 2020. Mycotoxins [Internet] Updated: 05/13/2020. https://www.efsa.europa.eu/en/topics/topic/mycotoxins.

World Health Organization & Joint FAO/WHO Expert Committee on Food Additives (83rd, 2017: Geneva, Switzerland), 2017. Evaluation of Certain Contaminants in Food: Eighty-Third Report of the Joint FAO/WHO Expert Committee on Food Additives WHO Technical Report Series; 1002 World Health Organization. https://apps.who.int/iris/handle/10665/254893.

FDA, 2017. Natural Toxins and Mycotoxins [Internet] Updated: 10/27/2017. https://www.fda.gov/food/chemicals-metals-pesticides-food/natural-toxins-and-mycotoxins.

Nutraceuticals: Health effects and clinical applications

Oliver Chen[a,b] **and Eunice Mah**[a], [a] Biofortis Research, Addison, IL, United States; and [b] Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of L. Galland, Functional Foods: Health Effects and Clinical Applications, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 366–371, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00130-6.

Introduction 501
Nomenclature of nutraceuticals 501
Health benefits of nutraceuticals 502
Brain health 502
Cardiometabolic health 503
Glucose regulation 503
Lipids 504
Body weight 504
Blood pressure 505
Immunity and infection 505
Urinary health 505
Eye health 506
Sleep 507
Bone health 507
Muscle health 507
Summary 508
References 508

Key points

- Nutraceuticals are products falling between nutritional and pharmaceutical product categories.
- There is no clear and comprehensive definition for nutraceuticals approved by international authorities, which is needed to regulate their safety or therapeutic efficacy.
- Nutraceuticals are generally comprised of ingredients derived from foods and herbs.
- Health benefits of nutraceuticals must be substantiated by clinical evidence, with preclinical data used to support mechanisms of action.

Introduction

Nutrition science has evolved from the provision of basic nutrition to support of essential functions and structures of the body for health promotion and disease prevention. Nutraceuticals are products derived from both food and non-food sources that are purported to provide health benefit in support of structure or function of the body, prevention and mitigation of diseases, delay of the aging process, and promotion of wellbeing. Many nutraceuticals are appreciated for their health benefits beyond basic nutrition in different systems of the human body, such as cardiovascular, immune, musculoskeletal, and nervous systems, via an array of mechanisms of action, including hypolipidemic, immune-modulating, antioxidant, anti-inflammatory, antitoxic or ergogenic effects. Popular nutraceuticals include cranberry, *Ginkgo biloba*, ginseng, green tea, glucosamine, lutein, micronutrients, omega-3 fatty acids, probiotics, and St. John's Wort. Although the pharmacological activity of most of these substances is well established in vitro or in small mammals, establishing clinical evidence poses a challenge for nutraceutical research. In this article, the health benefits of nutraceuticals with clinical evidence supported by meta-analysis are discussed.

Nomenclature of nutraceuticals

The term "nutraceutical" is a portmanteau of 2 words, "nutrient" and "pharmaceutical," first conceptualized in 1989 by Stephen L. DeFelice. More recently, a redefinition of this concept was made as "the phytocomplex if they derive from a food of vegetal origin, and as the pool of the secondary metabolites if they derive from a food of animal origin, concentrated and administered in the more suitable pharmaceutical form, which are capable of providing beneficial health effects, including the prevention and/or the

treatment of a disease" (Domínguez Díaz et al., 2020). Thus, nutraceuticals fall within the border of nutritional and pharmaceutical product categories. There is confusion related to the terminology of nutraceuticals, functional foods and dietary supplements. Nutraceutical is a term often used interchangeably with the term functional foods, but they are not truly interchangeable as the former refers to nearly any purified or isolated bioactive component delivered in a supplement form, while the latter includes products containing bioactive components in food form. Additionally, unlike functional foods, nutraceuticals may have a recommended daily dose.

According to FDA regulations, most nutraceuticals are categorized as dietary supplements under the authority of the Federal Food, Drug, and Cosmetic Act, even though they are not specifically defined by law in the US, and are widely used in the marketplace (Reader, 2019). The 1994 Dietary Supplement Health and Education Act (DSHEA) in the US defined dietary supplements as "a product taken by mouth that contains a dietary ingredient intended to supplement the diet." The dietary ingredients may include vitamins, minerals, herbs, botanicals, amino acids, and substances such as enzymes, organ tissues, glandulars, or metabolites,/ Importantly, dietary supplements in the US are not allowed to claim to treat, mitigate or cure diseases or conditions, but only to be marketed to support the structure or function of the body. The Food Directorate of Health Canada (Canada, 2002) defines nutraceuticals as products isolated or purified from foods that are sold in medicinal forms and demonstrated to have a physiological benefit or provide protection against chronic disease. There is no specific regulation for nutraceuticals in the US, Europe, or Japan providing a statutory definition. Thus, a clear and comprehensive definition approved by international authorities is needed to regulate their safety and health or therapeutic efficacy.

Health benefits of nutraceuticals

A diet composing diverse healthful foods and adequate essential nutrients is crucial for health promotion and prevention. Nutraceuticals can be part of a healthy diet supporting human health by reducing risk factors of chronic disease and promoting wellbeing and quality of life, particularly in those who have moderate or high risk for disease. On some occasions, under a physician's supervision, they can be added to therapeutic regimens when prescribed medicines do not suffice for treatment goals. For example, calcium nutraceutical products can be added to hormone replacement therapy for the management of osteoporosis, and red yeast rice to statin therapy for hypercholesterolemia. A benefit of nutraceutical and pharma combinations, for example, may be a reduction of side effects that may occur at high doses of prescribed medications. The efficacy of nutraceuticals in recommended dosage must be proven in clinical studies before they can be marketed to consumers.

Brain health

According to the US Centers for Disease Control and Prevention (CDC), a healthy brain is "one that can perform all the mental processes that are collectively known as cognition, including the ability to learn new things, intuition, judgment, language, and remembering" (CDC). Poor brain health may eventually manifest as cognitive impairment. Various nutraceuticals have been investigated in relation to cognition, with a few demonstrating promising beneficial evidence. Examples of common nutraceuticals that have been investigated for their ability to help maintain or enhance cognitive function in healthy individuals include *Bacopa monniera* (Brahmi), *Ginkgo biloba*, ginseng, vitamins, and omega-3 fatty acids. Several clinical trials on *B. monniera* in healthy adults have shown improvements in learning and intelligence (Downey et al., 2013; Morgan and Stevens, 2010), attention (Calabrese et al., 2008; Peth-Nui et al., 2012; Stough et al., 2008), memory (Barbhaiya et al., 2008; Calabrese et al., 2008; Kumar et al., 2016; Morgan and Stevens, 2010; Peth-Nui et al., 2012; Roodenrys et al., 2002), and processing speed (Calabrese et al., 2008; Stough et al., 2001). A meta-analysis of nine clinical trials involving healthy subjects and those with memory complaints suggested that *B. monniera* has the potential to improve attention speed, but not memory (Kongkeaw et al., 2014). Beneficial effects of *G. biloba* in healthy adults have been observed in several clinical trials on memory (Kaschel, 2011; Kennedy et al., 2002; Rigney et al., 1999; Silberstein et al., 2011; Steiner et al., 2016; Wesnes et al., 2000), attention (Kennedy et al., 2000; Kennedy et al., 2002), and processing speed (Cieza et al., 2003). However, a meta-analysis of 13 clinical trials showed non-significant effect sizes of *G. biloba* for memory, executive function, and attention (Laws et al., 2012). Improvements in memory and attention have been shown in healthy adults with ginseng (Kennedy et al., 2001; Reay et al., 2005; Reay et al., 2006; Scholey et al., 2010). A meta-analysis of 22 clinical trials reported no significant effect of B vitamins or antioxidant vitamins on cognitive function, and insufficient evidence to evaluate the effect of omega-3 fatty acids on cognition (Jia et al., 2008).

In addition to maintaining a healthy brain, nutraceuticals may be beneficial for brain-related diseases or disorders, but clinical data are limited. A meta-analysis of clinical trials involving Alzheimer's disease (AD) patients showed inconclusive evidence on whether ginseng was an effective treatment (Wang et al., 2016). Melatonin supplementation in AD patients has been shown to improve sleep and suppresses sundowning (a state of confusion occurring in the late afternoon and spanning into the night) in several clinical trials (Cardinali et al., 2010). Additionally, a meta-analysis of seven clinical trials reported improvement in mini-mental state examination (MMSE) score in mild AD patients receiving >12 weeks of melatonin treatment (Wang et al., 2017a,b).

The most commonly investigated nutraceuticals for depression are St. John's Wort, omega-3 fatty acids, and S-adenosyl-methionine. A meta-analysis of 27 clinical trials suggested that St. John's Wort had comparable efficacy and safety to selective serotonin reuptake inhibitors (the most commonly prescribed antidepressant) in patients with mild-to-moderate depression (Ng et al., 2017). A meta-analysis of 26 clinical trials showed that intake of omega-3 PUFAs with EPA $\geq$ 60% at a dose $\leq$ 1 g/d had beneficial effects

on depression (Liao et al., 2019). Finally, in a meta-analysis of eight clinical trials, the authors concluded that a firm conclusion on the use of S-adenosyl-methionine for treating depression in adults could not be made, due to the lack of high-quality evidence (Galizia et al., 2016).

Putative mechanisms by which nutraceuticals modulate brain function are numerous. A well-studied mechanism is amelioration of oxidative stress and inflammation in the brain. Additionally, neurodegenerative diseases such as AD and epilepsy, as well as brain injury, typically involve the accumulation of misfolded proteins, e.g., amyloid-β peptide, resulting in a cascade of events leading to increased inflammation and oxidative stress (Makkar et al., 2020). An important consideration for exogenous antioxidants is their ability to cross the blood-brain barrier (BBB). Bacosides, believed to be the active constituents of B. Monnieri, can be metabolized into aglycones (jujubogenin and pseudojujubogenin) and these metabolites are likely responsible for the diminished oxidative stress in the brain, as they have good BBB penetration (Ramasamy et al., 2015). In contrast to bacosides, melatonin readily passes through the BBB and has been shown to be protective against Aβ-induced lipid peroxidation, decreased inflammation in the brain, and prevention of amyloid-β peptide generation (Rosales-Corral et al., 2003; Shukla et al., 2017). Reduction in brain inflammation has been observed following omega-3 fatty acid treatment in animal models of AD and aging (Larrieu and Laye, 2018). Additionally, omega-3 fatty acids, especially EPA, are important in limiting the degradation of BBB integrity, which is implicated in neurodegenerative diseases and aging (Barnes et al., 2021).

Certain nutraceuticals, particularly those associated with psychiatric disorders, are believed to modulate neurotransmitter production, release, and synaptic concentration. For example, St. John's Wort alleviates symptoms of depression by inhibiting monoamine (e.g., serotonin and/or noradrenaline) reuptake into nerve terminals, which is a common mechanism of action of a variety of antidepressants (Butterweck, 2003). B. Monnieri has been shown to decrease plasma corticosterone concentration in stressed rats and to maintain brain levels of noradrenaline, dopamine, and serotonin, which were otherwise reduced with stress (Sheikh et al., 2007).

Together, evidence exists in support of potential benefits of nutraceuticals in maintaining cognitive function in healthy adults, as well as of their potential therapeutic role for neurodegenerative diseases and mental disorders, through various putative mechanisms. However, most clinical trials present inconsistent results for nutraceuticals. Indeed, no health claims have been approved by the FDA for brain-related outcomes. However, there is a qualified health claim for phosphatidylserine and reduced risk for cognitive dysfunction, where the FDA concluded that "the science provides minimal and preliminary evidence sufficient for qualified health claims about phosphatidylserine and reduced risk of these conditions." More rigorous and robust studies on the efficacy of nutraceuticals in regulating brain health are warranted to better understand the potential of nutraceuticals in maintaining and improving brain function and reducing the risk of neurodegenerative diseases.

Cardiometabolic health

Cardiometabolic disorders represent a cluster of interrelated cardiovascular risk factors, primarily insulin resistance, impaired glucose tolerance, dyslipidemia, hypertension, and central adiposity. Various nutraceuticals have been examined for their ability to regulate each of these risk factors, and those that are promising are described below.

Glucose regulation

Examples of common nutraceuticals that have been investigated for their beneficial effects on glucose homeostasis include the minerals zinc and chromium, fiber, and non-nutrient nutraceuticals, garlic, ginseng, and berberine. Two meta-analyses of clinical trials involving participants with type 2 diabetes (T2D), prediabetes, obesity, polycystic ovary syndrome (PCOS), metabolic syndrome or gestational diabetes showed that zinc improved fasting glucose, Insulin, 2 h postprandial glucose, homeostatic model assessment for insulin resistance (HOMA-IR), and HbA1c (Wang et al., 2019; Pompano and Boy, 2021). Dosages of both <25 mg/d and ≥ 25 mg/d have been effective in improving HOMA-IR, while lower dosages (<25 mg/d) were more beneficial for fasting blood glucose and higher dosages (≥ 25 mg/d) were better for HbA1c, a long-term marker of glucose regulation (Pompano and Boy, 2021). Chromium supplementation of >200 μg/d decreased HbA1c among individuals with baseline HbA1c $\geq$ 8% (indicative of uncontrolled diabetes) in a meta-analysis of 25 clinical trials (Suksomboon et al., 2014). Additionally, chromium was effective in reducing fasting blood glucose (Suksomboon et al., 2014).

Adding additional fiber (4–40 g/d) to habitual diet has been shown to reduce fasting blood glucose and HbA1c in T2D patients (Post et al., 2012). In a meta-analysis of nine clinical trials with T2D patients, significant reduction in fasting blood glucose, fructosamine, and HbA1c were observed following garlic supplementation (0.05–1.5 g/d) (Wang et al., 2017a,b). Ginseng supplementation has also been reported to improve fasting glucose, postprandial insulin, and HOMA-IR, but not postprandial glucose, fasting insulin, or HbA1c in a meta-analysis of 8 clinical trials with individuals with T2D or impaired glucose tolerance (Gui et al., 2016). In another meta-analysis of four trials with T2D patients, berberine (an isoquinoline derivative alkaloid isolated from Rhizoma Coptidis) with a co-intervention of lifestyle modification was better than lifestyle modification alone or placebo in improving fasting blood glucose, postprandial glucose, and HbA1c (Dong et al., 2012). Two meta-analyses of clinical trials report beneficial effects of turmeric and curcumin on HbA1c and HOMA-IR in adults with cardiometabolic disorders (e.g., T2DM, metabolic syndrome) (Tabrizi et al., 2018; Altobelli et al., 2021).

This clinical evidence shows that nutraceuticals appear to have a role in the management of blood glucose, particularly in people with impaired glucose regulation.

Lipids

Many nutraceuticals have been investigated for their lipid-lowering effects. The International Lipid Expert Panel (ILEP) reviewed several lipid-lowering nutraceuticals with the objective of providing recommendations for their optimal use to manage dyslipidemia in patients not on statin therapy, those on statin or combination therapy who have not achieved lipid goals, and patients with statin intolerance (Cicero et al., 2017). According to their review, red yeast rice extract, berberine, and omega-3 fatty acids showed evidence that treatment was beneficial, useful, and effective (Cicero et al., 2017).

Red yeast rice, obtained by the fermentation of yeast (*Monascus purpureus*) in rice, contains substances with lipid-lowering activities, including polyketides such as monacolin K, which is structurally identical to lovastatin. Two meta-analyses reported reduction in low density lipoprotein-cholesterol (LDL-C) and triglycerides (TG) as well as increases in HDL-C with red yeast rice extract supplementation in diverse populations, including myocardial infarction patients with borderline hypercholesterolemia, and patients with dyslipidemia, coronary heart disease, T2DM, hypertension, and non-alcoholic fatty liver disease (Sungthong et al., 2020). Red yeast rice-based therapy is safe and well tolerated independent of dose and duration of therapy and health condition of users (Fogacci et al., 2019). Additionally, EFSA concluded that a relationship exists between the administration of red yeast rice and the maintenance of plasma LDL-C, specifically using a dose of red yeast rice containing 10 mg of monacolin K (Efsa Panel on Dietetic Products and Nutrition and Allergies (NDA), 2011). Due to the similarity in structure between monacolin K and lovastatin, the primary cholesterol-lowering putative mechanism of action of red yeast rice is thought to be a reversible inhibitory action on 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) reductase (the key enzyme in endogenous cholesterol synthesis) (Cicero et al., 2017).

Berberine is a quaternary benzylisoquinoline alkaloid present in Coptis (*Coptis chinensis*, *Coptis japonica*), Hydrastis (*Hydrastis canadensis*), and Berberis (*Berberis aristata*, *Berberis vulgaris*, *Berberis croatica*). Two meta-analyses on berberine report decreases in total cholesterol (TC), TG, and LDL-C and an increase in HDL-C (Ju et al., 2018; Dong et al., 2013). Another meta-analysis comparing the effects of berberine alone or combined with statins reports that berberine was effective only in reducing TG when compared with simvastatin, and that the combination of berberine and simvastatin was more effective in decreasing TG and TC compared to simvastatin alone (Zhang et al., 2019). According to the International Lipid Expert Panel (ILEP), the use of berberine at doses ranging between 500 and 1500 mg was effective in lipid-lowering and relatively safe both in primary and secondary prevention. The effect of berberine on the reduction of TG and TC is partly related to its positive effect on insulin resistance and clearance of LDL-C, respectively (Cicero et al., 2017).

Several meta-analyses report reduction in TG following omega-3 fatty acid supplementation (Natto et al., 2019; Gao et al., 2020; Berge et al., 2014). Additionally, the lipid-lowering effects of omega-3 have been recognized by several authoritative organizations. The EFSA established a claim in 2010 indicating that the intake of at least 2 g/d of docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA) can maintain normal blood TG concentration (Efsa Panel on Dietetic Products and Nutrition and Allergies (NDA), 2010). The American Heart Association has also indicated that doses from 2 to 4 g/d of EPA/DHA reduce TG by 25%–30% (Miller et al., 2011). The hypolipidemic effects of omega-3 fatty acids are mediated through lowering synthesis of hepatic very-low-density lipoprotein and TG, and increasing beta-oxidation of fatty acids and the synthesis of phospholipids (Harris and Bulchandani, 2006).

Body weight

Supplements claiming to decrease body weight feature several putative mechanisms, such as delaying or limiting nutrient absorption, inducing satiety and decreasing appetite, or increasing energy expenditure (Watanabe et al., 2020). Nutraceuticals most actively investigated for their weight loss properties include green tea, caffeine, conjugated linoleic acid, and garcinia (Payab et al., 2020; Maunder et al., 2020; Batsis et al., 2021). Although the efficacy of these nutraceuticals in reducing body weight, fat mass, or waist circumference has been reported by some clinical trials, the reductions are modest, and their clinical significance is unclear.

A Cochrane meta-analysis of six clinical trials of at least 12 weeks' duration reported non-significant changes in body weight or waist-to-hip ratio in overweight or obese adults following green tea supplementation (Jurgens et al., 2012). A more recent meta-analysis on green tea supplementation in varied populations reported reduction in body weight and BMI (Lin et al., 2020). The decrease in body weight was most evident for dosages $<$500 mg/d for 12 weeks. Reduction in waist circumference was found for supplementation of green tea $\geq$800 mg/d for $<$12 weeks. Another meta-analysis on green tea catechins concluded that green tea catechins with caffeine were associated with significant reduction in body weight and waist circumference, but catechins alone do not show benefits (Phung et al., 2010). Thus, it is likely that caffeine in green tea extract or its additive, or synergistic interaction with catechins contribute to weight loss. This speculation is supported by one meta-analysis reporting reduction in body weight, BMI, and fat mass following caffeine consumption (Tabrizi et al., 2019). However, the inclusion of studies with other bioactive compounds, such as ephedrine, weakened the evidence of caffeine (Tabrizi et al., 2019). A separate meta-analysis of epidemiologic studies reported that higher coffee intake was significantly associated with modestly lower BMI and waist circumference in men only (Lee et al., 2019). Additionally, higher coffee intake was associated with overall reduced central adiposity, which was not significant when analyzed by sex subgroups (Lee et al., 2019). Similarly, another meta-analysis reported that green coffee bean extract significantly reduced body weight and waist circumference (Asbaghi et al., 2020). All these data support the effectiveness of polyphenols and caffeine in body weight management.

In a meta-analysis of long-term (at least 6 months) clinical trials, conjugated linoleic acid was reported to reduce body weight and fat mass, although authors cautioned that the decreases were small and, thus, their clinical significance is unclear (Onakpoya et al., 2012). A more recent meta-analysis also concluded that conjugated linoleic acid reduced body weight and fat mass and

increased lean body mass. Additionally, subgroup analyses indicated that conjugated linoleic acid was more effective in individuals >44 y and when provided for more than 12 weeks at dosages >3.4 g/d (Namazi et al., 2019 #80). One meta-analysis reported that *Garcinia cambogia* significantly reduced body weight, percent fat mass, and waist circumference (Golzarand et al., 2020). Dose-response analysis revealed a nonlinear association between *Garcinia cambogia* dosage and change in body weight, whereby peak reductions corresponded to a dose of ~2000 mg/d (Golzarand et al., 2020).

Blood pressure

Among foods, beetroot juice has the most convincing evidence of antihypertensive effect; among micronutrients, magnesium, potassium, and vitamin C can improve blood pressure, and among non-nutrient nutraceuticals, soy isoflavones, cocoa catechins, and melatonin have been proposed (Borghi et al., 2020; Cicero et al., 2019). However, the non-nutrient nutraceuticals reviewed were showed only limited evidence of effect (Cicero et al., 2019).

Nitric oxide is a potent vasodilator and vascular endothelium relaxant, and a reduction in its availability and/or bioactivity is linked to the development of hypertension, endothelial dysfunction, atherosclerosis, and CVD (Li and Forstermann, 2000). Beetroot juice is a concentrated source of inorganic nitrates, contributing substantially to nitric oxide availability via the nitrate-nitrite-nitric oxide pathway (Lidder and Webb, 2013). Two meta-analyses reported decreases in blood pressure (both systolic and diastolic) of normotensive/pre-hypertensive/mild hypertensive adults following beetroot juice supplementation (Bahadoran et al., 2017; Jackson et al., 2018). One of these also reported improvements in endothelial function assessed using flow-mediated dilation and reductions in arterial stiffness and platelet aggregation (Jackson et al., 2018). However, such hypotensive benefits were not observed in patients with chronic obstructive pulmonary disease (Alshafie et al., 2021). Cocoa catechins, known to regulate availability of nitric oxide, can significantly improve endothelial function and decrease systolic and diastolic blood pressure, with an optimal effect observed with 710 mg total flavanols, 95 mg (−)-epicatechin or 25 mg (+)-catechin (Sun et al., 2019; Jafarnejad et al., 2020).

Numerous mechanisms have been proposed to explain Mg^{2+}-induced BP reduction, including a calcium-channel blocking action and an increase in prostaglandin E and nitric oxide synthesis. Seven meta-analyses on oral magnesium and blood pressure/hypertension were found. Of those, one reported no effect (Jee et al., 2002), one reported lowering of diastolic BP (DBP) but not systolic BP (SBP) (Dickinson et al., 2006), four reported lowering both SBP and DBP (Zhang et al., 2016; Rosanoff and Plesset, 2013; Kass et al., 2012; Dibaba et al., 2017), and another reported population dependent effects (Rosanoff et al., 2021). Specifically, magnesium at >600 mg/d is effective to reduce BP in untreated adults with hypertension (≥140/90 mmHg) and at 240–607 mg/d in uncontrolled, treated adults (Rosanoff et al., 2021). However, oral magnesium did not affect BP in normotensive subjects or those with controlled hypertension (Asbaghi et al., 2021; Rosanoff et al., 2021).

Immunity and infection

Many micronutrients are well-known for their roles in the immune system (Maggini et al., 2018). Copper, iron, selenium, folate, zinc, vitamin A, vitamin B12, vitamin B6, vitamin C, and vitamin D all have approved EFSA health claims to support the normal function of the immune system based on their well-established biochemical roles and/or on deficiency symptoms involving the immune system (Efsa Panel on Dietetic Products and Nutrition and Allergies (NDA), 2016). Additionally, available clinical evidence suggests that some of these nutrients positively affect the incidence and severity of infections and immune-related medical conditions. Examples include vitamin D as prophylaxis for respiratory infection (Martineau et al., 2019; Martineau et al., 2017), as an adjuvant treatment for pneumonia (Yang et al., 2021), reducing the severity of atopic dermatitis in adults and children (Kim and Bae, 2016; Hattangdi-Haridas et al., 2019), and decreasing asthma exacerbations (Jolliffe et al., 2017) and vitamin C as a supplementary therapy along with antiviral regimens to relieve patients from the symptoms of the common cold (Ran et al., 2020). Probiotics have recently gained increasing attention for their potential immune-modulating properties, as evidenced by the results of several meta-analyses providing support for their beneficial effects on the respiratory infection (Yeh et al., 2018; Lei et al., 2017; Hao et al., 2015), gastroenteritis (Ansari et al., 2020), *Helicobacter pylori* infection (Shi et al., 2019; Lu et al., 2016), *Clostridium difficile* infection (Shen et al., 2017), and atopic dermatitis (Zuccotti et al., 2015 ; Li et al., 2019; Jiang et al., 2020; Amalia et al., 2020).

Urinary health

Urinary tract infections (UTI) are prevalent worldwide, particularly in women. Clinically, UTIs are categorized as uncomplicated or complicated, with the latter defined as those associated with factors that compromise the urinary tract or host defense. The recurrence rate in the 12 months following an initial uncomplicated UTI in women ranges from 25 to 44%. The typical treatment for UTI is antibiotics but, given the rapid rise of multi-drug resistant uropathogens, many complementary and integrative approaches to managing UTI have been proposed, including nutraceuticals, such as probiotics, D-mannose, and cranberry products.

Cranberry juice is perhaps best known for the prevention and treatment of UTI. This health benefit received a US FDA qualified health claim in 2020 (FDA, 2020), "consuming one serving (8 oz.) each day of an at least 27% cranberry juice beverage or 500 mg each day of cranberry dietary supplement may help reduce the risk of recurrent UTI in healthy women." However, cranberry-containing products, including cranberry-containing nutraceuticals, were effective in women with recurrent UTI, but not in populations with potentially complicated UTI such as those with neuropathic bladder, elderly patients, or pregnant patients (Wang et al., 2012; Fu et al., 2017). Additionally, a recent meta-analysis (Raguzzini et al., 2020) showed that cranberries were ineffective against UTI in patients with spinal cord injury, compared with controls. Although the mechanism by which constituents in cranberries

modulate UTI incidence remains to be elucidated, the anti-adhesion activities of cranberry constituents against uropathogenic *E. coli* are likely responsible for the benefit.

D-mannose is an inert monosaccharide metabolized and excreted in urine which can inhibit bacterial adhesion to the urothelium, representing a promising nonantibiotic prevention strategy for UTI. D-mannose has been used by some women in an attempt to prevent the recurrence of UTI. A meta-analysis study showed that D-mannose with dosing between 420 mg/d to 2 g three times/d could reduce the risk of UTI recurrence, compared to placebo, but the reduction was not statistically significant when compared to preventative antibiotics (Lenger et al., 2020). Additionally, D-mannose appears to be well tolerated with minimal side effects.

Probiotics are "live microorganisms that, when administered in adequate amounts, confer a health benefit on hosts." Probiotics, such as Lactobacillus and Bifidobacterium, are beneficial microorganisms that may block pathogenic bacteria adhesion to urothelial cells and defend against infections in the urogenital tracts. The concept originated from observation of an inverse association between *E. coli* colonization in women with recurrent UTI and Lactobacillus, the most common bacteria genus in the human vaginal microbiota. Nevertheless, mixed data have been reported on the efficacy of probiotics in preventing recurrent UTI in adults (Vagios et al., 2020). Similarly, a meta-analysis of 10 clinical trials showed a null effect of probiotic therapy on the incidence of UTI and its recurrence in children (Hosseini et al., 2017). A significant moderate benefit was only observed when probiotics were used as adjuvant therapy to antibiotics. Thus, using probiotics in the protection and prevention of UTI remains to be further researched.

Taken together, clinical evidence exists regarding the potential beneficial role of nutraceuticals in the prevention of recurrent UTI. Apparently, cranberry products are more promising than D-mannose and probiotics.

Eye health

Advancing age is associated with increased risk for ocular diseases, such as cataract, age-related macular degeneration, and dry eye syndrome, with the first two being the leading cause of vision impairment and blindness, and the final being the most common complaint in ophthalmologic practice (Mccusker et al., 2016). While treatments are available for age-related vision loss, such as surgical removal of cataracts, many causes of vision loss, such as dry age-related macular degeneration (AMD), remain poorly understood, and no treatments are currently available. The importance of nutrition in eye health has been well appreciated. In particular, results of the Age-Related Eye Disease Study (AREDS) studies demonstrating that high intake of vitamin A, vitamin, C, zinc, copper, and carotenoids could reduce the progression of AMD by approximately 25% (Francisco et al., 2020). Thus, nutraceuticals may help support eye health and protect against age-related eye diseases.

AMD involves damage to the macula, an area of the retina necessary for high-acuity vision. There are two forms, dry and wet AMD. The former is the most common and results from radical-induced damage to retinal pigmented epithelial cells and accumulation of extracellular protein- and lipid-containing deposits called drusen. Thus, antioxidants may protect against cellular damage in the retina by reducing sunlight-derived free radicals. However, results of a meta-analysis of five clinical trials showed that vitamin E supplements did not affect the development of AMD compared with placebo (Evans and Lawrenson, 2017a,b). Similarly, a null effect was noted with vitamin C or beta-carotene supplementation. Nevertheless, the AREDS formula containing a battery of antioxidants, vitamin A and C, zinc, copper, and carotenoids appears to be protective against AMD progression to the late-stage (Evans and Lawrenson, 2017a,b). Additionally, a small significant protective effect on AMD progression was found in people taking zinc alone. Nutraceuticals containing lutein and zeaxanthin that are heavily marketed for AMD have shown little or no impact on the progression of AMD (Evans and Lawrenson, 2017a,b). This result conflicts with the finding that lutein supplementation (10 or 20 mg per day) was associated with an increase in macular pigment optical density (MPOD), a marker of lutein accumulation in the macular, where high MPOD suggests larger protection against blue-light induced drusen formation (Feng et al., 2019). There was no evidence that omega-3 fatty acid supplements were protective against progression to advanced AMD (Lawrenson and Evans, 2015) even though there was an inverse relationship between dietary intake of omega-3 fatty acids and risk of developing AMD or progression to advanced AMD.

A cataract is the result of opacification in eye lenses that impair vision and cause blindness. While it is most commonly corrected with surgery to restore vision, dietary antioxidants may delay and/or prevent cataracts because oxidative damage plays a major role in its etiology. A meta-analysis of 12 observational studies showed that vitamin A, C, and E, β-carotene, and lutein were associated with reduced risk of age-related cataracts (Jiang et al., 2019). However, a meta-analysis result of 8 clinical trials showed that vitamin E or β-carotene interventions were not protective compared with placebo (Jiang et al., 2019). Thus, antioxidants may be protective against cataracts but need to be substantiated in human trials.

Dry eye disease is a chronic disease marked with irritation and burning that can cause inflammatory damage to the cornea and conjunctiva and is prevalent in the elderly population. The main cause of the disease is dysfunction of the Meibomian gland due to hyperkeratinization of the ductal epithelium of the gland and reduced production of meibum, as well as tear hyperosmolarity and tear film instability. Current therapeutic strategies for managing dry eye disease include enhancing tear volume and quality by using artificial eye drops and reducing ocular inflammation. However, two-thirds of dry eye sufferers remain symptomatic despite adherence to treatment (Downie and Keller, 2015). With their anti-inflammatory effect, omega-3 fatty acids may help mitigate the symptoms of dry eye disease. A meta-analysis study with 17 clinical trials demonstrated that compared with placebo, omega-3 fatty acid supplementation decreased dry eye symptoms and increased the tear breakup time (Giannaccare et al., 2019). Another meta-analysis of 13 clinical trials showed that without eye medication, polyunsaturated fatty acids, including α-linolenic, linoleic, eicosapentaenoic, or docosahexaenoic acid, improved tear breakup time, osmolarity and ocular surface disease index score, compared with the placebo (Chi et al., 2019). Thus, omega-3 fatty acid nutraceuticals may play a role in managing dry eye disease.

Sleep

Sleep is an essential and complex biological process that needs to be satisfied to maintain many physiological functions, e.g., muscle recovery and repair, neurological development, cardiac, immune, and metabolic functions, cognition, and mood (Watson et al., 2015). Additionally, inadequate sleep (i.e., duration and/or quality) is a threat to health because of the association of insufficient sleep with obesity, metabolic, immunological, and cardiovascular health, some types of cancer, pain, and mental illness. The prevalence of sleeping problems (during the past 12 months) has been increasing in developed and developing countries worldwide. There are different medications for sleep disorders, such as antihistamines, zolpidem (Ambien), zaleplon (Sonata), benzodiazepines, doxepin, eszopiclone, and emborexant. However, their use is accompanied by adverse side effects, ranging from dizziness during awake time to gastrointestinal discomfort, neurologic side effects, and long-term dependence. Thus, nutraceuticals can be an alternative or complement with pharmacological therapies without common serious side effects.

A recent study showed that a polyphenol botanical blend supplement improved self-reported sleep quality and daytime functioning and reduced insomnia severity in individuals with subclinical sleep disturbances, but did not compromise neurocognitive functioning during daytime (Tubbs et al., 2021). The gut-brain axis, extensive two-way interaction between the gastrointestinal tract and central nervous system, informs a new direction for developing strategies for nervous system-related health conditions, such as modifying the gut microbiota for sleep. Wong et al. (2015) reported that 6 weeks of probiotic (8 strains in *Bifidobacterium*, *Lactobacillus*, and *Streptococcus* genera) supplementation in men with irritable bowel syndrome (IBS) increased salivary melatonin, a hormone involved in the regulation of circadian rhythmicity. A meta-analysis study of 14 clinical trials showed that probiotic (live microbes)/paraprobiotics (inactivated or non-viable microbes) consumption induced positive changes to perceived sleep health in people suffering from inadequate sleep but did not affect objective parameters of sleep, i.e., efficiency and latency (Irwin et al., 2020).

Bone health

Bone comprises an organic phase (matrix) and an inorganic mineral phase, together with different cells responsible for a cycle of bone resorption and formation. The matrix consists of structural proteins, collagen, and mucopolysaccharide; hydroxyapatite containing calcium is the main mineral in the inorganic phase. Joints are comprised of synovium, cartilage, ligaments, bone, and muscles. Aging is accompanied by reduced bone thickness and density, and stiffer and less flexible joint movement. Osteoporosis is a progressive, systemic bone disease with loss of bone and its microscopic structure, resulting from dysregulation of proinflammation. Osteoarthritis, the most common form of arthritis among older people, is caused by the degradation and loss of articular cartilage due to the age-related increase in free radicals and inflammation. Thus, nutraceuticals that support bone structure and ameliorate oxidative stress and inflammation can benefit bone and joint health (Pandey et al., 2018).

Calcium supplements are commonly taken to maintain a calcium intake of at least 1000–1200 mg/d to prevent or treat osteoporosis. A meta-analysis study shows that increasing dietary calcium intake increases bone mineral density at the total hip, lumbar spine, femoral neck, and total body (Tai et al., 2015). Calcium supplements also increase bone mineral density. However, the magnitude of the increase from both types of interventions is small (<2%), which may not lead to a clinically significant reduction in risk of fracture. Vitamin D has been included in common regimens for osteoporosis management, due to its effect on calcium absorption in the small intestine and bone formation. However, the result of a meta-analysis study showed there was a small benefit of vitamin D supplementation at the femoral neck but no effect at any other body site (Reid et al., 2014). While vitamin K is known for its involvement in blood clotting, its contributions to bone formation are emerging. A meta-analysis of clinical trials showed that vitamin K decreased risk of clinical bone fracture, compared to controls, but bone mineral density was not altered (Mott et al., 2019). Thus, using nutraceuticals containing vitamin D or K to protect against osteoporosis seems unjustified at this time. However, there may be potential benefit of combining vitamin D and K on bone mineral density (Kuang et al., 2020). Soy isoflavones with estrogenic activity may modulate maintenance of the bone structure, supported by a meta-analysis showing that daily ingestion of an average 56 mg soy isoflavones moderately decreased bone resorption, but not bone formation (Taku et al., 2010). As short-chain fatty acids produced from microbes in the human gut can regulate bone loss (Zaiss et al., 2019), nutraceuticals that modulate the gut microbiota are anticipated to affect bone health. A meta-analysis of 5 clinical trials showed a significant positive association between probiotic supplementation and bone mineral density in the lumbar spine, but not in hips (Yu et al., 2021).

Osteoarthritis involves the deterioration of articular cartilage and underlying bone. Chondroitin sulfate is a major component of the extracellular matrix of many connective tissues, and glucosamine is a component of articular cartilage. Foods containing these 2 ingredients may reduce joint deterioration and related pain and disability. However, meta-analysis data show that chondroitin (1200 mg/day) or glucosamine had some effectiveness on pain reduction but no effect on cartilage volume (Knapik et al., 2019; Knapik et al., 2018). Inflammation is an underlying cause of pain and deterioration of joints. Nutraceuticals formulated with anti-inflammatory ingredients may be beneficial to arthritis. For example, omega-3 fatty acids can decrease joint pain intensity in people with rheumatoid arthritis (Goldberg and Katz, 2007). Thus, nutraceuticals containing chondroitin sulfate, glucosamine and/or omega-3 fatty acids may help pain management in people with arthritis, but may not be adequate to stop deterioration.

Muscle health

Loss of muscle mass, power, and strength, a pathogenic condition called sarcopenia, can impair physical function and increase the risk of frailty in the elderly population. Nutritional interventions maintaining muscle quality and quantity can support quality of

life and extend health span. Amino acids, omega-3 fatty acids, creatine, and β-hydroxy-β-methylbutyrate, are thought to affect muscle mass and strength (Martin-Cantero et al., 2021; Gielen et al., 2021). While protein is the primary source for muscle synthesis, a meta-analysis showed that protein supplementation alone did not benefit upper-limb, lower-limb, or handgrip strength, compared to resistance exercise training in healthy elderly adults (Labata-Lezaun et al., 2020). Nevertheless, individual amino acids can exert specific effects in muscle synthesis and function. For example, leucine is a building block and serves as a nutritional signal for protein synthesis. A meta-analysis of 16 clinical trials showed that, compared with controls, leucine supplemented in the range of 2–7.8 g/d significantly increased gain in lean body mass and body weight in older people; the effect size was even larger in those with manifested sarcopenia (Komar et al., 2015). β-hydroxy-β-methylbutyrate has been shown to promote turnover of muscle protein, potentially playing a role in ameliorating skeletal muscle wasting and weakness. A meta-analysis of 15 clinical trials showed favorable effects of 3 g/d β-hydroxy-β-methylbutyrate on increasing skeletal muscle mass and strength (Bear et al., 2019). Similar to β-hydroxy-β-methylbutyrate, omega-3 fatty acids, particularly at >2 g/d, can promote muscle mass gain (Huang et al., 2020). Creatine can modulate muscle mass via several mechanisms, including ATP production, glycogen synthesis, calcium reuptake, and protein turnover. Its supplementation (≥5 g/d) augments increases in lean muscle mass and strength from resistance training in older adults, compared to placebo (Forbes et al., 2021). Thus, nutraceuticals can be added to an exercise regimen to protect against loss of muscle mass and strength, a condition that can lead to frailty and morbidity of older populations.

Summary

The term "nutraceutical" is a portmanteau of 2 words, "nutrient" and "pharmaceutical," and they can be derived from vegetal and animal origins and are capable of providing beneficial health effects, including the prevention and/or the treatment of disease. However, there is no specific regulation for nutraceuticals in the US, Europe, and Japan, providing statutory definition. Additionally, nutraceuticals are categorized as dietary supplements in the US, which are not allowed to claim to treat, mitigate or cure disease or condition. Numerous nutraceuticals have proven clinical efficacy in a wide range of health indications, such as *Bacopa monniera*, *Ginkgo biloba*, ginseng, melatonin, St. John's Wort, and omega-3 fatty acids for brain health; garlic, berberine, turmeric, red yeast rice, omega-3 fatty acids, green tea, conjugated linoleic acid, garcinia, beetroot, cocoa catechins for the cardiometabolic system; probiotics for immunity and infection; cranberry, D-mannose, probiotics for urinary health; probiotics for sleep; calcium, vitamin D and K, chondroitin sulfate, glucosamine, and omega-3 fatty acids for bone and joints; leucine, β-hydroxy-β-methylbutyrate, and creatine for muscle. Thus, nutraceuticals may be used to improve health, delay the aging process, prevent chronic disease, increase health span, or to support the structure or function of the body. However, their effectiveness and safety must be substantiated by human studies, with supporting evidence on mechanisms of action from preclinical studies.

References

Alshafie, S., EL-Helw, G.O., Fayoud, A.M., Elrashedy, A.A., Gbreel, M.I., et al., 2021. Efficacy of dietary nitrate-rich beetroot juice supplementation in patients with chronic obstructive pulmonary disease (COPD): a systematic review and meta-analysis. Clin. Nutr. ESPEN 42, 32–40.

Altobelli, E., Angeletti, P.M., Marziliano, C., Mastrodomenico, M., Giuliani, A.R., et al., 2021. Potential therapeutic effects of curcumin on glycemic and lipid profile in uncomplicated type 2 diabetes-A meta-analysis of randomized controlled trial. Nutrients 13.

Amalia, N., Orchard, D., Francis, K.L., King, E., 2020. Systematic review and meta-analysis on the use of probiotic supplementation in pregnant mother, breastfeeding mother and infant for the prevention of atopic dermatitis in children. Australas. J. Dermatol. 61, e158–e173.

Ansari, F., Pashazadeh, F., Nourollahi, E., Hajebrahimi, S., Munn, Z., et al., 2020. A systematic review and meta-analysis: the effectiveness of probiotics for viral gastroenteritis. Curr. Pharm. Biotechnol. 21, 1042–1051.

Asbaghi, O., Sadeghian, M., Rahmani, S., Mardani, M., Khodadost, M., et al., 2020. The effect of green coffee extract supplementation on anthropometric measures in adults: a comprehensive systematic review and dose-response meta-analysis of randomized clinical trials. Complement. Ther. Med. 51, 102424.

Asbaghi, O., Hosseini, R., Boozari, B., Ghaedi, E., Kashkooli, S., et al., 2021. The effects of magnesium supplementation on blood pressure and obesity measure among type 2 diabetes patient: a systematic review and meta-analysis of randomized controlled trials. Biol. Trace Elem. Res. 199, 413–424.

Bahadoran, Z., Mirmiran, P., Kabir, A., Azizi, F., Ghasemi, A., 2017. The nitrate-independent blood pressure-lowering effect of beetroot juice: a systematic review and meta-analysis. Adv. Nutr. 8, 830–838.

Barbhaiya, H.C., Desai, R.P., Saxena, V.S., Pravina, K., Wasim, P., et al., 2008. Efficacy and tolerability of BacoMind on memory improvement in elderly participants—a double blind placebo controlled study. J. Pharmacol. Toxicol. 3, 425–434.

Barnes, S., Chowdhury, S., Gatto, N.M., Fraser, G.E., Lee, G.J., 2021. Omega-3 fatty acids are associated with blood-brain barrier integrity in a healthy aging population. Brain Behav. 11, e2273.

Batsis, J.A., Apolzan, J.W., Bagley, P.J., Blunt, H.B., Divan, V., et al., 2021. A systematic review of dietary supplements and alternative therapies for weight loss. Obesity 29, 1102–1113.

Bear, D.E., Langan, A., Dimidi, E., Wandrag, L., Harridge, S.D.R., et al., 2019. β-Hydroxy-β-methylbutyrate and its impact on skeletal muscle mass and physical function in clinical practice: a systematic review and meta-analysis. Am. J. Clin. Nutr. 109, 1119–1132.

Berge, K., Musa-Veloso, K., Harwood, M., Hoem, N., Burri, L., 2014. Krill oil supplementation lowers serum triglycerides without increasing low-density lipoprotein cholesterol in adults with borderline high or high triglyceride levels. Nutr. Res. 34, 126–133.

Borghi, C., Tsioufis, K., Agabiti-Rosei, E., Burnier, M., Cicero, A.F.G., et al., 2020. Nutraceuticals and blood pressure control: a European Society of Hypertension position document. J. Hypertens. 38, 799–812.

Butterweck, V., 2003. Mechanism of action of St John's wort in depression : what is known? CNS Drugs 17, 539–562.

Calabrese, C., Gregory, W.L., Leo, M., Kraemer, D., Bone, K., et al., 2008. Effects of a standardized *Bacopa monnieri* extract on cognitive performance, anxiety, and depression in the elderly: a randomized, double-blind, placebo-controlled trial. J. Altern. Complement. Med. 14, 707–713.

Canada, H., 2002. Nutraceuticals/Functional Foods and Health Claims on Foods. Available. https://www.canada.ca/en/health-canada/services/food-nutrition/food-labelling/health-claims/nutraceuticals-functional-foods-health-claims-foods-policy-paper.html. (Accessed 3 November 2021).
Cardinali, D.P., Furio, A.M., Brusco, L.I., 2010. Clinical aspects of melatonin intervention in Alzheimer's disease progression. Curr. Neuropharmacol. 8, 218–227.
Centers for Disease Control and Prevention. What is a Healthy Brain? New Research Explores Perceptions of Cognitive Health Among Diverse Older Adults. Available https://www.cdc.gov/aging/pdf/perceptions_of_cog_hlth_factsheet.pdf, 1 October 2021.
Chi, S.C., Tuan, H.I., Kang, Y.N., 2019. Effects of polyunsaturated fatty acids on nonspecific typical dry eye disease: a systematic review and meta-analysis of randomized clinical trials. Nutrients 11.
Cicero, A.F.G., Colletti, A., Bajraktari, G., Descamps, O., Djuric, D.M., et al., 2017. Lipid lowering nutraceuticals in clinical practice: position paper from an International Lipid Expert Panel. Arch. Med. Sci. 13, 965–1005.
Cicero, A.F.G., Grassi, D., Tocci, G., Galletti, F., Borghi, C., et al., 2019. Nutrients and nutraceuticals for the management of high normal blood pressure: an evidence-based consensus document. High Blood Press. Cardiovasc. Prev. 26, 9–25.
Cieza, A., Maier, P., Poppel, E., 2003. Effects of *Ginkgo biloba* on mental functioning in healthy volunteers. Arch. Med. Res. 34, 373–381.
Dibaba, D.T., Xun, P., Song, Y., Rosanoff, A., Shechter, M., et al., 2017. The effect of magnesium supplementation on blood pressure in individuals with insulin resistance, prediabetes, or noncommunicable chronic diseases: a meta-analysis of randomized controlled trials. Am. J. Clin. Nutr. 106, 921–929.
Dickinson, H.O., Nicolson, D.J., Campbell, F., Cook, J.V., Beyer, F.R., et al., 2006. Magnesium supplementation for the management of essential hypertension in adults. Cochrane Database Syst. Rev. CD004640.
Domínguez Díaz, L., Fernández-Ruiz, V., Cámara, M., 2020. An international regulatory review of food health-related claims in functional food products labeling. J. Funct. Foods 68, 103896.
Dong, H., Wang, N., Zhao, L., Lu, F., 2012. Berberine in the treatment of type 2 diabetes mellitus: a systemic review and meta-analysis. Evid. Based Complement. Alternat. Med. 2012, 591654.
Dong, H., Zhao, Y., Zhao, L., Lu, F., 2013. The effects of berberine on blood lipids: a systemic review and meta-analysis of randomized controlled trials. Planta Med. 79, 437–446.
Downey, L.A., Kean, J., Nemeh, F., Lau, A., Poll, A., et al., 2013. An acute, double-blind, placebo-controlled crossover study of 320 mg and 640 mg doses of a special extract of *Bacopa monnieri* (CDRI 08) on sustained cognitive performance. Phytother. Res. 27, 1407–1413.
Downie, L.E., Keller, P.R., 2015. A pragmatic approach to the management of dry eye disease: evidence into practice. Optom. Vis. Sci. 92, 957–966.
Efsa Panel on Dietetic Products, Nutrition and Allergies (NDA), 2010. Scientific opinion on the substantiation of health claims related to eicosapentaenoic acid (EPA), docosahexaenoic acid (DHA), docosapentaenoic acid (DPA) and maintenance of normal cardiac function (ID 504, 506, 516, 527, 538, 703, 1128, 1317, 1324, 1325), maintenance of normal blood glucose concentrations (ID 566), maintenance of normal blood pressure (ID 506, 516, 703, 1317, 1324), maintenance of normal blood HDL-cholesterol concentrations (ID 506), maintenance of normal (fasting) blood concentrations of triglycerides (ID 506, 527, 538, 1317, 1324, 1325), maintenance of normal blood LDL-cholesterol concentrations (ID 527, 538, 1317, 1325, 4689), protection of the skin from photo-oxidative (UVinduced) damage (ID 530), improved absorption of EPA and DHA (ID 522, 523), contribution to the normal function of the immune system by decreasing the levels of eicosanoids, arachidonic acid-derived mediators and pro-inflammatory cytokines (ID 520, 2914), and "immunomodulating agent" (4690) pursuant to Article 13(1) of Regulation (EC) No 1924/2006. EFSA J. 8, 1796–1838.
Efsa Panel on Dietetic Products, Nutrition and Allergies (NDA), 2011. Scientific opinion on the substantiation of health claims related to monacolin K from red yeast rice and maintenance of normal blood LDL-cholesterol concentrations (ID 1648, 1700) pursuant to Article 13(1) of Regulation (EC) No 1924/2006. EFSA J. 9, 2304–2319.
Efsa Panel on Dietetic Products, Nutrition and Allergies (NDA), 2016. Guidance on the scientific requirements for health claims related to the immune system, the gastrointestinal tract and defence against pathogenic microorganisms. EFSA J. 14, 4369–4382.
Evans, J.R., Lawrenson, J.G., 2017a. Antioxidant vitamin and mineral supplements for preventing age-related macular degeneration. Cochrane Database Syst. Rev. 7, Cd000253.
Evans, J.R., Lawrenson, J.G., 2017b. Antioxidant vitamin and mineral supplements for slowing the progression of age-related macular degeneration. Cochrane Database Syst. Rev. 7, Cd000254.
FDA, U., 2020. Qualified Health Claim for Certain Cranberry Products and Urinary Tract Infections.
Feng, L., Nie, K., Jiang, H., Fan, W., 2019. Effects of lutein supplementation in age-related macular degeneration. PLoS One 14, e0227048.
Fogacci, F., Banach, M., Mikhailidis, D.P., Bruckert, E., Toth, P.P., Lipid, Blood Pressure Meta-Analysis Collaboration Group, International Lipid Expert Panel, et al., 2019. Safety of red yeast rice supplementation: a systematic review and meta-analysis of randomized controlled trials. Pharmacol. Res. 143, 1–16.
Forbes, S.C., Candow, D.G., Ostojic, S.M., Roberts, M.D., Chilibeck, P.D., 2021. Meta-analysis examining the importance of creatine ingestion strategies on lean tissue mass and strength in older adults. Nutrients 13.
Francisco, S.G., Smith, K.M., Aragonès, G., Whitcomb, E.A., Weinberg, J., et al., 2020. Dietary patterns, carbohydrates, and age-related eye diseases. Nutrients 12.
Fu, Z., Liska, D., Talan, D., Chung, M., 2017. Cranberry reduces the risk of urinary tract infection recurrence in otherwise healthy women: a systematic review and meta-analysis. J. Nutr. 147, 2282–2288.
Galizia, I., Oldani, L., Macritchie, K., Amari, E., Dougall, D., et al., 2016. S-adenosyl methionine (SAMe) for depression in adults. Cochrane Database Syst. Rev. 10, CD011286.
Gao, C., Liu, Y., Gan, Y., Bao, W., Peng, X., et al., 2020. Effects of fish oil supplementation on glucose control and lipid levels among patients with type 2 diabetes mellitus: a meta-analysis of randomized controlled trials. Lipids Health Dis. 19, 87.
Giannaccare, G., Pellegrini, M., Sebastiani, S., Bernabei, F., Roda, M., et al., 2019. Efficacy of omega-3 fatty acid supplementation for treatment of dry eye disease: a meta-analysis of randomized clinical trials. Cornea 38, 565–573.
Gielen, E., Beckwée, D., Delaere, A., De Breucker, S., Vandewoude, M., et al., 2021. Nutritional interventions to improve muscle mass, muscle strength, and physical performance in older people: an umbrella review of systematic reviews and meta-analyses. Nutr. Rev. 79, 121–147.
Goldberg, R.J., Katz, J., 2007. A meta-analysis of the analgesic effects of omega-3 polyunsaturated fatty acid supplementation for inflammatory joint pain. Pain 129, 210–223.
Golzarand, M., Omidian, M., Toolabi, K., 2020. Effect of Garcinia cambogia supplement on obesity indices: a systematic review and dose-response meta-analysis. Complement. Ther. Med. 52, 102451.
Gui, Q.F., Xu, Z.R., Xu, K.Y., Yang, Y.M., 2016. The efficacy of ginseng-related therapies in type 2 diabetes mellitus: an updated systematic review and meta-analysis. Medicine 95, e2584.
Hao, Q., Dong, B.R., Wu, T., 2015. Probiotics for preventing acute upper respiratory tract infections. Cochrane Database Syst. Rev. CD006895.
Harris, W.S., Bulchandani, D., 2006. Why do omega-3 fatty acids lower serum triglycerides? Curr. Opin. Lipidol. 17, 387–393.
Hattangdi-Haridas, S.R., Lanham-New, S.A., Wong, W.H.S., Ho, M.H.K., Darling, A.L., 2019. Vitamin D deficiency and effects of vitamin D supplementation on disease severity in patients with atopic dermatitis: a systematic review and meta-analysis in adults and children. Nutrients 11.
Hosseini, M., Yousefifard, M., Ataei, N., Oraii, A., Mirzay Razaz, J., et al., 2017. The efficacy of probiotics in prevention of urinary tract infection in children: a systematic review and meta-analysis. J. Pediatr. Urol. 13, 581–591.
Huang, Y.H., Chiu, W.C., Hsu, Y.P., Lo, Y.L., Wang, Y.H., 2020. Effects of omega-3 fatty acids on muscle mass, muscle strength and muscle performance among the elderly: a meta-analysis. Nutrients 12.
Irwin, C., Mccartney, D., Desbrow, B., Khalesi, S., 2020. Effects of probiotics and paraprobiotics on subjective and objective sleep metrics: a systematic review and meta-analysis. Eur. J. Clin. Nutr. 74, 1536–1549.
Jackson, J.K., Patterson, A.J., Macdonald-Wicks, L.K., Oldmeadow, C., Mcevoy, M.A., 2018. The role of inorganic nitrate and nitrite in cardiovascular disease risk factors: a systematic review and meta-analysis of human evidence. Nutr. Rev. 76, 348–371.
Jafarnejad, S., Salek, M., Clark, C.C.T., 2020. Cocoa consumption and blood pressure in middle-aged and elderly subjects: a meta-analysis. Curr. Hypertens. Rep. 22, 1.

Jee, S.H., Miller III, E.R., Guallar, E., Singh, V.K., Appel, L.J., et al., 2002. The effect of magnesium supplementation on blood pressure: a meta-analysis of randomized clinical trials. Am. J. Hypertens. 15, 691–696.
Jia, X., Mcneill, G., Avenell, A., 2008. Does taking vitamin, mineral and fatty acid supplements prevent cognitive decline? A systematic review of randomized controlled trials. J. Hum. Nutr. Diet. 21, 317–336.
Jiang, H., Yin, Y., Wu, C.R., Liu, Y., Guo, F., et al., 2019. Dietary vitamin and carotenoid intake and risk of age-related cataract. Am. J. Clin. Nutr. 109, 43–54.
Jiang, W., Ni, B., Liu, Z., Liu, X., Xie, W., et al., 2020. The role of probiotics in the prevention and treatment of atopic dermatitis in children: an updated systematic review and meta-analysis of randomized controlled trials. Paediatr. Drugs 22, 535–549.
Jolliffe, D.A., Greenberg, L., Hooper, R.L., Griffiths, C.J., Camargo Jr., C.A., et al., 2017. Vitamin D supplementation to prevent asthma exacerbations: a systematic review and meta-analysis of individual participant data. Lancet Respir. Med. 5, 881–890.
Ju, J., Li, J., Lin, Q., Xu, H., 2018. Efficacy and safety of berberine for dyslipidaemias: a systematic review and meta-analysis of randomized clinical trials. Phytomedicine 50, 25–34.
Jurgens, T.M., Whelan, A.M., Killian, L., Doucette, S., Kirk, S., et al., 2012. Green tea for weight loss and weight maintenance in overweight or obese adults. Cochrane Database Syst. Rev. 12, CD008650.
Kaschel, R., 2011. Specific memory effects of *Ginkgo biloba* extract EGb 761 in middle-aged healthy volunteers. Phytomedicine 18, 1202–1207.
Kass, L., Weekes, J., Carpenter, L., 2012. Effect of magnesium supplementation on blood pressure: a meta-analysis. Eur. J. Clin. Nutr. 66, 411–418.
Kennedy, D.O., Scholey, A.B., Wesnes, K.A., 2000. The dose-dependent cognitive effects of acute administration of *Ginkgo bilobato* healthy young volunteers. Psychopharmacology 4, 416–423.
Kennedy, D.O., Scholey, A.B., Wesnes, K.A., 2001. Dose dependent changes in cognitive performance and mood following acute administration of Ginseng to healthy young volunteers. Nutr. Neurosci. 4, 295–310.
Kennedy, D.O., Scholey, A.B., Wesnes, K.A., 2002. Modulation of cognition and mood following administration of single doses of *Ginkgo biloba*, ginseng, and a ginkgo/ginseng combination to healthy young adults. Physiol. Behav. 5, 739–751.
Kim, G., Bae, J.H., 2016. Vitamin D and atopic dermatitis: a systematic review and meta-analysis. Nutrition 32, 913–920.
Knapik, J.J., Pope, R., Hoedebecke, S.S., Schram, B., Orr, R., et al., 2018. Effects of oral glucosamine sulfate on osteoarthritis-related pain and joint-space changes: systematic review and meta-analysis. J. Spec. Oper. Med. 18, 139–147.
Knapik, J.J., Pope, R., Hoedebecke, S.S., Schram, B., Orr, R., 2019. Effects of oral chondroitin sulfate on osteoarthritis-related pain and joint structural changes: systematic review and meta-analysis. J. Spec. Oper. Med. 19, 113–124.
Komar, B., Schwingshackl, L., Hoffmann, G., 2015. Effects of leucine-rich protein supplements on anthropometric parameter and muscle strength in the elderly: a systematic review and meta-analysis. J. Nutr. Health Aging 19, 437–446.
Kongkeaw, C., Dilokthornsakul, P., Thanarangsarit, P., Limpeanchob, N., Norman Scholfield, C., 2014. Meta-analysis of randomized controlled trials on cognitive effects of Bacopa monnieri extract. J. Ethnopharmacol. 151, 528–535.
Kuang, X., Liu, C., Guo, X., Li, K., Deng, Q., et al., 2020. The combination effect of vitamin K and vitamin D on human bone quality: a meta-analysis of randomized controlled trials. Food Funct. 11, 3280–3297.
Kumar, N., Abichandani, L.G., Thawani, V., Gharpure, K.J., Naidu, M.U., et al., 2016. Efficacy of standardized extract of *Bacopa monnieri* (Bacognize(R)) on cognitive functions of medical students: a six-week, randomized placebo-controlled trial. Evid. Based Complement. Alternat. Med. 2016, 4103423.
Labata-Lezaun, N., Ilurda-Almuzara, L., lópez-de-celis, C., Rodríguez-Sanz, J., González-Rueda, V., et al., 2020. Effectiveness of protein supplementation combined with resistance training on muscle strength and physical performance in elderly: a systematic review and meta-analysis. Nutrients 12.
Larrieu, T., Laye, S., 2018. Food for mood: relevance of nutritional omega-3 fatty acids for depression and anxiety. Front. Physiol. 9, 1047.
Lawrenson, J.G., Evans, J.R., 2015. Omega 3 fatty acids for preventing or slowing the progression of age-related macular degeneration. Cochrane Database Syst. Rev. 2015. Cd010015.
Laws, K.R., Sweetnam, H., Kondel, T.K., 2012. Is *Ginkgo biloba* a cognitive enhancer in healthy individuals? A meta-analysis. Hum. Psychopharmacol. 27, 527–533.
Lee, A., Lim, W., Kim, S., Khil, H., Cheon, E., et al., 2019. Coffee intake and obesity: a meta-analysis. Nutrients 11.
Lei, W.T., Shih, P.C., Liu, S.J., Lin, C.Y., Yeh, T.L., 2017. Effect of probiotics and prebiotics on immune response to influenza vaccination in adults: a systematic review and meta-analysis of randomized controlled trials. Nutrients 9 (11), 1175. https://doi.org/10.3390/nu9111175.
Lenger, S.M., Bradley, M.S., Thomas, D.A., Bertolet, M.H., Lowder, J.L., et al., 2020. D-mannose vs. other agents for recurrent urinary tract infection prevention in adult women: a systematic review and meta-analysis. Am. J. Obstet. Gynecol. 223, 265.e1–265.e13.
Li, H., Forstermann, U., 2000. Nitric oxide in the pathogenesis of vascular disease. J. Pathol. 190, 244–254.
Li, L., Han, Z., Niu, X., Zhang, G., Jia, Y., et al., 2019. Probiotic supplementation for prevention of atopic dermatitis in infants and children: a systematic review and meta-analysis. Am. J. Clin. Dermatol. 20, 367–377.
Liao, Y., Xie, B., Zhang, H., He, Q., Guo, L., et al., 2019. Efficacy of omega-3 PUFAs in depression: a meta-analysis. Transl. Psychiatr. 9, 190.
Lidder, S., Webb, A.J., 2013. Vascular effects of dietary nitrate (as found in green leafy vegetables and beetroot) via the nitrate-nitrite-nitric oxide pathway. Br. J. Clin. Pharmacol. 75, 677–696.
Lin, Y., Shi, D., Su, B., Wei, J., Gaman, M.A., et al., 2020. The effect of green tea supplementation on obesity: a systematic review and dose-response meta-analysis of randomized controlled trials. Phytother. Res. 34, 2459–2470.
Lu, M., Yu, S., Deng, J., Yan, Q., Yang, C., et al., 2016. Efficacy of probiotic supplementation therapy for *Helicobacter pylori* eradication: a meta-analysis of randomized controlled trials. PLoS One 11, e0163743.
Maggini, S., Pierre, A., Calder, P.C., 2018. Immune function and micronutrient requirements change over the life course. Nutrients 10.
Makkar, R., Behl, T., Bungau, S., Zengin, G., Mehta, V., et al., 2020. Nutraceuticals in neurological disorders. Int. J. Mol. Sci. 21.
Martin-Cantero, A., Reijnierse, E.M., Gill, B.M.T., Maier, A.B., 2021. Factors influencing the efficacy of nutritional interventions on muscle mass in older adults: a systematic review and meta-analysis. Nutr. Rev. 79, 315–330.
Martineau, A.R., Jolliffe, D.A., Hooper, R.L., Greenberg, L., Aloia, J.F., et al., 2017. Vitamin D supplementation to prevent acute respiratory tract infections: systematic review and meta-analysis of individual participant data. BMJ 356, i6583.
Martineau, A.R., Jolliffe, D.A., Greenberg, L., Aloia, J.F., Bergman, P., et al., 2019. Vitamin D supplementation to prevent acute respiratory infections: individual participant data meta-analysis. Health Technol. Assess. 23, 1–44.
Maunder, A., Bessell, E., Lauche, R., Adams, J., Sainsbury, A., et al., 2020. Effectiveness of herbal medicines for weight loss: a systematic review and meta-analysis of randomized controlled trials. Diabetes Obes. Metab. 22, 891–903.
Mccusker, M.M., Durrani, K., Payette, M.J., Suchecki, J., 2016. An eye on nutrition: the role of vitamins, essential fatty acids, and antioxidants in age-related macular degeneration, dry eye syndrome, and cataract. Clin. Dermatol. 34, 276–285.
Miller, M., Stone, N.J., Ballantyne, C., Bittner, V., Criqui, M.H., American Heart Association Clinical Lipidology, Thrombosis, and Prevention Committee of the Council on Nutrition, Physical Activity, and Metabolism, Council on Arteriosclerosis, Thrombosis and Vascular Biology, Council on Cardiovascular Nursing, Council on the Kidney in Cardiovascular Disease, et al., 2011. Triglycerides and cardiovascular disease: a scientific statement from the American Heart Association. Circulation 123, 2292–2333.
Morgan, A., Stevens, J., 2010. Does *Bacopa monnieri* improve memory performance in older persons? Results of a randomized, placebo-controlled, double-blind trial. J. Altern. Complement. Med. 16, 753–759.

Mott, A., Bradley, T., Wright, K., Cockayne, E.S., Shearer, M.J., et al., 2019. Effect of vitamin K on bone mineral density and fractures in adults: an updated systematic review and meta-analysis of randomized controlled trials. Osteoporos. Int. 30, 1543–1559.
Namazi, N., Irandoost, P., Larijani, B., Azadbakht, L., 2019. The effects of supplementation with conjugated linoleic acid on anthropometric indices and body composition in overweight and obese subjects: a systematic review and meta-analysis. Crit. Rev. Food Sci. Nutr. 59, 2720–2733.
Natto, Z.S., Yaghmoor, W., Alshaeri, H.K., Van Dyke, T.E., 2019. Omega-3 fatty acids effects on inflammatory biomarkers and lipid profiles among diabetic and cardiovascular disease patients: a systematic review and meta-analysis. Sci. Rep. 9, 18867.
Ng, Q.X., Venkatanarayanan, N., Ho, C.Y., 2017. Clinical use of *Hypericum perforatum* (St John's wort) in depression: a meta-analysis. J. Affect. Disord. 210, 211–221.
Onakpoya, I.J., Posadzki, P.P., Watson, L.K., Davies, L.A., Ernst, E., 2012. The efficacy of long-term conjugated linoleic acid (CLA) supplementation on body composition in overweight and obese individuals: a systematic review and meta-analysis of randomized clinical trials. Eur. J. Nutr. 51, 127–134.
Pandey, M.K., Gupta, S.C., Karelia, D., Gilhooley, P.J., Shakibaei, M., et al., 2018. Dietary nutraceuticals as backbone for bone health. Biotechnol. Adv. 36, 1633–1648.
Payab, M., Hasani-Ranjbar, S., Shahbal, N., Qorbani, M., Aletaha, A., et al., 2020. Effect of the herbal medicines in obesity and metabolic syndrome: a systematic review and meta-analysis of clinical trials. Phytother Res. 34, 526–545.
Peth-Nui, T., Wattanathorn, J., Muchimapura, S., Tong-Un, T., Piyavhatkul, N., et al., 2012. Effects of 12-week *Bacopa monnieri* consumption on attention, cognitive processing, working memory, and functions of both cholinergic and monoaminergic systems in healthy elderly volunteers. Evid. Based Complement. Alternat. Med. 2012, 606424.
Phung, O.J., Baker, W.L., Matthews, L.J., Lanosa, M., Thorne, A., et al., 2010. Effect of green tea catechins with or without caffeine on anthropometric measures: a systematic review and meta-analysis. Am. J. Clin. Nutr. 91, 73–81.
Pompano, L.M., Boy, E., 2021. Effects of dose and duration of zinc interventions on risk factors for type 2 diabetes and cardiovascular disease: a systematic review and meta-analysis. Adv. Nutr. 12, 141–160.
Post, R.E., Mainous III, A.G., King, D.E., Simpson, K.N., 2012. Dietary fiber for the treatment of type 2 diabetes mellitus: a meta-analysis. J. Am. Board Fam. Med. 25, 16–23.
Raguzzini, A., Toti, E., Sciarra, T., Fedullo, A.L., Peluso, I., 2020. Cranberry for bacteriuria in individuals with spinal cord injury: a systematic review and meta-analysis. Oxid. Med. Cell. Longev. 2020, 9869851.
Ramasamy, S., Chin, S.P., Sukumaran, S.D., Buckle, M.J., Kiew, L.V., et al., 2015. In silico and in vitro analysis of bacoside A aglycones and its derivatives as the constituents responsible for the cognitive effects of *Bacopa monnieri*. PLoS One 10, e0126565.
Ran, L., Zhao, W., Wang, H., Zhao, Y., Bu, H., 2020. Vitamin C as a supplementary therapy in relieving symptoms of the common cold: a meta-analysis of 10 randomized controlled trials. Biomed Res. Int. 2020, 8573742.
Reader, F., 2019. How the FDA Regulates Nutraceuticals. Available. https://www.fdareader.com/blog/how-the-fda-regulates-nutraceuticals. (Accessed 3 November 2021).
Reay, J.L., Kennedy, D.O., Scholey, A.B., 2005. Single doses of *Panax ginseng* (G115) reduce blood glucose levels and improve cognitive performance during sustained mental activity. J. Psychopharmacol. 19, 357–365.
Reay, J.L., Kennedy, D.O., Scholey, A.B., 2006. Effects of *Panax ginseng*, consumed with and without glucose, on blood glucose levels and cognitive performance during sustained "mentally demanding" tasks. J. Psychopharmacol. 20, 771–781.
Reid, I.R., Bolland, M.J., Grey, A., 2014. Effects of vitamin D supplements on bone mineral density: a systematic review and meta-analysis. Lancet 383, 146–155.
Rigney, U., Kimber, S., Hindmarch, I., 1999. The effects of acute doses of standardized *Ginkgo biloba* extract on memory and psychomotor performance in volunteers. Phytother Res. 13, 408–415.
Roodenrys, S., Booth, D., Bulzomi, S., Phipps, A., Micallef, C., et al., 2002. Chronic effects of brahmi (*Bacopa monnieri*) on human memory. Neuropsychopharmacology 27, 279–281.
Rosales-Corral, S., Tan, D.X., Reiter, R.J., Valdivia-Velázquez, M., Martínez-Barboza, G., et al., 2003. Orally administered melatonin reduces oxidative stress and proinflammatory cytokines induced by amyloid-beta peptide in rat brain: a comparative, in vivo study vs. vitamin C and E. J. Pineal Res. 35, 80–84.
Rosanoff, A., Plesset, M.R., 2013. Oral magnesium supplements decrease high blood pressure (SBP>155 mmHg) in hypertensive subjects on anti-hypertensive medications: a targeted meta-analysis. Magnes. Res. 26, 93–99.
Rosanoff, A., Costello, R.B., Johnson, G.H., 2021. Effectively prescribing oral magnesium therapy for hypertension: a categorized systematic review of 49 clinical trials. Nutrients 13.
Scholey, A., Ossoukhova, A., Owen, L., Ibarra, A., Pipingas, A., et al., 2010. Effects of American ginseng (*Panax quinquefolius*) on neurocognitive function: an acute, randomised, double-blind, placebo-controlled, crossover study. Psychopharmacology 212, 345–356.
Sheikh, N., Ahmad, A., Siripurapu, K.B., Kuchibhotla, V.K., Singh, S., et al., 2007. Effect of *Bacopa monniera* on stress induced changes in plasma corticosterone and brain monoamines in rats. J. Ethnopharmacol. 111, 671–676.
Shen, N.T., Maw, A., Tmanova, L.L., Pino, A., Ancy, K., Crawford, C.V., Simon, M.S., Evans, A.T., 2017. Timely use of probiotics in hospitalized adults prevents clostridium difficile infection: a systematic review with meta-regression analysis. Gastroenterology 152 (8), 1889–1900.e9. https://doi.org/10.1053/j.gastro.2017.02.003.
Shi, X., Zhang, J., Mo, L., Shi, J., Qin, M., et al., 2019. Efficacy and safety of probiotics in eradicating *Helicobacter pylori*: a network meta-analysis. Medicine 98, e15180.
Shukla, M., Govitrapong, P., Boontem, P., Reiter, R.J., et al., 2017. Mechanisms of melatonin in alleviating Alzheimer's disease. Curr. Neuropharmacol. 15, 1010–1031.
Silberstein, R.B., Pipingas, A., Song, J., Camfield, D.A., Nathan, P.J., et al., 2011. Examining brain-cognition effects of *Ginkgo biloba* extract: brain activation in the left temporal and left prefrontal cortex in an object working memory task. Evid. Based Complement. Alternat. Med. 2011, 164139.
Steiner, G.Z., Yeung, A., Liu, J.X., Camfield, D.A., Blasio, F.M., et al., 2016. The effect of Sailuotong (SLT) on neurocognitive and cardiovascular function in healthy adults: a randomised, double-blind, placebo controlled crossover pilot trial. BMC Complement. Altern. Med. 16, 15.
Stough, C., Lloyd, J., Clarke, J., Downey, L.A., Hutchison, C.W., et al., 2001. The chronic effects of an extract of *Bacopa monniera* (Brahmi) on cognitive function in healthy human subjects. Psychopharmacology 156, 481–484.
Stough, C., Downey, L.A., Lloyd, J., Silber, B., Redman, S., et al., 2008. Examining the nootropic effects of a special extract of *Bacopa monniera* on human cognitive functioning: 90 day double-blind placebo-controlled randomized trial. Phytother. Res. 22, 1629–1634.
Suksomboon, N., Poolsup, N., Yuwanakorn, A., 2014. Systematic review and meta-analysis of the efficacy and safety of chromium supplementation in diabetes. J. Clin. Pharm. Ther. 39, 292–306.
Sun, Y., Zimmermann, D., DE Castro, C.A., Actis-Goretta, L., 2019. Dose-response relationship between cocoa flavanols and human endothelial function: a systematic review and meta-analysis of randomized trials. Food Funct. 10, 6322–6330.
Sungthong, B., Yoothaekool, C., Promphamorn, S., Phimarn, W., 2020. Efficacy of red yeast rice extract on myocardial infarction patients with borderline hypercholesterolemia: a meta-analysis of randomized controlled trials. Sci. Rep. 10, 2769.
Tabrizi, R., Vakili, S., Lankarani, K.B., Akbari, M., Mirhosseini, N., et al., 2018. The effects of curcumin on glycemic control and lipid profiles among patients with metabolic syndrome and related disorders: a systematic review and metaanalysis of randomized controlled trials. Curr. Pharm. Des. 24, 3184–3199.
Tabrizi, R., Saneei, P., Lankarani, K.B., Akbari, M., Kolahdooz, F., et al., 2019. The effects of caffeine intake on weight loss: a systematic review and dos-response meta-analysis of randomized controlled trials. Crit. Rev. Food Sci. Nutr. 59, 2688–2696.
Tai, V., Leung, W., Grey, A., Reid, I.R., Bolland, M.J., 2015. Calcium intake and bone mineral density: systematic review and meta-analysis. BMJ 351, h4183.
Taku, K., Melby, M.K., Kurzer, M.S., Mizuno, S., Watanabe, S., et al., 2010. Effects of soy isoflavone supplements on bone turnover markers in menopausal women: systematic review and meta-analysis of randomized controlled trials. Bone 47, 413–423.
Tubbs, A.S., Kennedy, K.E.R., Alfonso-Miller, P., Wills, C.C.A., Grandner, M.A., 2021. A randomized, double-blind, placebo-controlled trial of a polyphenol botanical blend on sleep and daytime functioning. Int. J. Environ. Res. Public Health 18.
Vagios, S., Hesham, H., Mitchell, C., 2020. Understanding the potential of lactobacilli in recurrent UTI prevention. Microb. Pathog. 148, 104544.

Wang, C.H., Fang, C.C., Chen, N.C., Liu, S.S., Yu, P.H., et al., 2012. Cranberry-containing products for prevention of urinary tract infections in susceptible populations: a systematic review and meta-analysis of randomized controlled trials. Arch. Intern. Med. 172, 988–996.

Wang, Y., Yang, G., Gong, J., Lu, F., Diao, Q., et al., 2016. Ginseng for Alzheimer's disease: a systematic review and meta-analysis of randomized controlled trials. Curr. Top. Med. Chem. 16, 529–536.

Wang, J., Zhang, X., Lan, H., Wang, W., 2017a. Effect of garlic supplement in the management of type 2 diabetes mellitus (T2DM): a meta-analysis of randomized controlled trials. Food Nutr. Res. 61, 1377571.

Wang, Y.Y., Zheng, W., Ng, C.H., Ungvari, G.S., Wei, W., et al., 2017b. Meta-analysis of randomized, double-blind, placebo-controlled trials of melatonin in Alzheimer's disease. Int. J. Geriatr. Psychiatr. 32, 50–57.

Wang, X., Wu, W., Zheng, W., Fang, X., Chen, L., et al., 2019. Zinc supplementation improves glycemic control for diabetes prevention and management: a systematic review and meta-analysis of randomized controlled trials. Am. J. Clin. Nutr. 110, 76–90.

Watanabe, M., Risi, R., Masi, D., Caputi, A., Balena, A., et al., 2020. Current evidence to propose different food supplements for weight loss: a comprehensive review. Nutrients 12.

Watson, N.F., Badr, M.S., Belenky, G., Bliwise, D.L., Buxton, O.M., Buysse, D., et al., 2015. Joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society on the recommended amount of sleep for a healthy adult: methodology and discussion. J. Clin. Sleep Med. 11, 931–952.

Wesnes, K.A., Ward, T., Mcginty, A., Petrini, O., 2000. The memory enhancing effects of a *Ginkgo biloba*/*Panax ginseng* combination in healthy middle-aged volunteers. Psychopharmacology 152, 353–361.

Wong, R.K., Yang, C., Song, G.H., Wong, J., Ho, K.Y., 2015. Melatonin regulation as a possible mechanism for probiotic (VSL#3) in irritable bowel syndrome: a randomized double-blinded placebo study. Dig. Dis. Sci. 60, 186–194.

Yang, C., Lu, Y., Wan, M., Xu, D., Yang, X., et al., 2021. Efficacy of high-dose vitamin D supplementation as an adjuvant treatment on pneumonia: systematic review and a meta-analysis of randomized controlled studies. Nutr. Clin. Pract. 36, 368–384.

Yeh, T.L., Shih, P.C., Liu, S.J., Lin, C.H., Liu, J.M., et al., 2018. The influence of prebiotic or probiotic supplementation on antibody titers after influenza vaccination: a systematic review and meta-analysis of randomized controlled trials. Drug Des. Dev. Ther. 12, 217–230.

Yu, J., Cao, G., Yuan, S., Luo, C., Yu, J., et al., 2021. Probiotic supplements and bone health in postmenopausal women: a meta-analysis of randomized controlled trials. BMJ Open 11, e041393.

Zaiss, M.M., Jones, R.M., Schett, G., Pacifici, R., 2019. The gut-bone axis: how bacterial metabolites bridge the distance. J. Clin. Invest. 129, 3018–3028.

Zhang, X., Li, Y., Del Gobbo, L.C., Rosanoff, A., Wang, J., Zhang, W., et al., 2016. Effects of magnesium supplementation on blood pressure: a meta-analysis of randomized double-blind placebo-controlled trials. Hypertension 68, 324–333.

Zhang, L.S., Zhang, J.H., Feng, R., Jin, X.Y., Yang, F.W., et al., 2019. Efficacy and safety of berberine alone or combined with statins for the treatment of hyperlipidemia: a systematic review and meta-analysis of randomized controlled clinical trials. Am. J. Chin. Med. 47, 751–767.

Zuccotti, G., Meneghin, F., Aceti, A., Barone, G., Callegari, M.L., et al., 2015. Probiotics for prevention of atopic diseases in infants: systematic review and meta-analysis. Allergy 70, 1356–1371.

Nuts and seeds

J. Salas-Salvadó[a] **and M. Pascual-Compte**[b], [a] Department of Biochemistry & Biotechnology, School of Medicine, IISPV, Rovira i Virgili University, Reus, Spain; and [b] Universitat Rovira i Virgili, Tarragona, Spain

© 2023 Elsevier Ltd. All rights reserved.

This is an update of J. Gray, Nuts and seeds, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 329–335, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00208-7.

Types of nuts and seeds 514
Nuts 514
Almonds 514
Brazil nuts 514
Cashews 514
Chestnuts 514
Coconut 515
Hazelnuts 515
Macadamia nuts 515
Peanuts 515
Pecans 515
Pine nuts 515
Pistachio nuts 515
Walnuts 516
Seeds 516
Pumpkin seeds 516
Sesame seeds 516
Sunflower seeds 516
Macronutrient content of nuts and seeds 516
Fat 517
Carbohydrate 517
Protein 517
Micronutrient content 518
Fiber content 518
Polyphenols and phytosterols 519
Toxins and contaminants 519
Phytic acid 519
Intolerances/allergies to nuts 520
Nuts and seeds contaminants 520
Nuts and their role in the diet and health 520
Conclusion 521
References 521
Further reading 521

Key points

- To determine the nutritional composition of major types of nuts and seeds grown for human consumption.
- To update the health benefits of nuts and seeds according to the latest research.
- To learn some curiosities about nuts and seeds.

In botanical terms, the word "nut" is used to describe a wide range of seeds, mostly from trees; however, there are some exceptions. Peanuts are the seeds of a legume and other nuts grow inside of leathery fruits. Nuts include among others the almond, cashew, chestnut, Brazil nut, hazelnut, macadamia, peanut, pecan, pine nut, pistachio and walnut. Nuts and seeds come from a diverse range of different plants (vegetables, flowers, or crops grown for a variety of uses), so their nutritional composition is quite varied. Seeds, such as pumpkin seeds, sesame and sunflower, are sources of protein, unsaturated fatty acids, various micronutrients, and fiber (non-starch polysaccharides (NSP)). Nuts and seeds have a wide range of uses. In the typical Western diet or Mediterranean diet they tend to be used either as snack items or added as minor ingredient to savory and sweet dishes, but they have wider applications

in vegetarian and other plant-based diets as important sources of protein and other nutrients. Certain nuts and seeds are also made into spreads, for example peanut butter and tahini (sesame seed spread) or vegan "cheeses" (mainly made by cashew paste).

Types of nuts and seeds

The major types of nuts and seeds grown for human consumption are shown in **Table 1**. These are the most commonly consumed nuts and seeds by the worldwide consumers.

Nuts

Almonds

The almond (*Prunus amygdalis* var. *dulcis*), sometimes called the sweet almond, is one of the oldest nut crops have long been known as a source of essential nutrients. It is believed to have originated in Southeast Asia but is now grown more widely. Nowadays, the USA is the largest almond producer, followed by Spain and Australia.

Almonds are high in fiber, vitamin E, riboflavin and minerals such as calcium, iron, magnesium, phosphorus, potassium, zinc, copper and manganese (432/2012 n.d.432/2012 n.d.; 1924/2006 n.d.). They are mainly eaten raw, roasted and/or salted.

Almond hulls are the main almond by-product used for animal feeding and energy production, which is a good way to reduce their environmental impact. In addition, interest in these by-products has increased as they have beneficial properties (caused by their content in polyphenols and unsaturated fatty acids) which can be used for new ingredients such as food, cosmetic, and pharmaceutical industries (Mandalari et al., 2010).

Brazil nuts

The Brazil nut (*Bertholletia excelsa*) tree grows in non-flooding areas in the Amazon rainforest in South America, being Bolivia, Peru and Brazil the main producing countries. A Brazil nut tree is one of the tallest of the Amazon Basin's tropical rainforest, reaching up to 50 m in height and it can produce around 250 pounds or 113 kg of Brazil nut kernels. Brazil nuts are high in selenium and vitamin E (antioxidants), which contribute to the protection of cells from oxidative stress. In fact, Brazil nuts contain more selenium than any other food on the planet.

Cashews

Cashews (*Anacardium occidentale L.*) are native to northeast Brazil but is now cultivated extensively in all tropical areas, notably in India and East Africa. The main producing countries of cashews are India, Cote d'Ivoire and Vietnam.

The cashew fruit, which contains the seed or "nut", hangs at the end of a false fruit called cashew "apple". The cashew nuts are usually eaten raw or roasted and they are high in vitamin K and minerals such as iron, magnesium, phosphorus, zinc, manganese and copper.

Chestnuts

The sweet or Spanish chestnut (*Castanea sativa)* is a native tree of southern Europe. After pollination by insects, female flowers are converted into brown fruits wrapped in a green spiky case. Inside the fruit, the seeds or "nuts" can be found, which are covered with a outer coat. The shell of the nut is hard and inedible for that reason has to be cooked before being eaten, often by roasting or boiling.

Table 1 Major types of nuts and seeds grown for human consumption.

Almonds	Pecans
Brazil nuts	Pine nuts
Cashews	Pistachios
Chestnuts	Walnuts
Coconut	Pumpkin seeds
Hazelnuts	Sesame seeds
Macadamias	Sunflower seeds
Peanuts	

Coconut

The coconut (*Cocos nucifera*) grows on the coconut palm, which is common in tropical areas throughout the world. The coconut tree is a species of palm in the *Arecaceae* family; their native origin is unknown, as the nuts were easily dispersed between islands and continents by ocean currents and explorers.

An adult palm usually carries 30–40 paripinnate leaves in its crown and attains a height of about 15–30 m when fully mature. The white coconut "meat", which can be eaten either fresh or desiccated, is actually part of the endosperm of the seed. Coconut contains more water in comparison with other nuts and also has more saturated fat than unsaturated fat.

Coconut "milk" is found in the unripe nut and is commonly used in some Asian dishes.

Hazelnuts

The most widely grown hazelnut (*Corylus avellana*) is a native of Europe, although approximately 10 different species of *Corylus* grow throughout Europe, North America, and Asia. The main producing countries of hazelnuts are Turkey, Italy, Azerbaijan, Georgia, USA, Chile and Iran, being Turkey the largest producer. Hazelnuts are high in monounsaturated fat, which may help lower LDL-cholesterol levels (Orem et al., 2013). Hazelnuts are mainly eaten raw or roasted.

Macadamia nuts

The macadamia nut (*Macadamia integrifolia*, smooth-shelled; *Macadamia tetraphylla*, rough-shelled) is a member of the *Proteaceae* family and native to eastern tropical Australia. The main producing countries of macadamia nuts are South Africa, Australia and Kenya. Macadamia nuts are high in monounsaturated fat, fiber, vitamin B1, magnesium, manganese and copper. After harvesting, the nuts are dried, roasted, and sometimes roasted and salted. Macadamias can be used as snacks or as an ingredient in confectionery products. In addition, macadamia oil is used for cooking.

Peanuts

The peanut (*Arachis hypogaea*), sometimes referred to as the ground nut or monkey nut, originated in South America. Although referred to as a nut, it is in fact part of the legume family. Spread by European explorers, the plant reached Asia, Africa, and North America. The main producing countries nowadays are China, India, and the US Peanuts are high in plant protein, an ideal food for vegetarian or vegan diets. The peanut itself may be eaten raw, roasted and/or salted.

Pecans

The pecan (*Carya illinoinensis*) is a member of the walnut family, and the tree is classified botanically as a hickory. The tree is a native of North America, grown wild along the Mississippi River and its tributaries. Pecans are high in thiamin, also known as vitamin B1, which contributes to the normal function of the heart. Pecans are also high in monounsaturated fat, which may help to reduce the risk factors of cardiometabolic disease, such as dyslipidemia, insulin resistance and hypertension (McKay et al., 2018). The United States and Mexico are the largest producers of pecans, accounting for 92% of world production. The nut is similar to the walnut, but with a more mild and sweet flavor. The pecan nut kernel is eaten fresh and in the US, it is used widely in confectionery and baked goods, such as in the well-known Pecan Pie.

Pine nuts

Pine nuts or kernels are small edible seeds, which are extracted from the cones of various species of pine. The most commonly eaten variety is that from the European stone pine (*Pinus pinea*), which is native to northern Mediterranean regions. Pine nuts are high in vitamin K, which contributes to the maintenance of normal bones. They are widely used along Europe in confectionary and especially for the famous Pesto sauce.

Pistachio nuts

The pistachio nut is the seed of the pistachio tree (*Pistacia vera* L.). Originating from West-Central Asia, it was cultivated 3000 years ago, and it has also been cultivated for many years in Mediterranean regions and in California. Nowadays, the main producing countries are the USA, Iran, Turkey, Syria and Afghanistan. The green nut kernels are mainly eaten raw, roasted, and roasted-salted. Pistachio consumption has been positively associated with some cardiometabolic risk factors in adults with well-controlled type 2 diabetes (Hadi et al., 2021).

Walnuts

The walnut tree is one of the oldest fruit tree known for humans. The walnut (*Juglans* spp.) is the common name given to approximately 20 species of trees in this family but the most important species is *Juglans regia* – also known as the English or Persian walnut –, which is believed to have originated in Ancient Persia. The main producing countries of walnuts are USA, China, Ukraine, Iran, Chile and Turkey. Walnuts are high in polyunsaturated fat. Due to their nutritional composition, they also may help to reduce the risk of heart disease (Guasch-Ferré et al., 2021).

Seeds

Pumpkin seeds

The large flat and oval-shaped seeds of the members of the pumpkin family (*Cucurbita maxima*, *Cucurbita moschata*, *Cucurbita pepo* and related species) can be dried and eaten raw, used in both sweet and savory cooked dishes, or roasted. Pepitas are known by their Spanish name. They are native from South America and are mainly grown in Brazil. Pumpkin seeds are a natural source of antioxidants, vitamins and minerals.

Sesame seeds

The sesame plant (*Sesamum indicum*) belongs to the *Pedaliaceae* family, and is native from Africa, grows in tropical, subtropical regions and southern temperate areas of the world. The seeds are small and off-white in color. They may be eaten whole or ground. These seeds are very common to be used in confectionery and baked goods. In some countries, especially in Asian countries, it is very popular to use the sesame oil in cooking. The seeds can be also ground to a paste called tahini, used in all hummus recipes.

Sunflower seeds

The sunflower seed is the seed of the sunflower plant (*Helianthus annus* L.) which is a member of the *Compositae* family. Native from Central North America, the sunflower seeds were cultivated by the native Indians, and were introduced to Europe in the sixteenth century. The flat seeds may be dehusked and eaten raw or dried as a snack, but the plant is mainly cultivated for the sunflower oil, which is widely used for cooking and also for margarine manufacture.

Macronutrient content of nuts and seeds

Raw nuts may contain 50% or more water, but these nuts are usually dried or roasted for consumption and storage, so the water content of these nuts is between 2 to 7 g/100 g of product. The exceptions are fresh coconut and chestnuts, with a moisture content of 40% and 47%, respectively. The water, macronutrients, and energy content of the nuts and seeds more commonly consumed and discussed in this article are shown below, in **Table 2**.

Table 2 Water, macronutrient, and energy content of selected nuts and seeds (per 100 g, kernel only or edible part).

					Energy	
	Water (g)	*Protein (g)*	*Fat (g)*	*Carbohydrate (g)*	*kJ*	*kcal*
Almonds (unroasted)	4.4	21.2	49.9	21.6	2424	579
Brazil nuts (unroasted)	3.4	14.3	67.1	11.7	2759	659
Cashews (unroasted)	1.7	15.3	46.4	18.1	2403	574
Chestnut (roasted)	40.5	3.2	2.2	53	1025	245
Coconuts	47.0	3.3	33.5	15.2	1480	354
Hazelnuts (unroasted)	5.3	15	60.8	16.7	2629	628
Macadamia nuts (unroasted)	1.3	7.9	75.8	13.8	3006	718
Peanuts (unroasted)	6.9	25.2	48.8	16.5	2360	563
Pecans (unroasted)	3.5	9.2	72	13.9	2890	691
Pine nuts (dried)	2.3	14.0	68.4	13.1	2820	673
Pistachios (unroasted)	4.4	20.2	45.3	27.2	2340	560
Walnuts (unroasted)	4	15.2	65.2	13.7	2740	654
Pumpkin seeds (dried)	5.2	30.2	49	10.7	2340	559
Sesame seeds (dried)	4.7	17.7	49.7	23.4	2400	573
Sunflower seeds (dried)	4.7	20.8	51.5	20	2440	584

Source: U.S. Department of Agriculture, Agricultural Research Service. FoodData Central, 2019. fdc.nal.usda.gov.

Table 3 Total fat and fatty acid composition of selected nuts and seeds (g per 100 g, kernel only).

	Total fat	*Saturated fatty acids*	*Monounsaturated fatty acids*	*Polyunsaturated fatty acids (total)*	*cis n-6 polyunsaturated fatty acids*	*cis n-3 polyunsaturated fatty acids*
Almonds	49.9	3.8	31.6	12.3	12.3	0.003
Brazil nuts	67.1	16.1	23.9	24.4	23.9	0.018
Cashews	43.8	7.78	23.8	7.84	–[a]	–[a]
Chestnuts	1.81	0.266	0.945	0.468	–[a]	–[a]
Coconut	33.5	29.7	1.42	0.366	–[a]	–[a]
Hazelnuts	60.8	4.46	45.7	7.92	0	0
Macadamia nuts	75.8	12.1	58.9	1.5	0	0.2
Peanuts	49.2	6.28	24.4	15.6	–[a]	–[a]
Pecans	72	6.18	40.8	21.6	20.6	0
Pine nuts	68.4	4.9	18.8	34.1	33.2	0.1
Pistachios	45.3	5.91	23.3	14.4	14.1	0.2
Walnuts	65.2	6.13	8.93	47.2	38.1	9.08
Pumpkin seeds	49	8.66	16.2	21	20.7	0.1
Sesame seeds	49.7	6.96	18.8	21.8	21.4	0.4
Sunflower seeds	51.5	4.5	18.5	23.1	23	0

[a]No data available.

Source: U.S. Department of Agriculture, Agricultural Research Service. FoodData Central, 2019. fdc.nal.usda.gov.

Fat

Nuts and seeds are high in fat, especially, unsaturated fats (polyunsaturated and monounsaturated fatty acids). However, the amount of fat is quite variable, ranging from approximately 76% in the macadamia nut and 72% in the pecan to approximately 44–61% in nuts such as the almond, cashew, hazelnut, and pistachio, and as low as 2% in chestnuts. The fat content of the edible seeds is around 49–51%.

The different fatty acid fractions contained in these nuts and seeds are also quite variable, as shown below, in **Table 3**. In some nuts, such as the peanut, hazelnut, and macadamia nut, monounsaturated fatty acids predominate, whereas in the walnut and in sunflower seeds polyunsaturated fatty acids predominate. The exception is the coconut, in which saturated fatty acids constitute the major fat fraction.

Carbohydrate

The carbohydrate content of nuts and seeds is quite variable. With the exception of the starch-rich chestnut (almost 53% carbohydrate), the carbohydrate content of most nuts is relatively low at approximately 10–27%. From the seeds, are the pumpkin seeds the ones with a lowest content of carbohydrates (10%). Whereas, Brazil nuts and pine nuts are the nuts with the lowest content of carbohydrates in comparison with the other nuts (11 and 13%, respectively).

Protein

The protein content of nuts is quite variable, but most nuts are considered to be a good source of protein. Only the chestnut and coconut have a low content of protein (3%). For most other nuts, the protein content is high (14–25%), being an exception the macadamias and pecans with a medium content of protein (8–9%). Pumpkin, sesame, and sunflower seeds are also high in protein (17–30%).

However, all plant foods, also nuts and seeds, contain low proportions of indispensable amino acids in comparison from those needed in the human diet. The first limiting amino acids are lysine (Brazil nut, cashew nut, hazelnut, pine nut, and walnut), sulfur amino acids methionine and cysteine (almond), tryptophan (macadamia, pecan), and threonine (peanut). Thus, although the total amount of protein in nuts and seeds may be high, these foods must be complemented by other sources of plant protein, such as legumes, quinoa, tofu and/or animal protein (meat, fish, eggs, and dairy products), to ensure that the overall protein quality of the diet is adequate.

Table 4 Vitamin content of selected nuts and seeds (per 100 g, kernel only).

	Carotene (μg)	*Vitamin E (mg)*	*Thiamin (mg)*	*Riboflavin (mg)*	*Niacin (mg)*	*Vitamin B_6 (mg)*	*Folate (μg)*
Almonds	1	25.6	0.20	1.14	3.62	0.14	44
Brazil nuts	0	5.65	0.62	0.03	0.3	0.1	22
Cashews	0	0.85	0.42	0.05	1.06	0.41	25
Chestnuts	–[a]	–[a]	0.26	0.3	1.3	0.66	110
Coconut	0	0.24	0.06	0.02	0.54	0.05	26
Hazelnuts	11	15	0.64	0.11	1.8	0.56	113
Macadamia nuts	–[a]	0.54	1.2	0.16	2.47	0.28	11
Peanuts	0	6.56	0.65	0.13	12.4	0.34	239
Pecans	29	1.4	0.66	0.13	1.17	0.21	22
Pine nuts	17	9.33	0.36	0.22	4.4	0.09	34
Pistachios	305	2.86	0.87	0.16	1.3	1.7	51
Walnuts	12	0.7	0.34	0.15	1.12	0.5	98
Pumpkin seeds	9	2.18	0.27	0.15	4.9	0.14	58
Sesame seeds	5	0.25	0.79	0.24	4.5	0.79	97
Sunflower seeds	30	35.2	1.48	0.35	8.34	1.34	227

[a]No data available.

Source: U.S. Department of Agriculture, Agricultural Research Service. FoodData Central, 2019. fdc.nal.usda.gov.

Table 5 Mineral and trace element content of selected nuts and seeds (per 100 g, kernel only).

	Sodium (mg)	*Potassium (mg)*	*Calcium (mg)*	*Magnesium (mg)*	*Phosphorus (mg)*	*Iron (mg)*	*Copper (mg)*	*Zinc (mg)*	*Manganese (mg)*	*Selenium (μg)*
Almonds	1	733	269	270	481	3.7	1.0	3.1	2.2	4.1
Brazil nuts	3	659	160	376	725	2.4	1.7	4.1	1.2	1920
Cashews	12	660	37	292	593	6.7	2.2	5.8	1.7	19.9
Chestnuts	5	726	29	137	155	2.3	0.6	1.4	2.6	–[a]
Coconut	20	356	14	32	113	2.4	0.4	1.1	1.5	10.1
Hazelnuts	0	680	114	163	290	4.7	1.7	2.4	6.2	2.4
Macadamia nuts	5	368	85	130	188	3.7	0.7	1.3	4.1	3.6
Peanuts	10	690	89	171	380	2.5	1.1	4.4	1.7	7.1
Pecans	0	410	70	121	277	2.5	1.2	4.5	4.5	3.8
Pine nuts	2	597	16	251	575	5.5	1.3	6.5	8.8	0.7
Pistachio s	1	1020	105	121	490	3.9	1.3	2.2	1.2	7
Walnuts	2	441	98	158	346	2.9	1.6	3.0	3.4	4.9
Pumpkin seeds	7	809	46	592	1230	8.8	1.3	7.8	4.5	9.4
Sesame seed	11	468	975	351	629	14.6	4.1	7.7	2.4	34.4
Sunflower seeds	9	645	78	325	660	5.2	1.8	5.0	1.9	53

[a]No data available.

Source: U.S. Department of Agriculture, Agricultural Research Service. FoodData Central, 2019. fdc.nal.usda.gov.

Micronutrient content

The vitamin and mineral contents of the nuts and seeds discussed in this article are shown in Tables 4 and 5, respectively.

In general, nuts are a good source of the vitamins B complex, including folic acid (vitamin B9) and tocopherols (vitamin E). Nuts are also high in fiber and contain minerals such as, selenium, zinc, calcium, magnesium and phosphorus. To highlight, Brazil nuts are the richest source of selenium in the planet.

Seeds contain quite large amounts of many minerals. In particular, sesame seeds are a good source of calcium. Seeds are also high in potassium, magnesium, phosphorus, iron, and in trace elements such as copper, zinc and manganese.

Fiber content

Compositional values for the total amount of fiber, and the different fiber fractions are shown in Table 6 for the nuts and seeds discussed in this article. It can be seen that nuts and seeds are good sources of fiber, similar to the amounts found in vegetables

Table 6 Total dietary fiber, as measured by the Englyst method, and fiber fractions in selected nuts and seeds (g per 100 g, kernels only).

		Fiber fractions			
			Noncellulosic polysaccharide		
	Total fiber	*Cellulose*	*Soluble*	*Insoluble*	*Lignin*
Almonds	12.5	–[a]	–[a]	–[a]	–[a]
Brazil nuts	7.5	1.6	1.3	1.4	N
Cashews	3.3	0.6	1.6	1.0	N
Chestnuts	4.1	1.1	1.3	1.7	N
Coconut	9	0.8	1.0	5.5	N
Hazelnuts	9.7	2.2	2.5	1.8	N
Macadamia nuts	8.6	1.4	1.9	2.0	N
Peanuts	8.5	2.0	1.9	2.3	N
Pecans	9.6	1.2	1.5	2.0	N
Pine nuts	3.7	N	N	N	N
Pistachios	10.6	1.3	2.7	2.1	N
Walnuts	6.7	1.1	1.5	0.9	N
Pumpkin seeds	6	1.1	1.7	2.5	N
Sesame seeds	11.8	N	N	N	N
Sunflower seeds	8.6	1.4	1.8	2.8	N

[a]Estimated value.
[b]Nutrient present in significant quantities but no reliable information available on the amount.
Source: U.S. Department of Agriculture, Agricultural Research Service. FoodData Central, 2019. fdc.nal.usda.gov.

and fruits. Nuts and seeds contain more insoluble fiber than soluble, much of which is cellulose. Of the insoluble noncellulosic polysaccharides, arabinose predominates in most nuts, although the coconut contains large quantities of mannose. Most nuts and seeds are likely to contain quite large amounts of lignin, particularly those with a tough seed coat such as sesame seeds, although actual values are not available.

Polyphenols and phytosterols

Natural antioxidants present in nuts are in the form of nutrient and non-nutrient antioxidants. Nutrient antioxidants, namely vitamins A, C and E, and minerals such as selenium, there are also non-nutrient antioxidants, namely phenolics and carotenoids. Several studies have also reported that phenolic compounds possess much stronger antioxidant activities than nutrient antioxidants (Alasalvar et al., 2020b). Phytochemicals are defined as non-nutritive, naturally occurring, and biologically active compounds found in plants. They mainly consist of phenolics, carotenoids, nitrogen-containing compounds and alkaloids. These antioxidants serve as the primary defense against lipid peroxidation by protecting body's cells from free radical damage. In addition, nuts are rich in phytosterols that contribute to lower LDL cholesterol levels (Alasalvar et al., 2020a).

Toxins and contaminants

Phytic acid

Phytic acid (*myo*-inositol hexaphosphoric acid) is found in all seeds. In addition, nuts naturally contain a high amount of phytic acid. The phytate content of the commonly eaten nuts and seeds is quite variable. In general, the oil seeds, such as sesame and sunflower, and the tree nuts (almond, walnut, hazelnut …), have higher phytate levels than the leguminous peanut or the coconut and chestnut, which is particularly low. Phytic acid is a highly effective chelator, which forms insoluble complexes with mineral cations. Therefore, its presence in some plant foods, such as nuts and seeds, has led to concerns that it may reduce the bioavailability of various dietary minerals and trace elements, including calcium, magnesium, iron, zinc, and copper. The phytic acid present in nuts and seeds can have an inhibitory effect on mineral absorption but soaking nuts in water before eating them may reduce phytic acid levels.

It is likely that in a mixed diet of animal and plant foods, dietary phytate may be of less significance than among people consuming diets where plant foods are the basis source of nutrition (vegans). Current data suggest that the mineral status of most vegetarian or vegan adults is adequate, but because of increased requirements for growth, vegetarian children may be more

vulnerable to the reduced bioavailability of minerals and trace elements, notably zinc, which could be a consequence of the ingestion of large amounts of phytate-containing plant foods.

Intolerances/allergies to nuts

Nuts have become part of any healthy diets, and this enhanced consumption is reflected in an increased prevalence of nut allergies. Tree nuts are common food allergen sources that induce IgE-mediated allergic reactions. An allergic reaction to nuts can cause severe health problems or lethal results (McWilliam et al., 2015).

According to recent study (Midun et al., 2021), epidemiological data have shown that the prevalence for tree nuts is between 0.05% and 4.9% and for peanuts between 0.5% and 3%. Recently, it has been demonstrated that a significant proportion of nut allergic patients are able to tolerate other nuts. But avoidanceremains the key treatment of management of nut allergy. In case of the children population, researchers from the LEAP study (Learning Early About Peanut Allergy) observed that the introduction of peanuts into an infant's diet, prior to 11-months old, reduced the prevalence of peanut allergy by approximately 70–80%. In addition, they found that this tolerance to peanut remains after one year of not eating peanuts. In case of nut-allergic subjects, complete avoidance of nuts is the safest approach. On the other hand, nut intolerance is less frequent. The symptoms of a food intolerance usually occur several hours after eating the food and you need to eat a larger amount of food. In addition, symptoms are less serious than the allergy symptoms. For example, a food intolerance is never a life-threatening event, unlike an allergy case (Borres et al., 2022).

Nuts and seeds contaminants

Nuts and seeds can be contaminated by pathogens at any stage during production, processing, storage, and distribution. Certain molds produce secondary metabolites, which are toxic to humans and animals, known as the mycotoxins. Of these mycotoxins, the aflatoxins, notably aflatoxin B_1, are produced by three closely related species of mold: *Aspergillus flavus*, *Aspergillus parasiticus*, and *Aspergillus nomius*. These molds may contaminate tree nuts, but one of the most important crops to be affected is the peanut. Aflatoxins and ochratoxins, which are produced by *Aspergillus* species, have been found to contaminate nuts.

Some species of mold are able to proliferate within growing crops even before they are harvested. This relationship has been found to exist between *A. parasiticus* and peanuts. There are regulatory limits for the aflatoxin levels in foods. For example, in the UK, the sale of nuts for direct human consumption is prohibited if the aflatoxin content exceeds 4 $\mu g\,kg^{-1}$ or 10 $\mu g\,kg^{-1}$ for nuts, which are to be subjected to further processing before being sold. Nowadays, contaminants controls are very strict and have to accomplish the law of each country, in order these nuts can be sold. The numbers are low and their significance in public health terms is small.

Nuts and their role in the diet and health

Nuts and seeds are recognized as healthy foods because of their unique nutritional composition which can make a useful contribution to the dietary intake of macronutrients (notably plant protein, and unsaturated fatty acids), micronutrients (vitamins and minerals), dietary fiber, and energy. Although nuts and seeds are not in the basis of an average Western diet, they play a key role in the Mediterranean diet or diets of Western vegetarians, especially vegans.

All nuts are good sources of dietary fiber, which may help improve the digestive health. In addition, they provide macronutrients, micronutrients, fat-soluble bioactives and phytochemicals. Because nuts are an energy-dense food containing a high amount of fat, there is a concern regarding body weight increase (Julibert et al., 2020). Contrary, due to their fat composition (high in unsaturated fatty acids), nuts may help improve cardiovascular health and prevent CVD diseases and there is enough scientific evidence that frequent nut consumption does not lead to weight gain or increase in the risk of abdominal obesity when incorporated into healthy diets (Nishi et al., 2021). A frequent nut consumption has been associated with lower rates of coronary heart disease and total cardiovascular disease (CVD) incidence and mortality (Becerra-Tomás et al., 2019). Nut intake was also related to lower rates of sudden death as well as of peripheral artery disease, atrial fibrillation, and all-cause and CVD mortality. Nuts have a lipid-lowering effect, so this explains their protective effect on CVD. In addition, nut consumption may help improve endothelial function, reduce postprandial glycaemia and insulin resistance (Becerra-Tomás et al., 2021). Recent studies have also been shown that nut intake may change gut microbiota composition and metabolism.

In summary, a large body of scientific evidence suggests that individuals who regularly consume around 30–42 g/day of nuts, disclose lower rates of highly prevalent chronic non-communicable diseases such as cardiovascular diseases, diabetes and cognitive decline. Thus far, the FDA has approved three qualified health claims on nuts in general and walnuts and macadamias in particular, whereas the EFSA has approved one authorized health claim for walnuts and endothelial function. Nuts and seeds has been included in different dietary guidelines and association reports on goals for health promotion and disease reduction for 2020 (Alasalvar et al., 2020a).

Conclusion

Nuts and seeds are energy-dense foods, which stand out for their unique nutritional composition and health benefits. They are high in dietary fiber, plant protein, polyunsaturated and monounsaturated fatty acids, several vitamins and minerals. Due to its interesting nutrient composition, especially fat composition, nut consumption has been associated with cardiovascular disease prevention, as well as postprandial glycaemia and insulin resistance decrease, which helps reducing the risk of type 2 diabetes mellitus. There is a large body of scientific evidence, which suggests that a regular consumption of nuts (around 30 g per day) has been associated with lower rates of highly prevalent chronic non-communicable diseases such as cardiovascular diseases, diabetes and cognitive decline.

See Also: Fatty acids: Metabolism; Dietary fiber: Physiological effects and health outcomes; Folate/folic acid; Mycotoxins: Toxicity, occurrence, risk assessment and prevention; Nutrigenetics; Nutrigenomics; Protein quality and sources; Trans-fatty acids: Health effects, recommendations, and regulations

References

1924/2006, R. (EC) N., (n.d.). Regulation (EC) No 1924/2006 of the European Parliament and of the Council of 20 December 2006 on Nutrition and Health Claims Made on Foods. [WWW Document]. URL https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32006R1924&qid=1643014101669&from=ES (accessed 1.24.2022).

432/2012, C.R. (EU) N., (n.d.). Commission Regulation (EU) No 432/2012 of 16 May 2012. [WWW Document]. URL https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32012R0432&qid=1643014025142&from=ES (accessed 1.24.2022).

Alasalvar, C., Salas-Salvadó, J., Ros, E., Sabaté, J., 2020a. Health Benefits of Nuts and Dried Fruits, first ed. CRC Press.

Alasalvar, C., Salvadó, J.S., Ros, E., 2020b. Bioactives and health benefits of nuts and dried fruits. Food Chem. 314. https://doi.org/10.1016/J.FOODCHEM.2020.126192.

Becerra-Tomás, N., Paz-Graniel, I., Kendall, W.C., et al., 2019. Nut consumption and incidence of cardiovascular diseases and cardiovascular disease mortality: a meta-analysis of prospective cohort studies. Nutr. Rev. 77, 691–709. https://doi.org/10.1093/nutrit/nuz042.

Becerra-Tomás, N., Paz-Graniel, I., Hernández-Alonso, P., et al., 2021. Nut consumption and type 2 diabetes risk: a systematic review and meta-analysis of observational studies. Am. J. Clin. Nutr. 113, 960–971. https://doi.org/10.1093/ajcn/nqaa358.

Borres, M.P., Sato, S., Ebisawa, M., 2022. Recent advances in diagnosing and managing nut allergies with focus on hazelnuts, walnuts, and cashew nuts. World Allerg. Org. J. 15 (4), 100641. https://doi.org/10.1016/j.waojou.2022.100641.

Guasch-Ferré, M., Hernández-Alonso, P., Drouin-Chartier, J.-P., et al., 2021. Walnut consumption, plasma metabolomics, and risk of type 2 diabetes and cardiovascular disease. J. Nutr. 151, 303–311. https://doi.org/10.1093/jn/nxaa374.

Hadi, A., Asbaghi, O., Kazemi, M., et al., 2021. Consumption of pistachio nuts positively affects lipid profiles: a systematic review and meta-analysis of randomized controlled trials. Crit. Rev. Food Sci. Nutr. 1–14. https://doi.org/10.1080/10408398.2021.2018569.

Julibert, A., del Mar Bibiloni, M., Gallardo-Alfaro, L., et al., 2020. Metabolic syndrome features and excess weight were Inversely associated with nut consumption after 1-year follow-up in the PREDIMED-plus study. J. Nutr. 150, 3161–3170. https://doi.org/10.1093/jn/nxaa289.

Mandalari, G., Bisignano, C., D'Arrigo, M., et al., 2010. Antimicrobial potential of polyphenols extracted from almond skins. Lett. Appl. Microbiol. 51, 83–89. https://doi.org/10.1111/J.1472-765X.2010.02862.X.

McKay, D.L., Eliasziw, M., Chen, O., et al., 2018. A pecan-rich diet improves cardiometabolic risk factors in overweight and obese adults: a randomized controlled trial. Nutrients 10. https://doi.org/10.3390/NU10030339.

McWilliam, V., Koplin, J., Lodge, C., Tang, M., Dharmage, S., Allen, K., 2015. The prevalence of tree nut allergy: a systematic review. Curr. Allergy Asthma Rep. 15, 54.

Midun, E., Radulovic, S., Brough, H., Caubet, J.C., 2021. Recent advances in the management of nut allergy. World Allerg. Org. J. 14 (1), 100491. https://doi.org/10.1016/j.waojou.2020.100491.

Nishi, S.K., Viguiliouk, E., Blanco Mejia, S., et al., 2021. Are fatty nuts a weighty concern? A systematic review and meta-analysis and dose–response meta-regression of prospective cohorts and randomized controlled trials. Obes. Rev. 22. https://doi.org/10.1111/obr.13330.

Orem, A., Yucesan, F.B., Orem, C., et al., 2013. Hazelnut-enriched diet improves cardiovascular risk biomarkers beyond a lipid-lowering effect in hypercholesterolemic subjects. J. Clin. Lipidol. 7, 123–131. https://doi.org/10.1016/J.JACL.2012.10.005.

Further reading

Bolling, B.W., McKay, D.L., Blumberg, J.B., 2010. The phytochemical composition and antioxidant actions of tree nuts. Asia Pac. J. Clin. Nutr. 19 (1), 117–123.

Greenhawt, M.J., McMorris, M.S., Furlong, T.J., 2009. Self-reported allergic reactions to peanut and tree nuts occurring on commercial airlines. J. Allergy Clin. Immunol. 124 (3), 598–599.

Hartmann, H.T., Flocker, W.J., Kofranek, A.M., 1981. Plant Science. Growth, Development, and Utilization of Cultivated Plants. Prentice Hall, Englewood Cliffs, NJ.

Fruit and nuts. In: Holland, B., Unwin, I.D., Buss, D.H. (Eds.), 1992. The First Supplement to McCance & Widdowson's the Composition of Foods, fifth ed. The Royal Society of Chemistry, London.

Kuhnle, G.G., Dell'Aquila, C., Aspinall, S.M., Runswick, S.A., Mulligan, A.A., Bingham, S.A., 2008. Phytoestrogen content of beverages, nuts, seeds, and oils. J. Agric. Food Chem. 56 (16), 7311–7315.

Lopez-Uriarte, P., Bullo, M., Casas-Agustench, P., Babio, N., Salas-Salvado, J., 2009. Nuts and oxidation: a systematic review. Nutr. Rev. 67 (9), 497–508.

Macfarlane, B.J., Bezwoda, W.R., Bothwell, T.H., et al., 1988. Inhibitory effect of nuts on iron absorption. Am. J. Clin. Nutr. 47, 270–274.

Ministry of Agriculture, Fisheries and Food, 1994. The Dietary and Nutritional Survey of British Adults – Further Analysis. HMSO, London.

Moodley, R., Kindness, A., B Jonnalagadda, S., 2007. Elemental composition and chemical characteristics of five edible nuts (almond, Brazil, pecan, macadamia, and walnut) consumed in Southern Africa. J. Environ. Sci. Health Part B 42 (5), 585–591.

Morris, E.R., 1993. Phytic acid. In: Macrae, R., Robinson, R.K., Sadler, M.J. (Eds.), Encyclopaedia of Food Science, Food Technology and Nutrition. Academic Press, London, pp. 3587–3591.

Moss, M.O., 1996. Mycotoxins. Mycol. Res. 100, 524–526.
Phillips, K.M., Ruggio, D.M., Ashraf-Khorassani, M., 2005. Phytosterol composition of nuts and seeds commonly consumed in the United States. J. Agric. Food Chem. 53 (24), 9436–9445.
Sabate, J., Nuts, Y.A., health outcomes, 2009. New epidemiologic evidence. Am. J. Clin. Nutr. 89 (5), 1643S–1648S.
Young, V.R., Pellett, P.L., 1994. Plant proteins in relation to human protein and amino acid nutrition. Am. J. Clin. Nutr. 59, 1203S–1212S supplement.

Organic foods

Léon Guéguen and Gérard Pascal, French Academy of Agriculture, Paris, France

© 2023 Elsevier Ltd. All rights reserved.

This is an update of Guéguen L., and Pascal G. (2016) Organic Foods. Reference Module in Biomedical Sciences. Elsevier. 13-May-16 doi: 10.1016/B978-0-12-801238-3.99501-1.

Introduction 523
Scientific bases for the comparisons of food quality 524
Essential or beneficial nutrients 524
Plant products 524
Animal products 525
Undesirable components 526
Health effects 527
Clinical trials 527
Epidemiological studies 527
Organoleptic quality 528
Conclusions and prospects 528
References 529

Key points

- The chemical composition of organic foods is not significantly different from that of conventional foods for the main nutrients
- Higher levels of polyphenolic compounds in certain organic fruits and of omega-3 polyunsaturated fatty acids in organic milk from grass-fed cows do not contribute significantly to the total dietary intakes
- Pesticide residues in conventional plant products are almost always lower than the allowed maximum limit
- Organic foods also contain natural toxins and residues of allowed pesticides
- Some beneficial associations observed between the consumption of an organic diet and several pathologies are not causal relationships but are due to the better hygiene of life and the dietary behavior of "organic eaters"
- Better nutritional and health values should therefore no longer be claimed for choosing to consume organic foods
- Organic farming cannot be sustainable, especially for cereals, without the proximity of animal husbandry to provide the essential organic fertilizers.
- The lower productivity of organic agriculture require more land, at the expense of biodiversity and carbon storage, and is not sufficient to feed the future world population.

Introduction

Organic farming, also known as biological or ecological agriculture, should follow a well-defined regulation and the mode of production must be controlled by a certification body. Its main constraints are a minimal use of off-farm inputs, the ban of synthetic chemicals including fertilizers, pesticides, drugs, the use of organic fertilizers and natural pesticides, long crop rotations, the maintenance of organic matter and microbial life in the soil, the ban of genetically modified plants. For organic husbandry, regulations concern welfare, use of organic feeds and limitation of therapeutic interventions (especially antibiotics and hormones). When necessary, derogations are granted. The conversion time from conventional to organic farming varies from 2 to 3 years. Organic farming involves obligations in relation to means of production but not to results with respect to nutritional, health, or sensory properties of the products. Foods marketed under the label "organic" must contain at least 95% ingredients from organic agriculture, with possible food additives from a restrictive list. For processed foods, only biological and physical treatments, except irradiation, are allowed.

Certified organic agriculture represents currently approximately 1.5% of agricultural land worldwide, over two-thirds are grassland-grazing areas. This average hides large variations between continents and countries: for example, 0.3% in Africa and Asia, 1% in North America, 3.3% in Europe (8% in the European Union) with a maximum of 25% in Austria. In the last case, the mountain regions with predominance of grazing livestock are more conducive to organic farming. About half of the world organic land is in Australia but 97% is extensive pasture. In contrast, in sub-Saharan Africa, farmers who have no access to chemical fertilizers and pesticides unknowingly practice a form of non-certified organic farming.

The organic market is increasing everywhere. The demand is mainly motivated by health maintenance, protection of the environment, including water resources, but also for ethical reasons. Thus, fear of "chemical" and attraction to the "natural" are the main reasons for choosing organic foods. Productivity is much lower in organic agriculture, the yields being 20–50% lower than in conventional farming, especially in the case of cereals. As costs of production are higher, selling prices are also significantly higher (30–100% more in stores). The market for organic foods ranges from 1% to 8% of food purchased, according to country, but demand is growing and most of the major consuming countries are dependent on imports.

Although motivations of consumers' choice, whether ethical or ideological, are not discussed in this article, it is important to know whether the widespread belief of better nutritional and health quality for organic foods is supported by scientific studies.

Scientific bases for the comparisons of food quality

The chemical composition of foods varies depending on many factors which are not all directly linked to the mode of production. These are, for crops, species and variety, ripeness at time of harvest, yield, climate, season, and for animal products, breed, age, fat content, growth rate and feed intake. It is obvious that when these conditions are comparable, it is unlikely that tissue composition is different. The clearest example is that of milk, whose lipid composition depends strongly on the consumption of grass or fresh forage, regardless of the production system, being organic or conventional.

The differences between organic and conventional products have been the matter of numerous studies and reviews. Several studies failed to demonstrate a significant change in product composition attributable to the mode of production, either because the experimental design or characterization of food were not suitable or well defined, or because statistical interpretation was missing or impossible. This resulted in a lack of useable compositional results, particularly for some vitamins.

It is significant that most reviews published by organizations or associations of organic farming drawed a positive balance in favor of organic foods (Soil Association, 2000; Worthington, 2001; Organic Trade Association, 2008; Leifert and Niggli, 2009), whereas independent academic publications claimed that there was an overall lack of significant differences (Williams, 2002; Winter and Davis, 2006; Magkos et al., 2006; Williamson, 2007). The data summarized below result mainly from several reviews published before 2010 in Germany (Woëse et al., 1997), in UK by the Food Standards Agency (Dangour et al., 2009, 2010), in France in 2003 by the French Agency for Food Safety (Afssa, now Anses) with a 2010 update (Guéguen and Pascal, approximately 100 new references added), an US systematic review (Smith-Spangler et al., 2012), two Danish critical reviews (Jensen et al., 2012, 2013) and, more recently, three European comprehensive systematic reviews and meta-analyses (Baranski et al., 2014; Średnicka-Tober et al., 2016; STOA, 2016; Mie et al., 2017) and a French critical review (Guéguen, 2018).

Essential or beneficial nutrients

Plant products

A trend toward higher dry matter content in some organic vegetables has sometimes been reported. This difference is not systematic and depends greatly on the cultivar and stage of ripeness at harvest. No significant differences were observed for carbohydrate, protein, and total lipids of vegetables and fruits, particularly when expressed on a dry matter basis.

Organic cereals are generally poorer in protein, which raises, in the case of wheat, technological problems for bread production and requires the choice of adapted varieties. However, it seems that the balance of essential amino acids is sometimes better in organic compared to conventional wheat.

For minerals and trace elements, it can be concluded, from a large number of individual results on approximately 15 species of fruits and vegetables, that overall there is a lack of differences in levels. Occasionally, magnesium and iron contents have a trend to be somewhat higher in some organic vegetables, but no differences were observed for most minerals (calcium, potassium, sodium, copper, zinc, or selenium). Some extremely high levels of copper were found in potatoes, tomatoes, and grapes, due to repeated treatments with copper sulfate. The mineral composition of seeds is almost constant and no difference was recorded. Mineral and trace element contents of bread depend on the proportion of bran in the flour and not on the mode of production of wheat.

Too few studies have been done on vitamins A, E, and B group in vegetables and fruits, but the differences between organic and conventional products seem small. However, many studies have examined vitamin C, especially in tomatoes (Dangour et al., 2009), and a trend toward higher levels in some organic fruits and vegetables was confirmed but not systematically: for 43 individual comparisons, 19 (44%) were higher, 18 (42%) were equal, and 6 (14%) were lower (Guéguen and Pascal, 2010). This trend must be tempered by taking into account other factors that affect even more the level of vitamin C, especially the freshness of the product.

Several studies have shown the presence of larger amounts of phytomicrocomponents such as polyphenols and other phenolic compounds in organic plants. In the absence of phytopharmaceutical treatment, this increase can be explained by an increased defense reaction of the plant unprotected against attacks from insects or fungi. The production of secondary metabolites, including molecules with beneficial antioxidant properties, is then promoted. Low availability of soil nitrogen could also have the same effect. Thus, the levels of polyphenols, especially flavonoids, are sometimes 20–40% greater in organic vegetables and fruits, as mostly shown for tomato. However, this observation is not systematic because among 70 individual validated data across all products,

31 (54%) were higher in organic agriculture, 23 (40%) were equal, and three (5%) were lower (Guéguen and Pascal, 2010). In contrast, levels of carotenoids in organic fruits and vegetables are essentially lower or equal, especially for lycopene in tomatoes.

Several studies have shown that the antioxidant power of organic apples is 10–15% greater than that of conventional apples, but it remains to show that these small differences have any nutritional significance.

A systematic review (Baranski et al., 2014) on nutritional quality of conventional vs. organic foods, supported by the UK Food Standards Agency, has found significant differences only for nitrogen (protein in grains and nitrate in vegetables), being lower in organic products, and phosphorus, being lower in conventional products. However, it is important to remind that phosphorus is not a limiting factor in the human diets. Despite the use of very strict criteria for exclusion of publications, this review is currently the most comprehensive.

The data of this last meta-analysis were utilized by an international group of experts in a Science and Technology Options Assessment (STOA) report submitted in 2016 to the European Parliament (report condensed in the review of Mie et al., 2017), the conclusions of which are as follows: "In summary, collectively the published meta-analyses indicate a modestly higher content of phenolic compounds in organic crops, which is plausible. These compounds are believed to play a role in preventing several non communicable diseases in humans, although the detailed mechanisms are not generally well understood. It is important to bear in mind that in many cases the variation in the concentration of phenolic compounds is greater between different types and varieties of crops and between years, climates, soils, etc. than between production systems".

For all the very many other constituents analyzed (dry matter, fibers, minerals, trace elements, vitamins, nitrate, etc.), the differences are not significant or have no nutritional impact. The lower protein content of organic cereals, which affects the baking value of wheat, is confirmed by all systematic reviews. A better protein (e.g., amino acid balance) quality of organic cereals, in particular slightly higher contents of threonine and leucine in wheat and more globulin and albumin in triticale have sometimes been observed. However, these differences are not always significant and have not been confirmed by all recent reviews. As for mineral and trace elements, it is well known that their contents are not very variable in the grain and do not depend on fertilization.

Here is the general conclusion of the STOA European report on organic plants: "Most aspects of crop composition, including vitamins and minerals, are not affected by the agricultural management system. If they are it is only to a limited extent. From the perspective of nutritional guidelines, which are generally concerned with macronutrients, vitamins and minerals, there is no reason to prefer organic over conventional plant foods or vice versa".

Animal products

The diet of organically farmed animals is not very different from that of conventional animals, particularly for pigs and poultry, which eat mainly grains and oil-seed cakes. The only difference is that most feeds (95–100%) must come from organic farms. However, the chemical composition of these feeds, mainly seeds, depends little on production mode. A supplement of minerals, trace elements, and vitamins in organic diets is allowed.

For most constituents, the composition of meat depends little on the mode of production. Only lipid composition varies according to animal's breed and age. Thus, comparisons must be made on animals of the same age because the degree of adiposity varies accordingly. For example, an "industrial" 40-day-old chicken cannot be compared to an 80-day-old organic chicken. Feed can affect fatty acid composition, especially of unsaturated fatty acids in meat, and more in pigs and poultry than in ruminants. Thus, the meat of cattle mainly fed with grass or fresh forage, preferred in organic husbandry, is richer in polyunsaturated fatty acids (PUFA) omega-3 and sometimes in conjugated linoleic acid (CLA) than the meat of cattle fed with corn silage and concentrate. It is the same for the composition of eggs. However, similar compositions are obtained for extensive grazing livestock or for hens with an outdoor run.

Several recent and validated studies have focused on the composition of milk which, for most of its constituents (protein, lactose, minerals, and trace elements), is fairly constant. Only vary omega-3 PUFA levels and CLAs that depend heavily on the season and the proportion of grass in the diet (Butler et al., 2008). Similar results, even better, can be obtained by extensive livestock grazing or by incorporation of clover or flaxseed. Only two trace elements from milk, iodine and selenium, are influenced by dietary intake and their levels are increased by the use of a mineral supplement. Selenium content of feed strongly depends on soil concentration.

Modeling was done in the recent STOA report. Based on results from the meta-analysis by Średnicka-Tober et al. (2016), the increase in PUFA content of organic meat is not clearly demonstrated. Even under the high assumption of the all-organic food scenario, the maximum nutritional impact would be less than 7% of the requirement. Here are excerpts from the findings of this report on omega-3 in humans …"On average, replacing conventional with organic dairy products while keeping the diet constant will increase the intake of omega-3 PUFA by approximately 4%. Replacing conventional meat products with organic meat products may increase the omega-3 PUFA intake by an additional 6% … Accordingly, at this point, there is no strong evidence available that would support the existence of health benefits of a higher ruminant fatty acid content in organic compared to conventional milk … It is therefore not possible to conclude any specific health benefit offered by a modest increase in omega-3 PUFA intake from a change from conventional to organic milk and meat."

The desirable increase in the intake of omega-3 fatty acids in the diet is much more effective by using certain vegetable oils (flaxseed, rapeseed) or fatty fish, which simultaneously leads to a decrease in intake of saturated fatty acids.

Another conclusion of all the reviews is that organic milk is always poorer in iodine and selenium, two essential trace elements added to the mineral supplement in intensive conventional breeding.

Undesirable components

Nitrate is abundant in vegetables but not in fruits, cereals, milk, or meat. Nitrate levels in vegetables vary according to several factors, including sunlight, rainfall, and especially nitrogen fertilization. Nitrate levels are lower in greenhouse production, and in autumn compared to spring and summer for field production. They increase when nitrogen is provided by rapidly available soluble fertilizers. Organic farming uses mostly organic fertilizers such as guano, meat or blood meals, which are rapidly assimilated by the plant and thus also lead to high levels of nitrate. Nevertheless, the published comparisons show that organic vegetables (lettuce, spinach, rucola, carrot, beet, etc.) have usually lower nitrate levels than conventional vegetables.

Nitrate does not have a good reputation because of past accidents of methemoglobinemia in infants with poor food hygiene, notably contaminated feeding bottles in which microorganisms accelerated the reduction of nitrate to nitrite. Improvement of hygiene leads to a very low risk nowadays. In adults, the formation of carcinogenic nitrosamines has long been suspected. In fact, many studies have shown that nitrate of vegetables, which represent approximately 75% of ingested nitrate, have no negative effect on health in adults and their carcinogenic effect has not been demonstrated (European Food Safety Authority, 2008). It is noteworthy that water is also a vector of nitrate in the diet but the maximum limit (usually 50 mg L-1) is not a threshold of toxicity for adults. In addition, several studies (Katan, 2009; Hord et al., 2009) emphasize the beneficial effects of nitrate, thanks to the formation of nitrogen monoxide (NO), especially in the immune protection of the mouth and stomach, in the prevention of hypertension and cardiovascular disorders.

The fear of residues of synthetic pesticides in conventional plant products is by far the main reason for the choice of organic foods by consumers. It is true that the ban of their use in organic agriculture should logically lead to the absence of residues in food. Many surveys have been conducted on these residues in the US, UK, New Zealand, Netherlands, France, and elsewhere. For example, an annual report of the European Food Safety Authority (EFSA) compiles the results of legal controls in the Member States on a large number of plant products. For 2018, no detectable traces of synthetic pesticide residues were found in 58% of samples (all origins confounded), traces were detected in 40.5% but at levels below the maximum residue level (MRL), and only 1.5% samples exceeded the MRL. These results overlap well with those of other national studies. It is noteworthy that the tests are done on raw, unwashed, unpeeled products.

MRL is the pesticide residue level in a particular food following its production according to Good Agricultural Practice (GAP). There is an MRL for each pesticide and each plant species, and values adopted have been harmonized in the UE in 2008. Taking into account food habits of consumers, the sum of residues of a given pesticide from all food respecting MRL must not from a regulatory point of view exceed its acceptable daily intake (ADI). However, when a residue of one food is found above the MRL, the result does not automatically mean the levels of residue found are a risk to people's health. What matters is the sum on all residues in the total diet. ADI itself is calculated with a safety factor of at least 100, from the highest dose without effect observed in laboratory animals.

These assessments are based on a calculation of risk, taking into account the possible cumulative effects of several molecules of the same chemical group having the same mechanism of action, and leading to a "reasonable certainty of no harm".

If synthetic pesticides raise a distinct risk to health (e.g., skin disorders and hematopoietic cancers) for the farmer highly exposed and poorly protected, this should not be amalgamated with the negligible risk to consumers who ingest doses of residues of the order of one million times lower. Based on studies published by international bodies, residues of synthetic pesticides are without any risk to the consumer and the expected marginal benefit of eating more organic products is insignificant. Thus, according to an UE study, individual chronic exposure to pesticides would be between 0 and 0.2% of the acceptable daily intake (ADI). Other studies done in the US by the FDA also show pesticide exposures below 1% of ADI, or approximately 10,000 fold lower than the highest dose having no effect on the animal. A study from the University of Oxford found no difference in overall cancer risk when comparing 180,000 women aged 50 years or over who reported never eating organic food with around 45,000 women usually or always eating organic food.

To be on the safe side, it is advisable, especially for infant feeding, to wash and peel vegetables and, if possible, fruits. It should be noted that there is a consensus on the beneficial health effects of consumption of fruits and vegetables, whereas nearly half contain detectable residues of synthetic pesticides. Arguments based on pesticide residues (and nitrate in the case of vegetables) should not, therefore, be used as a pretext for decreasing the recommended consumption of fruits and vegetables.

It is true that organic plant products do not generally contain residues of synthetic pesticides. However, surveys sometimes reveal their presence at levels below the MRL, which is due to pollution, errors, faults, or derogations. Furthermore, organic products can contain residues of natural pesticides authorized as rotenone, pyrethrins, azadirachtin from neem oil (from *Azadiracta indica*), and particularly copper often heavily used. These residues are not taken into account in official inspections, although their safety is not guaranteed. Thus, rotenone is neurotoxic and is banned in the EU since 2008, whereas azadirachtin from neem oil, permitted in some countries (e.g., by derogation in France), is an endocrine disruptor. Copper in excess poisons the soil and is not without health consequences.

Like the beneficial antioxidants, toxic secondary metabolites can be formed as a defensive response to attacks by insects or fungi in untreated plants. The effects on human health of hundreds of natural toxins acting as insecticide or fungicide have not been well studied. Some such as cruciferous glucosinolates (sometimes beneficial), glyco-alkaloids in potatoes and tomatoes and celery furanocoumarins are well known. Others have not been identified or studied, and it may be important consider their effects on health.

Similarly, lipid transport proteins in Rosaceae (most edible fruits) are defense proteins responsible for severe allergies in children and adults. Studies have shown that they are more abundant in the skin of organic apples and plums whose consumption should not be recommended for allergic patients (Barré et al., 2009).

It would be logical to find higher levels of carcinogenic mycotoxins in organic cereals not protected by antifungal pesticides. In fact, if cases of severe contamination by mycotoxins have been found in organic cereals, the difference with the conventional grains is not systematic (Murphy et al., 2006). Thus, the presence of Fusarium mycotoxins in wheat depends on many factors and several recent studies in various UE countries (Germany, UK, Netherlands, Italy) showed that organic wheat is sometimes the least contaminated. The main factors of variation in mycotoxin levels are the year, the climate, and the storage conditions.

The risk of pollution by heavy metals, polychlorobiphenyls (PCBs), or dioxins is not different in the two modes of production but it depends on exposure to atmospheric deposition. All outdoor productions, animal or vegetable, are most at risk. This is usually the case for organic agriculture but also often for conventional agriculture, especially for the meat of grazing animals (Dervilly-Pinel et al., 2017). Thus, milk from cows on pasture, whether organic or conventional, is less well protected than the milk from cows in barn because the consumption of grass and soil may be an important vector of various pollutants. It is the same for eggs from hens with outdoor run, often more contaminated than eggs from hens reared in cages.

For similar reasons, the risk of bacterial or viral contamination is greatest in outdoor production, from hydrotelluric sources (e.g., *Clostridium botulinum*) or contact with wildlife fauna (e.g., *Campylobacter*).

Microbial contamination of plants can be enhanced by the use of organic fertilizers compared to mineral fertilizers. Thus, the use of manure, even composted, increases the risk of contamination of fruits and vegetables by *Escherichia coli*, *Salmonella*, or *Listeria monocytogenes*. On the one hand, poultry manure from organic livestock is often contaminated with *Campylobacter*, which may be an increased risk for eggs. However, according to several studies, these are only trends and, on the other hand, it appears that the resistance of these bacteria to various antibiotics is lower in organic farming.

Cases of mastitis in dairy cows are more frequent without that with antibiotic treatment. However, milk is not sold then and a limited use of antibiotics in livestock is also authorized in organic husbandry, if required. Internal parasites in sheep are conditions prevalent in organic farming, but the new UE regulation now allows the use of antiparasite treatments without limitation.

The use of hormones is universally prohibited in organic animal husbandry, but also in conventional husbandry in the UE. Several countries (USA, Canada, Australia, South Africa, New Zealand, Mexico, Chile) authorize steroid hormones (estradiol, progesterone, testosterone, zeranol, trenbolone acetate, and melengestrol acetate) to increase meat yield in beef cattle, or a protein hormone, rBST (recombinant bovine somatotropin) to increase milk production in dairy cows (prohibited in Canada based on concerns about health effects, including mastitis in treated animals). The Food Safety authorities of these countries ensure that the possible residues found in meat and milk do not present any health issue for the consumer (e.g., early puberty in girls, risk of breast cancer or allergy, endocrine disruption have been mentioned in countries where these treatments are not allowed) if the treatments are applied according to the regulation.

Health effects

Clinical trials

Given the many factors that determine the nutritional and health qualities of agricultural products, it is difficult to demonstrate significant differences resulting specifically from the production system, organic or conventional.

Ten controlled clinical studies comparing organic and conventional foods were included in the systematic review supported by the UK Food Standards Agency. Most were carried out on subjects consuming conventional or organic vegetables and fruits, and the antioxidant status of blood plasma was used as a biomarker. None of these studies could demonstrate a positive effect of diet on this blood parameter but this does not suffice, however, to characterize the health status. There is no published long-term study comparing the health effects of a diet exclusively organic or conventional, using several criteria relevant to health. This lack is unfortunate, but the small differences found in the composition of foods would leave little chance of finding a significant health effect.

In several trials, none of the numerous measured health biomarkers were affected by the cultivation system, particularly for the effect of organic fruits and vegetables on blood anti-oxidant status. In fact, the antioxidant capacity of the diet depends, for about half, on the polyphenolic compounds which are mostly provided, in Europe, by tea and coffee (50%), followed by cocoa and red wine, the contribution of fruits and vegetables being less than 10%. Therefore, an eventual increase of 15–30% of the content of polyphenolics in organic fruits cannot have a detectable effect on the blood anti-oxidant status.

Epidemiological studies

Observational epidemiological studies involving large cohorts of consumers receiving long-term organic or conventional food are scarce. This is primarily due to the fact that it is difficult to recruit participants who regularly consume a significant proportion of certified organic foods (which represent in France less than 5% of the foods) and that a fully organic diet is exceptional.

The largest prospective study was published by epidemiologists from Oxford University in 2014 (Bradbury et al., 2014). It was conducted over 9 years on more than 600,000 women older than 50 years. It showed that the consumption of organic food had no influence on the incidence of common cancers, except perhaps on non-Hodgkin lymphoma. This should, in this regard, exonerate the traces of residues of synthetic pesticides that can potentially be found in products from conventional agriculture.

A large French prospective study, NutriNet-Santé (and its BioNutriNet component), covers cohorts of 60,000 to 70,000 adult volunteers, 78% of whom are women, that are followed since 2009 and periodically respond to a questionnaire on their diet. This study first observed that the frequency of overweight, obesity and metabolic syndrome was lower in heavy organic consumers

(Kesse-Guyot et al., 2017). However, as the authors acknowledged, these are associations and not causal links. Many causes of bias and confounding factors, partially taken into account in the interpretation of the results, are involved in this study: on the one hand, the proportion of organic foods in the diet, evaluated by questionnaire, is imprecise and clearly overestimated and, on the other hand, it is well known that organic food eaters are more attentive to the balance of their diet and, more generally, to their lifestyle (tobacco, alcohol, physical exercise…).

The most publicized study (Baudry et al., 2018) concerned the association between organic food and cancer and its conclusions were translated in the medias by headlines such as "eating organic reduces cancer risk by 25%". This study was the subject of many critical reactions because, despite the precautions objectively taken by the authors of the article, this purported beneficial effect made public opinion. The study compared the 20% of the cohort who never consumed organic with the 20% for whom more than half of the food was organic. In fact, this association has only been observed in women and it is not for all types of cancers but only for postmenopausal breast cancer and some lymphomas. According to the authors, one explanation could be the lower ingestion of synthetic pesticide residues by organic consumers. The very low contribution of pesticide residues in conventional fruits and vegetables, not quantified in the study, certainly does not support such a hypothesis, especially since, when the statistical interpretation only takes into account the consumption of plant products, the only ones that may contain pesticide residues, the observed association is no longer significant for breast cancer. The statistical interpretation of data from NutriNet-Santé has also been severely criticized and therefore the conclusions of the studies strongly contested. In addition to the flaws in the statistical methodology, the poor evaluation of the share of organic in the diet does not allow any associations observed to be attributed to the differences between organic and conventional foods but to the behavior of the "organic eater", more vigilant about his hygiene of life and eating more balanced diets, including less meat and more fruits and vegetables. The noted associations concern, therefore, the effect of the lifestyle and eating behavior of organic eaters and not the favorable effect of organic foods alone. These would then be possible beneficial effects on the health of a flexitarian diet (less meat and more vegetable foods). This is objectively and clearly admitted by one of the co-authors of the article: "The main lesson is that going organic without reducing the share of meat in our consumption is useless, neither for health, nor for the environment".

Organoleptic quality

Organoleptic characteristics of organic foods have not been the subject of many comparative studies. Those that have been published do not highlight superior sensory qualities due to the production method, as results are variable and contradictory. For fruit and vegetables, sensory properties are mainly determined by the variety, ripeness and freshness. Organic farming often uses more hardy cultivars and, if production is local, the harvest can be later. For meat, taste properties depend mainly on breed, age, and degree of fatness. The mode of production, organic or conventional, is not a criterion of taste. For milk and eggs, no reproducible difference was obtained in the literature. In line with the nutritional values and health attributes, regulation of organic agriculture does not imply any better organoleptic quality of its products, but only an obligation of means of production.

Conclusions and prospects

Consumption of organic foods cannot have a nutritional effect in the overall diet. Even if we admit a slight superiority of organic foods for some nutrients in some foods, the difference would be insignificant in the global regime. For example, measures taken in France for "one organic meal per week in collective catering" cannot have significant influence on the quality of diet throughout the week. The health impact is also negligible because the effect of organic vegetables or fruits on blood antioxidant status could not be demonstrated, nitrate in vegetables are safe and chemical residues of conventional foods, including synthetic pesticides, are almost entirely below the allowed maximum limit and do not pose a risk to health. The few clinical or observational epidemiological studies have failed to attribute any beneficial effect to organic food on the incidence of various diseases, including cancer. Better nutritional and health values should therefore no longer be claimed for choosing to consume organic foods. The argument of preventing climate change or preserving biodiversity is also highly debatable. Indeed, the lower yields of organic farming (20–50%) lead to devoting more areas to the same production, at the expense of forests, fallows, wetlands and permanent meadows, and therefore biodiversity and carbon storage.

In most Western countries, organic foods account for less than 5% of foods consumed. Despite a strong trend of increasing demand, this share will be limited if a larger supply does not bring down the selling prices of organic foods. Significant imports from distant areas are not desirable because they are incompatible with the ecological spirit of organic agriculture. Organic production in industrialized and emerging countries will also be limited by local availability of organic fertilizers and especially by heavy yield losses (e.g., 50% for wheat in France). In developing countries, particularly in several parts of sub-Saharan Africa, food production could be increased through a wide application of the principles of organic agriculture for food crops associated with animal husbandry. However, this production will be self-consumed and may not be sufficient without recourse to a minimum of mineral fertilizers and plant protection products to reduce the very high crop losses. The increasing demand for organic foods, currently hampered by higher purchase prices, could explode if a large increase in local production or imports led to falling prices. However, this would be a vicious circle as farmers' income could drop and then discourage their conversion. In addition, throughout the world, intensive conventional farming will be subject to environmental constraints worldwide, including the use of pesticides.

Reductions of 30–50% of the amounts used presently are already planned in some countries. Under these conditions, the difference in food quality and ecological effects perceived by consumers will decrease and will not be market-friendly with organic foods.

Organic farming cannot be sustainable, especially for cereals, without the proximity of animal husbandry, including intensive breeding, to provide the essential organic fertilizers. An objective of the European "Green Deal project" to reach 25% of land in organic farming by 2030 seems paradoxical and unrealistic, while productivity is much lower and that, according to the FAO, it would be necessary to increase food production by 50–70% to feed the future world population.

Between the two extreme forms of agriculture, intensive conventional and organic farming, there are several intermediate modes of production that provide good quality food and preserve the environment, without sacrificing the high productivity required to maintain the food sovereignty of developed countries and to ensure, or even improve, global food security.

References

Afssa (Anses since 2010 French Food Safety Agency), 2003. Evaluation nutritionnelle et sanitaire des aliments issus de l'agriculture biologique. Afssa, Maisons-Alfort, France. Available at. http://www.afssa.fr/Documents/NUT-Ra-AgriBio.pdf.

Baranski, M., Srednicka-Tober, D., Volakakis, N., et al., 2014. Higher antioxidant and lower cadmium concentrations and lower incidence of pesticide residues in organically grown crops: a systematic literature review and meta- analyses. Br. J. Nutr. 112, 794–811.

Barré, A., Brulé, C., Borges, J.P., et al., 2009. Concentration des LTP dans la peau et la pulpe des fruits. Revue Française d'Allergologie 49, 166–169.

Baudry, J., Assmann, K.E., Touvier, M., et al., 2018. Association of frequency of organic food consumption with cancer risk in the NutriNet-Santé prospective cohort. JAMA Intern. Med. 178, 1597–1608.

Bradbury, K.E., Balkwill, A., Spencer, E.A., et al., 2014. Organic food consumption and the incidence of cancer in a large prospective study of women in the United Kingdom. Br. J. Cancer 110, 2321–2326.

Butler, G., Nielsen, J.H., Slots, T., et al., 2008. Fatty acid and fat-soluble antioxidant concentrations in milk from high- and low-input conventional and organic systems: seasonal variation. J. Sci. Food Agric. 88, 1431–1441.

Dangour, A.D., Dodhia, S.K., Hayter, A., et al., 2009. Nutritional quality of organic foods: a systematic review. Am. J. Clin. Nutr. 90, 680–685.

Dangour, A.D., Lock, K., Hayter, A., et al., 2010. Nutrition-related health effects of organic foods: a systematic review. Am. J. Clin. Nutr. 92, 203–210.

Dervilly-Pinel, G., Guérin, T., Minvielle, B., et al., 2017. Micropollutants and chemical residues in organic and conventional meat. Food Chem. 232, 218–228.

European Food Safety Authority (EFSA), 2008. Nitrate in Vegetables-Scientific Opinion of the Panel on Contaminants in the Food Chain. Question number EFSA-Q-2006-071.

Guéguen, L., 2018. Aliments bio: le vrai et le faux. In: Regnault-Roger, C. (Ed.), « Idées reçues et agriculture. Parole à la science ». Presses des Mines, Paris chap.6, 121-121.

Guéguen, L., Pascal, G., 2010. Le point sur la valeur nutritionnelle et sanitaire des aliments issus de l'agriculture biologique. Cahiers de Nutrition et de Diététique 45, 130–143.

Hord, N.G., Tang, Y., Bryan, N., 2009. Food sources of nitrates and nitrites: the physiological context for potential health benefits. Am. J. Clin. Nutr. 90, 1–10.

Jensen, M.M., Jorgensen, H., Halekoh, U., Olesen, J.E., Lauridsen, C., 2012. Can agricultural cultivation methods influence the healthfulness of crops for foods? J. Agric. Food Chem. 60, 6383–6390.

Jensen, M.M., Jorgensen, H., Lauridsen, C., 2013. Comparison between conventional and organic agriculture in terms of nutritional quality of food. A critical review. CAB Rev. 8 (045), 1–13.

Katan, M.B., 2009. Nitrate in foods: harmful or healthy? Am. J. Clin. Nutr. 90, 1–2.

Kesse-Guyot, E., Baudry, J., Assmann, K.E., et al., 2017. Prospective association between consumption frequency of organic food and body weight change, risk of overweight or obesity: results from the NutriNet-Santé Study. Br. J. Nutr. 117, 325–334.

Leifert, C., Niggli, U., 2009. Quantifying the Effect of Organic and Low Input Productionmethods on Food Quality and Safety and Human Health. Final report of QLIF, subproject 2 "Effects of production methods". http://www.qlif.org.

Magkos, F., Arvaniti, F., Zampelas, A., 2006. Organic food: buying more safety or just peace of mind? A critical review of the literature. Crit. Rev. Food Sci. Nutr. 46, 23–56.

Mie, A., Andersen, H.R., Gunnarsson, S., et al., 2017. Human health implications of organic food and organic agriculture: a comprehensive review. Environ. Health 16 (1), 111. https://doi.org/10.1186/s12940-017-0315-4.

Murphy, P.A., Hendrich, S., Landgren, C., Bryant, C.M., 2006. Food mycotoxins: an update. J. Food Sci. 71, R51–R65.

Organic Trade Association and The Organic Center, 2008. New Evidence Confirms the Nutritional Superiority of Plant-Based Organic Foods. OTA, Washington, DC.

Smith-Spangler, C., Brandeau, M., Hunter, G.E., et al., 2012. Are organic foods safer or healthier than conventional alternatives? A systematic review. Ann. Intern. Med. 157, 348–366.

Soil Association, 2000. Organic Farming, Food Quality and Human Health: A Review of the Evidence. Soil Association, Bristol, UK.

Średnicka-Tober, D., Baranski, M., Seal, C.J. et al., 2016. Higher PUFA and n-3 PUFA, conjugated linoleic acid, α-tocopherol and iron, but lower iodine and selenium concentrations in organic milk: a systematic literature review and meta-and redundancy analyses. Br. J. Nutr. 115, 1043–1060.

STOA (Science and Technology Options Assessments), Scientific Foresight Unit, and European Parliament Research Service, 2016. Human Health Implications of Organic Food and Organic Agriculture. Report PE 581.922. http://www.europarl.europa.eu/thinktank/fr/document.html?reference=EPRS_STU(2016)581922.

Williams, C.M., 2002. Nutritional quality of organic food: shades of grey or shades of green. Proc. Nutr. Soc. 61, 19–24.

Williamson, C.S., 2007. Is organic food better for our health? Nutr. Bull. 32, 104–108.

Winter, C.K., Davis, S.F., 2006. Organic foods. J. Food Sci. 71, R117–R124.

Woëse, K., Lange, D., Boess, C., Boögl, K.W., 1997. A comparison of organically and conventionally grown foods—results of a review of the relevant literature. J. Sci. Food Agric. 74, 281–293.

Worthington, V., 2001. Nutritional quality of organic versus conventional fruits, vegetables, and grains. J. Alternative Compl. Med. 7, 161–173.

Phytochemicals: Classification and occurrence

A. Cassidy and C. Kay, Norwich Medical School, University of East Anglia, Norwich, UK

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of A. Cassidy, C. Kay, Phytochemicals: Classification and Occurrence, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 39–46, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00226-9.

Introduction 530
Phytochemicals: General 531
Flavonoids 531
Flavonols 532
Flavones 532
Flavan-3-ols (Both Monomeric and Polymeric Forms) 533
Anthocyanins 533
Flavanones 533
Isoflavones 533
Current Estimates of Intake 534
Absorption and Metabolism of Flavonoids 534
Bioavailability of Flavonoids 535
Potential Mechanisms of Action 536
Potential Health Effects 536
Cardiovascular Health 536
Neuroprotective Effects 537
Cancer 537
Safety 537
Conclusions 537
Further Reading 537

Glossary

Aglycone A non-sugar compound remaining after replacement of a sugar by a hydrogen atom.
Conjugate Two or more compounds bound together.
Enzyme A protein with contains an active site which binds to a substrate to help catalyse a reaction.
Glucosidase An enzyme that cleaves the bond between glucose and another compound.
Glycoside Organic compounds formed when a monosaccharide is bound to another compound.

Abbreviations

CVD	Cardiovascular disease
eNOS	Endothelial nitric oxide synthase
NADPH oxidase	Nicotinamide adenine phosphate-oxidase
NF-κB	Nuclear factor kappa B
NO	Nitric oxide
NRF2	Nuclear factor (erythroid-derived 2)-like 2
SOD	Superoxide dismutase

Introduction

There is considerable evidence to suggest that populations that consume diets rich in fruits and vegetables, whole-grain cereals, and complex carbohydrates have a reduced risk of a range of chronic diseases. This has led to the suggestion that the diversity of substances found in food, particularly plant-derived foods, may underlie the protective effects that are attributed to diets high in

fruits and vegetables. Although fruits and vegetables are rich sources of micronutrients and dietary fiber, they also contain a wide variety of secondary metabolites, which provide the plant with color, flavor, UV protection, and antimicrobial and insecticide properties. Many of these substances have been attributed a wide array of biological properties but have yet to be recognized as nutrients in the conventional sense. These potentially protective plant compounds, termed phytochemicals, are receiving increasing attention and some have been termed phytonutrients, as they have been shown to exert numerous physiological functions in mammalian systems. However phytonutrients are not currently defined as 'nutrients' as they are not essential for human/animal growth and development but may help maintain health throughout life, including the prevention of chronic disease. Many of them are ubiquitous throughout plants and as a result are present in our daily diet in varying amounts. Among the most important classes are the flavonoids, which are classified based on their chemical and structural characteristics. The present article focuses on the different classes of phytochemicals and their potential relationships to human health.

Phytochemicals: General

Phytochemicals comprise a wide group of structurally diverse plant compounds, which are predominantly associated with the cell wall and widely dispersed throughout the plant kingdom. They are secondary metabolites in plants, characterized by having at least one aromatic ring with one or more hydroxyl groups attached. The nature and distribution of these compounds can vary depending on the plant tissue in which they are located, but they are mainly synthesized from carbohydrates via the shikimate and phenylpropanoid pathways. They range in chemical complexity from simple phenolic acids, such as caffeic acid, to complex high-molecular-weight compounds, such as the procyanidins/tannins, and they can be classified according to the number and arrangement of their carbon atoms. In plants, they are commonly found conjugated to sugars and organic acids and can be classified into two groups, flavonoids and nonflavonoids. Although there are many subclasses of bioactive phytochemicals, including flavonoids, stilbenes, lignans, phytosterols, sulfides, glucosinolates, and carotenoids (**Figure 1**), this article focuses mainly on the flavonoid subgroup, one of the most researched classes of phytochemicals to date; however, a brief introduction to the other classes is provided below:

- Stilbenes are present in a wide range of plants and one of the most studied stilbenes, *trans*-resveratrol, found in grapes and wine, is known for its effects on vascular activity and longevity.
- Lignans are phenolic compounds present in high concentrations in linseed (flaxseed) and studies suggest they have the potential to reduce the risk of cardiovascular disease (CVD) and cancer .
- Phytosterols are found in vegetable oils and are most notably known for their ability to decrease blood LDL-cholesterol levels and lower the risk of CVDs.
- Sulfides and glucosinolates are present in garlic and brassica vegetables and are most commonly investigated for their anticarcinogenic properties.
- Carotenoids are pigments (red, orange, yellow) common to many fruits and vegetables and their consumption has been linked to reduced risk of chronic degenerative diseases.

Flavonoids

Flavonoids constitute a large class of phytochemicals that are widely distributed in the plant kingdom, are present in high concentrations in the epidermis of leaves and skin of fruits, and have important and varied roles as secondary plant metabolites. More than 8000 varieties of flavonoids have been identified, many of which are responsible for the colors of fruits and flowers. They are found in fruits, vegetables, tea, wine, dark chocolate, grains, roots, stems, leaves, and flowers and are thus regularly consumed by humans. Although it has been widely known for centuries that chemical derivatives of plant origin possess a broad spectrum of biological activities, it was first suggested that flavonoids may be important for human health in the 1930s when it was observed that a fraction from lemon juice could decrease the permeability of arteries and partially prevent symptoms in scorbutic pigs. At the time, it was suggested that these compounds should be defined as a new class of vitamins, vitamin P, and the substance responsible for these effects was identified as the flavonoid rutin. However, the data were not generally accepted and the term vitamin P was abandoned in the 1950s. There was renewed interest in flavonoids when a potentially protective role of flavonoids in relation to heart disease in humans was reported. Since then, there has been a surge of interest in the potential role of flavonoids in human health, with research suggesting antioxidant effects, hormonal actions, anti-infectious actions, anti-inflammatory and cancer-preventative effects, the ability to induce chemical defense enzymes, and actions on blood clotting and the vascular system. However, concrete clinical evidence that they positively influence human health is lacking, and potential adverse effects have also been reported for a limited number of polyphenols.

Other flavonoid groups that are thought to be less important from a nutrition perspective are the dihydroflavones, flavan-3,4-diols, coumarins, chalcones, dihydrochalcones, and aurones. The basic flavonoid skeleton can have numerous constituents; hydroxyl groups are usually present at the 3-, 5-, and 7-positions. Sugars are very common, and the majority of flavonoids exist naturally as glycosides. The presence of both sugars and hydroxyl groups increases water solubility, but other constituents, such as methyl or isopentyl groups, increase flavonoids lipophilicity. Although many thousands of different flavonoids exist, they can be classified into a few different subclasses. The main subclasses that are important from a human health perspective are the flavones, flavonols, flavan-3-ols, flavanones, anthocyanins, and isoflavones (and their polymeric forms) (**Figure 2**). Extensive

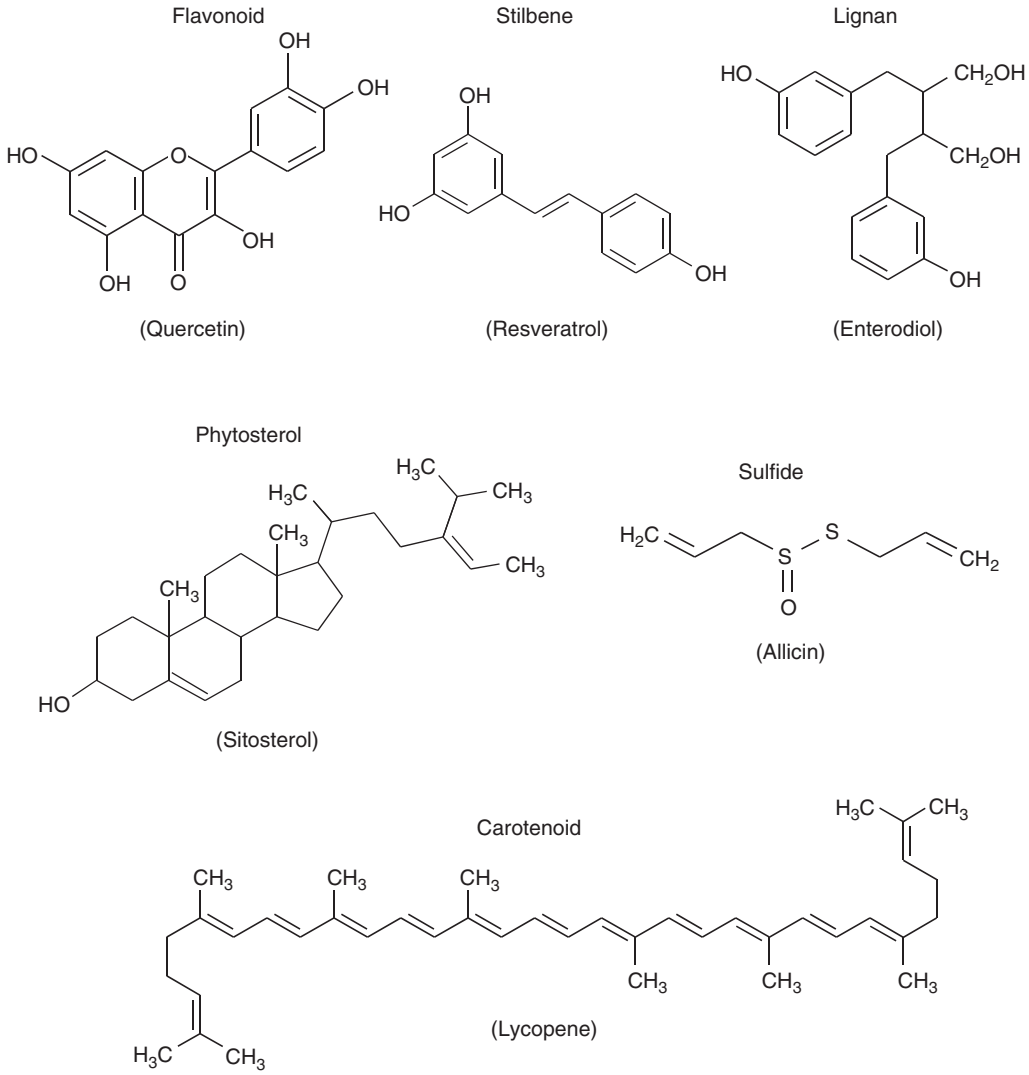


Figure 1 Chemical structure of flavonoid, stilbene, lignan, phytosterol, sulfide, and carotenoid.

information on the different flavonoids present in commonly consumed plant-based foods including fruits, vegetables, and drinks is available; however, there is wide variability in the levels present in specific foods, in part due to seasonal changes and varietal differences.

Flavonols

Flavonols are arguably the most widespread of the flavonoids because they are dispersed throughout the plant kingdom. The distribution and structural variations of flavonols are extensive and have been well documented. The most common flavonols are quercetin, kaempferol, myricetin, and isorhamnetin.

Flavones

Flavones have a close structural relationship to the flavonols, but unlike flavonols they are not widely distributed in plants. Their only significant dietary occurrences are in celery, parsley, and a few other herbs, and they predominantly occur as 7-*O*-glycosides (e.g., luteolin and apigenin). In addition, polymethoxylated flavones have been found in citrus fruits (e.g., nobiletin and tangeretin).

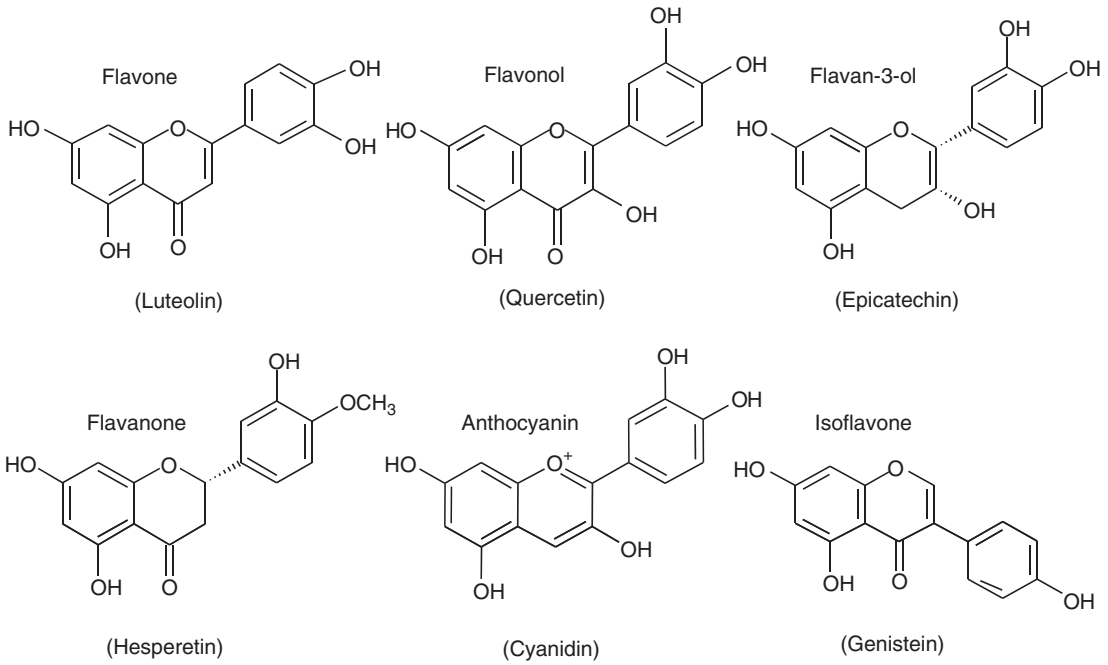


Figure 2 Structures of the major subclasses of flavonoids: flavones, flavonols, flavan-3-ols, flavanones, anthocyanins, and isoflavones.

Flavan-3-ols (Both Monomeric and Polymeric Forms)

Flavan-3-ols, often referred to as flavanols, are the most complex class of the flavonoids because they range in size from simple monomers (catechin and its isomer epicatechin) to the oligomeric and polymeric proanthocyanidins, which are also known as condensed tannins. Proanthocyanidins can occur as polymers of up to 50 units, and when hydroxylated they can form gallocatechins or undergo esterification to form gallic acid. Red wine contains oligometric proanthocyanidins derived mainly from the seeds of black grapes. Green tea is also a rich source of flavan-3-ols, principally epigallocatechin, epigallocatechin gallate, and epicatechin gallate. However, during fermentation of tea leaves the levels of catechins decline and thus the main components of black tea are high-molecular-weight thearubigins, whose structures are derived from flavonoids. The catechins are widespread, but the main sources in the diet come from processed plant-foods such as tea, wine, and chocolate.

Anthocyanins

Anthocyanins are widespread in nature, predominantly in fruits, leaves, and flower tissues, in which they are responsible for the red, blue, and purple colors. They are also found in stems, seeds, and root tissue. In plants, they protect against excessive light by shading leaf mesophyll cells and scavenging radicals. Additionally, they play an important role in attracting pollinating insects. The most common anthocyanins are cyanidin, pelargonidin, delphinidin, peonidin, petunidin, and malvidin, which are present in plants as sugar conjugates.

Flavanones

The flavanones are the first flavonoid products of the flavonoid biosynthetic pathway. They are characterized by the presence of a chiral center at C2 and the absence of the C2–C3 bond. The flavanone structure is highly reactive, and they have been reported to undergo hydroxylation, glycosylation, and *O*-methylation reactions. Flavanones are present in high levels in citrus fruits, with the most common glycoside known as hesperidin (hesperetin-7-*o*-rutinoside), which is present in citrus peel. Interestingly, flavanone rutinosides are tasteless, whereas the flavanone neohesperidoside conjugates (e.g., neohesperidin) from bitter orange and naringenin (naringenin-7-*o*-neohesperidoside) from grapefruit peel have an intensely bitter taste.

Isoflavones

Isoflavones are flavonoids, but they are also called phytoestrogens because of their ability to bind to estrogen receptors and exert weak oestrogenic activity. Apart from some basic structural similarities to mammalian estrogens, the key to their estrogenic effect is their unique structural orientation and the presence of the hydroxyl groups on the A and B rings. They are classified as both estrogen

agonists and antagonists because they compete with estrogen for their receptors. They have also been demonstrated to exert a wide number of biological effects that are independent of their estrogen receptor activity.

Current Estimates of Intake

Diets rich in plant-derived foods can provide more than 1 g of phenolic compounds per day (which includes phenolic acids, flavonoids, and their polymers), although there are major international and inter-individual differences in exposure. There are six main diet-derived flavonoid subclasses of present interest to human health; namely, flavonols, flavones, flavan-3-ols, flavanones, anthocyanins, and isoflavones, and their principal dietary sources are shown in (Table 1).

Given the differences in dietary intake, particularly for fruits and vegetables, between populations, it is not surprising that the relationships between the predominant flavonoids and their sources will vary between populations, nor is it unexpected that there will be wide inter- and intra-individual variations in intake of the individual flavonoid subclasses. It is only recently that comprehensive databases for estimating intakes of the diverse subclasses of flavonoids have become available. Previous estimates have been based on only a few subclasses and vary considerably among studies depending on which subclasses were included and which foods were considered in the assessment of intake. Two recent studies have estimated the range of intake of the different flavonoid subclasses in the USA and Spain (Table 2). In the USA, flavan-3-ols contributed the most to total flavonoid intake, although no data were available for estimating proanthocyanidin intakes in this study. In Spain, higher intakes of flavanones and anthocyanins were observed compared to the USA. The main sources of total flavonoid intake were apples, red wine, and other fruits in Spain, whereas tea and citrus fruits and juices were the main contributors in the USA. (Table 1).

Absorption and Metabolism of Flavonoids

Many flavonoids occur as glycosides in foods, and both flavonoid structure and the type of sugar moiety determine their primary site of absorption. Some flavonoid species are absorbed in the small intestine, where the sugar moiety becomes hydrolyzed by either lactase phloridzin hydrolase, which is located at the brush border membrane or via intracellular beta-glucosidase activity. The type of glycoside present has a significant impact on the site of flavonoid absorption; glucosides are primarily absorbed from the small intestine whereas rutinosides (a disaccharide) require hydrolysis by colonic bacteria before absorption. Methylation, glucuronidation, and sulfation are also common metabolic conjugates of flavonoid metabolism. In addition, the type of flavonoid structure also

Table 1 Principal dietary sources of flavonoids

Flavonoid	*Compound*	*Food source*
Flavonol	Quercetin, kempferol, myricetin	Onion, apple, broccoli, tea, olives, kale, cranberry, lettuce, beans (green, yellow)
Flavone	Luteolin, apigenin, tangeretin	Olives, celery, parsley, tangerines
Flavan-3-ol	Catechin, epicatechin, epigallocatechin	Tea, red wine, chocolate, apple
Flavanone	Naringenin, hesperidin	Citrus fruit
Anthocyanin	Cyanidin, delphinidin, malvidin, petunidin	Grapes, cherries, berries, blood orange juice
Isoflavone	Genistein, daidzein, glycitein	Soy and soy products

Source: Reproduced from Hollman PC and Katan MB (1997) Absorption, metabolism and health effects of dietary flavonoids in man. *Biomed Pharmacother* 51(8): 305–310, and Scalbert A and Williamson G (2000) Dietary intake and bioavailability of Polyphenols. *Journal of Nutrition* 130 (8S Supplement): 2073S–2085S, with permission from JCI.

Table 2 Estimated dietary intake of flavonoid subclasses in different countries

Flavonoid subgroup	*Estimated intake (mg day^{-1})*		
	USA	*Spain*	*Japan*
Flavonol	13	19	16
Flavone	2	3	<1
Flavan-3-ol	157	33	40
Flavanone	14	79	No data
Isoflavone	<1	<1	50
Anthocyanins	3	19	No data
Proanthocyanidins	No data	189	No data

impacts absorption as some forms of flavonoids, such as the isoflavones (glycosides), also require colonic hydrolysis of the sugar moiety before absorption. The absorption and metabolism of flavan-3-ols differs markedly depending on the chemical complexity of the species; for example, the location of absorption for monomeric species is primarily in the upper small intestine whereas absorption of higher-molecular-weight polymers requires prior bacterial metabolism in the large intestine before absorption of the metabolites. Absorption from the small intestine generally results in peak plasma concentrations within 1–3 h after ingestion, which is the case for most flavonoids, whereas absorption from the large intestine can take as long as 8–12 h before peak plasma concentrations of metabolites are observed.

Human and animal studies exploring the absorption and metabolism of flavonoids identify significant proportions of glucuronide, sulfate, and methylated metabolites in both the circulatory system and in tissues/organs such as the stomach, intestine, liver, brain, and eyes. Following absorption, flavonoids are readily metabolized in intestinal cells to form glucuronide and sulfate conjugates that appear in portal blood although additional conjugation (such as methylation) can occur in the liver. The conjugation of flavonoids in the small intestine and liver generally has a significant impact on polarity/water solubility and therefore rates of urinary excretion. Metabolism in other organs and tissues (such as the kidney) has also been reported. Flavonoids for the most part are processed in the body much the same as phenolic drugs, and their absorption and subsequent metabolism can also be affected by such factors as matrix, chemical composition, relative pH, age, gender, and host genetics.

Even though some species of flavonoids are absorbed intact (i.e., primarily nonglycosylated forms), substantial amounts of lower-molecular-weight products of flavonoid degradation (spontaneous) or microbial catabolism (via colonic micro flora) are also absorbed after initial biotransformation by the colonic microflora. The colon contains numerous microorganisms, and as a result has significant capacity for catalytic and hydrolytic reactions. These colonic bacteria produce enzymes that are capable of stripping flavonoid conjugates of their sugar moieties, thus enabling free aglycones to be absorbed. The enzymes produced by colonic bacteria can also break down the flavonoids into simpler compounds, resulting in the production of a range of derivatives. The main identified products of colonic metabolism are benzoic acids, phenylacetic acids, and phenylpropionic acids. In addition various lower-molecular-weight products of ring fission can occur where all of these above listed products may be subsequently reabsorbed from the colon and enter the systemic circulation or be eliminated in feces. These products of colonic metabolism are a very active area of current flavonoid research and some of these products may be common across many species of flavonoid and in the future may prove to be useful biomarkers of flavonoid intake. The bacterial transformation of flavonoids is also an important area for future research because these metabolic reactions may result in deactivation of bioactive compounds or activation of previously inactive compounds.

Bioavailability of Flavonoids

Bioavailability in a nutritional context is a term used broadly to include a full range of digestive and metabolic factors that influence the amount and type of compound that reaches the systemic circulation. Bioavailability or pharmacokinetics of flavonoids is based on data from absorption, distribution, metabolism, and excretion (ADME) studies conducted both in humans and animals. Available data suggest that the most abundant flavonoid compounds may not necessarily lead to the highest concentrations of biologically active metabolites in target tissues nor be the most biologically active in relation to specific health outcomes. Some subclasses are rather well absorbed; for example, following ingestion, isoflavones, the flavan-3-ol epicatechin, and the flavanones can reach micromolar concentrations in plasma. On the other hand, even large oral doses of anthocyanins result in only nanomolar plasma concentrations. Absorption rates vary significantly, with anthocyanins reaching peak concentrations within 1–2 h following ingestion compared to 6–8 h for isoflavones (depending on food matrix, compositional effects, and site of intestinal absorption). Relative urinary excretion ranges from 0.3% to 43% of the ingested dose depending on the subclass; isoflavone urinary yields are high, followed by some of the flavan-3-ols, flavanones, and quercetin glucosides (flavonols) but urinary excretion rates for procyanidins, galloylated tea catechins, and anthocyanins are low. However, it is worth noting that the emphasis of previous bioavailability studies has been on the intact parent flavonoids, and for some flavonoid subclasses it is possible that currently unidentified metabolites (arising from spontaneous breakdown or microbial catabolism) may be bioavailable and potentially responsible for significant health effects. The key role of the gut microflora in metabolism of some flavonoid compounds may explain why some exert biological effects *in vivo* even though their apparent bioavailability is low.

To date the isoflavone class has been most widely studied, and from the available evidence it is clear that in healthy adults, isoflavones are absorbed rapidly and efficiently. Following the consumption of either pure compounds, isoflavone-rich extracts or foods/beverages rich in isoflavones, the parent compounds and their metabolites can be detected in plasma and urine of human volunteers. After ingestion, isoflavones are hydrolyzed by intestinal glucosidases, which release the aglycones, daidzein, genistein, and glycitein. These may be absorbed or further metabolized to many specific metabolites including equol and *p*-ethyl phenol. Numerous studies attest to the fact that following ingestion, soy isoflavones attain maximal plasma concentrations within 4–8 h, and are then eliminated from the body through the bile and kidneys with a mean terminal elimination $t_{1/2}$ (half-life) that is approximately 8 h on average. There is evidence from several studies that high concentrations of isoflavones can be found in tissues: breast tissue of premenopausal women and in prostate glands of men. Our knowledge of the bioavailability of other flavonoid subclasses is less well-studied to date.

As with pharmacological compounds, demonstrating efficacy and understanding potential risks of flavonoids requires knowledge of their bioavailability. Further knowledge of how factors, including genetic determinants, food matrix, chemical composition, and age affect the bioavailability of flavonoids is an important area for future research.

Potential Mechanisms of Action

The effect of flavonoids on biological processes has been extensively studied, but few investigations have attempted to determine the actual flavonoid metabolites responsible for the observed effects. Much of the *in vitro* data assumes that biological activity originates from the parent/precursor flavonoids ingested without taking into consideration the biotransformation that may occur following ingestion and metabolism. It is well established that following ingestion they are transformed into a range of structurally distinct metabolic conjugates or degradation products. The majority of *in vitro* research has also been carried out with single flavonoids and few studies have investigated the relative effects of single compounds versus mixtures of compounds or the effects of factors such as matrix on absorption, metabolism, and bioactivity.

When interpreting the present mechanistic data, it is also important to note that little attention has been paid to physiologically relevant concentrations of flavonoids in *in vitro* model systems. Thus, in some instances, biological effects have been shown at concentrations that are unachievable *in vivo* following 'normal' habitual dietary consumption; therefore, the biological relevance of these mechanisms to humans is questionable. Despite these issues, there is considerable evidence to suggest that flavonoids have beneficial biological activities.

Epidemiological and experimental evidence supports the contribution of many dietary flavonoids to improving cardiovascular health, reducing cancer, and neurodegenerative disease risk (refer to bioactivity section below), with specific indications for improvements in vascular blood flow, hypertension, and cell cycle progression. As the pathophysiological processes leading to the development of CVD and cancer are so complex, there are numerous potential mechanisms by which bioactive plant compounds present in food could act and elucidating these underlying mechanisms is a key aim for nutrition research. Given that lipid peroxidation and oxygen free radicals are thought to be involved in conditions such as atherosclerosis, cancer, neurodegenerative diseases, and various inflammatory conditions, and that the flavonoid hydroxylated benzoid ring structure lends itself to radical scavenging, the primary health focus for flavonoids was traditionally thought to be through their direct antioxidant properties. Indeed *in vitro* studies have shown that flavonoids are efficient scavengers of free radicals; however, based on the *in vivo* bioavailability of flavonoids and their blood concentrations relative to other endogenous antioxidants, their mechanisms of action are unlikely to be the result of global radical scavenging. Therefore more recent research has focused on exploring other mechanistic actions of flavonoids using more physiologically relevant concentrations. Some recent examples of these mechanistic activities include: inhibition of cyclooxygenase, which in turn reduces platelet aggregation and thrombosis; inhibition of nicotinamide adenine dinucleotide phosphate-oxidase (NADPH oxidase) and regulation of nitric oxide synthase, which are involved with inflammation and vascular function; immunomodulation, anti-inflammatory, direct effects on enzyme function and gene and protein expression, selective interactions with protein kinase and lipid kinase signalling cascades and transcription factors such as Nuclear factor (erythroid-derived 2)-like 2(NRF2) and nuclear factor kappa beta(NF-kB).

Potential Health Effects

There is substantial epidemiological evidence that populations that consume diets rich in plant foods have a reduced risk of CVD, various cancers, and other age-related conditions. Identification of the role of flavonoids in the primary mechanisms that may protect against cellular damage may yield clues to slowing aspects of the aging process and postpone age-related diseases.

Although historically, the biological effects of flavonoids was attributed to their antioxidant actions, through their ability to scavenge reactive oxygen and other radicals, recent evidence suggests that this classic hydrogen-donating antioxidant activity cannot account for the reported *in vivo* bioactivity of flavonoids, and recently attention has been focussed on their other biological effects, including anti-inflammatory effects and effects on cell signalling pathways.

Cardiovascular Health

Intervention and experimental studies have demonstrated roles of some flavonoids in pathways involved in CVD initiation and progression, including improvement in vascular function, hypocholesterolemic effects, reduction of foam cell formation, thrombosis and inflammation, and protection against ischemia-reperfusion injury and arrhythmia. Differential effects of different subclasses have been observed, and some subclasses have been shown to exert beneficial effects on blood pressure by increasing endothelial derived nitric oxide (NO), either via modulation of endothelial nitric oxide synthase (eNOS) activity/expression, changes in eNOS substrate availability or through the prevention of radical induced NO conversion caused by enzymes such NADPH oxidase. Together these functions implicate the importance of specific flavonoids in a host of CVD-related conditions, including athersoclerosis, hypertension, congestive heart failure, cardiac hypertrophy, ischemic heart disease, and others.

Results from a recent meta-analysis of randomized controlled trials on flavonoids and flavonoid-rich foods provide evidence that some subclasses of flavonoids are associated with a significant reduction in blood pressure. For example, short-term interventions (1–18 week duration) with cocoa flavan-3-ols significantly reduced systolic and diastolic blood pressure. However, to date, for a number of flavonoid subclasses including anthocyanins, there are very few published studies to systematically examine their potential effects on CVD risk biomarkers, whereas for others the levels of flavonoids administered in the interventions were well beyond the range typically consumed in the diet. Further long-term trials are required to substantiate the potential cardioprotective effects of different flavonoid subclasses.

Neuroprotective Effects

Many flavonoids can cross the blood brain barrier, and therefore some flavonoids (or their metabolites) are present in the brain and have the potential to exert neuroprotective/neuroinflammatory effects. Ongoing research is therefore examining their relative effects on cognitive function and disorders such as Parkinsons disease.

In experimental studies, administration of flavonoids or flavonoid-rich foods (e.g., berries) protects dopamine neurons from oxidative damage and apoptosis and inhibits formation of α-synuclein fibrils. Other potential mechanisms for the effects of flavonoids in the brain include interactions with neuronal signalling pathways that are critical in controlling neuronal survival and differentiation and in modulating activity/expression of several oxidative-related enzymes (e.g., eNOS and superoxide dismutase (SOD)), and regulation of mitochondrial function or neuroinflammation. In a review of animal studies, oral administration of blueberry or strawberry extract consistently showed favorable neuroprotective effects including increased dopamine release, alleviating oxidative stress or suppressing neuroinflammation. However, flavonoid subclasses differ in their ability to cross the blood brain barrier and these differences depend in part on the lipophilicity and polarity of the flavonoid compound. During absorption flavonoids are extensively metabolized, with chemical transformations resulting in *O*-methylation and glucuronidation during phase II metabolism, which may have a significant impact on flavonoid bioavailability to the brain. It is therefore possible that the less polar *O*-methylated metabolites, for example *O*-methylated epicatechin metabolites, which are formed in the small intestine and liver, may be more bioavailable to the brain than their parent aglycones.

Cancer

Some flavonoids have been shown to suppress proliferation and induce apoptosis in cancer cell lines; they have been shown to delete aberrant epigenetic marks, resulting in the re-expression of abnormally silenced genes, and they can inhibit angiogenesis in established tumors. In experimental animals, some flavonoids have been shown to inhibit cancer at various stages in the cancer process from initiation to metastasis. However the available evidence from epidemiological studies remains inconclusive. In a recent systematic review of green tea and cancer prevention, which included 23 cohort studies and 27 case-control studies, the data were equivocal. Results were contradictory, particularly for cancers of the digestive tract although there was some limited evidence for protection against lung cancer.

Safety

Although flavonoids may have potential health effects, the function of many of these compounds in the plant is to discourage attack by fungal parasites, herbivores, and pathogens. As a result, it is not surprising that many are toxic and mutagenic at high concentrations in cell culture systems, and excessive consumption by animals or humans may hypothetically cause adverse metabolic reactions. However, the concentrations used in cell culture experiments in general tend to far exceed the levels that are achievable *in vivo* following dietary consumption. In addition, all of these compounds have short half-lives and do not appear to accumulate in tissues thus suggesting low toxicity. For the majority of the identified phytochemicals, there are currently limited data on the 'safe level' of intake or optimal level of intake for health benefits, and it is critical that these margins be more clearly defined in future research, particularly given the growing number of flavonoid supplements in the market.

Conclusions

There is increasing evidence that flavonoids may be protective against a number of age-related disorders. Data suggest that diets high in flavonoids may not only reduce the risk of CVD and cancer but also, by protecting against cellular damage, may slow aspects of the aging process and improve quality of life by postponing age-related diseases. There is still much to be uncovered about their bioavailability, mode of action, and optimal doses or, indeed, the actual compounds responsible for their health effects (i.e., metabolites). Given that there are still significant gaps in our knowledge base of flavonoid bioactivity, there are currently no formal recommended dietary intakes for these phytochemicals, but on the basis of the available evidence, people should consume a wide variety of foods that incorporate the various phytochemicals to maximize disease prevention. Further research is required to define optimal doses for potential health effects and to define safe levels of intakes. Many of these compounds should be viewed as pharmacologically active compounds because although they occur naturally, they still require the same levels of proof of efficacy and safety in use as synthetic pharmaceutical agents or dietary supplements.

Further Reading

Cassidy, A, Kay, C., 2010. Phytochemicals. In: Lanham-New, SA, Macdonald, IA, Roche, H (Eds.), Nutrition & Metabolism, The Nutrition Society Textbook Series, 2nd edn. Wiley Blackwell, UK.

Chun, OK, Chung, SJ, Song, WO., 2007. Estimated dietary flavonoid intake and major food sources of U.S. adults. Journal of Nutrition 137 (5), 1244–1252.

Del Rio, D, Borges, G, Crozier, A., 2010. Berry flavonoids and phenolics: bioavailability and evidence of protective effects. British Journal of Nutrition 104 (Suppl 3), S67–S90.

Erdman JW Jr, Balentine D, Arab L, *et al*. (2007) Flavonoids and heart health. In: *Proceedings of the ILSI North America Flavonoids Workshop*, 31 May–1 June, 2005. Washington, DC. *Journal of Nutrition* 137(3 Suppl 1): 718S–737S.

Hooper, L, Kroon, PA, Rimm, EB, et al., 2008. Flavonoids, flavonoid-rich foods, and cardiovascular risk: a meta-analysis of randomized controlled trials. American Journal of Clinical Nutrition 88 (1), 38–50.

Manach, C, Williamson, G, Morand, C, Scalbert, A, Rémésy, C, 2005. Bioavailability and bioefficacy of polyphenols in humans. I. Review of 97 bioavailability studies. American Journal of Clinical Nutrition 81 (1 Suppl), 230S–242S.

Pérez-Jiménez, J, Hubert, J, Hooper, L, et al., 2010. Urinary metabolites as biomarkers of polyphenol intake in humans: A systematic review. American Journal of Clinical Nutrition 92 (4), 801–809.

Schewe, T, Steffen, Y, Sies, H., 2008. How do dietary flavanols improve vascular function? A position paper. Archives of Biochemistry and Biophysics 476 (2), 102–106.

Spencer, JP., 2009. Flavonoids and brain health: Multiple effects underpinned by common mechanisms. Genes & Nutrition 4 (4), 243–250.

Williamson, G, Clifford, MN, 2010. Colonic metabolites of berry polyphenols: The missing link to biological activity? British Journal of Nutrition 104 (Suppl 3), S48–S66.

Williamson, G, Manach, C., 2005. Bioavailability bioefficacy of polyphenols in humans. II. Review of 93 intervention studies. American Journal of Clinical Nutrition 81 (1 Suppl), 243S–255S.

Wright, B, Moraes, LA, Kemp, CF, et al., 2010. A structural basis for the inhibition of collagen-stimulated platelet function by quercetin and structurally related flavonoids. British Journal of Pharmacology 159 (6), 1312–1325.

Zamora-Ros, R, Andres-Lacueva, C, Lamuela-Raventós, RM, et al., 2010. Estimation of dietary sources and flavonoid intake in a Spanish adult population (EPIC – Spain). Journal of the American Dietetic Association 110 (3), 390–398.

Phytochemicals: Health effects of proanthocyanidins and related compounds

Yi Lin, Ningping Zhan, and Dejian Huang, Department of Food Science and Technology, National University of Singapore, Queenstown, Singapore

© 2023 Elsevier Ltd. All rights reserved.

This is an update of H. Wiseman, Phytochemicals: Health Effects, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 47–51, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00227-0.

Introduction 539
Dietary sources 539
Absorption and metabolites 541
B-type PAC, poly(catechin), for cardiovascular health 541
Antibacterial activity of A-type PAC 542
PAC, poly(gallocatechin), and mitigation of starch digestion 542
Conclusions 543
References 543

Key points

- Proanthocyanidins are important phytochemicals with diverse dietary sources.
- Proanthocyanidins have broad activity in health promotion by mitigating hyperglycemia, inflammation, and bacterial infection.
- The bioactivity of proanthocyanidins are highly dependent on their specific structural variations.
- Although poorly absorbed, oligomeric proanthocyanidins exert their bioactivity by interacting with the biological targets in the gastrointestinal tract and may act as prebiotics and modulate gut microbiota.

Introduction

As non-nutrient components of plants dietary phytochemicals have been hotly studied for their health promotion effects as active ingredients in nutraceuticals and functional foods. Among them, dietary organosulfides, terpenoids (including saponins), resin glycosides, and polyphenolic compounds (particularly proanthocyanidins) found in fruits and vegetables have been found to function as radical scavengers, anti-inflammatory agents, and enzyme inhibitors (particularly digestive enzymes) thereby mitigating the chronic diseases such as cardiovascular disease, type-II diabetes, obesity, cancer, and cognitive disorder. In this chapter, the progress of the work related to polyphenolic compounds are accounted to cover the literature in the past decade.

Dietary sources

Proanthocyanidins (PAC) compose of dimeric and oligomeric flavan-3-ols. Furthermore, some proanthocyanidins also contain gallic acid esters. Depending on the type of interflavonyl bond linkage, there are A-type (two interflavonyl bonds) and B-type (one interflavonyl bond) proanthocyanidins (Le Bourvellec and Renard, 2019). Proanthocyanidins are found mostly in fruits such as berries (e.g., blueberry and cranberry), durian, mangosteen, persimmons, cocoa, coconut husk, apple, and chiku (Cádiz-Gurrea et al., 2017; Rauf et al., 2019; Wang et al., 2012). PAC concentration is high in the skin and in the seeds of the fruits. This is consistent with the biological role of PAC for plant protection. In contrast, PAC are much less common in vegetables with okra seeds being an exception (Lu et al., 2016). In addition, PAC are also found in cereals and pseudocereals including barley, sorghum, millet, amaranth, and buckwheat (Zhu, 2019). Some culinary spices and herbs also are known for PAC and the most commonly known ones are cinnamon barks and hops (Gunawardena et al., 2014; Li and Deinzer, 2006). The basic building blocks for PAC are (epi)afzelechin, (epi)catechin, and (epi)gallocatechin, PAC found in different plants have structural variations on the type of flavan-3-ols, interflavonyl linkages, and degree of polymerizations. **Fig. 1** highlighted some food sources of PAC. The intake of PAC in our daily diets is not known and vary greatly based on the individual dietary habits. The estimated mean daily proanthocyanidin intake in the United States is 53.6 mg/person/day excluding monomers and 57.7 mg/person/day including monomers. Proanthocyanidin intake also varies with age and gender (Aron and Kennedy, 2008).

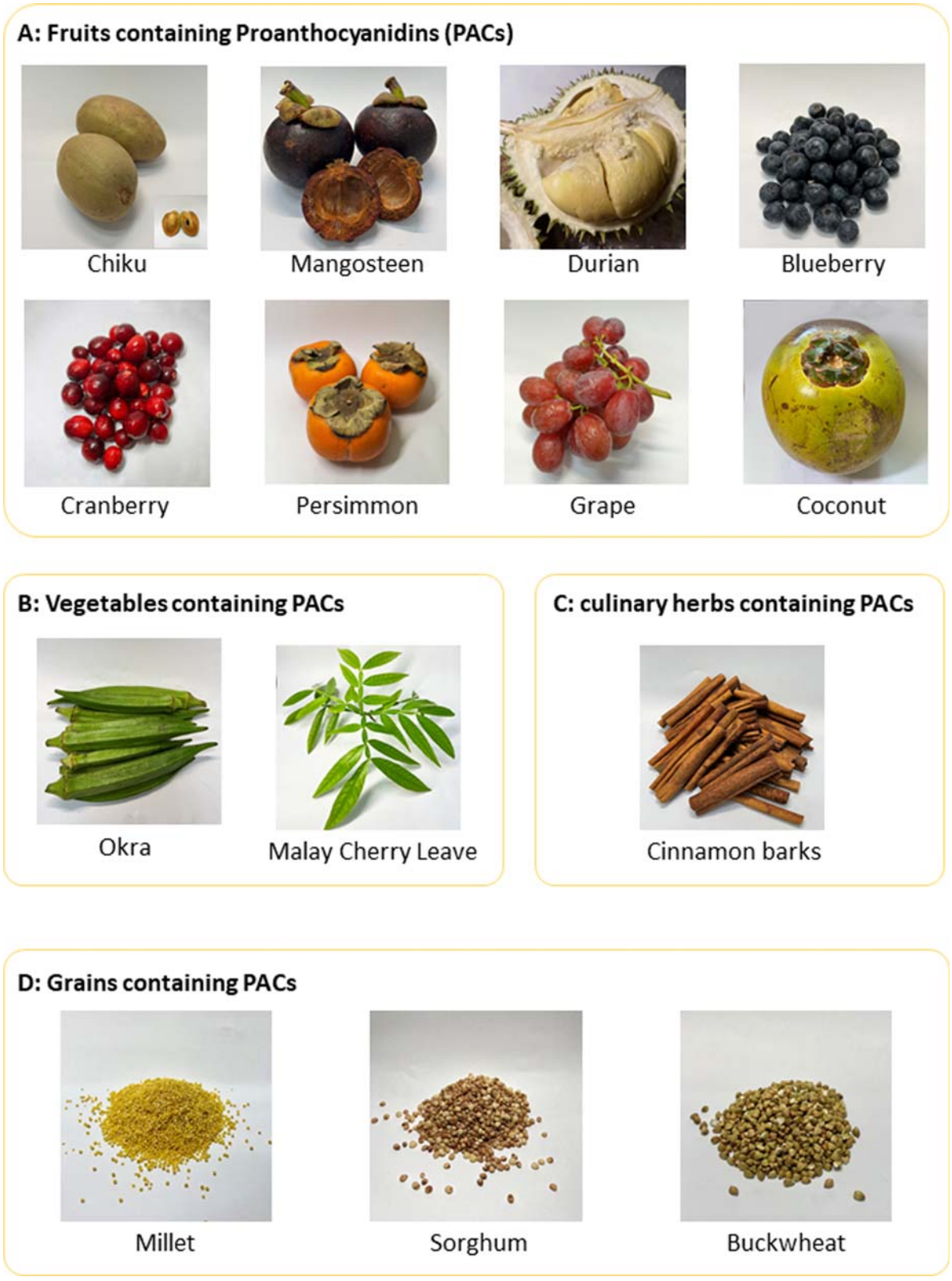
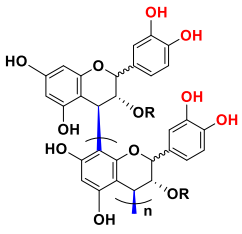


Fig. 1 Selected dietary sources of proanthocyanidins. Reclaiming PAC from agricultural by-products is an attractive topic for sustainable development of functional foods with PAC as active ingredients.

Absorption and metabolites

The monomers of PAC have high bioavailability. For example, (+)-catechin in plasma could reach 76.7 nM in humans after consumption of in 121 μmol catechin in dealcoholized red wine (Bell et al., 2000). Intake of 137 mg of (−)-epicatechin in PAC-rich chocolate could lead to a peak plasma (−)-epicatechin concentration of 260 nmol/L (Rein et al., 2000). However, flavan-3-ol dimers can only be absorbed ten times less. PAC with degree of polymerization equal or larger than four cannot be absorbed due to their high molecular weights and sizes and selectivity of the gut barrier (Vazquez-Flores et al., 2018). *In vitro*, PAC are subjected to acid mediated depolymerization into monomers in the presence of nucleophiles (such as thiols). However, under the acidic gastric juice media, little depolymerization was observed. The PAC are known to chelate metal ions such as ferrous ion, zinc, and calcium ion and reduce their bioavailability. Therefore, PAC are considered as antinutrients. In addition, binding to proteins and thus it can act as inhibitors of digestive enzymes such as pancreatic alpha-amylase and reduce the digestion rates of starch (Chen et al., 2022). This feature would be taken advantage of for development of starchy foods with low glycemic index. The monomers and oligomers (degree of polymerization <3) of flavan-3-ols could passively diffuse into the enterocytes are metabolized to their respective conjugates with sulfate, gluconates, and O-methylation on B-ring, PAC are not broken down in the small intestine and would reach the colon intact (Márquez Campos et al., 2020). On the other hand, the microbe in the gut were found to degrade extensively the PAC into fragments formed from the ring cleavage reactions (Márquez Campos et al., 2020). B-Type PAC underwent series of reactions including C-ring cleavage, A-ring oxidation and interflavan bond breakage followed by decarboxylation, dihydroxylation and lactonization to yield 5(3′4′-dihydroxy-phenyl)-γ-valerolactone (for procyanidins) in the case of grape seed extracts (fermented with human fecal microbiota) containing catechins and oligomeric PAC (Sánchez-Patán et al., 2011). Apparently the A ring of the flavan-3-ols were fragmented through unknown mechanisms in order to give such product (Ou and Gu, 2014). Procyanidin A1 [epicatechin-(4β→8, 2β→O→7)-epicatechin] and A2 were found to be absorbed in the small intestine of rats with absorption rates higher than B-type dimer (Appeldoorn et al., 2009). A2 dimer of PAC, found in cranberry, was detected in human plasma and urine (Zampariello et al., 2012). A study using ^{14}C labeled procyanidin B2 showed that after an oral dose, 58% ingested procyanidins were detected in 24 h urine and 40% were found in feces, suggesting high bioavailability of B2 dimer (Stoupi et al., 2010). However, besides B-type procyanidins, there are no study on the gut microbiota metabolites of A-type PAC and other oligomeric PAC such as polygallocatechins which are found in chiku and lady's finger seeds. As PAC may exert their bioactivity both through absorbed metabolites into human circulation systems, or act by modulation of gut microbiota. Full characterization are needed to understand the fates of ingested dietary PAC.

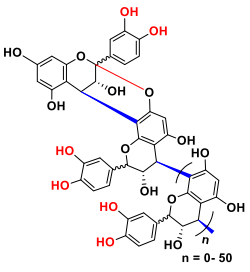
B-type PAC, poly(catechin), for cardiovascular health



n = 1- 50
B-Type Proanthocyanidins

PAC are credited for promoting cardiovascular health benefits of intake of food rich in PAC (Kruger et al., 2014). These include food and beverage (e.g., tea, wine, and dark chocolate) and dietary supplements (e.g., grape seed extracts and pine bark extracts) (Kruger et al., 2014). PAC are claimed to be the active ingredient in pine bark extracts that have been sold as pycnogenol (pine bark extracts) or grape seed OPCs in the marketplace for long time (Weber et al., 2007). However, despite good number of clinical studies, the recent systematic review on pine bark extract concluded that it the evidence of very weak for pine bark extract in lowering blood pressure, increasing HDL cholesterol, or decreases LDL cholesterol (Robertson et al., 2020). Grape seed PAC are slightly different from that of pine bark extracts because it contains significant amount galloyl esters through the C3-OH groups of the (epi)catechin units (Unusan, 2020), and therefore one would expect that they have similar bioactivity. Another B-type PAC is from cocoa, which was believed to be responsible for the extraordinary heart health of Kuna Indian whose diet included daily drinking five cups of high flavanols cocoa with flavonoid intake of 900 mg per day. A small but thoughtfully designed clinical study suggested that the epicatechin monomer was responsible for enhancing endothelial function measured by flow mediated dilation and decreased blood pressure of healthy young individuals (Rodriguez-Mateos et al., 2018). The Cocoa PAC with monomeric flavan-3-ol removed did not show such activity but was able to reduce total cholesterol level. Epicatechin metabolites, not the gamma-valerolactones from oligomeric PAC were suggested to be responsible for the health benefits of cocoa PAC. The European Food Safety Committee approved the health claims on cocoa flavanols for maintenance of normal endothelium-dependent vasodilation, which contributes to normal blood flow at daily dosage of 200 mg flavanols per day (Efsa Panel on Dietetic Products and Allergies, 2012).

Antibacterial activity of A-type PAC

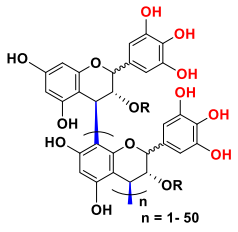


A-Type Proanthocyanidins

Traditionally, cranberry has been used for relieving urinary tract infection, which is a common problem in females. However, the components that are responsible for such activity have been under debate because the complex phytochemicals found in cranberry. Scientific evidences have suggested that A-type proanthocyanidins uniquely found in cranberry may be responsible for anti-urinary tract infection activity (Howell et al., 2010). There is only trace amount of PAC found in common cranberry juice and thus not possible to achieve effective dosage by drinking cranberry juice or eating cranberry fruits. High purity cranberry PAC (>18% based on DMAC assay) was suggested to be needed to achieve recommendated dosae of 36 mg PAC for preventing adhesion of bacteria on urinary tract. Studies found that the mannose found in cranberry might synergize with PAC to enhance the anti-adhesion activity. However, A-type PAC are poorly uptaken by Cacao-2 cells than "B-type" PAC (Ou et al., 2012). Therefore, it is unclear how mechanistically, the unabsorbed PAC could reduce the bacteria affinity in the urinary tract. Uropathogenic *E. coli* (UPEC) are the main cause of urinary tract infection. Evidences suggested that UPEC are opportunistic intracellular pathogens, rather than strictly extracellular organisms (Dhakal et al., 2008). Cranberry PAC may inhibit *E. coli* from adhering to uroepithelial cells or alter the pathogenic gut *E. coli* population and thus reduce the chance of infecting the urinary tract (Feliciano et al., 2015).

Recently, studies have suggested that cranberry PAC may be effective in prevention and management of periodontitis, an inflammatory disease of bacterial origin affecting tooth-supporting tissues. In such case, the bioavailability is not an issue (Nawrot-Hadzik et al., 2021; Andersen-Civil et al., 2021). Cranberry PAC may modulate gut microbiota via multiple interactions among them including facilitating the recruitment of immune cells and suppressing the amount of pro-inflammatory cytokines. Research work is needed to characterize the metabolites of the A-type PAC unique to cranberry by gut microbiota and investigate the antimicrobial activity of these metabolites.

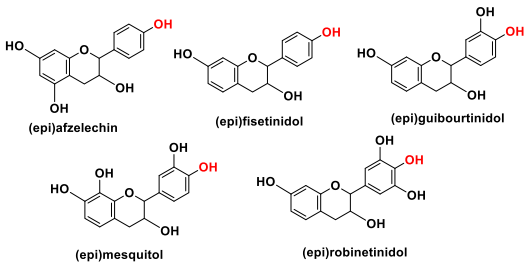
PAC, poly(gallocatechin), and mitigation of starch digestion



B-Type poly(gallocatechin)

PAC have been suggested to be active in modulating glucose homeostasis and the mechanisms could be due to absorbed monomer or dimers, and perhaps their metabolites, which may act *in vivo* as antioxidants and inhibitors of epithelial NADPH oxidase (Sun et al., 2020). PAC are also intensively studied and proven to be potent inhibitors of digestive enzymes, which could be the purpose of synthesizing PAC by plants for their protection against invaders including microbes and animals(Sun et al., 2020). The antinutrient activity of PAC could be utilized to mitigate starch digestibility of staple foods and reduce their glycemic index. To this end, PAC known for strong inhibitors of alpha-amylase include those isolated from persimmon, unripe *Manilkara zapota* (chiku), Malay cherry (*Lepisanthes alata*) leaves and fruits, and okra seeds (Lu et al., 2016; Wang et al., 2012; Zhang et al., 2016). The extracts from *L. alata* PAC is more potent than acarbose (based on equal weight), which is the most common antidiabetic drug working via inhibiting alpha-amylase in the gastrointestinal tract (Zhang et al., 2016). Acarbose is a better inhibitor against alpha-glucosidase in the small intestine surface, while PAC are better inhibitors against pancreatic alpha-amylase. This may offer advantage because intake of acarbose lead to side effects such as flatulence and diarrhea, which are due to accumulation of undegested fermentable oligo-saccharides. In theory, inhibition of alpha-amylase instead of alpha-glucosidase, would resulted in undigested starch, instead of fermentable sugar (dextrin, maltose, and maltotriose) that would stimulate gas formation by the microbiota in the gut. Among the various PAC, the ones show strong inhibitors of alpha-amylase share some common features i.e., they are B-type oligomers of gallocatechin with relatively high mean degrees of polymerization (Wang et al., 2012; Zhang et al., 2016). The PAC from Malay cherry leaves, chiku, and lady's finger are poly(gallocatechin) and this may explain their strong inhibition activity against alpha-amylase. Clinical study is thus needed to illustrate the effectiveness of the PAC for controlling of hyperglycemia.

Conclusions



Although monomeric dietary polyphenolic compounds have been recognized to play positive role in our health, PAC used to be treated as anti-nutrient factor in foods. Recent studies as we summarized herein suggested otherwise. Many biological targets for PAC are in our GI tract and this makes poor bioavailability of PAC not a concern. The structural diversity of PAC from different edible plants enable us to explore their property specifically for functional foods in controlling hyperglycemia, cardiovascular health, gut health, and anti-microbial activity. PAC are found abundantly in agri-food by products such fruit peels and seeds, grain shells (e.g., coconut shell), and leaves (e.g., Malay cherry leaves), exploiting PAC for human health promotion demands green and cost-effective extraction technology that can selectively separate PAC from the complex biomass matrix. At molecular level, flavanol monomers also include much less studied (epi)afzelechin, (epi)fisetinidol, (epi)guibourtinidol, (epi)mesquitol, and (epi)robinetinidol (Structure shown in the Figure). Their dimers and oligomers may have their unique bioactivity waiting for us to explore. We shall leave no stone unturned in the effort to find bioactive constituents for disease prevention through functional foods.

References

Andersen-Civil, A.I.S., Arora, P., Williams, A.R., 2021. Regulation of enteric infection and immunity by dietary proanthocyanidins. Front. Immunol. 12. https://doi.org/10.3389/fimmu.2021.637603.

Appeldoorn, M.M., Vincken, J.-P., Gruppen, H., Hollman, P.C.H., 2009. Procyanidin dimers A1, A2, and B2 are absorbed without conjugation or methylation from the small intestine of rats. J. Nutr. 139 (8), 1469–1473. https://doi.org/10.3945/jn.109.106765.

Aron, P.M., Kennedy, J.A., 2008. Flavan-3-ols: nature, occurrence and biological activity. Mol. Nutr. Food Res. 52 (1), 79–104. https://doi.org/10.1002/mnfr.200700137.

Bell, J.R.C., Donovan, J.L., Wong, R., Waterhouse, A.L., German, J.B., Walzem, R.L., Kasim-Karakas, S.E., 2000. (+)-Catechin in human plasma after ingestion of a single serving of reconstituted red wine. Am. J. Clin. Nutr. 71 (1), 103–108. https://doi.org/10.1093/ajcn/71.1.103.

Cádiz-Gurrea, M.D., Borrás-Linares, I., Lozano-Sánchez, J., Joven, J., Fernández-Arroyo, S., Segura-Carretero, A., 2017. Cocoa and grape seed byproducts as a source of antioxidant and anti-Inflammatory proanthocyanidins. Int. J. Mol. Sci. 18 (2). https://doi.org/10.3390/ijms18020376.

Chen, Y., Wang, J., Zou, L., Cao, H., Ni, X., Xiao, J., 2022. Dietary proanthocyanidins on gastrointestinal health and the interactions with gut microbiota. Crit. Rev. Food Sci. Nutr. 1–24. https://doi.org/10.1080/10408398.2022.2030296.

Dhakal, B.K., Kulesus, R.R., Mulvey, M.A., 2008. Mechanisms and consequences of bladder cell invasion by uropathogenic *Escherichia coli*. Eur. J. Clin. Invest. 38 (s2), 2–11. https://doi.org/10.1111/j.1365-2362.2008.01986.x.

Efsa Panel on Dietetic Products, Nutrition, & Allergies, 2012. Scientific opinion on the substantiation of a health claim related to cocoa flavanols and maintenance of normal endothelium-dependent vasodilation pursuant to Article 13(5) of Regulation (EC) No 1924/2006. EFSA J. 10 (7), 2809. https://doi.org/10.2903/j.efsa.2012.2809.

Feliciano, R.P., Krueger, C.G., Reed, J.D., 2015. Methods to determine effects of cranberry proanthocyanidins on extraintestinal infections: relevance for urinary tract health. Mol. Nutr. Food Res. 59 (7), 1292–1306. https://doi.org/10.1002/mnfr.201500108.

Gunawardena, D., Govindaraghavan, S., Münch, G., 2014. Chapter 30—anti-inflammatory properties of cinnamon polyphenols and their monomeric precursors. In: Watson, R.R., Preedy, V.R., Zibadi, S. (Eds.), Polyphenols in Human Health and Disease. Academic Press, San Diego, pp. 409–425.

Howell, A.B., Botto, H., Combescure, C., Blanc-Potard, A.-B., Gausa, L., Matsumoto, T., et al., 2010. Dosage effect on uropathogenic *Escherichia coli* anti-adhesion activity in urine following consumption of cranberry powder standardized for proanthocyanidin content: a multicentric randomized double blind study. BMC Infect. Dis. 10 (1), 94. https://doi.org/10.1186/1471-2334-10-94.

Kruger, M.J., Davies, N., Myburgh, K.H., Lecour, S., 2014. Proanthocyanidins, anthocyanins and cardiovascular diseases. Food Res. Int. 59, 41–52. https://doi.org/10.1016/j.foodres.2014.01.046.

Le Bourvellec, C., Renard, C.M.G.C., 2019. Interactions between polyphenols and macromolecules: effect of tannin structure. In: Melton, L., Shahidi, F., Varelis, P. (Eds.), Encyclopedia of Food Chemistry. Academic Press, Oxford, pp. 515–521.

Li, H.-J., Deinzer, M.L., 2006. Structural identification and distribution of proanthocyanidins in 13 different hops. J. Agric. Food Chem. 54 (11), 4048–4056. https://doi.org/10.1021/jf060395r.

Lu, Y., Demleitner, M.F., Song, L., Rychlik, M., Huang, D., 2016. Oligomeric proanthocyanidins are the active compounds in *Abelmoschus esculentus* Moench for its α-amylase and α-glucosidase inhibition activity. J. Funct. Foods 20, 463–471. https://doi.org/10.1016/j.jff.2015.10.037.

Márquez Campos, E., Jakobs, L., Simon, M.-C., 2020. Antidiabetic effects of flavan-3-ols and their microbial metabolites. Nutrients 12 (6). https://doi.org/10.3390/nu12061592.

Nawrot-Hadzik, I., Matkowski, A., Kubasiewicz-Ross, P., Hadzik, J., 2021. Proanthocyanidins and Flavan-3-ols in the prevention and treatment of Periodontitis—Immunomodulatory effects, animal and clinical studies. Nutrients 13 (1), 239. https://doi.org/10.3390/nu13010239.

Ou, K., Gu, L., 2014. Absorption and metabolism of proanthocyanidins. J. Funct. Foods 7, 43–53. https://doi.org/10.1016/j.jff.2013.08.004.

Ou, K., Percival, S.S., Zou, T., Khoo, C., Gu, L., 2012. Transport of cranberry A-type procyanidin dimers, trimers, and tetramers across monolayers of human intestinal epithelial Caco-2 cells. J. Agric. Food Chem. 60 (6), 1390–1396. https://doi.org/10.1021/jf2040912.

Rauf, A., Imran, M., Abu-Izneid, T., Iahtisham Ul, H., Patel, S., Pan, X., et al., 2019. Proanthocyanidins: a comprehensive review. Biomed. Pharmacother. 116, 108999. https://doi.org/10.1016/j.biopha.2019.108999.

Rein, D., Lotito, S., Holt, R.R., Keen, C.L., Schmitz, H.H., Fraga, C.G., 2000. Epicatechin in human plasma: *In vivo* determination and effect of chocolate consumption on plasma oxidation status. J. Nutr. 130 (8), 2109–2114. https://doi.org/10.1093/jn/130.8.2109S.

Robertson, N.U., Schoonees, A., Brand, A., Visser, J., 2020. Pine bark (*Pinus* spp.) extract for treating chronic disorders. Cochrane Database Syst. Rev. (9). https://doi.org/10.1002/14651858.CD008294.pub5.

Rodriguez-Mateos, A., Weber, T., Skene, S.S., Ottaviani, J.I., Crozier, A., Kelm, M., et al., 2018. Assessing the respective contributions of dietary flavanol monomers and procyanidins in mediating cardiovascular effects in humans: randomized, controlled, double-masked intervention trial. Am. J. Clin. Nutr. 108 (6), 1229–1237. https://doi.org/10.1093/ajcn/nqy229.

Sánchez-Patán, F., Monagas, M., Moreno-Arribas, M.V., Bartolomé, B., 2011. Determination of microbial phenolic acids in human faeces by UPLC-ESI-TQ MS. J. Agric. Food Chem. 59 (6), 2241–2247. https://doi.org/10.1021/jf104574z.

Stoupi, S., Williamson, G., Drynan, J.W., Barron, D., Clifford, M.N., 2010. A comparison of the *in vitro* biotransformation of (–)-epicatechin and procyanidin B2 by human faecal microbiota. Mol. Nutr. Food Res. 54 (6), 747–759. https://doi.org/10.1002/mnfr.200900123.

Sun, C., Zhao, C., Guven, E.C., Paoli, P., Simal-Gandara, J., Ramkumar, K.M., et al., 2020. Dietary polyphenols as antidiabetic agents: advances and opportunities. Food Front. 1 (1), 18–44. https://doi.org/10.1002/fft2.15.

Unusan, N., 2020. Proanthocyanidins in grape seeds: an updated review of their health benefits and potential uses in the food industry. J. Funct. Foods 67, 103861. https://doi.org/10.1016/j.jff.2020.103861.

Vazquez-Flores, A.A., Martinez-Gonzalez, A.I., Alvarez-Parrilla, E., Díaz-Sánchez, Á.G., de la Rosa, L.A., González-Aguilar, G.A., Aguilar, C.N., 2018. Proanthocyanidins with a low degree of polymerization are good inhibitors of digestive enzymes because of their ability to form specific interactions: a hypothesis. J. Food Sci. 83 (12), 2895–2902. https://doi.org/10.1111/1750-3841.14386.

Wang, H., Liu, T., Song, L., Huang, D., 2012. Profiles and α-amylase inhibition activity of proanthocyanidins in unripe *Manilkara zapota* (Chiku). J. Agric. Food Chem. 60 (12), 3098–3104. https://doi.org/10.1021/jf204715q.

Weber, H.A., Hodges, A.E., Guthrie, J.R., O'Brien, B.M., Robaugh, D., Clark, A.P., et al., 2007. Comparison of proanthocyanidins in commercial antioxidants: grape seed and pine bark extracts. J. Agric. Food Chem. 55 (1), 148–156. https://doi.org/10.1021/jf063150n.

Zampariello, C.A., McKay, D.L., Dolnikowski, G., Blumberg, J., Chen, C.Y.O., 2012. Determination of cranberry proanthocyanidin A2 in human plasma and urine using LC-MS/MS. FASEB J. 26 (S1). https://doi.org/10.1096/fasebj.26.1_supplement.124.8, 124.128-124.128.

Zhang, Y., Wong, A.I.C., Wu, J.e., Abdul Karim, N.B., Huang, D., 2016. *Lepisanthes alata* (Malay cherry) leaves are potent inhibitors of starch hydrolases due to proanthocyanidins with high degree of polymerization. J. Funct. Foods 25, 568–578. https://doi.org/10.1016/j.jff.2016.06.035.

Zhu, F., 2019. Proanthocyanidins in cereals and pseudocereals. Crit. Rev. Food Sci. Nutr. 59 (10), 1521–1533. https://doi.org/10.1080/10408398.2017.1418284.

Probiotics and prebiotics

Jose M. Saavedra, Johns Hopkins University School of Medicine, Baltimore, MD, United States

© 2023 Elsevier Ltd. All rights reserved.

Introduction 545
Probiotics 546
Definition 546
History 546
Probiotic microorganisms 547
Microbes in foods 547
Early identification of probiotic microbes 547
Commonly used probiotics 548
Mechanisms of action of probiotics 548
Health outcomes reported with probiotics 549
Safety of probiotics 551
Infectious risk concerns 551
Safety regulation 551
Regulation of probiotics 552
Prebiotics 553
Definition 553
History 553
Prebiotic substances 553
Fructo-oligosaccharides (FOS) 554
Galactooligosaccharides (GOS) 554
Human milk oligosaccharides (HMOS) 554
Prebiotic candidates 554
Saccharide prebiotic candidates 555
Non-saccharide prebiotic candidates 555
Health outcomes reported with prebiotics 555
Mechanisms of action of prebiotics 556
Safety 556
Regulation of prebiotics 556
Conclusion 557
References 557

Key points

- Prebiotics and probiotics are approaches to modulate the composition and function of the host's microbiota in ways that are beneficial to the host.
- Basic science, in-vitro, in-vivo, and observational studies provide clear and robust evidence of host-microbe interactions which support the rationale for probiotics and prebiotics to enhance health.
- Specific benefits with specific pre- and probiotics on particular health outcomes have been well documented.
- However, the strength and consistency of the clinical evidence in the field so far is extremely variable. Conclusions and guidance should be driven by specific outcomes with a specific probiotic candidate, in a particular population, and replicated beyond a single clinical trial.
- Pre- and probiotics in use today are safe for the general population, and specific benefits have been reported in some high-risk groups, including premature infants, immunocompromised individuals, and the elderly.
- Continued emphasis on rigorous research should help close the translational gaps that still exist between concept, science, and application.

Introduction

The human microbiota is a set of complex microbial ecosystems which develop and change over the life cycle. These microbes constantly interact with the host, driving and modulating a large set of physiologic responses which are major determinants of

host health. The microbiota of the respiratory tract, skin, vagina, and particularly that of the gastrointestinal (GI) tract play a major role in host immune development, protective and inflammatory responses, metabolic and gastrointestinal functions, as well as in neurocognitive function via the gut-brain axis.

An explosive increase in knowledge of the microbiota has been accompanied by a parallel interest in approaches to modify its composition and modulate its effects, via probiotic microbes and prebiotic substances, as a means of supporting health and preventing or managing disease processes. Such beneficial effects relate to infection, allergy, GI motility, inflammatory disorders, metabolic disorders (glucose control, dyslipidemia, obesity), and possibly mood and other central nervous system dysfunction. Other interventions that can modulate the microbiota and its interactions with the host include specific diets, fecal microbiota transplants, synbiotics, and postbiotics, which are not covered in this section.

This increased understanding of host microbial interactions comes from basic science, in-vivo and in-vitro experimentation, and the accelerated development of new methodologies, such as enhanced microbial genome sequencing techniques. Evidence from human clinical trials that specific pre- and probiotics can yield specific positive clinical outcomes has continued to mount. The need for improved rigor in developing clinical evidence, and the unclear and inconsistent regulatory frameworks for communicating health-related benefits, continue to be significant challenges to the broader application of pre-and probiotics for health and nutrition.

Probiotics

Definition

The most widely accepted definition of the term probiotics is: "Live microorganisms that, when administered in adequate amounts, confer a health benefit on the host." The first version of this definition was put forth in 2001, by consensus from an Expert Consultation of the Food and Agricultural Organization of the United Nations and the World Health Organization (FAO/WHO), and it was espoused in a statement from the International Scientific Association for Probiotics and Prebiotics (ISAPP) at an expert consensus in 2013. The definition implicitly excludes non-live or viable microorganisms and their metabolic products or derivatives, and microbes strictly used as food processing aids.

Most microorganisms described as potential probiotics in the scientific literature include multiple species of bacteria (mainly the genus *Lactobacillus* and *Bifidobacterium*) and yeasts (specifically the genus *Saccharomyces*). Probiotics are generally given orally to reduce the risk or manage specific diseases or related pathologic mechanisms. Today most available probiotic products are commercialized in oral supplements and foods containing probiotic microorganisms, although some products for topical -skin or vaginal-applications are available (FAO/WHO, 2001, Hill et al., 2014).

History

The fermentation of foods, including dairy products, grains, and other plants, using bacteria or yeast has been a technique of food preservation used for thousands of years. These foods have been consumed in all parts of the globe for their nutritional value and, in some instances, for their purported health benefits. However, the modern-day scientific link between the consumption of specific microorganisms and their potential role in health can be attributed to the Russian scientist Élie Metchnikoff. Based on laboratory and field observations while working at the Pasteur Institute, in 1907 he published his "Optimistic Studies" on the "Prolongation of Life". At the time, it was known that the activity of "putrefactive" microbes such as *Clostridium* produced "potentially toxic substances" including phenols, indols, and ammonia from protein digestion; and that milk fermented with lactic acid bacteria could inhibit the growth of these proteolytic bacteria due to the low pH resulting from lactose fermentation. Metchnikoff also noted that rural populations in Bulgaria and Russia, who consumed yogurt and fermented milks regularly, had longer lives. He postulated that phagocytes, which are important for health in earlier life, hasten cell death from vital organs in old age. And that the products of "putrefactive microbes" are responsible for "intestinal intoxication," which rendered cells particularly vulnerable to phagocytes. Moreover, he proposed that taking "skimmed milk which has been boiled and rapidly cooled, and on which pure cultures of the recently characterized *Bulgarian bacillus* (now *Lactobacillus bulgaricus*) have been sown", can inhibit these mechanisms of aging. He himself consumed milk soured with *L. bulgaricus* and believed it benefitted his health (Metchnikoff, 1907; Vaughan, 1965).

However, the term "probiotic" denoting "for life" was not used until decades later and traversed a circuitous path to its current definition. Until now, the first documented use of the term appears to be that proposed by W. Kollath in 1953 to refer to a broader set of "active substances that are essential for a healthy development of life," contrasting it to harmful substances such as "antibiotics." A more focused view was expressed by F. Vergin in 1954, referring to the detrimental effects of antibiotics and other antimicrobial substances on the gut microbial population, versus "probiotics," referring to factors favorable to gut microbial populations. H. Kolb and others formed an Association for Microbial Therapy. They advanced the concept in 1955, suggesting, "Antibiotic therapy causes flora damage. In such cases, we administer cultures of symbionts. In this way, deleterious effects of antibiotics are prevented by probiotic therapy" (Rusch, 2002; Hamilton-Miller et al., 2003).

In the '60s and '70s, the use of the term became more specific. Lilly and Stillwell (1965) used the term to describe protozoan microorganisms or factors which stimulated in vitro the growth of other protozoa. Fujii and Cook (1973) used the term for "compounds that build resistance to infection in the host but do not inhibit the growth of microorganisms in vitro," referring

to synthetic chemicals that protected mice against infection with *Staphylococcus aureus*. Parker (1974) used it to refer to "organisms and substances which contribute to intestinal microbial balance."

In 1989 Fuller proposed greater specificity to the use of the term as "a live microbial feed supplement which beneficially affects the host animal by improving its intestinal microbial balance", limiting it to only "live microorganisms" (not "substances") that benefit the host, and specifying an "improved microbial balance" as the attendant mechanism. Throughout the 90s several other research and consensus groups put forth definitions which included various criteria for a "probiotic" microbe, including restrictions on "origin" (i.e., human vs. animal), the capacity to "colonize" the intestinal lumen, demonstration or documentation of safety, and mechanism of action (e.g., "by promoting microbial balance," or "by an immune-modulating effect"). As the knowledge of the composition and the multiple effects and relationships between the host and the intestinal microbiota exploded in the last decades, such limiting definition criteria became less useful or relevant (Fuller, 1989, Hamilton-Miller et al., 2003).

Today's most widely accepted definition is that proposed by FAO/WHO in 2001–2002: "Live microorganisms that, when administered in adequate amounts, confer a health benefit on the host.", also adopted by the International Scientific Association for Probiotics and Prebiotics at an expert consensus in 2013. And although not explicit, the definition implies that such live microorganisms are well characterized taxonomically (including genus, species, and strain). This definition excludes non-viable microorganisms, their metabolites, or other substances, and also excludes previous criteria used to denote a probiotic, such as the original source from which it was isolated or identified (human or other), specific level of safety, or mechanism of action. In addition, the definition is expressly vague in terms of type or route of administration, type of microorganisms (phylum, species), and type of "health benefits." Nevertheless, it has been generally regarded as practical, "relevant, and sufficiently accommodating" for current and anticipated applications of probiotics (FAO/WHO, 2001, Hill et al., 2014).

Probiotic microorganisms

Microbes in foods

The current concept of probiotics needs to be differentiated from the presence of bacteria and yeast in foods primarily for the purposes of processing and preservation by fermentation. Used for millennia, bacterial and yeast fermentation allowed not only food preservation but the production of organoleptically and culturally desirable products. Species of *Lactobacillus, Streptococcus, Leuconostoc, Lactococcus, Weissella, Saccharomyces*, among others, have been used to ferment vegetables (sauerkraut, pickled vegetables, kimchi), legumes (tempeh, miso, soy sauce), cereals (bread, sourdough), fruits (wine, spirits), meats (sucuk and other sausages), tea (kombucha), and dairy (yogurt, kefir, buttermilk). It is estimated that more than 5000 varieties of fermented foods (and beverages) are currently produced and consumed globally. There is consensus today that probiotic microbes require appropriate identification (genus, species, and strain) and documentation of a specific beneficial effect on the host. In most of these traditionally or industrially made and commercialized products, live or viable microbes may or may not be present or viable at the end of processing or through shelf-life. In most cases, the amount or concentration of microbes will vary; in many cases the viability is unknown or changed by warming or cooking, the strain specificity will not be known or noted, and the benefit related to the bacterium of yeast itself may not be documented. So, while many of these food products may be nutritionally adequate or beneficial, it is accepted that for a "probiotic" effect, a benefit must be at least partially attributable to the microbes when consumed, and the benefit should extend beyond the nutritional benefit of the food matrix that contains it. Thus, the mere presence of microbes in traditional fermented foods would not qualify them as a "probiotic foods," and it would be preferable to describe some of these as a "food containing probiotics," assuming the specific bacterium or yeast in the food product fulfills the definition of a probiotic (Marco et al., 2021).

Early identification of probiotic microbes

Historically, the identification of specific microorganisms with potential health-related benefits started at the turn of the last century with the characterization of lactobacilli which produce lactic acid as a major metabolic end-product of carbohydrate fermentation—lactic acid bacteria (LAB). In 1900, Ernst Moro presented the first bacteriological characterization of *L. acidophilus*. Around that time also, S. Grigorov described *Lactobacillus bulgaricus* (today known as *Lactobacillus delbrueckii* subsp. *bulgaricus*), which was central in the work of E. Metchnikoff proposing beneficial health-related effects by specific LAB. Improved microbiologic methods gradually enhanced the characterization and refined the taxonomic designation of multiple species. And better-defined species and strains of lactobacilli and bifidobacteria began to be studied for potential health benefits.

In the 1930s in Japan, Minoru Shirota isolated and described *Lactobacillus casei shirota* (later reclassified as *L. paracasei shirota*). He documented growth inhibition of intestinal pathogens and other potential probiotic benefits with this bacterium, leading to the first commercially marketed fermented dairy drink, Yakult, still sold worldwide today. Around 1900, working at the Pasteur Institute, H. Tissier described the genus *Bifidobacterium*. First isolated from a breastfed infant in 1899, he named it *Bacillus bifidus communis*. Tissier found that bifidobacteria are the dominant species in the gut of breastfed babies and postulated they had a protective effect on the intestinal health of breastfed infants.

Aside from multiple lactobacilli and bifidobacteria, one specific strain of *Escherichia* (*E. coli nissle 1917*) was isolated by A. Nissle from feces of a german soldier who was not affected during an outbreak of shigellosis in 1917. Relatedly, the French microbiologist H. Boulard, during a cholera epidemic in 1920 in Indochina, noticed that individuals drinking tea from the outer skins of lychee and mangosteens appeared to be protected. From it, he isolated a saccharomyces (now *S. cerevisiae boulardii*), which was patented in

1947 for its potential medicinal use. Strains of *E. coli nissle 1917* and *S. boulardii* continued to be used and today are commercialized as probiotics (McFarland, 2010; Zommiti et al., 2020; ISAPP, 2019).

A significant leap in the understanding of probiotics has benefitted from the explosive interest and progress in the characterization of the human microbiota. This has been possible due to improvements in genomic sequencing technology and decreased costs. The first bacterial genomes were completely sequenced in the mid-'90s, and the first *Bifidobacterium* in 2002. Thanks to next-generation sequencing techniques, more than 200,000 bacterial and archeal complete or draft genomes, including hundreds of lactobacilli and bifidobacteria, have been uploaded to public databases. This has allowed for better identification of strains, maintaining adequate control of cultured microbes for commercial purposes, and identifying genetic characteristics to select candidate probiotics (Zhang et al., 2020).

Commonly used probiotics

Probiotic microorganisms should always be identified by their genus, species, subspecies (if applicable), and an alphanumeric strain designation. The nomenclature used for strain designation follows authentication and preservation by repositories such as the ATCC (American Type Culture Collection—US); CNCM (National Collection of Microorganisms Cultures -France); or NCIMB (National Collection of Industrial and Marine Bacteria—UK). Examples: *Lactobacillus* (genus) *reuteri* (DSM 17,938) (strain) or *Bifidobacterium* (genus) *animalis* (species) subsp. *lactis* (subspecies) (Bb-12) (strain) or.

In the last few decades, the great majority of species and specific strains studied and documented as having health-related benefits belong to the genus *Lactobacillus* and *Bifidobacterium*. These include strains of *Lactobacillus*, *Bifidobacterium*, and *Saccharomyces*, including those below, which are among those better studied. Other lactobacilli and bifidobacteria have also been proposed as probiotics; and in recent years, less well-studied species from the genus *Roseburia*, *Akkermansia*, *Propionibacterium*, and *Faecalibacterium* are emerging as probiotic candidates.

Most probiotics today are commercialized as supplements in a powder form, containing freeze-dried or lyophilized bacteria, which allows for maintaining viability for extended periods of time. This is the case for most probiotics found in capsules, sachets, or powdered products, including infant formulas. However, heat and humidity will gradually reduce bacterial viability over time, so it is critical that the target effective bacterial counts (colony forming units or CFUs) in the product be maintained throughout shelf life, and that storage conditions be followed. More novel methods, such as suspending the bacteria in lipid emulsions, allow extended shelf life in a liquid format. Most probiotics added to liquids, including milks and dairy products such as yogurt, have a much shorter shelf life than dry products and require refrigeration (Hill et al., 2014; Zommiti et al., 2020; Sanders et al., 2018).

Most commonly used probiotics used in clinical trials:

Lactobacilli	Bifidobacteria
• *L. acidophilus*	• *B. adolescentis*
• *L. casei*,	• *B. animalis* subsp. *lactis*
• *L. helveticus*	• *B. breve*
• *L. johnsonii*	• *B. infantis*
• *L. paracasei*	• *B. longum*
• *L. plantarum*	Saccharomyces
• *L. reuteri*	• *S. cerevisiae* boulardii
• *L. rhamnosus*	

Only the probiotic genus and species listed. Many of them have more than one specific subspecies or more than one strain with beneficial effects studied in clinical trials. Some clinical effects are expected to be strain specific.

Mechanisms of action of probiotics

The rationale for probiotics to deliver health benefits arises from the dramatic improvement in our understanding of the mechanisms by which the host microbiota is a major determinant of health. These mechanisms and effects of the microbiota have been documented in several key functional domains:

- Effects on the gastrointestinal luminal environment (e.g., microbial digestion of macronutrients—particularly carbohydrates, with production of short-chain fatty acids (SCFA), luminal pH, and change sin composition of gastrointestinal contents and stools)
- Maintenance of gut barrier function and host protection (nutrient and adhesion competition and bacteriocins to inhibit pathogens), enhancement of barrier function of gut epithelium (e.g., stimulating mucus and mucin production, enhancement of function of tight junctions),
- Development of gut-associated and systemic immune response of the host (e.g., stimulation of secretory antibody production, particularly IgA, modulation of systemic T and B cell responses)

- Metabolic effects (salvage of energy from the diet by absorbing bacterial fermentation products like SCFA, and effects on enterohepatic circulation, and lipid absorption and metabolism)
- Gut-brain axis interactions (e.g., bidirectional neuroendocrine signaling which affects gut motility, permeability, pain perception, and neurocognitive function).

Alterations in the development, composition, and function of the microbiota and its symbiotic relationship with the host are generically described as "dysbiosis", and this has been associated with adverse health consequences elated to the health-related functional domains mentioned above. There is no consensus on what constitutes a "dysbiotic" microbiota profile. However, low presence of bifidobacteria and lactobacilli and elevated clostridia appear to be one of multiple markers of microbiota profiles associated with adverse health consequences. Several "risk factors for dysbiosis" have been associated with these deleterious alterations of the microbiota at different times of the life cycle, including birth by cesarean section, lack of breastfeeding, use of antibiotics, "poor" bacterial exposure or experiences (e.g., urban instead of rural living), and inadequate diets. And these variations have, in turn, been associated with short- and long-term risk of adverse clinical conditions, including infections, allergic disease, immune disorders, chronic inflammatory conditions, and functional gastrointestinal disorders. Emerging evidence suggests dysbiosis may increase risks for metabolic conditions such as dyslipidemia, metabolic syndrome, obesity, as well as neurologic, behavioral and mood disorders. The concept of probiotics is essentially the introduction of specific microorganisms, non-indigenous to the host, to alter the health risks associated with dysbiosis (ISAPP, 2019, Bermudez-Brito et al., 2012; Zommiti et al., 2020).

Most effects described and documented by probiotics relate to their interaction with the host's microbiota (microbial communities) and the complex gut luminal environment (microbial metabolites, toxins, and other organic molecules), which together compose the broader "microbiome." (Note: the term microbiome is used here as the microbiota plus its environment. The term is used in other contexts as to refer to the combined genetic material of the microorganisms in a particular environment). The interaction of specific microbes, including probiotics, with other microbes and the host's tissues and cells, occurs through signaling mechanisms that are exerted by:

- Microbial structural elements (such as bacterial pili and bacterial components including lipids, exo-polysaccharides, lectins, glycoproteins, and other proteins and polysaccharides),
- Microbial metabolites (organic and inorganic signaling molecules, toxins), and
- Mobile microbial genetic elements between microorganisms

The clinical effects on the host result from microbial-host signaling though these mechanisms, which bring about physiologic changes and their clinical consequences. These molecular interactive mechanisms are specific to different microbes. While some mechanisms may be shared among members of a particular taxonomic group (e.g., fermentation, production of SCFA, competitive exclusion of pathogens), others can be strain-specific (e.g., immunologic and neuroendocrine signaling, production of antimicrobial and other bioactive molecules).

Table 1 summarizes the physiologic mechanisms of action that have been documented in in-vitro, in-vivo, and in clinical trials, which have been associated and linked to clinical benefits from the use of probiotics. Most of these parallel the types of interactions described between host resident microbiota and health. As expected, specific probiotic microorganisms will have particular signaling and interactive mechanisms, and therefore specific effects on the host. As a result, many interactions and effects are strain-specific. These effects and their magnitude will also vary by the characteristics of the host and its environment, including their diet. Therefore, the final clinical outcomes observed in the host will be a "sum" of the multiple microbe-host interactions, influenced by multiple external factors.

The intestinal microbiota is established very early in life by microorganisms that colonize the gut lumen. In each period of the life cycle, these "indigenous" or "commensal" communities remain relatively stable. Thus, most ingested probiotics will not permanently "colonize" or be integrated into the host's microbiota communities. In fact, probiotic microorganisms will only be present in intestinal contents as long as they are ingested, and most effects will be exerted only while they conform part of the host microbiome. Furthermore, because most human studies are limited to studying and analyzing fecal microbiota, there is still very little understanding of the direct and indirect mechanisms by which they modify the microbiome in different parts of the human digestive tract (Bermudez-Brito et al., 2012; Lebeer et al., 2018; Berg et al., 2020; ISAPP, 2019).

Health outcomes reported with probiotics

The identification and selection of a bacterium for its clinical use as a probiotic requires.

- The specification of the microorganism (genus, species, and strain),
- Viability ("live microorganisms") in the product containing the probiotic at the time of ingestion
- A defined amount or "dose" at time of production and consumption (shelf life)
- The characteristics of the host or population ingesting the probiotic (e.g., age, diet, disease risk factors).
- A "defined benefit" (e.g., prevention or treatment specific pathologic condition(s) in that population

Table 2 summarizes the most studied health outcomes reported with the use of specific probiotics. The best documented positive outcomes reported in children are for prevention and management of gastrointestinal infections, infant colic, antibiotic-associated diarrhea, and necrotizing enterocolitis in premature infants. In adults, there is efficacy reported for lactose intolerance, constipation,

Table 1 The mechanisms listed include effects at cell, tissue, and organ system level. Their documentation derives from in-vitro, in-vivo, and ex-vivo studies as well as clinical trials. Some mechanisms may apply to broad microbial taxa; some are more strain specific. And various mechanisms may act in concert to yield specific clinical effects. Documentation of specific mechanisms may explain, and not necessarily predict a clinical outcome in the host, which needs to be demonstrated in clinical trials.

Physiologic mechanisms documented with the use of probiotics, which can be associated to clinical outcomes
Gastrointestinal functions
Lactase production by specific probiotics, aiding in lactose digestion
Modulation of stool composition and consistency, GI motility and bowel movement frequency
Modulation of the composition and activity of the microbiota
Competitive pathogen displacement, competition for nutrients, competitive adhesion to intestinal mucosa
Intestinal pathogen inhibition through luminal content acidification (production of SCFA)
Production of bacteriocins and other antimicrobials, inhibitors of bacterial adherence and virulence
Alteration of intestinal metabolome and immune mediated responses (see below)
Enhancement of gut epithelial barrier function
Decrease in epithelial permeability
Enhanced mucus and mucin production
Enhancement of tight junction functionality between epithelial cells
Improved epithelial cell proliferation and inhibition of apoptosis
Modulation of immune development and response
Modulation of cell- mediated and humoral immune functions through multiple mechanisms, including interaction with pattern recognition receptors of the immune system such as toll receptors, interactions with monocytes, macrophages, and dendritic cells,
Enhancing B-cell activity and secretory immunoglobulin production, particularly IgA
Effects on cells of the innate and adaptive immune system, modulating the balance of T-helper and T-regulatory cells, and inducing regulatory T-cell function, thus enhancing protection but decreasing inflammatory responses or tissue damage
Effects on systemic metabolic responses
Metabolic responses in the gut including modulation of bile salt hydrolase activity and enterohepatic circulation
Energy scavenging by absorption of products of fermentation in the lumen
Systemic metabolic responses from endocrine mediators, through potential probiotic stimulation and modulation of gut hormonal secretion, e.g., effects on intestinal secretion of ghrelin, gastrin, CCK, peptide YY
Effects on the central nervous system (via the gut-brain axis)
Bidirectional signaling to the central nervous system (CNS), affecting intestinal secretion, gut motility, intestinal permeability, immune response, and pain perception
Effects on production of neurochemicals: short-chain fatty acids (SCFA), oxytocin, gamma- aminobutyric acid (GABA), serotonin, tryptamine, noradrenaline, dopamine, acetylcholine, and other neurotransmitters,
Effects on gut hormone secretion and CNS signaling, including modulation of ghrelin, gastrin, CCK, peptide YY secretion
Modulation of gut related inflammatory markers which can have CNS effects, e.g., inflammatory cytokines IL-8, CCL4, CCL2, TNF-alpha, INF, IL-6
Nociceptive effects, e.g., blocking pain receptors in the GI tract

C. difficile colitis, ulcerative colitis/pouchitis, and Traveler's diarrhea with certain probiotics. There is less evidence for other conditions, including respiratory infections, eczema, and *H. pylori* infection. Evidence is emerging for dyslipidemias, urogenital infections, and neurobehavioral or mood disorders (Merenstein et al., 2020; Sanders et al., 2018; Szajewska, 2016; Pobio, 2022, Zommiti et al., 2020)

The clinical research literature on the "efficacy" of probiotics continues to grow and accumulate, albeit with significant limitations. Early studies in the late 80s and 90s were often less rigorous in specifying the microorganisms used in clinical trials, nor were the amounts, dose, or viability rigorously controlled or documented. This has improved significantly in the last 20 years, and the microorganisms and dosing of viable organisms (CFUs in the product) are adequately reported in most clinical trials today. Still, some studies use products that are not always well manufactured or controlled.

Other limitations in the scientific literature come from the breadth of organisms used, and different health benefits studied, with few confirmatory studies for efficacy of a specific probiotic, for a particular benefit, in a particular population. Studies of a specific bacterium will also vary significantly by design, dose, duration of intervention, etc. And there are almost no studies comparing one probiotic intervention with another. In addition, the risk profile for the subjects for the outcome studied may often not be accounted for or controlled. Today, based on a better understanding of the development of the microbiota and variables that influence risk for dysbiosis and disease, we recognize factors that should be controlled for, and most often are not. For example, a clinical trial done in a population of children may not control for antibiotic use prior to the trial, or for vaginal birth vs. by cesarean section, which we know now are clear confounders in the potential for dysbiosis and a clinical response to the intervention. Lastly, some studies use combinations of bacteria, which tests the effect of the combination but makes it difficult to establish individual bacterial components as probiotics.

The way that the accumulated literature is assessed regarding "efficacy" also has major limitations. Multiple systematic reviews and meta-analyses have evaluate the "efficacy of probiotics" for a particular clinical outcome. However, these analyses combine

genus, species, strains, dose, duration, etc., as if the trials addressed a similar, consistent, or homogeneous intervention targeting a particular health outcome, which is simply not appropriate. Such analyses of the clinical trial literature may help identify potential benefits broadly and may suggest the potential for use of any or all probiotics for certain clinical benefits; however, they make it very difficult, and in fact hinder, the development of specific clinical recommendations.

The high level of safety associated with most lactobacilli and bifidobacteria in most populations has worked for and also against the study and use of probiotics. Their relative safety has allowed for a large number of trials, but often done with less rigor than usually applied to a potential drug or similar intervention. In addition, the less-than-ideal regulatory frameworks increase the difficulty for communicating disease-related benefits when there is evidence. On the one hand, poor regulation facilitates poorly informed use or recommendations, even when positive data is not strong or rigorous. On the other, in some regulatory environments, the level of evidence needed to communicate a benefit is as high as it is for a drug. These limitations have created both "hype" as well as skepticism in different circles of scientists, academics, businesses, and consumers. Ultimately, a specific outcome with a specific intervention, using a specific probiotic candidate, in a particular and reasonably homogeneous population and replicated beyond a single clinical trial, should drive recommendations (Merenstein et al., 2020; Sanders et al., 2018; ISAPP, 2019, Zommiti et al., 2020).

Safety of probiotics

There is a broad and robust consensus that most microorganisms used as probiotics today are safe for most populations. Most safety concerns arise from three potential risks: bacteremia or infection, production of toxins or metabolites of concern, and transfer of genetic material (e.g., for antibiotic resistance) from probiotic bacteria to potential pathogens. Reassurances come from a long history of safe use, clinical trials, epidemiologic assessments, and increased understanding of bacterial metabolism and their genetics.

Infectious risk concerns

Most species of lactobacilli and bifidobacteria used today have a very long history of use in the food supply and as supplements, consumed by millions of individuals globally every day. Only a few epidemiologic and surveillance studies have been done to assess safety concerns, but none have shown any evidence of increases in adverse events with increased consumption. Many genera, species, and some strains of probiotics in use are present in human commensal microbiota, and many probiotics have been originally isolated from healthy humans.

No significant concerns of adverse effects have been identified in hundreds of clinical trials, although most of these trials are focused on efficacy and not specifically designed for detecting safety issues. Clinical trials in high-risk populations, including preterm infants, the elderly, patients with HIV, cancer, and other immunocompromised conditions, have not identified any significant risks. Somewhat paradoxically, some of the better-documented benefits of probiotics have been shown in populations at particularly "high risk", such as premature infants. Specific probiotics have demonstrated significant reductions in risk of necrotizing enterocolitis, with no adverse effects from the probiotics used.

Other concerns include the effects of D-lactate produced by probiotic strains (and potential risk of metabolic acidosis), and consequences from deconjugation of bile salts. Similarly, outside very rare cases of acidosis reported in patients with short bowel syndrome, there has been no significant or reliable documentation of other adverse effects from probiotic metabolic activity. Another theoretical concern with probiotics has been their potential for genetically transferring resistance to potential pathogens. There have been no adverse events reported based on this potential risk. Finally, concerns have also been raised following adverse events due to contaminants in probiotic products, not related to the probiotic itself. This is in part associated with poor manufacturing practices of supplements and health products, and lax oversight or regulatory controls in many countries.

However, risk is never zero, and some populations are of greater concern than others. Very rare cases of bacteremia and sepsis have been reported in patients with serious underlying conditions, including reports with use of lactobacilli and bifidobacteria. The safety record is sparser for non-lactobacillus and non-bifidobacterial strains used as probiotics (e.g., *Enterococcus and Saccharomyces)*. The most commonly reported type of adverse event has been fungemia with *S. boulardii.* Patients with central line catheters appear to be at particular risk (Sanders et al., 2010; Doron and Snydman, 2015; ISAPP, 2019).

Safety regulation

FAO/WHO guidelines recommend that for probiotics used in foods, safety be always assessed for side effects in human studies, for metabolic activities of concern, e.g., if the strain under evaluation belongs to a species known to produce toxins in mammals, it must be tested for toxin production. They should also be assessed for antibiotic resistance patterns, and epidemiological surveillance for adverse incidents should be done in consumers (aftermarket surveillance). These guidelines have generally been adopted, although the level of global compliance varies significantly. The European Food Safety Authority (EFSA) has adopted a premarket system for safety assessment of microbial species used in food and feed production. Once positively assessed, they are granted a "Qualified Presumption of Safety" (QPS) status. Most of the bacterial species used as probiotics have QPS status. In the United States, safety for use of ingredients in foods, including probiotics, requires petitioning and qualification by FDA as having "generally recognized as safe" (GRAS) status if the probiotic was not marketed in the United States prior to 1994. The US FDA requires that dietary supplements be manufactured under "good manufacturing practice" or GMPs standards. It is the manufacturer's responsibility to meet safety requirements and to provide documentation when required. There is no indication that probiotic dietary supplements in

Table 2 The conditions listed are those for which there is published evidence of efficacy of probiotics. See sources below for specific strains and doses indicated. The level of evidence is variable for various conditions as well as for the specific probiotics for each condition.

Clinical conditions for Which positive outcomes have been reported with the use of probiotics		
Condition	*Population*	*Probiotic*
Gastrointestinal		
Acute infectious diarrhea (prevention[1] & management[2])	Infants, children & adults	*L. rhamnosus, B. lactis, L. casei, L. plantarum L. reuteri, S. boulardii,* others[3]
Antibiotic associated diarrhea (prevention)	Children & adults	*L rhamnosus, L. reuteri, L plantarum, L. casei, B. lactis S. boulardii,* others, combinations[4]
C. difficile colitis (management)	Children & adults	*L. rhamnosus, L. plantarum, L. casei, L acidophilus + L casei, S. boulardii,* combinations
Colic (management)	Infants	*L. reuteri*
Community acquired infectious disease (GI & respiratory)	Children & adults	*B. lactis, L. casei, L reuteri, L rhamnosus*
Constipation (management)	Adults	*B. lactis, L. rhamnosus, L. reuteri, L. casei Shirota*
H. pylori infection (management)	Children & adults	*L. rhamnosus, S. boulardii, L. casei shirota, L. reuteri, L. plantaum, L.johnsonii,* others, combinations
IBS (management)	Adults	*B. lactis, B. longum, L plantarum, L. rhamnosus + L. helveticus, Combinations*
Lactose intolerance (management)	Childen & adults	*L. bulgaricus + S. thermophilus, L. acidophilus, L. delbrueckii*
Necrotizing enterocolitis (prevention)	Preemies	*L. breve, L. reuteri, Others, Combinations*
Recurrent abdominal pain (management)	Children	*L. reuteri, L. rhamnosus*
Traveler's diarrhea (prevention)	Adults	*L. rhamnosus, S. boulardii*
Ulcerative colitis, pouchitis (management)	Adults	*L. rhamnosus, S. boulardii, E. coli Nissle*
Skin		
Eczema/Atopic dermatitis (management)	Infants & young children	*L. reuteri, L. rhamnosus, B. breve*
Mastitis (management)	Adult women	*L. fermentum*
Urogenital		
Bacterial vaginosis & vulvovaginal candidiasis (management)	Adult women	*L. rhamnosus + L. reuteri,* other combinations
Metabolic		
Dyslipidemia, (reduction in total cholesterol and LDL)	Adults	*L. acidophilus, B. lactis, L. reuteri,* others
Mood		
Depression & anxiety (management)	Adults	*B. longum + L helveticus, L casei Shirota,* others

(1) "Prevention" refers to risk reduction of development of a condition. (2) "Management" refers most often to adjuvant use of the probiotic in the clinical management of the condition, to improve signs and symptoms, not as sole o primary treatment. (3) "Others" refers to additional individual probiotics reported. (4) "Combinations" refers to positive outcomes reported with combinations of multiple probiotics.

Sources: Sanders et al. (2018), Merenstein et al. (2020), Zommiti et al. (2020), Probio (2022).

compliance with these GMPs pose a safety concern to the general population (FAO/WHO, 2001; Degnan, 2008, Koutsoumanis, 2020).

Regulation of probiotics

Beyond the regulatory aspects of safety of probiotics, the regulation of claims regarding the benefits of commercialized probiotic products remains a complex and unresolved issue. Challenges arise from the lack of adequate or consistent regional or national regulatory frameworks, compounded by the limitations in quality and quantity of the scientific documentation for a positive health outcome. Some challenges relate to the type of purported benefit, how it is stated, and how it is applied to different product categories (foods, supplements, drugs). In addition, independent of a scientifically documented clinical effect, there is no universal agreement as to what constitutes a "benefit", a "health benefit", or a "health claim."

As an example, in the European Union in Europe, the word "probiotic" in a supplement or food is itself understood to be a "health claim", not a "nutrition claim." Therefore, inclusion of the word "probiotic" on a product label requires a "health claim approval" by the European Food and Safety Authority (EFSA). Given the EFSA's high level of evidence requirements, since 2006, there have been no claims approved by the Authority, and the use of the word "probiotic" remains essentially unauthorized. However, there is variability among EU Member States in interpreting and enforcing this regulation. On the other hand, the US FDA allows the use of the term "probiotic", and restrictions on claims depend on the classification of the product (e.g., a dietary supplement, food ingredient, or a drug). Dietary supplement labels can claim how the product affects the structure or function. Structure-function claims, e.g., "to support gut health" or "improves microbial balance", can be made without FDA approval. But manufacturers cannot make "health claims", that indicate the use to prevent, treat, or mitigate a disease. The use of claims

related to a disease, even if well documented, would essentially require the product to be classified as a drug and meet very different regulatory requirements than foods or supplements, or undergo a very complex route to receive a "qualified health claim", which has not happened in the United States (Degnan, 2008; Nyanzi et al., 2021; Rijkers et al., 2011; de Simone, 2019).

Prebiotics

Definition

The most recent and increasingly used prebiotic definition is "a substrate that is selectively utilized by host microorganisms conferring a health benefit", proposed by the International Scientific Association for Probiotics and Prebiotics (ISAPP). The concept of prebiotics was initially proposed by Gibson and Roberfroid in 1995 as a way to modify the intestinal microbiome to benefit the host by using dietary substrates that are selectively utilized by potentially beneficial species in the host microbiota (Gibson and Roberfroid, 1995; Gibson et al., 2017).

History

The concept of prebiotics is in many ways tied to that of probiotics, and both follow the growth in knowledge of the relationship between a host and its microbiota as a major determinant of health [see section on Probiotics]. In 1921, Rettger & Cheplin published experiments in animals and humans demonstrating that lactose and dextrins (less absorbable sugars than glucose or sucrose) led to an "enrichment with lactobacilli following consumption." And "when given in sufficient amounts" together with *L. acidophilus* dextrins "bring about a marked transformation of the intestinal flora in which *B. acidophilus* assumes particular prominence and may even completely supplant all other types of bacteria", an observation which heralded the concepts or prebiotics and synbiotics (the combination of pre- and probiotics).

In the '70s and '80s, there was a significant development in the understanding of dietary fiber: edible substances that are not digested by human enzymes and will undergo varying levels of fermentation by bacteria in the distal gut. Beneficial outcomes with certain dietary fibers included effects on gastrointestinal health (e.g., improved intestinal regularity and risk reduction of colorectal cancer), lipid and metabolic status (e.g., lowering cholesterol, modulating blood sugar levels), and healthy weight maintenance. Research in the '90s showed that some of these non-digestible carbohydrates, particularly fructooligosaccharides found in common foodstuffs such as artichoke and chicory, were selectively utilized by bifidobacteria and lactobacilli. This gave way to the designation of these as "prebiotics" by Gibson and Roberfroid, who defined the term as "non-digestible food ingredients that beneficially affect the host by selectively stimulating the growth and/or activity of one or a limited number of bacteria already resident in the colon." This definition gradually evolved in the literature. Initially, it referred to "food ingredients", since these substances were mostly non-digestible carbohydrates of vegetable origin (e.g., inulin, fructooligosaccharides). Today various other substances, including human milk oligosaccharides, non-carbohydrate substances, and synthesized substances, are considered as having this selective effect on microbiota. Other definitions limited the term to "effects in the colon", or "in the gut." However, this did not address potential changes or effects on other microbial habitats such as skin, oral cavity, respiratory tract, and vagina. The type of effects of these substances on specific species has also evolved, from "selectively stimulating the growth and/or activity", or "selectively fermented", to "utilized." And finally, the "selective utilization" has, for the most part, referred to preferential substrate use by *Bifidobacterium* and *Lactobacillus* species. In practice, the term has been commonly and less rigorously used to describe "bifidogenic" substances. Today, emerging evidence suggests growth promotion of other species may be beneficial (Gibson et al., 2017; Davani-Davari et al., 2019).

Parallel to these developments, in the '50s, P. György and R. Kuhn described the presence in human milk of a "bifidogenic" substance, which they called "*L. bifidus* growth factor." They showed it was present in much greater concentrations in human versus cow milk and in higher concentrations in human colostrum. This explained the predominance of bifidobacteria in the distal intestine of healthy infants. Subsequent work showed that these were a large family of structurally complex unconjugated glycans which are not digested by human enzymes, eventually called human milk oligosaccharides (HMOS). HMOS are selectively utilized by bifidobacteria and have well-documented beneficial health effects, thus fulfilling the definition of prebiotics (Walsh et al., 2020).

The current ISAPP definition, "a substrate that is selectively utilized by host microorganisms conferring a health benefit," broadens the concept to include non-carbohydrate substances, categories other than food, and applications other than the gastrointestinal tract; but retains the requirement for documentation of a health benefit via selective microbiota-mediated mechanisms. On the one hand, this enlagest the concept for research, inclusion, and exploration of many substances as candidate prebiotics (Sanders et al., 2019; Gibson et al., 2017). On the other hand, it requires the demonstration, not only of a health benefit, but also documentation that such benefit is linked to a specific microbial-mediated mechanism. The strict demonstration of such a link has consequences in terms of classification, communication, and regulatory challenges for using the term.

Prebiotic substances

To be substrates "selectively utilized" by the microbiota, prebiotics are non-digestible, non-absorbable substances, mainly carbohydrates. Historically they have been associated with, or defined by, their ability to promote the growth of *Bifidobacterium* and *Lactobacillus* species; however, some more recent work suggests that other species of "potentially beneficial" microorganisms like

Faecalibacterium, *Akkermansia*, *Ruminococcus*, and *Roseburia*, could be increased by prebiotic substances. Prebiotics may be present naturally in foodstuffs, or they can be synthesized. Research has focused on components isolated from plants or synthesized substances rather than whole plant foods, allowing for tighter control of substance and dose. Most prebiotic substances studied can also be qualified as dietary fibers (all of which are non-absorbable, and thus available to gut luminal bacteria), particularly soluble fermentable fibers. However, some but not all soluble fibers have a prebiotic effect, and some proposed "candidate" prebiotics are not saccharides, such as polyphenols and polyunsaturated fatty acids. Structurally, most prebiotics are saccharides of varying size and complexity. The best-studied categories include fructans (fructooligosaccharides), galactans (galactooligosaccharides), and human milk oligosaccharides.

Fructo-oligosaccharides (FOS)

FOS are glycans composed of linear chains of fructose units linked by β(2-1)glycosidic bonds. The number of fructose units ranges from 2 to 60 and often terminate in a glucose unit.

FOS occur naturally in small concentrations in plants such as onion, garlic, asparagus, banana, and artichoke. However, most FOS used today in supplements or food ingredients are primarily inulin and inulin-derived fructooligosaccharides. Inulin is found at high concentrations in chicory root, artichoke, or agave, from which long it is extracted. Shorter chain fructooligosaccharides (FOS) are usually obtained through partial enzymatic hydrolysis of inulin and have a typical DP of less than 10 fructose units. FOS can also be manufactured by enzymatic synthesis using fructosyltransferases.

Galactooligosaccharides (GOS)

GOS, also called oligogalactose, oligolactose, or trans galactooligosaccharides (TOS), occur naturally in some mammals' milk, particularly the milk of marsupials, but is found in negligible concentrations in human milk. These galactose polymers typically have β(1-6) bonds (β GOS). GOS can also be naturally found in plants, particularly pulses, like chickpeas, in which the bonds are α bonds (α GOS), including α (1-6), α (1-4,6) type bonds.

GOS can also be obtained (α-GOS) by extraction from plants, mainly from pulses such as soybean, lentil, and chickpeas. However, most GOS used today as supplements or food ingredients are enzymatically synthesized from lactose using β-galactosidases. They typically have chains of 2–10 molecules of galactose, the final product and DP depending mainly on the type of galactosidase used.

Human milk oligosaccharides (HMOS)

After lactose and fat, HMOS constitute the third-largest solid component by weight in human milk; and their concentration is highest in colostrum and decreases in mature milk. They comprise approximately 200 structurally different structures, composed primarily of lactose (glucose and galactose connected by a β(1-4) glycosidic bond), and chains of varying complexity which include fucose, N-ethylglucosamine, and sialic acid. These glycans can be linear or branched, and the sequences can have fucose or sialic acid. They can be categorized into fucosylated, sialylated, and neutral core HMOS. Of all HMOS, the most abundant oligosaccharide is 2′-fucosyl-lactose (2′FL), which is present in the milk of most women. Some mothers have lower concentrations of 2′FL and other fucosylated HMOS in their milk. This is genetically determined by the absence of specific fructosyltransferases (FUT) in the mammary gland, particularly FUT 3.

Despite being a major solid component of human milk, due to their particular glycosidic bonds, HMOS are not digested by human enzymes, and thus have no nutritional value for the infant. 1% is absorbed into the circulation; the rest transit through the intestine and serve as substrate for bacteria in the gut lumen. HMOS are a highly selective substrate for a number of species of bifidobacteria, which are consequentially predominant in the microbiota of breastfed infants. This predominance is absent in infants fed infant formulas. Compared to FOS or GOS, HMOS are more selectively utilized by bifidobacteria. Some *Bifidobacterium* species, like *B. infantis*, have entire gene clusters that control the expression of specific glycosidases and transporters dedicated to HMO utilization.

About 20% of mothers have lower concentrations of 2′FL and other fucosylated HMOS in their milk. This is genetically determined by the absence of specific fructosyltransferases (FUT) in the mammary gland, particularly FUT 3, and has allowed for observational studies related to these differences. In otherwise healthy infants, a lower concentration of fucosylated HMOS, particularly low 2′FL in mother's milk, is associated with decreased bifidobacteria in the gut, and an increased risk of infectious diseases, and potentially atopic diseases. Thus, HMOS fulfill all the criteria to be considered a prebiotic. In addition, outside their effect on the microbiota, they also have direct effects related to pathogen protection, epithelial integrity, and immune modulation in the infant. Over the last few years, chemical and enzymatic synthesis have made possible the commercial production of some HMOS, like 2′-fucosyl-lactose, lacto-N-neotetraose, and 3- and 6-sialyllactose, which has allowed further research to document clinical benefits of specific HMOS (Sprenger et al., 2022; Wiciński et al., 2020).

Prebiotic candidates

Many other substances are being increasingly assessed for their potential effect as prebiotics. These include simple and complex non-digestible saccharides which behave like dietary fibers and have variable documentation of selective utilization by specific species. And they have varying levels of evidence on health benefits, including improvements in blood sugars and lipids, digestive health

benefits, laxation, and beneficial changes to immune markers. Others, like polyphenols and polyunsaturated fatty acids, are novel prebiotic candidates with alternate mechanisms of action. These include:

Saccharide prebiotic candidates

- Beta-glucans
- Resistant starch (RS)
- Guar gum.
- Polydextrose (PDX)
- Lactulose
- Xylooligosaccharides (XOS)
- Isomalto-oligosaccharide (IMO)
- Lactosucrose

Non-saccharide prebiotic candidates

- Polyphenols
- Polyunsaturated fatty acids

(Gibson et al., 2017; Wiciński et al., 2020; Sprenger et al., 2022; Davani-Davari et al., 2019; Hutkins et al., 2016).

Health outcomes reported with prebiotics

Many benefits reported in clinical trials with prebiotic substances parallel those documented with probiotics. Some of these benefits are associated with, and potentially driven by, increases in lactobacilli, bifidobacteria, or other beneficial species (an "indirect" probiotic effect). Some outcomes observed with these substances may also result from other potentially beneficial effects and mechanisms.

The best documented compositional effects of FOS, GOS, and HMOS on the host microbiota are growth promotion of bifidobacteria and lactobacilli, with variable decreases in other less desirable species such as *Clostridium*. Positive health outcomes reported in clinical trials using prebiotics are summarized in **Table 3**. The strength of the evidence so far is higher for digestive-related effects, e.g., constipation and stool regularity, which have been well documented for inulin, FOS, GOS, and other saccharides; and have been allowed as claims by some regulatory bodies. Lowering of total and LDL cholesterol and glycemic control have been shown with inulin, some FOS, and some GOS, as well as other soluble fibers. Improved calcium absorption

Table 3 The conditions listed are those for which there is published evidence of efficacy of prebiotics. With some exceptions, the overall level of evidence so far is quite limited and highly variable for most outcomes. See article text.

Clinical conditions for which positive outcomes have been reported with the use of prebiotics			
Condition	*Population*	*Effect*	*Prebiotic*
Gastrointestinal			
Constipation	Children & adults	Increase in stool weight, and frequency, softer stool consistency,	Inulin, FOS, GOS
Infectious diarrhea	Infants & adults	Risk reduction	FOS, GOS, HMOs
Necrotizing enterocolitis	Preemies	Decreased risk of developing NEC	FOS + GOS
Traveler's diarrhea	Adults	Risk reduction	GOS
Ulcerative colitis	Adults	Marginal improvement of symptoms	FOS
Respiratory			
Respiratory infections	Infants, children and adults	Risk reduction	FOS, GOS[a], HMOs
Skin			
Eczema/Atopic dermatitis	Infants	Risk reduction and amelioration of symptoms	FOS + GOS
Urogenital			
Bacterial vaginosis	Adult women	Improvement in efficacy of antibiotic treatment	GOS
Metabolic			
Dyslipidemia,	Adults	Reduction in total and LDL cholesterol	Inulin, FOS
Glucose control	Adults	Improved glucose control	Inulin, FOS
Bone health			
Improved calcium absorption	Children & adults	Improved bone mineralization Potential for managing osteoporosis in adults	Inulin, FOS

[a]Most often in combination with a probiotic.

Sources: Gibson et al. (2017), Carlson et al. (2018), Sanders et al. (2019).

has been demonstrated with some saccharide prebiotics. The evidence for other outcomes, including infections (gastrointestinal, respiratory, urogenital), and other inflammatory conditions (eczema, ulcerative colitis, etc.) is still quite limited, and results are inconsistent.

This literature on clinical health effects has consistently increased, but also has significant limitations. The number of prebiotic substances being studied, with different compositions (type of saccharides, degree and complexity of polymerization, physicochemical properties, etc.), is very large. However, with some exceptions, the number of confirmatory studies for specific individual prebiotic candidates is generally low. Studies also vary greatly in design, dose, duration of intervention, and definition of populations studied. As with probiotics, assuming part or all of the effect is driven by improvements in microbiome composition, adequate control for factors affecting the microbiota in the populations being studied is needed, and not always accounted for. Lastly, many systematic reviews and meta-analyses assess "prebiotic interventions" for a particular health outcome, assuming it is a single or homogeneous form of intervention. They report on the "efficacy of prebiotics" as a whole for a particular outcome. Such analyses do not help define the efficacy of a specific intervention, and actually, make consensus and guidance more difficult. Recommendations should be driven by evidence for a specific intervention, for a specific outcome, with a specific prebiotic candidate, in a specific, reasonably homogeneous population, and which is replicated beyond a single clinical trial (Gibson et al., 2017; Sanders et al., 2019; Carlson et al., 2018; Walsh et al., 2020).

Mechanisms of action of prebiotics

As with most soluble dietary fibers, saccharide-based prebiotics will generally be fermented by bacteria in the distal gut. This leads to production of SCFA and other organic acids, increased intestinal gas, lower gut lumen pH, increased water stool content, and decreased stool consistency. These mechanisms explain the effects of certain dietary fibers on stool consistency, and bowel movement regularity, which have been reported with inulin, FOS, and GOS. Emerging evidence on prebiotic effects on immunity may be indirectly related to similar mechanisms. How much of these effects relate to actual "selective" microbial utilization by host microbiota, as would be required to be defined as a prebiotic, is less well defined, and probably very variable.

Prebiotics like inulin and FOS have also been shown to help decrease fasting glucose levels and decrease LDL and total cholesterol in individuals with dyslipidemia. The mechanisms for reducing lipid levels, blood glucose control, and immune protection modulation attributed to some prebiotics are not yet totally clear and have mainly been studied in vitro and in animal models. It is postulated, for example, that the generation of particular SCFA by prebiotics, such as propionate, may modulate hepatic gluconeogenesis. SCFAs could also influence appetite regulation and food intake by triggering a release of the gut hormones like GLP-1 and PYY, and stimulating glucose-dependent insulin secretion, inhibiting glucagon release in the pancreas, slowing gastric emptying, as well as directly suppressing appetite in the brain.

The absorption of minerals, particularly calcium and magnesium, appears to be enhanced by some prebiotics. It has been proposed that SCFA production and lower luminal pH can increase the solubility of mineral salts and improve mucosal function and absorptive surface, increasing mineral absorption. The potential mechanisms for protective and immune-related effects of prebiotics on infection and inflammatory conditions such as allergy are still unclear. These include increased bacterial production of SCFA, which lowers pH and inhibits pathogens, and support gut mucosal integrity. An "indirect probiotic effect" could also enhance mucosal integrity, immune function, and regulate expression of pro- or anti-inflammatory cytokines.

HMOS can, in part, behave mechanistically like other saccharide-based prebiotics. They are utilized in a highly selective way by specific genera and species, lower pH, and increase SCFA in the distal gut lumen, which inhibits pathogens, modulates the microbiota composition, and enhances epithelial protective barrier function. In fact, they have been described as "the fiber of breast milk." However, it is also clear HMOS have multiple other direct effects on resident microbiota (e.g., serving as decoys for pathogens) and on epithelial and immune function, not directly related to their bifidogenic or fermentative capacity (e.g., interacting directly with toll receptors and dendritic cells in the gut mucosa.) (Carlson et al., 2018; Gibson et al., 2017; Wiciński et al., 2020; Sanders et al., 2019)

Safety

The safety profile for most naturally occurring and synthetic prebiotics or prebiotic candidates is very good, and for many of them, well supported by a long history of safe use. Inulin, FOS, and GOS were used in the EU before 1997 and are considered safe food ingredients. However, prebiotic substances created and introduced after 1997 (e.g., synthesized HMOS) are considered novel and require safety assessment for commercialization. In the US, the FDA permits food manufacturers to self-affirm GRAS (Generally Recognized as Safe) status for products included as supplements or food ingredients, which has been the case with most inulin, FOS, GOS, polydextrose, HMOS, and other potential prebiotic substances (Sanders et al., 2019; Davani-Davari et al., 2019).

Regulation of prebiotics

To date, no major regulatory body has established a definition for the term prebiotic. All related health claims allowed by the US Food and Drug Administration (FDA) and the European Food Safety Authority (EFSA) relate to specific dietary substances, whether they could be considered a prebiotic or not. For example, "Chicory inulin contributes to maintenance of normal defecation by increasing stool frequency" and "non-digestible carbohydrates contribute to a reduction in post-prandial glycemic response" are

claims allowed by EFSA. However, in Europe or Canada, the term "prebiotic" itself is considered a health claim. Thus, to be allowed, requires a very high level of evidence for addressing a disease or pathologic condition. Furthermore, any potentially beneficial modification to the intestinal microbiota resulting from prebiotic use is generally not acceptable as a health claim or benefit.

Similarly, the US FDA has allowed claims for specific dietary fibers (potential prebiotics), such as oat fiber and whole-grain barley, which contain beta-glucans for improving cholesterol levels and reducing risk of coronary heart disease. However, it is unclear how much selective growth of specific species is needed to achieve theses clinical outcomes (to be considered a "prebiotic effect"); and the term "prebiotic" remains undefined by the FDA. Any health claims of a substance related to prevention or management of disease must be submitted to the FDA through a health claims petition process. Currently, there are no health claims, i.e., no disease-related claims allowed specifically for any prebiotics. Nevertheless, the term is used in supplements and food ingredients in practice, with similar claims used for some dietary fibers, e.g., "for maintaining digestive health" and "helping increase levels of good bacteria in the large intestine." In Japan, the Foods for Specified Health Use (FOSHU) permits claims on some foods and beverages that provide health benefits if valid scientific proof is provided to support such claims. Prebiotics are not explicitly mentioned or defined but can be categorized as FOSHU, which includes some oligosaccharides, lactose, dietary fiber, polydextrose, and guar gum (Sanders et al., 2019; Hutkins et al., 2016).

Conclusion

The explosion of knowledge on the relationship between human health and its microbiota is mirrored by the volume of scientific work focused on approaches to modulate the composition of these microbial ecosystems in ways that are beneficial to the human host. Basic science and in-vitro studies provide clear and robust evidence of the host-microbe interactions which mediate these interactions and support the rationale for probiotics and prebiotics to enhance health. Specific benefits with specific pre- and probiotics on particular outcomes have been well documented. Pre- and probiotics in use today are safe for the general population, and specific benefits have been reported in some high-risk groups, including premature infants, immunocompromised individuals, and the elderly.

However, the strength and consistency of the evidence are extremely variable, and there are a great many trials, with varying levels of rigor, with many pre-and probiotics, rather than less, more rigorous studies, on specific microbes or prebiotic substances, for specific outcomes. And attempts to assess pre- or probiotic "efficacy" as a whole category are not helpful. This has hindered consensus on efficacy and made guidance more difficult. Less than ideal and inconsistent regulatory frameworks have also created challenges in communicating potential benefits when evidence exists and allowing communication of unsubstantiated benefits in some product categories.

Pre- and probiotic interventions to promote health have a strong basis, some benefits are well demonstrated, and much potential remains untapped. Continued emphasis on rigorous research should help close the translational gaps that still exist between concept, science, and application.

References

Berg, G., Rybakova, D., Fischer, D., Cernava, T., et al., 2020. Correction to: microbiome definition re-visited: old concepts and new challenges. Microbiome 8 (1), 119.

Bermudez-Brito, M., Plaza-Díaz, J., Muñoz-Quezada, S., Gómez-Llorente, C., Gil, A., 2012. Probiotic mechanisms of action. Ann. Nutr. Metab. 61, 160–174.

Carlson, J.L., Erickson, J.M., Lloyd, B.B., Slavin, J.L., 2018. Health effects and sources of prebiotic dietary fiber. Curr. Dev. Nutr. 2 (3), nzy005.

Davani-Davari, D., Negahdaripour, M., Karimzadeh, I., et al., 2019. Prebiotics: definition, types, sources, mechanisms, and clinical applications. Foods 8 (3), 92.

de Simone, C., 2019. The unregulated probiotic market. Clin. Gastroenterol. Hepatol. 17 (5), 809–817.

Degnan, F.H., 2008. The US Food and Drug Administration and probiotics: regulatory categorization. Clin. Infect. Dis. 46 (Suppl. 2), S133–S136. Discussion S144-51.

Doron, S., Snydman, D.R., 2015. Risk and safety of probiotics. Clin. Infect. Dis. 60 (Suppl. 2), S129–S134.

Food and Agriculture Organization of the United Nations & World Health Organization, 2001. Report of a joint FAO/WHO expert consultation on evaluation of health and nutritional properties of probiotics in food including powder milk with live active lactic acid bacteria. Argentina. In: FAO WHO Probiotics in Food. Health and Nutritional Properties and Guidelines for Evaluation. FAO Food and Nutrition Paper 85. Rome 2006. ISSN 0254-4725.

Fujii, A., Cook, E.S., 1973. Probiotics. Antistaphylococcal and antifibrinolytic activities of omega-guanidino acids and omega-guanidinoacyl-L-histidines. J. Med. Chem. 16 (12), 1409–1411.

Fuller, R., 1989. Probiotics in man and animals. J. Appl. Bacteriol. 66 (5), 365–378.

Gibson, G.R., Roberfroid, M.B., 1995. Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. J. Nutr. 125 (6), 1401–1412.

Gibson, G.R., Hutkins, R., Sanders, M.E., et al., 2017. Expert consensus document: the International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. Nat. Rev. Gastroenterol. Hepatol. 14 (8), 491–502.

Hamilton-Miller, J.M., Gibson, G.R., Bruck, W., 2003. Some insights into the derivation and early uses of the word "probiotic". Br. J. Nutr. 90 (4), 845.

Hill, C., Guarner, F., Reid, G., et al., 2014. Expert consensus document. The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. Nat. Rev. Gastroenterol. Hepatol. 11 (8), 506–514.

Hutkins, R.W., Krumbeck, J.A., Bindels, L.B., et al., 2016. Prebiotics: why definitions matter. Curr. Opin. Biotechnol. 37, 1–7.

International Scientific Association for Probiotics and Prebiotics, 2019. The ISAPP Quick Guide to Probiotics for Health Professionals: History, Efficacy, and Safety. https://isappscience.org/for-clinicians/resources/probiotics/. (Accessed 15 April 2022).

Koutsoumanis, K., Allende, A., Alvarez-Ordóñez, A., et al., 2020. Scientific opinion on the update of the list of QPS-recommended biological agents intentionally added to food or feed as notified to EFSA (2017–2019). EFSA Panel on Biological Hazards (BIOHAZ), EFSA J. 18 (2), e05966.

Lebeer, S., Bron, P.A., Marco, M.L., et al., 2018. Identification of probiotic effector molecules: present state and future perspectives. Curr. Opin. Biotechnol. 49, 217–223.

Lilly, D.M., Stillwell, R.H., 1965. Probiotics: growth-promoting factors produced by microorganisms. Science 147 (3659), 747–8.

Marco, M.L., Sanders, M.E., Gänzle, M., et al., 2021. The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on fermented foods. Nat. Rev. Gastroenterol. Hepatol. 18 (3), 196–208.
McFarland, L.V., 2010. Systematic review and meta-analysis of *Saccharomyces boulardii* in adult patients. World J. Gastroenterol. 16 (18), 2202–2222.
Merenstein, D.J., Sanders, M.E., Tancredi, D.J., 2020. Probiotics as a Tx resource in primary care. J. Fam. Pract. 69 (3), E1–E10.
Metchnikoff, E., 1907. The Prolongation of Life: Optimistic Studies (Translated from the French by P. Chalmers Mitchell). William Heineman, London.
Nyanzi, R., Jooste, P.J., Buys, E.M., 2021. Invited review: probiotic yogurt quality criteria, regulatory framework, clinical evidence, and analytical aspects. J. Dairy Sci. 104 (1), 1–19.
Parker, R.B., 1974. Probiotics, the other half of the antibiotic story. Anim. Nutr. Health 29, 4–8.
Probio, A.E., 2022. Clinical Guide to Probiotic Products Available in USA Applications, Dosage Forms and Clinical Evidence to Date—2022 Edition. http://www.usprobioticguide.com. (Accessed 15 April 2022).
Rijkers, G.T., de Vos, W.M., Brummer, R.J., Morelli, L., Corthier, G., Marteau, P., 2011. Health benefits and health claims of probiotics: bridging science and marketing. Br. J. Nutr. 106 (9), 1291–1296.
Rusch, V., 2002. Probiotics and definitions: a short overview. In: Heidt, P.J., Midtvedt, T., Rusch, V., van der Waaij, D. (Eds.), Old Herborn University Seminar Monograph 15: Probiotics: Bacteria and Bacterial Fragments as Immunomodulatory Agents. Herborn Litterae, Herborn-Dill, Germany.
Sanders, M.E., Akkermans, L.M., Haller, D., et al., 2010. Safety assessment of probiotics for human use. Gut Microb. 1 (3), 164–185.
Sanders, M.E., Merenstein, D., Merrifield, C.A., Hutkins, R., 2018. Probiotics for human use. Nutr. Bull. 43, 212–225.
Sanders, M.E., Merenstein, D.J., Reid, G., Gibson, G.R., Rastall, R.A., 2019. Probiotics and prebiotics in intestinal health and disease: from biology to the clinic. Nat. Rev. Gastroenterol. Hepatol. 16 (10), 605–616.
Sprenger, N., Tytgat, H., Binia, A., Austin, S., Singhal, A., 2022. Biology of human milk oligosaccharides: from basic science to clinical evidence. J. Hum. Nutr. Diet. 35 (2), 280–299.
Szajewska, H., 2016. What are the indications for using probiotics in children? Arc. Dis. Child. 101 (4), 398–403.
Vaughan, R.B., July 1965. The romantic rationalist: a study of Elie Metchnikoff. Med. Hist. 9 (3), 201–215. PMID: 14321564; PMCID: PMC1033501.
Walsh, C., Lane, J.A., van Sinderen, D., Hickey, R.M., 2020. Human milk oligosaccharides: shaping the infant gut microbiota and supporting health. J. Funct. Foods 72, 104074.
Wiciński, M., Sawicka, E., Gębalski, J., Kubiak, K., Malinowski, B., 2020. Human milk oligosaccharides: health benefits, potential applications in infant formulas, and pharmacology. Nutrients 12 (1), 266.
Zhang, Z., Wang, J., Wang, J., Wang, J., Li, Y., 2020. Estimate of the sequenced proportion of the global prokaryotic genome. Microbiome 8 (1), 134.
Zommiti, M., Feuilloley, M., Connil, N., 2020. Update of probiotics in human world: a nonstop source of benefactions till the end of time. Microorganisms 8 (12), 1907.

Protein quality and sources

Daniel Tomé, UMR PNCA, AgroParisTech, INRAE, Université Paris-Saclay, Paris, France

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A.V. Kurpad, Protein: Quality and Sources, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 123–130, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00241-5.

Introduction 560
Protein content and amino acid composition 560
Protein concentration 560
Protein/energy ratio 560
Amino acid analysis of food proteins 561
Assessment of protein quality in human and animal 561
Metabolic studies 561
Influence of energy intake on nitrogen balance 561
Assays in laboratory animals 562
Assessment of protein quality by amino acid scoring 562
Amino acid scores 562
Amino acid scoring reference pattern 562
Correction for protein digestibility 563
Protein quality and dietary sources 563
Differences in protein content, amino acid profile and digestibility 563
Improvement of protein concentration, amino acid profile and digestibility 564
Conclusion 566
References 566
Further reading 566

Key points

- The nutritive value of proteins depends both on the amount of protein provided, on the amino acid composition and concentration, and on the bioavailability of nitrogen and amino acids.
- Protein quality is assessed in humans by measuring nitrogen balance, by biological assay in animal models, or from amino acids using the amino acid scoring corrected by the digestibility.
- Protein quality matters because there are differences in the quality of protein from the different food sources and some forms of food storage and processing can also affect protein quality.

Glossary

Amino acid score (or "chemical score") Value of the limiting amino acid with the lowest score in a protein (i.e., the "most limiting amino acid"). A protein is assigned a percentage score of 100 (or a fractional score of 1.00) when one of its EAAs are limiting

Amino acid scoring pattern Amino acid composition of a hypothetical reference protein that contains all EAAs in the amounts necessary to satisfy requirements

Amino acid scoring procedure Calculation of the proportion of each EAA in a protein or diet relative to the scoring pattern. It can be expressed as percentage or as a fractional value

Essential amino acids (EAAs) Also called "indispensable." Amino acids that the diet must provide because humans cannot synthesize them from other components at a rate commensurate with normal body needs

Limiting amino acids EAAs in food proteins that are present in lower proportions than in the reference protein (i.e., with fractional value o1.00, relative to the reference protein)

Nitrogen balance The average amount of nitrogen that is retained or lost from the body. It is calculated from measurements of dietary, urinary, and fecal nitrogen and estimates of integumental (sweat, skin, nails, and hair) nitrogen losses

Protein (or nitrogen) or amino acid digestibility The proportion of dietary nitrogen or amino acid that is absorbed. "True" protein or amino acid digestibility is calculated correcting for endogenous or obligatory fecal or ileal nitrogen or amino acids losses (i.e., nitrogen or amino acid in epithelial cells, gastrointestinal secretions, and intestinal flora)

Nomenclature

CCVP Codex Committee on Vegetable Proteins
DIAAS Digestible Amino Acid Score
EAA Essential Amino Acid
FAO United Nations Food and Agriculture Organization
NPU Net Protein Utilization
PDCAAS Protein Digestibility-Corrected Amino Acid Score
P/E ratio Percentage of Protein Energy
PER Protein Efficiency Ratio
UNU United Nations University
WHO World Health Organization

Introduction

Proteins are essential nutrients in human diets, and an adequate protein supply is required to meet the body's nutritional needs (WHO/FAO/UNU, 2007). Protein quality is related to the capacity of protein from foods and diets to provide nitrogen and dietary essential (or indispensable) amino acids (EAAs) to achieve nitrogen balance and support body composition and functions for growth, maintenance, and in different physiological situations in humans. The nutritive value of proteins from food and diet depends both on the amount of protein provided, but also on the amino acid composition and concentration, and on the bioavailability of protein-derived nitrogen and amino acids. This quality influences the quantity of protein intake required to meet dietary nutritional needs—the lower is the quality, the higher is the required dietary protein intake. Protein quality matters because there are differences in the quality of protein from the different food sources and some forms of food storage and processing can also affect protein quality. This article examines the ways of assessing the protein quality of foods and diets and the quality inherent to various protein source.

Protein content and amino acid composition

Protein concentration

Protein concentration or density (i.e., the amount of protein per unit of food) is a factor of a food's protein quality. Protein content in food products and food ingredients is determined either by measuring nitrogen content, by measuring amino acid content, or by direct measurement of protein by colorimetric methods, spectrophotometric methods, or infrared spectroscopy (Tome et al., 2019). The different methods have some limitations and measuring nitrogen content with the Kjeldahl or Dumas methods and using a Nitrogen to Protein Conversion Factor remains the more frequently used approach for protein content in foods. The conversion factor used for a mixture of protein sources is 6.25, corresponding to a nitrogen content of 16%. Specific protein conversion factors range from 5.7 (17.5% nitrogen) to 6.4 (15.6% nitrogen) for the major protein sources in the diet.

Protein/energy ratio

The percentage of protein energy in the diet (P/E ratio) has been used to describe whether a diet provides adequate amounts of protein. The reasoning is that energy requirements are the main driving force for food intake. Therefore, a diet is adequate if it satisfies the requirements for all nutrients when it is eaten in amounts that will satisfy energy needs. P/E ratio is calculated by dividing the amount of metabolizable energy derived from dietary protein (grams of protein 16.7 kJ or 4 kcal) by the total amount of metabolizable energy in the diet, multiplied by 100 to avoid using fractional values. The P/E ratio indicates the amount of protein that the diet provides relative to energy but does not imply a constant relationship between protein and energy requirements. Obviously, the critical modifiable factor is the energy requirement, which is linked to the habitual energy expenditure level. Therefore, in adults with a lower energy requirement, such as the older subjects, the P/E requirement ratio will be higher, while in adults with a high energy expenditure and energy requirement the P/E requirement ratio will be lower. Diets, especially those eaten by adults, often provide protein in amounts that surpass requirements, which elevates the P/E ratio.

Amino acid analysis of food proteins

Modern methods that involve acid or alkaline hydrolysis of the protein followed by separation and quantification of the released amino acids by ion-exchange chromatography, gas chromatography, high-performance liquid chromatography and its ultra-fast variant, and electrophoresis, and other chemical and microbiological methods for specific amino acids, such as lysine, methionine, cysteine, and tryptophan, provide data with a repeatability within laboratory of approximately 5% and a reproducibility between laboratories of approximately 10% (Tome et al., 2019). Not all amino acids are fully released or are partially destroyed under conditions of standard acid hydrolysis and acid hydrolysis after performic acid oxidation or alkaline hydrolysis are performed for sulfur amino acids or tryptophan, respectively. Although several national and international food composition tables include amino acid contents of foods, it is preferable to use analytical results from a reliable laboratory owing to technical shortcomings in the preparation of some tables and to the considerable variability between the reported values, especially for tryptophan, cysteine, and methionine. Amino acid data are usually calculated as milligrams amino acid per gram of protein. If they are reported as milligrams amino acid per gram of nitrogen, they are converted to the protein equivalents by multiplying by specific protein conversion factors. To calculate the amino acid content of a combination of food proteins, as in a processed food based on several protein sources or in a mixed diet, a weighted mean of the published or analytical results of each component should be used.

Assessment of protein quality in human and animal

Metabolic studies

The most accurate assessment of protein quality of foods for humans is through clinical or metabolic studies that measure nitrogen balance (WHO/FAO/UNU, 2007). Different amounts of food protein are fed to a group of individuals and to calculate nitrogen losses for each amount excreta are collected and analyzed for their nitrogen content until a steady state is reached and integumental nitrogen losses generally estimated at approximately 5 mg N kg^{-1} day^{-1}. At that point, the relationship between nitrogen intake and nitrogen balance is evaluated (Fig. 1). The slope of the line before nitrogen balance reaches a plateau and the amount of dietary protein needed to attain zero nitrogen balance are indicators of protein quality; the steeper the slope and the lower the amount of dietary protein to achieve balance, the higher the quality of the protein being tested.

Influence of energy intake on nitrogen balance

When food energy intake is insufficient to satisfy energy needs, amino acid oxidation increases in an effort by the human body to satisfy energy requirements. This raises urinary nitrogen excretion and reduces nitrogen balance. However, increased energy intake may reduce amino acid oxidation and urinary nitrogen excretion, thereby improving N balance until it reaches a plateau. This response, known as the protein-sparing effect of dietary energy, can be attenuated if the quantity or quality of food protein intake is inadequate. It has been postulated that the protein-sparing effect of dietary carbohydrates is mediated by increased insulin secretion, which inhibits proteolysis, hepatic gluconeogenesis, and renal ammonia genesis. The protein-sparing effect of dietary fat may be due to a reduction of amino acid oxidation through an effect of free-fatty acid oxidation in the liver, whereby the increase in NADH/NAD inhibits branched-chained keto-acid dehydrogenase. For these reasons nitrogen

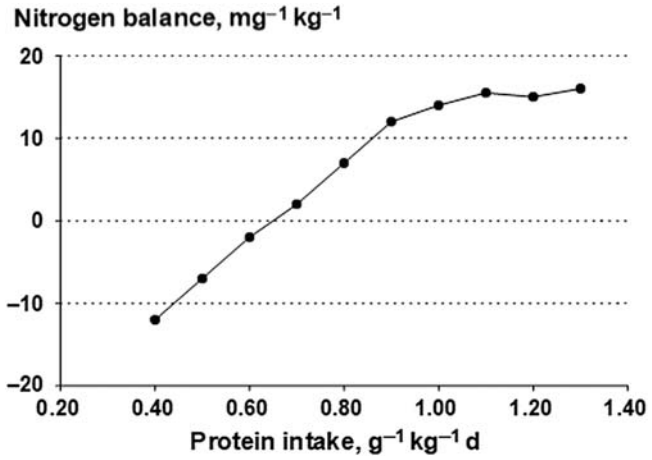


Fig. 1 Relation of N balance to protein intake.

balance must not be used to estimate protein quality when the amount of dietary energy is such that it produces weight loss or gain in an otherwise well-nourished individual. Because of their high cost and experimental complexity, metabolic studies are done mainly to evaluate new, nonconventional protein sources and novel food processes that may affect protein quality. Other methods that can predict protein quality for humans rapidly and at low cost are used to evaluate diets and conventional foods routinely.

Assays in laboratory animals

Biological assays in laboratory animals have been used to assess food protein quality, based either on a protein's ability to support growth in young rats (protein efficiency ratio (PER)) or on nitrogen retention (net protein utilization (NPU)). However, these assays often underestimate the quality of some vegetable and animal proteins for humans. For example, the proteins of pulses and milk casein have a lower quality for rats than for humans because rats have a higher requirement of sulfur-containing amino acids. Thus, application of rat assay for human nutrition can result in quantitative errors. The discrepancy usually has economic rather than public health implications because rat assays generally err by underestimating protein quality for humans, but the value of certain animal proteins can be overestimated because of higher efficiency of utilization by the rat. Nevertheless, the PER and NPU remain useful model for screening food protein quality and to validate theoretical models based on the amino acid composition of the protein in question.

Assessment of protein quality by amino acid scoring

Amino acid scores

The concept of assessing protein quality based on a protein's constituent amino acids was introduced in the late 1940s and was further discussed and developed (Lee et al., 2016; Shivakumar et al., 2020; WHO/FAO/UNU, 2007). Amino acid chemical scores are defined as the content of each of the 9 indispensable amino acid in the test protein over the content of the same amino acid in an indispensable amino acid scoring reference pattern.

Chemical score = EAA in the test protein/same EAA in the reference pattern.

It was later suggested that the calculations be corrected by the protein's digestibility. In the protein digestibility-corrected amino acid score (PDCAAS) used to assess protein quality, the amino acid score is corrected for true ileal or fecal nitrogen digestibility of the test protein. In the digestible indispensable amino acid score (DIAAS), more recently proposed, the amino acid score is corrected by true ileal digestibility of each indispensable amino acid (Lee et al., 2016; Tomé et al., 2014).

$$\text{PDCAAS} = \text{Chemical score x true faecal or ileal Protein digestibility}$$

$$\text{DIAAS} = \text{Chemical score x true ileal EAA digestibility}$$

Significant scientific and technological advancements now allow the use of an amino acid scoring procedure adjusted for digestibility as a good and practical predictor of protein quality for humans. This method is recommended by expert committees of WHO, FAO, UNU, and CCVP, as well as by regulatory agencies of several countries, for routine evaluation of protein quality for humans. The elements required for its application are knowledge about the amino acid composition and digestibility of the food protein(s) under evaluation and a scoring pattern based on human amino acid requirements.

Amino acid scoring reference pattern

Table 1 shows the internationally accepted patterns for amino acid scoring applicable to infants and to persons after 0.5 years of age (WHO/FAO/UNU, 2007; Shivakumar et al., 2020); the composition of high-quality animal foods is shown for comparison. For infants younger than 0.5 year, the scoring pattern is based on the amino acid composition of breast milk, even if some EAAs in human milk exceed minimum requirements for infants of this age. For example, infants consuming cow's milk proteins, which have less sulfur-containing amino acids than human milk, show adequate growth and nitrogen balance. Thus, although the use of a scoring pattern based on human milk composition may somewhat underestimate the composition of the maintenance requirement and of tissue deposition in children of different ages. This was because published nitrogen balance studies to determine the EAA requirement in older children and adults have experimental flaws. In adults, EAA requirements have now largely been determined by amino acid oxidation techniques, which are more accurate, and coupled with the adult protein requirement, allow for an adult scoring pattern to be established. Because proteins with amounts of EAAs that satisfy the needs of young children will probably be adequate for older children, the scoring pattern for preschool children is currently used for all children after 0.5 years of age. All EAAs present in proportions that exceed the value in the reference pattern are assigned a fractional score of 1.00 (or a percentage score of 100%), even if mathematical calculation gives a higher value. The EAA with the lowest value (i.e., the most limiting amino acid) determines the protein's amino acid score. The main EAAs that are likely to limit the protein

Table 1 Mean protein requirement and Amino acid scoring patterns for catch-up growth 0.5–4.9 year, infants under 0.5 year, preschool children, and adults (Shivakumar et al., 2020; WHO, 2007). Composition of animal proteins shown for comparison.

	Catch-up growth 0.5–4.9 y	*Infant 0–0.5 year*	*Preschool children 0.5–2 year*	*Adult*	*Egg, cow's milk, and beef protein*
			(g protein/day)		
Mean protein requirement	*2.82*	–	*0.86*	*0.66*	–
			Amino acid scoring pattern (mg amino acid per g protein)		
Histidine	24	20	18	15	22–34
Isoleucine	34	32	31	30	47–54
Leucine	70	66	63	59	81–95
Lysine	65	57	52	45	70–89
Methionine + cysteine	31	28	26	22	33–57
Phenylalanine + tyrosine	63	52	46	38	80–102
Threonine	36	31	27	23	44–47
Tryptophan	10	8.5	7.4	6	12–17
Valine	46	43	42	39	50–66

quality of mixed diets for humans are lysine, the sulfur-containing amino acids (methionine and cysteine), threonine, and tryptophan (WHO/FAO/UNU, 2007). Consequently, when information on all EAAs is not available, protein quality can be estimated based on its score for these four amino acids.

Correction for protein digestibility

A protein may have a good amino acid composition relative to the scoring pattern, but if it is not fully digested and its constituent amino acids are not absorbed, its capacity to provide nitrogen and EAAs for human function will diminish. Consequently, amino acid scores as predictors of protein quality must be adjusted for protein digestibility and amino acid availability. Protein and amino acid digestibility are a measure of the amount or proportion of dietary protein derived nitrogen and amino acids that are made available to the organism after digestion and absorption. Digestibility is determined by measuring the digestive losses of nitrogen or amino acids at the level of the terminal ileum (ileal digestibility) or in the feces (fecal digestibility). This apparent digestibility must be corrected by the amount of endogenous protein nitrogen and amino acid losses to calculate the true digestibility. As intestinal amino acid absorption is practically complete at the end of the small intestine, oro-ileal disappearance is considered to represent dietary protein-derived nitrogen and amino acid absorption more accurately than oro-faecal disappearance. The True Ileal Digestibility assay is the best currently available approach to assess nitrogen and amino acid absorption. Digestibility measurements at the ileal level may provide a better measure of amino acid digestibility, however this may pose significant challenges (Shivakumar et al., 2020). True Ileal amino acid digestibility is assessed by different invasive or minimally invasive procedures in human, or alternatively in animal pig or rat models (Devi et al., 2018; Fuller and Tomé, 2005; Gaudichon et al., 2002; Kashyap et al., 2018; Lee et al., 2016; Moehn et al., 2005; Shivakumar et al., 2020; Tomé et al., 2014). *In vitro* procedures have also been developed using combinations of trypsin, chymotrypsin, peptidase, and bacterial pro-tease. Further research is needed to validate their use as predictors of protein digestibility in humans.

Protein quality and dietary sources

Differences in protein content, amino acid profile and digestibility

Table 2 illustrates the protein concentration in different food protein sources. Animal product protein sources such as meat, milk, eggs, and some animal products are rich in protein with protein content of 30–70% dry weight. Legume seeds are a source of protein with protein content (dry weight) of 35–38% for soybeans and lupines, and 20–25% for peas and beans. Cereal seeds have a protein content of 15–20% dry weight. Among vegetables, pulses have the highest protein concentrations, ranging from 20% to 25% in most raw beans and peas to approximately 36% in soya beans. Most nuts and edible seeds contain 8–18% protein. Many oil seeds have 12–20% protein, and the cake that remains after oil extrusion can have as much as 30–40% protein. Evaluation of a food's protein concentration must be done for ready-to-eat preparations because food processing and cooking can result in significant changes relative to raw foods. Meats, poultry, and fish usually have a higher concentration of protein after cooking or frying, whereas vegetable food preparations contain more water and less protein than the raw products. Protein-dense foods are especially important for young infants, whose small gastric capacity limits the amount they can eat, and for older people with poor appetite.

Table 2 Protein content of various animal and plant products.

	Average protein content per food family	
	% Wet matter	*% Dry matter*
Animal product		
Meats, poultry	30	50–70
Fishes and amphibians	25	50–70
Insects	30	40–60
Cold cuts and cured meats	16	30–60
Eggs and derivatives	12	40–50
Cheeses	20	30–50
Milk and dairy products	3–15	20–50
Plant product		
Oil seeds and chestnuts	18	25–35
Legume seeds	9	20–25
Cereals and pasta	8	15–20
Potatoes and related products	3	–
Legumes	2	–
Fruits	1	–

Protein concentration in raw and ready-to-eat foods (g protein per 100 g wet food)		
	Raw	*Ready-to-eat*
Beef, lean	21.4	36.8 (cooked)
Fish	20.0	31.8 (fried)
Wheat flour	11.0	12.0 (white bread)
Egg, hen	11.3	11.3 (hard-boiled)
Lentils	23.7	7.1 (cooked)
Common beans	22.0	6.2 (cooked)
Maize	9.40	4.2 (tortilla)
Milk powder, cow	26.1	3.2 (12% in water)
Rice	7.20	2.5 (boiled)
Potato, no skin	1.80	1.1 (cooked)

Not all food proteins are digested, absorbed, and utilized to the same extent because of inherent differences in their source (e.g., inside vegetable cells with indigestible membranes), their physicochemical nature (e.g., protein configuration and amino acid binding), the presence of food constituents that modify digestion (e.g., dietary fiber, tannins, and other poly-phenols), the presence of antiphysiological factors that interfere with protein breakdown (e.g., trypsin inhibitors and lectins), and processing conditions that alter the nature or release of amino acids (e.g., Maillard reaction and formation of polyamino acids and methylmercaptan). Protein nitrogen digestibility values and more recently ileal amino acid digestibility values of specific foods and well-defined diets may be taken from reliable published data or must be determined, preferably in humans. When cost and practicality do not permit metabolic studies in humans to be performed, standardized methods in animal models are used (Shivakumar et al., 2020). Nevertheless, animal data must be used with caution for foods and diets that are known or suspected of being handled differently by the human and animal intestines. When data are not available for a mixed diet, a weighted average can be calculated from the true digestibility of its constituent protein sources.

Examples for PDCAAS and DIAAS values of different food protein are reported in **Table 3**. Foods of animal origin, such as milk and milk products, eggs, meats, poultry, and fish, have excellent amino acid composition, true protein or amino acid digestibility of 95–98%, and a score of 100%. In addition, their protein concentrations often increase after cooking. Almost all vegetable foods have one or more limiting amino acids. In general, proteins in natural vegetable foods have digestibility of 70–85%. Vegetable protein isolates, flours, and extruded products have higher digestibility. Pulses usually have limiting sulfur-containing amino acids. Cereals and cereal products are the largest sources of protein in most areas of the world. Cereal grains and flours contain approximately 7–12% protein with a quality that is limited by their lysine content and, in many instances, also by threonine or tryptophan. Although deficient in lysine and threonine, rice has one of the best amino acid compositions among cereals, whereas sorghum and native maize (i.e., not genetically improved) are among the lowest.

Improvement of protein concentration, amino acid profile and digestibility

Protein concentration can increase by genetic selection of protein sources, as in improved varieties of rice that have approximately 30% more protein than native rice, using nitrogen-concentrating fertilizers that can raise the protein contents of several cereals or by

Table 3 Examples of lowest non-truncated PDCAAS and DIAAS values calculated for adult for various animal and plant protein sources.

		PDCAAS		*DIAAS*	
Protein sources	*Limiting amino acid*	*Non-truncated*	*Truncated*	*Non-truncated*	*Truncated*
Animal product					
Whey protein isolate	His	97–112	97–100	107–109	100
Whey protein concentrate	His	99–107	99–100	97–107	97–100
Skimmed milk powder	SAA	112	100	105	100
Beef meat	Leu	114	100	112	100
Egg	His	115	100	113	100
Whole milk powder	SAA	116	100	116	100
Milk protein concentrate	SAA	121–125	100	118–120	100
Average animal	–	115	100	114	100
Plant product					
Corn grain, silage	Lys	47	47	42	42
Wheat	Lys	46–51	46–51	40–45	40–45
Sunflower cake	Lys	54	54	46	46
Triticale	Lys	55	55	50	50
Sunflower expeller	Lys	58	58	49	49
Rice	Lys	59	59	48–56	48–56
Barley	Lys	59	59	47	47
Wheat bran	Lys	60	60	49	49
Pea protein concentrate	SAA	71–86	71–86	62–82	62–82
Peas	SAA	78	78	65	65
Soya protein isolate	SAA	86–95	86–95	84–90	84–90
Rapessed expeller, cake	Ileu	92	92	70	70
Soya flour	SAA	93	93	89	89
Soyabean cake	SAA	99	99	97	97
Soyabean expeller	SAA	102	100	100	100
Soybean	SAA	102	100	100	100
Average plant	–	70	70	61	61

All EAAs present in proportions that exceed the value in the reference pattern are assigned a fractional score truncated to 1.00 (or a percentage score of 100%), even if mathematical calculation gives a higher value (non-truncated score). The EAA with the lowest value (i.e., the most limiting amino acid) determines the protein's amino acid score.
His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; SAA, sulfur-containing amino acids; PDCAAS, protein digestibility-corrected amino acid score; DIAAS, digestible indispensable amino acid score; EAA, essential amino acid.
Values from Ertl et al. (2016), Mathai et al. (2017), Sosulski and Sarwar (1973), Rutherfurd et al. (2015), Herreman et al. (2020).

industrial and home processing that reduce the water content of food preparations. Protein fortification by addition of concentrated protein sources, such as casein, soya protein isolates, soya flour, milk powder, or dehydrated egg, will also increase the protein concentration of foods and diets, as well as their amino acid score in some instances.

The amino acid profile of a food or diet can be improved by increasing the amount of constituent amino acids in its proteins, adding specific amino acids, or combining foods in proportions that result in a better amino acid pattern. In cereals genetic handling resulted in higher contents of the amino acids that limit their protein quality. For example, varieties of Opaque-2 corn have approximately 50% more lysine and 35% more tryptophan than native corn, both of which are limiting amino acids in this cereal. The addition of synthetic amino acids by fortification and enrichment eliminates or reduces the magnitude of limiting amino acids, for example, in lysine-enriched wheat flour. For complementation, the combination of a food that has one or more limiting EAAs with another food(s) that has a surplus of these amino acids results in an improved combined amino acid profile. A double complementation effect has been achieved in the formulation of vegetable mixtures based on protein sources in which one has a surplus of the EAA that is limiting in the other and vice versa (**Fig. 2**).

Various food processing procedures can improve protein digestibility by removing food constituents that reduce digestibility (such as dietary fiber), breaking down poorly digestible vegetable cell membranes, destroying or neutralizing antiphysiological factors, and increasing the food surface area in contact with gastrointestinal enzymes. For example, soya protein isolate, polished rice, and refined wheat flour have higher protein digestibility than soya flour, whole rice, and whole wheat, respectively. Food storage and processing in adverse circumstances can reduce protein quality by making some EAAs unavailable for use in the human body. These conditions should be avoided to preserve protein quality. Some examples are the storage of dried milk under mild to moderate heat and humidity, which renders lysine side chains unavailable after reacting with the reducing sugar, lactose (Maillard or "browning" reaction); the severe treatment of protein with alkali, which causes lysine and cysteine residues to react and form lysinoalanine; and the treatment of proteins with oxidizing agents, which can result in a loss of methionine. Severe heating conditions in the presence of reducing sugars or oxidized lipids can make some food proteins resistant to digestion, thereby reducing the availability of all their amino acids.

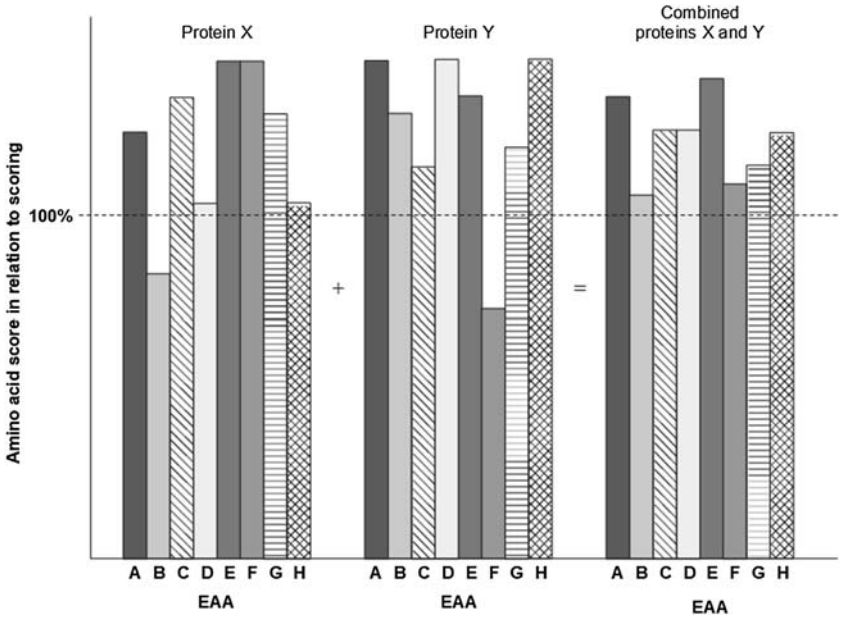


Fig. 2 The effect of combining proteins on the amino acid score of the dietary protein intake.

Conclusion

Different methods and approach have been used to assess the nutritive value of protein from food and diet. Significant scientific and technological advancements and development make the amino acid scoring procedure adjusted for digestibility as an accurate predictor of protein quality for humans. This method is recommended by most expert committees and regulatory agencies.

References

Devi, S., Varkey, A., Sheshshayee, M., Preston, T., Kurpad, A.V., 2018. Measurement of protein digestibility in humans by a dual-tracer method. Am. J. Clin. Nutr. 107, 984–991.
Ertl, P., Knaus, W., Zollitsch, W., 2016. An approach to including protein quality when assessing the net contribution of livestock to human food supply. Animal 10, 1883–1889.
Fuller, M.F., Tomé, D., 2005. In vivo determination of amino acid bioavailability in humans and model animals. J. AOAC Int. 88, 923–934.
Gaudichon, C., Bos, C., Morens, C., Petzke, K.J., Mariotti, F., Everwand, J., Benamouzig, R., Daré, S., Tomé, D., Metges, C.C., 2002. Ileal losses of nitrogen and amino acids in humans and their importance to the assessment of amino acid requirements. Gastroenterology 123, 50–59.
Herreman, L., Nommensen, P., Pennings, B., Laus, M.C., 2020. Comprehensive overview of the quality of plant- and animal-sourced proteins based on the digestible indispensable amino acid score. Food Sci. Nutr. 8, 5379–5391.
Kashyap, S., Shivakumar, N., Varkey, A., Duraisamy, R., Thomas, T., Preston, T., Devi, S., Kurpad, A.V., 2018. Ileal digestibility of intrinsically labeled hen's egg and meat protein determined with the dual stable isotope tracer method in Indian adults. Am. J. Clin. Nutr. 108, 980–987.
Lee, W.T., Weisell, R., Albert, J., Tomé, D., Kurpad, A.V., Uauy, R., 2016. Research approaches and methods for evaluating the protein quality of human foods proposed by an FAO expert working group in 2014. J. Nutr. 146, 929–932.
Mathai, J.K., Liu, Y., Stein, H.H., 2017. Values for digestible indispensable amino acid scores (DIAAS) for some dairy and plant proteins may better describe protein quality than values calculated using the concept for protein digestibility-corrected amino acid scores (PDCAAS). Br. J. Nutr. 117, 490–499.
Moehn, S., Bertolo, R.F., Pencharz, P.B., Ball, R.O., 2005. Development of the indicator amino acid oxidation technique to determine the availability of amino acids from dietary protein in pigs. J. Nutr. 135, 2866–2870.
Rutherfurd, S.M., Fanning, A.C., Miller, B.J., Moughan, P.J., 2015. Protein digestibility-corrected amino acid scores and digestible indispensable amino acid scores differentially describe protein quality in growing male rats. J. Nutr. 145, 372–379.
Shivakumar, N., Jackson, A.A., Courtney-Martin, G., Elango, R., Ghosh, S., Hodgkinson, S., Xipsiti, M., Lee, W.T., Kurpad, A.V., Tomé, D., 2020. Protein quality assessment of follow-up formula for young children and ready-to-use therapeutic foods: recommendations by the FAO Expert Working Group in 2017. J. Nutr. 150, 195–201.
Sosulski, F.W., Sarwar, G., 1973. Amino acid composition of oilseed meals and protein isolates. Can. Inst. Food Sci. Technol. J. 6, 1–5.
Tomé, D., Jahoor, F., Kurpad, A., Michaelsen, K., Pencharz, P., Slater, C., Weisell, R., 2014. Current issues in determining dietary protein quality and metabolic utilization. Eur. J. Clin. Nutr. 68, 537–538.
Tome, D., Cordella, C., Dib, O., Peron, C., 2019. Nitrogen and Protein Content Measurement and Nitrogen to Protein Conversion Factors for Dairy and Soy Protein-Based Foods: A Systematic Review and Modeling Analysis. World Health Organization, Geneva.
WHO/FAO/UNU, 2007. Protein and Amino Acid Requirements in Human Nutrition. World Health Organization.

Further reading

Bailey, H.M., Mathai, J.K., Berg, E.P., Stein, H.H., 2020. Most meat products have digestible indispensable amino acid scores that are greater than 100, but processing may increase or reduce protein quality. Br. J. Nutr. 124, 14–22.

Bodwell, C., Carpenter, K., Mcdonough, F., 1989. A collaborative study of methods of protein evaluation: introductory paper. Plant Foods Hum. Nutr. 39, 3–11.
Bornhorst, G.M., Paul Singh, R., 2014. Gastric digestion in vivo and in vitro: how the structural aspects of food influence the digestion process. Annu. Rev. Food Sci. Technol. 5, 111–132.
Bos, C., Juillet, B., Fouillet, H., Turlan, L., Daré, S., Luengo, C., N'tounda, R., Benamouzig, R., Gausserès, N., Tomé, D., Gaudichon, C., 2005. Postprandial metabolic utilization of wheat protein in humans. Am. J. Clin. Nutr. 81, 87–94.
Calvez, J., Benoit, S., Piedcoq, J., Khodorova, N., Azzout-Marniche, D., Tomé, D., Benamouzig, R., Airinei, G., Gaudichon, C., 2020. Very low ileal nitrogen and amino acid digestibility of zein compared to whey protein isolate in healthy volunteers. Am. J. Clin. Nutr. nqaa274.
Ghosh, S., Suri, D., Uauy, R., 2012. Assessment of protein adequacy in developing countries: quality matters. Br. J. Nutr. 108, S77–S87.
Ghosh, S., 2013. Assessment of protein adequacy in developing countries: quality matters. Food Nutr. Bull. 34, 244–246.
Gilani, G.S., Xiao, C.W., Cockell, K.A., 2012. Impact of antinutritional factors in food proteins on the digestibility of protein and the bioavailability of amino acids and on protein quality. Br. J. Nutr. 108, S315–S332.
Gilani, G.S., 1997. The protein digestibility–corrected amino acid score method overestimates quality of proteins containing antinutritional factors and of poorly digestible proteins supplemented with limiting amino acids in rats. J. Nutr. 127, 758–764.
Guerra, A., Etienne-Mesmin, L., Livrelli, V., Denis, S., Blanquet-Diot, S., Alric, M., 2012. Relevance and challenges in modeling human gastric and small intestinal digestion. Trends Biotechnol. 30, 591–600.
Hellström, P.M., Grybäck, P., Jacobsson, H., 2006. The physiology of gastric emptying. Best Pract. Res. Clin. Anaesthesiol. 20, 397–407.
Rand, W.M., Pellett, P.L., Young, V.R., 2003. Meta-analysis of nitrogen balance studies for estimating protein requirements in healthy adults. Am. J. Clin. Nutr. 77, 109–127.
Reynaud, Y., Buffière, C., Cohade, B., Vauris, M., Liebermann, K., Hafnaoui, N., Lopez, M., Souchon, I., Dupont, D., Rémond, D., 2021. True ileal amino acid digestibility and digestible indispensable amino acid scores (DIAASs) of plant-based protein foods. Food Chem. 338, 128020.
Sanchón, J., Fernández-Tomé, S., Miralles, B., Hernández-Ledesma, B., Tomé, D., Gaudichon, C., Recio, I., 2018. Protein degradation and peptide release from milk proteins in human jejunum. Comparison with in vitro gastrointestinal simulation. Food Chem. 239, 486–494.
Santos-Hernández, M., Alfieri, F., Gallo, V., Miralles, B., Masi, P., Romano, A., Ferranti, P., Recio, I., 2020. Compared digestibility of plant protein isolates by using the INFOGEST digestion protocol. Food Res. Int. 137, 109708.
Tessier, R., Khodorova, N., Calvez, J., Kapel, R., Quinsac, A., Piedcoq, J., Tomé, D., Gaudichon, C., 2020. 15N and ^{2}H intrinsic labeling demonstrate that real digestibility in rats of proteins and amino acids from sunflower protein isolate is almost as high as that of goat whey. J. Nutr. 150, 450–457.
Tome, D., 2012. Criteria and markers for protein quality assessment—a review. Br. J. Nutr. 108, S222–S229.
WHO, 1991. Protein Quality Evaluation: Report of the Joint FAO/WHO Expert Consultation. Food & Agriculture Org, Bethesda, Md. USA 4–8 December 1989.
Wu, G., Fanzo, J., Miller, D.D., Pingali, P., Post, M., Steiner, J.L., Thalacker-Mercer, A.E., 2014. Production and supply of high-quality food protein for human consumption: sustainability, challenges, and innovations. Ann. N. Y. Acad. Sci. 1321, 1–19.

Protein: Requirements and role in diet

D.J. Millward, University of Surrey, Guildford, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of D.J. Millward, Protein: Requirements and Role in Diet, In Encyclopedia of Human Nutrition (Third Edition), edited by Benjamin Caballero, Academic Press, 2013, https://doi.org/10.1016/B978-0-12-375083-9.00242-7.

Introduction 569
Terminology 569
Metabolic demands for amino-acids: the metabolic model 570
Obligatory and adaptive components of the metabolic demand (see) 570
Nitrogen balance studies in practice 572
Inherent difficulties with nitrogen balance studies 573
Energy balance and N balance 573
Adaptation to changes in protein intake 573
Adaptation and the efficiency of protein utilization 574
Interpretation of N balance studies in the current WHO recommendations 574
Dietary protein allowances for individuals and populations 574
Protein requirements for growth and special needs 575
Pregnancy requirements 576
Areas of uncertainty 576
Requirements for infants and children, pregnancy and lactation 576
The amino-acid pattern of the metabolic demand: plant versus animal sources 576
Conclusions: implications of adaptation for nutrition policy 577
References 578
Further Reading 578

Key points

- Dietary protein intakes can vary markedly throughout the life cycle within healthy diets according to the balance between plant-based or animal-source foods so that identifying the lower limit of the healthy range is challenging.
- The metabolic model for the protein requirement identifies the cellular metabolic demand, which includes obligatory and adaptive components. The dietary requirement is an intake which meets the demand after adjusting for digestibility and amino acid-dependent biological value. Dietary allowances are set at the upper end of the range of interindividual variation.
- Obligatory demands comprise amino acids used for synthesis of various nitrogenous metabolites, for net synthesis of proteins lost from the body and in children, pregnancy and lactation, the net protein deposition during tissue and fetal growth and breast-milk secretion.
- Adaptive components of the metabolic demand comprise amino acid oxidation at a rate set by the habitual protein intake which occurs to minimize increases in the potentially toxic branched-chain, aromatic and sulfur amino acids.
- The minimum protein requirement, defined as the lowest intake that will maintain nitrogen balance in healthy adults and provide for special needs of children or pregnant or lactating women, derive from a meta analysis of nitrogen balance studies in adult men and women, in which the median intake for nitrogen equilibrium in individual subjects, was independent of age gender or dietary protein composition. The range of overall variation was large, possibly due to non-intended variation in energy intakes between studies and in incomplete adaptation to the test diets.
- A factorial model of requirements in children, pregnancy and lactation assumed similar obligatory demands per kg body weight as the adult value plus special needs adjusted for an assumed efficiency of dietary protein utilization deriving from nitrogen balance studies. The efficiency values used are low and arguably underestimates of the true biological values so that intakes for special needs are likely to be overestimated, markedly so in pregnancy. Thus recommended intakes for pregnant women may be associated with health risks.
- The metabolic demand includes requirements for indispensable amino acids and dispensable amino acid nitrogen and can be met by mixtures of plant source proteins as well as animal source proteins. The amino acid profile of many traditional and novel plant proteins satisfies the amino acid requirement profile, with the main difference between animal and plant dietary protein sources reflecting a lower digestibility.

Introduction

Discussions in the area of nutritional recommendations are seldom without some degree of controversy and protein is an exemplar of the difficulties and debates which can occur. The amount and type of dietary protein needed to allow optimal genomic expression during growth and development is poorly understood. There are virtually no effective markers of protein deficiency in adults or children, and clinical manifestations long thought to reflect a protein-deficient diet, such as edematous infantile malnutrition, are now generally recognized as reflecting the combined effects of infection and micronutrient-deficiency. At the other end of the spectrum of public health issues, within health clubs worldwide, more money is spent on protein and amino-acid supplements than on any other ergogenic aid, even though the effectiveness of such expense is poorly documented to say the least.

Values for protein and amino-acid requirements derive from nitrogen (N) balance and growth studies which began early in the 20th century and identified estimates of dietary requirements and the pattern of amino-acid essentiality. From the 1970s research expanded to include stable isotope tracer studies of human protein and amino-acid metabolism and turnover. However the 2007 expert report from WHO on human protein and amino-acid requirements (World Health Organization, 2007) concluded that, of the methodologies available, N balance remains the only available method for estimating the requirement for total protein (N), and no current method is entirely reliable for determining the dietary requirement for indispensable amino-acids (IAAs). However because of these limitations, the latest recommendations for protein and amino-acid requirements are not entirely satisfactory, yet their limitations are by no means widely recognized.

It is the case, as indicated in Fig. 1, that dietary protein intakes can vary markedly even within diets which would be judged "healthy", and between population groups who choose plant-based rather than animal-source foods. In fact given $\geq$3 fold variation in dietary protein content indicated in Fig. 1, overall protein intakes can vary over a 5 fold range if variation in lifestyle and food energy intake is taken into account in ostensibly healthy individuals. The challenge is to identify the intake at the lower end of this range which satisfies the criteria of a requirement value.

In general terms the dietary protein requirement is the amount of protein in an otherwise nutritionally adequate diet, necessary to maintain the desired structure and function of the organism and to provide for any special needs. In a 2007 WHO report the minimum protein requirement, (MPR), was defined as: the lowest level of dietary protein intake that will balance the losses of nitrogen from the body, and thus maintain the body protein mass, in persons at energy balance with modest levels of physical activity, plus, in children or in pregnant or lactating women, the needs associated with the deposition of tissues or the secretion of milk at rates consistent with good health. This has the benefit of being conceptually straightforward with N-balance as its main focus. While several stable-isotopic approaches based on different paradigms have been described no consensus has been reached on any specific alternative approach, with one approach identified as deeply flawed (Millward and Jackson, 2012). However the nitrogen balance approach is beset by major design and practical difficulties of which the problem of metabolic adaptation is especially difficult as will be described here.

Terminology

Protein requirements can be discussed in terms of metabolic demand, dietary requirement, and dietary allowances. Metabolic demand is the cellular requirement for amino-acids of a pattern which matches their metabolic consumption in metabolic pathways which vary with the phenotype and the developmental and physiological state of the individual. The dietary protein requirement is the amount of protein that must be consumed in order to satisfy the metabolic demand. If digestion and absorption is incomplete (with dietary N lost in the feces), the digestibility is less than 100%, and if the absorbed amino-acid pattern does not exactly match cellular needs (with some absorbed amino-acid N lost as urea), the biological value is less than 100%. In each case the efficiency of

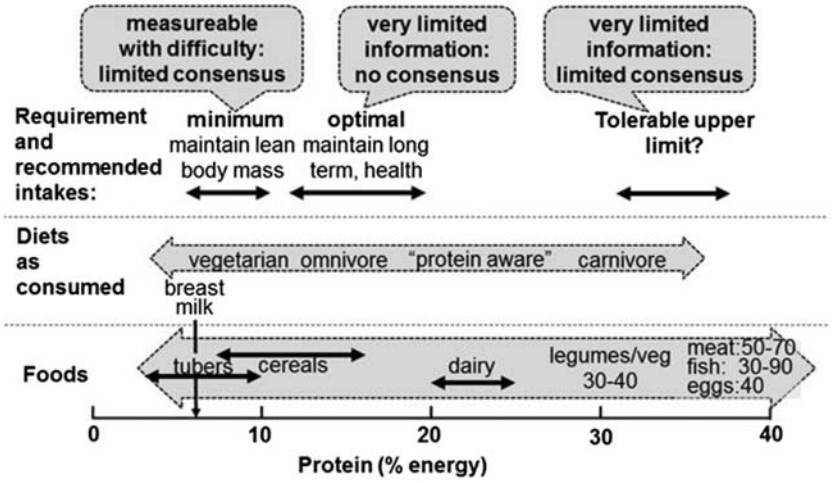


Fig. 1 Relationships between protein content of foods, dietary intakes and current status of knowledge about requirements (see Millward, 2013).

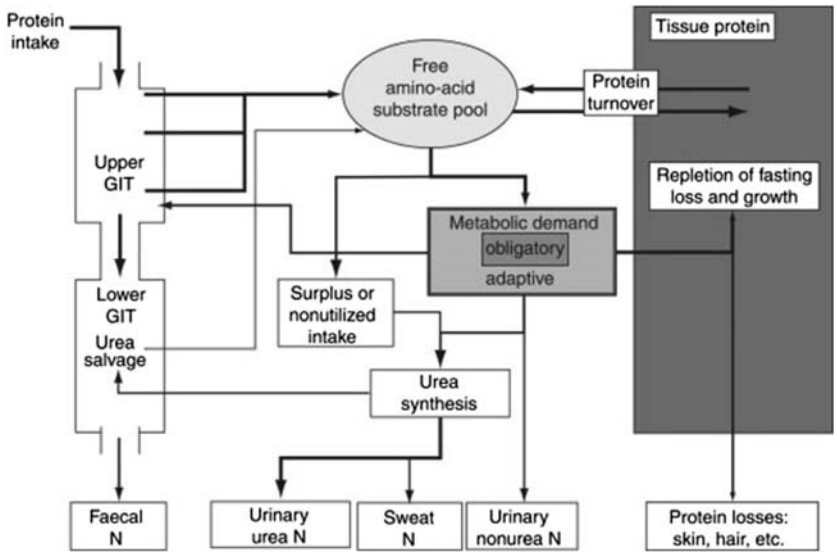


Fig. 2 Schematic representation of the metabolic demands for amino-acids.

utilization will be less than 100% and the dietary protein requirement will be greater than the metabolic demand. Dietary allowances are a range of intakes derived from estimates of individual requirements taking into account variability between individuals. They are designed to meet the dietary requirements of almost all individuals or the majority of the population. The most recent values of safe intakes for individuals and populations are those published by (WHO) (World Health Organization, 2007) values which have been reproduced in the 2012 EFSA report (European Food Safety Authority, 2012) (although the terminology varies).

Metabolic demands for amino-acids: the metabolic model

Current evidence supports the representation of the metabolic demands as in Fig. 2. The metabolic demand for amino-acids is to maintain tissue protein at appropriate levels and to provide for all amino-acid derived metabolites and any additional needs during growth, rehabilitation, pregnancy, and lactation. The ~20 K proteins encoded in the human genome are diverse, including structural or fibrous insoluble types and soluble globular species, with characteristic properties and functions that are determined by their amino-acid sequence. All proteins are in a dynamic state of constant turnover (i.e., breakdown to constituent amino-acids and resynthesis), although for the structural proteins this is slow or minimal. Nonprotein products include nucleic acids and a diverse range of smaller molecules, such as creatine, taurine, glutathione, hormones (e.g., catecholamines and thyroxine), neurotransmitters (serotonin and dopamine), and nitric oxide, a key regulator of blood flow and other physiological processes.

The metabolic demand is supplied from the free amino-acid pool, the size of which, for most amino-acids, is regulated within narrow limits. This regulation involves supply from three sources: dietary proteins after digestion and absorption from the upper gastrointestinal tract (GIT), tissue protein after proteolysis during protein turnover, and urea salvage in the large bowel, (discussed below). Within the free amino-acid pool there are also interconversions of several amino-acids, especially during transfer between organs and in the course of the absorption of dietary amino-acids and ammonia. Removal of free amino-acids occurs by reactions in which they act as substrates, and these reactions are shown as three pathways, one of which is the metabolic demand. This pathway involves a number of irreversible pathways, including net protein synthesis, and other irreversible metabolic transformations of individual amino-acids with the N ultimately excreted as urea and other end products. The second and quantitatively largest pathway is the removal for protein synthesis during protein turnover. However at N equilibrium, the removal of amino-acids for synthesis is balanced by their replacement through proteolysis so turnover does not exert a significant net metabolic demand. The third pathway is the irreversible removal of amino-acids by oxidation and nitrogen excretion provoked, for example, by the transient increases in some or all free amino-acids after a protein meal. This would represent an inefficient utilization. Terminal removal of N by urea synthesis is not entirely irreversible because of urea salvage in the lower GIT. Thus, the rate of urea synthesis is usually in excess of the rate of urea excretion because some urea enters the large bowel and is hydrolyzed by bacteria. Most of this nitrogen is utilized by bacteria, and since little is lost as fecal N, it is eventually returned to the systemic pool as mainly ammonia, which can contribute to the *de novo* synthesis of dispensable amino-acids, and possibly some amino-acids after bacterial death and proteolysis.

Obligatory and adaptive components of the metabolic demand (see Millward, 2003)

The obligatory component of the demand for amino-acids for subjects at N equilibrium, (i.e., at maintenance), comprises conversion of some individual amino-acids into important metabolites that are further transformed into nitrogenous end products, mainly urea and other compounds in urine, feces, or sweat, as well as net synthesis of proteins lost from the body as skin, hair, and any other secretions including that within the upper GIT involving mucin proteins lost into the large bowel. These diverse

biological demands for amino-acids for maintenance represent an essential but small intrinsic metabolic demand for an equivalent amount of protein, assumed empirically to be equal to the obligatory nitrogen loss (ONL). This is the sum of all N losses from the body observed in subjects fed a protein-free but otherwise nutritionally adequate diet after 7–14 days, by which time nitrogen losses have declined to a stable and reproducible low level with subjects losing body protein at a constant daily rate. In normal adults, the daily obligatory urinary, fecal, and subcutaneous and other N losses are approximately 32, 10.5, and 4.8 mg N kg^{-1}, respectively, i.e., 47.1 mg kg^{-1} d^{-1}, in total equivalent to about 0.3 g protein kg^{-1} d^{-1} tissue protein mobilized to meet such demands. The ONL is a function of body weight, is assumed not to vary with age, and can be assumed to derive from net tissue proteolysis at a rate determined by the metabolic consumption of the rate-limiting amino-acid, the amino-acid with the highest ratio of molar proportion in the metabolic demand to molar proportion in protein. In fact because the amino-acid pattern of the obligatory metabolic demand is unlikely to match that of tissue protein, the N content of the metabolic demand is likely to be less than the ONL because only some of tissue amino-acids mobilized to provide for the metabolic demand will serve useful functions with the remainder oxidized and their N excreted. Thus the ONL falls in response to feeding selective amino-acids, such as threonine, tryptophan, and methionine, and it is generally assumed that methionine is the rate-limiting amino-acid driving the ONL. In addition to these metabolic demands for maintenance, any net protein synthesis associated with growth, pregnancy, and lactation also constitutes an obligatory metabolic demand.

The adaptive component of the metabolic demand represents amino-acid oxidation at a rate varying with the habitual protein intake that occurs as a result of the adaptive changes in capacity and activity of the pathways of oxidation of amino-acids that regulate free amino-acid pool sizes. Although this aspect of amino-acid metabolism is least understood, it is likely to occur as a consequence of the fact that humans grow slowly or maintain constant weight on diets that contain protein considerably in excess of minimum needs. Thus, to be able to rapidly dispose of excess protein and maintain the very low tissue concentrations of those amino-acids that may be toxic at higher concentrations, (the branched chain, aromatic, and sulfur amino-acids), pathways of oxidative amino-acid catabolism adapt (change their V_{max}), enabling them to operate at the appropriate rate set by habitual protein intakes. Importantly, this adapted rate of amino-acid oxidation changes only slowly in response to either a change in dietary protein intake level or to feeding and fasting. This has two main consequences.

First, when intake falls below habitual intake mobilization of tissue protein occurs with a negative N balance for as long as it takes to adapt to the lower level of intake. This was previously identified as the labile protein reserves. It can be assumed that for intakes greater than the minimum protein requirement (MPR), full adaptation to the new level will include not only a change in the adaptive metabolic demand to match intake but also repletion of tissue N lost during the adaptive transition, i.e., a period of positive N balance.

Second, because the adaptive rate of amino-acid oxidation continues to some extent into the postabsorptive state, there are varying postabsorptive losses of tissue protein and nitrogen excretion with varying habitual intake—that is, a diurnal cycle of postabsorptive losses and postprandial gains with an amplitude that changes with changes in the habitual level of protein intake as shown in Fig. 3. As such, the adaptive metabolic demand includes a component of net protein synthesis that repletes postabsorptive losses. The magnitude of this varies in a complex way with eating pattern and with the amount and quality (amino-acid score) of the habitual protein intake.

There are several practical implications of the adaptive metabolic demands model.

1. The true MPR will be that associated with the lowest possible adaptive metabolic demand. However, it is not known with any certainty how long such adaptation would take. Evidence suggests it is likely to be considerably longer than the two-week periods employed in short-term balance studies. This implies our current estimate of the MPR, from short-term balances, is likely to be an overestimate and that some of the variability in protein requirements between studies may reflect variable completeness of adaptation to the test diets.

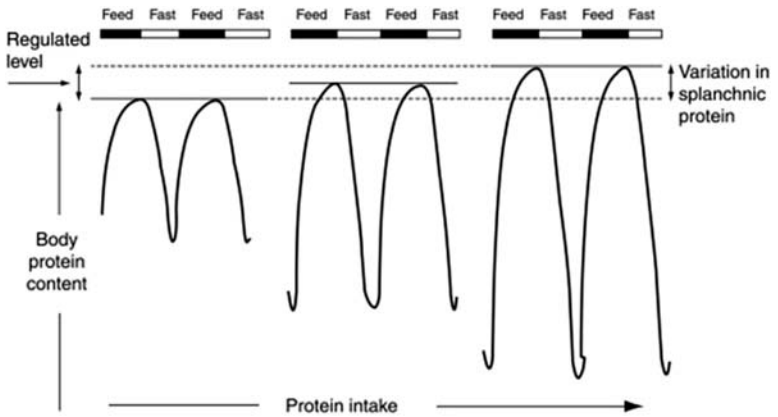


Fig. 3 Schematic representation of balance regulation throughout the diurnal cycle with increasing habitual protein intakes (see Millward, 1995).

2. Within an adaptive demand model, intakes and requirements are correlated, which has implications for the definition of risk of deficiency and safe intakes.
3. In N-balance studies the slope of the regression of N-intake on balance will underestimate the efficiency of protein utilization. This is discussed further in the next section.

Nitrogen balance studies in practice

N balance studies have been central to the definition of protein requirements within a model of nutrient balance: i.e., an intake which balances all losses observed in a healthy weight-stable subject. N is the proxy measure of protein intake and losses, and health is identified only in terms of an appropriate lean body mass at a healthy body weight. Thus, the MPR, is determined as the intake for N equilibrium, (zero balance), as indicated in N balance studies. There are no other measurable physiological or biochemical indicators of health which are exclusively and unequivocally indicators of protein intake and no disease states which have been identified as having a quantifiable relationship with protein intakes.

The aim of N balance studies is simple—to define the relationship between intake and all losses (urinary, fecal, and surface—mainly sweat, skin, hair, breath ammonia, nail clippings, etc.) in subjects fed 3 or preferably more intakes, usually at or below the expected intake for equilibrium. This allows the intake (I) which matches all losses (L) and maintains N equilibrium to be identified as the requirement (R). Thus,

$$\text{when } I = R, \quad \text{balance} = I - L = 0.$$

Subjects adapted to a protein-free diet excrete the ONL which equates to a daily loss of about 0.3 g protein/kg/day. On refeeding with protein, balance improves as the dietary protein provides for some of the metabolic demand, but N losses also increase so that the required intake for balance is more than the ONL. This is a crucial part of the problem of interpretation of N-balance studies. The increasing N losses with intake have been assumed to reflect an inefficiency of utilization of dietary protein to meet the metabolic demand, (quantified by the slope of the regression of intake on balance). However, such increased losses with intake would be expected as part of the adaptive metabolic demand which increases with intake, and these cannot be identified separately in N-balance studies from losses due to an actual inefficiency of intake, so the slope underestimates the true efficiency of utilization to an unknowable extent.

The 2007 WHO report adopted a meta-analysis of acceptable N balance studies in adults aimed at estimating both the ONL and the requirement (Rand et al., 2003). All the reported individual balance points from this meta analysis and the analytical principle is shown in Fig. 4. In this figure the protein intake which will balance all losses and produce N equilibrium, the maintenance requirement is identified by linear regression of balance against intake, as intercept/slope. In practice the currently accepted MPR for adults, (0.66 g/kg/day), is the median value of all individual subject linear regressions of the three or more levels of protein intake. Although the majority of the studies were young men studied with animal proteins, some studies with mixed plant proteins,

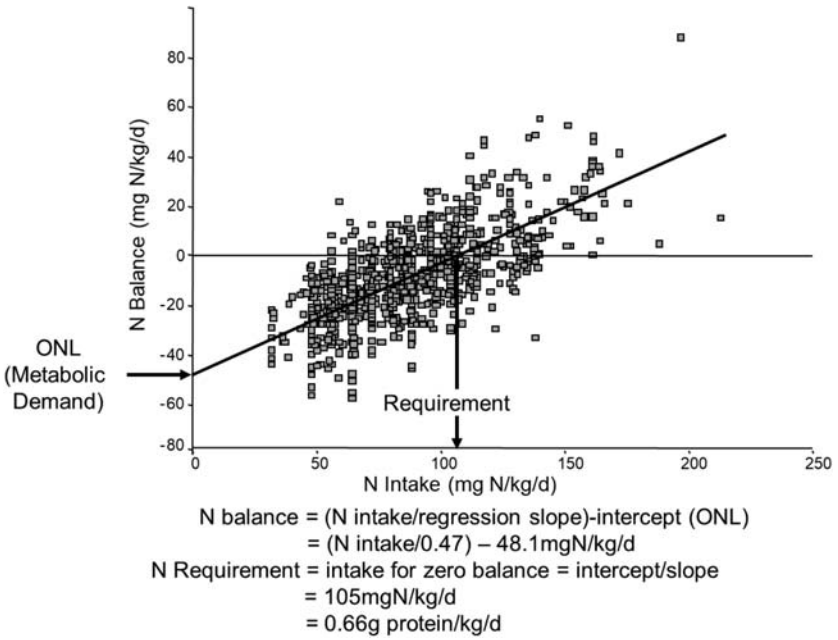


Fig. 4 Illustration of the results obtained and their analysis in a meta-analysis of N balance studies. Data from Rand et al. (2003).

Table 1 Potential problems relating to N balance methodology (see Millward, 2001).

Imprecise: balance is small difference between two much larger amounts
Systematic errors: intake overestimated, loss underestimated due to difficulty of accounting for all losses: e.g.,
Skin surface and secretions
Loss of nitrogen as gas
Expired ammonia
Endogenous NO production gives urinary nitrate, fecal ammonia, and nitrite (not always measured)
Changing size of the body urea pool
Design: balance influenced by
Dietary energy intake and physical activity
Extent of adaptation of amino-acid oxidation
Analysis: linear regression slope markedly underestimates efficiency of utilization

with women and with older adults were included. However no significant influences of gender, age or dietary protein source were observed. Subsequent to this meta-analysis, N balance studies examining age and gender specifically have confirmed that these factors do not influence the MPR.

Inherent difficulties with nitrogen balance studies

This apparently simple but laborious approach is beset with a large number of quite serious problems, as listed in **Table 1** (see Millward, 2001).

The lack of precision results in large overall errors and systematic errors appear to result in an overestimate of balance with unrealistic positive balances (protein gains) at high intakes as is apparent in **Fig. 4**. Of the various difficulties relating to the design and interpretation of N-balance studies the most important are (a) ensuring subjects are at energy balance, (b) allowing sufficient time for adaptation of amino-acid oxidation rates to the change in protein intake with the test diets, and (c) the interpretation of the slope of the regression as a true measure of the efficiency of protein utilization.

Energy balance and N balance

Ensuring that energy intakes match energy expenditure is the major difficulty with either weight and lean body mass loss or gain a consequence of any mismatch. This means that the protein requirement is a function of the state of energy balance with the influence quite marked (Millward, 2004). Thus the relationship between nitrogen balance (NB), intakes of energy (EI; kcal kg^{-1}) and of N (NI; mg N kg^{-1}) indicated

$$NB = 0:171NI + 1.006EI - 69:13.$$

This implies that the intake for N equilibrium (the requirement) will vary from 1.4 to 0.32 g kg^{-1} d^{-1} according to whether energy intakes are low (equal to the resting metabolic rate RMR), or high, (equal to twice the RMR). In fact, most, (85%), of the estimated between individual variability used to calculate the safe protein intake shown in **Fig. 5** could be accounted for by an error of only ±10% of basal metabolic rate (BMR) in estimating the true energy needs of a subject (Millward, 2004, 2012a). Such errors might not be observed in terms of weight change during the few days in which nitrogen balance is measured and it is likely that many of the positive N-balances shown in **Fig. 4** reflect excess energy intake.

Since N-balance varies as a function of energy intake, it may be argued that protein requirements can only be defined in terms of a specified energy intake level, but what is the appropriate energy intake? Should populations with low protein staples consume more energy to achieve body protein equilibrium? Will this predispose to obesity? To what extent does variation in energy intakes at energy balance (i.e., with increasing levels of physical activity) influence nitrogen balance? These are difficult and currently unanswered questions.

Adaptation to changes in protein intake

With the metabolic demand for amino-acids including both fixed and variable demands, the relationship between intakes and balance will be a function of time and the rate of adaptation, a major cause of the difficulty in the determination of human protein requirement by N balance. In practice, most balance studies are short term, i.e., 2 weeks at each intake studied and with diet periods randomized to minimize metabolic carryover of prior diets. The actual time taken for complete adaptation amino-acid oxidation and urea excretion rates to the new intakes is poorly understood but it is likely to be much longer than the time allocated in balance studies (Millward, 2003). Certainly studies in which the order of the test diets are varied indicate different N balances for the same intake according to the previous intake level. While two weeks allows N excretion to stabilize in subjects fed a protein-free diet, (i.e., during studies of the ONL), it appears that this adjustment to the extreme metabolic change of a protein-free diet occurs

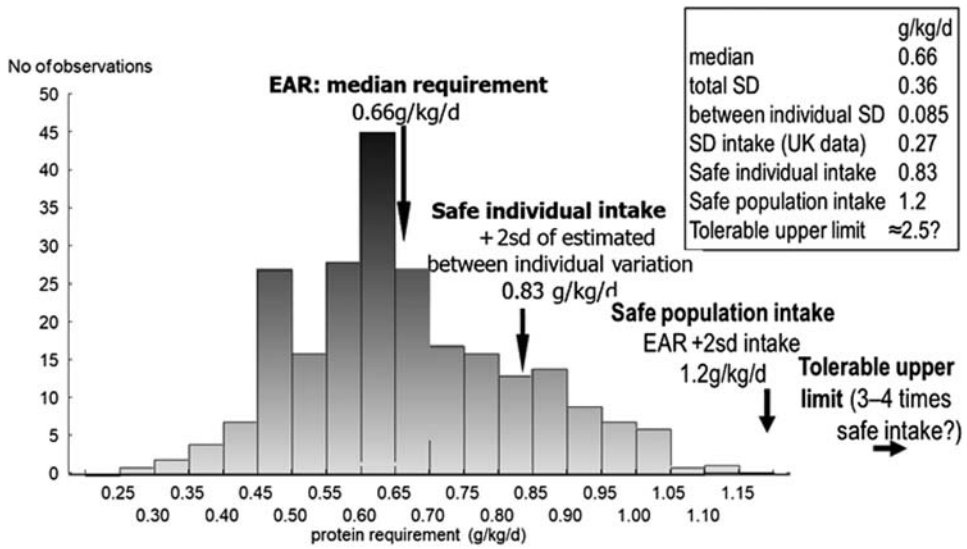


Fig. 5 Histogram showing distribution of individual reported values for the protein requirements assessed by nitrogen balance, with values for estimated average requirement (EAR), safe individual and population intakes and tolerable upper limit indicated. Requirements determined as intakes for nitrogen equilibrium observed in multilevel balance studies, analyzed by linear regression for each individual studied, ($n = 224$ individual subjects from $n = 32$ studies, after a 5% trim of outliers), expressed as protein equivalents. The values in the box indicate the magnitude of the total and between individual variation used in calculating the safe individual intake. Data from a meta analysis of N balance data reported by Rand et al. (2003). The minimum requirement after full adaptation is not known but likely to be less than the median value within the range shown.

more rapidly than the adjustment from one intake to another. The evidence suggests changes in N excretion continue over several months after reduced protein intakes. Incomplete adaptation to a reduced intake at the end of a two week dietary change will result in a more negative N-balance, increasing the apparent protein requirement.

Adaptation and the efficiency of protein utilization

Because of a lack of sufficient N-balance studies in children and pregnant and lactating women, their protein requirements have been derived with a factorial model: i.e., estimates of the additional needs during growth, pregnancy and lactation are added to the maintenance requirement assumed to be the same as for adults per kg body weight. The factorial model involves adjusting measured rates of protein deposition, in children or during pregnancy, or rates of breast milk production, by an efficiency factor to determine the dietary intake to meet these special needs. Historically these efficiency factors were decided somewhat arbitrarily at 0.7. However in the 2007 WHO report, much lower specific values were chosen from the slopes of the linear regressions of the limited N balance studies reported for the different population groups. Thus the calculated special needs added to the maintenance requirement are now much larger than previously identified.

Clearly most balance studies are performed with high quality protein sources and this raises the question as to why protein utilization should be so inefficient in healthy adults. In fact direct studies of the efficiency of utilization of proteins with ^{13}C-1 leucine oxidation balance studies have shown that the true efficiency of utilization of dietary protein is as would be expected, i.e., 95–100% for milk and high quality proteins and lower (60–70%) for wheat (Millward, 2003). The shallow slope (low efficiency) of the N-balance regression has been explained in terms of the adaptive metabolic demand, which results in increasing oxidative losses and consequent metabolic demand above the ONL (the intercept) with increasing intakes, i.e., the slope is unrelated to the efficiency of utilization. The slope of the regression is only of value in terms of defining the intake for equilibrium (i.e., the requirement = intercept/slope = intake for zero balance). This means that the special needs for children, pregnant and lactating women discussed below are all markedly overestimated in the current WHO report and these values have been adopted by other international and national bodies which invariably tend to adopt the WHO values.

Interpretation of N balance studies in the current WHO recommendations

Dietary protein allowances for individuals and populations

Nitrogen balance studies determine experimentally an individual MPR, while allowances for protein are calculated from the distribution and interindividual variability of these values within the relevant population as shown in Fig. 5. The EAR or average/median requirement, is the mid point of the range of individual requirements. The safe individual intake (Recommended Nutrient Intake in the UK) is defined as the 97.5th percentile of the distribution, nominally the average + 2SD. Thus any individual receiving such an

intake will have a very low, ($\leq$2.5%), risk of deficiency (intake $<$ requirement). However it is not generally appreciated that the safe intake for a population will differ from the safe individual intake. This is because calculation of a safe population intake, needs to take into account the distribution of individual intakes as well as requirements. Because the variability of intake is usually greater than that of the requirement, the safe population intake is greater than the safe individual intake, usually approximating to the average requirement +2SD of intakes or plus 3–5SD of requirements. In fact some argue that the EAR is the most important measure which can be used to judge the adequacy of intakes of populations because the prevalence of deficiency approximates to the proportion of the population with intakes below the EAR, with 2.5% usually an acceptable risk level, (this is called the cut point method of calculating the approximate deficiency prevalence). Thus dietary adequacy would involve most individuals with an intake above the EAR. A tolerable upper limit (TUL) has not been identified for protein but the evidence suggests that intakes of 3–4 times the safe individual intake are consumed without obvious harm. This would suggest that the TUL may be quite high and certainly higher than the value of 2$\times$ safe intake often assumed in the past.

The serious implication of lack of complete adaptation in short term multi-level balance studies is that because of the very wide range of protein intakes in the human diet, mainly through variable meat intake, the apparent requirement indicated in a study may still reflect the prior habitual diet: i.e., the actual metabolic demands are higher than minimum levels because of the adaptive component of amino-acid oxidation set to balance previous intakes. This may explain the very wide range of reported apparent requirements shown in Fig. 5 from below 0.4 to greater than 1.1 g protein per kg per day. If adaptation does account for the variability rather than variation in actual protein requirements than a quite different analytical model would be implied in which the MPR and safe intake are much lower. In this case risk of deficiency for fully adapted individuals would remain low until intakes fell to quite low levels. Such adaptive models pose difficult questions for public health nutrition.

Protein requirements for growth and special needs

For infants, children, and pregnant and lactating women, protein requirements are derived by a semi-factorial analysis of the maintenance and special needs, with an assumed efficiency of utilization, all adjusted for individual variation to give the safe intake. The main components of the protein requirements for infants children and adults are shown in Fig. 6. The metabolic demands are for maintenance, assumed to be the ONL as derived for adults (Rand et al., 2003), and assumed to be the same for all age groups, and the demand for growth derived from measured rates of protein accretion in infants and children. The additional amounts needed to account for the low dietary efficiency of utilization to meet these metabolic demands come from balance studies in adults and children which indicate efficiencies of 47% and 58% respectively. To account for interindividual variability, the safe intakes includes the addition of 2 SDs for maintenance and dietary growth needs, calculated from a coefficient of variation (CV) that is the weighted mean of the CVs for maintenance and growth. Overall the safe intake falls from 1.31 g kg^{-1} d^{-1} at 6 months to 0.83 g kg^{-1} d^{-1} in adults. The values for infants are validated against protein intakes of healthy breast-fed infants.

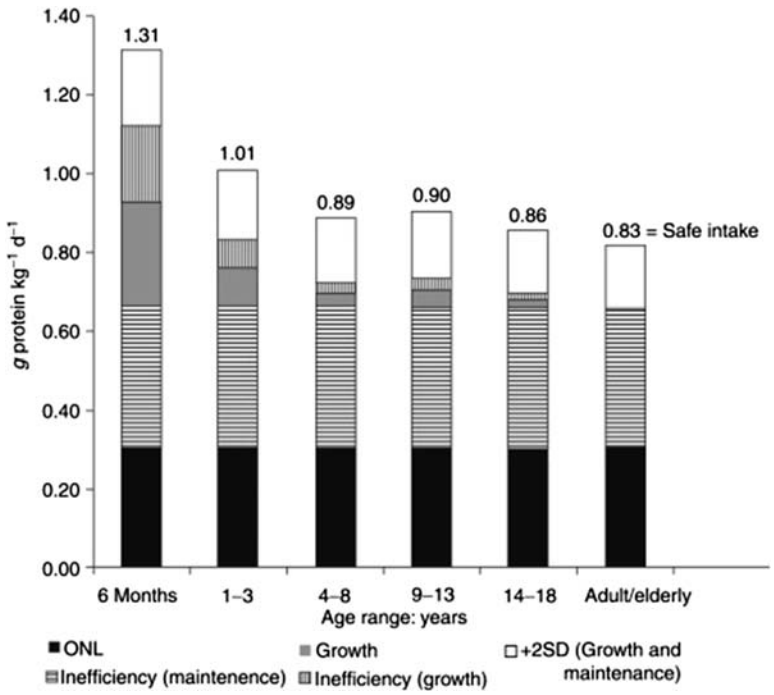


Fig. 6 Histogram showing components of the protein requirements throughout the lifecycle within the factorial model. Overall values are averages of the separate values for boys and girls. Values from WHO/FAO/UNU (2007).

Pregnancy requirements

The values reported by WHO/FAO/UNU allow for protein retention in the products of conception and in the maternal tissues associated with an average gestational weight gain in a healthy women of 13.8 kg, estimated from body composition analyses, and for the increased maintenance costs of the increased body weight. The efficiency of dietary protein utilization to meet these growth costs is assumed to be 42% a value observed in nitrogen balance studies in primiparous teenagers, and the safe intake is calculated assuming a CV of 12%. This equates to additions of 0.7, 9.6, and 31.2 g protein per person per day. Lactation requirements of 19 g d^{-1} and 12.5 g d^{-1} derived from the demand observed as average milk protein output of well-nourished women exclusively breastfeeding for the first 6 months and partially breastfeeding after this time adjusted for an efficiency of utilization of 47% as observed in the meta-analysis of N balance studies in adults (Rand et al., 2003).

Areas of uncertainty

Requirements for infants and children, pregnancy and lactation

Definition of protein requirements has historically been problematic and successive reports have never been without new areas of controversy. In the most recent report a major consequence of the difficulty associated with interpreting nitrogen balance studies is not only uncertainty about the true adult MPR but also the possibility of overestimation of requirements for the other population groups through calculations within the factorial model. The problem derives from application of very low efficiency figures of $\approx$50% with no biological explanation for their values other than their derivation in a few balance studies. As is apparent in Fig. 6 for children and adolescents the identifiable metabolic demand for maintenance and growth accounts for only half of the average requirement because the assumed efficiency of dietary utilization is only about 50%. An even lower efficiency is assumed in calculating the requirements for pregnancy. If the actual efficiency is higher, as assumed in previous reports, then the true requirement values would be lower. Although overestimation of requirements for infants and children is unlikely to result in harm, this may not be the case for pregnancy. In fact there is evidence that excessive protein intakes in pregnancy can have adverse effects on pregnancy outcomes so that if any additional dietary intakes are recommended for pregnant women these should comprise of a normal healthy diet. In the WHO/FAO/UNU report advice is given that the additional protein during pregnancy should consist of additional normal food, rather than high protein supplements. During the third trimester an extra 31 g d^{-1} of protein would represent an extra 3.6 MJ of a mixed diet assuming it contains 15% protein energy. This is considerably more additional energy than is recommended at this stage of pregnancy by any agency. Indeed given the concern for adverse influences of overweight and obesity on pregnancy outcomes and given that pregnant women often have successful pregnancy outcomes without any increase in food intake, it is likely that this amount of additional food would result in excessive weight gain.

The amino-acid pattern of the metabolic demand: plant versus animal sources

With continuous and extensive amino-acid interconversion, the pattern of amino-acids in dietary protein need not exactly match the amino-acid pattern of the demand. This is because of some interconvertibility of dietary amino-acids although this is not entirely straightforward. Firstly while the indispensability of nine of the 21 amino-acids in protein is clear (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine), as is the dispensability of five amino-acids, (aspartic acid, asparagine, glutamic acid, alanine, and serine), which can be replaced by nitrogen from other amino-acids or sources of nonessential N such as ammonium compounds, there is also an intermediate class of conditionally dispensable amino-acids (DAAs). These (cysteine, tyrosine, glycine, arginine, glutamine, and proline) can be formed from other amino-acids but under specific physiological or pathological conditions their formation may be inadequate. Secondly it appears that the indispensable amino-acids (IAAs) are a poor source of amino nitrogen for the DAAs after their oxidation and that dietary amino-acid mixtures rich in IAAs like egg protein are less effective in meeting the demand than mixtures with added non-essential N. Such diluted mixtures maintain overall balance with lower levels of IAAs, indicating that the metabolic demand for amino-acids includes an absolute need for non-essential N.

Historically growth studies with laboratory animals gave rise to the view of marked differences in the ability of plant compared with animal sources of protein to support growth resulting in classification of dietary proteins according to their nutritional value (quality) in terms of their IAA content. Thus compared with animal source proteins, cereal proteins generally have low levels of lysine and in some cases tryptophan because of the structure of the main storage proteins—the prolamins-e.g., gluten in wheat and zein in maize. In legumes while the main storage proteins, the globulins contain much higher levels of lysine than cereals, they often have relatively lower levels of the sulfur amino-acids, methionine and cysteine. However when different plant protein sources are combined in these assays, such as cereals and legumes, near maximal animal growth can be observed and this gave rise to the concept of complementation, in which the appropriate balance of essential amino-acids is provided from combinations of plant proteins. In fact, apart from cereals, legumes and some starchy roots, most other plant proteins, such as Irish potatoes and leaves such as spinach or oilseeds, and the novel protein sources such as marine microalgae, aquatic plants, blue-green algae and mycoprotein, have balanced amino-acid profiles.

However it is insufficiently appreciated that in human nutrition, with the rapid growth of the newborn infant slowing markedly after weaning, the nutritional demand for IAAs for tissue growth becomes a minor component of the metabolic demand after the first year of life. Little metabolic demand for amino-acids is generated by protein turnover because of amino-acid recycling. Some net protein synthesis is associated with skin and hair growth and with gastric secretions (e.g., threonine-rich mucus glycoproteins) that pass into the colon to be utilized for bacterial metabolism, and with the post prandial replacement of post absorptive tissue protein losses as shown in **Fig. 3**. The metabolic demand for maintenance of normal function and composition is a poorly understood pattern of amino-acids utilized in the various metabolic pathways other than protein synthesis, but it has long been known that this pattern differs from that required for growth (the tissue protein pattern) and contains a much lower overall amount of IAAs. The currently recommended amino-acid pattern for human adults, the maintenance pattern, contains 28% IAAs compared with 42% in tissue proteins. Furthermore nitrogen balance studies of different plant protein sources has consistently failed to show marked differences between plant-based compared with animal protein based diets. Indeed long term studies of healthy active adults on diets where wheat provided most of the protein have shown that body composition and physical fitness is maintained over periods of several months (see Millward, 2012b). This means that protein quality in terms of the amino-acid profile may be less relevant in human nutrition at least after the first few years of life when the rapid growth of infancy has subsided.

In fact it is likely that the main difference between animal and plant dietary protein sources reflect the lower digestibility of some plant protein sources due to a combination of tough plant cell walls which reduce the accessibility of plant proteins to digestive enzymes and antinutritional factors which can inhibit digestion.

The currently accepted amino-acid pattern of the maintenance requirement for IAAs defined in the 2007 WHO report is not entirely satisfactory (see Millward, 2012b). Each value was chosen from several values which derived from different methodologies on the basis of the consultations judgment as to which was likely to be the most accurate. While the chosen value for lysine (30 mg/kg per day) was derived from tracer studies using the 24 h indicator amino acid method, which were considered to provide the best stable isotope data currently available, it was also stated that the possibility of a lower value intermediate between 22 and 30 could not be ruled out. This means that considerable uncertainty remains about the amino-acid requirement pattern which is used to judge the quality of dietary protein sources.

Conclusions: implications of adaptation for nutrition policy

In general, protein requirements serve two purposes. One is as a basis for prescription (i.e., advice on safe diets through recommending appropriate dietary intakes). Adaptation implies a low but difficult to define MPR. Indeed, since natural diets, providing sufficient energy and other nutrients, usually provide considerably more than the currently defined MPR, especially the case for young children, so that the magnitude of the MPR becomes to some extent an issue of scientific curiosity only. Formulation of policy in relation to prescriptive matters will inevitably and correctly be most concerned with satisfying the upper range of demands for protein and, where there is uncertainty, include positive margins of error. In this case, it is arguably unwise to adopt an adaptive model and reduce the MPR, even if agreement could be reached on the likely lower limit of adaptation. Indeed, an adaptive model does not imply that protein is an unimportant nutrient for the maintenance of human health and well-being but that indicators other than balance (nitrogen, protein, or amino-acid) need to be identified. Thus, the most relevant measure is an optimal requirement allowing balance and supporting both optimal body function and minimum risk of chronic disease. Unfortunately, it remains the case that there is mixed evidence about the overall long term health effects of protein at intakes at or above the current MPR with no overall consensus. The most widely discussed issue is the suggestion that intakes above the MPR can protect the elderly from the muscle wasting-disease sarcopenia. This is based on the observed anabolic resistance of muscle protein synthesis to essential amino acids together with a number of dietary intake-sarcopenia studies. However the evidence base from these studies that increased protein intakes limit sarcopenia is minimal or based on studies in which the authors interpretation of the results has been identified as misjudged (Millward, 2012c). This results in a dilemma for those attempting to frame prescriptive dietary guidelines. From this perspective, it is probably wise to retain current values as an operational expedient until it becomes possible to quantify the benefits (and any risks) of protein intakes within the adaptive range.

The other purpose of requirement recommendations is as a diagnostic indicator of deficit risk, often within an epidemiological context in which population groups rather than individuals are considered. In this case, indicators used to estimate prevalence of disease states or deficit risk are carefully chosen so as to strike an acceptable balance between false positives and false negatives. The main implication of adaptation for estimating risk of deficiency as intakes become less than requirements is a dramatic reduction in the prevalence of risk for most populations compared with that assessed according to the traditional model, which does not account for adaptation. This occurs because the requirement and intake can be assumed to be correlated and because the actual MPR and safe intake calculated from it will be lower. As in the prescriptive context, this low risk of deficiency applies only to that of being unable to maintain nitrogen balance after full adaptation with otherwise nutritionally adequate diets satisfying the energy demands. Whether such populations, (likely to be rare), enjoy optimal protein-related health in terms of immune function, bone health, or any other function is a separate issue and needs to be addressed as such. From this perspective, it follows that maintenance of nitrogen balance can no longer be used as a surrogate of adequate long term protein-related health, and that current lack of quantifiable alternative indicators is no excuse for ignoring the issue of adaptation.

References

European Food Safety Authority, 2012. Scientific opinion on dietary reference values for protein. EFSA J. 10 (2), 2557–2559. https://doi.org/10.2903/j.efsa.2012.
Millward, D.J., Jackson, A.A., 2012. Protein requirements and the indicator amino acid oxidation method. Am. J. Clin. Nutr. 95 (2), 1498–1501. https://doi.org/10.3945/ajcn.112.037002.Mireia.
Millward, D.J., 1995. A protein-stat mechanism for regulation of growth and maintenance of the lean body mass. Nutr. Res. Rev. 8 (1), 93–120. https://doi.org/10.1079/nrr19950008.
Millward, D.J., 2001. Workshop on "protein and amino acid requirements and recommendations" methodological considerations. Proc. Nutr. Soc. 60, 3–5. https://doi.org/10.1079/PNS200064.
Millward, D.J., 2003. An adaptive metabolic demand model for protein and amino acid requirements. Br. J. Nutr. 90 (2), 249–260. https://doi.org/10.1079/bjn2003924.
Millward, D.J., 2004. Macronutrient intakes as determinants of dietary protein and amino acid adequacy. J. Nutr. 134 (6 Suppl), 1588–1596. https://doi.org/10.1093/jn/134.6.1588s.
Millward, D.J., 2012a. Identifying recommended dietary allowances for protein and amino acids: a critique of the 2007 WHO/FAO/UNU report. Br. J. Nutr. 108 (Suppl. 2), S3–S21. https://doi.org/10.1017/S0007114512002450.
Millward, D.J., 2012b. Amino acid scoring patterns for protein quality assessment. Br. J. Nutr. 108 (Suppl. 2), 30–43. https://doi.org/10.1017/S0007114512002462.
Millward, D.J., 2012c. Nutrition and sarcopenia: evidence for an interaction. Proc. Nutr. Soc. 71 (4), 566–575. https://doi.org/10.1017/S0029665112000201.
Millward, D.J., 2013. The use of protein:energy ratios for defining protein requirements, allowances and dietary protein contents. Publ. Health Nutr. 16 (5), 763–768. https://doi.org/10.1017/S1368980013000396.
Rand, W.M., Pellett, P.L., Young, V.R., 2003. Meta-analysis of nitrogen balance studies for estimating protein requirements in healthy adults. Am. J. Clin. Nutr. 77 (1), 109–127. https://doi.org/10.1093/ajcn/77.1.109.
World Health Organization, 2007. Protein and Amino Acid Requirements in Human Nutrition. Report of a Joint FAO/WHO/UNU Expert Consultation (WHO Technical Report Series 935). WHO, Geneva.

Further Reading

Campbell, W.W., Johnson, C.A., McCabe, G.P., Carnell, N.S., 2008. Dietary protein requirements of younger and older adults. Am. J. Clin. Nutr. 88, 1322–1329.
Food and Nutrition Board and Institute of Medicine, 2002. Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino-acids (Macronutrients). National Academy Press, Washington, DC.
Millward, D.J., Jackson, A.A., 2003. Reference protein: energy ratios of diets in relation to current diets in developed and developing countries: implications of proposed protein and amino-acid requirements values. Publ. Health Nutr. 7, 387–405.
Millward, D.J., 1998. Metabolic demands for amino-acids and the human dietary requirement: Millward and Rivers (1988) revisited. J. Nutr. 42, 367–393.
Millward, D.J., 1999a. Optimal intakes of dietary protein. Proc. Nutr. Soc. 58, 403–413.
Millward, D.J., 1999b. The nutritional value of plant based diets in relation to human amino-acid and protein requirements. Proc. Nutr. Soc. 58, 249–260.

Salt, hypertension and cardiovascular outcomes

Tilman B. Drüeke[a] **and Paolo M. Suter**[b], [a] CESP, Hôpital Paul Brousse, Paris-Sud University (UPS), Versailles Saint-Quentin-en-Yvelines University (Paris-Ile-de-France-Ouest University, UVSQ), Villejuif, France; and [b] Clinic for Endocrinology and Clinical Nutrition, University Hospital, Zurich, Switzerland

© 2023 Elsevier Ltd. All rights reserved.

Introduction 579
Salt (sodium) intake and blood pressure 581
Salt sensitivity 582
Epidemiological studies 583
Intervention studies 585
Salt intake and cardiovascular outcomes 588
Conclusions 593
Acknowledgments 593
References 593

Key points

- Sodium chloride (salt) is an essential nutrient required for numerous physiological functions. It is naturally present in many foodstuffs; it is also added in food processing, at the table and in cooking process.
- Urinary sodium excretion is generally considered as a useful surrogate parameter for sodium intake, with spot urine samples being less reliable than 24-h urine samples, and multiple urine samples more representative of a person's usual sodium intake than a single sample.
- Sodium chloride is but one among many factors which are involved in the control of blood pressure. Individuals are more or less sensitive to salt in terms of blood pressure.
- While numerous observational studies in general population have shown associations between salt intake and blood pressure, many others failed to demonstrate such an association. Differences in confounding factors probably account for the discrepant results.
- Observational studies are merely hypothesis generating. Causal relationships can only by proven by intervention studies. Unfortunately, the number of well conducted intervention studies is much smaller, and again reported findings have remained inconclusive, both as regards the effect of salt intake on blood pressure and its role in cardiovascular events and mortality.

Introduction

"Salt" is a commonly used term. It is composed of sodium and chloride although strictly speaking there are numerous other salts containing sodium or chloride such as sodium carbonate, sodium fluoride, potassium chloride, and magnesium chloride. Sodium does not naturally occur in isolated form. It represents only one of the two elements necessary to form the most common form of "salt" (NaCl). Both in the medical and general literature the terms sodium and salt are often used interchangeably.

Sodium is required for a large number of physiological functions, including the maintenance of body fluid volumes and their distribution, skeletal and cardiac muscle contraction, and nerve conduction velocity. Salt also plays a role in growth and development, general appetite, sexual performance, reproduction, and pregnancy, to name just a few. Chloride participates together with sodium in most of these functions. Chloride also is involved in the digestion of meals in the stomach via hydrochloric acid.

Increased needs of salt are caused by changes in external conditions and enhanced cutaneous sodium losses such as in hot climate and at high levels of physical activity. They can also be caused by disease states such as salt losing endocrinopathies and nephropathies. In other disease states there may be decreased needs of salt. Examples are salt-sensitive hypertension, chronic heart failure, chronic kidney disease, the nephrotic syndrome, and decompensated liver cirrhosis. The latter are clinical conditions associated with salt and water retention leading to symptoms of volume overload. In these conditions controlling salt intake is key for a successful therapeutic strategy. However, a number of randomized clinical trials have shown that salt restriction also stimulates sympathetic nervous system activity (thus stimulating sodium retention), increases serum cholesterol and triglyceride levels, activates the renin-angiotensin system (thus promoting cardiovascular damage and risk), decreases insulin resistance and alters aldosterone secretion (Graudal et al., 2020). These stimulations are not necessarily beneficial, they may be even harmful. Thus in each individual, potential benefits of a more or less severe salt restriction must be weighed against potential harms.

Dietary salt is naturally present in many foodstuffs, but salt is also added in food processing and at the table or cooking process (discretionary salt). Natural sources make up 1–1.5 g and discretionary salt 2–3 g of daily ingested sodium chloride. The main amount stems from salt added to food by food manufacturers. It represents roughly 3–4 g per day. Since the body has only limited reserves of sodium and chloride, it depends on the regular intake of salt. The minimum daily requirement compatible with life has been estimated to be 1–2 g sodium chloride. However, this minimum amount is not compatible with normal physical and mental activities. Total salt intake estimates derived from 24 h urine samples are in the range of 7–9 g per day in the Western world (Muntzel and Drüeke, 1992), with the majority of samples being comprised between 6 and 14 g per day in high and middle income countries according to one source (Intersalt Cooperative Research Group, 1988) and between 6.5 and 12 g (2.6–4.8 g sodium) per day according to another, more recent source (McCarron et al., 2013). **Table 1** shows an update of the amount of dietary salt intake in various countries worldwide. The relative importance of assumptions about both discretionary salt used by the cook, at the table, and salt added by the industry has to be stressed, keeping in mind that a significant proportion of the salt used during preparation of meals is not ingested, but discarded.

Urinary sodium excretion is generally considered as a useful surrogate parameter for sodium intake. Note that in hot climates or under sustained physical exercise up to 50% (or even more) of the absorbed sodium chloride can be lost by sweating, that is via the cutaneous route. This needs to be taken into account when relying on urinary sodium as a proxy of sodium consumption. There is an ongoing debate on the relative merits of 24 hr urine collections as compared to spot urine samples for the indirect assessment of sodium intake. A recent study compared the sodium excretion and blood pressure relationships from measured 24 h urinary sodium vs. formula-estimated spot urinary sodium values, using restricted cubic spline plots for adjusted multilevel linear models (Naser et al., 2021). According to this analysis all spot urine formulas overestimated 24 h sodium at lower levels but underestimated 24 h sodium at higher levels. There was a linear relationship between measured 24 h urinary sodium excretion and systolic blood pressure, while estimated sodium excretion from all 3 spot urine formulas had a J-shaped relationship with systolic blood pressure. Although 24 h urinary sodium excretion is more reliable the collection of spot urine is more convenient and less expensive (John et al., 2016). While single measurements lack accuracy on an individual level (Ginos and Engberink, 2020) repeat measurements better reflect average sodium intake (Sun et al., 2017), especially when used for population-level estimates in large epidemiologic studies (O'Donnell et al., 2020). This is true for both spot urine and 24 hr urine collections. Food questionnaires are not reliable in assessing dietary salt intake (Ginos and Engberink, 2020) since food tables are rather incomplete and not regularly revised and updated and thus not very useful.

Table 1 Mean salt intake according to Intersalt, Intermap and various other sources[a].

	Intersalt	*Intermap*	*Other sources*	*Extreme data*
Argentina	9.1	n.a.	<12	
Australia	n.a.	8.5	10.3/7.7	12
Austria	n.a.	n.a.		
Belgium	8.5	n.a.	10.4	
Brazil	n.r.	9.9	10.8	1.9
Canada	10.3	n.a.	8.5	
China	12.1	11.9	8.4/7.7	
Denmark	8.2	n.a.	10.6/7.1	
Finland	9.1	8.4		
France	n.a.	n.r.	7.9	
Germany	9.5	n.a.	9.0/6.5	
Hungary	11.7	n.a.		
Iceland	8.1	n.a.	9.3/6.5	
India	10.7	n.a.	9.9/9.0	
Italy	10.3	11.0	11.4	
Japan	11.0	12.6		23.5
Mexico	8.5	n.a.	6.3	
Netherlands (The)	8.8	7.1		15.4
Nigeria	n.a.	6.7	6.6	
Poland	11.1	n.a.	13.5	
Portugal	n.r.	n.a.	12.3	
RSA	n.a.	8.7	7.8/9.5	
Russia	n.a.	n.r.	13.4	
South Korea	12.2	n.a.	11.5	
Spain	8.9	n.r.	11/8	5.5
Switzerland	n.a.	n.a.	9.3	
UK	8.9	8.9	8.8	
USA	7.6	9.1	9.0	

n.a., not available; n.r., considered non reliable.

[a]Sources: Intersalt (Intersalt Cooperative Research Group, 1988); Intermap (Brown et al., 2009); and several additional articles identified in PubMed, including McCarron et al. (2013) and Bernstein and Willett (2010).

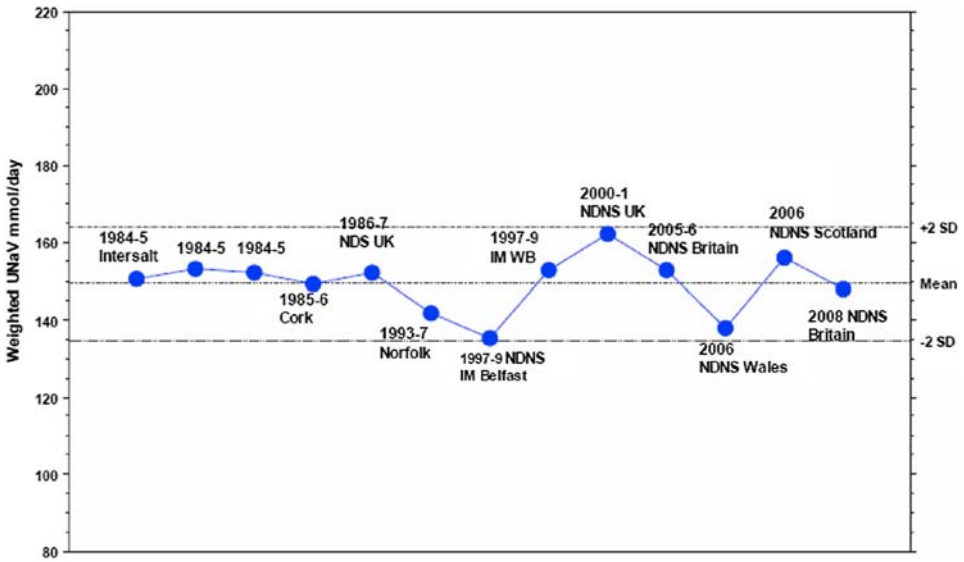


Fig. 1 Mean and ±SD 24 h UNaV from 13 published surveys in the UK between 1984 and 2008 with essentially equal representation of women and men ($n = 6343$). Trend line equation $y = -0.097x + 150.4$; $R2 = 0.0026$; UNaV, urinary sodium excretion; UK, United Kingdom; NDS, National Diet Survey; NDNS, National Diet and Nutrition Survey. From McCarron et al. (2009) (reproduced with permission).

In the past decades numerous attempts have been made to decrease salt intake in the general population. These attempts have not been successful in most instances. Dietary sodium intake has varied minimally in the UK over 25 years on the turn of the 20th century, encompassing several surveys from the UK and also the Intersalt data (Intersalt Cooperative Research Group, 1988), as reported by McCarron et al. (2009) (**Fig. 1**). This analysis did not support pronouncements by the British Foods Standards Agency that their national campaign directed at sodium reduction had achieved a significant reduction in the population. Similarly, according to a subsequent report by Bernstein and Willett based on the analysis of 38 studies dating from 1957 to 2003, sodium intake in the US adult population appeared to be well above current guidelines and did not decrease with time (Bernstein and Willett, 2010). It is interesting to note that the consumption of sodium and also that of potassium present specific seasonal variations, as demonstrated by a recent study from Switzerland (Marti- Soler et al., 2017). However, other studies, even in high salt consuming countries, did not show a large seasonal intake variations, or variations were only found in certain subgroups of the population (such as older women).

Salt (sodium) intake and blood pressure

As already mentioned, sodium and chloride are indispensable for maintaining body fluid volumes in the normal (physiological) range, both outside and inside the cells of the organism, that is in the extracellular and intracellular space, respectively. The extracellular volume comprises the blood space and the interstitial space. Maintaining blood volume within physiological limits is essential for the control of blood pressure, in concert with fine tuning of cardiac output (i.e. "stroke volume" × "heart rate") and peripheral vascular resistance, that is the degree of openness (vasoconstriction vs. vasodilatation) of the peripheral arteries and arterioles. The healthy organism (in contrast to patients with heart or kidney failure, who show an impaired natriuresis) closely adapts salt and water excretion to salt and water intake, guaranteeing a neutral balance. In most healthy people, the kidneys are capable of excreting large amounts of sodium when intake is high, and of excreting little sodium when intake is low. It is noteworthy that the kidneys' ability to excrete sodium is impaired in the presence of low potassium diets (Krishna and Kapoor, 1991). This tight control helps to prevent increases or decreases of blood pressure beyond the physiological range.

Sodium chloride is but one among the numerous factors which control blood pressure. The maintenance of normal blood pressure requires a subtle interplay between endogenous factors including age, gender, the cardiovascular system, the kidneys, the endocrine system, neurohumoral factors and genetically determined predisposition including ethnicity, and exogenous factors including dietary factors, physical activity, body mass index, and psychosocial stress. Among the dietary factors, sodium chloride plays an important, although frequently overemphasized role as compared to potassium and other factors. Although the accumulation of sodium chloride favors arterial hypertension and depletion favors hypotension its effects should not be regarded in isolation, especially not without considering the concomitant intake of potassium, the overall health status, body weight status, insulin resistance and many other still ignored factors such a circadian rhythm misalignment.

However, since the majority of studies aimed at examining the effect of sodium have been done without considering concomitant potassium intake and other determinants of salt sensitivity we shall present reports on sodium in the first place, and only briefly mention reports on potassium and sodium/potassium ratio in the second place.

Salt sensitivity

An important question is what is excessive and what is insufficient salt intake? The short answer is that this depends on the individual sensitivity to exogenous salt as regards its effect on blood pressure. **Table 2** summarizes the factors commonly incriminated in salt sensitivity and salt resistance, respectively. Salt sensitive individuals exhibit a blood pressure decrease in response to salt restriction, and a blood pressure increase in response to salt overload. Several different methods have been proposed to assess salt sensitivity but there is no commonly accepted definition. As an example, salt sensitivity can be assessed by putting subjects on a high or low salt diet for several days and measuring subsequent blood pressure changes. In the general population, approximately 40% of people are considered to be "salt-sensitive." Consequently, 60% among them are "salt-resistant." Young individuals are more often salt-resistant than older individuals. A minority of young people may even be "reverse reactors." They exhibit an increase in blood pressure in response to reduced salt intake. **Fig. 2** shows a histogram of salt sensitivity in normotensive individuals, that is changes in mean blood pressure in response to salt depletion (Weinberger et al., 1986).

Not only lay people but also many physicians continue to be confused about the effects of salt on blood pressure. As already mentioned, several disease states as well metabolic-neuroendocrine constellations are associated with a reduced sodium excretory capacity of the kidneys and arterial hypertension, and limiting salt intake in such disease states can reduce blood pressure. But those observations cannot be used to impose a reduction of salt intake on healthy subjects.

The assumption of a close relation between salt intake and blood pressure in the general population actually has a historical basis (Muntzel and Drüeke, 1992). First, in the early decades of the 20th century no efficient antihypertensive treatments were available other than a severe reduction of salt intake. The well-known amazing blood pressure reduction by the Kempner rice diet is most likely the first well documented evidence of the effect of a severe salt reduction on blood pressure as well as morbidity and mortality in highly selected patients. Usually it is ignored that while under this diet the patients lost more than 20 kg (obviously part of this weight loss was due to edema reduction and thus volume depletion). Nevertheless one can dispute whether the salt restriction or the body weight reduction was of greater physiological importance for their clinical improvement. In view of the global obesity pandemic weight reduction would most likely be favored as the main mechanism. Second, in the 1950s and 1960s Dahl developed the concept that dietary salt is a major factor responsible for increased blood pressure, based on experimental studies in salt-sensitive rats, with his first landmark report dating back to 1958 (Dahl, 1958).

Table 2 Salt sensitivity: variable individual response to changes in dietary salt intake.

Relative salt-sensitivity	* Old individuals * Black subjects * Some hypertensive subjects (30–40%) * Baseline level of blood pressure * Calcium deficiency * Expansion of extracellular volume and increase of Atrial Natriuretic Factor (ANF) * Stimulation of sympathetic nervous system * Biochemical and genetic markers: – Low plasma renin activity – Increased plasma ANF – Decreased urinary kallikrein – Decreased urinary endothelin-1 – Haptoglobin 1-1 phenotype – 17a-hydroxylase deficiency – Impaired nitric oxide synthase activity – Mutations of sodium channel genes – Genetic polymorphism of a_2-adrenergic receptor
Relative salt-resistance	* Young individuals * White subjects * Most hypertensive subjects (60 à 70 %) * Biochemical and genetic markers: – Elevated plasma renin activity – Low plasma ANF – Haptoglobin 1-2 and 2-2 phenotypes – Genetic polymorphism of a2-adrenergic receptor
Variable salt sensitivity during lifetime, potential factors	* Age * Baseline level of blood pressure * Obesity * Pregnancy * Several disease states * Some medications

Certain biochemical and genetic markers may be used to assess salt sensitivity. For example the activities of several plasma enzymes and the amounts of certain proteins in plasma and tissues are related to salt sensitivity.

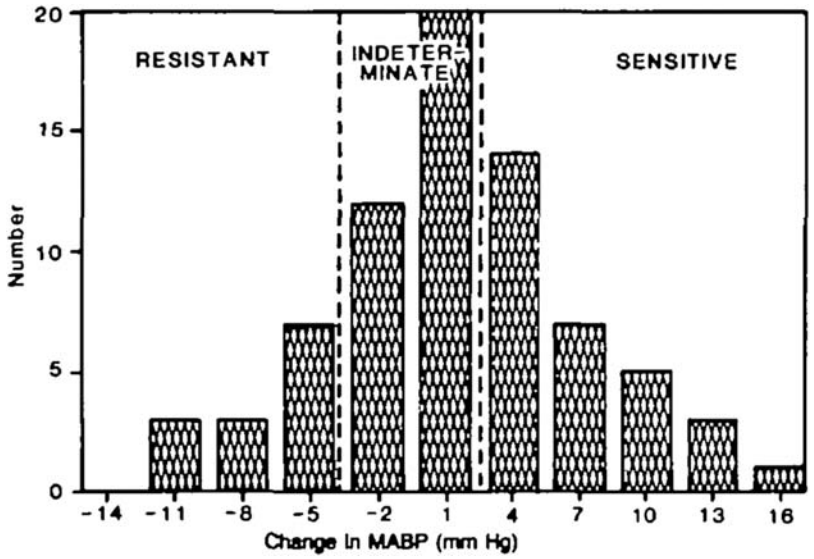


Fig. 2 Blood pressure response to salt depletion of normotensive subjects. Changes in Mean Arterial Blood Pressure (MABP) in normotensive adults after dietary sodium restriction. Sodium sensitivity was defined as a decrease of at least 3 mmHg, and sodium resistance as an increase of at least 3 mmHg. From Weinberger et al. (1986) (reproduced with permission).

Epidemiological studies

Following this experimental evidence, numerous epidemiological studies done in general population showed associations between salt intake and blood pressure. This started with reports of studies in Northern and Southern Japan, which suggested an association between high salt intake and high blood pressure, although the evidence was rather weak, if not entirely speculative (Muntzel and Drüeke, 1992). Since the 1980s, the results of better designed epidemiological studies were reported which aimed to definitively establish the presumed relationship between salt intake and blood pressure. The largest early effort in this field has been undertaken with the Intersalt study (Intersalt Cooperative Research Group, 1988). It included 10,079 volunteers in 52 participating centers worldwide and showed a weak positive association between urinary sodium excretion (reflecting salt intake) and blood pressure. Specifically, systolic blood pressure augmented by 2.2 mm/Hg for every 100 mmol per day increase in habitual sodium intake (6 g salt per day). However, the association disappeared when four centers of non-acculturated populations in Brazil, Kenya and New Guinea were excluded from the analysis, as shown in **Fig. 3**, which also shows associations after adjustment for various confounders. People of these non-acculturated populations had unusually low salt intakes and blood pressures, unlike the people in industrialized countries. Further they all had a normal body weight, a high potassium intake, plenty of physical activity and a complete lack of acculturation induced stress biomarkers. Moreover, the generally limited life span of these populations did not enable any information on the possible development of hypertension with age. Another observational study of large sample size was performed at same time in Scotland. It failed to identify a relationship between urinary sodium excretion and blood pressure (Smith et al., 1988).

Differences in several confounding factors probably account for the discrepant results. It is well known that when age, body mass index (BMI), physical activity, potassium intake and alcohol consumption are taken into account the relation between sodium intake and blood pressure becomes weaker or even disappears.

The largest study ever done in this field is the PURE trial (Mente et al., 2014). The authors studied 102,216 adults from 18 countries. Estimates of 24 h sodium excretion were made based on spot urine measurements as surrogates for sodium intake. Regression analyses showed increments of 2.11 mm Hg in systolic blood pressure and 0.78 mm Hg in diastolic blood pressure for each 1 g increment in estimated sodium excretion (**Fig. 4**). The slope of this association was steeper with higher sodium intake. Further, the slope of the association was steeper for persons with hypertension than for those without hypertension and was steeper with increased age.

A very recent observational study of large sample size showed that treatments with effervescent or soluble acetaminophen (paracetamol) formulations containing high amounts of sodium, as compared to sodium-free acetaminophen formulations, were associated with increased risks of cardiovascular disease and all-cause mortality (Zeng et al., 2022). This was true for both in individuals with hypertension (1-year risk, 5.6% vs. 4.6%) and without hypertension (1-year risk: 4.4% vs. 3.7%). It must be pointed out, however, that the effervescent and soluble formulations of 0.5 g acetaminophen contain 0.44 and 0.39 g of sodium, respectively. Thus, the intake of maximum daily dose (i.e. 4 g/day) of sodium-containing acetaminophen corresponds to the considerable amount of an additional ingestion of respectively 3.5 and 3.1 g of sodium (8.75 and 7.75 g of salt) per day. In this context it is important to mention that a randomized controlled trial recently showed that regular daily intake of 4 g acetaminophen (with a negligible sodium content of only 0.04 mg per capsule) increased systolic blood pressure per se in individuals with hypertension by $\approx$5 mm Hg when compared with placebo and hence increased cardiovascular risk (MacIntyre et al., 2022).

Several other studies showed that excessive alcohol consumption and obesity are more important lifestyle factors than salt intake for the development or degree of severity of hypertension and modulation of salt sensitivity but because of space limits some of

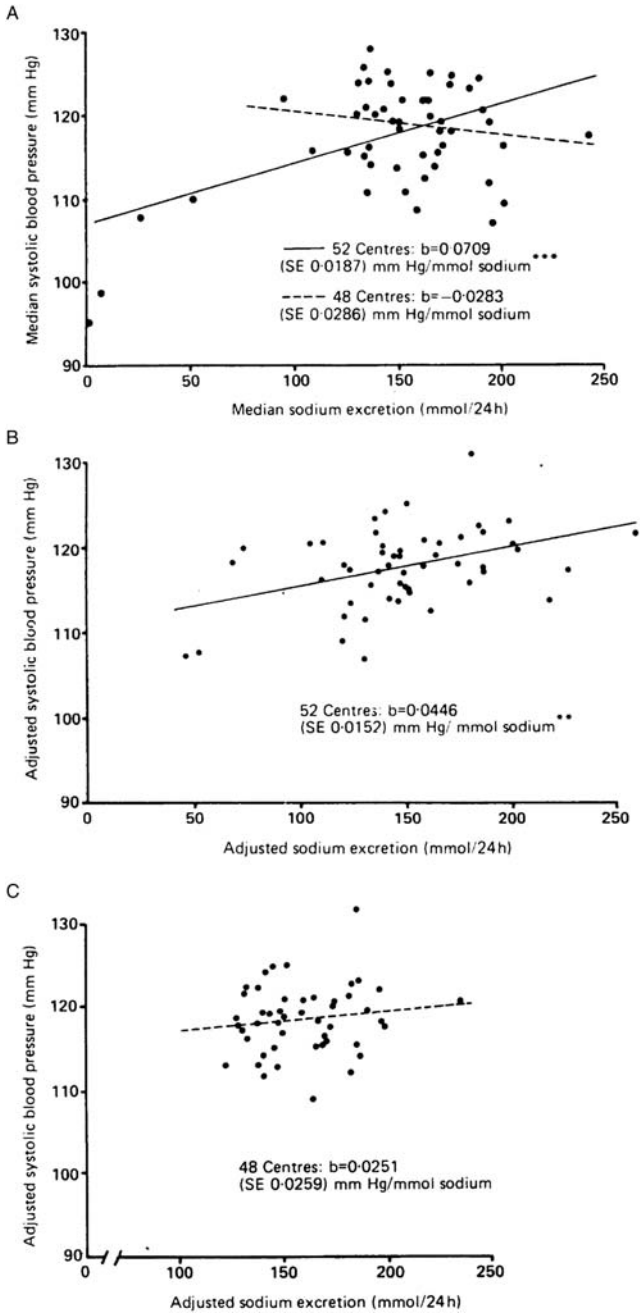


Fig. 3 INTERSALT 1988—within population studies. Cross center plots of median systolic blood pressure and median sodium excretion and fitted regression lines for 52 and 48 centers: (A) standardized for age and sex; (B) also adjusted for body mass index and alcohol intake (52 centers); (C) also adjusted for body mass index and alcohol intake (48 centers). *p < 005; **p < 0001; ***p < 0001. From Intersalt Cooperative Research Group (1988) (reproduced with permission).

them are merely mentioned here by referring to recent reviews and meta-analyses (Fuchs and Fuchs, 2021; Roerecke et al., 2017; Babu et al., 2018; Zhou et al., 2018; Arabshahi et al., 2014) but not discussed any further.

There is ample epidemiological evidence from the literature that high potassium intake is protective against hypertension. As an example, an observational cross-sectional study of 1996 done in 3239 elderly people in The Netherlands showed that an increase in dietary potassium intake of 1 g/day, based on food frequency questionnaires, was associated with a 1.2 mmHg lower systolic and a 0.8 mmHg lower diastolic blood pressure (Geleijnse et al., 1996). Note that dietary recalls are more reliable for the assessment of potassium intake than of sodium intake. A very recent cross-sectional analysis of the NHANES database in the US involving 16,684 adults aged >20 years confirmed a correlation between the highest quartile of potassium intake and hypertension (Cheteu Wabo et al., 2022). This was also true for the sodium/potassium ratio. The study with the largest sample size in this respect is PURE,

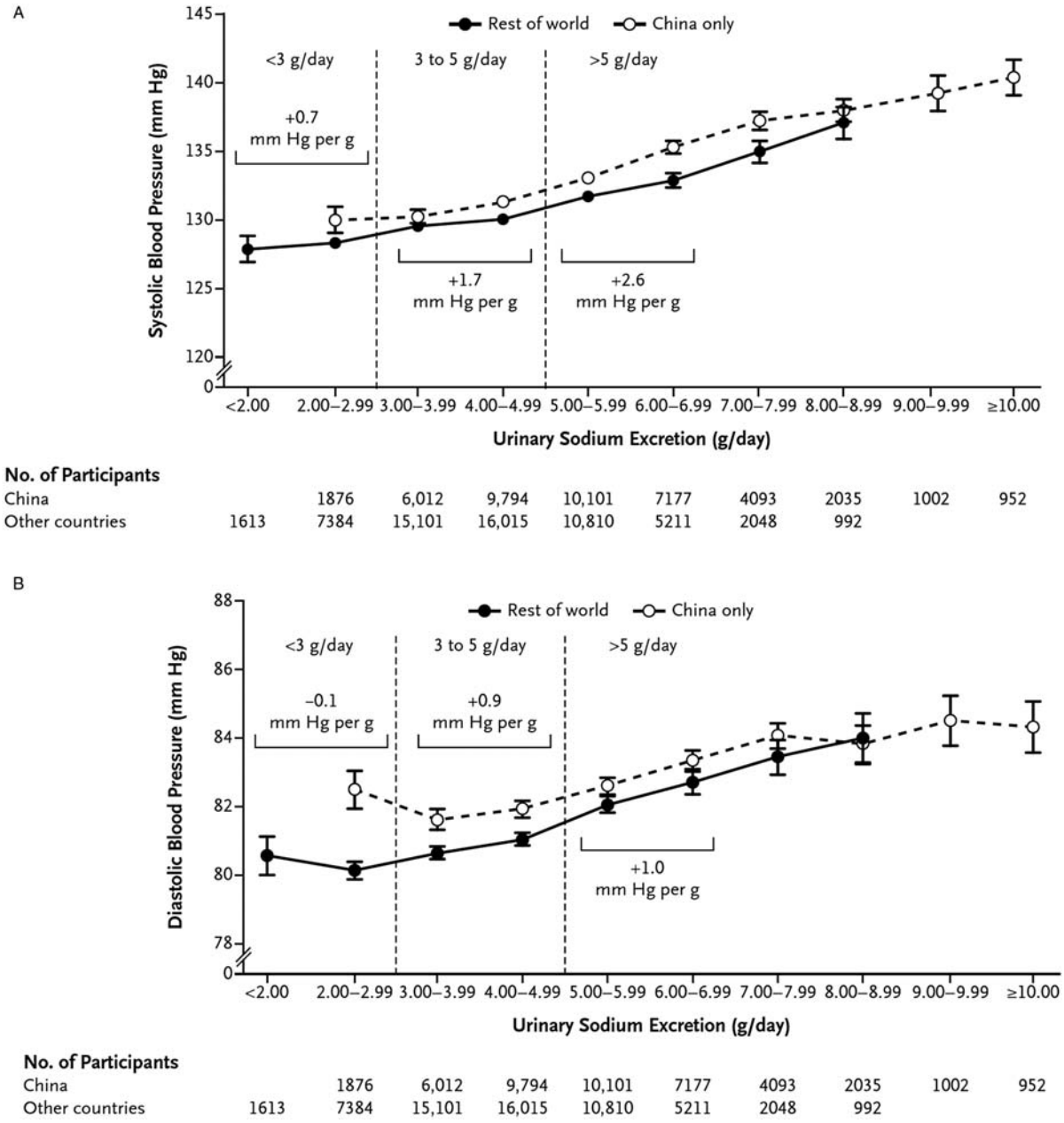


Fig. 4 Mean systolic and diastolic blood pressure according to sodium excretion (Mente et al., 2014). The analysis was adjusted for age, sex, body-mass index, educational level, alcohol intake, and geographic region. Changes in blood pressure are shown for sodium excretion of less than 3 g per day, excretion of 3–5 g per day, and excretion of more than 5 g per day. Persons with extremely low or extremely high sodium excretion are included in the figure. In China, 218 persons with excretion of less than 2 g per day were included in the group with excretion of 2.00–2.99 g per day, and 482 persons with excretion of more than 11 g per day were included in the group with excretion of 10.00 g or more per day. In other countries, 235 persons with excretion of 9.00–9.99 g per day and 112 persons with excretion of 10 g or more per day were included in the group with excretion of 8.00–8.99 g per day. I bars indicate 95% confidence intervals. From Mente et al. (2014) (reproduced with permission).

a worldwide observational study done in 102,216 adult subjects, mean age 50 (Mente et al., 2014). Based on spot urine samples, the authors found a significant inverse correlation between urinary potassium and blood pressure. For each increment of 1 g in estimated potassium excretion per day, there was a decrement of 0.75 mmHg in systolic blood pressure.

Intervention studies

Epidemiological studies are able identify associations between lifestyle factors and disease but cannot demonstrate causal relationships. Therefore, several intervention studies were subsequently performed, albeit their number is much smaller than that of

published epidemiological studies. Not surprisingly, even the intervention studies led to conflicting results in that some of them found significant reductions in both normotensive and hypertensive people when salt intake was reduced, whereas others failed to observe an effect. Potential explanations include differences in sample size, age, initial blood pressure, study duration, and last but not least compliance with dietary counseling and educational intervention. We present here three conflicting reports.

The TOHP-II trial is the largest to examine the effect of long-term sodium reduction on blood pressure, with a mean follow-up of 36 months (Anonymous, 1997). The trial was done in the 1990s. The authors recruited 2382 people, aged 30–54 years. Baseline diastolic blood pressure was 83–89 mmHg, and systolic blood pressure < 140 mm Hg. The participants had a body mass index representing 110%–165% of desirable body weight. A 2 × 2 factorial design was used to evaluate an intervention for weight loss concomitantly. Mean daily intake was 3.1 g sodium (7.8 g salt) at 18 months, and 3.2 g sodium (8 g salt) at 36 months despite targeting a sodium intake of <1.8 g (4.5 g salt) in the intervention group. This indicated that the target of <2.3 g (5.8 g salt) has not been achieved even in this controlled setting of a clinical trial in which substantial efforts were made to lower sodium intake. In the control group, the mean daily intake of sodium was 3.9 g (9.8 g salt) at 18 months, and 4.0 g (10 g salt) at 36 months. The mean difference in systolic blood pressure between sodium intake groups was 2.9 mmHg at 6 months, 2.0 mmHg at 18 months, and 1.2 mmHg at 36 months. In the intervention group, the reduction in prevalence of hypertension also diminished over time. The primary outcome measure in TOHP-II, mean change in diastolic blood pressure, was no different between sodium reduction and control groups at 36 months.

The TONE trial (Whelton et al., 1998) was another counseling/educational intervention. It included 975 elderly individuals, age 60–80, whose hypertension was controlled by a single antihypertensive drug. More than half of the participants were obese. The authors evaluated both sodium reduction and weight loss in 2 × 2 factorial trial design, with the inclusion of control subjects. In the non-obese group, the effect of sodium reduction was compared to corresponding controls. After 90 days, systolic blood pressure was significantly reduced by 3.4 mmHg in the intervention group. Antihypertensive therapy withdrawal began after 90 days of the intervention. Thirty months later, more patients in the sodium intervention group were off antihypertensive therapy or had a systolic blood pressure <150 mmHg and a diastolic blood pressure <90 mmHg. Specifically, average systolic and diastolic BPs were 131 and 74 mmHg, respectively, for the 127 remaining participants assigned to sodium reduction, as compared to 134 and 75 mm Hg for the 83 remaining participants assigned to usual care. Thus here again, achieved differences were rather modest. There was no evidence of a difference in cardiovascular disease.

That compliance is a major issue has been demonstrated in a controlled intervention trial conducted in two towns of Belgium. The trial was aimed at exploring the feasibility of reducing salt intake in the general population by means of simple and inexpensive intervention techniques (Staessen et al., 1988). A total of 2211 people accepted to participate at the study, which lasted 5 years. The inhabitants of the intervention town underwent multiple types of instruction how to reduce salt intake, whereas those of the control town remained on their usual unrestricted salt consumption. The intervention led to a reduction of urinary sodium excretion by 17 mmol/24 h (1.0 g salt/24 h) in adult women in the intervention town, which differed significantly from the concurrent trend of an increase by 8 mmol/24 h (0.5 g salt/24 h) in women of the control town. However, both systolic (-7.5 compared with -7.9 mmHg) and diastolic (-2.3 compared with -3.0 mmHg) blood pressures declined to the same extent in women of the two towns. A decrease in urinary sodium excretion of 12 mmol/24 h (0.7 g/24 h salt), and in systolic and diastolic blood pressures of 5.6 mmHg and 2.4 mmHg respectively was observed in adult men of the intervention town. However, these trends were the same in the control town, with a decrease in urinary sodium excretion of 12 mmol/24 h (0.7 g/24 h salt) and in blood pressure values 4.9 mmHg and 0.2 mmHg, respectively. The authors concluded that salt intake in the long-run cannot be restricted to less than 5 g/24 h. While a more moderate salt restriction might constitute a more realistic goal its influence on blood pressure in the community at large would probably be trivial.

Post-hoc analyses allow a more systematic examination of intervention trials although the quality of many among them is frequently questionable. In one post-hoc analysis of the 1990s (Law et al., 1991), salt reduction was efficacious in studies lasting 5 weeks or longer (33 of 78 trials analyzed) but not in those lasting less than 5 weeks (45 of 78 trials analyzed). However, in another analysis where the author took care to include only those studies that had an appropriate design, he reached the opposite conclusion (Swales, 1991). There was no effect of salt reduction in studies which lasted more than 4 weeks, a conclusion reinforced by three subsequent meta-analyses (Graudal et al., 1998; Hooper et al., 2002; Midgley et al., 1996). However, meta-analyses by other groups reached opposite conclusions (Dickinson et al., 2006; He and MacGregor, 2006; Strazzullo et al., 2009). Therefore one should exercise extreme caution when relying on observational data alone, given the complexity of the interaction with numerous other endogenous and exogenous confounding variables.

The results of the prospective randomized DASH-2 trial reported in 2001 were considered by its authors to solve this issue definitively (Sacks et al., 2001). The authors allocated 412 adult volunteers from USA to six different treatment groups. The subjects ingested for a 1-month period either a typical, high caloric US diet or a diet enriched in vegetables, fruits and low-fat dairy products ("DASH diet"). In addition, they were allocated to a salt intake of either 9 g, 6 g or 3 g per day. The most marked decrease in blood pressure was obtained with the DASH diet together with usual sodium intake. The additional blood pressure reduction induced by concomitant salt restriction was much smaller (**Fig. 5**). Specifically, the decrease of systolic blood pressure achieved with the DASH diet, as compared to that on standard US diet, was 5.9 mmHg. The decrease observed in response to a reduction of salt intake from 9 g to 6 g per day was only 1.3 mmHg, and the decrease in response to a reduction of salt intake down to 3 g per day was 1.7 mmHg, knowing that the latter drastic restriction is unacceptable for most people in the long run. It is also noteworthy that the DASH study has been performed in a selected cohort of US individuals with high salt sensitivity, i.e. a percentage of black and overweight subjects strikingly higher than in the population at large. Extrapolations from these findings to general population and to people outside the US may therefore not be justified. Finally, the primary criterion of evaluation was blood pressure, i.e. an intermediary

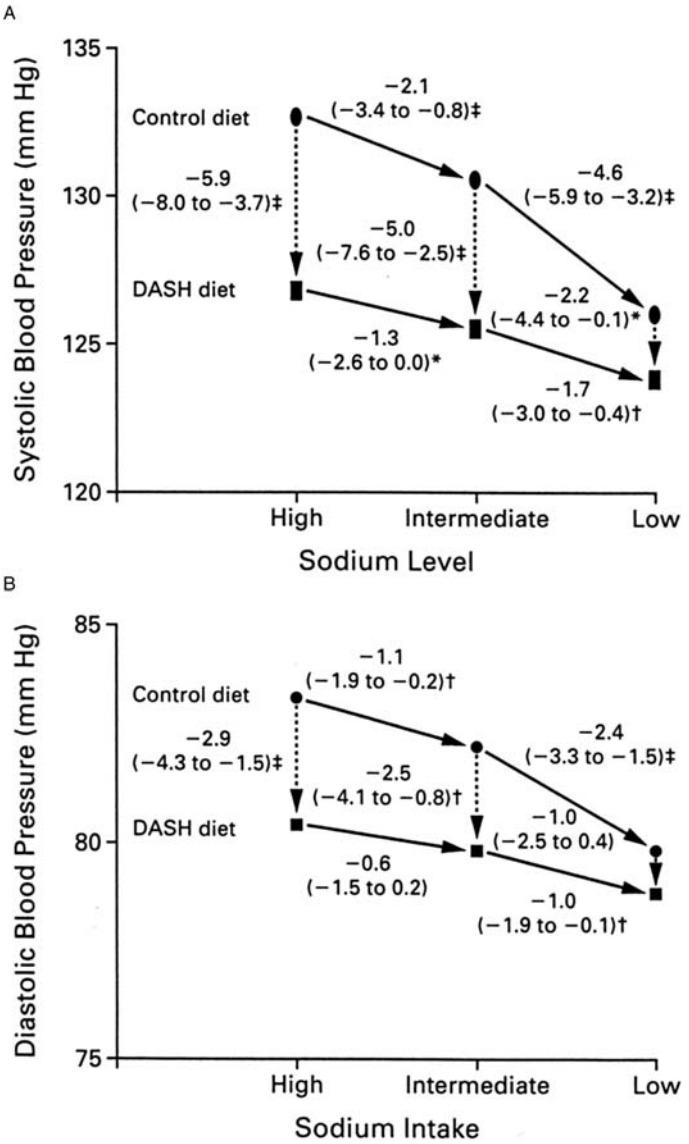


Fig. 5 Sodium reduction, the DASH diet, and changes in systolic blood pressure. Beneficial effects of the DASH diet and reduced intake of sodium on systolic and diastolic blood pressure in 412 patients with mild hypertension, age >45 years. The participants were randomly assigned to follow a DASH diet or a typical U.S. diet for 30 days. During that period, each group consumed, in a crossover design, three versions of the diet adjusted for daily sodium content: high, 8.8 g NaCl/day: intermediate, 5.8 NaCl/day; low, 3.0 NaCl/day. Body weight was held constant. The two downward-sloping arrows on the left depict the effect of intermediate sodium intake as compared with higher sodium intake, and the two downward-sloping arrows on the right depict the effect of lower sodium intake as compared with intermediate sodium intake. The dotted lines show the effect of the DASH diet as compared with the typical U.S. diet at each level of dietary sodium. Numbers shown represent the mean changes with 95% confidence intervals. From Sacks et al. (2001) (reproduced with permission).

endpoint, not a hard endpoint such as cardiovascular morbidity or mortality. DASH-2 compared a diet rich in fruits and vegetables, and consequently rich in potassium, to the usual US diet less rich in these nutrients. The observation of a more marked blood pressure decrease with the diet rich in fruits and vegetables as compared to the two reduction steps in salt intake is in line with the view that potassium may be a more important dietary factor than sodium in controlling blood pressure. In fact numerous prospective trials have shown that a high potassium intake is protective against hypertension, whereas a low intake favors its development or aggravates already existing hypertension, as summarized in a recent meta-analysis by Filippini et al. (2020).

Nutrients such as calcium and magnesium also contribute to the regulation of blood pressure (Cheteu Wabo et al., 2022). Sufficient to high intakes of these nutrients are possibly protective against high blood pressure but here again, literature reports are highly controversial. The role of diet in blood pressure regulation involves a number of interrelated nutrients and, as food intake varies widely, the arbitrary modification of one or the other among them may be beneficial or detrimental, especially when considering hard outcomes

such as cardiovascular events and mortality (see below). Several other exogenous factors play a role as well, either via an interaction of salt in controlling blood pressure or independently (see above). As an example, regular physical activity is protective whereas stress, obesity, and excessive alcohol consumption and many others favor arterial hypertension (Muntzel and Drüeke, 1992).

Salt intake and cardiovascular outcomes

In the last decade, attention has shifted from the relationship between salt intake and hypertension to that between salt intake and hard clinical outcomes, i.e. cardiovascular events, cardiovascular mortality and global mortality. Although randomized controlled trials with morbidity and mortality endpoints are the gold standard they are highly impractical to address this issue. Only few such trials have been conducted in studies with small sample size, relatively short duration, and most importantly poor protocol adherence. Before discussing these trials we shall present the major observational studies in this field. Some among them are presented individually, all the others are summarized in meta-analyses.

In 2010, Alderman examined the association of sodium consumption with clinical outcomes in 13 observational cohort studies, including more than 100,000 participants and more than 800 morbidity and mortality events (Alderman, 2010). The results were conflicting, as shown in **Table 3**. In two studies whose participants ingested high amounts of sodium (Finland, mean value 4600 mg/day; Japan, mean value 5428 mg/day corresponding to 11.6 and 13.7 g salt per day), there was a positive association between salt intake and cardiovascular disease events (Nagata et al., 2004; Tuomilehto et al., 2001). In the eleven other studies with lower mean sodium intakes comprised between 2070 and 3680 mg/day (corresponding to salt intakes between 5.2 and 9.3 g salt per day), two also detected a positive association between salt and cardiovascular disease events, but each of them was a post hoc subgroup analysis with findings not entirely consistent with the overall study results (Hooper et al., 2002). In five among the studies, there was no association between salt intake and clinical outcome. Remarkably, in the four remaining studies sodium intake was inversely associated with cardiovascular disease events. A J-shaped curve, in which the most favorable dietary sodium range would be around 3450 mg/day (8.5 g salt/day)—with possible cardiovascular risk above and below this level—might best explain these apparently contradictory observations (Alderman, 2007).

In another meta-analysis including several of the observational studies examined by Alderman, but without stratification taking into account differences in daily sodium intakes, Strazzullo et al. detected a positive association of sodium intake with stroke, a finding driven by studies in people with high salt intake (Strazzullo et al., 2009). However, the authors failed to observe an association between sodium consumption and cardiovascular events. Total mortality was not reported. Not surprisingly, given different populations, different diets and sodium intakes, different methods and conflicting results, more than one explanatory hypothesis has emerged.

Bibbins-Domingo et al. performed yet another meta-analysis at that time, based on the above mentioned and other observational data (Bibbins-Domingo et al., 2010). Using a computer-simulation model to estimate the effects of low sodium intake they postulated that a reduction of daily dietary salt intake by 3 g might reduce the annual number of new cases of chronic heart disease by 60,000–120,000, stroke by 32,000–66,000, and myocardial infarction by 54,000–99,000, and allow the annual number of deaths from any cause to fall by 44,000–92,000. However, these estimates rely on several unverified assumptions, in particular the assumption that reducing salt intake lowers blood pressure and hence the risk of stroke and chronic heart disease (Appel and Anderson, 2010).

Table 3 Observational studies linking dietary sodium to cardiovascular disease outcomes[a].

Lower salt intake associated with cardiovascular events
Worksite Hypertension Study, 1995
National Health and Nutrition Examination Survey
I, 1998
II, 2006
III, 2008
O'Donnell et al. (2014, 2019)
Salt intake had no association with cardiovascular events
Honolulu Heart Study, 1997
Scottish Heart Health Study, 1997
Health Professional Study follow-up, 1997
Multiple Risk Factor Intervention Trial, 2000
Increased salt intake associated with cardiovascular events
National Health and Nutrition Examination Survey, involving obese patients, 2000
Finnish Heart Study, 2001
Takayama, 2005
O'Donnell et al. (2014, 2019)
Ma et al. (2022)

[a]Unless otherwise indicated, all studies listed herein are referenced in Alderman (2007).

More recent observational studies found either a negative, a U-shaped, or a J-shaped association between salt consumption and mortality, both in general population and in patients with diabetes or heart failure. We discuss several among them here.

Stolarz-Skrzypek et al. found a negative association (Stolarz-Skrzypek et al., 2011). They examined 3681 participants from two European studies on the effects of genes vs. environment on hypertension. None of the participants had cardiovascular disease, 2096 were normotensive at baseline, and 1499 had blood pressure and urinary sodium excretion measured at baseline and after a median follow-up of 7.9 years. Cardiovascular disease deaths decreased across increasing tertiles of 24 h sodium excretion, from 50 deaths in the low, 24 in the medium, and 10 in the high sodium excretion group, resulting in respective death rates of 4.1%, 1.9%, and 0.8%. Baseline sodium excretion predicted neither total mortality nor fatal combined with nonfatal cardiovascular events. Remarkably, the risk of hypertension did not increase across increasing urine sodium tertiles. However, in multivariable-adjusted analyses, a 100-mmol increase in sodium excretion was associated with a significant, albeit only 1.71 mmHg increase in systolic blood pressure. Thus although systolic blood pressure changes over time associated with urinary sodium changes, this did not translate into a higher risk of hypertension or cardiovascular complications. Most importantly, lower sodium excretion was associated with higher cardiovascular mortality.

O'Donnell et al. observed a J-shaped association in a first observational study (O'Donnell et al., 2011). In this post hoc analysis of 28,880 patients with cardiovascular disease or diabetes, who had been enrolled in two different drug trials, they found that both high and low sodium intakes, based on spot urine sodium samples, were correlated with an increase in cardiovascular events, corresponding to a J-shaped curve. As compared to participants with a baseline daily sodium excretion of 4–6 g (10–15 g salt), those participants who excreted more than 6 g sodium (15 g salt) and those who excreted less than 4 g sodium (10 g salt) exhibited an increase in cardiovascular deaths, strokes, or heart attacks.

The same group of authors from McMaster in Canada subsequently devoted three other observational studies of very large sample size on this issue. In the first, O'Donnell et al. examined the association between urinary sodium and potassium excretion and the composite outcome of death and major cardiovascular events in 101,945 persons from 17 countries (O'Donnell et al., 2014). Daily sodium and potassium excretion was estimated based on morning fasting urine samples. Mean follow-up was 3.7 years. Compared with an excretion of 4.00–5.99 g sodium (10–15 g salt) per day (reference range), a higher excretion ($\geq$7.00 g sodium or 17.5 g salt per day) was associated with an increased risk of the composite outcome, as well as increased risks of death and major cardiovascular events (**Fig. 6**). The association between a high sodium excretion and the composite outcome was strongest among participants with hypertension, with an increased risk at an excretion of 6.00 g sodium (15 g salt) or more per day. As compared with the reference range, an excretion that was below 3.00 g sodium (7.5 g salt) per day was also associated with an increased risk of the composite outcome. As compared with a potassium excretion that was less than 1.50 g per day, higher potassium excretion was associated with a reduced risk of the composite outcome.

In the second study of the McMaster group, Mente et al. aimed to explore whether the association between sodium intake and cardiovascular disease events and all-cause mortality was modified by hypertension status (Mente et al., 2018). In a pooled analysis, they included 133,118 individuals (63,559 with and 69,559 without hypertension), median age 55 years, from 49 countries in four large prospective studies and estimated daily urinary sodium excretion based on spot urine sampling. Median follow-up was 4.2 years. They found that increased sodium intake was associated with greater increases in systolic blood pressure in hypertensive than in normotensive people, with a 1.22 mmHg change per g sodium (2.5 g salt). In the hypertensive patients a daily excretion of 7 g sodium (17.5 g salt) or more and less than 3 g sodium (7.5 g salt) were both associated with increased risk compared with a daily excretion of 4–5 g sodium (10–12.5 g salt). In the normotensive individuals, higher sodium excretion was not associated with risk of the primary composite outcome when compared with 4–5 g sodium (10–12.5 g salt) excretion, whereas an excretion of less than 3 g sodium (7.5 g salt) was associated with increased risk. These findings are again in keeping with the view of a U-shaped relationship between salt consumption and hard clinical outcomes.

In the third study from the McMaster group, O'Donnell et al. evaluated the joint association of sodium and potassium urinary excretion with cardiovascular events and mortality, in the context of current World Health Organization recommendations for daily intake (<2.0 g sodium, >3.5 g potassium). They used data from an international prospective cohort study of 103,570 adult individuals people who provided morning fasting urine samples (O'Donnell et al., 2019). Mean daily estimated sodium and potassium urinary excretion were 4.93 g and 2.12 g, respectively. After a median follow-up of 8.2 years, 7884 (6.1%) participants had died or experienced a major cardiovascular event. Increasing urinary sodium excretion was positively associated with increasing potassium excretion, and only 0.002% had a concomitant daily urinary excretion of <2.0 g sodium (<5 g salt) and >3.5 g potassium. The authors observed a J-shaped association of sodium excretion and inverse association of potassium excretion with death and cardiovascular events. For joint sodium and potassium excretion categories, the lowest risk of death and cardiovascular events occurred in the group with moderate sodium excretion and higher potassium excretion (21.9% of cohort). Compared with this reference group, the combinations of low potassium with low sodium excretion and low potassium with high sodium excretion were associated with the highest risk, followed by low sodium excretion and high sodium excretion among those with potassium excretion greater than the median. Higher potassium excretion attenuated the increased cardiovascular risk associated with high sodium excretion. The authors stated that the simultaneous target of low sodium intake (<2 g/day) with high potassium intake (>3.5 g/day) is extremely uncommon. They concluded that combined moderate sodium intake (3–5 g/day) with high potassium intake is associated with the lowest risk of mortality and cardiovascular events.

More recently, Messerli et al. again examined the hypothesis that dietary sodium intake is a risk factor for cardiovascular disease and premature death (Messerli et al., 2021). Their analysis was done in 181 countries worldwide, using several sources of data. Sodium intake was estimated according to a previous study of global, regional and national sodium intakes based on either urinary

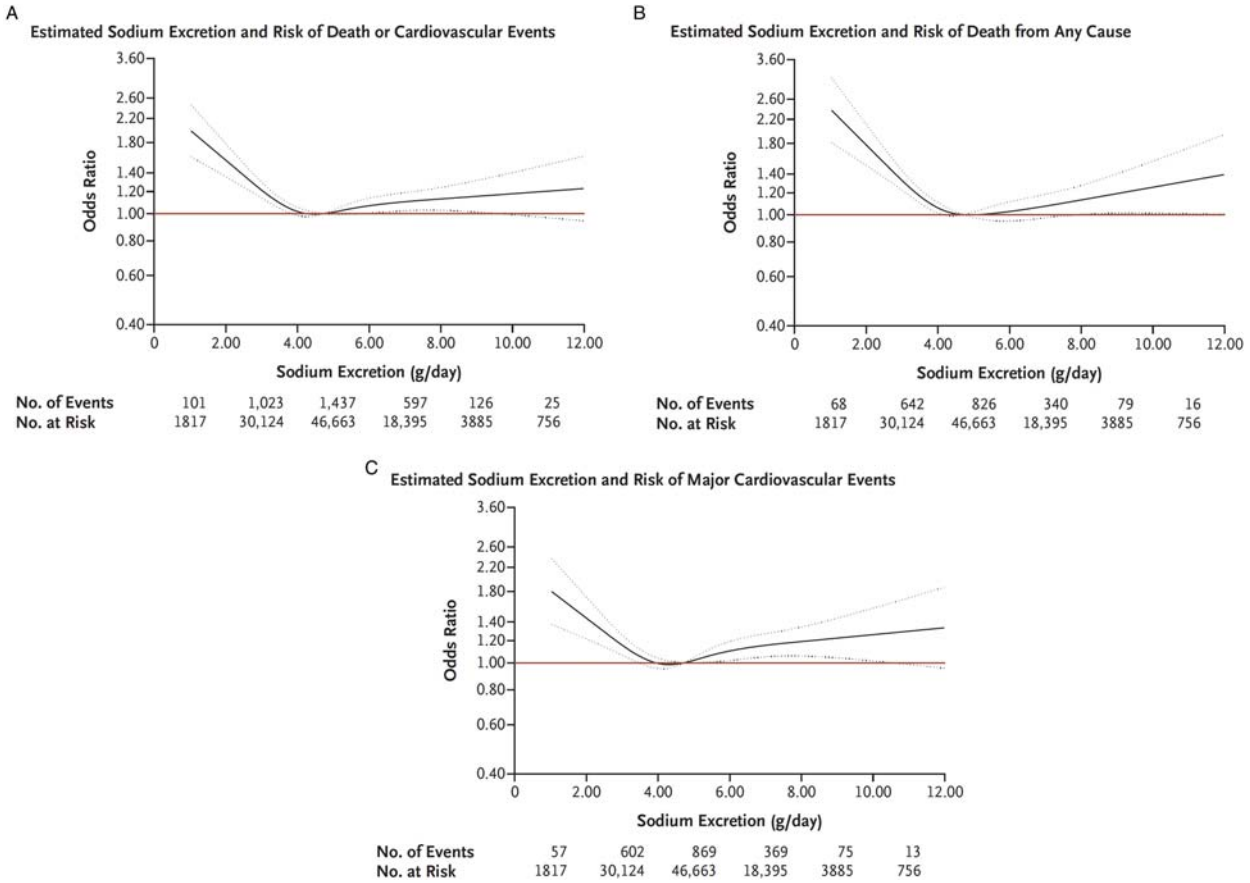


Fig. 6 Association of estimated 24 h urinary sodium excretion with risk of death and major cardiovascular events. Panel A shows a restricted-cubic-spline plot of the association between estimated 24 h urinary sodium excretion and the composite outcome of death from any cause and major cardiovascular events. The spline curve is truncated at 12.00 g per day (event rate among participants with sodium excretion >12.00 g per day, 8 events in 305 participants). Panel B shows a restricted cubic-spline plot of the association between estimated sodium excretion and death. The event rate among participants with sodium excretion of more than 12.00 g per day was 5 events in 305 participants. Panel C shows a restricted-cubic-spline plot of the association between estimated sodium excretion and major cardiovascular events (defined as death from cardiovascular causes, myocardial infarction, stroke, or heart failure). The event rate among participants with sodium excretion of more than 12.00 g per day was 6 events in 305 participants. All plots were adjusted for age, sex, geographic region, educational level, ancestry (Asian vs. non-Asian), alcohol intake, body-mass index, and status with respect to diabetes mellitus, history of cardiovascular events, and current smoking. Dashed lines indicate 95% confidence intervals. The median sodium excretion (4.72 g per day) was the reference standard, indicated by the red line. To convert the values for estimated sodium excretion to salt intake in grams per day, multiply by 2.5. From O'Donnell et al. (2014) (reproduced with permission).

sodium or dietary sodium recall data (Powles et al., 2013). The authors correlated age-standardized estimates of country-specific average sodium consumption with healthy life expectancy at birth and at age of 60 years, death due to non-communicable diseases, and all-cause mortality for the year of 2010, after appropriate adjustments. They found a positive correlation between sodium intake and healthy life expectancy at birth, as well as healthy life expectancy at age 60 but not for death due to non-communicable diseases. Conversely, all-cause mortality correlated inversely with sodium intake (Fig. 7). In a sensitivity analysis restricted to 46 countries in the highest income class, sodium intake continued to correlate positively with healthy life expectancy at birth and inversely with all-cause mortality. They concluded that the observation of sodium intake correlating positively with life expectancy and inversely with all-cause mortality worldwide, and also in high-income countries alone, argued against dietary sodium intake being a culprit of curtailing life span or a risk factor for premature death.

Two other studies of more modest sample size examined this issue in persons with type 1 and type 2 diabetes, respectively. Ekinci et al. measured 24 hr urinary sodium excretion at a single visit in 638 patients with type 2 diabetes who were subsequently followed for 10 years (Ekinci et al., 2011). Mean baseline urinary sodium excretion was 184 mmol (10.7 g salt) per day. At end of follow-up, they recorded 175 deaths with 43% among them having a cardiovascular cause. All-cause mortality was lower by 28% for every 24 hr sodium excretion higher by 100 mmol. Cardiovascular mortality also inversely associated with sodium excretion. This led to them to hypothesize that low salt intake might cause adverse effects. However, as pointed out by Lambers Heerspink et al. (2012), patient characteristics were not balanced since the low salt group had longer diabetes duration, lower body mass index, lower glomerular filtration rate, higher prevalence of macrovascular disease, and received less frequently renin-angiotensin-aldosterone blockade and insulin treatment.

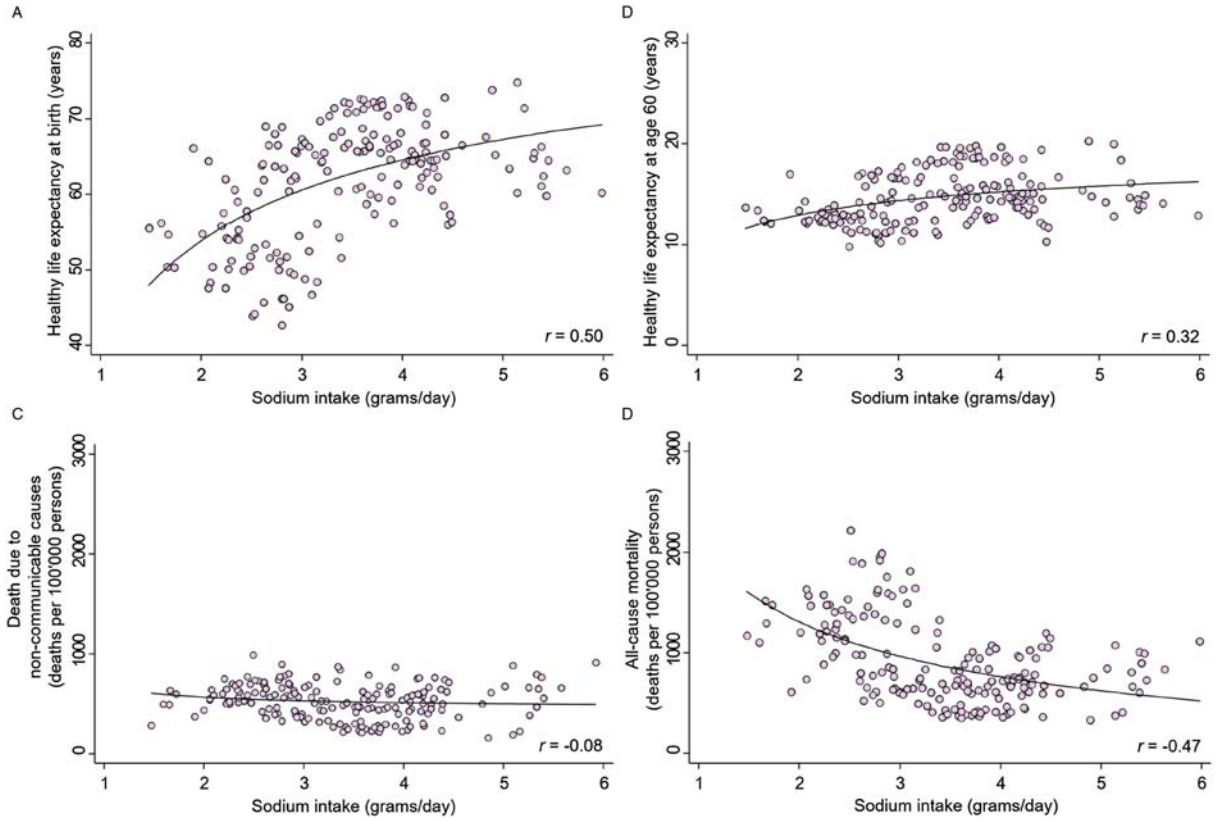


Fig. 7 Scatterplots of sodium intake (g/day) for the year of 2010 and healthy life expectancy at birth, healthy life expectancy at age 60, age-standardized death rates of death due to non-communicable diseases and all-cause mortality in 181 countries worldwide (A–D, respectively). From Messerli et al. (2021) (reproduced with permission).

Thomas et al. examined the role of salt intake in outcomes of 2807 adult patients with type 1 diabetes who participated at the Finn Diane cohort and were followed for 10 years (Thomas et al., 2011). The authors found that urinary sodium excretion was non-linearly associated with all-cause mortality in that both low and high urinary sodium excretion was associated with higher mortality. The nadir of mortality was at a daily consumption of 100 mmol sodium (5.8 g salt). The association persisted after adjustment for confounders including age, gender, diabetes duration, glomerular filtration rate and albuminuria. Lambers Heerspink et al. criticized this study as well, stating that the low salt cohort included more females, patients with lower BMI, less anti-hypertensive medication and more macrovascular disease, all findings being consistent with a higher cardiovascular risk (Lambers Heerspink et al., 2012).

Taken together, these observations demonstrate that studies of possible associations between salt intake and hard clinical outcomes done normotensive people of the population at large should not be lumped together with studies done in patients with hypertension, diabetes or other diseases. They further show that changes of cardiovascular and global outcomes in response to changes in salt intake do not necessarily parallel changes in blood pressure. As already mentioned, a major criticism of the studies from the McMaster group is that spot urine analyses are less reliable than 24 h urine collections since the former tend to overestimate sodium excretion at lower levels and to underestimate sodium excretion at higher levels (Naser et al., 2021).

In the year 2022, two additional major studies on the relation between sodium intake and cardiovascular outcomes were reported.

First of all, Ma et al. reported the results of an observational study on the relation between sodium intake and cardiovascular disease. They combined individual data from 10,709 participants of six different cohorts (Ma et al., 2022). Mean age was 51.5 years, 54.2% were women, and the majority among them had normal blood pressure at baseline. All participants had at least two 24 h urine samples. The authors ascertained 571 cardiovascular events during a median study follow-up of 8.8 years. Median 24 h urinary sodium excretion was 3270 mg (8.2 g salt). Higher sodium excretion, lower potassium excretion, and a higher sodium-to-potassium ratio were all associated with a higher cardiovascular risk in analyses that were controlled for confounding factors (Fig. 8). In analyses that compared quartile 4 of the urinary biomarker (highest) with quartile 1 (lowest), the hazard ratios were 1.60, 1.19 to 2.14 for sodium excretion, 0.69 for potassium excretion, and 1.62 for the sodium-to-potassium ratio. Each daily increment of 1000 mg sodium excretion (2.5 g salt excretion) was associated with an 18% increase in cardiovascular risk, and each daily increment of 1000 mg potassium excretion was associated with an 18% decrease in risk. Thus higher sodium and lower potassium

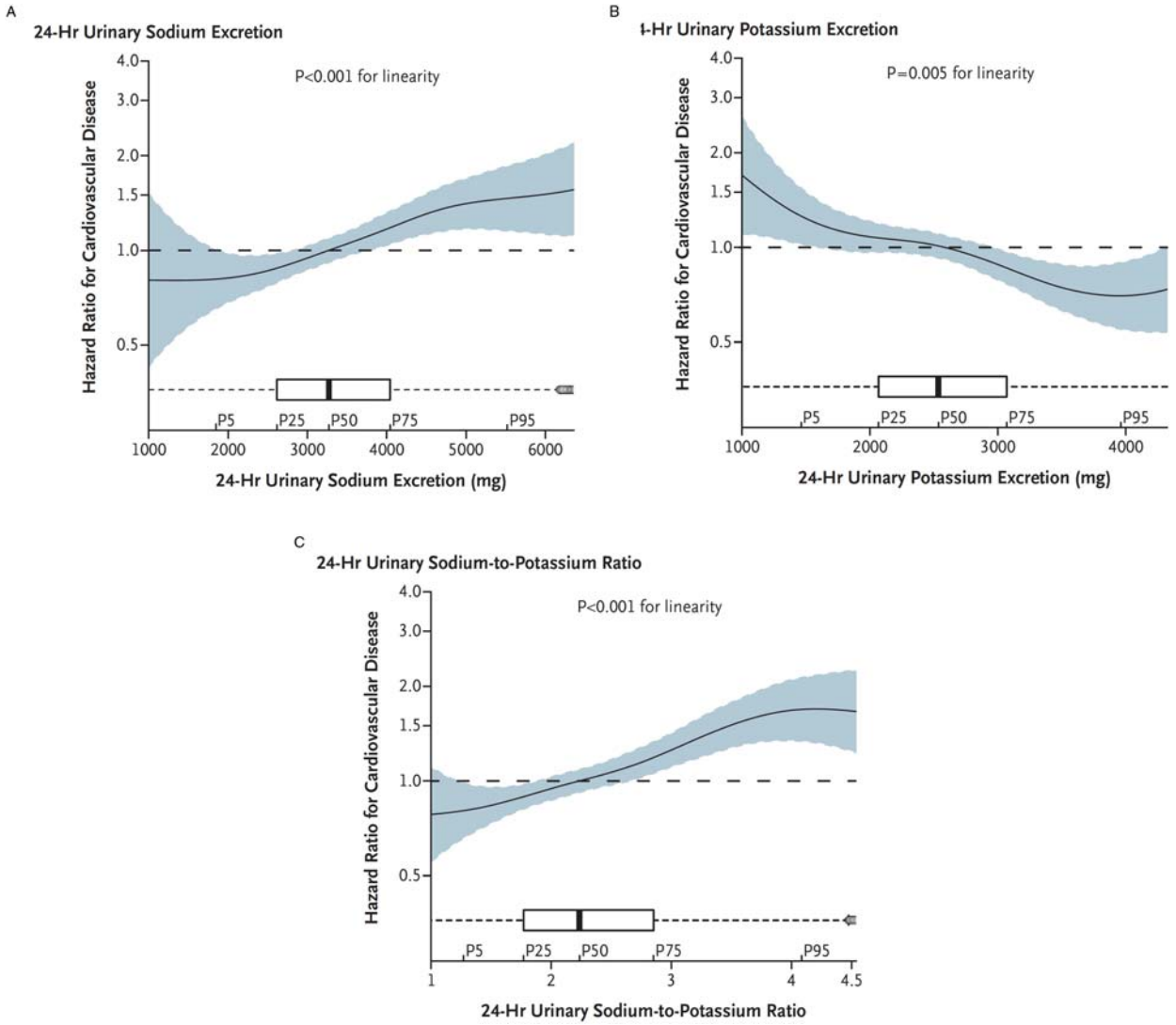


Fig. 8 Spline plots for the associations of 24 hour urinary sodium and potassium excretion and sodium to- potassium ratio with cardiovascular risk. The spline analysis of pooled data supported a linear association over the range of sodium excretion (Panel A; 5th to 95th percentile, 1846–5520 mg) and potassium excretion (Panel B; 5th to 95th percentile, 1462–3961 mg) within the overall study population. The sodium-to-potassium ratio (Panel C) was assessed on the basis of the sodium and potassium excretion measured in millimoles. Hazard ratios were estimated from Cox models stratified according to study cohort with adjustment for age, sex, race, educational level, height, body-mass index, alcohol consumption, smoking status, physical activity, history of diabetes and elevated cholesterol status, family history of cardiovascular disease, and mutual adjustment for 24 h urinary potassium and sodium excretions. Shaded areas indicate 95% confidence intervals, and the dashed line at 1.0 indicates the reference. Box plots at the bottom of the graphs show the distributions of the urinary biomarker. The vertical bar indicates the median, and the ends of the box the interquartile range; the whiskers (dashed lines) extend to values no farther than 1.5 times the interquartile range (which may be past the graphed area), and dots indicate values that are farther than 1.5 times the interquartile range. The 5th, 25th, 50th, 75th, and 95th percentiles (P5, P25, P50, P75, and P95, respectively) are shown at the bottom of each graph. From Ma et al. (2022) (reproduced with permission).

intakes, as measured in multiple 24 h urine samples, were associated in a dose–response manner with a higher cardiovascular risk. The authors concluded that their findings supported reducing sodium intake and increasing potassium intake from current levels. However, although the urinary sodium measurement method appeared to be optimal in the combined cohort studies examined by Ma et al. (2022), the post-hoc character of the analysis had major limitations including the inclusion of very heterogeneous study cohorts and an observation time period spanning 2 decades. Most importantly, the mean value of the lowest quartile of urinary sodium excretion was 2212 mg (5.5 g salt). This means that half of the participants in the lowest quartile had a daily sodium intake above the one recommended by WHO (<2300 mg sodium or 5.8 g salt). Thus the study by Ma et al. fails to inform the impact of the presently recommended very low salt intake on cardiovascular events and mortality. This means that the left side of a potential U-curve relationship has not been adequately assessed. It is noteworthy that the report by Ma et al. does not provide any information on the relation between sodium intake and blood pressure.

Second, Ezekowitz et al. reported the results of an open-label, randomized, controlled clinical trial to test the effects of dietary sodium reduction for patients with heart failure on the incidence of future clinical events (Ezekowitz et al., 2022). They randomly assigned 397 patients to a low sodium diet (<100 mmol/day [<3.8 g salt/day]) and 409 patients to usual care. Adherence to the diet was not based on urinary sodium measurements, but only estimated by 3-day food records. The authors found that dietary sodium restriction did not reduce the clinical composite outcome of all-cause mortality, cardiovascular-related hospitalization, or cardiovascular-related emergency department visits compared with usual care over 12 months. An improvement in the patient reported outcome of quality of life and clinician assessed New York Heart Association functional class was noted; however, no significant between-group difference was seen in 6 min walk distance. It must be noted that the study was prematurely stopped due to futility, and therefore only 806 (81%) of the estimated 992 patients needed were included in the trial. The authors concluded that in ambulatory patients with heart failure, a dietary intervention to reduce sodium intake did not reduce clinical events.

Conclusions

In agreement with O'Donnell et al. (2020) we would like to conclude that the aim of the WHO recommendation to reduce daily sodium (salt) intake to amounts below 2.3 g (5.8 g) in the long run seems unrealistic based on the results of numerous intervention trials conducted at the individual, community, or national level. More modest reductions of salt intake do not allow achieving clinically meaningful reductions of blood pressure, as compared to the efficacy of antihypertensive agents. Although a number of observational data suggest an association of salt intake with blood pressure, cardiovascular risk, and/or mortality risk convincing evidence for this claim is lacking to date. Moderate salt intake in association with higher potassium intake is probably achievable by many people. However, the association with low cardiovascular risk as suggested by some observational studies is at best hypothesis generating. Whether this translates to real life conditions in general population has to be shown in definitive intervention trials. It does not seem that a low sodium diet is able to reduce clinical events in patients with chronic heart failure. For the time being the advice to focus on the established classical risk factors and risk markers seems to be more promising.

Acknowledgments

Disclosures.

TBD reports having received advisor, consultant, speaker and/or travel fees from Astellas and Glaxo-Smith-Kline during the past 36 months (outside of present review article).

PS reports no conflict of interest.

References

Alderman, M.H., 2007. Presidential address: 21st scientific meeting of the international society of hypertension: dietary sodium and cardiovascular disease: the "J"-shaped relation. J. Hypertens. 25 (5), 903–907. https://doi.org/10.1097/HJH.0b013e3280c14394.

Alderman, M.H., 2010. Reducing dietary sodium: the case for caution. JAMA 303 (5), 448–449. https://doi.org/10.1001/jama.2010.69.

Anonymous, 1997. Effects of weight loss and sodium reduction intervention on blood pressure and hypertension incidence in overweight people with high-normal blood pressure. The Trials of Hypertension Prevention, phase II. The Trials of Hypertension Prevention Collaborative Research Group. Arch. Intern. Med. 157 (6), 657–667.

Appel, L.J., Anderson, C.A.M., 2010. Compelling evidence for public health action to reduce salt intake. N. Engl. J. Med. 362 (7), 650–652. https://doi.org/10.1056/NEJMe0910352.

Arabshahi, S., Busingye, D., Subasinghe, A.K., Evans, R.G., Riddell, M.A., Thrift, A.G., 2014. Adiposity has a greater impact on hypertension in lean than not-lean populations: a systematic review and meta-analysis. Eur. J. Epidemiol. 29 (5), 311–324. https://doi.org/10.1007/s10654-014-9911-6. PMID: 24838697.

Babu, G.R., Murthy, G.V.S., Ana, Y., Patel, P., Deepa, R., Neelon, S.E.B., Kinra, S., Reddy, K.S., 2018. Association of obesity with hypertension and type 2 diabetes mellitus in India: a meta-analysis of observational studies. World J. Diabetes 9 (1), 40–52. https://doi.org/10.4239/wjd.v9.i1.40. PMID: 29359028; PMCID: PMC5763039.

Bernstein, A.M., Willett, W.C., 2010. Trends in 24-h urinary sodium excretion in the United States, 1957–2003: a systematic review. Am. J. Clin. Nutr. 92 (5), 1172–1180. https://doi.org/10.3945/ajcn.2010.29367.

Bibbins-Domingo, K., Chertow, G.M., Coxson, P.G., et al., 2010. Projected effect of dietary salt reductions on future cardiovascular disease. N. Engl. J. Med. 362 (7), 590–599. https://doi.org/10.1056/NEJMoa0907355.

Brown, I.J., Tzoulaki, I., Candeias, V., Elliott, P., 2009. Salt intakes around the world: implications for public health. Int. J. Epidemiol. 38 (3), 791–813. https://doi.org/10.1093/ije/dyp139. PMID: 19351697.

Cheteu Wabo, T.M., Wu, X., Sun, C., et al., 2022. Association of dietary calcium, magnesium, sodium, and potassium intake and hypertension: a study on an 8-year dietary intake data from the National Health and Nutrition Examination Survey. Nutr. Res. Pract. 16 (1), 74–93. https://doi.org/10.4162/nrp.2022.16.1.74.

Dahl, L.K., 1958. Salt intake and salt need. N. Engl. J. Med. 258 (23), 1152–1157. https://doi.org/10.1056/NEJM195806052582305 contd.

Dickinson, H.O., Mason, J.M., Nicolson, D.J., et al., 2006. Lifestyle interventions to reduce raised blood pressure: a systematic review of randomized controlled trials. J. Hypertens. 24 (2), 215–233. https://doi.org/10.1097/01.hjh.0000199800.72563.26.

Ekinci, E.I., Clarke, S., Thomas, M.C., et al., 2011. Dietary salt intake and mortality in patients with type 2 diabetes. Diabetes Care 34 (3), 703–709. https://doi.org/10.2337/dc10-1723.

Ezekowitz, J.A., Colin-Ramirez, E., Ross, H., on behalf of the SODIUM-HF Investigators, et al., 2022. Reduction of dietary sodium to less than 100 mmol in heart failure (SODIUM-HF): an international, open-label, randomised, controlled trial. Lancet 399.

Filippini, T., Naska, A., Kasdagli, M.I., et al., 2020. Potassium intake and blood pressure: a dose-response meta-analysis of randomized controlled trials. J. Am. Heart Assoc. 9 (12), e015719. https://doi.org/10.1161/JAHA.119.015719.

Fuchs, F.D., Fuchs, S.C., 2021. The effect of alcohol on blood pressure and hypertension. Curr. Hypertens. Rep. 23 (10), 42. https://doi.org/10.1007/s11906-021-01160-7. PMID: 34762198.

Geleijnse, J.M., Witteman, J.C., den Breeijen, J.H., et al., 1996. Dietary electrolyte intake and blood pressure in older subjects: the Rotterdam Study. J. Hypertens. 14 (6), 737–741. https://doi.org/10.1097/00004872-199606000-00009.

Ginos, B.N.R., Engberink, R.H.G.O., 2020. Estimation of sodium and potassium intake: current limitations and future perspectives. Nutrients 12 (11), E3275. https://doi.org/10.3390/nu12113275.

Graudal, N.A., Galløe, A.M., Garred, P., 1998. Effects of sodium restriction on blood pressure, renin, aldosterone, catecholamines, cholesterols, and triglyceride: a meta-analysis. JAMA 279 (17), 1383–1391. https://doi.org/10.1001/jama.279.17.1383.

Graudal, N.A., Hubeck-Graudal, T., Jurgens, G., 2020. Effects of low sodium diet versus high sodium diet on blood pressure, renin, aldosterone, catecholamines, cholesterol, and triglyceride. Cochrane Database Syst. Rev. 12, CD004022. https://doi.org/10.1002/14651858.CD004022.pub5.

He, F.J., MacGregor, G.A., 2006. Importance of salt in determining blood pressure in children: meta-analysis of controlled trials. Hypertension 48 (5), 861–869. https://doi.org/10.1161/01.HYP.0000245672.27270.4a.

Hooper, L., Bartlett, C., Davey Smith, G., Ebrahim, S., 2002. Systematic review of long term effects of advice to reduce dietary salt in adults. BMJ 325 (7365), 628. https://doi.org/10.1136/bmj.325.7365.628.

Intersalt Cooperative Research Group, 1988. Intersalt: an international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion. Intersalt Cooperative Research Group. BMJ 297 (6644), 319–328. https://doi.org/10.1136/bmj.297.6644.319.

John, K.A., Cogswell, M.E., Campbell, N.R., et al., 2016. Accuracy and usefulness of select methods for assessing complete collection of 24-hour urine: a systematic review. J. Clin. Hypertens. 18 (5), 456–467. https://doi.org/10.1111/jch.12763.

Krishna, G.G., Kapoor, S.C., 1991. Potassium depletion exacerbates essential hypertension. Ann. Intern. Med. 115 (2), 77–83. https://doi.org/10.7326/0003-4819-115-2-77.

Lambers Heerspink, H.J., Navis, G., Ritz, E., 2012. Salt intake in kidney disease—a missed therapeutic opportunity? Nephrol. Dial. Transplant. 27 (9), 3435–3442. https://doi.org/10.1093/ndt/gfs354.

Law, M.R., Frost, C.D., Wald, N.J., 1991. By how much does dietary salt reduction lower blood pressure? I—analysis of observational data among populations. BMJ 302 (6780), 811–815. https://doi.org/10.1136/bmj.302.6780.811.

Ma, Y., He, F.J., Sun, Q., et al., 2022. 24-Hour urinary sodium and potassium excretion and cardiovascular risk. N. Engl. J. Med. 386 (3), 252–263. https://doi.org/10.1056/NEJMoa2109794.

MacIntyre, I.M., Turtle, E.J., Farrah, T.E., Graham, C., Dear, J.W., Webb, D.J., 2022. PATH-BP (paracetamol in hypertension–blood pressure) investigators*. Regular acetaminophen use and blood pressure in people with hypertension: the PATH-BP trial. Circulation 145 (6), 416–423. https://doi.org/10.1161/Circulationaha.121.056015. PMID: 35130054; PMCID: PMC7612370.

Marti-Soler, H., Pommier, C., Bochud, M., et al., 2017. Seasonality of sodium and potassium consumption in Switzerland. Data from three cross-sectional, population-based studies. Nutr. Metabol. Cardiovasc. Dis. 27 (9), 792–798. https://doi.org/10.1016/j.numecd.2017.06.012.

McCarron, D.A., Geerling, J.C., Kazaks, A.G., Stern, J.S., 2009. Can dietary sodium intake be modified by public policy? Clin. J. Am. Soc. Nephrol. 4 (11), 1878–1882. https://doi.org/10.2215/CJN.04660709.

McCarron, D.A., Kazaks, A.G., Geerling, J.C., Stern, J.S., Graudal, N.A., 2013. Normal range of human dietary sodium intake: a perspective based on 24-hour urinary sodium excretion worldwide. Am. J. Hypertens. 26 (10), 1218–1223. https://doi.org/10.1093/ajh/hpt139.

Mente, A., O'Donnell, M.J., Rangarajan, S., et al., 2014. Association of urinary sodium and potassium excretion with blood pressure. N. Engl. J. Med. 371 (7), 601–611. https://doi.org/10.1056/NEJMoa1311989.

Mente, A., O'Donnell, M., Rangarajan, S., et al., 2018. Urinary sodium excretion, blood pressure, cardiovascular disease, and mortality: a community-level prospective epidemiological cohort study. Lancet 392 (10146), 496–506. https://doi.org/10.1016/S0140-6736(18)31376-X.

Messerli, F.H., Hofstetter, L., Syrogiannouli, L., et al., 2021. Sodium intake, life expectancy, and all-cause mortality. Eur. Heart J. 42 (21), 2103–2112. https://doi.org/10.1093/eurheartj/ehaa947.

Midgley, J.P., Matthew, A.G., Greenwood, C.M., Logan, A.G., 1996. Effect of reduced dietary sodium on blood pressure: a meta-analysis of randomized controlled trials. JAMA 275 (20), 1590–1597. https://doi.org/10.1001/jama.1996.03530440070039.

Muntzel, M., Drüeke, T., 1992. A comprehensive review of the salt and blood pressure relationship. Am. J. Hypertens. 5 (4 Pt 1), 1S–42S. https://doi.org/10.1093/ajh/5.4s.1s.

Nagata, C., Takatsuka, N., Shimizu, N., Shimizu, H., 2004. Sodium intake and risk of death from stroke in Japanese men and women. Stroke 35 (7), 1543–1547. https://doi.org/10.1161/01.STR.0000130425.50441.b0.

Naser, A.M., He, F.J., Rahman, M., Campbell, N.R.C., 2021. Spot urine formulas to estimate 24-hour urinary sodium excretion alter the dietary sodium and blood pressure relationship. Hypertension 77 (6), 2127–2137. https://doi.org/10.1161/HYPERTENSIONAHA.120.16651.

O'Donnell, M.J., Yusuf, S., Mente, A., et al., 2011. Urinary sodium and potassium excretion and risk of cardiovascular events. JAMA 306 (20), 2229–2238. https://doi.org/10.1001/jama.2011.1729.

O'Donnell, M., Mente, A., Rangarajan, S., et al., 2014. Urinary sodium and potassium excretion, mortality, and cardiovascular events. N. Engl. J. Med. 371 (7), 612–623. https://doi.org/10.1056/NEJMoa1311889.

O'Donnell, M., Mente, A., Rangarajan, S., et al., 2019. Joint association of urinary sodium and potassium excretion with cardiovascular events and mortality: prospective cohort study. BMJ 364, l772. https://doi.org/10.1136/bmj.l772.

O'Donnell, M., Mente, A., Alderman, M.H., et al., 2020. Salt and cardiovascular disease: insufficient evidence to recommend low sodium intake. Eur. Heart J. 41 (35), 3363–3373. https://doi.org/10.1093/eurheartj/ehaa586.

Powles, J., Fahimi, S., Micha, R., et al., 2013. Global, regional and national sodium intakes in 1990 and 2010: a systematic analysis of 24 h urinary sodium excretion and dietary surveys worldwide. BMJ Open 3 (12), e003733. https://doi.org/10.1136/bmjopen-2013-003733.

Roerecke, M., Kaczorowski, J., Tobe, S.W., Gmel, G., Hasan, O.S.M., Rehm, J., 2017. The effect of a reduction in alcohol consumption on blood pressure: a systematic review and meta-analysis. Lancet Public Health 2 (2), e108–e120. https://doi.org/10.1016/S2468-2667(17)30003-8. PMID: 29253389; PMCID: PMC6118407.

Sacks, F.M., Svetkey, L.P., Vollmer, W.M., et al., 2001. Effects on blood pressure of reduced dietary sodium and the Dietary Approaches to Stop Hypertension (DASH) diet. DASH-Sodium Collaborative Research Group. N. Engl. J. Med. 344 (1), 3–10. https://doi.org/10.1056/NEJM200101043440101.

Smith, W.C., Crombie, I.K., Tavendale, R.T., Gulland, S.K., Tunstall-Pedoe, H.D., 1988. Urinary electrolyte excretion, alcohol consumption, and blood pressure in the Scottish heart health study. BMJ 297 (6644), 329–330. https://doi.org/10.1136/bmj.297.6644.329.

Staessen, J., Bulpitt, C.J., Fagard, R., Joossens, J.V., Lijnen, P., Amery, A., 1988. Salt intake and blood pressure in the general population: a controlled intervention trial in two towns. J. Hypertens. 6 (12), 965–973. https://doi.org/10.1097/00004872-198812000-00003.

Stolarz-Skrzypek, K., Kuznetsova, T., Thijs, L., et al., 2011. Fatal and nonfatal outcomes, incidence of hypertension, and blood pressure changes in relation to urinary sodium excretion. JAMA 305 (17), 1777–1785. https://doi.org/10.1001/jama.2011.574.

Strazzullo, P., D'Elia, L., Kandala, N.B., Cappuccio, F.P., 2009. Salt intake, stroke, and cardiovascular disease: meta-analysis of prospective studies. BMJ 339, b4567. https://doi.org/10.1136/bmj.b4567.

Sun, Q., Bertrand, K.A., Franke, A.A., Rosner, B., Curhan, G.C., Willett, W.C., 2017. Reproducibility of urinary biomarkers in multiple 24-h urine samples. Am. J. Clin. Nutr. 105 (1), 159–168. https://doi.org/10.3945/ajcn.116.139758.

Swales, J.D., 1991. Dietary salt and blood pressure: the role of meta-analyses. J. Hypertens. Suppl. 9 (6), S42–S46.

Thomas, M.C., Moran, J., Forsblom, C., et al., 2011. The association between dietary sodium intake, ESRD, and all-cause mortality in patients with type 1 diabetes. Diabetes Care 34 (4), 861–866. https://doi.org/10.2337/dc10-1722.

Tuomilehto, J., Jousilahti, P., Rastenyte, D., et al., 2001. Urinary sodium excretion and cardiovascular mortality in Finland: a prospective study. Lancet 357 (9259), 848–851. https://doi.org/10.1016/S0140-6736(00)04199-4.

Weinberger, M.H., Miller, J.Z., Luft, F.C., Grim, C.E., Fineberg, N.S., 1986. Definitions and characteristics of sodium sensitivity and blood pressure resistance. Hypertension 8 (6 Pt 2), II127–134. https://doi.org/10.1161/01.hyp.8.6_pt._2.ii127.

Whelton, P.K., Appel, L.J., Espeland, M.A., et al., 1998. Sodium reduction and weight loss in the treatment of hypertension in older persons: a randomized controlled trial of nonpharmacologic interventions in the elderly (TONE). TONE Collaborative Research Group. JAMA 279 (11), 839–846. https://doi.org/10.1001/jama.279.11.839.

Zeng, C., Rosenberg, L., Li, X., Djousse, L., Wei, J., Lei, G., Zhang, Y., 2022. Sodium-containing acetaminophen and cardiovascular outcomes in individuals with and without hypertension. Eur. Heart J. ehac059. https://doi.org/10.1093/eurheartj/ehac059. Epub ahead of print. PMID: 35201347.

Zhou, W., Shi, Y., Li, Y.Q., Ping, Z., Wang, C., Liu, X., Lu, J., Mao, Z.X., Zhao, J., Yin, L., Zhang, D., Tian, Z., Zhang, L., Li, L., 2018. Body mass index, abdominal fatness, and hypertension incidence: a dose-response meta-analysis of prospective studies. J. Hum. Hypertens. 32 (5), 321–333. https://doi.org/10.1038/s41371-018-0046-1. PMID: 29581553.

Sucrose: Dietary sucrose and disease

Philipp A. Gerber, Endocrinology, Diabetology and Clinical Nutrition, University Hospital Zurich (USZ), University of Zurich (UZH), Zurich, Switzerland

© 2023 Elsevier Ltd. All rights reserved.

This is an update of B. Caballero, Sucrose: Dietary Sucrose and Disease, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 231–233, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00257-9.

Introduction **596**
Sucrose as a caloric sweetener **597**
Sucrose and other added sugars 597
Sugar-sweetened beverages 597
Sucrose intake and reduced satiety 597
Conclusion 598
Metabolic effects of sucrose **598**
Metabolism of sucrose 598
Association of sucrose consumption with metabolic diseases 598
Direct effects of sucrose on glucose metabolism 599
Direct effects of sucrose on serum lipids and inflammation 600
Direct effects of sucrose on hepatic lipid metabolism 600
Conclusion 600
Interaction of sucrose consumption with the intake of other nutrients **600**
Sucrose consumption and micronutrient intake 600
Sucrose consumption and macronutrient intake 601
Conclusion 601
Sucrose metabolism by components of the microbiome **601**
Sucrose and dental health 601
Sucrose and the intestinal microbiome 602
Conclusion 602
Effects of naturally occurring sucrose **602**
Sucrose in fruits 602
Conclusion 603
Summary, conclusion and outlook **603**
Summary 603
Conclusion 603
Outlook 603
References **603**

Key points

- The worldwide ingestion of added sugar/sucrose is still very high
- Sucrose is an important source of calories and promotes weight gain
- Energy independent negative effects of sucrose on metabolism are likely to exist
- Added sugars are suspected to dilute the intake of other nutrients
- The oral and intestinal microbiome interacts with added sugar
- Sugar/sucrose consumption by the ingestion of fruits is not harmful
- A reduction of the intake of added sugar is recommended

Introduction

While the diet of our ancestors was not completely free of sugar, the amount of simple carbohydrates ingested was low, refined sugar was not available and only few natural sources provided the disaccharide sucrose as such (e.g., fruits), and even fewer sources contained some "free" sucrose as defined by the WHO (e.g., honey). In contrast, modern western diets provide a large amount of the

daily energy intake in the form of sugar, mostly sucrose. In Europe, where sucrose is by far the most important form of added sugar, recent summaries of analyses from different countries report a daily total sugar intake of 76–117 g/day, resulting in up to 20% of calories which are consumed as sugar (Azaïs-Braesco et al., 2017). The amount of added sugar is reported to figure between 42 and 76 g per day, with up to 12% of the daily energy intake provided by added sugar. 60–80% of the total amount of added sugar that is ingested is provided by sweet products and sugar-sweetened beverages.

Sucrose is a disaccharide, consisting of one molecule of glucose and one molecule of fructose that are linked via an ether bond. When ingested, sucrose is hydrolyzed by the enzyme sucrase, an intestinal α-glucosidase, to its component monosaccharides, which then are absorbed. While glucose and fructose contain the same caloric content, they enter different metabolic pathways and exert different effects regarding human health and disease.

In this article, different aspects of the impact of sucrose or its components on human disease are highlighted. First, sucrose provides a significant amount of calories, and thus, its effect as a caloric sweetener on metabolism is discussed. However, there are also calorie-independent effects that are proposed, which are highlighted in a second part. Third, it is of importance to understand how sucrose may influence the ingestion of other nutrients and is thus indirectly targeting health. In a fourth article, recent advances in the understanding of the interaction of dietary sucrose with the microbiome are highlighted. Finally, the impact of the ingestion of naturally occurring sucrose, in contrast to sucrose in the form of added sugar, is briefly discussed, with the latter being the main source of human sucrose intake nowadays.

Sucrose as a caloric sweetener

Sucrose and other added sugars

Sucrose, mostly derived from sugarcane and sugar beets, has been the main added sugar for centuries. During the past 50 years, the use of high-fructose corn syrup (HFCS) as caloric sweetener became popular as an alternative to the industrial use of sucrose as added sugar. HFCS is produced from cornstarch, which is enzymatically cleaved into glucose molecules and further processed by D-xylose isomerase in order to convert some of the glucose to fructose. The use of HFCS became popular because it is easier to handle and cheaper than sucrose.

Thus, like sucrose, high-fructose corn syrup consists of glucose and fructose, but the two monosaccharides are present in their "free" form. HFCS is mainly used as "HFCS 42" (containing 42% fructose) or "HFCS 55" (containing 55% fructose). Therefore, the ratio of fructose and glucose in sucrose and HFCS is comparable. The use of HFCS was suspected to have specific adverse effects, e.g., its consumption was linked to the development of obesity (Bray et al., 2004), and thus, the question about differences between this sweetener and sucrose became important. However, so far it was mostly concluded that there is no evidence for different effects of sucrose ingestion as compared to the ingestion of the corresponding monosaccharides glucose and fructose, as long as their ratio (50% each) is not dramatically altered (Melanson et al., 2007; Tappy and Lê, 2010).

Sugar-sweetened beverages

The consumption of sugar in the form of sugar-sweetened beverages (SSBs), providing a high amount of calories in liquid form, is suspected to have particular deleterious effects on health. Although the qualitative and quantitative sugar composition of SSBs varies (e.g., use of sucrose or HFCS, which may differ by region), the association of their consumption with weight gain and obesity was almost univocally reported by many observational studies from different places, independently of their country of origin and the time they were conducted. Such studies include cross-sectional studies that show an association of SSB consumption with higher body weight/BMI, but also prospective cohort studies demonstrating larger weight gain and obesity in adults as well as in children being associated with a higher intake of SSBs (Malik et al., 2006).

Furthermore, interventional trials were conducted in individuals of different age in order to explore a possible direct causal relationship of the intake of sugar-sweetened beverages with weight gain. In such a trial, overweight adult subjects were provided either sucrose sweetened beverages (80% of supplement) and food (20% of supplement) for 10 weeks, or similar amounts containing artificial sweeteners. As a result, body weight increased in the sucrose group and decreased in the sweetener group during the intervention (Sørensen et al., 2014). Another intervention was conducted in children for a total duration of 18 months. More than 600 children were randomly assigned to receive 250 mL of a beverage which was either artificially sweetened (with sucralose and acesulfame; sugar-free group) or a similar beverage containing 26 g of sucrose (104 kcal per day) (de Ruyter et al., 2012). As a result, in the sugar group statistically significant more weight gain as well as an increase in skinfold-thickness, waist-to-height ratio and fat mass was observed.

Sucrose intake and reduced satiety

The above-mentioned observed effects of caloric sweeteners as sucrose on weight gain/obesity led to the hypothesis that they are not able to induce effects of satiety that are sufficient to maintain total caloric intake at a constant level. This hypothesis was consecutively tested in interventional trials comparing the effects of different sugars (with the same content of calories) on satiety. In a recent

study with healthy young adults, different hormones that affect appetite and energy homeostasis were measured after ingestion of 75 g of either sucrose or glucose (Yunker et al., 2021), including the hormones glucagon-like peptide (GLP-1), peptide YY (PYY), and ghrelin. In addition to lower plasma insulin levels, also responses in GLP-1 and PYY (both providing satiety feelings) were significantly lower after sucrose ingestion as compared to glucose, providing evidence for a reduced effect of sucrose on satiety as compared to other sugars.

The reason for this effect of a reduced satiety induction was suggested to be based on the composition of sucrose, containing 50% fructose and 50% glucose. Many trials have shown that fructose has a satiating effect which is clearly inferior as compared to glucose. When MRI was performed in healthy adults who ingested either glucose or fructose, different changes in cerebral blood flow were observed in brain regions involved with appetite and reward pathways (Page et al., 2013). Similarly, in another interventional trial, it was shown that ingestion of fructose compared with glucose resulted in greater brain responses to food cues in the visual cortex and left orbital frontal cortex, and led to greater hunger and desire for food—and even greater willingness to give up long-term monetary rewards to obtain immediate high-calorie foods (Luo et al., 2015).

Thus, the reduced induction of satiety after sucrose consumption—as compared to pure glucose—is very likely to be a result of the composition of sucrose, containing 50% fructose, which may be the driver for a higher promotion of feeding behavior.

Conclusion

Sucrose, together with high-fructose corn syrup, is one of the most important added sugars and a source of a high amount of calories, contributing to weight gain and obesity. A reduced induction of satiety feelings may be one of the factors contributing to this effect.

Metabolic effects of sucrose

Metabolism of sucrose

As stated above, sucrose is hydrolyzed in the intestine into equal amounts of glucose and fructose, which then are absorbed by the intestinal epithelial cells separately. Glucose, on one hand, is the preferred substrate and energy source of most eukaryotic cells. Apart from its direct use for energy production, it can be stored in the form of glycogen, primarily in the cells of the liver and the skeletal muscle, or as lipids (mainly in the adipose tissue) after entering the pathways of lipogenesis (Wasserman, 2009). The uptake into these organs is mainly mediated by insulin. On the other hand, fructose cannot be directly metabolized in many cells. Thus, it is first metabolized in the liver by distinct pathways. Apart from the possibility of being converted to glycogen, it enhances fatty acid synthesis as a lipogenic substrate as well as being a potent inducer of lipogenic enzyme expression (Geidl-Flueck and Gerber, 2017).

Thus, one of the most important questions addressed by many studies during the past years was whether the ingestion of glucose or fructose (or its co-ingestion as sucrose or HFCS) is—due to these distinct metabolic pathways—resulting in specific, different effects on health and disease.

Association of sucrose consumption with metabolic diseases

With all the evidence that an increase in sucrose intake leads to weight gain, increased fat mass and obesity, it was concluded that these unfavorable effects were inevitably linked to consecutive metabolic disturbances resulting from obesity, as non-alcoholic fatty liver disease, dyslipidemia, or type 2 diabetes—regardless of the possibility of additional direct metabolic effects that are confined to sucrose due to its specific properties.

Like for obesity itself, the association of type 2 diabetes with sugar consumption was explored by many observational studies which mainly tested the hypothesis of this relationship by assessing the consumption of sugar-sweetened beverages and the concomitant diabetes incidence in selected populations. In a meta-analysis of such studies, including more than 300,000 Individuals and more than 15,000 cases of type 2 diabetes, a clear increase of the risk to develop type 2 diabetes was identified in people consuming a high amount of sugar-sweetened beverages (Malik et al., 2010). In contrast, a later meta-analysis which analyzed data of a similar number of individuals, but assessed specific differences of different sugar types, was not able to detect an association of a higher incidence of type 2 diabetes with total sugar, fructose, or sucrose consumption (Tsilas et al., 2017). However, this analysis concluded that confidence in the estimates was limited by evidence of serious inconsistency between studies, and serious imprecision in the pooled estimates for the different sugar categories.

Similarly, the assessment of effects of sugar intake on the development of non-alcoholic liver disease, another severe metabolic disturbance associated with a Western Diet, provided some insights, but there are also remaining questions. Current evidence of observational studies was summarized as supporting the hypothesis that sugar intake, but in particular fructose intake, increases hepatic lipid accumulation (Jensen et al., 2018). However, other authors concluded that the apparent association between indexes of liver health and fructose or sucrose intake appears to be confounded by excessive energy intake, thus rendering available evidence insufficiently robust to draw conclusions regarding effects of fructose or sucrose consumption on non-alcoholic fatty liver disease (Chung et al., 2014).

Data regarding an association of further cardiovascular risk factors with sugar consumption are also derived mostly from studies that search for associations of these risk factors with sugar sweetened beverage consumption. A systematic review including different prospective cohort studies reported that all included studies examining vascular risk factors found direct associations between SSB consumption and change in blood pressure, blood lipids or blood sugar (Keller et al., 2015). Such prospective cohort studies are supported by large cross-sectional studies as the Health Professionals Follow-up Study (HPFS) (de Koning et al., 2012) and Nurses' Health Study (NHS) cohorts (Yu et al., 2018), where an association between SSB consumption and higher plasma triglycerides, but also inflammatory cytokines and other cardiometabolic risk factors was detected. Regarding hypertension, a systematic review and meta-analysis found, despite heterogeneity of the studies that were examined, a risk of developing hypertension associated with SSB consumption (Jayalath et al., 2015).

Finally, direct assessment of cardiovascular disease was also performed by many studies. A recent report of the California Teachers Study presented the prospective association of baseline SSB consumption with incident cardiovascular disease in more than 100,000 women. After adjusting for potential confounders, an increased risk for the development of cardiovascular disease, revascularization and stroke was found to be associated with the consumption of SSB (Pacheco et al., 2020). Finally, one of the latest reports of the above-mentioned HPFS and NHS prospective cohort studies examined total and cause-specific mortality in almost 120,000 men and women. Here, after adjusting for major diet and lifestyle factors, consumption of SSBs was associated with a higher risk for total mortality. The association was observed for cardiovascular and, to a somewhat lesser, but statistically significant extent, for cancer related mortality (Malik et al., 2019).

Regarding cardiovascular disease, it is debated whether the consumption of specific sugars promotes inflammation independently of overfeeding (where an increase in inflammatory markers is probably associated with an expansion of adipose tissue). Indeed, there are observational studies that confirmed an elevation of CRP in heavy SSB drinkers even after adjusting for BMI (Lin et al., 2020).

Taken together, most observational data on the association of sugar consumption and cardio-metabolic risk factors (up to total mortality) is derived from studies on the consumption of sugar sweetened beverages. As described, most of these observational studies adjusted this risk for confounding factors as weight gain, suggesting a calorie-independent effect. However, the question remained whether such diseases are not only the result of the increase in (caloric intake driven) adipose tissue expansion, and not of a specific calorie-independent effect of the sugars that were investigated. To differentiate such effects of sucrose on metabolism (as compared to other nutrients), interventional trials using iso-caloric diets were needed. Such studies were conducted during the past decades and discussed below.

Direct effects of sucrose on glucose metabolism

Early studies conducted around 1980 used diets with large differences in sucrose content to explore the effects of sucrose consumption on metabolism in healthy subjects. In one of the first trials, a very high proportion (30% of total energy intake) of an isocaloric diet was either provided as sucrose or as wheat starch in order to assess the effect of these two different carbohydrates on glucose metabolism (Reiser et al., 1979). Indeed, fasting insulin and glucose levels were significantly elevated after 6 weeks of sucrose consumption compared to the consumption of starch, and the insulin response as well as the insulin to glucose ratio after a sucrose load were increased after sucrose consumption. As a confirmation of the isocaloric composition of these diets, there was no change in weight gain observed. Thus, a disturbance of glucose metabolism, mediated probably by a reduction in insulin sensitivity, was proposed, and confirmed in a second study applying isocaloric diets with different sucrose content (2–30% of total energy intake) in individuals who were already hyperinsulinemic at baseline. Again, an increase in fasting glucose as well as in glucose and insulin levels after a sucrose load was observed after a 6 week diet with high sucrose content compared to starch consumption (Reiser et al., 1981).

However, these early observations were challenged by later studies. In such a trial, somewhat smaller differences in sucrose consumption, which probably reflect better the usual span of sucrose intake in daily live, were investigated during a similar period of 6 weeks (Black et al., 2006). With an eucaloric diet over 6 weeks designed to result in weight maintenance but containing 10% or 25% of total daily energy intake as sucrose, no differences in glucose metabolism were observed in healthy subjects, despite a thorough investigation using euglycemic clamp technique as well as a continuous glucose monitoring system. Furthermore, a study conducted in mostly obese, already moderately insulin resistant subjects, could not detect any beneficial effect on insulin resistance with a lower sucrose intake of 5% (vs. 15%) of daily energy intake, despite a slightly lower fasting plasma glucose (Lewis et al., 2013). A study that compared different sugars regarding their effect on insulin sensitivity did see an effect of pure fructose feeding, but not of sucrose feeding (Aeberli et al., 2013). Thus, it might be hypothesized that fructose is the main driver of insulin resistance, and that the fructose content of sucrose is not sufficient to induce such insulin resistance—unless sucrose is provided in very high quantities.

In summary, these trials provided no clear evidence of biologically significant effects of sucrose consumption on glucose metabolism, at least when consumed in an eucaloric diet for some weeks, unless very high amounts of sucrose are ingested. Nevertheless, the question remained whether sucrose ingestion induces subtle metabolic changes that may have long-term negative effects on glucose metabolism that are not detectable after short-term ingestion.

Direct effects of sucrose on serum lipids and inflammation

Similar as for glucose metabolism, first interventional studies examining lipid metabolism were already conducted around 1980 and provided evidence for unfavorable effects of sucrose on serum lipids. Here again, the consumption of sucrose was compared to the consumption of wheat starch (30% of daily energy intake). While serum fatty acids were not changed, there was an increase in triglycerides (by 33%) as well as total cholesterol (Reiser et al., 1979). Again, the feeding with such very large dosages was adapted in later studies, when more moderate amounts of calories were replaced by sucrose. In a recent trial conducted in individuals with a different genetic background regarding the presence of the apolipoprotein E2 allele (which is involved in the metabolism of triacylglycerol-rich lipoproteins), 40 g/day of sucrose was used in an isocaloric adaptation of the usual diet for 8 weeks. No changes in any of the conventional serum lipid parameters (total, HDL- and LDL-cholesterol as well as triglycerides) were observed (Erkkilä et al., 2007). Another interventional cross-over study came to the same conclusion, when sugar-sweetened beverages containing glucose, fructose or sucrose were used (80 g per day of each) (Aeberli et al., 2011). However, in this study, additional and more subtle changes in LDL-cholesterol particles were investigated. Of interest, it could be shown that the intervention with sucrose and fructose, but not with glucose, promoted the generation of small, dense LDL-particles which are known to be associated with a high cardiovascular risk (Gerber and Berneis, 2012), again pointing to a specific unfavorable effect of fructose (as part of sucrose) on lipid metabolism.

Regarding a general inflammatory response as an additional cardiovascular risk factor, the effect on CRP (as one of the most important indicators of inflammation) of different sugars was summarized in a systematic review and meta-analysis of intervention studies. Here, no difference between the effects of different sugars (e.g., fructose, sucrose, glucose and HFCS) was detected (Della Corte et al., 2018). One possible reason for this might be that some intervention studies were conducted over a very short time (e.g., 1 week (Kuzma et al., 2016)), which was probably insufficient to reveal significant differences in an outcome as chronic inflammation.

Direct effects of sucrose on hepatic lipid metabolism

Due to the high prevalence of non-alcoholic fatty liver disease and the suspected association of sugar intake with this disease that was reported at least in some observational studies, interventional studies were conducted to investigate a possible causality of this association. A systematic review and meta-analysis identified more than 20 studies that reported data on the effects of the consumption of different sugars on prespecified indexes of liver health as liver fat or liver enzymes (Chung et al., 2014). One result of the study was, here again, that it is very difficult to differentiate effects of overfeeding from effects that are directly associated with the quality of the sugar ingested.

However, one trial assessed the effect of different test drinks containing either sucrose or semi-skim milk in isocaloric quantities, with an aspartame-sweetened drink and water. Of interest, the study was conducted for an extended period of 6 months. In the sucrose-sweetened beverage group, the fat content not only of the liver, but also of skeletal muscle and visceral fat was significantly increased (Maersk et al., 2012). An even closer look at sugar quality was provided by a recent study investigating the effects of isocaloric amounts of glucose, fructose and sucrose (80 g/day for each, ingested as sugar-sweetened beverages for 7 weeks) on hepatic de-novo lipogenesis using tracer-based methods (Geidl-Flueck et al., 2021). An increase in hepatic de-novo lipogenesis was observed after the consumption of fructose and sucrose, but not glucose.

Conclusion

In conclusion, observational as well as interventional studies provided varying results regarding the effect of added sucrose consumption on different metabolic parameters. In interventional studies, adverse effects on metabolic health were seen mostly when either large dosages of sugars were used, or when subtle changes in metabolism were examined. Thus, it can be hypothesized that adverse effects of sugar consumption probably exist, but that they need a certain time until they become clinically relevant.

Interaction of sucrose consumption with the intake of other nutrients

Considering the composition of food rich in added sugars as sucrose, there was increasing concern that a high intake of such food might compromise the intake of other macro- and micronutrients. Thus, many cross-sectional or prospective studies evaluated the association between the intake of sugar and diet quality as well as nutrient intake.

As for other questions regarding sucrose intake, most data available is derived from sugar-sweetened beverage intake, with sucrose being one of the most widely used sugar.

Sucrose consumption and micronutrient intake

Most systematic reviews that were conducted around 10 years ago concluded that the evidence available at the time did not allow for firm conclusions regarding the hypothesis of a general micronutrient dilution mediated by the ingestion of nutrients rich in added sugar (Gibson, 2007; Livingstone and Rennie, 2009; Rennie and Livingstone, 2007). Data from a recent European report on the

association of sugar-sweetened beverage consumption and micronutrient intake derived from 3-day food records suggested a decrease of calcium, iron, and magnesium intake with increasing consumption of SSBs (Mullie et al., 2018). Another study presenting data from the Australian Health Survey reported a peak intake of 16 out of 19 micronutrients when 5–15% of energy was consumed as free sugar. For higher sugar intake, a decreasing trend was seen, including the intake of fiber, folate and magnesium (Mok et al., 2018). One of the latest studies, a cross-sectional analysis in two Swedish adult populations, reported again a significant inverse association between the intake of added sugar and the intake of nine micronutrients (calcium, folate, iron, magnesium, potassium, selenium, vitamin C, vitamin D and zinc) in both populations (González-Padilla et al., 2020).

Despite these data, caution regarding the hypothesis of micronutrient dilution by sugar intake in general is advised since the effect of different sugar containing nutrients on micronutrient ingestion might be much more complex, given the large differences in individual diet patterns.

Sucrose consumption and macronutrient intake

In addition to adverse effects on micronutrient ingestion, it is also suggested that the ingestion of a high amount of calories in the form of sugar my adversely affect the intake of other important macronutrients as protein.

Indeed, such a reduction was observed in the past—cross-sectional studies reported a reduction of all macronutrients other than carbohydrates (e.g., fat, protein, fiber) being associated with added sugar intake (González-Padilla et al., 2020). Of importance, the reduction of protein intake was also found in a study performed in older people, where a sufficient supply with proteins is highly important (Jyväkorpi et al., 2017). Furthermore, the reduction of protein intake after the introduction of sucrose-sweetened beverages was also confirmed in interventional trials (Geidl-Flueck et al., 2021).

Conclusion

A reduced intake of important micro- and macronutrients due to an increased consumption of nutrients with a high content of added sugar is a serious concern, however, such risk is likely to be significantly influenced by individual diet patterns.

Sucrose metabolism by components of the microbiome

Sucrose and dental health

Sucrose has been known to affect dental health for a long time, and experimental data dates back to the time before 1950 when an association between sucrose consumption and dental caries was first demonstrated (Shafer, 1949). Dental caries is still one of the most prevalent noncommunicable chronic diseases, with estimates of a global prevalence of some 2.5 billion people in 2015 (Kassebaum et al., 2015), and with an enormous medical and economic impact on societies (Righolt et al., 2018). The prevalence of untreated caries in the USA was recently reported to be more than 20% in the adult population (Bashir, 2021).

The pathophysiology of caries is based on the fermentation of dietary sugars by bacteria of the oral flora as *Streptococcus mutans* or Lactobacillus species which are normally present in dental plaque. With the intake of high amounts of sugar, their presence and activity increases, producing a high amount of acid when metabolizing dietary carbohydrates, and in particular sugar. The landmark Vipeholm studies, conducted almost 70 years ago, underscored the importance of both the frequency of sugar consumption, as well as its consistency (Gustafsson, 1954; Gustafsson et al., 1954; Hojer and Maunsbach, 1954; Krasse, 2001). These data could be reproduced over the past decades (Touger-Decker and Van Loveren, 2003).

Early childhood caries is a particular worrisome consequence of high sugar intake in children. Multiple studies have shown a strong correlation of the intake of added sugar and the development of caries early in life. Here again, sugar-sweetened beverages are of particular importance in the development of dental caries. Despite the efforts to reduce childhood caries during the past decades, recent data from two German birth cohorts still showed a significantly increased caries burden in children/adolescents 10 and 15 years old being associated with the consumption of SSB (Pitchika et al., 2020). In addition, the results of a prospective study among Scottish young children provided evidence that the introduction of SSBs during the 1st year of life can put children in a trajectory of high levels of dental caries (Bernabé et al., 2020). Of importance, four categories of social determinants were identified to have an impact on added sugar intake in children. These are socioeconomic disadvantage, household dietary habits, the location of added sugar sources and peer influence (Chi and Scott, 2019). Thus, any efforts to reduce added sugar intake in children should address these important factors.

Despite a focus on the burden of caries in early life, it is important to acknowledge the fact that the association of sugar (as from SSBs) with dental disease is still present in adulthood. A recent report of cross-sectional data from the US National Health and Nutrition Examination Survey (NHANES) reported an adjusted prevalence for untreated decay of 30% to be associated with the consumption of 47 g of added sugar from SSBs compared to those reporting no sugar from SSBs, and the number of untreated decayed teeth increased with sugar intake from SSBs (Moss et al., 2021). A clear dose-response gradient regarding the association of SSB consumption and dental disease was also observed by a very recent systematic review and meta-analysis, underlining the suggestions that the association is likely to be causal (Valenzuela et al., 2021).

Finally, although there was much attention on the association of SSBs and caries during the past years, it is important to mention that added sugar also in solid form is associated with dental disease, as demonstrated by a recent analysis of data from the UK

Children's Dental Health Survey (CDHS), were the ingestion of drinks and foods with added sugar were analyzed separately (Hong et al., 2018). It was shown that the frequent consumption of foods and drinks with added sugars was associated with the presence of dental caries among 12- and 15-year-olds.

Sucrose and the intestinal microbiome

The intestinal microbiome consists of a very diverse consortium of different bacteria, archaea, fungi, protozoa, and viruses that inhabit the gut of all mammals. Many studies in humans, but also in other species have demonstrated that the microbiome plays a pivotal role in a broad variety of fundamental biological processes impacting human health and disease locally (e.g., gut epithelial health), but even more important systemically (e.g., control of energy homeostasis, metabolic processes, immunologic activity or neurobehavioral development) (Barko et al., 2018). Thus, the gut microbiome is an active participant in host physiology, and any disturbances regarding its composition may affect the whole individual.

When the interaction of sugar consumption with the microbiome is evaluated, we must be reminded that the disaccharide sucrose is enzymatically hydrolyzed into glucose and fructose, which then are metabolized in significant amounts by the microbiome. The effect of fructose on the microbiome is of particular interest and seems to exhibit significant differences as compared to the effect of glucose. Thus, the intake of sucrose, providing both glucose and fructose, is suggested to affect the microbiome quite differently as compared to the intake of food providing mainly glucose molecules (e.g., starch) or fructose alone (Rosas-Villegas et al., 2017).

So far, our understanding about possible interactions of different sugars with the microbiome is still incomplete and mainly derived from experimental data in rodents. One important question, which was addressed in many studies during the past years, is whether the impact of sucrose and fructose on hepatic lipid metabolism is, at least in part, mediated by interactions with the microbiome. In one of the first studies assessing this question, fructose (and sucrose), but not glucose, were shown to induce hepatic triglyceride accumulation in rodents (Bergheim et al., 2008). This effect was accompanied by a concomitant increase in portal vein endotoxins. Of interest, liver lipid accumulation could be reversed by the administration of an antibiotic therapy together with fructose consumption. This led to the hypothesis that a fructose-induced increased translocation of bacterial toxins might be the reason for this observation. While the importance of disruption of the gut vascular barrier as a prerequisite for non-alcoholic steatohepatitis was further emphasized (Mouries et al., 2019), the role of different sugars on this remained unclear, since further studies on this topic provided conflicting results. Another study showed that glucose, but not fructose, alters the intestinal paracellular permeability in association with gut inflammation and dysbiosis in mice (Zhang et al., 2021). Furthermore, and in contrast to these rodent studies, a clinical cross-over pilot study in obese humans who were provided 75 g of fructose or glucose (substituted isocalorically for complex carbohydrates) for 14 days did not show any difference between the glucose and fructose intervention regarding changes in the gut microbiome, metabolome, and permeability as well as endotoxemia (Alemán et al., 2021). However, the question remains whether obese people exhibit an already altered gut microbiome, which is not significantly changed by further interventions, and further studies are needed to clarify these questions. As another hypothesis, latest experimental evidence suggests that the interaction of the microbiome with ingested fructose regarding hepatic lipid accumulation relies probably on the conversion of dietary fructose to acetate by the gut microbiota, which then supplies lipogenic acetyl-CoA (Zhao et al., 2020). Depletion of the microbiota in this study potently suppressed this effect.

Conclusion

The causal effect of sucrose and other sugars on dental disease, mediated by oral microbiota, is well established. In contrast, the role of the intestinal microbiome in mediating adverse effects of sugars regarding the development of metabolic (as fatty liver) disease is still incompletely understood and targeted by many current studies.

Effects of naturally occurring sucrose

Sucrose in fruits

With the increasing evidence for harmful effects of fructose (and thus, sucrose, which provides fructose as well), the question about possible adverse effects of the consumption of high amounts of fruits, containing fructose and sucrose in different amounts, arose. However, corresponding evaluations that were performed in large meta-analyses of prospective cohort studies consistently provided evidence for a protective role of fruit consumption (at least up to 5 servings per day) regarding cardiovascular disease, cardiovascular mortality, and all-cause mortality (Wang et al., 2014; Zhan et al., 2017). Correspondingly, interventional studies could not show any benefit of a reduction of fructose/sucrose consumption by reducing the intake of fruits regarding cardiovascular risk factors (Aeberli et al., 2011).

It is hypothesized that a possible harm of the moderate amount of fructose and sucrose that is ingested when consuming fruits is clearly outweighed by the beneficial effects of fruits, providing many protective nutrients as fibers or vitamins.

Conclusion

So far, there is no evidence of negative effects of sugar consumption by the ingestion of fruits in moderate amounts.

Summary, conclusion and outlook

Summary

In most parts of the world, sugar consumption is very high, in particular when compared to the dietary sugar intake that was present in prehistorical diets. Most of the added sugar consumed nowadays consists of sucrose and thus, of a mixture (50/50) of the two monosaccharides glucose and fructose. High-fructose corn syrup is often used as added sugar as well, with a comparable glucose and fructose content, and most of the existing evidence suggests similar effects of sucrose and high-fructose corn syrup regarding health and disease.

Five important topics regarding sucrose consumption and human health must be considered.

(1) Sucrose is a source of energy. A reduced ability of sucrose to induce satiety after consumption is suggested, in particular because of its fructose component. Thus, overfeeding induced by sucrose consumption is likely to occur.
(2) Sucrose is very likely to have energy independent negative effects on human metabolism. Again, these effects are probably mainly mediated by the fructose component, but may be facilitated by glucose. They include the induction of lipogenic activity in the liver, a metabolic switch which may pave the way for further changes negatively affecting metabolic health.
(3) The consumption of food rich in sucrose may result in a reduced intake of other micro- and macronutrients which are important for maintaining health, but this effect depends on the individual diet pattern.
(4) Fermentation of sucrose and/or its components by the oral and intestinal bacterial flora results in either fermentation products or changes in the microbiome which may have adverse effects on health.
(5) The consumption of sugar (fructose, sucrose) containing fruits does not seem to be associated with adverse effects on health as compared to food containing added sugar.

Conclusion

Current evidence derived either from larger cross-sectional or prospective cohort studies, or from smaller interventional trials suggests that the consumption of sucrose in the form of added sugar contributes to the development of different diseases. The call to reduce the intake of free sugars to at least less than 10%, better less than 5%, as suggested by the WHO, should be emphasized.

Outlook

There is still a need to further clarify the role of sucrose regarding clinical endpoints. While conducting interventional studies that introduce additional sucrose to the diet are precluded by ethical considerations, interventional trials that reduce the intake of added sugar in a randomized manner would help to further clarify the impact of sucrose on human health and disease.

References

Aeberli, I., Gerber, P.A., Hochuli, M., Kohler, S., Haile, S.R., Gouni-Berthold, I., Berthold, H.K., Spinas, G.A., Berneis, K., 2011. Low to moderate sugar-sweetened beverage consumption impairs glucose and lipid metabolism and promotes inflammation in healthy young men: a randomized controlled trial. Am. J. Clin. Nutr. 94, 479–485.

Aeberli, I., Hochuli, M., Gerber, P.A., Sze, L., Murer, S.B., Tappy, L., Spinas, G.A., Berneis, K., 2013. Moderate amounts of fructose consumption impair insulin sensitivity in healthy young men: a randomized controlled trial. Diabetes Care 36, 150–156.

Alemán, J.O., Henderson, W.A., Walker, J.M., Ronning, A., Jones, D.R., Walter, P.J., Daniel, S.G., Bittinger, K., Vaughan, R., Macarthur, R., Chen, K., Breslow, J.L., Holt, P.R., 2021. Excess dietary fructose does not alter gut microbiota or permeability in humans: a pilot randomized controlled study. J. Clin. Transl. Sci. 5, e143.

Azaïs-Braesco, V., Sluik, D., Maillot, M., Kok, F., Moreno, L.A., 2017. A review of total & added sugar intakes and dietary sources in Europe. Nutr. J. 16, 6.

Barko, P.C., Mcmichael, M.A., Swanson, K.S., Williams, D.A., 2018. The gastrointestinal microbiome: a review. J. Vet. Intern. Med. 32, 9–25.

Bashir, N.Z., 2021. Update on the prevalence of untreated caries in the US adult population, 2017–2020. J. Am. Dent. Assoc. 153 (4).

Bergheim, I., Weber, S., Vos, M., Krämer, S., Volynets, V., Kaserouni, S., Mcclain, C.J., Bischoff, S.C., 2008. Antibiotics protect against fructose-induced hepatic lipid accumulation in mice: role of endotoxin. J. Hepatol. 48, 983–992.

Bernabé, E., Ballantyne, H., Longbottom, C., Pitts, N.B., 2020. Early introduction of sugar-sweetened beverages and caries trajectories from age 12 to 48 months. J. Dent. Res. 99, 898–906.

Black, R.N., Spence, M., Mcmahon, R.O., Cuskelly, G.J., Ennis, C.N., Mccance, D.R., Young, I.S., Bell, P.M., Hunter, S.J., 2006. Effect of eucaloric high- and low-sucrose diets with identical macronutrient profile on insulin resistance and vascular risk: a randomized controlled trial. Diabetes 55, 3566–3572.

Bray, G.A., Nielsen, S.J., Popkin, B.M., 2004. Consumption of high-fructose corn syrup in beverages may play a role in the epidemic of obesity. Am. J. Clin. Nutr. 79, 537–543.

Chi, D.L., Scott, J.M., 2019. Added sugar and dental caries in children: a scientific update and future steps. Dent. Clin. North Am. 63, 17–33.

Chung, M., Ma, J., Patel, K., Berger, S., Lau, J., Lichtenstein, A.H., 2014. Fructose, high-fructose corn syrup, sucrose, and nonalcoholic fatty liver disease or indexes of liver health: a systematic review and meta-analysis. Am. J. Clin. Nutr. 100, 833–849.

de Koning, L., Malik, V.S., Kellogg, M.D., Rimm, E.B., Willett, W.C., Hu, F.B., 2012. Sweetened beverage consumption, incident coronary heart disease, and biomarkers of risk in men. Circulation 125, 1735–1741. S1.

de Ruyter, J.C., Olthof, M.R., Seidell, J.C., Katan, M.B., 2012. A trial of sugar-free or sugar-sweetened beverages and body weight in children. N. Engl. J. Med. 367, 1397–1406.
Della Corte, K.W., Perrar, I., Penczynski, K.J., Schwingshackl, L., Herder, C., Buyken, A.E., 2018. Effect of dietary sugar intake on biomarkers of subclinical inflammation: a systematic review and meta-analysis of intervention studies. Nutrients 10.
Erkkilä, A.T., Schwab, U.S., Agren, J.J., Hallikainen, M., Gylling, H., Uusitupa, M.I., 2007. Moderate increase in dietary sucrose does not influence fasting or postprandial serum lipids regardless of the presence of apolipoprotein E2 allele in healthy subjects. Eur. J. Clin. Nutr. 61, 1094–1101.
Geidl-Flueck, B., Gerber, P.A., 2017. Insights into the hexose liver metabolism-glucose versus fructose. Nutrients 9.
Geidl-Flueck, B., Hochuli, M., Németh, Á., Eberl, A., Derron, N., Köfeler, H.C., Tappy, L., Berneis, K., Spinas, G.A., Gerber, P.A., 2021. Fructose- and sucrose- but not glucose-sweetened beverages promote hepatic de novo lipogenesis: a randomized controlled trial. J. Hepatol. 75, 46–54.
Gerber, P.A., Berneis, K., 2012. Regulation of low-density lipoprotein subfractions by carbohydrates. Curr. Opin. Clin. Nutr. Metab. Care 15, 381–385.
Gibson, S.A., 2007. Dietary sugars intake and micronutrient adequacy: a systematic review of the evidence. Nutr. Res. Rev. 20, 121–131.
González-Padilla, E., Dias, J.A., Ramne, S., Olsson, K., Nälsén, C., Sonestedt, E., 2020. Association between added sugar intake and micronutrient dilution: a cross-sectional study in two adult Swedish populations. Nutr. Metab. 17, 15.
Gustafsson, B.E., 1954. The Vipeholm dental caries study: survey of the literature on carbohydrates and dental caries. Acta Odontol. Scand. 11, 207–231.
Gustafsson, B.E., Quensel, C.E., Lanke, L.S., Lundqvist, C., Grahnen, H., Bonow, B.E., Krasse, B., 1954. The Vipeholm dental caries study; the effect of different levels of carbohydrate intake on caries activity in 436 individuals observed for five years. Acta Odontol. Scand. 11, 232–264.
Hojer, J.A., Maunsbach, A.B., 1954. The Vipeholm dental caries study: purposes and organisation. Acta Odontol. Scand. 11, 195–206.
Hong, J., Whelton, H., Douglas, G., Kang, J., 2018. Consumption frequency of added sugars and UK children's dental caries. Community Dent. Oral Epidemiol. 46, 457–464.
Jayalath, V.H., de Souza, R.J., Ha, V., Mirrahimi, A., Blanco-Mejia, S., Di Buono, M., Jenkins, A.L., Leiter, L.A., Wolever, T.M., Beyene, J., Kendall, C.W., Jenkins, D.J., Sievenpiper, J.L., 2015. Sugar-sweetened beverage consumption and incident hypertension: a systematic review and meta-analysis of prospective cohorts. Am. J. Clin. Nutr. 102, 914–921.
Jensen, T., Abdelmalek, M.F., Sullivan, S., Nadeau, K.J., Green, M., Roncal, C., Nakagawa, T., Kuwabara, M., Sato, Y., Kang, D.H., Tolan, D.R., Sanchez-Lozada, L.G., Rosen, H.R., Lanaspa, M.A., Diehl, A.M., Johnson, R.J., 2018. Fructose and sugar: a major mediator of non-alcoholic fatty liver disease. J. Hepatol. 68, 1063–1075.
Jyväkorpi, S.K., Pitkälä, K.H., Puranen, T.M., Björkman, M.P., Kautiainen, H., Strandberg, T.E., Soini, H., Suominen, M.H., 2017. High intake of nonmilk extrinsic sugars is associated with protein and micronutrient dilution in home-dwelling and institutionalized older people. J. Am. Med. Dir. Assoc. 18, 301–305.
Kassebaum, N.J., Bernabé, E., Dahiya, M., Bhandari, B., Murray, C.J., Marcenes, W., 2015. Global burden of untreated caries: a systematic review and metaregression. J. Dent. Res. 94, 650–658.
Keller, A., Heitmann, B.L., Olsen, N., 2015. Sugar-sweetened beverages, vascular risk factors and events: a systematic literature review. Publ. Health Nutr. 18, 1145–1154.
Krasse, B., 2001. The Vipeholm dental caries study: recollections and reflections 50 years later. J. Dent. Res. 80, 1785–1788.
Kuzma, J.N., Cromer, G., Hagman, D.K., Breymeyer, K.L., Roth, C.L., Foster-Schubert, K.E., Holte, S.E., Weigle, D.S., Kratz, M., 2016. No differential effect of beverages sweetened with fructose, high-fructose corn syrup, or glucose on systemic or adipose tissue inflammation in normal-weight to obese adults: a randomized controlled trial. Am. J. Clin. Nutr. 104, 306–314.
Lewis, A.S., Mccourt, H.J., Ennis, C.N., Bell, P.M., Courtney, C.H., Mckinley, M.C., Young, I.S., Hunter, S.J., 2013. Comparison of 5% versus 15% sucrose intakes as part of a eucaloric diet in overweight and obese subjects: effects on insulin sensitivity, glucose metabolism, vascular compliance, body composition and lipid profile. A randomised controlled trial. Metabolism 62, 694–702.
Lin, W.T., Kao, Y.H., Sothern, M.S., Seal, D.W., Lee, C.H., Lin, H.Y., Chen, T., Tseng, T.S., 2020. The association between sugar-sweetened beverages intake, body mass index, and inflammation in US adults. Int. J. Publ. Health 65, 45–53.
Livingstone, M.B., Rennie, K.L., 2009. Added sugars and micronutrient dilution. Obes. Rev. 10 (Suppl. 1), 34–40.
Luo, S., Monterosso, J.R., Sarpelleh, K., Page, K.A., 2015. Differential effects of fructose versus glucose on brain and appetitive responses to food cues and decisions for food rewards. Proc. Natl. Acad. Sci. U. S. A. 112, 6509–6514.
Maersk, M., Belza, A., Stødkilde-Jørgensen, H., Ringgaard, S., Chabanova, E., Thomsen, H., Pedersen, S.B., Astrup, A., Richelsen, B., 2012. Sucrose-sweetened beverages increase fat storage in the liver, muscle, and visceral fat depot: a 6-mo randomized intervention study. Am. J. Clin. Nutr. 95, 283–289.
Malik, V.S., Schulze, M.B., Hu, F.B., 2006. Intake of sugar-sweetened beverages and weight gain: a systematic review. Am. J. Clin. Nutr. 84, 274–288.
Malik, V.S., Popkin, B.M., Bray, G.A., Després, J.P., Willett, W.C., Hu, F.B., 2010. Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes: a meta-analysis. Diabetes Care 33, 2477–2483.
Malik, V.S., Li, Y., Pan, A., De Koning, L., Schernhammer, E., Willett, W.C., Hu, F.B., 2019. Long-term consumption of sugar-sweetened and artificially sweetened beverages and risk of mortality in US adults. Circulation 139, 2113–2125.
Melanson, K.J., Zukley, L., Lowndes, J., Nguyen, V., Angelopoulos, T.J., Rippe, J.M., 2007. Effects of high-fructose corn syrup and sucrose consumption on circulating glucose, insulin, leptin, and ghrelin and on appetite in normal-weight women. Nutrition 23, 103–112.
Mok, A., Ahmad, R., Rangan, A., Louie, J.C.Y., 2018. Intake of free sugars and micronutrient dilution in Australian adults. Am. J. Clin. Nutr. 107, 94–104.
Moss, M.E., Luo, H., Rosinger, A.Y., Jacobs, M.M., Kaur, R., 2021. High sugar intake from sugar-sweetened beverages is associated with prevalence of untreated decay in US adults: NHANES 2013–2016. Community Dent. Oral Epidemiol. https://doi.org/10.1111/cdoe.12725.
Mouries, J., Brescia, P., Silvestri, A., Spadoni, I., Sorribas, M., Wiest, R., Mileti, E., Galbiati, M., Invernizzi, P., Adorini, L., Penna, G., Rescigno, M., 2019. Microbiota-driven gut vascular barrier disruption is a prerequisite for non-alcoholic steatohepatitis development. J. Hepatol. 71, 1216–1228.
Mullie, P., Mertens, E., Charlier, R., Knaeps, S.S., Lefevre, J., Clarys, P., 2018. Relation between sugar-sweetened beverage consumption and micronutrient intake in a prospective study. Eur. J. Clin. Nutr. 72, 170–173.
Pacheco, L.S., Lacey, J.V., Martinez, M.E., Lemus, H., Araneta, M.R.G., Sears, D.D., Talavera, G.A., Anderson, C.A.M., 2020. Sugar-sweetened beverage intake and cardiovascular disease risk in the California Teachers study. J. Am. Heart Assoc. 9, e014883.
Page, K.A., Chan, O., Arora, J., Belfort-Deaguiar, R., Dzuira, J., Roehmholdt, B., Cline, G.W., Naik, S., Sinha, R., Constable, R.T., Sherwin, R.S., 2013. Effects of fructose vs glucose on regional cerebral blood flow in brain regions involved with appetite and reward pathways. JAMA 309, 63–70.
Pitchika, V., Standl, M., Harris, C., Thiering, E., Hickel, R., Heinrich, J., Kühnisch, J., 2020. Association of sugar-sweetened drinks with caries in 10- and 15-year-olds. BMC Oral Health 20, 81.
Reiser, S., Bohn, E., Hallfrisch, J., Michaelis, O.E., Keeney, M., Prather, E.S., 1981. Serum insulin and glucose in hyperinsulinemic subjects fed three different levels of sucrose. Am. J. Clin. Nutr. 34, 2348–2358.
Reiser, S., Hallfrisch, J., Michaelis, O.E., Lazar, F.L., Martin, R.E., Prather, E.S., 1979. Isocaloric exchange of dietary starch and sucrose in humans. I. Effects on levels of fasting blood lipids. Am. J. Clin. Nutr. 32, 1659–1669.
Rennie, K.L., Livingstone, M.B., 2007. Associations between dietary added sugar intake and micronutrient intake: a systematic review. Br. J. Nutr. 97, 832–841.
Righolt, A.J., Jevdjevic, M., Marcenes, W., Listl, S., 2018. Global-, regional-, and country-level economic impacts of dental diseases in 2015. J. Dent. Res. 97, 501–507.
Rosas-Villegas, A., Sánchez-Tapia, M., Avila-Nava, A., Ramírez, V., Tovar, A.R., Torres, N., 2017. Differential effect of sucrose and fructose in combination with a high fat diet on intestinal microbiota and kidney oxidative stress. Nutrients 9.
Shafer, W.G., 1949. The caries-producing capacity of starch, glucose, and sucrose diets in the Syrian hamster. Science 110, 143–144.
Sørensen, L.B., Vasilaras, T.H., Astrup, A., Raben, A., 2014. Sucrose compared with artificial sweeteners: a clinical intervention study of effects on energy intake, appetite, and energy expenditure after 10 wk of supplementation in overweight subjects. Am. J. Clin. Nutr. 100, 36–45.
Tappy, L., Lê, K.A., 2010. Metabolic effects of fructose and the worldwide increase in obesity. Physiol. Rev. 90, 23–46.

Touger-Decker, R., Van Loveren, C., 2003. Sugars and dental caries. Am. J. Clin. Nutr. 78, 881S–892S.

Tsilas, C.S., de Souza, R.J., Mejia, S.B., Mirrahimi, A., Cozma, A.I., Jayalath, V.H., Ha, V., Tawfik, R., Di Buono, M., Jenkins, A.L., Leiter, L.A., Wolever, T.M.S., Beyene, J., Khan, T., Kendall, C.W.C., Jenkins, D.J.A., Sievenpiper, J.L., 2017. Relation of total sugars, fructose and sucrose with incident type 2 diabetes: a systematic review and meta-analysis of prospective cohort studies. CMAJ 189, E711–E720.

Valenzuela, M.J., Waterhouse, B., Aggarwal, V.R., Bloor, K., Doran, T., 2021. Effect of sugar-sweetened beverages on oral health: a systematic review and meta-analysis. Eur. J. Publ. Health 31, 122–129.

Wang, X., Ouyang, Y., Liu, J., Zhu, M., Zhao, G., Bao, W., Hu, F.B., 2014. Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. BMJ 349, g4490.

Wasserman, D.H., 2009. Four grams of glucose. Am. J. Physiol. Endocrinol. Metab. 296, E11–E21.

Yu, Z., Ley, S.H., Sun, Q., Hu, F.B., Malik, V.S., 2018. Cross-sectional association between sugar-sweetened beverage intake and cardiometabolic biomarkers in US women. Br. J. Nutr. 119, 570–580.

Yunker, A.G., Luo, S., Jones, S., Dorton, H.M., Alves, J.M., Angelo, B., Defendis, A., Pickering, T.A., Monterosso, J.R., Page, K.A., 2021. Appetite-regulating hormones are reduced after oral sucrose vs glucose: influence of obesity, insulin resistance, and sex. J. Clin. Endocrinol. Metab. 106, 654–664.

Zhan, J., Liu, Y.J., Cai, L.B., Xu, F.R., Xie, T., He, Q.Q., 2017. Fruit and vegetable consumption and risk of cardiovascular disease: a meta-analysis of prospective cohort studies. Crit. Rev. Food Sci. Nutr. 57, 1650–1663.

Zhang, X., Monnoye, M., Mariadassou, M., Beguet-Crespel, F., Lapaque, N., Heberden, C., Douard, V., 2021. Glucose but not fructose alters the intestinal paracellular permeability in association with gut inflammation and dysbiosis in mice. Front. Immunol. 12, 742584.

Zhao, S., Jang, C., Liu, J., Uehara, K., Gilbert, M., Izzo, L., Zeng, X., Trefely, S., Fernandez, S., Carrer, A., Miller, K.D., Schug, Z.T., Snyder, N.W., Gade, T.P., Titchenell, P.M., Rabinowitz, J.D., Wellen, K.E., 2020. Dietary fructose feeds hepatic lipogenesis via microbiota-derived acetate. Nature 579, 586–591.

Supplementation: Dietary supplements

S.S. Percival, University of Florida, Gainesville, FL, USA

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of S.S. Percival, Supplementation: Dietary Supplements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 246–250, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00260-9.

Introduction 606
How is the Sales Data Derived? 606
What is a Dietary Supplement? How are they Regulated in Different Countries? 606
Product Quality and Safety Issues 608
Interactions 609
Label Claims 609
Potential Benefits of Dietary Supplements 610
Conclusions 610
Further Reading 611

Introduction

In 2004, global sales of dietary supplements represented a significant business. Worldwide sales have been estimated at $70–250 billion. The demand for herbal products worldwide increased at an annual rate of 8% from 1994 to 2001, although this growth has slowed in recent years.

Issues and controversies in the dietary supplement market are related to defining exactly what is a dietary supplement, understanding how sales and marketing data are derived, defining the regulatory environment, safety issues, product quality issues, labeling and health claim issues, and scientific evidence for benefit. This article describes some of these controversies and provides examples to illustrate these issues.

How is the Sales Data Derived?

Global sales have been estimated to be between $70 billion and $250 billion. This approximately three-fold difference in estimates is due to the variation in what products are actually included in product sales results. As will be discussed, the definition of dietary supplements varies greatly from country to country; therefore, deriving sales data is complex.

Another difficulty in assessing sales of dietary supplements is the source from which sales data are gathered. Many business surveys rely on only one or two of the following sales outlets to derive their results:

- Supermarkets and mass merchandisers
- Natural food and health food stores
- Direct sales from Internet, mail order, practitioners, and multilevel marketing
- Pharmacies and drugstore chains

What is a Dietary Supplement? How are they Regulated in Different Countries?

Each country has developed regulatory definitions and systems that place dietary supplements, particularly botanicals, into categories of drugs, traditional medicines, or foods. However, in the late 1980s, many countries launched major changes in regulations that may or may not have been approved at the time of this writing. Many regulations are still in draft form.

The US Congress defined the term 'dietary supplement' in the Dietary Supplement Health and Education Act (DSHEA) of 1994. A dietary supplement is a product, taken orally, that contains a 'dietary ingredient' that is intended to supplement the diet. The dietary ingredient includes vitamins, minerals, herbs or other botanicals, amino acids, a dietary substance for use by man to supplement the diet by increasing the total dietary intake (e.g., enzymes or tissues from organs or glands), or a concentrate, metabolite, constituent, or extract. Dietary supplements may be found in many forms, such as tablets, capsules, softgels, gelcaps, liquids, or powders. They may also be produced in other forms, such as a beverage, spread, or bar, in which case information on the label must clearly state that the product is a dietary supplement and it is not represented as a conventional food or a sole item of a meal or diet.

Whatever their form, DSHEA places dietary supplements in a special category under the general umbrella of 'foods,' not drugs, and requires that every supplement be labeled a dietary supplement and carry a Supplement Facts Label.

In UK, there is a distinct separation of food supplements and herbal medicines. The Food Standards Agency developed the Food Standard Act of 1999 and is responsible for protection of public health. The Food Supplement Directive 2002/46/EC, which harmonizes European Community legislation on food supplements, was published in 2002. This directive is stricter than existing UK standards and regulations but is relatively more liberal than that which exists in other European countries. The directive defines the term 'food supplements,' contains a list of vitamin and mineral sources that may be used in the manufacture of food supplements, states labeling requirements, and, in the future, will provide a framework for maximum and minimum levels for vitamins and minerals in food supplements. Herbals and botanicals are not discussed in this directive.

The Foods Supplement Directive defines a food supplement as any food the purpose of which is to supplement the normal diet and which is a concentrated source of a vitamin or mineral or other substance with a nutritional or physiological effect, alone or in combination, and is sold in dose form. Dose form means capsules, pastilles, tablets, pills, and other similar forms, and also powders, ampoules, drops, or other similar forms of liquids or powders, designed to be taken in small measured quantities. Because the directive defines a food supplement as something to supplement the diet, products that are not meant to supplement the diet (e.g., a weight loss product) are outside the scope of the regulations. There remains a complex legal area between food supplements and medicinal products, although the directive indicates that if a product is used for treating or preventing disease, or restoring, correcting, or modifying a physiological function, then it falls under the Medicines Directive 2001/83/EEC, Medicines Act 1968, or Medicines for Human Use Regulations 1994.

The Transatlantic Business Dialogue (TABD) approved a position statement regarding dietary supplements in 2002. The TABD is a group of corporations that promote closer commercial ties between the European Union and the United States. This position statement established industrywide consensus on standards and definition of permissible claims, as well as defining what is necessary for substantiation of those claims. In keeping with the Foods Supplement Directive, the TABD dealt only with vitamins and minerals, with the understanding that some of the conclusions may be revisited when warranted for herbals, botanicals, or other dietary supplements.

Herbal medicines, however, are regulated by the Medicine and HealthCare Products Regulatory Agency based in London. A herbal remedy is defined as

> a medicinal product consisting of a substance produced by subjecting a plant or plants to drying, crushing or any other process, or of a mixture whose sole ingredients are two or more substances so produced, or of a mixture whose sole ingredients are one or more substances so produced and water or some other inert substance.

There are two alternative regulatory routes in UK for herbal medicines: Licensing and exemption from licensing requirements:

- Licensed herbal medicines: To receive a product license before marketing, herbal medicines are required to meet safety, quality, and efficacy criteria in a similar manner to any other licensed medicine.
- Herbal remedies exempt from licensing requirements: The exemption applies to herbal remedies meeting certain conditions set out in Section 12 of the Medicines Act 1968. Section 12 allows a person to make, sell, and supply a herbal remedy during the course of his or her business provided the remedy is manufactured or assembled on the premises and that it is supplied as a consequence of a consultation between the person and his or her patient. Section 12 also allows the manufacture, sale, or supply of herbal remedies where the processing of the plant consists only of drying, crushing, or comminuting; the remedy is sold without any written specification as to its use; and the remedy is sold under a designation that only specifies the plant and the process and does not apply any other name to the remedy.

Canada has been estimated to have approximately 3% of the market share of the global nutritional market. Health Canada established the Office of Natural Health Products. Premarket assessment, labeling, licensing, and monitoring of herbal supplements are items in its mandate. The definition of a natural health product includes products for the use in 'diagnosis, treatment, mitigation, or prevention of a disease, disorder, or abnormal physical state or its symptoms in humans; restoring or correcting organic function in humans; or modifying organic functions in humans, such as modifying those functions in a manner that maintains or promotes health.' These products include homeopathic preparations, substances used in traditional medicine, a mineral or trace element, a vitamin, an amino acid, an essential fatty acid or other botanical-, animal-, or microorganism-derived substance. Foods are not included in this product category called natural health products. Canada's Food and Drugs Act of 1953 regulates foods and drugs but does not specifically deal with natural health products. Therefore, these types of products are regulated as either a food or a drug depending on the type and concentration of active ingredient and whether claims are made on the products.

Germany regulates vitamins and minerals as food if they are sold to complement the nutritive value of the diet and do not exceed safe levels. However, if the vitamin or mineral is used for disease treatment or prevention and is used at pharmacological levels, then it is considered a drug. Safety and efficacy of drugs must be established by clinical research. Medicinal plants are regulated differently depending on what plant and in what form it is sold. In general, extracts of plants are considered drugs and must be prescribed. Teas, however, are sold over-the-counter in pharmacies. Other teas, such as those that contain alkaloids, must be sold by prescription only. Beginning in 1980, an extensive analysis of the literature on more than 300 herbal remedies was undertaken by the German Kommission E. Approximately two-thirds of the herbals were listed as safe and at least minimally effective. The results were published as a series of monographs by the German Kommission E, and this body of work was summarized and translated into English

by the American Botanical Council. These substances are generally purchased at the pharmacy and are reimbursable through health insurance. One caveat regarding the German herbal preparations is that they are not likely to be the same preparations that are produced by other countries; thus, the safety and efficacy statements in the Kommission E are only for the preparations that are prepared in German pharmacies.

Australia regulates therapeutic goods under the Therapeutic Goods Act of 1989. Therapeutic goods include vitamins, minerals, plants and herbals, nutritional food supplements, naturopathic and homeopathic preparations, and some aromatherapy. The Therapeutic Goods Administration (TGA) developed the Office of Complementary Medicine to evaluate new substances and products. Basically, the TGA regulates these therapeutic goods as they do pharmaceutical products, and thus their criteria are more rigorous than the criteria of other countries. Most of the therapeutic goods are 'generally listed' rather than regulated. Listed medicines are considered to be relatively harmless, so the regulations allow for manufacturers to 'self-assess' their products in some situations. The majority of listed medicines are self-selected by consumers and used for self-treatment, and they are all manufactured with well-known established ingredients, such as vitamin and mineral products or sunscreens. These are assessed by the TGA for quality and safety but not efficacy. This does not mean that they do not work; rather, it means that the TGA has not evaluated them individually to determine if they work. It is a requirement under the act that sponsors have information to substantiate all of their product's claims.

The Japanese Ministry of Health and Welfare does not define or recognize a distinct category known as dietary supplements. Instead, there are only two classifications, food and drugs. In 1993, Japan defined a group of foods known as Foods for Specific Health Use (FOSHU). As of 2004, approximately 342 foods had been approved as FOSHU. The dietary ingredients are sold in the form of foods, not in the form of capsules, tablets, or powders.

The herbal supplements market in Japan has been strongly influenced by the practice of Kampo. Kampo (or Kanpo) is the adaptation of Chinese herb formulas to Japanese medicine. Approximately 25 years ago, the Japanese Ministry of Health formally recognized that certain traditional Chinese herb formulas (and a few formulas of similar nature developed in Japan) were suitable for coverage by national health insurance. These formulas are prepared in factories under strict conditions.

In summary, developing global data on dietary supplement sales depend on how they are defined. (**Table 1**) summarizes the differences in regulatory categories of different countries.

Product Quality and Safety Issues

Product quality is an issue derived from the explosive growth of the industry in the post-DSHEA world. Quality issues revolve around products that contain wrong ingredients, incorrect claims, contamination, or incorrect amounts – either too much or not enough.

An example plant misidentification was published in 1998 by Slifman *et al*. Two patients were admitted to hospital emergency rooms with palpitations, vomiting, nausea, and chest pressure, among other symptoms. Both individuals, having been admitted 1 month apart, had each consumed a program of dietary supplements, one containing 14 herbs, a tablet containing 11 herbs, liquid clay, a bulking powder, and capsules containing microorganisms. Of the five supplements, the one made up of 14 herbs tested positive for cardiac glycosides. The investigators determined that *Digitalis lanata* was present in the supplement. *D. lanata* contains cardiac glycosides, which resulted in the cardiac symptoms. Further investigation revealed that raw material labeled as plantain (genus *Plantago*) had been contaminated with *D. lanata* due to misidentification in the field.

Another quality issue that has safety manifestations was an incorrect claim on a product. PC-SPES, a combination of eight herbs, is claimed to be a nonestrogenic treatment for prostate cancer. However, several of the herbs used in this preparation do in fact have estrogenic activity. In 1998, DiPaola *et al.* showed a significant amount of estrogenic activity in both *in vitro* (yeast) and *in vivo*

Table 1 Regulatory categories of different countries

Country, act	*Definition*
United States, DSHEA	Vitamins, minerals, herbal, other botanical, amino acid, enzymes, organs, glands
Europe, Food Supplement Act	Vitamin and minerals
United Kingdom, Medicine and Health Care	Medicinal plants
Canada, Office of Natural Products	Mineral; trace element; vitamin; amino acid; essential fatty acid; botanical-, animal-, or microorganism-derived substances; homeopathic preparation; traditional preparations
Germany, Kommission E	Vitamin and mineral as both foods and drugs, botanicals (approved and not approved), teas as prescription and as over-the-counter
Australia, Therapeutic Goods Administration	Vitamin and mineral, plants, herbs, nutritional food supplements, naturopaths and homeopathic preparations, aromatherapies
Japan, Ministry of Health and Welfare	No definition of dietary supplements, regulations for foods, drugs, and Kampo

studies (mice and humans) with PC-SPES. Use of the supplement by men with prostate cancer resulted in similar side effects as would develop with estrogen therapy and theoretically could confound the results of standard therapy.

By law (DSHEA), the manufacturer is responsible for ensuring that its dietary supplement products are safe before they are marketed. Unlike drug products that must be proven safe and effective for their intended use before marketing, there are no provisions in the law for the US Food and Drug Administration (FDA) to 'approve' dietary supplements for safety or effectiveness before they reach the consumer. Also unlike drug products, manufacturers and distributors of dietary supplements are not required by law to record, investigate, or forward to the FDA any reports they receive of injuries or illnesses that may be related to the use of their products. Under DSHEA, once the product is marketed, the FDA has the responsibility to show that a dietary supplement is 'unsafe' before it can take action to restrict the product's use or remove it from the marketplace.

In 2003, the FDA banned all products containing ephedra alkaloids. Ephedra-containing products were, until the ban, marketed in conjunction with enhancing athletic performance and/or promoting weight loss. Recent studies provided enough additional evidence that ephedra presents a significant and unreasonable risk of illness and injury that the FDA banned all ephedra-containing products from the market and advised consumers to stop taking such supplements. Strong statements were issued cautioning about the use of ephedra-containing products, especially when strenuously exercising or in combination with other stimulants, such as caffeine.

Interactions

An issue that has become of concern is the interaction of dietary supplements with herbs and other dietary supplements, drugs, foods, lab tests, and diseases or other conditions. There are literally hundreds of potential interactions that have not yet been recognized. Both practitioners and consumers must be aware of the possibilities. In some cases, knowledge about interactions comes from documented reports. However, in other cases, the knowledge is theoretical, based on the pharmacological profile or mechanism of action of the supplement and the drug, food, test, or condition. For example, ginkgo biloba contains ginkgolides in the leaf that competitively inhibit platelet-activating factor (PAF). PAF inhibition decreases platelet aggregation among other many other physiological effects. Inhibition of PAF may increase cardiac contractility and coronary blood flow. Concomitant use of herbs and supplements that affect platelet aggregation could theoretically increase the risk of bleeding in some people due to ginkgo's effects on platelet aggregation. Spontaneous hematomas (broken blood vessels) and hemorrhaging in the anterior chamber of the eye have been reported in ginkgo users, although it is not known what other drugs or supplements these individuals were taking.

Herbs and supplements that promote platelet inhibition include angelica, anise, capsicum, celery, chamomile, clove, fenugreek, feverfew, fish oil, garlic, ginger, horse chestnut, horseradish, licorice, meadowsweet, onion, Panax ginseng, red clover, vitamin E, and willow. Similarly, concomitant administration of drugs, including aspirin, clopidogrel (Plavix), dalteparin (Fragmin), enoxaparin (Lovenox), heparin, indomethacin (Indocin), ticlopidine (Ticlid), and warfarin (Coumadin), may increase the risk of bleeding in some people. This is just one example of the interactions between drugs with herbals and herbals with other herbals. There may be an infinite number of interactions.

Currently, there are no mandated US federal guidelines to report adverse events or consumer health complaints associated with the use of dietary supplements. MedWatch reporting is voluntary. In 2004, the Life Sciences Research Office published a report, *Recommendations for Adverse Event Monitoring Programs for Dietary Supplements*.

Label Claims

Label claims regarding dietary supplements are a complex issue that varies from country to country. Yet no matter what specific claims are allowed or disallowed by a country, it is reasonable to assume that any global regulation requires that the claim be true, not misleading, and be clear to the consumer. A summary of US label claims follows.

The Nutrition Labeling and Education Act (NLEA) was passed in 1990 as a result of a pre-1984 FDA position that prohibited making any therapeutic or disease-related claims on a food or dietary supplement label. The NLEA permits certain claims describing a positive relationship between a supplement and a health-related condition (or disease). These claims are considered 'health claims' in order to distinguish them from nutrient content claims. A health claim must be authorized by the FDA, and the FDA can only authorize a claim if there is 'significant scientific agreement among qualified experts' or by the 1997 amendment that permits a manufacturer to rely on a statement from an 'authoritative scientific body' of the US government or the National Academy of Sciences. This is a rigorous assessment and only 14 claims have been authorized to date.

In addition to health claims, dietary supplement labels are permitted to have qualified health claims or structure-function claims. The rationale behind the development of a qualified health claim was the idea that the First Amendment should allow disclaimers to be considered as solutions to making claims nonmisleading (Pearson vs Shalala). In other words, the First Amendment does not allow the FDA to reject health claims unless it shows that disclaimers would fail to remedy harm from misleading statements. The criteria for a qualified health claim were released in 2003 and in this context the FDA will not take enforcement action against a manufacturer using the following specified qualifiers provided the FDA is satisfied that the qualifiers are not misleading:

- "Although there is scientific evidence supporting the claim, the evidence is not conclusive."
- "Some scientific evidence suggests…. However, FDA has determined that this evidence is limited and not conclusive."
- "Very limited and preliminary scientific research suggests…. FDA concludes that there is little scientific evidence supporting this claim."

Qualified health claims for dietary supplements recognized by the FDA as part of its enforcement discretion include such examples as the relationships between phosphatidylserine and cognitive function, B vitamins and cardiovascular disease, omega-3 fatty acids and cardiovascular disease, selenium and cancer, and antioxidant vitamins and cancer.

Dietary supplements are not permitted to carry labeling statements that imply such issues as 'cure,' 'mitigate,' 'treat,' or 'prevent disease' because these statements are considered within the definition of a drug and drugs are subjected to a rigorous premarket approval process. However, under DSHEA, structure-function claims are permitted on dietary supplements because dietary supplements may have effects on the structure or function of the body without the implication that they act as a drug and/or are related to disease. Structure-function claims include those that describe the role of the dietary supplement in affecting the structure or function in humans or the documented mechanism in which a dietary supplement acts to maintain such structure or function. In addition, dietary supplement label claims allow statements of benefits related to classical nutritional deficiency or statements regarding the general feeling of well-being derived from consumption.

Potential Benefits of Dietary Supplements

The 2000 *Dietary Guidelines for Americans* (new release due 2005) emphasizes choosing foods sensibly, maintaining a healthy weight, and exercising regularly. It acknowledges that some people may need a vitamin–mineral supplement to meet specific needs. Similarly, the Food and Nutrition Board and the American Dietetic Association also recognize that dietary supplements may be desirable for some nutrients and for some individuals. The following is a compilation of recommendations by these groups:

- Folic acid supplements for women of childbearing age due to the risk of neural tube defects
- Vitamin B_{12} supplements for people older than age 50 years due to inefficient absorption
- Vitamin B_{12} supplements for vegans who eat no animal products
- Calcium for people who seldom eat dairy products
- Vitamin D for elderly people who do not consume fortified dairy products and for others with little exposure to sunlight
- Iron supplementation for pregnant women
- Multivitamin-mineral supplement for people who are following a severely restricted weight-loss diet.

Specifically for athletes, the position of the American Dietetic Association, Dietitians of Canada, and the American College of Sports Medicine is that physical activity, athletic performance, and recovery from exercise are enhanced by optimal nutrition. These organizations recommend appropriate selection of food and fluids, timing of intake, and supplement choices for optimal health and exercise performance. In sports, athletes who are at greatest risk of micronutrient deficiencies are those who restrict energy intake or use severe weight-loss practices, eliminate one or more food groups from their diet, are sick or recovering from injury, or consume high-carbohydrate diets with low micronutrient density. In practice, athletes should consume diets that provide at least the Recommeded Dietary Allowances/Direct References Intakes for all micronutrients from food. It follows that, in general, no vitamin and mineral supplements are required if an athlete is consuming adequate energy from a variety of foods to maintain body weight. Supplementation may be necessary under conditions of inadequate food intake. Athletes, as for the general population, should follow supplementation recommendations unrelated to exercise, such as folic acid in women who may become pregnant.

Conclusions

One of the difficulties in assessing the nature of the worldwide dietary supplement industry and its regulations is largely in understanding what products are considered dietary supplements. In the United States, only pills, capsules, tablets, and the like are considered dietary supplements. Globally, it is sometimes difficult to discuss dietary supplements without discussing functional foods or nutraceuticals. Functional foods are similar in appearance to conventional foods but have demonstrated physiological benefits beyond the traditional nutritional value. Nutraceuticals may go so far as to declare not only health benefits, but also medical benefits that reduce the risk of chronic disease beyond basic nutritional functions. Canada regulates functional foods, nutraceuticals, and dietary supplements under one regulatory agency. The United States clearly distinguishes between foods and dietary supplements, although both fall under the category of food, which is distinct from drugs. UK distinguishes between herbal medicines and dietary supplements containing vitamins and minerals. Japan regulates functional foods as FOSHU and has no regulatory definition for dietary supplements as defined in the United States. Moreover, these regulations are in a constant state of flux as the industry changes and develop over time. Issues that must be monitored regarding dietary supplements consumption are product quality and potential harmful interactions among supplements, foods, and drugs. Health claims that have been approved by regulatory agencies worldwide stress that the claims be truthful, clear, and not misleading to the ultimate consumer. Current scientific

expertise acknowledges that dietary supplements, specifically some of the vitamins and minerals, have potential benefits in certain populations.

Further Reading

Anonymous, 2000. Nutrition and athletic performance. Position of the American Dietetic Association, Dietitians of Canada, and the American College of Sports Medicine. Journal of the American Dietetic Association 100, 1543–1556.

Huang, S.M., Lesko, L.J., 2004. Drug–drug, drug–dietary supplement, and drug–citrus fruit and other food interactions: What have we learned? Journal of Clinical Pharmacology 44 (6), 559–569.

Percival, S.S., Turner, R.E., 2000. Applications of herbs to functional foods. In: Wildman, R. (Ed.), Handbook of Functional Foods. CRC Press, Boca Raton, FL.

Turner, R.E., Degnan, F.H., Archer, D.L., 2005. Nutrition in Clinical Practice 20, 21–32.

U.S. Department of Agriculture, Dietary Guidelines for Americans US Department of Agriculture, Washington, DC (2000) Available at http.www.health.gov/dietaryguidelines/dgac

Sustainable diets: Their definition, measurement and promotion

Shauna M. Downs[a], **Alexandra Payne**[b], **and Wiktoria Staromiejska**[a], [a] Rutgers School of Public Health, Newark, NJ, United States; and [b] Amplify Studio, Urban Planning, Brooklyn, NY, United States

© 2023 Elsevier Ltd. All rights reserved.

Introduction 612
Defining sustainable diets 613
Rationale for promoting sustainable diets 613
Moving toward a global sustainable diet 613
Measuring sustainable diets 615
Shifting consumption toward sustainable diets: what are the challenges? 617
Solutions to help support sustainable diets 618
Consumer-facing solutions 618
Inform and empower 618
Guide and influence 620
Incentivize, discourage or restrict 620
Supply-side solutions 620
Advancing the sustainable diets agenda: the role of different stakeholder groups 621
The role of local government 621
The role of social movements and role models 621
Research gaps and recommendations for moving toward sustainable diets 622
Summary 622
References 623

Key points

- Sustainable diets encompass health, environmental, social and economic dimensions of the way food is produced and consumed
- There are often trade-offs between the different dimensions (health, environmental, social and economic) of sustainable diets
- Measuring sustainable diets in a comprehensive way is challenging, with the emphasis most often being placed on health and environmental aspects of diets
- Shifting global diets toward more plant-based foods (e.g., Planetary Health Diet) will be critical to achieving more sustainable consumption patterns
- Both consumer-facing and upstream food systems solutions will be needed to promote and support sustainable diets
- Social movements, local government actions and chefs have the potential to increase the awareness and adoption of sustainable diets

Introduction

The United Nations (UN) Sustainable Development Goals (SDGs) were adopted in 2015 with the view of providing a blueprint "*for peace and prosperity for people and the planet, now and into the future*" (United Nations, 2015). SDG 2—to "*end hunger, achieve food security and improve nutrition and promote sustainable agriculture*"—highlights the importance of food and nutrition in meeting the 2030 goals (United Nations, 2015). In addition to the SDGs, there are several global goals aimed at tackling food insecurity, suboptimal diets and malnutrition (e.g., World Health Organization (WHO) global target 2025 to improve maternal, infant and young child nutrition, WHO Non-communicable Disease targets). Alongside these global goals, the UN Food Systems Summit 2021 resulted in a myriad of commitments to transform the way the world produces and consumes food, highlighting the need to shift toward more sustainable consumption patterns. These international goals and actions highlight the growing prominence of sustainable diets on the global agenda. Global shifts toward sustainable diets could help to achieve SDG 2 and beyond; however, the world is currently off track to meet most of the global nutrition goals (FAO et al., 2021).

In 2020, between 720 and 811 million people were undernourished, a number that has been increasing since 2014 and is projected to increase even further due to the COVID-19 pandemic (FAO et al., 2021). At the same time, billions of people worldwide experience malnutrition in one or more of its forms: 149 million children under five are stunted, 45 million are wasted, over two

 https://doi.org/10.1016/B978-0-12-821848-8.00089-5

billion adults are overweight or obese, and approximately 2 billion people worldwide are deficient in key micronutrients (FAO et al., 2021). Moreover, 3 billion people globally cannot afford a healthy diet (FAO et al., 2021).

Consuming diverse, nutritious diets has the potential to help address the burden of malnutrition. However, current global diets are suboptimal and are among the top risk factors contributing to the global burden of disease (Afshin et al., 2019). While there are differences in dietary intakes across countries, much of the global population is under-consuming whole grains, legumes, nuts, fruits and vegetables while over-consuming red meat and starchy vegetables (Willett et al., 2019). However, this masks the inequality in dietary intakes and malnutrition across populations globally, and within given countries (Bell et al., 2021).

In many countries, the populations with the highest burden of poor-quality diets and malnutrition are employed within the food system. Subsistence farmers, particularly in Sub-Saharan Africa and Asia, are often malnourished and experience high levels of food insecurity. Their livelihoods are also at risk due to increased climate variability, including shocks. In high-income countries like the United States, people working in food production, harvesting, processing and service are often paid below a living wage with insufficient access to social safety nets, increasing their risk of food insecurity, poor quality diets and disease. In order to ensure that all populations have access to affordable, nutritious and sustainably produced food, real changes are needed in the way we produce food, how it moves through the food value chain, and how we consume it.

Defining sustainable diets

In 1986, Gussow and Clancy first introduced the concept of sustainable diets (Gussow and Clancy, 1986). They proposed that the nutrition community should go beyond focusing solely on the relationship between food and human health and also consider the health of the planet (Gussow and Clancy, 1986). While little progress was made to move this agenda forward for many years, over the past decade it has begun to receive considerable attention. While sustainable diets have been described in a variety of ways, the Food and Agriculture Organization (FAO) defines them as *"those diets with low environmental impacts that contribute to food and nutrition security and to healthy life for present and future generations. Sustainable diets are protective and respectful of biodiversity and ecosystems, culturally acceptable, accessible, economically fair and affordable, nutritionally adequate, safe, and healthy, while optimizing natural and human resources"* (FAO, 2012). The definition is comprehensive, including health, environmental, social and economic dimensions. However, given the breadth of dimensions of sustainable diets there will undoubtedly be trade-offs among them.

Fig. 1 provides an overview of the key principles of sustainable diets (FAO and WHO, 2019). While these principles are helpful in terms of providing insight into the elements of sustainable diets, it remains difficult to conceptualize what a sustainable diet looks like in practice, particularly among different populations and contexts. As such, there are challenges to providing an overarching definition of sustainable diets, given that so much of what makes diets sustainable are context specific.

Rationale for promoting sustainable diets

The food we consume, and how it is produced and moves along the food value chain, has serious implications for the health of people but also for the planet. The food system is responsible for approximately a third of greenhouse gas emissions, 30% of freshwater use and the largest contribution to biodiversity loss globally (Willett et al., 2019). It is putting immense pressure on the Earth's systems, leading us to surpass planetary boundaries (Willett et al., 2019).

Animal source foods, and ruminant meat (e.g., beef, lamb, etc.) in particular, are the main contributors to environmental degradation attributed to food systems (Willett et al., 2019). **Fig. 2** depicts the environmental effects of consuming different foods and food groups. While the environmental impact of animal source food consumption is already high, this is being exacerbated by increasing global intakes. **Fig. 3** describes the changes in animal source food availability (used as a proxy for consumption) over time based on socio-demographic Index (SDI), a metric that takes into account incomes per capita, average educational attainment, and fertility rates. While egg and fish availability has remained relatively stable since the early 1960s, milk and meat consumption has been growing particularly among higher SDI countries.

Given that ruminant meat production has the largest impact on the environment of any food group, and its intakes are currently increasing, it's clear that intakes need to decrease in order to stay within planetary boundaries. However, it is important to acknowledge the significant inequality in the availability of animal source food across countries globally. Approximately one-fifth of the global population consumes less than 10% of their energy from animal source foods (Bell et al., 2021). This is important given that animal source foods can be critical sources of key nutrients, particularly for young children. Increasing animal source food consumption among the lowest-income consumers globally, and decreasing consumption among the highest, has the most potential to lead to significant improvements in both health and environmental outcomes (Willett et al., 2019).

Moving toward a global sustainable diet

In recognition that global diets need to shift both from a health and an environmental perspective, the EAT Lancet Commission described a healthy reference diet that provides a blueprint for a global diet that sustains health while protecting the planet (Willett et al., 2019). The "Planetary Health Diet" emphasizes the consumption of plant foods, while limiting consumption of meat and

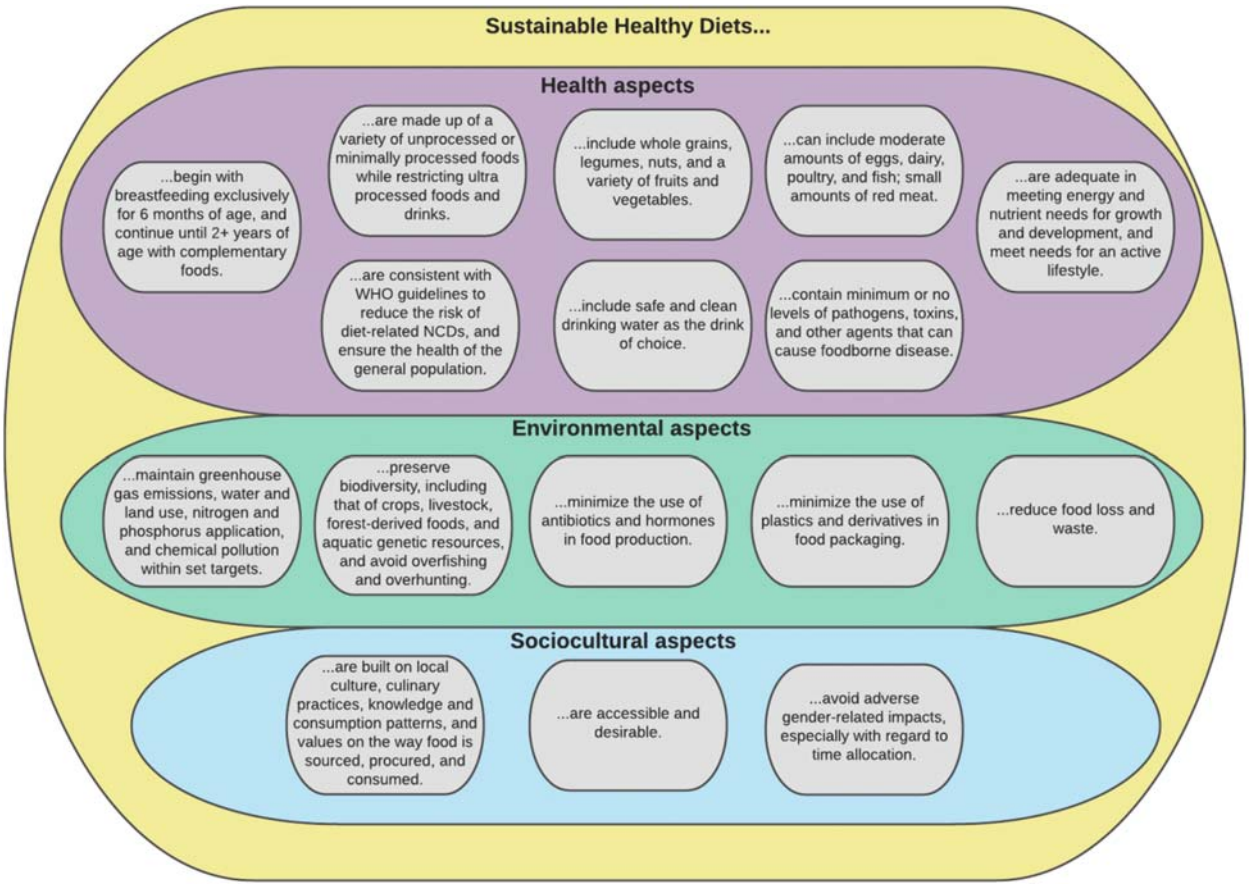


Fig. 1 An overview of the principles of a sustainable diet. Adapted from FAO and WHO (2019).

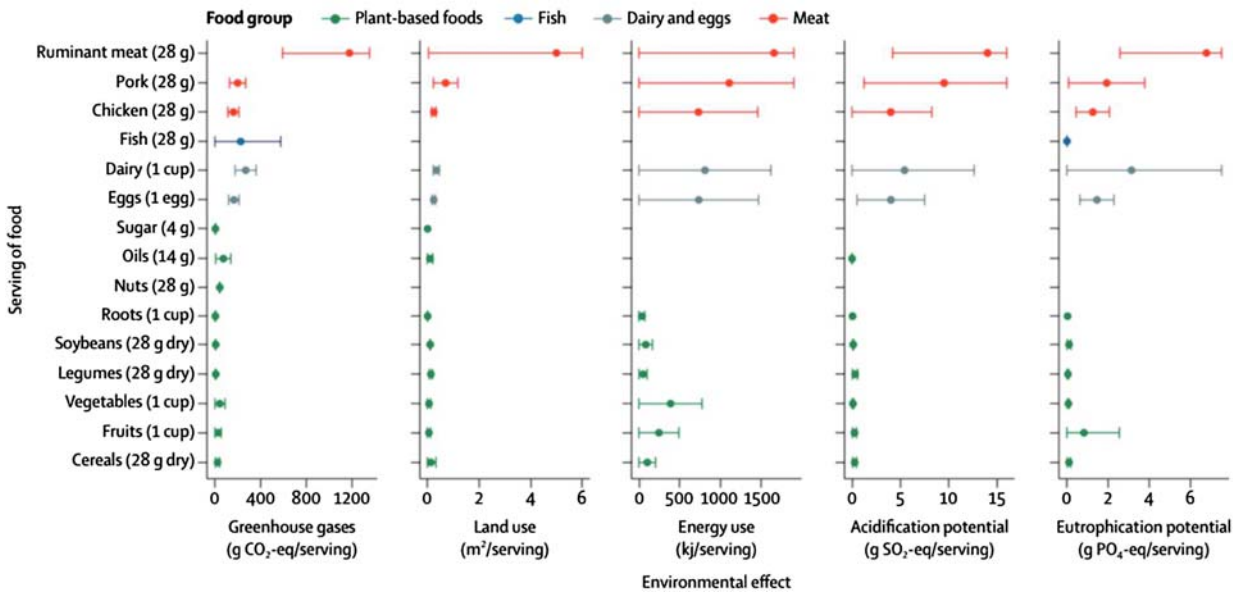


Fig. 2 Environmental effects per serving of food produced. Source: EAT Lancet. Permission has been obtained to reproduce.

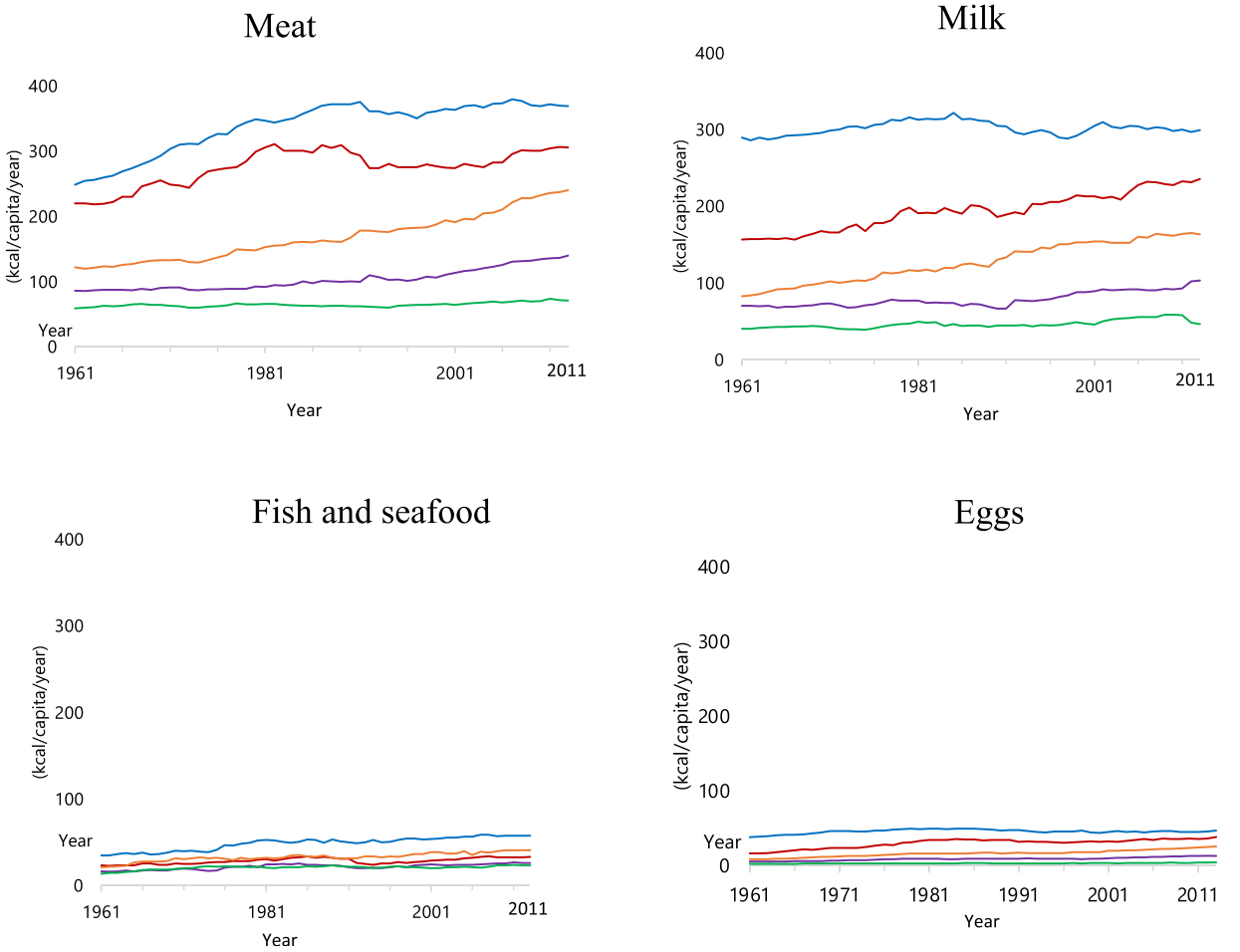


Fig. 3 Shifts in the availability of food over time (1961–2013) by SDI grouping. Source: FAOSTAT, https://www.fao.org/faostat/en/.

dairy (Fig. 4). It provides recommendations for the quantity of each food/food group to be consumed per day. While the diet is not designed to be prescriptive it provides guidance on a way forward to eat in a way that is healthy for people and for the planet. Current diets are not aligned to the "Planetary Health Diet," with most regions overconsuming animal source foods and starchy vegetables while under consuming fruits, vegetables, legumes, nuts and whole grains (Willett et al., 2019).

While the "Planetary Health Diet" takes into consideration the health and environmental principles of sustainable diets, it fails to sufficiently account for social and economic aspects of diets. An analysis of the affordability of the diet found that it would be affordable for the average household in high-income countries (Hirvonen et al., 2020). However, it would exceed household per capita income for 1.58 billion people globally (Hirvonen et al., 2020), highlighting the need for sustainable diets to be assessed in a more comprehensive way.

Measuring sustainable diets

Given the complexity of sustainable diets, and the breadth of elements that are included within its definition, identifying ways to measure them has been challenging. Most of the studies measuring sustainable diets that have been conducted to date have focused on the greenhouse gas emissions associated with dietary patterns (Jones et al., 2016; Eme et al., 2019). **Fig. 5** summarizes the ways in which sustainable diets have been measured in the existing literature. In a systematic review conducted in 2016, 73% of papers focused on environmental outcomes, with less than a third focused on social or economic outcomes (Jones et al., 2016). A review conducted in 2019 found a similar pattern where the majority of studies focused on human and/or health outcomes, with fewer than 15% focusing on social and economic outcomes (Eme et al., 2019). In addition, most studies have focused on high income countries (Jones et al., 2016; Eme et al., 2019).

While studies examining both health and environmental outcomes can help in terms of assessing the environmental sustainability of our diets, social and economic considerations are important components of food choice. Without comprehensive

EAT Reference Diet: Scientific targets for Planetary Health Diet

	Macronutrient intake grams per day (possible range)	Caloric intake kcal per day*
Whole grains (rice, wheat, corn and other)	232	811
Tubers or starchy vegetables (Potatoes and cassava)	50 (0-100)	39
Vegetables (all vegetables)	300 (200-600)	78
Fruits (all fruits)	200 (100-300)	126
Dairy foods (Whole milk or equivalents)	250 (0-500)	153
Protein sources Beef, lamb and pork Chicken & other poultry Eggs Fish Legumes Nuts	14 (0-28) 29 (0-58) 13 (0-25) 28 (0-100) 75 (0-100) 50 (0-75)	30 62 19 40 284 291
Added fats Unsaturated oils Saturated oils	40 (20-80) 11.8 (0-11.8)	354 96
Added sugars (all sugars)	31 (0-31)	120

*Based on 2500 kcal/day diet



Fig. 4 An overview of the EAT Lancet reference "Planetary Health Diet." Source: EAT Lancet Summary Report. Reference Diet table has been reproduced.

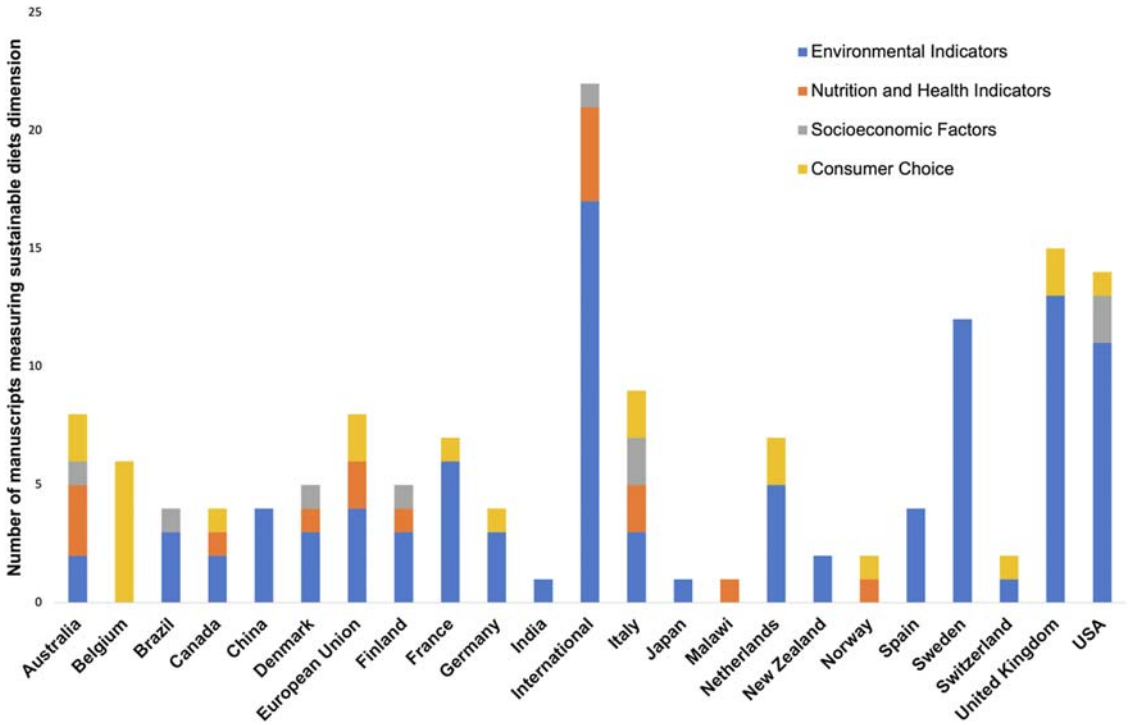


Fig. 5 An overview of the ways in which sustainable diets have been measured in the literature.

methodological approaches that include the social and economic dimensions of sustainable diets it will be difficult to weigh the trade-offs associated with the different dietary dimensions.

There are several gaps that remain regarding how to best measure sustainable diets in practice. Given the breadth of sustainable diets principles, an index that provides an overall composite score across the health, environmental, social and economic dimensions of sustainable diets could prove valuable. However, these types of indicators currently do not exist. Moreover, few studies have examined how to measure sustainable diets at the local level using comprehensive approaches that cut across the multiple dimensions of sustainable diets. However, the city of Milan has developed a monitoring framework that includes a myriad of indicators focused on enabling effective action (governance), sustainable diets and nutrition, social and economic equity, food production, food supply and distribution and food waste to monitor progress toward attaining its urban food policy pact. This monitoring framework provides a useful starting place for developing comprehensive assessments of local food and sustainability initiatives.

Shifting consumption toward sustainable diets: what are the challenges?

As with the difficulties in measuring sustainable diets, there are also challenges related to promoting their consumption among consumers. While there are many strategies that can be adopted across the food system to reorient food systems toward more sustainable production and consumption, dietary change will also be necessary. Consumers will need to make significant changes to the way they procure, prepare and consume food in order to shift diets toward sustainable consumption patterns, such as the "Planetary Health Diet." Promoting sustained shifts to diets is difficult. Even when there are direct benefits to personal health, individuals have a difficult time instituting long term lifestyle change. People have a difficult time evaluating and acting on trade-offs between present and future costs and benefits. Whereas most benefits of unhealthy options are certain and immediate, many potential costs are uncertain and far in the future.

Tackling embedded social and cultural norms within the everyday eating practices of individuals presents a plethora of challenges. Consumption is largely framed by the norms and social expectations of an individual's various perceived in-groups: religious, ethnic, social, etc. (Vermeulen et al., 2020). Foods and meals also play multiple important social functions, often serving as a central point of human bonding, cultural preservation and bridging social cohesion. Adapting to new dietary norms often presents a conflict to longstanding foodways and cultural norms. For example, in many societies the consumption of animal products is deeply entrenched in socio-cultural practices and has long been equated with affluence, success and stability (Klaudia and Wojciech, 2018). This has been highlighted in recent years by the way that the global growth of the middle class, and overall increases in incomes, has driven significant increases in meat consumption as well as highly processed foods. While the importance of mainstreaming more plant-based diets has been identified by experts as a key step toward sustainability, the deeper psychological underpinnings that influence food choice need additional consideration (Klaudia and Wojciech, 2018).

Decisions about which foods to consume are influenced not only by socio-cultural and psychological factors but also the food environments that we interface with. Individual capacity to access healthy diets from sustainable food systems differs drastically across the globe based on geographic, political, environmental, and economic contexts.

Even where healthy foods produced within sustainable food systems are physically available, ensuring equitable access presents an additional challenge due to economic constraints. Along with taste, cost is one of the most important factors influencing food choice. While plant-based foods are often less expensive than animal products, consumers often pay a premium for value added foods (e.g., local, organic, animal welfare considerations, etc.). Local foods are one example of "value-added" foods (see **Box 1**). Given that many low-income populations are already spending a higher proportion of their income on food, and their food choices are already economically constrained, the price premium of foods that are produced in a more sustainable way may be prohibitive.

Box 1 Local food: its definition and sustainability characteristics

"Local food" is often considered as a geographic concept, referring to the distance from production to consumption. However, the definition of "local" varies by country and state. In 2008 the Food, Conservation, and Energy Act in the United States required that a locally or regionally produced agricultural food product must travel less than 400 miles from its origin, or within the State in which it is produced. Meanwhile, a growing number of people globally (especially in the USA and European Union) consider local to denote a closer geographic range: 100 miles. The French Ministry of Agriculture officially defines "short circuit" food as coming from within 150 km but sees it in the form of marketing which is done either by direct sales from producers to consumers or through indirect sales through no more than one intermediary between the operator and the consumer. In an example of even stricter guidelines, the state of Vermont legally defines local food as food produced within 40 miles. Currently, many countries do not have established guidelines for defining local foods, and the concept has been more prominent in the global north. Population density also strongly influences perceptions of local: individuals in dense urban centers have a different perception of distances needed to travel to access foods than those in sparsely populated and/or rural areas. Additionally, when considering a diet, the question of the quantity of food which must or should be local also arises.

However, "local" and "sustainable" cannot be treated as equivalent. Local doesn't guarantee minimal environmental impact beyond transportation distance—which is not the only factor, nor is it one of the major factors, contributing to a food's carbon footprint. Furthermore, short distance food supply chains don't guarantee food security or social sustainability, as they are not always complex or dense enough to guarantee availability, or affordability (Stein and Santini, 2021).

At the crux of the issue related to equity and sustainable diets is that those with less income have more difficulty consuming foods which meet more stringent quality and environmental standards (organic, local or fresh vs. those from further fields and from increasingly industrial production methods). Moreover, consumers are often not paying the true cost of food—many of the externalities of food production such as environmental degradation or the health system costs associated with the food we consume are not reflected in the prices paid by consumers.

Consuming a sustainable diet, and promoting policies that support them, is further complicated by competing interests. The vested interests of agricultural organizations and commercial food companies are often in direct conflict with health and sustainability recommendations. Furthermore, in some countries such as the United States, the agencies providing dietary guidelines are the same agencies charged with supporting farmers and fostering stronger markets. This can create tensions in relation to advocating that populations reduce their meat consumption or for including sustainability considerations in dietary guidelines (Ahmed et al., 2019). Policy coherence across the food system as a whole is needed in order to ensure better alignment between food system incentives and dietary recommendations.

Solutions to help support sustainable diets

The most effective strategies to shift diets toward sustainable diets will involve multiple approaches that deliberately aim to influence consumers as well as to incentivize all actors in the food systems to produce foods in a better way, while taking into account multiple agendas and values (Vermeulen et al., 2020). Both consumer-facing strategies as well as those targeting the more upstream drivers of food production and consumption will be necessary.

Three main categories of consumer-facing interventions to promote sustainable diets were previously identified, which include: (1) informing and empowering, (2) guiding and influencing, and (3) incentivizing, discouraging or restricting (Bailey and Harper, 2015) (see Fig. 6).

Consumer-facing solutions

Inform and empower

Providing consumers with information about the sustainability of the foods they have access to, such as environmental labeling, can enable them to make more informed food choices (Bailey and Harper, 2015). Interventions that aim to inform and empower are the least intrusive strategies aimed at shifting consumers toward sustainable diet (Bailey and Harper, 2015). Two key ways in which policies and interventions have been used to inform and empower consumers to make more sustainable food choices is through the inclusion of sustainability in food based dietary guidelines and environmental food labeling.

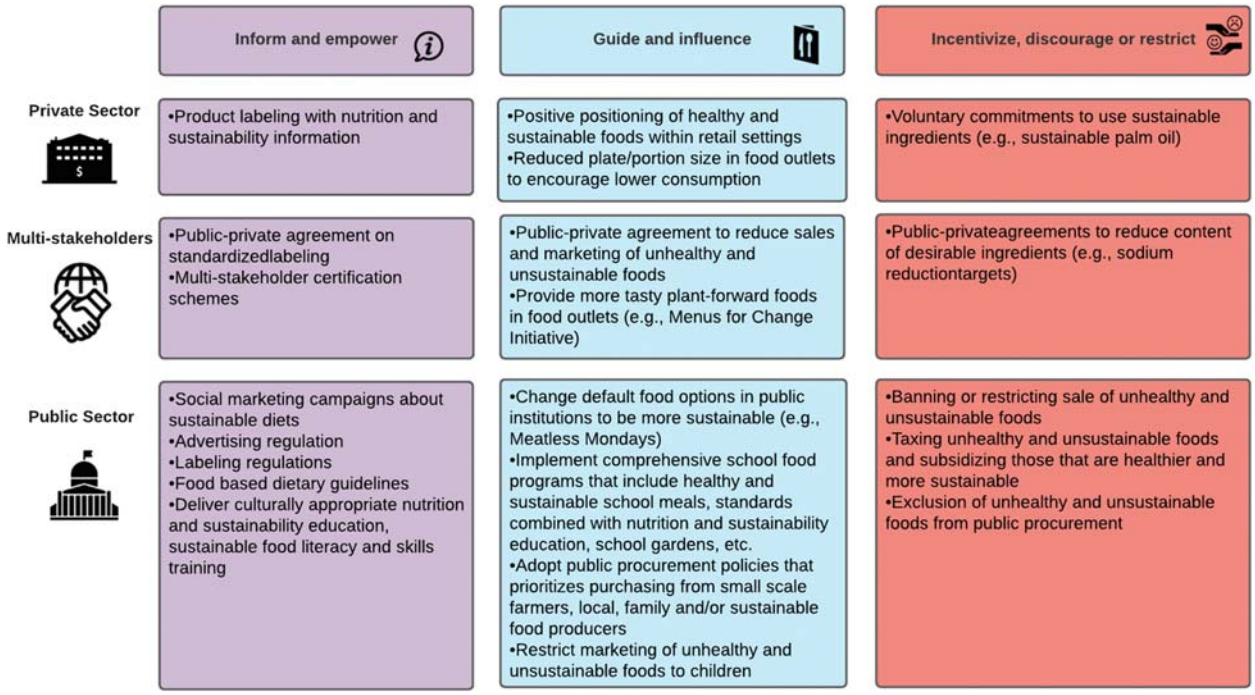


Fig. 6 An overview of the points for intervening to shift consumption toward sustainable diets. Adapted from Bailey and Harper (2015).

- *Food based dietary guidelines*

One of the ways in which sustainable diets have been incorporated into government policies to date are through dietary guidelines. The FAO/WHO have released guiding principles for sustainable healthy diets (see **Fig. 1**) "to be further translated into clear, non-technical information and messaging to be used by governments and other actors in policy-making and communications" (FAO/WHO, 2019).

While guiding principles for sustainable healthy diets have been developed at the global level, individual countries develop their own food based dietary guidelines. Only a few of the nearly 100 countries globally with dietary guidelines include aspects of sustainability within them (Ahmed et al., 2019). **Fig. 7** summarizes the sustainability considerations included in the dietary guidelines of countries globally (Rose et al., 2019). One of the most consistent messages across the dietary guidelines of countries that have included sustainability considerations is to consume a plant-based diet (Rose et al., 2019).

Given that dietary guidelines can have important implications for the implementation of government programs (e.g., school lunch programs), including sustainability considerations within them has the potential to make a significant impact on diets and environmental outcomes. Shifting toward adopting national food based dietary guidelines in their current state would lead to a reduction in environmental resource demand, including greenhouse gas emissions, by an average of 13% (Springmann et al., 2020). However, most dietary guidelines are incompatible with environmental targets in their current form (Springmann et al., 2020) and by including sustainability considerations they could have an even greater impact on environmental outcomes. For example, adopting the "Planetary Health Diet" would lead to a more than three times greater decline in greenhouse gas emissions and better attainment of environmental targets as compared to adopting current dietary guidelines (Springmann et al., 2020).

While it is clear that incorporating sustainability in dietary guidelines, particularly providing advice on limiting beef and dairy consumption, is important from an environmental perspective, it is often politically contentious. In the United States there was significant political backlash after the 2015–2020 Dietary Guidelines Advisory Committee included sustainability considerations in their report, leading to removal of sustainability considerations from the final guidelines. However, it may be possible to include aspects of sustainable diets in food based dietary guidelines without explicitly referring to sustainability in an effort to limit political opposition. A study examining the inclusion of ecological, economic, social and health dimensions of sustainable diets in food based dietary guidelines found that some countries that had not explicitly framed their dietary guidelines as including sustainability considerations still addressed many aspects of sustainable diets within their guidelines (Ahmed et al., 2019). For example, Brazil's dietary guidelines covered the widest range of dimensions of sustainable diets without explicitly referring to sustainability within them (Ahmed et al., 2019).

Given the global need to shift food choices toward more plant-based diets, there is a clear need for dietary guidelines to begin to include aspects of sustainable diets, whether explicit or not, within them. This then has the potential to influence public and institutional procurement policies providing incentives for changes to the food supply as well as food environments.

- *Environmental food labels*

Food labels have the potential to influence food choices; however, the evidence of health and nutrition labels leading to changes in purchasing behavior has been mixed (Hawkes et al., 2015). Labels appear to fill an information gap for people who already

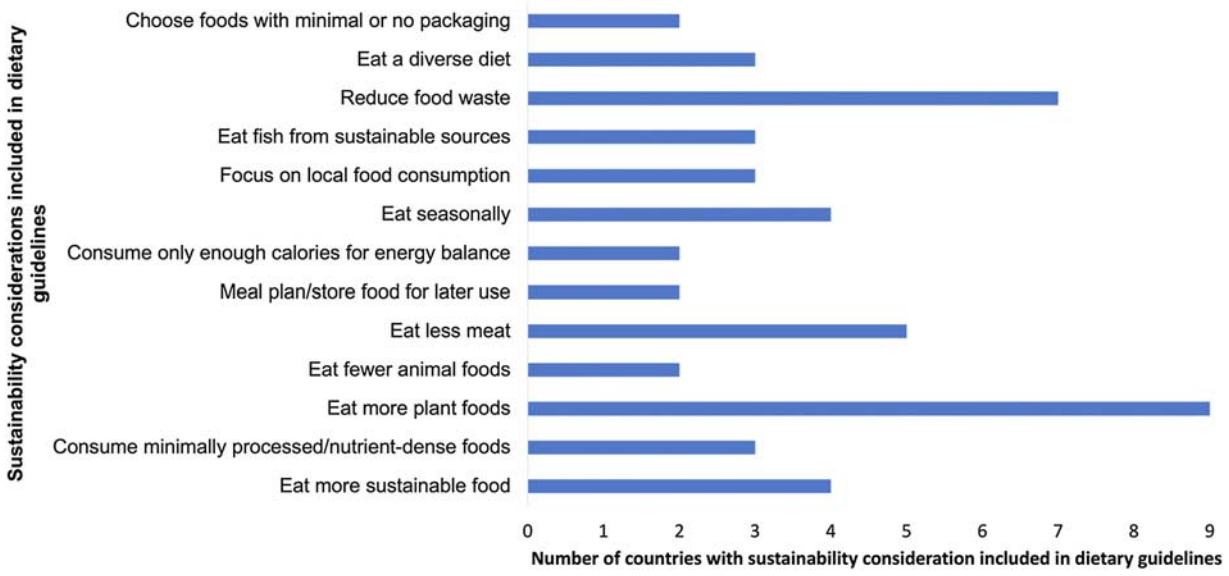


Fig. 7 Sustainability considerations in dietary recommendations in various countries. Adapted from Rose et al. (2019).

have healthy preferences but are limited by the lack of point-of-purchase information to make choices aligned with their preferences (Hawkes et al., 2015). Environmental sustainability or eco labeling has recently become more popular. However, unlike nutritional labeling, most environmental labeling schemes are certification initiatives rather than providing information about the environmental impacts of particular foods (Bailey and Harper, 2015). These labels assure consumers that the food meets specific minimum sustainability standards. While high quality studies are needed to evaluate the impacts of environmental labeling, there is some evidence to suggest that these approaches can influence consumer food choices (Potter et al., 2021). However, there are gaps in the current evidence base related to the types of labels that have the greatest impact on consumer food choices.

Guide and influence

Interventions that nudge or provide cues to consumers to consume specific foods could be employed to increase consumption of sustainable diets. Setting defaults as the healthier option have shown some success in terms of improving the nutritional quality of foods purchased. Institutions such as schools, hospitals and government buildings can be ideal settings for these types of interventions. Meatless Mondays and the Menus of Change, which has developed principles of healthy, sustainable menus, are other examples of ways in which consumers can be nudged toward more sustainable food choices.

Incentivize, discourage or restrict

Interventions that incentivize, discourage or restrict can help incentivize consumption of foods that are more sustainable, while discouraging those that are not. There is strong evidence to suggest that the use of taxes and subsidies can change consumer purchasing behavior (Thow et al., 2018). Overall, the strongest evidence relates to taxes on sugar-sweetened beverages and subsidies on fruits and vegetables (Thow et al., 2018). Given that beef contributes disproportionately to the environmental footprint of food production and consumption, adopting taxes aimed at reducing its consumption has the potential to improve environmental outcomes. While there is little evidence of the impact of fiscal policies targeting beef consumption, modeling studies have found that taxing red meat would lead to benefits from a health and environmental perspective (Springmann et al., 2018). Denmark is one of the only countries to date to have implemented a tax that would apply to red meat. In 2011, they implemented a tax on saturated fat that applied to foods high in saturated fat, including red meat (Jensen et al., 2015). While the tax was rescinded shortly after its implementation due to political backlash, it led to a 13–16% increase in the price of high fat varieties of minced beef and cream products which yielded a 4–6% decrease in intake of saturated fat from minced beef and regular cream (Jensen et al., 2015). While these findings imply that fiscal policy could be used to incentivize consumption of sustainable diets, there are several limitations of its use.

A sustainable diet includes a variety of different foods which makes it difficult to identify which foods should be targeted in fiscal policy approaches. While there have been studies that have modeled the effects of taxes across all foods based on GHG emissions (Latka et al., 2021), this type of approach may not be politically feasible. Moreover, to reach nutrition and environmental sustainability targets, considerably high tax levels are required (Latka et al., 2021). Sugar sweetened beverage taxes have led to shifts in consumption at relatively low levels (~10%), whereas modeling studies examining taxes on red meat have suggested that taxes would need to be upwards of 100% or higher (Latka et al., 2021; Springmann et al., 2018).

Another challenge associated with the use of taxation to incentivize the consumption of sustainable diets is that these taxes tend to be regressive (Thow et al., 2018). Red meat contains many key nutrients (e.g., iron, zinc, etc.), as do many other animal source foods. Moreover, consuming animal source foods could help to reduce stunting and other forms of malnutrition among populations living in low resource settings globally. There are thus ethical implications related to increasing the cost of accessing these key nutrients by low-income populations. This creates challenges in terms of the way in which taxes aimed at increasing consumption of sustainable diets are framed. Coupling taxes with subsidies on other nutrient-rich foods (e.g., fruits and vegetables) could help to at least partially address the regressivity of these types of taxes.

Supply-side solutions

The aforementioned consumer-facing strategies aimed at supporting sustainable diets need to be complemented with interventions throughout the food supply chain in order to make foods that comprise a sustainable diet more available, affordable and appealing. Innovative technologies will play a critical role in terms of reorienting food production, and the way it moves through supply chains, toward more sustainable practices. While many of the production practices that are currently used by subsistence farmers are more sustainable than industrial farming practices, such as increased diversity on farms, use of organic fertilizer, etc., there is a large yield gap. Introducing improved farming practices and increasing the adoption of technology and higher quality seed could help to address these gaps, increasing the sustainability of these production systems. Government incentives (including subsidies in some cases), investment in R&D and strengthening of extension services could also help to identify and implement farming, fishing and rearing practices that reduce the environmental impact of food production, including through improved feed, increased diversification of farms, the adoption of circular economy, etc. Improvements in food processing, packaging (i.e., biodegradable, fewer plastics, etc.), and distribution (i.e., use of electric vehicles) will also help increase the sustainability of foods that are available within the food environments that consumers interface with.

Advancing the sustainable diets agenda: the role of different stakeholder groups

There are various stakeholders that can help to support healthy diets from sustainable food systems. While UN agencies, government and the private sector have an important role to play, we highlight the particular influence of local government and social movements in this section.

The role of local government

Local governments are in a unique position to adopt innovative policies that support sustainable diets. There are many examples of cities around the globe that are working toward supporting more sustainable food systems in order to mitigate the negative environmental, health, and socioeconomic outcomes associated with current practices. Perhaps most notably, the mayor of Milan, Italy created an international agreement, Milan Urban Food Policy Pact (MUFPP), for cities and urban areas to adopt an action framework for food policy to measure its impact on diet quality and accessibility, as well as its impact on the environment and food supply chain (https://www.milanurbanfoodpolicypact.org). As of April 2021, 211 cities around the world had committed to adopting food policies to tackle at least one of the six areas of focus: governance, sustainable diets and nutrition, social and economic equity, food production, food supply and distribution, and/or food waste. While cities from all continents have signed on to the MUFPP, those who have already planned and activated a food strategy are concentrated in Europe, North- and Latin America. Two-hundred and eleven cities around the world have committed to adopting food policy to tackle at least one of the six areas of focus. While cities from all continents have signed on to the MUFPP, those who have already planned and activated a food strategy are concentrated in Europe, North- and Latin America.

While many of the examples of local government approaches to supporting sustainable diets and food systems are from high-income countries (see **Box 2**), there are also examples from low- and middle-income countries. Arusha, Tanzania joined the MUFP in 2015 and in 2018 launched its first food policy initiatives. The city has intentions of matching rapid urbanization with environmental protection policies in order to promote food safety while mitigating inefficient agriculture practices such as inappropriate use of crop protection products, like overuse of pesticides. By partnering with the MUFPP, partnerships with various NGOs have fostered the establishment of over 200 gardens and a wet market in order to promote biodiversity and food access, demonstrating the potential for cities around the world to put sustainable diets on the agenda and make a real impact on their communities.

The role of social movements and role models

Consumer organizations, grassroots movements and non-governmental organizations have an important role to play in promoting sustainable diets. Given the necessity of societal buy-in and bottom-up action to achieve dietary shifts, there is a strong potential role for social movements and local organizing to influence behavior and drive social change toward sustainable diets. Past environmental movements have relied heavily on the power of social movements to carry their messages, and to drive broad spread action, policy change and reframe social norms. In 2019, the UN Secretary-General directly named "people action"—including by youth, civil society, the media, the private sector, unions, academia and other stakeholders, to generate an unstoppable movement pushing for the required transformations—as one of the three necessary levels to accelerate sustainable solutions to key global issues.

Box 2 Examples of approaches to support sustainable diets and food systems by local governments

New York City (NYC) is one local municipality where multiple strategies have been underway to encourage sustainable practices, including: animal welfare practices and waste management, dietary recommendations and their relation to sustainability, reducing air pollution and greenhouse gas emissions from food transport and storage, promoting innovation in food and sustainability research and public procurement guidelines stipulating the inclusion of local and state grown produce. Additionally, since 2011 NYC releases annual Food Metrics reports to track and realign progress toward its food and sustainability goals. In February 2021, the city released a 10-year food policy plan that refines and expands the previous 2011 plan and includes a school education program focused on sustainability, as well as encouraging private sector partnerships with community stakeholders to promote sustainable food campaigns. Additional information can be found here: https://www1.nyc.gov/site/foodpolicy/index.page.

Similarly, the city of Toronto, Canada launched the Toronto Food Strategy in 2010 to continually develop innovations to promote resident and environmental health through their food system. The plan defines a sustainable diet as one that is able to provide people nutrients without negatively impacting the environment and ensuring equity. Additionally, Toronto has adapted a food system monitoring framework in order to keep track of their sustainability goals over time. The Toronto Food Strategy is unique in that it emphasizes a multifactorial approach toward sustainability goals focusing on not only ecological/environmental and health impacts, but also promoting social justice as a necessity to meet those goals. Additional information can be found here: https://www.toronto.ca/community-people/health-wellness-care/health-programs-advice/toronto-food-strategy/background/.

Birmingham, UK partnered with Pune, India to join the MUFP in 2016 as "Food Smart Cities" to utilize data and invest in innovation in order to make the food system more sustainable. The partnership between the two cities aims to develop policies that encourage the consumption of safe, healthy, and sustainable foods. By having similar approaches to issues such as COVID-19 and malnutrition, the cities are able to work together with compiled data to form their best mitigation approaches moving forward. Additional information can be found here: https://commonwealthsustainablecities.org/europe/nutrition-smart-cities-bindi-birmingham-india-nutrition-initiative-in-birmingham-and-pune/.

Social pressure is one of the strongest motivators when it comes to behaviors related to sustainability and the environment. For example, highlighting the increasing normality of eating less meat, telling people that most other people recycle, or use less energy (Vermeulen et al., 2020) have all been shown to have impacts on individual behavior when communicated by peers. Grassroots movements and social campaigns therefore present strong inroads to drive social change.

Starting in the 1980s, the Slow Food movement began modeling how food, and the choices made within and about the food system, can be used to protect communities, preserve culture and biodiversity and protect the environment. In subsequent decades we have seen the growth of these global social campaigns focused on changes to diet and food choices as drivers of human and environmental health. Increasingly, these movements, such as the Food Justice Movement, have focused on racial and economic disparities.

The resurgence of the urban agricultural and community garden movements in the global North presents an opportunity to promote personal and community food growing, and to adopt more neighborhood centric models of food production and distribution. The local food growing movement and its related platforms create strong entry points to conversations about changing dietary needs as well as a deeper understanding and interest in sustainable food production. Importantly, it also serves to normalize the idea and acceptability of urban food growing and engages people in a stronger understanding about where food comes from.

Food, Beverage and Hospitality have taken on a new status in the past decade with the exponential expansion of food media and content across all platforms (from socio-cultural cooking podcasts like "Good Foods" to food adventure travel like Anthony Bourdain's Parts Unknown, and Chef and Server Memoirs ala Bittersweet). This has increased the visibility and opportunities for chefs, restaurants and companies to inform and influence opinions and habits connected to food systems. Increasingly, celebrity chefs are highlighting the importance of sustainability, from NYC Blue Hill (NYC) and NOMA (Copenhagen) restaurants ongoing focus on regional, seasonal foods and sustainable sourcing to bolster local supply chains, to Eleven Madison Park's (NYC) post pandemic rebirth as a vegan based fine dining experience. Chefs are centering their menus on more sustainable models and introducing these ideas to consumers. We have seen the growth of the "farm-to-table" and "eat local" movements, which has increased the number of restaurants focusing on local, sustainable and seasonal foods, including some fast casual chains (e.g., Digg In and Sweetgreen in the United States). There are also encouraging opportunities to model behavior and drive change being brought about through corporate socio-environmental responsibility and a focus on more environmentally friendly and healthy foods. Examples of this include: Epicurious' move to no longer publishing new meat recipes; Burger King's launch of vegan "chicken" sandwich in the UK and meatless Impossible burgers throughout the US and UK; Chipotle's ongoing drive to increase its local and organic ingredients, etc. While these examples are mostly from the United States, this trend toward more environmental consciousness is also being seen in other parts of the world. However, it is important to note that much of the messaging around sustainability continues to target energy-dense foods of low nutritional value leading to concerns related to "greenwashing."

Research gaps and recommendations for moving toward sustainable diets

There are many gaps in the way in which sustainable diets are measured, particularly as it relates to the ways in which trade-offs among different dimensions of sustainable diets are considered. Moving forward, additional research is needed to provide guidance on the most appropriate ways to measure the multiple dimensions of sustainable diets, including how to weigh and balance trade-offs. Sustainability considerations have largely been missing from the food environment literature but given that it is the place within the food system that consumers interact with to make decisions about the foods to acquire and purchase, improved understanding regarding how best to measure sustainability properties within the food environment will become increasingly critical. Identifying a suite of indicators that could be used to measure the different dimensions of sustainable diets, at different scales (including national, regional and local level) would help to provide researchers and practitioners with the necessary measurement tools to comprehensively assess sustainable diets in different contexts as well as how they change over time.

Another key gap in the current evidence base that merits additional research relates to the most effective ways to influence populations to consume sustainable diets. Of the interventions aimed at increasing consumption of sustainable diets that have been conducted to date, few have included strong evaluations of their impact and implementation (Bailey and Harper, 2015). The majority of studies that have examined the impact of interventions aimed at increasing plant-based diets have focused on fruits and vegetables, which limits our understanding of the impact of interventions aimed at reducing consumption of animal source foods with high environmental footprints (Taufik et al., 2019). Changing consumer food choices is difficult. It is therefore imperative that consumer-facing interventions aimed at increasing demand for sustainable diets are combined with supply-side policies and interventions aimed at increasing the availability, affordability and appeal of sustainable healthy foods.

Summary

In summary, shifting global diets toward more plant-based foods (e.g., Planetary Health Diet) would lead to considerable reductions in the environmental impact of our diets. However, interventions aimed at shifting consumption toward plant-based diets need to also consider the economic and social considerations that influence their consumption. Interventions across the food system as a whole will be needed to incentivize food system actors and consumers to produce and consume a sustainable diet. While there is some evidence to suggest potential approaches to support and promote sustainable diets, significant research gaps remain.

Future research should focus on identifying better ways to measure sustainable diets in a comprehensive way and to develop innovative solutions targeting various food system actors to reorient food production and consumption toward more sustainable practices.

References

Afshin, A., et al., 2019. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 393 (10184), 1958–1972.

Ahmed, S., Downs, S., Fanzo, J., 2019. Advancing an integrative framework to evaluate sustainability in national dietary guidelines. Front. Sustain. Food Syst. 3, 76.

Bailey, R., Harper, D.R., 2015. Reviewing Interventions for Healthy and Sustainable Diets. Chatham House, The Royal Institute of International Affairs.

Bell, W., Lividini, K., Masters, W.A., 2021. Global dietary convergence from 1970 to 2010 altered inequality in agriculture, nutrition and health. Nat. Food 2 (3), 156–165.

Eme, P.E., Douwes, J., Kim, N., Foliaki, S., Burlingame, B., 2019. Review of methodologies for assessing sustainable diets and potential for development of harmonised indicators. Int. J. Environ. Res. Publ. Health 16 (7), 1184.

FAO and WHO, 2019. Sustainable Healthy Diets—Guiding Principles. Rome.

FAO, IFAD, UNICEF, WFP and WHO, 2021. The State of Food Security and Nutrition in the World 2021. Transforming Food Systems for Food Security, Improved Nutrition and Affordable Healthy Diets for All. FAO, Rome. https://doi.org/10.4060/cb4474en.

Food and Agriculture Organization of the United Nations, 2012. Sustainable Diets and Biodiversity: Directions and Solutions for Policy, Research and Action. FAO and Bioversity International, Rome, Italy.

Gussow, J.D., Clancy, K.L., 1986. Dietary guidelines for sustainability. J. Nutr. Educ. 18 (1).

Hawkes, C., Smith, T.G., Jewell, J., Wardle, J., Hammond, R.A., Friel, S., Thow, A.M., Kain, J., 2015. Smart food policies for obesity prevention. Lancet 385, 2410–2421.

Hirvonen, K., Bai, Y., Headey, D., Masters, W.A., 2020. Affordability of the EAT–Lancet reference diet: a global analysis. Lancet Glob. Health 8 (1), e59–e66.

Jensen, J.D., Smed, S., Aarup, L., Nielsen, E., 2015. Effects of the Danish saturated fat tax on the demand for meat and dairy products. Publ. Health Nutr. 19, 3085–3094.

Jones, A.D., Hoey, L., Blesh, J., Miller, L., Green, A., Shapiro, L.F., 2016. A systematic review of the measurement of sustainable diets. Adv. Nutr. 7 (4), 641–664.

Klaudia, M., Wojciech, P., 2018. Selected psychological aspects of meat consumption—a short review. Nutrients 10 (9), 1301.

Latka, C., Kuiper, M., Frank, S., Heckelei, T., Havlík, P., Witzke, H.-P., Leip, A., Cui, H.D., Kuijsten, A., Geleijnse, J.M., van Dijk, M., 2021. Paying the price for environmentally sustainable and healthy EU diets. Glob. Food Secur. 28, 100437.

Potter, C., Bastounis, A., Hartmann-Boyce, J., Stewart, C., Frie, K., Tudor, K., Bianchi, F., Cartwright, E., Cook, B., Rayner, M., Jebb, S.A., 2021. The effects of environmental sustainability labels on selection, purchase, and consumption of food and drink products: a systematic review. Environ. Behav. 53 (11), 001391652199547.

Rose, D., Heller, M.C., Roberto, C.A., 2019. In: Position of the Society for Nutrition Education and Behavior: the Importance of Including Environmental Sustainability in Dietary Guidance, vol. 51, pp. 3–15. https://doi.org/10.1016/j.jneb.2018.07.006.

Springmann, M., Mason-D'Croz, D., Robinson, S., Wiebe, K., Godfray, H.C., Rayner, M., Scarborough, P., 2018. Health-motivated taxes on red and processed meat: a modelling study on optimal tax levels and associated health impacts. PLoS One 13.

Springmann, M., Spajic, L., Clark, M.A., Poore, J., Herforth, A., Webb, P., Rayner, M., Scarborough, P., 2020. The healthiness and sustainability of national and global food based dietary guidelines: modelling study. BMJ m2322.

Stein, A.J., Santini, F., 2021. The sustainability of "local" food: a review for policy-makers. Rev. Agric. Food Environ. Stud. 1–13.

Taufik, D., Verain, M.C.D., Bouwman, E.P., Reinders, M.J., 2019. Determinants of real-life behavioural interventions to stimulate more plant-based and less animal-based diets: a systematic review. Trends Food Sci. Technol. 93, 281–303.

Thow, A.M., Downs, S.M., Mayes, C., Trevena, H., Waqanivalu, T., Cawley, J., 2018. Fiscal policy to improve diets and prevent noncommunicable diseases: from recommendations to action. Bull. World Health Organ. 96, 201–210.

United Nations, 2015. The 17 GOALS: Sustainable Development. https://sdgs.un.org/goals. (Accessed 14 March 2021).

Vermeulen, S.J., Park, T., Khoury, C.K., Béné, C., 2020. Changing diets and the transformation of the global food system. Ann. N. Y. Acad. Sci. 1478, 3–17. https://doi.org/10.1111/nyas.14446.

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F., Wood, A., Jonell, M., 2019. Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. Lancet 393 (10170), 447–492.

Sweeteners: Sensory properties, digestion, consumption trends, and health effects

Richard D. Mattes[a], **Sze-Yen Tan**[b], **and Robin M. Tucker**[c], [a] RDM Department of Nutrition Science, Department of Public Health, Purdue University, West Lafayette, IN, United States; [b] Institute for Physical Activity and Nutrition (IPAN), School of Exercise and Nutrition Sciences, Deakin University, Burwood, VIC, Australia; and [c] Department of Food Science and Human Nutrition, Michigan State University, East Lansing, MI, United States

© 2023 Elsevier Ltd. All rights reserved.

Introduction 625
Sweetener taste, digestion, and appeal 625
Receptor 625
Toxicology 627
Digestion and absorption 627
Sugar 627
Low calorie sweeteners 628
Gut signaling 628
Microbiome 628
Sugar 629
Low calorie sweeteners 629
Sweet hedonics 629
Sweeteners and appetite 629
The biological basis for sweetener intake 629
Consumption trends 630
Sugar 630
Low calorie sweeteners 631
Sweeteners and health 632
Sugar 632
Overweight and obesity 632
Cardiovascular disease (CVD) 633
Type 2 diabetes mellitus (T2DM) 633
Cancer 633
Dental caries 634
Low calorie sweeteners 634
Overweight and obesity 634
Type 2 diabetes 634
Cancer 635
Conclusion/summary/outlook 635
Conflict of interest 635
References 635

Key points

- There is one primary sweet taste receptor that is activated by both nutritive sweeteners and low calorie sweeteners (LCS).
- Sweeteners vary in the intensity of the sweetness they impart and LCS also elicit various side tastes, LCS are not sweeter than sucrose, they just achieve this sweetness intensity at low concentrations.
- Sweeteners vary in their digestion, microbial fermentation, and peripheral distribution with uncertain metabolic and health implications.
- Sweeteners are inherently pleasant, but their appeal and consumption are determined more by dietary experience and culture.
- While multiple health organizations have recommended that added/free sugar intake should not exceed 10% of energy intake, data suggest that many people fail to meet these recommendations.
- Use of LCS has grown globally and is projected to continue; there is considerable consumer interest in "natural" LCS compared to synthetic LCS.
- Several policy approaches have shown promise in reducing added sugar consumption.
- Sugar intake is associated with the risk of overweight, obesity, cardiovascular disease (CVD), type 2 diabetes mellitus (T2DM) (in the form of sweetened beverages), and dental caries. The association with cancer risk is inconclusive.

 https://doi.org/10.1016/B978-0-12-821848-8.00014-7

- Although LCS intake was positively associated with BMI in some prospective cohort studies, interventional studies consistently reported greater weight reduction with LCS use. There is a lack of association between LCS and CVD, T2DM, and cancer.

Introduction

Sweeteners comprise a wide variety of substances that produce a sensation of sweetness when consumed. While individuals will differ on their preferred concentration, sweetness is considered to be hedonically pleasing, and sweet foods are readily consumed. Given the associations between dietary intake and chronic disease, sweeteners have come under scrutiny for their role in conditions including obesity, diabetes, and cancer. This chapter discusses sweet taste biology, trends in both nutritive and low calorie sweetener consumption, and the associations between sweetener consumption and disease.

Sweetener taste, digestion, and appeal

Receptor

Sweetness is a sensation imparted by a wide range of chemicals (**Table 1**). Nevertheless, sweetness is transduced (converted from an external chemical signal to an internal electrical signal) primarily by a single sweet taste receptor. This receptor is a heterodimer comprised of the G-protein coupled receptors: taste receptor type 1 receptor 2 (T1R2) and taste receptor type 1 receptor 3 (T1R3). The receptor has three domains, the amino terminal domain (ATD), a cysteine-rich domain (CRD) and a transmembrane domain (TMD). The diverse chemicals that impart sweetness do so by binding to the different domains on each protein. For example, glucose binds primarily to the ATD of T1R3; aspartame, neotame and D-tryptophan bind to the ATD of T1R2; thaumatin and brassein bind to the CRD of T1R3; and cyclamate and neohesperidine dihydrochalcone bind to the TMD of T1R3 (DuBois, 2016). T1R3 alone can also serve as a sweet receptor but has lower sensitivity and primarily binds nutritive sweeteners (Damak et al., 2003). Following binding of a sweetener to a receptor on a taste cell, a cascade of cellular events occurs leading to the release of neurotransmitter resulting in the generation of an electrical signal that is conveyed by afferent nerves first to the nucleus of the solitary tract in the brainstem, next to the thalamus and finally to subcortical and cortical areas. These latter sites are reciprocally innervated with other brain centers allowing for integration of multiple sources of input and formation of a hedonic perception and early indication of the stimuli's safety (Chaudhari and Roper, 2010; Vincis and Fontanini, 2019).

An additional sweetener signaling system, that may not produce a cognitively apparent sensation, has also recently been identified and serves a different but inter-related purpose. Sensory stimulation alone by various sources (e.g., vision, olfaction, taste) elicit neurally-mediated physiological responses that mimic those occurring during food digestion and nutrient absorption and metabolism. These responses are termed first or cephalic-phase responses, and they reportedly prime the body to efficiently process ingested foods. For example, a cephalic phase insulin response (CPIR) is documented in humans to foods generally, but also to nutritive and low-calorie sweeteners (LCS) (see Dhillon et al., 2017). This response helps to modulate post-prandial glycemic responses (Calles-Escandon and Robbins, 1987) and is blunted or absent in patients with diabetes (Davies et al., 1994; Del Prato and Tiengo, 2001). Based on rodent studies, the CPIR may be activated by a transporter mechanism involving K_{ATP} channels. CPIR-activated taste cells responsive to sweeteners also release GLP-1, which acts as an incretin stimulating pancreatic insulin release (Kokrashvili et al., 2014).

Table 1 Sweetener types.

Class of stimuli	*Effective stimuli*	*Sweetener examples*
Carbohydrate	Monosaccharides	Glucose, fructose, galactose
	Disaccharides	Sucrose, lactose, maltose, isomaltulose
	Short chain oligosaccharides	Inulin, maltotriose
	Sugar alcohols	Mannitol, sorbitol, xylitol, lactitol, isomalt, maltitol, erythritol
Protein	Amino acids	Tryptophan, alanine, asparagine, glutamine, histidine, leucine, serine
	Dipeptides	L-aspartyl-L-phenylalanine methyl ester (aspartame), neotame advantame
	Polypeptide	Curculin, monellin, brazzein, miraculin, thaumatin
Miscellaneous	Low calorie sweeteners	Steviol, cyclamic acid, saccharin, asulfam-potassium, alitame, neohesperidine-dihydrochalcone
Heavy metals	Ions	Lead, beryllium

The sensation imparted by sweeteners is, by definition, sweetness, though some sweeteners also activate other taste receptors leading to additional taste qualities, most notably bitterness. For example, saccharin and acesulfame-potassium (Ace-K) bind to the bitter receptors coded by the TAS2R31 and TAS2R43 genes (Kuhn et al., 2004). Such undesirable taste notes can be minimized by reducing the concentration of the sweetener. Mitigating off-tastes is one reason for using blends of sweeteners. Additionally, there is marked individual variability in sensitivity to the bitterness of these sweeteners. At least part of the explanation for this lies with polymorphisms in bitter receptors (Allen et al., 2013). LCS can also stimulate oral, non-taste, receptors leading to metallic, liquorice, and irritancy sensations. Additionally, LCS vary markedly in the temporal pattern of sensations they elicit. Time-intensity profiling indicates the time from stimulus presentation to onset of sensation is: stevia > neotame > aspartame > sucralose, Ace-K. Because of differences in rise and decay times, their duration of sensation follows a pattern of neotame > sucralose and aspartame > stevia > Ace-K (Moraru, 2011; Tan et al., 2019).

The sweetness of LCS are often rank ordered relative to sucrose; examples are provided in **Table 2**. However, it is important to note that sweetness does not grow linearly with concentration. The relationship is better fit by a power function with an exponent of about 1.3 (Moskowitz, 1970). Thus, the actual sweetness relative to sucrose will vary with concentration. Additionally, the sweetness values for LCS are based on the concentration required to achieve a given level of sweetness. They are not actually capable of eliciting a sweetness intensity greater than high concentrations of sucrose. In humans, the curve relating sweetener concentration with sensation intensity asymptotes at about 10% sucrose in water. Introspection leads to the realization that no two sweeteners impart a sensation that differs by a hundred- or thousand-fold as tables often imply. Interestingly, when matched on sweetness intensity, some studies suggest nutritive sweeteners are more potent activators of brain reward centers (Han et al., 2019) and in taste receptor knock-out mice, ingestion of a sucrose solution is maintained over a sucralose solution (De Araujo et al., 2008). However, a recent systematic review failed to confirm robust differences between sweetener types on brain activation (Yeung and Wong, 2020).

Table 2 Relative sweetener taste intensity.

Sweetener	*Sweetness intensity relative to sucrose*
Acesulfam-K	200
Advantame	20,000
Agave syrup	1.56–1.92
Aspartame	180–250
Cyclamate	26–30
Erythritol	0.70
Fructose	1.17–1.75
Galactose	0.35
Glucose	0.74–0.8
Glycyrrhizin	50
High fructose corn syrup	1.2–1.6
Isomalt	0.5
Isomaltulose	0.5
Lactitol	0.4
Lactose	0.16–0.2
Maltitol	0.7
Maltodextrin	0.1
Maltose	0.33–0.5
Mannitol	0.7
Mannose	0.6
Monk fruit (Luo han Guo)	100–250
Neohesperidin dihydrochalcone	340
Neotame	7000–13,000
Saccharin	200–700
Sorbitol	0.6
Steviol glycosides	40–400
Sucralose	600
Sucrose	1
Sugar alcohols	0.25–1
Tagatose	0.8
Thaumatin	200–1000
Trehalose	0.6
Xylitol	1
Xylose	0.5

Toxicology

The safety of LCS has been an ongoing topic of debate. However, the regulatory bodies of many nations have thoroughly reviewed the toxicological evidence and concluded that, when consumed within the acceptable daily intake (ADI) level, they are safe. The ADI is generally calculated as the concentration that is one-hundredth of the no observable adverse effect level. However, the ADI does vary across nations. For example, for Ace-K, it is set at 9 mg/kg body weight by the European Food Safety Agency; 15 mg/kg body weight by the Food and Drug Administration and Joint Commission of Experts on Food Additives of the World Health Organization and the Food and Agriculture Organization; and 40 mg/kg body weight by the Danish Veterinary and Food Administration (Mattes and Popkin, 2009).

Digestion and absorption

The sweet taste receptor is referred to as a "taste" receptor because it was first characterized in the oral cavity and is the initial site for activation of the sense of taste. However, the identical receptor is present throughout the GI track and in a variety of peripheral tissues including bladder, pancreas, thymus, brain, testes, and bone (Behrens and Meyerhof, 2019). In these other locations, the receptor does not mediate a taste sensation. Rather, when activated, it stimulates the functions of the cells on which the receptor is located. Thus, binding of a sweetener to the receptor on an enteroendocrine or pancreatic islet cell leads to release of gut peptides or insulin, respectively. Consequently, the potential physiological effects sweeteners may exert will be, in part, a function of their digestion, absorption, and distribution. If a sweetener is not absorbed from the GI tract, it will not be available to bind to a receptor in the periphery. If it is degraded or fully absorbed in the upper small intestine, it will not be available further downstream as a substrate for the microbiota. There is considerable variability in GI processing even within a class of sweet compounds. For example, among the sugar alcohols, erythritol is largely absorbed in the small intestine whereas isomaltulose is only partially absorbed in the small intestine and lactitol is generally not absorbed in that region. The fate of selected sweeteners is outlined in **Table 3** and further described below:

Sugar

The predominant dietary nutritive sweeteners are the disaccharides, sucrose and lactose. Greater than 90% of these energy sources are absorbed. They are digested to their constituent monosaccharides via enzymatic degradation. This is accomplished by the brush border hydrolases, sucrase and lactase. Sucrose yields glucose and fructose while lactose is reduced to glucose and galactose. Glucose and galactose are transported primarily into jejunal enterocytes by the sodium-dependent hexose transporter, SGLT-1. At high glucose concentrations, the low-affinity, high-capacity transporter, GLUT-2, is also recruited. Fructose enters the enterocyte via the hexose transporter, GLUT-5. All three monosaccharides exit the enterocyte and enter the blood stream by GLUT-2. Based on

Table 3 Sweetener digestion, absorption and metabolism.

Sweetener	*Digestion-absorption-excretion*
Glucose	>90% absorbed by transporter from intestinal lumen and released as glucose by another transported into the circulation. Used by peripheral cells for energy.
Fructose	Absorbed by intestinal transporter with efficiency determined, in part, by dietary exposure. Variable amounts are converted to glucose in the enterocyte and the balance is released into circulation where it serves as an energy substrate.
Galactose	>90% absorbed by transporter from intestinal lumen and released as glucose by another transported into the circulation. Used by peripheral cells for energy.
Sucrose	Digested to glucose and fructose by brush border sucrase. Each monosaccharide is processed as above.
Lactose	Digested to glucose and galactose by brush border lactase. Each monosaccharide is processed as above.
Aspartame	Hydrolyzed by gut esterases and peptidases to ASP, PHE, and methanol.
Sucralose	Minimally digested and absorbed, 90% enters colon.
Saccharin	Not digested; absorbed and ~95% excreted in urine.
Acesulfam-potassium	Not digested; absorbed and 99% excreted in urine.
Stevioside & rebaudioside A	Enters the colon intact. In the colon, sugar moieties are removed and the steviol backbone is absorbed, conjugated with glucouronide, and excreted in urine.
Neohesperidin dihydrochalcone	Minimally absorbed. Fermented in the colon.
Neotame	Equally excreted in urine and feces.
Cyclamate	Small amount metabolized in the colon, mostly eliminated in feces.
Erythritol	90% absorbed in the small intestine and excreted in the urine.
Isomalt	Up to 10% absorbed in the small intestine; fermented in the colon.
Lactitol	Not absorbed in the small intestine; fermented in the colon.
Sorbitol	Partially absorbed in the small intestine; fermented in the colon.
Mannitol	Up to 25% absorbed in small intestine; fermented in the colon.

isotope tracer studies in mice administered low to moderate concentrations of glucose and fructose in a 1:1 ratio (as would occur with sucrose ingestion and be very similar with high fructose corn syrup intake), approximately 42% of this fructose is converted to glucose, glycerate, and organic acids (Jang et al., 2018). Only 14% of ingested fructose enters the blood intact. At higher fructose intakes, greater than 1 g/kg, fructose enters the circulation at higher concentrations, but some is passed downstream where it can serve as a substrate for the microbiota. The extent of this spill-over is based, in part, on customary fructose intake. High intake increases fructose absorption and concentrations in portal blood, but this is reversible with lower intake. Fructose derived from sucrose or high fructose corn syrup is processed similarly (Jang et al., 2018). Whether these recent findings in mice hold quantitatively for humans remains to be determined, but it is likely the trends are conserved across species.

Low calorie sweeteners

Reflecting their different chemical structures, low calorie sweeteners undergo diverse digestive and absorptive processes (Magnuson et al., 2016). Examples are provided below:

Aspartame is hydrolyzed in the upper small intestine to aspartic acid, phenylalanine, and methanol. These by-products, but not the parent compound, are absorbed into the circulation. The amino acids are available for protein synthesis and the methanol is oxidized in the liver. Thus, this sweetener does not enter the circulation so cannot bind to peripheral sweet receptors nor pass into the large intestine and serve as a substrate for the microbiota. Any effects of this sweetener on metabolism and health would have to occur by an alternate mechanism, (e.g., cephalic phase responses).

Ace-K is almost completely absorbed through the small intestine and enters the circulation. Absorption efficiency is not determined by level of exposure. Ace-K is not metabolized and is cleared rapidly, primarily by the kidneys, and excreted in the urine. However, low concentrations have also been detected in breast milk (Sylvetsky et al., 2015). Thus, this sweetener would theoretically have the capacity to serve as a ligand for peripheral sweet receptors.

Saccharin is largely (85%–95%) absorbed through the small intestine and enters the circulation. It is not metabolized and is cleared rapidly, primarily by the kidneys, and excreted in the urine. However, low concentrations have also been detected in breast milk (Sylvetsky et al., 2015). Thus, saccharin would have the capacity to serve as a ligand for peripheral sweet receptors. A small fraction (5%–15%) of saccharin passes into the colon. However, given the very low concentration initially ingested (due to its high potency as a sweetener) and efficient absorption, its concentration in the colon is extremely low. Thus, it is unlikely it serves as a meaningful substrate for the microbiota though a role in gut signaling is possible.

Sucralose is primarily (~90–95%) undigested. Seventy to ninety percent is lost in feces with about 14% excreted in urine. It is highly resistant to microbial fermentation or degradation in various tissues. It has been detected in breast milk (Sylvetsky et al., 2015), so it may also be excreted through lactation. Concentrations in the circulation will be low due to the limitations of its intake by its high potency sweet property and the dilution by blood. Sucralose is highly resistant to microbial metabolism. Frequency of exposure to sucralose does not alter its digestion, absorption or excretion. The small fraction entering the circulation would be available to serve as a ligand for peripheral sweet receptors. It may also exert a physiological effect in the GI tract and periphery via neural signaling.

Steviol Glycosides are digested by bacteria in the colon to steviol. Steviol is resistant to further bacterial digestion. The steviol moiety is absorbed slowly and is converted to the glucuronide in the liver. Steviol glucuronide is then transport via the bile duct back to the colon where 95% of an ingested amount is excreted. The balance is excreted in the urine. Concentrations will be limited by its high sweetness potency that will moderate ingestion and its concentration will be diluted in the circulation. Nevertheless, it will be available to bind to peripheral sweet receptors. It may also exert a physiological effect in the GI tract and periphery via neural signaling.

Gut signaling

It is well established that carbohydrate ingestion stimulates the release of gut peptides such as GLP-1, GIP, and PYY from enteroendocrine cells in the small intestine. These hormones reportedly modulate appetitive sensations, influence GI transit time, and act as incretins, stimulating the release of insulin from the pancreas. With the identification of sweet taste receptors on enteroendocrine K, L, and K/L cells and co-expression of α-gustducin (a component in the taste transduction system) in several animal models and humans, the possibility was raised that sweeteners could activate these cells with implications for glycemia, appetite, and energy balance (Bryant and Mclaughlin, 2016). Moreover, it was demonstrated that sweet receptor regulated SGLT1 expression and glucose absorptive capacity increase in response to luminal sugars and LCS in mice (Margolskee et al., 2007). Early reports based on cell culture (Jang et al., 2007) or rodent models (Mace et al., 2007) indicated that LCS could stimulate GLP-1 release and enhance glucose transport across enterocytes possibly raising post-prandial glycemia. Subsequent rodent studies indicated glucose was effective but not a range of LCS (Fujita et al., 2009). This was followed by several human trials that yielded no robust or consistent effects (e.g., Ma et al., 2010; Pepino et al., 2013; Brown et al., 2009; Steinert et al., 2011; Wu et al., 2012; Bryant et al., 2014; Bryant and McLaughlin, 2016). At present, the findings in rodent models do not appear to translate to humans.

Microbiome

The microbiome is responsive to short and longer-term dietary intake. Sensory and nutritive properties of foods and beverages influence ingestive decisions and activate neural and endocrine signaling pathways that influence digestion dynamics and, as a result, the

types and quantities of food components that reach the colon. These substrates are then differentially used by the various colonic bacterial species leading to changes in their balance and the metabolic products produced. Understanding of these process is incomplete, hampering predictions of how individual sweeteners will impact the microbiome and health-related outcomes.

Sugar

Except under pathological and specific non-pathological conditions (e.g., disaccharide intolerance), most nutritive sweeteners are efficiently digested and absorbed in the small intestine so are not present in appreciable amounts to serve as substrate for colonic bacteria. Exceptions include selected polyols (e.g., maltitol, xylitol, isomalt, lactitol). Erythritol, sorbitol, and mannitol do not affect the microbiota, and some sweet oligosaccharides are poorly digested (Ruiz-Aceituno et al., 2018). These may be added to foods for their prebiotic properties. They reportedly hold promise for a number of health outcomes (Belorkar and Gupta, 2016) that will require additional study to confirm.

Low calorie sweeteners

There are multiple, recent critical reviews of the evidence linking LCS consumption with the microbiome (Glendinning, 2018; Ruiz-Ojeda et al., 2019; Lobach et al., 2019). Evidence from animal trials reveals different patterns of effects for selected LCS. Alterations of microbial communities have been reported with saccharin, aspartame, sucralose, stevia, and ace-K. The mechanisms for these effects remain to be elucidated since aspartame does not reach the colon and saccharin and Ace-K do so in very minimal amounts. Additional concerns about these findings include the use of non-physiological dosing in some studies; failure to control the total diet, hampering attribution of effects to the sweetener; inconsistency of effects across trials; and extrapolation of such trials to humans with markedly different GI tracks and microbial populations compared to rodent models. Current evidence is mixed. One trial of 172 individuals who completed a food frequency questionnaire reported a positive association between LCS ingestion and various microbial species (Suez et al., 2014). In another trial of 31 individuals, 7 consumed aspartame, 7 consumed Ace-K, and 3 consumed both for 4 days (Frankenfeld et al., 2015). No effects on bacterial abundance were identified though LCS consumers versus non-consumers differed in microbial diversity. However, because background diet was not controlled, it is not possible to attribute even this difference to a sweetener. Well-designed and adequately powered human trials are lacking.

Sweet hedonics

It is widely accepted that there is an inherent liking for sweetness that is demonstrable through a stimulatory effect of in-utero exposure to saccharin on fetal drinking (De Snoo, 1937), as well as early post-natal mimetic reflexes (Steiner, 1973; Rosenstein and Oster, 1988) and sucking responses (Maone et al., 1990) to oral stimulation with sweet stimuli. There is a growing literature that early flavor exposure may hold implications for food acceptance later in life (Mennella, 2014). Very preliminary evidence in human infants is mixed on whether early sweetener exposure is a determinant of later sweet acceptance (Beauchamp and Moran, 1984; Liem and Mennella, 2002). However, dietary experience and cultural influences are the primary determinants of where sweetness is expected in the food supply, when such items are to be ingested, appropriate portion sizes, frequency of ingestion, as well as on the type and concentration of sweetener in an item. There is high inter-individual variability in the preferred sweetness levels of foods (Witherly et al., 1980). In that sweeteners are generally not consumed in isolation, hedonics is not closely associated with concentration, and the context in which sweeteners are ingested modifies their appeal. Taken together, the evidence suggests hedonic responses to sweetness have a biological basis, but are highly individual, context-specific and a reflection of recent exposure.

Sweeteners and appetite

It has been proposed that oral sweetener exposure, particularly with LCS, augments hunger (Blundell and Hill, 1986; Rogers et al., 1988; Tordoff and Alleva, 1990). However, the preponderance of evidence indicates nutritive sweeteners (Steinert et al., 2011) and LCS (Ford et al., 2011; Peters and Beck, 2016) suppress hunger or increase fullness, or, more commonly, have no independent impact on appetitive sensations (Sørensen et al., 2014; Fantino et al., 2018; Mattes and Popkin, 2009). A remaining question concerns the role of expectations on appetite ratings. Knowledge of the type of sweetener consumed may influence appetitive responses such that LCS could be expected by consumers to have a lesser impact on appetite.

The biological basis for sweetener intake

There is a widespread view that humans have a sensitivity to and preference for sweetness to facilitate the identification and ingestion of carbohydrate energy sources. This teleological perspective is supported by evidence that selected vertebrates (e.g., cats, sea lions, whales, pandas, western clawed frog, vampire bats) that do not rely on carbohydrate for energy have limited or no ability to detect sweetness. Additionally, hummingbirds, who feed on nectar, have repurposed the umami receptor to permit detection of carbohydrate again, while other birds have not (Baldwin et al., 2014; Zhao et al., 2010). However, this scenario has been questioned (Feng and Zhao, 2013) and there are other issues that do not support this view. First, the primary effective taste stimuli for sweetness are mono- and disaccharides, which would have been rare and only episodically encountered throughout most of human evolution. A sensory system tuned to starch, the overwhelming form of carbohydrate in the environment, would have held greater survival

value. There is some evidence humans can detect short chain oligosaccharides, but they are not detected by the sweet receptor and are not sweet (Lim and Pullicin, 2019). Second, unlike the case with fat and protein, where the effective stimulus is an essential nutrient, this is not the case for sweet stimuli. Thus, a sensory system tuned to sweeteners holds less salience. Third, sweetness is a very imperfect predictor or energy. Sweet fruits are generally energy dilute and potent sweeteners are consumed in such limited quantity that they contribute little to energy intake. Fourth, sweetness also is a poor indicator of safety. Heavy metals, such as lead, are sweet but neurotoxic. Complications from the ingestion of lead date back to the Romans who used a concoction of lead acetate to sweeten foods, and presently, children are at high risk for lead toxicity because of the appeal of sources like paint chips. These observations, the intake trends noted above, and the health consequences of sweetener ingestion outlined below, suggest sweetener intake is driven more by hedonics than homeostatic mechanisms.

Consumption trends

Sugar

Terminology surrounding sugar intake focuses on total, free, and added sugars. Total sugars refer to all sources of mono- and disaccharides in foods and beverages (Mela and Woolner, 2018). Free sugars are defined, per the World Health Organization, as "monosaccharides and disaccharides added to foods and beverages by the manufacturer, cook or consumer, and sugars naturally present in honey, syrups, fruit juices and fruit juice concentrates" (World Health Organization, 2021). The US Food and Drug Administration (FDA) and agencies in several other countries focus on added sugars rather than free sugars. The FDA defines added sugars as "sugars that are added during the processing of foods (such as sucrose or dextrose), foods packaged as sweeteners (such as table sugar), sugars from syrups and honey, and sugars from concentrated fruit or vegetable juices" (US Food and Drug Administration, 2021a). The FDA excludes naturally occurring sugars from this definition. Despite the slight differences in definitions, free and added sugars are the primary focus of public health recommendations surrounding sugar intake.

Free or added sugars in the diet come from a variety of sources, but beverages are significant contributors. Among American adults, sweetened beverages are the largest source of added sugar intake, followed by desserts (Bowman et al., 2019a). Tea and coffee are the third largest contributor of added sugar, followed by candy and sugars and breakfast cereals and bars. Broken down by age range, desserts are the primary contributor to added sugar intake among individuals ages 2–5 years (Bowman et al., 2019b) and 51–70 years, followed by sweetened beverages; while sweetened beverages are the leading contributor for all other age groups (US Department of Health and Human Services and US Department of Agriculture, 2020). Together, all five food groups mentioned contribute about 70% of the energy from added sugar while beverages provided nearly half. Similar trends have been noted among many other countries (Amarra et al., 2016; Lei et al., 2016; Promdee et al., 2007; Ruiz et al., 2017), and a global survey from 2010 reported that sugar sweetened beverage consumption was highest among Caribbean countries (Singh et al., 2015).

Public health messaging around sugar intake is largely consistent throughout the world, but there are some differences. For example, the World Health Organization (WHO) (World Health Organization, 2015) recommends limiting free sugar intake to less than 10% of total energy. These recommendations apply to both children and adults and were based on the effects of sugar on oral health (World Health Organization, 2015). Health Canada (Health Canada, 2019) also adopted the WHO recommendations. The US Dietary Guidelines align with WHO recommendations but are targeted to added, rather than free, sugar intake (Department and Services, 2020). Other countries, like the United Kingdom (National Health Service, 2018), recommend free sugar intakes of less than 5% of total energy. The WHO also conditionally supports a 5% limit, but this is based on low-quality data that suggest diets with less than 5% of energy coming from total sugars are even more beneficial for oral health than the 10% cut-off (World Health Organization, 2015). The US Dietary Guidelines advisory committee recommended reducing the previous recommendation of added sugar intake from 10% to 6%, but the recommendation was not accepted and remains unchanged in the 2020–2025 version of the guidelines (Thompson, 2020; US Department of Health and Human Services and US Department of Agriculture, 2020). The DGA's original 10% recommendation is based on the fact that meeting all other nutrient recommendations while achieving energy and sodium intake recommendations becomes exceedingly difficult if added sugars are not limited (US Department of Health and Human Services and US Department of Agriculture, 2020). Some argue the science does not fully support a limit as low as 6% (Thompson, 2020), and others doubt if such restrictions are achievable (Erickson and Slavin, 2015). Overall, recommendations to limit energy from free or added sugar to 5–10% of total energy intake are frequently noted.

Trend data on sugar consumption suggest reductions have occurred recently. In the US, increases in intake were observed between the 1970s and 1990s (Chun et al., 2010) where total sugar intake increased by 8% and added sugars by 12%. Nutritive sweetener intake, which includes sugar, honey, and corn syrup, appeared to peak in 1999, when per capita intake reached 422.6 kcals/day (US Department of Agriculture, 2021). However, intake has been gradually decreasing. In 2019, intake was estimated at 344.3 kcals/day (US Department of Agriculture, 2021), which represents a decrease of 18.5%. This trend was also reported based on analysis of multiple cycles of the National Health and Nutrition Examination Survey (7 cycles spanning 2003–2004 to 2015–2016) (Marriott et al., 2019). American children and adults have reduced their total sugar intake over time by 27% and 17%, respectively (Marriott et al., 2019). Reductions in the intake of sugar sweetened beverages (SSBs) over the study timeframe contributed substantially to this reduction. In addition, the US food supply has reduced the number of available products that contain added nutritive sweeteners, like sucrose and fructose (Popkin and Hawkes, 2016). Compared to 2000, when 63% of all food and beverage products sold in the US contained added nutritive sweeteners, by 2013, this was reduced to

55% (Popkin and Hawkes, 2016). Notably, this reduction in added sugars has not been accompanied by lower energy intake or body weight in the US, in part due to increased energy from high-quality carbohydrates, polyunsaturated fats, saturated fats, and plant protein (Shan et al., 2019).

Interest in reducing sugar intake has been documented in many nations. For example, a survey of Kenyan adults reported that 54% of respondents indicated they were trying to reduce sugar intake (Mwenda et al., 2018), and among Polish adults, nearly 100% of those surveyed claimed they were trying to limit sugar intake (Pielak et al., 2019). American adults are also attempting to reduce sugar intake; 80% of respondents in a recent survey indicated they were trying to either limit or avoid sugar intake (International Food Information Council Foundation, 2019). The same survey of over 1000 American adults noted that sugar reduction was the most commonly mentioned dietary change respondents had undertaken in the past 10 years (International Food Information Council Foundation, 2019). Clearly, the public is aware of health messaging regarding sugar.

Despite trends suggesting reduced sugar consumption, sugar intake frequently exceeds recommendations. For example, data from surveys of European countries indicate that total and added sugar intake of adults comprised between 15–21% and 7–11% of dietary energy, respectively; while total and added sugar intake among children ranged from 16 to 26% and 11 to 17% of total energy, respectively (Azaïs-Braesco et al., 2017). Data from the US indicates 54% of American adults and 66% of children consume more added sugars than recommended (Bowman et al., 2017). Broken down by age group, 50% percent of children ages 1–3 years consume more than 10% of their energy from added sugars (Dietary Guidelines Advisory Committee, 2020). Among children ages 4–8 years, 79% of boys and 75% of girls do not meet added sugar intake recommendations (Dietary Guidelines Advisory Committee, 2020). Results are similar among adolescents; 70% of boys and 75% of girls do not meet recommendations (Dietary Guidelines Advisory Committee, 2020). Average per capita intake of nutritive sweeteners was 344.3 kcals/day in the US population in 2019, which far exceeds recommendations of limiting added sugar intake to less than 10% of energy (US Department of Agriculture, 2021). These data suggest that intentions and intakes do not align.

Developing areas also struggle with high sugar intake. For example, a survey of eight Latin American countries reported average added sugar intakes of 13.2% of total energy with a range from 10.3% in Ecuador to 16.4% in Argentina (Fisberg et al., 2018). Among Gaza Palestinians, 44.1% of respondents consumed more than recommended amounts of added sugars (Jebril et al., 2020). Results from the Global School Health Survey indicated that over 45% of students ages 13–17 in Saint Lucia reported consuming sugar sweetened beverages at least once per day while over 55% of students in Guatemala and over 75% in Suriname answered similarly (Pan American Health Organization, 2021). Nearly 14% of adults in Kenya were classified as consuming a diet high in sugar (Mwenda et al., 2018). A high sugar diet was defined as adding "sugar to drinks already served with sugar or intake of processed foods or drinks high in sugar on a daily basis". Based on current public health recommendations, over consumption of added sugars is a global concern.

Low calorie sweeteners

One approach used by consumers to reduce sugar intake is to substitute LCS. LCS are substances that produce a sweet sensation but provide minimal energy due to their high potency (hence extremely low concentration) and/or poor digestibility. While different LCS are approved for use in different countries, six synthetic LCS have been approved by the Food and Drug Administration as food additives. These include: acesulfame potassium (Ace-K), advantame, aspartame, neotame, saccharin and sucralose. Three other natural products, steviol glycosides, thaumatin, and luo han guo fruit extracts are also approved. These compounds have undergone rigorous health testing and have been given Generally Recognized as Safe (GRAS) status (US Food and Drug Administration, 2021b).

LCS are widely consumed, and evidence suggests use continues to grow. Based on two 24 h diet recalls collected during the 2009–2012 National Health and Nutrition Examination Survey (NHANES), 41.4% of American adults and 25.1% of children consumed LCS (Sylvetsky et al., 2017). Among adults, individuals who were White, non-Hispanic, female, obese, and from the highest income tertile consumed more LCS (Sylvetsky et al., 2017). Compared to the number of adults (24.1%) and children (12.5%) who reported using LCS in 2007–2008 NHANES survey, numbers from the 2009 to 2012 NHANES survey represent a marked increase (Sylvetsky et al., 2012). The most popular reasons for using sweeteners by US consumers were to reduce sugar intake, lose weight, consume less energy, and control blood sugar (International Food Information Council Foundation, 2019). The most recent US Dietary Guidelines report suggests low- and no-calorie sweeteners may aid in reducing sugar intake and facilitate weight management, at least in the short term (US Department of Health and Human Services and US Department of Agriculture, 2020). From a global perspective, a recent review of worldwide exposure studies that evaluated daily intake of seven different LCS reported variable use by region and LCS, with no data available for African and limited data available for North American consumers in the format required and time period assessed (2008–2017) (Martyn et al., 2018). However, based on the available data, the authors concluded that there is little cause for concern that acceptable daily intakes are being exceeded at present (Martyn et al., 2018). Thus, LCS are commonly used and current exposures do not appear to exceed recommended intakes.

Consumers have several choices for discretionary LCS use. Splenda/sucralose is currently the most popular choice in the US, followed by Sweet'N Low (saccharin), Equal (blend of aspartame and Ace-K), and Stevia in the Raw (Statista, 2020). Twice as many people report using Splenda compared to each of the other top LCS (Statista, 2020). While the global LCS market is expected to grow through 2026 (Statista, 2021b), multiple surveys suggest that consumers are concerned that there are adverse health effects from consuming synthetic LCS, and as a result, some consumers seek to avoid these options (360 Market Updates, 2019; International Food Information Council Foundation, 2018; Strategy Online, 2017). These concerns have likely contributed to the

increasing popularity of naturally-derived LCS, like stevia, around the globe (Research and Markets LTD, 2020; Saraiva et al., 2020; Statista, 2021a; Strategy Online, 2017). While a consensus definition of "naturalness" is lacking, a systematic review of studies from 32 countries concluded that the concept is important to many consumers (Román et al., 2017).

In addition to using LCS to reduce sugar intake, other policy approaches have been implemented. These policies are meant to influence access to sugar-containing foods and beverages as well as LCS. Examples of policy approaches include levying taxes on SSBs and energy drinks (World Health Organization, 2017), bans on sales of SSBs in schools or businesses (Eneli et al., 2014; Epel et al., 2020), manufacturing guidelines to limit sugar in products (Enright and Eskensazi, 2018), and expanded labeling on products (Huang et al., 2019). While various studies reach differing conclusions, there is evidence to suggest these approaches can be effective. For example, an excise tax of approximately 10% reduced beverage purchases in Mexico by 12% over a two-year period (Colchero et al., 2016). A beverage tax was also effective in reducing consumption in the US (Falbe et al., 2016). A meta-analysis also concluded that SSB taxes reduced demand and the authors concluded that taxation improved BMI (Cabrera Escobar et al., 2013). Banning the sales of SSBs has also been associated with reduced consumption (Avery et al., 2015; Epel et al., 2020). Meta-analyses conclude that product reformulation to reduce sugar intake result in decreased sugar intake (Hashem et al., 2019), and that front of package labeling regarding sugar also reduces intake (Croker et al., 2020). In summary, while there is controversy regarding the best approaches to use to curb added sugar intake, there appear to be a variety of policies that can be used to address sugar intake in order to achieve public health recommendations.

Sweeteners and health

The effects of sweeteners, both nutritive and non-nutritive, on human health have been extensively investigated and the findings have been synthesized by multiple systematic reviews and meta-analyses. Evidence from these reviews related to overweight/obesity, cardiovascular disease (CVD), type 2 diabetes, cancer and dental caries are presented, first for nutritive sweeteners followed by LCS.

Sugar

There are concerns about the impact of sugar consumption on human health due to the acute effects sugar has on human ingestive behaviors and physiological responses (as discussed in the previous section). Whether these acute effects translate into long-term health consequences is summarized below.

Overweight and obesity

The relationship between free sugar intake and body weight was summarized in a systematic review that included 38 cohort studies and 30 interventional trials (Te Morenga et al., 2013). In adults, 11 of 16 cohort studies reported positive associations between free sugar and weight, BMI, and/or waist circumference. Only one cohort study noted a significant negative association. An increase in sugar intake over time was also associated with increased body weight. The associations from observational studies were confirmed by interventional trials, where a reduction in free sugar intake led to an average weight loss of 0.80 kg ($n = 5$ studies included in the meta-analysis). In contrast, increasing free sugar intake led to weight gain, and the magnitude of weight gain was positively related to study length (0.52 kg higher in studies up to 8 weeks, and 2.73 kg higher for trials >8 weeks). Meta-analysis of 13 studies that isocalorically replaced free sugars with other nutrients revealed no effect on body weight (mean difference = 0.04 kg, 95%CI −0.04–0.13 kg, $n = 13$ studies), suggesting that the effects of sugars on body weight may be adequately explained by their energy content. However, a systematic review did not observe an effect of reducing sugar intake on the body weight of children ($n = 5$ interventional trials). The authors indicated that this could be explained by the poor compliance to reduced free sugar diets in children. Overall, the evidence indicates that free sugar intake is positively associated with body weight in adults while the evidence in children is less clear.

A number of recent systematic reviews also specifically examined the relationship between SSB and body weight in adults and children. In adults, one review that included 11 observational studies reported 18% higher obesity risk (RR = 1.18, 95%CI 1.10–1.27, $n = 11$ studies) and 20% higher waist circumference (RR = 1.20, 95%CI 1.04–1.37, $n = 4$ studies) in SSB consumers than non-consumers (Ruanpeng et al., 2017). This was confirmed by another systematic review that included 10 studies (Luger et al., 2017), where eight of the studies did not overlap with those included in the previous review (Ruanpeng et al., 2017). In children, SSB intake has been associated with a higher incidence of adiposity by multiple observational studies (Te Morenga et al., 2013; Luger et al., 2017; Keller and Bucher Della Torre, 2015). Interventional trials conducted in children are limited ($n = 3$ studies). Among these, interventions were not SSB-specific in two studies, and two studies were secondary analyses (Luger et al., 2017). Considering these limitations, the evidence is insufficient to draw a conclusion (Dietary Guidelines Advisory Committee, 2020).

To summarize, the preponderance of evidence supports a positive association between sugar (free and via SSB) and body weight and risk of obesity in adults and, to a slightly lesser extent, in children. Few studies report a negative association. However, as highlighted by the scientific report compiled by the 2020 Dietary Guidelines Advisory Committee, the risk of bias in SSB studies is high, and generalizability is of concern. For these reasons, evidence relating SSB to obesity was rated by the DGAC as moderate in children, and limited in adults (Dietary Guidelines Advisory Committee, 2020).

Cardiovascular disease (CVD)

A recent systematic review of 24 prospective studies examined the relationships between free sugar intake and the incidence of CVD (median follow-up of 11 years), and CVD mortality (median follow-up of 13 years) (Khan et al., 2019). Sugar intake (total, sucrose, fructose) was not associated with the incidence of CVD. However, non-linear dose-response positive relationships were observed between total sugar intake (RR = 1.09, 95%CI 1.02–1.17, $n = 4$ studies) and fructose intake (RR = 1.08, 95%CI 1.01–1.15, $n = 2$ studies) and CVD mortality. Interestingly, sucrose was associated with decreased CVD mortality (RR = 0.94, 95%CI 0.89–0.99, $n = 2$ studies), while no association was noted for added sugar intake. These latter findings may be factual, but the authors suggested the unexpected lack of or negative relationships observed for sucrose and added sugar intake and CVD risk may be explained by reverse causation, where individuals with high risk of CVD may avoid sugars as a preventative measure.

The relationship between sugar intake and CVD was also tested specifically for SSB consumption. In a systematic review of prospective studies, every additional serving/day of SSB was associated with higher incident hypertension (RR = 1.10, 95%CI 1.06–1.15, $n = 6$ studies), coronary heart disease (RR = 1.16, 95%CI 1.06–1.27, $n = 4$ studies), and stroke (RR = 1.10, 95%CI 1.00–1.20, $n = 4$ studies) (Xi et al., 2015). In another review, the risk of myocardial infarction was also reported to be 22% higher (RR = 1.22, 95%CI 1.14–1.30, $n = 2$ studies) with incremental increases in SSB consumption (Narain et al., 2016). Despite these reported associations, the number of studies investigating the relationship between sugar and CVD is still very limited, therefore the 2020 DGAC concluded a firm conclusion cannot be drawn (Dietary Guidelines Advisory Committee, 2020).

Type 2 diabetes mellitus (T2DM)

The association between free sugar intake and T2DM has been summarized by a systematic review that included 15 prospective studies, with a median follow-up period of 12 years (Tsilas et al., 2017). When individuals with the highest vs. the lowest intake were compared, no association was observed between total sugar or fructose intake and T2DM. The lack of association was, partly, due to the significant heterogeneity in studies included in the meta-analyses ($I^2 = 76\%$ and 71% for total sugars and fructose respectively). Although an association was noted between sucrose intake and T2DM, the risk was lower in those with the highest consumption (RR = 0.89, 95%CI 0.80–0.98, $n = 8$ studies). The studies on sucrose had large study populations (almost 200,000 participants in total) and used validated dietary assessment tools. However, there were some differences between studies in term of study population (five studies included females and one study included males only, only some studies included younger participants) as well as T2DM outcomes (four studies relied upon medical diagnosis and the remaining were self-reported). Therefore, more studies are needed to confirm the relationship between free sugar intake and T2DM.

The evidence is more consistent relating SSB intake to T2DM. A number of meta-analyses have reported higher risk of T2DM (ranging from 18% to 30% higher) in individuals with higher SSB consumption (Wang et al., 2015; Malik et al., 2010; Greenwood et al., 2014; Imamura et al., 2015). Although these reviews included some similar studies, the outcome criteria varied across the meta-analyses, e.g., risk was assessed between the lowest vs. the highest intake (Malik et al., 2010; Wang et al., 2015); per 250 mL/day (Imamura et al., 2015) or 330 mL/day (Greenwood et al., 2014) increase; between geographical regions (Wang et al., 2015), and was also based on different age groups (Imamura et al., 2015). The higher risk was comparable between Asia, Europe, and the US (Wang et al., 2015). After adjusting for BMI, the relationships remained significant, although the reported risks were slightly lower. This is consistent with the scientific report of the 2020 Dietary Guideline Advisory Committee, where intake of added sugars, especially as SSB, was reported to increase the risk of T2DM in adults, though this was partly explained by BMI (Dietary Guidelines Advisory Committee, 2020). In another review that focused on fruit juice, the risk and incidence of T2DM was also reported to be higher, but this occurred only when the fruit juice was sweetened (RR = 1.28, 95%CI 1.04–1.59, $n = 2$ studies), and 100% fruit juice intake was not associated with T2DM (Xi et al., 2014). It should be noted that studies on fruit juice, to-date, were of mixed quality due to the lack of consistency in how 100% fruit juice was defined and quantified (Dietary Guidelines Advisory Committee, 2020).

To-date, there is evidence linking SSB intake to the risk of developing T2DM, which is not fully explained by adiposity. However, the relationship remains inconclusive for free sugars. The consistent relationship between SSB and metabolic diseases presented so far may be partly explained by the physical forms of sugar being ingested. In a liquid form, sugars are consumed faster, cause rapid fluctuations in blood glucose, and they have poor appetitive effects (Dhillon et al., 2016).

Cancer

A recent systematic review has summarized the findings of observational studies on the relationship between free sugar intake and cancer risk (Makarem et al., 2018). This review concluded that the evidence was mixed, where the association with cancer risk was not observed in 11 out of 29 studies (38%) that assessed total sugar and sucrose intake. In analyses involving fructose, 8 out of 14 studies (57%) did not observe an association, and mixed findings were reported in the remaining studies, i.e., 2 studies reported negative and 4 studies noted positive associations. Only 5 studies on added sugar were identified in this review. Two of these studies reported an association, but they were in opposite directions, possibly due to different target cancer types (RR = 0.72 for ovarian cancer vs. RR = 1.60 hematological cancer). The explanation for inconsistent findings is unclear as there was no evidence the outcomes were influenced by sample size and follow-up period. The authors of the review highlighted the challenges in accurately quantifying specific sugars in the diet, and the limitation of cross-sectional analysis of intake (not necessarily representative of long-term intake), which may explain the mixed findings (Makarem et al., 2018).

Inconsistent findings were also observed for trials examining fruit juice and SSB intake. In a systematic review of 64 observational studies, SSB intake was associated with higher risk of breast (RR = 1.14, 95%CI 1.01–1.30, $n = 7$ studies, $I^2 = 0\%$) and prostate (RR = 1.18, 95%CI 1.10–1.27, $n = 3$ studies, $I^2 = 0\%$) cancer (Llaha et al., 2021). A similar (but weaker) association was reported with fruit juice intake and prostate cancer (RR = 1.03, 95%CI 1.01–1.05, $n = 4$ studies, $I^2 = 0\%$). However, this review observed no association between SSB or fruit juice and other cancer types such as colorectal, bladder, renal cell, and pancreatic cancers.

Overall, only a very small number of studies reported a link; therefore, the association between sugar intake and cancer risk is considered weak. Additional evidence is required to confirm this conclusion.

Dental caries

A link between SSB intake and dental caries is well-established. In a recent systematic review of 38 cross-sectional studies, higher SSB intake in children was associated with higher risk of dental caries (OR = 1.57, 95%CI 0.38–1.26, $n = 16$ studies) and erosion (OR = 1.43, 95%CI 1.01–2.03, $n = 7$ studies). This review also reported that there were 0.82 more decayed, missing, and filled teeth in children consuming moderate vs. low, and high vs. moderate levels of sweeteners compared to children with low SSB intake (Valenzuela et al., 2021). Another systematic review of five cohort studies further suggested that higher risk of dental caries is linked to between-meal consumption of sugars (Hancock et al., 2020).

Low calorie sweeteners

Several mechanisms have been proposed to explain associations between LCS consumption and human health. Some are well-established, but most are still preliminary and require further investigation. These mechanisms relate to appetite regulation (Romo-Romo et al., 2016), energy intake (Rogers and Appleton, 2020), substrate oxidation (Chern and Tan, 2019), glucose metabolism (Romo-Romo et al., 2016; Tucker and Tan, 2017), advanced-glycation end product production (Deo et al., 2020), and effects on the gut microbiome (Walbolt and Koh, 2020).

The link between LCS intake and human health has been widely studied and several systematic reviews have been published. However, most of these reviews are narrative and fail to differentiate between the various LCS and how they are ingested. Some additional challenges in understanding the association between LCS and health include the extensive range of LCS-containing food products that may not be fully captured by nutrient analysis software, hence the lack of ability to accurately identify the types of sweeteners ingested, and to quantify their contributions to the food supply. This limits determination of potential dose-response analyses. The evidence presented below should therefore be interpreted within the context of these methodological limitations. Most trials focused on LCS consumed in beverages; the most common source of LCS. It is also important to note the high possibility of reverse causation in findings related to LCS and human health, where LCS products may be more frequently used by individuals attempting to manage disease risks.

Overweight and obesity

In a review of prospective observational studies, LCS beverage intake was positively, but weakly, correlated with the BMI of children and adults (mean correlation = 0.03, 95%CI 0.01–0.06, $n = 6$ studies) (Miller and Perez, 2014). However, no association was found between LCS use and body weight. The same review assessed clinical trials, where LCS interventions led to lower BMI (mean difference = -0.24 kg m^{-2}, 95%CI -0.41 to -0.07 kg m^{-2}, $n = 6$ studies), fat mass (mean difference = -1.10 kg, 95% CI -1.77 to -0.44 kg m^{-2}, $n = 6$ studies), and waist circumference (mean difference = -0.83 cm, 95%CI -1.29 to -0.37 cm, $n = 3$ studies) (Miller and Perez, 2014). The effects of LCS, compared to sugars, on body weight was further strengthened by two meta-analyses of interventional studies, where body weight reduction was seen with LCS (standardized mean difference SMD = -0.56 kg and -0.35 kg respectively) (Laviada-Molina et al., 2020; Rogers and Appleton, 2020). Lower weight in the LCS group was observed when the sweeteners were included in unrestricted diets (SMD = 0.47 kg, 95%CI 0.26–0.68 kg); were used by people with overweight/obesity (SMD = 0.51 kg, 95%CI 0.29–0.72 kg); and were consumed by adults but not children (SMD = 0.43 kg, 95%CI 0.22–0.64 kg, $n = 23$ studies) (Laviada-Molina et al., 2020). The totality of evidence indicates LCS use is associated with lower adiposity when they replace sugars in the diet. Although two reviews questioned the effectiveness of LCS on body weight management (Toews et al., 2019; Azad et al., 2017), the conclusions on these reviews were based on the inclusion of limited or inappropriate studies, as described by Sievenpiper et al. (2017). Finally, non-alcoholic fatty liver disease (NAFLD) is commonly associated with obesity and a systematic review (Green and Syn, 2019) that included 2 interventional trials concluded that LCS had no effect (Maersk et al., 2012), or reduced clinical indicators of NAFLD (Campos et al., 2015) when SSB was replaced with LCS-containing beverages.

Type 2 diabetes

Observational studies suggest elevated risk of type 2 diabetes with higher LCS intake when reported as servings per day (RR = 1.03, 95%CI 1.01–1.05, $n = 4$ studies) or for comparisons of lowest vs. highest quartile's of intake (RR = 1.14, 95%CI 1.05–1.25, $n = 9$ studies) (Azad et al., 2017). In a scoping review, positive associations were also reported between LCS consumed in beverages and the risk of T2DM, but findings were mixed when overall LCS intake was considered (Lohner et al., 2017). This highlights the importance of assessing the form of LCS being consumed. Nonetheless, this is an area of particular risk for reverse causation, where people with diabetes are more likely to use LCS to improve glycemic control. Findings are also subject to publication bias, inconsistency, and imprecision in pooled estimates (Sievenpiper et al., 2017).

In a meta-analysis of 26 feeding trials, LCS intake alone did not influence acute postprandial glucose (mean difference = −0.02 mmol/L, 95%CI −0.09–0.05 mmol/L) and insulin (mean difference = −2.39 pmol/L, 95%CI −11.83–7.05 pmol/L) concentrations, regardless of the types and dose of LCS ingested (Greyling et al., 2020). When LCS was compared to other nutritive sugars, there was a smaller or no increase in blood glucose and insulin due to a lower carbohydrate load (Tucker and Tan, 2017). In short-term interventional studies, LCS consumption also did not affect insulin resistance or glycated hemoglobin (HbA1c) (Azad et al., 2017). Considering all evidence and potential reverse causation from observational studies, LCS use does not appear increase the risk of T2DM.

Cancer

Governmental bodies such as the US Food and Drug Administration as well as its counterparts in many other nations have concluded that LCS are safe for consumption. This was supported by a scoping review that included 51 primary studies that reported limited evidence on the link between LCS use and cancer risk (Lohner et al., 2017).

Conclusion/summary/outlook

Despite strong recommendations to moderate sugar intake, reductions have been limited and the consumption of LCS is high and trending upward. This reflects the inherent appeal of sweetness and the reinforcing contribution of dietary experience and culture. Concern about the safety and health effects of high levels of consumption have prompted recommendations to moderate intake. To-date this has resulted in only modest success, but new approaches are emerging. While sweeteners share a common property, sweetness, they vary markedly in chemical structure resulting in differences in their taste properties, digestion, absorption, metabolism, and health effects. The strength of evidence implicating nutritive sweetener intake in the onset and manifestations of overweight/obesity, CVD, and diabetes ranges from limited to moderate with a stronger association for dental caries and inconclusive evidence for selected cancers. Much of this may be attributable to the energy contributed by these sweeteners. They are singled out as problematic for this contribution because they are often, though not necessarily, ingested though foods and beverages that provide limited other nutritive value. In contrast, there is a weak, but significant, inverse association between LCS use and overweight/obesity and no significant association with CVD, diabetes, or cancer.

Conflict of interest

RDM currently serves on the scientific advisory boards of the Grain Food Foundation and Mars, Inc. RMT currently serves on the scientific advisory board of the Central Bottling Company.

References

360 Market Updates 2019. Artificial Sweeteners Market—Growth, Trend and Forecast (2019–2024).
Allen, A.L., Mcgeary, J.E., Knopik, V.S., Hayes, J.E., 2013. Bitterness of the non-nutritive sweetener acesulfame potassium varies with polymorphisms in TAS2R9 and TAS2R31. Chem. Senses 38, 379–389.
Amarra, M.S.V., Khor, G.L., Chan, P., 2016. Intake of added sugar in Malaysia: a review. Asia Pac. J. Clin. Nutr. 25, 227–240.
Avery, A., Bostock, L., Mccullough, F., 2015. A systematic review investigating interventions that can help reduce consumption of sugar-sweetened beverages in children leading to changes in body fatness. J. Hum. Nutr. Diet. 28, 52–64.
Azad, M.B., Abou-Setta, A.M., Chauhan, B.F., et al., 2017. Nonnutritive sweeteners and cardiometabolic health: a systematic review and meta-analysis of randomized controlled trials and prospective cohort studies. Can. Med. Assoc. J. 189, E929–E939.
Azaïs-Braesco, V., Sluik, D., Maillot, M., et al., 2017. A review of total & added sugar intakes and dietary sources in Europe. Nutr. J. 16, 6.
Baldwin, M.W., Toda, Y., Nakagita, T., et al., 2014. Evolution of sweet taste perception in hummingbirds by transformation of the ancestral umami receptor. Science 345, 929–933.
Beauchamp, G.K., Moran, M., 1984. Acceptance of sweet and salty tastes in two-year-old children. Appetite 5, 291–305.
Behrens, M., Meyerhof, W., 2019. A role for taste receptors in (neuro) endocrinology? J. Neuroendocrinol. 31, e12691.
Belorkar, S.A., Gupta, A., 2016. Oligosaccharides: a boon from nature's desk. Amb. Express 6, 1–11.
Blundell, J.E., Hill, A.J., 1986. Paradoxical effects of an intense sweetener (aspartame) on appetite. Lancet 1, 1092–1093.
Bowman, S.A., Clemens, J.C., Martin, C.L., et al., 2017. Added Sugars Intake of Americans: What We Eat in America, NHANES 2013–2014.
Bowman, S.A., Clemens, J.C., Friday, J.E., et al., 2019a. Added Sugars in Adults' Diet: What We Eat in America, NHANES 2015–2016. Dietary Data Brief. U.S. Department of Agriculture.
Bowman, S.A., Clemens, J.C., Friday, J.E., et al., 2019b. Added Sugars in American Children's Diet: What We Eat in America, NHANES 2015–2016. Dietary Data Brief. U.S. Department of Agriculture.
Brown, R.J., Walter, M., Rother, K.I., 2009. Ingestion of diet soda before a glucose load augments glucagon-like peptide-1 secretion. Diabetes Care 32, 2184–2186.
Bryant, C., Mclaughlin, J., 2016. Low calorie sweeteners: evidence remains lacking for effects on human gut function. Physiol. Behav. 164, 482–485.
Bryant, C.E., Wasse, L.K., Astbury, N., et al., 2014. Non-nutritive sweeteners: no class effect on the glycemic or appetite responses to ingested glucose. Eur. J. Clin. Nutr. 68, 629.
Cabrera Escobar, M.A., Veerman, J.L., Tollman, S.M., et al., 2013. Evidence that a tax on sugar sweetened beverages reduces the obesity rate: a meta-analysis. BMC Publ. Health 13, 1072.
Calles-Escandon, J., Robbins, D.C., 1987. Loss of early phase of insulin release in humans impairs glucose intolerance and blunts thermic effect of glucose. Diabetes 36, 1167–1172.

Campos, V., Despland, C., Brandejsky, V., et al., 2015. Sugar-and artificially sweetened beverages and intrahepatic fat: a randomized controlled trial. Obesity 23, 2335–2339.
Chaudhari, N., Roper, S.D., 2010. The cell biology of taste. J. Cell Biol. 190, 285–296.
Chern, C., Tan, S.Y., 2019. Energy expenditure, carbohydrate oxidation and appetitive responses to sucrose or sucralose in humans: a pilot study. Nutrients 11, 1782.
Chun, O.K., Chung, C.E., Wang, Y., et al., 2010. Changes in intakes of total and added sugar and their contribution to energy intake in the U.S. Nutrients 2, 834–854.
Colchero, M.A., Popkin, B.M., Rivera, J.A., Ng, S.W., 2016. Beverage purchases from stores in Mexico under the excise tax on sugar sweetened beverages: observational study. BMJ 352, h6704.
Croker, H., Packer, J., Russell, S.J., et al., 2020. Front of pack nutritional labeling schemes: a systematic review and meta-analysis of recent evidence relating to objectively measured consumption and purchasing. J. Hum. Nutr. Diet. 33, 518–537.
Damak, S., Rong, M., Yasumatsu, K., et al., 2003. Detection of sweet and umami taste in the absence of taste receptor T1r3. Science 301, 850–853.
Davies, M., Rayman, G., Grenfell, A., et al., 1994. Loss of the first phase insulin response to intravenous glucose in subjects with persistent impaired glucose tolerance. Diabet. Med. 11, 432–436.
De Araujo, I.E., Oliveira-Maia, A.J., Sotnikova, T.D., et al., 2008. Food reward in the absence of taste receptor signaling. Neuron 57, 930–941.
De Snoo, K., 1937. Das trinkende kind im uterus. Gynecol. Obstet. Invest. 105, 88–97.
Del Prato, S., Tiengo, A., 2001. The importance of first-phase insulin secretion: implications for the therapy of type 2 diabetes mellitus. Diabetes Metab. Res. Rev. 17, 164–174.
Deo, P., Chern, C., Peake, B., Tan, S.Y., 2020. Non-nutritive sweeteners are in concomitant with the formation of endogenous and exogenous advanced glycation end-products. Int. J. Food Sci. Nutr. 71, 706–714.
U.S. Department of Agriculture, U.S. Department of Health and Human Services, 2020. Dietary Guidelines for Americans, 2020–2025. U.S. Department of Agriculture, U.S. Department of Health and Human Services, Washington, DC.
Dhillon, J., Running, C.A., Tucker, R.M., Mattes, R.D., 2016. Effects of food form on appetite and energy balance. Food Qual. Prefer. 48, 368–375.
Dhillon, J., Lee, J.Y., Mattes, R.D., 2017. The cephalic phase insulin response to nutritive and low-calorie sweeteners in solid and beverage form. Physiol. Behav. 181, 100–109.
Dietary Guidelines Advisory Committee, 2020. Scientific report of the 2020 dietary guidelines advisory committee: advisory Report to the Secretary of Agriculture and the Secretary of Health and Human Services. In: U.S. Department of Agriculture. Agricultural Research Service, Washington D.C.
Dubois, G.E., 2016. Molecular mechanism of sweetness sensation. Physiol. Behav. 164, 453–463.
Eneli, I.U., Oza-Frank, R., Grover, K., et al., 2014. Instituting a sugar-sweetened beverage ban: experience from a children's hospital. Am. J. Publ. Health 104, 1822–1825.
Enright, M., Eskensazi, L., 2018. The Children's Food & Beverage Advertising Initiative Category-Specific Uniform Nutrition Criteria. Council of Better Business Bureaus, Inc, Arlington, VA.
Epel, E.S., Hartman, A., Jacobs, L.M., et al., 2020. Association of a workplace sales ban on sugar-sweetened beverages with employee consumption of sugar-sweetened beverages and health. JAMA Intern. Med. 180, 9–16.
Erickson, J., Slavin, J., 2015. Total, added, and free sugars: are restrictive guidelines science-based or achievable? Nutrients 7, 2866–2878.
Falbe, J., Thompson, H.R., Becker, C.M., et al., 2016. Impact of the Berkeley excise tax on sugar-sweetened beverage consumption. Am. J. Publ. Health 106, 1865–1871.
Fantino, M., Fantino, A., Matray, M., Mistretta, F., 2018. Beverages containing low energy sweeteners do not differ from water in their effects on appetite, energy intake and food choices in healthy, non-obese French adults. Appetite 125, 557–565.
Feng, P., Zhao, H., 2013. Complex evolutionary history of the vertebrate sweet/umami taste receptor genes. Chin. Sci. Bull. 58, 2198–2204.
Fisberg, M., Kovalskys, I., Gómez, G., et al., 2018. Total and added sugar intake: assessment in eight Latin American countries. Nutrients 10, 389.
Ford, H.E., Peters, V., Martin, N.M., et al., 2011. Effects of oral ingestion of sucralose on gut hormone response and appetite in healthy normal-weight subjects. Eur. J. Clin. Nutr. 65, 508–513.
Frankenfeld, C.L., Sikaroodi, M., Lamb, E., et al., 2015. High-intensity sweetener consumption and gut microbiome content and predicted gene function in a cross-sectional study of adults in the United States. Ann. Epidemiol. 25, 736–742.e4.
Fujita, Y., Wideman, R.D., Speck, M., et al., 2009. Incretin release from gut is acutely enhanced by sugar but not by sweeteners in vivo. Am. J. Physiol. Endocrinol. Metab. 296, E473–E479.
Glendinning, J.I., 2018. Oral and post-oral actions of low-calorie sweeteners: a tale of contradictions and controversies. Obesity 26, S9–S17.
Green, C.H., Syn, W.K., 2019. Non-nutritive sweeteners and their association with the metabolic syndrome and non-alcoholic fatty liver disease: a review of the literature. Eur. J. Nutr. 58, 1785–1800.
Greenwood, D., Threapleton, D., Evans, C., et al., 2014. Association between sugar-sweetened and artificially sweetened soft drinks and type 2 diabetes: systematic review and dose-response meta-analysis of prospective studies. Br. J. Nutr. 112, 725–734.
Greyling, A., Appleton, K.M., Raben, A., Mela, D.J., 2020. Acute glycemic and insulinemic effects of low-energy sweeteners: a systematic review and meta-analysis of randomized controlled trials. Am. J. Clin. Nutr. 112, 1002–1014.
Han, P., Bagenna, B., Fu, M., 2019. The sweet taste signaling pathways in the oral cavity and the gastrointestinal tract affect human appetite and food intake: a review. Int. J. Food Sci. Nutr. 70, 125–135.
Hancock, S., Zinn, C., Schofield, G., 2020. The consumption of processed sugar-and starch-containing foods, and dental caries: a systematic review. Eur. J. Oral Sci. 128, 467–475.
Hashem, K.M., He, F.J., Macgregor, G.A., 2019. Effects of product reformulation on sugar intake and health—a systematic review and meta-analysis. Nutr. Rev. 77, 181–196.
Health Canada, 2019. Canada's Dietary Guidelines. Health Canada, Ottawa.
Huang, Y., Kypridemos, C., Liu, J., et al., 2019. Cost-effectiveness of the US Food and Drug Administration added sugar labeling policy for improving diet and health. Circulation 139, 2613–2624.
Imamura, F., O'connor, L., Ye, Z., et al., 2015. Consumption of sugar sweetened beverages, artificially sweetened beverages, and fruit juice and incidence of type 2 diabetes: systematic review, meta-analysis, and estimation of population attributable fraction. BMJ 351.
International Food Information Council Foundation, 2018. 2018 Food & Health Survey.
International Food Information Council Foundation, 2019. 2019 Food and Health Survey.
Jang, H.J., Kokrashvili, Z., Theodorakis, M.J., et al., 2007. Gut-expressed gustducin and taste receptors regulate secretion of glucagon-like peptide-1. Proc. Natl. Acad. Sci. U. S. A. 104, 15069–15074.
Jang, C., Hui, S., Lu, W., et al., 2018. The small intestine converts dietary fructose into glucose and organic acids. Cell Metab. 27, 351–361.e3.
Jebril, M., Liu, X., Shi, Z., et al., 2020. Prevalence of type 2 diabetes and its association with added sugar intake in citizens and refugees aged 40 or older in the Gaza Strip, Palestine. Int. J. Environ. Res. Publ. Health 17, 8594.
Keller, A., Bucher Della Torre, S., 2015. Sugar-sweetened beverages and obesity among children and adolescents: a review of systematic literature reviews. Child. Obes. 11, 338–346.
Khan, T.A., Tayyiba, M., Agarwal, A., et al., 2019. Relation of total sugars, sucrose, fructose, and added sugars with the risk of cardiovascular disease: a systematic review and dose-response meta-analysis of prospective cohort studies. Mayo Clin. Proc. 2399–2414.
Kokrashvili, Z., Yee, K.K., Ilegems, E., et al., 2014. Endocrine taste cells. Br. J. Nutr. 111, S23–S29.
Kuhn, C., Bufe, B., Winnig, M., et al., 2004. Bitter taste receptors for saccharin and acesulfame K. J. Neurosci. 24, 10260–10265.
Laviada-Molina, H., Molina-Segui, F., Pérez-Gaxiola, G., et al., 2020. Effects of nonnutritive sweeteners on body weight and BMI in diverse clinical contexts: systematic review and meta-analysis. Obes. Rev. 21, e13020.
Lei, L., Rangan, A., Flood, V.M., Louie, J.C.Y., 2016. Dietary intake and food sources of added sugar in the Australian population. Br. J. Nutr. 115, 868–877.

Liem, D.G., Mennella, J.A., 2002. Sweet and sour preferences during childhood: role of early experiences. Dev. Psychobiol. 41, 388–395.
Lim, J., Pullicin, A.J., 2019. Oral carbohydrate sensing: beyond sweet taste. Physiol. Behav. 202, 14–25.
Llaha, F., Gil-Lespinard, M., Unal, P., et al., 2021. Consumption of sweet beverages and cancer risk. A systematic review and meta-analysis of observational studies. Nutrients 13, 516.
Lobach, A.R., Roberts, A., Rowland, I.R., 2019. Assessing the in vivo data on low/no-calorie sweeteners and the gut microbiota. Food Chem. Toxicol. 124, 385–399.
Lohner, S., Toews, I., Meerpohl, J.J., 2017. Health outcomes of non-nutritive sweeteners: analysis of the research landscape. Nutr. J. 16, 1–21.
Luger, M., Lafontan, M., Bes-Rastrollo, M., et al., 2017. Sugar-sweetened beverages and weight gain in children and adults: a systematic review from 2013 to 2015 and a comparison with previous studies. Obes. Facts 10, 674–693.
Ma, J., Chang, J., Checklin, H.L., et al., 2010. Effect of the artificial sweetener, sucralose, on small intestinal glucose absorption in healthy human subjects. Br. J. Nutr. 104, 803–806.
Mace, O.J., Affleck, J., Patel, N., Kellett, G.L., 2007. Sweet taste receptors in rat small intestine stimulate glucose absorption through apical GLUT2. J. Physiol. 582, 379–392.
Maersk, M., Belza, A., Stødkilde-Jørgensen, H., et al., 2012. Sucrose-sweetened beverages increase fat storage in the liver, muscle, and visceral fat depot: a 6-mo randomized intervention study. Am. J. Clin. Nutr. 95, 283–289.
Magnuson, B.A., Carakostas, M.C., Moore, N.H., et al., 2016. Biological fate of low-calorie sweeteners. Nutr. Rev. 74, 670–689.
Makarem, N., Bandera, E.V., Nicholson, J.M., Parekh, N., 2018. Consumption of sugars, sugary foods, and sugary beverages in relation to cancer risk: a systematic review of longitudinal studies. Annu. Rev. Nutr. 38, 17–39.
Malik, V.S., Popkin, B.M., Bray, G.A., et al., 2010. Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes: a meta-analysis. Diabetes Care 33, 2477–2483.
Maone, T.R., Mattes, R.D., Bernbaum, J.C., Beauchamp, G.K., 1990. A new method for delivering a taste without fluids to preterm and term infants. Dev. Psychobiol. 23, 179–191.
Margolskee, R.F., Dyer, J., Kokrashvili, Z., et al., 2007. T1R3 and gustducin in gut sense sugars to regulate expression of Na+-glucose cotransporter 1. Proc. Natl. Acad. Sci. U. S. A. 104, 15075–15080.
Marriott, B.P., Hunt, K.J., Malek, A.M., Newman, J.C., 2019. Trends in intake of energy and total sugar from sugar-sweetened beverages in the United States among children and adults, NHANES 2003–2016. Nutrients 11, 2004.
Martyn, D., Darch, M., Roberts, A., et al., 2018. Low-/no-calorie sweeteners: a review of global intakes. Nutrients 10, 357.
Mattes, R.D., Popkin, B.M., 2009. Nonnutritive sweetener consumption in humans: effects on appetite and food intake and their putative mechanisms. Am. J. Clin. Nutr. 89, 1–14.
Mela, D.J., Woolner, E.M., 2018. Perspective: total, added, or free? What kind of sugars should we be talking about? Adv. Nutr. 9, 63–69.
Mennella, J.A., 2014. Ontogeny of taste preferences: basic biology and implications for health. Am. J. Clin. Nutr. 99, 704S–711S.
Miller, P.E., Perez, V., 2014. Low-calorie sweeteners and body weight and composition: a meta-analysis of randomized controlled trials and prospective cohort studies. Am. J. Clin. Nutr. 100, 765–777.
Moraru, C., 2011. Formulating in Sweetness. Prepared Foods.
Moskowitz, H.R., 1970. Ratio scales of sugar sweetness. Percept. Psychophys. 7, 315–320.
Mwenda, V., Mwangi, M., Nyanjau, L., et al., 2018. Dietary risk factors for non-communicable diseases in Kenya: findings of the STEPS survey, 2015. BMC Publ. Health 18, 1218.
Narain, A., Kwok, C., Mamas, M., 2016. Soft drinks and sweetened beverages and the risk of cardiovascular disease and mortality: a systematic review and meta-analysis. Int. J. Clin. Pract. 70, 791–805.
National Health Service, 2018. Sugar: The Facts [Online] [Accessed].
Pan American Health Organization, 2021. Global School Health Survey Results [Online] [Accessed].
Pepino, M.Y., Tiemann, C.D., Patterson, B.W., et al., 2013. Sucralose affects glycemic and hormonal responses to an oral glucose load. Diabetes Care 36, 2530–2535.
Peters, J.C., Beck, J., 2016. Low Calorie Sweetener (LCS) use and energy balance. Physiol. Behav. 164, 524–528.
Pielak, M., Czarniecka-Skubina, E., Trafiałek, J., Głuchowski, A., 2019. Contemporary trends and habits in the consumption of sugar and sweeteners—a questionnaire survey among poles. Int. J. Environ. Res. Publ. Health 16, 1164.
Popkin, B.M., Hawkes, C., 2016. The sweetening of the global diet, particularly beverages: patterns, trends and policy responses for diabetes prevention. Lancet Diabetes Endocrinol. 4, 174–186.
Promdee, L., Trakulthong, J., Kangwantrakul, W., 2007. Sucrose consumption in Thai undergraduate students. Asia Pac. J. Clin. Nutr. 16, 22–26.
Research and Markets Ltd, 2020. Global Stevia Market by Form, by End User, by Distribution Channel, by Region, Industry Analysis and Forecast, 2020–2026.
Rogers, P.J., Appleton, K.M., 2020. The effects of low-calorie sweeteners on energy intake and body weight: a systematic review and meta-analyses of sustained intervention studies. Int. J. Obes. 1–15.
Rogers, P.J., Carlyle, J.A., Hill, A.J., Blundell, J.E., 1988. Uncoupling sweet taste and calories: comparison of the effects of glucose and three intense sweeteners on hunger and food intake. Physiol. Behav. 43, 547–552.
Román, S., Sánchez-Siles, L.M., Siegrist, M., 2017. The importance of food naturalness for consumers: results of a systematic review. Trends Food Sci. Technol. 67, 44–57.
Romo-Romo, A., Aguilar-Salinas, C.A., Brito-Córdova, G.X., et al., 2016. Effects of the non-nutritive sweeteners on glucose metabolism and appetite regulating hormones: systematic review of observational prospective studies and clinical trials. PLoS One 11, e0161264.
Rosenstein, D., Oster, H., 1988. Differential facial responses to four basic tastes in newborns. Child Dev. 59, 1555–1568.
Ruanpeng, D., Thongprayoon, C., Cheungpasitporn, W., Harindhanavudhi, T., 2017. Sugar and artificially sweetened beverages linked to obesity: a systematic review and meta-analysis. QJM 110, 513–520.
Ruiz, E., Rodriguez, P., Valero, T., et al., 2017. Dietary intake of individual (free and intrinsic) sugars and food sources in the Spanish population: findings from the ANIBES study. Nutrients 9.
Ruiz-Aceituno, L., Hernandez-Hernandez, O., Kolida, S., et al., 2018. Sweetness and sensory properties of commercial and novel oligosaccharides of prebiotic potential. LWT 97, 476–482.
Ruiz-Ojeda, F.J., Plaza-Díaz, J., Sáez-Lara, M.J., Gil, A., 2019. Effects of sweeteners on the gut microbiota: a review of experimental studies and clinical trials. Adv. Nutr. 10, S31–S48.
Saraiva, A., Carrascosa, C., Raheem, D., et al., 2020. Natural sweeteners: the relevance of food naturalness for consumers, food security aspects, sustainability and health impacts. Int. J. Environ. Res. Publ. Health 17, 6285.
Shan, Z., Rehm, C.D., Rogers, G., et al., 2019. Trends in dietary carbohydrate, protein, and fat intake and diet quality among US adults, 1999–2016. J. Am. Med. Assoc. 322, 1178–1187.
Sievenpiper, J.L., Khan, T.A., Ha, V., et al., 2017. The importance of study design in the assessment of nonnutritive sweeteners and cardiometabolic health. Can. Med. Assoc. J. 189, E1424–E1425.
Singh, G.M., Micha, R., Khatibzadeh, S., et al., 2015. Global, regional, and national consumption of sugar-sweetened beverages, fruit juices, and milk: a systematic assessment of beverage intake in 187 countries. PLoS One 10, e0124845.
Sørensen, L.B., Vasilaras, T.H., Astrup, A., Raben, A., 2014. Sucrose compared with artificial sweeteners: a clinical intervention study of effects on energy intake, appetite, and energy expenditure after 10 wk of supplementation in overweight subjects. Am. J. Clin. Nutr. 100, 36–45.
Statista, 2020. US Population: Which Brands of Sugar Substitutes Do You Use Most Often? [Graph]. Statista: Statista, Inc.
Statista, 2021a. Estimated Size of the Stevia Market Worldwide From 2020 to 2026, by Category (In Million U.S. Dollars) [Graph]. Statista: Statista, Inc.
Statista, 2021b. Sugar Substitutes and Sweeteners in the US. [Graph]. Statista: Statista, Inc.

Steiner, J.E., 1973. The Gustofacial Response: Observation on Normal and Anencephalic Newborn Infants.
Steinert, R.E., Frey, F., Topfer, A., et al., 2011. Effects of carbohydrate sugars and artificial sweeteners on appetite and the secretion of gastrointestinal satiety peptides. Br. J. Nutr. 105, 1320–1328.
Strategy Online, 2017. Agreement With Statements About Sugar and Artificial Sweeteners Among Consumers in Canada in 2017 [Graph]. Statista: Statista, Inc.
Suez, J., Korem, T., Zeevi, D., et al., 2014. Artificial sweeteners induce glucose intolerance by altering the gut microbiota. Nature 514, 181–186.
Sylvetsky, A.C., Welsh, J.A., Brown, R.J., Vos, M.B., 2012. Low-calorie sweetener consumption is increasing in the United States. Am. J. Clin. Nutr. 96, 640–646.
Sylvetsky, A.C., Gardner, A.L., Bauman, V., et al., 2015. Nonnutritive sweeteners in breast milk. J. Toxicol. Environ. Health 78, 1029–1032.
Sylvetsky, A.C., Jin, Y., Clark, E.J., et al., 2017. Consumption of low-calorie sweeteners among children and adults in the United States. J. Acad. Nutr. Diet. 117, 441–448.
Tan, V.W.K., Wee, M.S.M., Tomic, O., Forde, C.G., 2019. Temporal sweetness and side tastes profiles of 16 sweeteners using temporal check-all-that-apply (TCATA). Food Res. Int. 121, 39–47.
Te Morenga, L., Mallard, S., Mann, J., 2013. Dietary sugars and body weight: systematic review and meta-analyses of randomized controlled trials and cohort studies. BMJ 346.
Thompson, D., 2020. New Dietary Guidelines for Americans Ignore Recommendations on Sugar, Alcohol [Online] [Accessed].
Toews, I., Lohner, S., De Gaudry, D.K., et al., 2019. Association between intake of non-sugar sweeteners and health outcomes: systematic review and meta-analyses of randomized and non-randomized controlled trials and observational studies. BMJ 364.
Tordoff, M.G., Alleva, A.M., 1990. Oral stimulation with aspartame increases hunger. Physiol. Behav. 47, 555–559.
Tsilas, C.S., De Souza, R.J., Mejia, S.B., et al., 2017. Relation of total sugars, fructose and sucrose with incident type 2 diabetes: a systematic review and meta-analysis of prospective cohort studies. Can. Med. Assoc. J. 189, E711–E720.
Tucker, R.M., Tan, S.Y., 2017. Do non-nutritive sweeteners influence acute glucose homeostasis in humans? A systematic review. Physiol. Behav. 182, 17–26.
U.S. Food and Drug Administration, 2021a. Added Sugars on the New Nutrition Facts Label [Online] [Accessed].
U.S. Food and Drug Administration, 2021b. Generally Recognized as Safe (GRAS) Notification Program [Online] [Accessed].
US Department of Agriculture, 2021. Caloric Sweeteners: Per Capita Availability Adjusted for Loss. Food Availability (Per Capita) Data System. US Department of Agriculture.
US Department of Health and Human Services & US Department of Agriculture, 2020. Dietary Guidelines for Americans, 2020–2025 [Online]. Available. https://www.dietaryguidelines.gov/sites/default/files/2020-12/Dietary_Guidelines_for_Americans_2020-2025.pdf. (Accessed 18 February 2021).
Valenzuela, M.J., Waterhouse, B., Aggarwal, V.R., et al., 2021. Effect of sugar-sweetened beverages on oral health: a systematic review and meta-analysis. Eur. J. Publ. Health 31, 122–129.
Vincis, R., Fontanini, A., 2019. Central taste anatomy and physiology. Handb. Clin. Neurol. 164, 187–204.
Walbolt, J., Koh, Y., 2020. Non-nutritive sweeteners and their associations with obesity and type 2 diabetes. J. Obes. Metab. Syndr. 29, 114.
Wang, M., Yu, M., Fang, L., Hu, R.Y., 2015. Association between sugar-sweetened beverages and type 2 diabetes: a meta-analysis. J. Diabetes Investig. 6, 360–366.
Witherly, S.A., Pangborn, R.M., Stern, J.S., 1980. Gustatory responses and eating duration of obese and lean adults. Appetite 1, 53–63.
World Health Organization, 2015. Guideline: Sugars Intake for Adults and Children. World Health Organization, Geneva.
World Health Organization, 2017. Report on Fiscal Policies to Reduce Consumption of Sugar-Sweetened Beverages and Other Regulatory Measures to Promote Healthy Diets in the Republic of Maldives. World Health Organization, Geneva, Switzerland.
World Health Organization, 2021. Reducing Free Sugars Intake in Children and Adults [Online] [Accessed].
Wu, T., Zhao, B.R., Bound, M.J., et al., 2012. Effects of different sweet preloads on incretin hormone secretion, gastric emptying, and postprandial glycemia in healthy humans. Am. J. Clin. Nutr. 95, 78–83.
Xi, B., Li, S., Liu, Z., et al., 2014. Intake of fruit juice and incidence of type 2 diabetes: a systematic review and meta-analysis. PLoS One 9, e93471.
Xi, B., Huang, Y., Reilly, K.H., et al., 2015. Sugar-sweetened beverages and risk of hypertension and CVD: a dose-response meta-analysis. Br. J. Nutr. 113, 709–717.
Yeung, A.W.K., Wong, N.S.M., 2020. How does our brain process sugars and non-nutritive sweeteners differently: a systematic review on functional magnetic resonance imaging studies. Nutrients 12, 3010.
Zhao, H., Zhou, Y., Pinto, C.M., et al., 2010. Evolution of the sweet taste receptor gene Tas1r2 in bats. Mol. Biol. Evol. 27, 2642–2650.

Traditional and indigenous foods for food security and sovereignty

Emily V. Merchant[a,b] **and James E. Simon**[a,b], [a] New Use Agriculture and Natural Plant Products Program, Department of Plant Biology, Rutgers University, New Brunswick, NJ, United States; and [b] Center for Agricultural Food Ecosystems, The New Jersey Institute for Food, Nutrition and Health, Rutgers University, New Brunswick, NJ, United States

© 2023 Elsevier Ltd. All rights reserved.

Introduction 639
Four pillars of food security and food sovereignty 642
Availability 642
Availability of AITFs in the natural and built environment 642
Availability of agricultural innovations for producers 642
Affordability and accessibility 643
Affordability and accessibility at the market 643
Pathways between production and food security 644
Acceptability 644
Nutrition 644
Enabling environments to promote consumption and production of AITFs 645
Sustainable food ecosystem 645
Tradition and indigenous foods promote ecosystem benefits and genetic diversity 646
Conclusion 646
Acknowledgment 646
References 647

Key points

- African, EurAsian, American, and Australian indigenous and traditional foods (AITFs) around the globe are an undervalued, under-recognized, under-leveraged source of nutrients that, depending upon the specific species and landrace, would be considered "high sources" of one or more nutrients (e.g., vitamins and minerals), which are often limited in the same region. A promotion of their production and consumption could help mitigate malnutrition and food insecurity.
- AITFs, often wild foraged, are generally resilient to changing environmental conditions, and therefore already adapted to the biotic and abiotic stressors found in the area, providing climate smart crops for consumption and production.
- AITFs can contribute toward increasing biodiversity among crop species globally, providing increased solutions to climate sensitivity and subsequently household hunger and food security.
- AITFs are a diverse group of plants that include annuals, perennials, and legumes. Some of these plants, such as the legumes, can fix nitrogen, enhance soil health, and reduce erosion, therefore contributing to environmental protection all while providing communities with fiber and food for household consumption and animal feed.
- AITFs can be used as economic drivers for job creation and income generating opportunities among vulnerable populations (e.g., women, youth) complimenting rather than replacing traditional agronomic staple crops.
- AITFs can be incorporated more formally into agro-forestry systems to enhance economic, environmental, and cultural sustainability.
- A multitude of strategies are urgently needed to address the core issues of food security and sovereignty. The promotion of AITFs could play a significant role in ensuring that nutrient rich foods are more available, affordable, accessible, and acceptable and adopted for production and household consumption.

Introduction

The world faces a formidable challenge: to feed an increasingly hot and hungry planet where extreme events such as the COVID-19 pandemic and climate shocks can contribute to household hunger and food insecurity. Despite some achievements in meeting the global food demand, the prevalence of malnutrition remains high in some low-to-middle income countries in sub-Saharan Africa (SSA), South Asia, and parts of Central and South America. In 2020, over 50% of the children under 5 were stunted in Asia (FAO, 2021). In addition to undernutrition, many low-to-middle income countries are facing the double burden of malnutrition, at both the household and individual level (Hoffman, 2019; Popkin et al., 2020). Disruptions to the global food supply chain, such as those experienced during the COVID-19 pandemic are expected to worsen under climate change.

Climate change is associated with increased temperatures, atmospheric carbon dioxide, and extreme weather events. Through several causal pathways, it is predicted that climate change will have a negative impact on global nutrition outcomes. Furthermore, the intensity of monoculture farming (e.g., land clearing, high use of irrigation, fertilizer, and pesticide application) is a threat to ecosystem services and a contributor to climate change. In return, climate change and the decline of ecosystem services (e.g., decline in pollinator species, carbon sequestration, pest control) impact agriculture production, creating a vicious cycle where agriculture and the ecosystem are at constant odds (Sunderland et al., 2019).

Agroecological issues are further exasperated by the production of fewer food crops. The availability of food supply worldwide is dependent on only a few crop species or "major crops." It is estimated that only 30 plant species, often widely and intensively cultivated, are used to meet 95% of the world's food requirements (Leff et al., 2004). Declining biodiversity among fruit and vegetable production worldwide is severely limiting the available food options for a sustainable, healthy food supply. There is a real need to increase biodiversity among food crops and to provide a more diverse diet to feed the ever-growing human population while avoiding dependency on a few major crops, especially under climate change.

The promotion of neglected and underutilized African, EurAsian, American, and Australian indigenous and traditional foods (AITFs) around the globe is one way that agriculture can diversify to better serve human nutrition and ecosystem needs. AITFs either originated in their respective continent or have a long history of being accidently or intentionally introduced and became naturalized into a new region. Either way, these plants are often wildcrafted and in some cases brought into cultivation and domesticated to grow locally and have become accepted through custom, habit, or tradition. Examples include African nightshade, amaranth, hibiscus, moringa, and spider plant (Li et al., 2020; Towns and Shackleton, 2018). While many of these plants may be considered neglected or underutilized on a global level (e.g., limited international trade or commerce, lack of scientific crop improvement research, little investment and, funding in their commercialization) they are important and highly valued on a local level. AITFs are often culturally preferred (Hoffman et al., 2018; Hunter et al., 2019; Simon et al., 2020, 2021) and nutritionally dense (Abukutsa-Onyango et al., 2010). Furthermore, AITFs are adapted to local environmental conditions and some are even considered "survivor plants" due to their tolerance to temperature and precipitation extremes (Chivenge et al., 2015).

Although AITFs hold significant potential to reduce global malnutrition while strengthening sustainable food ecosystems, they are not without their challenges. In some parts of the world, AITFs are grown as subsistence crops or wild foraged, and have historically carried the social stigma of "famine foods," which negatively impacts their adoption, promotion, and consumption (Towns and Shackleton, 2018). Addressing this stigma is particularly important in vulnerable communities that may not have access to affordable "Western" introduced vegetables (Hunter et al., 2019). Moreover, the success of any commodity is dependent on support from and linkages into the larger agricultural system, or in this context the food environment. Downs et al. (2020) defines the food environment as: "the consumer interface with the food system that encompasses the availability, affordability, convenience, promotion and quality, and sustainability of foods and beverages in wild, cultivated, and built spaces that are influenced by the sociocultural and political environment and ecosystems within which they are embedded." Unfortunately, AITFs are often in disabling food environments that do not promote their widespread production and trade (Ghosh-Jerath et al., 2021). For example, it is common for farmers to shift land allocation from household gardens growing nutritious traditional foods, toward commercial cereal crops for retail-sale (Abdoellah et al., 2020). In contrast to the "conventional food crops", the AITFs have not had the historical support and recognition by governments, international agencies, and research centers, which has drastically limited research, extension, industry, and public support.

Despite these challenges, AITFs possess strengths and opportunities such as significant unmet economic potential (Simon et al., 2020, 2021). A holistic enabling food environment that utilizes a systems approach is needed for global food security and food sovereignty under climate change conditions. Market-first, science-driven solutions need to incorporate a community resilience framework that rebuilds local economies, regenerates ecosystems, and mitigates climate impact (Simon et al., 2021; Weller et al., 2015). Models such as the Food-systems for Empowerment, Economic Development and Sustainability (FEEDS) (**Fig. 1**) have produced transformative changes across many regions and communities. This model engages farmers and consumers directly in the formulation of the research agenda by combining concepts from food security and food sovereignty discourse. Furthermore, this model actively engages farmers and consumers in the process of developing and disseminating technological innovation through hands-on demonstrations and storytelling that captures shared experiences, subsequently strengthening local community resiliency and improving nutrition outcomes.

"Double-duty" solutions that combine the food security and food sovereignty discourses are needed to create a global agricultural food ecosystem that harmonizes the environment and the world's food supply (Jarosz, 2014). While the food security and food sovereignty discourses share some similarities, namely that their primary goal is ensuring that all people have reliable access to sufficient quantity and quality of affordable, nutritious food (**Table 1**); it is important to look at their historical agendas. The first official definition of "food security" in 1974 was "the availability at all times of adequate world food supplies of basic foodstuffs to sustain a steady expansion of food consumption and to offset fluctuations in production and prices" (United Nations, 1975 cited in FAO, 2001). While today, these agendas are similar at heart, their primary difference can be seen in the historical definition: scale. Food security often includes a primary research and development agenda around exports and global markets with an emphasis on commercial farming. In contrast, food sovereignty tends to focus on small-holder farmers and is often regarded as a "peasant movement." The development work, research, and discourse around both terms are critical to feeding our global population and research and development work must commit to all scales of the food system.

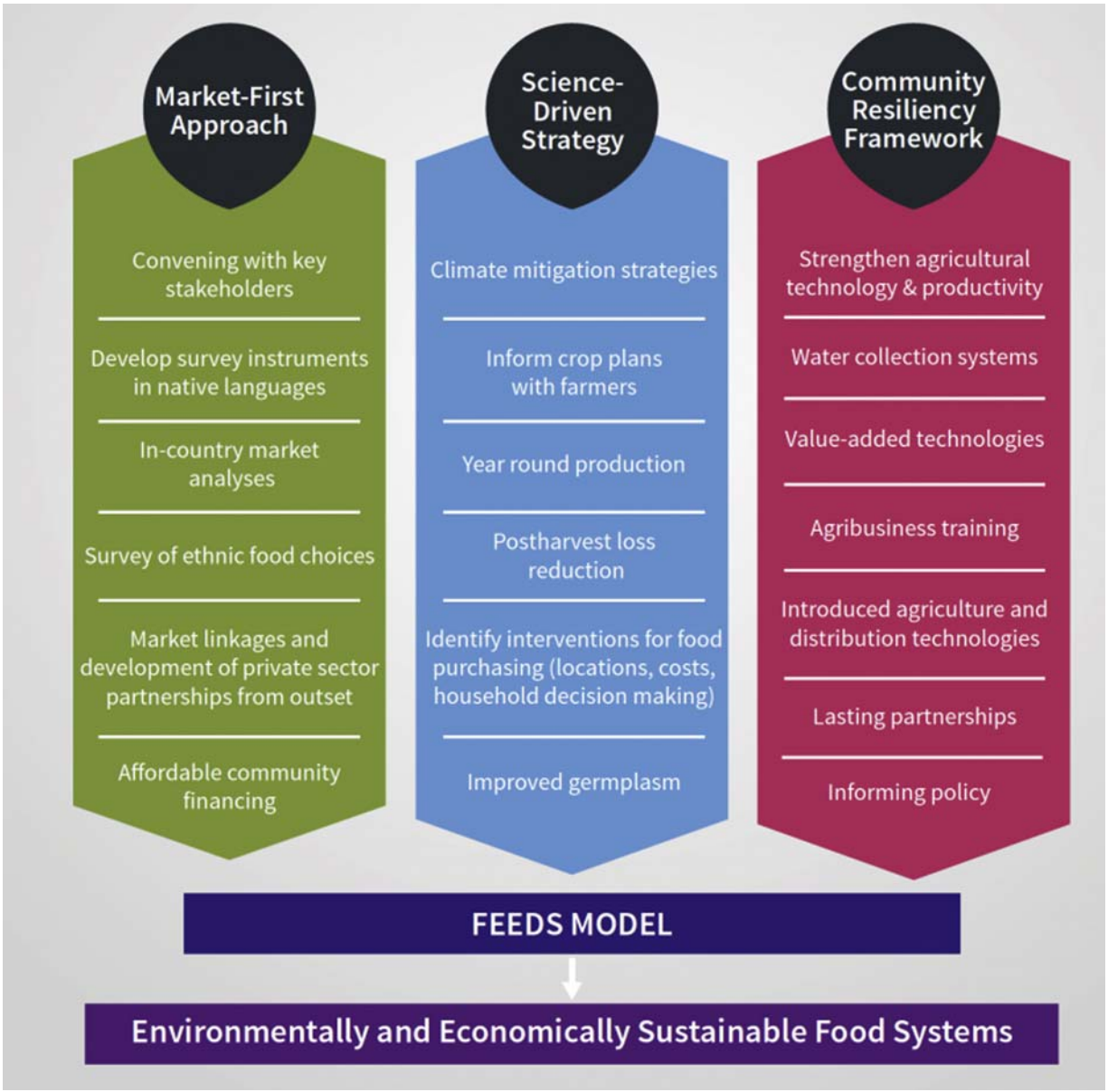


Fig. 1 Food-systems for Empowerment, Economic Development and Sustainability (FEEDS) model for global food security and food sovereignty (Simon et al., 2021).

Table 1 Current or widely used definition of food security and food sovereignty.

Term	*Current or widely used definition*	*Defining agency*
Food security	All people, at all times, have physical, social and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life. With the four pillars of food security: availability, access, utilization, and stability.	Food and agriculture Organization of the United Nations (FAO, 2001)
Food sovereignty	The right of peoples to healthy and culturally appropriate food produced through ecologically sound and sustainable methods, and their right to define their own food and agriculture systems.	La Vía Campesina (Patel, 2009)



Fig. 2 The four pillars of food security and food sovereignty within a sustainable food ecosystem.

Our paper will use the dual concept of food security and sovereignty as a framework for the promotion of consumption and production for various markets (e.g., local, national, international) of indigenous and tradition foods in EurAsia, Africa, the Americas, and Australia through four pillars: availability, affordability, accessibility, and acceptability (Fig. 2). We will also discuss and illustrate how AITFs serve as one pathway for a sustainable food ecosystem.

Four pillars of food security and food sovereignty

Availability

For AITFs to be affordable, accessible, and acceptable, they first need to be available. Availability is defined as a situation when there are sufficient quantities of AITFs, in the built or natural food environment, for household consumption and/or the availability of improved and appropriate technologies and inputs to produce them. This section examines the availability of traditional and indigenous foods within the food environment as well as the availability of agricultural innovations for producers.

Availability of AITFs in the natural and built environment

The availability of AITFs is critical for food security and sovereignty for many vulnerable populations such as rural, indigenous, and aboriginal communities. Ferguson et al. (2017) reported that among key informants traditional foods significantly contributed to dietary intake, particularly during food insecure times, among aboriginal communities in Australia. While these food types are culturally significant and contribute to healthy dietary intake, their availability, and often affordability, is influenced by factors in the natural and built environment. For example, seasonality and climate change variability can impact the varieties and yield of fruits and vegetables available, both on the farm and in the marketspace (Ochieng et al., 2016). While AITF production is vulnerable to environmental stressors, as a group of plants, they tend to be less sensitive compared to "Western" introduced crops, stabilizing the availability of a variety of AITFs throughout the year.

In addition to the natural environment, aspects of the built environment can further influence the availability of AITFs. A review by Hunter et al. (2019) reported that low recognition of cultural and market value can hinder the availability of AITFs, while limited and fragmented data can restrict the availability of technological and germplasm advances and subsequent training for smallholder farmers. In addition, limited capacity, due to poorly developed infrastructure as well as disabling policies that restrain the production of AITFs, can further limit the availability and increase risk along the value-chain. Despite these challenges, the promotion of cultivation or wild harvesting of AITFs through extension and other outreach services can increase production and harvest and subsequently availability for consumers and smallholder farmers.

Availability of agricultural innovations for producers

Empowering smallholder farmers with innovations in agriculture such as Good Agricultural Practices (GAP) (e.g., hands-on season-long training in soil preparation, fertilization, water management and irrigation, planting, harvest methods, and more); post-harvest handling that reduces food waste and maximizes post-harvest processing; and improved seeds can promote the growth and availability of high-quality products for rural communities worldwide. This concept is further exemplified when utilized in a market-first science-driven model that combines biocultural knowledge with scientific advances. Moreover, it is essential that these methods are

embedded into the engagement with farmers and subsequently accepted, adopted, and championed by the communities involved for successful and sustainable development. Despite continued population growth, prior to the COVID-19 pandemic, agricultural production was largely meeting global food demand, with the global prevalence of undernutrition declining (FAO, 2021). This achievement was driven largely by technological innovations such as improved germplasm resulting in higher-yielding higher nutrition agronomic and horticulture varieties; synthetic and organic high performing fertilizers and pesticides; and innovative low- and high-tech farm tools such as irrigation systems, tractors, and solar dryers (Byrnes et al., 2017; Simon et al., 2020). Agricultural innovations and subsequent trainings coupled to increase scientific understanding of the drivers and methods that promote sustainable longer-term behavior change, must be made available to farmers on a continual basis, just as they are in developed countries, to be maximized by producers.

Availability of extension or other outreach services and intervention trainings impact the adoption of advanced agronomic practices and technologies improving yield and in turn increasing affordability and accessibility of AITFs to consumers. For example, glutenous rice plays an important role in food and cultural security among smallholder farmers in rural Vietnam. A study by Sattaka et al. (2017) found that extension services, such as extension official site visits, demonstration plots, trainings, and multi-media messaging, played a key role in fostering the sustainable production of culturally preferred glutinous rice and helped to ensure local food and cultural security in Vietnam. While most of the farmers in the study were smallholder, it is essential that policy and programming promote the availability of improved agronomic practices and germplasm for both small- and large-scale farmers.

The availability of improved germplasm and seed stock for producers can improve production and yield and increase supply, having a direct effect on nutrition delivery to consumers. The genetic diversity among AITFs suggests that targeted breeding programs can enhance the agronomic and nutritional characteristics of plants (e.g., yield, flavor, and nutritional quality). For example, Byrnes et al. (2017) found that there was a genotype effect on iron content among several species of *Amaranthus*. The selective breeding efforts in the study led to the development of an *Amaranthus* varietal with relatively high iron. Yet, the selection and breeding of improved germplasm of AITFs as conducted by the World VegCenter, the leading international center in vegetables, can be far more impactful with a stronger commitment by the commercial seed industry and other actors along the value chain that can effectively disseminate, maintain, and make available the genetic materials to farmers. Market-first, science-driven breeding programs can narrow the information gap for effective varietal development serving the goals of both farmers and stakeholders, promoting the affordability, accessibility, and acceptability of indigenous and traditional vegetables, to reduce micronutrient deficiencies. The availability of agricultural innovations can promote the affordability and accessibility of AITFs. The inclusion of grower/farmer participation in the final selection of those AITFs that would be released is also helpful and leads to greater success in the new varieties adoption by those that participated in their field performance evaluation prior to its commercial release.

Affordability and accessibility

Affordability and accessibility of AITFs is defined by a situation where individuals have the needed resources to obtain adequate quantities of desired and culturally preferred AITFs economically and physically (respectively) and/or the required and appropriate inputs and technologies to produce them. This section examines the affordability and accessibility of AITFs in the market as well as how increased household production can improve food security and food sovereignty through two major pathways.

Affordability and accessibility at the market

A lack of economic affordability and market access is a major contributor to food insecurity and loss of food sovereignty. Approximately 25% of the world's population is classified as food insecure, and this number is often higher in vulnerable communities. A study by Lambden et al. (2006), in Arctic Canada, found that 40–70% of the indigenous women surveyed, from 44 tribes, could not afford enough food. In addition, food security and sovereignty are further exacerbated by crizes such as the COVID-19 pandemic. Food environment disruptors such as the extreme droughts and locust infestation, such as those currently afflicting east Africa, can cause market disruptions subsequently limiting access and affordability of nutritious foods (FAO, 2021).

The food security and food sovereignty discourse extend beyond sufficient caloric intake and considers the quality of calories ultimately linking the affordability of healthy diets with nutrition outcomes. In 2020, FAO (2021) estimated that 3 billion people, or 42% of the global population, was unable to afford a healthy diet. With a majority of communities in lower-and-middle income countries unable to afford adequate quantities of healthy foods, it is of no surprise that economic affordability is a significant barrier to culturally preferred nutritious foods.

While affordability and accessibility may be a barrier to purchase and consume AITFs, the promotion of household production may assist in increasing household consumption. In low-to-middle income countries, smallholder farming remains a primary livelihood for households, and therefore many rural consumers may serve a dual role also as producers. Nutrition-sensitive agriculture has demonstrated positive contributions to health outcomes for women and children. Constrained affordability and accessibility can limit an individual's ability to acquire production inputs and advanced technologies. Fortunately, AITFs often require fewer inputs compared to introduced vegetables. Since AITFs originated or have a long history of cultivation in the given locale, they are adapted to the local environmental conditions. Thus, this large group of under-utilized crops can provide real benefits to many environmentally vulnerable regions where increased heat, drought and biotic stressors associated with climate change are already noted and where local climate-resilient, culturally preferred, nutritious crops will become even more important. Furthermore, research and development initiatives have demonstrated the success of disseminating innovative production methods and technology through farmer engagement in extension settings (Sattaka et al., 2017; Simon et al., 2020, 2021).

Pathways between production and food security

Adequate production of nutritionally dense fruits, vegetables, and grains can address household food and nutrition insecurity through two causal pathways: (1) improved access to nutrient-rich foods; and (2) improved household finances either through generating income from the sale of produce or saving income from food expenditures (Korth et al., 2014).

AITFs are traditionally wild harvested or self-produced using home gardens. In both rural or urban settings, AITFs production in home gardens can contribute to household food security and sovereignty, and ecological sustainability through household and community autonomy within the food systems. A study by Márquez and Schwartz (2009) found that home gardens in Petén, Guatemala, which are highly diverse, rich, and productive, contributed to household income, nutrition delivery, and strengthened social bonds and networks. In addition to rural agriculture, urban agriculture can contribute to improved food security and food sovereignty even with limited space. A study by Gallaher et al. (2013) reported that sack gardening in urban communities increased social capital and alleviated household food insecurity improving household dietary diversity, particularly during times of food shortages.

In addition to direct consumption, household-level production of fruits and vegetables impact household economics by decreasing grocery expenditure and generating income from sales. AITFs are popular, highly valuable produce with unmet market potential. A study conducted by Senyolo et al. (2014) reported that due to their historical importance, demand and preference for African ITFs were high among survey participants. In addition, they found that most individuals (80%) were willing to pay a premium for African ITFs. Regional significance is important and can support production among rural and urban producers.

Furthermore, some AITFs, such as quinoa, have successfully expanded to the global market. Access to both regional and export markets is often dependent on supply chain infrastructure (e.g., processing, storage, transportation). Furthermore, as the global popularity for specialty crops increases the balance between production for household consumption and sales can shift favoring income generation over household food security. As research and development frameworks are designed for these markets and commodities, it is important to prioritize food sovereignty among producers.

As demand for AITFs in informal and formal markets increases, it is important that interventions and extension agents continue to emphasize production for household consumption. Furthermore, increased household production is only one pathway to increase the affordability and accessibility of nutritious diets. Affordability and accessibility hinge on stability in the food environment (Downs et al., 2020). Extreme climate events or poor seed stock can cause crop failure or low yields further driving demand and price, limiting the availability of food options for purchase. Moreover, political unrest or regional shocks, such as the COVID-19 pandemic, may limit an individual's ability to access the markets. Hence, policy-level change needs to be enacted to fully address this issue.

AITFs are often in disabling environments that do not promote the production and consumption of these crops. This can further hinder the acceptability of available and accessible agricultural innovations and culinary advances (Ghosh-Jerath et al., 2021). Moreover, in communities that celebrate AITFs with pride, advances may meet cultural resistance in a desire to adhere to traditional norms (Ochieng et al., 2019). To promote acceptability, market-first science-driven solutions need to incorporate a community resilience framework such as FEEDS, which involves farmers and consumers directly in the formulation of the research agenda and actively engages them in the process of technological innovation and dissemination through hands-on demonstrations and storytelling (**Fig. 1**) (Simon et al., 2021).

Acceptability

Acceptability of AITFs is defined by a situation when the available nutritious, culturally preferred foods are of sufficient quality (e.g., nutritional, freshness, and other attributes as defined by the community) and/or the production inputs and practices are of sufficient quality, and both are subsequently adopted. In addition to acceptability at the consumer level, it is also essential that AITFs exist in an enabling environment that promotes the production and consumption of AITFs. This section will examine how the superior quality (e.g., nutrition) in general of AITFS, compared to "Western" crops, and an how enabling environment can contribute to the acceptability of traditional and indigenous foods.

Nutrition

AITFs often contain significantly higher macro-, and micronutrients compared to "Western" introduced crops. This section provides examples of the general health attributes of AITFs and examples of how processing and value-added production can diversity and alter the nutritional properties.

AITFs provide valuable micro-and macronutrients such as vitamins, minerals, carbohydrates, proteins, fats, and bioactive compounds such as antioxidants (Hoffman et al., 2018; Hunter et al., 2019). In Africa, Abukutsa-Onyango et al. (2010) found that priority ITFs (African nightshade *Solanum scabrum*, amaranth *Amaranthus sp.*, spider plant *Cleome gynandra*, cowpea *Vigna unguiculata*, pumpkin leaf *Cucurbita moschata*, slender leaf *Crotalaria sp.*, and jute mallow *Corchorus olitorius*) contained over 70–80% of the recommended daily intake of iron and 18–54% of the recommended daily intake of protein. Furthermore, Ghosh-Jerath et al. (2021) found that traditionally consumed grains, leafy-vegetables, and fungi were rich in macro- and micronutrients such as protein, calcium, iron, folate, Vitamin A and C. For example, the commonly consumed green leafy-vegetable *Crotalaria juncea*, was found to be a high source of calcium and Vitamin A.

To minimize culinary monotony, researchers have assessed more nuanced aspects of nutrition delivery through recipe development and variation in preparation styles. This research is essential to provide communities with more targeted interventions that maximize nutrition delivery. A study by Habwe et al. (2009) reported that cooking (e.g., boiling, pan-frying) traditional leafy

vegetables significantly increased the iron content compared to raw consumption. Moreover, the study found that pan-frying the vegetables with complementary ingredients such as tomatoes and onions significantly increased the iron content. This supports the use of traditional recipes that serve green leafy-greens along-side other vegetables for a complete side-dish. Even more importantly, this study found that boiling the vegetables with lye, which is traditionally done to soften the fibrous leafy greens, significantly decreased the iron content. Context specific nutrition education should focus on the promotion of utilizing complimentary vegetables while minimizing the use of lye for preparation to maximize iron delivery.

Given that numerous AITFs are wild harvested or have little to no domestic cultivation, there may be some concerns as to their potential anti-nutritional properties such as concentrations of alkaloids, oxalates, glycoalkaloids, and saponins. While studies have affirmed the relatively low presence of these compounds, more research is needed in this area as additional value-added products are explored for market sales (Yuan et al., 2019, 2020). For example, varieties of nightshade, which are native to EurAsia, are commonly consumed in South America and Africa where it was introduced. In South America, it is common to consume only the berries and not the leaves, while in Africa only the leaves are consumed and not the berries. The berries contain high levels of glycoalkaloids, anti-nutritional compounds, that if not carefully prepared (e.g., cooked) can be harmful. Furthermore, through selective breeding, a cultivar with relatively low concentrations of these compounds could become available creating "new" products for the AITF supply chain (Yuan et al., 2019) while the "toxic" varieties may have their own value-added application such as uses in pesticides and dye. In addition to consumer acceptability, AITFs must exist in an enabling environment that promotes the production and consumption of AITFs.

Enabling environments to promote consumption and production of AITFs

The promotion and expansion of enabling environments that support the production and consumption of AITFs is critical to increasing the acceptability of AITFs at all levels of the food environment (e.g., local, national, global). Furthermore, enabling environments can help to mitigate barriers to household availability through production for consumption and sales; the accessibility of quality seeds and inputs; affordability and competitive market value; policies that promote the production and consumption of AITF; and women's empowerment.

An innate contribution to the enabling environment for AITFs is their food heritage. ITFs have a long history of cultivation and use in their respective areas with women often serving as the custodians of the cultural knowledge (Hunter et al., 2019). The promotion of AITFs through historical legacy and pathways, such as storytelling, can contribute to cultural security and reinvigorate the use of AITFs, providing a sense and source of "cultural pride" (Simon et al., 2021). These traditional practices also support ecological services and resources expanding the scope of the enabling environment beyond politics toward sustainability and food ecosystem wellness. However, it is of note that not all AITFs evoke thoughts of cultural pride. In some parts of the world, IFTs are grown or wild foraged as a coping mechanism for food shortages and carry the social stigma of "famine foods" (Towns and Shackleton, 2018). In addition, social folklore can contribute to hesitancy around consumption. For example, crops with dark purple pigmentation (e.g., some amaranths in East Africa) can be incorrectly associated as being poisonous. An enabling environment may help retell the story of AITFs under a different lens that promotes cultural pride.

Governments that support the promotion of AITFs, often through the establishment of initiatives and policy, foster an enabling environment through the development of a clear framework as well as establishing incentives and accountability that integrate AITFs at all levels. To promote the acceptability of AITFs, interventions must balance local needs (e.g., economic, nutritional), market capacity, and cultural practice by enhancing existing community interests and networks.

The Biodiversity for Food and Nutrition Project (BFN) is an example of a multinational program that mainstreamed biodiversity conservation and sustainable use of AITFs for improved nutrition and well-being in the local, national, and global food system. As summarized by Hunter et al. (2019) BFN is based in Brazil, Turkey, Sri Lanka, and Kenya and programmatic work was often built upon pre-existing national initiatives. In Brazil, the work built on several national initiatives such as, but not limited to, *National School Meals Program* and *Food Acquisition Program*, which mandated that school meals are partially soured from family farmers and paid an incentive for organic or agroecological produce from smallholder farmers.

Community participatory buy-in is important to promote policy-level change. Nutrition-sensitive agriculture interventions, at the household and community level, that take into consideration gender disparities may further promote an enabling environment for AITFs. Women serve as a major custodian in traditional knowledge and food practices. Furthermore, when women control the resources, they are more likely to use it for family consumption improving nutrition outcomes. A study by Ruel et al. (2010) found that nutrition-sensitive agriculture that contains complementary programming such as women's empowerment as well as sanitation, health, and water, had a greater impact on nutrition outcomes for women and children. Empowering women to use their food heritage and knowledge to produce foods for household consumption and sales can also extend the scope of the "enabling environment" to include sustainability and food ecosystem wellness, as often these behaviors are coupled, both consciously and as proxy, with protecting and promoting ecological services. Furthermore, linking nutrition-sensitive interventions and cultural pride with new and/or improved acceptable ways of preparing foods can create a more inclusive and sustainable approach to nutrition behavior-change.

Sustainable food ecosystem

Current industrial agricultural practices, climate change, and the decline of ecosystem services impact agriculture production. In the global agricultural system, genetic diversity is decreasing (Leff et al., 2004). In addition, the intensity of monoculture farming (e.g.,

large-scale land clearing, irrigation, fertilizer, and pesticide application) is a threat to many ecosystem services (e.g., pollination, pest control, and regulation of disease vectors) and a contributor to climate change. The global food supply chain's ability to feed the growing population depends on high amounts of the earth's natural resources such as land and water. The incorporation of AITFs in sustainable food ecosystems can contribute toward diverse, productive, and resilient agroecosystems (Altieri et al., 2012). This section will examine how AITFs can promote ecosystem services and the importance of genetic diversity for a stable food system.

Tradition and indigenous foods promote ecosystem benefits and genetic diversity

The incorporation of AITFs into agricultural systems poses several ecosystem benefits. AITFs have a low carbon footprint since are often gathered from cultivated or wild environments in traditional systems with minimal inputs compared to "Western" introduced crops. This is particularly important given that AITFs are adapted to marginal, complex, and difficult environments; environments that would require considerable inputs if they were used to cultivate "Western" introduced crops. In addition, traditional production systems often incorporate aspects of environmental management such as minimizing overharvesting and exploitation of crops as well as environmental benefits such as nitrogen-fixation (Sunderland et al., 2019). Furthermore, there is high genetic diversity among AITFs.

Agriculture diversification (both between and within species) can improve ecosystem services particularly during times of increased environmental stress (Altieri et al., 2012). Increased genetic diversity creates species redundancy, allowing stronger species to withstand environmental stress while sensitive species are lost from the system. This is particularly important for both human and environmental health. As climate variability continues to worsen, diverse genetic stock is critical for breeding nutritious, climate-smart, crops. In addition, crops, particularly those in wild and minimally cultivated environments serve many ecosystem niches such as but not limited to food and habitat. Diversified agricultural systems have been demonstrated to maintain greater biodiversity with high quality ecosystem services. Such climate "smart" crops may also provide greater benefits to communities during seasons and years of intense drought, floods, and/or pest infestations when the commodity crops may fail. This not only has positive ecological outcomes but also contributes to positive economic and social implications.

Conclusion

The success of holistic models that incorporate African, EurAsian, American, and Australian indigenous and traditional foods (AITFs) demonstrates the need to combine concepts from food security and food sovereignty and focus research and development work at all levels of the food system. These AITFs provide an undervalued, under-recognized, under-leveraged source of nutrients (in addition to tastes and flavors) that can be used as economic drivers for job creation and create income generating opportunities among vulnerable populations (e.g., women, youth) complimenting rather than replacing traditional agronomic staple crops, grown so often by smallholder farmers. In addition to human health, AITFs offer climate and environmental sustainability solutions. Overall, AITFs are adapted to the biotic and abiotic stressors of the area and resilient to changing environmental conditions, providing climate smart crops for collection and production. AITFs are a diverse group of plants which include annuals, perennials, and legumes. Some of these plant types can facilitate nitrogen fixation, enhance soil health, reduce erosion, and contribute to increased biodiversity therefore providing environmental protection particularly when incorporated into agro-forestry systems to enhance economic, environmental, and cultural sustainability. Approaches that blend modern agricultural science and indigenous knowledge systems are proving to enhance food security while conserving agrobiodiversity and ecosystem services throughout the world. A multitude of strategies are urgently needed to address the core issues of food security and sovereignty. The development of AITFs need champions as these crops could have serious and contributing roles to ensuring that nutrient rich foods are available, affordable, accessible, acceptable, and adopted for regular home consumption and production.

Acknowledgment

This work has been guided by our work over many years and thus is a synthesis of observations incorporated into the models under which we work. Prior research was made possible by the generous support of the American people through the US Agency for International Development (USAID). The contents are the responsibility of the Horticulture Innovation Lab and do not necessarily reflect the views of USAID or the US Government. This research was in part supported by the Horticulture Innovation Lab with funding from USAID (EPA-A-00-09-00004), as part of the US Government's global hunger and food security initiative, Feed the Future, for project titled "Improving nutrition with AIVs" in eastern Africa (Feed the Future, 2014). We thank Agilent Instruments, Shimadzu Instruments, the New Jersey Agricultural Experiment Station (HATCH project NJ12170), The New Jersey Institute for Food, Nutrition and Health and its Center for Agricultural Food Ecosystems (RUCAFE) and Center for Childhood Nutrition and Health; the World Vegetable Center (WorldVeg). We also thank the Academic Model Providing Access to Health Care (AMPATH), University of Eldoret, Moi University, University of Zambia, Cuttington University, Kenyan Agricultural and Livestock Organization (KALRO), AgriSmart-Zambia, Hantambo Women's Association, the Mitengo Women's Growers Association, and the Mirror of Hope CBO. We would also like to thank Drs. Steve Weller, Daniel Hoffman, Ramu Govindasamy, Xenia Morin, Rodolfo Juliani, and Shauna Downs. In addition we thank Dan Acquaye, Jerry Brown, Carol Wilson, John Bowman and Howard Shapiro as well as the many friends, faculty colleagues and importantly the communities and farmers with whom we have had the priviledge of working as each who contributed to the success of our collaborative projects.

References

Abdoellah, O.S., Schneider, M., Nugraha, L.M., Suparman, Y., Voletta, C.T., Withaningsih, S., Parikesit, A.A., Heptiyanggit, A., Hakim, L., 2020. Home garden commercialization: extent, household characteristics, and effect on food security and food sovereignty in Rural Indonesia. Sustain. Sci. 15, 797–815.

Abukutsa-Onyango, M., Kavagi, P., Amoke, P., Habwe, F., 2010. Iron and protein content of priority African indigenous vegetables in the Lake Victoria basin. Afr. J. Food Sci. 3 (12), 393–397.

Altieri, M.A., Funes-Monzote, F.R., Petersen, P., 2012. Agroecologically efficient agricultural systems for smallholder farmers: contributions to food sovereignty. Agron. Sustain. Dev. 32, 1–13.

Byrnes, D.R., Dinssa, F.F., Weller, S.C., Simon, J.E., 2017. Elemental micronutrient content and horticultural performance of various vegetable amaranth genotypes. J. Am. Soc. Hortic. Sci. 142, 265–271.

Chivenge, P., Mabhaudhi, T., Modi, A.T., Mafongoya, P., 2015. The potential role of neglected and underutilised crop species as future crops under water scarce conditions in sub-Saharan Africa. Int. J. Environ. Res. Public Health 12, 5685–5711. https://doi.org/10.3390/ijerph120605685.

Downs, S.M., Ahmed, S., Fanzo, J., Herforth, A., 2020. Food environment typology: advancing an expanded definition, framework, and methodological approach for improved characterization of wild, cultivated, and built food environments toward sustainable diets. Foods 9, 532.

FAO, 2001. State of Food Insecurity in the World 2001. Rome, Italy.

FAO, 2021. State of Food Security and Nutrition in the World 2021. Rome, Italy.

Ferguson, M., Brown, C., Georga, C., Miles, E., Wilson, A., Brimblecombe, J., 2017. Traditional food availability and consumption in remote Aboriginal communities in the Northern Territory, Australia. Aust. N. Z. J. Publ. Health 41, 294–298.

Gallaher, C.M., Kerr, J.M., Njenga, M., Karanja, N.K., WinklerPrins, A.M.G.A., 2013. Urban agriculture, social capital, and food security in the Kibera slums of Nairobi, Kenya. Agric. Hum. Val. 30, 389–404.

Ghosh-Jerath, S., Kapoor, R., Barman, S., Singh, G., Singh, A., Downs, S., Fanzo, J., 2021. Traditional food environment and factors affecting indigenous food consumption in Munda Tribal community of Jharkhand, India. Front. Nutr. 7, 1–18.

Habwe, F.O., Walingo, M.K., Abukutsa-Onyango, M.O., Oluoch, M.O., 2009. Iron content of the formulated East African indigenous vegetable recipes. Afr. J. Food Sci. 3, 393–397.

Hoffman, D.J., Merchant, E., Byrnes, D.R., Simon, J.E., 2018. Preventing micronutrient deficiencies using African indigenous vegetables in Kenya and Zambia. Sight Life 32 (2), 177–181.

Hoffman, D., 2019. The double burden of malnutrition: research agenda for reversing global trends. Ann. Nutr. Metabol. 75, 149–152.

Hunter, D., Borelli, T., Beltrame, D.M.O., Oliveira, C.N.S., Coradin, L., Wasike, V.W., Wasilwa, L., Mwai, J., Manjella, A., Samarasinghe, G.W.L., et al., 2019. The potential of neglected and underutilized species for improving diets and nutrition. Planta 250, 709–729.

Jarosz, L., 2014. Comparing food security and food sovereignty discourses. Dialogues Hum. Geogr. 4, 168–181.

Korth, M., Stewart, R., Langer, L., Madinga, N., Rebelo Da Silva, N., Zaranyika, H., van Rooyen, C., de Wet, T., 2014. What are the impacts of urban agriculture programs on food security in low and middle-income countries: a systematic review. Environ. Evid. 3, 21.

Lambden, J., Receveur, O., Marshall, J., Kuhnlein, H., 2006. Traditional and market food access in Arctic Canada is affected by economic factors. Int. J. Circumpolar Health 65, 331–340.

Leff, B., Ramankutty, N., Foley, J.A., 2004. Geographic distribution of major crops across the world. Global Biogeochem. Cycles 18, 1–27.

Li, X., Yadav, R., Siddique, K.H.M., 2020. Neglected and underutilized crop species: the key to improving dietary diversity and fighting hunger and malnutrition in Asia and the Pacific. Front. Nutr. 7, 1–12.

Márquez, A., Schwartz, N., 2009. Traditional home gardens of Petén, Guatemala: resource management, food security, and conservation. J. Ethnobiol. 28, 305–317.

Ochieng, J., Kirimi, L., Mathenge, M., 2016. Effects of climate variability and change on agricultural production: the case of small-scale farmers in Kenya. NJAS Wageningen J. Life Sci. 77, 71–78.

Ochieng, J., Schreinemachers, P., Ogada, M., Dinssa, F.F., Barnos, W., Mndiga, H., 2019. Adoption of improved amaranth varieties and good agricultural practices in East Africa. Land Use Pol. 83, 187–194.

Patel, R., 2009. Food sovereignty. J. Peasant Stud. 36, 663–706. https://doi.org/10.1080/03066150903143079.

Popkin, B.M., Corvalan, C., Grummer-Strawn, L.M., 2020. Dynamics of the double burden of malnutrition and the changing nutrition reality. Lancet 395, 65–74.

Ruel, M.T., Deitchler, M., Arimond, M., 2010. Developing simple measures of women's diet quality in developing countries: overview. J. Nutr. 140, 2048S–2050S.

Sattaka, P., Pattaratuma, S., Attawipakpaisan, G., 2017. Agricultural extension services to foster production sustainability for food and cultural security of glutinous rice farmers in Vietnam. Kasetsart J. Soc. Sci. 38, 74–80. https://doi.org/10.1016/j.kjss.2016.05.003.

Senyolo, G.M., Wale, E., Ortmann, G.F., 2014. Consumers' willingness-to-pay for underutilized vegetable crops: the case of African leafy vegetables in South Africa. J. Hum. Ecol. 47, 219–227.

Simon, J.E., Weller, S., Hoffman, D., Govindasamy, R., Morin, X., Merchant, E.V., Dinssa, F.F., Van Wyk, E., Byrnes, D., Odendo, M., et al., 2020. Improving income and nutrition of smallholder farmers in Eastern Africa using a market-first science-driven approach to enhance value chain production of African indigenous vegetables. J. Medicinally Active Plants 9 (4), 289–309.

Simon, J., Acquaye, D., Govindasamy, R., Asante-Dartey, J., Juliani, R., Diouf, B., Diatta, M., Langenhoven, P., Van Wyk, E., Hitimanana, N., et al., 2021. Building community resiliency through horticultural innovation. Sci. Global 1–10. https://doi.org/10.33548/SCIENTIA601.

Sunderland, T.C.H., O'Connor, A., Muir, G., Nerfa, L., Nodari, G.R., Wildmark, C., Bahar, N., Ichowitz, A., 2019. SDG2 zero-hunger—challenging the hegemony of monoculture agriculture or forests and people. In: Katila, P., Pierce Colfer, C.J., de Jong, W., Galloway, G., Pacheco, P., Winkel, G. (Eds.), Sustainable Development Goals. Cambridge University Press.

Towns, A.M., Shackleton, C., 2018. Traditional, indigenous, or leafy? A definition, typology, and way forward for African vegetables. Econ. Bot. 72 (4), 461–477.

United Nations, 1975. Report of the World Food Conference. Rome, Italy. November 5–16, 1974.

Weller, S.C., Van Wyk, E., Simon, J.E., 2015. Sustainable production for more resilient food production systems: case study of African indigenous vegetables in eastern Africa. Acta Hortic. 289–298.

Yuan, B., Byrnes, D., Dinssa, F.F., Simon, J.E., Wu, Q., 2019. Quantity assessment of polyphenols, glycoalkaloids and saponins in *Solanum scabrum* berries of different genetic sources and maturity by HPLC/UV-visible/MS methods. J. Sci. Food Agric. 99, 3578–3587.

Yuan, B., Dinssa, F.F., Simon, J.E., Wu, Q., 2020. Simultaneous quantification of polyphenols, glycoalkaloids and saponins in African nightshade leaves using ultra-high performance liquid chromatography tandem mass spectrometry with acid assisted hydrolysis and multivariate analysis. Food Chem. 312, 126030.

Trans-fatty acids: Health effects, recommendations, and regulations

S.K. Gebauer[a] **and D.J. Baer**[b], [a] US Food and Drug Administration, College Park, MD, United States; and [b] US Department of Agriculture, Beltsville, MD, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of S.K. Gebauer, D.J. Baer, *Trans*-Fatty Acids: Health Effects, Recommendations, and Regulations, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 288–292, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00102-1.

Introduction 648
Health effects 648
Cardiovascular disease 648
Coronary heart disease 648
Lipids and lipoproteins 649
Cancer 649
Insulin sensitivity and diabetes 649
Dietary recommendations, regulations, and dietary intake 649
***Trans* fat alternatives** 650
Conclusions 651
Disclaimer 651
References 651

Key points

- The adverse health effects of industrially-produced TFA are well documented in the scientific literature.
- Numerous countries have implemented effective measures to limit or remove TFA from the food supply.
- Intake of TFA has decreased worldwide as a result of public health efforts, although intake still remains higher than recommended in some countries that have not yet enacted policies to reduce TFA.

Introduction

TFA are a class of fatty acids that contain one or more double bonds in the *trans* configuration. The two dietary sources of TFA are those produced through the partial hydrogenation of oils (industrially-produced), and those that are naturally-occurring, which are produced by ruminant animals (ruminant TFA). Historically, vegetable oils were partially hydrogenated for use in margarines, commercial cooking and frying, and food manufacturing. The presence of TFA results in fats that are solid at room temperature, longer shelf-life of commercial products, and stability of oils at very high temperatures. Elaidic acid (C18:1, t9) typically is the major isomer contained in industrially-produced sources of TFA. Ruminant TFA are synthesized via bacterial metabolism of C18 unsaturated fatty acids in ruminant animals, and found in ruminant-derived products, such as beef, lamb, and dairy. Vaccenic acid (C18:1, t11), the predominant *trans*-monoene isomer in ruminant fats, comprises 50–80% of the total TFA in ruminant fat and is the precursor for the c9,t11 isomer of conjugated linoleic acid (CLA). CLA is a group of positional and geometric conjugated isomers of linoleic acid. The two predominant isomers of CLA with known bioactive properties are c9,t11 and t10,c12.

Health effects

Numerous studies have demonstrated the adverse effects of industrially-produced TFA on the risk of chronic diseases, including cardiovascular disease, cancer, and diabetes. However, the health effects of ruminant TFA in humans have not been as well-studied, and clinical studies comparing the two sources of TFA are more limited.

Cardiovascular disease

Coronary heart disease

Numerous epidemiological studies have demonstrated an association between intake of TFA and risk of coronary heart disease (CHD). In the early 1990s, results from the Nurses' Health Study, a prospective cohort study, demonstrated that intake of TFA isomers was directly related to risk of CHD (RR = 1.5, 95% CI 1.12–2.00, p for trend = 0.001) (Willett, 1993). In contrast, there

was a nonsignificant inverse association with intake of TFA from ruminant sources and risk of CHD. Two meta-analyses have reported similar findings demonstrating that intake of industrially-produced TFA increases total CHD and CHD mortality whereas intake of ruminant TFA does not (Bendsen, 2011; de Souza, 2015).

Overall, these data suggest that intake of ruminant TFA may have differential effects on CHD risk compared with those of industrially-produced TFA; however, both meta-analyses are based on a limited number of studies ($n = 6$). Further research is needed to determine the effects of ruminant TFA on CHD risk.

Lipids and lipoproteins

Clinical trials consistently demonstrate that consumption of TFA adversely affects lipids and lipoproteins. Studies have demonstrated that these effects are greater than those of saturated fatty acids. Compared with saturated fatty acids, consumption of TFA results in a lower concentration of high-density lipoprotein (HDL)-cholesterol and a higher concentration of low-density lipoprotein (LDL)-cholesterol (specifically small dense LDL particles), triglycerides, and Lp(a).

In a recent systematic review of nine studies, TFA from ruminant sources seemed to have less impact on HDL-cholesterol than TFA from industrially-produced sources (Verneque, 2020). However, for total- and LDL-cholesterol, results indicated that TFA from ruminant sources may increase blood concentrations to a greater degree than TFA from industrially-produced sources (Verneque, 2020) albeit there may be differences in the response of men and women. In some clinical studies, amounts of TFA consumed have been high (e.g., 7% of energy) compared to typical intake. Additionally, some studies suggest a greater effect in women, although these sex differences have not been well established.

Cancer

The association of TFA intake and cancer risk has been investigated for at least 19 types of cancer, with the most studied types of cancer being breast, prostate, and colorectal. One recent meta-analysis of total TFA intake (regardless of dietary source) demonstrated a positive association (significantly higher odds ratio) between TFA consumption and risk for prostate and colorectal cancer. No association was identified for breast, ovarian, or non-Hodgkin lymphoma (Michels, 2021). Data on the association between TFA intake and other cancer sites are very limited. In another recent meta-analysis, no association was found between intake of TFA and breast cancer risk in pre- and post-menopausal women; however, a positive association was found between serum concentration of TFA and breast cancer risk in post-menopausal women (Anjom-Shoae, 2020). There was no association in this meta-analysis between CLA intake and risk for breast cancer (Anjom-Shoae, 2020).

Insulin sensitivity and diabetes

Some studies suggest that TFA can exacerbate insulin resistance. However, a recent meta-analysis concluded that there was no effect of TFA (ruminant or industrially-produced sources) on fasting glycemia, insulin sensitivity, or insulinemia (Verneque, 2020). In the Nurses' Health Study, high intake of TFA (3% of energy) was associated with an increased risk of type 2 diabetes (Oh, 2005). Results from this study suggest that replacement of 2% of energy from TFA with polyunsaturated fatty acids would result in a 39% increase in risk of type 2 diabetes. However, in two other prospective studies, the Health Professionals Follow-Up Study (van Dam, 2002) and Iowa Women's Health Study (Meyer, 2001), there was no association with TFA intake and risk of diabetes.

Dietary recommendations, regulations, and dietary intake

Reflecting the growing body of scientific evidence on the adverse effects of industrially-produced TFA, numerous public health recommendations on TFA were issued in the early 2000s. In 2002, the US National Academies of Sciences, Engineering and Medicine recommended that TFA consumption should be as low as possible in a nutritionally adequate diet (Institute of Medicine). In 2003, the World Health Organization (WHO) recommended that TFA intake should be limited to less than 1% of overall energy consumption (World Health Organization). The *Dietary Guidelines for Americans* first included recommendations for TFA in the 2005 edition, with a key recommendation that TFA consumption should be as low as possible (U.S. Department of Health and Human Services). Numerous other national and international organizations have since issued similar recommendations. In a recent scoping review on available guidelines for dietary fat published between 2015 and 2020, most organizations included recommendations for industrially-produced TFA that spanned from complete avoidance to limiting intake to no more than 2% of energy intake (Schwingshackl, 2021).

Denmark became the first country to regulate TFA in the food supply, in 2003, by mandating that only 2% of fats and oils could be TFA (Stender, 2004). Also in 2003, the U.S. Food and Drug Administration (FDA) amended its regulations on food labeling, effective in 2006, to require that *trans* fat be declared on the Nutrition Facts label of foods (U.S. Food and Drug Administration, 2003). New York City took action in 2006 to remove TFA from foods sold in food service establishments, followed by the state of California and several U.S. cities and regions in 2007 and 2008 (e.g., Philadelphia, PA; Baltimore, MD; Boston, MA; and Montgomery County, Maryland). In 2015, the FDA took measures to remove industrially-produced TFA from the US food supply by determining that partially hydrogenated oils (PHOs) are no longer "Generally Recognized as Safe" or GRAS (U.S. Food and Drug Administration, 2015). Consequently, PHOs, which previously were the main source of TFA in the diet, have been largely

removed from foods in the United States. Other countries have also implemented regulations to eliminate PHOs (Government of Canada, 2018), and in 2018 the WHO called for the global elimination of industrially-produced *trans* fatty acids by 2023 (Ghebreyesus, 2018). The 2021 WHO report on global *trans* fat elimination summarizes the progress that has been made toward this goal. There are 57 countries that currently have mandatory TFA policies in place, however several countries that have the highest CHD burden attributed to TFA intake have yet to implement mandatory policies (World Health Organization, 2021a).

Overall, intake of TFA has declined worldwide in the past 25 years as a result of the public health efforts to remove industrially-produced TFA from foods. Consequently, in most countries where data are available, intake of ruminant TFA is now higher than intake of industrially-produced TFA (Wanders, 2017). Studies evaluating the effectiveness of regulations to limit TFA in the food supply, such as local regulations implemented in New York City, and regulations in other countries, have demonstrated a significant decrease in TFA consumption and disease burden, as reviewed by Li et al. (Li, 2019). When comparing nationally representative data in the US from 1999 to 2000 versus 2009–2010, which spanned the time period when US labeling requirements for *trans* fat were implemented as well as other actions such as local and state regulations to limit TFA in food-service establishments, plasma TFA concentrations decreased by an average of 61.9% in children and adolescents and by 54% in adults (Restrepo, 2020; Vesper, 2017). Data are still needed to determine the effectiveness of some of the more recent measures that have been implemented. For example, the data used in analyses of dietary intake and sources of TFA in the US were collected prior to when PHOs were removed from the food supply in the US (Wanders, 2017; Li, 2021).

Monitoring of TFA intake is a critical tool to not only evaluate the effectiveness of regulations and policies in areas where they have been implemented, but also to demonstrate the need for enacting measures to reduce TFA intake in countries where intake remains high. As previously discussed, although intake has declined in many countries where actions have been taken, intake still remains higher than recommended in countries that have not implemented regulations and policies to reduce TFA. To help reach the goal of elimination of *trans* fat from the global food supply by 2023, the WHO developed the REPLACE action package to help countries eliminate industrially-produced TFA (World Health Organization, 2021b). This approach highlights several areas where countries can focus their efforts, one of which involves determining and promoting healthier alternatives to industrially-produced TFA.

Trans fat alternatives

Trans fats became a prevalent ingredient in food products as a means for replacing saturated fats in the diet, as the adverse effects of saturated fats were well known. It is now understood that the effects of industrially-produced TFA on numerous health outcomes are more deleterious than those of saturated fatty acids. As TFA have been removed or significantly reduced in the food supply, it has led to the question—what alternatives to PHOs should be used as food manufacturers are reformulating products? There are multiple considerations when choosing alternatives for TFA, including taste and texture, as well as cost and availability. Moreover, the potential health effects of the alternative should be considered, so as to avoid introducing a harmful ingredient into the food supply.

One technique that has been utilized to replace industrially-produced TFA is interesterification of fats—a process in which fatty acids are rearranged on a triacylglycerol molecule. This technique is useful because the position of fatty acids on the glycerol molecule is critical in creating a fat with the proper physical properties required for the application. Stearic acid and palmitic acid, both saturated fatty acids, typically are used in interesterification. In a recent systematic review, interesterification with stearic and palmitic fatty acids (increasing their concentration in the s*n*-2 position), compared with low palmitic or stearic fatty acids in the s*n*-2 position, did not affect fasting concentrations of lipids or apolipoproteins (van Rooijen, 2020). The same authors also conducted a systematic review comparing the effects of stearic acid and palmitic acid and found that replacing palmitic acid with stearic acid lowered LDL-cholesterol (van Rooijen, 2020). Additional research is needed from both the fasted and post-prandial states to better understand the impact of interesterification on other biomarkers of health, especially cardiometabolic health.

Other TFA alternatives include oils that require little or no hydrogenation, such as tropical oils (palm oil, palm kernel oil, and coconut oil), as well as oils developed by manipulating the fatty acid composition of oil seeds (low-linoleic, midoleic, or high-oleic oils) to create products that exhibit increased oxidative stability during deep-frying and increased shelf-life due to increases in relatively stable fatty acids (i.e., oleic acid) and decreases in relatively unstable fatty acids (i.e., linolenic acid). Tropical oils, however, are high in saturated fatty acids, and are therefore not included in the group of oils that are recommended as part of a healthy dietary pattern (U.S. Department of Agriculture). When choosing TFA alternatives, an important consideration is whether they provide the functionality needed depending on the food application. For baking applications, where solid or semisolid fats are needed for consistency and feel, alternatives such as interesterified fats, tropical oils, and fully hydrogenated vegetable oil and can be used. For frying applications, oils such as soybean, canola, cottonseed, and high-oleic sunflower oil, for example, can be useful. A few dietary intervention studies have demonstrated lipid-lowering effects of high-oleic oils, including high-oleic canola oil (Gillingham, 2011; Jones, 2014) and high-oleic soybean oil (Lichtenstein, 2006), when compared to a variety of other oils. In 2018, FDA issued a qualified health claim for edible oils containing $\geq$70% of oleic acid per serving based on credible evidence that consumption of these oils may reduce the risk of CHD, when substituted for fats and oils that are higher in saturated fat (U.S. Food and Drug Administration). A more recent study was the first to compare high-oleic oils to alternative oils with similar functionality (Baer, 2021). In this study, consuming a diet prepared with high-oleic soybean oil, as well as a diet prepared with a blend of high-oleic soybean oil and fully hydrogenated soybean oil, resulted in improved lipid and lipoprotein profiles compared to a diet prepared with a blend of palm oil and palm kernel oil (Baer, 2021).

Soybean oil and canola oil, which are sources of the essential fatty acids linoleic acid and alpha-linolenic acid, are widely used in the food supply. While modified high-oleic oils have higher amounts of oleic acid than traditional oils, they also have lower amounts of linoleic acid and alpha-linolenic acid. Modeling has been used to estimate the effects of replacing these traditional oils with high-oleic oils on fatty acid intake at population levels. In US adults and children, replacement of soybean oil and canola oil with high-oleic varieties at multiple levels (10%, 25%, and 50% in adults; 20%, 40%, 60%, and 80% in children), were estimated to decrease saturated fatty acid intake, increase monounsaturated fatty acid intake (particularly oleic acid), and decrease polyunsaturated fatty acid intake (particularly linoleic acid and alpha-linolenic acid) (Raatz, 2018; Belury et al., 2022). Further, results suggest that replacement of soybean oil and canola oil with high-oleic soybean oil and high oleic canola oil may lead to inadequate intakes of these essential fatty acids in certain segments of the US adult population and in US children, at levels of replacement of $\geq$25% and $\geq$40%, respectively (Raatz, 2018; Belury et al., 2022). In contrast, previous modeling in the US population with different parameters suggests that adequate intake levels of linoleic acid and alpha-linolenic acid are met when estimating the effects of different uses of soybean oil varieties in the food supply (Lefevre, 2012; Crawford, 2011). While additional research is needed on the impact of using different TFA alternatives, these studies demonstrate the importance of using a balanced approach with a variety of different oils and fats in the food supply.

Conclusions

Overall, data from epidemiologic and clinical studies consistently demonstrate the adverse effects of industrially-produced TFA on CHD risk and lipids and lipoproteins; their effects on other chronic disease risk factors, such as insulin resistance, as well as risk of type 2 diabetes and cancer, are less consistent. Further research is needed to determine the effects of ruminant TFA at amounts typically consumed in the diet. Public health measures have been effective in reducing intake of industrially-produced TFA in countries where they have been implemented. As alternatives to TFA are used in foods, it is important to continue to study their effects on health outcomes. Studies suggest that high-oleic soybean oil, as well as blends of high-oleic soybean oil with fully hydrogenated soybean oil, may be functional alternatives that have beneficial effects on lipid profiles.

Disclaimer

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture.

The US Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or part of an individual's income is derived from any public assistance program. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720–2600 (voice and TDD). To file a complaint of discrimination, write to USDA, Director, Office of Civil Rights, 1400 Independence Avenue, S.W., Washington, DC 20250-9410, or call (800) 795–3272 (voice) or (202) 720–6382 (TDD). USDA is an equal opportunity provider and employer.

The views and opinions of the authors expressed herein do not necessarily state or reflect those of the US Food and Drug Administration.

References

Anjom-Shoae, J., Sadeghi, O., Larijani, B., Esmaillzadeh, A., 2020. Dietary intake and serum levels of *trans* fatty acids and risk of breast cancer: a systematic review and dose-response meta-analysis of prospective studies. Clin. Nutr. 39, 755–764.

Baer, D.J., Henderson, T., Gebauer, S.K., 2021. Consumption of high-oleic soybean oil improves lipid and lipoprotein profile in humans compared to a palm oil blend: a randomized controlled trial. Lipids 56, 313–325.

Belury, M.A., Raatz, S., Conrad, Z., 2022. Modeled substitution of traditional oils with high oleic acid oils decreases essential fatty acid intake in children. Am. J. Clin. Nutr. 115, 1180–1188.

Bendsen, N.T., Christensen, R., Bartels, E.M., Astrup, A., 2011. Consumption of industrial and ruminant *trans* fatty acids and risk of coronary heart disease: a systematic review and meta-analysis of cohort studies. Eur. J. Clin. Nutr. 65, 773–783.

Crawford, A.W., Wang, C., Jenkins, D.J., Lemke, S.L., 2011. Estimated effect on fatty acid intake of substituting a low-saturated, high-oleic, low-linolenic soybean oil for liquid oils. Nutr. Today 46, 189–196.

De Souza, R.J., Mente, A., Maroleanu, A., et al., 2015. Intake of saturated and *trans* unsaturated fatty acids and risk of all cause mortality, cardiovascular disease, and type 2 diabetes: systematic review and meta-analysis of observational studies. BMJ 351, h3978.

Ghebreyesus, T.A., Frieden, T.R., 2018. REPLACE: a roadmap to make the world *trans* fat free by 2023. Lancet 391, 1978–1980.

Gillingham, L.G., Gustafson, J.A., Han, S.Y., et al., 2011. High-oleic rapeseed (canola) and flaxseed oils modulate serum lipids and inflammatory biomarkers in hypercholesterolaemic subjects. Br. J. Nutr. 105, 417–427.

Government of Canada, 2018. Regulations amending certain regulations made under the food and drugs act (nutrition symbols, other labelling provisions, partially hydrogenated oils and vitamin D). Canada Gazette 152 (6), 279–369. https://gazette.gc.ca/rp-pr/p1/2018/2018-02-10/html/reg2-eng.html. (Accessed 11 February 2022).

Institute of Medicine, 2002. Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids. National Academies Press, Washington, D.C. United States.

Jones, P.J., Senanayake, V.K., Pu, S., et al., 2014. DHA-enriched high-oleic acid canola oil improves lipid profile and lowers predicted cardiovascular disease risk in the canola oil multicenter randomized controlled trial. Am. J. Clin. Nutr. 100, 88–97.

Lefevre, M., Mensink, R.P., Kris-Etherton, P.M., et al., 2012. Predicted changes in fatty acid intakes, plasma lipids, and cardiovascular disease risk following replacement of *trans* fatty acid-containing soybean oil with application-appropriate alternatives. Lipids 47, 951–962.
Li, C., Cobb, L.K., Vesper, H.W., Asma, S., 2019. Global surveillance of *trans*-fatty acids. Prev. Chronic Dis. 16, E147.
Li, C., Richter, P., Cobb, L.K., et al., 2021. Dietary sources of plasma *trans* fatty acids among adults in the United States: NHANES 2009–2010. Curr. Dev. Nutr. 5, nzab063.
Lichtenstein, A.H., Matthan, N.R., Jalbert, S.M., et al., 2006. Novel soybean oils with different fatty acid profiles alter cardiovascular disease risk factors in moderately hyperlipidemic subjects. Am. J. Clin. Nutr. 84, 497–504.
Meyer, K.A., Kushi, L.H., Jacobs Jr., D.R., Folsom, A.R., 2001. Dietary fat and incidence of type 2 diabetes in older Iowa women. Diabetes Care 24, 1528–1535.
Michels, N., Specht, I.O., Heitmann, B.L., et al., 2021. Dietary *trans*-fatty acid intake in relation to cancer risk: a systematic review and meta-analysis. Nutr. Rev. 79, 758–776.
Oh, K., Hu, F.B., Manson, J.E., et al., 2005. Dietary fat intake and risk of coronary heart disease in women: 20 years of follow-up of the Nurses' Health Study. Am. J. Epidemiol. 161, 672–679.
Raatz, S.K., Conrad, Z., Jahns, L., et al., 2018. Modeled replacement of traditional soybean and canola oil with high-oleic varieties increases monounsaturated fatty acid and reduces both saturated fatty acid and polyunsaturated fatty acid intake in the US adult population. Am. J. Clin. Nutr. 108, 594–602.
Restrepo, B.J., 2020. Intake of *trans*-fats among US youth declined from 1999–2000 to 2009–2010. Publ. Health Nutr. 23, 1103–1107.
Schwingshackl, L., Zahringer, J., Beyerbach, J., et al., 2021. A scoping review of current guidelines on dietary fat and fat quality. Ann. Nutr. Metab. 77, 65–82.
Stender, S., Dyerberg, J., 2004. Influence of *trans* fatty acids on health. Ann. Nutr. Metab. 48, 61–66.
U.S. Department of Agriculture, U.S. Department of Health and Human Services, 2020. Dietary Guidelines for Americans, 2020–2025, ninth ed. Available at. DietaryGuidelines.gov.
U.S. Department of Health and Human Services, U.S. Department of Agriculture, 2005. Dietary Guidelines for Americans, sixth ed. U.S. Government Printing Office, Washington, DC.
U.S. Food and Drug Administration, 2003. Food labeling: trans. Fed. Regist. 68, 41433–41506.
U.S. Food and Drug Administration, 2015. Final determination regarding partially hydrogenated oils. Fed. Regist. 80, 34650–34670.
U.S. Food and Drug Administration, 2018. Oleic Acid and Coronary Heart Disease (Corbion Biotech Petition). Retrieved from. https://www.fda.gov/media/118199/download.
Van Dam, R.M., Rimm, E.B., Willett, W.C., et al., 2002. Dietary patterns and risk for type 2 diabetes mellitus in U.S. men. Ann. Intern. Med. 136, 201–209.
Van Rooijen, M.A., Mensink, R.P., 2020. Palmitic acid versus stearic acid: effects of interesterification and intakes on cardiometabolic risk markers—a systematic review. Nutrients 12.
Verneque, B.J.F., Machado, A.M., De Abreu Silva, L., et al., 2020. Ruminant and industrial *trans*-fatty acids consumption and cardiometabolic risk markers: a systematic review. Crit. Rev. Food Sci. Nutr. 1–11.
Vesper, H.W., Caudill, S.P., Kuiper, H.C., et al., 2017. Plasma *trans*-fatty acid concentrations in fasting adults declined from NHANES 1999–2000 to 2009–2010. Am. J. Clin. Nutr. 105, 1063–1069.
Wanders, A.J., Zock, P.L., Brouwer, I.A., 2017. *Trans* fat intake and its dietary sources in general populations worldwide: a systematic review. Nutrients 9.
Willett, W.C., Stampfer, M.J., Manson, J.E., et al., 1993. Intake of *trans* fatty acids and risk of coronary heart disease among women. Lancet 341, 581–585.
World Health Organization, 2021a. Countdown to 2023: WHO Report on Global *Trans* Fat Elimination 2021. World Health Organization, Geneva. Licence: CC BY-NC-SA 3.0 IGO.
World Health Organization, 2021b. REPLACE *Trans* Fat: An Action Package to Eliminate Industrially Produced *Trans*-fatty Acids. ISBN: 9789240021105. https://www.who.int/publications/i/item/9789240021105.
World Health Organization/Food and Agriculture Organization (FAO), 2003. Diet, Nutrition and the Prevention of Chronic Diseases: Report of a Joint WHO/FAO Expert Consultation. WHO, Geneva. WHO Technical Report Series; No. 916.

Ultra-processed foods

Paula dos Santos Leffa, Nutrition Research Group, Federal University of Health Sciences, Porto Alegre, Brazil

© 2023 Elsevier Ltd. All rights reserved.

Introduction 653
Defining ultra-processed foods—NOVA classification system 654
Group 1. Unprocessed or minimally processed foods 654
Group 2. Processed culinary ingredients 655
Group 3. Processed foods 655
Group 4. Ultra-processed foods 655
Global participation of ultra-processed foods 656
Impact of ultra-processed foods on diet quality 656
Impact of ultra-processed foods on health 656
Obesity 656
Cardiovascular and metabolic diseases 657
Diabetes 657
Cancer 657
Mortality and aging 657
Other diseases and risk factors 658
Ultra-processed foods and chronic diseases: potential mechanisms of action beyond nutritional quality 658
Social determinants of ultra-processed foods consumption 658
NOVA in use 659
Conclusion 659
References 660

Key points

- The NOVA food classification system has been largely used and it is recognized as a valid tool for public health and nutrition research and policy;
- The most up-to-date evidence suggests that a major contributor of unhealthy dietary patterns and chronic non-communicable diseases is a high intake of ultra-processed foods;
- The article provides the mechanisms underlying the associations between ultra-processed foods and their impact on health;

Introduction

The food preparation and the techniques used for its cultivation and preservation allowed the evolution and adaptation of the human species (Wrangham, 2013). Historically, food processing has been slowly and gradually consolidated during the history of humanity. However, since the early 1980s, food processing has developed very rapidly and with intense changes through science and technology. Factors such as economic globalization, changes in food production, and consumption patterns have revolutionized the food patterns in most countries by increasing the production, availability, accessibility, and commercialization of processed food products and beverages (Ludwig, 2011; Swinburn et al., 2011). Traditional diets based on fruits and vegetables were replaced by diets with high energy density, rich in sugar and sodium, with increased consumption of fats, especially saturated and trans, refined carbohydrates and animal source foods (WHO, 2003). At the same time, obesity and its related conditions, such as hypertension and diabetes, have risen very rapidly. Such changes in dietary patterns are known as the "nutrition transition" (Popkin, 2004; Popkin et al., 2012), with increased consumption of industrially processed foods as a striking characteristic of these new dietary patterns.

Traditionally, epidemiological studies investigating dietary patterns and their impact on health have used energy and macro and micronutrient intake as the basis for assessing the diet. However, it is well-established that individuals do not ingest isolated nutrients, but meals composed of different foods that provide different nutrients and chemical compounds that interact synergically. Since 1998, the World Health Organization has suggested that food consumption should be assessed based on food rather than nutrients (World Health Organization, 1998), making it possible to measure the dietary pattern more reliably.

The impact of industrial processing and, particularly, techniques and ingredients developed by the food industry on the nature of food and the state of human health is still not widespread. Until recently, food processing was neglected in epidemiological and experimental studies on diet, nutrition, and health.

Recently, in 2010, the NOVA classification system of industrial food processing was developed based on the degree and purpose processing (Monteiro et al., 2010). The NOVA classification precisely defined ultra-processed food (UPF) and is now considered a valid tool for public health and nutrition research and policy. Moreover, there is a growing body of scientific evidence assessing the burden of disease attributable to specific UPF consumption. Therefore, understanding the use of UPF classification and its application is of great importance. The following sections describe the most recent NOVA update and a summary of its uses and the relationship between UPF consumption and nutritional quality and its impact as a risk factor for diseases.

Defining ultra-processed foods—NOVA classification system

Food processing in itself is not a problem as most foods are processed to some extent, if only to promote longer shelf-life and limit spoilage. Thus, it is essential to highlight that the terms "processing" and "industry" are very general and, therefore, not helpful. NOVA system classifies all foods and food products into four groups based on the extent, nature, and purpose of food processing they undergo. To that extent, the NOVA system considers all physical, biological, and chemical aspects used during the food manufacturing process, including additives (Monteiro et al., 2019b).

Foods may be consumed by themselves (fruits), or as the main item or accompanying items of dishes and meals (such as grains, flours, vegetables, meat, eggs); or as food products for cooking food and creating culinary preparations (such as oils, butter, sugar, salt); or they may be food products ready to consume or heat. NOVA groups are described below. **Table 1** presents detailed lists of foods of each of NOVA groups.

Group 1. Unprocessed or minimally processed foods

Unprocessed and minimally processed foods make up group 1 of NOVA. Unprocessed (or natural) foods are edible parts of plants (seeds, fruits, leaves, stems, roots) or animals (muscle, offal, eggs, milk), and also fungi, algae, and water, after separation from nature. Minimally processed foods are unprocessed foods altered by minimal industrial processes before being purchased and consumed, with the main objective of extending the useful life of the food, allowing its storage for prolonged use.

In this group, industrial processes include removal of inedible or unwanted parts, drying, crushing, grinding, fractioning, roasting, boiling, pasteurization, refrigeration, freezing, placing in containers, vacuum packaging, non-alcoholic fermentation or similar procedures, conserving much of the nutritional properties of the original food and increasing its availability and safety. None of

Table 1 The NOVA food groups with examples.

(1) Unprocessed or minimally processed foods	*(2) Processed culinary ingredients*	*(3) Processed foods*	*(4) Ultra-processed foods*
Fresh, frozen, or dried fruits or vegetables; grains (rice, corn, oat); legumes (beans, lentils, and chickpeas); starchy roots and tubers (potatoes, sweet potatoes, and cassava); fresh or chilled or frozen meat, poultry, fish and seafood; eggs; fresh or pasteurized milk; 100% pasteurized or fresh fruit juice; no-added-sugar plain yogurt; nuts and seeds without added sugar or salt; spices and dried herbs; tea, coffee, and drinking water. It also includes foods made up from two or more items in this group but contain no added ingredients (e.g., mixed dried fruits).	Vegetable oils (olives oil, coconut oil, soy oil); animal fat (butter and lard); vinegar; sugar, honey, and maple syrup; starches extracted from corn and other plants; and salt. It also includes foods consisting of two group 2 items (e.g., salted butter) and group 2 items with added vitamins or minerals, such as iodized salt.	Canned vegetables and legumes; salted or sugared nuts and seeds; canned fish; plain yogurt with added sugar; fruits in syrup (e.g., Jam made of fruit and sugar with or without added anti-oxidants; freshly made unpackaged bread and cheeses. They may contain additives (e.g., anti-oxidants).	Carbonated soft drinks, sweetened juice; juice powder; cookies; ice cream, sweet or savory packaged snacks; chocolate, candies (confectionery); "breakfast cereals"; cereal bars; cookies (biscuits), cakes and cake mixes; microwave popcorn; flavored granola bars with added sugar and preservatives; artificially flavored cheese; baby cereals; margarine, mayonnaize and dressing; sugary milk beverages; flavored and sugared yogurt; "energy drinks"; fish/chicken "nuggets"; fast food burgers, sausages, hot dogs, and other reconstituted meat products; frozen products ready to heat such as pasta dishes, pizzas, hamburgers, pre-prepared pies; powdered and packaged "instant" soups, noodles and desserts; industrially produced bread.

Adapted from Monteiro et al. (2019b).

these processes add salt, sugar, oils or fats, or other food substances to the original food. Group 1 items may infrequently contain additives that prolong product duration, protect original properties, or prevent the proliferation of microorganisms. Finally, in appropriate combinations and quantity, all foods in this group form the basis for healthy diets.

Group 2. Processed culinary ingredients

Processed culinary ingredients are substances derived from Group 1 foods or from nature by processes that include pressing, refining, grinding, milling, and drying. Examples are oils, butter, sugar and salt. Their use is in the preparation, seasoning, and cooking of group 1 foods. Processed culinary ingredients are not meant to be consumed by themselves but are normally used in combination with unprocessed and minimally processed foods to make meals and dishes. Group 2 items may contain additives used to preserve the product's original properties, prolong duration or prevent the proliferation of microorganisms.

Group 3. Processed foods

Processed foods are made essentially by adding salt, oil, sugar or other substances from Group 2 to Group 1 foods. Classic examples included bottled vegetables, canned fish, fruits in syrup, cheeses, and freshly made bread. They are relatively simple products—most processed foods have two or three ingredients and are recognizable as modified versions of Group 1 foods. Processes include various cooking or preservation methods, such as canning and bottling, and, in the case of bread and cheese, non-alcoholic fermentation. The purpose of processing in Group 3 is to increase Group 1 foods' durability or modify or enhance their sensory qualities. Processed food products usually maintain the identity and most proprieties of the original food. While some of these foods may contain additives and ingredients like oil, sugar, and salt, they can be part of a healthy diet in moderate amounts. Group 3 items may contain additives that prolong product duration, protect original properties, or prevent the proliferation of microorganisms.

Group 4. Ultra-processed foods

The fourth NOVA group is of UPF and drink products. These are formulations of ingredients, most of exclusive industrial use, that result from a series of industrial processes. UPF are not modified foods but formulations made mostly or entirely from substances derived from foods and additives and typically including little or no fresh food from Group 1. Processes enabling the manufacture of ultra-processed foods include fractioning whole foods into substances, chemical modifications of these substances, assembly of unmodified and modified food substances using industrial techniques such as extrusion, molding, and pre-frying. Processes end with sophisticated packaging, usually with synthetic materials.

Sugar, oils and fats, and salt are often ingredients used to make UPF, generally in high quantity. However, in terms of ingredients that are characteristics of ultra-processed foods and used only in the manufacture of these products, they can be divided into food substances of no or rare culinary use and classes of additives ("cosmetic additives"). Food substances of no or rare culinary use include high fructose corn syrup, maltodextrin, hydrogenated or interesterified oils, and protein isolates. Classes of additives found only in ultra-processed products include cosmetic additives whose function is to make the final product palatable or often hyper-palatable. These additives include flavors, flavor enhancers, colors, emulsifiers, sweeteners, thickeners, and anti-foaming, bulking, carbonating, foaming, gelling, and glazing agents. Other additives in ultra-processed foods include some also used in processed foods, such as preservatives, antioxidants and stabilizers that prolong product duration, protect original properties or prevent proliferation of microorganisms. Typical examples of UPF are soft drinks, sweet or savory packaged snacks, breakfast cereal, candies, chocolate, "instant" soups and noodles, processed meats, pre-prepared frozen dishes; and many other products.

The main objective of the processes and ingredients used for the manufacture of UPFis to create products ready to eat, drink, or heat, attractive (hyper-palatable), and highly profitable (low-cost ingredients) food products designed to displace all other food groups, especially from Group 1. Box 1 shows some ways to recognize ultra-processed food.

Box 1 How to identify an ultra-processed food?

One of the most practical ways to identify if a product is an ultra-processed foods is to check its list of ingredients:

- Typically, these foods have a long list of ingredients. A product containing more than five ingredients is likely to be ultra-processed.
- Among these ingredients, at least contains one item characteristic of the ultra-processed foods: Food substances of no or rare culinary use or cosmetic additives, some of them with unrecognizable ingredients.

Global participation of ultra-processed foods

The global consumption of ultra-processed foods has increased dramatically (Juul and Hemmingsson, 2015; Marrón-Ponce et al., 2018; Monteiro et al., 2011, 2018). In high-income countries such as the United States (Baraldi et al., 2018), United Kingdom (Rauber et al., 2018), and Canada (Moubarac et al., 2017), ultra-processed foods already represent a significant percentage of about 50–60% of the total dietary energy consumed. In low- and middle-income countries, the participation of ultra-processed foods in the diet is lower than in high-income countries such as Brazil, Mexico, and Chile, ranging from 20 to 30 % (Cediel et al., 2018; Louzada et al., 2015b; Marrón-Ponce et al., 2018). In Brazil, trends analysis from Household Budget Surveys evidenced that the caloric share of UPFin household purchases increased from 8.6% in 2002–2003 to 18.4% in 2017–2018. As intake of ultra-processed foods increased, intake of minimally processed foods also decreased.

While UPF contribution is highest in high-income countries, sales of UPF in low- and middle-income countries are rising at a disproportionate rate compared to high-income countries (Vandevijvere et al., 2019). For example, Euromonitor reported a sales growth in UPF by 30% in Brazil from 2000 to 2013, while in the same period, sales dropped in the United States and Canada (−9% and −7.3%, respectively) (PAHO, 2015). These shifts evidence that food supplies are now becoming part of a global food system increasingly dominated by ready-to-consume ultra-processed foods.

As the sale and consumption of ultra-processed foods increase worldwide, children are a vulnerable group of being their leading consumers. The literature describes a high contribution of ultra-processed foods among children, with values ranging from 19.7 to 47% of total energy intake (Karnopp et al., 2017; Leffa et al., 2020; Rauber et al., 2015; Sparrenberger et al., 2015). In fact, the consumption of these foods begins in early life, even before the recommended timing of food introduction for infants. In Brazil, a cross-sectional study revealed that almost one-quarter (23.3%) of infants under six months consumed at least one UPF on the previous day, with high participation of soft drinks and cookies or crackers (Spaniol et al., 2020). The consumption of UPF in childhood is especially alarming given that dietary patterns in this period often track into adulthood (Ventura and Worobey, 2013).

Impact of ultra-processed foods on diet quality

Since the NOVA food classification system was proposed in 2010, several studies conducted in different countries have assessed the impact of UPF consumption on diet quality, particularly focusing on nutrients associated with non-communicable diseases. The high contribution of UPF in dietary patterns is associated with diets of lower nutritional quality, characterized by an excess intake of energy, free/added sugar, unhealthy fats (saturated and trans), sodium, and a lower intake of proteins and fibers (Cornwell et al., 2018; Louzada et al., 2015b; Martínez Steele et al., 2017; Moubarac et al., 2017; Rauber et al., 2019). A report of the Food and Agriculture Organization of the United Nations (FAO) summarized evidence of UPF intake and the nutrient profile of the overall diet from analyses of nationally representative data sets collected in 15 countries (Monteiro et al., 2019a). Briefly, the findings revealed that regardless of country and its dietary pattern, ultra-processed foods make the diet nutritionally unbalanced.

In particular, there is an increasing concern about the impact of ultra-processed foods on free/added sugar intake. The World Health Organization guidelines recommend limiting free sugars intake to less than 10% of total energy intake to prevent excess body weight and dental caries and less than 5% for additional health benefits (WHO, 2015). In the United States, a study with a nationally representative sample of the population showed that nearly 90% of the calories of added sugars came from ultra-processed foods (Martínez Steele et al., 2016). Moreover, US Americans in the highest quintile of ultra-processed foods intake exceeded three times the recommended limit of 10% energy from added sugars compared to those in the lowest quintile (82.1% and 26.4%, respectively).

Finally, these studies document that ultra-processed foods are clearly associated with an overall deterioration of the nutritional quality of diets. This has a potential risk for increasing non-communicable diseases, highlighting the critical impact of these foods consumption on public health.

Impact of ultra-processed foods on health

New research had provided convincing evidence supporting a number of mechanisms to quantify the burden of disease attributable to specific UPF consumption. Rigorous studies have shown that the higher consumption of ultra-processed foods is associated with an increased risk of overweight, obesity, metabolic syndrome, and related diseases, in addition to the increased risk of mortality from all causes. The results of the main evidence on this topic are summarized below:

Obesity

In adults, studies with representative samples of populations from different countries demonstrated a positive and independent association between the greater participation of ultra-processed foods in the diet and obesity (Canhada et al., 2019; Filgueiras et al., 2019; Juul and Hemmingsson, 2015; Louzada et al., 2015a; Machado et al., 2020; Rauber et al., 2020). Juul et al. in a cross-sectional study with a nationally representative sample of US adults found that the highest consumption of ultra-processed foods was associated with 53% and 62% higher odds of being obese and abdominal obesity, respectively (Juul et al., 2018).

The association between UPF exposure and obesity-related outcomes has recently been confirmed in a randomized controlled trial (Hall et al., 2019). Adults received either ultra-processed or unprocessed diets for 2 weeks immediately followed by the alternate diet for 2 weeks. The diets had a similar nutritional profile in calories, energy density, macronutrients, sugar, sodium, and fiber. When exposed to the ultra-processed diet, the results showed that the participants consumed, on average, 500 calories more per day and increased, on average, 0.9 kg of body weight in two weeks. Significant changes were also observed in the hormones involved in regulating hunger and food intake during the study period. This study is an important milestone in the research field of ultra-processed foods as it was the first to determine a potential mechanism behind the relationship between UPF consumption and weight gain.

Finally, a recent systematic review and meta-analysis, including the results of 14 studies and 189,966 participants ranging between 10 and 64 years, reported the association between UPF with overweight and obesity (Askari et al., 2020). In sensitivity analysis, higher UPF intake was associated with an increase risk of 33% (pooled effect size: 1.33; 95% CI:1.18, 1.49, $p < 0.001$) of being overweight and 73% (pooled effect size: 1.73; 95% CI: 1.36, 2.20, $p < 0.001$) of obesity. Thus, both clinical and epidemiological studies implicate UPF intake with a higher risk of developing obesity.

Cardiovascular and metabolic diseases

With regards to metabolic diseases, a population-based cohort study in France (NutriNet-Santé) longitudinally evaluated the effect of ultra-processed foods consumption of more than 100,000 French adults. Over a five-year period, a 10% increase in the consumption of ultra-processed foods diet was associated with a significant increase in the risk of cardiovascular diseases (12%), coronary heart disease (13%), and cerebrovascular disease (11%) (Srour et al., 2019). These findings are consistent with the results of the Framingham Offspring Study (Juul et al., 2021), which followed more than 3000 adults free from cardiovascular diseases at baseline for up to two decades, indicated that for each additional daily serving of ultra-processed the risk of CVD risk increased by a further 7%, and the risk of CVD mortality increased by 9%. Investigating the impact on blood pressure, a large prospective cohort of adults Spanish (SUN project) with a mean follow-up period of 9.1 years evidenced that higher consumption of ultra-processed foods increased by 23% the risk of developing hypertension (Mendonça et al., 2017). Likewise, in 2 cross-sectional studies conducted in Brazil (Scarannide et al., 2021) and Canada (Nardocci et al., 2021), participants with the highest compared with those in the lowest intake of ultra-processed had 23% and 60% higher odds of hypertension, respectively. Other studies indicate an UPF link to metabolic syndrome in US adults (Martínez Steele et al., 2019) and Brazilian adolescents (Tavares et al., 2012).

Consumption of ultra-processed foods may affect metabolic health even in early life. A longitudinal study of Brazilian children found an association between consumption of UPF at pre-school age and higher levels of total and LDL cholesterol (Rauber et al., 2015) and waist circumference at school age (Costa et al., 2019). Additionally, Leffa et al. in a longitudinal study with 308 Brazilian children from a low-income population, found that higher UPF consumption at age 3 years was associated with increased levels of total serum cholesterol later and triglycerides later at age 6 years (Leffa et al., 2020).

Diabetes

Ultra-processed foods also appear to increase the risk for type-2 diabetes. NutriNet-Santé French prospective cohort verified the dietary consumption of more than 100,000 diabetes-free adults over six years (Srour et al., 2020). The risk for developing diabetes went up 13% for a 10%-point increase in the amount of UPF in the diet, even after several adjustments for metabolic comorbidities and diet quality. Recently, in the SUN Spanish cohort study (Llavero-Valero et al., 2021), participants who reported the highest consumption of ultra-processed foods had the highest risk of diabetes (HR: 1.53, 1.06–2.22). The same association has been found in a large-scale prospective study with UK adults (HR: 1.44, 1.04–2.02) (Levy et al., 2020).

Cancer

A diet based on ultra-processed foods may lead to increased cancer incidence. The NutriNet-Santé French cohort, researchers discovered that a 10% increase in consumption of ultra-processed foods was associated with a 12% higher risk for cancer in general and an 11% increased risk for breast cancer (Fiolet et al., 2018). Additionally, recent research observed that colorectal cancer was found to increase in proportion to the intake of ultra-processed foods in Spanish adults. Specifically, a 10% increase in the intake of ultra-processed foods and drinks was associated with an 11% increase in the risk of colorectal cancer (Romaguera et al., 2021).

Mortality and aging

Eating ultra-processed foods may leave people more prone to accelerate biological aging and early death. In the SUN Spanish cohort study, higher consumption of ultra-processed foods, categorized by >4 servings daily, was independently associated with a 62% relatively increased hazard for all cause mortality (Rico-Campà et al., 2019). Similar results have been found among US (Kim et al., 2019), Italian (Bonaccio et al., 2021) and French (Schnabel et al., 2019) adults. Analyzing the role of ultra-processed foods on DNA damage, a cross-sectional study analyzed the health data of around 900 people aged 55 or older in Spain (Alonso-Pedrero et al., 2020). Participants who consumed three or more servings per day of ultra-processed foods had an 82% chance of having shortened telomeres, a biomarker of an individual's biological age at a cellular level.

Other diseases and risk factors

A higher intake of UPF has also been tied to increased other diseases and risk factors, such as early childhood caries (de Souza et al., 2021), depression (Zheng et al., 2020), and, frailty syndrome (Sandoval-Insausti et al., 2020). A recent systematic review and meta-analysis for the first time assessed the association between UPF consumption and health status (Pagliai et al., 2021). The pooled analysis confirmed that greater UPF consumption worse cardiometabolic risk profile (increased risk of overweight/obesity, elevated waist circumference, reduced HDL-cholesterol levels and increased risk of the metabolic syndrome), and greater risk of all-cause mortality, cardiovascular disease, cerebrovascular disease and depression. Finally, this section evidences the key role of UPF consumption in the development of chronic diseases.

Ultra-processed foods and chronic diseases: potential mechanisms of action beyond nutritional quality

As outlined earlier, ultra-processed products are not modified foods, recognizable as such, but formulations of industrial sources of dietary energy and nutrients. While the association between UPF and an unbalanced nutritional composition is of great importance, current evidence suggests that ultra-processed foods may affect health through several complex mechanisms and interactions between many of their compounds and characteristics, which are not yet fully understood.

The high content of fat, salt, sugar, and artificial flavorings make ultra-processed foods hyper-palatable, for which endogenous satiety mechanisms may be altered and are sometimes even quasi-addictive (Schulte et al., 2015). Modifications to the food matrix by processes in the manufacture of ultra-processed foods are key players in health-damaged potential, even compared to minimally or non-ultra-processed foods with similar nutritional composition. The changes in the physical and structural characteristics of the original food can interfere with endogenous processes of satiety and control appetite. Evidence supports that ultra-processed foods are, on average, less satiating are more hyperglycemic than unprocessed or minimally processed foods (Fardet, 2016). The reason would that ultra-processed foods contains less satiating nutrients in their composition, such as fiber and protein, and many of these products are semi-solid or liquid, requiring less chewing, leaving a too short time for the complete stimulation of satiety hormones.

In addition, the consumption of ultra-processed foods can result in pathological eating behavior and may also enable greater energy intake in a shorter amount of time. Indeed, a study based on pooled data from five studies revealed that the average eating rate during the ultra-processed foods diet was two times higher than the unprocessed and minimally processed diet (69.4 ± 3.1 vs. 35.5 ± 4.4 kcal/min) (Forde et al., 2020).

Ultra-processed foods are designed to be convenient, practical, and portable. Generally, they are developed to be consumed anywhere, any time, while working, in front of the television, walking or driving, or when using cell phones—and frequently do not require plates and cutlery. Most of the time, they are sold as snacks, drinks, or dishes ready-to-consume or heat up. Therefore, they can easily replace freshly made meals based on unprocessed or minimally processed foods. In addition, the portion size of ultra-processed foods, such as "fast foods", has increased significantly in recent decades (Piernas and Popkin, 2011). Studies have demonstrated that energy intake is directly influenced by the size of the served portion, with larger portions resulting in notable increases in energy intake, which may play a role in weight gain (Steenhuis and Vermeer, 2009).

Other potential mechanisms for the link between ultra-processed diets and diseases may be related to the interaction of chemical foods additives on metabolic alterations on gut microbiota that have been linked to obesity and correlated conditions (Zinöcker and Lindseth, 2018). For example, Chassaing et al. in an experimental study, showed that carboxymethylcellulose and polysorbate-80 (P80), both emulsifiers largely used in ultra-processed foods manufacture, induced low-grade inflammation and obesity/metabolic syndrome in mice (Chassaing et al., 2015). In a recent study, polysorbate-80 (P80) has been shown to reduced microbiota diversity (Furuhashi et al., 2020). Furthermore, accumulating evidence suggests that low-calorie sweetener consumption interferes with the gut microbiota and plays a role in developing insulin resistance (Suez et al., 2014). It is worth note that human data on long term health impacts are still lacking, and potential effects remain largely unknown for most additives.

Regarding the packaging of ultra-processed foods, synthetic compounds like bisphenol A (BPA) are omnipresent in packages. Its exposition has been reported to disrupt several metabolic functions, including enhancing adipocyte cell differentiation (Heindel et al., 2015). A recent systematic review and meta-analysis, including the results of 10 studies, reported that BPA exposure increased the risk of obesity by 11% (Wu et al., 2020).

Social determinants of ultra-processed foods consumption

The presentation and marketing of ultra-processed foods make these products preferable to unprocessed or minimally processed foods, promoting their overconsumption. Food companies have used many persuasive and aggressive marketing strategies in commercials and food package design to attract consumers' attention and drive their preferences, especially among children and young people (Mallarino et al., 2013). Marketing strategies commonly target at children include cartoon characters, photos of celebrities, bright colors, childish lettering, use of health statements and claims (Giménez et al., 2017; Pulker et al., 2018). These advertisements are intended to build upon children's positive affects toward a brand, and also create sensory and hedonic

expectations, impacting their food choices (Ares and Deliza, 2010; Folkvord et al., 2016). Thus, these mechanisms support that the marketing and the non-interpretation of food labeling increase the likelihood of consuming ultra-processed foods, potentially increasing the risk of obesity.

Strong policies, legislation, and frameworks on regulations of the marketing of ultra-processed foods are essential foundations for support better nutrition. As a model, since 2016, Chile has executed the world's most comprehensive regulatory efforts with actions for increasing people's awareness of unhealthy foods (Corvalán et al., 2019). Briefly, Chilean law of food labeling and advertising is a set of policies, including mandatory front-of-package warning-label system in foods that contain large amounts of critical nutrients (calories, saturated fat, sugar, and sodium) marketing restriction and banned sales of ultra-processed foods in schools. After a short period of the policy's implementation, a consumer survey has identified wide variations on household beverage purchases from before and 18 months after policy implementation (Taillie et al., 2020). Purchases of beverages high-in critical nutrients dropped nearly 24% during the period in Chile and similar reductions were observed across the different education levels of the population.

Lastly, implementing a multi-pronged integrated approach is an effective public strategy for reductions in unhealthy foods consumption. With proved and convincing evidence of the effectiveness of food regulation models, it should become a priority for government officials and civil society groups to define national policies and regulations.

NOVA in use

In the past decade, the food classification system based on the degree of processing has been increasingly recognized in reports and comments from the Pan American Health Organization (PAHO, 2015) and the UN Food and Agriculture Organization (Monteiro et al., 2019a). The classification has also been used to orient the nutrient profile model developed by the PAHO for the region of the Americas.

In Brazil, the level of food processing was addressed to the official Dietary Guidelines in 2014, which did Brazil pioneering in adopting the NOVA classification for its food recommendations (Brazil, 2014). The main four recommendations and "golden rule" of the Brazilian guidelines are:

1. "Make unprocessed or minimally processed foods the basis of your diet."
2. "Use processed culinary ingredients in small amounts for seasoning and cooking foods and to create culinary preparations."
3. "Limit the use of processed foods, consuming them in small amounts as components of culinary preparations or as part of meals based on natural or minimally processed foods."
4. "Avoid ultra-processed products"**And the golden rule**: "Always prefer natural or minimally processed foods and freshly made dishes and meals to ultra-processed products."

The significance of food processing, more importantly, limiting the consumption of ultra-processed food, is now also adopted in dietary recommendations in several countries, including Uruguay, Ecuador, Peru, Canada, France, and Israel.

Conclusion

Given that most foods are processed to some extent, the understanding of food processing, as proposed by the NOVA, a classification of foods based on their degree and purpose of industrial food processing, is crucial. There is a clear consistency in the data that the industrial processing in the fourth group of NOVA, the UPF group, is the issue. As ultra-processed products are becoming dominant in the global food supplies, the focus on food and the general quality of the diet instead of isolated nutrients, as proposed by the classification system, can improve the understanding of the role of ultra-processed foods in diet quality, besides providing information to assess and monitor dietary patterns of the population. In addition, with established and clear criteria, NOVA classification can be applied in all countries all over the world.

Perhaps more important than these broad changes in dietary patterns is the impact of ultra-processed foods as a risk factor for chronic non-communicable diseases. It has become clear that the dramatic increase in epidemics of overweight and obesity, type 2 diabetes, and cardiovascular diseases are associated with increased consumption of ultra-processed foods. There is strong evidence provided by several studies that ultra-processed foods may affect health through many biological, social, and environmental mechanisms, beyond the traditionally recognized individual nutrients.

There is an urgent need for double and triple duty actions focused on minimizing the overall consumption of ultra-processed foods and reduce the health risk provided by these products. In order to reduce the consumption of ultra-processed food, this requires the implementation of a set of fiscal policies and regulations of ultra-processed product labeling, promotion, and advertising. Policy makers should consider actions to make unprocessed and minimally processed foods more valued, and more available, and affordable. In the mean time, the solid data leads governments to recommend prioritizing the consumption of unprocessed/minimally processed foods, and avoiding the consumption of UPF in order to promote the public health.

References

Alonso-Pedrero, L., Ojeda-Rodríguez, A., Martínez-González, M.A., Zalba, G., Bes-Rastrollo, M., Marti, A., 2020. Ultra-processed food consumption and the risk of short telomeres in an elderly population of the Seguimiento Universidad de Navarra (SUN) Project. Am. J. Clin. Nutr. 111 (6), 1259–1266.

Ares, G., Deliza, R., 2010. Studying the influence of package shape and color on consumer expectations of milk desserts using word association and conjoint analysis. Food Qual. Prefer. 21 (8), 930–937.

Askari, M., Heshmati, J., Shahinfar, H., Tripathi, N., Daneshzad, E., 2020. Ultra-processed food and the risk of overweight and obesity: a systematic review and meta-analysis of observational studies. Int. J. Obes. 44 (10), 2080–2091.

Baraldi, L.G., Martinez Steele, E., Canella, D.S., Monteiro, C.A., 2018. Consumption of ultra-processed foods and associated sociodemographic factors in the USA between 2007 and 2012: evidence from a nationally representative cross-sectional study. BMJ Open 8 (3), e020574.

Bonaccio, M., Di Castelnuovo, A., Costanzo, S., De Curtis, A., Persichillo, M., Sofi, F., Cerletti, C., et al., 2021. Ultra-processed food consumption is associated with increased risk of all-cause and cardiovascular mortality in the Moli-Sani Study. Am. J. Clin. Nutr. 113 (2), 446–455.

Brasil, 2014. Guia Alimentar Para a População Brasileira, second ed. Ministério da Saúde, Brasília.

Canhada, S.L., Luft, V.C., Giatti, L., Duncan, B.B., Chor, D., Fonseca, M.J.M.D., Matos, S.M.A., et al., 2019. Ultra-processed foods, incident overweight and obesity, and longitudinal changes in weight and waist circumference: the Brazilian Longitudinal Study of Adult Health (ELSA-Brazil). Publ. Health Nutr. 1–11.

Cediel, G., Reyes, M., da Costa Louzada, M.L., Martinez Steele, E., Monteiro, C.A., Corvalán, C., Uauy, R., 2018. Ultra-processed foods and added sugars in the Chilean diet (2010). Publ. Health Nutr. 21 (1), 125–133.

Chassaing, B., Koren, O., Goodrich, J.K., Poole, A.C., Srinivasan, S., Ley, R.E., Gewirtz, A.T., 2015. Dietary emulsifiers impact the mouse gut microbiota promoting colitis and metabolic syndrome. Nature 519 (7541), 92–96.

Cornwell, B., Villamor, E., Mora-Plazas, M., Marin, C., Monteiro, C.A., Baylin, A., 2018. Processed and ultra-processed foods are associated with lower-quality nutrient profiles in children from Colombia. Publ. Health Nutr. 21 (1), 142–147.

Corvalán, C., Reyes, M., Garmendia, M.L., Uauy, R., 2019. Structural responses to the obesity and non-communicable diseases epidemic: update on the Chilean law of food labeling and advertising. Obes. Rev. 20 (3), 367–374.

Costa, C.S., Rauber, F., Leffa, P.S., Sangalli, C.N., Campagnolo, P.D.B., Vitolo, M.R., 2019. Ultra-processed food consumption and its effects on anthropometric and glucose profile: a longitudinal study during childhood. Nutr. Metabol. Cardiovasc. Dis. 29 (2), 177–184.

de Souza, M.S., Vaz, J.D.S., Martins-Silva, T., Bomfim, R.A., Cascaes, A.M., 2021. Ultra-processed foods and early childhood caries in 0–3-year-olds enrolled at Primary Healthcare Centers in Southern Brazil–ERRATUM. Publ. Health Nutr. 24 (2), 377.

Fardet, A., 2016. Minimally processed foods are more satiating and less hyperglycemic than ultra-processed foods: a preliminary study with 98 ready-to-eat foods. Food Funct. 7 (5), 2338–2346.

Filgueiras, A.R., Pires de Almeida, V.B., Koch Nogueira, P.C., Alvares Domene, S.M., Eduardo da Silva, C., Sesso, R., Sawaya, A.L., 2019. Exploring the consumption of ultra-processed foods and its association with food addiction in overweight children. Appetite 135, 137–145.

Fiolet, T., Srour, B., Sellem, L., Kesse-Guyot, E., Allès, B., Méjean, C., Deschasaux, M., et al., 2018. Consumption of ultra-processed foods and cancer risk: results from NutriNet-Santé prospective cohort. Br. Med. J. 360, k322.

Folkvord, F., Anschütz, D.J., Boyland, E., Kelly, B., Buijzen, M., 2016. Food advertising and eating behavior in children. Curr. Opin. Behav. Sci. 9, 26–31.

Forde, C.G., Mars, M., de Graaf, K., 2020. Ultra-processing or oral processing? A role for energy density and eating rate in moderating energy intake from processed foods. Curr. Dev. Nutr. 4 (3), nzaa019.

Furuhashi, H., Higashiyama, M., Okada, Y., Kurihara, C., Wada, A., Horiuchi, K., Hanawa, Y., et al., 2020. Dietary emulsifier polysorbate-80-induced small-intestinal vulnerability to indomethacin-induced lesions via dysbiosis. J. Gastroenterol. Hepatol. 35 (1), 110–117.

Giménez, A., Saldamando, L. de, Curutchet, M.R., Ares, G., 2017. Package design and nutritional profile of foods targeted at children in supermarkets in Montevideo, Uruguay. Cad. Saúde Pública 33 (5). https://doi.org/10.1590/0102-311x00032116. Available at.

Hall, K.D., Ayuketah, A., Brychta, R., Cai, H., Cassimatis, T., Chen, K.Y., Chung, S.T., et al., 2019. Ultra-processed diets cause excess calorie intake and weight gain: an inpatient randomized controlled trial of Ad Libitum food intake. Cell Metabol. 30 (1), 67–77.e3.

Heindel, J.J., Newbold, R., Schug, T.T., 2015. Endocrine disruptors and obesity. Nat. Rev. Endocrinol. 11 (11), 653–661.

Juul, F., Hemmingsson, E., 2015. Trends in consumption of ultra-processed foods and obesity in Sweden between 1960 and 2010. Publ. Health Nutr. 18 (17), 3096–3107.

Juul, F., Martinez-Steele, E., Parekh, N., Monteiro, C.A., Chang, V.W., 2018. Ultra-processed food consumption and excess weight among US adults. Br. J. Nutr. 120 (1), 90–100.

Juul, F., Vaidean, G., Lin, Y., Deierlein, A.L., Parekh, N., 2021. Ultra-processed foods and incident cardiovascular disease in the Framingham Offspring Study. J. Am. Coll. Cardiol. 77 (12), 1520–1531.

Karnopp, E.V.N., Vaz, J.D.S., Schafer, A.A., Muniz, L.C., Souza, R.L., Santos, I.D., Gigante, D.P., et al., 2017. Food consumption of children younger than 6 years according to the degree of food processing. J. Pediatr. 93 (1), 70–78.

Kim, H., Hu, E.A., Rebholz, C.M., 2019. Ultra-processed food intake and mortality in the USA: results from the Third National Health and Nutrition Examination Survey (NHANES III, 1988–1994). Publ. Health Nutr. 22 (10), 1777–1785.

Leffa, P.S., Hoffman, D.J., Rauber, F., Sangalli, C.N., Valmórbida, J.L., Vitolo, M.R., 2020. Longitudinal associations between ultra-processed foods and blood lipids in childhood. Br. J. Nutr. 124 (3), 341–348.

Levy, R.B., Rauber, F., Chang, K., Louzada, M.L.C., Monteiro, C.A., Millett, C., Vamos, E.P., 2020. Ultra-processed food consumption and type 2 diabetes incidence: a prospective cohort study. Clin. Nutr. https://doi.org/10.1016/j.clnu.2020.12.018. Available at.

Llavero-Valero, M., Escalada-San Martín, J., Martínez-González, M.A., Basterra-Gortari, F.J., de la Fuente-Arrillaga, C., Bes-Rastrollo, M., 2021. Ultra-processed foods and type-2 diabetes risk in the SUN project: a prospective cohort study. Clin. Nutr. 40 (5), 2817–2824.

Louzada, M.L.C., Baraldi, L.G., Steele, E.M., Martins, A.P.B., Canella, D.S., Moubarac, J.C., Levy, R.B., et al., 2015a. Consumption of ultra-processed foods and obesity in Brazilian adolescents and adults. Prev. Med. 81, 9–15.

Louzada, M.L.C., Martins, A.P.B., Canella, D.S., Baraldi, L.G., Levy, R.B., Claro, R.M., Moubarac, J.C., et al., 2015b. Ultra-processed foods and the nutritional dietary profile in Brazil. Rev. Saude Publica 49. https://doi.org/10.1590/S0034-8910.2015049006132. Available at.

Ludwig, D.S., 2011. Technology, diet, and the burden of chronic disease. J. Am. Med. Assoc. 305 (13), 1352–1353.

Machado, P.P., Steele, E.M., Levy, R.B., da Costa Louzada, M.L., Rangan, A., Woods, J., Gill, T., et al., 2020. Ultra-processed food consumption and obesity in the Australian adult population. Nutr. Diabetes 10 (1), 39.

Mallarino, C., Gómez, L.F., González-Zapata, L., Cadena, Y., Parra, D.C., 2013. Advertising of ultra-processed foods and beverages: children as a vulnerable population. Rev. Saude Publica 47 (5), 1006–1010.

Marrón-Ponce, J.A., Tolentino-Mayo, L., Hernández-F, M., Batis, C., 2018. Trends in ultra-processed food purchases from 1984 to 2016 in Mexican households. Nutrients 11 (1). https://doi.org/10.3390/nu11010045. Available at.

Martínez Steele, E., Baraldi, L.G., Louzada, M.L.C., Moubarac, J.C., Mozaffarian, D., Monteiro, C.A., 2016. Ultra-processed foods and added sugars in the US diet: evidence from a nationally representative cross-sectional study. BMJ Open 6 (3), e009892.

Martínez Steele, E., Popkin, B.M., Swinburn, B., Monteiro, C.A., 2017. The share of ultra-processed foods and the overall nutritional quality of diets in the US: evidence from a nationally representative cross-sectional study. Popul. Health Metrics 15 (1), 6.
Martínez Steele, E., Juul, F., Neri, D., Rauber, F., Monteiro, C.A., 2019. Dietary share of ultra-processed foods and metabolic syndrome in the US adult population. Prev. Med. 125, 40–48.
Mendonça, R.D., Lopes, A.C.S., Pimenta, A.M., Gea, A., Martinez-Gonzalez, M.A., Bes-Rastrollo, M., 2017. Ultra-processed food consumption and the incidence of hypertension in a mediterranean cohort: the Seguimiento Universidad de Navarra Project. Am. J. Hypertens. 30 (4), 358–366.
Monteiro, C.A., Levy, R.B., Claro, R.M., Castro, I.R.R., Cannon, G., 2010. A new classification of foods based on the extent and purpose of their processing. Cad. Saúde Pública 26 (11), 2039–2049.
Monteiro, C.A., Levy, R.B., Claro, R.M., de Castro, I.R.R., Cannon, G., 2011. Increasing consumption of ultra-processed foods and likely impact on human health: evidence from Brazil. Publ. Health Nutr. 14 (1), 5–13.
Monteiro, C.A., Moubarac, J.C., Levy, R.B., Canella, D.S., Louzada, M.L.C., Cannon, G., 2018. Household availability of ultra-processed foods and obesity in nineteen European countries. Publ. Health Nutr. 21 (1), 18–26.
Monteiro, C.A., Cannon, G., Lawrence, M., Costa Louzada, M.L., Pereira Machado, P., 2019a. Ultra-Processed Foods, Diet Quality, and Health Using the NOVA Classification System. FAO, Rome.
Monteiro, C.A., Cannon, G., Levy, R.B., Moubarac, J.-C., Louzada, M.L., Rauber, F., Khandpur, N., et al., 2019b. Ultra-processed foods: what they are and how to identify them. Publ. Health Nutr. 22 (5), 936–941.
Moubarac, J.C., Batal, M., Louzada, M.L., Martinez Steele, E., Monteiro, C.A., 2017. Consumption of ultra-processed foods predicts diet quality in Canada. Appetite 108, 512–520.
Nardocci, M., Polsky, J.Y., Moubarac, J.-C., 2021. Consumption of ultra-processed foods is associated with obesity, diabetes and hypertension in Canadian adults. Can. J. Publ. Health 112 (3), 421–429.
Pagliai, G., Dinu, M., Madarena, M.P., Bonaccio, M., Iacoviello, L., Sofi, F., 2021. Consumption of ultra-processed foods and health status: a systematic review and meta-analysis. Br. J. Nutr. 125 (3), 308–318.
PAHO, 2015. Ultra-Processed Food and Drink Products in Latin America: Trends, Impact on Obesity, Policy Implications. PAHO, Washington, DC.
Piernas, C., Popkin, B.M., 2011. Food portion patterns and trends among U.S. children and the relationship to total eating occasion size, 1977–2006. J. Nutr. 141 (6), 1159–1164.
Popkin, B.M., Adair, L.S., Ng, S.W., 2012. Global nutrition transition and the pandemic of obesity in developing countries. Nutr. Rev. 70 (1), 3–21.
Popkin, B.M., 2004. The nutrition transition: an overview of world patterns of change. Nutr. Rev. 62 (7 Pt 2), S140–S143.
Pulker, C.E., Scott, J.A., Pollard, C.M., 2018. Ultra-processed family foods in Australia: nutrition claims, health claims and marketing techniques. Publ. Health Nutr. 21 (1), 38–48.
Rauber, F., Campagnolo, P.D.B., Hoffman, D.J., Vitolo, M.R., 2015. Consumption of ultra-processed food products and its effects on children's lipid profiles: a longitudinal study. Nutr. Metabol. Cardiovasc. Dis. 25 (1), 116–122.
Rauber, F., da Costa Louzada, M.L., Steele, E.M., Millett, C., Monteiro, C.A., Levy, R.B., 2018. Ultra-processed food consumption and chronic non-communicable diseases-related dietary nutrient profile in the UK (2008–2014). Nutrients 10 (5). https://doi.org/10.3390/nu10050587. Available at.
Rauber, F., Louzada, M.L.C., Martinez Steele, E., Rezende, L.F.M., Millett, C., Monteiro, C.A., Levy, R.B., 2019. Ultra-processed foods and excessive free sugar intake in the UK: a nationally representative cross-sectional study. BMJ Open 9 (10), e027546.
Rauber, F., Chang, K., Vamos, E.P., da Costa Louzada, M.L., Monteiro, C.A., Millett, C., Levy, R.B., 2020. Ultra-processed food consumption and risk of obesity: a prospective cohort study of UK Biobank. Eur. J. Nutr. https://doi.org/10.1007/s00394-020-02367-1. Available at.
Rico-Campà, A., Martínez-González, M.A., Alvarez-Alvarez, I., Mendonça, R.D., de la Fuente-Arrillaga, C., Gómez-Donoso, C., Bes-Rastrollo, M., 2019. Association between consumption of ultra-processed foods and all cause mortality: SUN prospective cohort study. Br. Med. J. 365, l1949.
Romaguera, D., Fernández-Barrés, S., Gracia-Lavedán, E., Vendrell, E., Azpiri, M., Ruiz-Moreno, E., Martín, V., et al., 2021. Consumption of ultra-processed foods and drinks and colorectal, breast, and prostate cancer. Clin. Nutr. 40 (4), 1537–1545.
Sandoval-Insausti, H., Blanco-Rojo, R., Graciani, A., López-García, E., Moreno-Franco, B., Laclaustra, M., Donat-Vargas, C., et al., 2020. Ultra-processed food consumption and incident frailty: a prospective cohort study of older adults. J. Gerontol. A Biol. Sci. Med. Sci. 75 (6), 1126–1133.
Scaranni, P.O.D.S., Cardoso, L.O., Chor, D., Melo, E.C.P., Matos, S.M.A., Giatti, L., Barreto, S.M., et al., 2021. Ultra-processed foods, changes in blood pressure and incidence of hypertension: the Brazilian Longitudinal Study of Adult Health (ELSA-Brazil). Publ. Health Nutr. 1–9.
Schnabel, L., Kesse-Guyot, E., Allès, B., Touvier, M., Srour, B., Hercberg, S., Buscail, C., et al., 2019. Association between ultraprocessed food consumption and risk of mortality among middle-aged adults in France. JAMA Intern. Med. 179 (4), 490.
Schulte, E.M., Avena, N.M., Gearhardt, A.N., 2015. Which foods may be addictive? The roles of processing, fat content, and glycemic load. PLoS One 10 (2), e0117959.
Spaniol, A.M., da Costa, T.H.M., Souza, A.M., Gubert, M.B., 2020. Early consumption of ultra-processed foods among children under 2 years old in Brazil. Publ. Health Nutr. 1–11.
Sparrenberger, K., Friedrich, R.R., Schiffner, M.D., Schuch, I., Wagner, M.B., 2015. Ultra-processed food consumption in children from a Basic Health Unit. J. Pediatr. 91 (6), 535–542.
Srour, B., Fezeu, L.K., Kesse-Guyot, E., Allès, B., Méjean, C., Andrianasolo, R.M., Chazelas, E., et al., 2019. Ultra-processed food intake and risk of cardiovascular disease: prospective cohort study (NutriNet-Santé). Br. Med. J. 365, l1451.
Srour, B., Fezeu, L.K., Kesse-Guyot, E., Allès, B., Debras, C., Druesne-Pecollo, N., Chazelas, E., et al., 2020. Ultraprocessed food consumption and risk of type 2 diabetes among participants of the NutriNet-Santé prospective cohort. JAMA Intern. Med. 180 (2), 283.
Steenhuis, I.H., Vermeer, W.M., 2009. Portion size: review and framework for interventions. Int. J. Behav. Nutr. Phys. Activ. 6, 58.
Suez, J., Korem, T., Zeevi, D., Zilberman-Schapira, G., Thaiss, C.A., Maza, O., Israeli, D., et al., 2014. Artificial sweeteners induce glucose intolerance by altering the gut microbiota. Nature 514 (7521), 181–186.
Swinburn, B.A., Sacks, G., Hall, K.D., McPherson, K., Finegood, D.T., Moodie, M.L., Gortmaker, S.L., 2011. The global obesity pandemic: shaped by global drivers and local environments. Lancet 378 (9793), 804–814.
Taillie, L.S., Reyes, M., Colchero, M.A., Popkin, B., Corvalán, C., 2020. An evaluation of Chile's Law of Food Labeling and Advertising on sugar-sweetened beverage purchases from 2015 to 2017: a before-and-after study. PLoS Med. 17 (2), e1003015.
Tavares, L.F., Fonseca, S.C., Garcia Rosa, M.L., Yokoo, E.M., 2012. Relationship between ultra-processed foods and metabolic syndrome in adolescents from a Brazilian Family Doctor Program. Publ. Health Nutr. 15 (1), 82–87.
Vandevijvere, S., Jaacks, L.M., Monteiro, C.A., Moubarac, J.-C., Girling-Butcher, M., Lee, A.C., Pan, A., et al., 2019. Global trends in ultraprocessed food and drink product sales and their association with adult body mass index trajectories. Obes. Rev. https://doi.org/10.1111/obr.12860. Available at.
Ventura, A.K., Worobey, J., 2013. Early influences on the development of food preferences. Curr. Biol. 23 (9), R401–R408.
WHO, 2015. Guideline: Sugars Intake for Adults and Children. World Health Organization, Geneva. Available at. https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=2033879. (Accessed 15 May 2021).
World Health Organization (Ed.), 1998. Preparation and Use of Food-Based Dietary Guidelines: Report of a Joint FAO/WHO Consultation. WHO, Geneva.

World Health Organization, 2003. Expert consultation on diet, nutrition, and the prevention of chronic diseases. In: WHO, FAO (Eds.), Diet, Nutrition, and the Prevention of Chronic Diseases: Report of a WHO-FAO Expert Consultation; [Joint WHO-FAO Expert Consultation on Diet, Nutrition, and the Prevention of Chronic Diseases, 2002, Geneva, Switzerland]. World Health Organization, Geneva. Presented at the Joint WHO-FAO Expert Consultation on Diet, Nutrition, and the Prevention of Chronic Diseases.

Wrangham, R., 2013. The evolution of human nutrition. Curr. Biol. 23 (9), R354–R355.

Wu, W., Li, M., Liu, A., Wu, C., Li, D., Deng, Q., Zhang, B., et al., 2020. Bisphenol A and the risk of obesity a systematic review with meta-analysis of the epidemiological evidence. Dose Response 18 (2), 1559325820916949.

Zheng, L., Sun, J., Yu, X., Zhang, D., 2020. Ultra-processed food is positively associated with depressive symptoms among United States adults. Front. Nutr. 7, 600449.

Zinöcker, M., Lindseth, I., 2018. The western diet-microbiome-host interaction and its role in metabolic disease. Nutrients 10 (3), 365.

Vegetarian diets

J. Dwyer and J. Harvey, Frances Stern Nutrition Center, Tufts Medical Center and Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of J. Dwyer, Vegetarian Diets, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 316–322, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00272-5.

Introduction 663
History 663
Vegetarian patterns and practices 665
Differences between vegetarian eating and vegetarianism 665
Nutritional features of well-planned vegetarian diets 666
Morbidity and mortality on vegetarian versus omnivorous diets 667
Nutritional adequacy 670
Adequate vegetarian dietary patterns 671
Possibly inadequate vegetarian dietary patterns 671
Special nutritional concerns for vegetarians 672
Key nutrients for vegetarians over the life cycle 672
Conclusions 673
References 673
Further reading 673

Key points

- To review current vegetarian eating patterns and practices in Western countries and their health effects.
- To define and describe different vegetarian eating patterns.
- To differentiate between vegetarian eating and vegetarianism.
- To understand nutritional benefits and the key nutrients of concern for followers of vegetarian eating patterns.

Introduction

Vegetarian eating patterns are the norm in many parts of the world, while in Western countries they are still the exception but are gaining in popularity. This article examines the healthfulness of those vegetarian eating patterns and practices, and their benefits and risks. Vegetarian patterns in other countries, such as India, Southeast Asia and some low-income developing countries, are quite different and are not reviewed here. Currently, approximately 5% of adults in the US and 4% of adults in Canada report that they follow some sort of self-described vegetarian diet, and a far lesser percent claim to eat no animal foods at all. Approximately a fifth of American adults report that they do without animal protein at more than four main meals a week (NHANES, 2009–2010).

Table 1 describes common vegetarian eating styles in Western countries by use of animal foods. However, those patterns also often differ in many other ways that have nutritional and health implications, and so some of the other differences are noted as well.

History

Vegetarian diets have been eaten for millennia and continue to be popular for cultural and economic reasons in many parts of the world. In contrast, in Western countries, until 50 years ago, these patterns were uncommon. The vegetarian eating styles tended to be largely among individuals or groups with philosophical, religious or moral motivations for their consumption or among immigrants with vegetarian cultural traditions. The dominant vegetarian eating pattern was lacto-ovo, with avoidance of animal flesh (meat and poultry) at a minimum. In those days, categorization of vegetarian patterns was relatively straightforward and consisted simply of differentiating between those who ate no animal foods at all (vegan vegetarians), those who consumed only milk and milk products (lacto vegetarians), and those who ate eggs as well (lacto-ovo vegetarians). This simple categorization scheme broke down in the 1960s and 1970s when new patterns of vegetarian eating emerged as a result of Westerners' greater exposure to the cuisines of other cultures, the influence of new variants of Eastern religions and philosophical systems with a vegetarian tradition, growing popularity of economic, environmental, and sustainability arguments for eating fewer animal foods, and evidence that some vegetarian eating patterns were associated with decreased risks of certain chronic degenerative diseases. Consumers are

Table 1 Common types of vegetarian dietary patterns in the western countries categorized by animal food use.

Pattern	*Animal food use and other characteristics*
Omnivorous	All culturally acceptable and economically available animal foods are consumed with no specific prohibitions on any food group.
Plant-based diet	A plant-based diet is usually defined as an eating pattern that focuses the consumption of foods primarily from plant sources (i.e., fruits, vegetables, nuts, seeds, whole grains, legumes, and beans). Meat, poultry, fish, and dairy products are still consumed but in limited quantities. However, the term is variously defined and thus confusing. According to the US Department of Agriculture's Economic Research Service, 70% of calories in US diets are from plant-based foods, 17% from meat, poultry and fish, and 13% from eggs and dairy.
DASH diet	The Dietary approaches to Stop Hypertension (DASH) diet is a plant-based diet high in low-fat-milk (2–3 servings per day), low in lean meat, poultry, fish (6 oz total per day), eggs, and very limited in red meat. It is high in fruits, vegetables, whole grains, nuts, and legumes, and limited in sugar-sweetened foods and beverages, sodium, added fats, and saturated/trans fats. DASH is one of the recommended eating patterns in the Dietary Guidelines for Americans 2020.
"Flexitarian" diet	"Flexitarian" patterns are plant-based and include meat, dairy, eggs, poultry, and fish/seafood on occasion or only in small quantities. Other food groups are rarely avoided entirely, but specific foods (e.g., those high in "added sugars", sodium, trans fat, saturated fat, or those considered "processed", "ultra-processed", "nonorganic", or that contain "chemical sounding" or certain specific ingredients etc.) may be avoided by some flexitarians.
Meat avoiding, meatless, or semi-vegetarian	Semi-vegetarians limit or avoid red meat and other flesh foods and may also restrict poultry, fish, and seafood. Diets are similar in most other respects to nonvegetarian diets.
Lacto-ovo vegetarian	Lacto-ovo vegetarians avoid all meat, poultry, and often fish, but milk products (especially low-fat products) and eggs are consumed. Iron, vitamin D, vitamin E, and choline may be limiting but are available from fortified cereals or supplements if those are eaten.
Lacto vegetarian	Lacto-vegetarians avoid all meat, fish, poultry, and eggs. Iron and choline may be low in these patterns.
Macrobiotic	Numerous restrictions include avoidance of all meat, poultry, milk, and eggs, with fish sometimes consumed in small amounts. Sugar, other refined sweeteners, members of the nightshade family (peppers, eggplant, tomatoes, and potatoes), and tropical fruits are avoided on some macrobiotic patterns. Current versions of the diet are possibly less restrictive than those of 40 years ago, but deficiencies of energy, iron, calcium, vitamin B_{12}, vitamin D, choline, omega 3 fatty acids and other nutrients may still arise in weanlings, pregnant women, young children, and adolescents during the growth spurt especially if diets are nutritionally unplanned and nutrient containing dietary supplements and fortified foods are avoided.
Vegan	Vegans' avoidances include all animal products including meat, fish, poultry, eggs, and dairy products. Some may also refuse to use animal products (leather, fur, lipstick) in daily life. Without careful planning, energy, vitamins B_{12} and D, choline, essential fatty acids, and bioavailable sources of iron and zinc may be low. Concentrated sources of energy-dense foods, such as sugars and fats are helpful for increasing energy intakes. Vitamins B_{12}, D and calcium are available in fortified soymilk, fortified cereals, and dietary supplements of these nutrients. Usually, protein is adequate as long as a variety of protein sources is consumed in sufficient amounts.
Vegetarian	The term refers to a range of different dietary patterns, all of which include avoidance of some or all animal foods to varying degrees, with or without additional restrictions, additions, or prohibitions on other foods or ingredients. Depending on their composition, vegetarian diets may be poor sources of protein of high biological value, highly bioavailable iron, vitamins A, D, B_2, B_6, and B_{12}, zinc, omega-3 fatty acids, choline, calcium, and sometimes iodine.
Vegetarianism	A larger belief system that is reflected not only in the eating pattern but also encompassing many other aspects of lifestyle and thought. Vegetarianism involves philosophy, values, beliefs, or religious convictions rather than simply food choices. Its adherents often hold strong, deeply held convictions about the moral, metaphysical, ethical, religious, or political appropriateness of vegetarian eating.
"Healthy vegetarian diet"	Healthy vegetarian diet is a term that refers to the food pattern recommended in the 2015 and 2020 Dietary Guidelines for Americans. In essence, it is a lacto-ovo vegetarian dietary pattern created by computer modeling using all foods in their most nutrient dense, lean or lower-fat forms, with recipes prepared using a minimal amount of added fats, sugars, refined starch, and sodium. Iron, vitamin D, vitamin E, and choline are somewhat limited but approach the levels suggested in the Recommended Dietary Allowances of the National Academies of Science, Engineering and Medicine.
Whole food	No regulatory definition of "whole food" exists but the term is often used to refer to foods that are minimally processed with no added ingredients, or any food that appears close to its original form.
Other patterns	Raw food eaters and "living food" eaters avoid animal foods and eat raw plant foods (fruits, vegetables, nuts, legumes, and cereals). They may also consume special foods that they believe to be especially healthful, such as wheatgrass or carrot juice. Fruitarians consume diets mostly of fruits, nuts, honey, and olive oil. Rastafarians eat a near-vegan diet and avoid alcohol, salt-preserved foods, and additives. Yogic groups (Transcendental Meditation, Hare Krishna, Bab Ram Das, etc.) vary in their eating patterns but are often lacto vegetarian.

now much better informed and concerned about nutrition than they were years ago. Those who follow vegetarian patterns today have complex and varied rationales that involve food preferences, environmental, economic, sustainability, and health concerns rather than solely religious or ethical reasons for adopting vegetarian eating patterns. Partial or semi-vegetarians often initially adopt a vegetarian diet for health or environmental sustainability reasons and in some instances their commitments then broaden or are augmented by other reasons for sustaining or further restricting their diets. Their eating patterns also include additions and restrictions on other foods and ingredients beyond animal foods alone, making it difficult to easily categorize either motivations or diets from the standpoints of their plant or animal constituents.

Many authoritative bodies now suggest that "plant-based" diets confer a health advantage, although the term "plant based" is rarely defined quantitatively. For example, evidence from large observational studies suggests that plant-based eating patterns rich in vegetables, beans, nuts, whole grains, fish, and plant oils have cardiovascular benefits. The mass media provides much favorable publicity about phytochemicals such as polyphenols, glucosinolates and other non-nutrient phytochemical bioactives in plant foods that supposedly have beneficial health effects on health, but evidence is still limited except for dietary fiber. At the same time, fears and concerns about the healthfulness of animal foods were retriggered in the late 20th century by the presence of an outbreak of bovine spongiform encephalopathy (BSE) in the United Kingdom and Europe, epidemics of hoof and mouth disease in cattle, and more recently severe acute respiratory syndrome (SARS), swine influenza, and today the COVID 19 epidemic that originated in animals and spread to people. Massive outbreaks of animal food borne illnesses, such as Salmonellosis caused by contamination of shell eggs and others, such as listeriosis, campylobacteriosis and *Escherichia coli* O157:H7 infections also stoke consumer concerns about animal foods (Richardson et al., 2021). What often goes unrecognized today is that many food borne illnesses including those caused by Salmonella, *E. coli*, Listeria and many viral illnesses can be contacted through contamination and failure to wash and store raw fruits and vegetables, as recent outbreaks of disease associated with sprouts and other raw foods have shown. Worries are also rife that "Western" diets rich in red meat, milk products, eggs and other animal foods contain saturated fat and cholesterol that increase risks of coronary artery disease, and that other components of animal foods cause certain cancers, type 2 diabetes, and other diseases. Over the past decade there has also been a growing emphasis in the mass media and by governments to pay more attention to environmental sustainability of national and global food supplies, and to the contributions of animal production agriculture to global warming. Such health and environmental concerns have probably also contributed to the increased prevalence of semi vegetarian and vegetarian eating styles. As a result of these influences, not only plant-based, but also meatless and vegetarian eating patterns have grown in popularity. Although "plant-based" diets do not conform strictly to the definition of a vegetarian diet, they are clearly different from usual intake patterns in that the type and amount of animal foods is substantially decreased (especi(especially red meat) (especially red meat) and many who eat them now consider themselves to be "vegetarians".

Vegetarian patterns and practices

The term "vegetarian diet" does not fully describe the variety in dietary and nutrient intakes and associated health status of followers of such eating patterns in Western countries. Today, myriad vegetarian eating patterns exist that are not easily described by focusing on a single dimension, such as animal food intake (see **Table 1**). The impact of these patterns on nutritional status and health requires more complete characterization of diet and other aspects of lifestyle than a simple description of which animal foods are left over to eat after others have been omitted from the diet. It is also important to characterize what other foods are eaten in greater detail if one is to obtain an accurate profile of the diet's nutrient adequacy. Traditionally, vegetarian diets were poor sources of energy, protein of high biological value, vitamins A, D, B_2, B_6, and B_{12}, zinc, bioavailable iron, omega-3 fatty acids, choline, calcium, and sometimes iodine. Today a wide variety of plant foods are available year around in most Western countries, along with fortified foods and nutrient containing dietary supplements as well as meat, poultry, milk and egg analogs that can fill these nutrient gaps. However, for a variety of reasons not all vegetarians are willing to use these products, and so risks of inadequacy still remain for some individuals.

Differences between vegetarian eating and vegetarianism

A vegetarian *eating style or pattern* is often confused with vegetarianism. In the world as a whole, most vegetarians have eating styles that are dictated chiefly by culture, habit, economic, food preferences, health cost, taste and environmental concerns rather than solely moral philosophical, religious, or ethics reasons. A small minority of vegetarians also subscribe to *vegetarianism*, consisting of one of many larger belief systems that are reflected in their eating pattern but that also encompass many other aspects of lifestyle and thought. Vegetarianism involves a variety of philosophies, values, beliefs, or religious convictions rather than simply food choices. Its adherents often hold strong, deeply held convictions about the moral, metaphysical, ethical, religious, or political appropriateness of vegetarian eating. Often those with the most restrictive avoidances of animal and other foods, such as vegans, have the most deeply held views. Advocates of vegetarianism have become more militant in recent years in Western countries. Some adherents, such as some but not all yogic groups, simply hope to enlist additional followers to what they believe is the optimal diet from their ethical, religious, health, and/or environmental standpoints. Those who adhere to the Seventh Day Adventist faith as well as to faiths such as Hinduism, Buddhism, and Jainism often follow vegetarian dietary patterns for religious reasons as well. Hinduism emulates non-violence and the cow is regarded as sacred, and so many Hindus are lacto-vegetarians and if meat is eaten

particular slaughtering methods must be used. Some Jains living in Western countries are vegans and also restrict root vegetables and tubers for religious reasons. The various Buddhist sects vary in their views on vegetarian eating although the religion emphasizes non-violence; some Buddhists are lactovegetarians, others vegan and still others are flexible and consume meat. Other adherents of vegetarianism are more vocal, strident, and occasionally violent advocates who proselytize broader agendas, including animal rights, anti-vivisection, reform of agricultural and animal husbandry practices, and nonuse of experimental animals in addition to health and environmental concerns. In the US these latter groups include People for the Ethical Treatment of Animals (PETA), the PETA foundation, the Animal Liberation Front, and the Physicians' Committee for Responsible Medicine (PCRM). They finance advertising, journal articles, and educational efforts to encourage adoption of their views. Extensive efforts are directed toward youth. Similar groups are also active in other countries.

Nutritional features of well-planned vegetarian diets

There are four dimensions of food and nutrition security that need to be considered when planning diets. The International Covenant on Economic, Social and Cultural Rights acknowledges having adequate food, and being free from hunger as rights of all human beings. However, adequate nutrition requires a combination of nutrients rather than simply enough food energy. Very few plant-based foods contain all the essential nutrients needed for survival. Therefore, the consumption of a variety of foods is important. Secondly, the cost and affordability of food can impact one's availability to food despite a food item being physically available for purchase. In the development of healthy diet recommendations, nutrition economics must be at the forefront. Third, healthy dietary recommendations need to be culturally accepted. And lastly, it is increasingly viewed as important to take the impact on the environment into consideration (Godfray et al., 2018). All of these desiderata apply to vegetarian as well as other eating patterns (Dwyer and Drewnowski, 2017).

The nutritional benefits of vegetarian diets on health are several. Plant foods are generally nutrient dense without being calorically dense, at least in their unprocessed state. They are good sources of vitamins B, E, C, potassium, magnesium, copper, manganese, dietary fiber, omega 6 polyunsaturated and monounsaturated fat, and complex carbohydrates. They are also rich in plant sterols and several other phytochemicals, such as the flavonoids, lignans, glucosinolates, and isothiocyanates, and possibly also compounds, such as those in curcumin and other spices that some vegetarians use in large amounts. Those bioactives may have beneficial health effects, although as yet there is not compelling evidence that they do. In addition to plant foods providing these potentially beneficial compounds, the foods in vegetarian diets are generally low in saturated and trans-fats and in cholesterol. These are all positive attributes from the health standpoint.

However, there are some limitations. While some plant foods such as legumes are high in protein, others are low or incomplete by themselves and so they need to be combined with each other to fill all the human body's needs for amino acids and nitrogen (Fussell et al., 2021). Although some of the proteins in single plant foods, such as cassava, are "incomplete" (that is, their amino acid profiles alone cannot meet human amino acid needs), combinations of different plant protein sources or small amounts of animal protein with the plant protein foods can complement each other, supply missing amino acids, and meet amino acid and nitrogen needs. Since most vegetarian diets in Western countries contain a number of different plant protein sources and they are usually available in generous amounts, adequacy or quality of protein is rarely a problem even on vegan diets if due care is paid to making wise dietary choices. Potentially some vegetarian diets may fall short in vitamins D, B_{12}, zinc, bioavailable iron and zinc, omega 3 fatty acids, choline, and iodine. Vitamin B_{12} is bound to proteins in food and is naturally present in foods of animal origin. It is not found in plant foods unless the product is fortified. Breakfast cereals and nutritional yeasts are common foods that are fortified with vitamin B_{12}. The bioavailability of vitamin B_{12} is dependent on dosage. The vitamin must bind with intrinsic factor, a protein secreted by the stomach, for it to be absorbed. Only a few foods naturally contain vitamin D. Fatty fish such as trout, salmon, tuna, and mackerel, and fish liver oils are the best sources of vitamin D. However, the animal's diet affects the amount of vitamin D that are in their tissues. In the US the milk supply is fortified with vitamin D. In addition to fatty fish, and fortified products, vitamin D can be obtained through sun exposure, especially below approximately 40 degree N latitude where exposure to sunlight is likely to be sufficient to fulfill bodily needs. Similar to vitamin D, the omega-3 fatty acids are found in fish (salmon, mackerel, tuna, sardines, bass, tilapia, cod, trout, shrimp, and lobster) and are in small amounts in beef and chicken. Some omega-3 fatty acids are found in a variety of plant foods and oils (flaxseed oil, chia seeds, walnuts, soybean oil, edamame, kidney beans, eggs, and milk). Zinc is found in a variety of foods; that in plants is not highly bioavailable. Oysters contain more zinc per serving than any other food. Other sources of zinc include beef, crab, lobster, pork and chicken. Zinc is found in whole grains, fortified breakfast cereals, dairy products, baked beans, and some nuts. Choline is an essential nutrient found most commonly in animal foods (i.e., beef, eggs, chicken, and milk). Although it is present as well in small amounts in soybeans, potatoes, wheat germ, and cruciferous vegetables. Iodine is a nutrient that may be lacking in a vegetarian's diet because most fruits and vegetables are poor sources of iodine, and the iodine that they contain depends on the soil, fertilizer, and irrigation practices used. Seaweed and kelp are highly variable in their iodine content .Ocean fish, other seafood, eggs, and iodized table salt are good sources of iodine. Milk is highly variable in iodine; today, with the decline in the use of iodophors in milking equipment, the amounts are much lower in cow milk than they were in the past. However, today it is easy for vegans and vegetarians to obtain nutrients that might otherwise be low or lacking in their diets. A wide variety of plant foods is at their disposal (Kyriakopoulou et al., 2021). There is also widespread availability of various soy, nut and other meat alternatives and meat analogs, and of fortified plant foods, including soy milks (with added vitamins B_{12} and D, as well as highly bioavailable calcium), calcium-fortified orange juice, and highly fortified breakfast cereals. Over the past several years, the availability of plant-based products has expanded in major grocery stores, and even the

major fast-food chains offer plant-based burger options. Milk alternatives have also increased and now include almond, soy, oat, pea, and rice milk. Dairy-free yogurts, cheeses, and ice creams are widely available. Meat alternatives now include breakfast sausage, bacon, chicken, hot dogs, burgers, and others. Although these products are easily accessible, some consumers may object to them because they eschew "processed" or "ultra-processed" foods as well as animal foods. Also, it should be noted that the nutrient profiles and costs of these products are often quite different and may not compare favorably to the nutrients in the animal products they substitute for. For example, some of the "plant milks" like soy have similar nutrient profiles to cow milk, while others such as rice milk, are quite different. Dietary supplements offer another alternative for remedying shortfalls in the micronutrients.

Some but not all of the dietary differences between the health of vegetarians and nonvegetarians are due to the types and amounts of essential nutrients in what they eat or avoid. There is less evidence supporting attributing the healthful aspects of vegetarian diets with respect to the bioactive constituents other than the nutrients that the diets contain or lack. For example, vegetarian diets are often claimed to be evolutionarily more in line with dietary patterns of Paleolithic times than are "Western" diets and therefore better suited to human genes. However, very little morphologic, bone, or artifactual evidence is available on the actual diets of early humans (Agoulniket al., 2021; Karlse et al., 2021). Most early hominids died young of infectious or other diseases rather than living into middle and old age when the chronic degenerative diseases so common today occur. What is clear is that the advent of farming and agriculture led to rapid population growth, and at least initially to increase in morbidity and mortality because of crowding, changes in patterns of disease transmission, and many other factors including diet, but not solely due to that.

Another rationale for vegetarian eating is that animal foods, particularly meat, increase risks of many infectious and chronic diseases. This is discussed elsewhere in this article. A third contention is that it is the synergy of foods that are plentiful in vegetarian diets and particularly in fruits and vegetables working together as a "package" (e.g., abundant fruits, vegetables, legumes, nuts, and other plant foods) that confers a special and added health advantage. Generally, Western vegetarian diets eaten today tend to be low in saturated fat and cholesterol and high in polyunsaturated fats, complex carbohydrates, dietary fiber, and contain a variety of plant proteins, which complement each other well and provide a protein mixture of good biological value, along with some animal protein as well as magnesium, potassium, folic acid, and antioxidant nutrients, such as vitamin E and selenium. Compared to common omnivorous patterns, the healthier vegetarian patterns also tend to be relatively low in food energy and relatively high in some of the nonnutritive phytochemical bioactives (phytosterols, flavonoids, and isothiocyanates) that may provide health benefits. Comparisons between a low-fat fast food nonvegetarian diet matched on type of fat, macronutrients, and dietary cholesterol, found that serum low-density lipoprotein (LDL) cholesterol levels were lower among those eating a vegetarian diet, suggesting that other components in the vegetarian diet might also have positive effects on this intermediary marker of coronary heart disease. However, the hypothesis that "food synergy" (e.g., the combined effects of plant foods and constituents working together in the vegetarian pattern) provides very large additional benefits over the health effects of each constituent alone or with other foods requires further verification. Finally, it must be remembered that many vegetarians also differ from nonvegetarians in other ways that may contribute to good health, including leading physically active lives, nonsmoking, and abstinence from alcohol and other illicit drugs.

The Dietary Guidelines for American's updated its' Healthy Vegetarian Diet in 2020 to provide an example of a nutritionally adequate lacto-ovo vegetarian eating pattern for those over age 2 years, and it is available at various calorie levels. Even with extensive modeling using foods alone, several nutrients still did not quite meet the Recommended Dietary Allowances (RDA) or adequate intake (AI) for several age and gender groups (iron, vitamins D, E, and choline. Also, the patterns did not consider cultural variation, and assumed that all foods chosen were the most nutrient-dense, lean. Low-fat versions prepared with recipes using minimal added fats, sugars, refined starches, and sodium. These considerations highlight the utility of dietary supplements or fortified foods that fill nutrient gaps and nutritional counseling by a registered dietitian/nutritionist who can personalize diets to eaters without sacrificing dietary adequacy and avoiding excess. **Table 2** provides some strategies that might be considered by vegetarians to prevent inadequacy.

Morbidity and mortality on vegetarian versus omnivorous diets

Well-planned vegetarian diets have nutritional profiles that are in line with recent expert recommendations. These patterns, if sustained, may reduce diet-related risks of some chronic degenerative diseases (coronary artery disease, hypertension, type 2 diabetes) and obesity associated with caloric excess (Laouli et al., 2021). The differences between vegetarians and omnivores in diet but also in other lifestyle characteristics that affect disease risk play a role in the varied patterns of illness and disease between them. For example, many vegetarians are health conscious, eschew tobacco, alcohol, and recreational drugs, avoid overexposure to the sun, and lead physically active lives, and so risk factors for diseases that are influenced by those factors are decreased. Other less likely differences that may be residual confounders of the differences between them are lesser exposure to second-hand smoke, excessive amounts of food energy and refined sugars along with increased consumption of whole grains, nuts, and legumes that are possibly protective.

In the past few decades, some but not all observational studies have found that vegetarians are healthier or live longer than nonvegetarians. While some of the differences, such as those in lung cancer, are likely due largely to the lesser use of tobacco among vegetarians, other differences favoring lower prevalence of disease may be attributable to the diet itself. Since those dietary factors or other differences between vegetarians and nonvegetarians influence disease risk they, rather than solely the use or nonuse of animal foods, are potentially responsible for some of the positive health effects (Mihrshai et al., 2017; Palacios and Maki, 2019). Therefore, the effects of all these factors need to be accounted for. With respect to morbidity, the mishandling of animal foods during

Table 2 Nutrients of concern with dietary strategies to counteract inadequacy and relevant dietary reference intakes (recommended dietary allowance (RDA) or adequate intake (AI).

		RDA/AI	
Nutrient	*Dietary strategies*	*Male*	*Female*
Calcium	Consume calcium fortified products such as orange juice, calcium fortified breakfast cereals and other products. Tofu made with calcium sulfate is a good protein source as well. Leafy greens such as spinach, turnip greens, kale, bok choi, and broccoli are good sources. Dietary supplements: tablets of calcium carbonate and citrate with vitamin D contain large amounts; multivitamin-mineral supplements contain less.	9–18 years: 1300 mg 19–70 years: 1000 mg 70+ years: 1200 mg	9–18 years: 1300 mg 19–50 years: 1000 mg 51+ years: 1200 mg
Choline	Dairy, eggs, cheese some beans such as soy or kidney beans, wheat germ and nuts are good sources of choline, as are many vegetables and fruits, and whole grains such as oats, pearl barley, buckwheat, quinoa and many vegetables such as brussels sprouts, broccoli, and cauliflower. Dietary supplements: single ingredient supplements prenatal multivitamins for pregnant women contain large amounts and some multivitamin mineral supplements contain smaller amounts	9–13 years: 375 mg 14+ years: 550 mg	9–13 years: 375 mg 14–18 years: 400 mg 19+ years: 425 mg
Iodine	Dairy products (such as milk, yogurt, and cheese are good sources. Fish (such as cod and tuna), shrimp, and other seafood, are generally rich in iodine. Use iodized salt and replace specialty salts, such as sea salt, kosher salt, Himalayan salt, etc., with it; specialty salts are not usually iodized. Iodine in seaweed (kelp) is present in varying amounts. Dietary supplements: many prenatal multivitamins now contain potassium or sodium iodide. Iodine is available usually in the form of potassium iodide or sodium iodide. Many multivitamin-mineral supplements contain iodine. Dietary supplements of iodine-containing kelp (a seaweed) are also available.	9–13 years: 120 mcg 14+ years: 150 mcg	9–13 years: 120 mcg 14+ years: 150 mcg
Iron	Consume iron-fortified breakfast cereals and breads. Increase consumption of legumes such as white beans, lentils, kidney beans, and peas. Add nuts or dried fruit like raisins to salads or as a garnish. Supplements: most multivitamin-mineral supplements or as a single nutrient vitamin. * Nonheme iron, the form of iron found in plants, is more readily absorbed when consumed with foods that contain vitamin C, such as citrus fruits, strawberries, sweet peppers, tomatoes, and broccoli.	9–13 years: 8 mg 14–18 years: 11 mg 19+ years: 8 mg	9–13 years: 8 mg 14–18 years: 15 mg 19–50 years: 18 mg 51+ years: 8 mg
Omega-3 fatty acids	Cook with plant oils such as flaxseed oil, canola oil, or soybean oil. Add chia seeds, flaxseed, or walnuts to smoothies or salads. Supplements: fish oil, krill oil, cod liver oil, and algal oil	9–13 years: 1.2 g 14+ years: 1.6 g	9–13 years: 1.0 g 14+ years: 1.1 g

Table 2 Nutrients of concern with dietary strategies to counteract inadequacy and relevant dietary reference intakes (recommended dietary allowance (RDA) or adequate intake (AI).—cont'd

Nutrient	*Dietary strategies*	*RDA/AI*	
		Male	*Female*
Vitamin B_2 (riboflavin)	Increase consumption of green vegetables such as asparagus, broccoli, and spinach. Eat riboflavin fortified cereals, bread, and grains and whole grain products. Supplements: most multivitamin-mineral supplements, B-complex vitamins, or as a single nutrient.	9–13 years: 0.9 mg 14+ years: 1.3 mg	9–13 years: 0.9 mg 14–18 years: 1.0 mg 19+ years: 1.1 mg
Vitamin B_{12} (cobalamin)	Eat B_{12} fortified breakfast cereals, nutritional yeasts, and other fortified products. Supplements: most multivitamins-mineral supplements such as B-complex vitamin products contain B_{12}, and a single product contain vitamin B_{12} higher levels in the form of cyanocobalamin. Other common forms include adenosylcobalamin, methylcobalamin, and hydroxycobalamin.	9–13 years: 1.8 mcg 14+ years: 2.4 mcg	9–13 years: 1.8 mcg 14+ years: 2.4 mcg
Vitamin D	Drink cow milk or plant milks fortified with vitamin D. Increase intake of vitamin D irradiated mushrooms. Consume vitamin D fortified breakfast cereals, orange juice, and milk substitutes. Supplements: most multivitamin-mineral supplements or as a single nutrient vitamin in the form of D_2 (ergocalciferol) and D_3 cholecalciferol). * Vitamin D is more readily absorbed when consumed with a meal or snack that contains fat.	9–70 years: 15 mcg (600 IU) 70+ years: 20 mcg (800 IU)	9–70 years: 15 mcg (600 IU) 70+ years: 20 mcg (800 IU)
Zinc	Increase consumption of beans and nuts by adding them to meals and salads. Consume whole grains and zinc fortified breakfast cereals. Supplements: Most multivitamin-mineral supplements, combined with calcium, magnesium, or other nutrients, or as a single nutrient vitamin. Forms of zinc found in supplements include zinc gluconate, zinc sulfate, and zinc acetate.	9–13 years: 8 mg 14+ years: 11 mg	9–13 years: 8 mg 14–18 years: 9 mg 19+ years: 8 mg

preparation and storage increases risks of food-borne illness, and historically caused many types of food-borne illness (Salmonellosis, Campylobacter food borne disease, Brucellosis, various worms, and other parasites) posed major risks in animal foods and could be avoided by not eating them. Now that animal food handling and preparation practices are more advanced, the dangers of these diseases have decreased, although continued vigilance is necessary.

Some conditions and disease risks are clearly lower in vegetarians; for example, constipation is less of a problem in vegetarians (especially vegans) than in omnivores, perhaps due to their much higher intakes of dietary fiber and more active lifestyles. Vegetarians and especially vegans generally tend to have lower body mass indices (BMIs) than do nonvegetarians, and consequently they have lower risks of the chronic diseases for which obesity is a risk factor, such as type 2 diabetes. There are now several epidemiological reports of effects of various animal foods, and particularly meat, on type 2 diabetes, although some studies implicate all meats, others only do so for processed meats (World Cancer Research Foundation, 2020; Zeraatkar et al., 2019). Whether meat is still associated with risk when the comparison group consists of healthy and weight conscious nonvegetarians remains to be determined.

Animal foods and especially red and other meats (saturated fats and cholesterol), poultry (unless lean and skinned), eggs, crustaceans like shrimp and lobster (rich in cholesterol), and whole fat dairy products (high in saturated and total fat) have been indicted as increasing risks of coronary artery disease. The foods' ill effects appear to stem primarily from their content of saturated fat, cholesterol, and particularly the high amounts of food energy that they contribute to the diets of eaters and increase risks. Also, meat eaters tend to have low intakes of nuts, fruits and vegetables, plant sterols, dietary fiber, and other dietary components that may contain other bioactives that lessen coronary artery disease risk. Total serum cholesterol and LDL cholesterol tend to be lower. Ischemic and coronary heart disease are often decreased among vegetarians, particularly among vegans compared to carnivores (Key

et al., 2019). Blood pressure and hypertension are also generally somewhat lower among vegetarians, likely due to their lower weights, lesser sodium, higher potassium and perhaps their higher nitrate rich vegetable intakes (Van der Voort et al., 2021).

Excessive food energy, alcohol, and possibly fat pose greater risks of some cancers than do other constituents of Western diets, and vegetarian diets low in the former might have beneficial effects. Cancer morbidity is lower in some vegetarians, such as Seventh-Day Adventists but it is primarily in the smoking and alcohol-related cancers, as might be expected because many vegetarians of that faith also do not use tobacco or drink alcohol. Modest (<10%) protection against some cancers seems to be associated with decreased weight. Western diets contain a great deal more meat than do those in many other parts of the world. The human data are mixed but some substances in certain forms of meat are carcinogenic at least in experimental animals. Processed and red meat have recently been singled out as a risk factor for colon cancer because its composition (often high fat, high saturated fat, high salt, and high in heme iron), preservation methods (forming N nitroso compounds or interacting with heme to form carcinogens), and styles of preparation (such as benzo(a)pyrene, heterocyclic amines and polycyclic aromatic compounds, adducts, and other possible carcinogens produced in meats broiled and fried at high temperatures) may expose eaters to carcinogens or cancer promoters. The World Cancer Research Foundation in 2018 concluded that intakes of red meat should be limited, and processed meat should be avoided based largely on epidemiological evidence (World Cancer Research Foundation, 2018). It suggested that meat, especially preserved and processed meat, had a modest effect on some cancers, particularly colorectal cancers. That conclusion has been criticized and more recent studies are mixed. One very large prospective study of older Americans suggested a modest association between meat intake and mortality while effects on breast cancer were not apparent. Trimethylamine N-oxide (TMAO), a molecule generated from choline, betaine, and carnitine via gut microbial metabolism contains constituents also present in meat. More and better studies are needed to definitively rule out meat, or specific types of meat, as one of the culprits in these cancers. The notion that dairy products and particularly milk contain sufficient estrogens or other cancer promoters continues to be debated, but the human and experimental animal evidence of such effects is weak. There is mixed evidence that osteoporosis and fracture risks are greater in vegans, perhaps due to their low calcium and vitamin D intakes but their more active lifestyles may often compensate for those risks (Thorpe et al., 2021). The associations between animal food intakes and other illnesses, including diverticular disease of the colon, gallbladder disease, and appendicitis are more speculative.

Fewer studies are available on specific and all-cause mortality and longevity among vegetarians. Smoking-related mortality is lower among groups of vegetarians who do not use tobacco, but most of these effects do not appear to be diet related. Obesity-related mortality is also reduced. Once these factors have been accounted for, other differences between vegetarians and nonvegetarians or between vegetarians who differ in the degree to which they restrict animal foods in mortality or longevity are not dramatic.

At present the evidence is insufficient to conclude that well-planned meatless or vegetarian diets have distinct health advantages over well-planned plant-based dietary patterns that include animal foods. Causal inference is difficult in making comparisons between vegetarians and nonvegetarians: residual confounding in diet and the many existing differences between vegetarians and health-conscious omnivores, including education, lifestyles, and many other health related behaviors, cannot all be controlled for in observational studies. However, it is abundantly clear that most Western vegetarians who subsist on well-planned regimens appear to be at least as healthy as their omnivorous counterparts, and that many omnivores eat diets that do not conform to expert recommendations and that contribute to ill health. Whether vegetarians have an advantage over health conscious nonvegetarians is still uncertain, but regardless vegetarians may prefer these patterns for many other reasons.

Nutritional adequacy

In English-speaking North America, dietary reference intake (DRI) standards that embody what is known about nutrient requirements have been issued by the Food and Nutrition Board of the National Academy of Sciences Engineering and Medicine and Health Canada. Similar but not identical reference standards are available in other Western countries to assess and plan nutritious diets. Nutritional adequacy is defined as meeting nutrient needs, such as the Recommended Dietary Allowances (RDAs) or the adequate intakes, while avoiding excess and staying below the tolerable upper levels of nutrient intakes (UL) and keeping within the acceptable macronutrient distribution ranges (AMDR) specified by expert groups (in the US, fat 20–35%, protein 10–35%, and carbohydrate 45–65% of calories). Most vegetarians have little difficulty in meeting these standards with a little planning.

From the nutritional standpoint the animal food groups (e.g., meat, fish/seafood, poultry/fowl, eggs, milk, and milk products) are generally nutrient-dense foods high in several micronutrients that are low or lacking in plant foods. In traditional vegetarian diets, depending on the particular animal food group under consideration, these nutrients that might fall short included protein of high biological value, highly bioavailable iron, zinc, calcium, vitamins A, D, B_2, B_6, and B_{12}, choline, omega 3 fatty acids, calcium, and iodine. If animal food groups are entirely eliminated, intakes of these nutrients often become deficient. Today, the widespread availability of special foods fortified with these nutrients and appropriate micronutrient-containing dietary supplements provide options that can help to make up the slack. Although dietary diversity defined as the number of food groups consumed is less among vegetarians, diversity within the remaining food groups (such as fruits, vegetables, nuts, and legumes) is often considerable among vegetarians and with careful choice this may suffice to provide adequate amounts of some of these nutrients. However, when vegetarians use dietary supplements as an option, they must make sure that the nutrients they obtain are those that are low in their diets, rather than employing a "shotgun" approach of a little more of many nutrients, and that the nutrient dose in the product is sufficient to meet nutrient needs. For example, for an older woman, a single glass of calcium-fortified orange juice and a serving of spinach, or a multivitamin-mineral supplement does not contain enough calcium to meet her nutrient needs, and a specific

combined calcium and vitamin D supplement might be more appropriate. Also, many vegetarians prefer food-based solutions and are reluctant to use dietary supplements because the source of the vitamin or the capsule is of animal origin. In some instances, such as among infants who are breastfed for several months with no other source of vitamin D, a dietary supplement of water-miscible vitamin D may be acceptable and it protects the infant from vitamin D deficiency.

Eating patterns appear to be more closely associated with health outcomes than are intakes of nutrients alone, perhaps because of the synergistic effects of nutrients working with each other or because other bioactives, such as phytochemicals, also have health effects. Vegetarians' intakes of bioactives rich in fruits and vegetables, such as the flavonoids, antioxidants, and dietary fiber are often much higher than those of omnivores. If these substances prove to have beneficial effects on health, such increased intakes may be important. However, to date only a few effects have been demonstrated.

From the nutritional standpoint there are some beneficial trade-offs from vegetarians' limited intakes of animal foods. In Western diets, animal foods are the major sources of dietary constituents that are excessive in Western diets, such as calories, fat, saturated and trans-fat, cholesterol, and sodium, and they are low in polyunsaturated fats and dietary fiber. Consumption of fewer animal foods, especially if it leads to lesser intakes of these constituents and greater intakes of other nutrients and bioactives, may have positive effects on overall nutritional status. The beneficial effects of vegetarian eating patterns on nutritional status vary, however, depending on the food group(s) avoided, the degree of limitation, substitutions of other rich food sources, use of fortified foods and dietary supplements containing the lacking nutrients, and other changes in dietary intake or lifestyles that occur at the same time. The nutritional goal in planning a vegetarian diet is to personalize the eating pattern to the individual's preferences and beliefs while maximizing the benefits and minimizing the health risks by a judicious choice of the type and amount of animal foods or other sources of needed nutrients that are acceptable to the eater.

Adequate vegetarian dietary patterns

Dietary practices among vegetarians are highly variable. Most vegetarians living in Western countries are at little or no risk of dietary inadequacy from their eating patterns and they do not merit special assessment although they may benefit from general dietary advice adopted to their eating preferences. For example, neither a man who regards himself as a semi-vegetarian (also referred to as a meat avoider) because he avoids red meat most of the time nor a lacto-ovo vegetarian woman with no other major dietary avoidances are likely to need further dietary assessment. In contrast, those at risk of , or whoever suffer from a chronic degenerative disease that requires dietary modification often benefit from counseling from a registered dietitian or other health profesional with expertise in the dietary modifications needed (Jardine et al 2021).

Some characteristics of sound and adequate vegetarian diet patterns include the following:

- Limited animal food avoidances, only sporadic avoidances, or abstinence rather than complete animal food avoidance.
- Consumption of a wide variety of animal food groups and a wide variety of foods within each group.
- Use of a nutritionally sound food guide for diet planning; sound vegan, and vegetarian food guides available that ensure that nutrient needs are adequate, balanced, and moderate
- If an individual follows a vegan pattern, or has multiple food avoidances, the regular use of foods fortified with nutrients likely to fall short in the diet and vitamin or mineral supplements in RDA (or Daily Reference Value (DRV or DV) on food labels) amounts to meet nutrient needs plus use of a nutritionally sound food guide
- Membership in a group (such as Seventh Day Adventists) or a family with a long cultural or religious tradition of adherence to healthy vegetarian eating styles and attitudes toward health, and use of fortified foods or micronutrient containing dietary supplements.

Possibly inadequate vegetarian dietary patterns

For vegetarians who are likely to be at high risk of dietary inadequacy, further and more complete dietary assessment and planning from a nutrition professional may be needed. The entire pattern of intake (including avoidances, substitutions, and additions of foods, and use of dietary supplements) must be examined by the counselor to obtain a full profile of nutrient adequacy or inadequacy. The presence or absence of other lifestyle practices with potentially beneficial health impacts (nonsmoking, abstinence from alcohol, high levels of physical activity) can also have impacts on health. Individual assessment is recommended for those at special risk in nutritionally vulnerable physiological groups [due to age, life stage (pregnancy, lactation, infancy, toddlerhood, pubertal growth spurt), or illness] or because they adhere to very limited patters [such as avoiding many animal food groups (vegans) and other food groups, or have multiple other food avoidances (e.g., all "processed", "ultra-processed", nonorganic, cooked, or canned foods, "nonwhite" foods, "foods with chemical ingredients" and "non-natural" foods)] so that only "whole", "non-processed", "organic" and "natural" foods are acceptable, further limiting options for planning an acceptable diet.

Characteristics that may indicate inadequate or unbalanced vegetarian diets and the need for further assessment include:

- Entire or very extensive avoidance of animal food groups.
- Refusal to eat fortified plant foods containing micronutrients likely to be low in vegetarian diets or to use micronutrient-containing dietary supplements, or processed foods.
- Refusal to eat "processed" or "nonorganic" foods.

- Frequent fasting, vomiting, purging, or drastically altered diet during illness.
- Lack or avoidance of conventional health and medical care (e.g., prescribed medications, vaccinations, mental, or dental health visits), use of only alternative medical care.
- Nutritional vulnerability due to age or physiological condition (infancy, weaning and toddlerhood, rapid growth, puberty, pregnancy, lactation, chronic illness, or recovery from illness, old age, frailty).
- Low weight for height, body mass index (BMI) $< 18.5\ kg\ m^{-2}$, >4–7 kg (10–15 lb.) unintentional weight loss, or rapid unintentional weight loss of >5% in a month.
- Deeply held beliefs in alternative philosophical or religious systems that rigidly restrict food choice and prescribe vegetarian diets that fail to meet the DRIs.

Special nutritional concerns for vegetarians

Of particular concern with respect to risk of inadequacy for vegetarians and especially vegans and others with extreme dietary restrictions are energy, vitamins B_{12} and D, riboflavin, omega 3 fatty acids, calcium, iron, zinc, choline, and iodine. Vegetarians of all types can meet current recommendations for these nutrients if they are willing to use nutrient sources in fortified foods (e.g., highly fortified cereals, calcium and vitamin D fortified soy milk or fruit juices, vitamin B_{12} fortified yeast, vitamin A fortified margarine) or specific micronutrient supplements containing enough of the micronutrient that is falling short. However, some vegetarians are unwilling to use these options, increasing risks of deficiency, and making dietary planning more difficult. With respect to macronutrients, vegetarian and particularly vegan diets tend to be low in energy, total fat, saturated fat, cholesterol, and sodium. If processed foods are avoided, added sugars are also low but the naturally occurring sources of sugars may still be quite high. Current recommendations for acceptable macronutrient distribution ranges (AMDR) in the US are for fat 20–55% of calories and for protein 10–35% of calories, with added sugars no more than 10% of calories and preferably lower, and the remainder from other carbohydrates.

Key nutrients for vegetarians over the life cycle

Well-planned vegan and vegetarian diets can meet nutritional needs at all stages of the life cycle including pregnancy, lactation, infancy, childhood, and adolescence (Melina et al., 2016). There are very few longitudinal studies of individuals on the more restrictive vegetarian diets; the notable exception is a Dutch cohort of macrobiotic vegetarians who were followed from birth to adolescence, and who showed very poor growth and continue to have some health problems (Salomé et al., 2019). More studies are needed so that long-lasting effects of diet early in life can be better ascertained.

Some vegetarian parents inadvertently feed their children with diets that are inadequate. The problem is not that nutritionally adequate diets cannot be planned, but that the eater's or cook's ideologies and concerns may get in the way (Bivi et al., 2021). Under such circumstances health problems have arisen and continue to do so, especially among infants and children on vegan diets that are limited in other foods as well.

Vegan diets present more problems of micronutrient adequacy than do other vegetarian diets all across the life cycle and particularly in infants and children because more food groups are eliminated, sources of vitamins B_{12}, D, choline and bioavailable iron, calcium, and zinc as well as omega 3 fatty acids may be lacking, the caloric density of the diet is lower, but bulk is higher and weaning from the breast is often very late. Vegan diets may also be low in calcium, iron, and zinc, and the forms of these minerals may not be highly bioavailable. Sources of riboflavin and choline also need to be identified.

Vegetarian infants are usually breastfed. They generally thrive until 4–6 months of age, and continue to do so if they remain at the breast and receive complementary feedings of nutritionally complete heat-treated cow's milk-based or fortified soy formulas, or other developmentally appropriate feedings that are high in nutrients, sufficient in energy, and low in bulk. Soy or other plant milks are not appropriate under 1 year of age. In countries where home prepared or commercial infant formulas and weaning foods do not provide adequate amounts of micronutrients, certain dietary supplements such as iron, choline, or vitamins B_{12} and D may be needed (Cofnas, 2019). Today, more fortified vegan foods are available for infants and other groups than in the past. Good food or supplemental sources of vitamin B_{12} and D (especially above approximately 40 degree N latitude where exposure to sunlight is not likely to be sufficient), linolenic acid (an omega 3 fatty acid to ensure that docosahexaenoic acid (DHA) intakes are satisfactory), riboflavin, calcium, choline and bioavailable forms of iron, and zinc must be included in the diets of weanlings and young children. Protein intakes are usually adequate if many different plant proteins are fed and energy intakes are sufficient (Salome et al., 2020). Growth monitoring is helpful to ensure that the diet is supportive of good growth. When soy milk is used later in childhood, especially if the child is a vegan, it should be fortified with vitamins D, $B_{12,}$ and calcium (Eaton et al., 2019). Other plant milks should be as well, but rarely are. For children during the pubertal growth spurt, energy, calcium, iron, vitamins D and B_{12}, as well as iron are of particular concern with respect to dietary adequacy. Pregnant and lactating vegetarian women have increased nutrient needs for these and other nutrients that can usually be dealt with by dietary planning. Choline and omega 3 fatty acid intakes as well as iron and vitamin B_{12} deserve particular attention. Older adults who are vegetarians may have problems meeting micronutrient needs because they have low energy requirements owing to physical inactivity, whereas their needs for calcium, vitamin D, and vitamins B_6 and B_{12} increase. Also, they may suffer from chronic diseases that alter nutrient metabolism. One example is atrophic gastritis in older persons, which affects absorption of vitamin B_{12}.

Conclusions

Vegetarian diets should be planned in accordance with the DRI or other authoritative nutritional recommendations so that they are healthful and nutritionally adequate. When vegetarian diets are unplanned, the nutrients that are likely to fall short usually differ somewhat from those on unplanned omnivorous diets. In some cases, these deficits can be easily remedied by dietary counseling and appropriate food choices. In others, differences between ideologies about life, diet, and nutrient needs are such that acceptable dietary strategies are difficult to achieve. Nutrition scientists and healthcare practitioners can help vegetarians who seek their advice by screening for nutritional status to identify high-risk individuals, by identifying acceptable food sources of specific nutrients that may be low in the diet for the eater, by suggesting dietary modifications that may be necessary to meet individual needs when intakes fall short, and by monitoring the vegetarian's progress.

See Also: Cancer: Epidemiology and associations between diet and cancer; Vitamin B_{12}: Physiology, dietary sources, and requirements

References

Agoulnik, D., Lalonde, M.P., Ellmore, G.S., McKeown, N.M., 2021. Part 1: the origin and evolution of the paleo diet. Nutr. Today 56 (3), 91–104.
Bivi, D., DiChio, T., Geri, F., Morganti, R., Goggi, S., Baroni, A.L., Mumolo, M.G., Bortoli, N., Peroni, D.G., Marchi, S., Bellini, M., 2021. Raising children on a vegan diet: parents' opinion on problems in everyday life. Nutrients 13, 1706. https://doi.org/10.3390/nu13061796.
Cofnas, N., 2019. Is vegetarianism healthy for children? Crit. Rev. Food Sci. Nutr. 59 (13), 2052–2060.
Dwyer, J.T., Drewnowski, A., 2017. Overview: food and nutrition security. In: Biesalski, H.K., Drewnowski, A., Dwyer, J.T., Weber, P., Eggersdorfer, M. (Eds.), Sustainable Nutrition in a Changing World. Springer International Publishing, Germany, pp. 3–24.
Eaton, J.C., Rothpletz-Puglia, P., Dreker, M.R., et al., 2019. Effectiveness of provision of animal-source foods for supporting optimal growth and development in children 6 to 59 months of age. Cochrane Database Syst. Rev. 2, 1–26.
Fussel, M., Contillo, A., Druehl, H., Rodrigue, N.R., 2021. Essential amino acid density: differences in animal and plant-based dietary patterns designed for older women. Nutr. Today 56 (2), 70–75.
Godfray, H.C.J., Aveyard, P., Garnett, T., Hall, J.A., Key, T.J., Lorimer, J., Pierrehumbert, R.T., Scarborough, P., Springman, M., Jebb, S.A., 2018. Meat consumption, health and the environment. Science 361, 243–251. https://doi.org/10.1126/science.aam5324.
Jardine, M.A., Kahleova, H., Levin, S.M., Ali, Z., Trapp, C.B., Barnard, N.D., 2021. Perspective: plant-based eating pattern for type 2 diabetes prevention and treatment: efficacy, mechanisms, and practical considerations. Adv. Nutr. nmab063. https://doi.org/10.1093/advances/nmab063. PMID: 34113961.
Karlsen, M.C., Livingston, K.A., Agoulnik, D., Miki, A., Lichtenstein, A.H., Gilhooly, C.H., Folta, S.C., Economos, C.D., McKeown, N.M., 2021. Part 2: theoretical intakes of modern-day paleo diets. Nutr. Today 56 (4), 158–168.
Key, T.J., Appleby, P.N., Bradbury, K.E., et al., 2019. Consumption of meat, fish, dairy products, and eggs and risk of ischemic heart disease. Circulation 2835–2845.
Kyriakopoulou, K., Keppler, J.K., van der Goot, A.J., Boom, R.M., 2021. Alternative to meat and dairy. Annu. Rev. Food Sci. Technol. 12, 29–50.
Laouli, N., Shah, S., MacDonald, C.J., Mahamat-Saleh, Y., El Fatouhi, D., Mancini, F., Fagherazzi, G., Boutron-Ruault, M.C., 2021. BMI inn the associations of plant-based diets with type 2 diabetes and hypertension risk in women: the E3N prospective cohort study. J. Nutr. nxab158. https://koi.org/10.1093/jn/nxab158.
Melina, V., Craig, W., Levin, S., 2016. Position of the Academy of Nutrition and Dietetics: vegetarian diets. J. Acad. Nutr. Diet. 116 (12), 1970–1980.
Mihrshahi, S., Ding, D., Gale, J., Allman -Farinelli, M., Banks, E., Bauman, A.E., 2017. Vegetarian diet and all-cause mortality: evidence from a large population based Australian cohort =the 45 and up Study. Prev. Med. 97 (4), 1–7. https://doi.org/10.1016/j.ypmed.2016.12.044.
Palacios, O.M., Maki, K.C., 2019. Vegetarian diet patterns and chronic disease risk: what we know and what we don't. Nutr. Today 54 (4), 132. https://doi.org/10.1097/NT.0000000000000350.
Richardson, L.C., Cole, D., Hoekstra, R.M., et al., 2021. Foods implicated in U.S. outbreaks differ from the types most commonly consumed. J. Food Protect. 84 (5), 869–875.
Salomé, M., de Gavelle, E., Dufour, A., et al., 2019. Plant-protein diversity is critical to ensuring the nutritional adequacy of diets when replacing animal with plant protein: observed and modeled diets of French adults (INCA3). J. Nutr. 536–545.
Salome, M., de Gavellle, E., Dufour, A., Dubuisson, C., Volatier, J.L., Fouillet, H., Huneau, J.F., Mariotti, F., 2020. Plant protein diversity is critical to ensuring the nutritional adequacy of diets when replacing animal with plant protein: observed and modeled diets of French adults (INCA3). J. Nutr. 150, 536–545.
Thorpe, D.L., Beeson, W.L., Knutsen, R., Fraser, G.E., Knutsen, S.F., 2021. Dietary patterns and hop fracture in the Adventist Health Study 2: combined vitamin D and calcium supplementation mitigate increased hip fracture risk among vegans. Am. J. Clin. Nutr. 114 (2), 488–495.
Van der Voort, CMT, Ten Haaf, DSM, Bongers, CCWG, Van Oorshot, F, Verdijk, LB, van Loon, LJC, Hoopman, MTE, 2021. Increasing Nitrate-Rich Vegetable Intake Lowers Ambulatory Blood Pressure in (pre)Hypertensive Middle-Aged and Older Adults: A 12-Wk Randomized Controlled Trial. J Nutr 151 (9), 2667–2679. https://doi.org/10.1093/jn/nxab157.
World Cancer Research Foundation, 2020. In: https://www.wcrf.org/dietandcancer/limit-red-and-processed-meat/.
Zeraatkar, D., Han, M.A., Guyatt, G.H., et al., 2019. Red and processed meat consumption and risk for all-cause mortality and cardiometabolic outcomes. A systematic review and meta-analysis of cohort studies. Ann. Intern. Med. 171 (10), 703–712.

Further reading

Craig, W.J., Mangels, A.R., Fresan, U., Marsh, K., Miles, F.L., Saunders, A.V., Haddad, E.H., Husky, C.E., Johnston, P., Larson-Meyer, E., Orlich, M, 2021. The safe and effective use of plant-based diets with guidelines for health professionals. Nutrients 11, 4144. https://doi.org/10.3390/nu13114144.
Dwyer, J.T., 1988. Health aspects of vegetarian diets. Am. J. Clin. Nutr. 489, 712–738.
Grossmann, L., Weiss, J., 2021. Alternative protein sources as techno functional food ingredients. Annu. Rev. Food Sci. Technol. 12, 93–117.
Jafari, S., Hezaveh, E., Jalipiran, Y., et al., 2021. Plant-based diets and risk of disease mortality: a systematic review and meta-analysis of cohort studies. Crit. Rev. Food Sci. Nutr. 1–13.
Kiely, M.E., 2021. Risks and benefits of vegan and vegetarian diets in children. Proc. Nutr. Soc. 80, 159–164.
Loveday, S.M., 2019. Food proteins: technological, nutritional, and sustainability attributes of traditional and emerging proteins. Annu. Rev. Food Sci. Technol. 10, 311–339.
Mangels, R., Messina, V., Messina, M., 2010. The Dietitian's Guide to Vegetarian Diets: Issues and Applications. Jones and Bartlett Learning, Ontario, Canada.

Mangels, R.R., Messina, V., Messina, M., 2021. The Dietitian's Guide to Vegetarian Diets: Issues and Applications, fourth ed. Jones and Bartlett Learning, New York.
Rees, K., Al-Khudairy, L., Takeda, A., Stranges, S., 2021. Vegan dietary pattern for the primary and secondary prevention of cardiovascular diseases. Cochrane Database Syst. Rev. 2, 1–31.
Sui, X., Zhang, T., Jiang, L., 2021. Soy protein: molecular structure revisited and recent advances in processing technologies. Annu. Rev. Food Sci. Technol. 12, 119–147.
US Department of Agriculture and US Department of Health and Human Services, 2020. Dietary Guidelines for ATE Increasing nitrate-rich vegetable intake lowers ambulatory blood pressure in (pre) hypertensive middle aged and older adults: a 12-wk. randomized controlled trial. J. Nutr. 2021, 1–13. https://doi.org/10.1093/jn/nxab147.
Vernooij, R.W.M., Zeraatkar, D., Han, M.A., et al., 2019. Patterns of red and processed meat consumption and risk of cardiometabolic and cancer outcomes. A systematic review and meta-analysis of cohort studies. Ann. Intern. Med. 171 (10), 732–743.

Whole grains and chronic disease risk

Caleigh M. Sawicki[a,c], **Nicola M. McKeown**[b], **Shatabdi Goon**[a], **and Shilpa N. Bhupathiraju**[a,c,1], [a] Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States; [b] Department of Health Sciences, Sargent College of Health and Rehabilitation Sciences, Boston University, Boston, MA, United States; and [c] Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of C.J. Seal, Whole Grains, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 422–430, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00285-3.

Introduction **676**
Structural components and composition of grains **676**
Bran 677
Endosperm 677
Germ 677
Definition of whole grains and health claims **677**
Recommendations for whole grain consumption and consumption patterns 678
Biomarkers of whole grain consumption 678
Whole grain consumption and disease risk **679**
Cardiovascular disease 679
Type 2 diabetes 679
Cancer 680
Colorectal cancer (CRC) 681
Breast cancer 682
Other cancers 683
Intermediate risk factors of disease and potential mechanisms of action **684**
Whole grains and adiposity 684
Whole grains and hypertension 685
Whole grains and inflammatory status 685
Whole grain, dietary fiber, and fermentable carbohydrates 685
Whole grains, antioxidants, and phytochemicals 686
Conclusion **686**
References **686**
Relevant websites **689**

Key points

- Although whole grains themselves are clearly defined, there has been variation in the definition of whole grain foods, recommendations, and health claims across different countries and organizations.
- Whole grains have been associated with lower risk of chronic diseases, including cardiovascular disease, type 2 diabetes, and cancers.
- Whole grains have also been associated with intermediate risk factors for chronic disease such as adiposity, hypertension, and inflammation.
- Components of whole grain including fiber, antioxidants, and phytochemicals may facilitate the health benefits of whole grains

Glossary

Bran Outer protective coat of a cereal grain that protects the seed from the environment. Rich in dietary fiber, micronutrients, and phytochemicals. 5–15% of grain weight depending on species

Endosperm Energy source for the growing seed. Contains carbohydrate in the form of starches and other oligosaccharides. Protein is found in the extracellular matrix. Source of B-vitamins. 60–85% of grain weight depending on species

[1]Dr. Bhupathiraju is a scientific consultant for LayerIV for work outside the submitted manuscript.

Germ The embryo of the seed, rich in protein, oils, fat soluble vitamins, micronutrients, and phytochemicals. 2–4% of grain weight depending on species

Phytochemicals Non-nutrient bioactive chemicals naturally found in fruits, vegetables, nuts, legumes, and grains that may have a positive impact on health

Whole grains Intact, ground, cracked, or flaked kernel after removal of inedible parts such as the hull and the husk. The starchy endosperm, bran, and germ are present in the same relative proportions as they were in the intact kernel

Introduction

It is universally accepted that eating whole grains is beneficial to health. Whole grains are nutrient-dense foods that, in addition to being dominant sources of carbohydrate and protein in the diet, contain a wealth of beneficial nutrients and phytochemicals. Consuming whole grains, whole grain food products, and foods prepared from whole grain flours is, therefore, recommended by government and health promotion agencies. Evidence demonstrating the health benefits of whole grains has come from a range of study designs, predominantly observational cohorts, but in more recent years there has been a growing number of intervention studies as well. This article provides an overview of this evidence, focusing on the association between whole grain intake and lower risk of morbidity and mortality from cardiovascular disease, type 2 diabetes, and cancer. Possible mechanisms of action including effects on intermediate risk factors such as body weight, blood pressure, inflammation, and the gut microbiota are described.

Structural components and composition of grains

Cereal grains are the seeds of the *Poaceae* (or *Gramineae*) family of grasses which are also referred to as kernels or grains. Cereal grains are a staple of many diets worldwide, with the dominant grains consumed being wheat, rice, maize, barley, rye, and oats. Other minor grains common in some countries include millet, sorghum, teff, triticale, and wild rice (Table 1). Although not part of the grass family, the "pseudocereals" amaranth, quinoa, and buckwheat are included in most definitions of cereal grains because they are similar in structure, nutrient composition, and in the way they are consumed as foods.

Cereal grains have three principle anatomical components, which relate to their function within the seed. Each component is complex and carries a unique nutrient profile (Fig. 1).

Table 1 Common grains that are considered whole grains.

Cereal type	*Scientific name*
Cereals	
Wheat, including spelt, emmer, faro, einkorn, kamut, durums	*Triticum* spp.
Rice	*Oryza sativa*
Maize (corn)	*Zea mays*
Barley, including hull-less or naked varieties but excluding pearled barley	*Hordeum* spp.
Rye, including hull-less or naked varieties	*Secale* spp.
Oats	*Avena* spp.
Millet	*Brachiaria* spp.; *Pennisetum* spp.; *Panicum* spp.; *Setaria* spp.; *Paspalum* spp.; *Eleusine* spp.; *Echinochloa* spp.
Sorghum	*Sorghum* spp.
Teff	*Eragrostis* spp.
Triticale	*Triticale* spp.
Canary seed	*Phalaris arundinacea* and *P. Canariensis*
Job's tears	*Coix lacryma-jobi*
Fonia, Black fonio and Asian millet	*Digitaria* spp.
Wild rice	*Zizania aquatic*
Pseudocereals	
Amaranth	*Amaranthus cordatus*
Quinoa	*Chenopodium quinoa*
Buckwheat	*Faopyrum* spp.

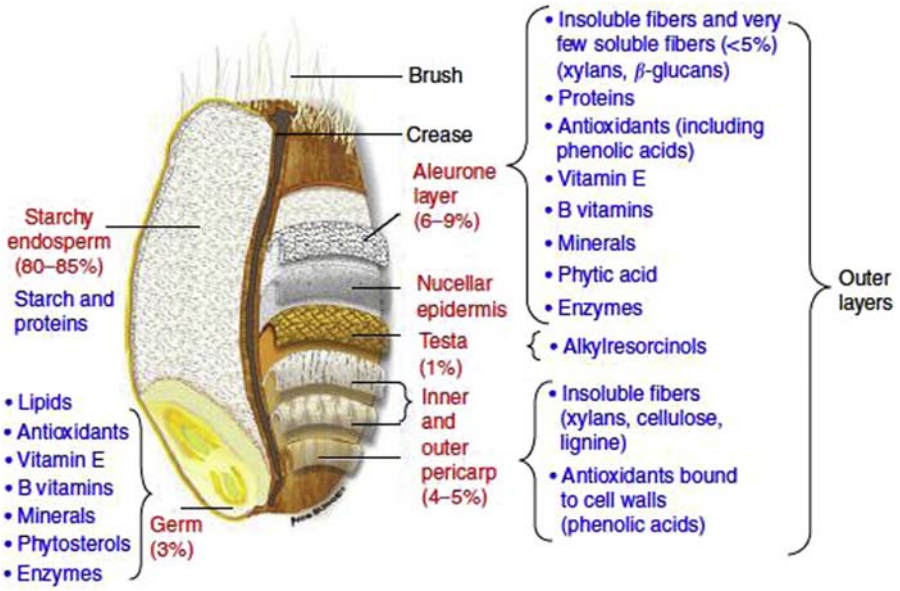


Fig. 1 Anatomical structure of a whole wheat grain. Reproduced with permission from Surget, A., Barron, C., 2005. Histologie du grain de blé. Industries des céréales 145, 3–7; Hemery, Y., Rouau, X., Lullien-Pellerin, V., Barron, C., Abecassis, J., 2007. Dry processes to develop wheat fractions and products with enhanced nutritional quality. J. Cereal Sci. 46(3): 327–347.

Bran

The bran is a tough outer layer that functions as a physical barrier to protect the seed. The bran is often referred to as a single component of the grain, but as Fig. 1 shows, there are several distinct layers including the pericarp and aleurone layers, which are each further subdivided. Modern milling techniques can be used to sequentially remove these layers. Removal of the bran is a vigorous, abrasive process known as de-hulling, pearling, or peeling. The principal nutrient in bran is dietary fiber, which is predominantly insoluble, although there can be significant quantities of soluble fibers such as arabinoxylan found in wheat and rye and beta-glucan in oats and barley. The bran layers are also a principal source of phytochemicals and contribute to the antioxidant potential of grain.

Endosperm

Quantitatively, the biggest part of the grain is the endosperm which comprises approximately 60–85% of the cereal grain dry weight, depending on species. The endosperm provides the necessary energy for the growing embryo. Nutritionally, the endosperm is mainly carbohydrate in the form of starches and other oligosaccharides such as fructans. Some protein is found in the extracellular matrix, as well as some B vitamins including pantothenic acid and riboflavin.

Germ

The germ is the smallest fraction of the grain at approximately 2.5% of the grain weight. The germ is the plant embryo which would form the new plant if germinated and is characterized by high lipid and protein content. The germ is also a rich source of minerals, particularly potassium, calcium, magnesium, and zinc, and both water- and fat-soluble vitamins including vitamin A, tocopherols, and tocotrienols. The oil content of the germ is prone to oxidation, which reduces the storage time for whole grain flours.

Definition of whole grains and health claims

Various definitions of "whole grain" exist and are largely based on the first definition proposed by the American Association of Cereal Chemists International (AACCI) in 1999:

> Whole grains consist of the intact, ground, cracked or flaked caryopsis (kernel) after the removal of inedible parts such as the hull and husk. The principal anatomical components—the starchy endosperm, germ, and bran—are present in the same relative proportions as they exist in the intact kernel.

More recently, to achieve consensus with respect to the definition of whole grains used in the European Union (EU), the HEALTHGRAIN Forum developed a similar definition in 2010 which additionally accounted for small losses (<2% of the grain or 10% of the bran) that can happen in modern industrial milling processes. This small loss is meant to account for the removal of the outermost layers of the bran, which is commonly done to eliminate mycotoxins, pesticides, or other contaminants that can be found in these layers. Intact whole grains must be processed before consumption to increase digestibility and improve the texture, flavor, and cooking characteristics of the products. During modern milling, the anatomical components of the grain are fractionated and recombined later to produce whole meal flour. Besides milling, less destructive processes such as rolling, flaking, and cracking of grains are also used. Intact whole grains are used in small quantities to add texture to some foods. Processing of the grains will, inevitably, result in changes to the gross morphology of the product and most likely the chemical composition. Although there is no evidence that recombined flours are nutritionally different than traditional stone-ground flours (where the grains are crushed without separation of the component fractions), the metabolic consequences of consuming whole grain foods prepared from heavily processed reconstituted flours compared to crushed, minimally processed flours have not been extensively investigated. Another characteristic known to change is the glycemic index (GI), which is higher in foods prepared from finely ground flours. This is because flours have a larger surface area for digestive enzymes to act on, thereby rapidly increasing blood glucose concentration.

Although there is general consensus for the definition of whole grains, defining whole grain *foods* becomes much more complex, particularly when a whole grain food contains additional ingredients including varying amounts of refined grain. There has been variation across countries and organizations regarding what should qualify as a whole grain food. In the US, the Food & Drug Administration (FDA) defines a whole grain food as a product containing 51% whole grain by weight per reference amount customarily consumed (RACC) per day, while the Whole Grain Council considers a product with 16 g of whole grain per serving eligible for a "100% Whole Grain" stamp. In Sweden, Norway, Denmark, the UK, and Taiwan, similar criteria of >50% of ingredients as whole grain have been used. In the Netherlands, 100% of the grain ingredients must be whole grain in breads, and in Germany more than 90% of ingredients must be whole grain. Others, such as Indonesia and Malaysia, allow certain whole grain or partial whole grain claims at lower amounts, around a minimum of 25%. Recently, the Whole Grain Initiative, consisting of global representatives from academia, government agencies, and industry, proposed a consensus definition:

> A whole-grain food shall contain at least 50% whole-grain ingredients based on dry weight. Foods containing 25–50% whole-grain ingredients based on dry weight, may make a front-of-pack claim on the presence of whole grain but cannot be designated "whole grain" in the product name.

The criteria for defining a whole grain food are important in determining what products can make whole grain related health claims. Since 1999, the FDA allows products meeting the 51% criteria to carry the statement "Diets rich in whole grain foods and other plant foods and low in total fat, saturated fat, and cholesterol may reduce the risk of heart disease and some cancers." Similar, equally cumbersome, claims in the UK and Europe followed. For example, the Swedish claim certified by the Swedish Nutrition Foundation states: "A healthy lifestyle and a well-balanced diet rich in whole grain products reduces the risk for (coronary) heart disease. The product X is rich in whole-grains (contains Y% of whole grain)." However, since 2010 the European Food Safety Authority made the decision to no longer approve whole grain health claims on the grounds that whole grain was "insufficiently characterized," particularly when the term was used to refer to all grains rather than a specific type of grain (i.e., wholegrain wheat or wholegrain rye).

Recommendations for whole grain consumption and consumption patterns

Many countries recommend or encourage greater intake of whole grains. Most provide qualitative statements such as "choose whole grains whenever possible," but only a few provide specific, quantitative recommendations. Since 2000, the US Dietary Guidelines have recommended that at least half of total grain-food intake should be as whole grains which, for those over 9 years of age, is a *minimum* of approximately 3 ounce-equivalents or 48 g per day. In Denmark, where there is already a stronger tradition of eating whole grain foods, the target level for consumption is 75 g of whole grain per 10 MJ (about 2400 kcal) energy consumed. Sweden has a similar recommendation, and in the Netherlands 90 g/day of brown bread or other whole grain products is recommended. However, most individuals consume amounts much lower than these recommendations. In the USA, for example, Americans consume a mean intake of <1 ounce-equivalent (or 16 g) of whole grains per day, according to data from the National Health and Nutrition Examination Survey (NHANES) (Shan et al., 2019). Mean consumption is of a similar magnitude in other major countries, such as France (14 g/d), Italy (16 g/d), the UK (20 g/d), and Australia (21 g/d) (Miller, 2020). Even in some Scandinavian countries, where whole grains are integral to traditional diets, mean consumption is still below their national recommendations, at around 55 g/d.

Biomarkers of whole grain consumption

Estimation of whole grain intake at the population level is necessary for development of public health initiatives and for exploring the health consequences of whole grain intake. However, methods for measuring whole grain intake vary and are prone to

measurement error. For example, in many earlier studies, dark bread was assumed to be whole grain bread, which may not be the case. Studies that collected data in the early 1990s were not originally designed to investigate whole grain intake (as it was not a focus of research interest at that time), and most research relied on the use of food frequency questionnaires to collect data on dietary intake. Some studies have reported intake in grams of whole grain per day, whereas others have reported intake as servings or as grams of whole grain food product, requiring additional assumptions on definitions of servings or whole grain foods. This creates a challenge for researchers attempting to estimate absolute whole grain intake and can confound observed associations, particularly when whole grain products contain a mixture of whole and refined grain. Although the estimation of absolute intake may vary depending on the method of measurement, relative associations observed between the highest and lowest consumers with chronic disease risk have been consistent.

More recently, there have been growing efforts to identify and utilize biomarkers of whole grain intake. Alkylresorcinols and their metabolites were the first proposed and are the most studied biomarkers specifically for whole grain wheat and rye intake (Ross, 2012; Sang, 2018). Other proposed biomarkers include benzoxazinoids (also for whole grain wheat, rye and sourdough rye), avenanthramides and avenacosides for whole grain oat intake, and enterolactone for total whole grain or cereal fiber intake (Landberg et al., 2019; Sang, 2018). New advances in targeted and untargeted metabolomic methods, which aim to identify and quantify metabolites in biological systems, are being used to study potential biomarkers for other grain varieties as well. These methods are also being utilized to identify potential mechanistic pathways by which whole grains may impact health. For example, in one randomized cross-over trial, untargeted metabolomics identified 15 metabolites associated with rye intake, including significantly lower plasma serotonin, which may be linked to metabolic control (Keski-Rahkonen et al., 2019).

Whole grain consumption and disease risk

Cardiovascular disease

Approximately one-third of global deaths can be attributed to cardiovascular disease (CVD), and morbidities associated with CVD account for significant proportions of health care costs. The inverse relationship between higher whole grain intake and lower risk of CVD has been documented in numerous observational studies, and for all types of CVD. Much of this data has been generated from large observational cohort studies carried out in North America and Europe. Meta-analyses that include data from multiple observational studies show significant and marked lowering of CVD risk (20–30%) between the lowest and highest consumers of whole grain intake. One meta-analysis estimated that 11.6% of all CVD mortality could be attributed to low whole grain intake (intake below the recommended amount) (Zong et al., 2016). Conversely, positive relationships between higher refined grain intake and higher risk of CVD have been observed in some observational studies, although others have found no significant association. These discrepancies may be due to differences in study design, background diet, or study population. It is important to note that higher whole grain intake is associated with many other indicators of a healthy lifestyle, such as lower smoking status, higher physical activity, use of nutritonal supplements and generally higher household income. Therefore, the relationships observed with whole grain intake may be due to other components of a generally healthier lifestyle. However, after mitigating these confounding factors through statistical adjustment, the inverse relationships between whole grain intake and CVD risk continue to remain significant. **Fig. 2** summarizes some of these data, using fully adjusted models from some of the large cohort studies investigating CVD incidence or mortality as endpoints. The range of lower risk, estimated by either hazards ratios or relative risk of incident CVD or CVD mortality, remains large—between 5% and 30%. Most of these analyses use data from large US cohort studies, including the Health Professionals Follow-up ($n = 42{,}850$); Nurses' Health Study ($n = 75{,}521$); Iowa Women's Health Study ($n = 34{,}491$); and Atherosclerosis Risk in Communities Study ($n = 15{,}972$). Other studies have also evaluated intermediate CVD risk factors in relation to whole grain intake. In a meta-analysis of 25 randomized controlled trials (RCTs), researchers observed improved total cholesterol (standardized mean difference [SMD] = −0.54, 95% CI −0.95 to −0.12) and LDL = −0.57, 95% CI −0.84 to −0.31), hemoglobin A1c (SMD = −0.33, 95% CI −0.61 to −0.04) and C-reactive protein (SMD = −0.22, 95% CI −0.44 to −0.00) with whole grain (all types) intake (Marshall et al., 2020). Functional measures of cardiac health are few but in one prospective cohort study involving postmenopausal women ($n = 229$), smaller reductions in minimum coronary artery diameter and a trend toward lower progression in mean percent stenosis with higher whole grain intake have been reported (Erkkilä et al., 2005).

Type 2 diabetes

Incidence of type 2 diabetes (T2D) has increased dramatically in line with the rise in obesity worldwide, and consequences of the epidemic represent a major challenge for health services. The etiology of the disease is complex, including interactions between genes, diet, and lifestyle. The underlying pathology is the development of insulin resistance, which has been linked to high intake of refined carbohydrates, along with low intake of dietary fiber. Many of the large cohorts used to examine the relationship between whole grain intake and CVD risk have also been used to explore similar relationships for T2D (**Fig. 3**). These observational studies suggest an inverse relationship between whole grain intake and risk of T2D (with a lower risk, ranging between 11% and 43%), as well as lower concentrations of glucose and insulin, where these measures are available. Prospective data from the National Institutes of Health (NIH)-AARP Diet and Health Study, which followed 367,442 participants for an average

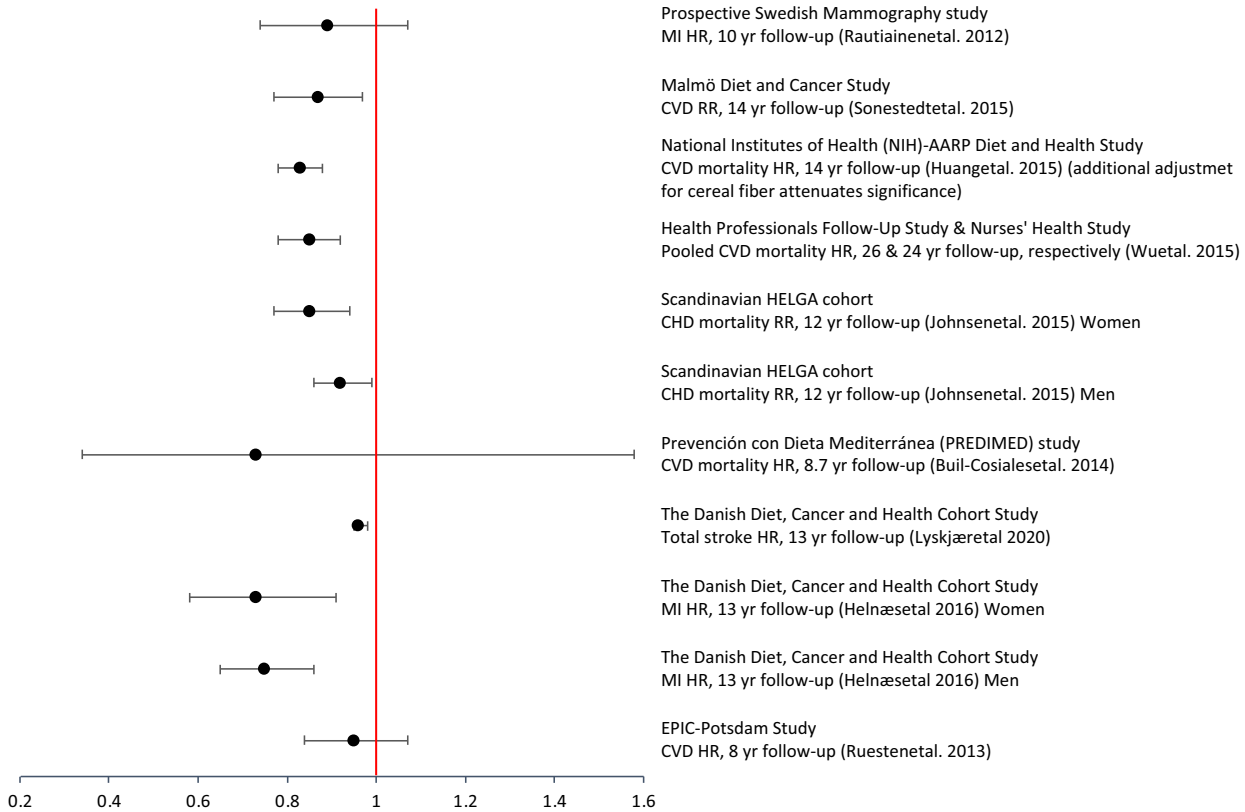


Fig. 2 Observational cohort studies on whole grain intake and cardiovascular disease (CVD). All models presented are the most fully adjusted unless otherwise stated. CHD = coronary heart disease, HR = hazard ratio, MI = myocardial infarction, RR = relative risk.

of 14 years with 46,067 total deaths, showed that the highest intake of whole grain (1.20 oz eq/day) vs. the lowest (0.13 oz eq/day) was associated with 43% lower risk of death from diabetes (HR 0.57 95% CI 0.37–0.89) after adjustment for potential confounders, including BMI (Huang et al., 2015). More recently, in a pooled analysis of 3 large cohorts with data from 158,259 women and 36,525 men (Nurses' Health Study I and II and Health Professionals Follow-up Study), with an average follow-up time of 24 years, and 18,629 incident cases of T2D, 29% lower risk of T2D was observed among those in the highest vs lowest category of whole grain intake (Hu et al., 2020). The greatest lowering of risk was observed at 2 servings (about 32 g) per day, after which further lowering of risk seemed to plateau.

A meta-analysis of 45 prospective cohort studies showed that, compared with people who never or rarely ate whole grain, those consuming 48–80 g/d (~3–5 servings/d) had ~26% lower risk of developing T2D [RR = 0.74 (95% CI: 0.69, 0.80)] (Ye et al., 2012). These authors also conducted a meta-analysis of 21 RCTs and observed that those whole grain interventions resulted in a weighted mean difference in fasting glucose of −0.93 mmol/L (95% CI: −1.65, −0.21) compared to control groups. However, many intervention studies investigating the effect of whole grain consumption on blood glucose and insulin resistance/sensitivity have yielded conflicting results. For example, in a meta-analysis of 14 RCTs, significant associations were observed between whole grain foods and acute changes in post-prandial glucose [iAUC (0–120 min) by −29.71 mmol min/L (95% CI: −43.57, −15.85 mmol min/L)] and insulin [iAUC (0–120 min) by −2.01 nmol min/L (95% CI: −2.88, −1.14 nmol min/L)] (Marventano et al., 2017). However, when the same study looked at 16 RCTs of medium or long duration, no significant effect was observed on fasting glucose, insulin, or insulin resistance. The discrepancies in these results may be due to differences in study populations, baseline disease risk, or the background diet that the study participants consumed.

Cancer

Unhealthy diet patterns are a major contributor to cancer, partially due to their direct and indirect effects on adiposity. On the other hand, consistent evidence from scientific literature shows that whole grains may protect against various types of cancer. In a 2020 meta-analysis of both case-control and cohort studies, whole grain consumption was associated with lower risk of total cancer mortality (Gaesser, 2020). Seven meta-analyses of prospective cohort studies indicated that whole grain intake was associated

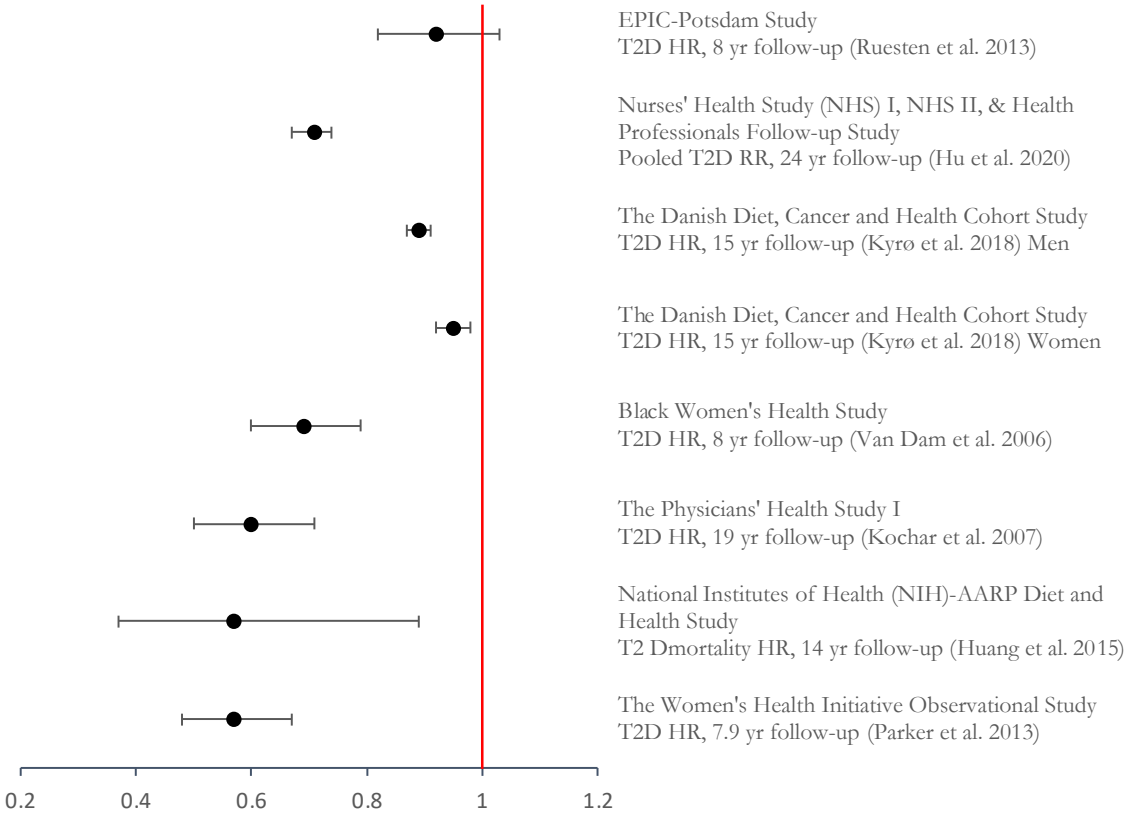


Fig. 3 Observational cohort studies on whole grain intake and Type 2 diabetes (T2D). All models presented are the most adjusted unless otherwise stated. HR = hazard ratio, RR = relative risk.

with 6%–12% lower risk of cancer mortality, when comparing highest vs. lowest intake groups. In dose-response analyses, each 30 g/day intake of whole grains was associated with ~7% lower risk of cancer mortality. For site-specific cancers, greater whole grain intake was associated with lower colorectal, gastric, pancreatic, and esophageal cancers (**Table 2**). The largest body of epidemiological evidence is for whole grain intake and risk of incident colorectal and breast cancers (**Table 3**).

Colorectal cancer (CRC)

In 2020, an estimated 1.9 million individuals were diagnosed with CRC, the second leading cause of cancer and the third leading cause of cancer death worldwide. Case-control studies, in particular, show lower risk of CRC with greater whole grain consumption; however, evidence from prospective studies is limited and inconsistent, with some studies showing protective associations between whole grain intake and CRC and others showing a null association. A meta-analysis of data from 10 individual prospective cohorts (Schwingshackl et al., 2018), reported 5% lower risk per 30 g serving of whole grains per day—translating into 20% lower CRC risk with whole grain intakes up to 120 g/day. More recently, two umbrella reviews graded the evidence from published meta-analyses of observational studies on the association of whole grain intake and incidence of CRC, and determined that the evidence was "possible/suggestive" for a protective impact of risk with higher intakes (Schwingshackl et al., 2018; Tieri et al., 2020). Estimates of whole grain intakes in servings per day or grams of whole grain products, rather than grams of whole grain ingredients, led to insufficient dose-response data. In addition, there was limited data on the impact of different sources of whole grains. Inconsistencies in associations may be due to a few incident cases, low habitual intake of whole grains in the sample, different lengths of follow-up, or lack of repeated measures to capture usual or long-term dietary intake. In addition to absolute whole grain intake, the source of whole grain and degree of processing may be important to consider with respect to CRC risk.

Total and individual whole grain *product* consumption among Danish men and women in the Diet, Cancer, and Health cohort were examined in relation to incident risk of colon and rectal cancer over a median follow-up of 10.6 years (Egeberg et al., 2010). Among 26,630 men, there were 461 incident cases of colon cancer and the incident relative risk (IRR) per 50 g of whole grain intake was 0.85 (95% CI 0.77, 0.94). There were 283 incident cases of rectal cancer with an IRR of 0.90 (95% CI 0.80, 1.01). There were no consistent associations between total or individual whole grain consumption and colon or rectal cancer risk among women. In men, whole grain wheat bread was the only whole grain food significantly associated with lower risk of colon cancer, where every 25 g per

Table 2 Whole grain intake and risk of cancer mortality.

	Highest vs. lowest intakes		*Dose response*		
Meta-analysis	*# Of cohorts or case-control studies included*	*RR or OR (95% CI)*	*# Cohorts or case-control studies included*	*Dose*	*RR or OR (95% CI)*
Jacobs et al., 1998	45	0.66 (0.60–0.72)			
Aune et al., 2016	6	0.89 (0.82–0.96)	6	90 g/day	0.85 (0.80–0.91)
Benisi-Kohansal et al., 2016	7	0.94 (0.91–0.98)	3	90 g/day	0.90 (0.83–0.98)
Chen et al., 2016	8	0.89 (0.84–0.95)	6	50 g/day	0.82 (0.69–0.86)
Wei et al., 2016	8	0.89 (0.82–0.96)	7	90 g/day	0.91 (0.84–0.98)
Zong et al., 2016	10	0.88 (0.83–0.94)	10	70 g/day	0.80 (0.72–0.89)
				50 g/day	
				30 g/day	0.85
				10 g/day	(0.76–0.94)
					0.89 (0.79–0.99)
					0.96 (0.91–1.01)
Zhang et al., 2018	14	0.94 (0.87–1.01)	14	28 g/day	0.97 (0.95–0.99)
Reynolds et al., 2019	5	0.84 (0.76–0.92)	7	15 g/day	0.95 (0.93–0.97)

RR = Relative Risk; OR=Odds Ratio; CI = confidence interval. CI = confidence interval; * all meta-analyses reported total cancer mortality except for the case-control meta-analysis of Jacobs et al. who reported total cancer risk for multiple sites combined. Unfilled field indicates that no meta-analyses were performed. g/day refers to the dose of whole grain intake associated with the corresponding relative risk or odds ratio in the dose-response analysis.
*Results from meta-analyses of observational cohort and case-control studies.
Adapted from Gaesser (2020).

day was associated with 11% lower risk of colon cancer (IRR 0.89; 95% CI 0.82, 0.97).Although not significant, a slight lowering in colon cancer risk was also observed for rye bread (IRR 0.94; 95% CI 0.88, 1.01). This response is interesting because intake of rye bread was approximately twice that of the wheat bread (63 g day compared with 31 g day), and rye was shown to lower colon cancer risk in women from the Swedish Mammography Cohort (Larsson et al., 2005).

Data from the Health Professionals Follow-up Study demonstrated that higher intake of whole grain and dietary fiber attributed to whole grain (i.e., cereal fiber) were associated with 25–27% lower CRC risk in men. However, no significant lowering of risk was observed in women in the Nurses' Health Study, which may be due to the lower intake of whole grains in women (He et al., 2019). More recently, the association between whole grains and fiber sources on risk of CRC, was examined in the largest cohort to date—478,994 US adults, 51–70 years of age, in the NIH-AARP Diet and Health Study (Hullings et al., 2020). Compared to those in the lowest quintile category of whole grain intake (daily median intake 0.2 servings per 1000 kcal), those in the highest quintile (1.3 servings per 1000 kcal) had 16% and 24% lower risk of developing CRC and rectal cancer, respectively. Interestingly, these associations remained statistically significant after adjustment for two mediating nutritional attributes of whole grains: dietary fiber and folate. This would suggest that other nutritional constituents of the whole grain may be responsible for the association. In this same analysis, of all potential fiber sources, only fiber from grains demonstrated a significant linear inverse association with CRC among those in the highest category of cereal fiber intake. There was no evidence that associations differed by sex. The most recent 2018 WCRF/AIRC Continuous Update Project concluded that there is strong evidence that consuming whole grains help protects against colorectal cancer.

Breast cancer

Based on data from the WHO in 2020, 2.3 million women were diagnosed with breast cancer and 685,000 died from the disease globally, with a large proportion of deaths being attributed to obesity and lifestyle related factors, including diet. However, evidence suggesting that a higher intake of whole grains is protective against breast cancer risk has been inconsistent.

Some observational studies show small but modest benefits of whole grain consumption (Adzersen et al., 2003; Chatenoud et al., 1998; Mourouti et al., 2016; Sonestedt et al., 2008; Tajaddini et al., 2015; Yun et al., 2010), but others show either no benefit (Egeberg et al., 2009; Farvid et al., 2016; La Vecchia et al., 1987; Levi et al., 1993) or, unexpectedly, higher risk in breast cancer risk (Nicodemus et al., 2001). The latter study used data from the Iowa Women's Health Study and was the first prospective cohort to consider grain sources on breast cancer among post-menopausal women in the United States. After multivariable adjustment, the

Table 3 Whole grain intake and risk of site-specific cancer: results from meta-analyses of observational cohort and case-control studies.

	Highest vs. Lowest intakes		*Dose response*			*Cancer site*
Meta-analysis	*# Of cohorts or case-control studies included*	*RR or OR (95% CI)*	*# Of cohorts or case-control studies included*	*RR or OR (95% CI)*	*Dose*	
Jacobs et al., 1998	7	0.79 (0.69–0.89)				Colorectal
Aune et al., 2016	4	0.79 (0.72–0.86)	6	0.83 (0.78–0.89)	90 g/day	Colorectal
Vieira et al., 2017			6	0.83 (0.79–0.89)	90 g/day	Colorectal
Schwingshackl et al., 2018	10	0.88 (0.83–0.94)	9	0.95 (0.93–0.97)	30 g/day	Colorectal
Reynolds et al., 2019	7	0.87 (0.79–0.96)	8	0.97 (0.95–0.99)	15 g/day	Colorectal
Zhang et al., 2020	25	0.89 (0.84–0.93)				Colorectal
Aune et al., 2011	5	0.82 (0.72–0.92)	4	0.86 (0.79–0.94)	90 g/day	Colon
Vieira et al., 2017			4	0.82 (0.73–0.92)	90 g/day	Colon
Schwingshackl et al., 2018	7	0.85 (0.77–0.93)	6	0.97 (0.95–0.99)	30 g/day	Colon
Aune et al., 2011	3	0.80 (0.59–1.07)	3	0.80 (0.56–1.14)	90 g/day	Rectal
Vieira et al., 2017			3	0.81 (0.54–1.20)	90 g/day	Rectal
Schwingshackl et al., 2018	5	0.80 (0.64–0.98)	5	0.94 (0.88–1.01)	30 g/day	Rectal
Jacobs et al., 1998	7	0.57 (0.47–0.67)				Gastric
Wang et al., 2020	5	0.87 (0.79–0.95)				Gastric
Xu et al., 2019	3	0.61 (0.40–0.83)				Gastric
Zhang et al., 2020	12	0.64 (0.53–0.79)				Gastric
Jacobs et al., 1998	4	0.70 (0.54–0.86)				Pancreatic
Lei et al., 2016	5	0.76 (0.64–0.91)				Pancreatic
Wang et al., 2015	8	1.13 (0.98–1.30)				Prostate
Reynolds et al., 2019	3	1.10 (1.02–1.19)	2	1.02 (0.98–1.05)	15 g/day	Prostate
Jacobs et al., 1998	2	0.86 (0.67–1.05)				Breast
Xiao et al., 2018	11	0.84 (0.74–0.96)	6	0.83 (0.73–0.93)	50 g/day	Breast
Jacobs et al., 1998	2	0.52 (0.09–0.95)				Esophageal
Zhang et al., 2020	7	0.54 (0.44–0.67)				Esophageal
Jacobs et al., 1998	4	0.57 (0.38–0.76)				Oral
Jacobs et al., 1998	2	0.67 (0.48–0.86)				Brain
Jacobs et al., 1998	3	0.55 (0.41–0.69)				Endometrial
Jacobs et al., 1998	2	0.41 (0.37–0.45)				Non-Hodgkin's lymphoma

RR = Relative Risk; OR=Odds Ratio; CI = confidence interval. Unfilled fields indicate that no meta-analyses were performed. g/day refers to the dose of whole grain intake associated with the corresponding relative risk or odds ratio in the dose-response analysis.
Adapted from Gaesser (2020).

risk of incident breast cancer was 20% *higher* for women with the highest, vs. lowest whole grain intake. However, when the data were further examined, it appeared that the incidence of breast cancer was not higher among women who had not undergone screening mammography before follow-up. This suggests that increasing whole grain intake was associated with more health-conscious behaviors such as attending cancer screening, thus increasing the detection rate in whole grain consumers. In a cohort of Danish post-menopausal women, no protective association was observed between whole grain intake and breast cancer risk, including all estrogen and progesterone receptor disease subtypes. It is possible that hormone replacement therapy may interfere with the mechanisms of lignans/enterolactone with higher whole grain intake.

A 2018 meta-analysis of data from 11 published studies (4 cohort and 7 case-control) suggested that, compared to the lowest intake, those with the highest whole grain intake had 16% lower risk of breast cancer (pooled RR 0.84, 95% CI: 0.74–0.96, $P = 0.009$) (Gaesser, 2020). When stratified by study design, the observed relationships were stronger for case-control than for cohort studies and, in some studies, sample size was cited as a potential limitation. Therefore, it remains inconclusive as to whether higher intake of whole grains is linked to a lower risk of breast cancer.

Other cancers

Research on whole grains and other cancers is limited. With respect to prostate cancer, results are mixed (Egeberg et al., 2011; Nimptsch et al., 2011). In one nested case-control study in Swedish men, no evidence of a protective association was observed

with the blood biomarker, plasma alkylresorcinol, and incident prostate cancer (Drake et al., 2014). The association between whole grain intake and other cancers such as endometrial (Aarestrup et al., 2012; Kasum et al., 2001), pancreatic (Lei et al., 2016), and stomach (McCullough et al., 2001) are limited to a few prospective cohort studies, with mixed findings and, thus, inconclusive evidence.

Intermediate risk factors of disease and potential mechanisms of action

There are various potential mechanistic pathways, many yet to be fully understood, by which whole grain consumption may lower chronic disease risk. When compared to refined grains, whole grains are higher in several nutrients and many potentially protective bioactive compounds found in the bran and germ layers, which are lost during the milling process. These include dietary fiber, magnesium, potassium, vitamin K, vitamin E, antioxidants, and other phytochemicals, all of which have been linked to cardiometabolic health. For example, dietary fiber and magnesium may reduce insulin resistance, and potassium has been associated with lower blood pressure. Although different grains have similar nutritional composition, they vary in type and amounts of specific nutrients and other bioactive components, which can lead to differential health benefits. For example, oats have a stronger favorable association with triglyceride, LDL, and total cholesterol compared to whole wheat or mixed grains (Hollænder et al., 2015), likely because oats are higher in soluble fiber, whereas wheat is higher in insoluble fiber. Emerging research utilizing metabolomics methods aim to identify metabolites that may further link whole grain intake to health benefits.

Whole grains and adiposity

A contributing factor in the development of many chronic diseases is excess body fat. Higher adiposity has been linked to greater risk of T2D, CVD, and cancer. In several observational studies, higher whole grain consumption was associated with lower body weight, BMI, and waist circumference (a measure of abdominal or central obesity), which may contribute to the observed protective association observed with greater whole grain intake in chronic disease development. Prospective analyses that measure change in weight over many years suggest that, although individuals tend to gain weight with age, those with higher whole grain intake gain less weight over time than those with lower intake. One recent dose-response meta-analysis of five prospective cohort studies found that for every additional 30 g/d of whole grain intake, the risk of overweight or obesity was 7% lower. In three other prospective cohorts, the risk of weight gain ($>$2 kg over 4 years) was $\geq$17% lower among the highest, vs. lowest whole grain consumers (Schlesinger et al., 2019). Similar results have been observed in the Nurses' Health Study (Liu et al., 2003) and Health Professionals Follow-up Study (Mozaffarian et al., 2011), as well as in the Black Women's Health Study (Boggs et al., 2013) cohorts from Spain (PREDIMED (Bautista-Castaño et al., 2013), and SUN (de la Fuente-Arrillaga et al., 2014)), and Australia (ALSWH (Quatela et al., 2017)).

Evidence from RCTs has been less consistent. One meta-analysis, that included both prospective cohorts and RCTs, found a significant association between higher whole grain intake and less weight gain over time among the cohort studies; however, no association with body weight was observed in the RCTs (Maki et al., 2019). Another recent meta-analysis of 21 RCTs, ranging from 2 to 24 weeks, also found no significant association between whole grain intake and body weight, BMI, body fat percentage, or other measures of obesity (Sadeghi et al., 2020). Conversely, another meta-analysis of 11 RCTs, ranging from 3 weeks to 18 months, found 0.62 kg lower weight among the highest, vs. lowest, whole grain consumers (Reynolds et al., 2019). There is heterogeneity across RCT designs, and each meta-analysis had different inclusion and exclusion criteria, possibly contributing to their different conclusions.

Emerging research shows that even among individuals with similar BMI, variability in body fat distribution may differentially impact metabolic disease risk. In particular, excess abdominal obesity, most easily measured by waist circumference, has been linked to higher risk of metabolic syndrome, T2D, and CVD, independent of overall BMI (Casanueva et al., 2010; Müller et al., 2012). Observational data support a potential beneficial association between whole grain intake and lower waist circumference. A recent analysis with 3121 participants of the Framingham Heart Study found that higher whole grain intake was significantly associated with less increase in waist circumference over time. Those in the highest quartile of whole grain intake had an increase of only 1.4 cm per 4-year period compared to a 3.0 cm increase in the lowest quartile (Sawicki et al., 2021). This seemingly small difference adds up over time, and even small increases in risk factors such as waist circumference have been shown to impact disease risk.

When specific adipose tissues are considered separately, visceral adipose tissue (VAT), the fat that surrounds organs, is more strongly related to metabolic risk than subcutaneous adipose tissue (SAT), the fat that lies under the skin. Greater VAT has been linked to insulin resistance, dyslipidemia, inflammation, and oxidative stress, all of which increase the risk of cardiometabolic diseases (Fox et al., 2007). Data from the Framingham Heart Study demonstrate that those in the highest whole grain consumption group had less SAT and less VAT. In contrast, those from the highest refined grain consumption group had more of both types of abdominal fat, especially VAT (McKeown et al., 2010). When SAT and VAT were mutually adjusted, only VAT remained significantly associated with whole and refined grain. More recently, a randomized, controlled 12-week intervention with 49 overweight participants in Japan reported that those randomized to a whole grain diet group lost a mean of 4 cm^3 in VAT, compared to no significant

change in the refined grain group (Kikuchi et al., 2018). However, other RCTs have not found a significant association between whole grain and VAT. A systematic review and meta-analysis of randomized controlled trials of whole grain and body weight changes concluded that whole grain may have a small beneficial effect on body fat, but not overall body weight, and that short duration and heterogeneity among study designs may contribute to the conflicting evidence (Pol et al., 2013).

Whole grains and hypertension

A small number of observational studies have shown an inverse relationship between hypertension and whole grain intake. A meta-analysis of 4 prospective observational studies reported that, for every 30 g/d higher whole grain intake, the risk of hypertension was lower by 8%. No significant association was observed with refined grain (Schwingshackl et al., 2017). It should be noted that two of the 4 included studies were in men only. A more recent prospective study in the Framingham Offspring Cohort found that high, vs. lower, whole grain intake was associated with smaller increases in systolic, but not diastolic, blood pressure over time (Sawicki et al., 2021). The effects of whole grain consumption on blood pressure have also been measured in intervention studies with varying results. A double-blind RCT with 33 overweight or obese men and women found that, after 8 weeks on a whole grain diet (50 g whole grain/1000 kcal per day) diastolic blood pressure decreased by 5.8 mm Hg (95% CI: −7.7, −4.0 mm Hg, $P = 0.01$ compared to control diet) (Kirwan et al., 2016). One RCT from the UK with more than 200 healthy volunteers, showed that consuming three servings per day of whole grain wheat or a 50:50 mix of whole grain wheat and whole grain oats for 12 weeks led to a significant reduction of 6 mmHg in systolic blood pressure and 3 mmHg in pulse pressure (Tighe et al., 2010). In another RCT of similar size, however, volunteers consumed either three servings of mixed whole grains per day for 16 weeks or three servings per day for 8 weeks followed by six servings per day for a further 8 weeks, with no changes in blood pressure (Brownlee et al., 2010). Heterogeneity of trials is emphasized by a meta-analysis that included 8 RCTs and did not show a statistically significant effect of WG on systolic blood pressure (Reynolds et al., 2019).

Whole grains and inflammatory status

It is now recognized that chronic subclinical inflammation is central in the progression of many diseases such as CVD and T2D. In several studies, relationships between whole grain consumption and markers of inflammatory status such as high-sensitivity C-reactive protein (hsCRP), tumor necrosis factor receptor-2 (TNF-R2), fibrinogen, plasminogen activator inhibitor-1 (PAI-1), or interleukin-6 (IL-6) have been investigated. For some, higher whole grain consumption was associated with lower inflammatory marker concentrations. In one cohort study of 259 healthy premenopausal women, participants who ate small amounts of whole grain (<one serving per day) for 2 years had lower hsCRP concentration (average 11.5%) than those who ate no whole grain at all (Gaskins et al., 2010). For women eating one or more servings of whole grain, hsCRP concentration was 12.3% lower. A more recent RCT showed improvement in IL-6 and CRP after 8 weeks on a whole grain intervention, compared to a refined grain control (Roager et al., 2017). Similar associations were found in two meta-analyses of RCTs (Hajihashemi and Haghighatdoost, 2019; Xu et al., 2018). While the underlying mechanisms are not completely understood, whole grains are rich in antioxidants and phytochemicals that may play a role in inflammatory pathways. Additionally, the fiber in whole grain may reduce postprandial hyperglycemia, thus reducing the formation of advanced glycation end products which are known to induce oxidative stress and inflammation.

Whole grain, dietary fiber, and fermentable carbohydrates

Whole grains are considered "complex carbohydrates" because they are good sources of dietary fiber. The variety of fibers, or non-starch polysaccharides, derived from cereal grains include cellulose, hemicellulose, arabinoxylans, lignin, fructans, such as inulin, and resistant starch. In many of the observational studies described above, health benefits associated with higher whole grain intake were similarly correlated with cereal fiber intake. The myriad of physiological effects of dietary fiber in the digestive tract have been linked to health benefits that may contribute to the reduction in disease risk associated with whole grain intake (McRorie and McKeown, 2017). These include direct effects on digestive characteristics such as viscosity, flow rate/transit time, and bulking. These changes then affect digestive processes, such as reducing stomach emptying, increasing satiety, reducing rates of carbohydrate and lipid digestion, blunting the postprandial glucose and insulin response, and lowering glycemic index/glycemic load. These effects may contribute to the observed associations between whole grain or cereal fiber and lower adiposity and better glycemic/insulinemic maintenance. Some dietary fiber may also bind and help remove bile acids in the small intestine, interrupting the enterohepatic circulation of cholesterol and lowering blood lipid concentrations. Similarly, dietary fiber may potentially reduce endogenous estrogen levels by binding estrogen and facilitating their fecal excretion.

The dietary fibers in whole grains also provide fermentable carbohydrates to the gut microbiota in the large intestine. The potential prebiotic effects of nondigestible oligosaccharides, such as fructans and arabinoxylan fractions, have also been suggested. Emerging research suggests that modulation of the gut microbiome composition can have physiological effects. Some, but not all, RCTs have shown that higher abundance of *Bifidobacterium* and *Lactobacillus* spp., promoted by fructans and galacto-oligosaccharides, is associated with improved metabolic health (So et al., 2018). In a recent RCT, healthy overweight adults with *Prevotella* abundance lost more body weight on an ad-libitum whole grain diet compared to refined grain diet, whereas those with low *Prevotella* abundance did not lose weight (Christensen et al., 2019). Fermentation of dietary fiber by the microbiota present

in the colon can also produce short-chain fatty acids (SCFAs), which are absorbed into the blood stream and can activate fatty acid oxidation and inhibit de novo synthesis and fat storage. SCFAs may also stimulate the production of gastrointestinal hormones that play a role in satiety and glucose metabolism, including PYY and GLP-1.

Whole grains, antioxidants, and phytochemicals

Phytochemicals are bioactive non-nutrient plant compounds in fruits, vegetables, whole grains, and other plant foods that have been associated with a lower risk of major chronic diseases. The most studied phytochemicals found in grains are phenolics. Phenolics are compounds with one or more aromatic rings and one or more hydroxyl groups, such as phenolic acids, plant lignans, alkylresorcinols, and flavonoids. Phenolics exist in cereal grains in two forms: soluble free and insoluble bound; the latter include compounds esterified to macromolecules which, as a result, may escape digestion in the upper gastrointestinal tract. It has been suggested that as much as 74% of phenolic compounds in wheat, maize, oats, and barley are found in the insoluble bound fraction. Phenolic compounds exist mainly as glycosides linked to various sugar moieties or as other complexes linked to organic acids, amines, lipids, carbohydrates, and other phenols.

These fractions may be released during the processing of the grain and be available in the small intestine, or they may be released during bacterial fermentation in the large intestine. Thus, there is potential for local effects in different parts of the gut or systemic effects once the compounds are absorbed. Phenolics and other compounds found in cereal grains can act as antioxidants because they can donate hydrogen atoms to free radicals. In principle, the higher the phenolics and other antioxidant compounds in cereal grains, the greater their potential antioxidant capacity, and this has been well documented *in vitro* using a variety of assays. However, their potential to affect antioxidant status *in vivo* is less certain, and the mechanisms by which antioxidants that cross the intestinal barrier protect the body remain unclear. Of note are lignans, a group of dietary phytoestrogen compounds found in the bran of cereals that have strong antioxidant capacity *in vitro* and may impact endogenous estrogen and/or sex hormones by binding or inhibiting enzymes involved in steroid hormone physiology that may promote health benefits, such as susceptibility to hormone-related cancers.

Conclusion

Evidence shows that whole grains contribute to a healthy diet. They have been linked to reduced risk of several chronic diseases, through diverse potential mechanistic pathways that are still active areas of research. Currently, most countries report that whole grain intake is below recommendations. Given increasing evidence for their health benefits, global efforts to increase whole grain intake are warranted.

References

Aarestrup, J., Kyro, C., Christensen, J., Kristensen, M., Wurtz, A.M., Johnsen, N.F., Overvad, K., Tjonneland, A., Olsen, A., 2012. Whole grain, dietary fiber, and incidence of endometrial cancer in a Danish cohort study. Nutr. Cancer 64 (8), 1160–1168. https://doi.org/10.1080/01635581.2012.723786.

Adzersen, K.H., Jess, P., Freivogel, K.W., Gerhard, I., Bastert, G., 2003. Raw and cooked vegetables, fruits, selected micronutrients, and breast cancer risk: a case-control study in Germany. Nutr. Cancer 46 (2), 131–137. https://doi.org/10.1207/S15327914NC4602_05.

Aune, D., Chan, D.S.M., Lau, R., Vieira, R., Greenwood, D.C., Kampman, E., Norat, T., 2011. Dietary fiber, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. BMJ 343, d6617.

Aune, D., Keum, N., Giovannucci, E., Fadnes, L.T., Boffetta, P., Greenwood, D.C., Tonstad, S., Vatten, L.J., Riboli, E., Norat, T., 2016. Whole grain consumption and risk of cardiovascular disease, cancer, and all cause and cause specific mortality: systematic review and dose-response meta-analysis of prospective studies. BMJ 353, i2716. https://doi.org/10.1136/bmj.i2716.

Bautista-Castaño, I., Sánchez-Villegas, A., Estruch, R., Martínez-González, M.A., Corella, D., Salas-Salvadó, J., Covas, M.I., Schroder, H., Alvarez-Pérez, J., Quilez, J., Lamuela-Raventós, R.M., Ros, E., Arós, F., Fiol, M., Lapetra, J., Muñoz, M.A., Gómez-Gracia, E., Tur, J., Pintó, X., et al., 2013. Changes in bread consumption and 4-year changes in adiposity in Spanish subjects at high cardiovascular risk. Br. J. Nutr. 110 (2), 337–346. https://doi.org/10.1017/S000711451200476X.

Benisi-Kohansal, S., Saneei, P., Salehi-Marzijarani, M., Larijani, B., Esmaillzadeh, A., 2016. Whole-grain intake and mortality from all causes, cardiovascular disease, and cancer: a systematic review and dose-response meta-analysis of prospective cohort studies. Adv. Nutr. 7 (6), 1052–1065. https://doi.org/10.3945/an.115.011635.

Boggs, D.A., Rosenberg, L., Rodríguez-Bernal, C.L., Palmer, J.R., 2013. Long-term diet quality is associated with lower obesity risk in Young African American women with normal BMI at baseline. J. Nutr. 143 (10), 1636–1641. https://doi.org/10.3945/jn.113.179002.

Brownlee, I.A., Moore, C., Chatfield, M., Richardson, D.P., Ashby, P., Kuznesof, S.A., Jebb, S.A., Seal, C.J., 2010. Markers of cardiovascular risk are not changed by increased whole-grain intake: The WHOLE heart study, a randomised, controlled dietary intervention. Br. J. Nutr. 104 (1), 125–134. https://doi.org/10.1017/S0007114510000644.

Casanueva, F.F., Moreno, B., Rodríguez-Azeredo, R., Massien, C., Conthe, P., Formiguera, X., Barrios, V., Balkau, B., 2010. Relationship of abdominal obesity with cardiovascular disease, diabetes and hyperlipidaemia in Spain. Clin. Endocrinol. 73 (1), 35–40. https://doi.org/10.1111/j.1365-2265.2009.03727.x.

Chatenoud, L., Tavani, A., La Vecchia, C., Jacobs, D.R., Negri, E., Levi, F., Franceschi, S., 1998. Whole grain food intake and cancer risk. Int. J. Cancer 77 (1), 24–28. https://doi.org/10.1002/(sici)1097-0215(19980703)77:1<24::Aid-ijc5>3.0.Co;2-1.

Chen, G.C., Tong, X., Xu, J.Y., Han, S.F., Wan, Z.X., Qin, J.B., Qin, L.Q., 2016. Whole-grain intake and total, cardiovascular, and cancer mortality: a systematic review and meta-analysis of prospective studies. Am. J. Clin. Nutr. 104 (1), 164–172. https://doi.org/10.3945/ajcn.115.122432.

Christensen, L., Vuholm, S., Roager, H.M., Nielsen, D.S., Krych, L., Kristensen, M., Astrup, A., Hjorth, M.F., 2019. Prevotella abundance predicts weight loss success in healthy, overweight Adults consuming a whole-grain diet ad libitum: a post hoc analysis of a 6-wk randomized controlled trial. J. Nutr. 149 (12), 2174–2181. https://doi.org/10.1093/jn/nxz198.

de la Fuente-Arrillaga, C., Martinez-Gonzalez, M.A., Zazpe, I., Vazquez-Ruiz, Z., Benito-Corchon, S., Bes-Rastrollo, M., 2014. Glycemic load, glycemic index, bread and incidence of overweight/obesity in a Mediterranean cohort: the SUN project. BMC Publ. Health 14, 1091. https://doi.org/10.1186/1471-2458-14-1091.

Drake, I., Sonestedt, E., Gullberg, B., Bjartell, A., Olsson, H., Adlercreutz, H., Tikkanen, M.J., Wirfalt, E., Wallstrom, P., 2014. Plasma alkylresorcinol metabolites as biomarkers for whole-grain intake and their association with prostate cancer: a Swedish nested case-control study. Cancer Epidemiol. Biomarkers Prev. 23 (1), 73–83. https://doi.org/10.1158/1055-9965.EPI-13-0878.

Egeberg, R., Olsen, A., Loft, S., Christensen, J., Johnsen, N.F., Overvad, K., Tjonneland, A., 2009. Intake of whole grain products and risk of breast cancer by hormone receptor status and histology among postmenopausal women. Int. J. Cancer 124 (3), 745–750. https://doi.org/10.1002/ijc.23992.

Egeberg, R., Olsen, A., Loft, S., Christensen, J., Johnsen, N.F., Overvad, K., Tjonneland, A., 2010. Intake of wholegrain products and risk of colorectal cancers in the Diet, Cancer and Health Cohort Study. Br. J. Cancer 103 (5), 730–734. https://doi.org/10.1038/sj.bjc.6605806.

Egeberg, R., Olsen, A., Christensen, J., Johnsen, N.F., Loft, S., Overvad, K., Tjonneland, A., 2011. Intake of whole-grain products and risk of prostate cancer among men in the Danish Diet, Cancer and Health cohort Study. Cancer Causes Contr. 22 (8), 1133–1139. https://doi.org/10.1007/s10552-011-9789-5.

Erkkilä, A.T., Herrington, D.M., Mozaffarian, D., Lichtenstein, A.H., 2005. Cereal fiber and whole-grain intake are associated with reduced progression of coronary-artery atherosclerosis in postmenopausal women with coronary artery disease. Am. Heart J. 150 (1), 94–101. https://doi.org/10.1016/j.ahj.2004.08.013.

Farvid, M.S., Cho, E., Eliassen, A.H., Chen, W.Y., Willett, W.C., 2016. Lifetime grain consumption and breast cancer risk. Breast Cancer Res. Treat. 159 (2), 335–345. https://doi.org/10.1007/s10549-016-3910-0.

Fox, C.S., Massaro, J.M., Hoffmann, U., Pou, K.M., Maurovich-Horvat, P., Liu, C.-Y., Vasan, R.S., Murabito, J.M., Meigs, J.B., Cupples, L.A., D'Agostino, R.B., O'Donnell, C.J., 2007. Abdominal visceral and subcutaneous adipose tissue compartments: association with metabolic risk factors in the Framingham Heart Study. Circulation 116 (1), 39–48. https://doi.org/10.1161/CIRCULATIONAHA.106.675355.

Gaesser, G.A., 2020. Whole grains, refined grains, and cancer risk: a systematic review of meta-analyses of observational studies. Nutrients 12 (12). https://doi.org/10.3390/nu12123756.

Gaskins, A.J., Mumford, S.L., Rovner, A.J., Zhang, C., Chen, L., Wactawski-Wende, J., Perkins, N.J., Schisterman, E.F., 2010. Whole grains are associated with serum concentrations of high sensitivity C-reactive protein among premenopausal women. J. Nutr. 140 (9), 1669–1676. https://doi.org/10.3945/jn.110.124164.

Hajihashemi, P., Haghighatdoost, F., 2019. Effects of whole-grain consumption on selected biomarkers of systematic inflammation: a systematic review and meta-analysis of randomized controlled trials. J. Am. Coll. Nutr. 38 (3), 275–285. https://doi.org/10.1080/07315724.2018.1490935.

He, X., Wu, K., Zhang, X., Nishihara, R., Cao, Y., Fuchs, C.S., Giovannucci, E.L., Ogino, S., Chan, A.T., Song, M., 2019. Dietary intake of fiber, whole grains and risk of colorectal cancer: an updated analysis according to food sources, tumor location and molecular subtypes in two large US cohorts. Int. J. Cancer 145 (11), 3040–3051. https://doi.org/10.1002/ijc.32382.

Hollænder, P.L.B., Ross, A.B., Kristensen, M., 2015. Whole-grain and blood lipid changes in apparently healthy adults: a systematic review and meta-analysis of randomized controlled studies. Am. J. Clin. Nutr. 102 (3), 556–572. https://doi.org/10.3945/ajcn.115.109165.

Hu, Y., Ding, M., Sampson, L., Willett, W.C., Manson, J.E., Wang, M., Rosner, B., Hu, F.B., Sun, Q., 2020. Intake of whole grain foods and risk of type 2 diabetes: results from three prospective cohort studies. BMJ 370. https://doi.org/10.1136/bmj.m2206.

Huang, T., Xu, M., Lee, A., Cho, S., Qi, L., 2015. Consumption of whole grains and cereal fiber and total and cause-specific mortality: prospective analysis of 367,442 individuals. BMC Med. 13, 59. https://doi.org/10.1186/s12916-015-0294-7.

Hullings, A.G., Sinha, R., Liao, L.M., Freedman, N.D., Graubard, B.I., Loftfield, E., 2020. Whole grain and dietary fiber intake and risk of colorectal cancer in the NIH-AARP Diet and Health Study Cohort. Am. J. Clin. Nutr. 112 (3), 603–612. https://doi.org/10.1093/ajcn/nqaa161.

Jacobs Jr., D.R., Marquart, L., Slavin, J., Kushi, L.H., 1998. Whole-grain intake and cancer: an expanded review and meta-analysis. Nutr. Cancer 30 (2), 85–96. https://doi.org/10.1080/01635589809514647.

Kasum, C.M., Nicodemus, K., Harnack, L.J., Jacobs Jr., D.R., Folsom, A.R., Iowa Women's Health Study, 2001. Whole grain intake and incident endometrial cancer: the Iowa Women's Health Study. Nutr. Cancer 39 (2), 180–186. https://doi.org/10.1207/S15327914nc392_4.

Keski-Rahkonen, P., Kolehmainen, M., Lappi, J., Micard, V., Jokkala, J., Rosa-Sibakov, N., Pihlajamäki, J., Kirjavainen, P.V., Mykkänen, H., Poutanen, K., Gunter, M.J., Scalbert, A., Hanhineva, K., 2019. Decreased plasma serotonin and other metabolite changes in healthy adults after consumption of wholegrain rye: an untargeted metabolomics study. Am. J. Clin. Nutr. 109 (6), 1630–1639. https://doi.org/10.1093/ajcn/nqy394.

Kikuchi, Y., Nozaki, S., Makita, M., Yokozuka, S., Fukudome, S., Yanagisawa, T., Aoe, S., 2018. Effects of whole grain wheat bread on visceral fat obesity in Japanese subjects: a randomized double-blind study. Plant Foods Hum. Nutr. 73 (3), 161–165. https://doi.org/10.1007/s11130-018-0666-1.

Kirwan, J.P., Malin, S.K., Scelsi, A.R., Kullman, E.L., Navaneethan, S.D., Pagadala, M.R., Haus, J.M., Filion, J., Godin, J.-P., Kochhar, S., Ross, A.B., 2016. A whole-grain diet reduces cardiovascular risk factors in overweight and obese adults: a randomized controlled trial. J. Nutr. 146 (11), 2244–2251. https://doi.org/10.3945/jn.116.230508.

La Vecchia, C., Decarli, A., Franceschi, S., Gentile, A., Negri, E., Parazzini, F., 1987. Dietary factors and the risk of breast cancer. Nutr. Cancer 10 (4), 205–214. https://doi.org/10.1080/01635588709513958.

Landberg, R., Hanhineva, K., Tuohy, K., Garcia-Aloy, M., Biskup, I., Llorach, R., Yin, X., Brennan, L., Kolehmainen, M., 2019. Biomarkers of cereal food intake. Genes Nutr. 14 (1), 28. https://doi.org/10.1186/s12263-019-0651-9.

Larsson, S.C., Giovannucci, E., Bergkvist, L., Wolk, A., 2005. Whole grain consumption and risk of colorectal cancer: a population-based cohort of 60,000 women. Br. J. Cancer 92 (9), 1803–1807. https://doi.org/10.1038/sj.bjc.6602543.

Lei, Q., Zheng, H., Bi, J., Wang, X., Jiang, T., Gao, X., Tian, F., Xu, M., Wu, C., Zhang, L., Li, N., Li, J., 2016. Whole grain intake reduces pancreatic cancer risk: a meta-analysis of observational studies. Medicine 95 (9), e2747. https://doi.org/10.1097/MD.0000000000002747.

Levi, F., La Vecchia, C., Gulie, C., Negri, E., 1993. Dietary factors and breast cancer risk in Vaud, Switzerland. Nutr. Cancer 19 (3), 327–335. https://doi.org/10.1080/01635589309514263.

Liu, S., Willett, W.C., Manson, J.E., Hu, F.B., Rosner, B., Colditz, G., 2003. Relation between changes in intakes of dietary fiber and grain products and changes in weight and development of obesity among middle-aged women. Am. J. Clin. Nutr. 78 (5), 920–927.

Maki, K.C., Palacios, O.M., Koecher, K., Sawicki, C.M., Livingston, K.A., Bell, M., Nelson Cortes, H., McKeown, N.M., 2019. The relationship between whole grain intake and body weight: results of meta-analyses of observational studies and randomized controlled trials. Nutrients 11 (6). https://doi.org/10.3390/nu11061245.

Marshall, S., Petocz, P., Duve, E., Abbott, K., Cassettari, T., Blumfield, M., Fayet-Moore, F., 2020. The effect of replacing refined grains with whole grains on cardiovascular risk factors: a systematic review and meta-analysis of randomized controlled trials with GRADE clinical recommendation. J. Acad. Nutr. Diet. 120 (11), 1859–1883.e31. https://doi.org/10.1016/j.jand.2020.06.021.

Marventano, S., Vetrani, C., Vitale, M., Godos, J., Riccardi, G., Grosso, G., 2017. Whole grain intake and glycaemic control in healthy subjects: a systematic review and meta-analysis of randomized controlled trials. Nutrients 9 (7), 769. https://doi.org/10.3390/nu9070769.

McCullough, M.L., Robertson, A.S., Jacobs, E.J., Chao, A., Calle, E.E., Thun, M.J., 2001. A prospective study of diet and stomach cancer mortality in United States men and women. Cancer Epidemiol. Biomarkers Prev. 10 (11), 1201–1205.

McKeown, N.M., Troy, L.M., Jacques, P.F., Hoffmann, U., O'Donnell, C.J., Fox, C.S., 2010. Whole- and refined-grain intakes are differentially associated with abdominal visceral and subcutaneous adiposity in healthy adults: the Framingham Heart Study. Am. J. Clin. Nutr. 92 (5), 1165–1171. https://doi.org/10.3945/ajcn.2009.29106.

McRorie, J.W., McKeown, N.M., 2017. Understanding the physics of functional fibers in the gastrointestinal tract: an evidence-based approach to resolving enduring misconceptions about insoluble and soluble fiber. J. Acad. Nutr. Diet. 117 (2), 251–264. https://doi.org/10.1016/j.jand.2016.09.021.

Miller, K.B., 2020. Review of whole grain and dietary fiber recommendations and intake levels in different countries. Nutr. Rev. 78 (Suppl. 1), 29–36. https://doi.org/10.1093/nutrit/nuz052.

Mourouti, N., Kontogianni, M.D., Papavagelis, C., Psaltopoulou, T., Kapetanstrataki, M.G., Plytzanopoulou, P., Vassilakou, T., Malamos, N., Linos, A., Panagiotakos, D.B., 2016. Whole grain consumption and breast cancer: a case-control study in women. J. Am. Coll. Nutr. 35 (2), 143–149. https://doi.org/10.1080/07315724.2014.963899.

Mozaffarian, D., Hao, T., Rimm, E.B., Willett, W.C., Hu, F.B., 2011. Changes in diet and lifestyle and long-term weight gain in women and men. N. Engl. J. Med. 364 (25), 2392–2404. https://doi.org/10.1056/NEJMoa1014296.

Müller, M.J., Lagerpusch, M., Enderle, J., Schautz, B., Heller, M., Bosy-Westphal, A., 2012. Beyond the body mass index: tracking body composition in the pathogenesis of obesity and the metabolic syndrome. Obes. Rev. 13 (Suppl. 2), 6–13. https://doi.org/10.1111/j.1467-789X.2012.01033.x.

Nicodemus, K.K., Jacobs Jr., D.R., Folsom, A.R., 2001. Whole and refined grain intake and risk of incident postmenopausal breast cancer (United States). Cancer Causes Contr. 12 (10), 917–925. https://doi.org/10.1023/a:1013746719385.

Nimptsch, K., Kenfield, S., Jensen, M.K., Stampfer, M.J., Franz, M., Sampson, L., Brand-Miller, J.C., Willett, W.C., Giovannucci, E., 2011. Dietary glycemic index, glycemic load, insulin index, fiber and whole-grain intake in relation to risk of prostate cancer. Cancer Causes Contr. 22 (1), 51–61. https://doi.org/10.1007/s10552-010-9671-x.

Pol, K., Christensen, R., Bartels, E.M., Raben, A., Tetens, I., Kristensen, M., 2013. Whole grain and body weight changes in apparently healthy adults: a systematic review and meta-analysis of randomized controlled studies. Am. J. Clin. Nutr. 98 (4), 872–884. https://doi.org/10.3945/ajcn.113.064659.

Quatela, A., Callister, R., Patterson, A.J., McEvoy, M., MacDonald-Wicks, L.K., 2017. Breakfast cereal consumption and obesity risk amongst the mid-age cohort of the Australian Longitudinal Study on Women's Health. Healthcare 5 (3), 49. https://doi.org/10.3390/healthcare5030049.

Reynolds, A., Mann, J., Cummings, J., Winter, N., Mete, E., Te Morenga, L., 2019. Carbohydrate quality and human health: a series of systematic reviews and meta-analyses. Lancet 393 (10170), 434–445. https://doi.org/10.1016/S0140-6736(18)31809-9.

Roager, H.M., Vogt, J.K., Kristensen, M., Hansen, L.B.S., Ibrügger, S., Mærkedahl, R.B., Bahl, M.I., Lind, M.V., Nielsen, R.L., Frøkiær, H., Gøbel, R.J., Landberg, R., Ross, A.B., Brix, S., Holck, J., Meyer, A.S., Sparholt, M.H., Christensen, A.F., Carvalho, V., et al., 2017. Whole grain-rich diet reduces body weight and systemic low-grade inflammation without inducing major changes of the gut microbiome: a randomised cross-over trial. Gut. https://doi.org/10.1136/gutjnl-2017-314786 gutjnl-2017-314786.

Ross, A.B., 2012. Present status and perspectives on the use of alkylresorcinols as biomarkers of wholegrain wheat and rye intake. J. Nutr. Metabol. 2012, 462967. https://doi.org/10.1155/2012/462967.

Sadeghi, O., Sadeghian, M., Rahmani, S., Maleki, V., Larijani, B., Esmaillzadeh, A., 2020. Whole-grain consumption does not affect obesity measures: an updated systematic review and meta-analysis of randomized clinical trials. Adv. Nutr. 11 (2), 280–292. https://doi.org/10.1093/advances/nmz076.

Sang, S., 2018. Biomarkers of whole grain intake. J. Agric. Food Chem. 66 (40), 10347–10352. https://doi.org/10.1021/acs.jafc.8b04110.

Sawicki, C.M., Jacques, P.F., Lichtenstein, A.H., Rogers, G.T., Ma, J., Saltzman, E., McKeown, N.M., 2021. Whole- and refined-grain consumption and longitudinal changes in cardiometabolic risk factors in the Framingham Offspring Cohort. J. Nutr. 151 (9), 2790–2799. https://doi.org/10.1093/jn/nxab177.

Schlesinger, S., Neuenschwander, M., Schwedhelm, C., Hoffmann, G., Bechthold, A., Boeing, H., Schwingshackl, L., 2019. Food groups and risk of overweight, obesity, and weight gain: a systematic review and dose-response meta-analysis of prospective studies. Adv. Nutr. 10 (2), 205–218. https://doi.org/10.1093/advances/nmy092.

Schwingshackl, L., Schwedhelm, C., Hoffmann, G., Knüppel, S., Iqbal, K., Andriolo, V., Bechthold, A., Schlesinger, S., Boeing, H., 2017. Food groups and risk of hypertension: a systematic review and dose-response meta-analysis of prospective studies. Adv. Nutr. 8 (6), 793–803. https://doi.org/10.3945/an.117.017178.

Schwingshackl, L., Schwedhelm, C., Hoffmann, G., Knuppel, S., Laure Preterre, A., Iqbal, K., Bechthold, A., De Henauw, S., Michels, N., Devleesschauwer, B., Boeing, H., Schlesinger, S., 2018. Food groups and risk of colorectal cancer. Int. J. Cancer 142 (9), 1748–1758. https://doi.org/10.1002/ijc.31198.

Shan, Z., Rehm, C.D., Rogers, G., Ruan, M., Wang, D.D., Hu, F.B., Mozaffarian, D., Zhang, F.F., Bhupathiraju, S.N., 2019. Trends in dietary carbohydrate, protein, and fat intake and diet quality among US adults, 1999–2016. JAMA 322 (12), 1178–1187. https://doi.org/10.1001/jama.2019.13771.

So, D., Whelan, K., Rossi, M., Morrison, M., Holtmann, G., Kelly, J.T., Shanahan, E.R., Staudacher, H.M., Campbell, K.L., 2018. Dietary fiber intervention on gut microbiota composition in healthy adults: a systematic review and meta-analysis. Am. J. Clin. Nutr. 107 (6), 965–983. https://doi.org/10.1093/ajcn/nqy041.

Sonestedt, E., Borgquist, S., Ericson, U., Gullberg, B., Landberg, G., Olsson, H., Wirfalt, E., 2008. Plant foods and oestrogen receptor alpha- and beta-defined breast cancer: observations from the Malmo Diet and Cancer Cohort. Carcinogenesis 29 (11), 2203–2209. https://doi.org/10.1093/carcin/bgn196.

Tajaddini, A., Pourzand, A., Sanaat, Z., Pirouzpanah, S., 2015. Dietary resistant starch contained foods and breast cancer risk: a case-control study in northwest of Iran. Asian Pac. J. Cancer Prev. 16 (10), 4185–4192. https://doi.org/10.7314/apjcp.2015.16.10.4185.

Tieri, M., Ghelfi, F., Vitale, M., Vetrani, C., Marventano, S., Lafranconi, A., Godos, J., Titta, L., Gambera, A., Alonzo, E., Sciacca, S., Riccardi, G., Buscemi, S., Del Rio, D., Ray, S., Galvano, F., Beck, E., Grosso, G., 2020. Whole grain consumption and human health: an umbrella review of observational studies. Int. J. Food Sci. Nutr. 71 (6), 668–677. https://doi.org/10.1080/09637486.2020.1715354.

Tighe, P., Duthie, G., Vaughan, N., Brittenden, J., Simpson, W.G., Duthie, S., Mutch, W., Wahle, K., Horgan, G., Thies, F., 2010. Effect of increased consumption of whole-grain foods on blood pressure and other cardiovascular risk markers in healthy middle-aged persons: a randomized controlled trial. Am. J. Clin. Nutr. 92 (4), 733–740. https://doi.org/10.3945/ajcn.2010.29417.

Vieira, A.R., Abar, L., Chan, D.S.M., Vingeliene, S., Polemiti, E., Stevens, C., Greenwood, D., Norat, T., 2017. Foods and beverages and colorectal cancer risk: a systematic review and meta-analysis of cohort studies, an update of the evidence of the WCRF-AICR Continuous Update Project. Ann. Oncol. 28 (8), 1788–1802. https://doi.org/10.1093/annonc/mdx171.

Wang, R.J., Tang, J.E., Chen, Y., Gao, J.G., 2015. Dietary fiber, whole grains, carbohydrate, glycemic index, and glycemic load in relation to risk of prostate cancer. Onco. Targets Ther. 8, 2415–2426. https://doi.org/10.2147/OTT.S88528.

Wang, T., Zhan, R., Lu, J., Zhong, L., Peng, X., Wang, M., Tang, S., 2020. Grain consumption and risk of gastric cancer: a meta-analysis. Int. J. Food Sci. Nutr. 71 (2), 164–175. https://doi.org/10.1080/09637486.2019.1631264.

Wei, H., Gao, Z., Liang, R., Li, Z., Hao, H., Liu, X., 2016. Whole-grain consumption and the risk of all-cause, CVD and cancer mortality: a meta-analysis of prospective cohort studies. Br. J. Nutr. 116 (3), 514–525. https://doi.org/10.1017/S0007114516001975.

Xiao, Y., Ke, Y., Wu, S., Huang, S., Li, S., Lv, Z., Yeoh, E.K., Lao, X., Wong, S., Kim, J.H., Colditz, G.A., Tamimi, R.M., Su, X., 2018. Association between whole grain intake and breast cancer risk: a systematic review and meta-analysis of observational studies. Nutr. J. 17 (1), 87. https://doi.org/10.1186/s12937-018-0394-2.

Xu, Y., Wan, Q., Feng, J., Du, L., Li, K., Zhou, Y., 2018. Whole grain diet reduces systemic inflammation. Medicine 97 (43), e12995. https://doi.org/10.1097/MD.0000000000012995.

Xu, Y., Yang, J., Du, L., Li, K., Zhou, Y., 2019. Association of whole grain, refined grain, and cereal consumption with gastric cancer risk: a meta-analysis of observational studies. Food Sci. Nutr. 7 (1), 256–265. https://doi.org/10.1002/fsn3.878.

Ye, E.Q., Chacko, S.A., Chou, E.L., Kugizaki, M., Liu, S., 2012. Greater whole-grain intake is associated with lower risk of type 2 diabetes, cardiovascular disease, and weight gain. J. Nutr. 142 (7), 1304–1313. https://doi.org/10.3945/jn.111.155325.

Yun, S.H., Kim, K., Nam, S.J., Kong, G., Kim, M.K., 2010. The association of carbohydrate intake, glycemic load, glycemic index, and selected rice foods with breast cancer risk: a case-control study in South Korea. Asia Pac. J. Clin. Nutr. 19 (3), 383–392.

Zhang, B., Zhao, Q., Guo, W., Bao, W., Wang, X., 2018. Association of whole grain intake with all-cause, cardiovascular, and cancer mortality: a systematic review and dose-response meta-analysis from prospective cohort studies. Eur. J. Clin. Nutr. 72 (1), 57–65. https://doi.org/10.1038/ejcn.2017.149.

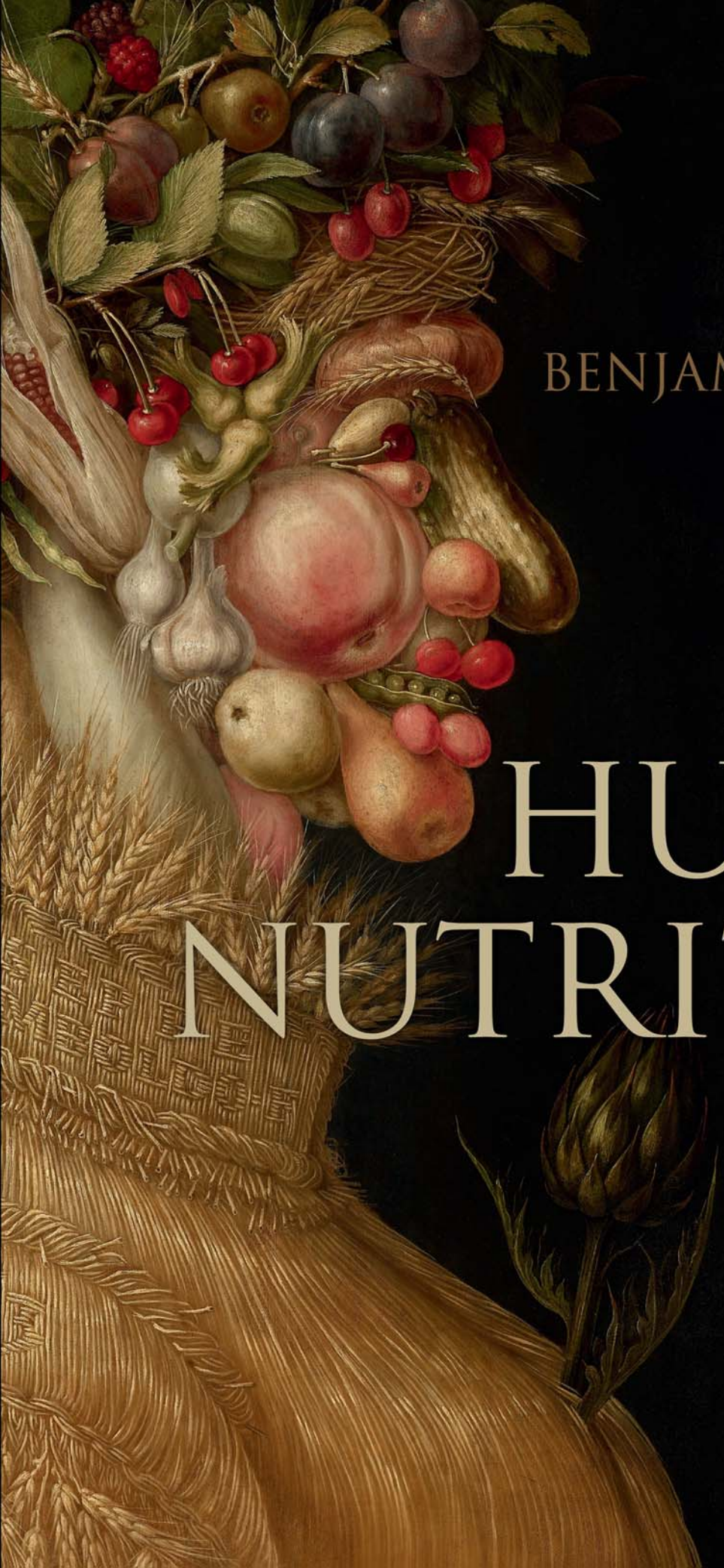
Zhang, X.F., Wang, X.K., Tang, Y.J., Guan, X.X., Guo, Y., Fan, J.M., Cui, L.L., 2020. Association of whole grains intake and the risk of digestive tract cancer: a systematic review and meta-analysis. Nutr. J. 19 (1), 52. https://doi.org/10.1186/s12937-020-00556-6.

Zong, G., Gao, A., Hu, F.B., Sun, Q., 2016. Whole grain intake and mortality from all causes, cardiovascular disease, and cancer: a meta-analysis of prospective cohort studies. Circulation 133 (24), 2370–2380. https://doi.org/10.1161/CIRCULATIONAHA.115.021101.

Relevant websites

The Whole Grains Council, http://www.wholegrainscouncil.org/.
The European Healthgrain Project, http://www.healthgrain.eu/pub/.
The Whole Grain Initiative, https://www.wholegraininitiative.org/.
UK National Health Service Live Well, http://www.nhs.uk/livewell/healthy-eating/Pages/Healthyeating.aspx.
USDA Dietary Guidelines for Americans, https://www.dietaryguidelines.gov/.
WCRF/AIRC Continuous Update Project, https://www.wcrf.org/diet-activity-and-cancer/risk-factors/wholegrains-vegetables-fruit-and-cancer-risk/

This page intentionally left blank



EDITOR IN CHIEF

BENJAMIN CABALLERO

VOLUME THREE

Encyclopedia of HUMAN NUTRITION

FOURTH EDITION

ENCYCLOPEDIA OF HUMAN NUTRITION

FOURTH EDITION

This page intentionally left blank

ENCYCLOPEDIA OF HUMAN NUTRITION

FOURTH EDITION

EDITOR IN CHIEF

Professor Benjamin Caballero,

Emeritus Professor, Bloomberg School of Public Health and School of Medicine,
Johns Hopkins University
Baltimore, MD, United States

VOLUME 3

Section 4: Disorders Directly Related to Inadequate Nutrient Intake

Edited by Katherine L. Tucker

Department of Biomedical and Nutritional Sciences, Center for Population Health,
University of Massachusetts Lowell, Lowell, MA, United States

Section 5: Nutrition in Disease States

Edited by Paolo M. Suter

Department of Endocrinology, Diabetology and Clinical Nutrition,
University Hospital, Zurich, Switzerland

ELSEVIER

AMSTERDAM ● BOSTON ● HEIDELBERG ● LONDON ● NEW YORK ● OXFORD
PARIS ● SAN DIEGO ● SAN FRANCISCO ● SINGAPORE ● SYDNEY ● TOKYO
Academic Press is an imprint of Elsevier

ACADEMIC PRESS

Academic Press is an imprint of Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge MA 02139, United States

Copyright © 2023 Elsevier Ltd. All rights reserved

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers may always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN 978-0-12-821848-8

For information on all publications visit our website at http://store.elsevier.com



Publisher: Oliver Walter
Acquisitions Editors: Emma Hayes and Michael Nicholls
Content Project Managers: Richard Berryman and Sindhuja Vadlamani
Associate Content Project Manager: N. Kiruthigadevi
Designer: Matthew Limbert

CONTENTS OF VOLUME 3

Contributors to Volume 3 *ix*

Editor Biographies *xv*

Preface *xix*

DISORDERS DIRECTLY RELATED TO INADEQUATE NUTRIENT INTAKE

Biochemical indices 1
CM Pfeiffer, RL Schleicher, and KL Caldwell

Body composition 20
M Ashby-Thompson, S Chung, and D Gallagher

Clinical examination 35
B Caballero

Dietary intake measurement: Methodology 39
AA Welch and AA Mulligan

Doubly labeled water 52
Michele N Ravelli, Timothy C Shriver, and Dale A Schoeller

Food choice: Behavioral aspects 62
Barbara J Rolls, Kathleen L Keller, John E Hayes, and Travis D Masterson

Glucose: Metabolism and homeostasis 70
Julie Miller Jones

Nutritional aspects of bone 79
AR Bournot, SA Lanham-New, M Alghamdi, and J Jalal

Nutritional assessment: Anthropometry 87
J Eaton-Evans

NUTRITION IN DISEASE STATES

Adipose tissue: Structure, function and metabolism 94
G Frühbeck and J Gómez-Ambrosi

Arthritis 109
F Cuesta Triana

Brain and nervous system: Biology, metabolism, and nutritional requirements 118
JD Fernstrom and MH Fenstrom

Cancer: Carcinogenic substances in food 125
AI Prieto, AM Troncoso, and AM Cameán

Cancer: Dietary management 136
Victoria Nelson and Louise Henry

Cancer: Epidemiology and associations between diet and cancer 146
Graham A Colditz and Hank Dart

Cancer: Epidemiology of lung cancer 154
Anthony J Alberg, Rajat Das Gupta, and Maxwell Akonde

Celiac disease 162
Balakrishnan S Ramakrishna

Cerebral palsy: Nutritional aspects 171
Patricia Miller

Colon: Structure, function, and disorders 184
Asim Maqbool

Cystic fibrosis 207
J Dowsett and O Tully

Dehydration 216
Shashank S Dhareshwar

Diabetes mellitus: Diagnosis and heterogeneity 227
Ram Jagannathan, Kosuke Tamura, and Priyathama Vellanki

Diabetes mellitus: Dietary management 234
M Lu, SJ Pilla, and SH Oh

Diabetes mellitus: Etiology and epidemiology 252
Yara S Beyh and KM Venkat Narayan

Diarrheal diseases 263
Phoebe Hodges and Paul Kelly

Diet and oral health 272
Anja Heilmann, Yanela Aravena Rivas, and Richard G Watt

Down's syndrome: Nutritional aspects 283
A Laverty

Eating disorders: Anorexia nervosa 290
Meredith Kells and Jennifer E Wildes

Eating disorders: Binge eating 296
MA Kalarchian and MD Marcus

Eating disorders: Bulimia nervosa 304
Heather A Davis and Jennifer E Wildes

Energy balance 314
Yves Schutz

Enteral and parenteral nutrition 324
Kent Aje, Sara Yacyshyn, Lindsay Dowhan, and Lisa Reitz

Food allergies: Diagnosis and management 337
TJ David

Food intolerance 345
B Caballero

Hyperactivity: Nutritional aspects 353
Klaus W Lange

Hyperlipidemia 361
Penny M Kris-Etherton, Lisa Sanders, Olivia Lawler, Terrence Riley, and Kevin Maki

Hyperlipidemia: Nutritional prevention and management 380
AH Lichtenstein

Hypertension: Dietary factors 391
LJ Appel

Hypoglycemia 401
RA Sulaiman, R Gama, and V Marks

Inborn errors of metabolism: Classification and biochemical aspects 412
DL Marsden

Inborn errors of metabolism: Nutrition management of phenylketonuria 424
DL Marsden, B Reineking, and FJ Rohr

Ketosis 432
DH Williamson

Lactose intolerance 441
DM Paige

Liver disorders: Nutritional management 449
Jenifer Thompson and Wikrom Karnsakul

Lung diseases 463
SA Unger

Malabsorption syndromes: Nutritional management 477
PM Tsai and C Duggan

Malnutrition 484
Joy Ngo and Lluis Serra-Majem

Malnutrition: Secondary, diagnosis and management 507
LH Allen

Nutrition and asthma 513
Charles B Stephensen and Nicholas J Kenyon

Nutrition and burn injury – manuscript 522
Marc G Jeschke, Sophie Brosset, and Anna Van Osch

Nutrition and HIV/AIDS 535
S Filteau and D Manno

Nutrition and parasitism 541
Olga P García, Gerardo A Zavala, Maiza Campos-Ponce, and Katja Polman

Nutrition and susceptibility to tuberculosis 549
Pranay Sinha and J Peter Cegielski

Nutritional support: Adults, enteral 556
Berkeley Limketkai, Octavia Pickett Blakely, and Gerard E Mullin

Nutritional support: In the home setting 564
M Elia and TR Smith

Nutritional support: Infants and children, parenteral 574
S Collier and C Lo

Obesity: Childhood obesity 580
G Bueno-Lozano and LA Moreno Aznar

Obesity: Complications 590
RL Atkinson

Obesity: Definition, etiology, and assessment 599
S Hawkesworth

Obesity: Prevention 604
TP Gill

Osteoporosis nutritional factors 612
Alexa Barad and Kimberly O O'Brien

Pediatric feeding disorders: Feeding children who can't or won't eat 622
Sarah E Fleet and Ryan D Davidson

Sarcopenia: Molecular mechanism and current nutritional approach 633
Kunihiro Sakuma, Akihiko Yamaguchi, and Muneshige Shimizu

Starvation and fasting: Biochemical aspects 645
Kathleen H McGrath, Wolfram Haller, and Julie E Bines

Stroke nutritional management 657
L Perry

The role of nutrition in inflammatory bowel disease: Disease associations, management of active disease and maintenance of remission 669
Gabrielle Wark, Dorit Samocha-Bonet, Simon Ghaly, and Mark Danta

Tuberculosis: Nutritional management 698
Pranay Sinha and J Peter Cegielski

Water security and nutrition 706
Joshua D Miller and Sera L Young

Weight management: Weight cycling 717
L Lissner and BL Heitmann

Weight management: Weight maintenance 724
Lauren A Griffiths, Anna M Jackson, Elizabeth T Anderson Steeves, and Hollie A Raynor

CONTRIBUTORS TO VOLUME 3

Kent Aje
The Cleveland Clinic Foundation, Cleveland, OH, United States

Maxwell Akonde
Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

Anthony J Alberg
Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

M Alghamdi
Joint Supervision Programme, King Abdulaziz University, Jeddah, Saudi Arabia

LH Allen
University of California, Davis, CA, USA

Elizabeth T Anderson Steeves
Department of Nutrition, University of Tennessee Knoxville, Knoxville, TN, United States

LJ Appel
Johns Hopkins University, Baltimore, MD, United States

Yanela Aravena Rivas
University College London, London, United Kingdom

M Ashby-Thompson
Columbia University, New York, NY, United States

RL Atkinson
Virginia Commonwealth University, VA, USA

Alexa Barad
Cornell University, Division of Nutritional Sciences, Ithaca, NY, United States

Yara S Beyh
Laney Graduate School, Rollins School of Public Health, Department of Global Health, Emory University, Atlanta, GA, United States

Julie E Bines
Department of Gastroenterology and Clinical Nutrition, The Royal Children's Hospital, Melbourne, VIC, Australia; Department of Pediatrics, The University of Melbourne, Melbourne, VIC, Australia; and Murdoch Children's Research Institute, Melbourne, VIC, Australia

Octavia Pickett Blakely
Division of Gastroenterology, University of Pennsylvania Perlman School of Medicine, Philadelphia, PA, United States

AR Bournot
University of Surrey, Guildford, UK

Sophie Brosset
Sunnybrook Research Institute, Toronto, ON, Canada; and Claude Bernard Lyon 1 University, Lyon, France

G Bueno-Lozano
Department of Pediatrics and Pediatric Endocrinology, University of Zaragoza, Hospital Clínico Universitario Lozano Blesa, Zaragoza, Spain; Instituto de Investigación Sanitaria de Aragón (IIS Aragón), Zaragoza, Spain; GENUD (Growth, Exercise, NUtrition and Development) Research Group, Zaragoza, Spain; and CIBER Fisiopatología Obesidad y Nutrición, Instituto de Salud Carlos III, Madrid, Spain

B Caballero
Johns Hopkins University, Baltimore, MD, USA

KL Caldwell
Centers for Disease Control and Prevention, Atlanta, GA, USA

AM Cameán
Departamento de Nutrición y Bromatología, Toxicología y Medicina Legal, Facultad de Farmacia, Universidad de Sevilla, Sevilla, Spain

Maiza Campos-Ponce
Faculty of Earth and Life Sciences, VU Amsterdam University, Amsterdam, the Netherlands

J Peter Cegielski
Emory University Rollins School of Public Health, Atlanta, GA, United States

S Chung
Department of Pediatrics, Konkuk University Medical Center, Konkuk University School of Medicine, Seoul, South Korea

Graham A Colditz
Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, MO, United States

S Collier
Harvard School of Public Health, Boston, MA, USA

Mark Danta
St Vincent's Clinical School, UNSW, Sydney, NSW, Australia; and Department of Gastroenterology and Hepatology, St Vincent's Hospital, Sydney, NSW, Australia

Hank Dart
Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, MO, United States

TJ David
University of Manchester, Manchester, UK

Ryan D Davidson
Harvard Medical School, Boston, MA, United States; and Boston Children's Hospital, Boston, MA, United States

Heather A Davis
Department of Psychiatry & Behavioral Neuroscience, University of Chicago, Chicago, IL, United States

Shashank S Dhareshwar
Department of Nephrology, St John's Medical College and Hospital, Koramangala, Bengaluru, India

Lindsay Dowhan
The Cleveland Clinic Foundation, Cleveland, OH, United States

J Dowsett
St Vincent's University Hospital, Dublin, Ireland

C Duggan
Harvard Medical School, Boston, MA, USA

J Eaton-Evans
University of Ulster, Coleraine, UK

M Elia
University of Southampton, Southampton, UK

MH Fenstrom
University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

JD Fernstrom
University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

S Filteau
London School of Hygiene and Tropical Medicine, London, UK

Sarah E Fleet
Growth and Nutrition Program, Division of Gastroenterology, Hepatology & Nutrition, Boston Children's Hospital, Boston, MA, United States; and Harvard Medical School, Boston, MA, United States

G Frühbeck
Clínica Universidad de Navarra, University of Navarra, CIBEROBN, Instituto de Salud Carlos III, IdiSNA, Pamplona, Spain

D Gallagher
Columbia University, New York, NY, United States

R Gama
School of Medicine and Clinical Practice, University of Wolverhampton, Black Country Pathology Services, Royal Wolverhampton NHS Trust, Clinical Chemistry, Black Country Pathology Services, Wolverhampton, United Kingdom

Olga P García
School of Natural Sciences, Universidad Autónoma de Querétaro, Querétaro, Mexico

Simon Ghaly
St Vincent's Clinical School, UNSW, Sydney, NSW, Australia; and Department of Gastroenterology and Hepatology, St Vincent's Hospital, Sydney, NSW, Australia

TP Gill
University of Sydney, NSW, Australia

J Gómez-Ambrosi
Clínica Universidad de Navarra, University of Navarra, CIBEROBN, Instituto de Salud Carlos III, IdiSNA, Pamplona, Spain

Lauren A Griffiths
Department of Nutrition, University of Tennessee Knoxville, Knoxville, TN, United States

Rajat Das Gupta
Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

Wolfram Haller
Department of Gastroenterology, Birmingham Women's and Children's Hospital, Birmingham, United Kingdom

S Hawkesworth
London School of Hygiene and Tropical Medicine, London, UK

John E Hayes
Department of Food Science, The Pennsylvania State University, University Park, PA, United States

Anja Heilmann
University College London, London, United Kingdom

BL Heitmann
Institute of Preventive Medicine, Frederiksberg, Denmark

Louise Henry
The Royal Marsden NHS Foundation Trust, London, United Kingdom

Phoebe Hodges
Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

Anna M Jackson
Department of Nutrition, University of Tennessee Knoxville, Knoxville, TN, United States

Ram Jagannathan
Division of Hospital Medicine, Emory University School of Medicine, Atlanta, GA, United States

J Jalal
Joint Supervision Programme, King Abdulaziz University, Jeddah, Saudi Arabia

Marc G Jeschke
Sunnybrook Research Institute, Toronto, ON, Canada; Department of Surgery, Division of Plastic Surgery, University of Toronto, Toronto, ON, Canada; Department of Immunology, University of Toronto, Toronto, ON, Canada; Ross Tilley Burn Centre, Sunnybrook Health Sciences Centre, Toronto, ON, Canada; and Hamilton Health Sciences, Deaprtment of Surgery Mc Master University, Hamilton, ON, Canada

Julie Miller Jones
St. Catherine University, St. Paul, MN, United States

MA Kalarchian
School of Nursing, Associate Provost for Interdisciplinary Research and Strategic Initiatives, Duquesne University, Pittsburgh, PA, United States

Wikrom Karnsakul
Department of Pediatric Gastroenterology, Hepatology and Nutrition, Johns Hopkins University School of Medicine, Baltimore, MD, United States

Kathleen L Keller
Department of Nutritional Sciences, The Pennsylvania State University, University Park, PA, United States; and Department of Food Science, The Pennsylvania State University, University Park, PA, United States

Meredith Kells
Department of Psychiatry & Behavioral Neuroscience, University of Chicago, Chicago, IL, United States

Paul Kelly
Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom; and Tropical Gastroenterology and Nutrition Group, University of Zambia School of Medicine, Lusaka, Zambia

Nicholas J Kenyon
Division of Pulmonary, Critical Care, Sleep Medicine, Department of Medicine, School of Medicine, University of California, Davis, Davis, CA, United States; and Veteran's Administration Northern California Health Care System, Mather, CA, United States

Penny M Kris-Etherton
Department of Nutritional Sciences, Penn State University, University Park, PA, United States

Klaus W Lange
Institute of Psychology, University of Regensburg, Regensburg, Germany

SA Lanham-New
University of Surrey, Guildford, UK

A Laverty
Learning Disability Team, Coleraine, Co. Londonderry, UK

Olivia Lawler
Department of Nutritional Sciences, Penn State University, University Park, PA, United States

AH Lichtenstein
Tufts University, Boston, MA, United States

Berkeley Limketkai
Division of Digestive Diseases, UCLA School of Medicine, Los Angeles, CA, United States

L Lissner
University of Gothenburg, Göteborg, Sweden

C Lo
Harvard Medical School, Boston, MA, USA

M Lu
The Johns Hopkins University School of Medicine, Institute for Clinical and Translational Research, Johns Hopkins Hospital Bloomberg Children's Center, Baltimore, MD, United States

Kevin Maki
Midwest Biomedical Research: Center for Metabolic and Cardiovascular Health, Addison, IL, United States; MB Clinical Research, Boca Raton, FL, United States; and Department of Applied Health Science, School of Public Health, Indiana University, Bloomington, IN, United States

D Manno
London School of Hygiene and Tropical Medicine, London, UK

Asim Maqbool
Program for Pancreatic Disorders, Division of Gastroenterology, Hepatology and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, PA, United States; and The University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, United States

MD Marcus
Departments of Psychiatry and Psychology, University of Pittsburgh School of Medicine, Western Psychiatric Hospital, Pittsburgh, PA, United States

V Marks
Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom

DL Marsden
Boston Children's Hospital, Boston, MA, United States

Travis D Masterson
Department of Nutritional Sciences, The Pennsylvania State University, University Park, PA, United States

Kathleen H McGrath
Department of Gastroenterology and Clinical Nutrition, The Royal Children's Hospital, Melbourne, VIC, Australia; and Department of Pediatrics, The University of Melbourne, Melbourne, VIC, Australia

Joshua D Miller
Department of Nutrition, University of North Carolina, Chapel Hill, NC, United States

Patricia Miller
Kennedy Krieger Institute, Baltimore, MD, United States

LA Moreno Aznar
Faculty of Health Sciences, Instituto Agroalimentario de Aragón (IA2), University of Zaragoza, Zaragoza, Spain; Instituto de Investigación Sanitaria de Aragón (IIS Aragón), Zaragoza, Spain; GENUD (Growth, Exercise, NUtrition and Development) Research Group, Zaragoza, Spain; and CIBER Fisiopatología Obesidad y Nutrición, Instituto de Salud Carlos III, Madrid, Spain

AA Mulligan
NIHR BRC Diet, Anthropometry and Physical Activity Group, MRC Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom

Gerard E Mullin
Division of Gastroenterology & Hepatology, Johns Hopkins University School of Medicine, Baltimore, MD, United States

KM Venkat Narayan
Laney Graduate School, Rollins School of Public Health, Department of Global Health, Emory University, Atlanta, GA, United States; and Emory University School of Medicine, Emory University, Atlanta, GA, United States

Victoria Nelson
The Royal Marsden NHS Foundation Trust, London, United Kingdom

Joy Ngo
NGO Nutrition Without Borders, Barcelona, Spain; and Nutrition Research Foundation (FIN), University of Barcelona Science Park, Barcelona, Spain

SH Oh
Research Nutrition Core, The Johns Hopkins University School of Medicine, Institute for Clinical and Translational Research, Johns Hopkins Hospital, Baltimore, MD, United States

Kimberly O O'Brien
Cornell University, Division of Nutritional Sciences, Ithaca, NY, United States

DM Paige
Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

L Perry
University of Technology Sydney, NSW, Australia

CM Pfeiffer
Centers for Disease Control and Prevention, Atlanta, GA, USA

SJ Pilla
School of Medicine, The Johns Hopkins University School of Medicine, Department of Medicine, Division of General Internal Medicine, Baltimore, MD, United States; and Welch Center for Prevention, Epidemiology & Clinical Research, Baltimore, MD, United States

Katja Polman
Department of Biomedical Sciences, Institute of Tropical Medicine Antwerp, Antwerp, Belgium

AI Prieto
Departamento de Nutrición y Bromatología, Toxicología y Medicina Legal, Facultad de Farmacia, Universidad de Sevilla, Sevilla, Spain

Balakrishnan S Ramakrishna
Institute of Gastroenterology, SRM Institutes for Medical Science, Chennai, Tamil Nadu, India

Michele N Ravelli
Department of Neurology, University of Wisconsin, Madison, WI, United States; and Biotechnology Center, University of Wisconsin, Madison, WI, United States

Hollie A Raynor
Department of Nutrition, University of Tennessee Knoxville, Knoxville, TN, United States

B Reineking
Ultragenyx Pharmaceutical Inc., Novato, CA, United States

Lisa Reitz
The Cleveland Clinic Foundation, Cleveland, OH, United States

Terrence Riley
Department of Nutritional Sciences, Penn State University, University Park, PA, United States

FJ Rohr
Met Ed Co, Boulder, CO, United States

Barbara J Rolls
Department of Nutritional Sciences, The Pennsylvania State University, University Park, PA, United States

Kunihiro Sakuma
Institute for Liberal Arts, Environment and Society, Tokyo Institute of Technology, Meguro-ku, Tokyo, Japan

Dorit Samocha-Bonet
St Vincent's Clinical School, UNSW, Sydney, NSW, Australia; and Garvan Institute of Medical Research, Sydney, NSW, Australia

Lisa Sanders
Midwest Biomedical Research: Center for Metabolic and Cardiovascular Health, Addison, IL, United States

RL Schleicher
Centers for Disease Control and Prevention, Atlanta, GA, USA

Dale A Schoeller
Biotechnology Center, University of Wisconsin, Madison, WI, United States; and Nutrition Science, University of Wisconsin, Madison, WI, United States

Yves Schutz
Department of Medicine, Physiology Unit, University of Fribourg, Fribourg, Switzerland

Lluis Serra-Majem
NGO Nutrition Without Borders, Barcelona, Spain; Institute of Biomedical and Health Sciences (IUIBS), University of Las Palmas de Gran Canaria (ULPGC), Las Palmas de Gran Canaria, Spain; and CIBER Fisiopatología de la Obesidad y Nutrición (CIBER OBN), Instituto de Salud Carlos III, Madrid, Spain

Muneshige Shimizu
Tokai University, Faculty of Marine Science, Department of Fisheries, Shimizu-ku, Shizuoka, Tokyo, Japan

Timothy C Shriver
Biotechnology Center, University of Wisconsin, Madison, WI, United States

Pranay Sinha
Boston University School of Medicine, Harvard University School of Public Health, Boston, MA, United States

TR Smith
University of Southampton, Southampton, UK

Charles B Stephensen
USDA Western Human Nutrition Research Center, Davis, CA, United States; and Department of Nutrition, University of California, Davis, CA, United States

RA Sulaiman
Department of Medical Genomics, Center for Genomic Medicine, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; and Alfaisal University, Riyadh, Saudi Arabia

Kosuke Tamura
Neighborhood Social & Geospatial Determinants of Health Disparities Laboratory, Population and Community Sciences Branch, Division of Intramural Research, National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD, United States

Jenifer Thompson
Department of Pediatric Gastroenterology, Hepatology and Nutrition, Johns Hopkins Hospital, Baltimore, MD, United States

F Cuesta Triana
Hospital Clinico San Carlos, Universidad Complutense, Madrid, Spain

AM Troncoso
Departamento de Nutrición y Bromatología, Toxicología y Medicina Legal, Facultad de Farmacia, Universidad de Sevilla, Sevilla, Spain

PM Tsai
University of California, San Francisco, CA, USA

O Tully
St Vincent's University Hospital, Dublin, Ireland

SA Unger
MRC International Nutrition Group, Medical Research Council Laboratories, The Gambia

Anna Van Osch
Ross Tilley Burn Centre, Sunnybrook Health Sciences Centre, Toronto, ON, Canada

Priyathama Vellanki
Division of Endocrinology, Metabolism and Lipids, Emory University School of Medicine, Atlanta, GA, United States

Gabrielle Wark
St Vincent's Clinical School, UNSW, Sydney, NSW, Australia; and Department of Gastroenterology and Hepatology, St Vincent's Hospital, Sydney, NSW, Australia

Richard G Watt
University College London, London, United Kingdom

AA Welch
Department of Epidemiology & Public Health, Norwich Medical School, University of East Anglia, Norwich, Norfolk, United Kingdom

Jennifer E Wildes
Department of Psychiatry & Behavioral Neuroscience, University of Chicago, Chicago, IL, United States

DH Williamson
Radcliffe Infirmary, Oxford, UK

Sara Yacyshyn
The Cleveland Clinic Foundation, Cleveland, OH, United States

Akihiko Yamaguchi
Department of Physical Therapy, Health Sciences University of Hokkaido, Kanazawa, Ishikari-Tobetsu, Hokkaido, Japan

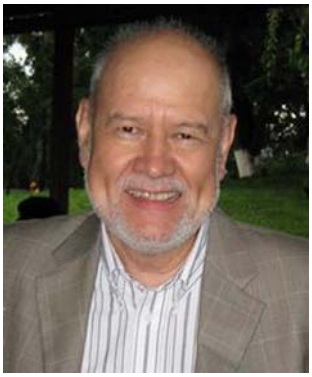
Sera L Young
Department of Anthropology, Northwestern University, Evanston, IL, United States; and Institute for Policy Research, Northwestern University, Evanston, IL, United States

Gerardo A Zavala
Department of Health Sciences, University of York, Heslington, United Kingdom

EDITOR BIOGRAPHIES

Editor in Chief

Professor Benjamin Caballero
Emeritus Professor, Bloomberg School of Public Health and School of Medicine, Johns Hopkins University, Baltimore, MD, United States.



Dr. Caballero is Professor Emeritus at the Department of International Health, Bloomberg School of Public Health, with joint appointment at the Department of Pediatrics, School of Medicine, Johns Hopkins University. He obtained his MD from the University of Buenos Aires, Argentina, and his PhD (in neuroendocrine regulation) from MIT, in Cambridge, Massachusetts. He started his academic career at Boston Children's Hospital, Harvard Medical School, and subsequently became the Founding Director of the Center for Human Nutrition at Johns Hopkins University.

Dr. Caballero has focused his research on child nutrition and health in developing countries. In particular, he has explored the combination of undernutrition and overweight that has become increasingly prevalent in low- and middle-income countries.

He is currently a member of the Council of the International Union of Nutritional Sciences. He has served on the Food and Nutrition Board of the US National Academy of Medicine and on a number of expert panels, including the Dietary Reference Intakes Committee, the Expert Panel on Macronutrient Requirements, and the Childhood Obesity Task Force. He was also a member of the U.S. Dietary Guidelines for Americans Advisory Committee, of the Scientific Advisory Board of the Food and Drug Administration, and of advisory committees of the National Institutes of Health and the Department of Agriculture. He is a Fellow of the American Society for Nutrition and of the Royal Society of Medicine (UK), and a member of the Spanish Academy of Nutritional Sciences.

He is the Editor-in-Chief of the *Encyclopedia of Food Sciences and Nutrition*, a 10-volume work on food production, consumption, and biological effects. He is also Editor-in-Chief of the *Encyclopedia of Human Nutrition*, which received the Book of the Year Award from the British Medical Association. His *Guide to Dietary Supplements* summarizes the current scientific basis for the use of mineral and vitamin supplements. He also co-edited a widely used textbook on human nutrition, *Modern Nutrition in Health and Disease*.

Section Editors

Section 1: *The Foundations of Human Nutrition*

Professor Angel Gil

Honorary President of the Iberoamerican Nutrition Foundation (FINUT), Emeritus Professor, Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology, Centre of Biomedical Research, University of Granada, Campus de la Salud, Avda. del Conocimiento, 18100 Armilla, Granada, Spain.



Angel Gil is an Emeritus Professor at the University of Granada and a Full Professor in the Department of Biochemistry and Molecular Biology. He is Doctor Honoris Causa at the Autonomous University of Nuevo León, Monterrey, México, and a visitant Professor of the University of Chile, Santiago de Chile, and University of San Miguel de Tucumán, Argentina. He got a major degree in Biology (1973) and PhD in Biochemistry and Molecular Biology (1978).

Prof. Gil is an internationally recognized authority in Food and Nutrition: His expertise extends from the study of human milk composition to the molecular effects of food bioactive compounds and probiotics and the design and development of novel products for infant and clinical nutrition. He conducted pioneering and innovative research, leading 7 international, 27 national, and more than 50 projects and 120 contracts; he has taught since 1981, supervising more than 50 PhD students.

Prof. Gil has several areas of interest that include evaluating the role of dietary nucleotides in early life and the development of infant nutrition products. Besides, the isolation, identification, and description of the mechanism of action of probiotics and the metabolic, molecular, and genetic factors involved in obesity and the early onset of metabolic syndrome (MS) in childhood; and the design, development, and evaluation of enteral clinical nutrition products. What describes Prof. Gil best is the variety of fields and problems he has faced during his professional carrier and his significant ability to combine his knowledge and expertise in Food Science and Human Biochemistry. This has allowed him to design, develop, innovate, and evaluate exclusive products for Human Nutrition, which are demonstrated in his published articles and his patents' impact.

The multi- and interdisciplinary nature of his work is reflected in the variety of international journals in which he has published 546 articles. Also, he has published 28 books and about 180 book chapters. His five volumes *Treatise of Nutrition*, 3rd Edition, Ed. Medica Panamericana, 2017, with more than 3500 pages, is the "bedside" book for the study of Nutritional Sciences in Spain and all Latin American countries.

He has also been the Chairman of the International Union of Nutritional Sciences (IUNS) 21st International Congress of Nutrition (2013) and the Executive Director of the 23rd International Congress of Nutrition (2017) and has been engaged in the organization of other renowned international congresses. He is a member of prestigious international and national nutrition societies and Honorary President of the Iberoamerican Nutrition Foundation (FINUT), a nonprofit organization promoted by the IUNS, in which the main goal is to contribute to the formation of young scientists in Food and Nutrition in the setting of Iberoamerica. He has received 42 National and International Awards for his contribution to Nutrition and Food Science, among them, the Class Fellow 2022 of the American Society of Nutrition; the Sir David Cuthbertson Lecture Award of the European Society of Clinical Nutrition and Metabolism for scientific achievement in clinical nutrition on 2021; the Award "Granada, City of Science and Innovation" 2021 to the Scientific Career; the Gregorio Marañón Award 2018 to the best Spanish Scientist in the field of Food Science and Nutrition; the Institute Danone Spain Award 2017; the Award of the Spanish Federation of Dairy Industries, 2015; the Nutra Excellence Award 2014, Nutra India Summitt; the UIB Honorary Award of 2013; and the NAOS Strategy Prize 2012, Special recognition for his extensive professional experience in the field of nutrition and obesity, Spanish Ministry of Health, Social Services and Equality (AESAN).

Section 2: *Molecular Mechanisms for the Interaction of Nutrients and Health*

Dr. Noel W. Solomons

Center for the Studies of Sensory Impairment, Aging, and Metabolism (CeSSIAM), Guatemala City, Guatemala, Guatemala



Noel W. Solomons was Assistant and Associate Professor from 1977 to 1984 in the former Department of Nutrition and Food Science of MIT. A resident of Guatemala since 1975, he cofounded the Center for Studies of Sensory Impairment, Aging and Metabolism (CeSSIAM) in 1985, where he has been Scientific Director ever since. He is an Adjunct Professor at Tufts University in both the Friedman School of Nutrition Science and Policy and the Department of Public Health and Community Medicine.

Dr. Solomons received his MD degree from Harvard Medical School and subsequently undertook clinical and research training in infectious diseases and in gastroenterology and clinical nutrition at the Universities of Pennsylvania and University of Chicago, respectively. He was Editor-in-Chief of the *Food and Nutrition Bulletin* from 2016–2020. He has 364 publications indexed on PUBMED. In addition, he has edited 2 books and contributed over 100 articles, reviews, editorials, and commentaries in nonindexed venues and over 50 book chapters. These are dedicated to the scientific and academic interests of his career including: clinical nutrition; human growth and body composition; lactose maldigestion; dietary intake, nutritional status, intestinal absorption, and food fortification related to various micronutrients (vitamins, trace elements, and essential fatty acids); complementary feeding; nutrition in aging and chronic disease; and the interaction of malnutrition and infection. Also, he has supervised doctoral dissertations for 12 PhD candidates from the USA, Canada, Germany, Spain, the UK and the Netherlands through CeSSIAM.

Section 3: *Diet Composition*
Professor Anura Kurpad
Department of Physiology, St John's Medical College, Bengaluru, India



Anura Kurpad works at St. John's Medical College, Bengaluru. He received his MBBS and MD from St. John's, and his PhD from Bangalore University in 1992. He was a Postdoctoral Fellow at the Rowett Research Institute and at Cambridge University, UK, and a visiting Scientist at MIT, Cambridge, United States. He is an elected Fellow of the Royal College of Physicians (London), Fellow of the Indian National Academy of Medical Sciences, Indian Academy of Sciences, and International Union of Nutritional Sciences. His interests are in human, clinical, and public health aspects of nutrition, applied throughout the life cycle. His research focuses on the physiology and clinical aspects of human energy and protein requirements and metabolism, with more recent interests in micronutrient status and metabolism. He is an Associate Editor of *The American Journal of Clinical Nutrition*. He was the President of the Nutrition Society of India from 2012 to 2016 and is the current President of the Asia Pacific Clinical Nutrition Society. He has been, and is, on many national and international advisory bodies, including being the Chair of the Indian Nutrient Requirements Committee.

Section 4: *Disorders Directly Related to Inadequate Nutrient Intake*
Dr. Katherine L. Tucker
Department of Biomedical and Nutritional Sciences, Center for Population Health, University of Massachusetts Lowell, Lowell, MA, United States



Katherine L. Tucker, PhD, is University Distinguished Professor of Nutritional Epidemiology in the Department of Biomedical and Nutritional Sciences, and Director of the Center for Population Health, at the University of Massachusetts Lowell. She holds an adjunct appointment at the University of Massachusetts Medical School. She received her PhD from Cornell University and her undergraduate degree from the University of Connecticut, both in nutritional sciences. Between these degrees, she spent 2 years as a Peace Corps volunteer in the Philippines. Before joining UMass Lowell, she was at the USDA Human Nutrition Research Center on Aging at Tufts University, and McGill University. Dr. Tucker has contributed to more than 450 articles in scientific journals. Her research focuses on dietary intake and risk of chronic disease, including osteoporosis, cognitive decline, obesity, metabolic syndrome, and heart disease, with an emphasis on health disparities. She is the PI of the Boston Puerto Rican Health Study, an ongoing cohort study, to examine the roles of diet, health behaviors, stress, and genetic predisposition in relation to chronic conditions, including heart disease, cognitive decline, and bone health; and is actively involved as a scientific advisor for the NHLBI Jackson Heart Study. She served two terms on the Food and Nutrition Board of the National Academies of Science and Engineering. She is a Fellow of the American Society for Nutrition (ASN), the Gerontological Society of America, and the American Society for Bone Mineral Research and is currently the Editor-in-Chief of *Advances in Nutrition*, the international review journal of the ASN, and Senior Editor of the forthcoming 12th edition of the textbook, *Modern Nutrition in Health and Disease*.

Section 5: *Nutrition in Disease States*
Dr. Paolo M. Suter
Department of Endocrinology, Diabetology and Clinical Nutrition, University Hospital, Zurich, Switzerland.



Paolo M. Suter, MD, MS, is a Professor of Medicine presently affiliated with the Department of Endocrinology, Diabetology and Clinical Nutrition at the University Hospital Zurich (Switzerland). He received his MD at the University of Zurich and an MS in Nutrition at Tufts University (Boston, USA). He specialized in Internal Medicine and was a Faculty Member at the Medical Policlinic of the University Hospital Zurich, where he was directing the well-known Hypertension and Obesity Outpatient Consultation. His research activities focused on vitamin nutriture in the elderly, alcohol metabolism and obesity, blood pressure and hypertension, as well as nutrition and lifestyle in non-communicable disease prevention. During his clinical work he focused on the ideal combination of pharmacological therapy with nonpharmacological strategies and on the therapy and prevention of diseases especially hypertension and obesity. Besides many research publications and review articles he authored a widely used textbook entitled *Checkliste Ernährung* (*Checklist Nutrition*). He served on different national and international boards in the area of his expertise.

Section 6: *Nutrition in the Life Cycle*
Professor Lawrence J. Cheskin
Department of Nutrition and Food Studies, George Mason University, Fairfax, VA, United States
And
Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, United States



Lawrence J. Cheskin, MD, FACP, FTOS, is Professor and Chair of the Department of Nutrition and Food Studies at George Mason University in Fairfax, VA, and an Adjunct Faculty Member of the Department of Medicine at the Johns Hopkins School of Medicine in Baltimore, MD, United States. He received his medical degree from Dartmouth and was a Postdoctoral Fellow in gastroenterology and liver diseases at Yale University-New Haven Hospital. Dr. Cheskin's work is at the intersection of public health nutrition and clinical medicine and has dedicated his career to research, education, and program building in the service of combating obesity, through both treatment and prevention. He founded the Johns Hopkins Weight Management Center, a multidisciplinary clinical research and treatment program. He was also Director of Clinical Research of the Global Obesity Prevention Center at Johns Hopkins and directed its Pilot Studies Core, which evaluated systems-focused proposals worldwide to study such areas as school policies and the built environment in preventing childhood obesity. He is co-PI of the Mason Cohort Study of entering college freshmen and has authored over 240 peer-reviewed journal articles and 8 books. He is committed to mentoring the next generation of public health and clinical scholars dedicated to nutrition and obesity.

Section 7: *Nutrition Epidemiology*
Dr. Manuel Franco
Associate Professor at the Department of Surgery, Medical and Social Sciences, University of Alcalá, Madrid, Spain, and adjunct faculty at the Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States.



Manuel Franco, MD, PhD, Epidemiologist, is an Associate Faculty at the University of Alcalá and an Adjunct Faculty at the Johns Hopkins Bloomberg School of Public Health.

He is the Principal Investigator of the European Research Council Starting Grant Heart Healthy Hoods (http://www.hhhproject.es) studying urban characteristics in relation to eating patterns, physical activity levels, smoking, and alcohol consumption. He is also the PI of the Participatory Project on food in the cities: Photovoice Villaverde (https://youtu.be/VIiFggKzVas).

Manuel Franco trained in Medicine both in Madrid and Berlin. As a Fulbright Scholar he joined the Johns Hopkins Bloomberg School of Public Health Department of Epidemiology to obtain his PhD and Postdoctoral Fellowship in the fields of social epidemiology and urban health.

His work focuses on the prevention of chronic diseases and their major risk factors as nutrition. He has published 94 international peer-reviewed articles and has led 15 studies as PI raising over 2.9 million € in competitive international research bids. His methodological interests include the measurement of urban characteristics related to chronic diseases and nutrition, the use of mixed methods, and the conduction of participatory action research methods.

Section 8: *Global aspects of Human Nutrition*
Professor Daniel J. Hoffman
Department of Nutritional Sciences and director of the Program in International Nutrition at Rutgers University, New Brunswick, NJ, United States



Daniel J. Hoffman, PhD, is a Professor in the Department of Nutritional Sciences and Director of the Program in International Nutrition at Rutgers University, New Brunswick, NJ, United States. He graduated from St. John's University, Collegeville, MN, United States, and received his PhD in Human Nutrition from the Tufts University Friedman School of Nutrition Science and Policy, Boston, MA, United States. For his doctoral training, Dr. Hoffman studied the metabolism of children living in the shantytowns of São Paulo, Brazil. He then completed a Postdoctoral Fellowship at Columbia University, New York, NY, United States, before joining the faculty at Rutgers University.

Dr. Hoffman's research focuses on the long-term impact of undernutrition early in life on body composition and metabolism later in life. This work is complemented by research on sustainable methods to improve diet and growth in lower- and middle-income countries, including Brazil, Mexico, Kenya, South Sudan, and Vietnam. He also directs the Center for Childhood Nutrition Research at the New Jersey Institute for Food, Nutrition, and Health where members conduct interdisciplinary research on childhood growth and development. Dr. Hoffman is an Associate Editor of the *Food and Nutrition Bulletin* and a Section Editor for the *Annals of Human Biology*. His other responsibilities include serving on the editorial board of the *Journal of Nutrition*, serving in leadership positions for the American Society for Nutrition, and advising international and national agencies on matters related to childhood nutrition. Aside from his academic life, Dr. Hoffman is a published documentary photographer whose portfolio includes work on the life of Roma in Europe, urban landscape of São Paulo, Brazil, gentrification of Times Square in New York City, and punk music as the "American Mosaic."

PREFACE

By the middle of last century, the science of nutrition had identified most of the essential nutrients and had provided evidence to propose specific dietary intake recommendations for many diet constituents, with the practical aim of preventing nutrient deficiencies. As chronic, noncommunicable diseases such as cardiovascular disease, cancer, etc., began to emerge as an important causal factor for disability and early death, scientists turned their attention to the potential effects of nutrients and diet patterns on chronic disease risk. Pioneering studies by Burkitt, Keys, Breslow, and others were followed by a large number of studies on the role of dietary patterns and constituents on certain chronic diseases. Many important studies were completed over the second part of the century, providing the evidence to support specific dietary recommendations to reduce disease risk.

The 21st century ushered the next transition in nutrition science, this time centered on the interrelationships between nutrients, dietary patterns, and the human genome. Over the past few decades, advances in our understanding of the human genome and on the molecular tools to explore it have permitted to probe those interactions in increasing detail. In turn, findings from nutrient–gene interaction studies have informed population-wide and clinical and metabolic studies, further advancing our understanding of the effects of diets on human health at the molecular level. This understanding of the links between genotype, phenotype, and nutrient/dietary intake became a key contributor to the emerging area of personalized nutrition/precision medicine.

All those phases of research emphasis, to different degree, continue to exist today and result in a vast, multidisciplinary, ever-expanding amount of information reaching the peer-reviewed literature. This massive amount of information needs to be organized and summarized in a way that makes it accessible to experts, teachers, and, as much as possible, the general public. This has been and continues to be the goal of the *Encyclopedia of Human Nutrition* since its first edition, over 20 years ago.

Such an ambitious task can only be achieved by the collective work of many people. We all have experienced the challenge of writing an article that combines focus and relevance with conciseness, so we are very appreciative of the work of our contributors. Their effort was backed up by an excellent editorial board, which reviewed and provided feedback on every manuscript. Finally, we must acknowledge the outstanding support of the Major Reference Works division at Elsevier. A publication like this *Encyclopedia* has a lot of moving parts, and it is a great privilege to be able to concentrate on the content, knowing that the other parts of the process are in the hands of excellent professionals.

We hope that this book will help satisfy the need for accurate and concise information to the many students and professionals who are committed to use nutrition science as a tool to improve people's quality of life.

Benjamin Caballero, MD, PhD
Editor In Chief

This page intentionally left blank

Biochemical indices

C.M. Pfeiffer, R.L. Schleicher, and K.L. Caldwell, Centers for Disease Control and Prevention, Atlanta, GA, USA

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of C.M. Pfeiffer, R.L. Schleicher, K.L. Caldwell, Biochemical Indices, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 156–174, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00198-7.

Introduction 2
Protein Nutritional Status 2
Essential Fatty Acid Status 5
Vitamin Nutritional Status 5
Vitamin A 5
Vitamin D 9
Vitamin E 9
Vitamin K 9
Thiamin (vitamin B_1) 9
Riboflavin (vitamin B_2) 10
Niacin (vitamin B_3) 10
Vitamin B_6 10
Folate 10
Vitamin B_{12} 11
Biotin 11
Pantothenic Acid 11
Vitamin C 11
Essential Mineral and Trace Element Nutritional Status 12
Sodium and Potassium 12
Calcium 12
Magnesium 12
Iron 12
Zinc 12
Copper 13
Selenium 15
Iodine 15
Choice of Laboratory Tests 17
Evaluation of Laboratory Indices 18
Conclusions and Future Directions 18
Further Reading 19
Relevant Websites 19

Glossary

Acute phase response Various physiological processes occurring soon after the onset of infection, trauma, inflammatory processes, and some malignant conditions; an increase or decrease in acute phase proteins in serum, fever, increased vascular permeability, and metabolic and pathologic changes are part of an acute phase response.

Biochemical test Laboratory test that measures the level of an analyte or metabolite in, for e.g., serum, blood or urine.

Chromatography assay Laboratory technique that separates a compound of interest from other compounds by either passing a liquid or a gas in which the compound of interest is dissolved (mobile phase) through a stationary phase; this process separates the compound of interest from other compounds in the mixture based on differential partitioning between the mobile and stationary phases.

Cutoff point A concentration or quantity that is selected in clinical practice for quantitative diagnostic tests and maximizes the specificity and sensitivity of the diagnostic test.

Immunoassay Laboratory assay that typically measures analytes in biological liquids such as serum and urine; the technique is based on the unique ability of an antibody to bind with high specificity to one or a very limited group of molecules, representing the antigen.

Nutritional deficiency Progression throughout various stages from adequate to inadequate nutritional body stores, leading to disturbances in metabolism and ultimately to clinical signs and symptoms.

Introduction

Biochemical methods are considered to be the most objective measures for the assessment of nutritional status of individuals or communities. Ideally, the method employed should cover various cutoff points specific and sensitive to depletion of the nutrient body pool or tissue store.

The pathophysiology of deficiency for most nutrients, particularly vitamins, progresses in successive stages. The first stage of deficiency is when nutrient body stores begin to be depleted; in this stage, urinary excretion of the nutrient decreases, whereas homeostatic regulation ensures that the level of nutrient in the blood or tissues does not change. In the next stage, depletion is more marked; nutrient excretion via urine continues to decrease and blood and other tissue concentrations are reduced. Evidence of an increase in compensatory or dependent metabolites or enzymes tends to characterize the third stage. In the last stage, morphological or functional disturbances are present, first reversible, then irreversible, which can manifest as clinical signs and symptoms when these biochemical effects accrue or worsen.

The static and functional tests most commonly used in nutritional status assessment in humans are discussed here. Static tests measure the content of nutrients, their active or inactive metabolites, or other related components in tissues or fluids. Functional tests measure the behavioral, physiological, or biochemical functions of the organism that are dependent on the adequate availability of a nutrient, or the organism's responses to the processes to maintain body stores or homeostasis. The availability of specific and reliable chromatography-based methods in recent times has allowed for more frequent use of static compared to functional tests. The choice of tissue or fluid depends on the information required (long-term status or short-term, body pool or tissue store), operational characteristics of the biochemical test, and on the condition of the subject.

Various factors influence biochemical test results. Some are related to the individual, such as age, sex, genetics, physiological, and hormonal status; others are environmental or situational, such as seasonality, elevation, and latitude and thus cannot be eliminated; still others are behavioral and more easily modified (e.g., alcohol intake, smoking habits, and use of medicines). Major factors influencing biochemical indicators are presented in **Table 1**. In addition, infection and inflammation are known to have confounding effects on several laboratory tests and this is discussed as part of each nutrient status section. Different approaches have been investigated to account for infection and inflammation in nutrition surveys (e.g., exclusion of participants with infection, adjustment of biomarker levels in those with infection). Although no consensus exists on the best approach, the concurrent serum determination of acute phase proteins such as C-reactive protein (CRP) and α1-acid glycoprotein (AGP) is highly recommended. Influential factors of a technical nature can usually be reduced or eliminated by adherence to standardized sample handling protocols. **Table 2** provides information on preanalytical factors (e.g., fasting, storage stability, freeze/thaw stability) that influence biochemical indicators used to assess nutritional status.

Protein Nutritional Status

Several serum or plasma proteins can be used to roughly measure the adequacy of protein intake and metabolism, including albumin and several transport proteins (transthyretin (TTR) involved in thyroid hormone transport and formerly called prealbumin, transferrin (TF), and retinol binding protein (RBP)). These proteins will decrease during acute phase response. Reduction of albumin levels may indicate a catabolic state, but is more common in chronic liver disorders. Serum albumin, easily measured by automated clinical analyzers, has a large body pool and a long half-life; it is therefore a less sensitive index of immediate nutritional status. TTR, complexed with RBP in the carriage of vitamin A, TF, and RBP have smaller pool sizes and shorter half-life than serum albumin, however, their specificity as an index of protein status is low and they are confounded by some nutritional deficiencies (vitamin A deficiency for RBP, iron deficiency for TF) and disease states. Serum transport proteins are measured by radial immunodiffusion or immunoassay including nephelometry and turbidimetry. Serum insulin-like growth factor I, or somatomedin C, is a regulator of anabolic properties. It has been proposed as a more sensitive indicator to changes in protein status than other serum proteins. It can be measured by immunoassay techniques.

Urinary creatinine, derived from the catabolism of creatine phosphate which is present mainly in muscle, can be used as a biochemical marker of muscle mass after adjustment for body mass, race, and gender. It is measured by direct colorimetric and enzymatic methods and available on automated clinical analyzers. A major difficulty in estimating urine creatinine excretion is ensuring that 24-h urine collections are complete. Various assumptions are required when estimating the muscle mass from urinary creatinine and influencing factors have to be taken into consideration. In the clinical setting, the creatinine/height index is used to assess the degree of depletion of muscle mass in children with the marasmic form of protein-energy malnutrition and to monitor the effects of long-term nutritional intervention on repletion of lean body mass in hospital patients. Measurement of the amino acid 3-methylhistidine in urine is used as an indicator of muscle protein turnover, however, available data on various populations is limited.

Metabolic changes that occur in protein-energy malnutrition are sometimes used as less specific indices of protein status. These relate to changes in free amino acid profiles in plasma, reduced urinary hydroxyproline excretion and increased urinary nitrogen excretion. Functional indices of protein status include muscle function, handgrip strength, and immunological testing.

Table 1 Factors known to influence biochemical indicators

Nutritional status	*Biochemical indicator*	*Influencing factors*
Protein	Serum albumin	↑ ↓ Age, sex ↑ Dehydration ↓ Acute phase infection and chronic inflammation, catabolic states, protein-losing diseases, hemodilution, zinc depletion
	Serum transport proteins: TTR, TF and RBP	↑ Chronic renal failure (for TTR and RBP), when estrogen is increased (for TF) ↓ Infection, protein-losing diseases, hemodilution, zinc depletion, vitamin A deficiency (for RBP), iron deficiency (for TF)
	Urinary creatinine	↑ ↓ Age, sex, diet, diurnal and day-to-day variations ↑ Intensive exercise, pregnancy, catabolic states, hypothyroidism ↓ Chronic renal failure, hyperthyroidism, diseases with decreased muscle mass
	Urinary 3-methylhistidine	↑ ↓ Age, sex, diet ↑ Intensive exercise, catabolic states ↓ Chronic renal failure
	Serum IGF-I	↓ Stress, hormonal diseases, hepatocellular diseases
Fatty acids	Plasma fatty acids	↑ ↓ Smoking, exercise, stress, pregnancy, estrogens, obesity, alcohol, diabetes, renal disease ↑ Age ↓ Fat malabsorption, catabolic states, depression
Vitamin A	Serum retinol, RBP, RDR and MRDR	↑ ↓ Sex, race ↑ Age, chronic renal disease, estrogens ↓ Protein–energy malnutrition, fat malabsorption, catabolic states, zinc deficiency, liver disorders
Carotenoids	Serum carotenes, cryptoxanthin, lutein, zeaxanthin, lycopene	↑ ↓ Age, sex, race, season, smoking, alcohol, BMI ↓ Fat malabsorption Diet and season, sex and age, infection, smoking, and drinking habits and circulating lipids are determinants of carotenoid concentrations.
Vitamin D	Serum 25OHD	↑ ↓ Season/latitude, age, sex, race ↑ Sun exposure, dairy consumption, estrogens ↓ Skin pigmentation, liver or renal diseases, smoking, drugs, fat malabsorption, obesity
Vitamin E	Serum α-tocopherol	↑ ↓ Race ↑ Age, hyperlipidemia, pregnancy ↓ Fat malabsorption, abetalipoproteinemia, premature infants
Vitamin K	Serum phylloquinone (K_1), menaquinones (K_2)	↑ ↓ Age, sex, season ↑ Plasma triglyceride levels ↓ Fat malabsorption, osteoporosis, liver disease, hemorrhagic disease of newborn, antibiotics and other drugs
Thiamin	Erythrocyte TDP	↓ Hemodialysis, storage of specimen
	Serum thiamin	↓ Alcoholism
	Urinary thiamin	↓ Starvation
	Erythrocyte EKT-AC	↑ Alcoholism, pernicious anemia ↓ Diabetes mellitus, polyneuritis, storage of specimen
Riboflavin	Erythrocyte or serum FAD, FMN	↓ Alcoholism, hypothyroidism, anorexia
	Urinary riboflavin	↑ Negative nitrogen balance, infection, drugs ↓ Starvation
	Erythrocyte EGR-AC	↑ Alcoholism, heterozygous beta thalassemia, iron-deficiency anemia, severe uremia, liver cirrhosis ↓ Glucose-6-phosphate dehydrogenase deficiency, pyridoxine deficiency, storage of specimen
Niacin	Urinary methylated niacin metabolites	↓ Drugs, alcoholism
Vitamin B_6	Serum PLP	↑ ↓ Age, sex ↑ Prolonged fasting ↓ Drugs, alcohol intake, smoking, malnutrition, pregnancy, various disease states
	Urinary 4-PA	↑ ↓ Drugs
	Erythrocyte EAST-AC	↑ ↓ Drugs and diseases that affect the liver and heart ↓ Alcohol intake, pregnancy

(Continued)

Table 1 Factors known to influence biochemical indicators—cont'd

Nutritional status	*Biochemical indicator*	*Influencing factors*
Folate	Serum folate	↑ ↓ Age, sex, race ↑ Hemolysis ↓ Pregnancy, smoking, alcoholism, antifolate drugs, malnutrition, malabsorption, storage of specimen
	RBC folate	↑ ↓ Age, sex, race ↑ Iron deficiency ↓ Pregnancy, smoking, malnutrition, malabsorption, vitamin B_{12} deficiency
	Plasma tHcy	↑ ↓ Age, sex, race, diet ↑ Deficiencies in B_2, B_6, and B_{12}, impaired renal function, smoking, alcohol intake, lack of exercise, various disease states, drugs, inborn errors affecting enzymes involved in lowering tHcy level, delayed processing of blood specimen ↓ Pregnancy
Vitamin B_{12}	Serum B_{12}	↑ ↓ Age ↑ Chronic renal failure, severe congestive heart failure, diabetes, organic diseases ↓ Drugs, alcoholism, pregnancy, smoking, vegetarianism, folate deficiency, bacterial overgrowth, disease states associated with alterations in the levels of vitamin B_{12}-binding proteins, malabsorption, worm infestation
	Serum MMA	↑ Impaired renal function, bacterial overgrowth, cobalamin genetic defects, classical methylmalonic acidemia
Vitamin C	Serum total ascorbic acid	↑ ↓ Age, sex, race ↓ Smoking, low socioeconomic status, catabolic states, obesity, improper specimen handling
Sodium and potassium	Urinary sodium, potassium	↑ ↓ Age, sex ↑ Renal disease, conditions in which urine is alkaline ↓ Renal diseases with decreased urine flow, diarrhea, or excessive sweating
Calcium	Serum ionized calcium	↑ ↓ Age ↑ Hyperparathryroidism, functional hypercalcemia, hemodialysis ↓ Hypoparathyroidism, vitamin D-deficient rickets
Magnesium	Serum magnesium	↑ ↓ Age, sex, race, diurnal variation, hypo-/hyperalbuminemia ↑ Renal failure, hemolysis, drugs (antacids, cathartics) ↓ Pregnancy, strenuous exercise, osteoporosis, drugs (diuretics, antibiotics), GI and renal disease
Iron	Serum ferritin	↑ ↓ Age, sex, race ↑ Acute phase infection and chronic inflammation, liver disorders, malignant diseases, acute leukemia, Hodgkin's disease, rheumatoid arthritis, thalassemia major, alcohol intake
	Serum sTfR	↑ ↓ Age, sex, race ↑ Autoimmune hemolytic anemia, sickle cell anemia, folate, or vitamin B_{12} deficiency ↓ Chronic renal failure
	Serum iron, TIBC, TS	↑ ↓ Age (mainly for iron), biological variation (mainly for iron) Chronic disease states with infection, inflammation (decrease in iron and TIBC, TS low) Decreased erythropoiesis due to folate or vitamin B_{12} deficiency (increase in iron, decrease in TIBC, TS high) Increased erythropoiesis in response to vitamin B_{12} and folate therapy, in hemolysis, in polycythemia (decrease in iron, increase in TIBC, TS low)
	Erythrocyte zinc protoporphyrin	↑ ↓ Age, sex ↑ Chronic disease states with infection, inflammation and some neoplastic diseases, lead poisoning, porphyrin disorders
	Hemoglobin	↑ ↓ Age, sex, race, biological variation ↑ Polycythemia, dehydration

Table 1 Factors known to influence biochemical indicators—cont'd

Nutritional status	*Biochemical indicator*	*Influencing factors*
		↓ Pregnancy, iron-deficiency anemia, deficiencies of vitamin A, B_2, B_6, B_{12}, folate and copper, parasitic infections, chronic infection and inflammation, chronic diseases that cause overhydration or plasma volume expansion, smoking
Zinc	Serum zinc	↑ ↓ Age, sex, diurnal variation, fasting ↑ Hemolysis, delayed separation of serum from red cells ↓ Pregnancy, acute infection and inflammation, estrogen-containing preparations, malabsorption syndromes, chronic disease states resulting from hypoalbuminemia
Copper	Serum copper and ceruloplasmin	↑ ↓ Age, sex, diurnal variation, certain disease states ↑ Pregnancy, estrogen-containing preparations, smoking, acute phase infection, chronic inflammation, stress, delayed separation of serum from red cells ↓ Malabsorption syndromes
Selenium	Serum selenium	↑ ↓ Age, diet ↓ Prematurity, pregnancy and lactation, smoking, genetic defects (maple sirup urine disease, PKU), certain disease states (disorders of digestive tract, muscle disorders, neurological diseases, inflammatory diseases, chronic renal failure, cancer, CVD)
Iodine	Urinary iodine	↑ ↓ Diurnal variation ↓ Pregnancy
	Serum or whole blood TSH	↑ Congenital hypothyroidism, exposure to iodine-containing antiseptics

↑ ↓, biomarker response increased or decreased; ↑, biomarker response increased; ↓, biomarker response decreased; 25OHD, 25-hydroxyvitamin D; 4-PA, 4-pyridoxic acid; BMI, body mass index; CRP, C-reactive protein; EAST-AC, erythrocyte aspartate aminotransferase activation coefficient; EGR-AC, erythrocyte glutathione reductase activation coefficient; ETK-AC, erythrocyte transketolase activation coefficient; FAD, flavinadenindinucleotide; FMN, flavinmononucleotide; IGF-1, insulin-like growth factor; MMA, methylmalonic acid; MRDR, modified relative dose response test; PLP, pyridoxal-5′-phosphate; RBC, red blood cells; RBP, retinol-binding protein; RDR, relative dose response test; sTfR, soluble transferrin receptor; TDP, thiamin diphosphate; TF, transferrin; tHcy, total homocysteine; TIBC, total iron binding capacity; TS, transferrin saturation; TSH, thyroid stimulating hormone; TTR, transthyretin; UIBC, unbound iron binding capacity.

Essential Fatty Acid Status

Essential fatty acids are those fatty acids that are not synthesized by the body but are vital for proper growth and development; linoleic and linolenic acids are essential and must be obtained through dietary sources. Fatty acids can be measured in plasma or serum, in the phospholipid fraction of plasma, in red blood cell membranes, in whole blood, or in tissues. *In vivo*, fatty acids are almost always combined with sterols, glycerol, or phospholipids although free fatty acids circulate bound to albumin (1 mol albumin: 20 mol fatty acids) and the flux of free fatty acids through the plasma is very large and varies with metabolic demands. Fatty acid data are expressed either as concentration or as a percentage of total fatty acids. Capillary gas chromatography (GC) is the technique most frequently used to separate fatty acids for quantitative analysis. Detection methods include flame ionization or electron capture negative chemical ionization mass spectrometry. Internal standards are used to correct for losses during sample preparation and improve the accuracy and precision of measurements. Most often, the diagnosis of essential fatty acid deficiency is made from clinical findings but deficiency may be detected weeks to months before clinical manifestations are present using fatty acid profiling. Unusual or large changes in total or individual plasma fatty acids are due to underlying disease states (liver disorders), medications, or inborn errors of metabolism. General fatty acid deficiency can be detected when values of linoleic and alpha-linolenic acids are low based on total plasma fatty acids or phospholipid extracts of plasma. In addition, the ratio of triene to tetraene fatty acids is the most common laboratory indicator of essential fatty acid deficiency (Holman index). The omega-3 (n−3) index, a measure of the amount of eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) in red blood cell membranes as a proportion of total fatty acids, has been suggested to be a marker of risk for sudden cardiac death.

Vitamin Nutritional Status

Vitamin A

Vitamin A (retinol) status can be assessed by testing liver, plasma, or serum. The best way to determine inadequate status is through hepatic biopsy but this procedure is invasive and unsuitable in population studies. Serum or plasma retinol is usually measured using high performance liquid chromatography (HPLC) with ultraviolet (UV) detection after separation from its carrier RBP. Serum retinol values do not always reflect total body status because of homeostatic control and therefore are often not useful for assessing the vitamin A status of individuals. Additional tests may be required to confirm vitamin A deficiency when 0.70 μmol l^{-1} is used as

Table 2 Preanalytical factors influencing biochemical indicators for nutritional status assessment

Nutritional status	*Biochemical indicator*	*Specimen collection requirements*	*Storage stability*	*Freeze/thaw stability*
Protein	Serum albumin	Fasting not required	Stable for a few days refrigerated; stable for several months frozen	Minimize freeze/thawing
	Urinary creatinine	Fasting not required; 24-h collection recommended (ideally for 3 consecutive days)	Stable for a few days at RT; stable for weeks refrigerated; stable for years frozen	Stable for at least 5 cycles
	Urinary 3-methylhistidine	Fasting not required; 24-h collection essential (ideally for 3 consecutive days)	Stable for a few days refrigerated; stable for several months frozen	No information
Fatty acids	Plasma fatty acids	Fasting essential	Stable for several years at −70 °C	Stable for at least 4 cycles
	RBC fatty acids	Fasting not required	Stable for 7 days at RT for an omega-3 index assessment; stable for at least 4 years at −80 °C	No information
Vitamin A	Serum retinol	Fasting not required	Stable for years at −70 °C; stable for weeks refrigerated	Little deterioration for at least 3 cycles
	Serum RBP	Fasting not required	Reported to be as or more stable than retinol	
Carotenoids	Serum carotenes, cryptoxanthin, lutein, zeaxanthin, lycopene	Fasting recommended	Stable for years at −70 °C; stable for weeks refrigerated	Little deterioration for at least 3 cycles
Vitamin D	Serum 25OHD	Fasting not required	Stable for years at −70 °C; stable for weeks at 37 °C	Little deterioration for at least 5 cycles
Vitamin E	Serum α- and γ-tocopherol	Fasting recommended; alternatively, normalization to cholesterol or lipids content	Stable for years at −70 °C; stable for weeks refrigerated	Little deterioration for at least 3 cycles
Vitamin K	Serum phylloquinone (K_1), menaquinones (K_2)	Fasting recommended; abstain from alcohol for 1 day before blood draw	Protect from light; stable for 6 months frozen	No information
Thiamin	Erythrocyte or whole blood TDP	Fasting not required; measure Hb when whole blood is used to correct for cell volume variability	Protect from light; transport frozen to avoid loss of TDP; whole blood stable at RT up to 2 days; hemolyzed erythrocytes and whole blood stable for at least 6 months at −70 °C	Stable for 3 cycles (whole blood hemolysates)
	Serum thiamin	Fasting not required	Protect from light; stable for 1 year frozen	No information
	Urinary thiamin	Fasting not required; 24-h collection recommended	No information	No information
	Erythrocyte ETK-AC	Fasting not required; packed erythrocytes must be washed and buffy coat removed	Rapid enzyme inactivation within 2 weeks stored at −20 °C; hemolysate is stable for 1 year at −70 °C	Avoid freeze/thawing
Riboflavin	Erythrocyte, whole blood or serum FAD, FMN	Fasting not required	Protect from light; stable for at least 5 h at RT in whole blood; endogenous vitamers stable for 14 days at RT in EDTA plasma (supplemented FAD is less stable); serum vitamers stable for 1 year frozen	No information
	Urinary riboflavin	Fasting not required; 24-h collection recommended	Protect from light; little deterioration when specimen in amber glass vial is exposed to fluorescent light at RT for up to 3 days; stable for up to 6 months refrigerated or frozen at −20 °C	No information
	Erythrocyte EGR-AC	Fasting not required; packed erythrocytes must be washed	If assay is not performed immediately, hemolysate has to be frozen and is stable for 1 year at −70 °C	No information
Niacin	Urinary methylated metabolites	Fasting not required; 24-h collection recommended; acidify with HCl	No information	No information

Table 2 Preanalytical factors influencing biochemical indicators for nutritional status assessment—cont'd

Nutritional status	*Biochemical indicator*	*Specimen collection requirements*	*Storage stability*	*Freeze/thaw stability*
Vitamin B6	Serum PLP	Fasting not required	Protect from light; stable for several hours at RT in the dark; stable for 1 month at −20 °C; stable for years at −70 °C	Stable for at least 3 cycles
	Urinary 4-PA	Fasting not required (morning fasting urine preferred)	Protect from light; stability of 4-PA in serum exceeds that of PLP and stability of 4-PA in urine is expected to be very good	No information
	Erythrocyte AST-AC	Fasting not required; packed erythrocytes must be washed	Hemolysate is stable for a few months at −70 °C	No information
Folate	Serum folate	Fasting essential for individual but probably not for population	Protect from light; stable for 1 week refrigerated; stable for a few years at −70 °C; ascorbic acid can be added (0.5% w/v) before storage to improve stability	Little deterioration for at least 3 cycles
	RBC folate	Fasting not required; measure Hct to correct for packed cells; use of serum folate level in calculation of RBC folate level preferred	Protect from light; whole blood stable for several days refrigerated; hemolysate with ascorbic acid (1% w/v) stable for several years at −70 °C	Little deterioration for at least 3 cycles
	Plasma tHcy	Fasting not required; separate plasma from red cells within an hour of collection to avoid artificial increase in tHcy	Stable for days at RT; stable for weeks refrigerated; stable for years frozen	Excellent stability
Vitamin B_{12}	Serum B_{12}	Fasting not required; avoid ascorbic acid	Protect from light; stable for several days refrigerated; stable for several years at −70 °C	Stable for at least 3 cycles
	Serum MMA	Fasting not required	Stable for days at RT; stable for weeks refrigerated; stable for years frozen	Excellent stability
Vitamin C	Serum total ascorbic acid	Fasting recommended; blood must be promptly processed and serum must be acidified (metaphosphoric acid) to stabilize ascorbic acid	Acidified serum is stable for at least 10 years at −70 °C	Stable for at least 3 cycles
Calcium	Serum free calcium	Fasting recommended; avoid heparin, triethanolamine, trypsin as they bind calcium; calcium standard should contain sodium and chloride at same level as test samples	Changes in pH of blood may alter measurement; adjust to pH 7.4 with CO_2 before measurement; perform measurement as soon as possible; serum can be stored anerobically refrigerated for several days or at −20 °C for 6 months	No information
Magnesium	Serum magnesium	Fasting recommended; avoid hemolysis; collect in metal-free container; separate red cells immediately	Serum stable for several days if refrigerated; stable for months if frozen	No information
Iron	Serum ferritin	Fasting not required	Stable for years at −70 °C	Stable for at least 3 cycles
	Serum sTfR	Fasting not required	Stable for years at −70 °C	Stable for at least 3 cycles
	Serum iron/TIBC	Fasting not required	Stable for at least 10 years at −70 °C	Stable for at least 3 cycles
	Erythrocyte zinc protoporphyrin	Fasting not required; avoid hemolysis	Stable for several days if refrigerated	Do not freeze
Zinc	Serum zinc	Fasting not required; collect in prescreened metal-free container; avoid contact with rubber stopper; avoid hemolysis; remove serum promptly from red cells	Stable for years at −70 °C	Stable for multiple cycles

(Continued)

Table 2 Preanalytical factors influencing biochemical indicators for nutritional status assessment—cont'd

Nutritional status	*Biochemical indicator*	*Specimen collection requirements*	*Storage stability*	*Freeze/thaw stability*
Copper	Serum copper	Fasting not required; collect in prescreened metal-free container; remove serum promptly from red cells	Stable for years at −70 °C	Stable for multiple cycles
Selenium	Serum selenium	Fasting not required; collect in prescreened metal-free container	Stable for years at −70 °C	Stable for multiple cycles
Iodine	Urinary iodine	Fasting not required; 24-h collection recommended for individual; measurement of urinary creatinine allows for adjustment in casual samples but could be problematic in malnutrition; avoid contamination	Stable for months if refrigerated; stable at least 10 years at −70 °C	Stable for at least 3 cycles

25OHD, 25-hydroxyvitamin D; 4-PA, 4-pyridoxic acid; AST-AC, erythrocyte aspartate aminotransferase activation coefficient; CRP, C-reactive protein; EGR-AC, erythrocyte glutathione reductase activation coefficient; ETK-AC, erythrocyte transketolase activation coefficient; FAD, flavinadenindinucleotide; FMN, flavinmononucleotide; Hct, hematocrit; MMA, methylmalonic acid; PLP, pyridoxal-5′-phosphate; RBC, red blood cells; RBP, retinol binding protein; RT, room temperature; sTfR, soluble transferrin receptor; TDP, thiamin diphosphate; tHcy, total homocysteine; TIBC, total iron binding capacity.

a cutoff. The distribution of serum retinol values in a population together with the prevalence of individuals with serum retinol values below a given cutoff point provide important information about the vitamin A status of a population. WHO recommends using the prevalence of serum retinol ≤0.70 μmol l^{-1} to define public health problems involving vitamin A deficiency as mild (2–9%), moderate (10–19%) or severe (≥20%). In the case of infection or inflammation, the degree of depression of circulating retinol can be roughly quantified by assessing the concentration of certain acute phase proteins such as CRP and AGP. Misclassification of vitamin A status is likely to occur unless subjects undergoing an acute phase response from infection, trauma, or a chronic inflammatory condition are identified and adjustments are made to account for this phenomenon.

RBP is a well-regulated transport protein for retinol. Nearly all circulating retinol is bound to a soluble RBP-transthyretin (TTR) complex in equimolar amounts. Because retinol is closely correlated with RBP, the measurement of this transport protein using enzyme-linked immunosorbent assay (ELISA) has been used to assess vitamin A status. In most populations, serum RBP has been shown to be a suitable surrogate for retinol. The molar ratio of RBP to TTR was introduced to detect vitamin A deficiency in the presence of inflammation but is rarely used nowadays because of unsatisfactory results in several population groups.

The deuterated-retinol-dilution (DRD) technique is used to indirectly assess total body vitamin A reserves. In this test, a dose of deuterium-labeled retinyl acetate is given orally. After allowing time to reach equilibration (3–21 days), deuterated and nondeuterated serum retinol levels are measured using GC with mass spectrometric detection (GC–MS) and this ratio is used to estimate total body stores of vitamin A. Because of technical requirements, this method is used mostly in research. In inflammation, the release of RBP from the liver is reduced, so the test is also unreliable during the acute phase response.

The provitamin (carotenes and cryptoxanthins) and nonprovitamin (lutein, zeaxanthin, and lycopene) compounds of vitamin A, the carotenoids, need consideration due to their independent and specific role in good health by preventing oxidation. Serum levels of carotenoids are correlated with vegetable and fruit intake. Lutein is the best indicator of green leafy vegetable consumption. Lycopene is a good measure of tomato-based product consumption. β-Carotene in industrialized countries is probably a biomarker of carrot consumption and in West Africa a good marker of red palm oil consumption. Serum or plasma levels of carotenoids are measured by HPLC-UV/vis; it is possible to measure in a single assay a panel of about a dozen fat-soluble micronutrients including several forms of vitamins A and E, and individual carotenoids.

Several functional tests have been developed to assess vitamin A reserves in the liver. These tests are based on the accumulation of unbound RBP within the liver when vitamin A intake is low and reserves of retinyl esters are depleted. Once retinol becomes available as a result of uptake from food or supplements or biosynthesis from carotenoids, it is released from the liver bound to RBP within hours of availability. Vitamin A functional tests, the relative dose-response (RDR) and the modified relative dose-response (MRDR) test, take advantage of this immediate release.

Other functional testing protocols involving dark adaptometry have been used to assess night vision as an indicator of vitamin A status. In the rapid dark adaptation test, the subject is light adapted and then while working in dim light, usually takes <10 min to correctly sort colored disks while adapting to darkness. The time required to achieve a perfect score is recorded. A simpler functional test is one in which pupillary dark adaptation is assessed; it requires minimal cooperation and is suitable for very young children

(≥2 years) who are most likely to be vitamin A deficient. The pupillary threshold test measures the tendency of the pupil to constrict in response to illumination. An impaired response is seen when vitamin A stores are depleted even if overt clinical signs of deficiency are absent.

Vitamin D

Vitamin D status is routinely assessed by measurement of serum or plasma 25-hydroxyvitamin D (25OHD). Typical methods are either antibody-based (e.g., radioisotope-, enzyme-linked- or chemiluminescence immunoassay), or chemistry-based (e.g., HPLC separation with UV or tandem mass spectrometry [MS/MS] detection). GC–MS has also been employed. Studies have shown that the chemistry-based methods are equivalent but that antibody-based methods may show significant bias compared to chemistry-based methods. Genetic variants near genes involved in cholesterol synthesis, hydroxylation, and vitamin D transport affect vitamin D status.

Several vitamin D functional tests have been suggested such as the measure of: (1) circulating parathyroid hormone (PTH); (2) bone resorption markers such as serum alkaline phosphatase activity, serum C-terminal telopeptide of type I collagen, urinary cross-linked N-telopeptides of type I collagen, and urinary deoxypyridinoline; (3) bone mineral density (BMD); and (4) intestinal calcium absorption. PTH appears to be a useful marker of vitamin D status when vitamin D is given without calcium but not with calcium. Markers of bone resorption are not good functional markers of vitamin D status, due in part to high inter-subject variability; however, bone resorption using a more sensitive calcium radioisotope method may prove to be useful. Increased BMD of some sites appears to be useful as a functional response to vitamin D supplementation in older persons but not adolescents. Intestinal calcium absorption has been shown to be responsive to vitamin D status.

Vitamin E

Vitamin E status can be assessed in serum, plasma, erythrocytes, platelets, and adipose tissue. The most common and practical measure is α-tocopherol in serum or plasma using HPLC with UV detection. Because α-tocopherol is bound to lipoproteins, circulating α-tocopherol is often expressed relative to serum cholesterol. The determination of α-tocopherol in adipose tissue biopsy provides information on long-term nutritional status, but this test is too invasive.

Vitamin E functional tests consist of the following assays: erythrocyte hemolysis, erythrocyte malondialdehyde (MDA) release, breath pentane or ethane, susceptibility of low density lipoprotein (LDL) to oxidation using diene conjugate second derivatives, and isoprostane formation. Susceptibility to erythrocyte hemolysis is inversely correlated with α-tocopherol concentration. *In vivo*, hemolysis (>20%) occurs when α-tocopherol concentration is <4.6–11.6 μmol l^{-1}. MDA is a breakdown product of lipid peroxidation and is measured colorimetrically. These two *in vitro* erythrocyte assays are the primary vitamin E functional tests; methodological limitations including the need for freshly prepared red blood cells make them disadvantageous.

Volatile hydrocarbons (pentane and ethane) in breath, generated from oxidized lipids, and the susceptibility of LDL to oxidation are general markers of oxidative stress and also have been investigated as functional tests of vitamin E deficiency, but technical difficulties, lack of specificity, and inconsistent findings were some of the problems. More recently, vitamin E has been shown to decrease isoprostane F2 in individuals with moderate hypercholesterolemia who exhibited oxidative stress. F2-isoprostanes are formed by free radical-mediated peroxidation of arachidonic acid, an omega-6 polyunsaturated fatty acid. Plasma concentrations of F2-isoprostanes were suppressed in these patients by 35–50% in a dose-dependent manner with maximum suppression at vitamin E intake of 1600–3200 IU d^{-1}.

Vitamin K

Vitamin K status assessment is usually performed in response to abnormal bleeding. Static tests are used to measure circulating concentrations of vitamins K_1 (phylloquinone) and K_2 (menaquinone). Serum or plasma vitamin K_1 is measured using HPLC with postcolumn chemical reduction followed by fluorometric detection or with electrochemical detection. Hemolysis should be avoided when the detection method is fluorescence. Measurement of K_1 has also been performed using GC–MS or HPLC-MS/MS. It has been suggested, although not yet commonly implemented, that K_1 concentration should be expressed as a ratio of the triglyceride concentration.

Determination of the serum undercarboxylated form of prothrombin (PIVKA-II) by ELISA and urinary γ-carboxyglutamic acid by HPLC with fluorometric detection have been proposed for assessment of vitamin K status. A vitamin K functional test that has been more widely used in recent years is the determination of serum undercarboxylated osteocalcin. This test is well correlated with other indicators of vitamin K status. A number of commercial ELISA kits are available.

Thiamin (vitamin B_1)

Thiamin status can be assessed through measurement of thiamin diphosphate (TDP) – also known as thiamin pyrophosphate (TPP) – in whole blood or erythrocytes by HPLC (preferably after postcolumn derivatization and fluorometric detection). TDP is the primary active form of vitamin B_1 and approximately 90% of thiamin in whole blood is present as TDP. Thiamin concentration in serum or plasma is small compared to that in erythrocytes and reflects recent intake rather than body stores. Urinary

thiamin excretion (preferably as 24-h urine samples) under basal conditions or after thiamin loading also reflects recent dietary intake, but the within-subject variation is high. Owing to limited body stores of thiamin, deficiencies can develop within a few weeks if intake is restricted. The response to thiamin therapy is usually rapid and a reliable test for thiamin deficiency. Infections that prevent normal absorption (diarrhea and dysentery) or increase the requirement (fever) can confound tests for thiamin status.

The erythrocyte transketolase activation coefficient (ETK-AC) test is a functional test that reflects the adequacy of body stores and is sensitive to marginal thiamin deficiency. It is measured by spectrophotometry. Because transketolase is a thiamin-dependent enzyme with a specific role in the glucose oxidative pathway, decreased enzyme activity is presumed to be due to the decrease of thiamin. However, the test is somewhat nonspecific, as factors other than thiamin status, such as genetic defects, may influence the enzyme activity and thus the test results. The ETK-AC test also suffers from imprecision problems and rapid loss of enzyme activity during frozen storage and repeated freezing and thawing. There is some debate whether the ETK-AC test – due to its slower response compared to the increase of TDP in whole blood – is a better biomarker to follow thiamin supplementation.

Riboflavin (vitamin B_2)

Riboflavin status can be assessed by urinary excretion of the vitamin in fasting, random, 24-h specimens (preferable), or by loading test, and by whole blood, erythrocyte, plasma, or serum flavin concentration. Riboflavin urinary excretion is indicative of recent dietary intake and is measured by HPLC using fluorometric detection, taking advantage of the inherent fluorescent properties of flavins. Erythrocytes are considered to be a more useful sample than plasma or serum because the riboflavin cofactors flavinadeninedinucleotide (FAD) and flavinmononucleotide (FMN) are concentrated in erythrocytes. They can be measured by HPLC with fluorometric detection. Whole blood FAD is considered a reliable indicator of long-term nutritional status, whereas FMN responds more quickly to changes in riboflavin intake. The light-sensitivity of riboflavin in particular and most B vitamins in general requires careful sample handling.

A riboflavin functional test that is commonly used is the erythrocyte glutathione reductase activation coefficient (EGR-AC) test measured by spectrophotometry. Because glutathione reductase is a flavoenzyme with FAD as a prosthetic group, the EGR-AC test is an indirect measure of FAD concentration in the erythrocytes and is considered a sensitive and robust index of riboflavin deficiency, but is less suitable for the assessment of riboflavin status at high riboflavin intake.

Niacin (vitamin B_3)

Niacin intake status can be assessed by measuring the excretion of methylated metabolites in urine by HPLC. Such metabolites are N'-methylnicotinamide (N'MN) and N'-methyl-2-pyridone-5-carboxamine (2-Py). Other biochemical markers include erythrocyte pyridine nucleotides, oral dose uptake tests, and plasma 2-pyridone derivative after an oral niacin load. Plasma concentrations of other niacin metabolites and of niacin are not useful markers of niacin status. The most reliable test for niacin deficiency is the patient's response to niacin therapy.

Niacin used as a drug has seen a sudden surge in popularity for treatment of lipid disorders and other associated clinical conditions for the prevention of cardiovascular risk. To clarify the role of metabolic pathways and evaluate pharmacokinetic studies, HPLC-MS/MS assays have recently been developed to measure the levels of niacin and its metabolites in various biological matrices.

Vitamin B_6

Vitamin B_6 status is typically assessed by measuring the level of one or more of the B_6 vitamers in serum or plasma. Serum pyridoxal-5′-phosphate (PLP) is generally viewed as the best single indicator of status. PLP is the active coenzyme form of vitamin B_6, reflects both dietary intake and tissue stores, and changes slowly in response to changes in dietary intake. 4-Pyridoxic acid (4 PA) is the end product of vitamin B_6 catabolism. 4 PA can be measured in serum, plasma, or urine and reflects recent intake. PLP and the B_6 vitamers are most commonly measured by HPLC using fluorescence detection. Chemical derivatization (sample, online, or post-column) is almost always used to enhance PLP fluorescence. HPLC-MS/MS methods for measuring B_6 vitamers are emerging. Plasma PLP can also be measured enzymatically, either by radioactive or nonradioactive assays.

Vitamin B_6 functional tests, such as the erythrocyte aspartate aminotransferase activation coefficient (AST-AC) test and the tryptophan load test, have been used more frequently in the past. Plasma total homocysteine (tHcy) in the absence of folate and vitamin B_{12} deficiencies can be considered indicative of vitamin B_6 status. For its determination, see the following discussion of folate functional tests.

Folate

Folate status can be assessed by serum or plasma folate, which provides information on recent intake, and erythrocyte folate, indicative of body folate stores and long-term nutritional status. Traditionally, folate has been measured by microbiologic assay, however, in clinical settings where high throughput is needed, commercial protein-binding assays on automated clinical analyzers are used. If folate vitamers are of interest, for e.g., the measurement of free folic acid in serum or the measurement of various methyl- and nonmethyl-folate forms in erythrocytes depending on MTHFR C677T genotype, chromatography-based separation techniques need to be employed. Nowadays, they are often coupled to mass spectrometry (HPLC-MS/MS), because this detection method

offers superior sensitivity, specificity, and selectivity compared to other detection methods such as fluorometric or electrochemical detection. Although the comparability of serum folate methods has been somewhat improved recently because serum-based standard reference materials have become available, diagnostic kits are not yet sufficiently standardized and no progress has yet been made in improving the comparability of assays for erythrocyte folate. Folate is the least stable of the B vitamins; careful sample handling and use of antioxidants are required to maintain sample integrity. Dried blood spots can also be used to measure folate by microbiologic assay. This presents a field-friendly alternative when prompt specimen processing cannot be performed or blood collection is limited to a finger stick.

In the absence of vitamin B_{12} and B_6 deficiencies, measurement of plasma tHcy is a sensitive functional test for folate status. Because an elevated plasma tHcy concentration is associated with an increased risk of cardiovascular diseases, the determination of this amino acid in plasma has become very common. Various methods are available for tHcy determination, but the most commonly used research methods are HPLC with fluorescence detection or coupled to mass spectrometry. They allow simultaneous measurement of other thiols in the same sample. Many fully-automated commercial kits are available on the basis of immunoassay and enzymatic methods. Prompt separation of the plasma from the red cells needs to be ensured to avoid artificial elevation of tHcy. Urinary formiminoglutamic acid (FIGLU) and lymphocyte deoxyuridine (dU) suppression assays are older functional tests for folate status that are no longer used routinely. Hypersegmentation of neutrophilic granulocytes are sometimes seen as a functional indicator during the examination of blood smears following routine cellular blood counts.

Vitamin B_{12}

Vitamin B_{12} status can be assessed by measuring serum or plasma total cobalamins and serum holo-transcobalamin II. The level of plasma vitamin B_{12} falls relatively late in depletion, limiting the utility of an isolated vitamin B_{12} measurement. Serum or plasma total cobalamins are commonly determined by competitive protein-binding assay, but microbiologic assays have also been used earlier. Holotranscobalamin II (holo TC) is the transport protein of absorbed cobalamin and has been considered as an early indicator of vitamin B_{12} deficiency and possibly a marker of cobalamin malabsorption. The availability of the holo TC assay is currently limited and only recently have reliable and sensitive methods for estimating holo TC become available. A new microparticle enzyme immunoassay is available on an automated immunoassay analyzer and can measure holoTC directly without sample pretreatment.

Vitamin B_{12} functional tests are the urinary or serum methylmalonic acid (MMA) and the plasma tHcy. MMA increases in vitamin B_{12} deficiency; the loading with valine or isoleucine produces a marked increase in both urine and serum. MMA is measured by GC–MS or HPLC-MS/MS. In the absence of folate and vitamin B_6 deficiencies, tHcy in plasma increases in vitamin B_{12} deficiency and decreases with B_{12} administration. For tHcy determination, see the discussion of folate functional tests.

Biotin

Biotin status has been assessed traditionally with bioassays and microbiologic assays. Modern methods rely upon the binding of biotin by either the protein avidin or streptavidin as part of competitive binding assays. To minimize the influence of interfering substances, prior separation and purification of biotin and its metabolites can be performed by HPLC. A low plasma biotin concentration is not a sensitive indicator of inadequate biotin intake. Urinary biotin excretion, particularly when extended to the excretion of its metabolites (that is, 3-hydroxyisovaleric acid [3-HIA] and 3-methylcrotonylglycine) is a more sensitive indicator of biotin status. Biotin is abnormally decreased, whereas 3-HIA is abnormally increased in urine in deficiency. The resolution of the signs and symptoms of deficiency in response to biotin supplementation is also important in the diagnosis of biotin deficiency.

Pantothenic Acid

Pantothenic acid intake status can be assessed by measuring whole blood concentrations or urinary excretion. The widespread occurrence of releasable pantothenic acid in food however makes a dietary deficiency unlikely.

Vitamin C

Vitamin C status can be assessed by measuring total ascorbic acid (oxidized and reduced) in serum or plasma, buffy-coat, or leukocytes. Ascorbic acid in plasma is considered as an index of the circulating vitamin available to tissues, and in leukocytes (particularly polymorphonuclear) it is believed to be a good indicator of tissue stores. Concentrations in leukocytes are much higher than in serum or plasma (14-fold). Isolation of specific cells is technically challenging limiting its usefulness. The urinary excretion of ascorbic acid is an index of recent intake; but because of instability of the collected sample, the determination is limited to special cases. Serum or plasma vitamin C is the most practical indicator of vitamin C status, however preanalytical requirements must be followed to promptly generate an acidified serum sample to stabilize ascorbic acid. Ascorbic acid is measured using HPLC coupled with electrochemical detectors. Newer methods have incorporated internal standards to improve accuracy and precision. Acute and chronic infections can depress markedly the serum ascorbic acid level due to a decrease in vitamin C reserves.

There are no reliable functional tests for vitamin C status.

Essential Mineral and Trace Element Nutritional Status

Sodium and Potassium

Sodium, potassium, and chloride in serum have little meaning in nutritional terms because they are tightly regulated. However, the excretion of sodium, potassium, and chloride in urine are a good indicator of intake. A 24-h specimen is needed to interpret concentrations in a person. Equations are available to calculate estimated electrolyte excretion from measured spot urine concentrations of electrolytes and creatinine. Electrolyte concentrations in serum and urine are typically measured by ion-selective electrodes (ISE), often available as add-ons to chemistry analyzers.

Calcium

Calcium intake cannot be assessed satisfactorily on a routine basis. Serum calcium concentrations (free plus bound calcium) are strongly homeostatically controlled and remain constant under most conditions. Serum or plasma free calcium (a.k.a., ionized calcium), the physiologically active form, is increasingly used to assess disturbances in calcium metabolism and is the most promising index of calcium status. It is measured by a calcium-selective electrode. Because one of the major roles of calcium is to be a structural component in bone and soft tissues, measurement of biochemical markers of bone remodeling (serum bone-specific alkaline phosphatase and osteocalcin; urine pyridinoline and deoxypyridinoline) or measurement of bone mass and bone density are indirect ways to assess calcium status.

Magnesium

Magnesium status can be assessed by measuring magnesium in serum, erythrocytes, leukocytes, and urine. Serum is the matrix most commonly used, mainly because of the ease of measurement by colorimetric methods or atomic absorption spectroscopy (AAS). Serum magnesium concentrations decrease rapidly in developing deficiency, followed by a slower decline of magnesium concentrations in erythrocytes. The validity of leukocyte magnesium concentrations as a biomarker of total body magnesium status is still under investigation. Urinary magnesium has been used as an indicator of magnesium status, primarily in association with a magnesium load test. However, this test is invasive and cumbersome and the protocol requires standardization. Measurement of serum ionized magnesium concentrations using ISE is promising, but its use in clinical disease states requires more investigation.

Iron

Iron status is assessed in relation to three stages of development of iron-deficiency anemia. In the first stage, to evaluate the size of body iron stores, serum ferritin can be measured using immunological methods (immunoturbidity, immunonephelometry, chemiluminescence, or ELISA). Commercial kits are available for most clinical analyzers. In the second stage, to determine the adequacy of iron supply to the erythroid marrow, the following biochemical indicators can be measured: serum iron (colorimetric methods, available as commercials kits), erythrocyte protoporphyrin (specific hematofluorometer), and serum soluble transferrin receptor (sTfR) (immunological methods, available as commercial kits). The transferrin saturation (TS) is calculated as the ratio of serum iron/TIBC (expressed as a percentage). In the third stage, iron-deficiency anemia develops, for which hemoglobin (Hb, spectrophotometry or automated with an electronic counter) is the most common indicator. Infection is an important confounder for iron status markers. Serum ferritin and erythrocyte protoporphyrin levels increase, whereas serum iron, serum iron binding capacity and Hb decrease.

Multiple indicators should be used to assess iron deficiency. The ferritin model has been used extensively in the past. It defines iron deficiency as an abnormal value for at least two of three indicators (serum ferritin, erythrocyte protoporphyrin, and TS). An approach for estimating body iron was developed more recently. It uses two indicators, serum ferritin and sTfR, and allows the full range of the iron status to be evaluated. Furthermore, sTfR is generally not influenced by infection, inflammation, and chronic diseases. This body iron model is currently applied to the National Health and Nutrition Examination Survey (NHANES) to assess the iron status of the US population.

Zinc

In healthy individuals, plasma or serum zinc are reliable markers of zinc status, mainly reflecting zinc intake. Because the effective regulation of zinc homeostasis buffers the functional response to dietary deficiency and excess, plasma zinc levels are generally considered a poor measure of marginal zinc deficiency. Urinary zinc excretion (24-h) and hair zinc can provide useful information on zinc status in zinc-supplemented persons, but whether these reflect zinc status in depleted persons is not clear. Zinc levels are typically measured by inductively coupled plasma mass spectrometry (ICP-MS), however, they can be assessed using AAS as well. Many precautions are required during sample collection, preparation, and storage to avoid contamination of the specimen (environmental exposure and hemolysis). Newer evidence suggests that platelet, mononuclear, and polymorphonuclear cell, and erythrocyte zinc levels are ineffective as biomarkers of zinc status.

Zinc functional tests are serum or plasma alkaline phosphatase, erythrocyte metallothionein (MT), monocyte metallothionein mRNA (MTmRNA), and serum thymulin assays. Alkaline phosphatase is a zinc metalloenzyme; rather than being indicative of zinc

Table 3 Biochemical indicators and analytical methods used in NHANES to assess the nutritional status of the US population during some or all years of 1999–2010

Nutritional status	*Biochemical indicator*	*Analytical method*
Protein	Serum albumin	Colorimetric on analyzer
	Urinary creatinine	Colorimetric Jaffé reaction on analyzer
Infection/inflammation	Serum CRP	Immunonephelometry on analyzer
Lipids	Serum total cholesterol, triglycerides, HDL and LDL cholesterol (calculated)	Enzymatic and colorimetric on analyzer
Fatty acid	Plasma or serum fatty acids	GC-MS
Vitamin A	Serum retinol and retinyl esters	HPLC-UV/VIS
Carotenoids	Serum carotenes, cryptoxanthin, lutein, zeaxanthin, lycopene	HPLC-UV/VIS
Vitamin D	Serum 25OHD	Radioimmunoassay (2000–2006)
	Serum $25OHD_2$, $25OHD_3$, epi-$25OHD_3$	LC-MS/MS (2007–2010)
Vitamin E	Serum α- and γ-tocopherol	HPLC-UV/VIS
Vitamin B6	Serum PLP	Enzymatic (2003–2004)
	Serum PLP and 4-PA	HPLC-FD (2005–2010)
Folate	Serum folate (total)	Radio protein binding assay (1999–2006)
	Serum folate (total)	Microbiologic assay (2007–2010)
	Serum folate (species)	LC-MS/MS (2007–2008)
	RBC folate (total)	Radio protein binding assay (1999–2006)
	RBC folate (total)	Microbiologic assay (2007–2010)
	Plasma tHcy	FPIA on analyzer
Vitamin B_{12}	Serum B12	Radio protein binding assay
	Serum MMA	GC-MS
Vitamin C	Serum total ascorbic acid	HPLC-ED
Sodium, potassium	Serum sodium and potassium	ISE on analyzer
Calcium	Serum total calcium	ISE on analyzer
Iron	Serum ferritin	Radioimmunoassay (1999–2002)
		Immunoturbidity on analyzer (2003–2010)
	Serum iron, TIBC, TS (calculated)	Colorimetric manual (1999–2002)
	Serum iron, UIBC, TS (calculated)	Colorimetric on analyzer (2003–2010); iron only (2007–2010)
	Serum sTfR	Immunoturbidity on analyzer
	Erythrocyte protoporphyrin	Fluorometric manual
Selenium	Serum selenium	ICP-MS
Iodine	Urinary iodine	ICP-MS

25OHD, 25-hydroxyvitamin D; $25OHD_2$, 25-hydroxyvitamin D_2; $25OHD_3$, 25-hydroxyvitamin D_3; 4-PA, 4-pyridoxic acid; CRP, C-reactive protein; epi-$25OHD_3$, 3-epimer-25-hydroxyvitamin D_3; FPIA, fluorescence polarization immunoassay; GC–MS, gas chromatography with mass spectrometry detection; HDL, high-density lipoprotein; HPLC-ED, high performance liquid chromatography with electrochemical detection; HPLC-FD, HPLC with fluorometric detection; HPLC-UV/VIS, HPLC with UV and visible detection; ICP-MS, inductively coupled plasma with mass spectrometry detection; ISE, ion selective electrode; LDL, low-density lipoprotein; LC-MS/MS, liquid chromatography with tandem mass spectrometry detection; MMA, methylmalonic acid; NHANES, National Health and Nutrition Examination Survey; PLP, pyridoxal-5′-phosphate; RBC, red blood cells; sTfR, soluble transferrin receptor; tHcy, total homocysteine; TIBC, total iron binding capacity; TS, transferrin saturation; UIBC, unbound iron binding capacity.

deficiency, it is considered to be of value after zinc supplementation but with contrasting results. A commercial kit is available for plasma alkaline phosphatase determination. Alkaline phosphatase activity has low specificity and is subject to many pathophysiological conditions. Erythrocyte MT decreases in moderate and severe zinc depletion and changes in response to elevated dietary zinc intake. Erythrocyte MT is measured by sandwich ELISA assay. MTmRNA is a new approach to zinc status assessment. It responds more rapidly to zinc supplements than erythrocyte MT. MTmRNA is measured in monocytes by competitive reserve transcriptase-polymerase chain reaction. An improvement of this method is the determination of MTmRNA on blood samples spotted onto filter paper. Confounding effects are limited to infection. MT and MTmRNA assays are very promising; further studies are needed because of the difficulty in their determination. Serum thymulin activity is decreased in zinc deficiency because it requires zinc to maintain its structure. There are some indications that erythrocytes, PMNCs, mononuclear cells, platelet zinc, and plasma alkaline phosphatase are not useful biomarkers of zinc status.

Copper

Serum copper is the most useful marker of copper status, effective in both replete and depleted persons. The tight homeostatic regulation of copper levels in circulation generally restricts major perturbations in levels to the extremes of dietary intake. Serum copper

Table 4 Available reference materials and external quality assessment programs for biochemical indicators of nutritional status

Nutritional status	*Biochemical indicator*	*Reference materials*	*Selected list of external quality assessment programs*
Protein	Serum albumin and transport proteins (TTR, TS and RBP)	ERM-DA470 (human serum, freeze-dried; one level; consensus value)	CAP General Chemistry Survey and Cal V/L Survey
Infection/ inflammation	Serum CRP	ERM-DA472 (human serum, frozen; one level; consensus value)	CAP CRP-Immunology Survey and Cal V/L Survey
	Serum AGP	ERM-DA470 (human serum, freeze-dried; one level; consensus value)	Not available at this time
Fatty acids	Fatty acids	NIST SRM 1950 (human plasma, frozen; one level; certified values for selected fatty acids)	Not available at this time
Vitamin A	Serum retinol	NIST SRM 968e (human serum, frozen; three levels; certified values)	NIST MMQAP; UK NEQAS
Carotenoids	Serum carotenes, cryptoxanthin, lutein, zeaxanthin, lycopene	NIST SRM 968e (human serum, frozen; three levels; certified values)	NIST MMQAP; UK NEQAS
Vitamin D	Serum 25OHD$_2$, 25OHD$_3$ and epi-25OHD$_3$	NIST SRM 972 (human serum, frozen; four levels; certified values for 25OHD$_2$, 25OHD$_3$, and epi-25OHD$_3$); NIST SRM 2972 (solvent-based; certified values for 25OHD$_2$ and 25OHD$_3$)	UK DEQAS; CAP Bone and Growth Survey and ABVD Survey; and NIST/NIH VitDQAP
Vitamin E	Serum α- and γ-tocopherol	NIST SRM 968e (human serum, frozen; three levels; certified values)	NIST MMQAP; UK NEQAS
Vitamin K	Serum phylloquinone (K_1)	UK KEQAS SRM-001 (human plasma; one level; consensus value)	UK KEQAS; NIST MMQAP
Vitamin B_6	Serum PLP	NIST SRM 3950 (human serum, frozen; two levels; certified values)	Not available at this time
Folate	Serum folate species	NIST SRM 1955 (human plasma, frozen; three levels; certified values for 5MTHF, reference values for FA, information values for TFOL and 5FTHF); NIBSC RM 03/178 (human serum, freeze-dried; one level; certified values for 5MTHF, FA, 5FTHF and TFOL)	CAP Ligand Assay General Survey and Cal V/L Survey for TFOL; UK NEQAS for TFOL
	Whole blood folate	NIBSC RM 95/528 (human whole blood hemolysate, freeze-dried; one level; consensus value)	CAP Ligand Assay General Survey and Cal V/L Survey for TFOL; UK NEQAS for TFOL
	Plasma tHcy	NIST SRM 1955 (plasma, frozen; three levels; certified values)	CAP Homocysteine Survey and Cal V/L Survey; DEKS
Vitamin B_{12}	Serum B12	NIBSC RM 03/178 (human serum, freeze-dried; one level; consensus values); NIBSC RM 81/563 (serum, freeze-dried; one level; consensus values)	CAP Ligand Assay General Survey and Cal V/L Survey; UK NEQAS
	Serum MMA	NIST SRM 1950 (human plasma, frozen; one level; certified value)	DEKS
Vitamin C	Serum total ascorbic acid	NIST SRM 970 (human serum, frozen; four levels; certified values)	NIST MMQAP
Sodium and potassium	Serum and urine sodium and potassium	NIST SRM 2201 (NaCl standard) and 2202 (KCl standard)	CAP General Chemistry Survey and Cal V/L Survey; CAP Urine Chemistry Survey and Cal V/L Survey
		NIST SRM 956c (human serum, frozen; three levels; certified value)	
Calcium	Serum total calcium	NIST SRM 956c (human serum, frozen; three levels; certified value)	CAP General Chemistry Survey and Cal V/L Survey
Magnesium	Serum magnesium	NIST SRM 956c (human serum, frozen; three levels; certified value)	Center de Toxicology Quebec
Iron	Serum ferritin	NIBSC RM 94/572 (human plasma, freeze-dried; one level; consensus value)	CAP Chemistry Survey and Cal V/L Survey; UK NEQAS
	Serum sTfR	NIBSC RR 07/202 (human serum, freeze-dried; one level; gravimetric/ spectrophotometric value assignment)	Not available at this time

Table 4 Available reference materials and external quality assessment programs for biochemical indicators of nutritional status—cont'd

Nutritional status	*Biochemical indicator*	*Reference materials*	*Selected list of external quality assessment programs*
	Serum iron	NIST SRM 937 (iron metal); NIST SRM 3126a (iron standard solution)	CAP Chemistry Survey and Cal V/L Survey; UK NEQAS
	Erythrocyte protoporphyrin	None available	State of New York Department of Health, Wadsworth Center
Zinc	Serum zinc	NIST SRM 1598a (bovine serum, frozen; one level; certified value)	CAP Chemistry Survey and Cal V/L Survey; Center de Toxicology Quebec
Copper	Serum copper	NIST SRM 1598a (bovine serum, frozen; one level; certified value)	CAP Chemistry Survey and Cal V/L Survey; Center de Toxicology Quebec
Selenium	Serum selenium	NIST SRM 1598a (bovine serum, frozen; one level; certified value)	CAP Chemistry Survey and Cal V/L Survey; Center de Toxicology Quebec
Iodine	Urine iodine	NIST SRM 3668 (urine-based; two levels) and NIST SRM 2668 (urine-based; two levels)	CDC EQUIP (Ensuring the quality of Urine Iodine Procedures)

25OHD, 25-hydroxyvitamin D; $25OHD_2$, 25-hydroxyvitamin D_2; $25OHD_3$, 25-hydroxyvitamin D_3; 5FTHF, 5-formyltetrahydrofolic acid; 5MTHF, 5-methyltetrahydrofolic acid; ABVD, Accuracy-based vitamin D; AGP, α1-acid glycoprotein; Cal V/L, Calibration Verification and Linearity Survey; CAP, College of American Pathologists; CDC, Centers for Disease Control and Prevention; CRP, C-reactive protein; DEKS, Danish External Quality Assessment Program; DEQAS, Vitamin D External Quality Assessment Scheme; epi-$25OHD_3$, 3-epimer-25-hydroxyvitamin D_3; EQUIP, Ensuring the quality of Urine Iodine Procedures; FA, folic acid; KEQAS, Vitamin K External Quality Assurance Scheme; MMA, methylmalonic acid; MMQAP, Micronutrients Measurement Quality Assurance Program; NEQAS, National External Quality Assessment; NIBSC, National Institute for Biological Standards and Control; NIH, National Institutes of Health; NIST, National Institute of Standards and Technology; PLP, pyridoxal-5′-phosphate; RBP, retinol-binding protein; RM, reference material; RR, reference reagent; SRM, standard reference material; sTfR, soluble transferrin receptor; TF, transferrin; TFOL, total folate; tHcy, total homocysteine; TTR, transthyretin; UK, United Kingdom; VitDQAP, Vitamin D Metabolites Quality Assurance Program.

is frequently measured by ICP-MS, although AAS is a common alternative. Levels of copper in other tissues or fluids are difficult to assess or are not considered valid indices of copper status.

Different copper functional tests are available: serum ceruloplasmin, erythrocyte superoxide dismutase (SOD), and leukocyte/platelet cytochrome c oxidase assays. Serum ceruloplasmin can be measured for its oxidase activity on various substrates or by radial immunodiffusion. The majority of serum copper is bound to ceruloplasmin, resulting in a similar response for these two markers. However, ceruloplasmin is not a useful marker in copper-replete adults and levels can be affected by a range of nondietary factors, mainly by infection because this is an acute phase reactant protein. Cu/Zn SOD is a cytosolic metalloprotein that catalyzes the reduction of superoxide to hydrogen peroxide and oxygen. Newer evidence suggests that it is not a useful marker of copper status. Cytochrome c oxidase activity in platelets or leukocytes is another marker under investigation, but the paucity of data does not allow any firm conclusions about its usefulness yet.

Selenium

Short-term selenium status is usually assessed by measuring plasma or serum selenium by ICP-MS or by AAS, a more commonly-available measurement procedure. Long-term selenium status is assessed by measuring selenium in whole blood or erythrocytes. Urinary selenium has been shown to be a reliable marker for recent selenium intake rather than a robust marker for selenium status. Insufficient data currently exist that would allow for the prediction of a health effect from the concentration of selenium in hair or nails. Also the presence of selenium in hair or nails may indicate both external and internal exposures and there is a lack of reference or background ranges to help frame the interpretation of the results.

Selenium functional tests are plasma, whole blood, and platelet glutathione peroxidase activity (GSH-px) assays. The plasma GSH-px is a useful marker of selenium status in populations with low selenium intake; it responds rapidly to supplementation. Erythrocyte GSH-px has a plateau above which it is independent of selenium status. In addition, erythrocyte GSH-px responds slowly to depletion and supplementation. Platelet GSH-px responds rapidly to selenium dietary changes, accordingly, it is considered to be a sensitive indicator of changing selenium status. GSH-px can be measured with an enzyme assay or ELISA; commercial kits are available. The determination of selenoprotein P, which accounts for 50% of selenium in blood, is also a useful and sensitive test for selenium status, at least in populations with relatively low-to-moderate selenium intakes. Selenoprotein P can be detected by isolating the protein with chromatography followed by detection of selenium with ICP-MS.

Iodine

Iodine status is most commonly assessed by measuring urinary iodine using highly specific ICP-MS methods or technically simpler spectrophotometric methods. Urinary iodine reflects iodine intake within the past few days, but the marker is generally not useful to

Table 5 Tentative cutoff points for interpretation of biochemical laboratory indices

Nutritional status	*Biochemical indicator*	*Deficiency*	*Excess*
Protein	Serum albumin[a]	<30 g l^{-1}	
	Serum TTR[a]	<0.11 g l^{-1} (severe) 0.11–0.16 g l^{-1} (moderate)	
	Serum TF[a]	<1.0 g l^{-1} (severe) 1.5–2.0 g l^{-1} (mild)	
	Serum RBP[a]	<25 mg l^{-1}	
Fatty acids	Plasma essential fatty acids Holman Index[b]	>0.2 (ratio of triene [C20:3n−9 mead acid] to tetraene [C20:4n−6 arachidonic acid] fatty acids)	
	RBC membrane fatty acids Omega-3 Index[c] (marker of risk for sudden cardiac death)	Amount of EPA and DHA as a proportion of total fatty acids: <4% (high risk) 4–8% (intermediate risk) >8% (low risk)	
Vitamin A	Serum retinol[a]	<0.35 µmol l^{-1} (severe) <0.70 µmol l^{-1} (moderate) <1.05 µmol l^{-1} (suboptimal)	
	Serum retinyl esters[a]		>10% of total vitamin A (fasting) (potential hypervitaminosis A)
	Serum RBP[a]	<0.70 µmol l^{-1}; validation in different populations needed	
	Serum RDR[a]	>20% (marginal status)	
	Serum MRDR[a]	>0.060 (marginal status)	
Vitamin D	Serum 25OHD[d]	<30 nmol l^{-1} (deficient) 30–50 nmol l^{-1} (inadequate)	>125 nmol l^{-1} (reason for concern, possibility of hypervitaminosis D)
Vitamin E	Serum α-tocopherol[a]	<11.6 µmol l^{-1}	
Thiamin	Erythrocyte TDP[a]	<120 nmol l^{-1} (high risk) 120–150 nmol l^{-1} (marginal)	
	Urinary thiamin[a]	<27 µg g^{-1} creat (high risk) 27–65 µg g^{-1} creat (medium risk)	
	Erythrocyte EKT-AC[a]	>1.25	
Riboflavin	Erythrocyte FAD[e]	<270 nmol l^{-1} RBC	
	Urinary riboflavin[a]	<27 µg g^{-1} creat (high risk) 27–79 µg g^{-1} creat (medium risk)	
	Erythrocyte EGR-AC[a]	>1.40 (high risk) 1.2–1.4 (medium risk)	
Niacin	Urinary N′MN, 2-Py[a]	<0.5 mg g^{-1} creat	
Vitamin B_6	Serum PLP[a]	<20 nmol l^{-1}	
	Erythrocyte EAST-AC[a]	>1.85 (deficient) 1.70–1.85 (marginal)	
Folate	Serum folate[a]	<6.8 nmol l^{-1} (negative balance)	
	RBC folate[a]	<317 nmol l^{-1} (used frequently)	
	Plasma tHcy[a]	>12–14 µmol l^{-1} (used frequently)	
Vitamin B_{12}	Serum B12[a]	<74 pmol l^{-1} (deficient) 100–150 pmol l^{-1} (moderate) 100–300 pmol l^{-1} (low-to-normal)	
	Serum MMA[f]	>271 nmol l^{-1}	
Vitamin C	Serum total ascorbic acid[a]	<11.4 µmol l^{-1} (deficient) 11.4–23 µmol l^{-1} (low levels)	
Iron	Serum ferritin[g]	<12 µg l^{-1} (<5 years) <12 µg l^{-1} (≥5 years)	>150 µg l^{-1} (females) >200 µg l^{-1} (males)
	Serum sTfR[a]	Assay specific cutoff values	
	Serum TS[a]	<16%	>70%
	Erythrocyte protoporphyrin[a]	>80 µmol/mol heme (severe) 60–80 µmol/mol heme (moderate)	
	Hemoglobin[a]	<110 g l^{-1} (6–59 months) <115 g l^{-1} (5–11 years) <120 g l^{-1} (12–14 years)	

Table 5 Tentative cutoff points for interpretation of biochemical laboratory indices—cont'd

Nutritional status	*Biochemical indicator*	*Deficiency*	*Excess*
		<120 g l^{-1} (nonpregnant women) <110 g l^{-1} (pregnant women) <130 g l^{-1} (men)	
Zinc	Serum zinc[a]	Children <10 years: 9.9 μmol l^{-1} (collected AM) 8.7 μmol l^{-1} (collected PM) Males ≥10 years: 11.3 μmol l^{-1} (collected AM fasting) 10.7 μmol l^{-1} (collected AM other) 9.3 μmol l^{-1} (collected PM) Females ≥10 years: 10.7 μmol l^{-1} (collected AM fasting) 10.1 μmol l^{-1} (collected AM other) 9.0 μmol l^{-1} (collected PM)	
Selenium	Serum selenium	<0.1 μmol l^{-1} (severely depleted)	
Iodine	Urine iodine[a] (used as population statistic only; not useful as individual measure)	<20 μg l^{-1} (severe) 20–49 μg l^{-1} (moderate) 50–99 μg l^{-1} (mild) <150 μg l^{-1} (pregnant women)	>300 μg l^{-1} (excessive) 200–299 μg l^{-1} (more than adequate) >250 μg l^{-1} (more than adequate for pregnant women)

2-Py, N′-methyl-2-pyridone-5-carboxamine; 25OHD, 25-hydroxyvitamin D; creat, creatinine; DHA, docosapentaenoic acid; EAST-AC, erythrocyte aspartate aminotransferase activation coefficient; EGR-AC, erythrocyte glutathione reductase activation coefficient; EPA, eicosapentaenoic acid; ETK-AC, erythrocyte transketolase activation coefficient; FAD, flavinadenin-dinucleotide; MMA, methylmalonic acid; MRDR, modified relative dose response test; N′MN, N′-methylnicotinamide; PLP, pyridoxal-5′-phosphate; RDR, relative dose response test; RBC, red blood cells; RBP, retinol-binding protein; sTfR, soluble transferrin receptor; TDP, thiamin diphosphate; TF, transferrin; tHcy, total homocysteine; TS, transferrin saturation; TTR, transthyretin.

[a]Gibson RS (2005) *Principles of Nutritional Assessment*, 2nd edn. New York: Oxford University Press.

[b]Institute of Medicine (2005) *Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids (Macronutrients)*. Washington, DC: The National Academies Press.

[c]Harris WS (2008) The omega-3 index as a risk factor for coronary heart disease. *American Journal of Clinical Nutrition* 87(suppl), 1997S–2002S.

[d]Institute of Medicine (2011) *Dietary Reference Intakes for Calcium and Vitamin D*. Washington, DC: The National Academies Press.

[e]Institute of Medicine (1998) *Dietary Reference Intakes: Thiamin, Riboflavin, Niacin, Vitamin B_6, Folate, Vitamin B_{12}, Pantothenic Acid, Biotin, and Choline*. Washington, DC: The National Academies Press.

[f]Allen RH, Stabler SP, Savage DG, Lindenbaum J (1990) Diagnosis of cobalamin deficiency I: Usefulness of serum methylmalonic acid and total homocysteine concentrations. *American Journal of Hematology* 34, 90–98.

[g]WHO. Serum ferritin concentrations for the assessment of iron deficiency in populations. *Vitamin and Mineral Nutrition Information System*. Geneva, World Health Organization, 2011 (WHO/NMH/NHD/MNM/11.2) [cited 2012]. Available at: http://www.who.int/vmnis/indicators/serum_ferritin.pdf

classify intake sufficiency or deficiency in a person, but rather to define the risk of a population. If 24-h urine cannot be collected, iodine excretion can be expressed per gram of creatinine but only in areas with very low inter and intraindividual variation in urinary creatinine. In clinical settings, the measurement of uptake of radioactive iodine is used.

Iodine functional tests relate to its impact on thyroid function. The measurement of thyroid hormones and of thyroglobulin – a key precursor in the production of thyroid hormones – is performed by specific competitive radioimmunoassay or immunofluorimetric methods. When thyroid disease is not present, serum thyroxine (T4) can be a useful marker of iodine status in most age groups, whereas serum pituitary thyroid-stimulating hormone (TSH) is a good marker in pregnant and lactating women, but not useful in children and adolescents. Serum 3,5,3′triiodothyroxine (T3) is not a useful biomarker for iodine status. Serum thyroglobulin is a useful marker in children and adolescents, but not useful during pregnancy and lactation. Infection has a confounding effect on iodine status because the synthesis of TTR – a thyroid hormone transporter – is markedly suppressed.

Choice of Laboratory Tests

The choice of laboratory tests depends on the type of study to be carried out. In field nutritional epidemiology studies, particularly in developing countries, the number and type of tests will be mainly limited by the specimen volume, the local laboratory infrastructure, and the availability of skilled personnel and financial resources. The lowest complexity profile of testing achievable with a spot urine sample and an EDTA whole blood sample from a finger stick covers representatives of iodine, iron, vitamin A, and folate

status, including parameters of infection and inflammation: urinary iodine (spectrophotometric method), Hb (portable point-of-care instrument), serum ferritin/sTfR/RBP/CRP/AGP (ELISA), and RBC folate (microbiologic assay). The measurement of additional vitamins and biochemical indicators generally requires a larger volume of blood that can only be obtained through venipuncture and requires the availability of more complex laboratory tests.

In population studies carried out in developed countries with high-level laboratory facilities, the selection of laboratory tests depends on the purpose of the study, specimen volume, and financial resources. Because certain plasma proteins and urine creatinine are part of standard biochemical profiles on fully-automated clinical analyzers, they can be easily determined. Fatty acid distribution profiles are rarely assessed in population studies. However, serum cholesterol, triacylglycerols and lipoprotein fractions are typically measured. These tests are available on fully automated clinical analyzers, have been largely standardized and are sometimes required for normalization of lipophilic compounds. The selection of micronutrient tests can be determined by the suspected deficiencies from previous dietary surveys and by the need to assess the impact of nutritional interventions such as fortification. A representation of micronutrients covered by the recent NHANES can be found in **Table 3**. Several analytical methods evolved in the last few years from less specific immunoassays to highly specific mass spectrometry-based techniques.

In a hospital setting, the selection of laboratory tests depends on the clinical condition of the patient at admission and during the subsequent course of injury or illness. Among hospital and institutionalized patients, deficiencies in proteins, vitamins, and trace elements are common due to underlying medical disease; suspicion of deficiency is based on history and physical examination.

Regardless of the setting, two other considerations deserve mention. First, the precision of the laboratory test largely influences the minimum detectable difference on repeat measurements or the ability to distinguish between healthy and diseased individuals or populations. It is therefore desirable to select laboratory tests with the highest achievable precision. Second, most biochemical indices do not have the required sensitivity and specificity to be used solely in the diagnosis of an abnormality. It is therefore recommended to combine findings from dietary intake assessment with static biochemical indices and functional tests whenever possible.

Most methods used to assess nutritional status have not been standardized yet which can lead to considerable differences among laboratories and methods. Where available, the selection of high-order reference methods should be favored or, in their absence, carefully validated methods with regards to sample collection, processing, and analysis should be used. An appropriate quality control (QC) system with internal and external verifications should be in place throughout the entire study. Particularly for longitudinal studies, large batches of in-house prepared QC pools are preferred to commercial QC samples where frequent lot changes can be expected. It is suggested to prepare two or three levels of QC pools and to analyze them in every assay together with the patient samples. Participation in external quality assessment programs or interlaboratory cross-comparisons as well as the regular use of reference materials for calibration verification is highly recommended. **Table 4** provides information on currently available international reference materials and gives a selected list of external quality assessment programs. The Centers for Disease Control and Prevention (CDC) also maintains an inventory of external quality assessment programs by country, some of which pertain to nutritional status indicators.

Evaluation of Laboratory Indices

In general, reference values are population specific; accordingly, each major laboratory in homogenous areas has to derive them from a clinically healthy reference population selected with very specific criteria. These values should preferably be given in percentiles. The National Report on Biochemical Indicators of Diet and Nutrition in the US Population 1999–2002 is a comprehensive CDC publication that offers nationally representative reference information for 27 nutritional indicators derived from NHANES. A second edition of the report, providing reference information for 58 nutritional indicators from NHANES 2003–2006, as well as information on prevalence of nutritional deficiencies for selected indicators, is released in 2012.

Ideally, cutoff points are derived by determining the biochemical values that correspond to the earliest determinable physiological, metabolic, functional, and morphological alterations. Because such an approach has been followed only in very few cases, most cutoff points have been derived statistically from reference values and should therefore be considered as tentative. **Table 5** presents a list of tentative cutoff points for interpretation of nutritional status tests. In some cases, different cutoff points are used for children, pregnant and lactating women, and the elderly. These values can be found in reference texts. It is important to remember that cutoff points as well as reference intervals can vary with the method used to measure the biochemical indices. Continued efforts to standardize methods are therefore needed.

Conclusions and Future Directions

Over the last few years, scientific and public health agencies initiated renewed efforts to better define biomarkers of nutritional status, to focus on analytical method standardization for accurate and precise measurement of biomarkers, and to develop systematic processes for the evaluation of biomarkers and surrogate endpoints. The Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health is leading an effort to build a consensus around biomarkers of nutrition for development (BOND). The goal of BOND is to promote the discovery, development, and use of biomarkers across a range of applications and to harmonize the global health community's decision-making about what biomarkers are best suited

for a given use under specific conditions and settings. In collaboration with CDC and NIH, the National Institute of Standards and Technology (NIST) has developed over the last several years reference methods and materials for nutritional biomarkers that are essential to enhance and promote high quality laboratory measurements. The development and validation of new dietary biomarkersare also constantly evolving. The emerging field of nutritional metabonomics – the study of metabolic responses – is receiving increasing attention as an analytical method to assess metabolic profiles as measures of dietary exposures and indicators of dietary patterns, dietary changes, or effectiveness of dietary interventions.

Yet, with all these efforts ongoing, the findings of the EURRECA project to assess potential biomarkers of micronutrient status by using a systematic review methodology were that far as fewer studies were available for biomarker assessment than initially predicted, and the risk of bias of included studies was greater than expected. The authors concluded that further research is needed to assess the usefulness of many potential biomarkers and that we still have to overcome gaps in our understanding of how well a biomarker works in particular population groups or in people with different baseline micronutrient status.

Further Reading

Bowman, B.A., Russell, R.M (Eds.), 2006. Present Knowledge in Nutrition, 9th edn. ILSI Press, Washington.

Burtis, C.A, Ashwood, E.R., Bruns, D.E. (Eds.), 2006. Tietz Textbook of Clinical Chemistry and Molecular Diagnostics, 4th edn. Saunders Elsevier, St. Louis.

Ciccimaro, E., Blair, I.A, 2010. Stable-isotope dilution LC-MS for quantitative biomarker analysis. Bioanalysis 2 (2), 311–341.

Eitenmiller, R.R, Landen, W.O (Eds.), 1999. Vitamin Analysis for the Health and Food Sciences. CRC Press, Boca Raton.

Gibson, R.S, 2005. Principles of Nutritional Assessment, 2nd edn. Oxford University Press, New York.

Hooper, L., Ashton, K., Harvey, L.J., Decsi, T., Fairweather-Tait, S.J., 2009. Assessing potential biomarkers of micronutrient status by using a systematic review methodology: methods. American Journal of Clinical Nutrition 89 (suppl), 1953S–1959S. This supplement entitled "Biomarkers of micronutrient status: EURRECA Workshop" contains eight additional biomarker articles.

Institute of Medicine (2010) Evaluation of biomarkers and surrogate endpoints in chronic disease, http://www.iom.edu/Reports/2010/Evaluation-of-Biomarkers-and-Surrogate-Endpoints-in-Chronic-Disease.aspx.

Jenab, M., Slimani, N., Bictash, M., Ferrari, P., Bingham, S.A., 2009. Biomarkers in nutritional epidemiology: applications, needs and new horizons. Human Genetics 125, 507–525.

Nordberg, G.F., Fowler, B.A., Nordberg, M., Friberg, L.T. (Eds.), 2007. Handbook on the Toxicology of Metals, 3rd edn. Academic Press, San Diego.

Raiten, D.J., Namaste, S., Brabin, B., et al., 2011. Executive summary: Biomarkers of nutrition for development (BOND): Building a consensus. American Journal of Clinical Nutrition (in press).

Wu, A.H.B, 2006. Tietz Clinical Guide to Laboratory Testing, 4th edn. Saunders Elsevier, St. Louis.

Young, D.S., 2007. Effects of Preanalytical Variables on Clinical Laboratory Tests, 3rd edn. AACC Press, Washington.

Zempleni, J., Rucker, R.B., McCormick, D.B., Suttie, JW (Eds.), 2007. Handbook of Vitamins, 4th edn. CRC Press Taylor & Francis Group, Boca Raton.

Relevant Websites

https://www.cdc.gov/mlp/pdf/EQA/eqa_list.pdf – CDC Inventory of external quality assessment programs by country.

http://www.cdc.gov/nutritionreport – CDC National Report on Biochemical Indicators of Diet and Nutrition in the US Population.

http://irmm.jrc.ec.europa.eu/reference_materials_catalogue/Pages/index.aspx – European Commission Joint Research Center - Institute for Reference Materials and Measurements (IRMM).

http://www.nibsc.ac.uk/products/biological_reference_materials.aspx – United Kingdom National Institute for Biological Standards and Control (NIBSC).

http://www.nist.gov/srm/index.cfm – United States National Institute of Standards and Technology (NIST).

Body composition

M. Ashby-Thompson[a], S. Chung[b], and D. Gallagher[a], [a] Columbia University, New York, NY, United States; and [b] Department of Pediatrics, Konkuk University Medical Center, Konkuk University School of Medicine, Seoul, South Korea

© 2023 Elsevier Ltd. All rights reserved.

This is an update of D. Gallagher, S. Chung, M. Akram, Body Composition, In Encyclopedia of Human Nutrition (Third Edition), edited by Benjamin Caballero, Academic Press, 2013, https://doi.org/10.1016/B978-0-12-375083-9.00027-1.

Introduction 21
Body composition determination 21
DXA 21
Hydrodensitometry/air plethysmography 22
Dilution techniques 22
Whole-body counting 22
Magnetic resonance imaging (MRI) 22
Quantitative magnetic resonance (QMR) 22
Bioimpedance analysis (BIA) 23
Anthropometry 23
Body mass index (BMI) 23
***In Vivo* neutron activation** 23
Three-dimensional photonic scanner (3DPS) 23
Models in body composition 24
Tissues and organs 25
Body composition applications during growth 25
Body composition applications during aging 27
Physiological application: two examples 28
Example 1 28
Example 2 31
Conclusion 32
Acknowledgment 32
References 32
Further reading 34

Key points

- Synopsis
- Introduction
- Body Composition Determination
- Dual-Energy X-Ray Absorptiometry (DXA)
- Hydrodensitometry/Air Plethysmography
- Dilution Techniques
- Whole-Body Counting
- Magnetic Resonance Imaging/Computed Tomography
- Quantitative Magnetic Resonance
- Bioimpedance Analysis (BIA)
- Anthropometry
- Body Mass Index (BMI)
- *In Vivo* Neutron Activation
- Three-Dimensional Photonic Scanner (3DPS)
- Models in Body Composition
- Tissues and Organs
- Body Composition Applications During Growth
- Body Composition Applications During Aging
- Physiological Application: Two Examples
- Example 1

- Example 2
- Conclusion
- Acknowledgments
- Further Reading

Introduction

Historically, the measurement of the body and its components centered around cadaver analyses where specific tissues and organs were extracted from the body for inspection (Forbes, 1987). The extraction of tissue samples from the living body was a step forward in allowing for the analyses of tissue morphology in a state more closely resembling the *in vivo* state. However, both cadaver and *in vitro* tissue analyses are subject to inaccuracies when extrapolations are being made to the living body. Nevertheless, much of our understanding of human body composition in both children and adults has roots in these approaches. During the twentieth century, significant advances were made in the development of *in vivo* methods of body composition analysis thanks to the disciplines of Physics, Engineering, and Medicine. Methodologies with minimal or no risk to the participant have allowed for the assessment of body composition in growth and development, aging, and disease. Refinements of existing methodologies continue now in the 21st century as do efforts to show linkages between the non-invasive quantification of adipose tissue depots and infiltration into tissues, skeletal muscle quality and quantity with metabolic health and with diseases rooted in metabolic dysregulation.

The physiological significance of knowing the composition of the body greatly depends on the question of interest. Common applications involving medical/clinical diagnoses include osteopenia/osteoporosis; muscle wasting; sarcopenia; lipodystrophy; altered states of hydration; malnutrition; obesity. There are also metabolic consequences (for example, insulin resistance) associated with high and low levels of body fat and where the fat is distributed. From a nutritional perspective, the interest in body composition has increased multifold with the global increase in the prevalence of obesity and its complications. This article will focus on our current state of body composition knowledge and how this knowledge was determined with the available most advanced methodologies.

Body composition determination

There is no single gold standard for body composition measurements *in vivo*. All methods incorporate assumptions that do not apply in all individuals and the more accurate models are derived using a combination of measurements, thereby reducing the importance of each assumption. The most commonly used technique today with good reproducibility in children and adults is dual-energy X-ray absorptiometry.

DXA

The DXA method evolved from earlier single and dual photon absorptiometry methods for evaluating bone mineral. DXA systems share in common an X-ray source that, after appropriate filtration, emits two photon energy peaks. The attenuation of the two energy peaks relative to each other depends on the elemental content of tissues through which the photons pass. Bone, fat, and lean soft tissues are relatively rich in calcium/phosphorus, carbon, and oxygen, respectively. DXA systems are designed to separate pixels, based on appropriate models and relative attenuation, into these three components. There are no known factors, including hydration effects that significantly influence the validity of DXA fat and bone mineral estimates. Excessive or reduced fluid volume would be interpreted as changes in lean soft tissue. The radiation exposure is minimal and can be used in children and adults of all ages. DXA measures in persons who fit within the DXA field-of-view have good reproducibility for total body and regional components. A method using half-body scans to measure persons whose body dimensions fall outside the DXA field-of-view has been described (Tataranni and Ravussin, 1995). In studies evaluating adults with obesity, significant differences between whole- and half-body scans were found for percentage fat mass and bone mineral content (BMC) (Lundqvist et al., 2009; Tataranni and Ravussin, 1995; Rothney et al., 2009). In one study comparing whole- and half-body scans to ADP before and after weight loss, whole- and half-body scans measured percent body fat differently compared to ADP before weight loss for percent body fat, however, that difference did not persist after weight loss when the participants fit entirely within the DXA field-of-view (Lundqvist et al., 2009). Among children and adolescents (ages 7–18 years) with obesity, there were no significant differences in percent fat, total mass, fat mass, lean mass and BMC when comparing whole- and half-body scans (Breithaupt et al., 2011). Among children and adolescents ages 4–20 years with wide BMI range and considering Tanner Stage, half-body scans produced very small differences for bone mineral density and bone mineral content (Ferreira et al., 2019). The difference was not considered clinically meaningful. Half-body scans have not been validated in normal weight adults. From three studies that compared scans from the right compared to the left side of the body, one study reported a difference in percent fat, fat mass, fat free mass and BMC (Ravussin and Bogardus,

1989); the second reported differences only in BMC (Rothney et al., 2009) and the third found no differences between the sides (Ferreira et al., 2019). The authors concluded that the differences were not large enough to affect accuracy and that the use of half-body scans could introduce bias, depending on whether scans of the right half versus the left half of the body is scanned. They agree that in theory, true anatomical differences between right and left side (lateral distribution and physical characteristics of visceral organs) may lead to bias.

Advancements in DXA analysis software has allowed for the estimation of visceral adipose tissue (VAT) from a whole-body DXA scan; versions are available from the major DXA manufacturers. Estimating VAT by DXA has the advantage of faster scan and analysis time, lower cost, less radiation exposure and the ability to accommodate a wider range of individuals compared to MRI or CT, the standards for measuring VAT. Presently, the tool produces variable results—underestimating at lower VAT volumes and overestimating at higher volumes among adults and adolescents with normal weight, overweight or obesity and overestimating volumes in children with normal weight, overweight and obesity. There is a lack of validity for measuring change in VAT.

Hydrodensitometry/air plethysmography

One of the oldest methods of measuring body composition, the determination of body volume by water displacement (Archimedes principle) allows for the estimation of fat-free mass (FFM) density (where an assumption is made that densities of fat and FFM are constant) from which percent body fat is calculated using a two-compartment body composition model. Today there are a number of additional methods for measuring body volume, including air displacement plethysmography. Limitations with this approach include the assumptions of stable densities of fat and FFM across the age range where this may not be true in older individuals and across race/ethnic groups (Schutte et al., 1984).

Dilution techniques

As fat is relatively anhydrous, the body's water is found primarily in the body's FFM compartment where approximately 73% of a healthy nonobese adults FFM compartment is water. The body's water pool can be measured using tracers which after administration, dilute throughout the body. Basic assumptions involved with tracer dilution for body composition determination include equal distribution throughout the pool of interest and dilution is complete within a specific period of time without any loss. Examples of commonly used tracers include deuterium oxide for total body water and sodium bromide for extracellular water. These isotope dilution techniques allow for the evaluation of fat and FFM without making the assumption that the hydration of FFM is constant and therefore stable.

Whole-body counting

A small constant percentage of total body potassium (TBK) is radioactive (^{40}K) and emits a g-ray. With appropriate shielding from background, this g-ray can be counted using scintillation detectors. As the ratio of ^{40}K–^{39}K is known and constant, ^{39}K or "total body potassium" can be estimated accordingly. All of the body's potassium is within the FFM compartment and the proportion of the body FFM compartment TBK/FFM ratio is relatively stable in the same subject over time and between different subjects. However, with increasing age or when comparing young versus elderly, the TBK to FFM ratio decreases. Although the specific mechanism(s) associated with this decrease in the TBK to FFM ratio with increasing age is/are unclear, it could be explained by a small but consistent increase in extracellular fluid compared to intracellular fluid.

Magnetic resonance imaging (MRI)

The use of MRI has resulted in important advances in body composition phenotyping. MRI studies are safe and instruments are available in most hospital or related facilities. Expense is a limiting factor. The importance of MRI is that this method acquires cross-sectional images of the body at predefined anatomic locations. Image analysis software then allows estimation of the adipose tissue, skeletal muscle, and organs based on pixel intensity. Acquiring images at predefined intervals and integrating the area between slices allows reconstruction of an entire organ of interest such as skeletal muscle mass. A significant advancement made possible by these imaging methods has been the characterization of a tissue distribution, such as adipose tissue where it is now possible to quantify visceral, subcutaneous, and intermuscular depots at the regional and whole-body level.

Quantitative magnetic resonance (QMR)

QMR (EchoMRI-AH; Echo Medical Systems, Houston, TX) is a non-imaging system that relies on proton nuclear magnetic resonance to measure human body composition (Gallagher et al., 2010). Using various pulse sequences, the QMR system provides estimates of fat mass, lean tissue mass, free water, and total body water. The QMR approach has important advantages over currently

available methods as it provides body composition estimates with high precision and without the use of ionizing radiation. The system for infants enables measurements in subjects from birth up to 12 kg, while the adolescent system optimizes measurement of subjects up to 80 kg. The adult system, moreover, can accommodate subjects up to 250 kg, almost double that of the widely used DXA approach. Human validation studies revealed systematic differences in body-fat estimates between the QMR and other available research methods (Napolitano et al., 2008; Gallagher et al., 2010; Andres et al., 2011). A validation study in infants found small differences between measured weights and fat, free water, and total water measured by QMR (Toro-Ramos et al., 2017). Further exploration is needed to identify these sources of and potential correction approaches for these differences. QMR accurately detected small changes in simulated body composition, and did not differ from deuterium dilution estimate of TBW.

Bioimpedance analysis (BIA)

BIA is a simple, inexpensive, and noninvasive body composition measurement method. BIA is based on the electrical conductive properties of the human body. Measures of bioelectrical conductivity are proportional to total body water and the body compartments with high water concentrations such as fat free and skeletal muscle mass. BIA assumes that the body consists of two compartments, fat and FFM (Body weight = Fat + FFM). Single frequency BIA (50 kHz) allows the estimation of TBW from which FM and percent body fat are derived. Single frequency BIA has also been used for estimating skeletal muscle mass. Multifrequency BIA uses different frequencies, ranging from low (1 kHz) to high (1000 kHz) to estimate TBW, FFM, FM, and intracellular water (detected by frequencies >50 kHz) and extracellular water (detected by frequencies 1–5 kHz) compartments. The equations developed to estimate body composition by any BIA system are population specific such that they are most valid in populations similar to the population in which a specific equation was developed. The validity of BIA in persons with severe obesity is questioned as TBW and extracellular water relative to TBW are both greater in subjects with obesity compared with normal-weight individuals.

Anthropometry

For routine clinical use, anthropometric measurements (circumference measures and skinfold thickness) have been preferred due to ease of measurement and low cost. Waist circumference and the waist-hip ratio measurements are commonly used surrogates of fat distribution, especially in epidemiology studies. Waist circumference is highly correlated with visceral fat and was recently included as a clinical risk factor in the definition of the metabolic syndrome. Specifically, waist circumferences greater than 102 cm (40 in) in men and greater than 88 cm (35 in) in women are suggestive of elevated risk.

Skinfold thicknesses which estimate the thickness of the subcutaneous fat layer are highly correlated with percent body fat. Because the subcutaneous fat layer varies in thickness throughout the body, a combination of site measures is recommended, reflecting upper and lower body distributions. Predictive percent body fat equations based on skinfold measures are age and sex specific in adults and children.

Body mass index (BMI)

The body mass index (BMI = weight (kg)/height (m^2)) continues to be the most commonly used index of weight status, where normal weight is a BMI 18.5–25.9 kg m^{-2}; overweight is a BMI 25.0–29.9 kg m^{-2}; obese a BMI > 30.0 kg m^{-2}. BMI is a commonly used index of fatness due to the high correlation between BMI and percent body fat in children and adults. The prediction of percent body fat is dependent on age (higher in older persons), sex (higher in males), and race (higher in Asian compared to African American and Caucasian).

In Vivo neutron activation

Nitrogen, carbon, hydrogen, phosphorus, sodium, chlorine, calcium, and oxygen are all measurable *in vivo* by method known as neutron activation analysis. A source emits a neutron stream that interacts with body tissues. The resulting decay products of activated elements can be counted by detectors and elemental mass established. Carbon, nitrogen, and calcium can be used to estimate total body fat, protein, and bone mineral mass using established equations. Neutron activation analysis is uniquely valuable in body composition research as there are no known age or sex effects of currently applied equations. However, facilities that provide these techniques are limited.

Three-dimensional photonic scanner (3DPS)

3DPS is noninvasive optical method used to generate a 3-dimensional (3-D) photonic image of an object. High-resolution 3DPS systems have four photonic image production units or cameras mounted on four corner poles (Ashby-Thompson et al., 2020). Each

unit has an eye-safe class-1 laser-light source (664 nm) and a high-speed, high-resolution digital camera. The scan field is 200 cm high, 100 cm wide, and 60 cm deep, generating over 2 million data points. Subjects must wear a tight-fitting cap to minimize air spaces between the hair and skull and well-fitting underwear that cling to the skin surface during scan acquisition. Subjects are positioned in a standardized position such that arms are abducted from the trunk, and no contact between the legs and must remain motionless during the 10 s of scan. 3DPS can generate hundreds of body circumferences, segment lengths, total and regional body volumes, total and regional body surface areas with high precision. High-resolution 3DPS systems have limited availability due to cost. Advances in technology and demand for personalized health assessment have resulted in a growth of laser-based imaging technologies and associated applications. These technologies show promise but cannot yet replace the high-resolution models.

Models in body composition

The use of models in the assessment of body composition allows for the indirect assessment of compartments in the body. Typically, a compartment is homogenous in composition (e.g., fat), however, the simpler the model the greater the assumptions made and the greater the likelihood of error. The sum of components in each model is equivalent to body weight (Fig. 1). These models make assessments at the whole-body level and do not provide for regional or specific organ/tissue assessments.

The basic two-compartment (2C) model (Table 1) is derived from measuring the density of FFM by hydrodensitometry and subtracting FFM from total body weight thereby deriving fat mass (body weight − FFM = fat mass). FFM is a heterogeneous compartment consisting of numerous tissues and organs. A 2C approach becomes inadequate when the tissue of interest is included within the FFM compartment. Nevertheless, the 2C model is routinely and regularly used to calculate fat mass from hydrodensitometry, total body water, and total body potassium.

A three-compartment (3C) model consists of fat, fat-free solids, and water. The water content of FFM is assumed to be between 70% and 76% for most species and results from cross-sectional studies in adult humans show no evidence of differences in the hydration of FFM with age. The fat-free solids component of FFM refers to minerals (including bone) and proteins. The 3C approach involves the measurement of body density (usually by hydrodensitometry) and total body water by an isotope dilution technique. Assumptions are made that both the hydration of FFM and the solids portion of FFM are constant. Because bone mineral content is known to decrease with age, the 3C approach is limited in its accuracy in persons or populations where these assumptions are incorrect.

A four-compartment (4C) model involves the measurement of body density (for fat), total body water, bone mineral content by DXA, and residual (residual = body weight − (fat + water + bone)). This model allows for the assessment of several assumptions that are central to the 2C model. The 4C approach is frequently used as the criterion method against which new body composition methods are compared in both children and adults.

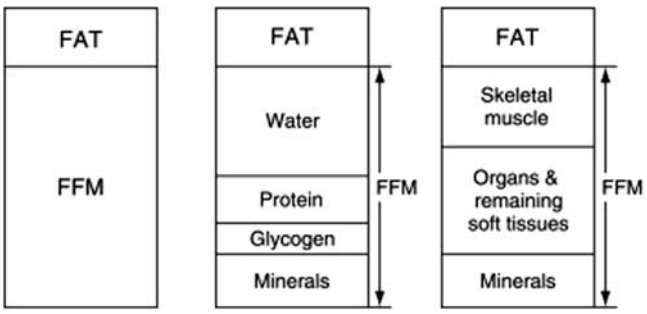


Fig. 1 Three different models for characterizing body composition compartments. Components are as labeled: FFM, fat-free body mass.

Table 1 Multicompartment body composition models.

Model	*Equations for % fat*	*References*
2C	$100 \times (4.971/D_b - 4.519)$	(Behnke et al., 1995)
3C	$100 \times (2.118/D_b - 0.78 \times (TBW/W) - 1.354)$	(Siri, 1961)
4C	$100 \times (2.747/D_b - 0.727 \times (TBW/W) + 1.146 \times (BMC/W) - 2.0503)$	(Boileau et al., 1985)
6C	$100 \times (2.513/D_b - 0.739 \times (TBW/W) + 0.947 \times (TBBM/W) - 1.79)$	(Heymsfield et al., 1996)

Db, body density; TBW, total body water; W, body weight; BMC, bone mineral content; TBBM, total body bone mineral.

The more complex 4C model involves neutron activation methods for the measurement of total body nitrogen and total body calcium, where total body fat = body weight − (total body protein (from total body nitrogen) + total body water (dilution volume) + total body ash (from total body calcium)). A six-compartment model is calculated as follows: fat mass (measured from total body carbon) = body weight − (total body protein + total body water + bone mineral + soft tissue mineral (from a combination of total body potassium, total body nitrogen, total body chloride, total body calcium) + glycogen (total body nitrogen) + unmeasured residuals). However, the availability of neutron activation facilities is limited and therefore the latter models are not readily obtainable by most researchers.

At the organizational level, a five-level model was developed where the body can be characterized at five levels (Wang et al., 1992). The following are the levels and their constituents: atomic = oxygen, carbon, hydrogen, and other (level 1); molecular = water, lipid, protein, and other (level 2); cellular = cell mass, extracellular fluid, and extracellular solids (level 3); tissue-system level = skeletal muscle, adipose tissue, bone, blood, and other (level 4); whole body (level 5).

Tissues and organs

The aforementioned models do not allow for subregion or specific organ and tissue measurements. For example, skeletal muscle mass (SM) is contained within the FFM compartment. SM represents the single largest tissue in the adult body and is equivalent to ~40% of body weight in young adults, decreasing to ~30% of young values at elderly ages. SM is one of the more difficult components to quantify. Estimates of SM are commonly derived from anthropometry, total body potassium, and DXA using modeling approaches previously described. The use of MRI in body composition research has allowed for a good estimation of SM, adipose tissue (AT), and select organs *in vivo*, in all age groups with no risk to the participant (Fig. 2). Moreover, AT distribution, including subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT), and intermuscular adipose tissue (IMAT) is also measurable using a whole-body multislice MRI protocol (Fig. 3). In studies relating body composition to energy expenditure, high metabolic rate organs including liver, kidneys, heart, spleen, and brain are also measurable using MRI.

Bone mineral content and bone mineral density of specific body sites (for example, radius, hip, lumbar spine) are most commonly measured using DXA. Bone mass and microarchitecture are important determinants of bone strength, with microarchitectural deterioration being one of the specific changes associated with osteoporosis. Using high-resolution microcomputed tomography (micro-CT) and computer software, detailed analysis of three-dimensional (3D) architecture is feasible and allows microstructural 3D bone information to be collected.

Body composition applications during growth

As many health and disease outcomes have origins in early childhood, measurement of body composition during infancy can provide important insights into the longitudinal associations between body composition, physiologic and metabolic processes and these health and disease outcomes throughout the lifespan. Body composition methods at this age include anthropometry (recumbent length, stature, weight, circumferences and skinfold thicknesses), BIA, MRI, multicompartment models including TBW by isotope dilution, body density by air-displacement plethysmography (ADP) and BMC by DXA. Limitations which may impact accuracy of these measurements in infancy have been identified and include population-specific prediction equations (anthropometry and BIA), movement during testing (ADP or MRI) and significant technical examiner expertize (MRI) (Gallagher et al., 2020). Emerging areas of relevance for measuring body composition in infancy include the use of metabolomics (interrogating relationships between maternal metabolome to newborn size and body composition), discovery-based genetics studies (to define the genetic architecture of body composition during childhood and the development of polygenic risk scores or other

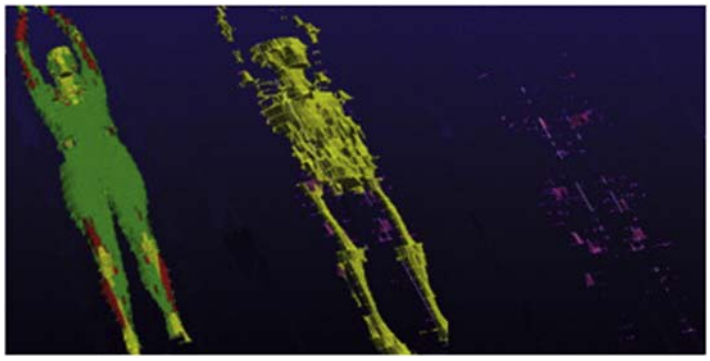


Fig. 2 3D reconstructed image of whole-body scan (from MRI). Skeletal muscle (red); adipose tissue (green); bone, organs, and residual (yellow); intermuscular adipose tissue (pink).

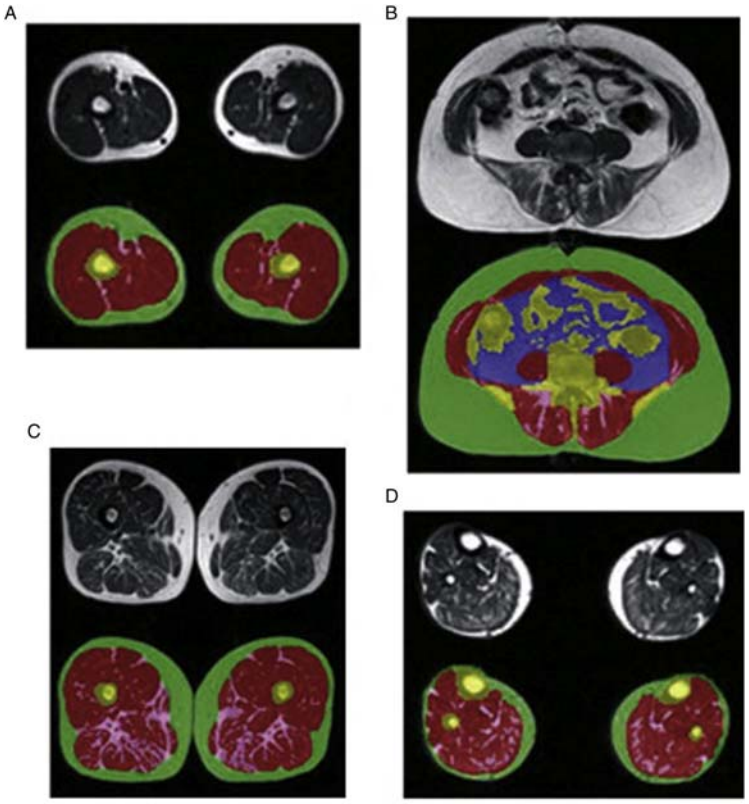


Fig. 3 Cross-sectional images from (A) upper arm, (B) trunk (L4–L5 level), (C) mid-thigh, and (D) mid-calf in an elderly female volunteer. IMAT, intermuscular adipose tissue (pink); SM, skeletal muscle (red); SAT, subcutaneous adipose tissue (green); VAT, visceral adipose tissue (blue).

approaches at birth to predict subsequent body composition and to identify children at risk for greater adiposity), epigenetics, and D3-creatine dilution as an index of muscle mass.

Skeletal muscle mass has a central role in intermediary metabolism, aerobic power, and strength. Its mass increases as a portion of body weight during growth, accounting for 21% at birth and 36% at adolescence. The essential role of skeletal muscle in many physiologic processes throughout the lifespan makes understanding of factors affecting it significant. The greater incidence of type 2 diabetes mellitus in adolescents in the US (particularly in girls from minority populations) and in Japan makes evaluation of race and sex differences in pediatric skeletal muscle mass (and adipose tissue or fat mass) especially important. Identification and characterization of differences could form the basis for further investigation of the associated metabolic implications.

Race differences in SM are known to exist as early as birth (Paley et al., 2016). African-Americans have greater limb lean tissue mass compared to Asian and Caucasian children, although Caucasian children have greater amounts than Asians throughout Tanner stages 1–5 (Song et al., 2002). Race differences in total body bone mineral content adjusted for total body bone area, age, height, and weight have been reported in prepubertal African–American, Asian, and Caucasian females and males. African-American children had greater total body bone mineral content than Asian and Caucasian children, although differences between Asian and Caucasian children are less clear. Collectively, these findings suggest that the proportions of specific FFM subcomponents may differ by race. Although mechanisms leading to bone and skeletal muscle differences between races are not well understood, endocrine factors may be involved.

Sex differences in FFM have been reported from birth throughout childhood with females having smaller amounts than males (Song et al., 2002). Total body bone mineral content is less in Tanner 1 females compared to males in African-Americans, Asians, and Caucasians. The mechanism for this sex difference in unclear. Gonadal steroids are significant mediators of adult sexual dimorphism of body composition, including fat-free soft tissues. Prepubertal females have higher concentrations of circulating estradiol than prepubertal males, and gonadotropin and gonadal steroids increase gradually in both males and females from the age of 5 years. Thus, prepuberty is a period with sex differences in circulating concentrations of sex steroids and of changes in these concentrations with advancing age. The earlier skeletal maturation of females, for example, has been attributed to the greater estradiol level in females compared to males. However, nonhormonal (possibly genetic) mechanisms may also play a role.

Fat or adipose tissue distribution is recognized as a risk factor for cardiovascular disease in both adults and children. An android or male fat pattern, with relatively greater fat in the upper body region, is associated with negative metabolic predictors whereas a gynoid or female fat pattern, with relatively greater fat in the hip and thigh areas, is associated with less metabolic risk. More and more studies are showing that the metabolic syndrome develops during childhood and is highly prevalent among overweight

children and adolescents. Although the concept of the metabolic syndrome referred initially to the presence of combined risk factors including VAT, dyslipidemia, hypertension, and insulin resistance in adults, it is now known to exist in children, especially where obesity and/or higher levels of VAT are present. Although sex-specific patterns of fat distribution had previously been thought to emerge during puberty, sex and race differences in fat distribution are now known to exist in prepubertal children. The implications are that a specific body composition pattern may differ by sex and race. An example is the relationship of blood pressure to central fat distribution in boys compared to girls where a significant positive relationship between trunk fat and blood pressure was reported in boys but not girls, and was independent of race, height, weight, and total body fat (He et al., 2002). Understanding the predictors of blood pressure in children is important because childhood blood pressure has been shown to track into adulthood in longitudinal studies. Children whose blood pressure levels were in the highest quintile, were two times more likely to be in the highest quintile 15 years later. Identification of clinically useful body composition measures would allow for the identification of children at increased risk for hypertension, who could benefit from monitoring.

Race differences in fat distribution among prepubertal Asians, African–Americans, and Caucasians also exist (He et al., 2004). Previous reports in adolescents have suggested significantly smaller hip circumferences in Asian females at all pubertal stages compared to Caucasians and Hispanics and greater trunk subcutaneous fat in Asian females compared to Caucasians. Differences in subcutaneous fat mass and fat distribution in Asian compared to Caucasian adults have also been described (Hoffman et al., 2005). Understanding the sex- and race-specific effects of puberty on regional body composition may help delineate the developmental timing of specific health risk associations.

Body composition applications during aging

During the adult life span, body weight generally increases slowly and progressively until about the seventh decade of life, and thereafter, declines into old age. An increased incidence of physical disabilities and comorbidities is likely linked to aging-associated body composition changes. Characterization of the aging processes has identified what is now considered a muscle disease (muscle failure), characterized by low muscle strength and associated with low muscle quantity and quality and collectively defined as "sarcopenia" (Cruz-Jentoft et al., 2019). This new definition places low muscle strength as the principal determinant because it is expected to facilitate prompt diagnosis and intervention in clinical practice and to better predict adverse outcomes. Low muscle quantity and quality remain the focus in research settings, however, there are differences in how these concepts are operationalized and they are difficult to measure accurately. Muscle quantity or mass can be estimated by MRI, CT, DXA, and BIA. Muscle quality, a relatively new term, refers to micro- and macroscopic changes in muscle structure and composition, and to muscle function delivered per unit of muscle mass. Imaging techniques such as MRI and CT and BIA phase angle measurements have been used to assess muscle quality. Muscle quality has also been defined by ratios of muscle strength to appendicular skeletal muscle mass or muscle volume.

Little is known about the overall rate at which sarcopenia develops in otherwise healthy elderly subjects, if this rate of progression differs between women and men, and the underlying mechanisms responsible for age-related sarcopenia. Peak SM mass is attained in the young adulthood years and slowly declines thereafter. During the latter adult years, SM decreases more rapidly as body fat becomes more centralized. Anthropometric equations have been developed for predicting appendicular skeletal muscle (ASM = SM of the limbs) in the elderly where sarcopenia was defined as ASM (kg)/height^2 (m^2) less than two standard deviations below the mean of the young reference group (Baumgartner et al., 1998). In the elderly men, the mean ASM/height^2 was approximately 87% of the young group. The corresponding value in women was approximately 80%. **Table 2** shows the estimated prevalences of sarcopenia in the same survey sample for each ethnic group, by age and sex. The same authors have reported that obese and sarcopenic persons have worse outcomes than those who are nonobese and sarcopenic.

Even in healthy, weight-stable elderly persons, changes in body composition over a 2-year period can include decreases in SM mass and bone mineral content with corresponding increases in IMAT and VAT, after adjusting for their baseline values, despite no detectable changes in physical function or food intake (Gallagher et al., 2000).

In adults, excess abdominal or VAT is recognized as an important risk factor in the development of coronary heart disease and noninsulin dependent diabetes mellitus. Waist circumference and the waist:hip ratio are commonly used to predict visceral fat

Table 2 Prevalence (%) of sarcopenia[a] in the New Mexico Elder Health Survey, by age, sex, and ethnicity, 1993–1995.

	Men		*Women*	
Age group (years)	*Hispanic (n = 221)*	*Non-hispanic whites (n = 205)*	*Hispanics (n = 209)*	*Non-hispanic whites (n = 173)*
<70	16.9	13.5	24.1	23.1
70–74	18.3	19.8	35.1	33.3
75–80	36.4	26.7	35.3	35.9
>80	57.5	52.6	60.0	43.2

[a]Appendicular skeletal muscle mass/height^2 (kg/m^2) less than two standard deviations below the mean value for the young adults from Gallagher et al. (1997), with permission from APS and OUP.

accumulation in epidemiological studies. However, waist circumference is unable to differentiate VAT from SAT. As a result, persons with similar waist circumferences could have markedly different quantities of VAT and abdominal SAT. Skinfold thickness has been used as a continuous variable grading adiposity or adipose tissue distribution within study populations.

The most accurate measurement of VAT requires imaging techniques (MRI and computed tomography (CT)), which are expensive and not readily available in many clinical settings. **Fig. 3**B shows an MRI-derived cross-sectional image at the L4–L5 level with adipose tissue depots identified. The AT located between muscle bundles (IMAT; **Fig. 3**) and visible by MRI and CT may be negatively associated with insulin sensitivity. In the elderly, greater IMAT (as suggested by lower skeletal muscle attenuation by CT) is associated with lower specific force production. Currently, there is no simple or clinic-based method to measure adipose tissue located between the muscle groups, defined in our laboratory as intermuscular adipose tissue (IMAT). IMAT has been reported to be significantly negatively correlated with insulin sensitivity and higher in type 2 diabetic women compared to nondiabetic women (Song et al., 2004).

Sex and race differences in body composition are well established in adults. Men acquire higher peak SM mass than women and some evidence exists suggesting that men may lose SM faster than women with age. Moreover, it is well established that women have a larger amount of total body fat or total adipose tissue than men. Among races, African–American adult men and women have larger amounts of SM than Asian and Caucasians even after adjusting for differences in body weight, height, age, and skeletal limb lengths (Gallagher et al., 1997).

Efforts are ongoing to better understand variations in IMAT as a function of age, race, and level of fatness. IMAT deposits appear comparable in size in adult African–Americans, Asians, and Caucasians at low levels of adiposity but accumulate as a greater proportion of TAT in African–Americans compared to Caucasians and Asians subjects (58 g IMAT/kg TAT in African–Americans; 46 g IMAT/kg TAT in Caucasians; 44 g IMAT/kg TAT in Asians) (Gallagher et al., 2005). Across race groups, VAT deposits also appear comparable in size at low levels of adiposity but with increasing adiposity VAT accumulates more in Asians and Caucasians compared to IMAT, although accumulation rates for IMAT and VAT do not differ in African–Americans (Gallagher et al., 2005). Although the association between greater amounts of abdominal or VAT and increased insulin resistance and the metabolic syndrome is well established compared to the peripherally located SAT, the role of the IMAT compartment in the metabolic alterations leading to the development of insulin resistance warrants further investigation, especially as it may influence race/ethnicity differences in dysglycemia. Collectively, sex and race differences exist in body composition in children and adults.

Physiological application: two examples

Example 1

Expressing heat production relative to body mass is required when comparing energy expenditure rates between individuals that differ in size. Age and gender-specific resting energy expenditure (REE) norms based on body weight and stature-derived were developed in the early 1900s by Kleiber and showed that adult mammals differing widely in body size had similar metabolic rates relative to body weight raised to the 0.75 power (Kleiber, 1961). Two components are usually considered as representative of whole-body metabolically active tissue, body cell mass (BCM), and FFM. BCM is typically estimated as the exchangeable potassium space that can be measured by total body potassium. The FFM component can be measured using two-component body composition methods.

In studies assessing REE, FFM is considered the principal contributor to energy requirements, and is commonly used as a surrogate for metabolically active tissue. However, this practice is inherently flawed as it pools together numerous organs and tissues that differ significantly in metabolic rate (Gallagher et al., 1998). The brain, liver, heart, and kidneys alone account for approximately 60% of REE in adults although their combined weight is less than 6% of total body weight or 7% of FFM. The skeletal muscle component of FFM comprises 40–50% of total body weight (or 51% of FFM) and accounts for only 18–25% of REE. REE varies in relation to body size across mammalian species. Within humans, REE per kg of body weight or FFM is highest in newborns ($\sim$56 kcal kg^{-1} day^{-1}), declines sharply until 4 years, and slowly thereafter reaching adult values ($\sim$25 kcal kg^{-1} day^{-1}). Among adults, REE is lower in the later adult years, to an extent beyond that explained by changes in body composition. That is, the loss of FFM cannot fully explain the decrease (5–25%) in REE in healthy elderly persons.

Recent attention has been given to modeling REE based on available information on organ- and tissue-specific metabolic rates combined (**Table 3**) with the mass of these tissues as determined by MRI. Whole-body REE can be calculated from organ-tissue mass (REE$_c$) and then compared to REE measured using indirect calorimetry (REE$_m$) for individuals or groups. REE (in kJ day^{-1}) of each organ-tissue component (subscript i) can be calculated using the following equation:

$$\mathrm{REE}_i = \mathrm{OMR}_i M_i \tag{1}$$

where (OMR) (organ metabolic rate) is the metabolic rate constant (in kJ per kg per day) for each organ-tissue component (**Table 3**) and M is the mass of the corresponding organ/tissue (kg). Whole-body REE (in kJ per day) is calculated as the sum of the seven individual organ-tissue REE:

$$\mathrm{REE}_c = \sum_{i=1}^{7} (\mathrm{REE}_i) \tag{2}$$

Table 3 Organ and tissue coefficients used in developing models.

	Weight (kg)[a]	*Density (kg L^{-1})*[a]	*Metabolic rate (kJ kg^{-1} day^{-1})*[b]
Skeletal muscle	28.0	1.04	55
Adipose tissue	15.0	0.92	19
Liver	1.8	1.05	840
Brain	1.4	1.03	1008
Heart	0.3	1.03	1848
Kidneys	0.3	1.05	1848
Residual	23.2	–	50

Residual mass was not assigned a density but was calculated as body mass minus sum of other measured mass components.

[a]Snyder et al. (1975), with permission from OUP and LWW.

[b]Elia (1992), with permission from OUP and LWW.

The whole-body REE equation is

$$REE_C = 1008M_{brain} + 840M_{liver} + 1848M_{heart} + 1848M_{kidneys} + 55M_{SM} + 19M_{AT} + 50M_{residual}$$

This approach has allowed for the hypothesis to be tested that the proportion of FFM as certain high metabolic rate organs, specifically liver and brain, is greater in children compared to young adults (Fig. 4). Findings thus far have shown that after accounting for this disproportion, the specific organ/tissue metabolic constants available in the literature (Table 3) are not adequate to account for REE in children. These results therefore imply that the decline in REE per kilogram body weight (or per kilogram FFM) during the growth years is likely due to both changes in body composition and changes in the metabolic rate of individual organs/tissues. When this approach was applied to young adults (31.2 ± 7.2 years), REE_c and REE_m were highly correlated, with no significant differences between them. When this approach was applied to persons over 70 years, both older men and women had significantly lower REE_m compared to REE_c, and the magnitude of the differences were 13% and 9.5%, respectively, for men and women. These findings suggest that even after adjustment for age-related organ and tissue atrophy in the elderly, whole body REE by indirect calorimetry continues to be lower than expected. The latter suggests that the metabolic rate constants used (Table 3) for specific organs and tissues may not be appropriate in the elderly.

At the individual or clinic level, the measurement of REE by indirect calorimetry is frequently unavailable. An alternate approach has been to estimate REE based on body weight, height, age, and sex. Many studies have examined the association between these basic and easily acquired measures and REE. A small number of studies have included FFM in their REE prediction equations. Table 4 lists published equations for the prediction of REE in healthy individuals.

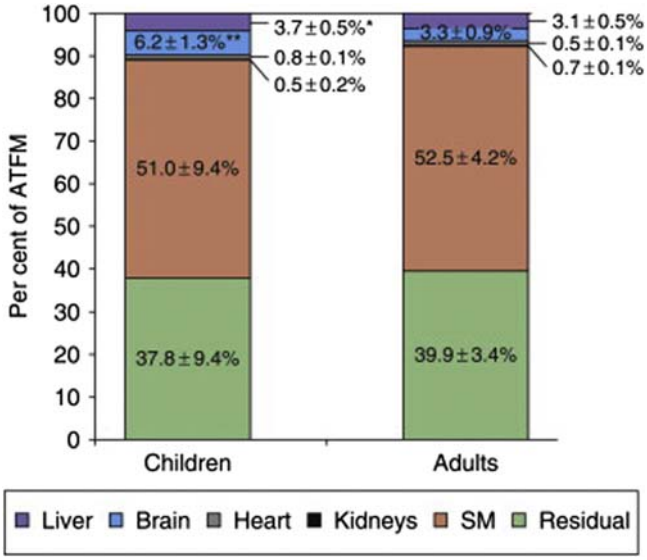


Fig. 4 Proportional contribution of each organ/tissue to Adipose Tissue Free Mass (ATFM). Liver ( ), brain ( ), heart ( ), kidneys ( ), skeletal muscle mass ( ), residual mass ( ), _ po.01 and __ po.001 for children vs. adults. Reproduced from Hsu, A., Heshka, S., Janumala, I., et al., 2003. Larger mass of high-metabolic-rate organs does not explain higher resting energy expenditure in children. Am. J. Clin. Nutr. 77: 1506–1511, with permission from JCI.

Table 4 REE prediction equations based on anthropometrics or body composition.

Authors	*Subjects/gender/nation*	*Weight status*	*Age (years)*	*Equation*
Harris and Benedict (1919)	239/M-F/USA	NW	29 ± 14 (X ± SD)	F: **BMR** = 9.5 wt (kg) + 1.9 ht (cm) − 4.7 age (years) + 655 M: **BMR** = 13.8 wt (kg) + 5.0 ht (cm) − 6.8 age (years) + 66
Robertson and Reid (1952)	2310/M-F/UK	NS	Range (3–80)	**RMR** = BSA (m^2) × 24 × age-specific value
Altman and Dittmer (1968)	>200/M-F/USA	NW	Range (3–16)	F: **REE** = 0.778 wt (kg) + 24.11 M: **REE** = 0.815 wt (kg) + 21.09
Dore et al. (1982)	140/F/UK	NW, OW, OB	Variable	**REE** = 8.24 wt (kg) + 0.02 FFM (kg) − 3.25 age (years) + 712
Bernstein et al. (1983)	202/M(154)/USA	OW, OB	40 ± 12 (X ± SD)	**RMR** = 7.48 wt (kg) − 0.42 ht (cm) − 3.0 age (years) + 844 **REE** = 22 FFM (kg) + 6.4 FM (kg) − 2.1 age (years) + 251
Garrow and Webster (1985)	104/F/UK	NW, OW, OB	Variable	**REE** = 24.2 FFM (kg) + 5.8 (% fat) + 310
Joint FAO/WHO/UN (1985)	11 000/M-F/Multi	NW, OW, OB	Variable	3–10 years F: **REE** = 22.5 wt (kg) + 499 3–10 years M: **REE** = 22.7 wt (kg) + 495 10–18 years F: **REE** = 17.5 wt (kg) + 651 10–18 years M: **REE** = 12.2 wt (kg) + 746 18–30 years F: **BMR** = 55.6 wt (kg) + 1397.4 ht (m) + 146 30–60 years F: **BMR** = 36.4 wt (kg) − 104.6 ht (m) + 3619 18–30 years M: **BMR** = 64.4 wt (kg) − 113.0 ht (m) + 3000 30–60 years M: **BMR** = 47.2 wt (kg) + 66.9 ht (m) + 3769
Schofield (1985)	7549/M-F/UK	NW, OW, OB	Range (<3 to >60)	Under 3 years F: **BMR** = 0.068 wt (kg) + 4.281 ht (m) − 1.730 Under 3 years M: **BMR** = 0.0007 wt (kg) + 6.349 ht (m) − 2.584 3–10 years F: **BMR** = 0.071 wt (kg) + 0.677 ht (m) + 1.553 3–10 years M: **BMR** = 0.082 wt (kg) + 0.545 ht (m) + 1.736 10–18 years F: **BMR** = 0.035 wt (kg) + 1.948 ht (m) + 0.837 10–18 years M: **BMR** = 0.068 wt (kg) + 0.574 ht (m) + 2.157 18–30 years F: **BMR** = 0.057 wt (kg) + 1.184 ht (m) + 0.411 18–30 years M: **BMR** = 0.063 wt (kg) − 0.042 ht (m) + 2.953 30–60 years F: **BMR** = 0.034 wt (kg) + 0.006 ht (m) + 3.530 30–60 years M: **BMR** = 0.048 wt (kg) − 0.011 ht (m) + 3.670 Over 60 years F: **BMR** = 0.033 wt (kg) + 1.917 ht (m) + 0.074 Over 60 years M: **BMR** = 0.038 wt (kg) + 4.068 ht (m) − 3.491
Owen et al. (1986)	44/F/USA	NW, OW, OB	29 ± 14 (X ± SD)	F: **RMR** = 7.18 wt (kg) + 795
Owen et al. (1987)	60/M/USA	NW, OW, OB	29 ± 14 (X ± SD)	M: **RMR** = 10.2 wt (kg) + 879
Owen (1988)	104/M-F/USA	NW, OW, OB	29 ± 14 (X ± SD)	**REE** = 23.6 FFM (kg) + 186
Ravussin and Bogardus (1989)	249/M-F/USA	NW, OW, OB	Variable	**REE** = 21.8 FFM (kg) + 392
Mifflin et al. (1990)	498/M-F/USA	NW, OW, OB	Range (19–78)	F: **RMR** = 9.99 wt (kg) + 6.25 ht (cm) − 4.92 age (years) − 161 M: **RMR** = 9.99 wt (kg) + 6.25 ht (cm) − 4.92 age (years) + 5 **REE** = 19.7 FFM (kg) + 413
Cunningham (1991)	Meta-analysis	NW, OW, OB		**REE** = 21.6 FFM (kg) + 370

Table 4 REE prediction equations based on anthropometrics or body composition.—cont'd

Authors	*Subjects/gender/nation*	*Weight status*	*Age (years)*	*Equation*
Maffeis et al. (1993)	130/M-F/Italy	NW, OW, OB	Range (6–10)	F: **REE** = (35.8 wt (kg) + 15.6 ht (cm) − 36.3 age (years) + 1552)/4.18 M: **REE** = (28.6 wt (kg) + 23.6 ht (cm) − 69.1 age (years) + 1287)/4.18
Hayter and Henry (1994)	2999/M/UK	NW, OW, OB	Range (18–30)	M: **RMR** = 51.0 wt (kg) + 3500
Pierro et al. (1994)	46 neonates/M-F/UK	NW	Range (1–126 days)	REE = −74.436 + 34.661 wt (kg) + 0.496 HR (beats/min) + 0.178 age (days) *infants were undergoing surgery for GI abnormalities
Molnar et al. (1995)	371/M-F/Hungary	NW, OB	Range (10–16)	M: RMR = 50.9 wt (kg) + 25.3 ht (cm) − 50.3 age (yr) + 26.9 F: RMR = 51.2 wt (kg) + 24.5 ht (cm) − 207.5 age (yr) + 1629.8 M + F: RMR = 50.2 wt (kg) + 29.6 ht (cm) − 144.5 age (yr) − 550 sex* + 594.3; *M = 0; F = 1
Piers et al. (1997)	39/M/Australia	NW, OW	Range (18–30)	M: **RMR** = 51.0 wt (kg) + 3415
van der Ploeg et al. (2001)	38/M/Australia	NW, OW	24.3 ± 3.3 (X ± SD)	18–30 years M: **RMR** = 48.2 wt (kg) + 25.8 ht (cm) − 49.6 age (years) + 113 18–30 years M: **RMR** = 21.0 wt (kg) − 56.2 age (years) + 76.1 FFM 4C (kg) + 2202
van der Ploeg and Withers (2002)	41/M/Australia	NW, OW	44.8 ± 8.6 (X ± SD)	30–60 years M: **RMR** = 41.92 wt (kg) + 13.79 ht (cm) − 14.89 age (years) + 1939 30–60 years M: **RMR** = 91.85 FFM 4C (kg) + 1463
Siervo et al. (2003)	157/F/Italy	NW, OW, OB	23.8 ± 3.8 (X ± SD)	F: **RMR** = 11.5 wt (kg) + 542.2
McDuffie et al. (2004)	502/M-F/USA		Range (6−11)	M: REE = 0.037 × weight − 4.67 × 1/height2 − 0.159 × race + 6.792 F: REE = 0.046 × weight − 4.492 × 1/height2 − 0.151 × race + 5.841 F: REE = 0.078 × fat-free mass + 0.026 × fat mass − 2.646 × 1/height2 − 0.244 × race + 4.8 F: REE = 0.101 × fat-free mass + 0.025 × fat mass + 0.293 × height 3 − 0.185 × race + 1.643
Almajwal and Abulmeaty (2019)	423/M-F/Saudi Arabia		Range (18–57)	REE = 3832.955 + 48.037 AdjWt (kg) − 30.642 ht (cm) + 141.268 sex − 4.525 age (years) M = 1; F = 0 AdjWt = (wt − IBW)/4 + IBW. IBW for M = (Ht (cm) − 152.4) × (1.0714) + 45.36 IBW for F = (Ht (cm) − 152.4) × (0.8928) + 45.36

M, male; F, female; NS, not specific; NW, normal weight; OW, overweight; OB, obesity; X, mean; SD, standard deviation; BSA, body surface area; wt, weight; ht, height; BMR, basal metabolic rate; RMR, resting metabolic rate; REE, resting energy expenditure; FFM, fat-free mass; FFM 4C, fat-free mass via the four-compartment body composition model; FM, fat mass. AdjWt, adjusted body weight; IBW, deal body weight.

Adapted from Almajwal and Abulmeaty (2019), Altman and Dittmer (1968), Bernstein et al. (1983), Cunningham (1991), Dore et al. (1982), Garrow and Webster (1985), Harris and Benedict (1919), Hayter and Henry (1994), Joint FAO/WHO/UNU Expert Consultation (1985), Maffeis et al. (1993), McDuffie et al. (2004), Mifflin et al. (1990), Molnar et al. (1995), Owen (1988), Owen et al. (1987, 1986), Pierro et al. (1994), Piers et al. (1997), Ravussin and Bogardus (1989), Robertson and Reid (1952), Schofield (1985), Siervo et al. (2003), van der Ploeg and Withers (2002), van der Ploeg et al. (2001).

Example 2

Prenatal exposures may increase long-term risk of developing obesity and related cardiometabolic diseases. Interventions to reduce adverse exposures in utero may impact risk of developing these conditions. Excessive maternal gestational weight gain (GWG) is one such exposure. A randomized controlled trial was developed to evaluate whether an intervention designed to prevent excessive GWG in women with overweight or obesity has a measurable effect on offspring body composition in the offspring at birth (less FM, greater FFM) (Gallagher et al., 2018).

Considerations for selection of measures depend on the resources available to the research team as well as parental preferences and concerns; however, the method(s) need to be sufficiently sensitive to detect small differences in FM and FFM between the intervention group and usual care group. In most studies, length and weight were most commonly measured because they are the least intrusive; generally require little skill, training, and equipment; and can be conducted quickly with minimal costs. However, length and weight and their relative indices such as weight-for-length and weight-for-age percentiles do not provide information on body

composition, specifically FM or FFM. Such indices assume that a higher percentile reflects additional or excess FM and fail to consider the contribution of the FFM compartment to weight. Skinfold thicknesses of the triceps, subscapular, and iliac crest can be used to assess between group differences in subcutaneous fat, a proxy for total body fat. While these methods are noninvasive and often acceptable to parents, they require skilled data collectors to conduct and additional time to acquire and can be burdensome to the parent and child.

Validated methods for measures of whole-body FM and FFM with high precision in the newborn are ADP (PEA POD), DXA and QMR. These instruments have higher cost, require trained/technical staff, and require more time to complete. ADP and DXA require that the infant be as still as possible during testing; movement is not an issue for QMR. For DXA, the infant is measured swaddled in a blanket. For ADP, the infant must be naked in a chamber where air temperature is comfortable. DXA also involves a small amount of radiation. Ultimately, ADP (PEA POD) was selected for measurement of the primary outcome, FM, and QMR for secondary outcomes FM, LM and TBW. Skinfold thicknesses at four sites, head circumference, length, and body weight and their indices (weigh-for-length and weight-for-age) were also collected as secondary measures to facilitate comparisons between other studies. Outcomes data were collected between 1 and 4 days after birth.

Results at baseline showed no between group differences in maternal characteristics (mean (±SD)): age 33.8 (±4.3) years, weight 81.9 (±13.7) kg, BMI 30.4 (±4.5) kg/m^2, gestational age at randomization 14.9 (±0.8) wk. GWG was less in the lifestyle intervention (LI) group by 1.79 kg ($p = 0.003$) or 0.0501 kg/wk ($p = 0.002$). Body weight and its lean component were significantly increased in LI measured by two independent methods. Compared to usual care (UC) group, LI infants had greater weight (131 ± 59 g $p = 0.03$), FFM (98 ± 45 g; $p = 0.03$) by PEAPOD, and lean mass (105 ± 38 g; $p = 0.006$) by QMR. FM and percentage fat were not significantly different. Head circumference, an index of greater brain growth, was similarly greater in LI. The QMR also measured greater total body water in the LI group.

This trial demonstrated that reliance only on infant body weight, length, and ponderal index would have resulted in an interpretation that LI produced larger and heavier babies, falsely presuming them to have greater adiposity and less lean mass per body weight. This rigorously controlled RCT highlights the importance of designing newborn trials with sensitive body composition measurement methods for adiposity and lean tissues.

Conclusion

The measurement of body composition allows for the estimation of body tissues, organs, and their distributions in living persons without inflicting harm. It is important to recognize that there is no single measurement method in existence that allows for the measurement of all tissues and organs and no method is error free. Furthermore, bias can be introduced if a measurement method makes assumptions related to body composition proportions and characteristics that are inaccurate across different populations. The clinical significance of the body compartment to be measured should first be determined before a measurement method is selected because the more advanced techniques are less accessible and more costly.

Acknowledgment

This work was supported by Grants UG3-DK128302, P30-DK026687-40S1 (Supplement to promote diversity), P30-DK026687, R01-HL70298, P01-DK42618 (Project 4), and R01-HD42187 from the National Institutes of Health.

References

Almajwal, A.M., Abulmeaty, M.M.A., 2019. New predictive equations for resting energy expenditure in normal to overweight and obese population. Internet J. Endocrinol. 2019, 5727496.

Altman, P., Dittmer, D. (Eds.), 1968. Metabolism. Federation of American Societies for Experimental Biology, Bethesda.

Andres, A., Gomez-Acevedo, H., Badger, T.M., 2011. Quantitative nuclear magnetic resonance to measure fat mass in infants and children. Obesity 19, 2089–2095.

Ashby-Thompson, M., Ji, Y., Wang, J., Yu, W., Thornton, J.C., Wolper, C., Weil, R., Chambers, E.C., Laferrere, B., Pi-Sunyer, F.X., Gallagher, D., 2020. High-resolution three-dimensional photonic scan-derived equations improve body surface area prediction in diverse populations. Obesity 28, 706–717.

Baumgartner, R.N., Koehler, K.M., Gallagher, D., Romero, L., Heymsfield, S.B., Ross, R.R., Garry, P.J., Lindeman, R.D., 1998. Epidemiology of sarcopenia among the elderly in New Mexico. Am. J. Epidemiol. 147, 755–763.

Behnke Jr., A.R., Feen, B.G., Welham, W.C., 1995. The specific gravity of healthy men. Body weight divided by volume as an index of obesity. 1942. Obes. Res. 3, 295–300.

Bernstein, R.S., Thornton, J.C., Yang, M.U., Wang, J., Redmond, A.M., Pierson Jr., R.N., Pi-Sunyer, F.X., Van Itallie, T.B., 1983. Prediction of the resting metabolic rate in obese patients. Am. J. Clin. Nutr. 37, 595–602.

Boileau, R.A., Lohman, T.G., Slaughter, M.H., 1985. Exercise and body composition of children and youth. Scand. J. Sports Sci. 7, 17–27.

Breithaupt, P., Colley, R.C., Adamo, K.B., 2011. Body composition measured by dual-energy X-ray absorptiometry half-body scans in obese children. Acta Paediatr. 100, e260–e266.

Cruz-Jentoft, A.J., Bahat, G., Bauer, J., Boirie, Y., Bruyère, O., Cederholm, T., Cooper, C., Landi, F., Rolland, Y., Sayer, A.A., Schneider, S.M., Sieber, C.C., Topinkova, E., Vandewoude, M., Visser, M., Zamboni, M., Writing Group for the European Working Group on Sarcopenia in Older People 2, The Extended Group for EWGSOP2, 2019. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing 48, 16–31.

Cunningham, J.J., 1991. Body composition as a determinant of energy expenditure: a synthetic review and a proposed general prediction equation. Am. J. Clin. Nutr. 54, 963–969.

Dore, C., Hesp, R., Wilkins, D., Garrow, J.S., 1982. Prediction of energy requirements of obese patients after massive weight loss. Hum. Nutr. Clin. Nutr. 36C, 41–48.

Elia, M., 1992. Organ and tissue contribution to metabolic rate. In: Kinney, J.M., Tucker, H.N. (Eds.), Energy Metabolism: Tissue Determinants and Cellular Corollaries. Raven Press, New York.
Ferreira, M.S., Marson, F.A.L., Wolf, V.L.W., Da Silva, M.T.N., Zambon, M.P., Antonio, M., Ribeiro, J.D., Mendes, R.T., 2019. Concordance between whole- and half-body scans to evaluate body composition in dual-energy X-ray absorptiometry in children and adolescents with different nutritional and pubertal conditions. Nutrition 66, 78–86.
Forbes, G.B., 1987. Human Body Composition: Growth, Aging, Nutrition, and Activity. Springer-Verlag, New York.
Gallagher, D., Visser, M., De Meersman, R.E., Sepulveda, D., Baumgartner, R.N., Pierson, R.N., Harris, T., Heymsfield, S.B., 1997. Appendicular skeletal muscle mass: effects of age, gender, and ethnicity. J. Appl. Physiol. 83, 229–239.
Gallagher, D., Belmonte, D., Deurenberg, P., Wang, Z., Krasnow, N., PI-Sunyer, F.X., Heymsfield, S.B., 1998. Organ-tissue mass measurement allows modeling of REE and metabolically active tissue mass. Am. J. Physiol. 275, E249–E258.
Gallagher, D., Ruts, E., Visser, M., Heshka, S., Baumgartner, R.N., Wang, J., Pierson, R.N., Pi-Sunyer, F.X., Heymsfield, S.B., 2000. Weight stability masks sarcopenia in elderly men and women. Am. J. Physiol. Endocrinol. Metab. 279, E366–E375.
Gallagher, D., Kuznia, P., Heshka, S., Albu, J., Heymsfield, S.B., Goodpaster, B., Visser, M., Harris, T.B., 2005. Adipose tissue in muscle: a novel depot similar in size to visceral adipose tissue. Am. J. Clin. Nutr. 81, 903–910.
Gallagher, D., Thornton, J.C., He, Q., Wang, J., Yu, W., Bradstreet, T.E., Burke, J., Heymsfield, S.B., Rivas, V.M., Kaufman, R., 2010. Quantitative magnetic resonance fat measurements in humans correlate with established methods but are biased. Obesity 18, 2047–2054.
Gallagher, D., Rosenn, B., Toro-Ramos, T., Paley, C., Gidwani, S., Horowitz, M., Crane, J., Lin, S., Thornton, J.C., Pi-Sunyer, X., 2018. Greater neonatal fat-free mass and similar fat mass following a randomized trial to control excess gestational weight gain. Obesity 26, 578–587.
Gallagher, D., Andres, A., Fields, D.A., Evans, W.J., Kuczmarski, R., Lowe Jr., W.L., Lumeng, J.C., Oken, E., Shepherd, J.A., Sun, S., Heymsfield, S.B., 2020. Body composition measurements from birth through 5 years: challenges, gaps, and existing & emerging technologies-A National Institutes of Health Workshop. Obes. Rev. 21, e13033.
Garrow, J.S., Webster, J., 1985. Are pre-obese people energy thrifty? Lancet 1, 670–671.
Harris, J.A., Benedict, F.G., 1919. A Biometric Study of Basal Metabolism in Man. Carnegie Institution, Washington, DC.
Hayter, J.E., Henry, C.J., 1994. A re-examination of basal metabolic rate predictive equations: the importance of geographic origin of subjects in sample selection. Eur. J. Clin. Nutr. 48, 702–707.
He, Q., Horlick, M., Fedun, B., Wang, J., Pierson Jr., R.N.,, Heshka, S., Gallagher, D., 2002. Trunk fat and blood pressure in children through puberty. Circulation 105, 1093–1098.
He, Q., Horlick, M., Thornton, J., Wang, J., Pierson Jr., R.N.,, Heshka, S., Gallagher, D., 2004. Sex-specific fat distribution is not linear across pubertal groups in a multiethnic study. Obes. Res. 12, 725–733.
Heymsfield, S.B., Wang, Z.M., Withers, R.T., 1996. Multicomponent molecular level models of body composition analysis. In: Roche, A.F., Heymsfield, S.B., Lohman, T.G. (Eds.), Human Body Composition. Human Kinetics, Champaign.
Hoffman, D.J., Wang, Z., Gallagher, D., Heymsfield, S.B., 2005. Comparison of visceral adipose tissue mass in adult African Americans and whites. Obes. Res. 13, 66–74.
Joint FAO/WHO/UNU Expert Consultation, 1985. Energy and protein requirements. Report of a joint FAO/WHO/UNU Expert Consultation. World Health Organ. Tech. Rep. Ser. 724, 1–206.
Kleiber, M., 1961. The Fire of Life, and Introduction to Animal Energetics. Wiley, New York.
Lundqvist, K., Neovius, M., Grigorenko, A., Nordenström, J., Rössner, S., 2009. Use of dual-energy X-ray absorptiometry in obese individuals: the possibility to estimate whole body composition from DXA half-body scans. Radiography 15, 12–19.
Maffeis, C., Schutz, Y., Micciolo, R., Zoccante, L., Pinelli, L., 1993. Resting metabolic rate in six- to ten-year-old obese and nonobese children. J. Pediatr. 122, 556–562.
Mcduffie, J.R., Adler-Wailes, D.C., Elberg, J., Steinberg, E.N., Fallon, E.M., Tershakovec, A.M., Arslanian, S.A., Delany, J.P., Bray, G.A., Yanovski, J.A., 2004. Prediction equations for resting energy expenditure in overweight and normal-weight black and white children. Am. J. Clin. Nutr. 80, 365–373.
Mifflin, M.D., St Jeor, S.T., Hill, L.A., Scott, B.J., Daugherty, S.A., Koh, Y.O., 1990. A new predictive equation for resting energy expenditure in healthy individuals. Am. J. Clin. Nutr. 51, 241–247.
Molnar, D., Jeges, S., Erhardt, E., Schutz, Y., 1995. Measured and predicted resting metabolic rate in obese and nonobese adolescents. J. Pediatr. 127, 571–577.
Napolitano, A., Miller, S.R., Murgatroyd, P.R., Coward, W.A., Wright, A., Finer, N., De Bruin, T.W., Bullmore, E.T., Nunez, D.J., 2008. Validation of a quantitative magnetic resonance method for measuring human body composition. Obesity 16, 191–198.
Owen, O.E., Kavle, E., Owen, R.S., Polansky, M., Caprio, S., Mozzoli, M.A., Kendrick, Z.V., Bushman, M.C., Boden, G., 1986. A reappraisal of caloric requirements in healthy women. Am. J. Clin. Nutr. 44, 1–19.
Owen, O.E., Holup, J.L., D'alessio, D.A., Craig, E.S., Polansky, M., Smalley, K.J., Kavle, E.C., Bushman, M.C., Owen, L.R., Mozzoli, M.A., et al., 1987. A reappraisal of the caloric requirements of men. Am. J. Clin. Nutr. 46, 875–885.
Owen, O.E., 1988. Resting metabolic requirements of men and women. Mayo Clin. Proc. 63, 503–510.
Paley, C., Hull, H., Ji, Y., Toro-Ramos, T., Thornton, J., Bauer, J., Matthews, P., Yu, A., Navder, K., Dorsey, K., Gallagher, D., 2016. Body fat differences by self-reported race/ethnicity in healthy term newborns. Pediatr. Obes. 11, 361–368.
Pierro, A., Jones, M.O., Hammond, P., Donnell, S.C., Lloyd, D.A., 1994. A new equation to predict the resting energy expenditure of surgical infants. J. Pediatr. Surg. 29, 1103–1108.
Piers, L.S., Diffey, B., Soares, M.J., Frandsen, S.L., Mccormack, L.M., Lutschini, M.J., O'dea, K., 1997. The validity of predicting the basal metabolic rate of young Australian men and women. Eur. J. Clin. Nutr. 51, 333–337.
Ravussin, E., Bogardus, C., 1989. Relationship of genetics, age, and physical fitness to daily energy expenditure and fuel utilization. Am. J. Clin. Nutr. 49, 968–975.
Robertson, J.D., Reid, D.D., 1952. Standards for the basal metabolism of normal people in Britain. Lancet 1, 940–943.
Rothney, M.P., Brychta, R.J., Schaefer, E.V., Chen, K.Y., Skarulis, M.C., 2009. Body composition measured by dual-energy X-ray absorptiometry half-body scans in obese adults. Obesity 17, 1281–1286.
Schofield, W.N., 1985. Predicting basal metabolic rate, new standards and review of previous work. Hum. Nutr. Clin. Nutr. 39 (Suppl. 1), 5–41.
Schutte, J.E., Townsend, E.J., Hugg, J., Shoup, R.F., Malina, R.M., Blomqvist, C.G., 1984. Density of lean body mass is greater in blacks than in whites. J. Appl. Physiol. Respir. Environ. Exerc. Physiol. 56, 1647–1649.
Siervo, M., Boschi, V., Falconi, C., 2003. Which REE prediction equation should we use in normal-weight, overweight and obese women? Clin. Nutr. 22, 193–204.
Siri, W.E., 1961. Body composition from fluid spaces and density: analysis of methods. In: Brozek, J., Hensch, E.A. (Eds.), Techniques for Measuring Body Composition. National Academy of Science, Washington, DC.
Snyder, W.S., Cook, M.J., Nasset, E.S., Karhausen, L.R., Parry Howells, G., 1975. Report on the Task Group on Reference Man. ICRP, Oxford.
Song, M.Y., Kim, J., Horlick, M., Wang, J., Pierson Jr., R.N., Heo, M., Gallagher, D., 2002. Prepubertal Asians have less limb skeletal muscle. J. Appl. Physiol. 92, 2285–2291.
Song, M.Y., Ruts, E., Kim, J., Janumala, I., Heymsfield, S., Gallagher, D., 2004. Sarcopenia and increased adipose tissue infiltration of muscle in elderly African American women. Am. J. Clin. Nutr. 79, 874–880.
Tataranni, P.A., Ravussin, E., 1995. Use of dual-energy X-ray absorptiometry in obese individuals. Am. J. Clin. Nutr. 62, 730–734.
Toro-Ramos, T., Paley, C., Wong, W.W., Pi-Sunyer, F.X., Yu, W.W., Thornton, J., Gallagher, D., 2017. Reliability of the EchoMRI infants system for water and fat measurements in newborns. Obesity 25, 1577–1583.

Van Der Ploeg, G.E., Withers, R.T., 2002. Predicting the resting metabolic rate of 30–60-year-old Australian males. Eur. J. Clin. Nutr. 56, 701–708.
Van Der Ploeg, G.E., Gunn, S.M., Withers, R.T., Modra, A.C., Keeves, J.P., Chatterton, B.E., 2001. Predicting the resting metabolic rate of young Australian males. Eur. J. Clin. Nutr. 55, 145–152.
Wang, Z.M., Pierson Jr., R.N., Heymsfield, S.B., 1992. The five-level model: a new approach to organizing body-composition research. Am. J. Clin. Nutr. 56, 19–28.

Further reading

Hall, K.D., Guo, J., Chen, K.Y., Leibel, R.L., Reitman, M.L., Rosenbaum, M., Smith, S.R., Ravussin, E., 2019. Methodologic considerations for measuring energy expenditure differences between diets varying in carbohydrate using the doubly labeled water method. Am. J. Clin. Nutr. 109, 1328–1334.
He, Q., Xia, B., Liu, A., Li, M., Zhou, Z., Cheung, E.C., Kuo, Z.C., Wang, B., Li, F., Tang, Y., Zheng, Z., Sun, R., Hu, Y.J., Meng, W., He, Y., Yuan, J., Zhang, C., 2021. Association of body composition with risk of overall and site-specific cancers: a population-based prospective cohort study. Int. J. Canc. 149 (7), 1435–1447.
Liu, B., Du, Y., Wu, Y., Snetselaar, L.G., Wallace, R.B., Bao, W., 2021. Trends in obesity and adiposity measures by race or ethnicity among adults in the United States 2011–18: population based study. BMJ 372, n365.
Ramos-Lopez, O., Milton-Laskibar, I., Martínez, J.A., San-Cristobal, C.R., Portillo, M.P., 2021. Precision nutrition based on phenotypical traits and the (epi)genotype: nutrigenetic and nutrigenomic approaches for obesity care. Curr. Opin. Clin. Nutr. Metab. Care 24, 315–325.

Clinical examination

B. Caballero, Johns Hopkins University, Baltimore, MD, USA

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of B. Caballero, Clinical Examination, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 233–235, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00199-9.

Introduction 35
Medical History 35
Physical Examination 36
Further Reading 38

Glossary

Beriberi Clinical syndrome resulting from thiamin (vitamin B-1) deficiency.
Keratomalacia Corneal lesion characteristic of vitamin A deficiency.
Kwashiorkor Undernutrition syndrome in young children characterized by severe protein depletion (edema, skin and hair changes, etc.).
Pellagra Clinical syndrome caused by niacin deficiency.

Introduction

The two most common settings for a clinical examination are the hospital (inpatient or outpatient) and the field health care unit. In the first situation, the physician or examiner may have access to resources that are usually not available in the field. Because of this and other constraints, the assessment of nutritional status in the field is frequently more narrowly aimed at identifying a specific clinical condition or set of signs and symptoms. In either case, it is essential that information on history and physical findings be collected in a standardized manner in terms of both format and procedures. The former is usually best achieved by the use of preprinted or computerized forms. Electronic forms can be programmed to perform immediate range checking as values are entered, thus alerting the operator when values out of range are entered. Procedures for examination must be clearly defined in writing, and any health worker should be able to follow the instructions and perform an acceptable measurement. Although many components of the examination are subjective, it is important to standardize as much as possible terms such as 'minor', 'average', and 'large' within the group of examining persons, attributing a numeric value whenever possible. If data entry requires selection from a numeric scale, they should also be standardized by cross-validation with experienced personnel or by means of photographs or models.

The two components of the clinical assessment are the medical history and the physical examination (Table 1).

Medical History

The medical history for nutritional assessment is no different from a general medical history, in which familial and past and present environmental factors and their possible association with specific diseases or disease risk are considered. For the purpose of nutritional assessment, this information will be used to determine if any nutritional finding or complaint may be caused by an underlying medical condition, particularly one that remains unrecognized at the time of the examination. Additionally, specific medical conditions and their current status are important factors altering nutrient requirements and dietary prescriptions.

One specific focus of medical history in a nutritional assessment context is the exploration of gastrointestinal function. Conditions such as chronic diarrhea, gastroesophageal reflux, and colonic disorders may be associated with reduced nutrient absorption or food avoidance that result in impaired nutritional status. Past history of gastrointestinal problems or surgery may also point to current alterations in nutrient digestion or absorption. Other important components of the medical history are history of weight loss or gain, past and present use of medications, use of special foods or formulas, changes in taste or smell, and food allergies and intolerances.

In children and adolescents, the medical history must also obtain information on neurodevelopmental stages, history of behavioral problems, and overall school performance. Food preferences must be noted, particularly in adolescence, when adoption of unconventional dietary practices is more likely to occur.

Table 1 Major components of a nutrition-oriented medical history

Medical history
History of weight loss or gain
Gastrointestinal symptoms (nausea, diarrhea, flatulence, pain, etc.)
History of changes in color or texture of skin, hair, conjunctiva, buccal mucosa
Use of medications
Physical activity level (work-related, leisure)
History of fatigue, shortness of breath, muscle cramps
Other lifestyle practices
Places of residence, travel (exposure to toxins, sunlight, food contaminants)
In children and adolescents
– Growth history
– Neurodevelopmental history
– General school performance
– Parental and siblings' body size (body mass index)
– Pubertal stage
– Food preferences, fads
Dietary history
Habitual dietary intake and preferences
Past diet history
Alcohol consumption
Food allergies and intolerances
Assessment of dietary intake
– 24-hour recall
– Food frequency questionnaire

A special component of the nutritional history is the assessment of habitual dietary intake. There are several approaches, all of them requiring substantial experience and standardization. These procedures, such as 24-hour dietary recall, or food frequency questionnaires, are discussed in detail elsewhere.

Physical Examination

As noted previously, anthropometric measurements are a key component of the physical examination. Measurement of weight and height is perhaps one of the most frequently performed nutritional measurements. Although its value is limited with regard to identifying specific nutrient deficiencies, it is invaluable to evaluate growth and adequacy of past and present diets in infants, children, and adolescents and to identify undernutrition and obesity in adults. Measurements should be done by trained personnel and following standard protocols. In addition to anthropometry, the physical examination focuses on signs of nutrient deficiency or excess. These signs usually appear only when the deficiency is advanced and are not to be expected in marginal deficiencies. Furthermore, the time that it takes for a deficient intake of a given nutrient to cause clinical manifestation of deficiency varies considerably, depending on whether the nutrient is stored in the body and on the initial status of the reserves. Typical signs for selected nutritional deficiencies are presented in Table 2. Virtually none of these signs, with the exception of Bitot's spots, are pathognomonic for one specific deficiency. However, they are useful in indicating a specific nutrient impairment and prompting further evaluation.

The physical examination should start with a general visual assessment of the patient. In children, state of alertness, willingness to engage in play, or resisting examination are important clues to energy level and physical strength. A generalized loss of fat depots, or excess adiposity as in the obese, is readily identifiable in most circumstances. A general overview can also identify pallor, loss of muscle mass, and skin changes.

Numerous signs of nutritional deficiencies can be identified in the skin and hair. Because skin exhibits a relatively rapid turnover, impairments in protein synthesis can result in fragile, flaky, and discolored skin. Vitamin A deficiency typically causes a dry, hyperkeratotic skin. The dermatitis of pellagra consists of patchy areas of hypo- or hyperpigmentations, usually in sun-exposed body regions, eventually progressing to hardened, broken surfaces. In protein–energy malnutrition, hair may become brittle, thin, and easily pluckable. Fluctuations in the rate of synthesis of hair protein may result in band discoloration, where pale and normal colors alternate, resulting in the 'banner sign', typical of kwashiorkor. Petechiae or hematomas may result from protein–energy malnutrition or vitamin K or vitamin E (in the newborn) deficiencies.

One of the most specific signs of nutritional deficiency can be identified in the eye. Vitamin A deficiency produces a series of alterations in the conjunctiva and the cornea that not only indicate a deficiency of this nutrient but also help grade its severity.

Table 2 Typical clinical signs of selected nutritional deficiencies

Deficiency	*Signs*
Protein–energy malnutrition	Hair: depigmentation, thinning, pluckability
	Edema in lower extremities (generalized in severe cases)
	Muscle wasting
	Decreased subcutaneous fat
	Skin: diffuse depigmentation, flaky dermatosis
	Liver enlargement
Vitamin A	Bitot's spot
	Conjunctival xerosis
	Corneal xerosis
	Keratomalacia
	Night blindness
Riboflavin	Angular stomatitis
	Cheilosis
	Scrotal (vulvar) dermatosis
	Red tongue
	Corneal vascularization
Thiamin	Edema
	Hyporeflexia
	Muscle tenderness
	Cardiac enlargement
	Tachycardia
Niacin	Pellagroid dermatosis
	Scarlet, raw, fissured tongue
	Malar and supraorbital pigmentation
Vitamin C	Bleeding, spongy gums
	Petechiae
	Ecchymoses
	Epiphyseal enlargement
	Atrophy of lingual papillae
	Follicular hyperkeratosis
Vitamin D	Active rickets: rib beading, epiphyseal enlargement, persistently open fontanelle, craniotabes, hypotonia
	Residual rickets: frontal or parietal bossing, bowlegs, knock-knees, thorax deformities
Iron	Pale conjunctiva
	Atrophy of lingual papillae
	Koilonychia
Folic acid, B_{12}	Usually associated with pallor of anemia
	Peripheral neuropathy (B_{12})
Iodine	Thyroid enlargement

The most commonly used classification of vitamin A deficiency is primarily based on eye findings, from Bitot's spots to perforated keratomalacia. Conjunctival pallor has been a classic sign of anemia, but its sensitivity varies substantially depending on ethnicity, ambient lighting, and experience of the observer.

The mouth and tongue are also areas where typical manifestations of deficiency can be detected. A red tongue is a classic sign of riboflavin deficiency but has also been associated with niacin deficiency; the latter may also include fissures. Conversely, a pale tongue may indicate iron deficiency. Glossitis, with or without color changes, has been linked to pyridoxine deficiency. A similar condition, including pain and intense red color, has been associated with biotin deficiency. Angular stomatitis and ulcerations and other lip lesions are associated with riboflavin or ascorbic acid deficiencies. In the latter, extensive involvement of the gums (swelling and bleeding) is also typical. Atrophy of the papillae occurs in vitamin B_{12}, niacin, and folate deficiencies. Excess vitamin A intake may result in discoloration of the gingival mucosa.

Rib beading (also known as rickets rosary) is a typical sign of vitamin D deficiency in children, but a similar manifestation may appear in vitamin C deficiency (scurvy). Epiphyseal enlargement and bowlegs are other classic signs of rickets. A distended abdomen is characteristic of protein–energy malnutrition in children. In the lower limbs, inspection must ascertain the presence of edema, which is also associated with protein–energy malnutrition.

Peripheral neuropathies such as those associated with beriberi or vitamin B_{12} deficiencies may result in visible impairment of limb movements, such as the 'foot drop' of dry beriberi.

In preadolescents and adolescents, assessment of sexual maturation (usually following the Tanner staging) is an important component of the physical examination, although it is not always feasible due to cultural and practical reasons. Alternatively, more limited information may be obtained in girls by self-reported menarcheal status. Self-assessment of Tanner stage by comparison with photographs is another useful alternative, but use of these photographs with children may not be acceptable in some communities.

To obtain a unified rating of a person's nutritional status, it is desirable to integrate clinical, laboratory, and functional data into a single scoring system. Several approaches to achieve this have been proposed, and their use will depend primarily on the target population and the intended use of the score. The Subjective Global Assessment is an approach that relies primarily on data from the physical examination and thus can be readily performed after this examination has been completed. Other scoring systems, such as the Prognostic Nutritional Index or the Instant Nutritional Index, rely to variable degrees on combinations of clinical and laboratory data.

Further Reading

Heimburger, D, McLaren, DS, Shils, ME, 2006. Clinical manifestations of nutrient deficiencies and toxicities. In: Shils, M, Shike, M, Ross, Caballero, B, Cousins, R (Eds.), Modern Nutrition in Health and Disease, 10th edn. Lippincott, Philadelphia, pp. 595–614.

McLaren, DS, 1992. A Colour Atlas and Text of Diet-Related Disorders. Wolfe, London.

Sardesai, VM, 2003. Introduction to Clinical Nutrition, 2nd edn. Marcel Dekker, New York.

Dietary intake measurement: Methodology

A.A. Welch[a] **and A.A. Mulligan**[b], [a] Department of Epidemiology & Public Health, Norwich Medical School, University of East Anglia, Norwich, Norfolk, United Kingdom; and [b] NIHR BRC Diet, Anthropometry and Physical Activity Group, MRC Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A.A. Welch, Dietary intake measurement: Methodology, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 65–73, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00075-1.

Introduction 40
Dietary intake measurements 40
Methods for measuring food consumption at the national level 42
Food balance sheets 42
Methods for estimating dietary intake at the household level: household budget surveys 43
Food account method 43
Inventory method 43
Household record 43
List recall methods 43
Individual dietary intake methods 43
24-hour recalls 44
Diet history 44
Food frequency questionnaires 44
Weighed food record inventory and estimated food record 45
Duplicate sample technique 45
Further information 45
Clinical practice 45
Factors affecting individual ability to report intake accurately 45
Interviewers 45
Computerized interview procedures 46
Using dietary methods in different populations 46
Recall of remote diet 46
Reproducibility of dietary methods 46
Use of data and conversion of reported intake to nutrients and food types 46
Qualitative analysis 46
Quantitative analysis 46
Data processing and computing dietary intake 47
Estimated food quantities 47
Data entry and nutrient calculation systems 47
Data processing errors 48
Issues associated with measurement of dietary intake 48
Measurement error 48
Measurement error in data collection and processing 48
Systematic bias 49
Interviewer bias 49
Social desirability bias 49
Impact of measurement error 49
Validation of dietary methods 49
Use of calibration methods to adjust for measurement error 50
Future developments 50
Conclusion 50
References 50
Further reading 51
Relevant websites 51

Key points

- Measuring dietary intake can be achieved using a number of different methods
- Dietary intake may be measured using direct methods of dietary assessment or using biomarkers of nutrition
- Measurement error can occur when measuring dietary intake
- Validation and calibration methods can be used to account for measurement error that occurs when measuring dietary intake

Glossary

Calibration A method of adjusting for systematic error. An example would be to use a more precise measure of dietary intake or a biomarker to adjust for measurement error with a less precise dietary method

Measurement error Error that can occur when measuring information in studies. An example would be error that occurred when estimating dietary intake of an individual

Qualitative analysis A technique for less structured or unstructured information and uses more open ended information such as open-ended survey responses or other non-numerical information such as photographs

Quantitative analysis A technique of analyzing numerical information or data and is used in statistical techniques that provide numerical outcomes

Validation A method of identifying the type and scale of measurement error. An example would be to use a biological marker such as plasma vitamin C to compare with dietary intake of vitamin C

Introduction

Dietary intake measurements are used to assess food, nutrient, or bioactive intake of individuals, groups, or populations. They may also be used to assess dietary patterns, such as the Mediterranean diet, or to study dietary behaviors, such as skipping breakfast. The purpose of collection of measurements varies from individual assessments in clinical situations (nutrition screening) or the adequacy of intake of population groups (nutrition surveillance) to use in research relating diet to health status, particularly in epidemiology. Measurements are also used to establish exposure to food-borne contaminants, in the evaluation of nutritional intervention programs, and to develop nutritional guidelines for governmental health policy. This article describes the dietary intake measurements available, issues associated with data collection, conversion to nutrients and food types, measurement error when using dietary intake methods, validation and calibration of dietary methods, and future developments. Dietary assessment methods can be broadly grouped into two categories, objective and subjective. Objective methods independently record the diet and do not rely on input from an individual. Direct observation, the analysis of duplicate diets and the use of nutritional biomarkers are the main types of objective methods. Subjective methods rely on the dietary information provided by an individual, with the aim of obtaining data that is as accurate as possible, by minimizing bias. Weighed and estimated food records, 24-h recalls, FFQs, diet checklists and diet histories are all types of subjective methods.

Dietary intake measurements

Table 1 describes the advantages and limitations of the main types of dietary methods, which are suitable for different purposes.

In all methods, "foods" refers to consumption of foods, beverages, and snacks both inside and outside of the home.

Of the individual methods, weighed records, estimated food records, 24-h recalls (24-HR), and dietary histories are more intensive. The quantity of food consumed may be weighed directly or estimated using household measures such as cups and spoons, photographs, standard units, or average portions (see Table 2). More recent methods in development utilize photographs of portion size recorded with mobile phone or digital camera technology. For all methods, the amount consumed can be measured or described either including or excluding wastage material usually discarded during food preparation, e.g., outer leaves and peel from vegetables or bones from cuts of meat.

Data should be derived from weighed intakes, government surveys, and research groups in populations similar to the one to be studied.

Some considerations when choosing a dietary method are shown in Table 3.

Table 1 Names and characteristics of dietary methods used for estimating food and nutrient intake.

Name of method		*Advantages*	*Limitations*
National level			
Food balance sheets		Available for 200 countries; suitable for monitoring change	Per caput not individual intake derived from estimates of use of commodities within a country or region; intake overestimated as nutrient losses during storage and preparation not accounted for; should not be used to provide estimates of nutritional adequacy of particular regions
Household level			
Food account method		Low respondent burden; relatively inexpensive	No estimates of change in larder stocks; measurements confined to food brought into the home (unless method modified to measure food consumed outside the home, which can be quite large); consumption of confectionery, alcoholic, and soft drinks excluded
Inventory method		Low respondent burden; relatively inexpensive	Consumption of confectionery, alcoholic, and soft drinks excluded
Household record		Suitable for populations with high proportion of homemade foods; useful if literacy levels are low; provides direct measure of food available for consumption	High input from field workers or interviewers
List recall methods		Relatively rapid and inexpensive; only one interview required; suitable for populations with higher proportion of purchased than home-produced food	Advance warning of interview may distort food consumption patterns; subject may fail to record items from memory; no record of foods eaten outside the home
Individual level			
Retrospective methods	24-hour recall (24-HR) (single or multiple days)	If interviewed, respondent literacy not important; not reliant on long-term memory; providing not forewarned, individuals do not alter food consumption; interview length 20–45 min	Single 24-HR should not be used for estimating intake of individuals but can be used for group assessments
	Diet history	Respondent literacy not required	Report of past intake is influenced by current diet; trained interviewers required; average interview length 1–1.5 h; high processing costs
	Food frequency questionnaire (FFQ) (if includes portion estimates termed semiquantitative FFQ)	Useful for large numbers; relatively straightforward to complete; administration simpler and less costly than other individual methods; more rapid data processing	Needs to be developed for specific population group to ensure important food items are covered and requires updating to accommodate changes to supply of foods; less flexible for later analysis as food lists are fixed; responses governed by cognitive, numeric, and literacy abilities of respondents also by length and complexity of the food list
Current methods	Weighed food record (weighed inventory technique)	No requirement for memory retrieval as it records current intake; food intake weighed so estimates of quantity consumed not required	Literate, cooperative respondents required as burden is high; possible that respondents change usual eating patterns to simplify the record; high processing costs
	Food record with estimated weights	No requirement for memory retrieval as it records current intake	Literate, cooperative respondents required as burden is high; possible that respondents change usual eating patterns to simplify the record; necessary to find values for estimates of quantity of food consumed; high processing costs

(Continued)

Table 1 Names and characteristics of dietary methods used for estimating food and nutrient intake.—cont'd

Name of method	*Advantages*	*Limitations*
Duplicate analysis	Greater accuracy	Highly labor intensive; requires laboratory to do food composition analysis; limited applicability in population studies
Records using electronic equipment e.g., mobile phones, digital cameras	Visual records of foods. Avoids need for paper records. Data can be sent to investigators electronically	Currently involves labor intensive programmes to convert to useable data i.e., quantities and types of foods, although systems are in development to deal with this Limited use in older people who experience difficulties with using newer technology

Table 2 Types of portion used for methods using estimated portions.

Portion types
Average or small, medium, large portions, weights – available from studies of weighed intake
Photographs (ideally there should be five or more representing the population range of intake)
Household measures (spoons, cups, mugs, liquid measures)
Standard units (1 apple, 1 banana)
Food models/replicas (three-dimensional models representing foods)

Table 3 Factors determining choice or suitability of method.

Size and scale of the data collection
Screening, clinical, research, surveillance purposes?
Literacy or numeracy of the population
Age of the individual or population (the very young or very old may need assistance with completion)
Intended or potential use of the data (immediate short-term assessment vs. prospective research)
Requirement for group or individual estimates for nutrient intake
Requirements from the data for nutrients, food groups, or bioactive nutrients
Detail and comprehensiveness of the information to be extracted for analysis (if information only required for particular nutrient or food type, shortened questionnaires may be administered)
Has repeatability of the method been assessed?
Have previous validation studies been performed on the method by other researchers in similar population groups to the one being studied?
Availability of resources for interviewers, including training
Availability of suitable coding and/or processing program (record and recall methods require greater resources than frequency methods but frequency programs are more complex to develop)

Methods for measuring food consumption at the national level

Food balance sheets

The Food and Agriculture Organization (FAO) publishes food balance sheets (FBSs) for approximately 200 countries. FBSs present a comprehensive picture of the pattern of a country's food supply during a specified reference period. Food balance sheets may also be termed national food accounts, food moving into consumption, food consumption statistics, food disappearance data, and consumption level estimates, reflecting differences in the method of calculation but providing similar information.

The supply available during a period is calculated from the total quantity of foodstuffs produced in a country, added to the total quantity imported and modified for any change in stocks that may have occurred. Calculation of quantities used for purposes other than human consumption (exports, livestock, used for seed, nonfood uses) and losses during storage and transportation are made. The per capita supply of each food item available for human consumption is calculated by dividing the total of available food by the number of the population actually consuming it and expressed in terms of quantity and nutrients. Estimates from FBSs include household wastage material, plate waste, and food fed to pets. Nutrient losses during storage, preparation, and cooking are not calculated and so figures for available food are greater than those reported by individual dietary surveys.

Food balance sheets can be used to formulate agricultural policies concerned with production, distribution, and consumption of foods and as a basis for monitoring changes and forecasting food consumption patterns, as well as to provide inter-country comparisons of available supplies.

Methods for estimating dietary intake at the household level: household budget surveys

Techniques for estimating intake at the household level include the food account method, the inventory method, the household record, and the list recall method. These methods measure all foods and beverages available for consumption by a household or family group during a specified time period of between 1 and 4 weeks, although some last for 2–3 months. Wastage factors are sometimes applied. Household surveys provide data for per capita consumption of foods or nutrients, not intake for specific individuals. Data are calculated irrespective of the age and gender distribution in the household. These methods provide population data for annual mean food consumption and selection patterns and are used for analyzing trends in intake. Household budget surveys are used more widely in Europe than elsewhere. As countries may not produce compatible data, the Data Food Networking Project (DAFNE) has developed the methodology to allow the data from 11 European countries to be combined and compared.

Food account method

A record is made by a respondent of details of all quantities of food entering the household (purchased, home grown, or received), usually over a period of 7 days. Changes in larder stocks are not estimated, as on average, some households will gain and some will use up stocks. Estimates of losses and wastage during preparation are made. This method is used for Living Costs and Food (LCF) module of the Integrated Household Survey (IHS) (until 2001 the LCF was called the National Food Survey and from 2001 to 2008 the UK Expenditure and Food Survey), and has included consumption of food, confectionery, soft drinks, and alcohol outside the home since 1992. As consumption outside the home now accounts for a substantial proportion of dietary intake in the UK the method was modified in 2001 to include the use of till receipts and individual 2-week diaries for each household member aged 7 years or older. This method can be used to measure seasonal variation in intake over 1 year.

Inventory method

The inventory method is similar to the food account method and respondents record all foods coming into the household. A wastage factor is often applied and a larder inventory is included at the beginning and end of the survey period.

Household record

Foods available for consumption (either raw or processed) are weighed or estimated. Foods for each meal are recorded separately to give a total for the household. Waste is measured directly or estimated. Interviewers visit the household early in the day to determine the quantity of food used to prepare the first meal and the number of individuals who consumed it. The midday meal may be weighed or recorded using estimated measures. A further interview is required later in the day. This method is appropriate for use in preindustrial societies where literacy is low and units for buying foods not standardized.

List recall methods

The respondent is asked by a trained interviewer to recall the amount and cost of food obtained for household use over a period, usually of 1 week. The method takes into account food use, purchases, and acquired food, but not waste. Quantities consumed are weighed or estimated using household measures. The interview can take up to 2.5 h. Response rates are usually high. Information on the age and sex of people in the household and the number of meals eaten both in and outside the home, income, and other socioeconomic characteristics may be collected. It is helpful to notify the respondent in advance so that records of purchases can be kept before the interview. This method was used by the United States Department of Agriculture (USDA) National Food Consumption Survey between 1931 and 1988.

Individual dietary intake methods

Many methods are available for estimating individual dietary intake measures and can be divided into two types: retrospective measures of intake such as 24-HR, dietary history or food frequency questionnaires (FFQs), or current measures of intake such as weighed or estimated food records. Qualitative information is available from all methods but quantitative estimates for nutrient

consumption are possible only if data for weighed or estimated portion weights are available. Most methods may be either self-completed or completed by a surrogate. Surrogates may be required if study individuals are too young, old, or infirm but data will be less reliable than when reported directly.

24-HR and FFQs may be self-completed or interview administered either face-to-face or by telephone and can be mailed. Data collection costs can be reduced if questionnaires can be self-completed or mailed.

The number of days of report required for adequate measures of nutrients using 24-HR, weighed, or estimated records varies depending on the day-to-day variability of nutrient consumption. The number of days is partly dependent on the variation in nutrient concentration in foodstuffs. The concentration of macronutrients such as protein and carbohydrate in foods varies less than micronutrients such as vitamin C or iron. The number of days required to classify individuals into the correct third of the percentage distribution for usual intake, for 80% of individuals, has been calculated in British and Swedish populations. Up to 7 days of recall would be required for energy, protein, sugars, and calcium. Nutrients with greater variability and requiring between 4 and 14 days of records were alcohol, vitamin C, riboflavin, and iron. More recent analysis for the number of days required to estimate energy intake, using doubly labeled water estimates of energy over 14 days as the reference, suggest that 3 days of record are optimal with no improvement with 4 or more days of records. Also, one of the 3 days of record should include a weekend day.

24-hour recalls

24-HRs determine intake during the preceding 24 h. Interviews can be recorded on paper or using interactive computerized software. The multiple pass approach has been developed in an effort to obtain a more accurate record of foods and drinks consumed over the previous day. The exact stages or passes may vary between tools. An example of a 5-stage approach is given below.

1. Quick list - to collect a list of foods and drinks consumed during the previous day
2. Forgotten foods – probe for any items that may have been forgotten during the quick list
3. Time and occasion – record the time and eating occasion for each item
4. Detail cycle – record detailed description, amount, cooking methods, use of condiments/associated foods, brands of shop-bought items, and any other information as a review of the day
5. Final probe – probe for anything else consumed, such as commonly forgotten items, e.g. drinks, fruit, biscuits, sauces

Day-to-day variability in nutrient intake is large and a single day will not categorize individuals correctly within a distribution of intake. Therefore, single 24-HR are better used for group assessments than estimates for individuals. However, multiple 24-HR can be used to overcome this problem. The sampling protocol for studies should include an equal proportion of all days of the week and coverage of all four seasons. Newer methods for on-line recording of 24-HR recalls currently exist and include ASA24 (https://asa24.nci.nih.gov/), myfood24 (https://www.myfood24.org/), Intake 24 (https://intake24.co.uk/) and the Oxford WebQ (https://www.ceu.ox.ac.uk/research/oxford-webq).

Diet history

The diet history consists either of an interview administered 24-HR or establishing usual eating pattern over a 1-week period, followed by a frequency questionnaire to provide additional information. The dietary history provides a representative pattern of usual intake and is interview administered only.

Food frequency questionnaires

FFQs consist of a list of specific foods or food types associated with frequency of consumption. They are termed semiquantitative if portions are included. Most questionnaires specify a frequency response in relation to an average or medium portion but some request records of specific portions. The period of record is usually the previous month or year. FFQs provide an indication of usual intake and can be used to obtain population estimates of frequency of consumption of food types. Guidance on the development, validation and use of FFQs for different study designs is available (Cade et al., 2002).

FFQs need to be developed for specific population groups otherwise important foods may be missed. FFQs may become outdated if the supply of foodstuffs changes. FFQs consist of a fixed food list, which may be a disadvantage for prospective studies as hypotheses to be tested are limited by the list. The FFQ EPIC Tool for Analysis (FETAs) is a tool to calculate nutrient and food group data from the EPIC-Norfolk FFQ (https://www.epic-norfolk.org.uk/for-researchers/feta-download/) but it has been designed so that it may be customized for different study populations. Factors that affect the response to FFQs are the literacy and numeracy of respondents, as some mathematical ability is necessary to calculate relative frequencies, the length and complexity of the food list, and the influence of current diet. Not all respondents will relate frequency to portion size accurately.

In the US, examples of FFQs are the Block and Willett questionnaires. In Europe, FFQs were developed for the European Prospective Investigations into Cancer and Nutrition (EPIC) study in the Netherlands, Germany, Greece, Italy, Denmark, France, and the UK.

A short FFQ, usually without portion size questions, may be used as a screening tool, to collect data about specific dietary components, such as intakes of fruits and vegetables, wholegrain foods, dairy foods or red/processed meats.

Weighed food record inventory and estimated food record

For weighed food records (WRs) all food consumed over a period is weighed and recorded with details of food type and method of preparation, on preprinted forms or booklets, to obtain consumption over a period of days. Portable scales need to be supplied. WRs may include some estimated items eaten out of the home. Leftover food should be weighed and deducted. The recommended time period for records is 4–7 days or more, although the number of days depends on the nutrient of interest, study population, and objectives of the study. As some populations have different eating habits at weekends, weekend days should be included proportionately.

For estimated food records, all foods consumed over a period are recorded with details of food type, method of preparation, and estimated portions over a period of days (see **Table 2**). If recorded over 7 days, this may be called a "7-day diary."

Both these methods have a high respondent burden and need cooperative, literate respondents. Respondents require training in the level of detail needed to describe foods. It is also possible that respondents will change usual eating patterns to simplify the process of the record. It is also beneficial to include a review of weighed records during the period of recording either after the first day or at the end.

Duplicate sample technique

Duplicate samples of all foods consumed are made and the nutrient content analyzed. This method is used for metabolic studies and though providing greater accuracy than other methods, its use is not feasible for most purposes.

Further information

Although nomenclature for dietary methodology is reasonably consistent, care should be taken when reading the literature as methods with the same name may have been applied differently. The final decision over which method to choose will depend on the aims of the study, the population for study, the potential burden on respondents, and the resources available. Household surveys and food balance sheets provide data for per caput but not individual intake. In general, individual and the more intensive methods are associated with higher costs and respondent burden, whereas household methods are more economical and have a lower respondent burden.

Clinical practice

Dietary methodology for clinical practice requires rapid assessments of nutritional intake in order to prescribe dietary change or to improve nutritional status. Traditionally, 24-HR of "usual" intake or diet histories have been used for this purpose. Food frequency questionnaires and weighed or estimated food records are not generally used, due to the more intensive burden on respondents and on the resources required for coding and processing the data.

There is considerable discussion over the optimum method to use for establishing individual dietary intake and studies designed to measure the validity of methods suggest that those that are more intensive and detailed lead to greater measurement precision, justifying the greater cost. Confirmation of these findings is required. Despite these potential benefits, if resources are unavailable, less intensive methods tend to be used.

Factors affecting individual ability to report intake accurately

Factors governing individual accuracy and quality of reports are respondent's literacy and numeracy skills; preconceived ideas on the purpose of the inquiry and, for list-based methods, the interpretation and meaning of food names (Smith, 1993). Individuals may make errors when measuring and recording food weights or estimating weights of foods consumed. There is also respondent variation in the perception of the size of portions represented by photographs.

Interviewers

The aim of using interviewers with dietary methods is to obtain a complete, accurate, and detailed record of what respondents eat. Therefore, it is important for interviewers to be well trained and have an awareness of food composition and preparation

techniques. Ideally, interviewers should be educated in nutrition (dietitians or nutritionists), although non-nutritionists can be trained to standardized techniques, and come from the same cultural or ethnic background as the study population. Interviewer protocols should be developed.

Computerized interview procedures

Computerized interview systems can aid interviewers by prompting for specific questions to elicit sufficient and specific detail and reduce the burden on interviewers. Examples are the Minnesota Nutrition Data System and the EPIC-SOFT systems, used in the US and a number of European countries. Although computerized interviews have advantages in improving accuracy and standardization, and in saving time and effort when recording and coding data, interviewers do have to be competent with computers and the resources required to develop systems are high.

Using dietary methods in different populations

Ethnic subpopulations may consume different food types than a main population and baseline surveys will be required to establish what types of foods and method of preparation are common. This information would be required before list-based methods such as the FFQ could be developed.

Recall of remote diet

Investigators may wish to recall diet in the remote past, perhaps of many years. However, interpretation of remotely recalled dietary data is complex as recalled diet is heavily influenced by current dietary habit. Some studies have found that the correlations between recalled past diet and current diet were higher than the correlations between actual past diet and recall of past diet. The onset of diseases such as cancer may affect the appetite and dietary intake of study participants and as recall of remote diet is strongly related to current diet, may affect recall of remote diet. As diet before the onset of disease is the measure of interest, it is preferable to collect dietary information prospectively, that is before disease onset. Case–control studies in which the diet of cases with disease is compared with controls may be affected by altered perception of recalled diet, particularly by cases.

Reproducibility of dietary methods

The reproducibility of a method may also be referred to as reliability, repeatability, or precision and is a measure of the extent to which the same results can be obtained when repeated under the same conditions. Repeated measures provide an estimate of the within-person variability of intake. However, interpretation of the repeatability of measures is difficult as a lack of consistency may be due to genuine change over a time period or a lack of sensitivity or specificity of the method used to measure intake.

Use of data and conversion of reported intake to nutrients and food types

Qualitative analysis

Dietary method data can be used qualitatively, for instance during the process of reviewing nutritional intake for the purpose of dietary treatment as in clinical practice. Data on frequency of consumption may also be collected and analyzed by the FFQ method without conversion to nutrient intakes. However, even for qualitative analyses, it is likely that paper-based dietary methods will require conversion to an electronic format. The majority of uses of dietary methods are targeted toward quantitative analyses.

Quantitative analysis

The data collected by dietary methods are converted into food and nutrient consumption by calculating the amount of food eaten and linking this to a database with values for the nutrient composition of foods.

The databases of nutrient composition of foods are provided by the governments of many countries. They consist of nutrient composition data for the average composition of commonly consumed foodstuffs and are usually available as printed publications, computerized databases, or as part of software packages. Values in nutrient composition databases are expressed as either per 100 g of food or per common household measure. Nutrient databases vary in the coverage and comprehensiveness of the foods and nutrients. They are revised periodically to cover newer foods of different nutrient compositions or to modify or extend the nutrient coverage. Some issues concerning the choice of nutrient databases are shown in Table 4. It is important to read the information distributed with the printed or electronic versions of databases to determine the uses and limitations of the data.

Table 4 Factors to consider when choosing a nutrient database to calculate nutrient intakes.

Comprehensiveness of food item and beverage coverage?
Does the database contain entries for important foods consumed by the population to be studied?
How comprehensive is the coverage of nutrients?
Does the database contain data for mixed or multiple ingredient recipes or dishes?
What analytical techniques were used to derive nutrients in the database? (There can be differences in nutrients measured by different techniques.)
Are the data officially evaluated?
What compilation methods were used to construct the database?
Which conversion factors are used to calculate metabolizable energy content of foods for protein, fat, carbohydrate, and alcohol?
What proportion of missing values exists within the database? (Missing values are counted as zero in calculations and so result in systematic underestimates of intake.)
For international studies or comparisons, how do the analytical methods for determining nutrient composition and compilation techniques affect the resulting data?

Several steps are involved in calculating nutrient intake (also known as coding or processing). The first is to choose an item in the database, which corresponds most closely with the food consumed. If the food consumed is not in the database a suitable alternative can be chosen by considering food type, general characteristics, and likely nutrient profile. Once the food has been chosen, the nutrient composition of the food quoted in the database is multiplied by the amount of food eaten, e.g., for 60 g food the nutrients would be multiplied by 0.6 (where nutrients are expressed per 100 g of food).

To calculate daily intake for an individual, the contribution of each food is calculated and all the foods for a day summated. If more than one day's data have been collected, it is usual to calculate the average of the number of days recorded. Data from FFQs are usually computed to consumption per day but can also be computed per week.

Although it is possible to compute intake by hand, using a calculator and a printed copy of a nutrient database, this is very labor intensive and in practice for most purposes has been superseded by computerization.

The UK Nutrient Databank (UKNDB) (https://beta.ukdataservice.ac.uk/datacatalogue/doi/?id=6533#!#7), contains over 5600 foods. The UKNDB is commissioned by Public Health England as part of the National Diet and Nutrition Survey (NDNS), is available in electronic format as an integrated dataset, and contains up-to-date nutrient composition data. Data in the UKNDB are very similar to the UK McCance and Widdowson's food composition tables (https://www.gov.uk/government/publications/composition-of-foods-integrated-dataset-cofid) but the UKNDB includes a larger range of processed foods and composite dishes, and missing values have been replaced with plausible values. It is maintained as part of NDNS.

Data processing and computing dietary intake

The same care as that invested in data collection should be applied to data processing as errors of great magnitude may be introduced.

Estimated food quantities

To obtain quantitative information for nutrients or food groups, actual or estimated food weights are used. For methods using estimated food weights, values also need to be found for foods described, such as standard units, average portions, or household measures. Sources of data are national publications, surveys of weighed dietary intakes, and food manufacturers. Data may also be included in nutrient calculation programs. Portion weights need to be population specific and, if unavailable, studies to establish values will be needed. Intensive methods used for large-scale surveys will require databases of more than 20 000 values for portion weights.

Data entry and nutrient calculation systems

A number of computerized data entry systems and nutrient calculation programs exist (Fitt et al., 2015; Welch et al., 2001); factors that need to be considered when choosing a system are given in **Table 5**. The features required depend on the intended use of the data but, as a minimum, should include a list of foods, weights of portions, and a nutrient composition database. Ideally, systems should enable entry of data in sufficient detail to fulfill hypotheses for investigation and include measures to ensure consistent entry by staff such as defaults for inadequately reported foods, portions, or mixed component foods. They should also include a method for entering newer foods with different nutrient composition from the existing nutrient database. This is particularly important, as the range of new foodstuffs and products with different nutritional characteristics is ever increasing.

Table 5 Factors to consider when choosing a computerized entry or interviewing program.

Speed of the assessment
Requirements of the study for detailed or general data
Food composition database used
Food portion database used
Cost of the system
Facilities for organization of data
Ability to extract nutrients or food groups from the system
How up to date are the nutrient composition databases included in the system?
Commercial availability

Computerized systems and nutrient databases become outdated and for large-scale prospective studies, it is desirable to develop systems with a flexible approach to updating by using database technology.

Data processing errors

Errors arising during the coding (data entry) and processing of individual dietary methods (24-HR, diet history, weighed and estimated records) need to be avoided. Misclassification can arise due to human error if incorrect foods are chosen during coding, for instance, if milk was consumed in the full-fat form but was coded for skimmed milk. This may also arise where a food has local or alternative food names, which may be unknown to the coder. It is important to have a qualified nutritionist available to develop a protocol for training staff, answering queries, and dealing with ambiguities. Coders should have knowledge of food composition and food preparation techniques. It is difficult to control entry of incorrectly matched foods but careful checking and staff training are crucial in preventing this. Other potential errors are entry of incorrect quantities or multiplication factors for portion weights and missed items, problems that can occur even with structured computer programs. So, systematic post-entry checks to identify extremes of portion weights or nutrient values and the verification and correction of data are necessary.

Issues associated with measurement of dietary intake

Measurement error

There is potential for the occurrence of measurement error with the measurement of any exposure such as when using dietary methods to measure nutritional intake. Errors may arise as a result of flaws in the design of the measurement instrument or during data collection or processing. Measurement error may also occur as a result of individual characteristics of participants in studies. Measurement error can be defined as the difference between the measured exposure (or measure of dietary intake) and the true exposure. All measurement of dietary exposures is subject to some degree of measurement error making it difficult to achieve measurements of true intake.

Efforts to reduce measurement error during data collection and processing should be introduced into the protocol of all studies, however, even if preventative measures are taken, it is impossible to eliminate it altogether. It is difficult to identify the type and structure of measurement error associated with dietary intake. Measurement error may occur because of inaccurate reporting by respondents. It may also vary according to dietary method, for instance, food items within record methods may be intentionally or unintentionally omitted and with FFQs, frequency of consumption may be inaccurately reported. Systematic bias, interviewer bias, recall bias, and social desirability bias have been identified but there are likely to be other sources of error. (Bias can be defined as the modification of a method of measurement by a factor, which influences the measurement in one or more directions.) Measurement error associated with dietary methods may consist of one or more types of error.

Measurement error in data collection and processing

Self-report dietary assessment instruments are affected by two main types of error or bias, systematic and random error, which must be understood and addressed in order to avoid misleading results.

Random error can occur for a variety of reasons. Examples of random error relating to dietary assessment include unclear explanations from investigators, an individual's error in estimating portion size, dietary coding errors, or limitations with a food composition table. Random error causes the estimated values to deviate from the true intake, due to chance, affecting the reliability of a method. It may be reduced by increasing the number of days of dietary assessment, including quality control and assurance checks or carrying out pilot studies, especially when creating a new questionnaire.

Systematic bias

Systematic bias is a systematic mismeasurement of data and can occur, for instance, if equipment such as weighing scales under- or overestimates values or if an interviewer consistently fails to use questions to probe for consumption of snacks and additional foods. If systematic bias can be identified, solutions can be found, for instance, by calibrating equipment or training and monitoring interviewers.

Interviewer bias

The behavior of an interviewer can influence the response of interviewees leading to interviewer bias. The degree of rapport between interviewer and respondent also influences results. Bias may occur if interviewers omit responses or record them incorrectly. Trained interviewers should ask open-ended questions in a neutral or non-leading manner, and not imply that a food or beverage should or should not have been consumed and avoid value judgments.

Social desirability bias

Social desirability bias can influence dietary measures as respondents strive to report what they think is required not what was actually consumed, for example, reporting less alcohol consumption than is the case or greater consumption of foods with perceived health benefits such as fish, fruit, or vegetables. This is likely to be the cause of misreporting, under-reporting, or low energy reporting, which occurs in certain respondents. It is possible to predict how much energy a respondent should report, as this is the amount required to maintain a stable weight. (Weight will be either gained or lost if more or less energy is consumed than required.) As energy intake should equate to energy expenditure, expenditure effectively measures intake. Techniques for measurement of energy expenditure such as whole body calorimetry and doubly labeled water can be used. Using these techniques, those individuals classified as low energy reporters are likely to be older, more overweight, and of lower educational and socioeconomic status than the rest of the population. Low energy reporters tend to have lower consumption of foods in certain food groups such as cookies, cakes, puddings, confectionery (candy), and sugary foods and, in some populations, lower consumption of spreads, cooking fats, and potato chips. Interviewers should be aware of low energy reporting, aim to be entirely nonjudgmental, and also request participants make complete records of food intake.

Impact of measurement error

As the proportion of error within a measurement increases, the accuracy of the measurement decreases and the results using the measurement will become less interpretable. Hence, greater measurement error reduces the likelihood that the truth has been measured with accuracy and increases the likelihood that analyses relating diet to disease status will tend toward null results. The effect of measurement error is to misclassify an individual within a range of intake.

Validation of dietary methods

Validation is used to quantify the measurement error that occurs when measuring dietary intake exposures. It requires two measures: a main measurement and a second measurement that is subject to less measurement error than the first. The errors of the two measurements should be independent. Validation is used to estimate the proportion of measurement error within the main method by modeling the differences between the main and the secondary measurements. It had been considered that dietary methods had errors independent of each other and that record methods such as 24-HR could be used, but it is now known that the errors are not independent as individuals report in the same way with different methods. Therefore, it is better to use biological variables measurable in blood or urine (also known as biomarkers) as the second measure for dietary validation (Jenab et al., 2009; Potischman and Freudenheim, 2003).

Nutritional biomarkers may be classified using more than one method. This classification is comprised of four types: recovery, concentration, predictive and replacement biomarkers, although the classification is not mutually exclusive. Recovery biomarkers allow quantitative estimates of intake over a specific time period. Concentration biomarkers measure relative concentration across a distribution. Predictive biomarkers can be used to assess the validity of a dietary assessment method but have lower overall recovery. Replacement biomarkers are closely related to concentration biomarkers and often the distinction between them is difficult to make - their differentiating characteristic is that they refer specifically to compounds for which information in food composition databases is unsatisfactory or unavailable. Examples of recovery biomarkers are urinary excretion over 24 h of nitrogen, potassium and sodium (Park et al., 2018). Examples of concentration biomarkers are measures in blood of vitamins such as vitamin C and carotenoids, minerals, and individual fatty acids. Urinary sucrose (Kuhnle et al., 2015)and fructose are examples of predictive biomarkers. Examples of replacement biomarkers are some aflatoxins and phytoestrogens (Grace et al., 2004) and some of the recent biomarkers identified through metabolomics.

Examples of validation studies are those performed within EPIC-Europe and the Observing Protein and Energy Nutrition Study (OPEN) in the US Work is ongoing to extend the number of biomarkers available and to define further and elicit the structure of measurement error. The doubly labeled water technique has been used to estimate total energy utilization in order to validate estimates of energy intake. However, this technique is extremely costly which reduces its use in large populations. Also, independent estimates of energy intake should be done in periods of weight stability.

Newer methods of identifying biomarkers have been developed recently. These include the use of metabolomic techniques (Brennan et al., 2015) which can identify and measure small molecules, called metabolites, which can provide detailed information on metabolic pathways and biological processes. The metabolome has been measured in biological samples of blood and urine and newer candidates for nutritional biomarkers identified. A recent study (Playdon et al., 2017) found strong associations between healthy eating patterns and biomarkers, identified using metabolomics, in blood. However, whether or not metabolomic biomarkers will enable us to achieve a more accurate self-reported assessment of dietary intake or proximates, such as carbohydrate or fat, remains to be seen. These techniques have so far identified components/metabolites on metabolic pathways of nutrient metabolism, which have mainly been specific to particular nutrients or bioactive compounds in foods. It is not possible to estimate intakes of foods from these methods, only to "validate" consumption of certain types of foods or not e.g. strawberries, citrus fruit.

Use of calibration methods to adjust for measurement error

In contrast to validation, which attempts to identify the type and scale of measurement error, calibration is designed to adjust for systematic over- or under-estimation in dietary intakes within populations. It may also be used at the individual level, in an attempt to correct for attenuation bias (or dilution) in relative risk due to errors in dietary measurements. Calibration of data has been proposed for large multicenter nutritional studies that have used different dietary methods to capture population-specific diets. Calibration studies require a highly standardized second dietary measure to be used in a representative subsample from each cohort to form a common reference measurement across populations. An example of this approach has been used by the European EPIC Study using a computerized, standardized 24-HR in 10 countries (Slimani et al., 1999).

Future developments

A number of technology-based tools which currently exist for the assessment of dietary intake have been evaluated (Eldridge et al., 2019; Vu et al., 2017). Future developments in methodology involve using computing, digital, and Internet technology, such as videos of food eaten and online programs for self-reported intake. Use of Dictaphones and combinations of weighing and other recording equipment are also possible. Development of statistical techniques to combine the optimal properties of different dietary methods is also ongoing and is a promising avenue for the future.

The number of foodstuffs available, particularly of manufactured foods and readymade meals, will continue to increase, presenting challenges for those attempting to estimate nutrient intake. Nutrient databases will continue to be expanded and updated to incorporate newer food items and nutrient measurements available using improved analytical techniques.

In some populations more than 40% of individuals have been shown to consume supplements and as very few comprehensive databases of vitamin and mineral supplements exist these need to be developed, as supplements can make a major contribution to nutrient intakes (Lentjes et al., 2011).

Conclusion

A wide variety of methods for measuring dietary intake exists, categorized as subjective or objective. It is essential to understand their advantages and limitations, and how to deal with measurement error which may occur, both for use in clinical practice and research, and when reading and interpreting the scientific literature.

References

Brennan, L., Gibbons, H., O'Gorman, A., 2015. An overview of the role of metabolomics in the identification of dietary biomarkers. Curr. Nutr. Rep. 4, 304–312. https://doi.org/10.1007/s13668-015-0139-1.

Cade, J., Thompson, R., Burley, V., Warm, D., 2002. Development, validation and utilisation of food-frequency questionnaires – a review. Publ. Health Nutr. 5, 567–587. https://doi.org/10.1079/phn2001318.

Eldridge, A.L., Piernas, C., Illner, A.K., Gibney, M.J., Gurinović, M.A., de Vries, J.H.M., Cade, J.E., 2019. Evaluation of new technology-based tools for dietary intake assessment—an Ilsi Europe dietary intake and exposure task force evaluation. Nutrients 11. https://doi.org/10.3390/nu11010055.

Fitt, E., Cole, D., Ziauddeen, N., Pell, D., Stickley, E., Harvey, A., Stephen, A.M., 2015. DINO (Diet in Nutrients Out)-An integrated dietary assessment system. Publ. Health Nutr. 18, 234–241. https://doi.org/10.1017/S1368980014000342.

Grace, P.B., Taylor, J.I., Low, Y.L., Luben, R.N., Mulligan, A.A., Botting, N.P., Dowsett, M., Welch, A.A., Khaw, K.T., Wareham, N.J., Day, N.E., Bingham, S.A., 2004. Phytoestrogen concentrations in serum and spot urine as biomarkers for dietary phytoestrogen intake and their relation to breast cancer risk in European Prospective Investigation of Cancer and Nutrition-Norfolk. Cancer Epidemiol. Biomarkers Prev. 13, 698–708.

Jenab, M., Slimani, N., Bictash, M., Ferrari, P., Bingham, S.A., 2009. Biomarkers in nutritional epidemiology: applications, needs and new horizons. Hum. Genet. 125, 507–525. https://doi.org/10.1007/s00439-009-0662-5.

Kuhnle, G.G.C., Tasevska, N., Lentjes, M.A.H., Griffin, J.L., Sims, M.A., Richardson, L., Aspinall, S.M., Mulligan, A.A., Luben, R.N., Khaw, K.T., 2015. Association between sucrose intake and risk of overweight and obesity in a prospective sub-cohort of the European Prospective Investigation into Cancer in Norfolk (EPIC-Norfolk). Publ. Health Nutr. 18, 2815–2824. https://doi.org/10.1017/S1368980015000300.

Lentjes, M.A.H., Bhaniani, A., Mulligan, A.A., Khaw, K.T., Welch, A.A., 2011. Developing a database of vitamin and mineral supplements (ViMiS) for the Norfolk arm of the European prospective investigation into cancer (EPIC-Norfolk). Publ. Health Nutr. 14, 459–471. https://doi.org/10.1017/S1368980010002867.

Park, Y., Dodd, K.W., Kipnis, V., Thompson, F.E., Potischman, N., Schoeller, D.A., Baer, D.J., Midthune, D., Troiano, R.P., Bowles, H., Subar, A.F., 2018. Comparison of self-reported dietary intakes from the Automated Self-Administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers. Am. J. Clin. Nutr. 107, 80–93. https://doi.org/10.1093/ajcn/nqx002.

Playdon, M.C., Moore, S.C., Derkach, A., Reedy, J., Subar, A.F., Sampson, J.N., Albanes, D., Gu, F., Kontto, J., Lassale, C., Liao, L.M., Mannistö, S., Mondul, A.M., Weinstein, S.J., Irwin, M.L., Mayne, S.T., Stolzenberg-Solomon, R., 2017. Identifying biomarkers of dietary patterns by using metabolomics. Am. J. Clin. Nutr. 105, 450–465. https://doi.org/10.3945/ajcn.116.144501.

Potischman, N., Freudenheim, J.L., 2003. Biomarkers of nutritional exposure and nutritional status: an overview. J. Nutr. 133, 873–874. https://doi.org/10.1093/jn/133.3.873s.

Slimani, N., Deharveng, G., Charrondière, R.U., Van Kappel, A.L., Ocké, M.C., Welch, A., Lagiou, A., Van Liere, M., Agudo, A., Pala, V., Brandstetter, B., Andren, C., Stripp, C., Van Staveren, W.A., Riboli, E., 1999. Structure of the standardized computerized 24-h diet recall interview used as reference method in the 22 centers participating in the EPIC project. Comput. Methods Progr. Biomed. 58, 251–266. https://doi.org/10.1016/S0169-2607(98)00088-1.

Smith, A.F., 1993. Cognitive psychological issues of relevance to the validity of dietary reports. Eur. J. Clin. Nutr. 47, S6–S18.

Vu, T., Lin, F., Alshurafa, N., Xu, W., 2017. Wearable food intake monitoring technologies: a comprehensive review. Computers 6, 1–28. https://doi.org/10.3390/computers6010004.

Welch, A., McTaggart, A., Mulligan, A., Luben, R., Walker, N., Khaw, K., Day, N., Bingham, S., 2001. DINER (Data into Nutrients for Epidemiological Research) – a new data-entry program for nutritional analysis in the EPIC–Norfolk cohort and the 7-day diary method. Publ. Health Nutr. 4, 1253–1265. https://doi.org/10.1079/phn2001196.

Further reading

Bingham, S., 1987. The dietary assessment of individuals; methods, accuracy, new techniques and recommendation. Nutrition Abstracts and Reviews, Series A 57, 705–742.

Cameron, M.E., Van Staveren, W.A., 1988. Manual on Methodology for Food Consumption Studies.

Gibson, R.S., 1990. Principles of Nutritional Assessment. Oxford University Press.

Margetts, B.M., Nelson, M. (Eds.), 1997. Design Coccepts in Nutritional Epidemiology, second ed. Oxford University Press, Oxford.

Pao, E.M., Cypel, Y.S., 1996. Estimation of dietary intake. In: Zeigler, E.H., Filer, L.J. (Eds.), Present Knowledge in Nutrition, seventh ed. ILSI, Nutrition Foundation, Washington, DC, pp. 498–507.

Penn, L., et al., 2010. Assessment of dietary intake: NuGO symposium report. Genes Nutr. 5 (3), 205–213. https://doi.org/10.1007/s12263-010-0175-9.

Thompson, F.E., Byers, T., 1994. Dietary assessment resource manual. J. Nutr. 124 (11 Suppl.). https://doi.org/10.1093/jn/124.suppl_11.2245s.

Willett, W., 1998. Nutritional Epidemiology, second ed. Oxford University Press, Oxford.

Relevant websites

https://asa24.nci.nih.gov/. ASA24 tool, which is a self-administered 24 hour recall.

https://beta.ukdataservice.ac.uk/datacatalogue/doi/?id=6533#!#7. The UK Nutrient Databank (UKNDB) contains over 5600 foods, is available in electronic format as an integrated dataset, and contains up-to-date nutrient composition data and no missing nutrient values.

https://www.ceu.ox.ac.uk/research/oxford-webq. The Oxford WebQ is a validated web-based 24 hour dietary assessment tool, developed for repeated administration in large prospective studies.

https://www.epic-norfolk.org.uk/for-researchers/feta-download/. The FFQ EPIC Tool for Analysis (FETA) is a tool to calculate nutrient and food group data from FFQs.

http://www.eurofir.net/. EuroFIR Project website with information on European food composition databases.

http://www.fao.org. INFOODS information for nutrient database compilers and suppliers.

https://www.gov.uk/government/publications/composition-of-foods-integrated-dataset-cofid. McCance and Widdowson's 'The Composition of Foods' – the UK food composition tables.

https://inddex.nutrition.tufts.edu/international-dietary-data-expansion-project. The International Dietary Data Expansion (INDDEX) Project, with information on critical issues that have long impaired effective food, nutrition, and agricultural policy and programming in low- and middle-income countries (LMICs).

https://intake24.co.uk/. Intake24 is an open-source, self-completed, computerised dietary recall system based on the multiple-pass 24-hour recall.

https://www.myfood24.org/. myfood24 (Measure Your Food on One Day) is an online 24-h dietary assessment tool.

Doubly labeled water

Michele N. Ravelli[a,b], **Timothy C. Shriver**[b], **and Dale A. Schoeller**[b,c], [a] Department of Neurology, University of Wisconsin, Madison, WI, United States; [b] Biotechnology Center, University of Wisconsin, Madison, WI, United States; and [c] Nutrition Science, University of Wisconsin, Madison, WI, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of T.C. Shriver, N.M. Racine, D.A. Schoeller, W.A. Coward, Energy Expenditure: Doubly Labeled Water, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 164–169, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00093-3.

Introduction 52
Methodology 55
Instrumentation 56
Uses for the doubly labeled water method 57
Conclusion 60
References 60
Relevant website 61

Key points

- The DLW is an accurate (0.5%) and precise (8%) method to measure total daily energy expenditure and total body water.
- The method is based on the difference between the elimination rates of water labeled with two stable isotopes (^{2}H and ^{18}O).
- The Doubly labeled water method was first validated and applied to measure the energy expenditure of wild animals and later validated in humans and utilized in hundreds of studies under free-living conditions.
- The DLW method revolutionized the study of energy metabolism through applications including energy requirements throughout the life-course, physical activity, weight management, and effects of disease and diet and lifestyle on energy expenditure.

Glossary

Body composition Partitioning of body tissues into compartments. A common two-compartment model separates the body into fat and fat-free mass

Doubly labeled water (DLW) A tracer kinetic method to measure CO_2 production using a combination of two stable isotopes of water-i.e., deuterium and oxygen-18

Indirect calorimetry The measurement of respiratory gas exchange during the oxidation of fuel in the body producing CO_2 and consuming O_2 from which the heat released during that oxidation is calculated

Respiratory exchange ratio (RER) The ratio of CO_2, output, and O_2 consumption that occurs during respiration

Total energy expenditure (TEE) The sum of resting metabolic rate, the thermic effect of meals, and physical activity expenditure

Introduction

Chemically defined, water is composed of one oxygen atom bonded to two hydrogen atoms. This molecule has unique physical and chemical properties that are vital to support life. The molecule is polar, having a slightly negative end and a slightly positive end. This characteristic allows water to easily dissolve other polar molecules and also creates surface tension, capillary action, and is a medium for electrical conductivity. From the largest mammal to the smallest bacteria, the existence of life on Earth depends on water. Water fills cells providing cell volume and intracellular structural support. Without water, organisms would not have a means to store genetic information as water's polarity is what facilitates DNA to form helical coils from which the genetic information can be translated into the synthesis of proteins. In multicellular organisms, water is a critical transport medium for polar molecules between cells whose membranes are lined with channels making them permeable to water. The metabolic processes taking place within cells in the form of photosynthesis or respiration require water as a chemical solvent and often also as a reactant or product. Water also provides the essential functions of lubrication and transport for nutrients and waste. Because of these roles and many others, water is central to life known on Earth.

Water is usually the largest component of the four basic molecular entities that together comprise the body: protein, lipid, mineral, and water. Because these are the only basic components, measurement of any of these can provide considerable insight into an individual's body composition. Moreover, nutritionists frequently combine protein, water, and mineral into a single compartment that is termed fat-free mass. This allows an investigator to describe the body in terms of fat and fat-free mass, which is useful in determining body energy stores and in assessing the health status of an individual in terms of malnutrition and obesity. Several methods are available to determine fat or fat-free mass. Dual-energy X-ray absorptiometry (DXA) uses X-rays to differentiate fat mass, fat-free mass, and bone mineral mass. Underwater weighing is a way to measure body volume by water displacement. This procedure helps to calculate body density or mass per unit volume, and from density, the body fat percentage can be estimated using based on the difference in density of fat and fat-free mass. Finally, one of the oldest and still commonly used methods of body composition analysis is to measure total body water (TBW). Water is found only in the fat-free compartment because fat is anhydrous. In humans and other mammals, TBW has been found to be relatively constant at approximately 73% of the body's fat-free mass. Thus, a measured TBW allows us to calculate Fat mass $=$ Body mass $-\left(\frac{TBW}{0.73}\right)$.

The common method for measuring TBW is isotope dilution, which is based on the principle of dilution using a small amount of tracer with a known volume to measure a large unknown volume into which it is mixed. The principle of dilution can be illustrated by analogy. Consider a highly concentrated dye in a small volume will have a very dark appearance (high light absorbance). If this is diluted in a large beaker of water, the mixture will have a lighter color (Fig. 1). If the amount of the concentrated dye added into the small beaker is known and the concentration of the dye in the larger mixture is measured, then the volume of the diluting water in the larger beaker can be calculated. Isotope dilution uses the same principle, but instead of a dye, an isotopically labeled water is added to the water. Total body water, or more correctly, isotope dilution space (N), which is calculated as:

$$N = \frac{\text{Dose}}{\text{Isotope concentration}}$$

For this purpose, minor isotopes of interest in a water molecule typically include ^{2}H (deuterium) and ^{18}O, which are stable (nonradioactive) isotopes, and ^{3}H (tritium), which is a radioactive isotope. Because stable isotopes do not undergo nuclear decay, they do not release harmful radiation; yet, as isotopes, they act very much like their more common isotope in chemical processes. These stable, minor isotopes occur in nature and make up a relatively constant percentage of the total pool of each element in nature. Because there is no risk to radiation, they constitute a safe option for use as tracer molecules, except for ^{2}H in water which begins to cause some changes in biological reaction rates when the enrichments exceed 10–15% body water. These enrichments are more than 1000 times those employed in the doubly labeled water (DLW) method.

When a loading or pulse of isotope-labeled water is consumed, however, the tracer does not remain in body water after TBW is measured. Water is constantly being lost from the body. This loss, if unchecked, further demonstrates the importance of water for human life. A water loss (~3% of TBW) as little as 2% body mass will cause dry mouth, headaches, and decreased urine production. A water loss of 5% of body mass will further cause increased heart rate, respiration, fatigue, muscle cramps, and even further decreased urine production. Severe dehydration, a loss of 10% of body mass or more after several days without water

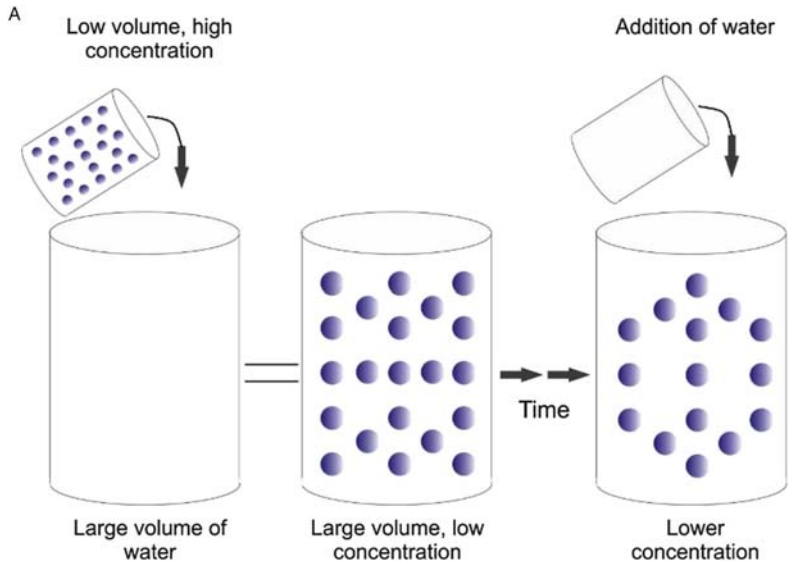


Fig. 1 Adding a concentrated dye to a beaker of water is analogous to a person adding a small amount of isotopic tracer to their body's water pool. Adding water to the mixture created is analogous to that person then ingesting water over time and further diluting the mixture.

consumption, will cause seizures, loss of vision, high pulse, muscle spasms, and eventually death. Repletion through drinking and food moisture along with the required electrolytes replaces lost water. In normal daily living, this loss and replacement occur well before dehydration occurs and creates a dynamic state in which part of the human water pool turns over on a daily basis. This daily loss of labeled water and replacement with unlabeled water further dilutes the isotopically labeled body water. Returning to the analogy of the dye solution, the mixture's color will become lighter with each loss of labeled water and replacement with unlabeled water (**Fig. 1**B).

Because the method uses water labeled with two isotopes, ^{2}H and ^{18}O, hence the name DLW. After dosing these labeled water molecules readily mix with unlabeled water and can therefore follow or trace the physiological and biochemical pathways of water in the body. Lifson was the first who reported that the oxygen in body water exchanges and equilibrate with those in carbon dioxide (Speakman, 1997). Because of this, the two isotopes do not behave identically. The ^{2}H is excreted as water, while the ^{18}O is excreted as both water and carbon dioxide. The turnover rate of ^{18}O is, therefore, greater than that of ^{2}H and the difference between the turnover rates is a measure of the CO_2 flux through the body (**Fig. 2**A). Because CO_2 is an end product of energy metabolism, one can use the difference between the two turnover rates to measure CO_2 production (**Fig. 2**B), and from that to calculate energy expenditure.

The DLW method is, therefore, a form of indirect calorimetry because it measures a product produced during the oxidation of fuel in the body and not a direct measure of the heat released during that oxidation. This form of indirect calorimetry has become the method of choice for the measurement of energy metabolism in free-living animals including humans. The isotope dilution allows for the dynamic measurement of each of the water isotopes over time. Because ^{2}H exits the body almost exclusively as water, it is possible to calculate the water elimination rate (rH_2O):

$$rH_2O = N_D k_D$$

where N_D is the deuterium dilution space and k_D is the fractional turnover rate of $^{2}H_2O$, which equals:

$$k_D = \frac{(Ln[D]_2 - Ln[D]_1)}{(t_2 - t_1)}$$

where the slope of deuterium enrichment is calculated by the difference between isotopic abundances ($Ln[D]_2$ and $Ln[D]_1$) in the samples across the time interval after the dose administration (t_2 and t_1). Similarly, ^{18}O's elimination rate ($rH_2O + 2rCO_2$) can be calculated:

$$rH_2O + 2rCO_2 = N_O k_O$$

where N_O is the ^{18}O dilution space and k_O is the fractional turnover rate of $H_2{}^{18}O$. However, note that provides us with two equations and two unknowns, which when combined allows us to solve for rCO_2:

$$rCO_2 = \frac{1}{2}(N_O k_O - N_D k_D)$$

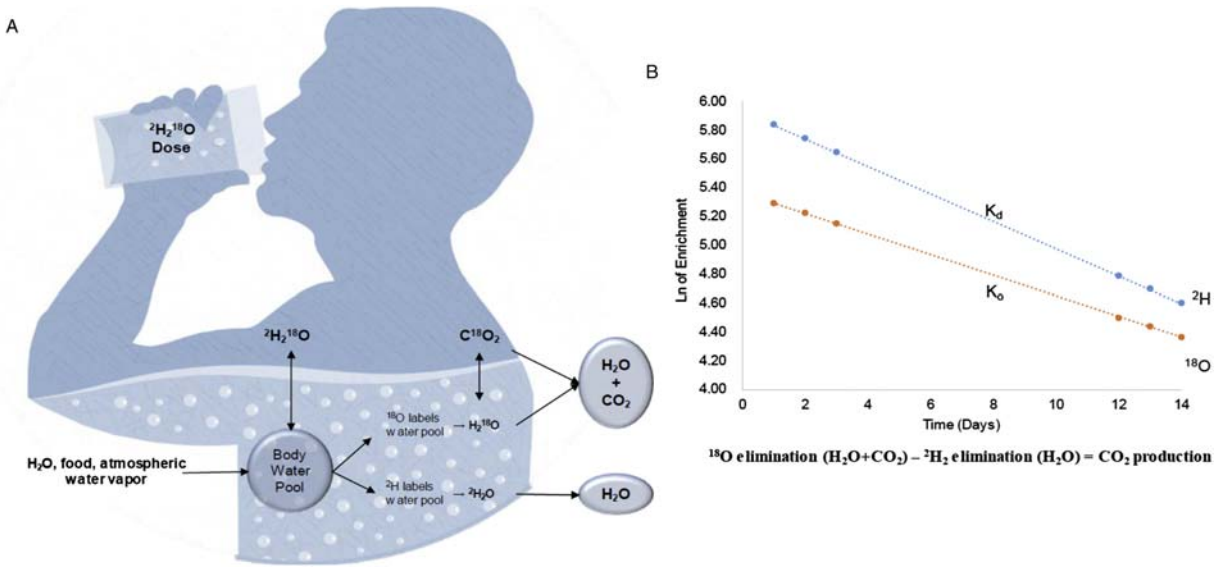


Fig. 2 Schematic image of the doubly labeled water method. (A) After ingesting the dose of DLW, the labeled compounds dilute in the body water pool and then equilibrate with the H_2O and CO_2. The washout kinetics of both isotopes and their concentrations decrease exponentially toward natural abundance levels due to the elimination of water and CO_2. (B) The difference in the turnover rates of ^{2}H and ^{18}O is proportional to CO_2 production.

However, this equation is not exact because it does not take into account isotope fractionation. Fractionation occurs when one isotope is discriminated against during a change in state (liquid to water vapor) or exchange between molecules (water to CO_2). The fractionation factor is <1 when the heavy isotope is discriminated against, =1 when there is no discrimination, or >1 when the light isotope is discriminated against. At 37 °C, the isotope fractionation factors are as follows: 0.946 for ^{2}H and 0.991 for ^{18}O between liquid to water vapor and 1.038 for ^{18}O between water and CO_2.

In addition, to further increase the accuracy, this equation needs to take into account the difference between dilution spaces (N_D and N_O) and TBW (N) across the lifespan. With regard to the former, a small amount, which can vary slightly particularly during the first few years after birth, of both ^{2}H (~4.3%) and ^{18}O (~0.7%) exchange with nonaqueous compounds creating an isotope dilution space that is a little larger than the TBW pool.

Over the past 40 years, a classical equation proposed by Schoeller et al. (1986), has been presented with small slight variations (Coward and Prentice, 1985; Racette et al., 1994; Schoeller, 1988; Speakman, 1997; Speakman et al., 1993) in the combination of the four basic parameters to measure the isotope elimination rates and, subsequently, CO_2 production, has been used in nutrition science studies. The variability on energy data using these slightly different equations is not relevant when results are compared within an investigation, however, it can significantly affect research outcomes when the TEE values are compared across studies, such as between cultures and lifestyles (Speakman et al., 2021).

To address these issues, a group of experts in stable isotopes and energy metabolism recently derived a new equation based on data of 6621 DLW measurements from 23 countries, which were compiled in the International Atomic Energy Agency (Speakman et al., 2019). This assessment aimed to obtain a more accurate estimative of energy expenditure values, besides promoting gathering and comparing data of future studies more efficiently (Speakman et al., 2021). To this end, the experts proposed a correction to the propagation of the isotope exchange variation with nonaqueous atoms, which even small, is another critical parameter that affected the CO_2 calculation.

Two new approaches were derived and recommended for use in DLW future studies (Speakman et al., 2021). First, the original equation proposed by Schoeller et al. (Schoeller et al., 1986), was modified taking into account the constant of 1.036 as the average dilution space ratio ($DSR = N_D/N_O$) to mitigate the potential error induced by the nonaqueous isotope exchange variation (Sagayama et al., 2016). This updated approach is indicated for individuals >2 and <96 years and achieved an accuracy of $-0.4 \pm 7.6\%$ when compared to values from indirect calorimetry. The equation was simplified as follow, considering the isotope fractionation and exchange constants into it:

$$rCO_2 = 0.455 \times N \times [(1.007k_O) - (1.041k_D)]$$

Second, the experts' group validated a new equation for babies and infants under 2 years of age with a body mass of 0–10 kg, which accuracy of $0.64 \pm 11.9\%$. In this approach, the equation considered the weight dependency observed for the isotopes DSR, and then combined it to the new standard equation above, as simplified below:

$$rCO_2 = [0.45859 \times N \times (K_O - (DSR \times k_D))]$$

Where $N = N_O$ and DSR is defined from an asymptotic exponential model, considering the body mass BM in kg, as follow:

$$DSR = 1.036 - 0.05 \times \exp(-0.5249 \times BM)$$

Finally, the investigators (Speakman et al., 2021) provided a free-to-use website (http://dlw.som.cuanschutz.edu/) where investigators can input isotope data to obtain rCO_2 and TEE using the recommended approaches.

Methodology

The initial distribution and equilibration of the two isotopes across TBW after ingestion and subsequent elimination from the body provides the basis for the total energy expenditure calculation. Because the isotope analysis is for aqueous samples, various body fluids (urine, saliva, and plasma) can be used. Urine is the usual fluid of choice as it is easily acquired in sufficient quantities and requires minimal cleaning to prepare it for isotope ratio mass spectrometric analysis. The procedure starts with the participant in the fasted state and providing a baseline sample, after which the dose is administered. Three more samples are obtained typically over 3–6 h post-dosing depending on other requirements of the study protocol and specific population characteristics. At this point, the subject is free to leave the testing center and resume their normal lifestyle and experimental protocol. Urine samples are again required at the end of the measurement period. The measurement period is typically between 0.5 and 3 elimination half-lives of the ^{2}H in body water. In most adults, this is between 3 days and 3 weeks, although to best account for the weekend and weekday behaviors, periods of 7 or 14 days after the dose was administered are recommended. The specimens should be collected at the same time of the day that the post-dose samples were obtained. The measurement is usually shortened under conditions of high water turnover, such as extreme physical activity, high ambient temperature, or infancy (Schoeller et al., 1986). This minimal sampling protocol is one of the attributes of the DLW method because it allows the subject to engage in the activities of their normal lifestyle. Unlike other methods of calorimetry that require frequent or continuous sample collection or monitoring.

This modest sampling requirement is possible because the DLW method transforms the human body into a metabolic recorder. By measuring the tracer concentrations in the body at the start and end of a 1- or 2-week metabolic period, the amount of tracer that left the body can be calculated and from that how much tracer was eliminated. This provides a measure of the total water flux and CO_2 flux through the body during that period, which, when divided by the number of days, yields the average daily flux.

To calculate total energy expenditure (TEE), one more piece of information is needed. The respiratory exchange ratio (RER) is the ratio of CO_2 output and O_2 input that occurs during respiration. The RER is dependent on what type of substrate (fat vs. carbohydrate) is being used to produce energy on a cellular level. The RER can be estimated based on diet. For example, the typical Western diet of 30–35% of energy from fat yields a food quotient of approximately 0.86. When the subject is close to the energy balance, the RER is assumed to equal the food quotient and this is used as an estimate of the RER in the standard indirect calorimetric equations for calculating energy expenditure from CO_2 production. The following modified Weir equation is the simplest and thus the most commonly used:

$$TEE\ \left(\text{kcal.day}^{-1}\right) = 22.26 \times\ rCO_2 \times (1.10 + 3.90\ /\ RER)$$

Where 22.26 is the gas volume constant in l mole of CO_2. The gas volume constant for CO_2 differs slightly from that of the perfect gas. The value for rCO_2 is derived from the tracer measurements as discussed above, and RER is estimated from diet or diet plus any change in body fat and protein stores.

Instrumentation

Unlike radioisotopes, which can be detected by the amount or the form of radiation products they produce, stable isotopes must be identified and differentiated on the basis of their mass. Several types of mass spectrometers are currently available that separate and measure isotope concentrations based on their mass-to-charge ratio (m/z), where m is the mass of the ion and z is its charge. In basic terms, a sample is introduced into the instrument in the gas phase, ionized, separated by m/z, and then detected. Briefly, magnetic sector mass analyzers create and then accelerate ions through a magnetic field. Because all ions have equal initial kinetic energies, they are separated by their velocities through the tube as a lighter isotope will follow a more curved path than the heavier isotope. The quadrupole mass analyzer contains four parallel rods that use a combination of a direct current and an alternative current voltage at radio frequencies to guide ions of selected mass toward the detector. Ions in the sample whose m/z ratios are not selected by the detector simply collide with the rod and do not reach the detector. Although quadripole instruments may be more compact and less expensive to purchase and operate, for the purposes of analyzing DLW, the isotope ratio mass spectrometer (IRMS) is used. This is a magnetic sector instrument that offers superior precision for isotope ratio analysis of the light elements like H and O. The use of IRMS allows investigators to use lower doses of isotopic tracer in their studies and thus reduces the cost of ^{18}O administered. Typically, 10% atom percent excess (APE) rather than 95–99% APE normalized ^{18}O is mixed with 99% APE $^{2}H_2O$ for dosing, thus making the application of the DLW method more cost-efficient, typically between $2.50 to $3.50 per kg of body weight. In this context APE is the percentage of water molecules that contain the heavy stable isotope in excess of that present in natural water. This fact alone has led to the use of this type of analysis in human experimental trials than any other. The sample is introduced into the IRMS as a gas (CO_2 or H_2) which is then ionized by electron impact under vacuum and accelerated from the source as an ion beam (Fig. 3). The ionized gas sample is then accelerated through an evacuated flight tube by an electrical gradient, during which the ions of different mass are separated by a magnetic field. The ion flight for each gas species is then terminated on impact with a faraday cup or collector. An array of these cups are configured to measure the current for each of the charged species of a gas and the isotope ratio is then measured for each sample and then compared with a reference of known isotope abundance, typically being either CO_2 or H_2.

Traditionally the introduction of the sample is through a dual-inlet system. This allows an unknown specimen to be measured against a known reference pure gas and the isotope abundance expressed using the delta notation (δ_i) and permil unit ($^0/_{00}$). The delta permil value is essentially the relative difference of the isotopic ratios for each of the two gases. Originally introduced by geochemists to express isotope abundances for natural samples collected from around the world, delta value in permil is calculated as follows:

$$\delta_i = \frac{(R_S - R_R)}{R_R} \times 1000,\ ‰$$

Where R_S is the ratio of the minor (heavy) to the major (light) isotope for the sample and R_R is the same for the reference. For example, the isotope ratio for an unknown sample of carbon dioxide enriched with ^{18}O is calculated by measuring the mass-to-charge ratio of 46/44 (a molecule of $^{12}C{-}^{16}O{-}^{18}O^+$, $m/z = 46$ over that of a molecule $^{12}C{-}^{16}O{-}^{16}O^+$, $m/z = 44$). The reference is calibrated and reported relative to an international isotopic standard. Similarly, isotope ratios can be measured by introducing the gas sample into the IRMS *via* a continuous flow system. This method utilizes a noninterfering carrier gas (typically He) to transport the sample into an ion source. Once in the source, the sample with the carrier is subjected to the isotope ratio analyses. The detector selects the mass of the sample and therefore the carrier is not detected. Other analytical methods also exist that are used for the analysis of isotopically stable water. Fourier transform infrared spectrometry (FTIR) is a technique that utilizes the absorption of infrared light to measure the deuterium content of a water sample. Based on the absorption of light specific to the vibration energies of the ^{18}O and ^{2}H in the O–H bonds, the concentration of deuterium can be measured, but FTIR does not provide the same precision for isotope abundance measures and thus requires the use of a larger and more costly dose of DLW. A relatively new method, laser absorption spectroscopy, has an analytic precision close to that of IRMS, and costs less than IRMS, and is easier to operate. The cavity ring-down spectroscopy (CRDS) instruments measure the light remaining nanoseconds after a pulse of laser light

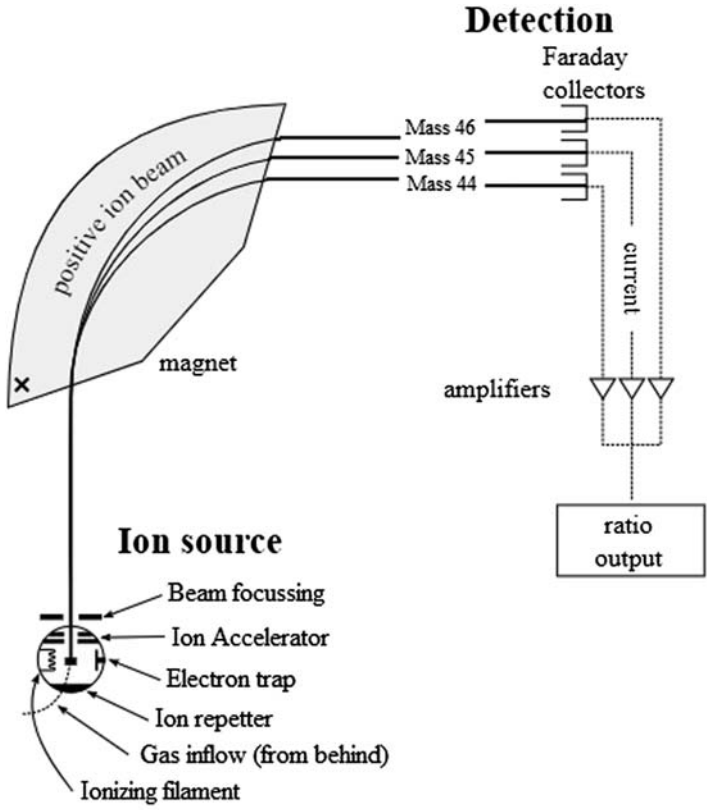


Fig. 3 Schematic of an IRMS measuring CO_2 at the US Geological Survey (https://en.wikipedia.org/wiki/Isotope-ratio_mass_spectrometry).

as it is absorbed by a gas sample in an optical cavity, while the off-axis integrated cavity output spectroscopy (OA-ICOS) devices provide a direct measurement of the absorbing substances, rather than only a cavity decay time, by scanning the laser light wavelength in the gas mixture. The better performance of the OA-ICOS is related to its non-dependence of hyper-critical optical alignment, making this new technology less vulnerable to vibrations, small physical shocks, and changes in temperature and pressure. An advantage of CRD and OA-ICOS is its ability to measure both $^{2}H_2O$ and $H_2{}^{18}O$ simultaneously, potentially increasing user throughput and reducing operating costs.

Uses for the doubly labeled water method

When introduced as a novel method more than 50 years ago, the DLW dosage was large, rendering the method as a novel, but expensive method to measure the total energy expenditure. Advances in measuring isotope abundances, however, have reduced the dose requirements and the DLW turned into a revolutionary tool in nutrition science, which is now used to measure energy expenditure in thousands of humans. Its reliability and accuracy in measuring the energy requirement in free-living conditions made it the method of choice against which others are validated. Therefore, the DLW technique is being largely used to answers questions related to TEE and its components, body composition, energy requirements at the different stages of life, and also under conditions of health, such as obesity or chronic disease, and to validate dietary assessment and physical activity instruments (**Fig. 4**).

Due to financial limitations and subject burdens, in the clinical assessment, clinical epidemiological investigations, often rely on the use of questionnaires for data collection. It is crucial that such self-reported energy intake and physical activity questionnaires provide accurate estimations to accurately assess the relationship between diet, physical activity, and chronic disease. Therefore, the validation of self-reported questionnaires is a common application of the DLW technique. Studies comparing such questionnaires against DLW revealed them to be subject to errors or misreporting (Burrows et al., 2019; Lee et al., 2011). A recent review of various dietary intake methods established that, regardless of the type of dietary survey instrument used, participants regularly underreported their caloric intake by anywhere from 15 to 34% (Park et al., 2018). Not dissimilarly physical activity monitors have been proposed as objective methods to assess physical activity energy expenditure to avoid the misreporting errors using questionnaires, however, have shown a high variability of results obtained by Jeran et al.(2016). For this reason data on relationships between diet or physical activity, from self-reported questionnaires and in some cases objective monitors such as accelerometers should be interpreted cautiously.

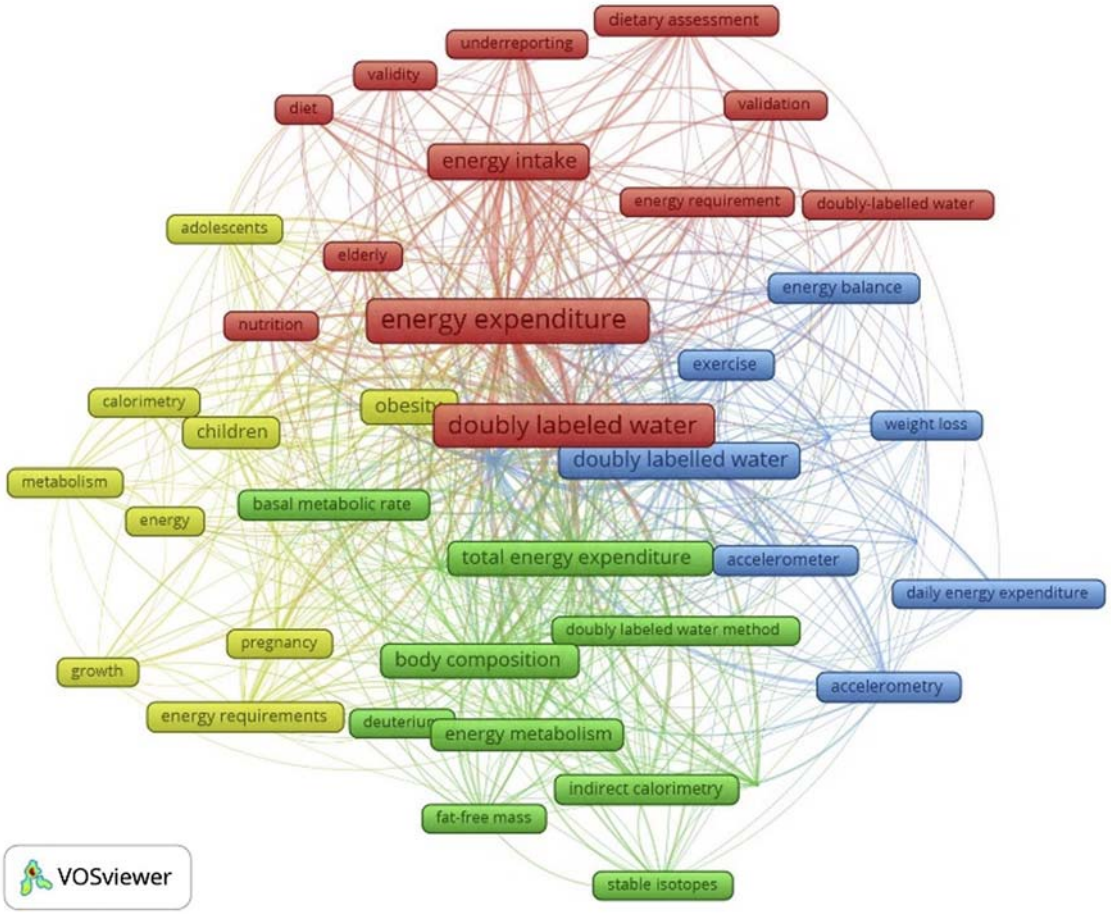


Fig. 4 Cluster map based on bibliometric data of author keyword with a minimum of 15 occurrences. Four clusters, distinguished by colors, described the main fields of application of the DLW technique. The frames' size represents the frequency of the keywords' appearance, and the distance between them is related to their correlation strength. The analyses were conducted using original research articles indexed in the SCOPUS database to study the field of knowledge about the doubly labeled water method and energy metabolism. The results of the search were analyzed by the VOSviewer software (version 1.16.16) that generated a network map based on the co-occurrence matrix.

Historically, energy metabolism and its requisite recommendations have been based on dietary intake data provided largely by various self-report questionnaires, however, as indicated above, the DLW technique has proven that these tools are prone to misreporting (Ravelli and Schoeller, 2020) and thus, are not a reliable basis to establish the human energy requirement guideline. Therefore, one of the most significant roles DLW has played has been its influence in revising the dietary reference intake energy requirements for humans. Indeed, the most recent dietary reference standards for energy requirements were revised using DLW-measured energy requirements instead of dietary intake estimates, the energy requirements of healthy individuals were established by national and international organizations have been reassessed yielding higher values. Based on the DLW data, adult human energy intake requirements were increased by 10% compared with estimated energy requirements published before 1990. An exception is the energy requirement for infants and neonates during the first two years of life, which were decreased by 10–15%.

As obesity has taken on the epidemic and global status, it is not surprising that DLW has been essential in evaluating its impact and studying its etiology. Of particular interest was the finding that obese individuals had on average a higher TEE than their age- and height-matched lean counterparts (**Fig. 5**). This was contrary to the popular hypothesis of the 1990s that obesity developed in individuals due to low energy expenditure as compared with their corresponding age- and height-matched lean counterparts based on energy intake data. Another theory that was not supported by data from the DLW technique was that weight regain after an intentional weight loss was due to a metabolic adaptation leading to dramatically low energy expenditure among post-obese patients. After assessing changes in the TEE of participants with obesity who underwent bariatric surgery, and thus achieved substantial weight loss, researchers found that TEE was decreased at 6 months after surgery due to a combination of a metabolic adaptation and smaller body size. At one year after the procedure, however, the metabolic adaptation no longer existed (Ravelli et al., 2019; Wolfe et al., 2018). In a larger cross-sectional data analysis based on DLW assessment of energy expenditure, it was possible to conclude that the rapid increase in human obesity over the last 50 years was largely an effect of increases in dietary intake and not a product of lower energy expenditure (Swinburn et al., 2009).

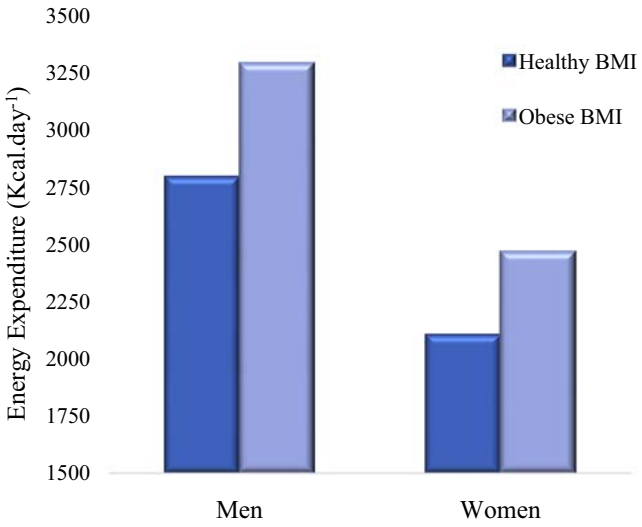


Fig. 5 Comparison of energy expenditure among men and women in individuals with healthy (18.5–24.9 kg m^{-2}) vs obese (>30.0 kg m^{-2}) BMI as measured by DLW. Data are from the International Atomic Energy Agency DLW database, with more than 3700 accumulated subjects in studies performed in 23 countries (Speakman et al., 2019).

In contrast to the application of DLW to studies to understand the obesity epidemic, DLW has also assessed humans' limits on human energy expenditure during conditions of extreme physical activity and endurance events. For example, cyclists during the Giro d'Italia (Plasqui et al., 2019), athletes over one athletic season (Silva et al., 2017), tri-athletes in international ironman competitions (Cuddy et al., 2010), and synchronized swimmers (Ebine et al., 2000) have all utilized DLW (**Fig. 6**). Additionally, this method has been used to define the energy and water requirements of those individuals who have participated in mountain climbing (Westerterp et al., 2000), Alaska hunting expeditions (Coker et al., 2018), as well as various military training exercises of the US Special Forces (Johnson et al., 2018; Margolis et al., 2014).

Not constrained to only a single physical event, nutritional epidemiologists have used DLW to track body composition changes and energy expenditures in groups of individuals across countries, continents, societies, and economic conditions. A meta-analysis of 4972 data from 98 studies, on which 14 were from low or middle human development index, investigated the differences of energy-expenditure variables between countries' development status and their populations' body composition (Dugas et al., 2011). Authors showed despite the differences in body sizes between populations living in developing and industrialized countries, no significant difference was observed between adjusted TEE for weight and age or physical activity level. These findings again suggest that diet rather than a decline in physical activity appears to be the main contributor to the current rapid rise in the prevalence of obesity in developed and developing countries. Additionally, the DLW technique can be

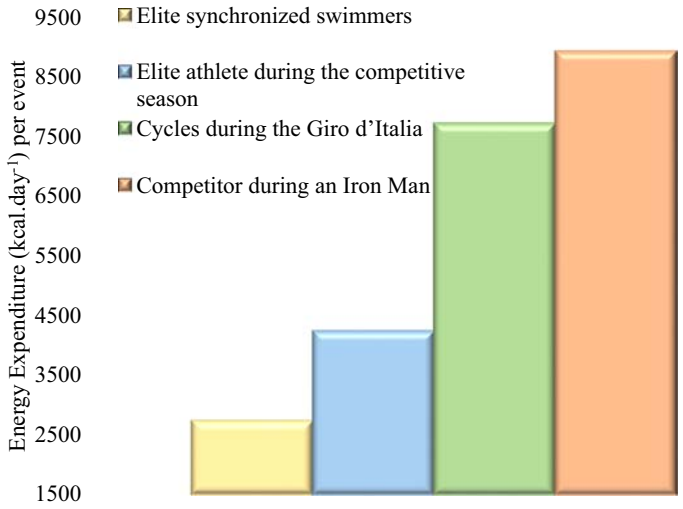


Fig. 6 Energy expenditures of four high-endurance activities as measured by DLW.

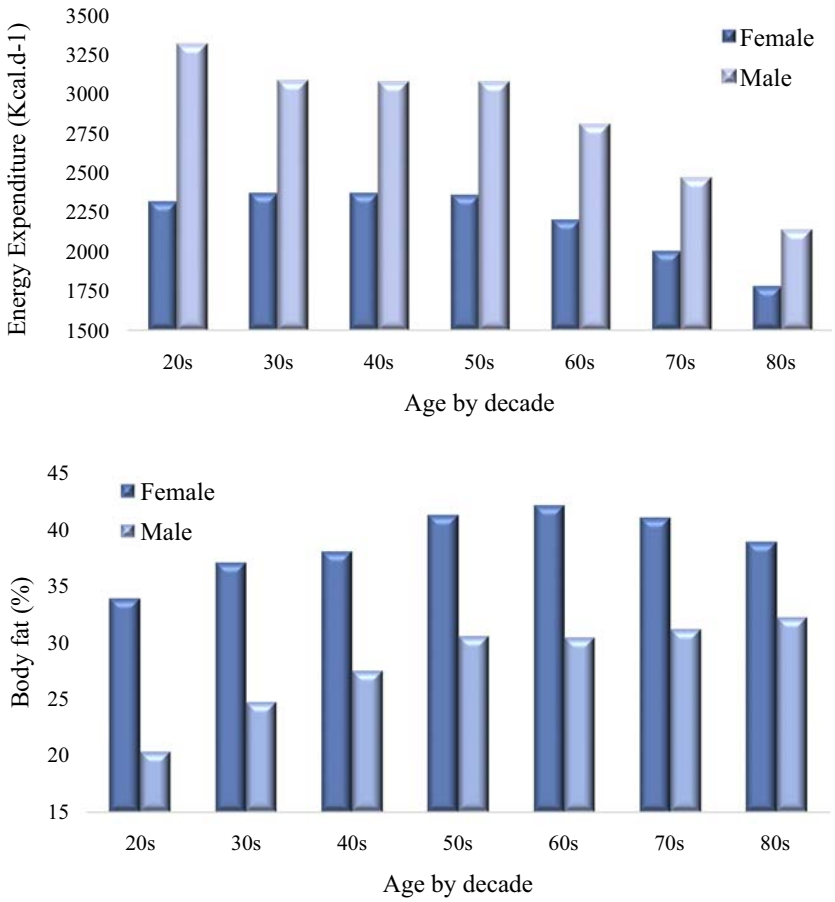


Fig. 7 Comparison between men and women of the effects of age on energy expenditure and body composition. Data are from the International Atomic Energy Agency DLW database, with more than 4800 accumulated subjects in studies performed in 23 countries (Speakman et al., 2019).

utilized *via* meta-analysis to corroborate or refute long-held hypotheses. Combining more than 5000 DLW study participants, the effects of age and sex on body composition and energy expenditure can now be analyzed using the recently compiled IAEA DLW database (Speakman et al., 2019). These data show the interactions between age, fatness, and energy expenditure for men and women (see **Fig. 7**).

The DLW method has been an important tool for researchers in a wide range of applications and has made an impact on the understanding of the human body. As obesity continues to increase globally, the use of DLW to measure body composition and energy metabolism will become even more paramount to health-related research in the future. In countries subjected to the dual burden of obesity and undernutrition, the DLW method has also become an important tool in the study of undernutrition and repletion.

Conclusion

After an analytical improvement of isotopic ratio mass spectrometry over the last 60 years, the application of the doubly labeled water method went from those in small animals to humans with high accuracy (0.5%) and precision (8%). It still has limited availability due to the moderate cost per measurement and the limited number of quality analytical laboratories. The high quality of results, and low burden to the subjects, however, have been instrumental in the development of accurate human energy requirements in countries across the globe. It is recognized as a revolutionary method to study energy metabolism, the DLW is an important tool in understanding the causes and treatment of obesity and malnutrition, and other aspects of human bioenergetics.

References

Burrows, T.L., Ho, Y.Y., Rollo, M.E., Collins, C.E., 2019. Validity of dietary assessment methods when compared to the method of doubly labeled water: a systematic review in adults. Front. Endocrinol. 10, 22.

Coker, R.H., Coker, M.S., Bartlett, L., et al., 2018. The energy requirements and metabolic benefits of wilderness hunting in Alaska. Phys. Rep. 6 (21), e13925.

Coward, W.A., Prentice, A.M., 1985. Isotope method for the measurement of carbon dioxide production rate in man. Am. J. Clin. Nutr. 41 (3), 659–663.
Cuddy, J.S., Slivka, D.R., Hailes, W.S., Dumke, C.L., Ruby, B.C., 2010. Metabolic profile of the ironman world championships: a case study. Int. J. Sports Physiol. Perform. 5 (4), 570–576.
Dugas, L.R., Harders, R., Merrill, S., et al., 2011. Energy expenditure in adults living in developing compared with industrialized countries: a meta-analysis of doubly labeled water studies. Am. J. Clin. Nutr. 93 (2), 427–441.
Ebine, N., Feng, J.Y., Homma, M., Saitoh, S., Jones, P.J., 2000. Total energy expenditure of elite synchronized swimmers measured by the doubly labeled water method. Eur. J. Appl. Physiol. 83 (1), 1–6.
Jeran, S., Steinbrecher, A., Pischon, T., 2016. Prediction of activity-related energy expenditure using accelerometer-derived physical activity under free-living conditions: a systematic review. Int. J. Obes. 40 (8), 1187–1197.
Johnson, C.D., Simonson, A.J., Darnell, M.E., et al., 2018. Energy expenditure and intake during Special Operations Forces field training in a jungle and glacial environment. Appl. Physiol. Nutr. Metab. 43 (4), 381–386.
Lee, P.H., Macfarlane, D.J., Lam, T.H., Stewart, S.M., 2011. Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): a systematic review. Int. J. Behav. Nutr. Phys. Activ. 8, 11.
Margolis, L.M., Murphy, N.E., Martini, S., et al., 2014. Effects of winter military training on energy balance, whole-body protein balance, muscle damage, soreness, and physical performance. Appl. Physiol. Nutr. Metab. 39 (12), 1395–1401.
Park, Y., Dodd, K.W., Kipnis, V., et al., 2018. Comparison of self-reported dietary intakes from the Automated Self-Administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers. Am. J. Clin. Nutr. 107 (1), 80–93.
Plasqui, G., Rietjens, G., Lambriks, L., Wouters, L., Saris, W.H.M., 2019. Energy expenditure during extreme endurance exercise: the Giro d'Italia. Med. Sci. Sports Exerc. 51 (3), 568–574.
Racette, S.B., Schoeller, D.A., Luke, A.H., et al., 1994. Relative dilution spaces of 2H- and 18O-labeled water in humans. Am. J. Physiol. Endocrinol. Metab. 267 (4 Pt 1), E585–E590.
Ravelli, M.N., Schoeller, D.A., 2020. Traditional self-reported dietary instruments are prone to inaccuracies and new approaches are needed. Front. Nutr. 7, 6.
Ravelli, M.N., Schoeller, D.A., Crisp, A.H., et al., 2019. Influence of energy balance on the rate of weight loss throughout one year of Roux-en-Y Gastric Bypass: a Doubly Labeled Water Study. Obes. Surg. 29 (10), 3299–3308.
Sagayama, H., Yamada, Y., Racine, N.M., Shriver, T.C., Schoeller, D.A., 2016. Dilution space ratio of 2H and 18O of doubly labeled water method in humans. J. Appl. Physiol. 120 (11), 1349–1354.
Schoeller, D.A., Ravussin, E., Schutz, Y., et al., 1986. Energy expenditure by doubly labeled water: validation in humans and proposed calculation. Am. J. Physiol. Regul. Integr. Comp. Physiol. 250 (5), R823–R830.
Schoeller, D.A., 1988. Measurement of energy expenditure in free-living humans by using doubly labeled water. J. Nutr. 118 (11), 1278–1289.
Silva, A.M., Matias, C.N., Santos, D.A., et al., 2017. Energy balance over one athletic season. Med. Sci. Sports Exerc. 49 (8), 1724–1733.
Speakman, J.R., Nair, K.S., Goran, M.I., 1993. Revised equations for calculating CO2 production from doubly labeled water in humans. Am. J. Physiol. Endocrinol. Metab. 264 (6), E912–E917.
Speakman, J.R., Pontzer, H., Rood, J., et al., 2019. The International Atomic Energy Agency International Doubly Labeled Water Database: aims, scope and procedures. Ann. Nutr. Metab. 75 (2), 114–118.
Speakman, J.R., Yamada, Y., Sagayama, H., et al., 2021. A standard calculation methodology for human doubly labeled water studies. Cell Rep. Med. 2 (2), 100203.
Speakman, J., 1997. Doubly Labeled Water: Theory and Practice. Springer Science & Business Media.
Swinburn, B.A., Sacks, G., Lo, S.K., et al., 2009. Estimating the changes in energy flux that characterize the rise in obesity prevalence. Am. J. Clin. Nutr. 89 (6), 1723–1728.
Westerterp, K.R., Meijer, E.P., Rubbens, M., Robach, P., Richalet, J.P., 2000. Operation Everest III: energy and water balance. Pflügers Arch. 439 (4), 483–488.
Wolfe, B.M., Schoeller, D.A., McCrady-Spitzer, S.K., et al., 2018. Resting metabolic rate, total daily energy expenditure, and metabolic adaptation 6 months and 24 months after bariatric surgery. Obesity 26 (5), 862–868.

Relevant website

http://www-pub.iaea.org/MTCD/publications/PDF/Pub1370_web.pdf, IAEA Human Health Series No. 3: Assessment of Body Composition and Total Energy Expenditure in Humans Using Stable Isotope Techniques.

Food choice: Behavioral aspects

Barbara J. Rolls[a], **Kathleen L. Keller**[a,b], **John E. Hayes**[b], **and Travis D. Masterson**[a], [a] Department of Nutritional Sciences, The Pennsylvania State University, University Park, PA, United States; and [b] Department of Food Science, The Pennsylvania State University, University Park, PA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of D.R. Just, B. Wansink, Food Choice: Behavioral Aspects, In Encyclopedia of Human Nutrition (Third Edition), edited by Benjamin Caballero, Academic Press, 2013, https://doi.org/10.1016/B978-0-12-375083-9.00115-X.

Introduction 62
Questionnaires to assess an individual's motives for food choices 63
The assessment of food choices 63
Sensory-specific satiety: an example of the utility of laboratory-based studies 64
Taste as a driver of food choice 64
The influence of liking, wanting, and motivation on food choice 64
Techniques used by food producers to understand consumer needs and food choices 65
Food choices in the "real world" 66
Technology and innovation in the study of food choice behavior 66
Summary: changing food choices to improve nutrition and health 67
Funding 67
References 68

Key points/objectives

- The goal of this review is to provide an understanding of the competing factors that influence an individual's food choices
- A summary is provided of methods used to understand psychological, cognitive, and biological factors that influence food choices
- The effects of taste, liking, wanting and motivation on food choices are described
- Consideration is given to techniques used by food producers to understand consumer needs and food choices
- Advances in technology for the study of food choice behavior are described

Introduction

Determinants of food choice are broad and span multiple levels of the socioecological model (Contento, 2008). At the outermost level, food choices are influenced by policies, prices, accessibility, and regulatory practices. Working from this level, food choices are influenced by settings and context (e.g., restaurants and workplaces), as well as by interpersonal factors, including family, cultural, and peer groups (Leng et al., 2017). Each of these factors plays a role in determining both the types and amounts of foods chosen. For a comprehensive picture, the reader is directed to a systematic review by (Chen et al., 2020), which includes available conceptual models on food choice and summarizes consistencies across models on the role of food-related influences, individual differences, and societal factors. The focus of the present review is to provide an overview of various methods used to understand psychological, cognitive, and biological factors that influence food choices at the individual level so that strategies for improving diet quality can be developed and improved.

A recent conceptual analysis, meant to reach consensus on terms used in the study of dietary behavior, defined food choice as an "umbrella term for behaviors and other factors occurring before food actually reaches the mouth" (Stok et al., 2018). Some behaviors included within this definition include monetary aspects (i.e., purchasing, share of income, and willingness to pay), food preparation, intended use, and overall food preferences. However, as such decisions are made prior to consumption, they may not reflect the foods chosen or the amounts of foods consumed once eating commences. Even if a specific food is highly liked by an individual, this does not necessarily mean that they will eat that food when presented with multiple competing foods at a meal or snack. While pre-meal decisions certainly play a key role in determining the types and amounts of foods selected, food choices during consumption of a multi-item meal evolve as the hedonic appeal of foods changes and biological mechanisms engage. Therefore, in this review, behavioral aspects of food choice will include those occurring both before and during consumption.

At the population or community level, standard methods include sophisticated demographically weighted surveys such as The National Health and Nutrition Examination Survey (NHANES) that ask a large number of individuals to record their intake over

several days (Ahluwalia et al., 2016). Such tools, even with problems of inaccurate reporting, can provide a general understanding of the average food choices of various groups of people and can be used to develop broad policies and recommendations to improve diet quality. To understand individual behavioral determinants of food choices, more controlled and specific methods are required.

Questionnaires to assess an individual's motives for food choices

While questionnaires to assess aspects of food choice such as those related to attitudes and beliefs had been used previously, the first psychometrically validated instrument specifically designed to assess motives for food choices came in 1995 with The Food Choice Questionnaire (FCQ) (Steptoe et al., 1995). This questionnaire consists of 37 items separated into nine factors. Four of the nine factors – sensory appeal, health, convenience, and price – emerged as most important for understanding food choice. The other five factors – weight control, familiarity, mood, natural content, and ethical concern – were reported less frequently. The FCQ remains a standard method for measuring motives for food choices and has been extended to include additional dimensions such as those related to culture, ethical concerns, and sustainability (Onwezen et al., 2019).

While there have been extensions of the original FCQ, a recent study sought to facilitate broader use and reduce participant burden by shortening it. The shortened FCQ retains the original nine factors but includes a single question for each. This short instrument was shown to have comparable predictive validity to the original FCQ and can be used to understand context-specific motives for food choices such as those in home versus restaurant settings (Onwezen et al., 2019). To investigate individual differences in food choice motives, the most suitable version of the FCQ can be administered along with anthropometric measures and other validated questionnaires such as the Three Factor Eating Questionnaire (Stunkard and Messick, 1985). While these questionnaires provide an overview of the motives that typically underlie an individual's food choices, additional measures are needed to provide insight into the types and nutritional characteristics of foods individuals choose.

The assessment of food choices

In addition to the questionnaires that have been developed to assess food choice motives, there are instruments that assess the types of foods individuals say they prefer, along with frequency of consumption. For example, the Fat Preference Questionnaire is a validated instrument used previously to show that the reported preference for high-fat foods declined after participation in a weight loss trial emphasizing fat reduction (Ledikwe et al., 2007). Validation of this instrument was done by comparing responses on the questionnaire to self-reported intake in a free-living cohort, and in participants in a weight-loss trial. Responses on the questionnaire were also compared over two days to foods selected and consumed from lab-based buffets offering foods similar to those on the survey. More generally, buffets that offer participants an array of foods provide a simple and frequently used method to assess food choice. While the number of foods that can be offered in buffet paradigms is limited by cost and logistics, the selection can be tailored to answer specific questions related to food choice. For example, if the goal is to determine whether consumption of a high-fat preload (first course) leads to a shift away from other high-fat foods at a subsequent course, a buffet offered for that course would provide a variety of foods varying in fat content (Hetherington and Rolls, 2018). An advantage of such paradigms is that food intake can be directly measured along with food selection, thus providing broader insight into eating behavior than questionnaires.

As buffet studies with real foods can be costly and labor-intensive, investigators may instead choose to use alternatives, such as a Fake Food Buffet (Bucher et al., 2012). In this paradigm, food replicas are offered to participants who are asked to select a plate of food or snacks that would represent what they would typically eat within a given context. While this results in a food selection paradigm similar to the use of real foods, the use of fake food precludes assessment of actual intake. Similar to buffets with real foods, only a limited set of food options can be presented to participants; however, the use of fake foods does reduce food and labor costs as well as food waste.

When the ability to conduct in-person studies is limited, or other study demands make assessment in the laboratory impractical, remote assessment of food choice can be useful. For example, the Fake Food Buffet has been adapted to a Web Buffet, in which participants are asked to choose the types and amounts of foods they would like for a meal from a set of images (Bucher and Keller, 2015). Other web-based tools have been developed such as the Food Choice Task. This task presents individuals with a series of images and asks them to select one. A benefit of this type of task is that it allows for easy calculation of scores related to an individual's food preferences, such as a fat preference score, based on the types of foods selected during the task (Foerde et al., 2018). While easy to use, it remains unclear how representative these tools are of actual food choices or intake. However, another computer task, The Multiple Food Test, has been shown to be related to actual food choices (Schreiber et al., 2020). The development of additional computerized tools that have validity to predict both real food choices and intake will facilitate the collection of more reproducible food choice data in the future.

This section is not meant to be comprehensive; rather, it is intended to illustrate the types of tasks that have been used to assess food choices. One caveat is that the tasks discussed here were developed and tested primarily with industrialized Western diets. They could, however, be adapted and validated to fit a variety of cultures and contexts. It is also important to note that such controlled studies may not reflect habitual diets, but are instead intended to test specific hypotheses about contextual and individual influences on food choice.

Sensory-specific satiety: an example of the utility of laboratory-based studies

Controlled laboratory-based studies can be used to extend findings from observational studies to gain a deeper understanding of robust influences on both food selection and food intake. For example, while the pleasantness or liking of a food has been shown to be a key determinant of food choice (Hayes, 2020), the degree to which a food is liked is not stable across contexts, or even within a meal. As a food is eaten, the pleasantness evoked by its sensory properties – including taste, texture, appearance, and odor – declines while the pleasantness of other uneaten foods generally remains unchanged. This decrease in the palatability of a food as it is eaten, relative to other foods, is termed sensory-specific satiety (Rolls, 1986). Behaviorally, sensory-specific satiety leads to the termination of consumption of the specific food being eaten and promotes switching to other foods, especially those with different sensory properties. Thus, sensory-specific satiety promotes the selection of a wider variety of foods and can help to ensure that a balance of nutrients is consumed; however, it can also lead to greater intake if multiple types of energy dense foods are available (Raynor and Vadiveloo, 2018; Rolls, 1986). Sensory-specific satiety helps to explain the excess energy intake associated with the availability of variety of tasty, energy dense foods at a restaurant buffet (Johnson and Wardle, 2014). These studies demonstrate why the assessments of food choice should not be limited to pre-consumption measures, and should also consider objectively assessed food intake.

Taste as a driver of food choice

According to the Food Choice Process Model, taste is one of the five factors that influences the decision to consume a food (Furst et al., 1996). In consumer surveys, taste is reliably the most important driver of food choice for most individuals (IFIC, 2020). Therefore, genetically influenced variations in taste perception likely play a role in food choice. Some taste preferences, such as liking for sweetness and disliking of bitterness and sourness, are innate (Desor and Lawrence, 1977; Mennella and Beauchamp, 1996). Evidence on salty taste is more mixed: while it is widely claimed infants do not acquire a preference for saltiness until a few months after birth (Bernstein, 1990), other data indicate some newborns may prefer salty stimuli within a few days of birth (Zinner et al., 2002). Preferences for fat, on the other hand, are learned in childhood (Birch, 2009). These biologically driven preferences are thought to guide humans toward food sources that provide them with energy and protect them from ingesting potentially toxic substances (which are often bitter in taste). What this means for food choice, is that some foods, like bitter fruits and vegetables, are harder to learn to like than others. However, with repeated exposure individuals can learn to like most foods, making experience a critical factor that shapes food choice (Cines and Rozin, 1982).

There are also genetically driven differences in taste receptors that have implications for food choice, food consumption, and even chronic disease risk. The best studied of these phenotypes is the ability to taste bitter thiourea compounds, like phenylthiocarbamide (PTC) and 6-n-propylthiouracil (PROP). Individuals who are less sensitive to the taste of PTC and PROP typically show increased acceptance of bitter and strong-tasting foods. In a recent meta-analysis, variants of *TAS2R38* (the gene responsible for the ability to taste PTC and PROP) (Kim et al., 2003), were associated with perceived bitterness of vegetables, berries, and wine, as well as preferences for sweet taste (Diószegi et al., 2019), consistent with earlier reports (Duffy et al., 2010). However, *TAS2R38* is only one of 25 bitter receptor genes in humans, some of which have been shown to influence bitterness perception, and potentially food choices. For example, variation in the *TAS2R31* gene is associated with perception of bitter side tastes in some sweeteners and with liking for grapefruit juice (Hayes et al., 2015). In addition to genes implicated in bitter perception, there is also convincing evidence that polymorphisms in the *CD36* gene, a fatty acid translocase, are related to the perception of fat and preferences for certain high-fat foods (Keller et al., 2012; Pepino et al., 2012; Sayed et al., 2015). For additional discussion of chemosensory gene variants and food choices, the reader is directed to (Nolden and Feeney, 2020).

The influence of liking, wanting, and motivation on food choice

While innate predispositions for specific tastes clearly exist, the extent to which a food is liked is modified by learning via prior experiences (Prescott, 2020)). Liking plays a key role in determining whether a food is chosen or avoided (Hayes, 2020). However, in the late 1990s, Berridge and Robinson challenged the presumed importance of hedonic responses by differentiating between the concepts of liking and wanting. In their framework, "liking" is an affective reaction to a food based on a combination of prior experience, sensory properties, and internal physiological state, while "wanting" refers to one's drive and motivation to get a food at a given time (Berridge, 1996). Because wanting relates to how motivated someone is to receive a food, it is arguably a more important driver of food choice than liking. Liking is thought to be a more stable hedonic response to a food but wanting varies due to both internal (e.g., hunger, satiety) and external (e.g., advertising, environment) stimuli. For example, one might have a high liking for spaghetti Bolognese, but the motivation to consume this dish (i.e., "wanting") first thing in the morning is probably quite low.

A variety of methods have been used to measure liking and wanting. For example, explicit measures ask participants to report on scales the extent to which they both like and want to eat a particular food item (Pool et al., 2016). However, to improve the ability to differentiate these constructs, some studies have developed implicit association tasks. For example (Finlayson et al., 2007), developed a computer-based paradigm where participants choose which of a pair of foods with different sensory properties (e.g., low-fat, savory, sweet) they would "most like to eat right now." By asking the individual to make choices in response to the same food

multiple times, paired with different foods on each presentation, a relative ranking of wanting can be determined. These types of rankings can be used to make inferences about the foods someone would choose in a variety of situations. For example, a systematic review from (Oustric et al., 2021) suggested liking and wanting for high-fat, energy-dense foods tended to decrease with weight loss, regardless of the type of intervention (e.g., pharmacological, behavioral, surgical). Although the relationship of liking and wanting to weight outcomes needs confirmation, these approaches will allow investigators to develop more targeted interventions in the future.

In addition to liking and wanting, the value placed on a food also depends on the availability of other foods or activities. For example, peanut butter cookies might be highly liked, but if given the choice of a peanut butter cookie and premium Belgian chocolate, the chocolate might ultimately be chosen. In order to assess the relative motivation to obtain a food, one can measure how hard someone will work to gain access to it. Epstein and colleagues adapted methods from Behavioral Economics to develop the Relative Reinforcing Value of food task (RRV task) to operationally measure how hard someone will work to obtain a food relative to some other food or activity (e.g., watching TV, exercise) (Epstein et al., 2007). In the RRV task, "work" is measured as the number of times participants will press a keyboard key to obtain portions of a desired snack relative to an alternative (either another food or activity). The number of key presses required to obtain a desired food increases on a progressive ratio with each portion earned. By comparing foods that vary in sensory or nutritional qualities, one can determine how hard someone will work to obtain a food, and from this, can infer food choice under a variety of physiological and context-related situations.

Because food choices are not constant and are influenced by both the context and the other types of foods available, environmental settings can be engineered to optimally affect dietary behaviors. These manipulations in choice architecture can be used to nudge consumers into making healthier food choices (Szaszi et al., 2017). For example, by creating environments where fruits and vegetables are easily accessible, and identifying appropriate nonfood reinforcers to substitute for high-energy dense foods (e.g., music, reading), one can engineer environments that are optimally suited for making healthier choices (Epstein, 2020). The long-term goal of these nudge strategies is to develop more personalized approaches to improve chronic health.

Techniques used by food producers to understand consumer needs and food choices

Substantial research on food choice is performed by food and beverage companies that make high volume, low margin products like snacks, soft drinks, candy, breakfast cereals, and ready to eat meals. The majority of this work is not reported publicly, as the results are proprietary and are seen as a competitive advantage. The methods used can be divided into those meant to assist in product optimization (i.e., classic sensory evaluation techniques, including blind tests with naïve consumers) and those meant to understand the broader needs of the consumer (i.e., consumer insights). These methods are reviewed in this section, along with their implications for understanding food choice.

One mainstay technique for product optimization – the 9-point hedonic scale – was first developed by researchers focused on testing military rations (e.g. (Meiselman and Schutz, 2003)). As they noted in 1955, "Even foods that are extremely well-liked, but by only a small proportion of the consumers, are unsuited for military use (Jones et al., 1955)." The 9-point hedonic scale, which was developed seven decades ago, is still widely used in the food industry, in part due to its simplicity for both investigators and consumers. Ratings from this category scale can be used in food preference surveys (to determine which foods are most liked by a group, as in the original ration selection studies), or to quantitatively compare the degree of liking for foods and beverages prepared/processed in different ways (Hayes et al., 2014), or to allow investigators to understand the optimal preferences of the consumer (Wang et al., 2018).

However, industrial research by food and beverage producers is not limited to blind taste tests with naïve consumers, as it also includes qualitative and quantitative methods that are used to understand food choices, and the motivations that underlie those choices. Many of these techniques are borrowed from the social sciences (e.g., ethnography, focus groups, and semi-structured interviews) or market research and experimental economics (e.g., surveys, conjoint analysis, non-hypothetical auctions). Critically, such studies are not conducted blind, as they are focused on the interaction of the participant with the food product as well as information about the food (e.g., nutritional information, food labeling). Participants are usually given some relevant details about the product (e.g., "low in fat", "a protein-rich food"), and perhaps the context in which the product will be consumed. For example, rather than giving a participant two beer samples and simply asking which they prefer, as in a blind taste test, participants might be prompted to imagine they are at a boardwalk café after a long day at the beach, and asked to evaluate which beer is more thirst quenching. In this way, industrial researchers are able to look beyond simple blind tastings to better assess consumer needs and wants.

Qualitative techniques like focus groups may provide key insights when performed by an experienced moderator. For example, in one such study, a canned food manufacturer was interested in a label change to highlight that vegetables in the can were grown locally. Over several focus groups, it quickly became clear consumers did not care if the vegetables were grown in-state or several states away; rather, they cared that farmers who grew the vegetables had been working with the canner for 2, 3, or even 4 generations. Likewise, academic researchers have used focus groups and semi-structured interviews to better understand beliefs and attitudes about salt use and sodium intake in a high-risk rural population (Smith et al., 2006). They found food lacking salt was seen as "fresh", and critically in this usage, the descriptor "fresh" was in fact pejorative, as "fresh" also implied bland and flavorless to their participants. Use of a traditional quantitative survey to assess beliefs about salt would have missed this key insight.

A method developed by market researchers in the 1970s with substantial value in assessing food choice, particularly when considering competing tradeoffs between product attributes, is conjoint analysis. This technique has become increasingly popular recently for internet-based research. The use of online conjoint analysis has substantially reduced the cost of collecting large samples required for more generalized and informative data. Multiple variants of conjoint analysis exist but, in general, product attributes are grouped into silos, and individual elements within the silo are varied. For example, Brodock and colleagues (Brodock et al., 2021) explored how sweetener type, added sugar, protein content, and fat level, influenced product choices for chocolate milk in adults. In this case, within a specific silo, such as added sugar, the stated amount of sugar was varied across choices. By systematically varying elements from each silo, an exhaustive set of product vignettes was generated and presented to participants and specific product aspects that play a role in consumer food choice were quantified. This provided insight into factors that influence food choice, even when participants were not aware of why they preferred one food over another. From this example, Brodock and colleagues were able to identify three distinct groups of consumers with respect to chocolate milk preferences (Brodock et al., 2021). Notably, these groups, formed solely based on product choices, also differed in measures of several health related behaviors. This type of convergent validity (i.e., revealed preferences in a choice task, and psychometrically validated eating behavior questionnaires) supports the view that such online tasks can provide valid and useful insights into food choices.

In summary, industrial research is not limited to blinded tests with consumers to optimize products, but also includes research on the interaction of consumers with products within a specific context or use case to better understand consumer needs, under the presumption that better understanding of such needs and wants will provide a competitive advantage in the market.

Food choices in the "real world"

Outside of the laboratory setting, food choices can be monitored in environments like grocery stores or restaurants to assess the array of contextual variables that can influence selection. These approaches are useful for the study of consumer behavior in response to environmental manipulations such as menu labels, portion size, and the variety of available choices. However, collecting data in these settings is less useful when trying to determine how food choices relate to individual characteristics of consumers. The collection of personal data on individuals requires informed consent and obtaining such consent largely compromises the ecological validity of such studies.

To simulate a "real-world" experience while still offering the control afforded by the laboratory, some investigators have established test restaurants or supermarkets that allow food selection to be observed in an ecologically relevant setting. In such facilities, a combination of food photography and weighing the foods selected can be used to answer questions as to how food choices vary with both individual and food characteristics (Hinton et al., 2013). However, such simulation of the "real-world" can be expensive and not practical for widespread application. A more sustainable approach could be to incorporate hypothesis-driven studies related to food choice into the food development and testing done, for example, at institutes for culinary training, some of which include an experimental restaurant or partnerships with restaurants, school canteens, or bars in a variety of locations (dos Santos et al., 2020; Miele et al., 2021).

Technology and innovation in the study of food choice behavior

Computerized tools can extend the assessment of food choice beyond the laboratory. Tools such as the Web Buffet (Bucher et al., 2012), the Food Choice Task (Foerde et al., 2018; Masterson et al., 2019; Hare et al., 2009), or the Multiple Food Test (Schreiber et al., 2020) provide participants with a variety of food choice situations. These paradigms allow for multiple comparisons to be made within a short period of time. Additionally, specific food comparisons (e.g., low-vs. high-energy dense foods) can be tested to better understand the sensory and nutritional characteristics that drive food choice. An advantage of these online tools is that they can be administered to large numbers of participants at a low cost. Such large-scale data collection can aid in the identification of individual variability related to food choice (e.g., (Brodock et al., 2021), and this could be used to develop tailored interventions to improve dietary intake.

One method originally developed by addiction researchers to study behavior in a participant's usual environment, Ecological Momentary Assessment (EMA), has promise for the study of food choice. Broadly, EMA allows intensive data collection about a participant's behavior to be collected in real time during their daily routine (Shiffman et al., 2008). The evaluation of real-world food choices via EMA helps to minimize recall and response biases (Shiffman et al., 2008). Further, computerized food choice tools inherently provide quantitative metrics (e.g., reaction time) that may give additional insight into the cognitive processes underlying food choice (Masterson et al., 2019). For example, one study that used a Food Choice Task in combination with functional magnetic resonance imaging (fMRI) demonstrated that weight conscious women with higher self-control were able to make choices more quickly between healthy and palatable, energy dense foods (Sullivan et al., 2015). Additionally, using the same task, it has been shown that individuals demonstrating higher levels of dietary restraint make healthier food choices, and they do so more quickly, than those with lower levels of restraint. Use of these digital food choice assessments alongside EMA paradigms may provide further insight into how external (e.g., food marketing exposure) and internal factors (e.g., emotional state) can modify dietary self-control and affect food choice over the course of a day. Overall, these digital tools can provide investigators increased flexibility, insight, and potential to understand food choice in the real world.

Hand-held devices and smartphones also show promise for deploying and monitoring interventions designed to help individuals make better food choices. For example, these devices have been used to deliver personalized interventions and educational materials that positively modify individuals' food choices (see (Chen et al., 2020) for a detailed review). Specifically, personalized programs have been shown to increase the intake of fruit, vegetables, and dairy, as well as decrease the intake of dietary fat and total energy intake (Acharya et al., 2011; Ambeba et al., 2015; Atienza et al., 2008; Burke et al., 2011; Olson et al., 2008). Additionally, dietary and physical activity self-monitoring applications on smartphones can help users focus on their daily food choices, leading them to modify their intake to achieve dietary goals (Duncan et al., 2014; Ipjian and Johnston, 2017; Mummah et al., 2017). Future use of these applications will provide interventionists and participants more precise and immediate feedback on the participant's cognitive state (i.e., current level of dietary restraint) and how that may influence their food choices in the moment.

Another emerging technology with promise for studying food choice is immersive virtual reality (iVR). Using Immersive iVR, an individual can enter a computer-generated environment where they can physically move around and interact with objects like foods and plates. Recent computational improvements provide a sense of realism not possible previously, improving the ecological relevance of data that are collected (Parsons and Rizzo, 2008). Additionally, researchers can control and manipulate environmental factors that potentially influence decision making within the iVR environment, holding these variables constant across individuals (Cheah et al., 2020; Cheah et al., 2019). Also, behavioral variables are measured continuously and automatically. Such measurements include the order, timing, and amount of food selected, along with psychophysiological variables of interest (e.g., eye movement, brain response, etc.) (Marcum et al., 2018). Additionally, these variables can be collected and analyzed without the need to record and manually annotate videos. Recent studies have demonstrated the utility of iVR food buffets. For example, food selections made in iVR were similar to those made in fake food buffets (Ung et al., 2018) and actual food buffets (Cheah et al., 2020; Persky et al., 2018), and the method is also being used to examine food choice in virtual supermarkets (Ruppert, 2011). In addition to making similar food choices to real world situations, participants in these studies also report that the iVR environment feels natural and that decisions made during these paradigms represent their real-world behavior (Lombart et al., 2019; Ruppert, 2011; Siegrist et al., 2019). Additionally, recent work has shown that physiological responses to food in iVR environments are similar to responses observed when participants are viewing food pictures or actual foods (Gorini et al., 2010; Ledoux et al., 2013; van der Waal et al., 2021).

While techniques like EMA and iVR have only recently been adapted for food choice research, they show promise to increase understanding of factors that drive food choices. These technologies provide new ways to investigate food choice behavior under conditions that allow for greater flexibility and reach than traditional laboratory studies. Specifically, they can provide information about a wide range of foods tested under a variety of conditions, and with participants that span the globe. A goal for these methods is that they will expand understanding of the influences on food choice under ecologically relevant conditions.

Summary: changing food choices to improve nutrition and health

This review has covered a variety of tools and methods that can be applied to help investigators understand food choice behavior. Validated questionnaires are widely used to characterize an individual's food choices in a variety of contexts and the motives for making these choices. Such instruments can be administered online or by electronic devices, making it feasible to collect data on large numbers of individuals in a wide variety of contexts. Data derived from large-scale data collection procedures can be integrated with traditional controlled laboratory studies to test specific hypotheses related to how food choice behavior affects food intake. Overall, the practical goal of most studies on food choice in academic settings is to improve diets and nutrition-related health. Similarly industrial researchers frequently focus on improved nutritional profiles for food products (i.e., reduction in added sugar or salt), but this must be balanced against a need to make and sell products that remain liked by consumers to be competitive with alternatives.

Many strategies have been used to encourage healthier food choices. Environmental approaches involve making healthy foods taste better, be more accessible, and more affordable. Behavioral nudges that remind consumers to choose heathier options are becoming easier to test with technological and methodological advances (Stuber et al., 2021). An umbrella review of systematic reviews on food choice found that nutrition education is the most common strategy used to alter food choices, followed by nutrition labels (Perez-Cueto, 2019). However, nutrition education alone has limited impact on food choices, suggesting that newer methods are needed to help individuals improve their food choices. Many opportunities exist to test the effectiveness of integrated approaches that involve methods discussed in this review. An understanding of the competing factors that influence an individual's food choices, from basic biology to interactions with the food environment, will facilitate a personalized approach to improving food choices.

Funding

BJR receives salary support from the National Institutes of Health, National Institutes of Diabetes and Digestive and Kidney Disorders [R01DK082580]. KLK receives salary support from the National Institute of Food and Agriculture and Hatch Act Appropriations [Project PEN04708] and from the National Institutes of Health, National Institutes of Diabetes and Digestive and Kidney Disorders [R01DK110060]. JEH receives salary

support from the United States Department of Agriculture via the National Institute of Food and Agriculture and Hatch Act Appropriations [Project PEN04708 and Accession # 1,019,852], and from the National Institutes of Health Office of the Director [NIH OD], and the National Institutes of Deafness and Other Communications Disorders [U01DC019573]. TDM receives salary support from the National Institute of Child Health and Human Development [R01HD092604]. The findings and conclusions in this publication are those of the authors, and do not represent the views of the U.S. Department of Agriculture, or the National Institutes of Health, and do not represent any U.S. Government determination, position or policy.

References

Acharya, S.D., Elci, O.U., Sereika, S.M., Styn, M.A., Burke, L.E., 2011. Using a personal digital assistant for self-monitoring influences diet quality in comparison to a standard paper record among overweight/obese adults. J. Am. Diet Assoc. 111, 583–588.

Ahluwalia, N., Dwyer, J., Terry, A., Moshfegh, A., Johnson, C., 2016. Update on NHANES dietary data: focus on collection, release, analytical considerations, and uses to inform public policy. Adv. Nutr. 7, 121–134.

Ambeba, E.J., Ye, L., Sereika, S.M., Styn, M.A., Acharya, S.D., Sevick, M.A., Ewing, L.J., Conroy, M.B., Glanz, K., Zheng, Y., Goode, R.W., Mattos, M., Burke, L.E., 2015. The use of mHealth to deliver tailored messages reduces reported energy and fat intake. J. Cardiovasc. Nurs. 30, 35–43.

Atienza, A.A., King, A.C., Oliveira, B.M., Ahn, D.K., Gardner, C.D., 2008. Using hand-held computer technologies to improve dietary intake. Am. J. Prev. Med. 34, 514–518.

Bernstein, I.L., 1990. Salt preference and development. Dev. Psychol. 26, 552–554.

Berridge, K.C., 1996. Food reward: brain substrates of wanting and liking. Neurosci. Biobehav. Rev. 20, 1–25.

Birch, L.L., 2009. Children's preferences for high-fat foods. Nutr. Rev. 50, 249–255.

Brodock, J.L., Hayes, J.E., Masterson, T.D., Hopfer, H., 2021. Differences in preferred fat level, sweetener type, and amount of added sugar in chocolate milk in a choice task relate to physical activity and orthorexia. Appetite 163, 105214.

Bucher, T., Keller, C., 2015. The web-buffet - development and validation of an online tool to measure food choice. Publ. Health Nutr. 18, 1950–1959.

Bucher, T., van der Horst, K., Siegrist, M., 2012. The fake food buffet - a new method in nutrition behaviour research. Br. J. Nutr. 107, 1553–1560.

Burke, L.E., Conroy, M.B., Sereika, S.M., Elci, O.U., Styn, M.A., Acharya, S.D., Sevick, M.A., Ewing, L.J., Glanz, K., 2011. The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial. Obesity 19, 338–344.

Cheah, C.S.L., Barman, S., Vu, K.T.T., Jung, S.E., Mandalapu, V., Masterson, T.D., Zuber, R.J., Boot, L., Gong, J., 2020. Validation of a virtual reality buffet environment to assess food selection processes among emerging adults. Appetite 153, 104741.

Cheah, C.S.L., Kaputsos, S.P., Mandalapu, V., Tran, T., Barman, S., Jung, S.E., Vu, K.T.T., Masterson, T.D., Zuber, R., Boot, L., Gong, J., 2019. Neurophysiological variations in food decision-making within virtual and real environments. 2019 IEEE EMBS International Conference on Biomedical and Health Informatics.

Chen, Y., Perez-Cueto, F.J.A., Giboreau, A., Mavridis, I., Hartwell, H., 2020. The promotion of eating behaviour change through digital interventions. Int. J. Environ. Res. Publ. Health 17, 7488.

Cines, B.M., Rozin, P., 1982. Some aspects of the liking for hot coffee and coffee flavor. Appetite 3, 23–34.

Contento, I.R., 2008. Nutrition education: linking research, theory, and practice. Asia Pac. J. Clin. Nutr. 17, 176–179.

Desor, J.A., Lawrence, S.G., 1977. Infants, Children, and Adults. Taste and development: the genesis of sweet preference, 77. 1068. US Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health.

Diószegi, J., Llanaj, E., Ádány, R., 2019. Genetic background of taste perception, taste preferences, and its nutritional implications: a systematic review. Front. Genet. 10, 1272.

dos Santos, Q., Perez-Cueto, F.J.A., Rodrigues, V.M., Appleton, K., Giboreau, A., Saulais, L., Monteleone, E., Dinnella, C., Brugarolas, M., Hartwell, H., 2020. Impact of a nudging intervention and factors associated with vegetable dish choice among European adolescents. Eur. J. Nutr. 59, 231–247.

Duffy, V.B., Hayes, J.E., Davidson, A.C., Kidd, J.R., Kidd, K.K., Bartoshuk, L.M., 2010. Vegetable intake in college-aged adults is explained by oral sensory phenotypes and TAS2R38 genotype. Chemosens. Percept. 3, 137–148.

Duncan, M., Vandelanotte, C., Kolt, G.S., Rosenkranz, R.R., Caperchione, C.M., George, E.S., Ding, H., Hooker, C., Karunanithi, M., Maeder, A.J., Noakes, M., Tague, R., Taylor, P., Viljoen, P., Mummery, W.K., 2014. Effectiveness of a web- and mobile phone-based intervention to promote physical activity and healthy eating in middle-Aged males: randomized controlled trial of the manup study. J. Med. Internet Res. 16, e3107.

Epstein, L.H., 2020. Choice is relative: Reinforcing value of food and activity in obesity treatment. Am. Psychol. 75, 139.

Epstein, L.H., Leddy, J.J., Temple, J.L., Faith, M.S., 2007. Food reinforcement and eating: a multilevel analysis. Psychol. Bull. 133, 884–906.

Finlayson, G., King, N., Blundell, J.E., 2007. Liking vs. wanting food: importance for human appetite control and weight regulation. Neurosci. Biobehav. Rev. 31, 987–1002.

Foerde, K., Gianini, L., Wang, Y., Wu, P., Shohamy, D., Walsh, B.T., Steinglass, J.E., 2018. Assessment of test-retest reliability of a food choice task among healthy individuals. Appetite 123, 352–356.

Furst, T., Connors, M., Bisogni, C.A., Sobal, J., Falk, L.W., 1996. Food choice: a conceptual model of the process. Appetite 26, 247–266.

Gorini, A., Griez, E., Petrova, A., Riva, G., 2010. Assessment of the emotional responses produced by exposure to real food, virtual food and photographs of food in patients affected by eating disorders. Ann. Gen. Psychiatr. 9, 30.

Hare, T.A., Camerer, C.F., Rangel, A., 2009. Self-control in decision-making involves modulation of the vmPFC valuation system. Science 324, 646–648.

Hayes, J.E., 2020. Influence of sensation and liking on eating and drinking. In: Meiselman, H.L. (Ed.), Handbook of Eating and Drinking: Interdisciplinary Perspectives. Springer Nature, Switzerland, pp. 1–25.

Hayes, J.E., Feeney, E.L., Nolden, A.A., McGeary, J.E., 2015. Quinine bitterness and grapefruit liking associate with allelic variants in TAS2R31. Chem. Senses 40, 437–443.

Hayes, J.E., Raines, C.R., DePasquale, D.A., Cutter, C.N., 2014. Consumer acceptability of high hydrostatic pressure (HHP)-treated ground beef patties. LWT - Food Sci. Technol. (Lebensmittel-Wissenschaft -Technol.) 56, 207–210.

Hetherington, M.M., Rolls, B.J., 2018. Favouring more rigour when investigating human eating behaviour is like supporting motherhood and apple pie: a response to Robinson, Bevelander, Field, and Jones (2018). Appetite 130, 330–333.

Hinton, E.C., Brunstrom, J.M., Fay, S.H., Wilkinson, L.L., Ferriday, D., Rogers, P.J., De Wijk, R., 2013. Using photography in "The Restaurant of the Future": a useful way to assess portion selection and plate cleaning? Appetite 63, 31–35.

IFIC [International Food Information Council], 2020. The 2020 Food and Health Survey. https://www.foodinsight.org/2020-food-and-health-survey/. (Accessed 9 March 2021).

Ipjian, M.L., Johnston, C.S., 2017. Smartphone technology facilitates dietary change in healthy adults. Nutrition 33, 343–347.

Johnson, F., Wardle, J., 2014. Variety, palatability, and obesity. Adv. Nutr. 5, 851–859.

Jones, L.V., Peryam, D.R., Thurstone, L.L., 1955. Development of a scale for measuring soldiers food preferences. J. Food Sci. 20, 512–520.

Keller, K.L., Liang, L.C., Sakimura, J., May, D., Van Belle, C., Breen, C., Driggin, E., Tepper, B.J., Lanzano, P.C., Deng, L., Chung, W.K., 2012. Common variants in the CD36 gene are associated with oral fat perception, fat preferences, and obesity in African Americans. Obesity 20, 1066–1073.

Kim, U.K., Jorgenson, E., Coon, H., Leppert, M., Risch, N., Drayna, D., 2003. Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. Science 299, 1221–1225.

Ledikwe, J.H., Ello-Martin, J., Pelkman, C.L., Birch, L.L., Mannino, M.L., Rolls, B.J., 2007. A reliable, valid questionnaire indicates that preference for dietary fat declines when following a reduced-fat diet. Appetite 49, 74–83.

Ledoux, T., Nguyen, A.S., Bakos-Block, C., Bordnick, P., 2013. Using virtual reality to study food cravings. Appetite 71, 396–402.
Leng, G., Adan, R.A.H., Belot, M., Brunstrom, J.M., De Graaf, K., Dickson, S.L., Hare, T., Maier, S., Menzies, J., Preissl, H., Reisch, L.A., Rogers, P.J., Smeets, P.A.M., 2017. The determinants of food choice. Proc. Nutr. Soc. 76, 316–327.
Lombart, C., Millan, E., Normand, J.M., Verhulst, A., Labbé-Pinlon, B., Moreau, G., 2019. Consumer perceptions and purchase behavior toward imperfect fruits and vegetables in an immersive virtual reality grocery store. J. Retailing Consum. Serv. 48, 28–40.
Marcum, C.S., Goldring, M.R., McBride, C.M., Persky, S., 2018. Modeling dynamic food choice processes to understand dietary intervention effects. Ann. Behav. Med. 52, 252–261.
Masterson, T.D., Brand, J., Lowe, M.R., Metcalf, S.A., Eisenberg, I.W., Emond, J.A., Gilbert-Diamond, D., Marsch, L.A., 2019. Relationships among dietary cognitive restraint, food preferences, and reaction times. Front. Psychol. 10, 2256.
Meiselman, H.L., Schutz, H.G., 2003. History of food acceptance research in the US Army. Appetite 40, 199–216.
Mennella, J., Beauchamp, G., 1996. The early development of human flavor preferences. In: Capaldi, E.D. (Ed.), Why We Eat what We Eat. The Psychology of Eating,. American Psychological Association, Washington, D.C., pp. 83–112
Miele, N.A., Giboreau, A., Almli, V.L., 2021. The temporality of eating behavior as a discriminant tool to characterize consumers: temporal dominance of behavior applied to bread consumption during a restaurant meal in France. Food Qual. Prefer. 92, 104224.
Mummah, S., Robinson, T.N., Mathur, M., Farzinkhou, S., Sutton, S., Gardner, C.D., 2017. Effect of a mobile app intervention on vegetable consumption in overweight adults: a randomized controlled trial. Int. J. Behav. Nutr. Phys. Activ. 14, 1–10.
Nolden, A.A., Feeney, E.L., 2020. Genetic differences in taste receptors: implications for the food industry. Annu. Rev. Food Sci. Technol. 11, 183–204.
Olson, A.L., Gaffney, C.A., Lee, P.W., Starr, P., 2008. Changing adolescent health behaviors. The healthy teens counseling approach. Am. J. Prev. Med. 35, S359–S364.
Onwezen, M.C., Reinders, M.J., Verain, M.C.D., Snoek, H.M., 2019. The development of a single-item food choice questionnaire. Food Qual. Prefer. 71, 34–45.
Oustric, P., Beaulieu, K., Casanova, N., O'connor, D., Gibbons, C., Hopkins, M., Blundell, J., Finlayson, G., 2021. Food liking but not wanting decreases after controlled intermittent or continuous energy restriction to $\geq$5% weight loss in women with overweight/obesity. Nutrients 13, 1–14.
Parsons, T.D., Rizzo, A.A., 2008. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. J. Behav. Ther. Exp. Psychiatr. 39, 250–261.
Pepino, M.Y., Love-Gregory, L., Klein, S., Abumrad, N.A., 2012. The fatty acid translocase gene CD36 and lingual lipase influence oral sensitivity to fat in obese subjects. JLR (J. Lipid Res.) 53, 561–566.
Perez-Cueto, 2019. An umbrella review of systematic reviews on food choice and nutrition published between 2017 and-2019. Nutrients 11, 2398.
Persky, S., Goldring, M.R., Turner, S.A., Cohen, R.W., Kistler, W.D., 2018. Validity of assessing child feeding with virtual reality. Appetite 123, 201–207.
Pool, E., Sennwald, V., Delplanque, S., Brosch, T., Sander, D., 2016. Measuring wanting and liking from animals to humans: a systematic review. Neurosci. Biobehav. Rev. 63, 124–142.
Prescott, J., 2020. Development of food preferences. In: Meiselman, H.L. (Ed.), Handbook of Eating and Drinking: Interdisciplinary Perspectives. Springer Nature, Switzerland, pp. 199–217.
Raynor, H.A., Vadiveloo, M., 2018. Understanding the relationship between food variety, food intake, and energy balance. Curr. Obes. Rep. 7, 68–75.
Rolls, B.J., 1986. Sensory-specific satiety. Nutr. Rev. 44, 93–101.
Ruppert, B., 2011. New directions in the use of virtual reality for food shopping: marketing and education perspectives. J. Diabetes Sci. Technol. 5, 315–318.
Sayed, A., Šerý, O., Plesnik, J., Daoudi, H., Rouabah, A., Rouabah, L., Khan, N.A., 2015. CD36 AA genotype is associated with decreased lipid taste perception in young obese, but not lean, children. Int. J. Obes. 39, 920–924.
Schreiber, M., Bucher, T., Collins, C.E., Dohle, S., 2020. The Multiple Food Test: development and validation of a new tool to measure food choice and applied nutrition knowledge. Appetite 150, 104647.
Shiffman, S., Stone, A.A., Hufford, M.R., 2008. Ecological momentary assessment. Annu. Rev. Clin. Psychol. 4, 1–32.
Siegrist, M., Ung, C.Y., Zank, M., Marinello, M., Kunz, A., Hartmann, C., Menozzi, M., 2019. Consumers' food selection behaviors in three-dimensional (3D) virtual reality. Food Res. Int. 117, 50–59.
Smith, S.L., Quandt, S.A., Arcury, T.A., Wetmore, L.K., Bell, R.A., Vitolins, M.Z., 2006. Aging and eating in the rural, southern United States: beliefs about salt and its effect on health. Soc. Sci. Med. 62, 189–198.
Steptoe, A., Pollard, T.M., Wardle, J., 1995. Development of a measure of the motives underlying the selection of food: the food choice questionnaire. Appetite 25, 267–284.
Stok, F.M., Renner, B., Allan, J., Boeing, H., Ensenauer, R., Issanchou, S., Kiesswetter, E., Lien, N., Mazzocchi, M., Monsivais, P., Stelmach-Mardas, M., Volkert, D., Hoffmann, S., 2018. Dietary behavior: an interdisciplinary conceptual analysis and taxonomy. Front. Psychol. 9, 1689.
Stuber, J.M., Hoenink, J.C., Beulens, J.W.J., Mackenbach, J.D., Lakerveld, J., 2021. Shifting toward a healthier dietary pattern through nudging and pricing strategies: a secondary analysis of a randomized virtual supermarket experiment. Am. J. Clin. Nutr. 114, 628–637. https://doi.org/10.1093/ajcn/nqab057.
Stunkard, A.J., Messick, S., 1985. The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. J. Psychosom. Res. 29, 71–83.
Sullivan, N., Hutcherson, C., Harris, A., Rangel, A., 2015. Dietary self-control is related to the speed with which attributes of healthfulness and tastiness are processed. Psychol. Sci. 26, 122–134.
Szaszi, B., Palinkas, A., Palfi, B., Szollosi, A., Aczel, B., 2017. A systematic scoping review of the choice architecture movement: toward understanding when and why nudges work. J. Behav. Decis. Making 31, 355–366.
Ung, C.Y., Menozzi, M., Hartmann, C., Siegrist, M., 2018. Innovations in consumer research: the virtual food buffet. Food Qual. Prefer. 63, 12–17.
van der Waal, N.E., Janssen, L., Antheunis, M., Culleton, E., van der Laan, L.N., 2021. The appeal of virtual chocolate: a systematic comparison of psychological and physiological food cue responses to virtual and real food. Food Qual. Prefer. 90, 104167.
Wang, G., Hayes, J., Ziegler, G., Roberts, R., Hopfer, H., 2018. Dose-response relationships for vanilla flavor and sucrose in skim milk: evidence of synergy. Beverages 4, 73.
Zinner, S.H., McGarvey, S.T., Lipsitt, L.P., Rosner, B., 2002. Neonatal blood pressure and salt taste responsiveness. Hypertension 40, 280–285.

Glucose: Metabolism and homeostasis

Julie Miller Jones, St. Catherine University, St. Paul, MN, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of V. Marks, Glucose: Metabolism and Maintenance of Blood Glucose Level, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 387–392, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00135-5.

Introduction 71
Glucose in the body 71
Body pool 71
Glucose: breakdown, storage and synthesis 72
Glucose breakdown to alcohol during fermentation reactions 72
Glycogen 73
Glycogen synthesis 73
Glycogen utilization 73
Glycogen and health 73
Gluconeogenesis 74
Hormones and glucose homeostasis 74
Insulin and blood glucose 75
Glucose in the urine 77
Conclusion 77
References 77
Further reading 77

Key points

- The glucose body pool and blood glucose homeostasis mechanisms are discussed.
- Metabolic reactions that involve the breakdown of glucose to produce energy through aerobic and anaerobic glycolysis and its related pathways are delineated.
- The metabolic and hormonal response to high and low blood sugar, including storage of extra energy as glycogen and fat, and the production of glucose when energy is needed.
- The synthesis and storage of glycogen by glycogenesis and its breakdown in glycogenolysis, and the synthesis of glucose from precursors such as amino acids and lactic acid in gluconeogenesis are explored.
- The various hormones that regulate and impact blood glucose are introduced and the interrelationships between them, various tissue sensors and levels of metabolites are discussed.

Glossary

Cephalic phase The start of the digestive process which begins gastric secretion even before food enters the stomach. It results from thoughts, aroma, sight, or taste of food

Glucagon A polypeptide hormone, secreted by pancreatic α-cells in the islets of Langerhans, that raises the concentration of glucose in the bloodstream by stimulating liver glycogenolysis. It also stimulates insulin secretion

Glucose A ubiquitous monosaccharide that provides 60–80% of dietary energy as the *simple sugar*, combined with fructose or *galactose* in the form of sucrose and lactose, respectively, or predominantly as its polymer, starch

Gluconeogenesis The process by which simple precursors such as lactic acid and glycerol can be built up to produce new glucose molecules in the liver

Glucose pool An expression of the amount of glucose in the body to which glucose molecules can be added from the food after eating and from the liver during fasting and removed into the tissues under the influence of insulin

GLUT4 The principal glucose transporter protein mediating glucose uptake and playing a pivotal role in regulating glucose homeostasis

Glycogenesis The enzymatic process by which glycogen is formed from glucose

Glycogenin The center of a glycogen molecule

Glycogenolysis The breakdown of polymeric storage form of glucose, glycogen, to produce glucose that is released into the bloodstream in times of glucose need

Homeostasis The concept explains how the body maintains vital functions such as temperature, blood sodium, and glucose concentration within narrow ranges despite large changes in the environment

Incretin One of several hormones released by glandular cells in the intestinal mucosa in response to the ingestion of food that stimulate insulin secretion only in the presence of a higher than fasting blood glucose concentration

Insulin A protein hormone secreted by the *B*-cells of the pancreatic islets of Langerhans necessary for glucose entry into insulin-sensitive cells. It is antagonistic to glucagon secretion

Insulin-like growth factor I (IGF-I) A polypeptide hormone structurally similar to insulin that can bind to insulin receptors and can stimulate glucose transport and with growth hormone promotes normal growth

Introduction

Glucose is the primary energy source of human cells; therefore, glucose must reach all tissues when they need it. Despite wide variation in glucose supply and demand, glucose in the bloodstream and body pool must not fluctuate wildly. Glucose homeostasis is controlled by a complex, carefully balanced integration of nervous impulses, counterregulatory hormones, and metabolic signals that activate and inhibit glucose movement, breakdown, storage and mobilization. Three pathways maintain blood glucose: (1) the breakdown of absorbed dietary carbohydrate through glycolysis; (2) the release of glycogen from stores in the liver or muscle through glycogenolysis; and (3) the endogenous synthesis of glucose from glucogenic amino acids and other carbohydrate sources through gluconeogenesis.

The brain modulates blood glucose by controlling various aspects of metabolism, such as food intake, energy expenditure, insulin secretion, hepatic glucose production and glucose/fatty acid metabolism in adipose tissue and skeletal muscle. Any defects in signal pathways between the brain and organs charged with supplying energy and maintaining glucose homeostasis may contribute to development chronic disease. This entry will highlight the metabolism of absorbed dietary carbohydrates by following the delivery of glucose to various tissue and breakdown for energy production by glycolysis. Energy storage of glucose as glycogen and generation of glucose via gluconeogenesis and glucose metabolism by the brain and other various tissues under aerobic and anaerobic conditions will be explored. The complicated interplay of control sensors in the brain and gut, and various hormones and rate-limiting enzymes and metabolites that control glucose movement, breakdown, synthesis and storage to maintain glucose homeostasis will be outlined.

Glucose in the body

Dietary carbohydrates are the primary source of glucose. Therefore, carbohydrate intakes are recommended to range between 45 and 65% of energy (E). While these percentages far exceed minimum carbohydrate requirement, which is based on energy needs of tissues that depend on ATP generated from glycolysis (Sünram-Lea and Owen, 2017), the brain prefers glucose as its fuel and needs at least 50–100 g d^{-1} of glucose. Thus, the Institute of Medicine of the US National Academy of Sciences (2005) recommends ~130 g d^{-1} of glucose (200–500 kcal d^{-1}). While the brain can use alternative fuels such as lactate and ketone bodies, their transfer through the blood brain barrier and ability to substitute for all glucose functions is limited (Bentsen et al., 2019). Thus, glucose remains the preferred fuel.

Blood glucose results from digestion of dietary carbohydrate. Three main pathways maintain blood glucose concentration: (1) the movement of glucose into tissues for energy production by glycolysis or energy storage as glycogen or as part of lipid synthesis; (2) the release glucose from glycogen stores by glycogenolysis; and (3) the endogenous synthesis of glucose from glucogenic amino acids and carbohydrate sources such as lactic acid through gluconeogenesis.

The glucose hungry brain is only 2% of body weight but consumes half of the body's glucose at rest. More is required during challenging mental tasks. The average adult needs ~110 g d^{-1} of glucose and ~75 g d^{-1} in young children. Glucose adequacy in the brain is so important than when supplies are low, the brain blocks glucose entry and utilization by other organs. For some brain regions, small glucose drops not only impair cognition, but also may damage tissue. During fasting or very low carbohydrate diets, the brain is forced to shift and metabolize β-hydroxybutyrate and other ketones in place of glucose (The benefits and drawbacks of diets promoting ketones as fuel for dieting, brain functioning, and exercise performance are subjects of much research and debate).

The testes, renal medulla, and erythrocytes are also totally dependent for correct functioning on a steady glucose supply to generate ATP from glycolysis. These tissues are incapable of utilizing ketones. The marrow uses glucose but relies on fat as its primary fuel.

Body pool

The total amount of glucose in blood and body fluids comprises "the body pool of glucose". In adults, this ranges from 8 to 28 g and corresponds to blood glucose concentrations of 3.5–5 mmol L^{-1}. The body pool stays within this narrow range despite enormous

fluctuations in glucose supply and demand. For example, ingestion of a large, carbohydrate-rich meal may raise blood glucose concentration as high as 10 mmol L^{-1}, but the glucose body pool changes little, due to carefully orchestrated interactions among multiple hormones, sensors, and metabolites. Further this produces only small perturbations in blood glucose in normal individuals because the rate of glucose removal from the pool rises to match glucose input. During an overnight fast in a normal adult, ~9 g of glucose enter and leave the body pool every hour. This marks the only way glucose leaves the pool in healthy individuals. However, in situations where blood glucose remains high, excess glucose is forced into the urine causing polyuria and measurable urinary glucose, which are classic symptoms of type 2 diabetes or impaired glucose tolerance. Glucose in the urine is usually followed up in the clinic with the oral glucose tolerance test (OGTT) described in **Box 1**.

Glucose: breakdown, storage and synthesis

Glucose, the product of carbohydrate digestion, is absorbed by the intestinal enterocyte and enters portal blood for delivery to the liver. The liver retains about 30% of the glucose for further metabolism. The remaining 70% travels to various tissues to provide energy. For most cell types, particularly muscle and adipose, insulin is required for glucose uptake and initiation of energy production by glycolysis. However, some tissues, such as brain, kidney and red blood cells do not require insulin.

Glycolysis is a multistep pathway key to energy production not only in body tissues, but also in organisms populating the microbiome. If energy is needed, insulin stimulates enzymes of glycolysis. The first step requires the addition of a high-energy phosphate group from ADP to glucose by a hexokinase. The result is glucose-6-phosphate (G-6-P). This is the "currency" needed to enter further metabolism or transfer to other tissues. G-6-P is then isomerized to fructose-6-phosphate, then phosphofructokinase adds a second high-energy phosphate group to make fructose-1,6-bisphosphate. This energy-rich, 6-carbon sugar is then split into two 3-carbon entities, glyceraldehyde-3-phosphate and dihydroxyacetone phosphate. These both can be converted to glyceraldehyde-3-phosphate (or may serve as the glycerol backbone for triglyceride synthesis if energy is not needed immediately). These 3-carbon carbohydrates undergo another phosphorylation to form two 1,3-bisphosphoglycerate. These compounds then transfer their energy-rich phosphate compounds to ADP to yield 2 ATPs and pyruvate. Thus, the glycolysis of a single glucose molecule utilizes two high-energy phosphates but generates four, creating a net energy gain.

The pyruvate produced continues to generate energy as it undergoes further metabolism. However, both the amount of energy produced, and reactions involved, depend on the level of oxygen available. Under aerobic conditions, each pyruvate enters the mitochondria and joins with coenzyme A to yield one acetyl CoA, one CO_2, and one reduced niacin (NADH). The high-energy electrons held in NADH can enter the electron transport chain. Acetyl CoA continues through the citric acid cycle (CAC) interacting with various carbohydrate intermediates crucial to continuation of the cycle. However, these intermediates may also be used in a variety of important metabolic reactions and used for gluconeogenesis. Under conditions of reduced or limited oxygen, as occurs during heavy exercise, pyruvate cannot follow the aerobic pathway through CAC and electron transport. Instead, pyruvate is converted to lactic acid by lactate dehydrogenase using NADH. Lactate is shuttled to the liver for conversion back to glucose in gluconeogenesis.

Mitochondria are necessary for aerobic glycolysis, but erythrocytes have none, so their energy needs are supplied through the pentose phosphate pathway (PPP). Glucose-6-phospate produced in the first step of glycolysis is shunted to PPP. In PPP's anaerobic phase, sugars such as fructose and pentoses (ribose-5-phosphate), vital for building nucleic acids, are produced. In its aerobic phase, it generates NADPH, a biologically active niacin, and provides both the energy and reducing power for synthesizing nucleotides, fatty acids, sterols, and non-essential amino acids. Further, NADPH can react with glutathione to protect cells from reactive oxygen species. Thus, PPP is pivotal for both redox homeostasis and glucose metabolism.

Glucose breakdown to alcohol during fermentation reactions

Ethyl alcohol is the main product of carbohydrate fermentation in food. Specifically, microbial fermentation of foods such as yeast-leavened doughs or mashes for wine and beer, produces CO_2 and ethyl alcohol.

However, only in extremely rare instances does this reaction occur in the human body. If it occurs, it follows a course of strong antibiotics that fosters conditions for fungal overgrowth, (usually candida), called "auto-brewery syndrome". Despite its rare

Box 1 The oral glucose tolerance test (OGTT)

Clinically the OGTT follows glucose uptake and delivery to the tissues, e.g., rise and fall of blood glucose, and is used to identify those with abnormal glucose tolerance and diabetes. After an 8 h fast, the subject drinks a test syrup containing 75 g of glucose. Blood glucose is measured during the postprandial (usually 2 h) period. In healthy individuals, blood glucose stays in the range of 60–140 mg dL^{-1} and returns to fasting levels within 2 h.

The same tool is used as in research to aid the understanding of glucose absorption and utilization from daily meal patterns. However, the disposition of glucose absorbed after a carbohydrate-rich meal by healthy subjects varies widely from individual to individual, depending not only on the glucose tolerance of the individual, but also many characteristics and composition of the "eating event", even previous ones. Thus, the actual rate and pattern of glucose utilization after a carbohydrate-rich meal can be hard to accurately predict.

occurrence, some writers justify their claims that ingestion of sugars and carbohydrates is addictive by erroneously citing that alcohol is produced.

Glycogen

Humans do store glucose, but capacity is capped. The storage molecule glycogen is structurally like its plant cousin amylopectin, but is more branched, compact, and smaller, with only 2000–60,000 glucose units. The main role of liver glycogen is to maintain blood glucose concentration and supply all tissues with glucose, whereas muscle glycogen supplies energy the needs of cardiac and skeletal muscles only.

Liver glycogen has more glucose units than that in muscle glycogen and occupies 5–6% of the liver's cell volume. In muscle, glycogen occupies only 1–2% of muscle cell volume, but total muscle glycogen is four times higher than that found in liver because of large total muscle mass. Skeletal muscle glycogen is distributed in defined pools, which vary depending on factors such as exercise intensity, degree of training or immobilization, and fiber phenotype (Ørtenblad and Nielsen, 2015).

Glycogen is also found in small amounts in kidney, adipose, and erythrocytes, and even smaller amounts in white blood cells and brain. In brain, glycogen plays a role in memory formation, learning, and other cognitive functions, and serves as an emergency cerebral energy source (Duran et al., 2019). Active glycogen metabolism appears allow some tolerance to hypoxia and help brain functioning.

Glycogen synthesis

In the fed state, when current body energy needs are satisfied, insulin upregulates multiple enzymes involved in hepatic glycogen synthesis including glycogen synthase and glycogen branching enzyme. Glycogen synthesis requires cooperative actions using the glycogenin core. The rate-controlling step glucokinase phosphorylation of G-6-P (Nozaki et al., 2020) is followed by phosphoglucomutase isomerization of G-6-P to glucose-1-phosphate. After a high energy uridine-diphosphate is added, glucose moieties conjugate to form a chain using glycogen synthase to become the glycogenin center that controls synthesis of the main glycogen chain by attaching UDP glucose onto itself to create a chain of 8–12 glucose moieties with α-1,4 links (Zeqiraj and Sicheri, 2015). With glycogenin still at the core, the resulting oligosaccharide continues to add glucose, using glycogen synthase and glycogen branching enzyme, which add α-1, 6 branch points every 12–13 glucose moieties. This results in compact, globular glycogen granules that accumulate in the liver. Skeletal muscle builds glycogen at highest rates after exercise, showing the importance of refueling after intense exercise (Adeva-Andany et al., 2016).

Not all the excess glucose can be stored as glycogen, so insulin also activates lipogenic enzymes to capture energy and metabolites generated from glycolysis, CAC, and electron transport to store it as fat. Specifically, pyruvate is transformed to acetyl CoA to become building blocks for synthesis of long-chain fatty acids. Glyceraldehyde formed during glycolysis becomes glycerol and serves as the triglyceride backbone. High energy metabolites available from CAC and electron transport, such as NADH or $FADH_2$, drive lipogenesis.

Glycogen utilization

Glycogen is utilized in the fasting (postabsorptive) state, which occurs after all ingested foods have been digested and absorbed or during fasts of 6–8 h or longer. During an overnight fast, glycogen drops to ~40 g (range 15 g–80 g) because the brain and erythron continue to utilize glucose, then blood glucose drops as no glucose is entering the bloodstream. This causes glucagon release, which upregulates enzymes of glycogenolysis—glycogen phosphorylase to split glucose 1-phosphate from the linear chains of glycogen and the glycogen-debranching enzyme too split glucose at the branch points (Kanugo et al., 2018).

Glucagon also activates enzymes involved in gluconeogenesis (see next section). New glucose release frequently equals the rate of glucose uptake into tissues, which stabilizes blood glucose concentration between 4.0 and 5.5 mmol L^{-1} (70–100 mg dL^{-1}). Hepatic glycogen stores may even be partially replenished. If heavy exercise causes liver glycogen to drop as much as 80%, glucose utilization and removal from blood outpace its production. During weeks of starvation or extremely low-carbohydrate (<50 g d^{-1}) diet, blood glucose concentration rarely falls below 3 mmol/L, and usually only when gluconeogenesis is impaired.

Muscle glycogen supplies its own energy needs, even when glucose is needed elsewhere. However, the breakdown of muscle glycogen during exercise inhibits glucose uptake from the bloodstream, thereby aiding glucose supply to other tissues. Muscles lack glucose-6-phosphatase, which means that glucose released from glycogen cannot be phosphorylated to the form needed to travel in the bloodstream. Instead, anaerobic glycolysis in the muscle produces lactate. This enters the bloodstream and is delivered to the liver, where it is reconverted to glucose by gluconeogenesis.

Glycogen and health

Low glycogen often occurs in early morning hours before breakfast, but timing depends on factors such as meal size and time since the last food ingestion, food characteristics and physical nature (food matrix and fluidity), macronutrient and fiber content. For those with impaired glucose tolerance, low blood glucose in the morning causes hypoglycemia and can have serious ramifications.

A variety of inborn errors of metabolism affecting enzymes involved in the synthesis or breakdown of glycogen can result in the inability to access glycogen or result in the accumulation of excess glycogen. There are at least 13 identified disorders that affect a wide range of organs and systems including neuronal loss (Duran and Guinovart, 2015), muscle weakness or obesity.

Gluconeogenesis

When blood glucose is low, gluconeogenic enzymes for *de novo* glucose synthesis are activated. "New" glucose enters the bloodstream to meet energy needs of the brain and other tissues and partially replenishes liver glycogen stores. Most gluconeogenesis occurs in the liver, but some occurs in the renal cortex, especially in diabetes and prolonged fasting. Regardless of location, gluconeogenesis requires a ready supply of glucose precursors. During exercise, lactate produced in the muscle from the anaerobic glycolysis is transferred to the liver and converted to glucose, which is returned to the muscle as part of the Cori cycle. Physical training can optimize the use of lactate and fatty acids as fuel alternatives.

During prolonged fasting or starvation, the release of glucogenic amino acids from proteolysis of skeletal muscle is needed for glucose synthesis. Alanine has a special role in gluconeogenesis, not only for providing three-carbon carbohydrate backbones to the liver as glucose precursors, but also for shuttling all amino acids released by muscle catabolism to the liver. The following are glucose building blocks: alanine, pyruvate, glycerol, and other glucogenic amino acids including glutamine/glutamate. The latter function in kidney gluconeogenesis in response to acidosis. The glycerol portion of triglycerides and the short-chain fatty acid propionate are the only fat precursors used in gluconeogenesis; no other fatty acids serve glucose precursors.

Gluconeogenesis is inhibited by eating, so insulin released in the fed state inhibits gluconeogenic enzymes, such as the rate-limiting phosphoenolpyruvate carboxykinase. Alcohol ingestion inhibits gluconeogenesis, so individuals consuming substantial energy as alcohol may develop severe hypoglycemia.

Hormones and glucose homeostasis

Glucose supply to the body tissues and in the bloodstream is regulated by numerous hormones, signal metabolites, sensors, and allosteric enzymes, often with counterbalancing effects. While the central nervous system aided by the pancreas and the gut form the command center, the liver has the important job of balancing glucose release, glycogen synthesis and degradation, and gluconeogenesis. If these hepatic processes are disrupted, they can contribute to hyperglycemia in the fasted and postprandial states. The body also has many neural and hormonal mechanisms at its disposal to correct or overcome any fall in blood glucose to below the critical level necessary for maintenance of normal body and brain functioning.

Insulin and glucagon have major roles in meeting energy needs, supplying a steady amount of glucose to the bloodstream, and promoting glucose movement from the bloodstream and body pool into peripheral tissues as needed. However, these hormones are intricately enmeshed to other key hormones, metabolites, and feedback control loops.

Insulin is a glucose-utilizing anabolic hormone that both works in tandem with, and is antagonistic to, glucagon. Insulin is among the smallest proteins and has two chains joined by disulfide bonds. This protein is released by pancreatic β-cells upon sensing a rise in blood glucose concentration, although food thoughts and smells can cause its release in the cephalic phase. Insulin's main function is to lower blood glucose and encourage glucose utilization. In that way it differs from other hormones, except somatostatin and IGH-1, in that others raise blood glucose.

Insulin removes glucose from the bloodstream or encourages glucose utilization in the following ways.

- It stimulates the expression of GLUT4, a principal glucose transporter protein in muscle and adipose. This enables glucose uptake from the bloodstream into these tissues (Exercise and hyperglycemia increase GLUT4 action).
- It promotes glucose breakdown by stimulating enzymes of glycolysis.
- It facilitates energy storage by stimulating glycogen synthase and fatty acid synthetase and lipogenesis.
- It inhibits gluconeogenesis by decreasing the expression of rate-limiting enzymes involved in glucose synthesis.

Glucagon, a 29- amino acid peptide hormone, is a glucose mobilizing catabolic hormone. It is released by α-cells (that comprise 40% of the pancreatic islet cells) when they sense low blood glucose. Glucagon acts by:

- Activating liver glycogen phosphorylase to start glycogenolysis.
- Activating enzyme machinery associated with gluconeogenesis to raise blood-glucose. Thus, it promotes muscle breakdown to produce glucogenic aminos acids that can be deaminated to serve as substrates for glucose synthesis. Further, it promotes utilization of pyruvate, lactate, and other carbohydrate intermediates.
- Blocking insulin action to increase or maintain plasma glucose and stimulate ketogenesis to provide alternate fuels if needed.
- Causing insulin release to enable utilization of glucose generated by its actions.

In abnormal glucose tolerance or diabetes, glucagon response to hypoglycemia is defective and results in chronic hyperglucagonemia.

Somatostatin, released by the anterior pituitary, controls growth. It decreases hormones associated with growth, and thereby inhibits somatotrophin release and suppresses glucagon's response to low blood glucose. Somatostatin also slows digestion and nutrient (glucose) absorption by inhibiting gastrin and the subsequent release of stomach acid.

Cortisol is a steroid hormone released by the adrenal cortex during physical and emotional stress. It provides glucose needed, in the short term, to address the stressful situation. Cortisol not only increases hepatic glucose output by stimulating gluconeogenic enzymes, but also acts to keep glucose in the bloodstream by reducing insulin secretion.

Constant stress causes continuous cortisol release and leads to both increased blood sugar levels, elevated inflammation, and insulin resistance. **ACTH** is released from adrenal glands and stimulates cortisol, which in turn, stimulates release of fatty acids from adipose tissue, so the glycerol portion can provide glycerol for gluconeogenesis.

Norepinephrine and epinephrine are released after stimulation by the sympathetic and parasympathetic nervous systems - adrenaline by the adrenal medulla and noradrenaline from nerve terminals in the liver. Strong emotions (e.g., fear or anger) trigger their release to increase heart rate, muscle strength, blood pressure, and glucose metabolism.

Both work in similar ways and increase blood glucose by:

- Inhibiting postprandial glucose uptake in insulin-dependent tissues.
- Stimulating liver glycogenolysis.
- Activating lipolysis in muscles to release glycerol, which can be used in gluconeogenesis, and free fatty acids to fuel muscle.

Growth hormone is released by the hypothalamus. It is antagonistic to insulin action by inhibiting liver glucose uptake and promoting lipolysis and gluconeogenesis. However, it works with insulin to promote use of amino acids to build tissue.

Thyroxine is released by the thyroid gland and stimulates carbohydrate metabolism by:

- Promoting intestinal glucose absorption.
- Enhancing insulin-dependent glucose entry into cells.
- Stimulating glucose generation through gluconeogenesis and glycogenolysis.

Many other hormones also control glucose by acting with insulin and glucagon. **Insulin-like growth factor-I (IGF-1)**, a polypeptide, structurally akin to insulin, is synthesized in liver and some peripheral tissues. Under the direction of growth hormone, it promotes normal growth and development of bone and tissue. Because nearly half the amino acid sequence is homologous with insulin, IGF-1 binds to insulin receptors stimulating glucose transport into fat and muscle. It acts in tandem with insulin to remove glucose from the bloodstream and promotes glucose homeostasis and glucose tolerance. In the brain IGF-1 works with insulin to modulate glucose metabolism, but faulty pathways are thought to contribute to neurological disorders (Fernandez et al., 2017; Zheng and Tong, 2017).

Food ingestion activates specialized endocrine cells distributed along the length of the intestinal epithelium. These include the activation of L-cells that triggers release of **glucagon-like peptide-1 (GLP-1)** and K-cells that triggers release **gastric inhibitory polypeptide (GIP)**. These two incretin hormones affect glucose metabolism by binding to their respective glucose-coupled receptor proteins (G-proteins) in fat, bone, brain, and pancreas. They bind to β-cells to enhance insulin secretion, especially under hyperglycemic conditions. They simultaneously suppress glucagon release. Both hormones promote β-cell proliferation and inhibit their apoptosis. In addition, they may impact memory formation, gastric emptying, and appetite control. They can sense and signal information related to the rate of nutrient absorption, the intestinal lumen contents, or integrity of the epithelial barrier and play a role in signaling glycogen breakdown and gluconeogenesis in the liver. However, some actions of the two hormones are antagonistic:

- GIP enhances postprandial glucagon response (and bone formation), while GLP-1 suppresses them.
- GIP, but not GLP-1, facilitates fat deposition in adipose.

Vasopressin (VP) is secreted by the posterior pituitary gland and facilitates glucose homeostasis in several ways. Through receptors in the pancreas and other site, it regulates:

- Hepatic glucose production,
- Insulin signaling,
- Aldosterone secretion.

High VP is a risk factor for glucose metabolism disorders in humans (Nakamura et al., 2017).

Adrenocorticotropic hormone (ACTH), growth hormone, and **prolactin** are released by the anterior pituitary gland. All four can cause glucose release by enhancing glycogenolysis in the liver.

Amylin, formed primarily in pancreatic islet β cells, is co-secreted with insulin in response to caloric intake. Amylin helps maintain glucose homeostasis by suppressing glucagon release in response to energy intake, slowing gastric emptying, and activating satiety signals from the brain to limit energy intake.

Table 1 gives a short synopsis about some hormones that affect blood glucose.

Insulin and blood glucose

Insulin release from pancreatic *B*-cells depends on the glucose concentration in the blood perfusing the islets. In the fed state, insulin concentration in peripheral blood ranges from ~150–600 pmol L^{-1}, which markedly increases uptake of absorbed glucose from arterial blood into peripheral tissues, helping blood glucose return to basal concentration. As glucose is removed, plasma insulin falls. Fasting adults with normal glucose tolerance have insulin concentration in peripheral blood of ~30 pmol L^{-1}, making glucose uptake by striatal muscle nearly zero and fatty acid release from lipocytes a major energy source.

Table 1 Hormones that affect blood glucose.

Hormone	*Stimuli*	*Inhibitors*	*Main effect on glucose homeostasis*
Insulin	Cephalic triggers Hyperglycemia Incretins- GIP & GLP-1 Glucagon Arginine and leucine Long chain fatty acids Vagus nerve/parasympathetic G-protein-coupled receptors (GPCRs)	Hypoglycemia Sympathetic Adrenaline Somatostatin GPCRs Short chain fatty acids	↓ Blood glucose by: ↓ Glycogenolysis ↓ Gluconeogenesis ↑ Peripheral glucose uptake and utilization
Glucagon	Hypoglycemia Adrenaline Most amino acids Fatty acids Cold	Insulin Hyperglycemia Ketones	↑ Liver glycogenolysis ↑ Gluconeogenesis
Adrenaline	Hypoglycemia, by sympathetic nerve stimulation Physical/mental stress		↑ Glycogenolysis in liver and peripheral tissues, ↓ Insulin secretion ↑ Glucagon secretion ↓ Adipose and muscle sensitivity to glucose ↓ Peripheral glucose utilization ↑ Lipolysis (i.e., ↑ plasma non-essential fatty acids levels)
Cortisol	Hypoglycemia through hypothalamic release of ACTH		↓ Peripheral glucose uptake ↑ Insulin resistance Permits hepatic glycogenesis
Growth hormone	Hypoglycemia through hypothalamus Ghrelin from stomach following food ingestion	Hyperglycemia Somatostatin Alcohol	↓ Peripheral glucose uptake ↑ Adipocytes ↓ Lipolysis by inhibiting lipoprotein lipase
Vasopressin	Hypoglycemic stress Dehydration	Hypo-osmolality alcohol	↑ Hepatic glycogenolysis
Gut incretin hormones			
Glucose-dependent insulinotropic peptide, also Gastric inhibitory polypeptide (GIP)	Actively absorbed sugars e.g., glucose and galactose, actively absorbed fats especially polyunsaturated	Glucagon Insulin	↑ Insulin secretion under hyperglycemia
Glucagon-like peptide- 1 (GLP-1)	Ingested food (absorbed and unabsorbed)	Glucagon Insulin	↑ Insulin secretion during hyperglycemia ↓ Glucagon secretion ↓ Gastric emptying ↓ Appetite

When blood glucose reaches ~3.5–4.0 mmol L^{-1}, insulin secretion continues, but at a reduced rate. In individuals without diabetes, insulin concentration is low enough to permit release of glucose by the liver and fatty acids by adipocytes. When *B*-cells fail and insulin secretion ceases, gross hyperglycemia and ketosis occur. Both are hallmarks of insulin-dependent diabetes.

During prolonged starvation (20 days or more without food), small amounts of insulin reach the liver. The amounts reaching adipocytes are, however, insufficient to prevent rapid lipolysis leading to hyperketonemia comparable with that seen in diabetic ketoacidosis (~10–20 mmol L^{-1}). However, small amounts of insulin secreted during starvation affect gluconeogenesis and glycogenolysis to help blood glucose remain within normal range during the prolonged fast. This is in sharp contrast to elevated blood glucose observed with *B*-cell malfunction and impaired glucose tolerance.

Insulin's ability to lower blood glucose depends not only on serum insulin concentration, but also whether the insulin is **exogenous** or **endogenous** (pancreatic). Proteases in pancreatic *B*-cells cleave the C-peptide from pro-insulin activating it prior to its release into the bloodstream, however C-peptide accompanies insulin to peripheral tissues. This creates a synergistic duo that causes a key difference between the functioning of endogenous and exogenous insulin. The endogenous insulin C-peptide complex reaches the liver at higher concentration than in peripheral tissues, but without the attendant C-peptide a higher concentration of exogenous insulin is found peripheral tissues. Both the lack of C-peptide and higher exogenous insulin in the periphery are thought to contribute to vascular complications of diabetes.

Glucose in the urine

Each day, more than 100 g of glucose are filtered from the bloodstream through the glomeruli of the kidneys. Under normal conditions, more than 99% of the glucose is reabsorbed by the kidney tubules. As a result, healthy people lose less than 150 mg of glucose in their urine each day, an amount too small to be detected in most screening procedures. When the glucose pool increases above a threshold level (~10 mmol of glucose/L), the amount of glucose filtered by the glomeruli exceeds the reabsorption capacity of the tubules, causing glucose to appear in the urine. Glucosuria is defined as urinary glucose concentration >0.15 g L^{-1} (0.8 mmol L^{-1}), but concentrations can be many times greater than in the blood. This causes increased excretion of water, sodium, chloride, and potassium resulting in osmotic diuresis associated with hallmarks of hyperglycemia and diabetes.

While temporary increases in glucose pool size (hyperglycemia) are not immediately harmful, decreases in glucose pool size (hypoglycemia) are. In pregnancy and other conditions that may impair glucose transport through the renal tubules, normal blood glucose concentration may result in glycosuria. However, under conditions with normal tubular glucose reabsorption when blood flow through the glomeruli is reduced or glomerular filtration is impaired, gross hyperglycemia may not produce glycosuria. Many mechanisms have evolved to prevent or overcome such potentially dangerous conditions.

Conclusion

Supplying energy in the form of glucose is critical to the functioning of tissues in the body, and therefore is under the control of a complex array of hormones, incretins, sensors, and organs. Many have countervailing effects. Some, such as insulin, initiate glucose breakdown by glycolysis and utilization of its metabolites to create ATP and other high-energy compounds; others, such as glucagon, inhibit these reactions. Insulin also orchestrates the storage of energy not immediately needed for future use by upregulating enzymes responsible for glycogenesis and lipogenesis. Countervailing hormones, such as glucagon, promote glycogenolysis and gluconeogenesis. While insulin and glucagon are key, the array of feedback loops and reactions and hormones that impact glucose concentration in the blood and body pool are many and complicated. The number of hormones, allosteric enzymes and their counteracting reactions, and alternate glucose pathways such as the Cori cycle, illustrate the complicated and interrelated mechanisms for meeting the body's energy needs while maintaining the glucose body pool and blood glucose. Homeostasis in these many reactions must occur despite wide ranges of carbohydrate intake and rates glucose of utilization, because it is key to good health.

References

Adeva-Andany, M.M., González-Lucán, M., Donapetry-García, C., Fernández-Fernández, C., Ameneiros-Rodríguez, E., 2016. Glycogen metabolism in humans. BBA Clin. 5, 85–100.

Bentsen, M.A., Mirzadeh, Z., Schwartz, M.W., 2019. Revisiting how the brain senses glucose-and why. Cell Metab. 29, 11–17.

Duran, J., Gruart, A., López-Ramos, J.C., Delgado-García, J.M., Guinovart, J.J., 2019. Glycogen in astrocytes and neurons: physiological and pathological aspects. Adv Neurobiol. 23, 311–329. https://doi.org/10.1007/978-3-030-27480-1_10. PMID: 31667813; PMCID: PMC7315007.

Duran, J., Guinovart, J.J., 2015. Brain glycogen in health and disease. Mol. Aspect. Med. 46, 70–77.

Fernandez, A.M., Hernandez-Garzon, E., Perez-Domper, P., Perez-Alvarez, A., Mederos, S., Matsiu, T., et al., 2017. Insulin regulates astrocytic glucose handling through cooperation with IGF-I. Diabetes 66 (1), 64–74. https://doi.org/10.2337/db16-0861.

Kanungo, S., Wells, K., Tribett, T., El-Gharbawy, A., 2018. Glycogen metabolism and glycogen storage disorders. Ann. Transl. Med. 6, 474.

Nakamura, K., Velho, G., Bouby, N., 2017. Vasopressin and metabolic disorders: translation from experimental models to clinical use. J. Intern. Med. 282, 298–309.

Nozaki, Y., Petersen, M.C., Zhang, D., Vatner, D.F., Perry, R.J., et al., 2020. Metabolic control analysis of hepatic glycogen synthesis in vivo. Proc. Natl. Acad. Sci. U. S. A. 117, 8166–8176.

Ørtenblad, N., Nielsen, J., 2015. Muscle glycogen and cell function—location, location, location. Scand. J. Med. Sci. Sports 25 (Suppl. 4), 34–40.

Sünram-Lea, S.I., Owen, L., 2017. The impact of diet-based glycaemic response and glucose regulation on cognition: evidence across the lifespan. Proc. Nutr. Soc. 76, 4466–4477.

Zeqiraj, E., Sicheri, F., 2015. Getting a handle on glycogen synthase—its interaction with glycogenin. Mol. Aspect. Med. 46, 63–69.

Zheng, P., Tong, W., 2017. IGF-1: an endogenous link between traumatic brain injury and Alzheimer disease? J. Neurosurg. Sci. 61, 416–421.

Further reading

Dimitriadis, G.D., Maratou, E., Kountouri, A., Board, M., Lambadiari, V., 2021. Regulation of postabsorptive and postprandial glucose metabolism by insulin-dependent and insulin-independent mechanisms: an integrative approach. Nutrients 13, 159.

Gribble, F.M., Reimann, F., 2019. Function and mechanisms of enteroendocrine cells and gut hormones in metabolism. Nat. Rev. Endocrinol. 15, 226–237.

Güemes, M., Rahman, S.A, Hussain, K., 2016. What is a normal blood glucose? Arch. Dis. Child. 101, 569–574.

Hædersdal, S., Lund, A., Knop, F.K., Vilsbøll, T., 2018. The role of glucagon in the pathophysiology and treatment of type 2 diabetes. Mayo Clin. Proc. 93, 217–239.

Haeusler, R.A., McGraw, T.E., Accili, D., 2018. Biochemical and cellular properties of insulin receptor signalling. Nat. Rev. Mol. Cell Biol. 19, 31–44.

Holst, J.J., Gribble, F., Horowitz, M., Rayner, C.K., 2016. Roles of the gut in glucose homeostasis. Diabetes Care 39, 884–892.

Klip, A., McGraw, T.E., James, D.E., 2019. Thirty sweet years of GLUT4. J. Biol. Chem. 294, 11369–11381.

Murray, B., Rosenbloom, C., 2018. Fundamentals of glycogen metabolism for coaches and athletes. Nutr. Rev. 76, 243–259.

Röder, P.V., Wu, B., Liu, Y., Han, W., 2016. Pancreatic regulation of glucose homeostasis. Exp. Mol. Med. 48, e219.
Roh, E., Song, D.K., Kim, M.S., 2016. Emerging role of the brain in the homeostatic regulation of energy and glucose metabolism. Exp. Mol. Med. 48, e216.
Sharma, V.K., Singh, T.G., 2020. Chronic stress and diabetes mellitus: interwoven pathologies. Curr. Diabetes Rev. 16, 546–556.
Tokarz, V.L., MacDonald, P.E., Klip, A., 2018. The cell biology of systemic insulin function. J. Cell Biol. 217, 2273–2289.
Wu, L., Butler, N.J.M., Swanson, R.A., 2019. Technical and comparative aspects of brain glycogen metabolism. Adv. Neurobiol. 23, 169–185.

Nutritional aspects of bone

A.R. Bournot[a], M. Alghamdi[b], J. Jalal[b], and S.A. Lanham-New[a], [a] University of Surrey, Guildford, UK; and [b] Joint Supervision Programme, King Abdulaziz University, Jeddah, Saudi Arabia

© 2023 Elsevier Ltd. All rights reserved.

This is an update of S.A. Lanham-New, M. Alghamdi, J. Jalal, Nutritional Aspects of Bone, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 220–226, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00029-5.

Introduction 79
Calcium 80
Peak bone mass attainment 81
Postmenopausal bone loss 81
Vitamin D 81
Vitamin D and calcium studies on bone 82
Vitamin D and risk of falling 82
Protein intake and bone 82
Vitamin K 82
Dietary alkali/potassium consumption 82
Importance of acid–base homeostasis to health 82
Skeletal link to acid–base maintenance 83
Observational and intervention studies 83
Isoflavones and bone health 83
Folate, vitamin B_{12}, and bone health link 83
Nutrients adversely affecting bone health 84
Vitamin A 84
Sodium 84
Impact of physical activity on bone health 84
Introduction 84
Importance of exercise to bone 84
Conclusions 85
Concluding remarks 85
References 86

Key points

- The pathogenesis of poor bone health across the lifecycle is multifactorial.
- Musculo-skeletal health is determined by the development of peak bone mass in the younger population and the rate of bone loss in the older population.
- Musculo-skeletal health is affected by a combination of genetic, endocrine, mechanical, and nutritional factors, with evidence of extensive interactions within and between these groups.
- Low calcium intake (<400 mg per day) has been associated with inadequate calcium retention and poor peak bone mass attainment. It is also associated with increased bone loss in the ageing Caucasian population.
- Vitamin D deficiency (<25 nmol/l) across the lifecycle has been shown to be detrimental to musculo-skeletal health in the long-term.
- Acid-base homeostasis may play a key role in reducing bone loss in the ageing population. Studies demonstrate that a high intake of potassium salts may reduce bone resorption but more longer-term studies are required.
- More research is needed on the importance of other micronutrients (including the trace elements, vitamin K) to bone health in the long term.
- Physical activity (impact loading) across the life-course is key to bone health.

Introduction

The pathogenesis of poor bone health is multifactorial. Both the development of peak bone mass in the younger population and the rate of bone loss in postmenopausal women and the elderly men and women are determined by a combination of genetic, endocrine, mechanical, and nutritional factors, with evidence of extensive interactions within and between these groups (Fig. 1).

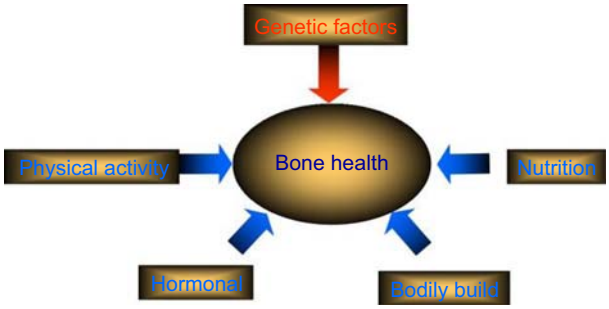


Fig. 1 Modifiable (endogenous) versus nonmodifiable (exogenous) factors affecting bone health.

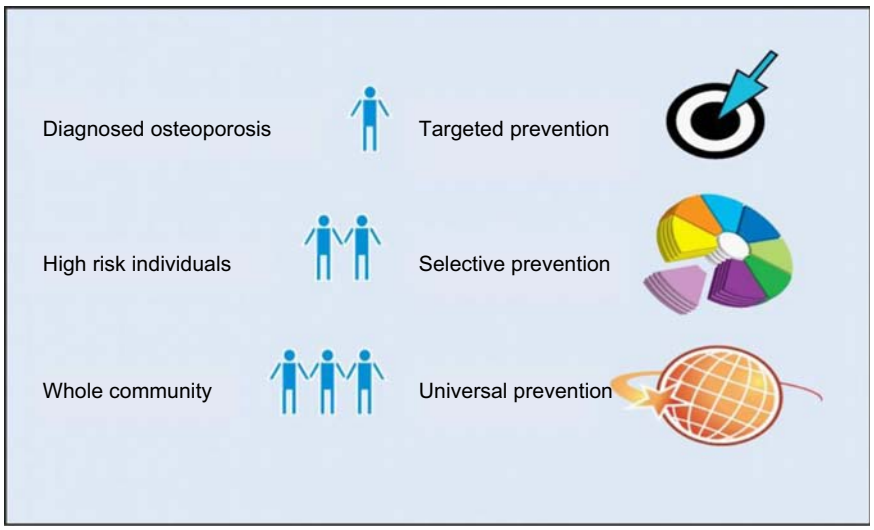


Fig. 2 Nutritional approaches to osteoporosis prevention.

Endogenous factors have a critical influence on the skeleton. Monozygotic and dizygotic twin research and mother/daughter pairs show a genetic influence on bone health in the region of 75%. In addition, the skeletal determinants of osteoporotic fracture risk, such as areal bone mineral density (BMD), bone geometry, and bone turnover, are all subject to strong genetic influences (Boudin et al., 2016).

Nutritional advice, which is based on sound scientific evidence, is of paramount importance to ensure and enable the optimization of bone health throughout the life cycle. Even a small or modest effect on bone health is likely to have a significant impact on fracture prevention – for example, an increase in BMD by one standard deviation unit is likely to result in a 50% reduction in fracture rates. For effective strategies, emphasis needs to be given, in combination, at three different levels as shown in Fig. 2: (1) Universal primary prevention, (2) selective prevention for high-risk groups and, (3) targeted prevention for individuals.

Calcium

It is well considered that calcium is critical to health. Calcium is the most abundant mineral element in the body and has two key roles. The first role is structural and the second is regulatory. For its structural role, bone consists of protein matrix encased in a crystalline mineral. Approximately 1 kg of calcium is contained within the skeleton (99% of Ca is contained in bones and teeth) and it is this mineral part that contributes to the strength of bone. Bone mineral provides a huge reserve of Ca, behaving as a large "ion exchanger" allowing interaction between ions in body fluids and bone. Ca is the most abundant mineral element in the body. For its regulatory role, plasma concentrations of calcium are maintained within very narrow limits (90–110 mg L^{-1}) and although only 1% of calcium is found in soft tissues and body fluids, this calcium is required for a number of key functions including cellular structure; inter- and intracellular metabolic function and signal transmission; muscle contractions including heart muscle, nerve function, activities of enzymes, and normal clotting of blood. The regulation of serum calcium levels is carefully maintained by calciotrophic hormones (parathyroid hormone, calcitriol (hormonally active metabolite of vitamin D, 1,25 $(OH)_2D$) and calcitonin).

Peak bone mass attainment

Peak bone mass (PBM), the highest level of bone mass attained during normal growth, is one of the key factors determining bone mass and fracture risk later in life. The age at which peak mass is attained ranges from approximately 17–35 years and varies for different skeletal sites.

Calcium absorption and bone calcium deposition rates peak (approximately five times that of adulthood) in girls shortly before menarche. Thus, low calcium during growth and late menarcheal age may affect peak bone mass and consequently fracture risk later in life. There are data to show that adolescent girls are less likely than boys to meet the current recommended dietary levels for calcium and that calcium intake in girls may begin to decline around the time of puberty. The demand to provide calcium to the fetus and neonate during pregnancy and lactation is considerable, and the results of a few recent small studies on the effect of adolescent pregnancy on bone turnover and later risk for osteoporosis as well as fetal bone development highlight the need for adequate calcium intake in pregnant teens (Lanham-New et al., 2007).

To date, clinical trials investigating the effect of increased calcium intake (either through foods or supplements) on peak bone mass development have been of relatively short duration, and it has been difficult to determine whether the positive effects of calcium supplementation on bone are maintained. In a very important calcium and bone study, Matkovic et al. (2005) looked at the effects of calcium intake over an extended period. They reported the results of a 4-year randomized clinical trial that involved 354 girls at stage 2 of puberty. The study was optionally extended for a further 3 years. The mean intake of calcium over the 7-year period was 830 mg day^{-1}, with calcium-supplemented participants receiving an additional 670 mg day^{-1} of calcium. The results indicated that calcium supplementation significantly influenced bone accretion in girls during the pubertal growth spurt. The effect diminished in young adulthood, but in tall girls, the significant effects remained at the metacarpals and at the forearm. Thus, calcium requirements for maximum skeletal development may be associated with bone size. While calcium intake from supplements, fortified foods, and dairy can increase skeletal mass and density measures, the positive effects may differ depending on the form that it is consumed in. A 2-year randomized controlled trial in Finland including girls aged 10–12 years who had low dietary calcium intake at inclusion found that increasing dietary calcium intake by consuming cheese was more beneficial for cortical bone mineral mass accrual than consuming a similar amount of calcium in tablet form (1000 mg day^{-1}) (Cheng et al., 2005). This may be due to differences in the absorption of calcium, the distribution of intake during the day, and the absorption of other nutrients from dairy products compared to tablets. Therefore, while calcium supplementation may be useful for those with inadequate intakes, further studies are needed to establish potential differences in calcium intake from dairy and supplements on PBM.

Postmenopausal bone loss

There are good data to suggest that calcium supplements are effective in reducing bone loss in late menopausal women (45 years postmenopause), particularly in those with low habitual calcium intake (400 mg day^{-1}) (Lanham-New, 2008). In addition, a meta-analysis that includes 15 trials indicates that calcium supplementation at levels between 500 and 2000 mg day^{-1} reduces postmenopausal bone loss (Shea et al., 2002). However, a secondary analysis of a randomized controlled trial involving 1471 healthy postmenopausal women (mean age 74 years) reported that calcium supplementation of 1000 mg day^{-1} was associated with an increased rate of cardiovascular events (Bolland et al., 2008). Therefore, further studies are needed to investigate the benefits of calcium supplements against their potentially deleterious effects in postmenopausal women. The findings of calcium supplementation studies in the early stages of the menopause are conflicting, and this is an area for further investigation (Lanham-New, 2008).

Vitamin D

Vitamin D is derived from both endogenous (skin) and exogenous (diet) sources. The main source of vitamin D for most people is through exposure to sunlight on the skin containing sufficient ultraviolet B (UVB) radiation. However, the relative contributions of these two sources are thought to vary widely among individuals and between different geographical areas. In northern latitudes, there is no UVB radiation of the appropriate wavelength (290–315 nm) for vitamin D synthesis from the end of September to the end of March; and for the remaining months of the year, 60% of the effective UVB radiation occurs between 11:00 a.m. and 3:00 p.m. The extent of UVB exposure not only depends on the latitude and time, but also cloud cover, ozone, and atmospheric pollution. There are relatively few dietary sources of vitamin D—the major providers being fish, eggs, meat, liver, and some fortified foods, such as fat spreads and cereals (Charoenngam et al., 2019).

Although it is well documented that vitamin D synthesis from sunlight is affected by the aging process, there is a re markable lack of awareness of this public health nutrition message. Several studies indicate low nutritional status of vitamin D in at-risk groups, manifesting in the re-emergence of rickets in children in some communities. Lack of sufficient sunlight exposure, especially but not only in Northern countries during winter months, poor nutrition, or low milk consumption or vegetarian diet, urban residence, and poverty are key factors that have been cited for the low vitamin D levels. Ethnic groups with darker skin may also be at a higher risk of vitamin D deficiency due to skin pigment melanin which reduces the penetration of UVB in the skin (Buttriss et al., 2022).

Vitamin D and calcium studies on bone

Vitamin D is involved in calcium and phosphate homeostasis, playing an important role in bone metabolism. Prolonged and severe vitamin D deficiency can lead to rickets in children and osteomalacia in adults, although these diseases are rare in most developed countries. Vitamin D deficiency may also be associated with osteoporosis and higher incidence of falls or fractures (Hill and Aspray, 2017). However, meta-analyses of clinical trials investigating the effect of vitamin D supplementation on BMD, fractures, and falls have yielded conflicting results (Bouillon et al., 2022). A meta-analysis of 23 trials by Reid et al. (2014) found a small benefit in femoral neck BMD, but no effect at other sites. Several studies have also reported that fracture risk does not appear to be influenced by vitamin D alone. The 2022 systematic umbrella review of meta-analyses by Chakhtoura et al. (2022) investigating vitamin D supplementation (with or without calcium) found no effect on fracture risk reduction in studies exclusively evaluating community-dwelling adults and younger adults (aged 50–65 years), or in studies on vitamin D alone compared to control. However, combined vitamin D and calcium supplementation overall reduced the risk of hip and any fractures, potentially driven by findings from older and institutionalized adults. Additionally, a meta-analysis of 15 trials by Liu et al. (2020) found that combined calcium and vitamin D significantly reduced the risk of hip fracture, and significantly increased BMD at several sites in postmenopausal women (50–79 years). Interestingly, this study found femoral neck BMD was significantly increased only when vitamin D intake was no more than 400 IU d^{-1}.

Vitamin D and risk of falling

Although there have been mixed results, low vitamin D status has been implicated in an increased risk of falling. A meta-analysis by Bischoff-Ferrari et al. (2009) showed that vitamin D supplementation of 700–1000 IU per day resulted in a reduction of falls in the elderly by around 20%, particularly in those with low vitamin D levels. Mechanisms of action need further elucidation, but certainly muscle weakness, which can affect balance and mobility, has been implicated (Bouillon et al., 2022). Vitamin D (and calcium) supplementation may also be helpful in reducing falls by improving body sway and by normalizing blood pressure.

Protein intake and bone

There is controversy concerning the relationship between dietary protein and bone metabolism. Although dietary protein has been shown to result in urinary calcium loss, negative calcium balance and increased bone loss, cross-sectional and longitudinal epidemiological studies examining the effect of protein intake on BMD, bone loss, and risk of fracture show mixed results: with some studies showing protein intake to be detrimental to bone health, whereas others demonstrating a beneficial effect. Conversely, protein-energy undernutrition is considered to be a risk factor for bone loss and osteoporosis. Low protein intake is related to low bone mass and increased risk of fracture, and protein supplementation has been shown to improve recovery from hip fracture (Darling et al., 2021).

Ecological studies have shown that worldwide *per capita* consumption of animal protein has been associated with a higher risk of hip fracture in women aged over 50 years. More recently, the correlation has been shown to be stronger with the ratio of animal protein to vegetable protein, a study which has adjusted for important cultural differences (Darling et al., 2009). It is important to note, however, that in these correlational studies, the unit of measurement is country and not individual and as such, these types of studies have a number of limitations, which must be considered in the interpretation of such data. In the most recent meta-analysis, total protein was found to have an overall beneficial effect on bone health throughout the lifecycle (Darling et al., 2019).

Vitamin K

Vitamin K has an important function for the skeleton; it acts as a cofactor in the posttranslational carboxylation of several bone proteins, with osteocalcin being the most abundant. Deficiency of vitamin K results in the synthesis of under-carboxylated osteocalcin (ucOC). Vitamin K_1, which is also known as phylloquinone, is a component of the Photosystem I of plants and is present in foods of plant origin. Alfalfa and green leafy vegetables are good sources of Vitamin K_1. Vitamin K_1 is also a component of some dietary supplements. Vitamin K_2 is a bacterial form of the vitamin and is also known as menaquinone.

There are observational data to show that low serum concentrations of both vitamin K_1 and ucOC are associated with an increased risk for osteoporotic fractures (Lanham-New, 2008). Studies examining the association between vitamin K and BMD have been inconsistent, and the most recent meta-analysis of vitamin K on measures of musculo-skeletal health do not support a protective effect (Mott et al., 2019).

Dietary alkali/potassium consumption

Importance of acid–base homeostasis to health

Acid–base homeostasis is absolutely critical to health. It is well documented that extracellular fluid pH remains between 7.35 and 7.45 and thus it is a major requirement of our metabolic systems to ensure that hydrogen ion concentrations are maintained

between 0.035 and 0.045 mEq. On a daily basis, humans eat substances that both generate and consume protons and as a net result, adult humans on a normal Western diet generate B1 mEq per kg body weight of acid per day. Of course, the more acid precursors a diet contains, the greater the degree of systemic acidity. As humans become older, their overall renal function declines, which includes their ability to excrete acid. Hence, with increasing age, humans become slightly but significantly more acidic.

Skeletal link to acid–base maintenance

The theoretical considerations of the role of alkaline bone mineral may play in the defense against acidosis date back as far as the late 1880s/early nineteenth century. The fundamental concepts were established in the late 1960s/early 1970s—a number of studies published during this period provided evidence that in natural (e.g., starvation), pathological (e.g., diabetic acidosis), and experimental (e.g., ammonium chloride ingestion) states of acid loading and acidosis, an association exists with both hypercalciuria and negative calcium balance.

There are clear mechanisms for a deleterious effect of acid on bone. Novel work in the 1980s by Arnett and Dempster demonstrated a direct enhancement of osteoclastic activity following a reduction in extracellular pH. This effect was shown to be independent of the influence of parathyroid hormone. Furthermore, osteoclasts and osteoblasts appear to respond independently to small changes in pH in the culture media in which they are growing (New, 2002).

Observational and intervention studies

A variety of population-based studies published in the latter part of the twentieth century and more recently between 2001 and 2011 have demonstrated a beneficial effect of fruit and vegetable/potassium intake on indices of bone health in young boys and girls, premenopausal women, perimenopausal women, postmenopausal women, and elderly men and women. Further support for a positive link between fruit and vegetable intake and bone health can be found in the results of the DASH (Dietary Approaches to Stopping Hypertension) and DASH-Sodium intervention trials. In DASH, diets rich in fruit and vegetables were associated with a significant fall in blood pressure compared with baseline measurements. However, of particular interest to the bone field were findings that increasing fruit and vegetable intake from 3.6 to 9.5 daily servings decreased the urinary calcium excretion from 157 mm day^{-1} to 110 mg day^{-1}. This study is the first population-based fruit and vegetable intervention trial showing a positive effect on calcium economy (albeit a secondary finding) (New, 2002). Research is now required to determine the long-term clinical impact of the DASH diet on bone health and fracture risk as well as clarification of the exact mechanisms involved with respect to this diet on skeletal protection. In the most recent meta-analysis, potassium supplementation was found to have a beneficial effect in reducing bone resorption but more longer-term studies are required (Lambert et al., 2015).

Isoflavones and bone health

Soy protein consumption may also help to explain why it is so difficult to find a clear-cut answer to whether there are bone health differences between populations who follow a vegetarian-based diet and those following a nonvegetarian diet. Soy isoflavones have a chemical structure similar to that of estradiol and have been shown to possess a certain degree of weak estrogenic activity. In the animal model, comparable favorable bone effects have been shown between 17*X* -estradiol and soy protein isolate, genistein or Daidzein. Although there are studies that support a beneficial effect of soy protein isolates on bone mass in both pre- and perimenopausal women, more data are urgently required. In the most recent trials, no effect of soy isoflavones on indices of bone health have been found. Furthermore, the available epidemiological studies looking at the association between soy product consumption and hip fracture rates are conflicting (Cashman, 2007).

Folate, vitamin B_{12}, and bone health link

The homocysteine–fracture risk link suggests the potential for a positive effect of vitamin B complex on the skeleton. There is increasing evidence at the experimental, clinical, and epidemiological level that raised homocysteine levels are associated with increased fracture risk. These levels can be reduced simply and cheaply through folic acid supplementation. In the absence of specific folic acid/fracture reduction trials, we cannot yet say whether this is an effective strategy for osteo porosis prevention but it is certainly an area for urgent research. There is also some evidence to show that vitamin B_{12} deficiency *per se* adversely affects bone health through the criticality of vitamin B to the collagen crosslinks but further randomized controlled trials are urgently required (Swart et al., 2013).

Nutrients adversely affecting bone health

Vitamin A

Vitamin A refers to a family of essential, fat-soluble compounds called retinoids. Retinol is the principal dietary form of vitamin A. The main natural sources of vitamin A are provided by animal foods such as liver, meat, milk products, eggs, and fatty fish. In addition, some foods are fortified with vitamin A. Cod liver oil (which is commonly used in a number of population groups, particularly postmenopausal women/elderly) also contains high levels of vitamin A. There are approximately 600 or so carotenoids, with approximately 10% of these having provitamin A activity.

Vitamin A is necessary for normal bone growth. However, intakes of vitamin A >1500 μg of retinol equivalents (Res) have been associated with lower BMD and higher fracture risk in populations in the US and Sweden (Nieves, 2005). In both these countries dairy products and cereals are generally fortified with vitamin A. It should also be noted that high doses of pure cod liver oil can provide as much as 1200 mg RE of vitamin A in a 10 mL dose.

Sodium

There are insufficient data to make the claim that salt is a significant risk factor for osteoporosis. However, high salt intakes have been associated with increases in urinary calcium loss. For example, it has been estimated that a 100 mmol increment in daily sodium intake is associated with an average loss of urinary calcium of approximately 1 mmol in free-living normocalciuric healthy populations. This loss has not been specifically correlated to bone loss, however. Some studies have shown an effect (i.e., higher sodium intake and higher bone turnover), but other studies have found no difference (Lanham-New et al., 2007).

Impact of physical activity on bone health

Introduction

More than a century ago, a German scientist call Julius Woolf stated the theory, which is now called "Woolf's Law": "Bone accommodates the forces applied to it by altering its amount and distribution of mass". More recently, this concept has been refined to a general theory of bone mass regulation, known as the mechanostat model. It is well known that in the absence of weight-bearing exercise bone loss will occur at both axial and appendicular skeletal sites. Although the exact mechanism whereby mechanical loading affects bone remains to be clarified, the scientific literature supports a positive relationship between physical activity, physical fitness, muscle strength, and bone mass at the lumbar spine and femoral neck sites (Wilson-Barnes et al., 2022).

Importance of exercise to bone

There is evidence that supports a positive relationship between weight-bearing exercise and bone mass in all age groups, including children, adults, and the elderly. Data also support a specific role for high-impact exercise to increase bone density in premenopausal women (Wilson-Barnes et al., 2022). Very little information is available on the relationship between exercise and fracture and is an area for further research. Although clearly exercise is of benefit to the skeleton, what remains undefined is exactly the type, intensity, and duration of weight-bearing physical activity required for optimum bone health. Furthermore, exercise may be of benefit in the prevention of osteoporosis, not necessarily via the mechanism of increasing bone mass but instead by increasing muscle strength, coordination, flexibility, and balance and thus reducing the tendency to fall.

Although bone mass has been shown to be higher in athletes involved in different sports, including tennis players, skaters, rowers, and volleyball players, there is increasing concern for the bone health of women engaged in high-in- tensity physical training, for whom amenorrhea is a common characteristic. Often, these types of sports also demand extremely low body weights and there is high reported incidence of anorexia nervosa among participants. The combination of amenorrhea (and/or anorexia) is detrimental to bone mass and there is now good evidence to show that they "under achieve" their peak bone mass (PBM) potential and thus are at a considerably increased risk of osteoporosis and indeed, by World Health Organization criteria, are often diagnosed as having the disorder. This picture of under-nutrition, amenorrhea, and osteoporosis is defined as the "female athletic triad" and in 1997, the American College of Sports Medicine published a position stand to "encourage the prevention, recognition and management of this syndrome" (American College of Sports Medicine, 1997) and this has now been updated and is referred to as Relative Energy Deficit Syndrome (REDS) (**Fig. 3**) (Mountjoy et al., 2018).

The exact mechanisms involved in PBM reduction remain unclarified. There are data to suggest that there is a sup pression of the osteoblasts rather than an increase in osteo clastic activity, a finding that is further supported by the finding that hormone replacement therapy (HRT) is not as effective in reducing bone loss in elite sportswomen as it is in young women with primary ovarian failure. Of further interest is the finding that in gymnasts, despite a high prevalence of oligo- and amenorrhea, bone mass shows a higher than predicted value (American College of Sports Medicine, 1997). Clearly this is an important group to study because it might provide insight into the type of mechanical loads that are most osteogenic.

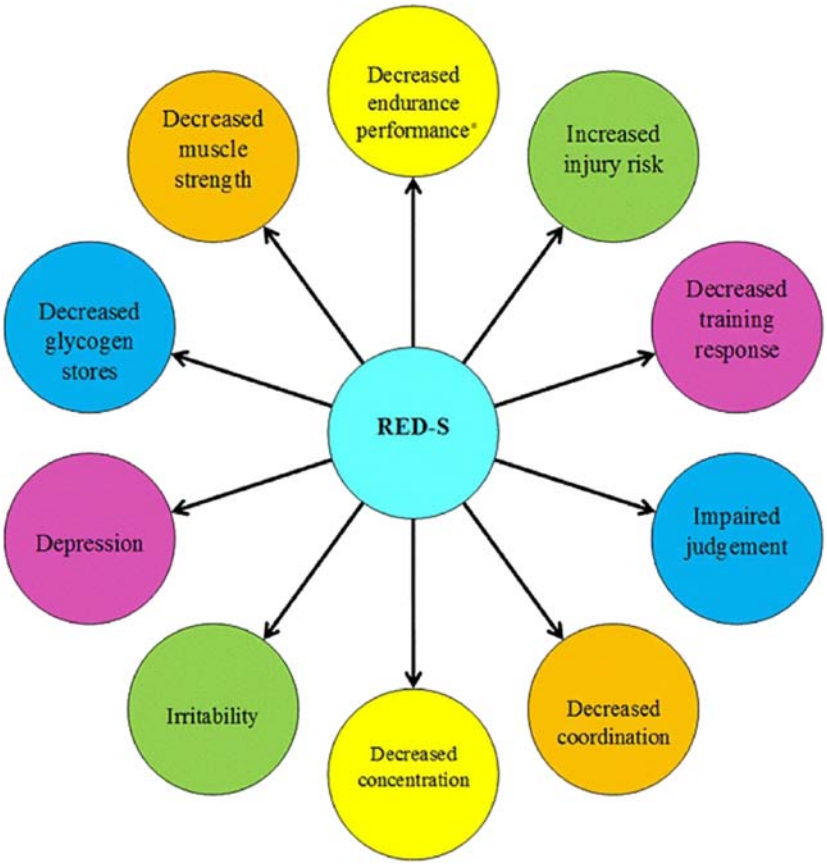


Fig. 3 Potential performance effects of Relative Energy Deficiency in Sport (REDS).

Conclusions

The effects of nutrition on the skeleton are key. There is evidence to suggest that such effects begin *in utero* and remain in place throughout the entire life span, providing target audiences on whom nutrition and bone health messages can be focused. On the dietary front, calcium and vitamin D are important nutrients for optimum bone health. At all costs, we must try to prevent population groups from the potential of suboptimum intakes and vitamin D and look to dietary strategies of fortification for particularly vulnerable groups such as the elderly, postmenopausal women, adolescent females, and amenorrheic women. Given that the decline in bone mass is seen as a natural aging phenomenon, we must encourage sufficient protein:energy nutrition in our aging population. Newer nutritional ideas, with a sound evidence-base for a positive effect are appearing, including dietary alkali. Although there are plausible mechanisms for the effect of other micronutrients on bone health, such as magnesium, trace elements, vitamin C, more research is required on the specific supplementation effects of these nutrients on markers of bone health.

Concluding remarks

It is common ground that there are genetic, environmental, lifestyle, and dietary determinants of risk of osteoporotic fracture as well as interactions between them. The key to secondary prevention is the understanding of how these components can be integrated into an effective assessment of the major risks after a previous fracture has occurred. The key to primary prevention is to understand both the pathological and physiological basis of bone fragility. It is not unreasonable to suppose that in a western lifestyle our limited and stereotypic patterns of locomotion from middle age onwards may offer considerably less protection than, for example, the more physically demanding activity of subsistence farming.

It is absolutely critical, given the fact that by 2030, 1:4 in the adult population will be elderly, that special attention is given to nutritional strategies for the optimization of bone health throughout the life cycle. Clearly, calcium and vitamin D nutrition are of great importance. Recent data suggest that calcium works synergistically with physical activity to enhance peak bone mass development and both should be on the agenda as recommended strategies for maximizing peak bone mass attainment during growth. At the other end of the age spectrum, calcium and vitamin D have been shown to be effective strategies for fracture prevention in the elderly, particularly for those populations where vitamin D insufficiency is rife.

Data continue to accumulate showing a positive impact of dietary potassium/fruit and vegetables on skeletal integrity. Although intervention trials are urgently required, it is sensible for the clinician to promote a high potassium intake in the patient's diet, because potassium has been shown to conserve calcium. A high intake of fruit and vegetables is likely to have numerous other health-related benefits.

Further data are urgently required to enable a fuller understanding of the complex interaction between dietary factors and bone health. However, there are promising data showing a positive impact of vitamin K on reducing post-menopausal bone loss. Furthermore, caution needs to be given to the overconsumption of carbonated soft drinks in our younger population, particularly if dairy products in the diet are displaced.

In this era of functional genomics, data are now urgently required to characterize the key nutrient:gene interactions that exist and that are likely to affect bone health in both our younger and older population groups. Targeting dietary advice at those genetically susceptible to osteoporosis is likely to become a useful toolset for the practising clinician.

References

American College of Sports Medicine, 1997. Position stand on the female athletic triad. Med. Sci. Sports Exerc. 29, i–ix.

Bischoff-Ferrari, H.A., Dawson-Hughes, B., Willet, W.C., et al., 2009. Fall prevention with supplemental and active forms of vitamin D: a meta-analysis of randomised controlled trials. Br. Med. J. 339 (b3692).

Bolland, M.J., Barber, A.P., Doughty, R.N., et al., 2008. Vascular events in healthy older women receiving calcium supplementation: randomised controlled trial. Br. Med. J. 336, 262–266.

Boudin, E., Fijalkowski, I., Hendrickx, G., et al., 2016. Genetic control of bone mass. Mol. Cell. Endocrinol. 432, 3–13.

Bouillon, R., Manousaki, D., Rosen, C., et al., 2022. The health effects of vitamin D supplementation: evidence from human studies. Nat. Rev. Endocrinol. 18, 96–110.

Buttriss, J.L., Lanham-New, S.A., Steenson, S., et al., 2022. Implementation strategies for improving vitamin D status and increasing vitamin D intake in the UK: current controversies and future perspectives: proceedings of the 2nd Rank Prize Funds Forum on vitamin D. Br. J. Nutr. 127 (10), 1567–1587.

Cashman, K.D., 2007. Diet, nutrition, and bone health. J. Nutr. 137 (11 Suppl. I), 2507S–2512S.

Chakhtoura, M., Bacha, D.S., Gharios, C., et al., 2022. Vitamin D supplementation and fractures in adults: a systematic umbrella review of meta-analyses of controlled trials. J. Clin. Endocrinol. Metab. 107 (3), 882–898.

Charoenngam, N., Shirvani, A., Holick, M.F., 2019. Vitamin D for skeletal and non-skeletal health: what we should know. J. Clin. Orthop. Trauma 10 (6), 1082–1093.

Cheng, S., Lyytikäinen, A., Kroger, H., et al., 2005. Effects of calcium, dairy product, and vitamin D supplementation on bone mass accrual and body composition in 10-12-y-old girls: a 2-y randomized trial. Am. J. Clin. Nutr. 82 (5), 1115–1126.

Darling, A.L., Millward, D.J., Torgerson, D.J., et al., 2009. Dietary protein and bone health: a systematic review and meta-analysis. Am. J. Clin. Nutr. 90, 1674–1692. Epub 4 Nov 2009. Editorial attached to paper. Kerstetter J. The American Journal of Clinical Nutrition. 90:1142-1143.

Darling, A.L., Manders, R.J.F., Sahni, S., et al., 2019. Dietary protein and bone health across the life-course: an updated systematic review and meta-analysis over 40 years. Osteoporos. Int. 30 (4), 741–761.

Darling, A.L., Millward, D.J., Lanham-New, S.A., 2021. Dietary protein and bone health: towards a synthesized view. Proc. Nutr. Soc. 80 (2), 165–172.

Hill, T.R., Aspray, T.J., 2017. The role of vitamin D in maintaining bone health in older people. Ther. Adv. Musculoskelet. Dis. 9 (4), 89–95.

Lanham-New, S.A., 2008. Importance of calcium, vitamin D and vitamin K for osteoporosis prevention and treatment. Proc. Nutr. Soc. 67, 163–176.

Lanham-New, S.A., Thompson, R.L., More, J., et al., 2007. Importance of vitamin D, calcium and exercise to bone health with specific reference to children and adolescents. Nutr. Bull. 32, 364–377.

Lambert, H., Frassetto, L., Moore, J.B., et al., 2015. The effect of supplementation with alkaline potassium salts on bone metabolism: a meta-analysis. Osteoporos. Int. 26, 1311–1318.

Liu, C., Kuang, X., Li, K., et al., 2020. Effects of combined calcium and vitamin D supplementation on osteoporosis in postmenopausal women: a systematic review and meta-analysis of randomised controlled trials. Food Funct. 11 (12), 10817–10827.

Matkovic, V., Goel, P.K., Badenhop-Stevens, N.E., et al., 2005. Calcium supplementation and bone mineral density in females from childhood to young adulthood: a randomised controlled trial. Am. J. Clin. Nutr. 81, 175–188.

Mott, A., Bradley, T., Wright, K., et al., 2019. Effect of vitamin K on bone mineral density and fractures in adults: an updated systematic review and meta-analysis of randomised controlled trials. Osteoporos. Int. 30 (8), 1543–1559. Epub 2019 May 10. Erratum in: Osteoporosis International. 2020; 31(11): 2269-2270.

Mountjoy, M., Sundgot-Borgen, J.K., Burke, L.M., et al., 2018. IOC consensus statement on relative energy deficiency in sport (RED-S): 2018 update. Br. J. Sports Med. 52, 687–697.

New, S.A., 2002. The role of the skeleton in acid-base homeostasis. Proc. Nutr. Soc. 61, 151–164.

Nieves, J.W., 2005. Osteoporosis: the role of micronutrients. Am. J. Clin. Nutr. 81, 1232S–1239S.

Reid, I.R., Bolland, M.J., Grey, A., 2014. Effects of vitamin D supplementation on bone mineral density: a systematic review and meta-analysis. Lancet 383 (9912), 146–155.

Shea, B., Wells, G., Cranney, A., et al., 2002. VII. Meta-analysis of calcium supplementation for the prevention of postmenopausal osteoporosis. Endocr. Rev. 23 (4), 552–559.

Swart, K.M.A., van Schoor, N.M., Lips, P., 2013. Vitamin B12, folic acid, and bone. Curr. Osteoporos. Rep. 11, 213–218.

Wilson-Barnes, S.L., Lanham-New, S.A., Lambert, H., 2022. Modifiable risk factors for bone health & fragility fractures. Best Pract. Res. Clin. Rheumatol. https://doi.org/10.1016/j.berh.2022.101758.

Nutritional assessment: Anthropometry

J. Eaton-Evans, University of Ulster, Coleraine, UK

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of J. Eaton-Evans, Nutritional Assessment: Anthropmetry, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 227–232, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00197-5.

Uses of Anthropometric Measurements **87**
Advantages and Limitations of Anthropometric Measurements **88**
Errors of Anthropometric Measurements **88**
Anthropometric Measurements **88**
Height 88
Weight 89
Head Circumference 89
Midupper Arm Circumference (MUAC) 89
Skinfold Thickness 89
Waist and Hip Circumferences 90
Elbow Width 90
Nutritional Indices 90
Body Mass Index (BMI) 91
Weight-for-Height 92
Weight-for-Age 92
Growth Velocity 92
Head Circumference-for-Age 92
Midupper Arm Circumference-for-Age 92
Midupper Arm Circumference-for-Height 92
Skinfold Thickness-for-Age 92
Midupper Arm Muscle Circumference (MUMAC) and Upper Arm Muscle Area (AMA) 92
Arm Fat Area (AFA) 93
Waist-to-Hip Ratio (WHR) 93
Reference Values **93**
Adults 93
Children 93
Further Reading **93**

Glossary

Kyphosis Abnormal curvature of the spine, resulting in a reduced height.

Occiput The most prominent part of the back of the cranium. Used as a reference to place the measuring tape to obtain head circumference.

Stadiometer Calibrated board used to measure height.

Stunting, stunted Chronic undernutrition, resulting in a reduced height-for-age and weight-for-age.

Uses of Anthropometric Measurements

In adults and children, anthropometric measurements can be used to estimate body fat and lean body mass, and assess their distribution and change over time. Body fat includes storage fat, found inter- and intramuscularly, around the organs and gastrointestinal tract and subcutaneously, as well as lipids in bone marrow, central nervous tissue, mammary glands, and other organs. Normal-weight men and women have approximately 10% and 20% body fat, respectively. Lean body or fat-free mass is approximately 73% water and protein with relatively small amounts of glycogen and minerals. Inadequate diets are associated with low body fat stores and reduced lean body mass in adults and growth failure of children. Consumption of food greater than requirements results in excessive body fat stores in adults and children. Body fat stores, which are too low or too high, are associated with increased risk of morbidity and mortality. The proportion and distribution of fat and fat-free mass vary with age, sex, genetics, disease, some hormones, and some drug treatments.

Different anthropometric measurements and combinations of measurements provide information on body composition and fat distribution and, therefore, nutritional status. The choice of measurements depends on the purpose of the assessment, the equipment available, the subjects being measured, and the skills of the observer making the measurements. Measurements can be made in laboratories, clinics, and hospitals using fixed, precision equipment with a high degree of accuracy, or in the field, including people's homes or rural centers, with portable equipment.

Advantages and Limitations of Anthropometric Measurements

Anthropometric measurements are quick and relatively easy to obtain, and require inexpensive equipment. For most measurements, there are adequate measurement protocols and reference standards for comparison. Limitations include inability to detect small differences in body composition, dependency on subject's cooperation (problem with small children and handicapped adults), and require an operator with a certain amount of training and experience, particularly for skinfold measurements.

Errors of Anthropometric Measurements

All anthropometric measurements should be made as accurately as possible. Measurement errors may result in the misclassification of subjects' nutritional status or may lead to changes in nutritional status over time being over- or underestimated. Very precise and accurate measurements are needed for nutrition research and in some clinical situations. The same degree of precision may not be possible in nutritional screening and surveillance programs in field studies. Errors in making measurements arise from the equipment, the physical state and age of the subjects, the time of day when the measurements are made, misreading of measurements by the observer, and as a result of rounding up or down to the nearest half or whole integer. These technical errors of measurement (TEM) vary with the age of the subjects, the measurements being made, and between (inter-) and within (intra-) observers. Values for a particular anthropometric measurement of a group of people by age and sex can be considered accurate if the inter- and intraobserver error is close to a reference value for TEM in a series of repeated measurements, and if there are no biases in the measurement. For measurements of subjects outside the age range, the coefficient of (R) can be calculated as $R=1-[(\text{TEM})^2/(\text{SD})^2]$, where SD is the total inter-subject variance including measurement error. It has been recommended that an R of 0.90, i.e., a measurement 90% error-free, is an acceptable lower limit of accuracy, although an intraobserver R of 0.95 might be more realistic in some circumstances.

TEM can be minimized by careful training of all observers and by making measurements using appropriate equipment in triplicate and then calculating the mean. If measurements for a research study are to be made by more than one person, the interobserver measurements made must be comparable. R can be calculated for interobserver variability by making a series of measurements.

Anthropometric Measurements

Height

Height or stature is measured in adults and children over the age of 2 years using a stadiometer, a portable anthropometer, or a moveable headboard on a vertical measuring rod. The measuring device should be checked for accuracy using a standard 2-m steel tape. Subjects should be measured to the nearest 0.1 cm. Subjects, in minimal clothing with bare heads and feet, should stand straight, arms hanging loosely to the side, feet together and with heels, buttocks and shoulder blades in contact with the vertical surface of the stadiometer. Errors occur if subjects do not stand straight, do not keep heels on the ground, or overstretch. Diurnal variation results in people being 0.5–1 cm shorter in the evening than in the morning.

Height cannot be measured accurately in adults with severe kyphosis of the spine and in those who are bed- or chair-ridden. Because knee height is highly correlated with stature, height in such adults can be estimated from the measurement of knee height, using a sliding calliper. The regression equations, derived from a nonrandom sample of American people over the age of 60 years, are as follows: Height (cm) for men=(2.02×knee height, cm)−(0.04×age, years)+64.19; height (cm) for women=(1.83×knee height, cm)−(0.24×age, years)+84.88. Variations in the proportion of limb length to trunk length can lead to a standard error in the estimate (SEE) of height from knee height of ±8 cm. Demispan, which is the distance between the sternal notch of the right collar bone and the left finger root of the middle and ring finger when the subject's arm is horizontal and in line with the shoulders, can also be used to estimate height.

Length, rather than height, is measured in infants and children under the age of 3 years. Length is measured by laying a child face upwards on a measuring board with the head against the fixed headboard, and moving another board up to and resting against the child's heels with the legs straight (**Figure 1**). Small changes in length (±0.5 cm) may not be significant as it is a difficult measurement to make. Children wriggle and will not stretch out their legs. Length measurements are 1–2 cm longer than height.

Height (stature) or length indicates attained size or growth of adults and children. Long periods of inadequate food intake or increased morbidity result in a slowing of skeletal growth and individuals being short for their age, or stunted. Consecutive

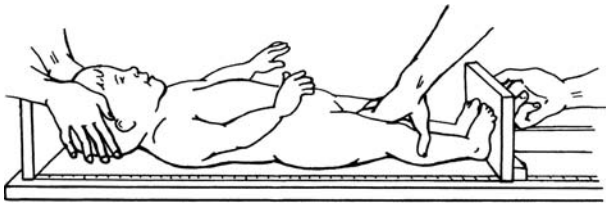


Figure 1 Measurement of recumbent length in children younger than 3 years of age. The head should be in contact with the fixed headboard, with child facing straight up. With legs fully extended, the mobile footboard should be placed firmly against the infant's heels. Reproduced with permission from Frisancho AR (1990) *Anthropometric Standards for the Assessment of Growth and Nutritional Status*. Ann Arbor: University of Michigan Press.

measurements of height every 3–6 months can be used to assess growth velocity in children and to indicate the timing of the adolescent growth spurt.

Weight

Weight is measured with digital weighing scales, using a pan, basket, sling, standing platform, or chair, depending on the age and mobility of the people being measured. Weighing scales must be set on a hard, level, and even surface. Scales should be accurate, sensitive, and robust. They must be carefully maintained, calibrated, regularly checked for accuracy using known weights, and always set at zero before use. Weight is usually measured to the nearest 0.1 kg for adults and 0.01 kg for infants.

Weight measures total body mass but does not provide information on the proportions of fat, water, protein, and minerals. Adults can be heavy for height if very muscular, over-fat, and big framed. With accurate scales, small changes in weight are detectable but may not necessarily reflect change in body fat or lean body mass. In healthy persons, day-to-day variation in body weight is usually small (±0.5 kg). Consecutive measurements of weight can be used to monitor the effects of treatment such as weight loss on reduction diets or weight gain with nutritional interventions and supplementation. Weight changes are assumed to reflect changes in the amount of body fat. However, changes in body weight may also result from differences in hydration, edema, tumor growth, and trauma, as well as from factors such as the amount of food in the gastrointestinal tract and the fullness of the bladder. Weight may remain constant if the loss of muscle mass is masked by increased fat as seen in sarcopenia, the age-related loss of muscle, or by increased fluid retention.

Weight-for-height (or length) can be used to indicate body composition in adults and is an age-independent measure of body composition in children. Growth can be measured in children by consecutive measurements of weight over time (growth velocity) or by weight-for-age if the children's ages are known.

Head Circumference

Head circumference is measured in infants and young children, to the nearest 0.1 cm, with a narrow flexible nonstretch tape laid over the supraorbital ridges and the part of the occiput, which gives the maximum circumference. The head circumference of infants increases rapidly in the first 2 years of life. Increase in head circumference in the first 2 years of life is affected by nutritional status and nonnutritional problems, including some diseases, genetic variation, and cultural practices.

Midupper Arm Circumference (MUAC)

MUAC is measured in adults and children, to the nearest 0.1 cm, using a flexible nonstretch tape laid at the midpoint between the acromion and olecranon processes on the shoulder blade and the ulna, respectively, of the arm (**Figure 2**). MUAC is a measure of the sum of the muscle and subcutaneous fat in the upper arm. In severe malnutrition both fat and muscle are reduced in the upper arm. Edema may increase a limb's circumference but it is not usually a problem of the upper arm. MUAC can be used as a indicator of body composition in adults and children. Since MUAC increases little between the age of 6 months and 5 years, it can be used in preschool children as an age-independent screening tool for severe malnutrition. An MUAC less than 12.5 cm suggests malnutrition; an MUAC greater than 13.5 cm is normal.

Skinfold Thickness

Precision skinfold thickness callipers are used to measure the double fold of skin and subcutaneous fat to the nearest millimeter. The usual sites of measurement are at the triceps (TSFT), the midpoint of the back of the upper arm (**Figure 3**); the biceps (BSFT) at the same level as the TSFT but to the front of the upper left arm; the subscapular (SSFT) just below and laterally to the left shoulder blade (**Figure 4**), and the suprailiac (SISFT) obliquely just above the left iliac crest. Skinfold thicknesses can also be measured at the midthigh, midcalf, and abdomen.

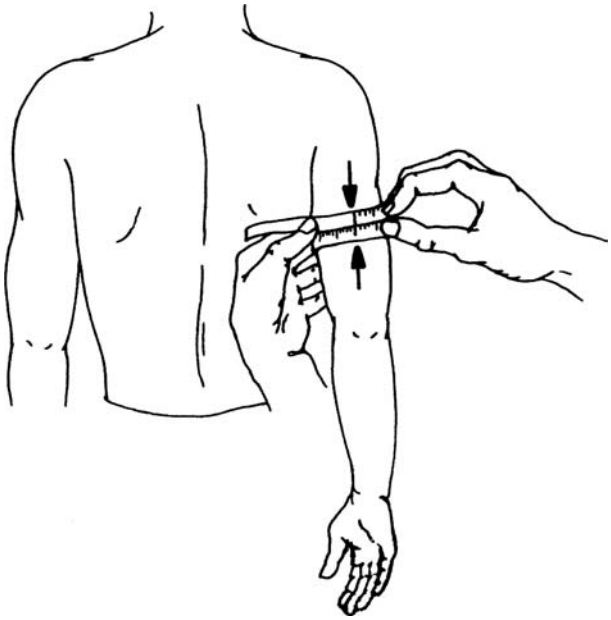


Figure 2 Measurement of upper arm circumference at the midpoint of the upper arm. Reproduced with permission from Frisancho AR (1990) *Anthropometric Standards for the Assessment of Growth and Nutritional Status*. Ann Arbor: University of Michigan Press.

Skinfold thicknesses are difficult measurements to make with precision and accuracy without rigorous training. It is difficult to pick up a consistent fold of skin and subcutaneous fat; in the very obese, the skinfold may be bigger than the callipers can measure; the fold of skin and fat compresses with repeated measurements; and the careless use of the callipers causes pain, bruising, and skin damage to subjects. There is, therefore, likely to be considerable inter- and intraobserver error in the measurements.

Skinfold thicknesses measure subcutaneous body fat and, therefore, indicate body composition. TSFT and SSFT indicate subcutaneous fat on the limbs and body trunk, respectively. Skinfold thickness measurements mistakenly assume that subcutaneous fat, measured at one or more selected sites, measures total body fat stores. However, subcutaneous fat at one site may not reflect fat stores at another site, and may not be positively correlated with the amount of visceral fat deposited around the internal organs of the body. Subcutaneous fat, and therefore skinfold thicknesses at the different sites, changes at varying rates with age, weight change, with diseases such as diabetes, and in women during pregnancy, postpartum, and at the menopause. Skinfold thicknesses are not useful for monitoring short-term change in fat stores. If only one skinfold thickness measurement is made, TSFT is most commonly selected. TSFT correlates with estimates of total body fat in women and children. SSFT is better than TSFT as an indicator of total body fat in men. SSFT has been shown to be a predictor of blood pressure in adults independently of age and racial group.

Waist and Hip Circumferences

Waist and hip circumferences are measured to the nearest 0.1 cm using a flexible narrow nonstretch tape in adults wearing minimal clothing, standing straight but not pulling in their stomach. Waist circumference is measured halfway between the lower ribs and the iliac crest, whereas hip circumference is measured at the largest circumference around the buttocks. Measurement error occurs if the tape is pulled too tight or loose, or if subjects wear clothes with belts or full pockets.

With increase in waist circumference there is a decrease in insulin sensitivity, and an increase risk for cardiovascular disease.

Elbow Width

Elbow width is the width of the epicondyles of the humerus with the elbow flexed at 90°. Sliding callipers are used to measure elbow width in adults to the nearest 0.1 cm. Elbow width is a measure of bone size. Frame size can be determined by comparison with reference values either by age or by height and sex.

Nutritional Indices

Most single anthropometric measurements do not in themselves assess nutritional status. Nutritional indices are derived either by combining two or more anthropometric measurements, shown in laboratory studies to be predictive of body composition, or by

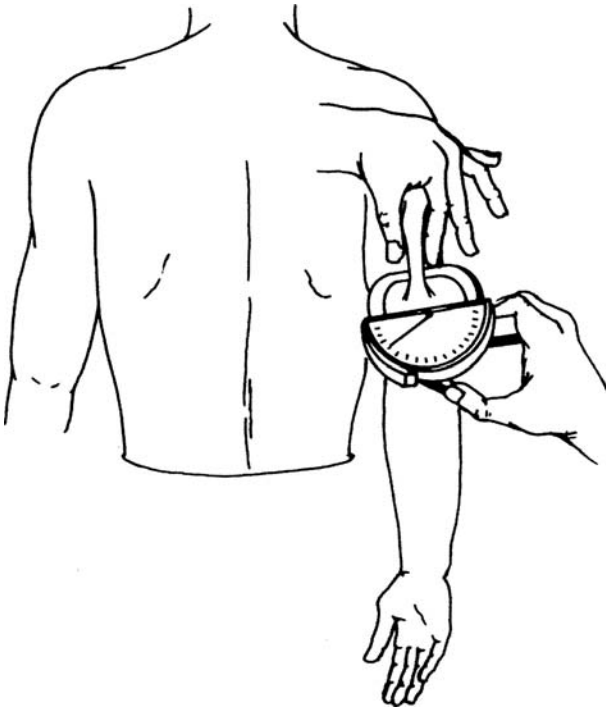


Figure 3 Measurement of triceps skinfold using a Lange caliper. With the subject's arm in a relaxed position, the skinfold is picked with thumb and index fingers at the midpoint of the arm. Reproduced with permission from Frisancho AR (1990) *Anthropometric Standards for the Assessment of Growth and Nutritional Status*. Ann Arbor: University of Michigan Press.

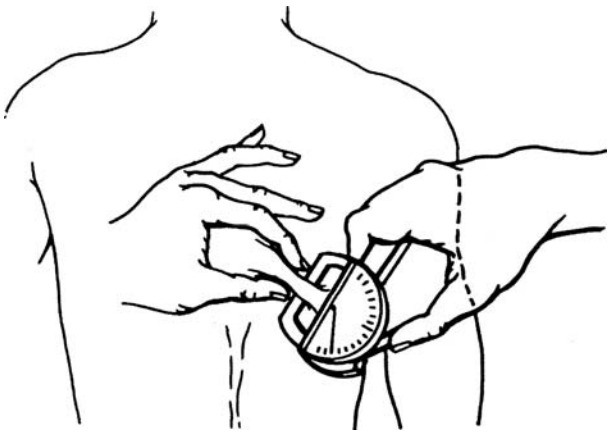


Figure 4 Measurement of subscapular skinfold using a Lange caliper. With subject's arm and shoulder relaxed, a horizontal skinfold is picked approximately 1 cm below the tip of the scapula with thumb and index fingers. The caliper is applied 1 cm from fingers. Reproduced with permission from Frisancho AR (1990) *Anthropometric Standards for the Assessment of Growth and Nutritional Status*. Ann Arbor: University of Michigan Press.

comparison of the anthropometric measurements with reference values of healthy, well-fed populations. A combination of these methods can also be used.

Body Mass Index (BMI)

BMI relates weight (kg) with height (m) by a simple calculation to indicate body composition (BMI=weight/height^2). It is the most commonly used screening measurement for both obesity and underweight, and to track growth patterns. By consensus, a healthy range of BMI for adults has been defined as the interval between 18.5 and 24.9 kg m^{-2}. Overweight or 'mild' obesity is defined from 25 to 29.9, obesity from 30 to 39.9, and morbid or severe obesity as >40. In children, because weight and height change proportions at different stages of growth, a normalized percentile distribution is used, defining the range of desired BMI for age and gender as

between the 5th and 95th percentiles, In most but not all age groups, there is a good correlation between BMI and body fat in the population, but predicting body fat from BMI at an individual level can be misleading.

Weight-for-Height

A deficit of weight relative to height indicates recent (acute) weight loss or undernutrition. Coupled with other rapid assessment methods (such as midarm circumference), a reduced weight-for-height may alert on acute, severe undernutrition requiring prompt response.

Weight-for-Age

Weight-for-age may indicate either acute or chronic weight deficit, and thus cannot be used to discriminate between these two conditions. The most common cause of a reduced weight for age is chronic undernutrition, also called stunting. It is usually accompanied by a reduced height-for-age, resulting in a child who is of small body size for his/her age, but otherwise healthy and with a BMI within the normal range.

Growth Velocity

Growth velocity, or change in weight or height over time, can be used to assess growth in children when compared with reference values by age and sex. Growth rates decline in the first few years of life and then increase with the pubertal growth spurt. Premature and small-for-dates children and those recovering from malnutrition and severe infections tend to have higher growth velocities (catch-up growth). Growth velocities are useful to monitor growth and assess the response to therapy including nutritional supplementation.

Head Circumference-for-Age

Head circumference-for-age by sex is used by pediatricians to identify children up to 2 years of age with severe chronic malnutrition pre- and postpartum and the need for further medical investigations. It is not a good indicator of children's nutritional status.

Midupper Arm Circumference-for-Age

MUAC-for-age indicates body composition (upper arm fat and muscle) in adults and children when used with measurements of weight and height. MUAC measurements are compared with reference values by age and sex. Because the rate of change of arm circumference is slow, it cannot be used to assess growth or monitor the response to therapy.

Midupper Arm Circumference-for-Height

Midupper arm circumference-for-height (the QUAC stick) is a cheap, quick, and age-independent screening tool for children with malnutrition. It is a vertical stick on which are inscribed the 80% and 85% median reference values for MUAC and height, respectively. A child is considered malnourished if the MUAC is less than 80% of the MUAC expected for height.

Skinfold Thickness-for-Age

Measurement of BMI with skinfold thicknesses can identify people who are heavy owing to excess fat or muscle mass. A high BMI and low TSFT and SSFT indicate a large muscle mass; a high BMI and high TSFT and SSFT indicate a high subcutaneous body fat.

Midupper Arm Muscle Circumference (MUMAC) and Upper Arm Muscle Area (AMA)

MUAMC and upper AMA are estimates of upper arm muscle and, therefore, body composition. They can be used as indicators of muscle mass and protein stores. Both MUAMC and AMA are calculated from measurements of MUAC and TSFT on the mistaken assumption that the arm is cylindrical, the subcutaneous fat is equally distributed, the bone atrophies in proportion to muscle wastage in malnutrition, and the cross-sections of neurovascular tissue and bone are small. The formula, with MUAC and TSFT in millimeter, is as follows: MUAMC=MUAC−(π×TSFT)AMA can be calculated from revised formulae, which take account of errors resulting from the noncircular nature of muscle and the inclusion of nonskeletal muscle with MUAC and TSFT in centimeters. For men: AMA = [MUAC−(π×TSFT)/4π]^2−10.0. For women: AMA = [MUAC−(π×TSFT)/4π]^2−6.5. MUAMC and AMA can be compared with reference values by age and sex. AMA cannot be used to monitor change in muscle stores because of the problems in making this measurement. The ratio of AMA to total body muscle mass changes with age and certain diseases.

Arm Fat Area (AFA)

AFA can be derived from measurements of MUAC and TSFT. AMA is a better indicator of total body fat but not percentage body fat, than TSFT alone. The formula used to calculate AFA (with MUAC and TSFT in millimeter) is AMA $= [(\text{TSFT} \tilde{n} \text{MUAC}/2) - (\pi \tilde{n} \text{TSFT}/4)]^2$. AMA can be compared with reference values by age and sex. Theoretically, limb fat area can be calculated for other limbs and the body trunk, but there are no reference values available.

Waist-to-Hip Ratio (WHR)

The WHR in adults discriminates between those with upper body or intraabdominal obesity (WHR greater than 1 in men and 0.8 in women) and those with lower body or peripheral obesity. Genetics, sex, and age partly determine body fat distribution. A high WHR is associated with an increased risk of premature mortality and morbidity.

Reference Values

Adults

Reference values for adults have been historically based on the relationship between body size and mortality risk. Since the early twentieth century, life insurance companies used actuarial tables to adjust premiums based on risk of premature death. This risk curve is J-shaped, and the BMI interval comprising the lowest portion of the curve was subsequently used to identify the desired or healthy BMI range. By consensus that interval was eventually defined as between BMI of 18.5 and 24.9, which is considered healthy or 'normal' BMI for adults. The segment from 25 to 29.9 is sometimes called overweight or mild obesity, from 30 to 39.9 obesity, and above, severe obesity.

There is ongoing debate on the limitations of using BMI to predict adverse health events. While the link of diseases such as type 2 diabetes and dislipidemias with obesity (i.e., high BMI) is well established, it is less clear if this indicator can be used to track risk for a number of other disorders. Furthermore, the risk level-BMI association varies with ethnicity, making difficult to adopt a universal cutoff point, as has been attempted.

Children

Traditional reference growth charts are descriptive, meaning that they reflect the body size of the population selected as 'reference'. Given the secular trends of 'healthy' populations to increase their body size, reference values derived from recent measurements would yield undesirably high BMIs. Therefore, reference charts, such as the CDC 2000, have used older datasets, collected at a time when excess weight was less prevalent.

The World Health Organization (WHO) released in 2006 the first prescriptive reference charts for children 0–5 years. These charts reflect the growth patterns of children raised on a healthy environment and fed and stimulated according to recommended practices. The WHO subsequently released charts for 5–19 years, following the more traditional descriptive approach but introducing a number of adjustments to optimize use of the chart across countries.

Further Reading

Centers for Disease Control and Prevention (2005) *Growth Charts Dataset*. Available from www.cdc.gov/nchs/about/major/nhanes/growthcharts/datafiles.htm

Frisancho, A.R., 1990. Anthropometric Standards for the Assessment of Growth and Nutritional Status. University of Michigan Press, Ann Arbor.

Must, A., Dallal, G.E., Dietz, W.C., 1991. Reference database for obesity: 85th and 95th percentiles of body mass index (wt/ht 2) and triceps shilnfold thickness. American Journal of Clinical Nutrition 53, 839–846.

World Health Organization, 1995. Physical Status: The Use and Interpretation of Anthropometry. World Health Organization, Geneva. WHO Technical Report Series No. 854.

World Health Organization: Child Growth Standards (2006) Available from http://www.who.int/childgrowth/en/ (accessed 12 January 2011).

Adipose tissue: Structure, function and metabolism

G. Frühbeck and J. Gómez-Ambrosi, Clínica Universidad de Navarra, University of Navarra, CIBEROBN, Instituto de Salud Carlos III, IdiSNA, Pamplona, Spain

© 2023 Elsevier Ltd. All rights reserved.

This is an update of G. Frühbeck, J. Gómez-Ambrosi, Adipose Tissue: Structure, Function and Metabolism, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 1–13, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00001-5.

Introduction 95
Development 96
Structure 99
Distribution 101
Function 102
Regulation of metabolism 104
Conclusion 107
References 107
Further reading 108

Key points

- Identify the adipose tissue as an extremely active endocrine organ involved in multiple biological functions
- Distinguish the different types and functions of adipose tissue as well as the diverse distribution characteristics
- Consider the physiological depot- and gender-specific differences of adipose tissue and their derangement in pathophysiological circumstances
- Recognize the implication of dysfunctional excess adiposity in the development of the comorbidities accompanying obesity

Glossary

Adipogenesis Development or formation of fat

Angiogenesis Formation of new blood vessels by branching morphogenesis

Autocrine A secreted substance which acts on surface receptors of the same cell

Catecholamines Any of various amines (e.g., epinephrine, norepinephrine, and dopamine) that contain a dihydroxy benzene ring, are derived from tyrosine and operate as hormones and/or neurotransmitters

Fatty-acid-binding proteins (FABPs) Family of carrier proteins for fatty acids and other lipophilic substances such as eicosanoids and retinoids. These proteins are thought to facilitate the transfer of fatty acids between extra- and intra-cellular membranes. Some family members are also believed to transport lipophilic molecules from the plasma membrane to certain intracellular receptors such as PPAR

Lipogenesis Normal deposition of fat or the conversion of carbohydrate or protein to fat

Lipolysis Triacylglycerol breakdown to yield fatty acids and glycerol

Paracrine Mode of hormone action in which a hormone binds to receptors on and affects the function of cells near to the cell that produced it

Peroxisome proliferator-activated receptors (PPARs) Group of nuclear receptor proteins that function as transcription factors regulating the expression of genes. PPARs play essential roles in the regulation of cellular differentiation, development, and metabolism (carbohydrate, lipid and protein) of higher organisms

Pleiotropic Producing many effects; multiple effects from a single gene; the control by a single gene of several distinct and seemingly unrelated phenotypic effects

List of Abbreviations

ALBP/FABP4/aP2 Adipocyte fatty acid binding protein

BAT Brown adipose tissue

BMP Bone morphogenetic protein

C/EBPs CCAAT/enhancer binding proteins
CETP Cholesteryl ester transfer protein
CRP C-reactive protein
HSL Hormone-sensitive lipase
IL Interleukin
LPL Lipoprotein lipase
MCP-1 Monocyte chemoattractant protein-1
PGAR/FIAF Peroxisome proliferator-activated receptor angiopoietin related protein/fasting-induced adipose factor
PGC-1α Peroxisome proliferator-activated receptor-γ coactivator-1α
PPAR Peroxisome proliferator-activated receptor
PRDM16 Positive regulatory domain containing 16
Pref-1 Preadipocyte factor-1
RBP4 Retinol binding protein-4
UCP Uncoupling protein
WAT White adipose tissue

Introduction

The role of white adipose tissue (WAT) in storing and releasing lipids for oxidation by skeletal muscle and other tissues became so firmly established decades ago that a persistent lack of interest hindered the study of the extraordinarily dynamic behavior of adipocytes (Frühbeck et al., 2001; Scherer, 2019). However, disentangling the neuroendocrine systems, which regulate energy homeostasis and adiposity has jumped to a first-priority challenge, with the recognition of obesity as one of the major public health problems. Strictly speaking, obesity is not defined as an excess of body weight but as an increased adipose tissue accretion, to the extent that health may be adversely affected. Therefore, in the last decades, adipose tissue has become the research focus of biomedical scientists for epidemiological, pathophysiological, and molecular reasons (van Baak and Mariman, 2019). Although the primary role of adipocytes is to store triglycerides during periods of caloric excess and to mobilize this reserve when expenditure exceeds intake, it is now widely recognized that adipose tissue lies at the heart of a complex network participating in the regulation of a variety of quite diverse biological functions (Fig. 1).

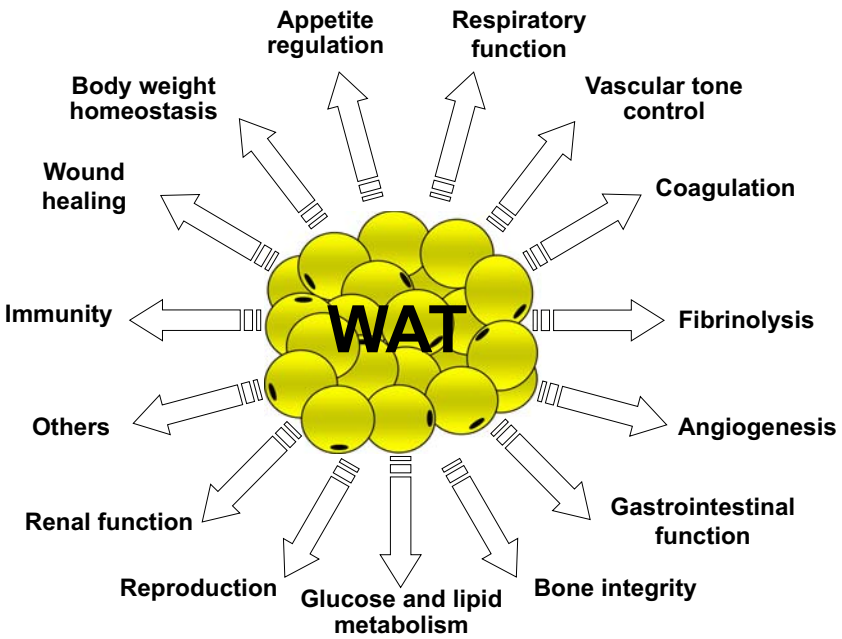


Fig. 1 Dynamic view of white adipose tissue based on the pleiotropic effects on quite diverse physiological functions. WAT, white adipose tissue.

Development

During fetal development WAT emerges at midgestation in humans or postnatally in rodents. Although the evolutionary and developmental features of WAT and brown adipose tissue (BAT) already suggested that they are quite distinct tissues, until recently, white and brown adipocytes were thought to be derived from the same precursor cell (Frühbeck et al., 2009b). Elegant *in vivo* fate mapping experiments in mice have recently provided a clear evidence that brown adipocytes arise from a separate and distinct population of progenitors (Fig. 2).

Brown fat cells are now known to exhibit a "myogenic" signature and share a common mesenchymal origin with skeletal muscle. In brief, mesenchymal stem cells can enter several cell lineages which culminate in the formation of bone, muscle, and adipose tissue, among others. The precursor cells destined to become white adipocytes first differentiate into adipoblasts and then preadipocytes through carefully-timed exposure to key regulators. Two members of the family of bone morphogenetic proteins (BMP), specifically BMP2 and BMP4, as well as PPARγ and C/EBPs are pivotal to drive these different phases. *Myf5*-expressing precursors give rise to skeletal muscle and brown adipose tissue, but not white adipose tissue (Fig. 2). BMP7 singularly drives the brown fat cell fate in both mesenchymal progenitor cells and committed brown preadipocytes suppressing early adipogenic inhibitors, such as necdin, *Pref-1*, and *WNTs*, at the same time inducing the key molecular determinant positive regulatory domain containing 16 (PRDM16) that triggers the activation of the complete brown adipogenesis program and blocks the induction of myotube-specific genes such as *Myf5*, *MyoD* and *myogenin*. PRDM16 binds and coactivates PPARγ with subsequent induction of key features to specify a brown fat fate, i.e., increased mitochondrial biogenesis and expression of UCP1, among others.

Pericytes, the cells that surround the endothelium tubes of the microvasculature, have been also identified as progenitor cells that become committed to the white adipocyte lineage either prenatally or in the early postnatal period. Adipose tissue has long been

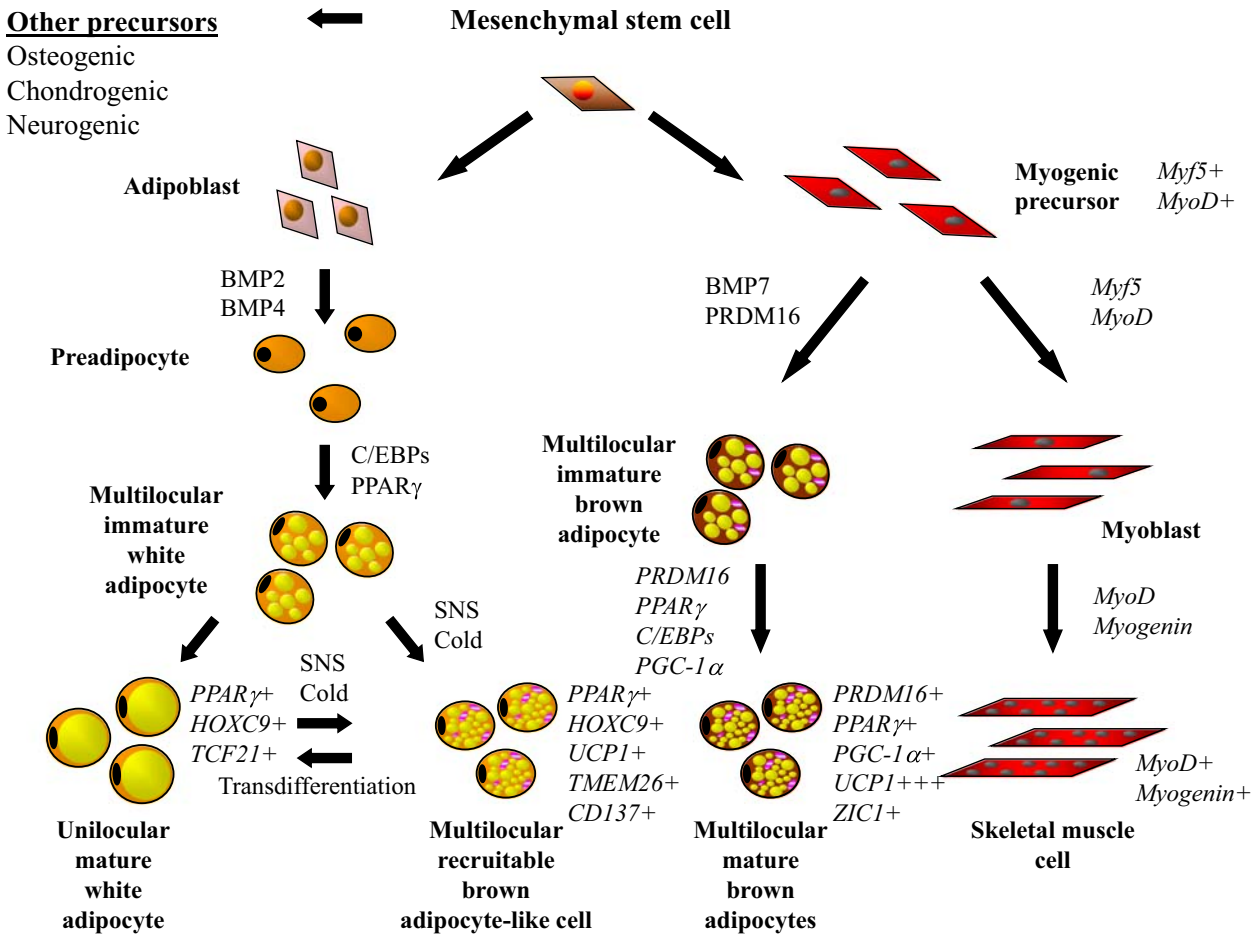


Fig. 2 Schematic diagram of the histogenesis of white and brown adipocytes. BMP, bone morphogenetic protein; C/EBPs, CCAAT/enhancer binding proteins; PGC-1α, peroxisome proliferator-activated receptor-γ coactivator-1α; PPAR, peroxisome proliferator-activated receptor; PRDM16, positive regulatory domain containing 16; UCP1, uncoupling protein 1.

recognized to expand in conjunction with its vasculature, but these new findings suggest that the blood vessels may actively direct the process (Bouloumie et al., 2002; Tang et al., 2008; Yang et al., 2013; Crewe et al., 2017). Thus, as well as serving as a progenitor niche, blood vessels may also produce signals for adipocyte development.

The determination phase results in the conversion of the stem cell to a preadipocyte, which still shares some morphological features with its precursor cell but has lost the potential to differentiate into other cell types (Fig. 3). In the subsequent phases of terminal differentiation, the committed preadipocyte takes on the characteristics of the white mature adipocyte by acquiring all the machinery needed for lipid transport, synthesis, and mobilization, hormonal responsiveness and the secretion of adipocyte-specific proteins. The morphological and functional changes that take place in the course of adipogenesis represent a shift in transcription factor expression and activity leading from a primitive, multipotent state to a final phenotype characterized by alterations in cell shape and lipid accumulation (Ghaben and Scherer, 2019). Various redundant signaling pathways and transcription factors directly influence fat cell development by converging the upregulation of PPARγ, which embodies a common and essential regulator of adipogenesis as well as of adipocyte hypertrophy. Among the broad panoply of transcription factors C/EBPs and the basic helix–loop-helix family (ADD1/SREBP-1c, adipocyte determination and differentiation factor-1/sterol regulatory element binding protein-1c) also stand out together with their link with the existing nutritional status. The transcriptional repression of adipogenesis includes both active and passive mechanisms. The former directly interferes with the transcriptional machinery, whereas the latter is based on the binding of negative regulators to yield inactive forms of known activators.

Adipose tissue develops extensively in homeotherms with the proportion to body weight varying greatly between species. There are two processes of adipose tissue formation. In the primary fat formation, which takes place relatively early (in human fetuses the first traces of a fat organ are detectable between 14 and 16 weeks of prenatal life), gland-like aggregations of epitheloid precursor cells, called lipoblasts or preadipocytes are laid down in specific locations and accumulate multiple lipid droplets. The secondary fat formation takes place later in fetal life (after the 23rd week of gestation) as well as in the early postnatal period, whereby the differentiation of other fusiform precursor cells that accumulate lipid to ultimately coalesce into a single large drop per cell leads to the dissemination of fat depots formed by unilocular white adipocytes in many areas of connective tissue. Adipose tissue may be

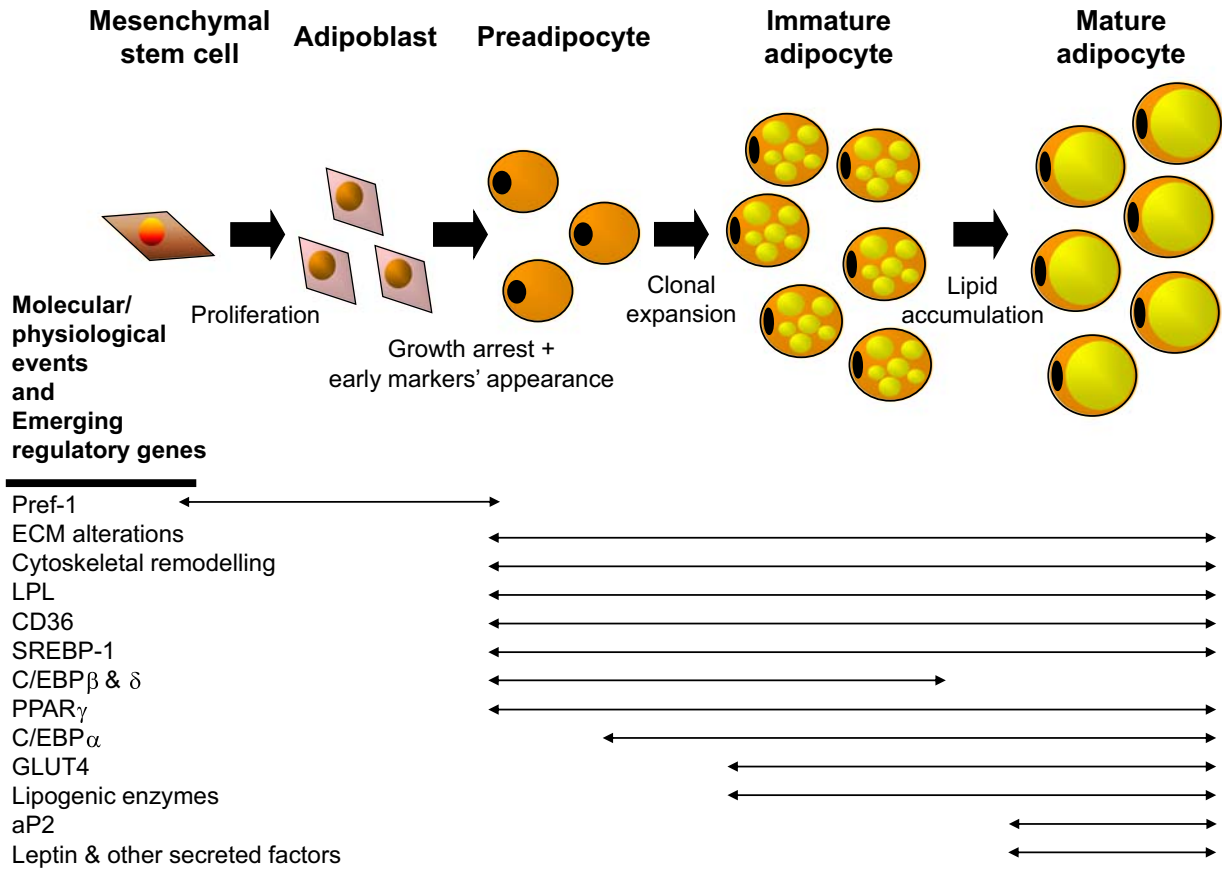


Fig. 3 Multistep process of adipogenesis together with events and participating regulatory elements. aP2, adipocyte fatty acid binding protein; CD36, fatty acid translocase; C/EBPs, CCAAT/enhancer binding proteins; ECM, extracellular matrix; GLUT4, glucose transporter type 4; LPL, lipoprotein lipase; PPAR, peroxisome proliferator-activated receptor; SREBP-1, sterol regulatory element binding protein-1.

partitioned by connective tissue septa into lobules. The number of fat lobules remains constant, whereas in the subsequent developmental phases the lobules' size grows continuously. At the sites of early fat development, a multilocular morphology of adipocytes predominates, reflecting the early developmental stage. Microscopic studies have shown that the second trimester may be a critical period for the development of obesity in later life. At the beginning of the third trimester, adipocytes are present in the main fat depots but are still relatively small. During embryonic development it is important to emphasize the temporo-spatial tight coordination of angiogenesis with the formation of fat cell clusters. At birth, body fat has been reported to account approximately for 16% of total body weight (with brown fat constituting 2–5%) with an increase in body fat from approximately 0.7 to 2.8 kg during the 1st year of life.

Adipogenesis, i.e., the development of adipose tissue, varies according to sex and age (Wajchenberg, 2000; Karastergiou et al., 2013; Sanchez-Gurmaches and Guertin, 2014). Two sensitive periods for changes in adipose tissue cellularity with peaks of accelerated adipose mass enlargement have been established, namely after birth and from 9 to 13 years. The capacity for cell proliferation and differentiation is highest during the 1st year of life, whereas it is less pronounced in the years before puberty. Childhood-onset obesity is characterized by a combination of fat cell hyperplasia and hypertrophy, whereas in adult-onset obesity a hypertrophic growth predominates. Initially, excess energy storage starts as hypertrophic obesity, resulting from the accumulation of excess lipid in a normal number of unilocular adipose cells (Jo et al., 2009). In this case, adipocytes may be four times their normal size. If the positive energy balance is maintained, a hyperplasic or hypercellular obesity characterized by a greater than normal number of cells is developed. It has been recently shown that adult humans are capable of new adipocyte formation with samples containing a significant proportion of cells with the ability of undergoing differentiation. Multipotent stem cells and adipoblasts, which are found during embryonic development, are still present postnatally.

Hormones, cytokines, growth factors, and nutrients influence the dynamic changes related to adipose tissue mass as well as its pattern of distribution (Fig. 4). The responsiveness of fat cells to neurohumoral signals may vary according to peculiarities in the adipose lineage stage at the moment of exposure. Moreover, the simultaneous presence at specific threshold concentrations of some adipogenic factors may be a necessary requirement to trigger terminal differentiation.

By measuring the relative abundance of ^{14}C in genomic DNA from adipocytes it has been recently clarified that in individuals approximately 10% of adipocytes experience apoptosis, whereas a comparable proportion are renewed, each year. Thus, WAT turns out to be a more dynamic tissue than was previously assumed. These findings are consistent across a wide range of body mass index (BMI), including subjects with early-onset obesity, and following weight loss. The adipocyte number has been shown to be a major determinant of fat mass in the adult; the number of adipocytes in both people with normal weight and obesity appears to be set during childhood. However, it remains possible that the common scenario of gradual but significant weight gain throughout adult life may be underpinned by an initial increase in triacylglycerol loading until an adipocyte size threshold is reached, when additional new adipocytes are recruited from committed precursor cells or mesenchymal stem cells.

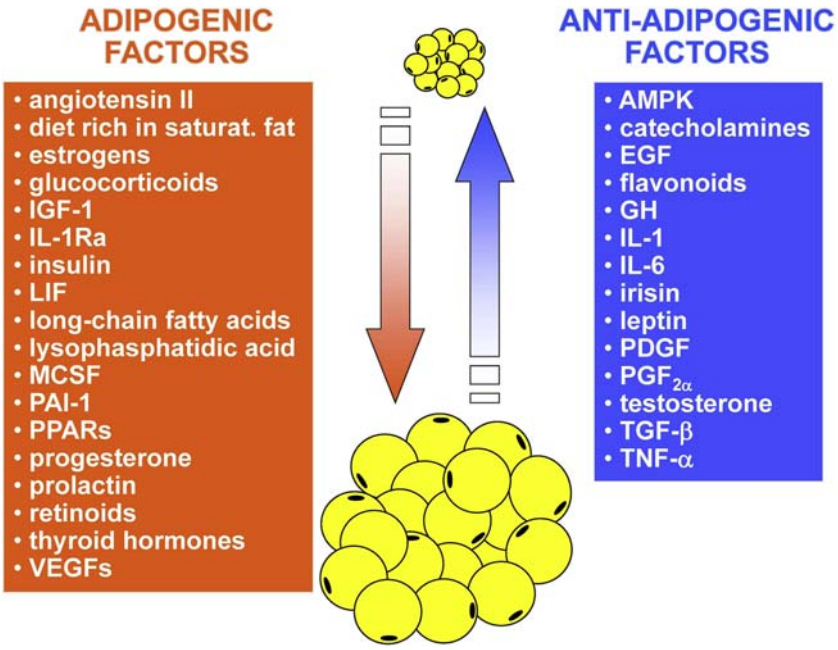


Fig. 4 Factors exerting a direct effect on adipose mass. AMPK, AMP-activated protein kinase; EGF, epidermal growth factor; GH, growth hormone; IGF-1, insulin-like growth factor 1; IL-1, interleukin-1; IL-1Ra, interleukin-1 receptor antagonist; IL-6, interleukin-6; LIF, leukemia inhibitory factor; MCSF, macrophage colony stimulating factor; PAI-1, plasminogen activator inhibitor-1; PDGF, platelet-derived growth factor; PGF2α, prostaglandin F2α; PPARs, peroxisome proliferator-activated receptors; TGF-β, transforming growth factor-β; TNF-α, tumor necrosis factor-α.

Structure

Adipose tissue is a special loose connective tissue dominated by adipocytes. The name of these cells is based on the cell's content of a large lipid droplet with "adipo" being a combining form derived from the Latin *adeps* meaning "pertaining to fat." In adipose tissue, fat cells are individually held in place by delicate reticular fibers clustering in lobular masses bounded by fibrous septa surrounded by a rich capillary network. In adults adipocytes may comprise approximately 90% of adipose mass accounting only for roughly 25% of the total cell population. Thus, adipose tissue itself is composed of not only adipocytes, but also by other cell types, termed as the stroma-vascular fraction, comprising blood cells, endothelial cells, pericytes as well as adipose precursor cells among others (**Fig. 5**), which account for the remaining 75% of the total cell population, representing a wide range of targets for an extensive autocrine–paracrine cross-talk.

Adipocytes, which are typically spherical and vary enormously in size (20–200 μm in diameter, with variable volumes ranging from a few picoliters to approximately 3 nL), are embedded in a connective tissue matrix and are uniquely adapted to store and release energy (Smorlesi et al., 2012). Surplus energy is assimilated by adipocytes and stored as lipid droplets. The stored fat is composed of mainly triacylglycerols (approximately 95% of the total lipid content comprised principally by oleic and palmitic acids) and to a smaller degree diacylglycerols, phospholipids, unesterified fatty acids, and cholesterol (Frühbeck et al., 2014). To accommodate the lipids adipocytes are capable of changing their diameter 20-fold and their volumes by several thousand-fold. However, the increase in size of fat cells is not indefinite. Once a maximum capacity is attained, which in humans averages 1000 pL, the formation of new adipocytes from the precursor pool takes place. The interior of adipocytes appears unstained since the histological techniques of standard tissue preparation dissolve the lipids, leaving a thin rim of eosinophilic cytoplasm that typically loses its round shape during tissue processing, thus contributing to the sponge-like appearance of WAT in routine preparations for light microscopy (**Figs. 6** and **7**). Owing to the fact that approximately 90% of the cell volume is a lipid droplet, the small dark nucleus becomes a flattened semilunar structure pushed against the edge of the cell and the thin cytoplasmic rim is also pushed to the periphery of the adipocytes. Mature white adipose cells contain a single large lipid droplet and are described as unilocular. However, developing white adipocytes are transiently multilocular containing multiple lipid droplets before these finally coalesce into a single large drop (**Fig. 8**). The nucleus is round or oval in young fat cells, but is cup-shaped and peripherally displaced in mature adipocytes. The cytoplasm is stretched to form a thin sheath around the fat globule, although a relatively large volume is concentrated around the nucleus. A thin external lamina called basal lamina surrounds the cell. The smooth cell membrane shows no microvilli but has abundant smooth micropinocytotic invaginations that often fuse to form small vacuoles appearing as rosette-like configurations (**Fig. 9**). Mitochondria are few in number with loosely arranged membranous cristae. The Golgi zone is small and the cytoplasm is filled with free ribosomes, but contains only a limited number of short profiles of the granular endoplasmic reticulum. Occasional lysosomes can also be found. The coalescent lipid droplets contain a mixture of neutral fats, triglycerides, fatty acids, phospholipids, and cholesterol. A thin interface membrane separates the lipid droplet from the cytoplasmic matrix.

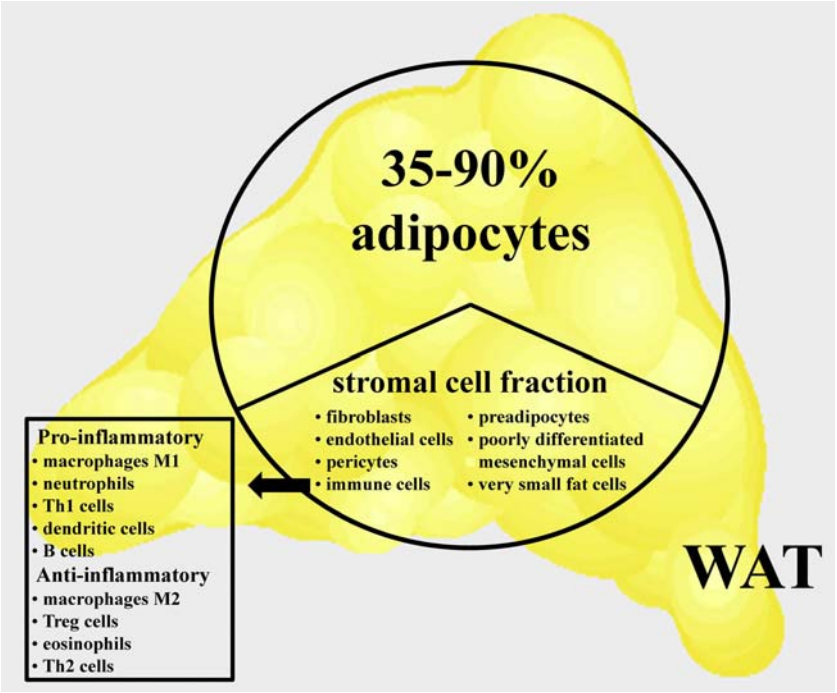


Fig. 5 Schematic representation of cell types present in adipose tissue.

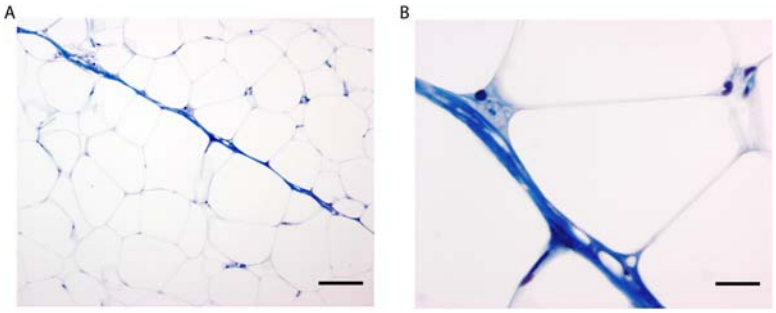


Fig. 6 (A) Human subcutaneous white adipose tissue with a Masson trichrome staining (10×; bar = 100 μm). (B) Same tissue at a higher magnification (40×; bar = 25 μm). Courtesy of Dr. MA Burrell and Dr. M Archanco, University of Navarra.

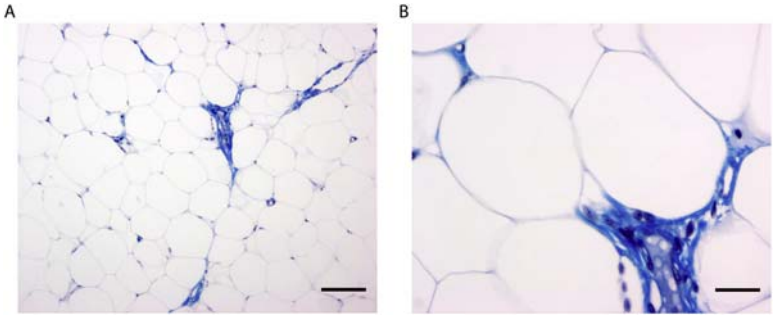


Fig. 7 (A) Human omental white adipose tissue with a Masson trichrome staining (10×; bar = 100 μm). (B) Same tissue at a higher magnification (40×; bar = 25 μm). Courtesy of Dr. MA Burrell and Dr. M Archanco, University of Navarra.

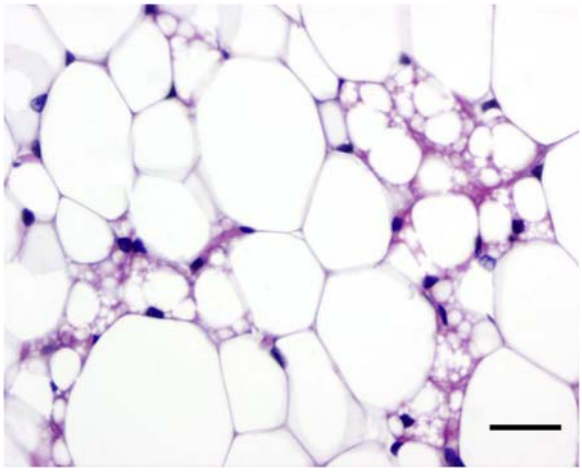


Fig. 8 Paraffin section of rat abdominal white adipose tissue with a hematoxylin and eosin stain showing the simultaneous presence of uni- and multilocular adipocytes (40×; bar = 25 μm). Courtesy of Dr. MA Burrell and Dr. M Archanco, University of Navarra.

Peripheral to this membrane is a system of parallel meridional thin filaments. Because of the size of these cells, relative to the thickness of the section, the nucleus (accounting for only one-fortieth of the cell volume) may not always be present in the section. Unilocular adipocytes usually appear in clumps near blood vessels, which is reasonable because the source and dispersion of material stored in fat cells depends on transportation by the vascular system.

Brown fat is a specialized type of adipose tissue that plays an important role in body temperature regulation (Frühbeck et al., 2009a). It is present in significant amounts in rodents and hibernating animals. In the newborn brown fat is well developed in the neck and interscapular region. Until recently, it was generally accepted that BAT involutes steadily during the first few months, with clearly recognizable depots having essentially disappeared within the 1st years after birth. In normal adults, only occasional brown adipocytes were thought to be scattered through white fat masses. However, recent findings using positron emission tomography (PET) with fluorodeoxyglucose (a marker of metabolic activity) have shown in adults symmetrical areas of increased tracer

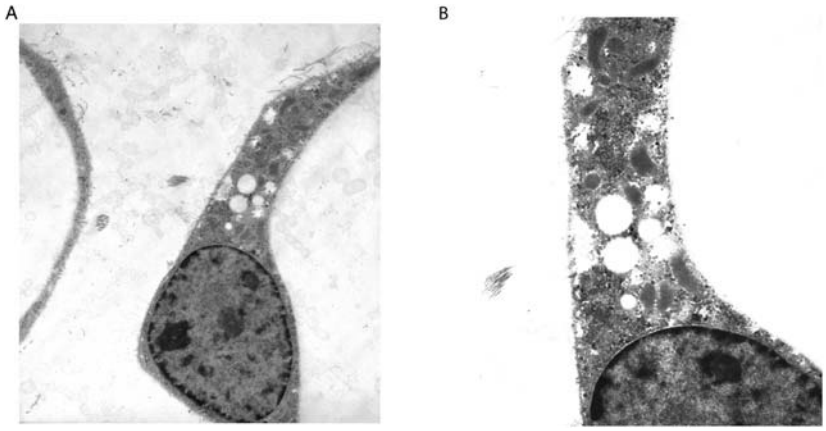


Fig. 9 (A) Transmission electron micrographs with the characteristically displaced nucleus to one side and slighly flattened by the accumulated lipid. The cytoplasm of the fat cell is reduced to a thin rim around the lipid droplet (7725×). (B) The cytoplasm contains several small lipid droplets that have not yet coalesced. A few filamentous mitochondria, occasional cisternae of endoplasmic reticulum and a moderate number of free ribosomes are usually visible (15,000×). Courtesy of Dr. MA Burrell and Dr. M Archanco, University of Navarra.

uptake in the upper parts of the body, broadly corresponding to the distribution of BAT in lower mammals and in human neonates. The main BAT depots in adults are found in the supraclavicular region and neck, with additional activity in the paravertebral, mediastinal, para-aortic, and suprarenal areas. The activity of this BAT can be acutely enhanced by cold exposure and stimulated by the sympathetic nervous system (Frühbeck et al., 2014).

The brown color of BAT is derived from a rich vascular network and abundant mitochondria and lysosomes. The individual multilocular adipocytes are frothy-appearing cells due to the fact that the lipid, which does not coalesce as readily as in white fat cells, is normally stored in multiple small droplets, has been leached out during tissue processing (**Fig. 10**). The spherical nuclei are centrally or eccentrically located within the cell. Compared to the unilocular white adipocytes, the cytoplasm of the multilocular brown fat cell is relatively abundant and strongly stained because of the numerous mitochondria present. The mitochondria are involved in the oxidation of the stored lipid, but because they exhibit a reduced potential to carry out oxidative phosphorylation, the energy produced is released in the form of heat due to the uncoupling activity of UCP and not captured in adenosine triphosphate (ATP). Therefore, brown adipose tissue is extremely well vascularized so that the blood is warmed when it passes through the active tissue.

Distribution

WAT may represent the largest endocrine tissue of the whole organism especially in people with overweight and obesity (Ahima and Flier, 2000). The anatomical distribution of individual fat pads dispersed throughout the whole body and not connected to each other collides with a classic organ-specific localization. WAT exhibits clear, regional differences in its sites of predilection (**Table 1**). The hypodermal region invariably contains fat, except in a few places such as the eyelids and the scrotum. Adipocytes also accumulate around organs like kidneys and adrenals, in the coronary sulcus of the heart, in bone marrow, mesentery, and omentum. Unilocular fat is widely distributed in the subcutaneous tissue of humans but exhibits quantitative regional differences that are influenced by age and sex. In infants and young children there is a continuous subcutaneous fat layer—the panniculus adiposus, over the whole body. This layer thins out in some areas in adults but persists and grows thicker in certain other regions. The sites

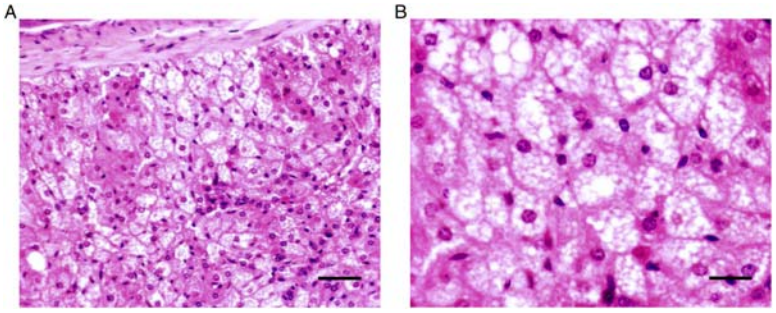


Fig. 10 (A) Paraffin section of rat brown adipose tissue with a hematoxylin and eosin stain (20×; bar = 50 μm). (B) Same tissue at a higher magnification (40×; bar = 25 μm). Courtesy of Dr. MA Burrell and Dr. M Archanco, University of Navarra.

Table 1 Distribution of human adipose tissue depots.

White
Subcutaneous (approx. 80%; deep + superficial layers)
Truncal
Cervical
Dorsal
Lumbar
Abdominal
Gluteo-femoral
Mammary
Visceral (approx. 20%; thoracic-abdominal-pelvic)
Intrathoracic (extra-intrapericardial)
Intraabdominopelvic
Intraperitoneal
Omental (greater and lesser omentum)
Mesenteric (epiplon, small intest., colon, rectum)
Umbilical
Extraperitoneal
Peripancreatic (infiltrated with brown adipocytes)
Perirenal (infiltrated with brown adipocytes)
Intrapelvic
Gonadal (parametrial, retrouterine, retropubic)
Urogenital (paravesical, para-retrorectal)
Intraparenchimatous (physiologically or pathologically)
Inter-intramuscular and perimuscular (inside the muscle fascia)
Perivascular
Paraosseal (interface between bone and muscle)
Ectopic (steatosis, intramyocardial, lipodystrophy, etc.)
Brown/beige or brite (contraction of brown and white)
Cervical, supraclavicular, axillary, mediastinal, paraspinal and abdominal
Interscapular and perirenal only in newborns
Pink
Mammary gland (during pregnancy and lactation the subcutaneous depots of the mammary glands experiment a transformation involving mainly the parenchyma through the development of milk-secreting lobulo-alveolar glandular structures)

differ in their distribution among sexes being responsible for the characteristic body form of males and females, termed android and gynecoid fat distribution. In males, the main regions include the nape of the neck, the subcutaneous area over the deltoid and triceps muscles, and the lumbosacral region. In females, subcutaneous fat is most abundant in the buttocks, epitrochanteric region, anterior and lateral aspects of the thighs as well as the breasts. Additionally, extensive fat depots are found in the omentum, mesenteries, and the retroperitoneal area of both sexes. In well-nourished, sedentary individuals, the fat distribution persists and becomes more obvious with advancing age with males tending to deposit more fat in the visceral compartment. Depot-specific differences may be related not only to the metabolism of fat cells but also to their capacity to form new adipocytes. Additionally, regional differences may result from variations in hormone receptor distribution as well as from specific local environmental characteristics as a consequence of differences in innervation and vascularization.

Regional distribution of body fat is known to be an important indicator for metabolic and cardiovascular alterations in some individuals (Hamdy et al., 2006; Goossens, 2017). The observation that the topographic distribution of adipose tissue is relevant to understanding the relation of obesity to disturbances in glucose and lipid metabolism was formulated before the 1950s. Since then numerous prospective studies have revealed that android or male-type obesity correlates more often with an elevated mortality and risk for the development of type 2 diabetes mellitus, dyslipidemia, hypertension, and atherosclerosis than gynoid or female-type obesity. Obesity has been reported to cause or exacerbate a large number of health problems with a known impact on both life expectancy and quality of life. In this respect, the association of increased adiposity is accompanied by important pathophysiological alterations, which lead to the development of a wide range of co-morbidities (Fig. 11).

Function

Although many cell types contain small reserves of carbohydrate and lipid, the adipose tissue is the body's most capacious energy reservoir. Because of the high energy content per unit weight of fat as well as due to its hydrophobicity, the storage of energy in the form of triglycerides is a highly efficient biochemical phenomenon (1 g of adipose tissue contains approximately 800 mg triacylglycerol and only approximately 100 mg of water). It represents quantitatively the most variable component of the organism,

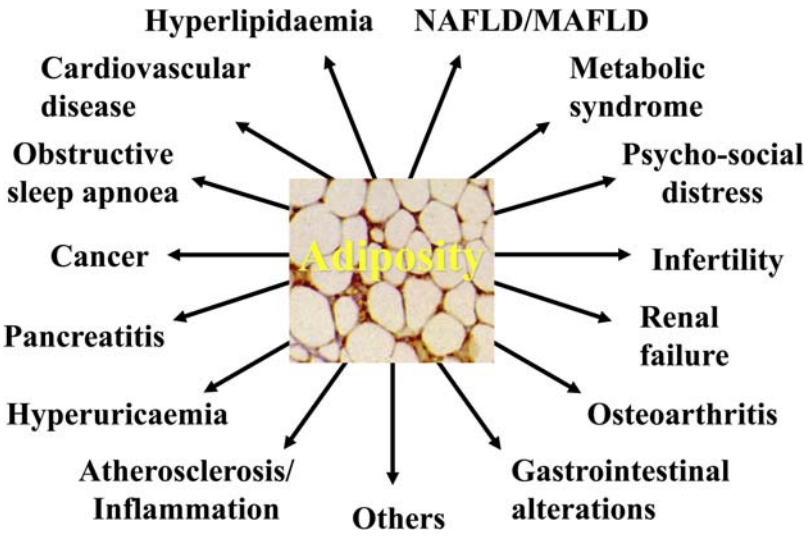


Fig. 11 Main comorbidities associated with increased adiposity. NAFLD, non-alcoholic fatty liver disease; MAFLD, metabolic-associated fatty liver disease.

varying from a few percent of body weight in elite athletes to more than half of the total body weight in people with severe obesity. The normal range is approximately 10–20% body fat for males and approximately 20–30% for females, accounting approximately for an energy reserve of 25–50 days in men and 40–60 days in women. During pregnancy most species accrue additional reserves and functionality of adipose tissue to help support the development of the fetus and to further facilitate the lactation period, which is identified as "pink fat" (**Table 1**).

Energy balance regulation is an extremely complex process composed of multiple interacting homeostatic and behavioral pathways aimed at maintaining constant energy stores. It is now evident that body weight control is achieved through highly orchestrated interactions between nutrient selection, organoleptic influences, and neuro-endocrine responses to diet as well as being influenced by genetic and environmental factors. The concept that circulating signals generated in proportion to body fat stores influence appetite and energy expenditure in a coordinated manner to regulate body weight was proposed almost 70 years ago. According to this model, changes in energy balance sufficient to alter body fat stores were signaled via one or more circulating factors acting in the brain to elicit compensatory changes in order to match energy intake to energy expenditure. This was formulated as the "lipostatic theory" assuming that as adipose tissue mass enlarges, a factor that acts as a sensing hormone or "lipostat" in a negative feedback control from adipose tissue to hypothalamic receptors informs the brain about the abundance of body fat, thereby allowing feeding behavior, metabolism, and endocrine physiology to be coupled to the nutritional state of the organism. The existing body of evidence gathered in the last decades through targeted expression or knockout of specific genes involved in different steps of the pathways controlling food intake, body weight, adiposity, or fat distribution has clearly contributed to unraveling the underlying mechanisms of energy homeostasis. The findings have fostered the notion of a far more complex system than initially thought, involving the integration of a plethora of factors.

The identification of adipose tissue as a multifunctional organ as opposed to a passive organ for the storage of excess energy in the form of fat has been brought about by the emerging body of evidence gathered during the last decades. This pleiotropic nature is based on the ability of fat cells to secrete a large number of hormones, growth factors, enzymes, cytokines, complement factors, and matrix proteins, collectively termed as adipokines or adipocytokines (**Table 2**, **Fig. 12**), at the same time as expressing receptors for most of these factors (**Table 3**), which warrants an extensive cross-talk at a local and systemic level in response to specific external stimuli or metabolic changes (Rodríguez et al., 2015). The vast majority of adipocyte-derived factors have been shown to be dysregulated in alterations accompanied by changes in adipose tissue mass such as overfeeding and lipodystrophy, thus providing evidence for their involvement in the etiopathology and co-morbidities associated with obesity and cachexia.

WAT is actively involved in cell function regulation through a complex network of endocrine, paracrine, and autocrine signals, which influence the response of many tissues, including the hypothalamus, pancreas, liver, skeletal muscle, kidneys, endothelium, and immune system. Adipose tissue serves the functions of being a store for energy reserve, insulation against heat loss through the skin, and a protective padding of certain organs. A rapid turnover of stored fat can take place, and with only a few exceptions (orbit, major joints as well as palm, and foot sole), the adipose tissue can be used up almost completely during starvation. Adipocytes are uniquely equipped to participate in the regulation of other functions such as reproduction, immune response, blood pressure control, coagulation, fibrinolysis, and angiogenesis. The advent of microarray technology has dramatically changed the study of the pattern of gene expression by enabling the simultaneous analysis of thousands of genes in a single experiment. Interestingly, the high number and ample spectrum of genes found to be expressed in WAT together with the changes observed in samples from obese patients substantiates the view of an extraordinarily active and plastic tissue (Sun et al., 2011). The complex and

Table 2 Relevant factors secreted by adipose tissue to the bloodstream.

Molecule	*Function/effect*
Adiponectin/ACRP30/AdipoQ	Plays a protective role in the pathogenesis of type 2 diabetes and cardiovascular diseases
Adipsin	Possible link between the complement pathway and adipose tissue metabolism
Angiotensinogen	Precursor of angiotensin II; regulator of blood pressure and electrolyte homeostasis
ASP	Influences the rate of triacylglycerol synthesis in adipose tissue
Chemerin	Regulates adipocyte differentiation and glucose uptake. It is potentially involved in the inflammatory response
FFA	Oxidized in tissues to produce local energy. Serve as a substrate for triglyceride and structural molecules synthesis. Involved in the development of insulin resistance
Glycerol	Structural component of the major classes of biological lipids and gluconeogenic precursor
IGF-I	Stimulates proliferation of a wide variety of cells and mediates many of the effects of growth hormone
IL-6	Implicated in host defense, glucose and lipid metabolism and regulation of body weight
Leptin	Signals to the brain about body fat stores. Regulation of appetite and energy expenditure. Wide variety of physiological functions
NO	Important regulator of vascular tone. Pleiotropic involvement in pathophysiological conditions
Omentin	Enhances insulin-stimulated glucose uptake
PAI-1	Potent inhibitor of the fibrinolytic system
PGI_2 & $PGF_{2\alpha}$	Implicated in regulatory functions such as inflammation and blood clotting, ovulation, menstruation and acid secretion
Resistin	Putative role in insulin resistance. Participates in inflammation
SAA	Acute phase protein involved in inflammation and HDL metabolism
TNF-α	Interferes with insulin receptor signaling and is a possible cause of insulin resistance in obesity
Vaspin	Exhibits insulin-sensitizing effects
VEGFs	Stimulation of angiogenesis
Visfatin/PBEF/NAMPT	Catalyzes the biosynthesis of nicotinamide adenine dinucleotide. Regulates vascular smooth muscle and immune cells function. Potentially involved in the regulation of insulin sensitivity

complementary nature of the expression profile observed in obese adipose tissue reflects a pleiad of adaptive changes affecting crucial physiological functions that may need to be further explored through genomic and proteomic approaches (Scherer, 2019).

The endocrine activity of WAT was postulated almost 40 years ago alluding to the tissue's ability for steroid hormone interconversion. In recent years, especially since the discovery of leptin, the list of adipocyte-derived factors has been increasing at a phenomenal pace (Galic et al., 2010). Another way of addressing the production of adipose-derived factors is by focusing on the functions in which they are implicated (**Fig. 12**). One of the best known aspects of WAT physiology relates to the synthesis of products involved in lipid metabolism such as perilipin, fatty acid-binding protein 4 (ALBP, FABP4 or aP2), CETP, and retinol binding protein (RBP). Adipose tissue has been also identified as a source of production of factors with immunological properties participating in immunity and stress responses as is the case of ASP and metallothionein. More recently, the pivotal role of adipocyte-derived factors implicated in cardiovascular function control such as angiotensinogen, adiponectin, peroxisome proliferator-activated receptor γ angiopoietin related protein/fasting-induced adipose factor (PGAR/FIAF), and C-reactive protein (CRP) has been established. A further subsection of proteins produced by adipose tissue concerns other factors with an autocrine–paracrine function like PPARγ, IGF-1, monobutyrin, and the UCPs.

BAT is specialized for heat production; its lipid stores turn over rapidly, and the liberated fatty acids are oxidized by the brown adipocyte's mitochondria in a process that generates heat directly. In neonatal mammals, hibernators, and rodents, a crucial function of BAT is the maintenance of body temperature through cold-induced thermogenesis. In addition, BAT thermogenesis is activated during overeating—an important aspect of diet-induced thermogenesis. In humans, as is the case in other larger mammals, the functional capacity of brown adipose tissue decreases because of the relatively higher ratio between heat production from basal metabolism and the smaller surface area encountered in adults. In addition, clothing and indoor life have reduced the need for adaptive nonshivering thermogenesis. However, it has been recently shown that human WAT can be infiltrated with brown adipocytes expressing UCP-1. The prevalence of active BAT in normal adults can be only estimated indirectly, but is thought to be present in approximately 10% of the general population. BAT, therefore, has the potential to play a role in normal energy balance and could become a pharmacological target for new drugs to treat obesity (Frühbeck et al., 2009a).

Regulation of metabolism

The control of fat storage and mobilization has been marked by the identification of a number of regulatory mechanisms in the last decades. Isotopic tracer studies have clearly shown that lipids are continuously being mobilized and renewed even in individuals in energy balance. Fatty acid esterification and triglyceride hydrolysis take place continuously. The half-life of depot lipids in rodents is

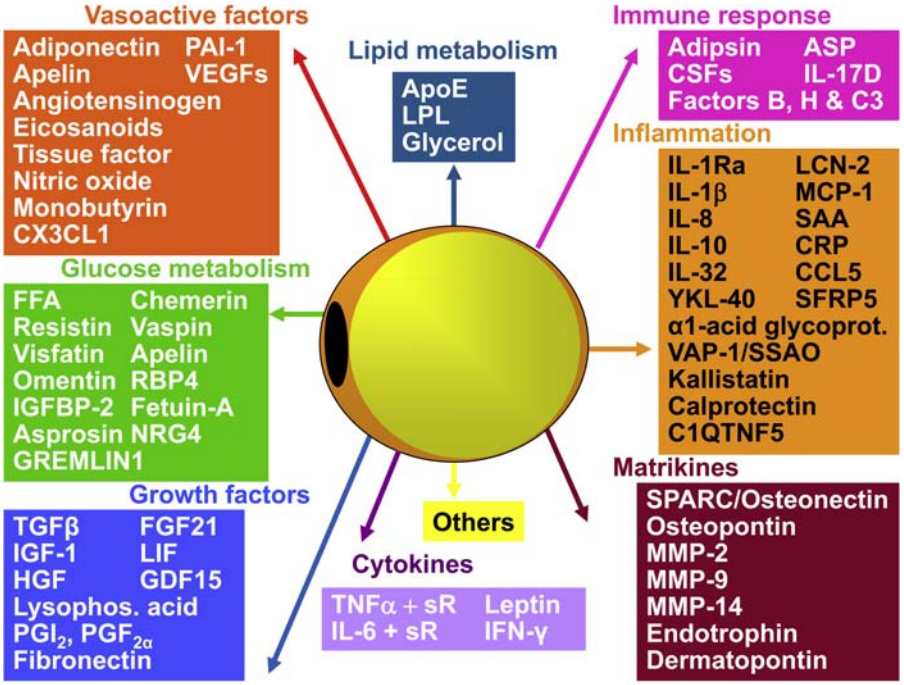


Fig. 12 Factors secreted by white adipose tissue, which underlie the multifunctional nature of this endocrine organ. Although owing to their pleiotropic effects some of the elements might play more than one physiological role, they have been included only under one function for clarity of the figure. ApoE, apolipoprotein E; ASP, acylation-stimulating protein; C1QTNF5, C1q and TNF related 5; CCL5, C-C motif chemokine ligand 5; CHI3L1 (YKL-40), chitinase-3-like protein 1; CSF, colony-stimulating factor; CX3CL, C-X3-C motif chemokine ligand 1; FGF21, fibroblast growth factor 21; GDF15, growth differentiation factor 15; GREMLIN1, GREM1, DAN family BMP antagonist, HGF, hepatocyte growth factor; IGF-1, insulin-like growth factor-1; IGFBP-2, insulin-like growth factor binding protein-2; IL-1, interleukin-1; IL-1β, interleukin-1 β; IL-6, interleukin-6; IL-8, interleukin-8; IL-10, interleukin-10; IL-17, interleukin-17; IL-1Ra, interleukin-1 receptor antagonist; IFN-γ, interferon-γ; LCN-2, lipocalin-2; LIF, leukemia inhibitory factor; LPL, lipoprotein lipase; MCP-1, monocyte chemoattractant protein-1; MMP-2, matrix metalloproteinase-2; MMP-9, matrix metalloproteinase-9; NRG4, Neuregulin 4; PAI-1, plasminogen activator inhibitor-1; PGF2α, prostaglandin F2α; PGI2, prostacyclin; RBP4, retinol binding protein-4; SAA, serum amyloid A; SFRP5, secreted frizzled related protein 5; SPARC, secreted protein acidic and cysteine rich; TGF-β, transforming growth factor-β; TNF-α, tumor necrosis factor-α; VAP-1/SSAO, vascular adhesion protein-1/semicarbazide-sensitive amine oxidase; VEGF, vascular endothelial growth factor.

approximately 8 days, meaning that almost 10% of the fatty acid stored in adipose tissue is replaced daily by new fatty acids. The balance between lipid loss and accretion determines the net outcome on energy homeostasis.

The synthesis of triglycerides, also termed lipogenesis, requires the supply of fatty acids and glycerol. The main sources of fatty acids are the liver and the small intestine. Fatty acids are esterified with glycerol phosphate in the liver to produce triglycerides. Since triglycerides are bulky polar molecules that do not cross cell membranes well, they must be hydrolyzed to fatty acids and glycerol before entering fat cells. Serum very low-density lipoproteins (VLDLs) are the major form in which triacylglycerols are carried from the liver to WAT. Short-chain fatty acids (16 carbons or less) can be absorbed from the gastrointestinal tract and carried in chylomicra directly to the adipocyte. Inside fat cells glycerol is mainly synthesized from glucose. In WAT fatty acids can be synthesized from several precursors, such as glucose, lactate, and certain amino acids, with glucose being quantitatively the most important in humans. In the case of glucose, GLUT4, the principal glucose transporter of adipocytes, controls the entry of the substrate into the adipocyte. Insulin is known to stimulate glucose transport by promoting GLUT4 recruitment as well as increasing its activity. Inside the adipocyte, glucose is initially phosphorylated and then metabolized both in the cytosol and in the mitochondria, to produce cytosolic acetyl-CoA with the flux being influenced by phosphofructokinase and pyruvate dehydrogenase. Glycerol does not readily enter the adipocyte, but the membrane-permeable fatty acids do. Once inside the fat cells, fatty acids are re-esterified with glycerol phosphate to yield triglycerides. Lipogenesis is favored by insulin, which activates pyruvate kinase, pyruvate dehydrogenase, acetyl-CoA carboxylase, and glycerol phosphate acyltransferase. When excess nutrients are available insulin decreases acetyl-CoA entry into the tricarboxylic acid cycle while directing it toward fat synthesis. This insulin effect is antagonized by growth hormone. The gut hormones glucagon-like peptide 1 and gastric inhibitory peptide also increase fatty acid synthesis, whereas glucagon and catecholamines inactivate acetyl-CoA carboxylase, thus decreasing the rate of fatty acid synthesis.

The release of glycerol and free fatty acids by lipolysis plays a critical role in the ability of the organism to provide energy from triglyceride stores (Frühbeck et al., 2014). In this sense, the processes of lipolysis and lipogenesis are crucial for the attainment of body weight control. For this purpose adipocytes are equipped with a well developed enzymatic machinery, together with a number of non-secreted proteins and binding factors directly involved in the regulation of lipid metabolism. The hydrolysis of triglycerides

Table 3 Main receptors expressed by adipose tissue.

Receptor	*Main effect of receptor activation on adipocyte metabolism*
Hormone-cytokine receptors	
Adenosine	Inhibition of lipolysis
Adiponectin (AdipoR1 and AdipoR2)	Regulation of insulin sensitivity and fatty acid oxidation
Angiotensin II	Increase of lipogenesis. Stimulation of prostacyclin production by mature fat cells. Interaction with insulin in regulation of adipocyte metabolism
Chemerin	Regulation of adipocyte differentiation and glucose uptake. Potential role in inflammatory response
GH	Induction of leptin and IGF-I expression. Stimulation of lipolysis
Ghrelin	Stimulation of adipogenesis and lipogenesis. Induction of glucose uptake
IGF-I and -II	Inhibition of lipolysis. Stimulation of glucose transport and oxidation
IL-6	LPL activity inhibition. Induction of lipolysis
Insulin	Inhibition of lipolysis and stimulation of lipogenesis. Induction of glucose uptake and oxidation. Stimulation of leptin expression
Leptin (OB-R)	Stimulation of lipolysis. Autocrine regulation of leptin expression
NPY-Y1 and Y5	Inhibition of lipolysis. Induction of leptin expression
PDGF (PDGFRα and PDGFRβ)	Control the differentiation balance of adipocytes and stromal cells. Involved in the regulation of fibrosis
Prostaglandin	Strong antilipolytic effects (PGE_2). Modulation of preadipocyte differentiation ($PGF_{2\alpha}$ and PGI_2)
TGF-β	Potent inhibition of adipocyte differentiation
TNF-α	Stimulation of lipolysis. Regulation of leptin secretion. Potent inhibition of adipocyte differentiation. Involvement in development of insulin resistance
VEGF	Stimulation of angiogenesis
Catecholamine-nervous system receptors	
Endocannabinoids CB_1	Stimulation of adiponectin expression induction of glucose uptake and GLUT4 translocation. Inhibition of lipogenesis
Muscarinic	Inhibition of lipolysis
Nicotinic	Stimulation of lipolysis
α_1-AR	Induction of inositol phosphate production and PKC activation
α_2-AR	Inhibition of lipolysis. Regulation of preadipocyte growth
β_1-, β_2- and β_3-AR	Stimulation of lipolysis. Induction of thermogenesis. Reduction of leptin mRNA levels
Nuclear receptors	
Androgen	Control of adipose tissue development (antiadipogenic signals). Modulation of leptin expression
Estrogen	Control of adipose tissue development (proadipogenic signals). Modulation of leptin expression
Glucocorticoids	Stimulation of adipocyte differentiation
PPARδ	Regulation of fat metabolism. Plays a central role in fatty acid-controlled differentiation of preadipose cells
PPARγ	Induction of adipocyte differentiation and insulin sensitivity
RAR/RXR	Regulation of adipocyte differentiation
T_3	Stimulation of lipolysis. Regulation of leptin secretion. Induction of adipocyte differentiation. Regulation of insulin effects
Lipoprotein receptors	
HDL	Clearance and metabolism of HDL
LDL	Stimulation of cholesterol uptake
VLDL	Binding and internalization of VLDL particles. Involvement in lipid accumulation

ACRP30, adipocyte complement-related protein of 30 kDa; ASP, acylation-stimulating protein; FFA, free fatty acids; GH, growth hormone; HDL, high density lipoprotein.; IGF, insulin-like growth factor; IL-6, interleukin 6; LDL, low density lipoprotein; LPL, lipoprotein lipase; NO, nitric oxide; NPY-Y1 and -Y5, neuropeptide receptors Y-1 and -5; OB-R, leptin receptor; PAI-1, plasminogen activator inhibitor -1; PDGFR, platelet-derived growth factor receptor; PGE_2, prostaglandin E_2; $PGF_{2\alpha}$, prostaglandin $F_{2\alpha}$; PGI_2, prostacyclin; PPAR, peroxisome proliferator-activated receptor; RAR, retinoic acid receptor; RXR, retinoid x receptor; T3, triiodothyronine; TGF-β, transforming growth factor-β; TNF-α, tumor necrosis factor-α; VEGF, vascular endothelial growth factor; VLDL, very low density lipoprotein; α_1- & α_2-AR, α_1-and α_2-adrenergic receptors; β_1-, β_2-and β_3-AR, β_1-, β_2-and β_3 adrenergic receptors.

from circulating VLDL and chylomicrons is catalyzed by lipoprotein lipase (LPL). This rate-limiting step plays an important role in directing fat partitioning. Although LPL controls fatty acid entry into adipocytes, fat mass has been shown to be preserved by endogenous synthesis. From observations made in patients with total LPL deficiency it can also be concluded that fat deposition can take place in the absence of LPL. A further key enzyme catalyzing a rate-limiting step of lipolysis is hormone-sensitive lipase (HSL), which cleaves triacylglycerol to yield glycerol and fatty acids within adipocytes. Some fatty acids are re-esterified, so that the fatty acid–glycerol ratio leaving the cell is usually less than the theoretical 3:1. Increased concentrations of cAMP activate HSL as well

as promote its movement from the cytosol to the lipid droplet surface. Catecholamines and glucagon are known inducers of the lipolytic activity, whereas the stimulation of lipolysis is attenuated by adenosine and prostaglandin E2. Interestingly, HSL deficiency leads to male sterility and adipocyte hypertrophy, but not to obesity, with an unaltered basal lipolytic activity suggesting that other lipases may also play a relevant role in fat mobilization.

The lipid droplets contained in adipocytes are coated by structural proteins, such as perilipin, that stabilize the single fat drops and prevent triglyceride hydrolysis in the basal state. The phosphorylation of perilipin following adrenergic stimulation or other hormonal inputs induces a structural change of the lipid droplet that allows the hydrolysis of triglycerides. After hormonal stimulation, HSL and perilipin are phosphorylated and HSL translocates to the lipid droplet. ALBP, also termed aP2, then binds to the N-terminal region of HSL, preventing fatty acid inhibition of the enzyme's hydrolytic activity.

Adipose tissue has been shown to contain 0.6–1.6 mg of cholesterol per gram wet weight. When expressed per unit of protein or organ mass, fat tissue contains more cholesterol than most other organs or membranes. The cholesterol content of adipose tissue increases with age and weight. The specific activity of adipose cholesterol exceeds that of plasma three-to five-fold. The half-life disappearance time of adipose tissue cholesterol is approximately 1 month, which is consistent with its function as a slowly turning over storage pool. The function of CETP is to promote the exchange of cholesterol esters and triglycerides between plasma lipoproteins. Fasting, high cholesterol diets as well as insulin stimulate CETP synthesis and secretion in WAT. In plasma CETP participates in the modulation of reverse cholesterol transport by facilitating the transfer of cholesterol esters from high-density lipoproteins (HDL) to triglyceride-rich apoB containing lipoproteins. VLDLs, in particular, are converted to low-density lipoproteins (LDLs), which are subjected to hepatic clearance by the apoB/E receptor system. Adipose tissue probably represents one of the major sources of CETP in humans. In obesity the activity and protein mass of circulating CETP is increased showing a negative correlation with HDL concentrations at the same time as a positive correlation with fasting glycemia and insulinemia suggesting a potential link with insulin resistance.

Synthesis and secretion of RBP by adipocytes is induced by retinoic acid and shows that WAT plays an important role in retinoid storage and metabolism. In fact, RBP mRNA is one of the most abundant transcripts present in both rodent and human adipose tissue. Hepatic and renal tissues have been regarded as the main sites of RBP production, whereas the quantitative and physiological significance of the WAT contribution remains to be fully established.

Conclusion

The processes participating in controlling energy balance as well as the intermediary lipid and carbohydrate metabolism are intricately coupled by neurohumoral mediators. The coordination of the implicated molecular and biochemical pathways underlies, at least in part, the large number of intracellular and secreted proteins produced by WAT with autocrine, paracrine, and endocrine effects. The realization that WAT secretes a plethora of pleiotropic adipokines at the same time as expressing receptors for a huge range of compounds has led to the development of new insights into the functions of adipose tissue at both the basic and clinical level. At this early juncture in the course of adipose tissue research, much has been discovered. However, much more remains to be learned about its physiology and clinical relevance. Given the adipocyte's versatile and ever-expanding list of secretory proteins, additional, and unexpected consequences are sure to emerge. The growth, cellular composition and gene expression pattern of adipose tissue is under the regulation of a large selection of central mechanisms and local effectors. The exact nature and control of this complex cross-talk has not been fully elucidated representing an exciting research topic.

References

Ahima, R.S., Flier, J.S., 2000. Adipose tissue as an endocrine organ. Trends Endocrinol. Metab. 11, 327–332.

Bouloumie, A., Lolmede, K., Sengenes, C., Galitzky, J., Lafontan, M., 2002. Angiogenesis in adipose tissue. Ann. Endocrinol. 63, 91–95.

Crewe, C., An, Y.A., Scherer, P.E., 2017. The ominous triad of adipose tissue dysfunction: inflammation, fibrosis, and impaired angiogenesis. J. Clin. Invest. 127, 74–82.

Frühbeck, G., Gómez-Ambrosi, J., Muruzábal, F.J., Burrell, M.A., 2001. The adipocyte: a model for integration of endocrine and metabolic signaling in energy metabolism regulation. Am. J. Physiol. Endocrinol. Metab. 280, E827–E847.

Frühbeck, G., Becerril, S., Sáinz, N., Garrastachu, P., García-Velloso, M.J., 2009a. BAT: a new target for human obesity? Trends Pharmacol. Sci. 30, 387–396.

Frühbeck, G., Sesma, P., Burrell, M.A., 2009b. PRDM16: the interconvertible adipo-myocyte switch. Trends Cell Biol. 19, 141–146.

Frühbeck, G., Méndez-Giménez, L., Fernández-Formoso, J.A., Fernández, S., Rodríguez, A., 2014. Regulation of adipocyte lipolysis. Nutr. Res. Rev. 27, 63–93.

Galic, S., Oakhill, J.S., Steinberg, G.R., 2010. Adipose tissue as an endocrine organ. Mol. Cell. Endocrinol. 316, 129–139.

Ghaben, A.L., Scherer, P.E., 2019. Adipogenesis and metabolic health. Nat. Rev. Mol. Cell Biol. 20, 242–258.

Goossens, G.H., 2017. The metabolic phenotype in obesity: fat mass, body fat distribution, and adipose tissue function. Obes. Facts 10, 207–215.

Hamdy, O., Porramatikul, S., AL-Ozairi, E., 2006. Metabolic obesity: the paradox between visceral and subcutaneous fat. Curr. Diabetes Rev. 2, 367–373.

Jo, J., Gavrilova, O., Pack, P., Jou, W., Mullen, S., Sumner, A.E., Cushman, S.W., Periwal, V., 2009. Hypertrophy and/or hyperplasia: dynamics of adipose tissue growth. PLoS Comput. Biol. 5, e324.

Karastergiou, K., Fried, S.K., Xie, H., Lee, M.J., Divoux, A., Rosencrantz, M.A., Chang, R.J., Smith, S.R., 2013. Distinct developmental signatures of human abdominal and gluteal subcutaneous adipose tissue depots. J. Clin. Endocrinol. Metab. 98, 362–371.

Rodríguez, A., Ezquerro, S., Méndez-Giménez, L., Becerril, S., Frühbeck, G., 2015. Revisiting the adipocyte: a model for integration of cytokine signaling in the regulation of energy metabolism. Am. J. Physiol. Endocrinol. Metab. 309, E691–E714.

Sanchez-Gurmaches, J., Guertin, D.A., 2014. Adipocytes arise from multiple lineages that are heterogeneously and dynamically distributed. Nat. Commun. 5, 4099.

Scherer, P.E., 2019. The many secret lives of adipocytes: implications for diabetes. Diabetologia 62, 223–232.

Smorlesi, A., Frontini, A., Giordano, A., Cinti, S., 2012. The adipose organ: white-brown adipocyte plasticity and metabolic inflammation. Obes. Rev. 13 (Suppl. 2), 83–96.
Sun, K., Kusminski, C.M., Scherer, P.E., 2011. Adipose tissue remodeling and obesity. J. Clin. Invest. 121, 2094–2101.
Tang, W., Zeve, D., Suh, J.M., Bosnakovski, D., Kyba, M., Hammer, R.E., Tallquist, M.D., Graff, J.M., 2008. White fat progenitor cells reside in the adipose vasculature. Science 322, 583–586.
Van Baak, M.A., Mariman, E.C.M., 2019. Mechanisms of weight regain after weight loss—the role of adipose tissue. Nat. Rev. Endocrinol. 15, 274–287.
Wajchenberg, B.L., 2000. Subcutaneous and visceral adipose tissue: their relation to the metabolic syndrome. Endocr. Rev. 21, 697–738.
Yang, S., Eto, H., Kato, H., Doi, K., Kuno, S., Kinoshita, K., Tsai, C.H., Chou, W.T., Ma, H., Yoshimura, K., 2013. Comparative characterization of stromal vascular cells derived from three types of vascular wall and adipose tissue. Tissue Eng. Part A 19.

Further reading

Cohen, P., Kajimura, S., 2021. The cellular and functional complexity of thermogenic fat. Nat. Rev. Mol. Cell Biol. 22, 393–409.
Hammarstedt, A., Gogg, S., Hedjazifar, S., Nerstedt, A., Smith, U., 2018. Impaired adipogenesis and dysfunctional adipose tissue in human hypertrophic obesity. Physiol. Rev. 98, 1911–1941.
Hildreth, A.D., Ma, F., Wong, Y.Y., Sun, R., Pellegrini, M., O'Sullivan, T.E., 2021. Single-cell sequencing of human white adipose tissue identifies new cell states in health and obesity. Nat. Immunol. 22, 639–653.
Kusminski, C.M., Bickel, P.E., Scherer, P.E., 2016. Targeting adipose tissue in the treatment of obesity-associated diabetes. Nat. Rev. Drug Discov. 15, 639–660.
Marcelin, G., Silveira, A.L.M., Martins, L.B., Ferreira, A.V., Clement, K., 2019. Deciphering the cellular interplays underlying obesity-induced adipose tissue fibrosis. J. Clin. Invest. 129, 4032–4040.
Morigny, P., Boucher, J., Arner, P., Langin, D., 2021. Lipid and glucose metabolism in white adipocytes: pathways, dysfunction and therapeutics. Nat. Rev. Endocrinol. 17, 276–295.
Rosen, E.D., Spiegelman, B.M., 2014. What we talk about when we talk about fat. Cell 156, 20–44.
Sun, K., Tordjman, J., Clement, K., Scherer, P.E., 2013. Fibrosis and adipose tissue dysfunction. Cell Metab. 18, 470–477.
Sun, W., Modica, S., Dong, H., Wolfrum, C., 2021. Plasticity and heterogeneity of thermogenic adipose tissue. Nat. Metab. 3, 751–761.
Unamuno, X., Gómez-Ambrosi, J., Rodríguez, A., Becerril, S., Frühbeck, G., Catalán, V., 2018. Adipokine dysregulation and adipose tissue inflammation in human obesity. Eur. J. Clin. Invest. 48, e12997.
Vijay, J., Gauthier, M.F., Biswell, R.L., Louiselle, D.A., Johnston, J.J., Cheung, W.A., et al., 2020. Single-cell analysis of human adipose tissue identifies depot and disease specific cell types. Nat. Metab. 2, 97–109.
Vishvanath, L., Gupta, R.K., 2019. Contribution of adipogenesis to healthy adipose tissue expansion in obesity. J. Clin. Invest. 129, 4022–4031.

Arthritis

F. Cuesta Triana, Hospital Clinico San Carlos, Universidad Complutense, Madrid, Spain

© 2023 Elsevier Ltd. All rights reserved.
This is an update of L.A. Coleman, R. Roubenoff, Arthritis, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 116–121, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00018-0.

Introduction 109
Definitions and etiology 109
Prevalence 110
Clinical features 110
Role of diet in the management of inflammatory arthritis 111
Nutritional assessment in RA 111
Diet and RA activity 111
Supplementation with dietary fatty acids 111
Vitamin, mineral supplementation and probiotics 113
Fasting and vegetarian diets 113
Conclusion 114
Role of diet in the management and prevention of degenerative arthritis 114
Conclusion 116
References 116
Further reading 116
Relevant websites 117

Key points

- To describe the most frequent forms of arthritis and the physiopathology.
- To understand the subjacent proinflammatory status, fundamental to know the objectives of dietary strategies.
- To review the most important dietary modifications useful to slow or control the disease activity according to recent evidence.

Introduction

The group of arthritis includes diverse forms of joint disease with or without some degree of inflammation. In this article these entities will be reviewed from a physiopathological and clinical perspective. Among the reviewed pathologies RA and OA are the most prevalent. It is obvious that dietary recommendations are not a substitute of treatments, but they are integrated into the multidisciplinary intervention. The review of the evidence is necessary to propose those diets or dietary modifications able to improve the prognosis of the patients, not only from a clinical point of view but also considering quality of life. Diet is particularly relevant by modulating the inflammatory response and even modifying the cardiovascular risk. There is plenty of information about diet, but the validated studies are scarce. The main approaches have been supplementation and elimination. Elimination diets suppress dietary components that may be related to immune response. The supplementation approach is also interesting, and omega 3 fatty acids are one the most investigated elements. This article also includes some ideas about the positive effect of exercise combination.

Definitions and etiology

Over a hundred types of arthritis are currently recognized. Among the degenerative arthritides, OA is the most common form, and the prototype of this group. In OA there is inflammation within the joint, but there is no evidence of whole-body inflammation, a key feature in distinguishing OA from the inflammatory arthritides. Clinically the patient presents pain, stiffness and functional disability with a reduced quality of life.

In general, OA is monoarticular or oligoarticular and affects a few joints, usually the large weight-bearing joints of the lower extremities, such as the knees and hips, but can be polyarticular. OA can also affect the hands, especially in women, but without the systemic illness that characterizes inflammatory diseases such as RA. The etiology of OA is unknown. The process starts with the micro rupture of the cartilage with a progressive inflammatory cascade that affects synovium, cartilage, and the surrounding structures. This evolution origin a loss of joint space and bony overgrowth, causing pain first with weight-bearing, then with passive

motion, and finally at rest. Obesity is considered a factor leading to a greater load over articular cartilage. Some publications have described the possible causal effect of vitamin D and antioxidant deficiency.

In contrast, in an inflammatory arthritis, such as RA, there is a systemic autoimmune disease with inflammatory polyarthritis, usually affecting the small joints of the hands, wrists, and feet, often spreading to include the knees and hips. Untreated inflammation may lead to bone and cartilage erosions with join destruction and functional impairment. There is evidence of a systemic immune response, with initial activation of clones of autoreactive T cells into the joint space with development of a thickened synovial membrane that releases proinflammatory cytokines, including interleukin (IL)-1β, tumor necrosis factor alpha (TNF)-α, IL-6, and proteases that induce bone erosions. There is also activation of the acute-phase response, with reduced albumin synthesis and increased production of fibrinogen, C reactive protein (CRP), and other acute-phase reactants. The systemic inflammation leads to altered energy and protein metabolisms and wasting of body cell mass and muscle mass, process described as "rheumatoid cachexia," that comprises the effects of undernutrition on motor and sensory functions with a combination of weakness, muscle atrophy and functional impairment.

Prevalence

OA is the most common joint affliction, and its prevalence increases dramatically with age. Radiographic evidence of OA is seen in 70% of people aged over 65 years, but symptoms do not necessarily correspond with X-ray changes. OA of the hip has been reported in 7–25% of adults aged 55 years and older, whereas knee OA is approximately twice as common, and OA of the hands three times as common. After 50 years of age there is an increased incidence in females. RA, however, affects 1–2% of the population, but generally attacks many more joints than OA and is associated with a twofold or higher increased risk of death. Other inflammatory arthritides, such as the seronegative spondyloarthropathies, are much less common.

Clinical features

The term "arthritis" means the presence of pain and inflammation (heat, swelling, redness) in a joint. Joint pain without inflammation is "arthralgia" and may be due to disease within the joint or in the surrounding soft tissues, ligaments, and tendons. Degenerative arthritis, such as OA, is generally a disease of the large weight-bearing joints of the lower extremities. In addition, OA commonly strikes the distal interphalangeal (DIP) and first carpometacarpal joints of the hands, especially in women. The affected joints have pain on motion, stiffness, mild swelling, and sometimes intra-articular effusions or swelling. As the disease progresses, bony overgrowth becomes clinically apparent, coinciding with the development of osteophytes on radiographic examination. These osteophytes, together with loss of joint space, are the radiographic hallmarks of OA, and reflect new bone formation at the joint margins. Over time, the range of motion in the joint is restricted, with loss of function, first by pain, later by loss of joint space, and finally by the osteophytes. Treatment of OA is essentially symptomatic. Nonpharmacologic therapies include exercise as a useful intervention for OA of the hands, hips and knees. The evidence is limited for topical creams, cognitive-behavioral therapy and acupuncture. Diverse pharmacologic interventions have been used: analgesics (topical and systemic), duloxetine and nonsteroidal anti-inflammatory drugs (NSAIDs) to reduce pain and limit the intra-articular inflammation. However, this treatment is seldom completely satisfactory, and occasionally the utility is limited by secondary effects. Finally, progression of the disease is usually seen, although prognosis is variable depending on the extent of the disease. Joint replacement surgery has revolutionized the care of end-stage OA, allowing return of function of joints that are otherwise immobile.

In inflammatory arthritis the situation is quite different. RA is a symmetric, additive polyarthritis involving up to several dozen small joints of the hands, wrists, and feet, often with involvement of the knees, hips, and ankles, and sometimes the elbows, shoulders, and cervical spine, distal interphalangeal joints are spared. There is pain, swelling, and warmth in the affected joints and morning stiffness or after prolonged immobility that can last for several hours. Unlike OA, in RA there is evidence of whole-body inflammation with constitutional symptoms such as fatigue, low grade fever or weight loss and often extraarticular manifestations. This leads to suppression of albumin gene expression and upregulation of the production of acute-phase proteins such as CRP, transferrin, and fibrinogen. In addition, there is suppression of serum iron, reduced zinc, and increased whole-body protein breakdown and resting metabolic rate. Treatment begins with rest, physical therapy, and use of NSAIDs to reduce pain and mild inflammation. Low-dose oral corticosteroids, equivalent to 5–10 mg day^{-1} of prednisone, are often necessary to control inflammation. However, these therapies do not alter the natural history of the disease and frequently is necessary to use nonbiologic disease-modifying antirheumatic drugs (DMARD) such as methotrexate, and biologic DMARD such as TNF-α inhibitors (infliximab, etanercept, adalimumab), and others (tocilizumab, rituximab) that have been shown to prevent erosions. It should be noted that some of these medications may also affect the nutritional status of individuals with RA via either altered appetite, blood sugar, plasma lipids, absorption, or protein metabolism.

Role of diet in the management of inflammatory arthritis

Nutritional assessment in RA

It is important to recognize that patients with RA do not have a normal nutritional status. It is useful to define the concomitant nutritional alterations based on the physiopathological mechanisms involved in the clinical practice: cachexia, when predominate chronic disease-related inflammatory conditions, protein-caloric malnutrition in case of malnutrition associated to acute inflammatory conditions and starvation-related malnutrition, where inflammation is not present. Cachexia is generally seen in the presence of hypermetabolism (elevated resting energy expenditure) and hypercatabolism (elevated protein breakdown), along with reduced physical activity. Clinical expression of malnutrition associated to rheumatic diseases can be different in the patient with low weight or overweight but is common the existence of lean body mass depletion with functional impact and impaired quality of life. On the other hand, the increase in body fat mass, or even the existence of sarcopenic obesity, is related with greater cardiovascular risk.

These metabolic abnormalities are linked to increased production of the catabolic cytokines IL-1β, IL-6 and TNF-α. The problems are further exacerbated by reduced physical activity, which minimizes the anabolic stimulus to muscle. In general, is described an increased baseline energy expenditure, muscle proteolysis, insulin resistance, anemia of chronic disease, osteoporosis, endothelial dysfunction, atherogenesis and even metabolic syndrome.

The combined effect of these metabolic alterations in a proinflammatory state leads to anorexia, unintentional weight loss, decreased muscle mass with reduced muscle strength and functional limitations, poor quality of life and shorter life expectancy. It is important do not forget the effect of potential side effects of the drugs used for treatment of rheumatologic diseases.

Diet and RA activity

Although many foods or food components have been considered as possible treatments for RA, most studies have focused on either supplementation (particularly the use of fish oil) or the use of an elimination diet, especially fasting or a vegetarian regimen (Tedeschi and Costenbader, 2016). Certain nutrients (omega-3 fatty acids, beta-hydroxymethylbutyrate and leucine) have regulatory properties over the immune and inflammatory response, typically present in these diseases. The investigation about the role of diet and microbiota in the development of RA is growing (Alpízar-Rodríguez et al., 2020). Microbiota and the intestinal barrier can be the link between various nutritional factors and the development of RA, although more clinical trials are necessary. The most recent systematic review published in 2020 was based on the Disease Activity Score in 28 joints (DAS28). The positive effects on RA disease activity were demonstrated for Mediterranean diet (MD), spices, antioxidants (quercetin and ubiquinone) and probiotics (*Lactobacillus casei*) (Nelson et al., 2020).

The recommended diet includes a low-fat low-sodium MD rich in vegetables, whole grains, fruits and poor in sugar-sweetened beverages, red and processed meat and trans fats (Genel et al., 2020). Another important aspect would be the supplementation with omega-3 fatty acids, essential amino acids and probiotics. A recent review evaluated a food pyramid for patients with RA (Rondanelli et al., 2021). This pyramid is similar to the recommended for the general population with some differences: 5 portions of fruit and vegetables are at the base, instead of carbohydrates (preferable gluten free). The seeds have been considered as daily consumption. At the apex two warnings, to avoid the excessive intake of salt and simple sugars and the recommendation to take a supplement of vitamin D, omega 3 and antioxidants in most cases (**Fig. 1**).

Supplementation with dietary fatty acids

Various dietary fatty acids have shown numerous immunomodulatory effects. Some fatty acids are considered essential because human body cannot synthesize them. Is the case of linoleic acid (LA) and alfa-linolenic acid (ALA). LA is a n-6 polyunsaturated fatty acid (PUFA) precursor of Arachidonic acid (AA), synthesized in mammalian tissues. ALA is a n-3 PUFA precursor. Both are essential components of phospholipid membranes and precursors of inflammatory mediators, such as prostaglandins (PG) and leukotrienes (LT). LA has a pro-inflammatory profile and ALA offers an anti-inflammatory response. In clinical practice is more interesting to explain the role of a dietary pattern than to focus on a specific nutrient. MD typically contains a balanced n-6/n-3 PUFA ratio and is a reasonable diet intervention (Chehade et al., 2019). A higher adherence to the MD is associated with a more diverse microbiota and elevated levels of fecal short chain fatty acids (SCFA), characterized by their anti-inflammatory properties. In patients under treatment with glucocorticoids, sugar and salt intake should be restricted.

More specifically, Eicosapentaenoic acid (EPA) derived eicosanoids result in decreased platelet aggregation, reduced neutrophil chemotaxis, and anti-inflammatory effects. Some meta-analysis have demonstrated a significant reduction in leukotriene B4 (LTB_4) levels after n-3 PUFA supplementation in RA patients. This lipid mediator has a relevant participation in the recruitment of leukocytes in cases of inflammation.

EPA and docosahexaenoic acid (DHA) supplementation causes modest improvement in the number of tender joints and fatigue among patients with RA, although clinical benefits have generally been small, subjective, and transient (Petersson et al., 2018). Possible mechanisms for this improvement in clinical symptoms of inflammation include altered neutrophil membrane lipid composition, reduced IL-1 production, or a change in the α-tocopherol content of the diet. Overall, findings suggest that clinical benefits of dietary supplementation with PUFAs are more commonly observed among patients consuming higher dosages of fish oil (minimum required dose 2.6 g/day of EPA + DHA), for longer periods than those previously studied. Indeed, beneficial

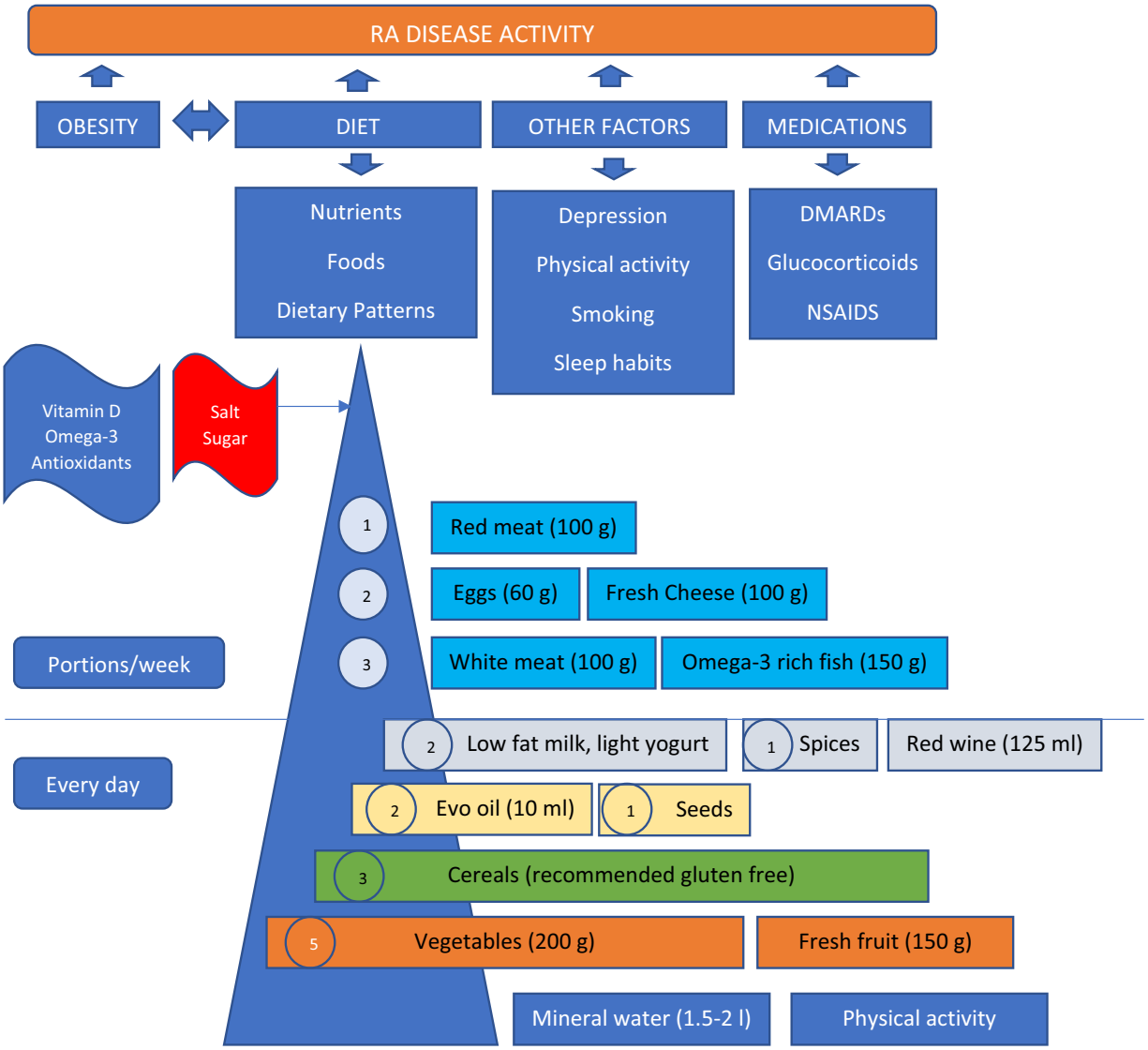


Fig. 1 Ideal food pyramid for RA. Number of portions (in circle). Not recommended (red flag). NSAIDS (nonsteroidal anti-inflammatory drugs). DMARDs (disease-modifying antirheumatic drugs). Modified from Rondanelli, M., et al., 2021. Ideal food pyramid for patients with rheumatoid arthritis: a narrative review. Clin. Nutr. 40(3), pp. 661–689. https://doi.org/10.1016/j.clnu.2020.08.020.

clinical effects have been observed for as long as 1 year among patients with RA ingesting 2.6 g daily of n-3 PUFA supplements. In terms of optimal dosage of supplementation, however, a level of 130 mg kg^{-1} day^{-1} (9 g of n-3 PUFAs in a person weighing 70 kg) has been shown to result in no additional improvement compared with patients receiving doses ranging from 3 to 6 g daily. Therefore, although the optimal level of fish oil supplementation is yet to be determined, there does appear to be an upper limit beyond which no additional benefit exists for patients.

Although some studies seem to suggest modest clinical improvements as a result of dietary fish oil supplementation in patients with RA, the effectiveness of n-3 PUFA supplementation combined with biologic DMARDs is not clear. On the other hand, the role in the RA comorbidities is under investigation (Navarini et al., 2017).

NSAIDs are known to inhibit the cyclooxygenase enzyme system, which is the same pathway that seems to be inhibited by EPA and DHA. It is therefore possible that in studies of fish oil supplementation, where patients are simultaneously maintained on NSAIDs, the effect of EPA is diminished, because the cyclooxygenase pathway is already inhibited by concurrent treatment with NSAIDs. Several studies have attempted to address this issue and have demonstrated a modest effect of n-3 fatty acid supplementation in both, patients who are treated with NSAIDs and those who are not, suggesting that concurrent treatment with NSAIDs does not seem to diminish the effect of n-3 fatty acids. A meta-analysis published in 2007 indicated that n-3 PUFA supplementation was associated with a decrease in NSAID consumption.

Although most studies regarding manipulation of dietary fatty acids have focused on fish oil supplementation, other fatty acids have also been studied. The use of ALA, the precursor of EPA and DHA, has not been shown to be of any benefit in RA. However,

γ-linolenic acid, found in blackcurrant seed, evening primrose, and borage seed oils, has resulted in clinically important reductions in the signs and symptoms of disease activity in patients with RA, perhaps via a reduction in PGE_2, IL-1, and IL-6. Because these oils do not cause an unpleasant fishy taste and odor in recipients, they may be preferred to fish oils for chronic treatment.

In summary, most studies of dietary supplementation with n-3 fatty acids suggest a modest improvement in clinical symptoms associated with RA, which are to some extent dose- and time-dependent. The most consistent clinical benefits have been reductions in tender joint counts and morning stiffness. Studies do not suggest that benefits are great enough to warrant discontinuing patients' other medications. However, the use of fish oil supplements, or diets rich in marine fish, may further improve clinical symptoms among patients with RA. Beyond the possible benefits in terms of controlling inflammatory symptoms of RA, increases in n-3 PUFAs are also associated with reduced risk of cardiovascular disease, increase of muscle mass in patients with primary and secondary sarcopenia and other health benefits.

Vitamin, mineral supplementation and probiotics

Most studies involving vitamin or mineral supplementation in RA have focused on either the antioxidant nutrients (vitamin C, vitamin E, beta carotene, selenium) or B vitamins. In general, results from randomized controlled trials of vitamin E supplementation have been of relatively short duration and have led to conflicting results so that there continues to be a lack of concrete evidence to support vitamin E supplementation at a particular dosage. Nonetheless, patients with RA could certainly be encouraged to increase their intake of vitamin E-rich foods, including edible vegetable oils (sunflower, safflower, canola, olive), unprocessed cereal grains, and nuts. Similarly, the effect of dietary sources of other antioxidant nutrients, such as selenium and vitamin C, on inflammatory symptoms in RA has also been ambiguous. Selenium is a nuclear component of antioxidant enzymes that intervenes in the oxidative and inflammatory state affecting disease activity. It should be emphasized that providing individual nutrient supplements does not necessarily offer the same overall benefit as when nutrients are obtained from whole foods. It is possible that the combinations of nutrients that are present in whole foods, or even some unidentified components of a food, are responsible for any observed beneficial effects, and that supplementing a typical diet with individual nutrients will not provide the same benefit.

In previous studies plasma levels of pyridoxal-5-phosphate (PLP), the metabolically active form of vitamin B_6, were lower in patients with RA compared to control subjects. Furthermore, plasma levels of PLP were inversely associated with TNF-α production by peripheral blood mononuclear cells, suggesting that abnormal vitamin B_6 status may be contributing to inflammation in RA. However, there is no evidence to support the efficacy of oral vitamin B_6 supplements for treating the symptoms of RA at this time. Furthermore, large doses of vitamin B_6 can be toxic; therefore, as with the antioxidant nutrients, patients with RA would obtain the greatest benefit by increasing dietary sources of vitamin B_6, consistent with the dietary reference intake (DRI) for this nutrient. If supplementation is considered, it should not exceed twice the DRI level. The active metabolite of vitamin D shows immunomodulatory effects mediated by vitamin D receptor expressed in multiple immune cells. Some authors describe the possible role of vitamin D deficiency in the progression from preclinical to clinical RA. In the Women's Health Initiative clinical trial higher levels of vitamin D intake were associated to an increased risk of RA.

Various authors have investigated the application of probiotics on disease activity and metabolic status of RA patients. There was a reduction in the levels of pro-inflammatory cytokines (TNF, IL-5, IL 12) with increased levels of anti-inflammatory IL-10. In the most recent meta-analysis (2018), the effect over disease related outcomes was considered weak in RA patients. Some observational and interventional studies suggest that diverse fermented foods (foods and beverages produced under controlled microbial growth) can have a protective role against immune mediated disorders such as RA. The benefit is based on microbiota modulation secondary to bioactive metabolites such as SCFA. Polyphenols are plant derived molecules with anti-inflammatory properties. Their use can be useful for OA treatment, but more clinical trials are needed.

Fasting and vegetarian diets

An alternative approach to alleviating the symptoms associated with chronic inflammation is elimination of various foods or food components, most often by fasting or assuming a lactovegetarian diet. Some studies have demonstrated a significant improvement in various objective and subjective measures of disease activity, including number of tender and swollen joints, Ritchie's articular index, morning stiffness duration, erythrocyte sedimentation rate (ESR), CRP, grip strength, and score on health assessment questionnaires, among patients with RA, 6 weeks-2 years after initiating a vegetarian diet. Furthermore, these clinical improvements were accompanied by changes in biochemical and immunological parameters consistent with a substantial reduction in inflammatory activity. Other studies included in the last Cochrane review (2009), have demonstrated no clinical improvement among patients with RA following different dietary interventions. Some meta-analysis described long-term benefits in subjects with RA that include fasting programs followed by lactovegetarian diets.

Several possible mechanisms have been proposed to explain the impact of elimination diets on clinical symptoms in RA. One possibility is that RA might be the result of hypersensitivity to environmental toxins or specifically to foods or food-related products, resulting in a food allergy of sorts that exacerbates symptoms of RA. Gluten free diets followed for 1 year were associated with a decrease of disease activity. However, true food intolerance, involving a systemic humoral immune response against food items, appears to be relatively uncommon among patients with RA. Another possible mechanism that has been proposed includes an alteration in the fatty acid content of the diet. Vegetarian diets contain more LA, but less AA, EPA, and DHA than omnivorous diets.

Therefore, the eicosanoid precursors (AA, EPA, and DHA) must be produced endogenously from LA and ALA. It has been hypothesized that if this endogenous production cannot compensate for the absence of AA in the diet, then the precursor of the proinflammatory eicosanoids would be reduced, perhaps explaining the beneficial effect of vegetarian diets in patients with RA. Furthermore, it has also been demonstrated that fasting for 7 days resulted in decreased release of LTB_4 from neutrophils, in addition to reductions in morning stiffness, articular index, and ESR, but that this reduction in LTB_4 occurred despite an increased AA content of the serum, platelets, and neutrophils. These findings suggest that perhaps fasting may impair a metabolic step of AA conversion. Caloric restriction could have beneficial effects on RA disease activity by inhibiting NLRP3 inflammasome (component of the innate immune system that mediates secretion of proinflammatory cytokines).

Other potential mechanisms include the possible effect of a vegetarian diet on antioxidant status acting as scavengers of free radicals, or on other dietary practices frequently associated with vegetarianism. Plant-based foods are naturally high in antioxidant nutrients (vitamin C, vitamin E, and beta-carotene) and low serum antioxidant levels have been associated with an increased risk of developing RA, although the specific mechanism involved remains unknown. Different investigations have described oxidative stress, lipid peroxidation, DNA damage, increased formation of reactive oxygen species and reduced antioxidant protective systems. Certainly, RA is associated with increased production of reactive oxygen species; these compounds seem to contribute to the inflammatory process, so a diet high in antioxidants could limit damage via their anti-inflammatory properties. Although changes in fatty acid composition or antioxidant status seem to be the most plausible explanations for the potential benefit of adhering to a vegetarian diet, there are other possible mechanisms as well. Fasting, for example, suppresses inflammation and frequently a period of fasting is recommended before initiating an elimination or vegetarian diet; it is possible that this fasting period contributes to the reduction in inflammation among patients with RA following a vegetarian diet.

Conclusion

In summary, the notion that food sensitivity reactions contribute significantly to clinical symptoms associated with RA remains controversial. However, it seems that at least a small subgroup of patients with RA may benefit from individualized dietary manipulation involving elimination of specific foods or food components, in combination with other medical therapies. However, fasting and other elimination diets should be used with caution in light of the prevalence of rheumatoid cachexia in this population. Such patients are prone to further loss of cell mass during restrictive diets, and the net effect may be to do more harm than good.

Nutritional alterations are prevalent in patients with rheumatologic disease. The initial diagnosis is essential, with application of instruments that provide an objective evaluation of body composition, particularly lean body mass. Comparison studies on handgrip strength and other related measures could explain the difference between hand inflammation and muscle function. Of the two primary approaches to the dietary management of inflammatory arthritis—supplementation and elimination diets—it appears that dietary supplementation with fish oil may result in the most consistent clinical benefits, although improvements still remain modest. Elimination diets, including fasting and vegetarian regimens, may provide some benefit for a limited number of patients with RA, but consistent alleviation of disease activity by objective clinical measures has not been demonstrated.

In neither case does the use of dietary management warrant discontinuing a patient's medical regimen; rather, diet may be useful as an adjunct to other more substantiated therapies. Perhaps the most prudent approach for patients with RA interested in attempting to control their disease activity is to adopt a lifestyle modification program based on diet and physical activity. More definitive research demonstrating consistent, objective clinical benefits is needed before specific dietary manipulations for patients with RA can be recommended. The physical activity program requires an active lifestyle with erobic and resistance exercise. Low levels of physical activity and glucocorticoids use are associated with nutritional complications. The French Society for Rheumatology published in 2021 specific dietary recommendations for patients with Chronic Inflammatory Rheumatic Diseases (Daien et al., 2021). **Table 1** resume the main aspects (here).

Role of diet in the management and prevention of degenerative arthritis

Much less is known about the role of diet in the treatment of OA and other degenerative arthritides (Thomas et al., 2018). The above discussion regarding n-3 PUFAs in RA may also pertain to OA, although the strength of the effect has not been studied as thoroughly. However, the same eicosanoid metabolism occurs in OA as in RA, with the exception that the disorder is limited to the joint rather than involving the whole body. Thus, fish oils may well be of benefit in OA. The evidence for dietary lipid modification is low but can improve metabolic health. Antioxidant intervention with vitamin E may also be effective in OA, with several studies showing an effect comparable with those of NSAIDs. Although not strictly nutrients, glucosamine and chondroitin sulfate, which are two of the constituents of normal cartilage that decline with arthritis, have been shown to be useful when given as an oral supplement, especially in patients with early OA. Glucosamine sulfate decreased the risk of developing radiographic knee OA over 2.5 years in overweight middle-aged women.

In contrast to RA, where diet's main role is in the treatment and little is known about prevention, there is more known about dietary components that lead to OA than about nutritional management of OA (Berenbaum et al., 2018).

It is clear that OA of the lower extremities is largely a problem brought on by obesity, especially OA of the knee (and hip, to a much lesser extent), suggesting that obesity seems to be a mechanical rather than systemic risk factor. Thus, maintaining body

Table 1 Dietary recommendations for patients with chronic inflammatory rheumatic diseases.

Recommendations	*Level of evidence*
In patients who are overweight or obese, weight loss support should be proposed to control chronic inflammatory rheumatic disease activity; weight loss also has beneficial cardiometabolic and psychological effects	C
A gluten-free diet should not be proposed to control chronic inflammatory rheumatic disease activity, in the absence of confirmed celiac disease	C
Fasting or vegan diets should not be proposed to control the activity of chronic inflammatory rheumatic diseases	D
Eliminating dairy products should not be proposed for managing chronic inflammatory rheumatic disease	C
Supplementation with polyunsaturated fatty acids, mainly omega-3, of more than 2 g per day, could be proposed for symptomatic relief in patients who have rheumatoid arthritis and likely for those suffering from other chronic inflammatory rheumatic diseases.	A
A Mediterranean-type diet could be proposed to patients who have rheumatoid arthritis and likely to those affected by other chronic inflammatory rheumatic diseases given its effects on joint symptoms and the cardiometabolic diseases	C
To control the activity of chronic inflammatory rheumatic disease, there is no indication for proposing vitamin (B9, D, E, K) or trace element (selenium and/or zinc) supplementation	B
Given that the data on the effectiveness of probiotics is insufficient and disparate, they are not recommended for controlling chronic inflammatory rheumatic disease activity.	B

General principles

- Nutritional advice is not a substitute for pharmacological treatment of chronic inflammatory rheumatic diseases.
- The nutritional advice given to patients affected by chronic inflammatory rheumatic diseases should be based on data from the scientific literature.
- Nutritional support is integrated into the overall care of patients affected by chronic inflammatory rheumatic diseases
- Broaching dietary habits can help patients get actively involved in the overall care of their chronic inflammatory rheumatic disease
- The nutritional advice given to patients affected by chronic inflammatory rheumatic diseases must take into account the intra- and extra-articular effects, particularly cardiometabolic and bone
- The nutritional advice must take the cultural and socioeconomic context into account
- Nutritional advice is indissociable from the promotion of exercise.
- If other nutritional recommendations exist that are specific to a disease, clinical condition or associated treatment, these continue to apply (e.g., undernutrition, obesity, sarcopenia, osteoporosis, etc.)

Modified from Daien, C., et al., 2021. "Dietary recommendations of the French society for rheumatology for patients with chronic inflammatory rheumatic diseases". Joint Bone Spine p. 105319. https://doi.org/10.1016/j.jbspin.2021.105319.

weight within the recommended ranges is probably the most important nutritional intervention to prevent OA. Weight loss leads to reduction in joint stress, and often reduces symptoms. In fact, recent studies have suggested that if all overweight and obese individuals reduced their body weight by 5 kg, or until their body mass index (BMI) was within the desirable range, 24% of surgeries for knee OA could be avoided. A recent meta-analysis described that a substantial loss of weight is necessary to reduce knee pain and joint stiffness improving physical function. Furthermore, studies have demonstrated that exercise can improve OA symptoms even independently of weight loss, presumably by increasing muscle strength and thus improving the shock-absorbing power of the muscles, hence sparing the cartilage and joint (Alrushud et al., 2017). However, patients with OA have a great deal of difficulty with exercise, and their sedentary life style is reinforced by their joint pain, generally leading to weight gain after the onset of OA, which in turn exacerbates the disease, creating a vicious cycle. Adults with knee OA are at high risk of sarcopenia. Some studies have demonstrated the combined effect of protein-rich supplementation and resistance exercise training over physical activity level and muscle gain (Liao et al., 2021). Exercise programs that increase physical activity and strengthen the muscles surrounding afflicted joints clearly improve symptoms in OA. Thus, OA can be thought of as a disease of overnutrition, whereas RA is generally

a disease of undernutrition. Interestingly, recent twin studies have examined the role of genetic versus environmental factors as mediators of the obesity–OA relationship, and have suggested that shared genetic factors are not as important as environmental factors in mediating the obesity-OA relationship. Dietary modification leading to weight loss is a critical component of the management of OA. MD influences inflammatory and cartilage degradation biomarkers, although longer effects require more investigation. A lower prevalence of OA is observed in subjects with a high adherence to MD. Obesity and metabolic syndrome have a role in the development of OA, including hand OA. The link between the two conditions can be a chronic low-grade inflammation related to obesity and gut dysbiosis. Another risk factor is vitamin D deficiency. In this point the adherence to a Mediterranean dietary pattern may counteract this proinflammatory state. Its consumption has been related to decreased pain, disability, reduce depressive symptoms and a clear improvement in knee cartilage (Dyer et al., 2017). A greater adherence to the DASH (Dietary Approaches to Stop Hypertension) dietary pattern is also associated with lower prevalence of OA.

Conclusion

The treatment of OA, with no existent disease-modifying drugs, is essential in Rheumatology so dietary modifications, and an additional exercise program, are basic to modify risk factors and preventing or limiting the consequences. Many epidemiological studies focus on the link between metabolic syndrome and OA, and include the presence of some potentially modifiable factors, such as obesity and dietary habits. The actual evidence is suggesting an important role for MD.

See Also: Cytokines: Metabolic and nutritional aspects; Starvation and fasting: Biochemical aspects; Vegetarian diets

References

Alpízar-Rodríguez, D., et al., 2020. The role of nutritional factors and intestinal microbiota in rheumatoid arthritis development. Nutrients 13 (1). https://doi.org/10.3390/nu13010096.

Alrushud, A.S., et al., 2017. Effect of physical activity and dietary restriction interventions on weight loss and the musculoskeletal function of overweight and obese older adults with knee osteoarthritis: a systematic review and mixed method data synthesis. BMJ Open 7 (6), e014537. https://doi.org/10.1136/bmjopen-2016-014537.

Berenbaum, F., et al., 2018. Modern-day environmental factors in the pathogenesis of osteoarthritis. Nat. Rev. Rheumatol. 14 (11), 674–681. https://doi.org/10.1038/s41584-018-0073-x.

Chehade, L., et al., 2019. Lifestyle modification in rheumatoid arthritis: dietary and physical activity recommendations based on evidence. Curr. Rheumatol. Rev. 15 (3), 209–214. https://doi.org/10.2174/1573397115666190121135940.

Daien, C., et al., 2021. Dietary recommendations of the French society for rheumatology for patients with chronic inflammatory rheumatic diseases. Joint Bone Spine 105319. https://doi.org/10.1016/j.jbspin.2021.105319.

Dyer, J., et al., 2017. Effect of a Mediterranean type diet on inflammatory and cartilage degradation biomarkers in patients with osteoarthritis. J. Nutr. Health Aging 21 (5), 562–566. https://doi.org/10.1007/s12603-016-0806-y.

Genel, F., et al., 2020. Health effects of a low-inflammatory diet in adults with arthritis: a systematic review and meta-analysis. J. Nutr. Sci. 9, e37. https://doi.org/10.1017/jns.2020.31.

Liao, C.-D., et al., 2021. Effects of protein-rich nutritional composition supplementation on sarcopenia indices and physical activity during resistance exercise training in older women with knee osteoarthritis. Nutrients 13 (8). https://doi.org/10.3390/nu13082487.

Navarini, L., et al., 2017. Polyunsaturated fatty acids: any role in rheumatoid arthritis? Lipids Health Dis. 16 (1), 197. https://doi.org/10.1186/s12944-017-0586-3.

Nelson, J., et al., 2020. Do interventions with diet or dietary supplements reduce the disease activity score in rheumatoid arthritis? A systematic review of randomized controlled trials. Nutrients 12 (10). https://doi.org/10.3390/nu12102991.

Petersson, S., et al., 2018. The Mediterranean diet, fish oil supplements and Rheumatoid arthritis outcomes: evidence from clinical trials. Autoimmun. Rev. 17 (11), 1105–1114. https://doi.org/10.1016/j.autrev.2018.06.007.

Rondanelli, M., et al., 2021. Ideal food pyramid for patients with rheumatoid arthritis: a narrative review. Clin. Nutr. 40 (3), 661–689. https://doi.org/10.1016/j.clnu.2020.08.020.

Tedeschi, S.K., Costenbader, K.H., 2016. Is there a role for diet in the therapy of rheumatoid arthritis? Curr. Rheumatol. Rep. 18 (5), 23. https://doi.org/10.1007/s11926-016-0575-y.

Thomas, S., et al., 2018. What is the evidence for a role for diet and nutrition in osteoarthritis? Rheumatology 57 (Suppl._4), iv61–iv74. https://doi.org/10.1093/rheumatology/key011.

Further reading

Bortoluzzi, A., et al., 2018. Osteoarthritis and its management—epidemiology, nutritional aspects and environmental factors. Autoimmun. Rev. 17 (11), 1097–1104. https://doi.org/10.1016/j.autrev.2018.06.002.

Dourado, E., et al., 2020. Diet as a modulator of intestinal microbiota in rheumatoid arthritis. Nutrients 12 (11). https://doi.org/10.3390/nu12113504.

Forsyth, C., et al., 2018. The effects of the Mediterranean diet on rheumatoid arthritis prevention and treatment: a systematic review of human prospective studies. Rheumatol. Int. 38 (5), 737–747. https://doi.org/10.1007/s00296-017-3912-1.

García-Morales, J.M., et al., 2020. Effect of a dynamic exercise program in combination with Mediterranean diet on quality of life in women with rheumatoid arthritis. J. Clin. Rheumatol. 26 (7S Suppl. 2), S116–S122. https://doi.org/10.1097/RHU.0000000000001064.

Gioia, C., et al., 2020. Dietary habits and nutrition in rheumatoid arthritis: can diet influence disease development and clinical manifestations? Nutrients 12 (5). https://doi.org/10.3390/nu12051456.

Li, Y., et al., 2016. Diet-intestinal microbiota Axis in osteoarthritis: a possible role. Mediat. Inflamm. 2016, 3495173. https://doi.org/10.1155/2016/3495173.

Miles, E.A., Calder, P.C., 2012. Influence of marine n-3 polyunsaturated fatty acids on immune function and a systematic review of their effects on clinical outcomes in rheumatoid arthritis. Br. J. Nutr. 107 (Suppl. 2), S171–S184. https://doi.org/10.1017/S0007114512001560.

Morales-Ivorra, I., et al., 2018. Osteoarthritis and the Mediterranean diet: a systematic review. Nutrients 10 (8). https://doi.org/10.3390/nu10081030.
Oliviero, F., et al., 2015. How the Mediterranean diet and some of its components modulate inflammatory pathways in arthritis. Swiss Med. Wkly. 145, w14190. https://doi.org/10.4414/smw.2015.14190.
Zhang, Y., et al., 2020. Adherence to DASH dietary pattern is inversely associated with osteoarthritis in Americans. Int. J. Food Sci. Nutr. 71 (6), 750–756. https://doi.org/10.1080/09637486.2020.1722075.

Relevant websites

Arthritis Foundation web. Provides a Multidisciplinary Intervention Over Different Types of Arthritis. It is Classified on Three Parts: Treatment, Healthy Living and Drug Guide. It Covers Different Age Groups.
http://www.euro.who.int/nutrition.
https://www.arthritis.org/health-wellness/detail?content=healthyliving&filter=nutrition.
World Health Organization website. CINDI (Countrywide Integrated Noncommunicable Disease Intervention) Dietary Guide. Provides a Guide for Healthy Eating and Healthy Lifestyles, as Suggested by the World Health Organization.

Brain and nervous system: Biology, metabolism, and nutritional requirements

J.D. Fernstrom and M.H. Fenstrom, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of J.D. Fernstrom, M.H. Fenstrom, Brain and Nervous System: Biology, Metabolism, and Nutritional Requirements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 200–206, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00030-1.

Design of the Nervous System 119
The Blood–Brain Barrier (BBB) 119
Energy Substrates 119
Amino Acids and Protein 120
Fatty Acids and Choline 121
Fatty Acids 121
Choline 121
Vitamins 121
Water-Soluble Vitamins 122
Fat-Soluble Vitamins 123
Minerals 123
Further Reading 124

Glossary

Blood–brain barrier Now not known really to be a 'barrier', but a set of transporter mechanisms located in the cell membranes of the endothelial cells that make up the brain capillaries, which promote or retard the exchange of certain small and large water-soluble molecules between the blood and the brain.

Central nervous system (CNS) The portion of the nervous system consisting of the brain, retina, and the spinal cord.

Glia Non-neuronal cells of the nervous system that provide physical and metabolic support for neurons, and insulate axons and nerve terminals, to ensure privacy in signaling.

Ketone bodies Small breakdown products of fatty acids, such as acetoacetate and beta-hydroxybutyrate, made in the liver when circulating fatty acids increase, such as during starvation, which are then taken up by brain (and other tissues) and used to generate ATP. Ketone bodies are a key source of energy for the brain during starvation.

Neuron The principal functional cells of the nervous system, which are organized into very complex circuits (analogous to the components of electronic circuits) that manage all of the housekeeping functions of the body and elaborate behavior and cognitive functions.

Neurotransmitter A molecule that is released by neurons that couples with specific receptors on other cells and induces a functional response, such as depolarization (a neuron), contraction (a muscle cell), or secretion (an endocrine or exocrine cell).

Parkinson's disease A disease of unknown etiology in which neuronal circuits in the brain that control certain features of muscle function and coordination degenerate. Patients, who are usually older, gradually develop muscle tremor, a bent posture, and difficulty in initiating voluntary movements, such as walking.

Peripheral nervous system The portion of the nervous system consisting of neurons and glial cells lying outside the CNS, which either supply sensory information to the CNS or send commands from the CNS to effector cells in the rest of the body, such as muscle and gland cells.

Rhodopsin The light-responsive pigment of the eye, which elicits phototransduction when struck by photons of light. It is synthesized in retina photoreceptor cell membranes through the covalent linkage of a protein, opsin, and a metabolite of vitamin A, 11-*cis*-retinaldehyde. Rhodopsin breaks down to opsin and a metabolite of 11-*cis*-retinaldehyde after phototransduction but is quickly regenerated.

Spina bifida A malformation of the spinal cord that occurs early during fetal development, when the spinal cord is forming, which when serious can compromise body functions controlled by nerves entering and exiting the spinal cord below the lesion. The incidence of spina bifida is markedly reduced when women take a folic acid supplement before and during pregnancy.

Design of the Nervous System

The nervous system has two principal cell types, neurons and glia. Neurons (like wires) conduct electrical signals and are organized into circuits to perform specific functions. They have a unique cellular architecture: small cellular extensions (dendrites), which receive chemical and electrical signals from other neurons; a longer extension (the axon, which can be up to a meter in length), which sends electrical signals down its length to one or more nerve terminals. Nerve terminals contain neurotransmitters, molecules released by arriving electrical signals that modify the electrical activity of the adjacent neurons. Neurons have considerable energy needs; indeed the brain, which is approximately 2% of body weight, consumes 15–20% of the body's daily energy intake. Glial cells, which make up approximately 60% of the brain's cell mass, provide physical and metabolic support for neurons and insulate axons and nerve terminals to insure privacy in electrical signaling. The glial cells found in peripheral nerves serve the same functions.

The nervous system is broadly divided into two parts, the central nervous system (CNS) and peripheral nervous system. The CNS consists of the brain, retina, and the spinal cord. It also contains complex neuronal circuits that control body functions (e.g., blood pressure, breathing, hunger, and movement). The peripheral nervous system consists of groups of neurons that mostly lie outside of the CNS and either supply sensory information to the CNS or send CNS commands to effector cells, such as muscle and gland cells.

The Blood–Brain Barrier (BBB)

Each portion of the nervous system is separated from the blood (and thus the rest of the body) by a metabolic 'barrier,' which modulates the access of nutrients to and the removal of metabolites from the neurons and glia within it. For the brain and spinal cord, this barrier is termed the 'blood–brain barrier' (BBB; there is also a blood–cerebrospinal fluid (CSF) barrier; CSF is made from blood); for the retina, it is the 'blood–retinal barrier,' and for peripheral neurons, the 'blood–nerve barrier.' The functions of these barriers are very similar; the focus of the discussion will be the BBB, because it has been the most studied.

The BBB is located in the endothelial cells that make up the brain's capillaries. Unlike the capillaries elsewhere in the body, the endothelial cells of the brain capillaries are tightly joined, such that nothing passes into (or out of) the brain without passing through these cells. The BBB thus presents a continuous lipid barrier to molecules. One implication is that the ease with which molecules in the blood gain access to brain should depend on their lipid solubility: The more lipid soluble, the greater the accessibility to brain by diffusion. However, most molecules of biologic importance to brain are not lipid soluble and thus do not easily diffuse across lipid membranes into brain. Examples include glucose, amino acids, and water-soluble vitamins. Consequently, endothelial cell membranes must be more than just lipid barriers; in fact, embedded in them are specific transport carriers that mediate the brain uptake of most nutrients.

Energy Substrates

The brain uses glucose as its primary energy substrate. Glucose is not lipid soluble and thus requires a BBB transporter. The glucose transporter has a maximal transport capacity for glucose of 1.4 mol min^{-1} g^{-1} of brain or approximately 1200 g d^{-1} for the entire brain (a human brain weighs 1400 g). The human brain consumes 15–20% of the body's oxygen consumption; brain glucose utilization is therefore about 100 g d^{-1}. The BBB transporter thus has a maximal capacity for transporting glucose well in excess of the daily requirements of the brain.

Inside the brain, glucose is rapidly taken up into the neurons by a cellular glucose transporter. Within the neuron, glucose enters the glycolytic pathway. The initial enzyme, hexokinase, has a very high affinity for glucose and is fully saturated at normal brain glucose concentrations. Hence, overall, each step in the glucose pipeline from the blood to brain neurons is designed to maximize glucose supply for neuronal energy production. It only fails when the blood glucose supply is abruptly curtailed, such as when a diabetic patient injects too much insulin and blood glucose levels rapidly fall (the transporter cannot compensate for such abrupt drops in blood glucose). The effect is dramatic: confusion, delirium, seizures, coma, and finally death occur as blood glucose drops to very low levels. Such effects are most rapidly reversed by the infusion of glucose, suggesting that no other compound in blood readily substitutes for glucose as the brain's primary energy substrate.

Normally, the body carefully maintains blood glucose concentrations. During starvation, however, blood glucose falls enough to cause the brain to recruit an additional energy source, ketone bodies. Ketone bodies are liver-produced by-products of the breakdown of stored fat and provide an extended supply of energy when the input of food-derived energy is low. The brain uses ketone bodies whenever their blood levels rise; blood ketone body concentrations rise during starvation. The BBB ketone body transporter (ketone bodies are not lipid soluble) is induced during starvation, enhancing the flow of ketone bodies into the brain. During prolonged starvation, more than half of the energy used by the brain is derived from ketone bodies. However, continued use of some glucose appears obligatory and is supplied via liver gluconeogenesis.

The chronic ingestion of high-fat diets also elevates blood ketone body concentrations, promoting their use by brain for energy production. However, extremely high levels of fat must be consumed and such diets are unpalatable. Hence, diet is not thought normally to influence cerebral energy production via dietary fat manipulation of ketone body supply to brain. Very high fat diets are occasionally used clinically to treat intractable seizures. Although the beneficial effect is linked to circulating ketone bodies levels, the mechanism is presently unknown.

Amino Acids and Protein

Neurons and glial cells in brain use amino acids to produce proteins. In addition, certain amino acids are used to produce small functional molecules such as neurotransmitters. Does diet influence amino acid flow into brain and their use in generating proteins and transmitters? The path from diet to brain begins with amino acid absorption from the gastrointestinal tract, insertion into the circulation, and extraction by the brain. This extraction process involves the BBB, which contains a number of transporters of amino acids. The properties of these transporters dictate how much of each amino acid enters (and exits) the brain. Currently, six carriers have been identified. Of special interest are two carriers: (1) The large neutral amino acid (LNAA) carrier – this carrier is shared by several amino acids (some are precursors for neurotransmitters, namely phenylalanine, tyrosine, tryptophan (TRP), and histidine). This carrier is competitive, allowing changes in the plasma concentration of any one LNAA to affect not only that amino acid's BBB transport but also that of each of its transport competitors. Glutamine, an LNAA present in the brain in high concentrations, drives the brain uptake of the other LNAA, by serving as the principal amino acid counter transported from the brain to blood each time an LNAA is taken up into brain. (2) The acidic amino acid carrier, which transports glutamic and aspartic acids. This carrier primarily transports glutamate (GLU) and aspartate from the brain to the circulation. The other transporters include one selective for basic amino acids; two selective for subgroups of the small, neutral amino acids; and one selective for taurine.

The carriers that move amino acids into brain are those that primarily transport essential amino acids (the large, neutral and basic amino acids), whereas those that move amino acids out of brain are those transporting nonessential amino acids (the acidic and small neutral amino acids). A small, net influx of the essential amino acids into brain no doubt reflects their consumption in brain by biosynthetic and metabolic pathways. The net efflux of the nonessential amino acids, notably aspartate, GLU, glycine, and cysteine may serve to remove from brain the amino acids that act directly as excitatory transmitters or cotransmitters. The brain carefully compartmentalizes these amino acids metabolically, because they excite neurons, and a mechanism to remove them from brain may be a component of this compartmentalization design.

Changes in dietary protein intake have no effect on brain protein synthesis in adults. Indeed, the chronic ingestion of very low levels of dietary protein does not depress brain protein synthesis; brain cells may thus be efficient in retaining and reusing amino acids released during intracellular protein breakdown. In neonatal and infant animals, however, low levels of protein intake are associated with below normal rates of protein synthesis in the brain. But, the presumed mechanism of this association, reduced uptake of essential amino acids into brain, and abnormally low brain concentrations of these amino acids has not been proven. Hence, at present, there is no convincing evidence linking dietary protein intake and brain protein synthesis via a limitation of amino acid availability to brain. For neurotransmitters, the evidence of this diet–brain link is more certain and provides interesting examples of the fundamentally different manner in which the brain uses transport carriers to handle amino acids that are neurotransmitter precursors and those that are neurotransmitters themselves. Good examples are TRP (an LNAA) and GLU (an acidic amino acid), which have been most extensively studied.

TRP is the precursor for the neurotransmitter serotonin (5-HT). The TRP concentration in brain rapidly influences the rate of 5-HT synthesis: Raising brain TRP concentrations increases 5-HT synthesis, whereas lowering brain TRP decreases 5-HT synthesis. Brain TRP uptake and concentrations are directly influenced by the plasma concentrations of TRP and its BBB LNAA transport competitors. The plasma concentrations of TRP and the other LNAA are readily modified by food intake, thereby linking diet to brain 5-HT synthesis. Dietary proteins and carbohydrates are the food components that change brain TRP and 5-HT: Carbohydrate ingestion increases plasma TRP, while lowering the plasma concentrations of its LNAA competitors (an effect dependent on the release of insulin), causing BBB TRP uptake, brain TRP concentrations, and 5-HT synthesis all to increase. The effect of ingesting a meal containing protein depends on the protein: A meal containing α-lactalbumin, a milk protein rich in TRP, causes the plasma concentrations of TRP to rise much more than those of its LNAA competitors. As a consequence, TRP gains a sizeable advantage in the competition for BBB transport, and brain TRP concentrations and 5-HT production increase (considerably more than that seen after carbohydrates are ingested). In contrast, consuming a meal containing zein, a corn protein very low in TRP, causes a marked decline in plasma TRP concentrations, while the plasma concentrations of its LNAA competitors rise. The result is considerable reduction in brain TRP uptake and 5-HT synthesis. The ingestion of meals containing proteins, such as casein, which contain moderate levels of TRP and other LNAA, modifies plasma concentrations of TRP and the other LNAA in a manner that results in no change in the competitive transport of TRP into brain. Consequently, brain TRP levels and 5-HT production are unchanged. Hence, a key feature of the LNAA transporter, its competitive nature, explains the impact of meals containing carbohydrates with or without protein on the production of a molecule important to normal brain function (5-HT).

Although there are many implications of the previously mentioned findings, one earlier hypothesis, based on the now erroneous idea that the ingestion of carbohydrate – but not protein – would raise brain TRP and stimulate 5-HT production, is no longer tenable. This hypothesis argued that a phenomenon termed 'carbohydrate craving' occurred in individuals who experienced no rise in brain TRP levels or 5-HT synthesis when ingesting carbohydrates (due to insulin resistance, which occurs in type-2 diabetics and in insulin-resistant obese subjects). Because increases in brain 5-HT release are known to suppress appetite, the argument was, such individuals would not experience the normal 5-HT-mediated suppression of hunger on eating carbohydrates and continue consuming them (thus 'carbohydrate craving'). Clearly, because the ingestion of even modest amounts of common dietary proteins is now known also to raise brain TRP and 5-HT (the α-lactalbumin in milk and also egg protein), a person seeking to raise brain 5-HT by ingesting carbohydrates could just as easily accomplish this by consuming these proteins. Thus carbohydrate craving, at least as envisioned in this earlier hypothesis, should not exist.

Chronic dietary effects are also observed. For example, chronic ingestion of diets containing proteins by rats with high proportions of one or more LNAA relative to TRP cause brain TRP and 5-HT concentrations to decline. The chronic ingestion of diets low in protein also causes the plasma concentrations of all LNAA to decline (including TRP), and brain TRP and 5-HT. In this case, brain TRP falls not because of a change in BBB competition but simply because the BBB uptake of all LNAA declines with falling plasma concentrations (the transporter becomes unsaturated, eliminating competition).

Other LNAA are neurotransmitter precursors in substrate-driven pathways in brain. Phenylalanine and tyrosine are substrates for catecholamine synthesis and histidine is the precursor of histamine. Like TRP, the brain concentrations of these amino acids are influenced by their competitive BBB uptakes from the circulation, and thus the diet. However, dietary effects for these amino acids are generally less noteworthy than for TRP.

The nonessential amino acid GLU is an excitatory neurotransmitter, causing neurons that express GLU receptors to depolarize. Because GLU is excitatory, responsive neurons can become overexcited, when subjected to prolonged GLU exposure, and die. The term 'excitotoxicity' was coined to describe this effect and led to the concern that GLU ingested in food (as a constituent of dietary proteins or as a flavoring agent) might cause the brain to become flooded with GLU, causing widespread neurotoxicity. The BBB acidic amino acid transporter prevents this from occurring: It primarily transports GLU out of the brain and not into it. Consequently, the BBB functions as a 'barrier' to GLU penetration from the blood.

Another mechanism also protects brain neurons from excessive exposure to GLU. Glial cells rapidly remove GLU from brain extracellular fluid and convert it to an electrically inert amino acid, glutamine. Although glial cells efficiently absorb neuronal GLU, they also readily clear any GLU that might stray into the brain from the circulation.

Fatty Acids and Choline

Fatty Acids

The brain uses fatty acids to synthesize the complex fat molecules that form neuronal and glial cell membranes. This process is more active in growing animals than in adults. The brain synthesizes some fatty acids from smaller molecules, but their uptake from the circulation is also an important source and is the only source for certain fatty acids (the essential fatty acids, which cannot be manufactured in the body). The details of the uptake process are not well understood.

From the nutritional perspective, diet influences essential fatty acid availability to the brain, with potentially important functional consequences. In almost all mammals, there are two essential fatty acids, linoleic acid and α-linolenic acid (termed polyunsaturated fatty acids (PUFAs)). In the nervous system (as elsewhere), linoleic and α-linolenic acid are incorporated into phospholipid molecules and inserted into cellular membranes, where they influence membrane fluidity and membrane-associated functions (e.g., the functionality of receptors and transporters). In addition, the linoleic acid in membrane lipids can be released and converted into arachadonic acid, a key precursor in the synthesis of prostaglandins and leukotrienes, which are families of important signaling molecules. α-Linolenic acid can be converted into docosahexanoic acid, a molecule found in very large amounts in the rods and cones of the retina and in the nerve terminal membranes of brain. Docosahexanoic acid is thought to be a key component of phototransduction and has been demonstrated to have important effects on vision. Dietary modifications in essential fatty acid intake might therefore be expected to influence membrane functions in brain, leading to alterations in brain function (as has been demonstrated for vision).

Choline

Choline occurs in the body as a constituent of lipid molecules in cell membranes, as a source of methyl groups, and as a precursor for the neurotransmitter acetylcholine (ACh). Choline is not an essential nutrient in humans, and deficiencies are rarely seen, because it is ubiquitous in the diet. However, in recent decades, dietary choline has been a focus of interest because of the possibility that changes in choline intake could influence neuronal ACh synthesis. ACh is a neurotransmitter; its synthesis and release by brain neurons is influenced by choline availability, which in turn can be altered by dietary choline intake, either in the form of free or fat-bound choline (phosphatidylcholine). In this context, oral choline and phosphatidylcholine have found some application in human diseases thought to involve ACh. For example, they have been used successfully to treat movement disorders such as tardive dyskinesia, a drug-induced muscular disorder in schizophrenic patients linked to low ACh function. However, they proved to be of little value in controlling abnormal muscle movements associated with Huntington's disease (also linked to low ACh function). Dietary choline and phosphatidylcholine supplements have also been studied as potential memory enhancers because CNS ACh neurons play an important role in memory. Patients with Alzheimer's disease have been most studied, but in general, the disappointing outcome has been that neither choline nor phosphatidylcholine has afforded much improvement in memory.

Vitamins

Neurons and glia have the same functional demands for vitamins as do other cells in the body. Their access to brain is thus an important consideration, particularly given the existence of the BBB. Water-soluble vitamins are transported across the BBB, and in some cases, the blood–CSF barrier, most often by non-energy-requiring carriers. After they are taken up into the neurons and

glial cells, most are rapidly converted into their biologically active derivatives, namely cofactors in enzyme-mediated reactions. Because cofactors are recycled, dietary deficiencies in one or another vitamin do not immediately lead to brain dysfunction, inasmuch as cofactor pools may take extended periods of time to become depleted. Although fat-soluble vitamins are lipid soluble, their passage through the BBB most likely involves more than simply diffusion.

Water-Soluble Vitamins

Folic acid is transported into brain as methylenetetrahydrofolic acid, the major form of folic acid in the circulation. It is then transported rapidly into the neurons and glia from the CSF or extracellular fluid. Once inside the cells, folates are polyglutamated. Methylenetetrahydrofolate is used by neurons and glia in reactions involving single-carbon groups, such as in the conversion of serine to glycine or homocysteine to methionine. Once methylenetetrahydrofolate is consumed in these reactions, folic acid is transported out of the brain into the circulation. Folate has become an issue of neurologic concern because of a link between folate deficiency and abnormal CNS development. The incidence of spina bifida, a serious spinal cord abnormality, rises above the population mean in the children of women who are folate deficient during pregnancy. Moreover, the incidence of spina bifida can be reduced by folic acid supplementation during pregnancy, beginning before conception. Initiating supplementation before conception is essential because the basic design of the CNS is laid down during the first trimester. At present, the mechanism(s) by which folic acid deficiency leads to the improper formation of the spinal cord is unknown. Folate deficiency may also be linked to depression in adults, and occasional studies suggest that folate supplementation can be a mood elevator in depressed patients. The mechanism(s) by which folate modifies mood is presently unknown.

Ascorbic acid (vitamin C) is actively transported into the brain extracellular fluid through the blood–CSF barrier, from which it is actively transported into the cells. Brain ascorbate pools show minimal fluctuations over a wide range of plasma ascorbate concentrations, which presumably explains the absence of CNS signs in ascorbate deficiency. To date, the only defined biochemical function of ascorbic acid in brain is as a cofactor for the enzyme that converts dopamine to norepinephrine (although ascorbate is thought by some to function as an antioxidant).

Thiamine (vitamin B_1) is taken up into brain by a BBB transporter; small amounts also gain entry via transport from blood into CSF. It is then transported into neurons and glia; conversion to thiamine pyrophosphate effectively traps the molecule within the cell. In nervous tissue, thiamine functions as a cofactor in important enzymes of energy metabolism. Severe thiamine deficiency in animals reduces thiamine pyrophosphate levels and the activities of thiamine-dependent reactions. It causes loss of the coordinated control of muscle movement; the exact biochemical mechanism is unsettled. The functional deficits are rapidly corrected with thiamine treatment, suggesting that neurons have not been damaged or destroyed. Thiamine deficiency in humans (beri-beri and Wernicke's disease) produces similar deficits in the control of muscle movements and also mental confusion. Korsakoff's syndrome, which occurs in almost all patients with Wernicke's disease, involves short-term memory loss and mental confusion. Severe thiamine deficiency in humans appears to produce neuronal degeneration in certain brain regions. The motor abnormalities can be corrected with thiamine treatment, but the memory dysfunction is not improved.

Riboflavin enters brain via a saturable BBB transport carrier. It is then transported into neurons and glia and trapped intracellularly by phosphorylation and converted to flavin adenine dinucleotide. Flavin adenine dinucleotide functions as a cofactor in carboxylation reactions. The brain contents of riboflavin and its derivatives are not notably altered in states of dietary riboflavin deficiency or excess.

Pantothenic acid is transported into brain by a BBB transport carrier. Neurons and glial cells take up pantothenic acid slowly by a mechanism of facilitated diffusion. Inside the cell, the vitamin becomes a component of coenzyme A, the coenzyme of acyl group transfer reactions. Relative to other tissues, the brain contains a high concentration of pantothenate, mostly in the form of coenzyme A. Brain coenzyme A concentrations are not depleted in pantothenate-deficiency states.

Niacin (vitamin B_3) is transported into brain as niacinamide, primarily via the BBB. Most niacin in brain is derived from the circulation, although brain may be able to synthesize small amounts. Niacin is taken up into neurons and glia and rapidly converted to nicotinamide adenine dinucleotide. The half-life of nicotinamide adenine dinucleotide in brain is considerably longer than in the other tissues. Nicotinamide adenine dinucleotide and nicotinamide adenine dinucleotide phosphate are involved in numerous oxidation–reduction reactions. Dietary niacin deficiency in the presence of a low intake of TRP causes pellagra in humans, a deficiency disease that includes mental depression and dementia, loss of motor coordination, and tremor. The mechanism(s) for these effects have not been identified.

Pyridoxine (vitamin B_6) is taken up into brain via a transport carrier that has not been well described. The vitamin can be transported in any of its nonphosphorylated forms (pyridoxine, pyridoxal, and pyridoxamine). Once within the brain extracellular fluid, the vitamin is readily transported into the neurons and glia and phosphorylated (primarily to pyridoxal phosphate or pyridoxine phosphate). Pyridoxal phosphate is a cofactor in a variety of neurotransmitter reactions, such as aromatic-L-amino acid decarboxylase (an enzyme of monoamine biosynthesis), glutamic acid decarboxylase (the enzyme of γ-amino butyric acid (GABA) synthesis), and GABA transaminase (the enzyme which catabolizes GABA). In humans, pyridoxine deficiency is rare because of its widespread occurrence in food. However, when identified, it has been associated with increased seizure activity, an effect dissipated by pyridoxine treatment. This effect may be linked to the production of GABA, an inhibitory neurotransmitter.

Biotin is transported into brain by a BBB carrier. It is a coenzyme for a variety of key carboxylation reactions in gluconeogenesis, fatty acid synthesis, and amino acid metabolism. Normally, biotin is recycled in cells during protein (enzyme) turnover, but not in brain; brain cells are thus more immediately dependent than the other cells on circulating biotin availability. Biotin deficiency is

rare; when it occurs, it can involve CNS symptoms (depression or sleepiness); the underlying basis for these effects is presently unknown.

Cobalamin (vitamin B_{12}) is thought to be transported into brain by a carrier-mediated mechanism. Little is known about this process or about the function of vitamin B_{12} in the nervous system. Vitamin B_{12} deficiency is associated with neurologic abnormalities, which are presumed to be derived from the demyelinization of CNS axons seen in advanced deficiency cases. These effects are reversed if vitamin B_{12} treatment is provided early enough; when it is left untreated, axonal degeneration occurs. Vitamin B_{12} may be important in neuronal repair mechanisms, which may become compromised in deficiency states. Nervous system damage associated with vitamin B_{12} deficiency can occur at any age.

Fat-Soluble Vitamins

Of the fat soluble vitamins, vitamin A (retinol) has been the most studied in relation to the CNS. The others have been considerably less examined, although vitamin E is currently of some interest because of its function as an antioxidant. The CNS is not thought to be a major focus of action for vitamins D and K, and thus little information is available regarding their roles in brain function.

The principal role of vitamin A in the CNS is as a component of the photoreceptive pigment of the eye, rhodopsin. In the blood, vitamin A circulates bound to the retinol-binding protein and transthyretin (prealbumin). Its transport into retinal cells occurs at the blood–retinal barrier (the retinal pigmented epithelial (RPE) cells), after the retinol–protein complex binds to retinol-binding protein receptors. Once bound, retinol is released into the RPE cell. The retinol-binding protein and transthyretin molecules are released back into the circulation. Inside the RPE cell, retinol binds to a specific protein and ultimately is esterified to a fatty acid. This molecule serves as the substrate for the conversion of retinol into the visually active form of the molecule, 11-*cis*-retinaldehyde, which then finds its way into the photoreceptor cell to be bound to opsin to form rhodopsin, the light-responsive pigment of the eye. When light strikes rhodopsin, phototransduction occurs and 11-*cis*-retinaldehyde is isomerized to all-*trans*-retinaldehyde, hydrolyzed from opsin, and released by the photoreceptor into the extracellular space (the opsin is retained and reused). The all-*trans*-retinaldehyde is shuttled into the RPE cell, where it is reconverted into 11-*cis*-retinaldehyde, and then recycled to the photoreceptor cells again to form rhodopsin.

From the nutritional perspective, retinal cells have an efficient system for managing and maintaining vitamin A pools. Hence, depletion of retinal vitamin A pools secondary to dietary deficiency only occurs over an extended time period. Deficiency appears functionally as 'night blindness,' as rhodopsin levels decline. Extended vitamin A deficiency leads to a loss of photoreceptor elements and eventually of the photoreceptor cells themselves. The cause of this cellular degeneration is not well understood.

Vitamin E is an antioxidant and a free radical scavenger that protects fatty acids in cellular membranes. It is transported in blood associated with lipoproteins. The mechanism of its transfer into nervous tissue is unknown. Dietary vitamin E deficiency is extremely rare in humans. It occurs in association with certain abnormalities of vitamin E transport and fat absorption and sometimes in individuals with protein-calorie malnutrition. The neurological manifestations are peripheral nerve degeneration, spinocerebellar ataxia, and retinopathy. Vitamin E has been proposed to play a role in a number of CNS diseases linked to oxidative damage. One example is Parkinson's disease, a movement disorder caused by the degeneration of certain groups of brain neurons. Evidence of oxidative damage is present in the brains of Parkinsonian patients, although controlled clinical trials of vitamin E supplementation have proved to be ineffective. Such negative findings question the likelihood of a vitamin E link to the etiology of the degenerative changes. A second example is Alzheimer's dementia, which is associated with a progressive, ultimately catastrophic degeneration of the brain. Several types of oxidative damage have been found in the brains of Alzheimer's patients, although it is presently unclear if this damage is the cause or effect. Vitamin E supplementation can slow the progression of Alzheimer's disease. However, such findings do not indicate if vagaries in vitamin E intake over an extended period of time are a cause of the disease.

Minerals

All of the essential minerals are important for cellular functions in brain, as they are elsewhere in the body. These are sodium, potassium, calcium, magnesium, iron, copper, zinc, manganese, cobalt, and molybdenum. Although most function as cofactors in enzymatic reactions, sodium and potassium are key ions in electrical conduction in neuronal membranes, calcium functions as a secondary messenger within neurons, and magnesium is an important component of certain neurotransmitter receptors. The diet normally provides more than adequate amounts of almost all minerals, except possibly for calcium, iron, magnesium, and zinc. The BBB permeability to most metals is quite low. For example, although the brain extracts 20–30% of the glucose in blood in a single capillary transit, it extracts <0.3% of any metal. The mechanisms of transport into brain for most metals are unknown. However, some details regarding the transport and/or functions of iron, calcium, and copper are available.

Iron circulates bound to a protein, transferrin. Iron uptake into brain occurs primarily at the BBB and involves a transferrin receptor-mediated endocytosis of the iron–transferrin complex by capillary endothelial cells. Iron dissociates from transferrin inside the cell and is delivered into the brain interstitial fluid; the transferrin is returned to the circulation. Brain iron associates with ferretin, a protein, and is stored intracellularly. The bulk of the iron–ferretin stored in brain resides in glial cells and is laid down early in postnatal life. Marked regional differences in iron and ferretin concentrations occur in brain; levels in some areas are as high as those in the liver. However, this distribution does not correlate with the density of transferrin receptors in brain capillaries; it is

presently unknown how or why the unequal distribution of iron develops. Numerous enzymes in brain are iron-requiring, including several hydroxylases involved in neurotransmitter production, and a key metabolic enzyme, monoamine oxidase.

Iron deficiency can cause impairments in attention and cognition in children. Similar effects are seen in animals. In iron-deficient rats, brain iron concentrations decline, with newborn and infant animals showing more rapid declines than older animals. Iron repletion in brain occurs in infant and adult rats with iron supplementation but not in animals depleted at birth. While outside of the brain, the activities of many iron-dependent enzymes are depressed by iron deficiency, whereas their activities are unaffected inside the brain. However, a reduction in certain dopamine receptors occurs, along with aberrations in dopamine-dependent behaviors (dopamine is a CNS neurotransmitter). The inability of brain iron stores to recover in rats made them iron deficient as newborns coincides with a persistence of dopamine-linked behavioral deficits, despite normal repletion of iron stores elsewhere in the body. Restoration of normal behavior with iron supplementation, along with brain iron stores, is seen in animals made iron deficient at other ages.

Iron deficiency also interferes with myelinization. Since marked glial proliferation and myelin formation occur early in infancy, iron deficiency during this period could prevent the optimal development of neuronal communications (glial cells provide insulation for axons and synapses). This effect could account for some of the behavioral deficits associated with neonatal iron deficiency.

Calcium is actively transported into the CNS, primarily via the blood–CSF barrier and is not sensitive to vitamin D. Because calcium concentrations in the circulation are regulated, under most circumstances, this process should also help to maintain brain calcium uptake and levels in the face of vagaries in calcium intake. Deficiencies in brain calcium should thus be a relatively rare occurrence.

Copper functions as a cofactor for numerous enzymes, including dopamine β-hydroxylase (DBH), which converts dopamine to norepinephrine. Dietary copper deficiency in humans is fairly rare. When produced in animals, it leads to reduced DBH activity in neurons and cells anywhere in the nervous system that synthesize norepinephrine. The mechanism of copper transport into the brain is presently unknown. Copper deficiency occurs as an X-linked genetic disease of copper transport in Menkes disease, in which tissue and brain copper levels become extremely low and produce neurodegeneration. Children with Menkes disease die at a very young age.

Further Reading

Choi, S., DiSilvio, B., Fernstrom, M.H., Fernstrom, J.D., 2009. Meal ingestion, amino acids and brain neurotransmitters: Effects of dietary protein source on serotonin and catecholamine synthesis rates. Physiology and Behavior 98, 156–162.

Davson, H., Zlokovic, B., Rakic, L., Segal, M.B., 1993. An Introduction to the Blood–Brain Barrier. CRC Press, Boca Raton, 335 pp.

Fernstrom, J.D., Garattini, S. (Eds.), 2000. International Symposium on Glutamate. Journal of Nutrition 130(4S): 891S–1079S.

Innis, S.M., 2008. Dietary omega 3 fatty acids and the developing brain. Brain Research 1237, 35–43.

Kamat, C.D., Gadal, S., Mhatre, M., Williamson, K.S., Pye, Q.N., Hensley, K., 2008. Antioxidants in central nervous system diseases: Preclinical promise and translational challenge. Journal of Alzheimers Disease 15, 473–493.

Moos, T., Nielsen, T.R., Skjørringe, T., Morgan, E.H., 2007. Iron trafficking inside the brain. Journal of Neurochemistry 103, 1730–1740.

Spector, R., 1989. Micronutrient homeostasis in mammalian brain and cerebrospinal fluid. Journal of Neurochemistry 53, 1667–1774.

Squire, L.R., Berg, D., Bloom, F.E., du Lac, S., Ghosh, A., Spitzer, N.C. (Eds.), 2008. Fundamental Neuroscience, 3rd edn. Academic Press, Burlington, MA. 1280 pp.

Cancer: Carcinogenic substances in food

A.I. Prieto, A.M. Troncoso, and A.M. Cameán, Departamento de Nutrición y Bromatología, Toxicología y Medicina Legal, Facultad de Farmacia, Universidad de Sevilla, Sevilla, Spain

© 2023 Elsevier Ltd. All rights reserved.

This is an update of D. Anderson, J.C. Phillips, Cancer: Carcinogenic Substances in Food, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 235–241, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00040-4.

Introduction 126
Naturally occurring carcinogens 126
Carcinogenic substances present in edible plants 126
Carcinogenic substances produced by fungi and microorganisms 127
Contaminants present in food: inorganic chemicals, organic chemicals and radioactive compounds 128
Carcinogens produced by food processing 129
Mechanisms of carcinogenicity 130
Carcinogenicity tests 131
Animal bioassays 131
Short-term predictive tests 132
Monitoring and control of hazards 133
Conclusion/summary/outlook 134
References 134
Further reading 135

Key points

- To present the main carcinogenic substances in food in relation to their origin: natural, contaminants, derived from food processing
- To highlight the importance of the mechanisms of carcinogenicity, including genotoxic and non-genotoxic substances.
- To describe the main *in vivo* assays for carcinogenicity evaluation
- To highlight the *in vitro* and *in vivo* short-term tests to predict carcinogenicity
- To emphasize that monitoring and control of cancer should be based on the risk analysis methodology

Glossary

Benchmark Dose (BMD) The minimum dose of a substance that produces a clear, low level health risk, usually in the range of 1–10% change in a specific toxic effect such as cancer induction

Bioassay Test of biological activity using living organisms

Carcinogen Agent capable of initiating the development of malignant changes leading to tumors

Carcinogenicity Cancer-causing property of a substance when an animal or human is exposed to it

Epigenetic Heritable change in the activity of a gene without affecting structure

Genotoxicity Processes that alter the structure, information content, or segregation of DNA including those that can cause DNA damage at the replication level

Hazard Potential of an agent to cause an adverse effect

Margin of Exposure (MOE) The ratio of two factors which assesses for a given population: the dose at which a small but measurable adverse effect is first observed and the level of exposure to the substance considered. It indicates if exposure should be considered of concern for public health

Mutation Heritable genetic changes that may be manifested at the phenotypic level or underlying DNA modifications (including specific base pair changes and chromosomal translocations)

Risk Probability that an (adverse) event will occur

Introduction

Carcinogenic substances in food may have different origins. They occur naturally in the physical environment and are found in a very large number of higher plants, fungi, and microorganisms, many of which are part of the human diet. The presence of carcinogenic environmental pollutants such as pesticides, industrial waste, drugs, etc. in food is significantly increasing due to human activity and industrialization worldwide. In addition, some carcinogens can be formed as a result of the traditional cooking, processing and preserving practices of food.

Since 1971, The International Agency for Research on Cancer (IARC) works in identifying the environmental chemicals that are carcinogenic hazards to humans classifying them into four groups: Group 1 (Carcinogenic to humans); Group 2A (Probably carcinogenic to humans); Group 2B (Possibly carcinogenic to humans) and Group 3 (Not classifiable as to its carcinogenicity to humans) (https://www.iarc.who.int/) (International Agency for Research on Cancer, 2021); Organization for Economic Co-operation and Development (OECD). In this article we focused on the compounds classified in Groups 1, 2A and 2B which may be present in foods, and they have been selected according to their nature: natural occurring, environmental contaminants, and produced by food preservation, processing and cooking.

Although carcinogens act through a wide variety of mechanisms, a substantial number have a common mechanism of action since they react with the genetic material of the body, DNA. These so-called genotoxic carcinogens generally require metabolic activation by the "host" animal to express their carcinogenicity. Although substantial efforts are being made to develop short-term, non-animal tests to predict the carcinogenicity of chemicals, animal bioassays remain the only reliable method for establishing the potential of a chemical to be a carcinogen (Anderson and Phillips, 2013).

Naturally occurring carcinogens

Naturally occurring substances identified as carcinogens in animals and humans present in foods may include natural components that are part of the food itself (plants) and natural compounds or toxins produced by fungi and microorganisms (**Table 1**). In addition, different chemical products (inorganic chemicals, organic chemicals and radioactive compounds) can be present in foods due to contamination (**Table 2**). They are present in the environment either as naturally occurring minerals or as a result of natural processes acting in the environment such as combustion, radioactive decay, or biodegradation of plant materials to oils.

Carcinogenic substances present in edible plants

Although the acute toxicity of many plant species has been known since written records first appeared, only comparatively recently has the carcinogenicity of plant-derived products been recognized. They are usually secondary metabolites, often present as part of the plant's natural defense mechanism against predation (i.e., natural pesticides), and are widespread in fruits, vegetables, herbs, and spices (Anderson and Phillips, 2013). However, the list of confirmed animal carcinogens present in plants is still relatively short, and only few of them are currently confirmed or suspected human carcinogens classified by IARC (**Table 1**).

Areca nut or betel nut is the seed of the tropical palm tree *Areca catechu* which is consumed in chewed form. Due to its stimulating and addictive properties, this natural product is used across the globe (over 600 million of people, WHO). Moreover, it is the primary ingredient in all betel quid preparations and is classified, together with areca nut, in the group 1 by IARC. Among their chemical constituents the alkaloids cause the addictive and carcinogenic effects (oral and pharynx cancer), therefore the consumption of areca nut in any form should be treated as a "neglected global public health emergency". Aristolochic acids are abundant in wild ginger and *Aristolochia* genus plants are widely used in traditional medicines for their therapeutic effects, although their intake has been directly linked to cancers of the urinary tract and hepatocellular carcinomas. On the other hand, the presence of aristolochic acids in edible parts of crops, originated from aristolochic acids-contaminated soil, has been demonstrated being considered as one of the major pathways by which humans become exposed to these compounds. Methoxsalen (8-methoxypsoralen) is a furanocoumarin found in a variety of natural products (celery, parsley, carrot, orange, lemon etc.). It has been used to treat epidermal proliferative diseases (psoriasis, vitiligo, eczema). The joint action of Methoxsalen plus ultraviolet A is required for the induction of carcinogenic effects.

Other natural compounds present in plants are classified in group 2B by the IARC. Amon them, cycasin (methylazoxymethanol-/3-glucoside), a colon carcinogen, is a naturally occurring compound that has been isolated from genus cycad plants, being its metabolite (methylazoxymethanol) formed by action of enzymes present in the intestinal microflora the active carcinogenic compound. Safrole, component of different species of plants such as *Ocimum basilicum*, *Piper nigrum* or *Myristica fragrans*, is mainly considered as liver carcinogen. Methyleugenol is widely distributed in plants, its intake is mainly through spicy foods and herbal beverages. Thus, methyleugenol-containing herbal beverages show an increasing consumption trend in countries like Indonesia; nonetheless it should be limited to about two weeks a year during a lifetime to be considered safe.

One of the first classes of naturally present toxic compounds identified in plants were the pyrrolizidine alkaloids from the genus *Senecio*. Subsequently, more than 650 related compounds have been isolated from numerous families and plant species, many of which are potent liver toxins and liver carcinogens affecting livestock, wildlife, and humans (Anderson and Phillips, 2013). In

Table 1 Naturally occurring carcinogenic substances classified according to IARC in plants, fungi and microorganisms.

Carcinogenic agent	*Classification group according to IARC*	*Food where it can be found*	*Producer organisms*
Plants			
Areca nut	1	Betel palm fruit by chewed	*Areca catechu*
Aristolochic acid	1	Traditional Chinese herbal therapy	*Aristolochia* and *Asarum* genus (wild ginger)
Betel quid with or without tobacco	1	Betel leaves and areca nut mixed with other aromatic substances	*Piper betle*
Cycasin	2B	Seeds	*Cycad circinalis* and *Cyca revoluta*
Dihydrosafrole	2B	Sassafras oil and root bark tea	*Sassafras albidum*
Ginkgo biloba extract	2B	Infusion of the leaves	*Ginkgo biloba*
8-Methoxypsoralen (plus ultraviolet A)	1	Celery, parsley, carrot, orange, lemon etc.	Apiaceae family
Methyleugenol	2B	Clove oil, basil, nutmeg, cinnamon, ginger etc.	*Syzygium aromaticum. Ocimum basilicum*, *Cinnamomum verum*, *Myristica fragrans*, *Zingiber officinale* etc.
Pyrrolizidine alkaloids: Lasiocarpine, monocrotaline and riddelliine	2B	Honey, tea, herbal infusions and food supplements	Asteraceae, Boraginaceae, Fabaceae, Orchidaceae and Apocynaceae families
Saffrole	2B	Sassafras oil and root bark tea	*Sassafras albidum*, Sassafras officinalis and *Cinnamomum camphora*
Fungi and microorganisms			
Aflatoxins (B1, B2, B2A, G1, G2 and G2A)	1	Cereals, oilseeds, fruits, vegetables, nuts, dried fruits, coffee beans, cocoa beans and spices	*Aspergillus* spp. and others
Aflatoxin M1	2B		*Aspergillus* spp. and others
Daunomycin	2B	Antineoplastic agents	*Streptomyces peucetius*
Fumonisin (B1, B2 and C)	2B	Cereals, oilseeds, fruits, vegetables, nuts, dried fruits, coffee beans, cocoa beans and spices	*Fusarium graminearum* and *Fusarium verticillioides*
Microcystin-LR	2B	Fish, molluscs and vegetables	Microcystis, Plankthotrix, Anabaena, Nostoc, Aphanizomenon, Anabaenopsis, Rivularia and Fisherella, among other genera
Mitomycin C	2B	Antitumor, antibiotic drugs	*Streptomyces* spp.
Ochratoxin A	2B	Cereals, oilseeds, fruits, vegetables, nuts, dried fruits, coffee beans, cocoa beans and spices	*Aspergillus ochraceus* and *Penicillium verrucosum*
Sterigmatocystin	2B	Cereals, animal feed, hard cheese, pecan nuts and green coffee beans	*Aspergillus* spp.
Streptozotocin	2B	Antibiotic drugs	*Streptomyces* spp.

particular, 1,2-unsaturated compounds are considered as genotoxic and carcinogenic substances due to their potential to undergo metabolic activation into reactive pyrroles. Among them, the IARC classified lasiocarpine, monocrotaline and riddelliine in the group 2B, while other pyrrolizidine alkaloids assessed were not classifiable (category 3) due to the limited information available. Humans can be exposed to pyrrolizidine alkaloids through herbal teas and milk honey, eggs or spices and food supplements; their presence in these foods, even at low levels, may be of concern for human health due to their hepatotoxic and genotoxic properties. Globally, most of these natural compounds are ingested through plant infusions. People perceive herbal beverages as safe although they contain carcinogenic compounds that may put consumers at risk due to over consumption of them. Further law enforcement to restrict the sale and use of products containing natural occurring chemicals is required to protect human health.

Carcinogenic substances produced by fungi and microorganisms

A number of naturally occurring organic chemicals or toxins produced by lower plants, such as fungi, and by microorganisms may contaminate and bioaccumulate in food (Table 1).

Mycotoxins are small-molecular natural products, toxic secondary metabolites mainly produced by filamentous fungi of *Aspergillus, Fusarium, Penicillium* and *Alternaria* genus. These toxins include aflatoxin B_1, B2, B2A, G1, G2 and G2A classified into the group 1 by IARC and are of the most potent carcinogens known. Moreover, with variable molecular structures, fumonisins, ochratoxin A and sterigmatocystin (a precursor of aflatoxin B1) (group 2B), are also of interest. Human exposure to such mycotoxins occurs

Table 2 Carcinogenic environmental contaminants classified according to IARC in foods.

Carcinogenic agent	*Classification group according to IARC*	*Food where it can be found*
Inorganic chemicals		
Arsenic	1	Vegetables, cereal, fish and seafood products
Beryllium	1	
Cadmium	1	
Chromium (VI) compounds	1	
Cobalt	2B	
Lead	2B	
Nitrate or nitrite	2A	
Nickel	2B	
Organic chemicals		
Dichlorvos	2B	Paddy, soybean, wheat, mustard, castor, groundnut, cucurbits and fish
Dieldrin	2A	
Dimethylarsinic acid	2B	Vegetables, fish and seafood products
Methylmercury compounds	2B	Fish and seafood products
Parathion	2B	Contaminated food
Polychlorinated biphenyls (dioxin-like)	1	Vegetables, meat etc.
Radioactive compounds		
Iodine-131	1	Vegetables and fish
Radium	1	
Radon-222	1	

mainly as a result of cereal crops and nuts stored in humid conditions (Anderson and Phillips, 2013). Food and feed contamination by mycotoxins is extremely common worldwide (about 25% of cereals according to the Food and Agriculture Organization of the United Nations, FAO) and it is very much related to the regional and climatic conditions. Recently, different studies have been focused on the occurrence of mycotoxins in different foods and feed showing that in Europe the content of fumonisins (mainly B1) is in general high, ochratoxin A is relatively low and aflatoxins are lowering. A continuous adjustment of the maximum legal maximum limits of these toxins together with the establishment of monitoring teams and improved agricultural production methods seems necessary to protect human health and food trade.

Microcystin-LR (MC-LR) is the most toxic and widespread microcystins (MCs), cyanotoxins produced by several genera of cyanobacteria. Among the structural multifarious variants of MCs (more than 246 MCs) (Hinojosa et al., 2019), only MC-LR is classified into the group 2B. The potential effects of this toxin on carcinogenesis include primary liver, colorectal and prostate carcinomas. MC-LR intake from contaminated water and food is a growing and an important source of human exposure is related to its bioaccumulation in various edible aquatic organisms, plants, and algae-based food supplements.

Contaminants present in food: inorganic chemicals, organic chemicals and radioactive compounds

Exposure to carcinogenic compounds through ingestion of contaminated food can be frequent. Many metallic contaminant elements are present in food, derived from a range of sources including the water used in food processing, soil residues, packaging, and cooking equipment. Several metals and some of their salts have been shown to be carcinogenic in animals and humans, particularly to the lungs (Anderson and Phillips, 2013). These include arsenic, beryllium, cadmium, chromium, cobalt, lead, and nickel which are classified into group 1 or 2B by IARC (see **Table 2**).

Methylmercury (CH_3Hg^+) is the most common organic component of Hg in the food chain. It is produced by direct chemical reaction or by the action of bacteria. To protect consumers against toxicity of CH_3Hg^+ and achieve the benefits of fish consumption, different international organizations and health authorities recommend that the consumption of fish/seafood species with a high content of mercury should be limited.

The formation of dimethyl arsenic acid (DMAA), commonly named cacodylic acid, occurs as a result of methylation of inorganic As by soil microorganisms, mainly from the genera *Aeromonas*, *Aspergillus*, *Escherichia* and *Saccharomyces*. The extensive use of As-derived herbicides during the 20th century led to a significant increase of anthropogenic DMAA acid into the environment, and severe accumulation of organic arsenic in soil and crops.

In relation to pesticides, dichlorvos (organochlorine insecticide) and dieldrin (organophosphorus pesticide) have been classified into group 2B by IARC. Dichlorvos is one of the most used pesticides due to its ample application in a high variety of crops (paddy, soybean, wheat, mustard, castor, groundnut, cucurbits). Moreover, the contamination of water bodies by dichlorvos has now become a global phenomenon. Along the food chain, pesticide residues are transferred from lower to higher trophic levels, and residues are also biomagnified in the process.

Polychlorinated biphenyls (PCBs) (Group 1, IARC) are a group of compounds of industrial origin present in plants, animals, and humans, and although huge efforts have been made to stop or limit their formation in developed countries, they are still widely spread and found due to their bioaccumulative nature and long-lasting existence mainly in industrialized regions. Much attention has been paid to research on human exposure risks posed by these compounds.

Regarding radioactive elements, some of them are also carcinogenic for humans, particularly to the lungs. These include iodine-131, radium, and radon gas and may act by damaging DNA directly or by increasing oxidative damage as a result of an increase in reactive radical species.

Carcinogens produced by food processing

A number of procedures used in food preparation/processing can introduce significant amounts of carcinogens into the food chain or the local environment. The most widely studied of these processes are the preservation of meats and fish by salting or smoking, grilling or broiling of meats, and cooking in vegetable oils (Anderson and Phillips, 2013). Moreover, frying, roasting, and baking foods rich in sugars, or even food fermentation may origin derived carcinogens as well.

Traditional methods for preserving meat and fish involve either salting or smoking. Epidemiological evidence has found an association between an increased incidence of cancer of the mouth and pharynx and the intake of salted meat and fish. It seems likely that a reaction between sodium nitrate or nitrite used for preserving the meat and alkylamides present in the meat results in the formation of N-nitrosamines, nitrosamides, and nitrosoguanidines. These compounds have been shown to be potent carcinogens in animal experiments to the mouth, pharynx, and other sites. Plenty of nitrosamines are found in different foods (**Table 3**), including N-nitrosodimethylamine (NDMA), N-nitrosodiethylamine (NDEA), etc. although not all of them show carcinogenic effects. Humans may be exposed to two different forms of Nitrosocompounds (NOCs): (i) exogenous, derived from different sources (e.g., tobacco products, diet, etc); (ii) endogenous nitrosamines and nitrosamides, generated by the reaction of nitrite with the products of amino acids' degradation in the stomach, accounting for up to 75% of the total NOC exposure (Cameán et al., 2006). Although dose levels required to induce tumor formation in animal studies are substantially higher than those likely to be ingested by humans, there is a concern that the presence of nitrosamines in food implies a significant risk. Several epidemiological and preclinical studies conducted on animal models reported a strong link between endogenous NOCs and colorectal cancer.

Preservation of meats and fish by smoking, and also their grilling can generate polycyclic aromatic hydrocarbons (PAHs). They are formed by incomplete combustion or pyrolysis and generally occur in complex mixtures. Both carcinogenic and noncarcinogenic compounds have been identified, being the PAHs of greatest concern those that are genotoxic and carcinogenic. For nonsmokers the major route of exposure to PAHs is food. The main PAH (over 100 identified) classified by the IARC is the benzo [a]pyrene (BaP) (Group 1). EFSA Panel on Contaminants in the Food Chain (CONTAM Panel) concluded that among PAHs, eight of them (**Table 2**) either individually or in a combination were the possible indicators of the carcinogenic potency of PAHs in food. The combination of two (BaP + chrysene = PAH2), four (PAH4) and eight PAH8 of these compounds represented around 34%, 60% and 80%, respectively of the amount of 15 PAHs found in food. BaP is not a suitable indicator of the occurrence of PAHs in food, being PAH4 and PAH8 the most adequate ones, with PAH8 not providing much added value compared to PAH4. Occurrence and toxicological data for individual PAHs, as well as oral carcinogenicity data with mixtures relevant for dietary exposure are needed for a more realistic risk assessment approach.

Heating oil during food cooking may generate a range of carcinogenic chemicals, including PAHs. However, many of the compounds produced are volatile and may therefore represent more of a risk to the cook than to the food consumer. The IARC considers that emissions from high temperature frying are classified in the 2A group (Anderson and Phillips, 2013).

In addition, practices such as frying or grilling of meats and fish have been found to generate significant quantities of heterocyclic nitrogenous compounds (HCAs), which contain methyl and/or amine groups in several positions. These heterocyclic amines, include more than 21 compounds, some of them are mentioned in **Table 2**. Their concentration in cooked food could oscillate considerably. They are some of the most potent bacterial mutagens known and have been shown to induce a wide range of tumors in animals and are classified in the group 2. They are particularly implicated in the induction of liver and gastrointestinal tract (GI) cancer (Caméan et al., 2006).

Moreover, acrylamide is formed in some carbohydrate rich foods, such as potatoes, grains, during high-temperature cooking, such as frying, roasting, and baking. Acrylamide is generated from natural reducing sugars and the amino acid asparagine in foods, and is not typically associated with meat, dairy, or seafood products. It is classified in the 2A group of IARC. Neurotoxicity, adverse effects on male reproduction, developmental toxicity and carcinogenicity were identified as possible critical endpoints for acrylamide from experimental animal studies, although the data from human studies were inadequate for dose-response assessment. Furan and methylfurans are formed in foods during thermal processing from several precursors such as ascorbic acid, amino acids, carbohydrates, unsaturated fatty acids and carotenoids, and are found in a variety of foods including coffee and canned and jarred foods. They induced neoplasic effects in animals, but only few data are available on the effect of furans in humans.

Sometimes, the toxic is formed during fermentation processes or during storage, such is the case of ethyl carbamate, that can occur naturally in fermented foods and beverages. It can be formed from various substances, including hydrogen cyanide, urea, citrulline, and other N-carbamyl compounds. The compound is a genotoxic classified in the group 2A and it could represent a health concern, particularly with respect to consumers of particular brands of stone fruit brandies.

Table 3 Carcinogens produced by food processing, following the IARC classification (https://monographs.iarc.who.int/agents-classified-by-the-iarc/).

Carcinogenic agent	*Classification group according to IARC*	*Food where it can be found*
Heterocyclic amines		
IQ (2-amino-3-methylimidazo[4,5-f]quinoline)	2A	Grilled fish and meat, meat extracts, bacon, fried or panfried meat, casein pyrolysates and products pyrolysis of proteins
MeIQ (2-amino-3,4-dimethylimidazo[4,5-f] quinoline)	2B	
Glu-P-1 (2-amino-6-methyldipyrido[1,2-a:3′,2′-d] imidazole)	2B	
Glu-P-2 ((2-aminodipyrido[1,2-a:3′,2′-d]imidazole)	2B	
Trp-P-1 (3-amino-1,4-dimethyl-5H-pyrido[4,3-b] indole)	2B	
Trp-P2 (2-aminodipyrido[1,2-a:3′,2-d]imidazole)	2B	
MeIQx (2-amino-3,8-dimethylimidazo[4,5-f] quinoxaline)	2B	
PhIp (2-amino-1-methyl-6-phenylimidazo[4,5-b] pyridine)	2B	
A-alpha-C (2-amino-9H-pyrido[2,3-b]indole	2B	
MeA-alpha-C (2-amino-3-methyl-9H-pyrido[2,3-b] indole)	2B	
N-Nitroso compounds (NOCs): N- Nitrosamines and nitrosamides		
NDMA (N-nitrosodimethylamine)	2A	Smoked fish and meat, ham, hot dogs, bacon, sausage, beef, etc.
NDEA (N-nitrosodiethylamine)	2A	
NDBA (N-nitrosodibutylamine)	2B	
NPIP (N-nitrosopiperidine)	2B	
NPYR (N-nitrosopyrrolidine)	2B	
NMOR (N-nitrosomorpholine)	2B	
NSAR (N- nitrososarcosine)	2B	
Polycyclic aromatic hydrocarbons (PAHS)		
Benzo[a]pyrene (BaP) (PAH2, PAH4, PAH8)	2A	Meats (barbequed, grilled, smoked), fish (preserved, processed, smoked), coffee, cereals, etc.
Chrysene (PAH2, PAH4, PAH8)	2B	
Benz[a]anthracene (PAH4, PAH8)	2A	
Benzo[b]fluoranthene (PAH4, PAH8)	2B	
Benzo[k]fluoranthene (PAH8)	2B	
Benzo[ghi]perylene (PAH8)	3	
Dibenz[a,h]anthracene (PAH8)	2A	
Indeno[1,2,3-cd]pyrene (PAH8)	2B	
Others		
Acrylamide	2A	Potatoes (especially French fries and potato chips), grain products (such as breakfast cereal, cookies, and toast), coffee
Ethyl carbamate (urethane)	2A	Spirits, wine, beer, bread, soy sauce and yoghurt
Furan and methylfurans (2-methylfuran, 3-methylfuran and 2,5-dimethylfuran)	2B	Ready-to-eat meals, toasted coffee, toasted bread

Globally, the scientific interest for derived toxicants in food is increasing in the last years in comparison to natural compounds and several international organizations have recommended different mitigation measures in order to prevent or reduce their formation.

Mechanisms of carcinogenicity

There are a wide range of mechanisms by which carcinogens can cause cancer in human and it is well established that cancer is a multi-step process. The genotoxic carcinogens, interact directly with DNA, resulting in a permanent heritable change to a cell following replication (i.e., an altered genotype). In contrast, nongenotoxic (so-called epigenetic) carcinogens do not interact directly with DNA but cause cancer by other mechanisms (Anderson and Phillips, 2013). Specifically, ten key characteristics of carcinogens have been described so far (Smith et al., 2016): they are electrophilic in nature or can be metabolically activated to an electrophile specie; they are genotoxic; they alter DNA repair or causes genomic instability; induce epigenetic alterations; induce oxidative stress; induce chronic inflammation; they could have immunosuppressive action; they modulate receptor-mediated effects; cause

immortalization and alter cell proliferation, cell death, or nutrient supply. These criteria have been introduced by IARC to facilitate systematic consideration of mechanistic evidence in carcinogens' evaluations.

Chemicals that react with DNA are invariably electrophiles, that target the nucleophilic (electron-rich) sites in the DNA. The electrophilic center may be present in the molecule itself (activation independent) as in β-propiolactone, dimethyl sulfate, and α,β-unsaturated aldehydes or be generated following metabolism (activation dependent) in the target species. Examples of classes of compounds that are converted to reactive electrophiles by oxidative metabolism include nitrosamines, chlorinated alkanes, hydrazines, and polycyclic aromatic hydrocarbons (PAHs) (Anderson and Phillips, 2013).

The mixed function oxidase system, cytochrome P-450, is the main enzymatic system involved in the activation of chemicals to carcinogenic species. This enzyme has a widespread distribution in the animal kingdom and it is known that different isoenzymes may activate the same compound in different species.

Most chemical carcinogens appear to be substrates of one particular isoenzyme called CYP1A1 thus including PAHs, aflatoxin, and 9-hydroxyellipticine, whereas the related isoenzyme CYP1A2 activates arylamines and amides such as 2-acetylaminofluorene and the cooked food mutagens. Other subfamilies include CYP2E1, which act on a wide range of small molecules, such as dialkylnitrosamines, urethane, vinyl monomers and haloalkanes, acrylamide, and CYP3A, which also activates PAHs, aflatoxins, and cooked food mutagens (Anderson and Phillips, 2013).

The chemistry of the activation process varies with the type of carcinogen. The oxidation of aflatoxin B_1, for example, results in the formation of the 8,9-epoxide in a single step whereas the activation of PAHs, such as benzo(a)pyrene, is a multi-step process involving an epoxide that is converted to a diol by epoxide hydrolase, which is then converted to the proximate carcinogenic species, a diol-epoxide. Activation of arylamines and amides to DNA reactive species, in contrast, frequently involves an initial oxidation step to an *N*-hydroxy derivative, which is then further metabolized to a highly reactive N—O-ester. This latter reaction is catalyzed by a transferase enzyme, usually sulfotransferase or acetyltransferase for arylamines and glucuronotransferase for arylamides. Other oxidative reactions result in the formation of unstable compounds that decompose spontaneously to the ultimate carcinogenic species. Thus, simple nitrosamines are oxidized by CYP2E1 to an α-hydroxy intermediate, which breaks down to the electrophilic alkyldiazonium ion (Anderson and Phillips, 2013). This isoenzyme CYP2E1 is involved in the formation of glycidamide (GA) from acrylamide by epoxidation, which is considered the route underlying the genotoxicity and carcinogenicity of acrylamide.

Enzyme systems other than the mixed function oxidase system may also be involved in the metabolic activation of carcinogens: prostaglandin H synthetase in the case of aflatoxins; for arylamines, oxidation may be carried out by prostaglandin peroxidase, myeloperoxidase, or by flavin-containing monooxygenases.

Regarding nongenotoxic or epigenetic carcinogens, there is not a common mechanism describing their action mode. One group of epigenetic carcinogens naturally present in many fruits produce renal tumors in the rat by binding to and preventing the degradation of a specific kidney protein, alpha-2-microglobulin, which requires a prior metabolic activation for carcinogenic activity. Similarly, a wide range of food contaminants induce liver tumors in rodents due to their ability to induce the proliferation of hepatic peroxisomes (phthalate diesters, hypolipidemic drugs, chlorinated herbicides). Other mechanisms of some nongenotoxic compounds are promotion (phorbol esters, barbiturates, chlorinated hydrocarbons), endocrine modulation (androgens and estrogens, antithyroid agents), immunosuppression (cyclosporine), tissue specific toxicity (metals such as arsenic and beryllium) and cytotoxicity (metal chelators and branched chain hydrocarbons) (Anderson and Phillips, 2013). Sometimes, several indirect mechanisms are involved in the carcinogenic mode of action of some toxicants, such is the case of furans, including epigenetic changes, oxidative damage to DNA and regenerative hyperplasia. The contributing factors in carcinogenesis are likely to vary according to dose, duration of exposure and degree of severity of liver cellular damage induced, inflammation and compensatory proliferation.

Carcinogenicity tests

Animal bioassays

As the mechanism of carcinogenesis in both humans and animals is not well understood, the only acceptable procedure for determining whether a chemical is likely to be a carcinogen is the examination of experimental animals exposed to the suspect material under carefully controlled conditions (Anderson and Phillips, 2013).

The procedures of these bioassays are conducted under rigorous conditions defined by the Code of Good Laboratory Practice (GLP). Although many regulatory authorities have guidelines for carcinogenicity evaluation, globally, the most commonly used *in vivo* carcinogenic studies follow the guidelines supported by the Organization for Economic Co-operation and Development (OECD), especially, the guideline OCDE 451 (2018). Alternatively, some international authorities (i.e., EFSA in the case of testing food additives) use a combined protocol to study chronic toxicity and carcinogenicity in the same experiment (OECD 453). The combined test provides greater efficiency in terms of time and cost compared to conducting two separate studies, without compromising the quality of the data in either the chronic phase or the carcinogenicity phase, and in this case the principles of dose should be carefully chosen.

The objectives of carcinogenicity studies (OECD 451) are: the identification of the carcinogenic properties of a chemical, resulting in an increased incidence of neoplasms, increased proportion of malignant neoplasms or a reduction in the time of appearance of neoplasms, compared with concurrent control groups; the identification of target organ(s) of carcinogenicity; the identification of

the time of appearance of neoplasms; characterization of the tumor dose-response relationship; identification of a no-observed-adverse-effect level (NOAEL) or point of departure for establishment of a Benchmark Dose (BMD); extrapolation of carcinogenic effects to low dose human exposure levels; provision of data to test hypotheses regarding mode of action.

The basic approach for carcinogenicity testing involves administering the test material to two suitable animal species for a considerable proportion of their natural lifespan. Rats and mice have been the experimental models of choice because of their relatively short life span, their widespread use in pharmacological and toxicological studies, their susceptibility to tumor induction, and the availability of sufficiently characterized strains. The study should preferably be carried out in animals from the same strain and source as those used in preliminary toxicity study(ies) of shorter duration.

To examine the carcinogenic potential of food components, the test substance is usually added in the diet, although in some circumstances administration may be in the drinking water or by gavage. As a minimum, 50 animals of each sex are allocated at random to each of the experimental groups. The duration of the study will be normally 24 months in the rat and 18 or 24 months in the mouse. During the study, the animals' clinical state is regularly monitored and at the end of the study a complete necropsy is performed on all surviving animals. Any tumors found are classified as either neoplastic or non-neoplastic and attempts are made to determine whether any of the observed tumors were the cause of (early) death of the animal (fatal tumors) or were unrelated to the death (incidental tumors).

The carcinogenicity bioassay aims to determine whether the administration of the test chemical has resulted in an increase in the incidence of tumors at one or more sites as compared with the normal background level. In order to accomplish this analysis, two major confounding factors need consideration. The first is the effect of differences in mortality rates between the control and treated groups and the second is the effect of differences in food intake and its consequence on body weight. Both factors can substantially alter the tumor pattern observed in different groups. Early deaths may prevent the animals reaching tumor-bearing age, and reduced food intake and the associated reduction in body weight may result in a considerable reduction in tumor incidence (Anderson and Phillips, 2013).

The interpretation of the results from a bioassay is complex but most authorities work to the "weight of evidence" principle. This evidence is taken in the light of the "adequacy" of the bioassay, which is dependent on some of the factors previously discussed. Rare or unusual tumors at a site would be given added weight. When data from bioassays are considered in human risk assessment, other factors must clearly also be taken into consideration such as evidence of genotoxicity in short-term tests and data on metabolism and potential human exposure. Furthermore, a measure of risk at doses substantially below the bioassay dose may be needed, and this may require an extrapolation using mathematical models (Anderson and Phillips, 2013).

In the event of a carcinogenic response being demonstrated in the study, additional mechanistic information together with good data on toxicokinetics are usually essential for risk assessment, both with respect to extrapolation to humans and possible determination of a threshold for non-genotoxic carcinogens (EFSA, 2012a).

Short-term predictive tests

A large number of test systems have been developed to detect damage to the genetic material of cells in an attempt to predict carcinogenic potential and thereby reduce the reliance on animal tests (**Table 4**).

In vitro assays for detecting genotoxicity include tests to detect gene mutation using bacterial or mammalian cells, and the so-called indicator tests that detect mechanistic changes associated with the formation of mutations, such as the binding of foreign molecules to the DNA bases (Anderson and Phillips, 2013). The most commonly used *in vitro* and *in vivo t*ests supported by the OCDE are shown in **Table 4**. Chromosome damage is considered in several tests and includes chromosome and chromatid aberrations. The *in vitro* micronucleus (MN) assay is a test for the detection of MN in the cytoplasm of interphase cells, and both aneugens and clastogens can be detected. Micronuclei represent a damage that has been transmitted to daughter cells, whereas chromosome aberrations scored in metaphase cells may not be transmitted (OECD 487). While the Mouse lymphoma assay (MLA) has been widely used for regulatory purposes, the TK6 has been used much less frequently. The alkaline comet assay detects DNA damage, mainly DNA breaks, in eukaryotic cells.

Since many of the cell systems used are unable to metabolically activate the majority of test chemicals, an exogenous mammalian metabolizing system, the so-called S-9 mix, is incorporated into the assay (Anderson and Phillips, 2013).

In vitro tests are backed up by short-term *in vivo* tests to confirm that the effects seen *in vitro* appear in the whole animal. These tests are mainly undertaken in rats, mice (ordinary and transgenic). A short-term *in vivo* assay measuring unscheduled DNA synthesis (UDS) in rat liver or gut is recommended by most regulatory authorities if there is a positive response in any *in vitro* assay and a negative response in an *in vivo* cytogenetics assay. Other test methods and end points are under consideration by regulatory authorities as indicators of genotoxic potential including the Comet assay (Anderson and Phillips, 2013). To fulfill animal welfare requirements, in particular the reduction in animal usage (3Rs—Reduction, Refinement, Replacement—principles), the *in vivo* Comet assay can also be integrated with other toxicological studies, e.g., repeated dose toxicity studies, or the endpoint can be combined with other genotoxicity endpoints such as the *in vivo* mammalian erythrocyte micronucleus assay (OECD 489).

During the last two decades extensive efforts have been directed to determine whether short-term tests are suitable for predicting carcinogenic potential, and comparative analysis and reports on their applicability are published.

Although many regulatory authorities have released guidelines for carcinogenicity evaluation, including short-term tests, they all still require animal studies as the ultimate test for carcinogenicity. However, the use made of short-term tests varies. In the US, the FDA recommends a battery of short-term tests for all 'additives' for which cumulative dietary intake is expected to exceed 1.5 μg per

Table 4 Short-term test systems for predicting carcinogenic potential.

Test system	*Cells used/experimental model*	*End point*	*OECD*
In vitro			
Bacterial mutation	*Salmonella typhimurium* TA strains *Escherichia coli* WP2	Reversion to histidine Independence	TG 471 (2020)
Mammalian cell gene mutation	Chinese hamster ovary (CHO), Chinese hamster lung (CHL) and Chinese Hamster lung (V79) lines cells, L5178Y mouse lymphoma cells, and TK6 human lymphoblastoid cells	Loss of Hprt[a], Xprt[a] mutations	TG 476 (2016)
	Mouse lymphoma with L5178Ycells (MLA assay)[c], human transformed lymphoblastoid cell line (TK6)	TK[b] gene mutations	TG 490 (2016)
Chromosomal damage	Chromosome aberrations (CA) *in vitro*: CHL, CHO, Chinese Hamster lung V79 cells, human peripheral blood lymphocytes (PBL)	Structural chromosome/chromatid aberrations (gaps, breaks, deletions)	TG 473 (2016)
	Micronucleus assay (MN) *in vitro* in a variety of cultures of cell lines or primary cell cultures, such as: Human peripheral blood lymphocytes (PBL) or human lymphoblastoid (TK6) cells	Micronuclei induction in the cytoplasm of interphase cells	TG 487 (2016)
Primary DNA damage	Comet assay (standard and enzyme-modified comet assay	Single and double strand breaks, oxidation bases of DNA	
In vivo			
Chromosomal damage	Chromosome aberrations (CA) Micronucleus assay (MN) in bone marrow erythrocytes	Micronucleus formation in erythrocytes sampled either in the bone marrow or peripheral blood cells of animals	TG 474 (2016)
DNA damage	Unscheduled DNA synthesis (UDS) in rat livers	Uptake of labeled nucleosides in cells that are not undergoing scheduled (S-phase) DNA synthesis	TG 486 (1997)
	- Alkaline Comet assay The assay is most often performed in rodents (many tissues, such as liver, stomach, etc), although it has been applied to other mammalian and non-mammalian species. - Enzyme-modified comet assay	Single and double strand breaks. The enzyme-modified comet assay could detect oxidation bases of DNA	TG 489 (2016)
Transgenic mutations in rodents (TGR)	Transgenic rodent somatic and germ cell gene mutation assay: LacZ bacteriophage mouse (MutaMouse); lacZ plasmid mouse; gpt delta (gpt and Spi−) mouse and rat; lacI mouse and rat (Big Blue®),etc	Point mutations, insertions, small deletions, large genome rearrangements	TG 488 (2020)

[a]Hypoxanthine-guanine phosphoribosyl transferase gene (Hprt in rodent cells, HPRT in human cells; collectively referred to HPRT test), and the xanthine-guanine phosphoribosyl transferase transgene (gpt) (referred to as the XPRT test).
[b]Thymidine kinase gene.
[c]MLA: Mouse lymphoma assay.

person per day to assist in the interpretation of animal feeding studies. The EFSA (2012a) recommended in the evaluations of food additives a battery of *in vitro* tests: a bacterial reverse mutation assay (OECD 471), and an *in vitro* mammalian cell micronucleus test (OECD 487). In the case of inconclusive, contradictory or equivocal results from *in vitro* testing, further testing *in vitro* will be conducted. In the case of positive results, it is appropriate to perform suitable *in vivo* testing (Micronucleus test, Comet assay, Transgenic rodent assay). Transgenic mouse models are a refinement, although not a complete replacement, to the rodent 2-year cancer bioassay, and may result in a significant reduction in the use of experimental animals (EFSA, 2012a). Some expert bodies, such as IARC, use short-term tests as an adjunct to animal carcinogenicity studies, thus giving added weight in the assessment process of likely human hazard to an animal carcinogen that is also positive in short-term tests (Anderson and Phillips, 2013).

Monitoring and control of hazards

Foods are complex mixtures of chemical compounds that contain not only nutrients but also many other substances among them toxic agents and specifically animal carcinogens, which are widely distributed in the environment. Therefore, the presence of carcinogenic compounds in foods while not desirable is likely to happen.

Monitoring and control of carcinogenic substances in foods is based on the risk analysis methodology, in which the first step is independent scientific risk assessment. As a result of risk assessment, it is necessary to punt into place the resulting measures as well as appropriate risk communication.

The identification and characterization of a carcinogenic hazard is mainly based on the reported animal carcinogenicity data and effects on human health from case reports and epidemiological studies. Other information, such as *in vitro* or *in silico* data, is increasingly being used to establish carcinogenic potential. These data should be assessed together with data on genotoxicity and any other relevant information to understanding the mode of action (MOA) by which the chemical compound causes cancer in humans or in experimental animals (COC, 2020). Dose-response assessment methodology is classically based on the determination of the relationship between the magnitude of exposure (dose) to an agent and the severity and/or frequency of the adverse health effects (response). Toxicological tests and experimental data are considered to derive health-based guidance values. But for the specific case of carcinogens (mainly those that are genotoxic in nature) any level can potentially cause adverse effects and thus no threshold for toxic effect or health-based guidance can be set. In these assumptions an alternative approach for risk assessment is the so called Margin of Exposure (MOE) approach, which supposes a qualitative description for a possible prioritization of risks based on the benchmark dose derived from dose response modeling (EFSA, 2012b), comparing animal dose-response data with human exposure. The MOE is in general more difficult to interpret in terms of health risk. The Benchmark Dose (BMD) is the dose corresponding to a specific change in effect over background (e.g., tumor formation). To perform mathematical modeling, sufficient experimental data are required together with the characteristics of the response. BMDL (Benchmark dose lower confidence limit) is then defined as the lower confidence limit of a point on the dose-response curve that presents an adverse effect, to consider uncertainty in the data. Hence a $BMDL_{10}$ give rise to a 10% of effect level and the BMDL cannot be considered as a safety reference value when dietary exposure is below it. MOE is then calculated by dividing $BMDL_{10}$ by the estimated dietary exposure. In summary, risk assessment based on MOE offers a relative indication of the level of health concern without quantifying the risk. In general, small MOEs indicate high concern, and large MOEs low concern. A MOE of 10,000 (based on a $BMDL_{10}$ from animal studies) supposes a low public health concern. At the same time, MOE should always be accompanied by an explanation on the background and the uncertainties in the reference point and exposure estimates.

On the other hand, it is well known that incorrect and unhealthy diet is one of the main causes of cancer. According to WHO, between 30 and 50% of the most common cancers could be prevented through healthy diet and an adequate level of physical activity. But it is a difficult task to identify specific dietary carcinogens due to the very complex nature of diets and the different epidemiological methods used. As a result, it is not very likely that epidemiological data link specific carcinogens present in food with carcinogenic outcomes. These compounds are usually present at very low concentrations in food and may only produce small increases in tumor incidence. But taking into account dietary habits, certain associations at the epidemiological level have been made as follow, between poor dietary fiber intake and colon cancer or excess of dietary fat and colon and breast cancer, even though specific agents involved have not been consistently identified (World Cancer Research Fund, 2018). Moreover, the exposure to both carcinogens and non-carcinogens is now being considered since some substances may only show carcinogenicity when present with other triggers (including life-style factors such as obesity) (COC, 2020).

Safe food is a topic of global concern that arouses interest among citizens and administrations due to the increasing international trade. The huge progress in analytical chemistry methods allows for the detection and quantification of contaminants present in foods at very low levels, among them carcinogens, helping to minimize risks and increasing consumers' protection through consistent risk management programs.

Although the current scientific approaches for assessing carcinogenic risk of food substances may seem adequate, there are still serious limitations. Hence, there is a need for proposal of methods to implement next generation assays together with new carcinogenic frameworks into food safety assessments.

Conclusion/summary/outlook

In summary, cancer prevention is needed because the global burden of cancer is high and continues to increase as a result of population growth and aging and upward trends in some exposures. In this context, the identification of carcinogens presents in food from different origins (natural, contaminants or produced during processing) is relevant as well as to conduct a risk assessment in order to protect public health.

References

Anderson, D., Phillips, J.C., 2013. Cancer: carcinogenic substances in food. In: Caballero, B., Allen, L., Prentice, A. (Eds.), Encyclopedia of Human Nutrition, third ed. Elsevier Ltd, Amsterdam, pp. 235–241.

Cameán, A.M., Jos, A., Moreno, I.M., Pichardo, S., Repetto, M., 2006. Tóxicos formados durante el procesado, preparación y almacenamiento de los alimentos. In: Cameán, A.M., Repetto, M. (Eds.), Toxicologia Alimentaria. Ed. Diaz de Santos, Madrid, pp. 493–516.

Committee on Carcinogenicity of Chemicals in Food, Consumer Products and the Environment (COC), 2020. COC Guidance Statement G01—Version 5.0. A Strategy for the Risk Assessment of Chemical Carcinogenicity. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/944877/G01_Strategy_for_the_RA_of_Chem_Carcinogenicity.pdf.

EFSA, 2012a. Guidance for submission for food additive evaluations. EFSA Panel on food additives and nutrient sources added to food (ANS). EFSA J. 10 (7), 2760.

EFSA, 2012b. Scientific Committee; Scientific opinion on the applicability of the margin of exposure approach for the safety assessment of impurities which are both genotoxic and carcinogenic in substances added to food/feed. EFSA J. 10 (3), 2578. https://doi.org/10.2903/j.efsa.2012.2578.
Hinojosa, M.G., Gutiérrez-Praena, D., Prieto, A.I., Guzmán-Guillén, R., Jos, A., Cameán, A.M., 2019. Neurotoxicity induced by microcystins and cylindrospermopsin: a review. Sci. Total Environ. 668, 547–565. https://doi.org/10.1016/j.scitotenv.2019.02.426.
International Agency for Research on Cancer, 2021. Agents Classified by the IARC. IARC, Lyon. Monograph Volumes No. 1–129. https://monographs.iarc.who.int/agents-classified-by-the-iarc/.
OECD, Guidelines for the Testing of Chemicals, Section 4 -Health Effects. OCDEiLibrary. https://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788. Last accessed (06 June 2022).
Smith, M.T., Guyton, K.Z., Gibbons, C.F., Fritz, J.M., Portier, C.J., Rusyn, I., DeMarini, D.M., Caldwell, J.C., Kavlock, R.J., Lambert, P.F., Hecht, S.S., Bucher, J.R., Stewart, B.W., Baan, R.A., Cogliano, V.J., Straif, K., 2016. Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. Environ. Health Perspect. 124, 6. https://doi.org/10.1289/ehp.1509912.
World Cancer Research Fund/American Institute for Cancer Research, 2018. Diet, Nutrition, Physical Activity, and Cancer: A Global Perspective. Continuous Update Project Expert Report.

Further reading

Ang, L.P., Wei Ng, P., Lean, Y.L., Kotra, V., Kifli, N., Goh, P.H., Lee, K.S., Sarker, M.R., Al-Worafi, Y.M., Ming, L.C., 2021. Herbal products containing aristolochic acids: a call to revisit the context of safety. J. Herb. Med. 28, 100447. https://doi.org/10.1016/j.hermed.2021.100447.
Chan, W., Pavlović, N.M., Li, W., Chan, C.-K., Liu, J., Deng, K., Wang, Y., Milosavljević, B., Kostić, E.N., 2016. Quantitation of aristolochic acids in corn, wheat grain, and soil samples collected in Serbia: identifying a novel exposure pathway in the etiology of balkan endemic nephropathy. J. Agric. Food Chem. 64, 5928–5934.
Díez-Quijada, L., Hercog, K., Stampar, M., Filipič, M., Cameán, A.M., Jos, A., Zegura, B., 2020. Genotoxic effects of cylindrospermopsin, microcystin-LR and their binary mixture in human hepatocellular carcinoma (HepG2) cell line. Toxins 12, 778. https://doi.org/10.3390/toxins12120778.
EFSA, 2007. Opinion of the scientific panel on contaminants in the food chain on a request from the European Commission on ethyl carbamate and hydrocyanic acid in food and beverages. EFSA J. 551, 1–44. https://doi.org/10.2903/j.efsa.2007.551.
EFSA, 2015a. Scientific opinion on acrylamide in food. EFSA J. 13 (6), 321. https://doi.org/10.2903/j.efsa.2015.4104, 4104.
EFSA, 2015b. Statement on the benefits of fish/seafood consumption compared to the risks of methylmercury in fish/seafood. EFSA J. 13 (1), 3982. https://doi.org/10.2903/j.efsa.2015.3982.
EFSA, 2017a. Risks for human health related to the presence of pyrrolizidine alkaloids in honey, tea, herbal infusions and food supplements. EFSA Panel on Contaminants in the Food Chain (CONTAM). EFSA J. 15 (7), 4908. https://doi.org/10.2903/j.efsa.2017.4908.
EFSA, 2017b. Risks for public health related to the presence of furan and methylfurans in food. EFSA J. 15 (10), 5005. https://doi.org/10.2903/j.efsa.2017.5005.
Gasecka, M., Drzewiecka, K., Magdziak, Z., Piechalak, A., Budka, A., Waliszewska, B., Szentner, K., Golinski, P., Niedzielski, P., Budzynska, S., Mleczek, M., 2021. Arsenic uptake, speciation and physiological response of tree species (*Acer pseudoplatanus*, *Betula pendula* and *Quercus robur*) treated with dimethylarsinic acid. Chemosphere 263, 127859. https://doi.org/10.1016/j.chemosphere.2020.127859.
Gupta, A.K., Tulsyan, S., Thakur, N., Sharmaa, V., Sinha, D.N., Mehrotra, R., 2020. Chemistry, metabolism and pharmacology of carcinogenic alkaloids present in areca nut and factors affecting their concentration. Regul. Toxicol. Pharmacol. 110, 104548. https://doi.org/10.1016/j.yrtph.2019.104548.
Gutiérrez-Praena, D., Jos, A., Pichardo, S., Moreno, I.M., Cameán, A.M., 2013. Presence and bioaccumulation of microcystins and cylindrospermopsin in food and the effectiveness of some cooking techniques at decreasing their concentrations: a review. Food Chem. Toxicol. 53, 139–152. https://doi.org/10.1016/j.fct.2012.10.062.
Ham, J.R., Choia, R.-Y., Yee, S.-T., Hwang, Y.-H., Kim, M.-J., Lee, M.-K., 2017. Methoxsalen supplementation attenuates bone loss and inflammatory response in ovariectomized mice. Chem. Biol. Interact. 278, 135–140. https://doi.org/10.1016/j.cbi.2017.10.014.
International Agency for Research on Cancer, 2019. IARC Monographs Preamble. Lyon: France. https://monographs.iarc.who.int/wp-content/uploads/2019/07/Preamble-2019.pdf.
Kalac, P., Kaltner, F., 2021. Pyrrolizidine alkaloids of European Senecio/Jacobaea species in forage and their carry-over to milk: a review. Anim. Feed Sci. Technol. 280, 115062. https://doi.org/10.1016/j.anifeedsci.2021.115062.
Kawai, K., Furukawa, H., Hirono, I., 1995. Genotoxic activity in vivo of the naturally occurring glucoside, cycasin, in the Drosophila wing spot test. Mutat. Res. 346, 145–149. https://doi.org/10.1016/0165-7992(95)90046-2.
Kim, S.K., 2020. Trophic transfer of organochlorine pesticides through food-chain in coastal marine ecosystem. Environ. Eng. Res. 25, 43–51. https://doi.org/10.4491/eer.2019.003.
Kirkland, D., Levy, D.D., LeBaron, M.J., Aardema, M.J., Beevers, C., Bhalli, J., et al., 2019. A comparison of transgenic rodent mutation and in vivo comet assay responses for 91 chemicals. Mutat. Res. Genet. Toxicol. Environ. Mutagen 839, 21–35.
Luo, S., Du, H., Kebede, H., Liu, Y., Xing, F., 2021. Contamination status of major mycotoxins in agricultural product and food stuff in Europe. Food Control 127, 108120.
Pan, C., Zhang, L., Meng, X., Qin, H., Xiang, Z., Gong, W., Luo, W., Li, D., Han, X., 2021. Chronic exposure to microcystin-LR increases the risk of prostate cancer and induces malignant transformation of human prostate epithelial cells. Chemosphere 263, 128295.
Spoof, L., Catherine, A., 2017. Appendix 3. In: Meriluto, J., Spoof, L., Codd, G. (Eds.), Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis, pp. 526–537. https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119068761.
Suparmi, S., Junico, G., Mariyama, S., Wesseling, S., Rietjensa, I.M.C.M., 2019. Levels of methyleugenol and eugenol in instant herbal beverages available on the Indonesian market and related risk assessment. Food Chem. Toxicol. 125, 467–478. https://doi.org/10.1016/j.fct.2019.02.001.
White, P.A., Luijten, M., Mishima, M., Cox, J.A., Hanna, J.N., Maertens, R.M., Zwart, E.P., 2019. In vitro mammalian cell mutation based on transgenic reporters: a report of the International Workshop on Genotoxicity Testing (IWGT). Mutat. Res. Genet. Toxicol. Environ. Mutagen 847, 403039. https://doi.org/10.1016/j.mrgentox.2019.04.002. https://www.wcrf.org/diet-and-cancer/.
Zaynab, M., Fatima, M., Sharif, Y., Sughra, K., Sajid, M., Ali Khan, K., Sneharani, A.H., Li, S., 2021. Health and environmental effects of silent killers organochlorine pesticides and polychlorinated biphenyl. J. King Saud Univ. Sci. 33, 101511. https://doi.org/10.1016/j.jksus.2021.101511.
Zhu, M., Yuan, Y., Yin, H., Guo, Z., Wei, X., Qi, X., Liu, H., Dang, Z., 2022. Environmental contamination and human exposure of polychlorinated biphenyls (PCBs) in China: a review. Sci. Total Environ. 805, 150270. https://doi.org/10.1016/j.scitotenv.2021.150270.

Cancer: Dietary management

Victoria Nelson and Louise Henry, The Royal Marsden NHS Foundation Trust, London, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of C. Shaw, Cancer: Dietary Management, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 242–246, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00039-8.

Nutrition cancer (introduction) 137
Guidelines 137
Malnutrition 137
Nutritional needs of specific patient groups 138
Children, teenagers, and young adults with cancer 138
Nutritional interventions 138
Practical management of eating difficulties 139
Anorexia (loss of appetite) 139
Taste changes (dysgeusia/ageusia) 139
Nausea and vomiting 139
Dysphagia (difficulty swallowing) 139
Mucositis and stomatitis 140
Xerostomia (dry mouth) 140
Gastrointestinal fistulas 141
Constipation 141
Diarrhea 141
Intestinal failure 142
Bowel obstruction 142
Palliative care 143
Living with and beyond cancer 143
Alternative and complementary diets 144
Conclusion 144
References 144
Further reading 145

Key points

- Nutrition has an important role to play throughout a person's cancer pathway
- Malnutrition and loss of muscle mass are common in those going through cancer treatment and negatively impact on clinical outcomes
- It is important to understand the role of diet, possible treatment side effects which may impact nutritional status and the role of nutrition interventions throughout a cancer pathway

Glossary

Anorexia Loss of appetite. In cancer patients it is frequently secondary to chemotherapy, pain medication, or tumor factors

Cachexia Severe weight loss (characterized by loss of lean body mass as well as fat mass) associated with disease states (catabolic state), more commonly seen in advanced cancer patients

Dysphagia Difficulty to swallow. May be related to tumor location or as a consequence of surgery in the esophagus, and as a temporary symptom secondary to Oro/pharyngeal/esophageal treatment related mucositis or infection

Enteral feeding Refers to specialized form of feeding in which nutrition (usually a special formula) is delivered directly into the stomach or intestine via a tube

Nutrition Support Dietary or nutrition interventions that focus on optimizing food/nutrient intake when a patient is suffering from reduced intake or weight/muscle loss. Interventions can include food fortification, using commercially available nutritional supplement sip feeds and other similar products as well as enteral or parenteral nutrition

Parenteral Nutrition (PN) (also referred to as Total parenteral nutrition -TPN) Intravenous nutrition support (in the form of glucose, lipid, amino acids, some minerals, trace elements and vitamins) provided bypassing the GI tract. Generally delivered by central venous access although can be provided via peripheral IV route but not usually via a cannula

Nutrition cancer (introduction)

Diet and nutrition are important considerations for most cancer patients. Weight loss, reduced food intake and poor appetite are common at presentation for some tumor types e.g., lung carcinoma or iatrogenic (as a consequence of treatment side effects) (Pinho et al., 2018). The impact of poor nutrition on clinical outcome is difficult to measure however it is widely observed that poorly nourished patients can struggle to tolerate treatment, require breaks in treatment, have prolonged hospital admissions and ultimately may not be able to receive their full course of treatment and therefore compromise efficacy of planned treatment (Mantzorou et al., 2017). Patients need excellent management of "nutrition impact symptoms" to optimize their nutritional status and hopefully improve or maintain a good quality of life (Kubrak et al., 2010; Omlin et al., 2013). Patients with advanced disease may need modified dietary advice and several patients will explore alternative or complimentary diets as a means of exerting some control over their diet and treatment in the hope of improving their prognosis.

As cancer treatment develops and improves, we are now seeing a significant proportion of patients surviving and needing to explore the optimal diet following treatment with the aim to promote long term survival and optimum quality of life (Aubrey et al., 2019).

Guidelines

Throughout this article we refer to ESPEN guidelines (Muscaritoli et al., 2021; Arends et al., 2016, 2017) which are practical guidelines based on current science on nutrition in cancer. ESPEN makes recommendations for various health care professionals and grades the level of evidence for the various recommendations. The recent guidelines contain easy to follow evidence-based flow diagrams (Muscaritoli et al., 2021) to aid clinical decision making for various cancer related nutritional consequences.

Malnutrition

Malnutrition can be defined as a state of poor nutrition; that can result from insufficient or excessive or unbalanced diet or from inability to absorb foods.

Within malnutrition there are separate terminologies to help differentiate between the various types of malnutrition which are defined in **Table 1** (Arends et al., 2017).

Sarcopenia & sarcopenic obesity have been shown to be associated with higher incidence of chemotherapy toxicity, shorter time to tumor progression, poorer outcomes of surgery, physical impairment, and shorter survival (Daly et al., 2018).

It is important that we screen for malnutrition regularly when someone is going through cancer treatment using a validated malnutrition screening tool to help identify those that would benefit from nutritional interventions (Shaw et al., 2014).

Once a patient is screened, they may be flagged as at risk of malnutrition or malnourished and hence referred for nutritional assessment. This assessment should include food intake, muscle mass, physical performance, and systemic inflammation. Muscle mass can be evaluated though dual X-ray absorptiometry (DEXA), computed tomography (CT) scans or bioimpedance analysis (BIA) but in practice is unlikely to be assessed due to time and financial constraints (De Las Peñas et al., 2019). Physical performance can be measured using dynamometry, gait speed and questionnaires, such as the Karnofsky performance status. Systemic inflammation can be estimated by serum C-reactive protein (CRP) and albumin (De Las Peñas et al., 2019).

As part of the nutritional assessment and intervention it is important that adequate nutritional requirements are calculated, and a plan put in place to meet these individual requirements. Total energy expenditure (TEE) may be predicted by assuming TEE to be 25–30 kcal/kg depending on the patient's performance status (Arends et al., 2016). However, these rough estimates are likely to be overestimating in obese and underestimating in severely malnourished patients. Protein intake should be above 1 g/kg/day and if possible, up to 1.5 g/kg/day (Arends et al., 2016). It is important that not only macronutrients, but micronutrients (vitamins and minerals needed by the body in small amounts) are considered to ensure nutritional adequacy.

Table 1 Terminology in cancer related malnutrition.

Anorexia	Associated with poor food intake which may be due to altered central nervous system appetite signals with the addition of symptoms from cancer treatment such as nausea or fatigue. It may also be due to physical limitations such as swallowing difficulties
Cachexia	Complex metabolic syndrome associated with underlying illness and is characterized by loss of muscle with or without loss of fat mass
Sarcopenia	Condition characterized by loss of skeletal muscle mass and function.

Nutritional needs of specific patient groups

Children, teenagers, and young adults with cancer

Cancer remains relatively rare in this patient group, however those patients with a cancer diagnosis often have complex nutritional needs. Nutritional requirements will need to encompass the additional energy, protein and micronutrient requirements required to support growth and development. A child's developing relationship with food can be disrupted by treatment side effects leading to increased levels of food aversions, restrictive eating, and behavioral eating difficulties which in turn lead to imbalanced diets. Many treatment regimens require intensive nutrition support and the wide use of enteral feeding tubes to support nutrition, hydration and the administration of medication while facilitating patients being at home. However, in contrast in the more common diagnosis in the patient group, Acute Lymphoblastic Leukemia, Lymphoma, and central nervous system tumors (brain tumors) the high level of steroid use in such treatment protocols leads to weight gain. This has now increasingly been identified as a risk factor for the long-term survival of patients as well as possible implications for health and wellbeing on treatment (Todatri et al., 2019).

Nutrients that need specific considerations in this group: energy, protein, calcium and Vitamin D.

Nutritional interventions

Many cancer patients will require some form of nutritional interventions during their illness. Nutrition Support will be required by patients who experience an eating difficulty or weight loss. The first course of action is to assess their oral intake: if patients can eat, then they should be given appropriate advice to maximize their food and fluid intake (**Table 2**).

The measures in **Table 2** should be considered to help prevent weight loss or encourage weight gain. It must be remembered that energy and protein requirements may be elevated due to the physiological effects of malignancy.

If patients are unable to achieve adequate nourishment by mouth to maintain their weight and nutritional status an enteral tube feed should be considered. The type of tube placed will depend on the factors outlined in **Box 1**.

Box 1 Factors for considerations when choosing enteral feeding route

(1) The anticipated length of time the feed will be required -e.g., NG tubes are generally designed for short term feeding -if likely to be needed for more than 6 weeks, consider whether longer term feeding tube solution is feasible and beneficial to the patient
(2) The physical state of the patient; for example, a nasogastric tube or percutaneous endoscopically placed gastrostomy tube may not be suitable for patients with complete esophageal obstruction. A jejunostomy tube may be preferred following upper gastrointestinal tract surgery. Timing of chemotherapy may preclude the patient from having an invasive procedure due to limited time for wound healing
(3) The wishes of the patient concerning the physical appearance of different tubes and the invasiveness of the procedure required to place them.
(4) Practicalities around placement of the tube in a timely manner

Numerous types of commercially produced enteral feeds are available. Most cancer patients will require complete, whole protein feeds providing 4–6 kJ ml^{-1} (1–1.5 kcal ml^{-1}) or if experiencing mucositis/radiation enteritis may benefit from a peptide-based formulation. Only in cases of severe malabsorption, gastrointestinal fistula, or pancreatic insufficiency should an elemental or low-fat feeds be necessary. The choice of feeding regimen will depend on the patient's oral intake, their mobility and care needs a and on the volume of feed tolerated. It may be administered via a feeding pump (for feeding during the day and/or overnight), gravity feeding, and bolus feeding.

PN is required where the gastrointestinal tract cannot be used, such as in patients with complete bowel obstruction or severe malabsorption. It should not be used as a first line of nutrition support when the enteral route is functioning and available. The relative risks and benefits of PN should be weighed up when planning nutrition care.

Table 2 Dietary advice to help prevent weight loss.

Fortify food with cream, butter, milk powder, cheese, vegetable oils, ground nuts or seeds, honey, etc.
Have small, frequent snacks.
Use full-fat and full-sugar products.
Avoid large amounts of lower energy foods (e.g., fruit and vegetables).
Try proprietary supplements, such as milky or juice drinks.
Altering the portion sizes, consistency or timing of food or drink

Table 3 Methods of nutritional support when oral intake is inadequate.

Method	*Route*
Enteral tube feeding	Nasogastric (NG) or nasojejunal tube (NJ)
	Percutaneous endoscopically guided Gastrostomy (PEG) surgically placed gastrostomy
	Low profile "button" style gastrostomy (long term)
	Radiologically inserted gastrostomy (RIG)
	Percutaneous gastrostomy with a jejunal extension
	Jejunostomy
Parenteral nutrition (PN)	Central line
	Peripheral line

Table 4 Dietary management of anorexia.

- Small frequent meals and snacks, served on a small plate
- Provide familiar and preferred foods
- Fortify foods to maximize energy and protein content and avoid "filling up" with low nutrient dense foods e.g., plain salad, vegetables, tea, and coffee
- Separate drinks from meals if patient reports early satiety
- Try to eat in a pleasant environment e.g., out of bed
- Explore options for psychological support

Practical management of eating difficulties

Anorexia (loss of appetite)

Anorexia (loss of appetite) is often associated with other eating difficulties, such as nausea and taste changes. Addressing these problems may improve the patient's appetite. Anorexia may occur as a side effect of medication such as chemotherapy, antimicrobial agents as well as pain. Optimizing symptom management e.g., pain management through effective analgesia, is essential in providing an appropriate diet (Table 4). For patients who have severe anorexia, an appetite stimulant such as corticosteroids, progestins, androgens, cannabinoids or alcohol may be considered, however evidence of efficacy is limited and most come with significant side-effects (Turcott et al., 2021).

Taste changes (dysgeusia/ageusia)

Cancer patients may suffer from lack of taste or abnormal taste. This can severely reduce the desire for food and is usually associated with chemotherapy agents, some antimicrobials, oral candida and other infections, radiotherapy to the oropharyngeal area and a dry mouth (xerostomia). For most cancer patients, the taste changes are temporary. They may find that foods taste metallic, bitter or excessively salty or sweet or they develop a strong preference for a certain type of flavor. Depending on the taste change experienced, it is often worth excluding certain foods from the diet or using certain flavorings to try to ameliorate abnormal taste (Table 5). If food groups are excluded, the nutritional quality of the diet should be assessed by an appropriately trained nutrition health care professional. Patients should be encouraged to perform regular mouth care.

Nausea and vomiting

Often related to medication e.g., chemotherapy agents, oral potassium supplements, opiates but can also be associated with thickened secretions, bowel obstruction, severe constipation. Etiology of the symptom should be determined and for the majority of patients first line treatment should be administration of antiemetic drugs. Some dietary suggestions may help patients with food choice when they are feeling nauseous (Table 6).

Dysphagia (difficulty swallowing)

Dysphagia may occur because of obstruction from tumors, mucositis, or neurological changes. It may occur with solid food, semisolid foods, such as porridge, or liquids. For the person who cannot manage solid food but is able to eat semisolids, altering the consistency of the food may be the only dietary change needed for example encouraging food with extra sauce, soft puddings, and

Table 5 Suggestions for overcoming taste changes.

Taste change	*Suggestions*
Excessively sweet	Reduce sugar content of food and drink
	Add salt to drinks and puddings
Excessively salty	Avoid packet soups, gravy, and sauces
	Avoid salted snacks (e.g., crisps and nuts) or try unsalted varieties
	Avoid bacon and other cured or tinned meat
	Add a pinch of sugar to sauces or soups
Metallic taste	Marinate meats (e.g., vinegar and wine) or consider other protein sources such as white meat, fish, eggs, and cheese, nuts, beans, and pulses
	Avoid tea, coffee, and chocolate
	Trial use nonmetallic cutlery
Taste blindness	Use extra flavorings: Salt, pepper, pickles, mustard, herbs, and spices.
	Eat highly flavored food (e.g., curry)
	Encourage experimentation with the odor of food
	General: try sucking boiled sweets and hard candies or chewing gum

Table 6 Suggestions for food and fluids when person has nausea.

Have cold food and drink in preference to hot because these have less odor
Sip fizzy drinks or flat fizzy drinks
Try drinking through a straw (to minimize exposure to odor)
Try ginger flavors (e.g., ginger ale and ginger biscuits).
Eat small, frequent snacks to avoid the stomach from becoming completely empty-carbohydrate foods are often preferred
Reduce greasy or very fatty foods as they may delay gastric emptying

nourishing drinks. For the patient who is only able to swallow fluids, close attention must be paid to their intake and dietary supplements are likely to be necessary. Some patients who can only manage liquids choose to liquidize their food; this dilutes the nutrients, so meals should be fortified to add energy and protein by adding butter, cream, honey, cheese, nut butters or beans. If there is complete dysphagia to both solids and liquids, feeding by an enteral tube should be considered (see **Table 3**). In some instances, people can swallow solid food but aspirate liquids. Patients should undergo a complete assessment from speech and language therapist/pathologist (SLT) to ascertain which textures are safe to swallow if any. A SLT may assess a patients swallow as unsafe and place them nil by mouth, therefore to maintain hydration and nutrition alternative feeding routes need to be explored through provision of an enteral tube.

Mucositis and stomatitis

Mucositis and stomatitis arise as a consequence of treatment such as high dose chemotherapy, radiotherapy (including total body irradiation) or infection.

If the mouth or throat is sore it can limit the physical ability to eat by affecting ability to chew and swallow. It can also affect taste. Analgesia taken before eating and drinking can help ease the pain and enable the person to eat a little more.

Modifying the diet is also helpful (**Table 7**).

Xerostomia (dry mouth)

Xerostomia (dry mouth) may be a long-term side effect of cancer treatment such as oropharyngeal radiotherapy or surgery affecting the salivary glands. Opiates can also cause a dry mouth.

Table 7 Suggestions to relieve mucositis and stomatitis.

Avoid citrus fruits and drinks.
Avoid salty, spicy food, vinegar, pickles, and other irritants and rough textured foods.
Carbonated fizzy drinks can be irritating
Experiment with the temperature of foods and drink. Tepid food and drinks can be soothing while others find iced drinks soothing
Avoid dry foods that need extra chewing (e.g., toast).
Eat soft food and use extra sauce.
Use proprietary nutritional supplements or milkshakes
Consider artificial nutrition support if adult patients are likely to have severe mucositis for >5 days.

Table 8 Dietary advice for Xerostomia.

- Add extra sauces to foods
- Have sips of drinks between mouthfuls of food
- Suck on ice cubes or popsicles
- Try chewing sugar free gum (although this may be unhelpful if there is no saliva production)
- Avoid dry foods such as bread as these form a dry bolus when chewed and can cause gagging or retching.
- Avoid alcohol and any products containing alcohol such as some mouthwashes as this can have a "drying" effect
- Try sucking tinned fruit, fresh pineapple, acidic sweets, and candy

Good dental hygiene is particularly important because saliva protects the mouth against infection and tooth decay. Patients should be talked through various saliva sprays and mouth gels until they find one suitable for them. Xerostomia can affect appetite by causing taste changes (Table 8).

Gastrointestinal fistulas

The etiology of a fistula must be considered as this determines the management options. A fistula can develop as a side effect of some cancer treatments including radiotherapy and surgery or be surgically formed (stoma) anywhere in the gastrointestinal tract. The site and whether it is an internal or enterocutaneous fistula will determine the dietary management (Table 9). Internal fistulas usually require the approach of nil by mouth, placing an enteral feeding tube lower than the fistula or commencing PN to help the healing of the fistula.

Constipation

The etiology of constipation must be considered when formulating a nutrition care plan. If it is due to a tumor pressing on the bowel (e.g., cancer of the ovary or colon), a low-fiber diet may be helpful. Low-fiber food is less bulky and may pass through the bowel more easily, particularly if accompanied by appropriate laxatives (i.e., stool softener).

If constipation is due to dietary imbalance and lack of fiber in the diet, then an increase in fiber and fluid intake will be helpful. Where constipation is due to analgesia, then appropriate laxatives need to be used in conjunction with any changes in the diet. In addition to fiber, a good fluid intake must be maintained, approximately 2 L per day is recommended.

Diarrhea

Before formulating a nutrition plan for diarrhea, it is essential to consider the etiology of the symptom. Diarrhea can be "overflow" secondary to severe constipation and needs to be managed with an intensive laxative regimen.

Table 9 Sites of fistulas and their management.

Fistula	*Site*	*Management*
Enterocutaneous	Neck, salivary fistula	"Nil by mouth" and enteral tube feed until healed
Internal	Chyle leak (e.g., in neck)	Low-fat diet initially; if unsuccessful, a low-fat, medium-chain triglyceride enteral tube feed If unsuccessful, consider parenteral nutrition
Enterocutaneous	Large bowel	Low-residue diet or elemental enteral tube feed
Enterocutaneous	Small bowel	See Table 10

There are several causes of diarrhea which include bowel disease, medications, or infection. Once the cause is investigated antibiotics, antivirals and anti-diarrheal agents may be prescribed. Patients may be advised to avoid excessive intakes of high-fiber foods, which can increase bowel transit time. If the cause for diarrhea is considered to be malabsorption, the patient should also be investigated for signs of pancreatic insufficiency, bile salt malabsorption or small intestine bacterial overgrowth (Andreyev et al., 2021).

When diarrhea is severe, it is important to replace the fluid lost to prevent dehydration, oral rehydration sachets are useful to replace fluid loses. Diarrhea caused by pelvic area radiotherapy needs to be controlled with drugs, and a low-fiber diet is not thought to be helpful in this instance (Wedlake et al., 2017).

Intestinal failure

A long-term side effect of pelvic radiotherapy may be enteritis resulting in intestinal failure. Extensive gastrointestinal surgery leaving less than 100 cm of small bowel, or an enterocutaneous fistula in the small bowel causing high stoma losses, may also cause intestinal failure. Previous chemotherapy that may affect the function of the bowel can contribute to this condition. Intestinal failure is more likely to occur when the patient does not have a functioning colon (e.g., in the case of ileostomists or when the ileo–cecal valve is absent). Dietary manipulation can greatly alleviate the symptoms of intestinal failure, such as thirst, dehydration, and high stoma losses (>1.5 L/24 h) or large volumes of diarrhea (**Table 10**).

An oral rehydration solution consisting of 20 g glucose, 3.5 g sodium chloride, 2.5 g sodium bicarbonate, and 1000 mL water provides 90 mmol of sodium per liter. It may be used chilled and to dilute weak fruit squashes/cordials. If the patient remains dehydrated despite following the advice detailed in **Table 10**, intravenous fluid replacement is necessary. Medication may be given to decrease gut transit time or reduce fluid losses. If medication is in the form of capsules, these should be opened, and the drugs given 60 min before meals. Suitable drugs include those that slow transit such as codeine phosphate, loperamide and those medications that reduce secretions for example omeprazole, and octreotide. In the longer term, due to the impact on micronutrient absorption the following should be monitored: plasma electrolytes, ferritin, vitamin D, serum albumin, magnesium, zinc, calcium, phosphate, alkaline phosphate, folate and vitamin B_{12} concentrations, prothrombin time, body weight, and urinary sodium concentration.

Bowel obstruction

Bowel obstruction may be partial or complete. In cases of complete bowel obstruction, the clinical condition of the patient must be considered. If it is anticipated that the obstruction will resolve, or if aggressive treatment such as surgery is planned, PN support

Table 10 Dietary management to reduce gut losses in intestinal failure.

Restrict fluids to 500–1000 mL daily, increasing to 1500 mL.
Avoid drinks for 30 min before and 45 min after meals.
Avoid foods that are particularly high in fiber.
Sprinkle salt liberally on food.
Consider fat restriction if patient has a colon and there is evidence of steatorrhea.
Take salt and carbohydrate foods together to help sodium absorption.
If gut losses are 1000 mL or more, part or all of fluid intake should consist of an oral rehydration solution.

should be considered. PN may not always be appropriate and needs a multidisciplinary approach to discuss the aims before commencing (Muscaritoli et al., 2021).

Depending on the degree of obstruction, in cases of subacute/partial obstruction, the following action may be taken under medical supervision:

- First day: sips of clear fluid, approximately 10 mL h^{-1}
- Second day: 30 mL h^{-1} clear fluid
- Third day: 60 mL h^{-1} clear fluid
- Fourth day: free clear fluids
- Fifth day: free fluids, including milk, low-fiber soup, custard, and jelly
- Sixth day: low-fiber diet, avoiding all fruit and vegetables, nuts, pulses, and whole grain cereals, whole meal bread, etc.

A patient who starts to vomit should return to the diet prescribed for the preceding day. If symptoms of bowel obstruction, such as abdominal pain and indigestion, remain controlled, fruit and vegetables may be introduced as tolerated, starting with small amounts.

Palliative care

In some patients, cancer will not be cured which in turn can mean they may have a short prognosis where there are no longer any treatment options, or they may go through palliative treatment and live many years with cancer. Palliative care focuses on the relief of symptoms rather than aggressive curative treatment. Most people receiving palliative care will suffer from at least one eating difficulty. Much of the advice detailed previously for overcoming dietary problems is relevant but depending on the individual's prognosis they may have varying nutrition priorities (Muscaritoli et al., 2021). If patients are unconcerned about their poor dietary intake, it may be appropriate not to offer any advice; conversely, for those who are very concerned, the problem should be addressed seriously, and it might be that enteral tube feeding and PN are appropriate in individual cases.

Living with and beyond cancer

Living with and beyond cancer covers a wide variety of circumstances, including people at diagnosis, during and post-treatment, and those with recurrence, therefore including people with potentially differing nutritional needs. People can be struggling with a poor nutritional intake because of treatment side effects or may need to optimize their nutritional status preoperatively or require advice post-treatment for optimum recovery and rehabilitation. People often want to change their dietary intake after cancer, and actively seek information, therefore it is important to provide evidence-based information appropriate for the various time points in an individual's cancer journey.

Prehabilitation can help people with cancer prepare for treatment through promoting healthy behaviors by addressing exercise, nutrition, and psychological needs. The aim of prehabilitation is often to empower individuals to maximize resilience to treatment and improve long term health. Providing advice on increasing physical activity and tailored nutritional advice can help manage individuals' symptoms, prevent, or reduce treatment side effects, improve quality of life and clinical outcomes, and reduce risk of recurrence (Veen et al., 2019).

Table 11 Lifestyle recommendations for those living with and beyond cancer.

• Be a healthy weight, aiming to keep weight within a healthy body mass index (BMI 18.5–25 kg/m^2)
• Be physically active, aiming for 150 min of moderate exercise or 75 min of vigorous activity a week
• Eat a diet rich in wholegrains, vegetables, fruit, and beans, aiming to provide 30 g of fiber per day
• Limit consumption of "fast food's" and other processed foods high in fat, starches, or sugars
• Limit consumption of red meat to three portions a week (350–500 g cooked red meat) and consume very little processed meat
• Limit consumption of sugar sweetened drinks
• Limit alcohol consumption, ideally to no alcohol but if consumed within the recommendations of 14 units a week

For those living with and beyond cancer, nutrition advice should be tailored to individual needs depending on symptoms as discussed throughout this article. Gastrointestinal symptoms have been reported as the most common of the chronic physical side effects of cancer treatment and may require input from a gastroenterology team (Andreyev et al., 2014). General advice can be provided where appropriate, **Table 11** has a list of recommendations that can be suggested to people living with and beyond cancer.

Alternative and complementary diets

With increasing use of the internet for nutrition information, there has been an increase in patients wanting to explore specific diets for their purported anti-tumor effect. Such diets may be followed alongside (complementary) or in place of (alternative) conventional treatment. Often the diets have not been tested or demonstrated to be effective in scientifically rigorous clinical trials with many relying on pre-clinical or in vitro studies alone.

Dietary regimens may share common features: they are mainly vegetarian or vegan or macrobiotic, promote organic food, raw foods and juices and limit unprocessed foods, sugar, and fat. Such diets often promote the use of high doses of vitamins and minerals.

Much interest has been expressed in the therapeutic role of micronutrients, but they can potentially lead to drug-nutrient interactions and affect treatment e.g., antioxidant could reduce efficacy of free-radicle forming chemotherapy agents or even promote tumor growth as was seen in a study using β-carotene supplementation in patients with lung cancer (Alpha-tocopherol Beta Carotene Cancer Prevention Study Group, 1994).

Proponents of such diets tend to advocate these diets based on research into the causes of cancer but showing that a particular food group may be beneficial in preventing cancer formation e.g., high fruit and vegetable intake may help in reducing the risk of colon cancer, but this is not to say that such a diet can conversely assist in curing colon cancer. The popularity of these different diets varies over time with current trends toward ketogenic and alkaline diets.

Nutritional inadequacies may arise particularly if the patient has a poor appetite. The diets may cause weight loss and are restrictive, and time-consuming to prepare. Some ingredients may be difficult to obtain and are often costly.

Studies appear to show no difference in survival rates between patients following complementary therapies and patients receiving conventional treatment alone. Patients who use complementary therapies, however, do report psychological benefits, such as feelings of hope and optimism. However, patients should have enough information about the possible advantages and disadvantages before embarking on strict complementary or alternative diets.

Conclusion

Diet and Nutrition are an important part of the treatment of the cancer patient. Providing patients with timely, effective nutrition advice and interventions may help overcome the effect of nutrition related impact symptoms and support patients in tolerating treatment and living with and beyond their diagnosis.

References

Alpha-tocopherol, Beta Carotene Cancer Prevention Study Group, 1994. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. N. Engl. J. Med. 330 (15), 1029–1035.

Andreyev, J., Ross, P., Donnellan, C., 2014. Guidance on the management of diarrhoea during cancer chemotherapy. Lancet Oncol. 5 (10), e447–e460. https://doi.org/10.1016/S1470-2045(14)70006-3.

Andreyev, H.J.N., Lalji, A., Mohammed, K., Muls, A.C.G., Watkins, D., Rao, S., Starling, N., Chau, I., Cruse, S., Pitkaaho, V., Matthews, J., Caley, L., Pittordou, V., Adams, C., Wedlake, L., 2021. The FOCCUS study: a prospective evaluation of the frequency, severity and treatable causes of gastrointestinal symptoms during and after chemotherapy. Support Care Cancer 29 (3), 1443–1453. https://doi.org/10.1007/s00520-020-05610-x.

Arends, J., Bachmann, P., Baracos, V., Barthelemy, N., Hartmut, B., Bozzetti, F., Fearon, K., Hutterer, E., Isenring, E., Kaasa, S., Krznaric, Z., Laird, B., Larsson, M., Laviano, A., Muhlebach, S., Muscaritoli, M., Oldervoll, L., Ravasco, P., Solheim, T., Strasser, F., Schueren, M., Preiser, J.C., 2016. ESPEN guidelines on nutrition in cancer patients. Clin. Nutr. 1–38. https://doi.org/10.1016/j.clnu.2016.07.015.

Arends, J., Baracos, V., Bertz, H., Bozzetti, F., Calder, P.C., Deutz, N.E.P., Erickson, N., Laviano, A., Lisanti, M.P., Lobo, D.N., McMillan, D.C., Muscaritoli, M., Ockenga, J., Pirlich, M., Strasser, F., Schueren, M., Gossum, A.V., Vaupel, P., Meimann, A., 2017. ESPEN expert group recommendations for action against cancer-related malnutrition. Clin. Nutr. 36 (5), 1187–1196.

Aubrey, V., Hon, Y., Shaw, C., Burden, S., 2019. Healthy eating interventions in adults living with and beyond colorectal cancer: a systematic review. J. Hum. Nutr. Diet 32, 501–511. https://doi.org/10.1111/jhn.12651.

Daly, L.E., Ni Buhuachalla, E.B., Power, D.G., Cushen, S.J., James, K., Ryan, A.M., 2018. Loss of skeletal muscle during systemic chemotherapy is prognostic of poor survival in patients with foregut cancer. J. Cachexia Sarcopenia Muscle 9, 315–325. https://doi.org/10.1002/jcsm.12267.

De Las Peñas, R., Majem, M., Perez-Altozano, J.P., Virizuela, J.A., Cancer, E., Diz, P., Donnay, O., Hurtado, A., Jimenez-Fonseca, P., Ocon, M.J., 2019. SEOM clinical guidelines on nutrition in cancer patients (2018). Clin. Transl. Oncol. 21 (1), 87–93.

De Pinho, N.B., Martucci, R.B., Rodrigues, V.D., D'Almeida, C.A., Thuler, L.C.S., Saunders, C., 2018. Malnutrition associated with nutrition impact symptoms and localization of the disease: results of a multicentric research on oncological nutrition. Clin. Nutr. 38, 1274–1279. https://doi.org/10.1016/j. clnu.2018.05.010. PMID: 29853223.

Kubrak, C., Olson, K., Jha, N., Jensen, L., McCargar, L., Seikaly, H., 2010. Nutrition impact symptoms: key determinants of reduced dietary intake, weight loss, and reduced functional capacity of patients with head and neck cancer before treatment. Head Neck 32, 290–300. https://doi.org/10.1002/hed.21174. PMID: 19626639.

Mantzorou, M., Koutelidakis, A., Theocharis, S., Giaginis, C., 2017. Clinical value of nutritional status in cancer: what is its impact and how it affects disease progression and prognosis? Nutr. Cancer 69 (8), 1151–1176. https://doi.org/10.1080/01635581.2017.1367947.

Muscaritoli, M., Arends, J., Bachmann, P., Baracos, V., Barthelemy, N., Bertz, H., Bozzetti, F., Hütterer, E., Isenring, E., Kaasa, S., Krznaric, Z., Laird, B., Larsson, M., Laviano, A., Mühlebach, S., Oldervoll, L., Ravasco, P., Solheim, T., Strasser, F., Schueren, M., Preiser, J.C., Bischoff, S.C., 2021. ESPEN practical guideline: clinical Nutrition in cancer. Clin. Nutr. 40, 2898–2913.

Omlin, U., Blum, D., Wierecky, J., Haile, S.R., Ottery, F.D., Strasser, F., 2013. Nutrition impact symptoms in advanced cancer patients: frequency and specific interventions, a case–control study. J. Cachexia Sarcopenia Muscle 3, 455–461. https://doi.org/10.1007/s13539-012-0099-x. PMID: 23307589.

Shaw, C., Fleuret, C., Pickard, J.M., Mohammed, K., Black, G., Wedlake, L., 2014. Comparison of a novel, simple nutrition screening tool for adult oncology inpatients and the malnutrition screening tool (MST) against the patient-generated subjective global assessment (PG-SGA) support. Care Cancer 23, 47–54. https://doi.org/10.1007/s00520-014-2319-8.

Totadri, S., Trehan, A., Mahajan, D., Viani, K., Barr, R., Ladas, E.J., 2019. Validation of an algorithmic nutritional approach in children undergoing chemotherapy for cancer. Pediatr. Blood Cancer 66, 12. https://doi.org/10.1002/pbc.27980.

Turcott, J.G., Zatarain-Barrón, Z.L., Cárdenas Fernández, D., Castañares Bolaños, D.T., Arrieta, O., 2021. Appetite stimulants for patients with cancer: current evidence for clinical practice. Nutr. Rev. https://doi.org/10.1093/nutrit/nuab045.

Van Veen, M.R., Mols, F., Bours, M.J.L., Weijenberg, M.P., Kampman, E., Beijer, S., 2019. Adherence to the World Cancer Research Fund/American Institute for Cancer Research recommendations for cancer prevention is associated with better health–related quality of life among long-term colorectal cancer survivors: results of the PROFILES registry. Support Care Cancer 27 (12), 4565–4574.

Wedlake, L., Shaw, C., McNair, H., 2017. Randomized controlled trial of dietary fiber for the prevention of radiation-induced gastrointestinal toxicity during pelvic radiotherapy. Am. J. Clin. Nutr. 106 (3), 849–857.

Further reading

Arends, J., Baracos c, V., Bertz, H., et al., 2017. ESPEN expert group recommendations for action against cancer related malnutrition. Clin. Nutr. 36, 1187–1196.

Arends, J., Strasser, F., Gonella, S., Chasen, M., Ripamonti, C.I., 2021. Cancer cachexia in adult patients: ESMO clinical practice guidelines. ESMO Open 6, 3. https://doi.org/10.1016/j.esmoop.2021.100092.

Bozzetti, F., 2013. Nutritional support of the oncology patient. Crit. Rev. Oncol. Hematol. 87 (2), 172–200.

Fearon, K., Strasser, F., Ankeer, S.D., et al., 2011. Definition and classification of cancer cachexia: an international consensus. Lancet Oncol. 12 (5), 489–495.

Jones, L., Watling, R.M., Wilkins, S., Pizer, B., 2010. Nutritional support in children and young people with cancer undergoing chemotherapy. Cochrane Database Syst. Rev. 7, CD003298.

Mazzotta, P., Jeney, C.M., 2009. Anorexia–cachexia syndrome: a systematic review of the role of dietary polyunsaturated fatty acids in the management of symptoms, survival, and quality of life. J. Pain Symptom Manag. 37 (6), 1069–1077.

Muscaritoii, M., Arerends, J., Bachmann, P., et al., 2021. ESPEN practical guideline: clinical Nutrition in cancer. Clin. Nutr. 40, 2898–2913.

World cancer research Fund, 2018. World Cancer Research Fund, American Institute for Cancer Research Survivors of Breast and Other Cancers Continuous Update Project Expert Report.

Cancer: Epidemiology and associations between diet and cancer

Graham A. Colditz and Hank Dart, Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, MO, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of G.A. Colditz, H. Dart, Cancer: Epidemiology and Associations Between Diet and Cancer, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 247–252, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00036-2.

Introduction 147
Sources of evidence linking diet and cancer 147
Descriptive studies 147
Special-exposure groups 148
Evidence from descriptive studies 148
Time trends within countries 148
Migrant studies 148
Analytical studies 149
Cohort studies 149
Case–control studies 149
Mendelian randomization studies 150
Intervention studies 150
Meta-analysis 150
Epidemiological issues in the study of diet and cancer 151
Resolved and unresolved issues 151
Summary of known relations between diet and cancer 152
Conclusion 153
References 153
Further reading 153

Key points

- Dietary factors account for a large proportion of human cancers, particularly overnutrition that results in weight gain and obesity.
- Evidence linking diet and human cancer began to grow in the middle part of the 20th century, and has become stronger and more refined with modern epidemiological methods, including case–control studies, cohort studies, Mendelian randomization studies, randomized controlled trials, and meta-analyses.
- Overall, an eating pattern to lower cancer risk is one that is largely plant-based, rich in vegetables, fruits, whole grains, and legumes, and low in red and processed meat, salt, sugar, and alcohol.

Glossary

Case–control study A study that compares two groups of people: those with the disease or condition under study (cases) and a very similar group of people who do not have the disease or condition (controls). Researchers study the medical and lifestyle histories of the people in each group to learn what factors may be associated with the disease or condition. For example, one group may have been exposed to a particular substance that the other was not. Also called retrospective study

Cohort study A research study that compares a particular outcome (such as lung cancer) in groups of individuals who are alike in many ways but differ by a certain characteristic (e.g., female nurses who smoke compared with those who do not smoke)

Ecologic study A study that compares large groups of people instead of individuals for differences in things such as cancer rates. The groups can differ by location (e.g., city, county, or country). They can also differ by time (a few days, years, or decades). Groups can be immigrants (compared with people who are native to the country) or people with different types of jobs

Mendelian randomization study A type of observational study in which a gene variant, or collection of variants, serves as a proxy for a modifiable factor (e.g., alcohol intake)

Meta-analysis A type of systematic review that combines published results from individual studies into summary findings, or reanalyzes individual participant data, which can have greater statistical power and precision than those of the individual studies alone

Randomized clinical trial A study in which the participants are assigned by chance to separate groups that compare different treatments; neither the researchers nor the participants can choose which group. Using chance to assign people to groups means that the groups will be similar and that the treatments they receive can be compared objectively. At the time of the trial, it is not known which treatment is best. It is the patient's choice to be in a randomized trial

Relative risk A measure of the risk of a certain event happening in one group compared to the risk of the same event happening in another group. In cancer research, relative risk is used in prospective (forward looking) studies, such as cohort studies and clinical trials. A relative risk of one means there is no difference between two groups in terms of their risk of cancer, based on whether or not they were exposed to a certain substance or factor, or how they responded to two treatments being compared. A relative risk of greater than one or of less than one usually means that being exposed to a certain substance or factor either increases (relative risk greater than one) or decreases (relative risk less than one) the risk of cancer, or that the treatments being compared do not have the same effects. Also called risk ratio

Introduction

Dietary factors account for a large proportion of human cancers, particularly overnutrition that results in weight gain and obesity. The evidence linking diet and cancer has been developing since the early 1900s, beginning with early laboratory studies and ecologic studies of cancer rates in different parts of the world. Modern epidemiologic methods, including case–control studies, cohort studies, Mendelian randomization studies, randomized controlled trials, and meta-analyses have helped define in detail the associations between specific dietary factors and cancer. Factors with strong links to an increased risk of various cancers include: overnutrition/obesity, alcohol, red and processed meat, and salt. Factors with strong links to a decreased risk include: fruits, vegetables, whole grains, fiber, dairy, calcium, and coffee.

Laboratory scientists have known since the early 20th century that various nutritional manipulations can influence the occurrence of tumors in animals. Despite this discovery of the relationship between diet and cancer in animals, widespread interest in the study of diet and cancer in humans did not develop until decades later, when large international differences in cancer rates were correlated with variations in dietary factors. For example, investigators found strong correlations between estimated per capita fat consumption and breast cancer rates internationally, raising the possibility that dietary fat could have an important role in the etiology of breast cancer. Other observations such as those demonstrating that migrating populations adopted, sooner or later, the cancer rates of their new host population strengthened the evidence that international differences were the result not of genes, but of noninherited factors, including diet.

This chapter will provide an introduction to the study designs used to investigate diet and cancer, including their history, their strengths and weaknesses, and finally their combined evidence on specific associations between dietary and related factors and individual cancers.

Sources of evidence linking diet and cancer

Descriptive studies

Rates of cancer show large differences between countries for many malignancies not merely due to different age structures of the populations. International correlations compare disease rates with lifestyle factors such as per capita consumption of specific dietary factors.

Age-adjusted rates of colon and breast cancer are many times higher in some global regions than others. For example, south Asia (8.1), western sub-Saharan Africa (9.0) and central sub-Saharan Africa (9.1) have the lowest incidence rates of colorectal cancer per 100 000; in contrast, Australasia (which includes Australia and New Zealand) (46.4), high-income Asia Pacific (41.9), and North America (39.1) have the highest (Collaborators, 2019).

Strong nutritional correlates exist for specific cancers. These studies, also known as ecological studies, use the country or other geographic area as the unit of measure rather than the individual. Diet is estimated from national food supply or disappearance and not by individual level measures of consumption. For example, Armstrong and Doll in 1975 compared per capita total fat intake and national breast cancer mortality rates among women and found a correlation of 0.89: Countries with higher fat intake had higher breast cancer mortality. They also compared per capita fat intake and mortality from colon cancer and observed a correlation of 0.85 for men and 0.81 for women.

One important strength of these correlation studies was that the contrasts in dietary intake were very large. For example, the range of fat intake within a population tends to be small compared with the range of fat intake between different populations.

Although correlation studies opened the door to new leads in the study of diet and cancer, certain limitations prevented them from advancing past the level of hypothesis generation. First and foremost, there are many factors other than dietary differences that distinguish countries with a high incidence from those with a low incidence. This makes it difficult to identify dietary factors as the primary explanation for the differences in the etiology of cancers. For example, besides consuming a diet with a higher proportion of

energy from fat, populations of countries that are more industrialized will also have shifted from an agrarian to an urbanized, sedentary society with lower total energy expenditure. Therefore, with increasing industrialization, exposure to many aspects of life will decrease exercise and increase fat intake. Consider the example of colon cancer. The international correlation between fat and colon cancer mortality in men is 0.85, and for meat it is 0.85 (Armstrong and Doll, 1975). There is also a correlation between gross national product and colon cancer mortality (0.77 for men); more industrialized countries have higher economic production and higher rates of cancer. Owing to the many factors that are associated with industrialization it is not possible to separate out which factor is important in the etiology of colon cancer, lack of physical activity or increased consumption of fat or meat. Studies with data on lifestyle factors at the individual level are needed to clarify which of these variables is important (see **Analytical studies** below).

Special-exposure groups

Within populations there are groups that have unique dietary patterns which may provide valuable information in the probe for further information on the relationship of diet and cancer. These groups are called special-exposure groups and are often defined by ethnic or religious characteristics. In addition to offering many of the advantages of correlation studies, the number of alternative explanations for any observations may be reduced if the special-exposure group lives in the same area as the comparison group.

As a largely vegetarian group, the Seventh-day Adventists have been used in studies of meat eating and cancer. Studies of these groups, however, are limited in the same ways that other ecological studies are limited. For example, although lower rates of colon cancer have been observed among Seventh-day Adventists—supporting the hypothesis that meat is related to colon cancer—there are other lifestyle choices that characterize the group, such as low rates of tobacco use and alcohol intake, which could also modify their rates of colon cancer.

Evidence from descriptive studies

In 1981, Doll and Peto made an estimate based largely on descriptive studies that 35% of cancers in the USA may be attributable to dietary factors; but reflecting uncertainty in the sources of data used for this estimate, they noted that the range of possible dietary contribution was from as low as 10% to as high as 70% (Doll and Peto, 1981). The marked variation in the rates of most cancers among countries is evidence that dietary factors may influence the development of cancer. Though their analysis focused largely on potential mechanisms through which diet could impact cancer risk, Doll and Peto highlighted a range of potential dietary factors that could influence risk, including overnutrition, dietary fat, carotenoids, cooking method, and aflatoxins. Despite the fact that descriptive studies provide an excellent source of hypotheses, it is necessary to conduct analytical studies to collect data that will provide more definitive evidence.

Time trends within countries

The analysis of cancer trends over time can lead to useful findings in the study of diet and cancer. By looking at the change in cancer rates in a specific population over time and comparing these rates with changes in specific factors over the same period (e.g., changes in dietary habits), investigators can uncover possible associations supporting the dietary factors hypotheses. For example, in 1991 researchers examined vital statistics for Japanese natives and US whites that revealed changes in cancer mortality and related antecedent patterns of lifestyle in the two populations (Wynder et al., 1991). These investigations uncovered that animal fat consumption in Japan steadily increased from a daily level of 6.5 g per person in 1955 to 27.6 g in 1987; at the same time the Japanese rate of colon cancer in men rose at a rapid pace; in fact, the mortality rates owing to colon cancer in men almost tripled over this time. This evidence lent more support to the hypothesis that mortality from colon cancer in men is influenced by high dietary fat consumption.

Similar data were collected in Singapore to determine trends in the incidence of breast cancer: in 1996, an average annual increase in breast cancer incidence of 3.6% over a 25-year period for all women was reported (Seow et al., 1996). The most convincing evidence that the observed trend was real was that it was clearly cohort-related rather than period-related. The risk was observed to increase in successive birth cohorts from the 1890s to the 1960s. Changes in dietary consumption patterns (e.g., the adoption of a more Western diet) among other factors, such as decreasing parity, are cited as having a possible effect on the continuing increase in rates of breast cancer among women in Asia. Like descriptive studies, time-trend studies are a valuable source for hypotheses generation, but more definitive evidence is required from analytical epidemiology to uncover any real associations between dietary factors and cancer rates.

Migrant studies

Migrant studies examine the rates of specific diseases in migrating populations. These studies have been important in addressing the possibility that observed correlations in ecological studies are owing to genetic factors. Generally, results from migrant studies have found that the migrating group takes on the rate of cancer of the new country. Hence genetic factors are excluded as the dominant cause for varying rates of cancer between countries. A landmark example of this is seen in breast cancer incidence rates of Asian migrant populations to the USA. Though rates are increasing, Asia has historically had low rates of breast cancer relative to the

USA. Two studies in the 1990s documented that these rates among certain Asian migrants to the USA moved toward the higher US rates (Stanford et al., 1995; Ziegler et al., 1993). The increased risk of breast cancer among migrants occurs primarily in later generations, leading investigators to believe that the causal factors operate early in life. Investigators also consider major changes in the rate of disease that occur within a population over time as evidence that nongenetic factors play an integral role in the etiology of cancer. The limitations of migrant studies are similar to those of ecological studies.

Analytical studies

Cohort studies

Cohort studies involve the collection of information from healthy participants who are followed over time and observed for the occurrence of new cases of disease (incident cases). During or at the end of follow-up, the disease frequency within a cohort may be measured as either a cumulative incidence rate (the number of cases divided by the entire base population) or an incidence density rate (the number of cases divided by the total follow-up time accumulated by all members of the population, or "person–time" follow-up). The relative risk is the rate of disease (cumulative incidence rate or incidence density rate) in the exposed (e.g., those with a high intake of dietary fat) divided by the rate of disease in the unexposed (e.g., those on a low-fat diet). A relative risk of 2 implies that the exposed group has twice the rate of disease compared with the unexposed group.

For illustration, in a study of 121 700 women, a group of participants who completed dietary questionnaires and had no previous diagnosis of cancer in 1980, were followed through 1988 to address the hypothesis that dietary fat increases and fiber intake decreases the risk of breast cancer (Willett et al., 1992). This outcome was defined by histologically confirmed cases of breast cancer. In one analysis, the primary exposure of interest was energy-adjusted intake of total dietary fiber. Among the women in the highest quintile of energy-adjusted dietary fiber intake there were 299 cases of breast cancer compared with 305 cases among the women in the lowest quintile. This gave a relative risk (with adjustment for established breast cancer risk factors) of 1.02 (95% CI 0.85–1.23) for those in the highest quintile of energy-adjusted dietary fiber intake compared with those in the lowest quintile. In this cohort study, dietary fiber was not associated with breast cancer.

During the same period, there was also a growing body of evidence from cohort studies for the assessment of dietary fat intake and breast cancer in developed countries. In one pooled analysis based on individual level data from seven prospective studies with at least 150 incident breast cancer cases each ($n = 4980$) and a large comparison series (i.e., noncases), the average relative risk was 1.05 (95% CI 0.94–1.16) comparing highest quintile of total fat intake with the lowest, which was not statistically significant (Hunter et al., 1996).

The use of cohort studies can be advantageous in many ways when studying the relationship between diet and cancer. A cohort study allows the assessment of multiple effects of a given dietary exposure. Dietary data can be updated during follow-up and the temporal relation between diet and cancer can be addressed. For example, the potentially beneficial effects of moderate alcohol intake in reducing the risk of outcomes like gallstones, ischemic stroke, and coronary heart disease, and the potentially deleterious effects of alcohol on outcomes like cancer, bone fractures, and hemorrhagic stroke can be weighed against each other in a cohort study (Colditz et al., 2016; Mostofsky et al., 2016). It is also possible to measure the absolute rates of disease according to the level of food or nutrient intake.

Among the limitations of cohort studies is the concern that current practice, usage, or exposure may change over the duration of the follow-up, limiting the ability to come to any relevant conclusions in studies of diet and cancer that have measured exposure just once at the beginning of the study. Controlling for extraneous variables such as smoking, which are related both to risk of cancer and to dietary intake, and separating the effects of specific dietary factors from those that exist together, also limit the range of knowledge that can be extracted from cohort studies.

Some investigators believe that the large number of subjects required to study rare disease and the high expense of management and maintenance also limit the usefulness of cohort studies. Others believe that the larger overall monetary investment most cohort studies require can be advantageous: more variables can be studied and in the long run further hypotheses can be generated and more conclusions produced than in a single case–control study that relies on recall of past habits.

Case–control studies

In case–control studies information is obtained from diseased participants and compared with information provided by disease-free controls with respect to a possible risk factor (e.g., level of a dietary factor). Data collected from these studies can be used to evaluate the hypothesis that the risk factor is a cause of the disease. The cases are selected from a defined population, such as a country population. The population represents those at risk of developing the disease under study. Each time someone in the defined population is diagnosed with the disease during the duration of the study, this individual joins the case series. As each case arises from the population, one or more controls should be sampled to estimate the prevalence of the exposures among those remaining free from disease. The controls may be chosen from any population of individuals that provides valid information about those at risk for the disease. It is important to choose controls so that their probability of selection is unrelated to the exposure being studied.

In the study of the relationship of diet and cancer, case–control studies may be used to evaluate the hypotheses that individual or multiple dietary factors are the cause of the cancer under investigation. For example, a study in 1977 identified all cases of lung cancer diagnosed during an 18-month period from 1972 in three Singapore hospitals (MacLennan et al., 1977). Controls were chosen from other hospital patients free of any smoking-related diseases. There were a total of 233 cases and 300 controls

interviewed regarding their frequency of consumption of dark-green leafy vegetables and food preparation habits. The investigation found a substantially increased risk of lung cancer among those reporting a low consumption of dark-green leafy vegetables.

Case–control studies are better suited to the study of rare diseases because in cohort studies tens of thousands of individuals must be followed in order to study the most common cancers. It is also thought that case–control studies may be quicker and less expensive to conduct because they require fewer subjects, and they are therefore often employed as an alternate mode of investigation to cohort studies.

Among the limitations of case–control studies is the comparability of information between the cases and the controls. Although in a cohort study the exposure of interest is measured before the onset of disease, in case–control studies the exposure is assessed in individuals who (in most cases) already know their own disease status. Often the person collecting the data will also know the disease status of the patient. This may influence the accuracy of the data collected, either through differential recall by cases and controls, or by an interviewer being more persistent in questioning cases than controls. For rapidly fatal cancers, a biased subset of cases may be well enough to participate. In cohort studies neither the participant nor the investigator knows whether or not the subject will be a case or noncase by the end of the follow-up period, and typically all cases are ascertained for analysis.

Mendelian randomization studies

Mendelian randomization studies are a relatively new method of assessing the relationship between select dietary factors and chronic disease, including cancer (Qi, 2009). In Mendelian randomization, a gene variant, or collection of variants, serves as a proxy for a modifiable dietary factor. Because such gene variants should be distributed randomly in populations, the method may help limit the potential confounding and related issues that can exist with other observational studies, and therefore may help provide evidence of causation. However, three key assumptions need to be met in Mendelian randomization studies (Yarmolinsky et al., 2018b). Briefly: the gene variant should be associated with the dietary factor of interest; the gene variant should not be associated with confounders; and, finally, the gene variant should not be associated with the cancer of interest. Because of this, Mendelian randomization analyses can currently only be applied to a select number of modifiable dietary factors. But with the continued growth of genome-wide association studies, the list of applicable factors is likely to grow as well.

The study of links between selenium and prostate cancer risk provides a good illustration of Mendelian randomization analysis. As a follow-up to the SELECT selenium and vitamin E supplementation trial, Yarmolinksy et al. studied the association between circulating selenium levels and prostate cancer by using a proxy of eleven single nucleotide polymorphisms (SNPs) strongly linked to serum selenium levels (Yarmolinsky et al., 2018a). In line with previous results, the Mendelian randomization analysis found no statistically significant link between blood selenium and overall risk of prostate cancer (RR 1.01, 95% CI 0.89–1.13).

As with all study methods, Mendelian randomization analyses have both strengths and weaknesses. While they can limit biases and help provide evidence of causation using observational data, for their results to be reliable, they must also satisfy a number of assumptions, which can be difficult to meet, and which also limit the number of factors that are able to be studied.

Intervention studies

In principle, the most powerful means of determining the effects of dietary factors on cancer risk is an intervention study (i.e., a randomized trial). In randomized trials bias is removed because of the equal distribution of risk factors in each group. For example, it had been proposed that a randomized trial of fat reduction could help uncover the mystery of the relationship between dietary fat intake and breast cancer. The Women's Health Initiative was started in 1992 by the US National Institutes of Health with the goal of enrolling and randomizing over 48,000 postmenopausal women, half of whom would be trained to follow a diet deriving less than 20% of energy from fat. Given its focus, the trial was limited in the hypotheses it could test. For example, it would not be able to assess the potential impact that dietary fat reduction at an early age could have on breast cancer risk several decades later. Other problems with such a randomized trial included the difficulty of maintaining compliance with a diet incompatible with prevailing food consumption habits, and the gradual secular population decline in total fat consumption which reduced the size of the comparison of fat intake between the intervention group and the control groups. The Women's Health Initiative Trial counseled the women in the intervention group to adopt a diet that was high in fruits, vegetables, and grain products as well as low in total and saturated fat, therefore making it more difficult to distinguish between the effect of the fat reduction and that of increasing intake of fruits, vegetables, and grain products. The trial failed to show any significant benefit of reducing fat intake on breast cancer risk, both within the initial intervention period of 8.3 years and a post-intervention follow-up of 5.2 years (Prentice et al., 2006; Thomson et al., 2014).

All in all, while intervention studies may, in principle, have a great chance of determining effects of dietary factors on cancer risk, trials of sufficient duration and size may not be feasible because of issues such as long-term compliance and cost.

Meta-analysis

A meta-analysis is a type of systematic review that combines published results from individual studies into summary findings, which can have greater statistical power and precision than results in the individual studies (Deeks et al., 2021). It can be a particularly useful method when single studies on a particular topic are underpowered and/or have inconclusive results.

As an illustration, in a 2014 meta-analysis assessing the association of calcium intake and risk of colorectal cancer, 5 of 15 included cohort studies found modest, statistically significant protective effects with calcium intake (Keum et al., 2014). Results from the ten other studies were non-significant, two of which had positive associations. When combined, the meta-analysis included over 12,000 cases of colorectal cancer and approximately 1.4 million total participants. The summary does-response result

showed that for each 300 mg per day increase in calcium intake, the risk of colorectal cancer decreased by 8% (RR 0.92 95% CI 0.89–0.95).

A number of factors determine the quality of a meta-analysis, including the quality of the individual studies included, the heterogeneity of the results included, and consideration of publication bias, among others. To overcome some of these limitations, investigators often collaborate to reanalyze the original individual participant data in a common analysis generating a pooled summary result. This is considered highest quality meta-analysis (Tierney et al., 2021).

Epidemiological issues in the study of diet and cancer

Resolved and unresolved issues

Some of the issues that researchers have encountered in their attempt to uncover the mystery of the dietary factors linked to cancer include the difficulty of distinguishing the importance of parts of dietary factors from the overall effect of each dietary factor. Looking back, in a meta-analysis in 1990 of 12 case–control studies of dietary fat intake and cancer, four studies observed a significant positive association, six uncovered nonsignificant positive associations, and two saw inverse associations. When the data were analyzed together there was a positive association observed for both total fat intake and saturated fat intake. Investigators must ask themselves which factor has larger implications in the study of diet and cancer, as not all studies have included analyses of the individual types of fats along with their data on overall fat consumption. In the study of the influence of dietary fiber intake (which includes crude fiber and many soluble fiber fractions) on cancer rates, there is still ongoing debate about specific definitions of fiber and the most appropriate method of biochemical analysis for determining fiber content of individual foods (Fuller et al., 2016; Jones, 2014). This same issue arises with the study of most dietary factors and could affect important advances in the study of diet and cancer.

Biochemical indicators of food and nutrient intake have two fundamental uses in epidemiological studies. Most often they serve as a "surrogate" for actual dietary intake in studies of disease occurrence. For nutrients that vary widely in concentration within foods and for which food composition tables are inaccurate, biochemical indicators may be the most feasible way of measuring intake. Within-food variation may occur owing to differences in food storage, processing, or preparation, or may be owing to geographical differences in soil nutrient content. For example, it has been found that selenium content in US soil can vary by as much as 100-fold, which in turn causes the selenium content of swine muscle to vary more than 15-fold.

Like most exposures driving risk of chronic disease, nutrient exposures relevant to disease are usually long-term. As the promotion period for cancers may be years or decades, it is usually desirable that a biomarker indicates the cumulative effect of diet over an extended period of time. There are a few possible methods to surpass the barrier of an indicator that is only sensitive to short-term intake, and to overcome the day-to-day intake fluctuations that occur with most nutrients: (1) experimental studies, in which nutrient levels are manipulated; (2) Mendelian randomization studies, which can assess by proxy lifetime nutrient exposure and (3) sampling levels in individuals longitudinally. Biomarkers of nutrient levels in blood or other tissues can provide a useful assessment of intake of certain nutrients, although the above considerations must be acknowledged, and careful attention must be given to specimen collection, storage, and analysis in order to avoid misclassification or bias. With an expanding array of biochemical indicators that have been validated as measures of dietary intake, their use in nutritional epidemiology will continue to grow.

The limited range of diet within most populations adds its own set of complexities to the epidemiological study of nutrition and cancer. For example, in the majority of populations where foods high in fat are readily available, very few individuals consume less than 30% of their energy from fat. This makes it difficult to study the impact of reducing fat intake to less than 30% of total energy intake. At the same time, some individuals of a relatively homogeneous population may have very different dietary patterns: For example, a range of dietary fat intake from 25% to 40% of total energy was seen within a cohort of 52,000 male health professionals in the USA.

Given that most neoplasms have a long induction period (the time from an exposure to a carcinogen to the development of cancer), often spanning several decades, accurate measure of long-term dietary intake is of utmost importance in the study of the implications of diet on cancer. Therefore, short-term methods of dietary assessment such as 24 h recalls are usually insufficient. In the context of case–control studies these short-term methods are inappropriate because they measure current diet, and it has been found that individuals alter their diet after the diagnosis of cancer. Mendelian randomization studies can measure lifetime exposures, but reliable links between gene variants and dietary exposures are currently limited. The most feasible method of measuring long-term intakes in large numbers of individuals is the food frequency questionnaire: These questionnaires measure the usual frequency of a selected list of foods.

Food frequency questionnaires to assess dietary intake need to be carefully designed. First of all, the food items on the questionnaire must represent the major source of nutrients of interest within the study population. Depending on the consistency of the concentration of a nutrient in a given food, the precision of dietary questionnaires varies. Food frequency questionnaires may provide rankings by level of intake, but they do not quantify actual intake. A dietary questionnaire may efficiently distinguish between participants with low-fiber and high-fiber intakes in a given population, but it will not necessarily provide a precise assessment of the absolute fiber intake. In the case of larger studies, it is possible for a random sample of participants to provide a more comprehensive assessment of intake by keeping several weeks of dietary records. This additional information will provide a more

precise quantification of dietary intake by helping estimate true dose–response relationships between a nutrient and diet expressed in absolute intake.

Summary of known relations between diet and cancer

With the wealth of studies since the 1970s and important advances in analytical methods, there are clearly documented and strong associations between diet and cancer (**Table 1**) (World Cancer Research Fund/American Institute for Cancer Research, 2018b; Bouvard et al., 2015; Lauby-Secretan et al., 2016; World Cancer Research Fund/American Institute for Cancer Research, 2018a). Convincing evidence based on consistent findings from epidemiological studies conducted in diverse populations now shows that diet is an established cause of multiple cancers, including two of the most common cancers worldwide, breast and colorectum (Sung et al., 2021). With these rich epidemiological data we can more confidently conclude that close to 30% of cancer is attributable to diet (Colditz et al., 2012).

Public health officials have taken the accumulated evidence and developed strategies for minimizing cancer risk. Among these recommendations is a largely plant-based diet high in vegetables, fruits, whole grains, and legumes and low in red and processed meat, salt, sugar, and alcohol. Fats that are consumed should be polyunsaturated or monounsaturated, examples being canola oil and olive oil. Regular physical activity, in addition to independently lowering the risk of multiple cancers, can help with weight control, particularly when combined with a healthy eating pattern.

Table 1 Strong and convincing evidence for increased and decreased risk of major forms of cancer and excess body weight by foods and drinks, micronutrients, dietary nonnutrients, and nutrition-related indicators.

	Strong evidence	*Convincing evidence*
Increased risk		
Foods and drinks		
Red meat	Colorectum, pancreas, prostate (mainly advanced)	
Processed meat	Stomach	Colorectum
Western diet and fast food	Excess weight	
Sugar-sweetened beverages		Excess weight
Alcohol	Breast (premenopausal), stomach	Oral cavity, pharynx, larynx, colorectum, breast (postmenopausal), liver, esophagus
Micronutrients		
Salt (NaCl)	Stomach	
Nutritional covariates		
Overweight/obesity	Oral cavity, pharynx, larynx, stomach (gastric cardia), gallbladder, ovary, prostate (advanced), meningioma, thyroid, multiple myeloma	Colorectum, breast (postmenopausal), endometrium, kidney, esophagus, liver, pancreas
Height	Endometrium, kidney, pancreas, prostate, melanoma	Breast (premenopausal), breast (postmenopausal), colorectum, ovary
Decreased risk		
Foods and drinks		
Fruits and vegetables	Aerodigestive	
Whole grains	Colorectum	
Dairy	Colorectal	
Mediterranean diet	Excess weight	
Coffee	Endometrium, liver	
Alcohol	Kidney	
Micronutrients		
Calcium	Colorectum	
Nonnutrients		
Fiber	Colorectum, excess weight	
Nutritional covariates		
Overweight/obesity	Breast (premenopausal)	
Physical activity	Breast (premenopausal), breast (postmenopausal), endometrium, excess weight	Colorectum

Based on: Bouvard, V., Loomis, D., Guyton, K.Z., et al., 2015. Carcinogenicity of consumption of red and processed meat. Lancet Oncol. 16, 1599–1600. Lauby-Secretan, B., Scoccianti, C., Loomis, D., et al., 2016. Body fatness and cancer–viewpoint of the IARC working group. N. Engl. J. Med. 375, 794–798. World Cancer Research Fund/American Institute for Cancer Research, 2018a. Continuous Update Project: Diet, Nutrition, Physical Activity and the Prevention of Cancer. Summary of Strong Evidence. [Online]. Available: wcrf.org [Accessed July 30, 2021]. World Cancer Research Fund/American Institute for Cancer Research, 2018b. Diet, Nutrition, Physical Activity and Cancer: A Global Perspective. Continuous Update Project Expert Report 2018.

Conclusion

Over the past decades our understanding of the links between what we eat and our risk of cancer has become increasingly refined. Overall, an eating pattern to lower cancer risk is one that is largely plant-based, rich in vegetables, fruits, whole grains, and legumes, and low in red and processed meat, salt, sugar, and alcohol. Identifying such links between diet and cancer, though, is in many ways only an initial step toward improving the health of individuals and the broader public. Many different factors determine eating patterns, and successfully promoting healthy changes to eating patterns can be difficult and take sustained efforts on multiple levels using multiple methods. So, while it is important to continue to uncover and clarify connections between diet and cancer, it is also important that we continue to apply that knowledge by exploring, implementing, and promoting effective, evidence-based efforts in prevention.

See Also: Cancer: Epidemiology of lung cancer; Dietary surveys: Surveys of food intake in groups and individuals; Dietary fiber: Physiological effects and health outcomes; Vegetarian diets

References

Armstrong, B., Doll, R., 1975. Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. Int. J. Cancer 15, 617–631.
Bouvard, V., Loomis, D., Guyton, K.Z., et al., 2015. Carcinogenicity of consumption of red and processed meat. Lancet Oncol. 16, 1599–1600.
Colditz, G.A., Philpott, S.E., Hankinson, S.E., 2016. The impact of the Nurses' Health Study on population health: prevention, translation, and control. Am. J. Publ. Health 106, 1540–1545.
Colditz, G.A., Wolin, K.Y., Gehlert, S., 2012. Applying what we know to accelerate cancer prevention. Sci. Transl. Med. 4, 127rv4.
GBD Colorectal Cancer Collaborators, 2019. The global, regional, and national burden of colorectal cancer and its attributable risk factors in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol. Hepatol. 4, 913–933.
Deeks, J.J., Higgins, J.P.T., Altman, D.G., 2021. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., Welch, V.A. (Eds.), Cochrane Handbook for Systematic Reviews of Interventions Version 6.2. Cochrane.
Doll, R., Peto, R., 1981. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. J. Natl. Cancer Inst. 66, 1191–1308.
Fuller, S., Beck, E., Salman, H., Tapsell, L., 2016. New horizons for the study of dietary fiber and health: a review. Plant Foods Hum. Nutr. 71, 1–12.
Hunter, D.J., Spiegelman, D., Adami, H.O., et al., 1996. Cohort studies of fat intake and the risk of breast cancer–a pooled analysis. N. Engl. J. Med. 334, 356–361.
Jones, J.M., 2014. Codex-aligned dietary fiber definitions help to bridge the 'fiber gap'. Nutr. J. 13, 34.
Keum, N., Aune, D., Greenwood, D.C., et al., 2014. Calcium intake and colorectal cancer risk: dose-response meta-analysis of prospective observational studies. Int. J. Cancer 135, 1940–1948.
Lauby-Secretan, B., Scoccianti, C., Loomis, D., et al., 2016. Body fatness and cancer–viewpoint of the IARC working group. N. Engl. J. Med. 375, 794–798.
Maclennan, R., Da Costa, J., Day, N.E., et al., 1977. Risk factors for lung cancer in Singapore Chinese, a population with high female incidence rates. Int. J. Cancer 20, 854–860.
Mostofsky, E., Mukamal, K.J., Giovannucci, E.L., et al., 2016. Key findings on alcohol consumption and a variety of health outcomes from the Nurses' Health Study. Am. J. Publ. Health 106, 1586–1591.
Prentice, R.L., Caan, B., Chlebowski, R.T., et al., 2006. Low-fat dietary pattern and risk of invasive breast cancer: the Women's Health Initiative randomized controlled dietary modification trial. JAMA 295, 629–642.
Qi, L., 2009. Mendelian randomization in nutritional epidemiology. Nutr. Rev. 67, 439–450.
Seow, A., Duffy, S.W., Mcgee, M.A., et al., 1996. Breast cancer in Singapore: trends in incidence 1968–1992. Int. J. Epidemiol. 25, 40–45.
Stanford, J.L., Herrinton, L.J., Schwartz, S.M., Weiss, N.S., 1995. Breast cancer incidence in Asian migrants to the United States and their descendants. Epidemiology 6, 181–183.
Sung, H., Ferlay, J., Siegel, R.L., et al., 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 71, 209–249.
Thomson, C.A., Van Horn, L., Caan, B.J., et al., 2014. Cancer incidence and mortality during the intervention and postintervention periods of the Women's Health Initiative dietary modification trial. Cancer Epidemiol. Biomarkers Prev. 23, 2924–2935.
Tierney, J.F., Stewart, L.A., Clarke, M., 2021. Chapter 26: individual participant data. In: Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., Welch, V.A. (Eds.), Cochrane Handbook for Systematic Reviews of Interventions Version 6.2. Cochrane.
Willett, W.C., Hunter, D.J., Stampfer, M.J., et al., 1992. Dietary fat and fiber in relation to risk of breast cancer: an 8-year follow-up. J. Am. Med. Assoc. 268, 2037–2044.
World Cancer Research Fund/American Institute for Cancer Research, 2018a. Continuous Update Project: Diet, Nutrition, Physical Activity and the Prevention of Cancer. Summary of Strong Evidence [Online]. Available. (Accessed 30 July 2021).
World Cancer Research Fund/American Institute for Cancer Research, 2018b. Diet, Nutrition, Physical Activity and Cancer: A Global Perspective. Continuous Update Project Expert Report 2018.
Wynder, E.L., Fujita, Y., Harris, R.E., et al., 1991. Comparative epidemiology of cancer between the United States and Japan. A second look. Cancer 67, 746–763.
Yarmolinsky, J., Bonilla, C., Haycock, P.C., et al., 2018a. Circulating selenium and prostate cancer risk: a Mendelian randomization analysis. J. Natl. Cancer Inst. 110, 1035–1038.
Yarmolinsky, J., Wade, K.H., Richmond, R.C., et al., 2018b. Causal inference in cancer epidemiology: what is the role of Mendelian randomization? Cancer Epidemiol. Biomarkers Prev. 27, 995–1010.
Ziegler, R.G., Hoover, R.N., Pike, M.C., et al., 1993. Migration patterns and breast cancer risk in Asian-American women. J. Natl. Cancer Inst. 85, 1819–1827.

Further reading

Doll, Peto, 1981. The causes of cancer: quantitative estimates of avoidable risks in the United States today. J. Natl. Cancer Inst. 66, 1191–1308.
Willett, W., 2013. Nutritional Epidemiology. Oxford University Press, Oxford.
World Cancer Research Fund/American Institute for Cancer Research, 2018. Diet, Nutrition, Physical Activity and Cancer: A Global Perspective Continuous Update Project Expert Report. dietandcancerreport.org.

Cancer: Epidemiology of lung cancer

Anthony J. Alberg, Rajat Das Gupta, and Maxwell Akonde, Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A.J. Alberg, Cancer: Epidemiology of Lung Cancer, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 259–264, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00038-6.

Introduction 154
Lung cancer histopathology 155
Risk factors 155
Diet 156
Dietary hypotheses and mechanisms 156
Dietary associations with lung cancer 157
Fruit 157
Vegetables 157
Micronutrients 157
Phytochemicals 158
Fat 158
Body mass index 158
Beverages 158
Meat and fish 159
Diet and prevention 159
Chemoprevention trials 159
Observation versus experiment 159
Conclusions 160
References 160

Key points

- Briefly review the epidemiology of lung cancer and non-dietary causes of lung cancer.
- Provide an overview of methodologic challenges in studying the potential influence of dietary factors on lung cancer risk, with emphasis on the ubiquitous challenge of confounding by cigarette smoking. Further, epidemiology relies mostly on evidence from observational studies, and even when randomized trials are possible there are important differences in research methodology.
- Provide an overview of the evidence of dietary factors that have been widely studied in relation to lung cancer.

Glossary

Biomarker A compound measured in human tissue as a marker of the body's exposure; for example, measurement of concentrations of beta-carotene in blood to measure dietary intake of beta-carotene

Case-control study An observational (nonexperimental) study that starts with patients of disease (lung cancer) and controls without disease and attempts to measure exposures that occurred in the past

Confounding When an extraneous variable (e.g., cigarette smoking) affects the association between the independent variable (e.g., dietary factor) and dependent variable (e.g., lung cancer)

Epidemiology The study of health and disease in human populations

Phytochemical Low molecular weight molecules produced by plants

Prospective study An observational (nonexperimental) study that starts with measurement of exposure and follows participants up over time for occurrence of the disease (lung cancer)

Introduction

Lung cancer was a rare condition in the early 1900s but since then its occurrence increased markedly so that it is now accounts for 18% of all deaths globally, making it the leading worldwide cause of cancer death. Lung cancer is the third most common cancer

globally. Overall, an estimated 2.2 million new cases of lung cancer, constituting 11.4% of all cancer cases, were reported in 2020. Lung cancer is the most common cancer in males (14% of all male cancer cases) and the third most common cancer in females (8% of all female cancer cases) (Sung et al., 2021). One predominant cause is the major contributor to the global lung cancer epidemic: cigarette smoking. The link between cigarette smoking and lung cancer is so strong that population patterns of cigarette smoking are the primary determinant of the occurrence of lung cancer at the population level (Malhotra et al., 2016; Pesch et al., 2012).

Distinct patterns are present in the occurrence of lung cancer globally. The age-adjusted incidence rates and mortality rates for lung cancer rates are presently declining in high-income countries primarily due to decreases in the prevalence of cigarette smoking. The majority (62%) of lung cancer cases now occur in low- and middle-income countries. Western and central Africa have the lowest lung cancer rates in both men and women, primarily due to the low prevalence of cigarette smoking. Eastern Asia, Central and Eastern Europe currently experience the highest lung cancer rates in men, whereas northern Europe and North America have the highest lung cancer rates in women. The trends are mostly driven by historic cigarette smoking patterns (McIntyre and Ganti, 2017; Sung et al., 2021). In other countries such as China, and several parts of Africa where cigarette smoking prevalence increased more recently, lung cancer rates are increasing and will continue to do so unless reductions in the prevalence of cigarette smoking are achieved (Bray and Weiderpass, 2010; McIntyre and Ganti, 2017; Sung et al., 2021).

In additional to the high lung cancer incidence rate, an important contributor to the high lung cancer mortality rate is a poor survival rate. Due to poor survival from lung cancer, the population patterns in lung cancer mortality rates tend to closely mirror the incidence rates. For example, the 5-year survival rate is 21% and 13% in the United States and United Kingdom, respectively (Allemani et al., 2018; American Cancer Society, 2021). Lung cancer survival rates are especially low for late-stage disease (Bade and Cruz, 2020).

In addition to cigarette smoking and other forms of combustible tobacco, many other environmental risk factors and clinical risk indicators for lung cancer have been identified. A substantial body of evidence has been generated on the potential influence of dietary factors on lung cancer risk.

Lung cancer histopathology

The four most common histological types of lung cancer are adenocarcinoma, squamous cell carcinoma, large cell carcinoma, and small cell carcinoma. All four histologic types are caused by cigarette smoking (Alberg and Nonemaker, 2008; Warren et al., 2020). For clinical purposes these diagnoses are dichotomized as non-small cell lung cancer, which comprise approximately 85% of the total, with the remaining 15% comprised of small cell lung cancer. This classification is used because surgery is not a primary curative treatment option for small cell lung cancer due to its high metastatic potential to the central nervous system.

Characterizing lung tumors based on molecular characteristics plays an increasingly important role in classifying and treating lung cancer using precision medicine approaches. For example, mutations and genetic events in *EGFR*, *ALK*, and *ROS1* are routinely tested for in the US to guide treatment plans for adenocarcinoma of the lung.

Risk factors

Cigarette smoking is the major cause of lung cancer, and the most important factor responsible for the worldwide lung cancer epidemic. Cigarette smoking causes approximately 85% of all lung cancer cases in regions where the cigarette smoking epidemic is fully mature. In smokers, the risk of lung cancer increases with both the duration of smoking and the number of cigarettes smoked. There is no known safe level of smoking, as even the secondhand tobacco smoke involuntarily inhaled by nonsmokers is causally associated with lung cancer. Lung cancer risk decreases in those who quit smoking compared to persistent smokers, but not to the level of those who never smoked (Alberg and Nonemaker, 2008; U.S. Department of Health and Human Services, 2020).

In addition to cigarette smoking, many other causes of lung cancer have been established. Numerous occupational lung carcinogens have been identified; the substances involved include radon (found in underground mines), arsenic, asbestos, chromium, chloromethyl ethers, nickel, and polycyclic aromatic hydrocarbons (Shankar et al., 2019; Steenland et al., 1996). Synergism with smoking has been shown for several of these agents, such as asbestos and radon. Many other agents are suspected occupational carcinogens (Alberg et al., 2013; Errenet al., 1999; Shankar et al., 2019).

Air pollution is associated with increased lung cancer risk. Outdoor air pollution increases lung cancer risk through inhalation of air contaminants from combustion sources that generate polycyclic aromatic hydrocarbons and radionuclides. Carcinogens in indoor air vary with the setting but may include radon, tobacco smoke, smoke from wood or coal burning, and cooking fumes (Alberg et al., 2013; Eckel et al., 2016).

The observed familial aggregation of lung cancer suggests that genetic factors may influence susceptibility. Data from genome-wide association studies (GWAS) have provided promising leads. The results of four GWAS have been remarkably consistent in identifying genetic variants within a region on the long arm of chromosome 15 that are associated with lung cancer risk. For example, those with at least one variant allele of a specific SNP in this region (rs8034191) had a 1.3-fold greater risk of lung cancer than those homozygous for the wild type allele (Wei et al., 2011).

Diet

Dietary hypotheses and mechanisms

Epidemiological research on diet and lung cancer has been both hypothesis-driven and descriptive, exploring associations between foods, nutrient indexes, or biomarkers and lung cancer risk. Interest in macronutrients has emphasized indices of dietary fat, which has long been known to have the capacity to act as a tumor promoter. Micronutrients have been extensively studied, spurred initially by the pioneering studies of Bjelke and the original vitamin A and β-carotene hypotheses. Bjelke and subsequent researchers originally focused on vitamin A because of its role in cellular differentiation, but this line of inquiry was subsequently expanded to include antioxidant micronutrients, with an emphasis on β-carotene. The more general hypothesis has been advanced that antioxidant micronutrients may protect against oxidative damage to DNA and thereby protect against cancer. Hypotheses concerning specific beverages have also been proposed; for example, animal studies have shown a link between alcohol and changes in lung lipids, including surfactant, and in levels of enzymes that can activate procarcinogens and mutagens.

Another epidemiologic approach, empirical rather than hypothesis-driven, has been to explore the intakes of several specific foods or food groups for associations with lung cancer risk. The evidence documenting inverse associations between fruit and vegetable consumption and lung cancer stem from this inductive approach.

Certain methodological issues are relevant to the topic of diet and lung cancer. Foremost among these is the major challenge posed by the potential confounding effects of cigarette smoking due to the potent causal role of cigarette smoking, combined with the fact that smokers tend to eat less healthful diets than nonsmokers. This makes it very difficult to disentangle the potential impact of cigarette smoking on any observed diet–lung cancer association. Thus, even when efforts are made to attempt to control for smoking, residual confounding of diet–lung cancer associations may persist. Complicating matters further is that cigarette smoke can directly affect nutritional biomarkers (Fig. 1); for example, smokers tend to have lower levels of circulating antioxidant micronutrients even after accounting for differences in dietary intake. Similar associations have even been noted for secondhand smoke exposure.

Characteristics of epidemiological studies in general further introduce challenges to clear-cut inferences concerning observed associations between dietary factors and lung cancer. Approaches to dietary assessment are not fully standardized, so there may be differences between studies in the number of foods queried, the measurement of serving sizes, and the data collection approach employed. There is also uncertainty as to the biologically relevant exposure window for lung cancer, and dietary agents may plausibly act in early or later stages of carcinogenesis. Clinically diagnosed lung cancer reflects a series of complex molecular genetic events that occur over many years, and the relevant windows for dietary exposures are uncertain. Case–control studies usually measure past diet during some reference period, whereas cohort studies tend to focus on current diet. Case–control studies have commonly been employed to study diet and lung cancer; many of these studies focus on diet during the 5 years preceding diagnosis. These studies provide direct information concerning dietary factors in the later stages of carcinogenesis. To the extent that such measures reflect usual adult (or lifetime) diet, these studies may also be relevant to the earlier stages of carcinogenesis. However, as lung cancer tends to be rapidly fatal, many case–control studies include data collected from deceased subjects' next-of-kin. Data from surrogate respondents are likely to be less accurate than self-reported data, and therefore introduce substantial misclassification.

Evidence concerning relationships between lung cancer and fruits, vegetables, micronutrients, phytochemicals, fat, body mass index, beverages, and meat intake is described in the section on dietary associations with lung cancer. To provide a guide for assessing the evidence for each dietary factor, evidence ratings from an objective assessment of the world's evidence on these topics, summarized in the continuous update report of the World Cancer Research Fund (WCRF), are used for factors that were assigned

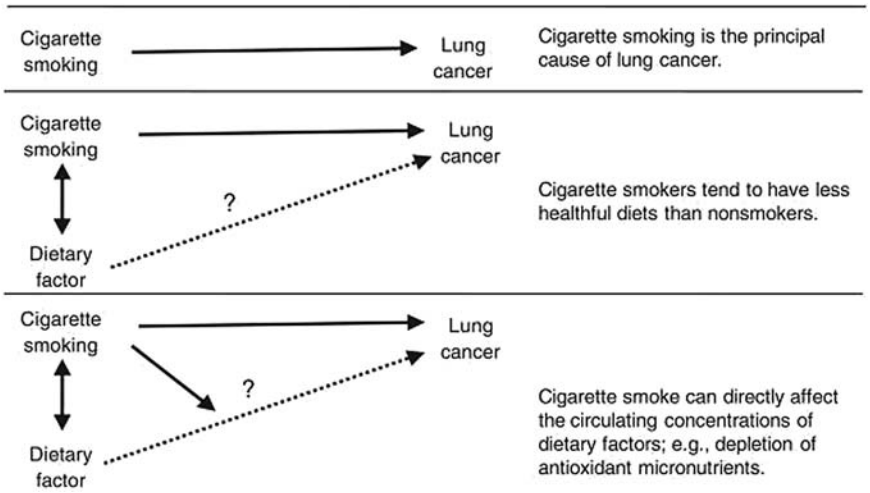


Fig. 1 Cigarette smoking complicates the study of diet and lung cancer.

evidence ratings. The rating scale used included evidence ratings of "convincing," "probable," "limited–suggestive", and "limited–no conclusion" for whether a dietary factor was associated with increased or decreased risk of lung cancer (World Cancer Research Fund, 2021).

Dietary associations with lung cancer

Fruit

In total, the evidence points toward greater levels of fruit consumption being inversely associated with lung cancer risk. The WCRF meta-analysis of 23 studies showed a statistically significant 8% decrease in the risk of lung cancer for every 100 g of fruit intake per day. Evidence of a dose response relationship was found for fruit intakes up to 200–300 g per day. The WCRF report rated the overall evidence as "limited–suggestive" that fruit consumption is inversely associated with lung cancer risk (World Cancer Research Fund, 2021). No clear pattern emerges when studies have examined specific fruits or classes of fruits. For example, apples and citrus fruits are associated with reduced risk of lung cancer in some studies but not in others (Wang et al., 2021).

Vegetables

Evidence for an inverse association between vegetable consumption and lung cancer risk parallels the evidence for fruit consumption. The WCRF meta-analysis of 20 studies showed a statistically significant 6% decrease in the risk of lung cancer for every 100 g of vegetable intake per day. Evidence of a dose-response relationship was found for vegetables intakes up to 300–400 g per day (World Cancer Research Fund, 2021). Consequently, the WCRF report rated the overall evidence as "limited–suggestive" that higher levels of vegetable intake are associated with decreased lung cancer risk.

In addition to vegetable intake as a whole, the results for a number of specific vegetables, such as carrots and cruciferous vegetables, have been consistently observed to be inversely associated with lung cancer risk. The association with cruciferous vegetable intake has tended to remain strong and robust even in studies that have carefully controlled for cigarette smoking (Lam et al., 2009). As discussed below, the growing evidence of an inverse association between cruciferous vegetable intake and lung cancer risk has bolstered interest in isothiocyanates as a promising chemopreventive agent.

Micronutrients

Two different strategies have been used to evaluate the relationship of micronutrients to lung cancer. One approach has been to use food-frequency questionnaires to measure micronutrient intake. A second approach is to use biomarkers, assaying the circulating concentrations of micronutrients. The former approach provides a better average measure of micronutrient "exposure," whereas the latter approach has the advantage of measuring micronutrient concentrations closer to cellular level, where the biologic effect is postulated to occur. However, a single assay of circulating micronutrient concentrations may fail to capture the biologically appropriate window of exposure.

The strongest evidence for the biomarker approach is generated from prospective cohort studies, where blood is collected from a population that is initially cancer-free and the population is then followed for the occurrence of lung cancer. The results of such prospective studies bolster the evidence supporting the premise that in general, higher circulating concentrations of carotenoids (α-carotene, β-carotene, β-cryptoxanthin, lutein, lycopene, and total carotenoids), are associated with reduced lung cancer risk. Circulating concentrations of retinol, tocopherol, and selenium have not been associated with a reduced risk of lung cancer in most studies. In contrast to prospective studies, biomarker studies of micronutrients based on measurements made after a patient is diagnosed with lung cancer are limited because a clinical diagnosis of lung cancer and its treatment and concomitant changes in diet can lead to decreases in circulating micronutrient levels, introducing the potential for reverse causality.

For dietary intake, the evidence is most abundant for vitamins A, C, and E, and carotenoids. The evidence relating measures of retinol intake to lung cancer risk provide "limited–suggestive" evidence that retinol is actually associated with increased risk of lung cancer (World Cancer Research Fund, 2021), contrary to Bjelke's initial hypothesis. However, studies of dietary intake of carotenoids and vitamin C point more consistently toward an inverse association. For example, in a systematic review of the evidence from prospective studies, both dietary intake and circulating concentrations of total carotenoids were associated with 20–30% lower risk of lung cancer in the highest-versus-lowest exposure categories. Associations of similar magnitude were observed for the provitamin A carotenoids β-carotene, α-carotene, and β-cryptoxanthin and for the nonprovitamin A carotenoids lycopene and lutein. Based on these data, the WCRF concluded that foods containing carotenoids were "probable" protective factors for lung cancer (World Cancer Research Fund, 2021).

Thus, prospective studies of both biomarkers and of dietary intake favor a protective association between carotenoids and lung cancer. However, uncertainty remains about whether a generally protective association is specific to these carotenoids or whether carotenoid intake merely serves as a marker of the intake of other protective substances or healthier dietary habits in general. The evidence for vitamin C and vitamin E is suggestive of a protective association (Huang et al., 2020; Luo et al., 2014), whereas the data on vitamin A, and selenium have yielded null findings (Fritz et al., 2011). Conversely, the use of vitamin B6 and B12 supplements were observed to increase lung cancer risk by 30%–40% among men in a prospective cohort study (Brasky et al., 2017).

Phytochemicals

Phytochemicals are low molecular weight molecules produced by plants. Of the many classes of phytochemicals, those studied in relation to lung cancer include phytoestrogens, flavonoids, and glucosinoids. The tumor promoting effects of steroid hormones can be blocked by phytoestrogens. Soya beans are a primary source of a specific class of phytoestrogens known as isoflavonoids. The relatively few studies to date of isoflavonoids in relation to lung cancer have not provided evidence of a link (World Cancer Research Fund, 2021).

Flavonoids are polyphenolic compounds found in many foods derived from plants; flavonoids often exhibit potent antioxidant activity. Some fruits contain high levels of flavonoids, such as apples (quercetin) and white grapefruit (naringin). Flavonoid intake has been at least weakly associated with reduced risk of lung cancer in many, but not all, of the studies to date.

Isothiocyanates are metabolites of the class of phytochemicals known as glucosinolates. Isothiocyanates could exert anticancer effects by blocking carcinogens *via* induction of phase II detoxification enzymes, such as glutathione *S*-transferase. Cruciferous vegetables contain high concentrations of glucosinolates, and hence consumption leads to higher endogenous isothiocyanate levels. As with cruciferous vegetables, lung cancer risk is also consistently lower with higher intakes or urinary levels of isothiocyanates.

A postulated biologic relationship between isothiocyanates and a common polymorphism in the *GSTM1* gene provides an example of a potential gene–diet interaction relevant to lung carcinogenesis. A focus in cancer epidemiology is to characterize inter-individual susceptibility to cancer by studying polymorphisms in genes involved in carcinogenic pathways, including how these genetic markers interact with environmental exposures to contribute to cancer risk. The role of glutathione *S*-transferase as a phase II detoxification enzyme has made a common polymorphism in the glutathione *S*-transferase M1 (*GSTM1*) gene of interest in relation to lung cancer. Compared to persons with the *GSTM1* present genotype, those with the *GSTM1* null genotype have a small but statistically significantly higher risk of lung cancer.

When isothiocyanates have been studied in combination with *GSTM1*, the decreased risk of lung cancer associated with isothiocyanates has been especially pronounced in persons with the *GSTM1* null genotype. This association could represent the cancer blocking activity of isothiocyanates being allowed to play an enhanced role in *GSTM1* null individuals because they are not being metabolized as quickly as in those with the *GSTM1* present genotype. This example illustrates the potential interactions between genetic and dietary factors. Integrating genetic and epigenetic markers into the study of nutritional factors provides a mechanistically-based approach that holds promise for advancing understanding of the complex role of diet in the etiology of lung cancer.

Fat

Evidence that dietary fat may facilitate tumor growth was reported as early as 1940. Correlation exists between international or regional dietary fat consumption and lung cancer mortality. In case–control studies, total fat intake is consistently associated with lung cancer risk, with less consistent results for saturated fat, unsaturated fat, and cholesterol intake. In a pooled analysis of data from ten prospective cohort studies, lung cancer risk was weakly positively associated with total fat and saturated fat intake, whereas a weak inverse association was observed for polyunsaturated fat intake (Yang et al., 2017). In the WCRF report the evidence was rated as "limited–no conclusion" that total dietary fat is associated with increased lung cancer risk (World Cancer Research Fund, 2021). With respect to specific foods, the same level of evidence was applied to butter.

Body mass index

In contrast to the situation for most types of cancer, prospective cohort studies consistently show a strong inverse relationship between body mass index (BMI) and lung cancer risk (Wang et al., 2018). These remarkably strong, consistent findings clearly demonstrate that leanness is statistically associated with lung cancer risk. The key remaining question is whether this association is genuine or whether it is indirect. Confounding by cigarette smoking is a viable explanation for these findings because cigarette smoking is strongly associated both with the risk of lung cancer and with leanness. However, the inverse association between BMI and lung cancer has been observed among never smokers (Zhu and Zhang, 2018). Therefore, research is needed to further test the hypothesis that leanness is a susceptibility factor for lung cancer, and if so to advance understanding of the underlying biological mechanisms.

Beverages

Many beverages, including alcohol, coffee, tea, and milk have been studied for a possible link to lung cancer, although the issue of potential confounding by cigarette smoking recurs. The majority of studies of alcohol drinking in relation to lung cancer risk that have adjusted for age and cigarette smoking have observed either null or weak associations. After adjusting for cigarette smoking, the WCRF meta-analysis found a significant dose-response relationship between ethanol drinking and increased lung cancer risk. The association was not significant for specific alcoholic beverages, but a significant inverse association was observed for wine. The mechanistic basis for an association between alcohol drinking and lung cancer is plausible based on factors such as the presence of carcinogens such as acetaldehyde in alcoholic beverages and the ability of alcohol to lead to increased cellular penetration of other carcinogens. The WCRF report concluded the evidence is "limited–suggestive" that alcoholic drinking is associated with increased lung cancer risk (World Cancer Research Fund, 2021).

Some studies have observed heavy coffee consumption to be associated with an elevated risk of lung cancer after adjustment for cigarette smoking, but a host of case–control studies have generated findings that fluctuate around the null (Galarraga and Boffetta, 2016). The issue of confounding between coffee drinking and other health behaviors, particularly cigarette smoking, has not been addressed adequately, indicating that much stronger evidence is needed for coffee drinking to be considered a risk factor for lung cancer. Despite numerous *in vitro* and *in vivo* studies that have observed potential tumor-inhibitory effects of tea, the epidemiologic evidence provides inconsistent link between tea drinking and lung cancer risk. A meta-analysis of prospective cohort studies showed a greater risk of lung cancer risk among tea drinkers ($\geq$2 cups/day) who were non-smokers. The biological link modulating this risk needs to be explored (Zhu et al., 2021). Based on evidence such as this, the WCRF report concluded the evidence is "limited–no conclusion" that tea is associated with lung cancer risk (World Cancer Research Fund, 2021).

The associations observed between milk drinking and lung cancer depend on milk fat content. Milk drinking is not strongly associated with lung cancer risk when milk fat content is ignored. The associations between whole milk and lung cancer tend to be either null or in the direction of increased risk, whereas the associations for reduced fat or nonfat milk tend to be either null or in the protective direction (Yang et al., 2016). Perhaps milk consumption, including type of milk, is merely serving as a marker of fat intake, which as noted above tends to be associated with increased lung cancer risk. Consistent with the equivocal nature of the evidence and concerns about confounding by cigarette smoking, the WCRF report did not provide evidence ratings for any of these beverages in relation to lung cancer risk (World Cancer Research Fund, 2021).

Drinking water can be a route of exposure to environmental contaminants. This is exemplified by the clear increase in lung cancer risk associated with drinking water that is contaminated with high levels of arsenic (World Cancer Research Fund, 2021). Based on studies conducted in geographic regions where drinking water is contaminated with high concentrations of arsenic, the WCRF report rated the evidence as "convincing" that high concentrations of arsenic in drinking water is a risk factor for lung cancer (World Cancer Research Fund, 2021).

Meat and fish

A pooled meta-analysis found an increased risk of lung cancer associated with higher intakes of red meat and processed meat (Xue et al., 2014). The cooking method may play a role, as heterocyclic amines from cooked meat may contribute to an increased lung cancer risk. Based on the slight trending of the results toward increased risk, the WCRF report rated the evidence for both red meat intake and processed meat intake to be "limited–suggestive" of increased risk (World Cancer Research Fund, 2021). A meta-analysis provided evidence that fish consumption may be inversely associated with lung cancer risk (Song et al., 2014), but further research is needed to determine if this association is genuine. The WCRF report rated this evidence as "limited-no conclusion" (World Cancer Research Fund, 2021).

Diet and prevention

Chemoprevention trials

Three large-scale, randomized, double-blind, placebo-controlled trials were undertaken to test the hypothesis that β-carotene supplementation protects against lung cancer (World Cancer Research Fund, 2021). All three studies indicated that β-carotene supplementation in later adulthood does not protect against lung cancer. In fact, β-carotene supplementation was unexpectedly associated with an increased risk of lung cancer among the high-risk populations of heavy smokers in the Alpha-Tocopherol, Beta-Carotene (ATBC) Cancer Prevention Study (Omenn et al., 1996) and among the smokers and asbestos-exposed workers in the Carotene and Retinol Efficacy Trial (CARET) Study (Virtamo et al., 2003). The WCRF thus rated this strong, consistent evidence from two randomized controlled trials as "convincing" that β-carotene supplement increase lung cancer risk in current smokers (World Cancer Research Fund, 2021). These experimental results thus not only failed to corroborate the evidence from observational studies, but clearly demonstrated that β-carotene supplementation increased risk in groups at the highest risk of lung cancer. The combined results of multiple randomized controlled trials of vitamin E supplements are clearly consistent with no effect on lung cancer risk.

Observation versus experiment

The ATBC and CARET studies enrolled older, high-risk individuals who had high cumulative exposure to tobacco smoke or asbestos (Virtamo et al., 2003; Omenn et al., 1996). The results therefore presumably apply mainly to the latter stages of carcinogenesis. The doses administered were far higher than the normal dietary range, and the dose–response relationship for preventive effects, anticipated from the observational evidence, may not be applicable. Because antioxidant nutrients may exert their protective effect in the earlier stages of carcinogenesis, β-carotene may have been administered too late to halt the evolution of cellular changes that lead to lung cancer. Alternatively, compounds present in fruits and vegetables other than the micronutrients studied in the trials may protect against lung cancer. The protective associations for fruit and vegetable consumption were allied to the micronutrient hypothesis, but the results of the chemoprevention trials of both β-carotene and vitamin E raise questions about the potential pay-off from large trials designed to test single micronutrients, unless there is a strong mechanistic basis combined with substantial observational evidence pointing to an individual micronutrient as the primary protective agent. Indeed, fruits and vegetables contain an

abundance of antioxidants and phytochemicals with diverse anticarcinogenic activities. Then again, fruit and vegetable intake may be acting as a marker of a healthier lifestyle that is associated with lower cancer risk.

Conclusions

Knowledge of the relationship between diet and lung cancer has increased tremendously during past 50 years. Promising leads suggest that nutritional factors could have a substantial impact on lung cancer risk in humans. In general, persons who eat more fruits and vegetables have a lower risk of lung cancer than persons who consume less of these foods. The specific constituents of fruits and vegetables that may confer protection are unknown. An important unanswered question is whether fruits and vegetables directly confer protection against lung cancer or whether estimates of fruit and vegetable consumption are indicators of differences between individuals who eat healthy and unhealthy diets that are leading to uncontrolled confounding. Nevertheless, the protective association noted for fruit and vegetable consumption has the potential to contribute to prevention. A diet adequate in fruit and vegetables is already known to be prudent for preventing chronic diseases in general.

Even for factors such as fruit and vegetable consumption, the highest category of intake is usually associated with at most a halving in the risk of lung cancer. An association of this magnitude could result from residual confounding by cigarette smoking. Future research that provides the strictest possible control for cigarette smoking, such as studies that match cases and controls in the study design or studies limited to never smokers, will help to resolve longstanding questions about dietary factors and lung cancer by addressing head-on the persistent concern about residual confounding by cigarette smoking.

Advances in understanding of the role of diet in lung cancer etiology should not obscure the fact that cigarette smoking is the predominant cause of lung cancer. Many important questions about the complex relationship between diet and lung cancer remain, but the primary way that the lung cancer epidemic will be controlled is to prevent the uptake of cigarette smoking among children and effectively assist addicted smokers to stop smoking cigarettes.

References

Alberg, A.J., Nonemaker, J., 2008. Who is at high risk for lung cancer? Population-level and individual-level perspectives. Semin. Respir. Crit. Care Med. 29 (3), 223–232.

Alberg, A.J., Brock, M.V., Ford, J.G., Samet, J.M., Spivack, S.D., 2013. Epidemiology of lung cancer: diagnosis and management of lung cancer: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 143 (5), e1S–e29S.

Allemani, C., Matsuda, T., Di Carlo, V., et al., 2018. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet 391 (10125), 1023–1075.

American Cancer Society, 2021. 2021. Cancer Facts & Figures. Available at. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2021.html. (Accessed 11 June 2021).

Bade, B.C., Cruz, C.S.D., 2020. Lung cancer 2020: epidemiology, etiology, and prevention. Clin. Chest Med. 41 (1), 1–24.

Brasky, T.M., White, E., Chen, C.L., 2017. Long-term, supplemental, one-carbon metabolism–related vitamin B use in relation to lung cancer risk in the Vitamins and Lifestyle (VITAL) Cohort. J. Clin. Oncol. 35 (30), 3440–3448.

Bray, F.I., Weiderpass, E., 2010. Lung cancer mortality trends in 36 European countries: secular trends and birth cohort patterns by sex and region 1970–2007. Int. J. Cancer 126 (6), 1454–1466.

Eckel, S.P., Cockburn, M., Shu, Y.H., et al., 2016. Air pollution affects lung cancer survival. Thorax 71 (10), 891–898.

Erren, T.C., Jacobsen, M., Piekarski, C., 1999. Synergy between asbestos and smoking on lung cancer risks. Epidemiology 405–411.

Fritz, H., Kennedy, D., Fergusson, D., et al., 2011. Selenium and lung cancer: a systematic review and meta-analysis. PLoS One 6 (11), e26259.

Galarraga, V., Boffetta, P., 2016. Coffee drinking and risk of lung cancer—a meta-analysis. Cancer Epidemiol. Prev. Biomarkers 25 (6), 951–957.

Huang, J., Weinstein, S.J., Yu, K., Männistö, S., Albanes, D., 2020. A prospective study of serum vitamin E and 28-year risk of lung cancer. JNCI 112 (2), 191–199.

Lam, T.K., Gallicchio, L., Lindsley, K., et al., 2009. Cruciferous vegetable consumption and lung cancer risk: a systematic review. Cancer Epidemiol. Prev. Biomarkers 18 (1), 184–195.

Luo, J., Shen, L., Zheng, D., 2014. Association between vitamin C intake and lung cancer: a dose-response meta-analysis. Sci. Rep. 4 (1), 1–7.

Malhotra, J., Malvezzi, M., Negri, E., La Vecchia, C., Boffetta, P., 2016. Risk factors for lung cancer worldwide. Eur. Respir. J. 48 (3), 889–902.

McIntyre, A., Ganti, A.K., 2017. Lung cancer—a global perspective. J. Surg. Oncol. 115 (5), 550–554.

Omenn, G.S., Goodman, G.E., Thornquist, M.D., et al., 1996. Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. N. Engl. J. Med. 334 (18), 1150–1155.

Pesch, B., Kendzia, B., Gustavsson, P., et al., 2012. Cigarette smoking and lung cancer—relative risk estimates for the major histological types from a pooled analysis of case–control studies. Int. J. Cancer 131 (5), 1210–1219.

Shankar, A., Dubey, A., Saini, D., et al., 2019. Environmental and occupational determinants of lung cancer. Transl. Lung Cancer Res. 8 (Suppl. 1), S31.

Steenland, K., Loomis, D., Shy, C., Simonsen, N., 1996. Review of occupational lung carcinogens. Am. J. Ind. Med. 29 (5), 474–490.

Song, J., Su, H., Wang, B.L., Zhou, Y.Y., Guo, L.L., 2014. Fish consumption and lung cancer risk: systematic review and meta-analysis. Nutr. Cancer 66 (4), 539–549.

Sung, H., Ferlay, J., Siegel, R.L., et al., 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA 71 (3), 209–249.

U.S. Department of Health and Human Services, 2000. Synergy Between Asbestos and Smoking on Lung Cancer Risks. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease, Atlanta, GA, p. 48.

Virtamo, J., Pietinen, P., Huttunen, J.K., et al., 2003. Incidence of cancer and mortality following alpha-tocopherol and beta-carotene supplementation: a postintervention follow-up. JAMA 290 (4), 476–485.

Wang, J., Xu, H., Zhou, S., et al., 2018. Body mass index and mortality in lung cancer patients: a systematic review and meta-analysis. Eur. J. Clin. Nutr. 72 (1), 4–17.

Wang, J., Gao, J., Xu, H.L., et al., 2021. Citrus fruit intake and lung cancer risk: a meta-analysis of observational studies. Pharmacol. Res. 166, 105430.

Warren, G.W., Alberg, A.J., Cummings, K.M., Dresler, C., 2020. Smoking cessation after a cancer diagnosis is associated with improved survival. J. Thorac. Oncol. 15 (5), 705–708.

Wei, C., Han, Y., Spitz, M.R., et al., 2011. A case–control study of a sex-specific association between a 15q25 variant and lung cancer risk. Cancer Epidemiol. Prev. Biomarkers 20 (12), 2603–2609.
World Cancer Research Fund, 2021. Diet, Nutrition, Physical Activity and Lung Cancer. Available at. https://www.wcrf.org/wp-content/uploads/2021/02/Summary-of-Third-Expert-Report-2018.pdf. (Accessed 11 June 2021).
Xue, X.J., Gao, Q., Qiao, J.H., Zhang, J., Xu, C.P., Liu, J., 2014. Red and processed meat consumption and the risk of lung cancer: a dose-response meta-analysis of 33 published studies. Int. J. Clin. Exp. Med. 7 (6), 1542–1553.
Yang, Y., Wang, X., Yao, Q., Qin, L., Xu, C., 2016. Dairy product, calcium intake and lung cancer risk: a systematic review with meta-analysis. Sci. Rep. 6 (1), 1–8.
Yang, J.J., Yu, D., Takata, Y., et al., 2017. Dietary fat intake and lung cancer risk: a pooled analysis. J. Clin. Oncol. 35 (26), 3055.
Zhu, H., Zhang, S., 2018. Body mass index and lung cancer risk in never smokers: a meta-analysis. BMC Cancer 18 (1), 1–10.
Zhu, J., Smith-Warner, S.A., Yu, D., et al., 2021. Associations of coffee and tea consumption with lung cancer risk. Int. J. Cancer 148 (10), 2457–2470.

Celiac disease

Balakrishnan S. Ramakrishna, Institute of Gastroenterology, SRM Institutes for Medical Science, Chennai, Tamil Nadu, India

© 2023 Elsevier Ltd. All rights reserved.

This is an update of V. Nehra, E.V. Marietta, J.A. Murray, Celiac Disease, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 298–306, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00046-5.

Introduction 162
Pathogenesiss 162
Genetics 163
Epidemiology 164
Pathology 164
Clinical features 165
Associations 166
Diagnosis 167
Treatment 168
Complications 168
Conclusion 169
References 169
Further reading 170

Key points

- Celiac disease is a disease caused by sensitivity to gluten protein in dietary wheat, barley or rye in genetically susceptible individuals
- The HLA-DQ genes have a strong genetic influence on development of celiac disease
- The disease primarily affects the small intestine, although it may manifest in a number of different ways
- The diagnosis is made through use of serological tests followed, in most cases, by duodenal biopsy
- Treatment is lifelong and complete withdrawal of wheat from the diet

Introduction

Celiac disease is a disorder primarily affecting the small intestine, which manifests in genetically predisposed individuals who consume wheat, barley or rye. The disease is due to an aberrant immune response to storage proteins found in these cereals, and provides an example of the interaction between host genetics and the environment. The disease is rare or uncommon in areas of the world where wheat is not consumed, for instance in South East Asian countries where rice is the staple cereal or in sub-Saharan Africa where maize is the staple cereal. On the other hand, it is very prevalent in areas such as Europe and North America where wheat is regularly consumed in the diet. From the evolutionary point of view, celiac disease can be considered as a consequence of the change to a pastoral life and cultivation of cereals. The name celiac disease is attributed to Aretaeus, a Greek physician in the first century CE, who described a similar disease and called it "koeliakos" (Paveley, 1988). No major recorded descriptions of the disease exist thereafter until the detailed description of the disease in children in 1888 published by Dr. Samuel Gee, a British pediatrician (Paveley, 1988). The recognition that wheat ingestion is responsible for celiac disease is credited to the Dutch pediatrician Willem-Karel Dicke (Paveley, 1988). Celiac disease was recognized to be quite prevalent in Europe, and subsequently there has been recognition that the disease is equally prevalent in North America. The availability of serological tests to screen for the disease has resulted in recognition of the magnitude of disease burden in many other parts of the world. The diagnosis of celiac disease is initially made serologically by testing for the presence of antibody to tissue transglutaminase, the finding of characteristic changes on small intestinal biopsy, and ultimately by a response (clinical, biochemical and histological) to complete withdrawal of gluten from the diet.

Pathogenesiss

The pathogenesis of celiac disease involves interactions between host genetics and environmental factors, the most significant of the latter being ingestion of wheat, barley or rye (Kagnoff, 2007). Celiac disease occurs due to an immune reaction to the storage proteins of wheat, rye and barley, which are closely related members belonging to the Triticeae tribe of the grass family. Their storage

proteins are rich in proline and glutamine. The disease-inducing proteins in wheat are called gluten (consisting of two major protein types gliadins and glutenins) while the related proteins in rye and barley are called hordeins and seccalins. Oats do not belong to the Triticeae tribe and have proteins called avenins which have a much lower proline content, and are only rarely associated with celiac disease. Rice, maize, sorghum and millets belong to very distantly related subfamilies of grass and their storage proteins are not associated with celiac disease.

Celiac disease develops only in genetically susceptible individuals. The strongest genetic association of celiac disease is with the HLA system (Kagnoff, 2007). The HLA complex, located on chromosome 6p21, contains more than 200 genes and over 3000 known alleles. The HLA class II molecules—DP, DQ and DR-are involved in exogenous peptide antigen presentation to T cells. The expression of either HLA-DQ2 or of HLA-DQ8 antigen is a pre-requisite for the development of celiac disease. The HLA DQ molecules are proteins that are found on the cell surface of antigen-presenting cells mainly macrophages, dendritic cells and B cells. They are composed of two chains, the α and the β chains. These are encoded by the genes HLA-DQA1 and HLA-DQB1 respectively. Many persons have two α and two β chain variants resulting in the presence of four different DQ proteins in the same individual. The β chain of HLA-DQ2 protein is encoded by one of two alleles either HLA-DQB1*0201 or HLA-DQB1*0202. The α chain of HLA-DQ2 is encoded by either HLA-DQA1*0201 or HLA-DQA1*0501. The HLA-DQ8 protein is encoded by the gene HLA-DQB1*0302.

HLA-DQ2 and HLA-DQ8 can bind proline-rich peptides containing deamidated glutamine residues. An enzyme called tissue transglutaminase, probably produced by fibroblasts, binds to gliadin-derived peptides and deamidates glutamine to glutamic acid, resulting in negatively charged residues and improved binding of the gliadin peptides to pockets in the DQ2 or DQ8 molecules on antigen presenting cells. There may be up to 50 peptides in wheat, 60 in rye and 35 in barley that can bind to HLA-DQ2 or -DQ8. $CD4^+$ T lymphocytes in the lamina propria of the small intestine are activated when these peptides bind to HLA molecules. The release of tissue transglutaminase into the inflamed gut also results in a strong autoimmune response to tissue transglutaminase with high levels of circulating IgA specific for tissue transglutaminase (tTG-IgA) present in untreated celiac patients. Thus, the main characteristics of celiac disease are the DQ2/DQ8 restricted responses to gliadin peptides, the strong intestinal T cell response to deamidated gliadin peptides, and the production of tTG-IgA (Kagnoff, 2007).

Gluten peptides reach the lamina propria of the intestine through paracellular and transcellular pathways (Kagnoff, 2007). After binding to and presentation by the HLA-DQ2 or DQ8 proteins, they activate gluten-specific $CD4^+$ T helper 1 cells in the lamina propria which secrete the cytokines and chemokines necessary to induce inflammation with crypt hyperplasia and villus atrophy that are characteristic of the disease. Apart from the adaptive immune response involving T helper 1 cells, it is likely that there is also an innate immune response to certain proteins from wheat that results in epithelial cell activation and increased intraepithelial lymphocytes (IEL). The innate immune activation of IEL results in the expression of MICA (MHC Class I Chain-related A protein), which binds to and activates a receptor found on NK cells, γδ T cells and on some $CD4^+$ and $CD8^+$ T cells. This results in antigen-dependent cell-mediated cytotoxicity with release of cytokines as well as cytotoxic granules containing perforin and granzymes that enter target epithelial cells and trigger apoptosis leading to cell death. Interleukin-15 (IL-15) is the cytokine that is implicated in linking innate to adaptive immunity in celiac disease. This cytokine is produced by dendritic cells and macrophages in response to exposure to gluten peptides, and activates intraepithelial lymphocytes. The exact nature by which epithelial cell destruction occurs and the roles of cytotoxic and regulatory T cells is not yet fully defined. The cytokines secreted by T cells may be important in driving matrix remodeling, that leads to villus atrophy and crypt hyperplasia. Interferon-γ from activated T cells appears to be involved as an agent of damage, and this may activate macrophages that secrete tumor necrosis factor-α and matrix metalloproteinases MMP-12 and MMP-13 which lead to proteolysis of the matrix followed by villus-crypt remodeling.

Genetics

The combination of HLA-DQA1*0501 and DQB1*0201 that encodes the HLA-DQ2 heterodimer can be encoded by alleles on the same haplotype, or by one subunit each from paternal and maternal haplotypes (Trinka et al., 2001). Homozygosity for the above combination or possession of a second DQB1*02 allele increases the susceptibility to celiac disease in European populations. This "gene dose" effect is supported by evidence that T lymphocyte proliferation and cytokine responses to gluten depend on the DQ type and gene dose. An extended haplotype comprising A1-B8-DR3-DQ2 is found in about 10% of European subjects and is associated with other autoimmune diseases including type 1 diabetes and autoimmune thyroid disease. In India, celiac disease may be detected in up to 11% of patients with type 1 diabetes mellitus.

HLA-DQ2 and DQ8 proteins are expressed in 30% of the population, yet only 2–5% of the appropriate HLA gene carriers develop celiac disease, suggesting that there are either other genes or other environmental factors or both that contribute to the development of celiac disease in the susceptible host. The evidence suggests that the HLA background leads to a usually suppressed T cell response to gluten in DQ2/DQ8 carriers, and that this suppression is removed in some individuals due to either other genes or environmental conditioning leading to overt celiac disease.

Several lines of evidence suggest that genetic influences outside the HLA system are also at work in determining susceptibility to celiac disease. When monozygotic twins are compared to HLA allele-identical dizygotic twins (i.e., same gene dose in both), the concordance for celiac disease in dizygotic twins is less than among monozygotic twins. Similarly, while HLA-DQ2 prevalence is equally high among Sardinians and the Saharawi tribes, the prevalence of celiac disease in Sardinia is only one-fifth of that in the Saharawi. The identification of genes outside the HLA complex that contribute to development of CD has been facilitated by

genome-wide association studies (GWAS) in multiple populations. Several non-HLA risk loci have been identified for celiac disease through GWAS (Kumar et al., 2021), the strongest association being with a locus close to the IL-2/IL-21 locus on Chromosome 4q27. Modifier genes may influence expression of celiac disease through mechanisms such as the release of specific cytokines or other molecules that up-regulate HLA-DQ2 and HLA-DQ8 proteins on antigen presenting cells. Together, these non-HLA genes account for only 3–4% of the genetic susceptibility to celiac disease compared to 30–40% for the HLA genes.

Epidemiology

Celiac disease was originally identified to affect around 1% of individuals in Northern Europe, but has now been recognized to occur around the world (Lebwohl and Rubio-Tapia, 2021). Meta-analysis of studies around the world indicate that the global prevalence of biopsy-confirmed celiac disease is around 0.7% being highest in Europe and Oceania. There are major geographical differences in prevalence that appear to be driven by differences in diet and by the population prevalence of HLA-DQ2 and -DQ8. The disease is rare in Southeast Asia and South India where rice is the predominant dietary cereal, and in sub-Saharan Africa where maize constitutes the main dietary cereal. On the other hand, prevalence is high in Europe, North America and North India where wheat is regularly consumed in the diet. The prevalence of celiac disease among the Saharawi, an inbred North African tribe with high prevalence of DQ2 and high wheat intake, is close to 5%.

The prevalence of celiac disease in different populations around the world mirrors also the genetic susceptibility to celiac disease. The most prominent factor hitherto associated with genetic susceptibility to CD is the HLA system. The HLA complex is a highly polymorphic region located on chromosome 6p21, and containing more than 200 genes and over 3000 known alleles. The HLA class II molecules—DP, DQ and DR-are involved in exogenous peptide antigen presentation to T cells. The expression of either HLA-DQ2 or of HLA-DQ8 is considered to be a necessary pre-requisite for the development of celiac disease. The population prevalence of CD in Europe, North America, North Africa, the Middle East, and North India averages 1%. In these populations, either HLA-DQ2 or HLA-DQ8 is present in approximately 40% of the population. Among the Saharawi tribes of Africa, who are of mixed Arab-Berber ancestry with a high degree of inbreeding, there is a 5% prevalence of CD probably the highest in the world. In this population, carriage of HLA-DQ2 or HLA-DQ8 is around 60% of the population. On the other hand, celiac disease is virtually unknown in Japan and in Burkina Faso where most of the population does not express either HLA-DQ2 or HLA-DQ8.

Another evidence for a strong genetic effect comes from twin studies, with concordance for celiac disease being observed in 75% of monozygotic twins but only in 10% of dizygotic twins. There is a familial clustering of celiac disease with a 10% prevalence of celiac disease in first degree relatives compared to the general population. The relative risk for developing celiac disease in a sibling of an index patient varies from 20 to 60.

The prevalence of celiac disease in many regions has steadily increased in the past six decades, even after adjusting for increased physician and patient awareness and increased sensitivity of the diagnostic tests. This suggests the presence of an altered environment contributing to disease expression, as the genetic makeup of the population is unlikely to have changed. Putative environmental factors could include increased consumption of wheat, alteration in type or content of gluten in wheat, or factors affecting the immune response. Evidence indicates that the modern hexaploid wheat, *Triticum aestivum*, is more immunogenic to T cells than diploid or tetraploid wheats cultivated a couple of centuries ago. Thus, human agricultural practices may have contributed to the increasing incidence of celiac disease. Changes in human dietary practices are probably also responsible. The Saharawi traditionally had prolonged breastfeeding, and later consumed camel's milk, dates, and meat and very little cereal. Over time, breastfeeding duration reduced and wheat flour products especially bread became the staple, leading to a high incidence of celiac disease. The time of exposure to gluten-containing foods appears to be also important. Celiac disease incidence is lower in Denmark and Estonia compared to Sweden, and this has been attributed to higher infant consumption of gluten is the latter country. In an apparent natural experiment in Sweden, a significant increase of celiac disease was noted in children who were weaned between the years 1984–1996, the incidence during this period being nearly three times as much as in the periods before or after (Myléus et al., 2009). This coincided with changes in the wheat protein content of infant foods and early weaning during the period under consideration, and led to the recommendation that complementary feeds should not be introduced before four months of age and not much later than 6 months. Breast feeding at the time of introduction of wheat is now considered to have a protective effect against development of the disease, suggesting that it may induce tolerance to dietary gluten. The "hygiene hypothesis" has also been advanced as an explanation for an increasing incidence of celiac disease in the modern world. In this hypothesis, reduced childhood exposure to intestinal infections or alterations in the gut microbiome may alter T cell conditioning away from the normal balance that is maintained between T helper 1, T helper 2 and T regulatory mechanisms, leading to a dysregulated immune system and increased gluten sensitivity. On the other hand, case-control studies suggest a mild increase in risk for celiac disease in children who had multiple enteric infections in infancy, the speculation being that these increase intestinal permeation of the antigens that induce celiac disease.

Pathology

The pathology of celiac disease was first appreciated after the introduction of small intestinal capsule biopsy by Margot Shiner in 1956. The disease affects mainly the proximal small intestine, but the changes may be patchy. The pathology of jejunal capsule

Table 1 Histological classification of small bowel changes in celiac disease.

	IEL/100 enterocytes				
Marsh type	*Jejunum*	*Duodenum*	*Crypt length*	*Villus height*	*Comments*
0	<40	<30	Normal	Normal	Normal
1	>40	>30	Normal	Normal	Celiac in remission, other enteropathies
2	>40	>30	Increased	Normal	Treated celiac, dermatitis herpetiformis
3a	>40	>30	Increased	Mild atrophy	Symptomatic celiac disease
3b	>40	>30	Increased	Marked atrophy	
3c	>40	>30	Increased	Complete atrophy	

IEL = intraepithelial lymphocytes.
Original Marsh classification modified by Oberhuber et al. (1999).

biopsies was described in detail by Michael Marsh, who originated a classification of histological severity (Oberhuber et al., 1999). Currently the histology is evaluated on mucosal biopsies taken from the duodenum or jejunum during fiberoptic endoscopy of the upper gastrointestinal tract. The Marsh classification, as modified by Oberhuber, remains widely used in the reporting of duodenal mucosal biopsies from patients with celiac disease (**Table 1**). In brief, the presence villous atrophy as well as crypt elongation or hypertrophy, along with increased intraepithelial cells, remains the cornerstone of diagnosis. Type 1 and type 2 in the Marsh classification are not considered compatible with the diagnosis of untreated celiac disease, but may occur in patients on gluten free diet. The Type 2 Marsh change may be seen in dermatitis herpetiformis. Complete villous atrophy (**Fig. 1**), seen in some patients with celiac disease, is seldom if never not found in other small bowel mucosal diseases.

Clinical features

Celiac disease may present in a wide variety of ways. In children, the disease is diagnosed usually between the first to seventh year of life, with the child remaining quite healthy until the introduction of cereals into the diet. Young children may develop chronic diarrhea, failure to thrive, muscle wasting, abdominal distension, vomiting, and abdominal pain. Older children may present with anemia, short stature, rickets, behavioral disturbances, or poor performance in school. In some children constipation, pseudo-obstruction, and intussusception may be seen. Dental enamel defects involving secondary dentition as well as neurological syndromes and epilepsy with intracranial calcification have been reported in children. In adults, celiac disease may be overt in presentation with classic gastrointestinal symptoms of diarrhea, weight loss, and abdominal pain. These patients are categorized as having typical or classical celiac disease. Recurrent oral ulcers may be a feature of the disease. The presence of diarrhea or steatorrhea, which occurs in approximately 50% of patients, indicates severe disease and malabsorption.

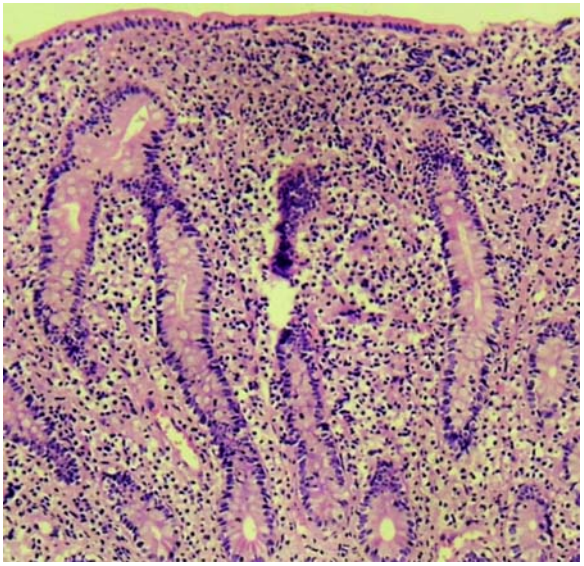


Fig. 1 Endoscopic duodenal biopsy from a patient with celiac disease showing complete loss of villi and marked elongation of crypts along with inflammatory infiltration of the lamina propria.

Celiac disease may also be diagnosed in individual with non-gastrointestinal symptoms, particularly in adults. Here the presentation may include anemia, abnormal liver tests, osteopenic bone disease, neurological symptoms, or menstrual abnormalities. These individuals are categorized as atypical celiac disease. Anemia may be secondary to iron and/or folate deficiency. Six to ten percent of patients with unexplained iron deficiency anemia may have celiac disease in the absence of any other features suggestive of the disease. They have predominant extra-gastrointestinal symptoms, but test positive to antibody to tissue transglutaminase, and show changes consistent with celiac disease on biopsy. Other symptoms present in this group include musculoskeletal signs and symptoms, such as bone pains, proximal muscle weakness, osteopenia, osteoporosis, and fractures; neurological manifestations including intractable epilepsy, cerebellar ataxia, peripheral neuropathy, dementia, myoclonus and myelopathy; gynecological disorders including late menarche, unexplained infertility, or unexplained abortions; and dental enamel defects.

A third group of individuals remains asymptomatic, but are identified by screening serology, and their duodenal biopsies show changes characteristic of celiac disease. Yet another group of asymptomatic individuals may be identified by positive serology during population surveys, but will have normal duodenal mucosal biopsies. These individuals are considered to have potential celiac disease. Thus, celiac disease has been characterized as an iceberg disease (**Fig. 2**). Typical celiac disease tends to be more common in children, while atypical disease is more common in adults. Although the term "celiac iceberg" has been used, the proportion of the disease burden that remains below the surface varies with the age of the cohort under study and with the population studied. Thus, most children with the disease are probably symptomatic, while many or most adults may remain asymptomatic. The clinical significance of potential celiac disease (positive serology, no histologic change or symptoms) remains unclear with some studies suggesting long term risks including malignancies, while others found no obvious long term risks for these individuals.

Associations

Celiac disease may be considered a multi-system disorder, originating in and primarily affecting the gastrointestinal tract but also affecting a number of other systems (Therrien et al., 2020). Underlying many of these associations is a thread of autoimmunity, hence the term celiac autoimmunity is also used to denote the spectrum. Recent studies indicate that the mechanism may involve molecular mimicry of gliadin antigens by microbially derived peptides. Conditions that have been associated with celiac disease are listed in **Table 2**. In these groups of patients, screening is sometimes indicated to diagnose celiac disease. Screening is done by serological testing, and patients who test positive are advised small bowel biopsy. The high-risk groups include first-degree relatives of confirmed cases of celiac disease, those with type 1 diabetes mellitus, Down's syndrome, Turner's syndrome, and unexplained dental enamel deficits, and children with unexplained short stature.

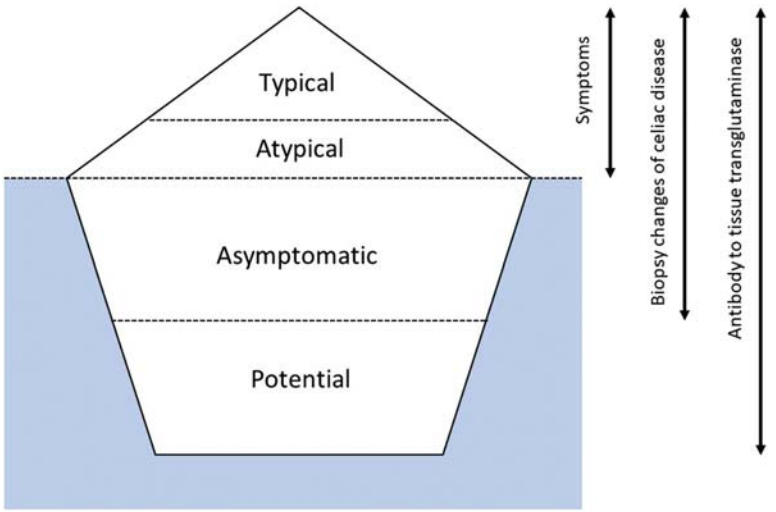


Fig. 2 Schematic representation of the celiac iceberg showing the different phenotypes. The proportions of typical, atypical, asymptomatic and potential celiac disease likely vary from population to population.

Table 2 Associations of celiac disease.

Hematological	Iron deficiency anemia
	Folate deficiency anemia
	Pure red cell aplasia
Endocrine	Type 1 diabetes mellitus
	Graves disease
	Hypothyroidism
	Autoimmune polyglandular syndrome type II
Reproductive	Unexplained infertility
	Recurrent miscarriage
	Male infertility
Liver	Cryptogenic cirrhosis
	Autoimmune hepatitis
	Primary biliary cirrhosis
	Autoimmune cholangitis
	Non-cirrhotic portal hypertension
Dermatological	Dermatitis herpetiformis
	Lichen sclerosus
	Psoriasis
	Oral aphthous ulcers
Rheumatological	Pediatric rheumatological disease
	Metabolic bone disease
Dental	Enamel hypoplasia
Neuropsychiatric	Drug resistant epilepsy
	Myoclonus ataxia
	Autism spectrum disorders

Diagnosis

Anemia due to iron deficiency or folate deficiency, hypoalbuminemia, abnormal transaminases, and evidence of metabolic bone disease may be present. The specific diagnosis of celiac disease has been facilitated by the development of serological tests, in particular the identification of IgA antibodies to human tissue transglutaminase (IgA anti-tTG). These enzyme linked immunosorbent assays use capture antigens isolated from human erythrocytes or antigens produced by recombinant technology. These tests are very sensitive and, in patients with a high pre-test probability of celiac disease, they are quite specific (Berry et al., 2021). When used to screen the general population, where the pre-test probability of celiac disease is low, they lose specificity. When the prevalence of selective IgA deficiency in a population is significant, it is necessary to measure serum IgA levels as the IgA anti-tTG test may be negative. Approximately 2% of patients with celiac disease may have selective IgA deficiency while 5–11% of IgA-deficient individuals have celiac disease. Measuring antibodies to gliadin, previously used to diagnose celiac disease, is not specific to the diagnosis and is now rather used to follow the response to a gluten free diet. In children under the age of two years, it has been suggested that measuring IgA antibody to deamidated gluten peptides is more useful than IgA anti-tTG to diagnose celiac disease. IgA or IgG antibody to endomysial antigen is considered to be very specific for the diagnosis of celiac disease. It is usually detected by immunofluorescence using monkey esophagus or human umbilical cord as a substrate. Unlike the anti-tTG and anti-gliadin assays it is confounded by subjective interpretation and lack of wide availability.

The next step in diagnosis usually involves getting a biopsy of the small intestine, usually now accomplished by upper gastrointestinal endoscopy. The classic endoscopic feature is scalloping of the duodenal mucosa and evidence of mucosal and villous atrophy which is visible using white light endoscopy but clearly highlighted during narrow band imaging. These findings also occur in other diseases that cause villous atrophy.

Small bowel biopsy remains the gold standard for diagnosis of celiac disease. Biopsies are taken from the duodenum (including the duodenal bulb in children) and jejunum. The diagnostic biopsy changes have been discussed in an earlier section. There has been debate about the need for small bowel biopsy in the diagnosis, particularly in children where invasive tests are preferably avoided. Several investigators have shown that high titers of anti-tTG correlate well with histological damage to the small intestinal mucosa, and have suggested that biopsy may be avoided in such patients.

Imaging is primarily useful in excluding infiltrative, inflammatory and neoplastic diseases. Contrast enhanced computed tomography (CT) and CT enterography may show nonspecific features of malabsorption including fluid filled loops, dilatation, dilution of positive contrast, and conformation. Celiac disease is characterized by excessive gas in the colon and a reversal of the mucosal fold pattern between jejunum and ileum referred to as jejunization of the ileum.

HLA-DQ typing is sometimes useful in diagnosis. Almost or nearly all patients with celiac disease have the genes that code for HLA-DQ antigens 2 or 8. HLA-DQ testing is usually carried out by molecular typing using polymerase chain reaction and hybridization with sequence specific oligonucleotide probes. When used in patients with a high suspicion of celiac disease, the absence of HLA-DQ 2 or -DQ 8 generally negates the diagnosis of celiac disease.

Finally, the diagnosis of celiac disease is confirmed by the response to a gluten free diet. Response may be documented clinically (e.g., resolution of diarrhea, resumption of growth), in laboratory parameters (e.g., normalization of hemoglobin or iron status), and histologically (restoration of normal histology in duodenal biopsies). The extent to which such resolution needs to be documented remains contextual. Prior to advent of serological tests, it was mandatory to document histological improvement in small bowel biopsies on gluten withdrawal, and recurrence of changes upon gluten rechallenge. These measures are now considered unnecessary in the majority of patients.

Treatment

Treatment, with rare exceptions, should ideally be commenced only after completing the investigative process including the duodenal biopsy. Detailed counseling is necessary before starting treatment. The lifelong nature of the disease and the need for strict gluten exclusion need to be emphasized to the patient and, in the case of children, to the parents. Some patients are unable to come to terms with such strict dietary restrictions, but most comply with these once they understand the central nature of gluten to their symptoms. Patients should be referred for professional dietary advice on how to achieve a gluten-free life style, and should be encouraged to join a local or national support group.

It is important to identify and to correct nutritional deficiencies, prominent among these being those of iron, folate, calcium, and the fat-soluble vitamins (A, D, E, K). In occasional cases additional measures, such as red cell transfusion or iron carboxymaltose infusion to correct anemia, and aggressive treatment of metabolic bone disease may be necessary. Osteomalacia presents with bone pain and pseudofractures, is associated with elevated alkaline phosphatase, and responds to a gluten-free diet and calcium and vitamin D supplementation. Osteoporosis is common in adults with celiac disease, and is diagnosed by testing for bone mineral density testing with a *T*-score less than 2.5 standard deviation below mean peak value in young adults. The primary treatment for osteoporosis in a celiac is a strict gluten-free diet with adequate calcium and vitamin D3, and all patients with celiac disease are recommended to take calcium and vitamin D3 supplements.

Intensive nutritional support and fluid replacement may be needed in very ill patients. Coexisting malignancy or autoimmune disease should be considered especially in elderly or ill patients. Follow up of patients to ensure response to gluten-free diet and compliance is crucial to ensure long-term compliance as well as in detecting potential complications of the disease. Screening of at-risk family members should be considered.

Complete avoidance of gluten, maintaining a gluten-free diet (GFD) is the single most important therapeutic intervention in celiac disease. Grains that should be totally avoided in the GFD include wheat, barley, rye, spelt, and kamut. The role of oats in genesis or perpetuation of celiac disease remains controversial. Recent studies have clearly demonstrated that oats are nontoxic for most patients with celiac disease; however, there is the concern that contamination of oat products by gluten takes place during the growing, milling, or processing of oat products. Thus, the inclusion or restriction of oats needs to be considered on an individual basis. Gluten intake in populations that consume wheat may range from 10 to 30 g per day. It is important to inquire if a food has any ingredients that are in any way derived from, or processed with, wheat, barley, or rye. This part needs expert counseling from a dietitian who is interested in celiac disease therapy and by interacting with celiac disease support groups which may be local, national or international. Many foods such as bread, cookies, biscuits, and pasta are obvious sources of gluten. Overlooked sources of gluten may include non-food items such as medications and religious offerings, foods containing malt or malt flavoring, hydrolyzed vegetable protein, modified food starch, natural flavorings, vegetable gum and fat substitutes. Food labeling is extremely important, and the treating dietitian must have access to up-to-date listings of commercially available processed foods that are gluten free, tailored to geographic location. Patients should not rely on the self-test of reaction to gluten as a means of detecting gluten in foods as symptoms may be delayed.

The response to GFD occurs within weeks, and improvement in overall well-being occurs early. Antibody titers may be used to follow recovery; they decline substantially within a year and may return to normal values after 2 years. Histological changes may persist despite this and may sometimes need to be assessed. Non-response to GFD, noted in about 10% of patients, is most often due to inadvertent ingestion of gluten. As little as 30 mg gluten per day for three months may trigger histological changes and symptoms in patients with celiac disease, highlighting the need for strict avoidance of gluten. Of concern are those foods that may contain hidden gluten, and the role of the dietitian becomes crucial here.

Maintaining a life-long gluten-free diet is challenging and this is an area where celiac disease support groups are especially helpful. There is ongoing research to look for alternative treatment strategies including enzymatic degradation of the immunogenic gluten peptides after a gluten meal in celiac patients, blocking the binding of gluten peptides to HLA-DQ2, and restoration of immunologic tolerance to gluten.

The approach to diagnosing and treating a patient suspected to have celiac disease is outlined in **Fig. 3**.

Complications

Microscopic colitis, either lymphocytic or collagenous, is sometimes associated with celiac disease and may cause continuing watery diarrhea. In addition to a strict GFD, symptomatic treatment with anti-diarrheal agents such as loperamide, or with topical steroids (budesonide) may be considered. Disaccharidase deficiency with lactose intolerance is usually secondary to mucosal injury and may

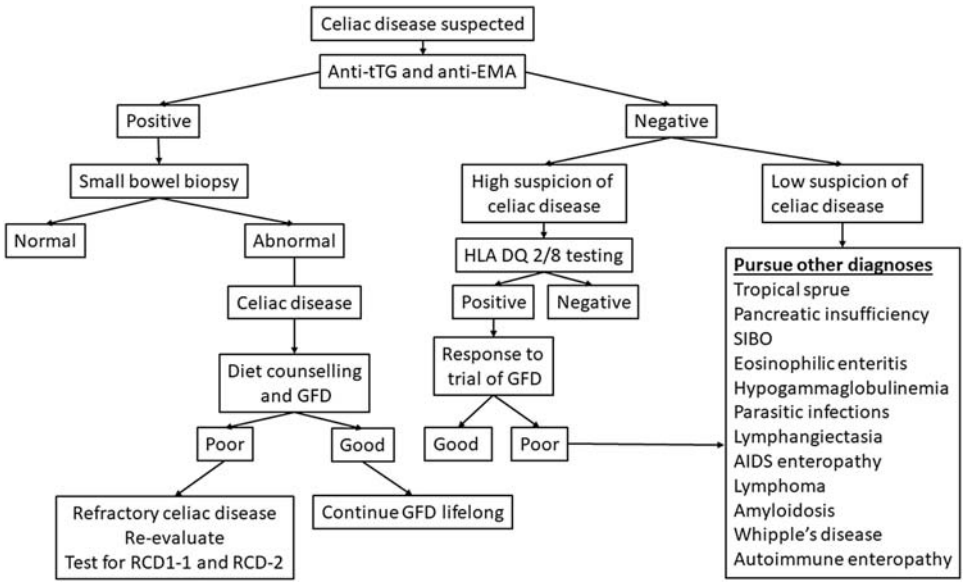


Fig. 3 Flow chart for diagnosing and treating celiac patients. The proper procedure for diagnosing and managing a patient who potentially has celiac disease. tTG = tissue transglutaminase; EMA = endomysial antibody; GFD = gluten free diet; RCD = refractory celiac disease.

resolve with GFD, unless there is a primary hypolactasia. Small intestinal bacterial overgrowth may occur secondary to steatorrhea and reduced small bowel motility and may require use of non-absorbed antibiotics such as rifaximin. A small number of patients will fail to respond to all measures, and develop a condition called refractory celiac disease, which is classified into two types (Penny et al., 2020). In type 1 refractory celiac disease, intraepithelial lymphocytes have a normal phenotype (expressing CD3, CD4 and CD8 on their surface), while in type 2 disease the intraepithelial lymphocytes show a clonal proliferation and lose these surface markers but still express CD3 in the cytoplasm. Type 1 refractory celiac disease without a clonal proliferation of T cells responds much better to immunosuppression, and may represent a self-perpetuating autoimmune process within the intestine. Type 2 disease responds poorly to immunosuppressive treatment and progresses over a period of 5 years to a T cell lymphoma (enteropathy associated T cell lymphoma). Collagenous sprue has features similar to those of celiac disease but is characterized by a thick layer of collagen just below the intestinal epithelium and responds poorly to all therapies. These patients require long-term nutritional support. Adenocarcinoma of the small intestine may occur in patients with celiac disease. The prime reasons are the chronic inflammation of celiac disease, and the presence of defects of the DNA mismatch repair genes. Survival with aggressive surgical therapy may be better than that for sporadic small bowel adenocarcinoma.

Conclusion

Celiac disease, caused by sensitivity to dietary gluten in genetically susceptible individuals, is a disease on the rise in many parts of the world. While the HLA-DQ genes have the strongest genetic influence on development of the disease a number of other causes—including non-HLA genes, infections, hygiene, and time of introduction of wheat into the diet in relation to breast feeding—may all influence expression of the disease. The disease primarily affects the small intestine, but affects many other systems due to immune and non-immune effects. The diagnosis is made through use of serological tests followed, in most cases, by duodenal biopsy. The disease is best treated by lifelong and complete withdrawal of wheat from the diet.

References

Berry, S., McMahon, N., Henderson, J., 2021. Guidelines for diagnosing coeliac disease: European Society Paediatric Gastroenterology, Hepatology and Nutrition. Arch. Dis. Child Educ. Pract. Ed. 106, 362–366.

Kagnoff, M.F., 2007. Celiac disease: pathogenesis of a model immunogenetic disease. J. Clin. Invest. 117, 41–49.

Kumar, V., Wijmenga, C., Withoff, S., 2021. From genome-wide association studies to disease mechanisms: celiac disease as a model for autoimmune diseases. Semin. Immunopathol. 34, 567–580.

Lebwohl, B., Rubio-Tapia, A., 2021. Epidemiology, presentation, and diagnosis of celiac disease. Gastroenterology 160, 63–75.

Myléus, A., Ivarsson, A., Webb, C., Danielsson, L., Hernell, O., Högberg, L., Karlsson, E., Lagerqvist, C., Norström, F., Rosén, A., Sandström, O., Stenhammar, L., Stenlund, H., Wall, S., Carlsson, A., 2009. Celiac disease revealed in 3% of Swedish 12-year-olds born during an epidemic. J. Pediatr. Gastroenterol. Nutr. 49, 170–176.

Oberhuber, G., Granditsch, G., Vogelsang, H., 1999. The histopathology of coeliac disease: time for a standardized report scheme for pathologists. Eur. J. Gastroenterol. Hepatol. 11, 1185–1194.

Paveley, W.F., 1988. From Aretaeus to Crosby: a history of coeliac disease. BMJ 297, 24–31.

Penny, H.A., Baggus, E.M.R., Rej, A., Snowden, J.A., Sanders, D.S., 2020. Non-responsive coeliac disease: a comprehensive review from the NHS England National Centre for refractory coeliac disease. Nutrients 12, 216.
Therrien, A., Kelly, C.P., Silvester, J.A., 2020. Celiac disease: extraintestinal manifestations and associated conditions. J. Clin. Gastroenterol. 54, 8–21.
Trinka, G., Wijmenga, C., van Heel, D.A., 2001. A genetic perspective on coeliac disease. Trends Mol. Med. 16, 537–550.

Further reading

Baggus, E.M.R., Hadjivassiliou, M., Cross, S., Penny, H., Urwin, H., Watson, S., Woodward, J.M., Sanders, D.S., 2019. How to manage adult coeliac disease: perspective from the NHS England rare diseases collaborative network for non-responsive and refractory coeliac disease. Frontline Gastroenterol. 11, 235–242.
Bai, J.C., Verdú, E.F., 2019. Evolving paradigms in the diagnosis of adult patients with celiac disease. Am. J. Gastroenterol. 114, 854–857.
Brown, I., Bettington, M., Rosty, C., 2021. The role of histopathology in the diagnosis and management of coeliac disease and other malabsorptive conditions. Histopathology 78, 88–105.
Catassi, C., Gatti, S., Lionetti, E., 2015. World perspective and celiac disease epidemiology. Dig. Dis. 33, 141–146.
Celdir, M.G., Jansson-Knodell, C.L., Hujoel, I.A., Prokop, L.J., Wang, Z., Murad, M.H., Murray, J.A., 2020. Latitude and celiac disease prevalence: a meta-analysis and meta-regression. Clin. Gastroenterol. Hepatol. S1542–3565, 31382–31383.
Hujoel, I.A., Murray, J.A., 2020. Refractory celiac disease. Curr. Gastroenterol. Rep. 22, 18.
Kahaly, G.J., Frommer, L., Schuppan, D., 2018. Celiac disease and endocrine autoimmunity—the genetic link. Autoimmun. Rev. 17, 1169–1175.
Lázár-Molnár, E., Snyder, M., 2018. The role of human leukocyte antigen in celiac disease diagnostics. Clin. Lab. Med. 38, 655–668.
Muhammad, H., Reeves, S., Jeanes, Y.M., 2019. Identifying and improving adherence to the gluten-free diet in people with coeliac disease. Proc. Nutr. Soc. 78, 418–425.
Rubio-Tapia, A., Murray, J.A., 2010. Celiac disease. Curr. Opin. Gastroenterol. 26, 116–122.

Cerebral palsy: Nutritional aspects

Patricia Miller, Kennedy Krieger Institute, Baltimore, MD, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of J. Krick, P. Miller, Cerebral palsy: Nutritional Aspects, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 317–325, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00048-9.

Introduction 171
Definition and etiology 172
Classification 172
Associated disabilities/deficits 172
The interdisciplinary team 172
Nutritional assessment 172
Growth 173
Growth charts 174
Ideal body weight 175
Body composition 175
Bone mineral density 175
Energy needs 175
Protein, nutrient and fluid needs 176
Diet/feeding history 176
Oral motor dysfunction and dysphagia 177
Other related problems 178
Gastrointestinal issues (GI) 178
Muscle tone and positioning 178
Orthopedic surgeries 178
Medications 179
Dental 179
Pain and discomfort 179
Mealtime behaviors 179
Intervention strategies 179
Maximizing oral nutritional intake 179
Oral motor considerations 180
Diet/texture modifications 180
Alternative feeding routes 180
Coordinated services 181
Summary 182
References 182

Key points

- The reader will understand what cerebral palsy is and how it affects children with regard to their growth, diet and nutritional needs.

Introduction

Nutritional issues such as diet, body composition and growth are integral aspects of medical care for children with cerebral palsy (CP). The first section defines CP and describes its causes, prevalence, and classification types. Associated deficits related to CP are also explored. The topic of nutritional assessment of children with CP includes discussions on growth, body composition, and energy, protein, fluid, and nutrient needs. Feeding and swallowing problems and the influence of muscle tone on the ability to eat safely are discussed in-depth, as are alternative feeding routes. The interdisciplinary approach is emphasized throughout as the ideal model to provide services to children with CP in order to ensure quality of life in the community.

Definition and etiology

Cerebral palsy is a term that refers to a number of non-progressive disorders of movement and posture that result from an injury to the central nervous system (CNS) during early brain development (Stavsky et al., 2017). CP is the most common motor disability in childhood and prevalence estimates range from 1.5 to more than 4 per 1000 live births (CDC, 2018). Risk factors for CP during intrauterine, intrapartum or postnatal periods include premature birth, low birthweight, disruption of blood and oxygen supply to the developing brain, maternal infection, congenital malformations, multiple gestations, and genetic disorders (Graham et al., 2019). Acquired CP can also occur after the first month of life due to brain injury or infection. CP results in issues of muscle control such as uncoordinated movements or muscle stiffness, which can lead to poor balance, difficulty in mobility and problems completing basic daily living skills such as getting dressed, walking or feeding oneself. CP is a lifelong condition with no cure, but interdisciplinary treatment and therapeutic interventions can lessen the effects and assist the individual with CP in successfully functioning in their environment.

Classification

Cerebral Palsy is classified in several different ways related to topographical distribution, level of severity, physiological classification and now most commonly by gross motor skill function. There are 4 types of CP classified by motor type including spastic, dyskinetic, ataxic and mixed, with spastic being the most common (CDC, 2018). These types are further characterized by the part of the body or number of limbs affected. The Gross Motor Functional Classification System (GMFCS) is a standardized functional classification system based on an ordinal grading system (Palisano et al., 2008). Classification systems have also been developed for upper extremity function (Manual Ability Classification System – MACS), communication (Communication Function Classification System – CFCS) and eating/drinking (Eating and Drinking Ability Classification System – EDACS) (Paulson and Vargus-Adams, 2017). **Fig. 1** depicts the classification of CP by motor types, topography and function (McGrath, 2016). The type of CP and the degree of involvement play an important part in nutritional assessment and treatment.

Associated disabilities/deficits

Associated deficits of CP are important to note since they affect nutritional status. Injury to the developing brain can affect multiple areas placing children with CP at increased risk for intellectual or learning disabilities. Sensory deficits, neurobehavioral and emotional issues are common (CDC, 2018). Seizures occur frequently and in addition to medical management, the ketogenic diet is often prescribed to treat seizures. It is a diet high in fat with limited nutrients and fluid and requires implementation and monitoring by both a neurologist and a nutritionist trained in the diet. Feeding problems including oral motor dysfunction, dysphagia, digestive problems and respiratory issues have a direct effect on nutritional status.

The interdisciplinary team

Treatment of children with CP is best achieved with an interdisciplinary team whose members participate, share expertize and make decisions together to allow for successful functioning of the child in their environment (Patel et al., 2008). The team members working with children with CP may include but is not limited to Neurology, Developmental Pediatrics, Orthopedics, Gastroenterology, Physical Therapy, Occupational Therapy, Speech Language Pathology, Nutrition, Nursing, Social Work, Child Life and Psychology. The six F's framework – function, family, fitness, fun, friends and future – should be included as participation goals in the treatment of CP by all disciplines involved (Graham et al., 2019).

Nutritional assessment

The goal for nutritional assessment and intervention is to have healthy, alert, interactive individuals who are able to take advantage of all that the environment has to offer. Each child must be able to participate to his or her capacity in the learning and therapeutic habilitative processes and in social, community, and leisure activities. **Table 1** lists the common factors, which can affect the nutritional status of children with CP. Those children who present with poor weight gain at an early age, have significant motor impairments and have feeding and swallowing difficulties are at greatest risk of having significant nutritional problems (Rempel, 2015).

The Subjective Global Nutrition Assessment (SGNA) evaluates nutrition status based on a practitioner's judgment of nutritional indices, rather than based only on the quantitative measurements (Secker and Jeejeebhoy, 2012). It includes questions related to nutrition history and physical exam, signs of fat or muscle wasting and edema in conjunction with measures of weight, height and triceps skin fold (TSF). Minocha et al. (2018) found the SGNA questionnaire to be a simple, comprehensive, non-invasive and cost effective tool for screening for malnutrition in children with CP. The recent consensus statement from the Academy of Nutrition and Dietetics and the American Society for Parenteral and Enteral Nutrition, which standardized the definition and identification criteria for diagnosing pediatric malnutrition, has not been validated in children with CP and should be used with caution in this population (Becker et al., 2014).

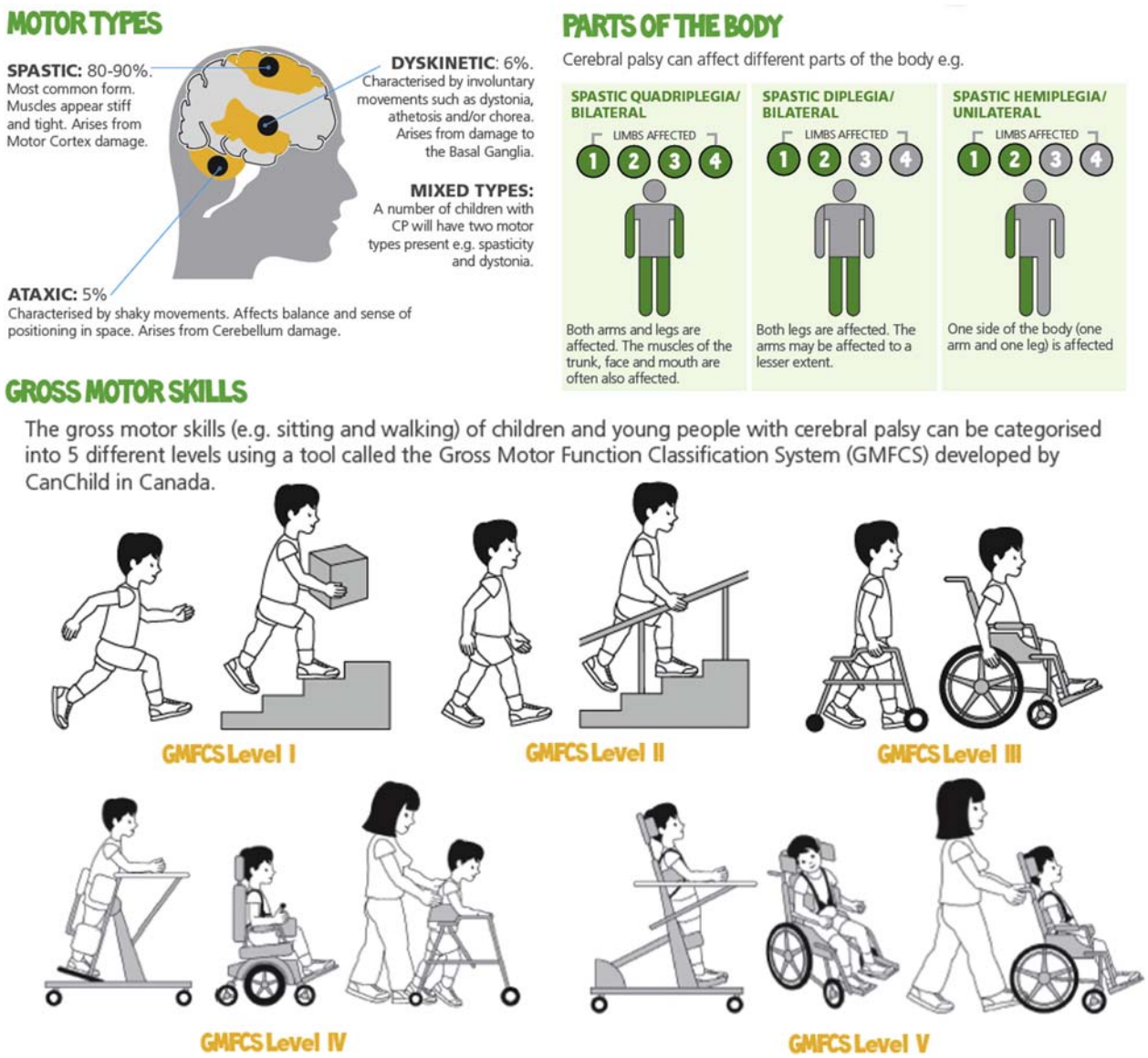


Fig. 1 Classifications of cerebral palsy. Reproduced from CP world day McGrath (2016).

Table 1 Factors affecting nutritional status.

• Associated disabilities	• Required physical and cognitive support
• Oral motor dysfunction/dysphagia	• Increased or decreased energy needs
• Abnormal tone and motor patterns	• Food refusal/selectivity
• Behavioral problems	• Psychosocial needs
• GI issues – emesis and reflux	• Constipation
• Positioning requirements	• Frequent surgical procedures
• Seizure disorder	• Use of multiple medications
• Sensory deficits	• Caregiver knowledge/skill
• Resources including access to food	• Dental problems

Growth

Adequacy of growth may be the single most valuable indicator of nutritional status in children. Healthy, typically developing children grow at predictable rates, with patterns reflective of certain age categories. Children with CP have altered growth patterns and tend to be shorter and lighter than the reference standard (Day et al., 2007). Growth patterns can be affected by gender, age, pubertal

Table 2 Equation for estimation of stature using segmental measures.

Age 0–12 years – children with CP		
Segmental measure	*Equation to estimate stature (S) (cm)*	*SE of estimate (cm)*
Tibial length (TL)	$S = (3.26 \times TL) + 30.8$	1.7
Knee height (KH)	$S = (2.69 \times KH) + 24.2$	1.4
Age 12–18 years – typically developing children, validated in a small group of children with CP		
Race and gender	*Equation to estimate stature (S) (cm)*	*SE of estimate (cm)*
White males	$S = (2.22 \times KH) + 40.54$	4.21
Black males	$S = (2.18 \times KH) + 39.6$	4.58
White females	$S = (2.15 \times KH) + 43.21$	3.90
Black females	$S = (2.02 \times KH) + 46.59$	4.39

SE, standard error.
Adapted from Samson-Fang and Bell (2013).

stage, race, genetic potential, hormonal variations and dietary intake. Early identification and treatment of growth problems is paramount to ensuring overall health and development especially in children with CP.

Both nutritional and non-nutritional factors influence growth in children with CP. Non-nutritional factors influencing growth in children with CP include central nervous system abnormalities, muscle tone/contractures, immobilization/limited physical activity, lack of weight bearing, repeated orthopedic surgeries, and growth hormone deficiency/dysregulation. Children with asymmetric involvement of CP have decreased growth and fat mass on the affected side, with the difference increasing with age and functional severity, thereby supporting the impact of non-nutritional and non-endocrine factors (Rempel, 2015).

Obtaining accurate anthropometric measurements in children with CP can be very challenging secondary to joint contractures, scoliosis, ambulatory status and poor cooperation. Recumbent length should be measured for those under 2 years of age, and for those unable to stand where segmental measurements can be used as an estimation of height. Alternative measures for estimating height in those with CP who are not able to stand include tibial length (TL) and knee height (KH). These measurements require special calipers and techniques to be completed properly. Clinicians should be trained in obtaining these measurements to ensure accuracy and reproducibility. Equations to estimate height using these measurements based on age, gender and measurement site can be found in **Table 2** (Sampson-Fang and Bell, 2013). Discrepancies in leg length should be noted, with the same side (longer side) measured each time. The estimate of stature from these equations have a degree of error and so caution should be exercised with the use of these estimations when calculating body mass index (BMI), for instance, as the standard of error will be magnified. Weight should be measured on a digital scale with the child wearing only light clothing, with a dry diaper if applicable. If the child is unable to independently stand on the scale, a sling scale or wheelchair scale can be used. All braces, splints and/or items on the wheelchair should be removed before measuring to eliminate confounding error.

The use of *z* scores for length-for-age, weight-for-age, and weight-for-length promotes an accurate evaluation of discrete changes from one measurement date to another. Percentile tables describe ranges, and consequently detection of minor movement within the range is difficult to describe. The *z* score denotes standard deviation units from the median and allows the practicing clinician and investigator to pinpoint precisely any given measurement.

Growth charts

Early growth charts for children with CP were limited to those with quadriplegia who were between the ages of 1–10 years (Krick et al., 1996). Following the development of the GMFCS, new charts developed by Day et al. (2007), were stratified by gross motor levels. These charts also differentiated between those who were orally fed versus tube fed in the groups with more severe limitations (GMFCS levels 4 and 5). Both charts show that the rate of growth in children with CP is slower, the difference from the norm becomes greater as age increases, and there is further deviation as gross motor function declines. Enhanced charts produced by Brooks et al. (2011) demonstrate the risk of comorbidities based on weight-for-age for those at each GMFCS level. Individuals whose weight-for-age falls in the shaded area on the lower end of the chart have a higher risk of mortality than those outside of this area. The growth charts from Day et al. (2007) and Brooks et al. (2011) can be found at www.lifeexpectancy.org (2011). It is important to note that all growth charts for children with CP describe growth patterns irrespective of nutritional status and therefore are not considered ideal standards. The growth charts for children with CP should be used with clinical judgment in conjunction with the CDC growth charts as well as with other anthropometric measurements.

Ideal body weight

The estimation of an ideal body weight (IBW) allows the clinician to assess progress and effectiveness of intervention, while providing the child and family a goal to attain. The IBW should be aimed at maintaining adequate fat and muscle stores to endure repeated surgeries or a common virus while facilitating daily physical care and management. Weight-for-length and BMI are indicators of nutritional status, which obscure the issue of chronological age and address whether the individual is proportionate. IBW can be expressed as this ratio. Those with cerebral palsy should attain and maintain an IBW that takes into account their age, level of physical ability, and their independence. Measurement of arm anthropometry will provide a description of body composition and support clinical judgments related to IBW. In the absence of population specific guidelines, the 10th %ile weight for height or BMI on the CDC growth charts or the 50%ile BMI on the CP charts have been used historically. The growth charts developed by Brooks et al. (2011) can also be used in determining an appropriate weight above the shaded area on the graph, which is associated with improved health outcomes.

Body composition

Children with CP have altered body composition with decreased muscle mass when compared to their age matched peers (Caselli et al., 2017). Analysis of body composition in the clinical setting can be challenging, and not all methods have been validated for use in children with CP.

Measurement of arm anthropometry can include but is not limited to TSF, subscapular skin fold (SSSF) and Mid arm-upper arm circumference (MUAC). Patient cooperation with the measuring techniques required for accuracy and safety, may be difficult to obtain or maintain. Assessment of these measurements can assist with classification of nutritional status, particularly related to adequacy of muscle and fat mass. Gurka et al. (2009) found that a simple correction factor to the commonly used Slaughter equations substantially improves the ability to estimate percentage of body fat from TSF and SSSF in children with CP. The corrected equations can be found in **Table 3**. Dual-energy x-ray absorptiometry (DXA) scans are frequently performed to assess bone mineral density and this can provide measures of fat and fat free mass to assist with evaluation. Children with CP tend to have increased truncal fat with depleted extremity stores, which can make assessment challenging, requiring the clinician's clinical interpretation to assess nutritional status.

Bone mineral density

Low bone mineral density, osteopenia and osteoporosis are found in most children with CP, particularly in those who are immobile. Fractures are common and many of those who sustain one fracture will sustain a repeated fracture, with the femur being the most common site (Kuperminc and Stevenson, 2008). Severity of the neurological impairment, Calcium and Vitamin D intake, limited weight bearing during growth, temporary immobilization associated with orthopedic surgery, use of anticonvulsants and lower TSF (fat stores) all independently contribute to the lower bone mineral density. It is important to provide physical therapy for weight bearing activities in addition to nutritional therapy including adequate Calcium and Vitamin D intake with supplementation if needed based on serum Vitamin D levels. There is promising work with the use of bisphosphonate medications and whole-body vibration in children with CP to improve bone density (Jesus and Stevenson, 2020).

Energy needs

Children with moderate to severe CP have energy needs, which differ significantly from typically developing children. Equations that are frequently used to predict energy requirements were developed using healthy children and adults in usual environmental and physical activity conditions and do not provide an accurate assessment of the needs of those with CP. Factors affecting energy expenditure in children with CP include age, muscle tone, physical activity (with or without assistive devices), repeated surgical interventions, illness and the additional need for catch-up growth or wound healing. Most equations typically overestimate needs leading to overfeeding, particularly in those who are non-ambulatory and tube fed. Culley and Middleton (1969) historically described calorie needs based on height in centimeters, with those who are ambulatory requiring 13.9 kcal/cm and those who are non-ambulatory requiring 11.1 kcal/cm. More recently, energy needs for non-ambulatory children with CP have been estimated to be 60–70% of calculations using typical methods (Bell and Samson-Fang, 2013). Those with profound gross motor impairment could need even as little as 10–15 kcal/kg. Energy intake should be adjusted as needed to achieve weight goals, while care should be exercised to provide adequate protein, nutrients and fluid despite these very low calorie needs. The use of any approach should be a guidepost to start and with careful monitoring and modifications based on observed clinical changes in growth. For those who are tube fed, establishing tolerance at a lower level of calories and then gradually increasing intake can help achieve better tolerance, avoid overfeeding and limit gastroesophageal reflux (GER) and emesis. Increases or decreases in energy intake can be made in 10–20% increments, with frequent monitoring to achieve the desired goal.

Table 3 Original Slaughter equations and corrections for children with cerebral palsy.

Population	*Original Slaughter equation for predicting percentage body fat*
Sum (triceps, subscapular)≤35 mm	
Males	
Prepubescent[a] white	% Body fat = 1.21(tri + sub) − 0.008(tri + sub)^2 − 1.7
Prepubescent black	% Body fat = 1.21(tri + sub) − 0.008(tri − sub)^2 − 3.2
Pubescent white	% Body fat = 1.21 (tri + sub) − 0.008(tri − sub)^2 − 3.4
Pubescent black	% Body fat = 1.21(tri + sub) − 0.008(tri − sub)^2 − 5.2
Postpubescent white	% Body fat = 1.21(tri + sub) − 0.008(tri − sub)^2 − 5.5
Postpubescent black	% Body fat = 1.21(tri + sub) − 0.008(tri − sub)^2 − 6.8
Females (all)	% Body fat = 1.33(tri + sub) − 0.013(tri + sub)^2 − 2.5
Sum (triceps, subscapular) >35 mm	
Males (all)	% Body fat = 0.783(tri + sub) + 1.6
Females (all)	% Body fat = 0.546(tri + sub) + 9.7
	Cerebral-palsy-specific corrections to Slaughter-estimated percentage body fat[b]
Overall correction	+12.2
Additional correction for	
Males	−5.0
More severe GMFCS	+5.1
Black race	−3.1
Pubescent	+2.0
Postpubescent	−4.6
Sum (triceps, subscapular) >35 mm	−3.2

[a]Prepubescent, Tanner stage 1,2; pubescent, Tanner stage 3; postpubescent, Tanner stage 4,5.

[b]Instructions for using these corrections on a given child with CP: always add 12.2 to the Slaughter-estimated percentage body fat. Then, if the individual falls within each of the additional categories, add that respective correction as well. For example, for a black pubescent male at GMFCS level 1 whose sum (triceps, subscapular) <35 mm, the predicted percentage body fat = Slaughter percentage body fat + 12.2 − 5.0 − 3.1 + 2.0. Tri + sub, triceps skinfold + subscapular skinfold. GMFCS, Gross Motor Function Classification System.

From Gurka et al. (2009).

Protein, nutrient and fluid needs

Protein and nutrient needs are based on the Dietary Reference Intakes (DRIs) of the general population. There is no evidence to support altered requirements of vitamins, minerals or trace elements in children with CP. As noted earlier, care must be taken to ensure that those with very low energy needs receive supplementation of protein and nutrients when needed. The undernourished child may require as much as 2 g protein/kg/day to promote catch-up growth (Bell and Samson-Fang, 2013). Children undergoing orthopedic surgeries or those with wounds also require additional protein and nutrients for stress and healing. It is important to monitor for and evaluate possible protein and nutrient deficiencies through a nutrition focused physical exam assessing subcutaneous tissue for edema, skin integrity, hair texture/color and muscle or adipose wasting. Children with CP are frequently deficient in calcium, iron, zinc, selenium and Vitamins C, D and E and may require supplementation (Romano et al., 2017).

Fluid needs are calculated based on body weight using the Holliday-Segar nomogram rather than based on calorie intake. For weights ranging from 0 to 10 kg, the fluid need is 100 mL/kg/day; from 10 to 20 kg the fluid need is 1000 mL plus 50 mL/kg for each kilogram of body weight more than 10; over 20 kg the fluid need is 1500 mL plus 20–25 mL/kg for each kilogram more than 20 (Holliday and Segar, 1957). Monitoring of hydration in a clinical setting is done by checking a urine specific gravity, while caregivers at home can monitor urine for quantity, color and odor. Additional fluid should be provided during periods of stress, illness, or significant environmental changes. For children receiving enteral tube feedings, total formula provided is counted entirely toward fluid intake, assuming there are no other health or renal concerns.

Diet/feeding history

Obtaining accurate estimates of oral intake particularly in children with CP can be challenging. Overestimating intake in children with CP by caregivers is common, and care should be taken when assessing intake from food recalls or records, as these have not been found to be reliable or valid methods in this population (Walker et al., 2011). A 3-day weighted food record provides the most

Table 4 Factors to consider in feeding history.

- Frequency of meals and snacks
- Meal length
- Positioning at mealtimes
- Modifications to food or liquid texture
- Food loss/spillage
- Use of adaptive feeding utensils
- Assistance required a mealtimes
- Use of supplements or calorie boosters
- Food preferences and dislikes
- Number of caregivers able to feed child

accurate measure of energy intake in children with CP, but it requires training and considerable effort of the caregiver to complete (Scarpato et al., 2017). It is likely that length of mealtime, effort by the child and caregiver and oral motor losses contribute to overestimating intake. Food records and recalls can be used qualitatively to assess variety of food intake, versus quantitatively (Sampson-Fang and Bell, 2013). Meal observation to note actual intake including loss of solids and liquids in addition to mealtime interactions can be very useful. Factors to consider in the feeding history are listed in **Table 4**.

Oral motor dysfunction and dysphagia

Oral motor dysfunction and dysphagia are common in children with CP, and the severity increases with increasing levels on the GMFCS having a direct effect on nutritional status (Benfer et al., 2017). Alterations in muscle tone result in abnormal movements of the lips, tongue and jaw which manifest as lip/tongue retraction, tongue tip elevation, tongue/jaw thrust, tonic bite, and instability of the lips, cheek and jaw (Arvedson, 2013). Problems may present as poor intake, inefficient and lengthy mealtimes, abnormal oral-motor patterns, inappropriate progression of feeding skills, and physiological compromise with feeding. A strong gag reflex, tactile hypersensitivity in the oral area and drooling, can also complicate feeding. Eating skills are acquired in a sequential pattern so that a developmental history is important to understand in planning any intervention. Sensory, cognitive and language deficits may also complicate the feeding process.

Aspiration, when swallowed food or liquids enter the airway or lungs and can result in coughing, choking, pneumonia or chronic lung changes over time. Aspiration can also occur from below when regurgitated gastric material enters into the esophagus, and depending upon how adequately the airway is protected, can lead to direct aspiration into the lungs. Aspiration can result in discomfort, arching, irritability, food refusal, esophagitis, apnea, frequent respiratory compromise and drooling. Aspiration can also be silent, with no overt signs of distress to the child such as coughing or choking. Children with spastic CP are most at risk for dysphagia and aspiration. Indicators of feeding and swallowing dysfunction or aspiration are listed in **Table 5**. Children who exhibit any of these symptoms may require radiological evaluation via Modified Barium Swallow Study or Fiber-optic Endoscopic Evaluation of Swallowing to assess for aspiration. These modalities can not only detect aspiration, but can provide more detailed information about the oral structures and the competency of the oral, pharyngeal, and esophageal phases of swallowing. They also provide information helpful in determining appropriate solid and liquid textures as well as proper head and neck positioning. Fatigue may occur in the child who is not able to sustain the work involved with feeding and may be expressed by an increase in respiratory rate, diaphoresis, or increased work of breathing. The causes may be muscular, respiratory, or cardiac, and they

Table 5 Indicators of feeding and swallowing dysfunction/aspiration.

- Facial weakness
- Decreased sensation
- Congestion, noisy “wet” sounds
- Multiple swallows to clear a bolus
- Unexplained fevers
- Unexplained irritability
- Coughing/choking/gagging before, during or after a swallow
- Food refusal
- Pain with swallowing
- Difficulty managing secretions, drooling
- History of upper respiratory infections, pneumonia
- Apnea during feeding
- Failure to thrive or maintain weight

may increase the risk of aspiration or hypoxia. The work required to eat a meal is accomplished at a higher physiological cost to the child, thereby increasing caloric needs. Setting time limits for meals and snacks is often suggested.

There are numerous scales developed which assess feedings skills, oral intake, safe intake and nutritional status in children with neurodevelopmental disabilities. Some are used for certain age groups, and others require special training to be able to administer. Benfer et al. (2012) determined that the Schedule for Oral Motor Assessment (SOMA) and the Dysphagia Disorders Survey (DDS) had the strongest clinical utility in pre-school children with neurodevelopmental disabilities. The EDACS mentioned earlier, is based on functional abilities including key features of safety and efficiency related to eating and drinking (Sellers et al., 2014). A recent screening tool developed by Bell et al. (2019) consists of a 4-item tool for feeding/swallowing difficulties and undernutrition in children with CP that can be used independently with parents or caregivers.

The importance of a feeding evaluation in conjunction with a Speech Language Pathologist or Occupational Therapist who is skilled in assessing feeding problems cannot be underscored in assessing function, safety and providing guidance in feeding techniques. An interdisciplinary team evaluation is essential for the assessment, development of appropriate goals and facilitation of a treatment plan that respects the developmental progression.

Other related problems

Gastrointestinal issues (GI)

Constipation is a chronic problem for most children with CP due to alterations in the enteric nervous system, which requires input from the CNS to coordinate intestinal motility. This can be exacerbated by changes in muscle tone, skeletal muscle incoordination, skeletal deformities and prolonged immobility (Elawad and Sullivan, 2001). Constipation is frequently underdiagnosed and therefore under treated. Factors such as loss of sensation, limited physical activity, medication side effects and inadequate dietary fiber and or fluid can also play a role. Oral motor dysfunction also plays a role in diminished intake as well as fluid and food losses. Modified fluid and food textures may also yield less free water and fiber. Discomfort associated with constipation may decrease appetite and increase GER and/or emesis. Dietary intervention may be limited and medical management may be necessary. Children with CP typically benefit from use of both a stool softener and stimulant with the goal of treatment promoting a soft, daily bowel movement.

GER is common in children with CP related to muscle tone, positioning, transient relaxation of the lower esophageal sphincter and increased intra-abdominal pressure (Andrew et al., 2012). This can result in irritability or arching during or after feeding. Treatment for GER includes medical therapy with antacids, acid blockers/proton pump inhibitors, mediations to increase gut motility, reduction in feeding rate, positioning, thickening food or liquids, smaller, more frequent feedings or surgical intervention. Children who are fed enterally via a feeding tube and fail medical treatment may require a Nissen Fundoplication procedure to prevent reflux or post pyloric/jejunal feedings to by-pass the stomach completely. Jejunal feedings are also of assistance for those with delayed motility. Refusal of food by a child with CP likely arises from dysphagia and chronic GER, and should not be confused with behavioral avoidance (Asgarshirazi et al., 2017).

Superior Mesenteric Artery Syndrome (SMA) is a condition in which the third portion of the duodenum is intermittently compressed by the overlying superior mesenteric artery, resulting in gastrointestinal obstruction (Neuman et al., 2014). Symptoms include recurrent vomiting, abdominal distention, weight loss and postprandial distress. Children with CP are at high risk for SMA particularly if there has been severe weight loss eliminating the mesenteric fat pad, prolonged supine positioning and/or scoliosis surgery. An upper gastrointestinal evaluation (UGI) is used to define the problem and non-surgical treatments include gastric aspiration, naso-jejunal (NJ) or gastro-jejunal (GJ) feedings distal to the obstruction, slow continuous feedings, weight gain and positioning.

Muscle tone and positioning

It is important to understand the influences of muscle tone and proper positioning on the ability to eat safely and efficiently in this population. Increased or decreased muscle tone contributes to difficulty preserving a patent airway, compromised self-feeding skills, poor rib cage expansion and esophageal motility, and difficulty in maintaining a stable supported base for seating (Andrew et al., 2012). Fluctuating muscle tone leads to involuntary movements and limited postural stability. Despite the type of muscle tone, optimal positioning is crucial for feeding and swallowing. The proper feeding position includes neutral alignment of head and neck, midline orientation, symmetrical trunk position, 90 degrees pelvic/femoral alignment, and symmetrical arm position with neutral shoulders (Physiopedia, 2010). Consultations with orthopedists and/or rehabilitation physicians to address current and potential musculoskeletal problems, physical and occupational therapists for functional assessment, orthoptists for deformity management, and durable medical equipment specialists to customize standard wheelchair components are valuable.

Orthopedic surgeries

Orthopedic procedures are common in children with CP and each surgery should be preceded by an evaluation of nutritional status and assessment of the child's ability to physically heal and recover quickly from the trauma (Leonard et al., 2020). Many children,

who are marginal oral feeders, decompensate, lose weight and have a difficult time healing because of a cascade of events including pain, poor positioning for safe feeding, worsening constipation, minimal intake, lethargy, and increased medications for pain, which may have a sedative effect. Some children require supplemental tube feedings prior to surgery or during the postoperative period to promote weight gain and circumvent poor oral intake.

Medications

Drug-nutrient interactions should be considered for children receiving long-term medications for seizure disorders, alterations of muscle tone, attentional deficits, GI disorders and/or other chronic conditions. One drug or the combination of multiple drugs may affect nutrition in many ways including decreased appetite, interference with absorption of specific nutrients, nausea/vomiting and constipation (Ptomey and Wittenbrook, 2015). Medications used frequently for the treatment of seizures or reflux can result in decreased nutrient absorption or losses of Vitamins D, K, B6, B12, Folate, Iron and Calcium.

Medications used to treat spasticity and dystonia such as Valium or Baclofen can have significant effects on oral motor function, GI motility and constipation. Baclofen can be administered through oral routes or via an intrathecal pump. The surgically implanted pump delivers a continuous dose of medication directly into the spinal fluid in the lower back. As tone is significantly reduced in the children for whom intrathecal baclofen pumps are used, so is the energy requirement (Liu et al., 2008). With the use of the baclofen pump, weight gain can be seen in children who have been on tube feedings with a constant intake over time, and adjustments in the calorie level may be necessary. This decrease in tone is also associated with decreased GI motility and increased constipation.

Dental

Increased incidence of cavities and erosion frequently occur due to poor oral hygiene and teeth grinding in children with CP (Ahmad et al., 2019). Hypersensitivity in the oral area and hyperplasia of the gums from long-term use of phenytoin may also be seen. Malocclusion is a common musculoskeletal problem and contributes to drooling, which can negatively affect daily oral care.

Pain and discomfort

Inquiry into pain and discomfort as it relates to eating such as face grimacing, crying, arching, sweating, or stopping a preferred activity is an important aspect of a total assessment. Contributors to pain are multifactorial and can include hip subluxation, muscle spasms, dystonia, GER and constipation (Graham et al., 2019).

Mealtime behaviors

Parent–child interactions can also influence feedings. Ineffective communication, lack of bonding, the absence of social interaction or poor interactive skills, family dysfunction, and decreased environmental stimuli can exacerbate feeding difficulties or lead to frustration and anxiety with subsequent food refusal or parental withdrawal. Aversion to oral feeds may result as an outcome of medical complications, such as esophagitis and GER, or lack of feeding experience at critical milestones secondary to prolonged tube feedings. Behavioral treatment should only be undertaken after thorough medical, nutritional, and neurodevelopmental assessments are completed.

Intervention strategies

Maximizing oral nutritional intake

The benefits of nutrition intervention include improved alertness and attention span/learning, enhanced health and immune function, decreased episodes of illness and hospitalizations, better respiratory muscle strength, reduced risk of decubitus ulcers, improved wound healing, reduced stress on the family and enhanced quality of life for the family and child. The feeding plan should be safe, promote growth or weight maintenance without excessive energy expenditure, and meet the needs of the family. It should reflect their resources in time and skill, and it should address their concerns and expectations. The goals for treatment once feeding and swallowing problems are identified include preventing aspiration and thereby respiratory compromise; providing adequate calories, protein, vitamins, minerals, and fluid; and educating caregivers regarding nutritional requirements.

For those children who are safe to eat by mouth, energy intake should be maximized. This can be done either by increasing the caloric density of the diet or by increasing total intake through smaller, more frequent meals. In children who are not able to take in more by mouth, intake can be boosted by increasing the concentration of formula if appropriate, addition of calorie boosters to meals, use of high-calorie foods or the addition of oral supplements. **Table 6** provides ideas to maximize energy intake.

Table 6 Foods to use for increasing energy intake.

- Powdered skim milk
- Cooked eggs
- Cheeses
- Ice cream
- Mayonnaize or salad dressing
- Cream cheese
- Gravy
- Baby cereals or potato flakes
- Fruit in heavy syrup
- Meats cooked in oil with breading/added gravy
- Carnation breakfast essentials
- Butter, margarine, oils
- Peanut butter
- Wheat germ
- Sour cream
- Heavy cream/sweetened condensed milk
- Concentrated juice
- Pudding, yogurt, custard and milkshakes
- Vegetables in cheese or cream sauces
- Commercial oral formulas/supplements

Oral motor considerations

Management strategies for daily mealtime feeding include positioning, modification of the sensory properties of the food, oral motor facilitation techniques, and equipment adaptations (Bell and Samson-Fang, 2013). It is important to acknowledge the inability to change the underlying feeding problem while providing a method of circumventing the problem to allow adequate nutrition and growth. For example, facilitative techniques to minimize excessive jaw movement may entail the feeder providing physical jaw control/support; a change in the food consistency, texture, temperature, or taste to improve the ability to propel a bolus through the oropharynx; the careful selection of adaptive feeding equipment to assist with self-feeding and/or increased intake; and an appropriate seating system (Rempel, 2015).

Diet/texture modifications

Changes in the texture of solids and liquids may be necessary to ensure safety, but may also improve efficiency, decrease effort and therefore result in increased nutritional intake. Once again, guidance from our interdisciplinary colleagues in Speech Language Pathology and Occupational Therapy who specialize in oral motor therapy is necessary to determine the appropriate texture of solids and liquids. **Table 7** lists some common thickening agents for foods and liquids. The International Dysphagia Diet Standardization Initiative (IDDSI) Framework provides a common terminology to describe food textures and liquid thickness (IDDSI, 2019). There are resources for pediatrics, providing guidelines for how to assess foods at each level, as well as what to include and avoid. See **Fig. 2** for a diagram of the IDDSI levels from thin liquids to regular texture solids. **Table 7** lists some common thickening agents for foods and liquids.

Alternative feeding routes

Supplemental or complete tube feeding may be necessary for the child who is not safe to eat or drink by mouth secondary to aspiration, for those who continue to fail to grow despite dietary interventions, or for those with prolonged, stressful mealtimes. In addition, those with evidence of low body fat stores or low weight in respect to height/length should be considered for supplemental feedings (Romano et al., 2017). A recent study noted that epilepsy, poor motor function, trunk muscle tone disorder

Table 7 Thickening agents for foods and liquids.

- Pureed or blenderized fruits and vegetables
- Dry baby cereals
- Yogurt
- Soft tofu
- Instant pudding
- Potato flakes
- Wheat germ
- Bread crumbs
- Graham cracker crumbs
- Corn starch
- Unflavored gelatin
- Commercial powder or gel thickeners

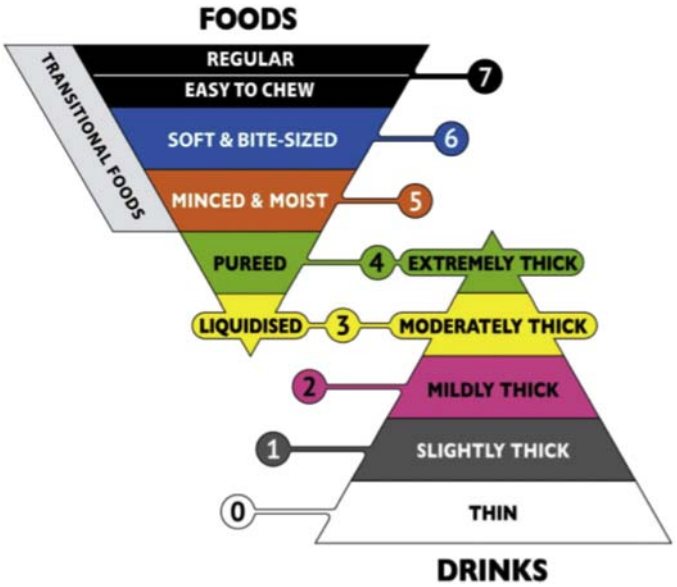


Fig. 2 The complete IDDSI framework. Reproduced from IDDSI (2019).

and male gender were accurate, sensitive and specific factors associated with the need for tube feedings (Bertoncelli et al., 2020). Enteral nutrition support can help improve growth, nutritional status, hydration, bowel function and allow for easier administration of medications.

Naso-gastric, naso-jejunal, gastrostomy, gastrostomy–jejunal, and jejunostomy tubes are options available for providing supplemental or total nutrition. The degree of GER and risk of aspiration determine where the tube is placed, whereas the length of time needed for tube feedings determines whether a naso-enteral or surgically placed tube is required. The decision regarding continuous, intermittent, or combination tube feeds is dependent on the individual needs and tolerance of the patient.

Tube feedings should be considered a tool to improve nutritional status rather than failure of the child's ability to eat. Based on the medical diagnosis and developmental stage of the child, the prognosis for return to oral feeding varies, and the length of time to achieve this goal is extremely variable. For some children, the goal of returning to full or partial oral feeding is not realistic. In a study evaluating the health of children with CP, Liptak et al. (2001) describes those who were tube fed as having the lowest mental age, requiring the most health care resources, using the most medications, and having respiratory problems. These children were characterized as especially frail and required numerous health-related resources and treatments. Oral motor therapy should focus on maintaining existing oral motor skills, encouraging pleasurable oral experiences, and tolerance of oral hygiene practices. Children who do not eat by mouth due to safety issues should receive non-nutritive oral motor stimulation to maintain their swallowing function. Benefits of non-nutritive stimulation include maintaining oral sensation and tolerance, saliva production and swallowing, and the facilitation of parent-child interactions. Parents of children who do not eat by mouth may require psychosocial support to process their feelings of guilt, fear and loss of normalcy with regard to eating (Craig, 2013).

Parenteral nutrition should only be used when the gastrointestinal tract is dysfunctional. When initiating feedings in patients with significant recent weight loss or failure to thrive, whether enteral or parenteral nutrition is used, it is important to be aware of the refeeding syndrome. This syndrome refers to phosphorus depletion and alterations in potassium, magnesium, and glucose metabolism, resulting in severe metabolic and physiological complications. It is imperative to increase calorie delivery slowly with close laboratory monitoring.

Coordinated services

The provision of nutrition services and prevention of further disabling conditions can be done in a variety of health care, school, vocational, home, and community settings. Participation in physical fitness activities targeting muscle strength and cardiorespiratory fitness should be encouraged to prevent secondary conditions such as chronic pain, fatigue, and osteoporosis. It is the responsibility of the family in concert with the health care team to promote nutrition care planning in these settings. More than 90% of children with CP live to adulthood; however, their life expectancy is less than that of the general population. Similar to the general population, individuals with CP are susceptible to developing obesity, hypertension, diabetes and heart disease and therefore require awareness and adherence to prudent dietary guidelines. The chronicity of nutrition problems for individuals with CP is recognized and has in part created a need for care coordination and integrated service planning to provide meaningful and cost-effective services.

Summary

Nutritional challenges are common in children with CP. The benefits of nutrition intervention can enhance overall health and development and improve the quality of life for the child and family. All interdisciplinary plans including nutrition should be safe, promote growth or weight gain without excessive energy expenditure and reflect the family's resources in time and skill, addressing their concerns and expectations. Early detection, assessment and treatment of nutritional deficiencies is imperative to avoid complications and improve outcomes in children with CP.

References

Ahmad, R., Rahman, N.A., Hasan, R., Yaacob, N.S., Ali, S.H., 2019. Oral health and nutritional status of children with cerebral palsy in northeastern peninsular Malaysia. Spec. Care Dent. 40 (1), 62–70.

Andrew, M.J., Parr, J.R., Sullivan, P.B., 2012. Feeding difficulties in children with cerebral palsy. Arch. Dis. Child. Educ. Pract. Ed. 97 (6), 222–229. Available at: https://pubmed.ncbi.nlm.nih.gov/22293504/. (Accessed 12 July 2021).

Arvedson, J.C., 2013. Feeding children with cerebral palsy and swallowing difficulties. Eur. J. Clin. Nutr. 67 (S2), S9–S12. Available at: https://www.nature.com/articles/ejcn2013224. (Accessed 13 April 2019).

Asgarshirazi, M., Farokhzadeh-Soltani, M., Keihanidost, Z., Shariat, M., 2017. Evaluation of feeding disorders including gastro-esophageal reflux and oropharyngeal dysfunction in children with cerebral palsy. J. Fam. Reprod. Health 11 (4), 197–201. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6168757/. (Accessed 12 July 2021).

Becker, P., Carney, L.N., Corkins, M.R., Monczka, J., Smith, E., Smith, S.E., Spear, B.A., White, J.V., 2014. Consensus statement of the Academy of Nutrition and Dietetics/American Society for parenteral and enteral nutrition. Nutr. Clin. Pract. 30 (1), 147–161.

Bell, K.L., Samson-Fang, L., 2013. Nutritional management of children with cerebral palsy. Eur. J. Clin. Nutr. 67 (S2), S13–S16.

Bell, K.L., Benfer, K.A., Ware, R.S., Patrao, T.A., Garvey, J.J., Arvedson, J.C., Boyd, R.N., Davies, P.S., Weir, K.A., 2019. Development and validation of a screening tool for feeding/swallowing difficulties and undernutrition in children with cerebral palsy. Dev. Med. Child Neurol. 61.

Benfer, K.A., Weir, K.A., Boyd, R.N., 2012. Clinimetrics of measures of oropharyngeal dysphagia for preschool children with cerebral palsy and neurodevelopmental disabilities: a systematic review. Dev. Med. Child Neurol. 54 (9), 784–795.

Benfer, K.A., Weir, K.A., Bell, K.L., Ware, R.S., Davies, P.S.W., Boyd, R.N., 2017. Oropharyngeal dysphagia and cerebral palsy. Pediatrics 140 (6), e20170731.

Bertoncelli, C.M., Altamura, P., Vieira, E.R., Bertoncelli, D., Latalski, M., Berthet, S., Solla, F., 2020. Predictive model for gastrostomy placement in adolescents with developmental disabilities and cerebral palsy. Nutr. Clin. Pract. 35 (1), 149–156. Available at: https://pubmed.ncbi.nlm.nih.gov/31134674/. (Accessed 12 July 2021).

Brooks, J., Day, S., Shavelle, R., Strauss, D., 2011. Low weight, morbidity, and mortality in children with cerebral palsy: new clinical growth charts. Pediatrics 128 (2), e299–e307. Available at: http://www.lifeexpectancy.org/articles/NewGrowthCharts.pdf. (Accessed 12 January 2020).

Caselli, T.B., Lomazi, E.A., Montenegro, M.A.S., Bellomo-Brandão, M.A., 2017. Assessment of nutritional status of children and adolescents with spastic quadriplegic cerebral palsy. Arq. Gastroenterol. 54 (3), 201–205. Available at: https://pubmed.ncbi.nlm.nih.gov/28723982/. (Accessed 13 July 2021).

CDC, 2018. Data and Statistics for Cerebral Palsy [online]. Centers for Disease Control and Prevention. Available at: https://www.cdc.gov/ncbddd/cp/data.html.

Craig, G.M., 2013. Psychosocial aspects of feeding children with neurodisability. Eur. J. Clin. Nutr. 67 (S2), S17–S20.

Culley, W.J., Middleton, T.O., 1969. Caloric requirements of mentally retarded children with and without motor dysfunction. J. Pediatr. 75 (3), 380–384. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0022347669802623. (Accessed 12 July 2021).

Day, S.M., Strauss, D.J., Vachon, P.J., Rosenbloom, L., Shavelle, R.M., Wu, Y.W., 2007. Growth patterns in a population of children and adolescents with cerebral palsy. Dev. Med. Child Neurol. 49 (3), 167–171. Available at: https://pubmed.ncbi.nlm.nih.gov/17355471/. (Accessed 12 July 2021).

Elawad, M.A., Sullivan, P.B., 2001. Management of constipation in children with disabilities. Dev. Med. Child Neurol. 43 (12), 829–832. Available at: https://pubmed.ncbi.nlm.nih.gov/11769270/. (Accessed 12 July 2021).

Graham, D., Paget, S.P., Wimalasundera, N., 2019. Current thinking in the health care management of children with cerebral palsy. Med. J. Aust. 210 (3), 129–135.

Gurka, M.J., Kuperminc, M.N., Busby, M.G., Bennis, J.A., Grossberg, R.I., Houlihan, C.M., Stevenson, R.D., Henderson, R.C., 2009. Assessment and correction of skinfold thickness equations in estimating body fat in children with cerebral palsy. Dev. Med. Child Neurol. 52 (2), e35–e41.

Holliday, M.A., Segar, W.E., 1957. The maintenance need for water in parenteral fluid therapy. Pediatrics 19 (5), 823–832. Available at: https://pubmed.ncbi.nlm.nih.gov/13431307/. (Accessed 12 July 2021).

IDDSI, 2019. IDDSI [online]. Available at: https://iddsi.org/.

Jesus, A.O., Stevenson, R.D., 2020. Optimizing nutrition and bone health in children with cerebral palsy. Phys. Med. Rehabil. Clin 31 (1), 25–37.

Krick, J., Murphy-Miller, P., Zeger, S., Wright, E., 1996. Pattern of growth in children with cerebral palsy. J. Am. Diet Assoc. 96 (7), 680–685. Available at: https://pubmed.ncbi.nlm.nih.gov/8675911/. (Accessed 12 July 2021).

Kuperminc, M.N., Stevenson, R.D., 2008. Growth and nutrition disorders in children with cerebral palsy. Dev. Disab. Res. Rev. 14 (2), 137–146.

Leonard, M., Dain, E., Pelc, K., Dan, B., De Laet, C., 2020. Nutritional status of neurologically impaired children: impact on comorbidity. Arch. Pediatr. 27 (2), 95–103. Available at: https://pubmed.ncbi.nlm.nih.gov/31791829/. (Accessed 19 July 2021).

Liptak, G.S., O'Donnell, M., Conaway, M., Chumlea, W.C., Worley, G., Henderson, R.C., Fung, E., Stallings, V.A., Samson-Fang, L., Calvert, R., Rosenbaum, P., Stevenson, R.D., 2001. Health status of children with moderate to severe cerebral palsy. Dev. Med. Child Neurol. 43 (06), 364.

Liu, L.-F., Moyer-Mileur, L., Gooch, J., Samson-Fang, L., 2008. The contribution of tone to resting energy expenditure in children with moderate to severe cerebral palsy: evaluation utilizing intrathecal baclofen injection. J. Pediatr. Rehabil. Med. 1 (2), 163–167. Available at: https://pubmed.ncbi.nlm.nih.gov/21791760/. (Accessed 19 July 2021).

McGrath, A., 2016. What is Cp – Infographic Poster – English (USA) – World Cp Day [online] World CP Day. Available at: https://worldcpday.org/download/what-is-cp-infographic-poster-english-usa/. (Accessed 12 July 2021).

Minocha, P., Sitaraman, S., Choudhary, A., Yadav, R., 2018. Subjective global nutritional assessment: a reliable screening tool for nutritional assessment in cerebral palsy children. Indian J. Pediatr. 85 (1), 15–19. Available at: https://www.ncbi.nlm.nih.gov/pubmed/29022211. (Accessed 14 October 2019).

Neuman, A., Desai, B., Glass, D., Diab, W., 2014. Superior mesenteric artery syndrome in a patient with cerebral palsy. Case Rep. Med. Available at: https://www.hindawi.com/journals/crim/2014/538289/. (Accessed 15 July 2021).

Palisano, R.J., Rosenbaum, P., Bartlett, D., Livingston, M.H., 2008. Content validity of the expanded and revised gross motor function classification system. Dev. Med. Child Neurol. 50 (10), 744–750. Available at: https://www.onlinelibrary.wiley.com/doi/full/10.1111/j.1469-8749.2008.03089.x. (Accessed 19 September 2019).

Patel, D.R., Pratt, H.D., Patel, N.D., 2008. Team processes and team care for children with developmental disabilities. Pediatr. Clin. 55 (6), 1375–1390.

Paulson, A., Vargus-Adams, J., 2017. Overview of four functional classification systems commonly used in cerebral palsy. Children 4 (4), 30.

Physiopedia, 2010. Feeding the Child With Cerebral Palsy - A Speech Therapists Perspective [online]. Available at: https://www.physio-pedia.com/Feeding_the_Child_with_Cerebral_Palsy_-_a_speech_therapists_perspective.

Ptomey, L.T., Wittenbrook, W., 2015. Position of the Academy of Nutrition and Dietetics: nutrition services for individuals with intellectual and developmental disabilities and special health care needs. J. Acad. Nutr. Diet. 115 (4), 593–608.

Rempel, G., 2015. The importance of good nutrition in children with cerebral palsy. Phys. Med. Rehabil. Clin 26 (1), 39–56.

Romano, C., van Wynckel, M., Hulst, J., Broekaert, I., Bronsky, J., Dall'Oglio, L., Mis, N.F., Hojsak, I., Orel, R., Papadopoulou, A., Schaeppi, M., Thapar, N., Wilschanski, M., Sullivan, P., Gottrand, F., 2017. European society for paediatric gastroenterology, hepatology and nutrition guidelines for the evaluation and treatment of gastrointestinal and nutritional complications in children with neurological impairment. J. Pediatr. Gastroenterol. Nutr. 65 (2), 242–264.

Samson-Fang, L., Bell, K.L., 2013. Assessment of growth and nutrition in children with cerebral palsy. Eur. J. Clin. Nutr. 67 (S2), S5–S8.

Scarpato, E., Staiano, A., Molteni, M., Terrone, G., Mazzocchi, A., Agostoni, C., 2017. Nutritional assessment and intervention in children with cerebral palsy: a practical approach. Int. J. Food Sci. Nutr. 68 (6), 763–770.

Secker, D.J., Jeejeebhoy, K.N., 2012. How to perform subjective global nutritional assessment in children. J. Acad. Nutr. Diet. 112 (3), 424-431.e6.

Sellers, D., Mandy, A., Pennington, L., Hankins, M., Morris, C., 2014. Development and reliability of a system to classify the eating and drinking ability of people with cerebral palsy. Dev. Med. Child Neurol. 56 (3), 245–251. Available at: https://pubmed.ncbi.nlm.nih.gov/24344767/. (Accessed 12 July 2021).

Stavsky, M., Mor, O., Mastrolia, S.A., Greenbaum, S., Than, N.G., Erez, O., 2017. Cerebral palsy—trends in epidemiology and recent development in prenatal mechanisms of disease, treatment, and prevention. Front. Pediatr. 5 (21).

Walker, J.L., Bell, K.L., Caristo, F.M., Boyd, R.N., Davies, P.S.W., 2011. A review of energy intake measures used in young children with cerebral palsy. Dev. Med. Child Neurol. 53 (6), 569. Available at: https://pubmed.ncbi.nlm.nih.gov/21506998/. (Accessed 12 July 2021).

www.lifeexpectancy.org, 2011. Life Expectancy for CP, VS, TBI and SCI [online]. Available at: https://www.lifeexpectancy.org/articles/GrowthCharts.shtml. (Accessed 27 July 2021).

Colon: Structure, function, and disorders

Asim Maqbool[a,b], [a] Program for Pancreatic Disorders, Division of Gastroenterology, Hepatology and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, PA, United States; and [b] The University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A. Maqbool, Colon: Structure, Function, and Disorders, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 378–396, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00059-3.

Introduction **185**
Structure and function **185**
Gross morphology 185
Vasculature 185
Innervation 186
Histology 187
Function **188**
Electrolyte transport: ion channels 188
Fluid transport 189
The ENS and gastrointestinal motility 191
Colonic motility 191
Defecation 191
Colonic immune function and colonic microflora 192
The colonic microbiome 192
The colon and energy metabolism 194
The colon and nitrogen metabolism 194
Disorders of the colon **195**
Diarrhea 195
Protein losing enteropathy (PLE) 196
Infections and enteric parasites 196
Polyps 197
IBD 197
Epidemiology 197
Etiology 197
Genetics 199
Environmental influences 200
Pathogenesis 201
Pathology 202
Inflammatory bowel disease and the terminal ileum 203
Extraintestinal manifestations 204
Nutritional consequences of inflammatory bowel disease 205
Treatment of IBD 205
Conclusion **205**
References **206**
Further reading **206**

Key points

To provide the reader with an understanding of:

- The anatomy and structure of the colon
- The enteric nervous system
- The intestinal microbiota, probiotics and prebiotics
- Intestinal disorders, including diarrhea, polyps, protein-losing enteropathy, infections, inflammation, and inflammatory bowel disease

Glossary

Diarrhea Defined as either a decrease in stool consistency or an increase in stool frequency and volume

Enteric nervous system The ENS operates both in conjunction with and independent of the peripheral nervous system

Inflammatory bowel disease A chronic inflammatory condition of unclear etiology, categorized by location, extent, severity of disease, in addition to being a strictly mucosal process vs. a transmural process

Polyps Intestinal polyps are intraluminal protuberant tumors characterized by their gross morphological appearance, location(s), numbers, size, and presence (pedunculated) or absence (sessile) of the stalk

Prebiotics Naturally occurring nutrients that support the intestinal microbiome

Probiotics Bacteria present in nature and in naturally fermented foods that have benefits for the host following ingestion

Protein losing enteropathy Protein losses occurring from the colon, usually of an exudative process, and not only related to inflammatory conditions, but also to perturbations in oncotic and hydrostatic pressure equilibrium

The intestinal microflora and the microbiome The complement of bacteria that normally resides in the colon

Introduction

The colon is not a static structure but only serves as a reservoir and conduit for stool and expulsion, but rather, is a dynamic organ involved in the absorption of salts, fluids and nutrients, protein and energy balance. The colon hosts immune tissue involved in the dynamic interplay with intestinal micro flora, important in both health and disease. This article reviews up-to-date understanding of the structure and function of the colon, both in health and disease states.

Structure and function

Gross morphology

The colon is a continuous structure originating at the ileocecal valve and extending to the anus. The cecum is the first part of the colon, which lies in a posterior position at the right iliac fossa, and has an ovoid-like shape. This cavity is more generous in proportion than other compartments of the colon. The appendix (a blind-ending outpouching) originates in the cecum and its opening is usually visible during colonoscopy.

The ascending colon runs cephalad and anteriorly from the cecum to just inferior to the liver, to the hepatic flexure, emerging into the peritoneum. The transverse colon continues from the hepatic flexure to the splenic flexure, from where it travels distally, and once again posteriorly to the sigmoid colon, an S-shaped, tortuous, narrow peritoneal structure. At the peritoneal reflection the rectum arises and, closely following the sacral curve leads to the anal canal. The rectum is a vault-like structure, which can distend in order to accommodate fecal load. The anal canal bears two sphincters, an internal and an external anal sphincter. The internal sphincter is comprised of inner circular smooth muscle fibers, and a distal external fiber on the other side of a muscular pelvic diaphragm. The fibers of the external sphincter are intertwined with those of the levator ani, tethered anteriorly and posteriorly to the perineal body and the coccyx, respectively (Fig. 1).

With respect to colonic mobility within the abdominal, peritoneal, and pelvic cavities, the cecum and flexures are less mobile, with the sigmoid colon being the most mobile. The transverse colon supports the greater omentum and has a variable degree of mobility.

Cross-sectionally, the colon has an external longitudinal muscle and an inner layer of circular musculature, the former has a coalescence of fibers forming band-like structures known as teniae. These teniae are particular to the large intestine, and are located at one-third of the circumference from each other, and run continuously from one end of the colon to the other. Haustra are hemilunar-like outpouchings, which are present between teniae. The more proximal rectal tenial fibers surround the rectum; the inner fibers form the internal anal sphincter. The external fibers are intertwined with those of the levator ani, and sandwiched between fibers running anterior to posterior, from the peroneal body to the coccyx, forming the external sphincter.

Vasculature

The ascending colon and portions of the transverse colon are perfused by branches of the superior mesenteric artery, with the remainder of the colon receiving arterial blood from tributaries of the inferior mesenteric artery. Distal iliac arterial branches perfuse the anal canal. Venous drainage is achieved via the superior and inferior mesenteric veins lying in close approximation to their arterial counterparts, and subsequently dumping into the portal vein.

Additional gross morphologic structures include "lymphatic vessels," in close approximation to the vasculature, leading to lymph nodes in the celiac, superior, and inferior preaortic regions. Perianal drainage is via the inguinal lymph nodes.

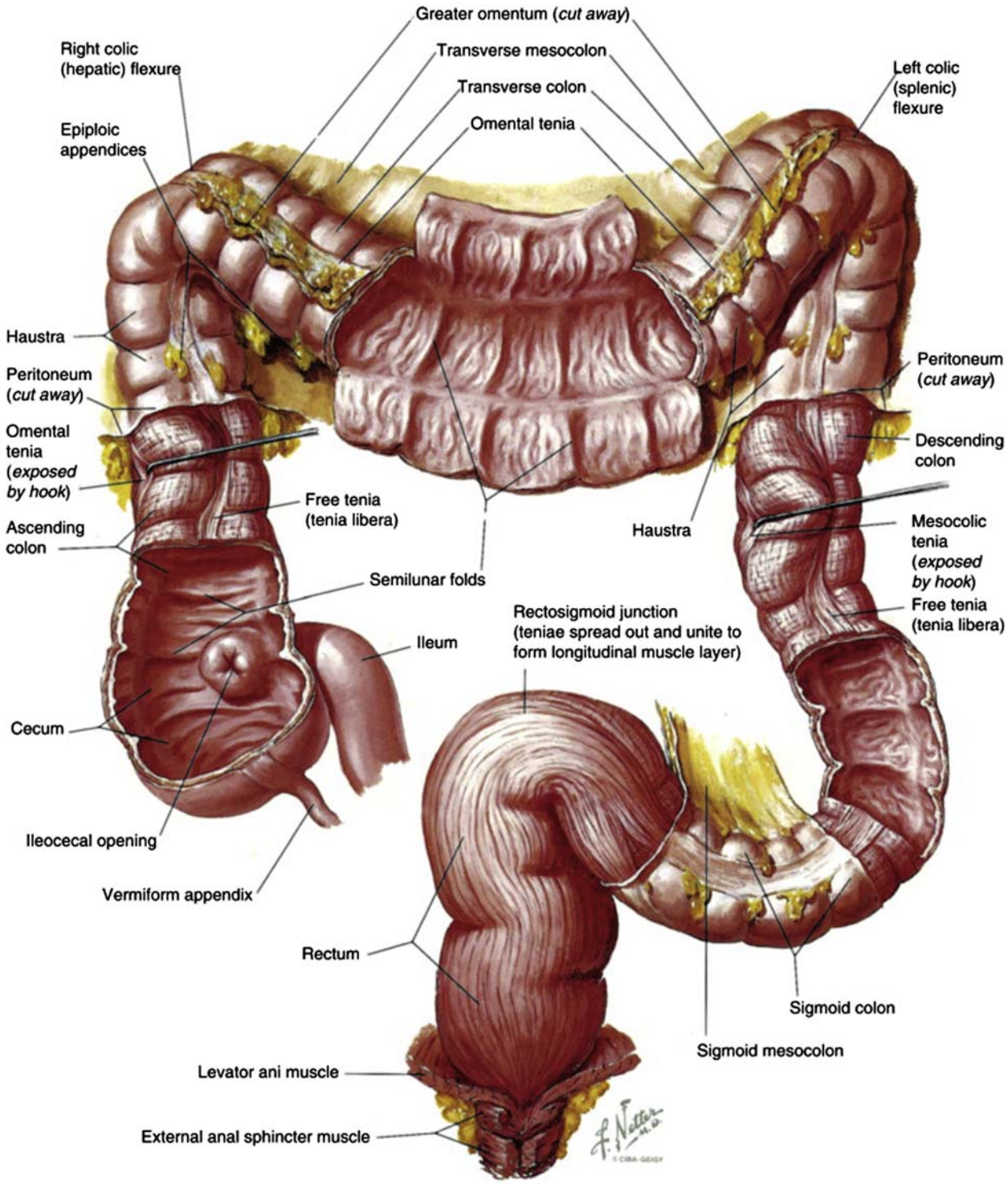


Fig. 1 Mucosa and musculature of the large intestine. Reproduced from Netter, F., 1995. Atlas of Human Anatomy. Geneva, Ciba-Geigy, with permission from Anatomy Atlas.

Innervation

Parasympathetic innervation to the proximal colon is provided via the vagus nerve; the distal colon and rectum are innervated via pelvic parasympathetic fibers. The sympathetic nervous system innervates the proximal colon via lower thoracic fibers, and the distal colon and rectum via lumbar fibers. Prevertebral sympathetic vertebrae receive fibers from neurons projecting out of the gut.

Histology

Cross-sectionally, the intestinal wall is divided into four layers, with the serosa, a monolayer of mesothelial cells comprising the outermost, followed by the muscularis externa. These muscle layers comprise an external longitudinal layer and an internal circular layer. Sandwiched in-between these two layers lies Auerbach's (myenteric) plexus. The submucosa is the next more medial layer; a rich admixture of cells, including structural elements such as fibroblasts and dense connective tissue, immunologically important cells (plasma cells, lymphocytes, macrophages, eosinophils, mast cells) in addition to vascular tissue and innervation to Meissner's plexus (ganglion cells), and lymphatics comprise this layer. The muscularis mucosa, a thin sheet of smooth muscle, separates the deeper submucosa from the mucosa. The lamina propria runs interior to this layer, and is composed of connective tissue, and is lined by the luminal epithelium (Fig. 2A).

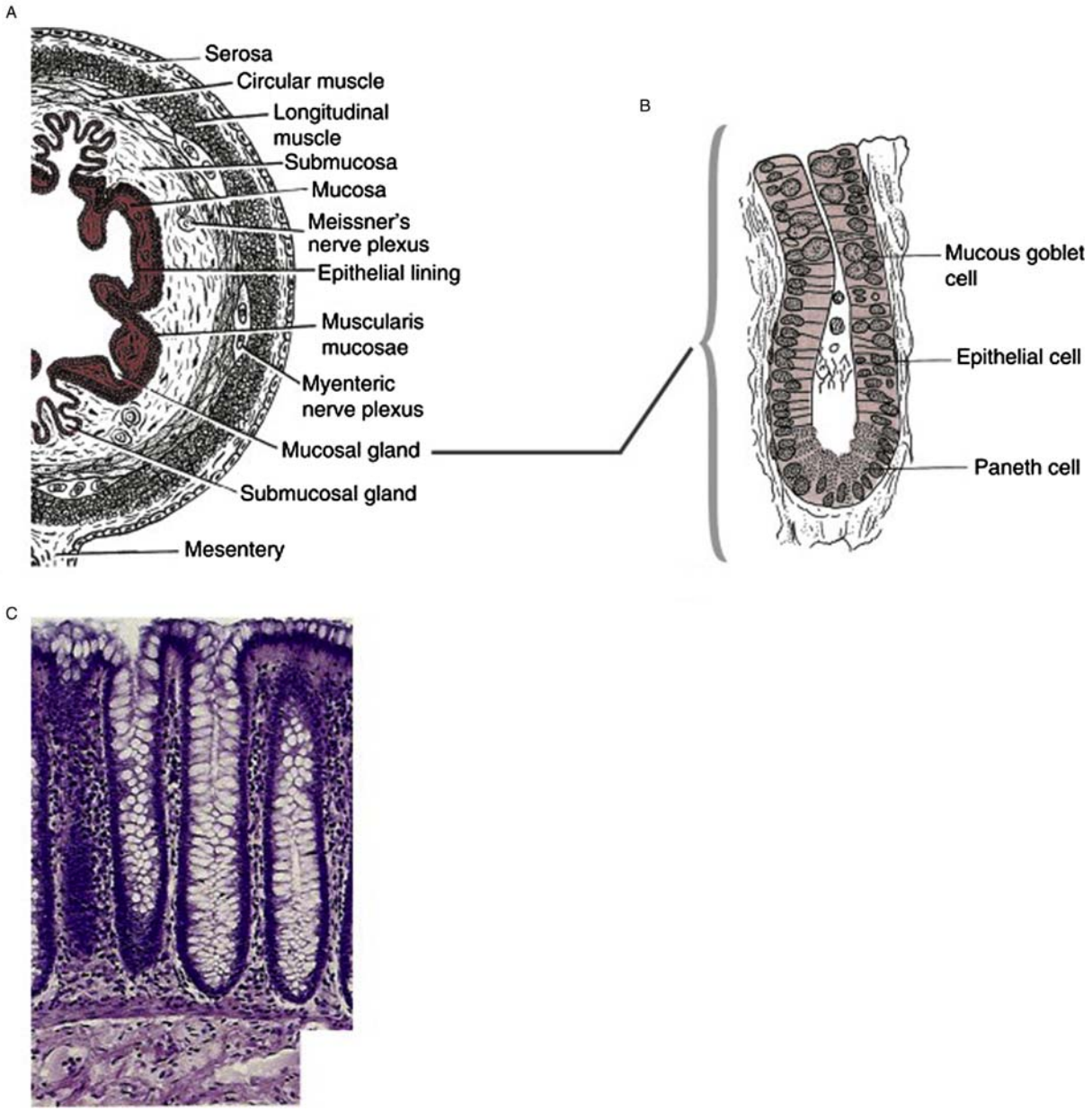


Fig. 2 (A) and (B): Cross sections of the gut and a colonic crypt. Reproduced from Guyton, A.C., 1991. Textbook of Medical Physiology, eighth ed. Philadelphia, WB Saunders Company. (C) H and E stain of a typical colonic crypt. Reproduced from Burkitt, H.G., Young, B., Heath, J.W., 1993. Wheater's Functional Histology, third ed. London, Churchill Livingstone.

Table 1 Colonic cell types.

Cell type	*Location*	*Function(s)*
Stem cells	Crypt (base) • Nonmigratory until differentiated	Pluripotent
Undifferentiated crypt cell	Crypt	Secrete water and chloride into intestinal lumen
Paneth cells	Crypt base • Nonmigratory • Basophilic cytoplasm • Proximal one-third of colon only	• Growth factor secretion and digestive enzyme synthesis • Antimicrobial peptide synthesis and release
Goblet cells	Colonic crypt • Most common cell type in the colon	Mucin release
Enteroendocrine cells	Mostly in small intestine	Receptor-mediated epithelial cell function modulators
Enterocytes	• Basolateral membrane Predominantly small intestinal; present in the colon	• Digestive enzyme synthesis (small intestine) • Ion transporters and channels involved in fluid and electrolyte transport
M cells	Small and large intestines	Bind, process, and present antigens to components of the mucosal lymphoid immune system
Intraepithelial lymphocytes	• Overlying lymphoid follicles Small and large intestines • Basolateral membranes	• Memory T cells • Mucosal immune defense

The intestinal epithelium is a tight monolayer of cells that functions to absorb nutrients, electrolytes, and liquids, as well as to secrete mucus and fluids. The epithelial surface is punctuated by numerous tightly packed crypts, which contain epithelial precursor cells, enteroendocrine cells, other undifferentiated cells, and Paneth cells. Goblet cells, which secrete mucin, are also located in the crypt (**Fig. 2**; **Table 1**). As undifferentiated and precursor cells mature, they migrate superiorly to the surface and to the monolayer of absorptive cells present in crypts. The average lifespan of a colonocyte is 3–6 days.

The absorptive colonocyte develops short microvilli while in the colonic crypt, which elongate during its migration to the surface. The hydrophobic lipid bilayer of the colonocyte epithelium prevents passive transport of charged particles. The epithelial membrane contains specific protein transporters, carrier proteins, and channels allowing electrolyte transport. The electrochemical gradient formed by active transport facilitates passive flow across cell membranes.

Function

Electrolyte transport: ion channels

Fluids and electrolytes are absorbed via one of two pathways, transcellular vs. paracellular. Active and passive transport systems exist via both these pathways.

There is a clear polarity to the distribution of protein transporters, channels, and pumps distinguishing the apical from the basolateral membrane. Active transport utilizes transcellular, energy-driven protein pumps or channels to facilitate passage of electrolytes from an area of low concentration to one of high concentration/electrochemical gradient. A prime example of this is the Na-K-ATPase pump, the principal pump present along the basolateral membrane. The net effect of the three Na ions expelled for every two K ions accepted into the cell is a lowered intracellular Na content, and resultant net negative charge. The negative charge formed by this active transport creates an electrochemical gradient facilitative to the passive flow for other ions across the cell membrane, a process known as secondary active transport (**Figs. 3** and **4**).

Ion transporters may be additionally subclassified into symporters, in which ions move in the same direction, or antiporter, in which case ions move in opposite directions across the cell membrane. Cotransport of ions with other molecules, such as that of Na and glucose. The intracellular concentration of glucose is regulated both by uptake at the apical surface as well as by exit through the basolateral membrane, allowing for conditions favorable to uptake from the lumen. The Na-glucose transporter system allows for therapeutic interventions, such as the use of oral rehydration solution (ORS) in cases of severe diarrhea related to cholera or other processes. Similar cotransporters exist linked to the transport of bile salts and amino acids (**Table 2**).

Although sodium is the primary cation involved in ion transport, short-chain fatty acids (SCFAs) constitute the primary anion in the colon and primary metabolic fuel for colonocytes. Their transport is postulated to be linked to Na-H transporters and pH, specific bicarbonate linked transporters, and the concentration gradient across cell membranes. Chloride transport occurs via both active and passive processes, and is the major intestinal anion involved in the intestinal secretion of fluids.

Colonic smooth channels also possess ion channels, and are involved in active and secondary ion transport processes involving calcium. The electrochemical gradient formed by the activity of these ion channels facilitates the function of smooth muscle action potential generation on depolarization. With the generation of smooth muscle action potentials attaining threshold voltage, contractility of the smooth muscle is possible. The efflux of calcium into these active transport channels activates the process of

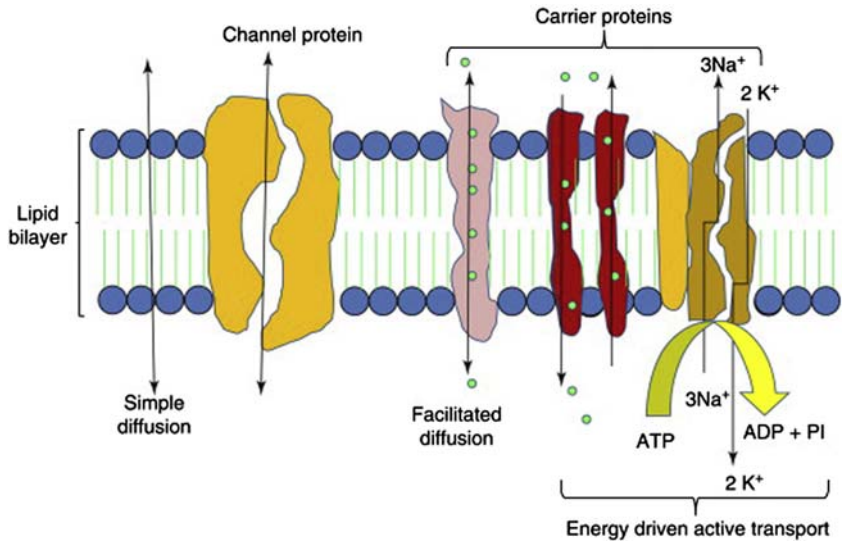


Fig. 3 Electrolyte transporters and the cell membrane.

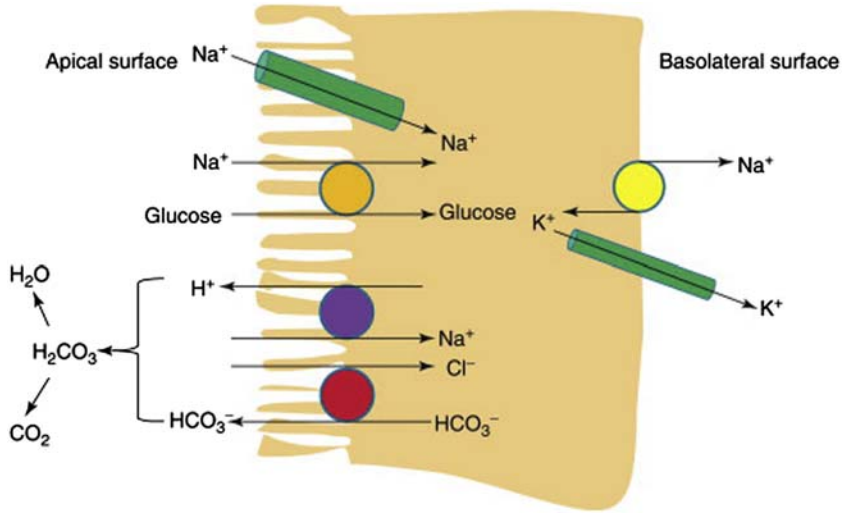


Fig. 4 Electrolyte transport at the colonocyte level.

contraction. Interaction with the enteric nervous system (ENS) stimulates the release of calcium ions stored in intracellular stores. The function of ion channels can be modified by calcium channel-blocking drugs. This contractile activity, when occurring in a coordinated fashion and modulated by neurotransmission, effects peristalsis and colonic motility, which are discussed in this article.

Fluid transport

There is an evident heterogeneity to the mucosal epithelium dependent on location in the alimentary canal, in several aspects. The type, variety, and number of ion transporters, channels, and carrier proteins vary from region to region, i.e., from jejunum to colon. Additionally, the nature of interepithelial cell junctions varies from the proximal to distal intestinal tract, influencing the "leakiness" of the respective regions. Finally, a clear gradient in cell composition and function between colonic crypt cells and those on the surface exists. Physiologic heterogeneity follows the aforementioned patterns, defining tissue function in these respective areas. For example, the colonic crypts have more of a secretory function, whereas the villus structures seen most notably in the jejunum exhibit a greater absorptive function. This heterogeneity is key in understanding changes in intraluminal osmolality and fluid shifts that occur in the intestine.

Approximately 98% of the daily fluid load handled by the intestine is reabsorbed—approximately 9 L day^{-1}. Of this, the jejunum absorbs 85% and the colon absorbs approximately 13% or 1.5 L.

Table 2 Electrolyte transport: examples.

Ion	*Transporter*	*Location*	*Type*	*Function(s)*
Na	Na, K-ATPase	Basolateral membrane	Active; antiport	Principal ion involved in water absorption
	Na-H exchangers	Apical and basolateral membrane	Secondary; antiport	
Na and Cl	NaCl	Apical	Antiport; passive; electrochemically neutral	
Cl	Protein channel	Apical	Diffusion; passive (secretion) and some active transport proteins at the apical surface (absorption and secretion), including CFTR	Principal ion involved in water secretion Basal rate of secretion influenced by several mediators (endocrine, paracrine, neural, luminal, etc.)
Cl	Protein channel			
K	Protein channel		• Antiport active transport (basolateral membrane) • Active secretion at the apical membrane; linked to Cl transport function • Absorptive active apical K- ATPase pumps in distal colon	
HCO3		Apical and basolateral channels;	• Alkaline phosphatase linked • Passive transport mechanisms • Na-HCO_3 cotransporter postulated • CFTR-synchronized apical channel and Cl-HCO3 exchanger postulated	
Short-chain fatty acids	Apical		• Postulated link to Na-H ion transport	Principal anion of the colon

CFTR, cystic fibrosis transmembrane conductance regulator.

Passive reabsorption of water occurs in the intestines, regulated primarily by electrolyte transport, i.e., following an osmotic gradient. Na-driven/related transport mechanisms are the primary driving force allowing water absorption. This osmotic gradient facilitates water absorption via both transcellular and paracellular pathways.

Transcellular water transport mechanisms such as aquaporins, or water channels, have been described. The paracellular pathway of water transport has been studied extensively, a process often described as "solvent drag" (Fig. 5).

The leakiness of paracellular pathways that varies by location in the lower alimentary tract (more prominent in the jejunum, with subsequent decrease distally), and the magnitude of the osmotic gradient (also effected by dietary Na content) are important factors affecting solvent drag. The nature of the intercellular junctions in a particular region of the colon determines the permeability, or

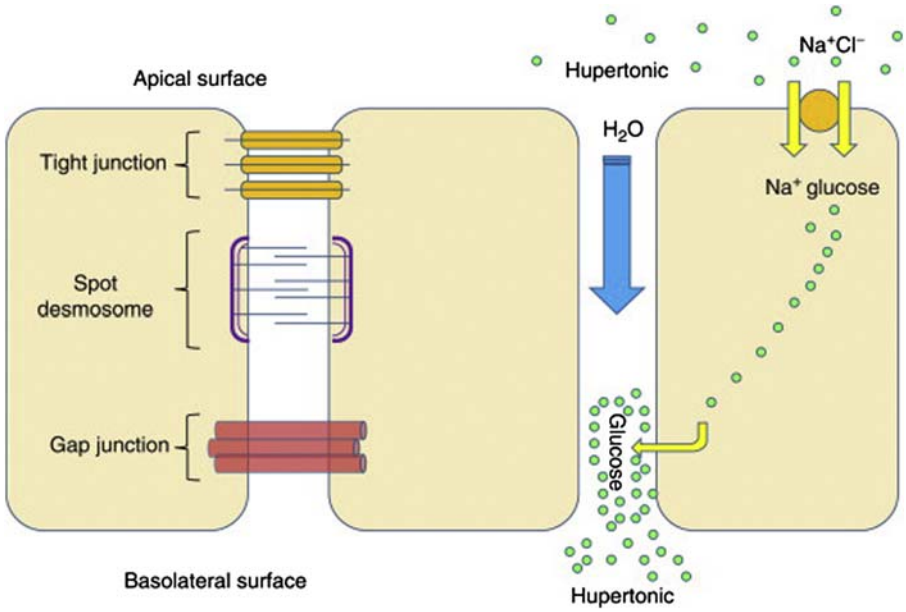


Fig. 5 Intercellular junctions and fluid transport across the cell membrane.

"leakiness" of that particular epithelial area. Several intercellular structures have been described, including the *zona occludens* (tight junction), desmosomes (connections between cells), and the *zona adherens*, the latter function in cell adhesion and therefore contribute to maintaining cellular polarity across the membrane. *Zona occludens* are more apical in location and form junctional complexes between cells. It has been postulated that these junctional complexes may be more dynamic than previously believed, responding to signaling mechanisms and subject to regulation, thereby influencing their function and resultant permeability characteristics (Fig. 5).

The ENS and gastrointestinal motility

The ENS operates both in conjunction with and independent of the peripheral nervous system. Nerve plexi exist within the bowel wall, with Auerbach's plexus sandwiched between longitudinal and circular muscle layers, and Meissner's plexus located more medially in the submucosa.

The ENS is the largest component of the autonomic nervous system, based on nerve cell number.

Interstitial cells of Cajal, a cell type unique to the alimentary tract, are present medial to the inner smooth muscle layer. These specialized cells interact with myenteric neurons, and are thought to exhibit independent electrical activity, generating and transmitting slow waves to smooth muscle, functioning as pacemakers for colonic motility. The ENS is capable and does functions independent of the central nervous system (CNS), with reflex activity, in response to luminal stimuli, including muscle contraction and coordination—i.e., motility, blood flow, and glandular secretion. Modulation of the ENS is via the sympathetic and parasympathetic nervous system.

Colonic motility

The colon functions to delay passage of luminal contents to allow for water absorption, and to allow for mixing of luminal contents with the mucosa, to store fecal matter before defecation, and to propel contents forward during defecation.

The frequency and duration of propagative, high-pressure waves in the colon in part is determined by pressure exerted by the intraluminal contents (mechanical) and degree of stretch stimulation, (chemical) composition of the contents, and by other stimuli interacting with the colon.

The gastrocolic reflex, an anterograde postperistaltic process, occurs following a meal, originating proximally and propagating anterograde. Both the caloric content and the fat composition of the meal bear influence on colonic peristalsis. Likewise, gastric distention by food contents, water, or gas also has a stimulatory effect. Gastrointestinal hormones secreted in response to a meal, such as cholecystokinin, are thought to mediate peristaltic responses to a (fatty) meal. Irritant laxatives also stimulate peristalsis, even when administered rectally. Opiates are known to inhibit the ENS, and, as a consequence, retard peristalsis. Colonic motility diminishes significantly during sleep, resuming on awakening.

Motor activity varies by colonic region, in degree, frequency, amplitude, velocity, being propagative vs. nonpropagative (the latter are more common in the distal colon than the former), relative distance of propagation, and direction of propagation (anterograde vs. retrograde, the latter most commonly seen in the proximal colon). Approximately one-third of these colonic peristaltic waves are propulsive, and the ones associated with propulsion of stool tend to be slower, yet greater in amplitude.

Defecation

Defecation involves the integration of peristaltic activity in most colonic regions, and not solely in the anorectal region. In the predefecatory phase, approximately an hour before actual defecation, the majority of the colon exhibits an increase in propulsive peristaltic waves, first in the proximal colon, then advancing distally.

The sensation of defecatory urge is not evident until approximately 15 min before defecation. At that time, there is a marked increase in propagative peristaltic activity, originating more distally in the colon. Each of these late propagative waves successively originates "proximate" to its preceding one, with "greater" amplitude, and present over a "greater" distance of colonic length.

Stool contact with the receptors in the upper anal canal can effect relaxation of the inner anal sphincter. In addition, stretch receptor stimulation of the rectal-vault walls results in the urge to defecate. Failure of relaxation of the external anal sphincter (which is under voluntary control) results in retrograde passage of stool into the rectum, with subsequent diminishing of more proximal peristaltic propagative waves, thereby maintaining continence when immediate defecation is not desirable or convenient.

Evacuation of the rectum and defecation require correcting the angle of the anal canal in the anterior–posterior plane; this is accomplished by assuming a squatting position. Contraction of the abdominal musculature and of the diaphragm, with a relaxed pelvic floor facilitates defecation, even in the absence of colonic peristalsis.

Stool size and consistency vary based on diet and water intake, and transit time determines size and consistency of the stool, as well as bacterial content (a major component of stool). Higher water content tends to result in larger, softer stools. The more fusiform-shaped the stool is, the less likely that its passage is associated with straining. The transit time through the colon is inversely related to the stool's water content and, hence, its consistency (Fig. 6).

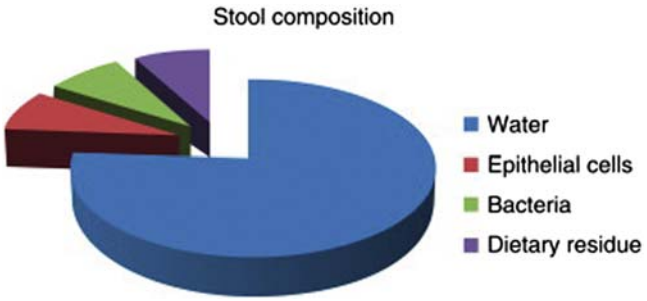


Fig. 6 Fecal composition.

Colonic immune function and colonic microflora

The immune system of the gastrointestinal tract functions in defending against infection (bacterial, viral, and parasitic) and luminal antigens ingested/formed by bacteria. Nonspecific and specific mechanisms exist.

The mucin secreted by colonic goblet cells functions as a barrier for the mucosal surface. Mucosal integrity serves an important barrier to luminal pathogens. Interepithelial cell junctions function both to control permeability as pertains to fluid and electrolyte absorption, as well as to preventing pathogen access beyond this layer.

The enteric immune system is vast and complex; it functions to interact with the rest of the immune system, as well as with luminal contents. Gut-associated lymphoid tissue consists of both discretely organized tissue, such as Peyer's patches (lymphoid follicles with proliferative potential in response to antigen presentation) containing M cells, and the more diffuse lymphocytes and macrophages distribution among the submucosa, mucosa, and lamina propria. M cells function in antigen sampling of intraluminal contents by binding antigens, endocytosis, antigen processing, and subsequent interaction with lymphocytes and macrophages within Peyer's patches, eliciting host responses. The lymphocyte complement of Peyer's patches originates in either the bone marrow or the thymus, enters the systemic circulation to migrate to Peyer's patches, interacts, returns to the intestinal mucosa or, via mesenteric lymph nodes, reenters the systemic circulation to reach other organs.

The gastrointestinal tract houses up to 80% of the body's immunoglobulin-producing cells. Intraepithelial T lymphocytes, plasma cells, macrophages, dendritic cells, eosinophils, and mast cells also function in a specialized manner.

Secretory immunoglobulin A (IgA) is an important host immune-defensive mechanism. Unlike the monomeric, systemic form of IgA, intestinal secretory IgA is polymeric (specifically, dimeric) in nature. This dimeric immunoglobulin is secreted by B lymphocytes situated in the lamina propria, and contains a unique "J" chain instrumental in polymer formation. This IgA binds to the Ig receptor of the epithelial cell on the basolateral membrane, and, following endocytosis and transport across the cell, is secreted from the apical side.

Secretory IgA binds to intraluminal antigens, including dietary ones, and functions in preventing their absorption. Additionally, secretory IgA has the ability to bind to microorganisms, hence preventing adherence, colonization, and invasion. Secretory IgA is secreted in breast milk, and in the breastfed neonate and infant confers a degree of passive immunity to infection by limiting luminal contents from interacting with, or directly binding to/invading the mucosa.

Interaction of intraluminal bacteria with the immune system may affect intestinal permeability, and may modulate the intestinal immune system. Certain bacterial species are believed to interact with other enteric flora as well as with the host immune system to effect a healthier gastrointestinal tract and enhanced nutrient digestion; organisms studied include Lactobacillus, Vibrio species, and Saccharomyces (commonly referred to as probiotics).

Regulation of quantity of bacteria, in addition to the specific profile of bacterial species present is dependent on a host of factors, including gastric acid output, gastrointestinal motility, luminal contents, and the milieu created therein. Additionally, the intraluminal environmental milieu is affected by the specific properties of different species of bacteria, their interactions with other luminal species, and with the host itself.

The colonic microbiome

The colon accommodates the largest number of enteric flora, on the order of 10^{10}–10^{12} greater than 100,000 the number of flora and more than 100-fold greater diversity of species than any other location in the alimentary canal. Efflux of bacteria into the ileum is hindered by the ileocecal valve, which functions to restrict several of these bacterial species from entering large intestine. The majority of these colonic bacteria are anaerobic in nature (Table 3).

The enteric flora plays several important roles, including interaction with the enteric immune system, effecting cellular immune activity, associated with the size and number of Peyer's patches present, influencing intestinal motility, and nutritively important functions, including bile salt deconjugation (facilitates enterohepatic circulation of bile salts), bilirubin metabolism (deconjugation and urobilin formation, allowing excretion), mucin degradation, and lipid metabolism (generation of SCFAs). Androgens and estrogens are hydrolyzed facilitating resorption and conservation of these sterols, whereas cholesterol is processed into coprostanol, a nonabsorbed sterol. Ammoniagenesis via protein and urea degradation may play a role in hepatic encephalopathy (Table 4).

Table 3 Colonic enteric flora.

Bacterial genus	*Prevalence (%)*	*Total count (CFUg^{-1} or ml)*
Anaerobes		10^{10}–10^{12}
		10^{9}–10^{12}
• *Bacteroides*	100	
• *Porphyromonas*	100	
• *Bifidobacterium*	30–70	
• *Lactobacillus*	20–60	
• *Clostridium*	25–35	
• *Peptostreptococcus*	–	
• *Peptococcus*	–	
• Methanogens	–	
Facultative aerobes		10^{2}–10^{9}
• *Enterococcus*	100	
• *Escherichia coli*	100	
• *Staphylococcus*	30–50	
• Other Enterobacteriaceae	40–80	

Table 4 Examples of biochemical reactions by intestinal flora.

Reaction type	*Reaction*	*Example substrate*
Hydrolysis	Amides	Methotrexate
	Glucuronides	Estradiol-3-glucuronide
Dehydroxylation	Decarboxylation	Amino acids
	Deamination	Amino acids
	Dehydrogenase	Bile acids, cholesterol
Reduction	Double bonds	Unsaturated fatty acids
	Acetylation	Histamine

Consumption of lipids, carbohydrates, and protein also occurs by colonic bacteria, in addition to that of vitamins (vitamin B_{12} and folic acid are consumed; vitamin K and biotin are produced by these bacteria).

The development of the colonic microbiome is influenced by many factors, including environmental factors such as site of birth (home vs. hospital, rural vs. urban hospital setting), of antibiotic exposure perinatally, of oxygen deprivation, and by mode of birth (vaginal vs. cesarian). Although it is well established that the microbiome bacterial profile is largely determined after birth and with the development of immunological tolerance within the first 2–3 day of life, with a preponderance of facultative anaerobes, it is also recognized that the dietary composition plays an important ongoing role in the maintenance and character of the microbiome. Differences in the microbiome may exist in children who are breast vs. formula fed, and the composition of the breast milk in turn is modulated by maternal dietary intake. Dietary fiber in particular is very important throughout the life cycle, as bacterial fermentation produces SCFAs, which are the preferred colonocyte energy substrate. Butyrate is one such SCFA of particular importance.

The composition of the intestinal microbiome has been associated to specific disease states, including obesity, inflammatory bowel disease (IBD), and cystic fibrosis (CF). In all these conditions, in addition to the microbiome being varied from their respective healthy counterparts, it is thought to be reduced in terms of the diversity.

In the case of obesity, a varied proportion of the intestinal microflora phyla (more Firmicutes and less Bacteroidetes) have been observed in the obese vs. lean individuals. From a microbial metagenomic perspective, these bacteria are enriched in genes associated with energy harvest. What is not clear at present is if this association is causative or reflective of being obese or not. What these observations do is to raise interesting questions regarding whether if the intestinal microbiome contributes significantly to energy balance in their respective human hosts. The intestinal microflora had previously not been considered to contribute to host energy balance in a clinically significant manner.

In IBD and CF, the microbiome differs from that of healthy individuals. In the case of IBD, there is a decrease in Bacteroidetes and Firmicutes and increase in Proteobacteria—i.e., there is a decrease in the protective flora and a corresponding increase in detrimental flora. This microbiome is thought to play a causative role in disease pathogenesis in susceptible individuals. Interestingly, in comparison to healthy individuals, the colonic butyrate production in individuals with ulcerative colitis (UC) is reduced. This may be indicative of both dietary differences between these two groups (a known risk factor for the development of IBD in genetically susceptible individuals) and possibly in the pathogenesis of the disease process.

Mucus secreted by colonic goblet cells forms a physical barrier to bacterial pathogens known as the unstirred layer, and is routinely cleared by the colon. In CF, as with pulmonary secretions, colonic secretions are underhydrated, and this renders the unstirred mucus layer thicker and less able to clear trapped bacteria, which is thought to contribute to colonic inflammation frequently seen in this condition.

Modulation of the intestinal microbiome is possible by dietary changes, the use of antibiotics, and of probiotics. In the case of IBD, the colonic microflora has often been a target for therapy, with the use of antibiotics useful as therapy for the disease, while colonic diversion is another modality sometimes used. Another strategy studied in digestive and inflammatory disorders of the colon involves use of probiotics and of prebiotics (mostly dietary oligofructose) to modulate and change the profile of the microbiome. Specifically, supplementation of single and multiple species of these probiotics has been studied in the prevention and treatment of antibiotic-associated diarrhea, bacterial overgrowth in short bowel syndrome, rotaviral infections, refractory post IBD resection-related pouchitis, irritable bowel syndrome and necrotizing enterocolitis, and for the treatment/prevention of recurrent *Clostridium difficile* colitis. In CF animal models, the use of probiotics and bowel hydration retention agents such as polyethylene glycol 3350 has been associated with a decrease in colonic markers of inflammation, suggesting that altering the intraluminal microbiome and intestinal milieu may have significant impact on health and disease states. The use of probiotics in other conditions that extend beyond the gastrointestinal system (such as allergic conditions and pulmonary disease) remains an active area of research. It is highly likely that the relationship between host and microbiome is dynamic and bidirectional, and that the effects extend beyond the colon, from a nutritional, metabolic, and immunological basis. However, many studies of probiotics have not yielded clinically significant results, and caution in interpreting these findings and in consideration to the roles and types of probiotics in use should be exercised. The human host develops immunotolerance within the first 2–3 days of life, and so if probiotic therapy is started, it may need to be continued to maintain/sustain microbiome changes; cessation of supplementation may allow the microflora to return to its preprobiotic supplementation profile. Some specificity of the type of probiotic for different conditions may need to be considered. Probiotic use may be contraindicated in immunocompromised individuals where the risks of bacterial translocation and sepsis are increased, and are currently contraindicated for use in pancreatitis, where their use has been associated with adverse outcomes.

The colon and energy metabolism

Although the role of the colon in fluid and electrolytes transport is well known, until recently, less was known about the colon and energy metabolism. The intestinal microflora-mediated fermentation of dietary fiber and utilization of SCFAs (acetate, propionate, and butyrate) are important in the maintenance of the colonic microbiome. SCFAs are also absorbed and contribute approximately 200 kcal d^{-1} to the human host.

The role of the colon in energy retention is of increased importance in cases of short bowel syndrome related to resection, and in relation to the small bowel length. In cases of small bowel resection, hypertrophy of colonic tissue occurs which enhances fluid and electrolyte retention capacity; however, malabsorption of long-chain fatty acids and carbohydrates may still occur. Although it is understood that the sites of small bowel resection and amount of resection are important (in addition to the status of the ileocecal valve, an important antibacterial barrier for the small bowel), the role of colonic retention may also be important. In this setting, colonic energy retention of fermentable carbohydrates—dietary fiber—into SCFAs and their absorption increases. In individuals with small bowel resections, dietary modification to increase fermentable carbohydrates may be an important strategy in decreasing malabsorption and maldigestion.

The colon and nitrogen metabolism

Colonic nitrogen metabolism involves both protein and nonprotein sources. Protein catabolism in the gastrointestinal tract can account for 10% of total body protein metabolism. The intestinal microbiome contributes much of the fecal matter and intraluminal nitrogen content and, as such, the majority of fecal nitrogen losses. Colonic ammonia specifically is the main nonprotein source of nitrogen in the colon. This ammonia is obtained by the hydrolysis of urea by colonic bacteria, and is utilized for bacterial protein synthesis. Intraluminal ammonia is also enterohepatically recirculated, and is reconverted to urea by the liver. Interestingly, increasing dietary protein intake does not increase fecal nitrogen excretion; however, increase in dietary fiber is associated with increased luminal bacterial utilization of the nitrogen (ammonia) for protein synthesis, thereby increasing fecal nitrogen excretion.

Altering the intestinal nutrient exposure, bacterial load, and pH can disrupt this enterohepatic recycling of urea and trapping it in the intestinal lumen. This is a desirable strategy employed in liver failure to decrease blood and hepatic ammonia levels. Specifically, lactulose, a nonabsorbable sugar is enterally administered to lower intestinal pH and to drive bacterial fermentation and protein synthesis, thereby trapping the nitrogen in the fecal material.

Disorders of the colon

Diarrhea

Diarrhea is defined as either a decrease in stool consistency or an increase in stool frequency and volume. It results from a complex interplay between colonic epithelial cell function, luminal factors, intestinal motility, and other factors.

Intestinal motility also influences stool volume and consistency. The ENS, with some modulation by the autonomic nervous system, is the primary regulator of gastrointestinal motility. Neuropeptides, gastrointestinal hormones, and luminal stimuli, such as dietary factors and interactions with bacteria, influence colonic motility. Disruptions in these systems can and do influence stool consistency and frequency.

Mechanisms of diarrhea can also be viewed from the perspective of absorptive capacity of the small intestine and colon; diarrhea results when this threshold is exceeded.

From a pathophysiological perspective, four mechanisms of diarrhea are traditionally described as of one of the following four types:

- Osmotic
- Secretory
- Motility
- Inflammatory

A degree of overlap occurs between these different types of diarrhea.

Osmotic diarrhea occurs when there is a failure to absorb a solute (usually a carbohydrate) in the proximal small intestine, thus not rendering the fluid isotonic, as would regularly occur, but, rather, hypertonic. Although electrolytes may be reabsorbed, the carbohydrate is not; rather, a portion of it is metabolized by enteric flora to SCFAs, carbon dioxide, hydrogen, and methane. With sodium and other electrolytes absorbed readily by the colon, and resultant low sodium concentration in the lumen, compounded by the presence of nonabsorbed carbohydrate, the high osmotic gradient draws fluid into the lumen and results in diarrhea. This type of diarrhea is characterized by a significant osmotic gap that can be calculated; an additional clinically significant feature of this type of diarrhea is that it diminishes on cessation of enteral intake. Malabsorbed carbohydrate and its metabolites also effect a lowering of the pH of the stool. Lactose deficiency is a good example of osmotic diarrhea in both children and adults. Ingestion of nonabsorbable sugars, such as sorbitol, can also lead to osmotic diarrhea. Excessive intake of carbohydrate-rich and, in particular, simple-sugar-rich beverages can contribute to osmotic diarrhea in children, and can exacerbate pre-existing diarrheal disease. Osmotic diarrhea resulting from excessive carbohydrate and/or simple sugar intake usually improves to resolves with reduction to cessation of dietary intake of these particular nutrients.

Secretory diarrhea occurs when the net secretion of fluids and electrolytes from the colon exceeds their absorption. This type of diarrhea exists independent of eating, and is not influenced by fasting or bowel rest. The prototypical example of pure secretory diarrhea (i.e., in the absence of inflammation or blood present in the stool) is of congenital chloride transport defects, and of gastrointestinal hormonal disorders, such as in Zollinger–Ellison syndrome and in disorders of vasoactive intestinal peptide or in other neuroendocrine tumors (Fig. 7).

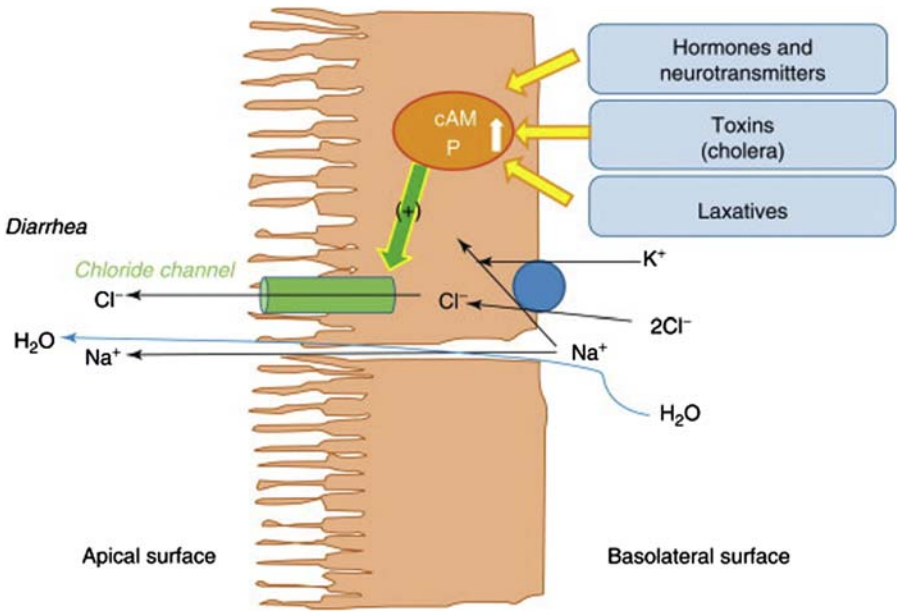


Fig. 7 Diarrhea and chloride transport across the small intestine epithelium.

Cholera occurs when the toxin interacts with the colonocyte stimulating chloride, potassium, and bicarbonate secretion, via toxin A stimulation of cyclic adenosine monophosphate (cAMP); some degree of inflammation may accompany this. ORS, which contributes fluid, sodium, and glucose relies on cellular mechanisms to effect rehydration, and is the mainstay of therapy.

Motility disorders influence intestinal function as pertain to absorption; whereas decreased transit enhances absorption of nutrients, significant decreases in motility can result in stasis. Deconjugation of bile acids by enteric flora can result in malabsorption and inflammation. Increases in motility can occur in the clinical picture of an inflamed colon, which can occur in infants and adults alike. Acute hormonal influences are more common in the adult population, such as those seen with thyrotoxicosis and carcinoid syndrome. Pharmacological agents or substance abuse can also influence motility.

Inflammatory diarrhea results in secretion of mucus, and, typically, with the presence of blood in the lumen, which in itself is a cathartic agent. The integrity of the epithelial barrier is often compromised, with resultant exudation of water and proteins. Bacterial invasion of the mucosa may occur, and is one example of inflammatory diarrhea. Additional disorders that can cause inflammatory diarrhea include allergic colitis and IBD.

Lastly, diarrhea can be categorized clinically into acute and chronic forms, with the latter being defined in persistence of symptoms for more than 3 weeks. Each type of diarrhea can be further clinically divided based on age with respect to likelihood of cause.

Protein losing enteropathy (PLE)

Albumin is the main protein maintaining oncotic pressure in the vascular compartment. Intracellular fluid status is maintained by an equilibrium of the oncotic pressure and hydrostatic pressure. Under normal circumstances, intact mucosal barrier and intercellular junctions function to prevent protein losses from the extracellular spaces. Protein losses can occur via the intestines, and can be categorized as conditions associated with mucosal disruption (functional or structural), or those with lymphatic obstruction plus increased hydrostatic pressure in the lymphatic vessels driving the process. The conditions associated with mucosal disruption can be further subdivided into inflammatory, infectious, and noninfectious processes. Another way to characterize these conditions are whether or not they are associated with normal vs. low serum albumin levels; specifically, low albumin states occur in conditions where protein synthesis cannot maintain equilibrium with turnover and needs. Stool alpha-1 antitrypsin status determination is the preferred method of diagnosis of PLE, as it is a protease inhibitor, which is not absorbed by the intestine, but is secreted into the gastrointestinal lumen, where it is not degraded by the intestinal microflora, and is readily measureable and quantifiable. Examples of conditions associated with PLE are presented in Table 5.

Infections and enteric parasites

In order for viral and bacterial agents to cause inflammatory disease involving the gastrointestinal tract, nonspecific host defense factors of gastric acidity, gastrointestinal motility, enteric flora, barrier functions of mucus secretion and mucosal integrity (in some cases), and of specific enteric mucosal immunity and systemic immune mechanisms have to be overcome.

Table 5 Classification of PLE.

- Inflammatory
  - Non-infectious
    - Eosinophilic gastroenteropathy
    - Cow's milk protein allergy/intolerance
    - Celiac disease
    - Graft vs. host disease
    - Crohn's disease
    - Intestinal polyposis syndromes
    - Necrotizing enterocolitis
    - Malnutrition
    - Anastomotic ulceration
  - Infectious
    - CMV gastritis (Menetriere's disease)
    - Rotavirus
    - *Helicobacter pylori*
    - Giardia lamblia
    - *Clostridium difficile*
    - *Salmonella*
    - *Strongyloides stercoralis*
  - Lymphatic obstructive
    - Lymphangiectasia
    - Cardiac disease/postsurgical (Fontan)

Table 6 Bacterial pathogens grouped by pathogenic mechanism.

Adherent	*Invasive*	*Toxigenic*	*Cytotoxic*
Enteropathogenic *E. coli*	*Shigella*	*Shigella*	*Shigella*
Enterohemorrhagic *E. coli*	*Salmonella*	Enterotoxigenic *E. coli*	Enteropathogenic *E. coli*
Enteroaggregative *E. coli*	*Yersinia enterocolitica*	*Yersinia enterocolitica*	Enterohemorrhagic *E. coli*
Diffuse-adherent *E. coli*	*Campylobacter jejuni*	*Aeromonas*	*Clostridium difficile*
	Vibrio parahemolyticus	*Vibrio cholerae*	

These infections can result in symptoms of vomiting, diarrhea, and abdominal pain, in addition to systemic effects such as fever. Clinical symptoms vary according to the pathogen.

Bacterial virulence is facilitated by enterotoxin secretion (which may be site specific in its action, secreted before introduction, or while within the lumen), adherence and invasion of the mucosa, and cytotoxin production, which function to disrupt mucosal and cellular function.

Bacteria can be classified based on their pathological mechanism (**Table 6**) as well as by their site of activity, and the nature of clinical signs and symptoms manifested. Signs and symptoms vary significantly by pathogen and age at presentation, with some forms presenting as crampy abdominal pain with watery diarrhea of relatively short duration, to frankly bloody diarrhea, and systemic signs and symptoms of inflammation with frank sepsis and shock possible. Common bacterial, viral, and parasitic infections involving the colon are outlined in **Tables 7** and **8**.

Polyps

Intestinal polyps are intraluminal protuberant tumors characterized by their gross morphological appearance, location(s), numbers, size, and presence (pedunculated) or absence (sessile) of the stalk (**Fig. 8**). Additional salient features include specific histological features used to discriminate between types and aid in predicting malignant potential. Extraintestinal manifestations are also associated with specific polyposis syndromes. Age at occurrence is important with respect to clinical significance and malignant potential; family history of polyps or of polyposis syndromes can also be predictive of disease evolution and aid in screening and surveillance of family members.

In children, juvenile polyps are the most frequently occurring kind, accounting for approximately 90% of colonic polyps. There are also many other types of polyps and polyposis syndromes, which are reviewed in **Tables 9** and **10**. The age of the subject, family history and inheritance patterns, number and location of polyps, and histology guide to the frequency of surveillance colonoscopy. Symptoms of rectal bleeding usually bring these patients to the attention of a physician. Polyps can cause clinically significant—yet often painless—bleeding so as to cause anemia, and can be linked to abdominal pain, rectal prolapse, or lead points associated with intussusceptions.

IBD

The phrase IBD encompasses UC and Crohn's disease. Indeterminate colitis is a diagnosis attributed to a condition in which a clear distinction cannot be made between the two aforementioned forms of IBD, as opposed to a heterogeneous group of diseases that present a wide clinical and histological spectrum.

Epidemiology

IBD presents in a bimodal fashion as pertains to age, first in late adolescence or early adulthood, and a smaller peak in the fifth decade of life. Overall, the sexes are equally affected for UC; in adults, the incidence of Crohn's disease is 20–30% higher in women.

The second half of the 20th century has seen the incidence of UC remain stable over time; Crohn's disease has demonstrated a marked increase across all age groups since 1950. Although IBD can affect all races, Caucasians are affected markedly more than Africans, or people of African origin. Ashkenazi Jews have a markedly increased risk of IBD compared with other Jewish groups. The incidence in the Ashkenazi Jewish population roughly parallels that of the respective geographical community in which they reside, albeit at a level which can be three-to four-times that of that general population, suggesting a genetic predisposition. The majority of individuals affected by both disorders are in North America and Northern Europe. The remainder of Europe, Latin America, and Australia has lower incidence rates, with rare cases occurring in Africa and Asia.

Etiology

The exact etiology of IBD is unclear and is an area of active research. A multifactorial interaction between genetic predisposition, environmental stimuli, endogenous triggers, immunological dysregulation, and modifying factors is postulated, and is discussed below.

Table 7 Bacterial enteric infections.

Name	*Epidemiology and pathogenesis*	*Clinical features*	*Diagnosis and treatment*
Shigella *S. dysentriae* *S. sonnei* *S. flexneri* *S. boydii*	• Acute infection • Highly contagious; low infective dose (10–100 organisms)	• Bacterial dysentery - Crampy abdominal pain and watery stools; - Progressive to bloody, mucoid, pus- laden stools - Tenesmus - Fever - Meningismus - Febrile seizures in younger patients • Hemolytic Uremic Syndrome (HUS possible)	Diagnosis • Crypt abscesses • Lymphatic hypertrophy • Necrosis • Elevated WBC count • Stool culture Treatment • Fluid and electrolyte replacement • Hand washing to prevent transmission • Limited role of antibiotics
Salmonella *S. typhi* *S. paratyphi* *S. enteritidis*	• Infective load: 10^3–10^5 organisms • Reservoirs: poultry and eggs, lizards, amphibians • Raw/undercooked foods • Mucosal invasion (jejunum and colon) • Inflammatory response with active secretion	• Five clinical presentations - Acute gastroenteritis (12–72 h incubation) - Focal, nonintestinal infection - Bacteremia - Asymptomatic carrier state - Enteric fever, abdominal cramping, nausea, vomiting, bloody, mucoid stools, rose spots on the trunk, leukopenia, prolonged excretion possible, variable by age, carrier state not uncommon	Diagnosis • Stool culture Treatment • Supportive • Very limited role for antibiotics
Campylobacter	• Transmission by contaminated foods (poultry, eggs, milk; water; and domestic animals) • Initial site(s): jejunum • Colon	• Incubation period of 2–11 days • Fever prodrome • Severe diarrhea, tenesmus, and abdominal pain	Diagnosis • Incubation for culture Treatment • Supportive • Role for antibiotics for limiting excretion period and duration of illness
Clostridium *C. difficile*	• Enteric flora • Toxin A (enterotoxin: alters permeability; inflammation mediator)	• Antibiotic-associated diarrhea	Diagnosis • Stool *C. difficile* toxins A and B ELISA • Endoscopy and histology: • Pseudomembranes (mucin, fibrin, polymorphonuclear lymphocytes, necrotic debris) • Erythema • Edema • Friability • Apthous ulcers
	• Toxin B (cytotoxin) • Pseudomembranes	May be restricted to the right colon	Treatment • Supportive therapy • Cessation of offending antibiotic • Metronidazole as first-line agent • Vancomycin secondary agent • Probiotics useful in relapse prevention
Yersinia *Y. enterocolitica*	• Transmission - Contaminated pork, - Enterotoxin elaboration		Diagnosis • Cultures are not very accurate • Endoscopy: mucosal ulcerations, friability throughout colon and terminal ileum possible • Histology: lamina propria infiltration B inflammatory cells; ulcerative and necrotic areas • Dilated crypts Treatment • Antibiotic

Table 7 Bacterial enteric infections.—cont'd

Name	*Epidemiology and pathogenesis*	*Clinical features*	*Diagnosis and treatment*
Aeromonas Hydrophila	Water contaminants	Three symptoms • Mild watery diarrhea • Bloody diarrhea • Persistent diarrhea	Diagnosis • Stool culture Treatment • Antibiotics
E. coli Enteropathogenic *E. coli* (EPEC)	• Localized adherence to enterocytes • Signal transduction • Intimate adherence and effacement	• Diarrhea • Vomiting • Malaise • Fever • Mucoid, nonbloody stools • Two-week duration	Diagnosis • Presence of adherent organisms on small intestinal/rectal biopsy Treatment • Antibiotics
Enterotoxigenic *E.coli* (ETEC)	Enterotoxin elaboration • Heat-labile (LT) toxin • Heat-stabile (ST) toxin • Fimbriae-based attachment • Stimulate adenylate cyclase • (LT) and guanylate cyclase • (ST) to secrete fluid	• Nausea • Abdominal pain • Watery diarrhea • Traveler's diarrhea	Diagnosis • Bioassays • Immunoassays • Gene probes for ST or LT Treatment • Supportive • Antibiotics decrease duration of excretion; not recommended for children
Enteroinvasive *E. coli* (EIEC)	• Colonize colon • Invade tissue • Replicate within cells • Secretory enterotoxins	• *Shigella*-like - Watery diarrhea - Bloody mucoid, pus-laden diarrhea - Tenesmus and fever	Diagnosis • Bioassays • Serotyping • ELISA Treatment • Supportive • Limited antibiotic role
Enterohemorrhagic *E. coli* (EHEC)	• Part of normal enteric flora in healthy animals • Cytotoxin similar to shiga toxin • Adherence • Transmission - Contaminated, undercooked meat - Unpasteurized apple cider - Children and the elderly more prone to HUS	• Hemorrhagic colitis - Crampy abdominal pain - Watery diarrhea progressing to bloody stools - Absence of fever • Hemolytic Uremic Syndrome (HUS)	Diagnosis • Serotyping • Serum antibody tests • Cytotoxin bioassays • DNA hybridization • PCR-based tests • ELISA Treatment: • No effective therapy • Supportive care - Dehydration correction - Management of electrolyte abnormalities - Blood transfusions as necessary
Enteroaggregative *E. coli* (EAEC)	• Localized adherence likely (HEp-2 or HeLa cells) • Enterotoxin • Increased intestinal mucus secretion	• Diarrhea - Watery - Mucoid - Persistent	Diagnosis • DNA probes
Diffuse adherent *E. coli*	• Diffuse adherence (HEp-2 or HeLa cells) likely	• Diarrhea	Diagnosis • DNA probes

Genetics

A positive family history in a first-degree relative confers significant risk, between 10% and 20% risk of disease occurrence of either disease type in a first-degree relative. Northern European, North American populations, and in particular globally the Ashkenazi Jewish population have the highest risk of the disease.

A high rate of concordance among Swedish monozygotic twins vs. dizygotic twins has been reported in Crohn's disease (44% vs. 3.8%). In the same study, the incidence rates observed in monozygotic twins for UC were (6.3%). These data, although supportive of a genetic role, show less than 100% penetrance, suggesting that whereas genetics are more important in Crohn's disease than in UC, environmental influences play a significant role. Simple Mendelian models of inheritance are inadequate to address the complex inheritance patterns of IBD. Candidate gene studies have suggested modest HLA associations, which differ in different populations. Systemic genome searches more recently employed in families with several members have IBD employing linkage analyses. Evidence of the *NOD2* gene on chromosome 16 being involved with cases of Crohn's disease has led to it being labeled as the IBD1 gene locus. This gene is; involved with the encoding of a protein associated with monocytic nuclear factor-κB; this protein and pathway are involved with the interaction of monocytes with bacterial peptidoglycans. Note that only approximately 30% of individuals with Crohn's disease test positive for this particular gene mutation.

Table 8 Additional colonic pathogens.

Name	*Pathogenesis*	*Clinical symptoms*	*Diagnosis and treatment*
Amoeba *Entamoeba histolytica*	• Travel to endemic areas a risk factor. Large intestinal commensal organism • Transmission: - Person to person contact - Contaminated food/water (cysts) - Cysts transform into trophozoites at the terminal ileum - Invade mucosa and submucosa	• Acute onset • Fulminant colitis - Bloody, mucoid diarrhea - Abdominal distention - Abdominal pain • Perforation possible • Hepatic abscesses possible	Diagnosis • Histopathology - Hyperemia and edema - Acute inflammation - Microulceration - Flask ulcer formation • (Fresh) stool examination for cysts or trophozoites Treatment • Lodoquinol • Metronidazole
Helminths *Trichuris trichura* (whipworm)	• Primarily colonic	• Heavy infestations associated with (bloody) diarrhea • Rectal prolapsed	Diagnosis • Stool assays Treatment • Thiabendazole • Mebendazole
Schistosomiasis *S. mansoni*	• Snail as pathogen • Fresh water contamination	• Dysenteric-like illness • Bloody diarrhea • Perianal fistulae	Diagnosis • Endoscopic - Focal and diffuse fibrosis - Intraluminal - Granulomatous masses (bilharziomas) • Stool exam for viable eggs Treatment • Praziquantel

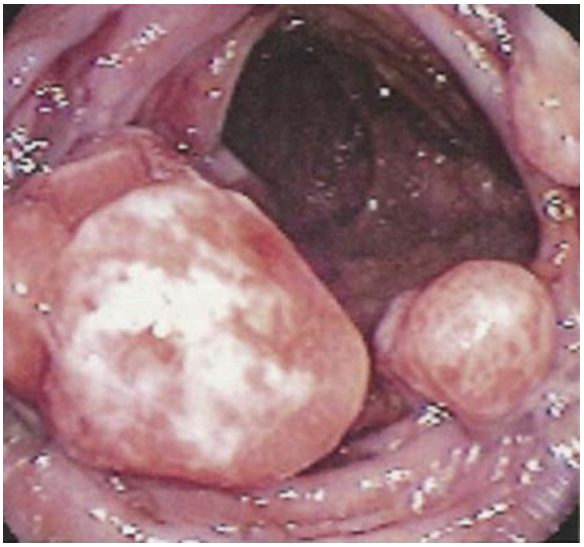


Fig. 8 Endoscopic view of a colonic polyp.

Environmental influences

The rapid increase of Crohn's disease over the past 50 years, increasing trends in immigrant populations, as well as incomplete genotype–phenotype associations have promoted attention to environmental factors. In particular, the search to identify an antigenic trigger to the enteric immune system has been pursued by several investigators. Postulated microbial intraluminal triggers have included mycobacterium and viruses. Dietary antigens or toxins have not been identified; diet westernization has been explored and remains an active area of research. Environmental exposures early in the life cycle such as birth environment, nutritive factors (breast vs. formula fed, with the former thought to have a protective effect) may alter the risk for developing disease in susceptible individuals/groups. Additional modulating factors include smoking and the use of oral contraceptives.

The incidence and prevalence of IBD are more in the developed world, and the global distribution is inverse to that of where geohelminthic worm infections are endemic.

Table 9 Hamartoma to us intestinal polyps.

Syndrome	*Location of polyps*	*Pathology*	*Extraintestinal abnormalities*	*Cancer risk*
Juvenile polyposis	Colon; some small intestinal involvement	• Up to 3 cm in size • Mucus retention and inflammatory cells in the lamina propria cysts • Mostly pedunculated		Colonic: low risk
Peutz-Jeghers	Mostly small intestinal; some gastric and colonic involvement	• 1–3 cm in size • Either sessile or pedunculated • Glandular epithelium and smooth muscle branching	Macular pigmentation on hands, lips, and mouth	Up to 18x vs. the general population; lower than other polyposis syndromes
Cowden's syndrome	Colon and stomach	• Multiple polyps • Hamartomatous	• Lipomas • Papillomas • Orocutaneous hamartomas	• Fibrocystic or fibroadenomatous, ductal breast cancer • Nodular thyroid hyperplasia or follicular adenoma

Table 10 Polyposis syndromes.

Type/syndrome	*Location(s)*	*Histology*	*Clinical features*	*Cancer risk*
Familial polyposis coli	Colonic: fundic gland hyperplasia (stomach)	• Thousands of adenomas • Elevated ornithine decarboxylase levels • APC gene	• Apparent after puberty • Diarrhea as most common symptom • Abdominal pain • Hypertrophic retinal lesions	• Thyroid cancer • Pancreatic cancer Risk of colon approximately reaches 100% by 55 years of age
Gardner's syndrome	Colon, stomach, duodenum, small intestine	• 2–5 mm sessile • Adenomas in the antrum and periampullary regions • More than 1000 at a time	• Triad of: - Polyps - Osteomas - Soft tissue tumors - Dental abnormalities	• Duodenal tumors at highest risk • Associated risk of: - Pancreatic carcinoma - Ampullary cancer - Hepatoblastoma
Turcot's syndrome	Colonic	Adenomatous polyps	• Present in adolescents with cancer; family history • Autosomal recessive	• Associated neural tumors - Medulloblastoma - Gliomas
Cronkhite-Canada syndrome	• Throughout gastrointestinal tract	• Adenomatous lesions within adenomatous polyps	• Alopecia • Nail dystrophy • Brown macular skin lesions • Edema related to protein losing enteropathy	• Five percent of cases evolve into gastrointestinal carcinomas
Inflammatory polyposis	Colonic: pseudopolyps	• Pleomorphic regenerative tissue	• Systemic signs and symptoms of inflammation (Section IBD)	• Colonic; risk of cancer from inflammatory bowel disease

Furthermore, there are nutritional factors also implicated in the increasing incidence and prevalence of IBD in the developing world. Transition from traditional to the modern western diet has been identified as a risk factor for many noncommunicable diseases, including IBD. In particular, decreasing fruit, vegetable, and fiber intake, increasing terrestrial animal protein intake, and decreased marine and plant-based protein intake, and changing profiles of types and amount of dietary fat intake (i.e., less ω-3 polyunsaturated fatty acids, higher intakes of ω-6 polyunsaturated fatty acids) are among implicated risk factors for the development of IBD in susceptible individuals.

Pathogenesis

The interactions between the enteric immune system and the intestinal lumen are dynamic; some degree of inflammation in response is ever present in the normal mucosal lamina propria of the colon and small intestine, which see a very large antigenic load daily. An intact mucosal barrier, in addition to normally functioning immunoregulatory mechanisms prevent this interaction progressing to the level at which tissue injury occur.

Current chronic, inflammatory relapsing disease processes may represent either an inappropriate persistent immune response to a luminal antigen/stimulus, vs. an appropriate immune response to a persistent, abnormal stimulus, vs. perhaps a prolonged immune response to a ubiquitous stimulus.

Enteric flora may play a role in this process, although no evidence to date points strongly to a single pathogen. Defective mucosal barrier function and increased intestinal permeability—the latter being documented in patients with IBD and in up to 10% non-affected first-degree relatives—may also be involved.

The immune response is primarily T-cell mediated, of a Th-1 nature. Abnormalities of interleukin-12 (IL-12), interferon gamma (INF-γ), and tumor necrosis factor alpha (TNF-α) have also been implicated in the immunological pathogenesis of this disease. White blood cells respond to these inflammatory mediators and proliferate the immune response. These recruited cells synthesize agents such as arachidonic acid metabolites, platelet activating factor, proteases, free radicals such as reactive oxygen species—all of which can and do cause direct injury to cells and the mucosa.

Pathology

Pathology differs between these two disorders, both in terms of anatomical distribution and tissue involvement.

Ulcerative colitis

UC is limited to the colon and rectum, usually beginning distally in the rectum and extending to varying lengths proximally, by definition, in a continuous fashion (**Fig. 9**). Usually, a clear distinction can be made where disease ends and normal mucosa can be appreciated grossly, or endoscopically. The gross appearance of the mucosa is dependent on the severity of the disease process. Mild disease presents with a diffuse erythema and loss of the characteristic appearance of the vasculature. Numerous small, superficial ulcerations, exudates, and bleeding are seen in moderate disease; larger, deeper ulcerations increased exudates, and the development of pseudopolyps, loss of normal gross architectural landmarks such as the folds diminish. Microscopically, UC is limited to the mucosa; with more severe disease, deeper layers may show a degree of involvement, with inflammatory cell infiltrates, shortening, branching, and decreases in the number of crypts as well as crypt abscesses can also be seen (**Fig. 10**).

Crohn's disease

Crohn's disease can involve any part of the alimentary tract from the mouth to the anus, and frequently does so in a discontinuous fashion, leaving "skip areas"—regions which are grossly and histologically normal (**Fig. 11**); in the colon, this lends a cobblestoned appearance. Macroscopically, wall thickening is evident in long-standing disease. This disease, by definition, is a transmural process (**Fig. 12**). With chronic disease, fibrostenosis occurs, narrowing the intestinal lumen. Stricturizing disease may follow fibrosis of superficial and deeper layers of the intestinal wall, which are evident on radiographic studies (**Fig. 13**).

The mesentery may also demonstrate inflammation, with adhesion and fixation of the colon a consequence of the transmural inflammatory process. Adjacent loops of bowel may become matted together. As luminal diameter narrows, intraluminal pressure may increase; in the face of unabating inflammation, this transmural process may lead to fistula formation. Enteroenteric fistulas are limited to the bowel; enterovaginal, enterovesicular, and enterocutaneous fistulization may occur. Inflammatory intra-abdominal masses called phlegmons may also form by this fistulization process.



Fig. 9 Distribution of ulcerative colitis.

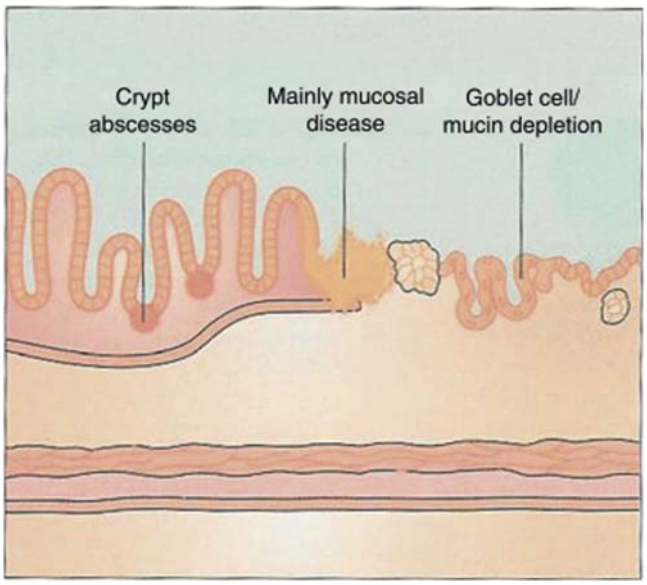


Fig. 10 The mucosa in ulcerative colitis. Reproduced from Kelly, D.A., Booth, I.W., 1996. Pediatric Gastroenterology and Hepatology. London, Mosby-Wolfe.

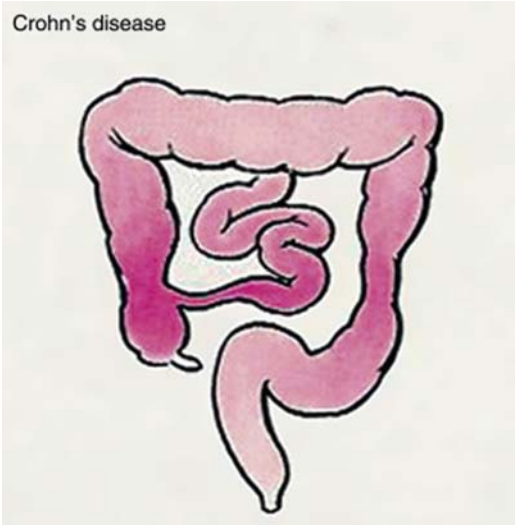


Fig. 11 Distribution of Crohn's disease.

The endoscopic appearance of Crohn's disease varies both by location and by time relative to the disease evolution. Intestinal Crohn's disease may initially present with aphthous ulceration overlying Peyer's patches in the colon. Ulcerations eventually grow in diameter and progress in depth, with frankly friable, exudative lesions.

Histological findings of affected areas include intense inflammatory cell infiltrates extending into the crypts, with shortening and forking of these structures, and associated abscesses. The inflammation is transmural; fibrosis and histiocyte proliferation are also seen. Noncaseating granulomatous submucosal and mucosal lesions, which are a hallmark of this disease, are not found in a majority of biopsy specimens. Granulomas can also be seen in intestinal infections such as in intestinal tuberculosis and sarcoidosis.

Even macroscopically normal-appearing tissue may yield histological findings of inflammation compatible with Crohn's disease, thus indicating that examination of the entire alimentary canal with surveillance biopsies is required before arriving at a diagnosis.

Inflammatory bowel disease and the terminal ileum

The terminal ileum is one of the most commonly involved sites in the intestine for Crohn's disease, originating often at the lymphoid follicle; strictures may form. A phenomenon of ileal involvement has been postulated in cases of apparent UC that

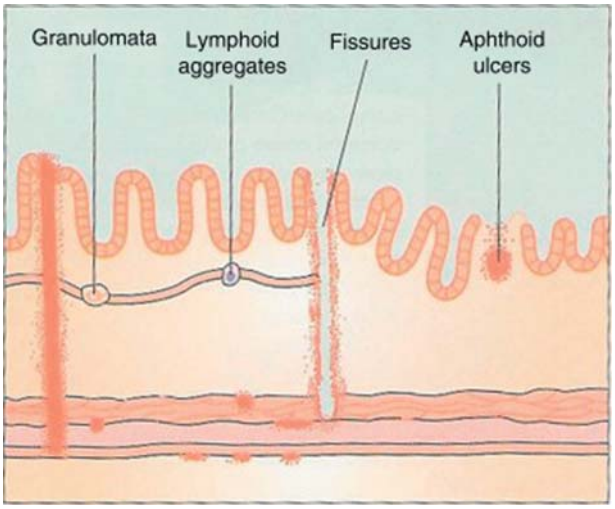


Fig. 12 The mucosa in Crohn's disease. Reproduced from Kelly, D.A., Booth, I.W., 1996. Pediatric Gastroenterology and Hepatology. London, Mosby-Wolfe.



Fig. 13 Small bowel stricturing noted on small bowel follow through series in a child with Crohn's disease. Reproduced from Kelly, D.A., Booth, I.W., 1996. Pediatric Gastroenterology and Hepatology. London, Mosby-Wolfe.

involve the ileum, in which cecal inflammation is postulated to "backwash" into the ileum; this finding being consistent with UC is controversial.

Extraintestinal manifestations

Extraintestinal manifestations are common in both Crohn's disease and UC, including ophthalmologic (uveitis), joint involvement (arthralgias and arthritis of the large joints), the skin, hepatobiliary system, pancreas, renal system, and vascular system, most commonly. Anemia and weight loss are common at the time of presentation. Growth and pubertal delay are very common at the time of presentation in children; short stature occurs in up to 50% of children. Some of these findings relate to the inflammatory process itself; others are linked to malnutrition associated with IBD.

Perianal disease with fistulization and skin tags are perhaps the most common extraintestinal abnormality associated with Crohn's disease.

Nutritional consequences of inflammatory bowel disease

Malnutrition includes acute weight loss, partly attributable to anorexia associated with inflammation, and partly to the disease process itself, i.e., inadequate intake as well as of excessive (malabsorptive) losses. An example delineating all of these mechanisms is of anemia, which can result from frank blood loss from associated gastrointestinal bleeding, anemia of chronic disease mediated by the inflammatory mediators, anorexia with decreased dietary iron intake, and, as in the case of duodenal and jejunal disease activity (as can occur in Crohn's disease), with decreased absorption.

Intestinal disease can result in both decreased nutrient absorption as well as disruption of the mucosal barrier resulting in exudation of proteins, a process known as PLE. The latter can result in hypoalbuminemia; third spacing of fluids as a result of decreased intravascular oncotic pressure can occur. Increased energy expenditure as a consequence of inflammation is noted, particularly in the febrile state, or with sepsis. Inflammation and discomfort also contribute to decreased enteral intake—factors contributory to a catabolic state.

In addition to iron, other mineral and trace element deficiencies are noted in IBD. Iron deficiency has been discussed above. Zinc is intimately associated with gut mucosa, and is susceptible to deficiency; low albumin levels resulting from PLE and increased intestinal epithelial cell turnover probably represent a significant source of zinc depletion. Vitamin B_{12} and folic acid deficiencies have also been documented among the water-soluble vitamins, particularly when the terminal ileal disease is noted. Vitamin D deficiency is the most common among the fat-soluble vitamins.

Treatment of IBD

Medical approaches: Several *anti*-inflammatory treatment modalities have been employed in the treatment of IBD. Their use is dictated by disease type, location, extent, and severity. Steroids provide the cornerstone of initial therapy for acute inflammation. Five aminosalicylate derivatives, use of antimetabolites such as azathioprine and 6-mercaptopurine methotrexate, and newer biological agents including *anti*-TNF-α are currently employed.

Nutritional therapies including exclusive to near exclusive enteral nutrition have a role in inducing and maintaining remission of Crohn's disease involving the small bowel in particular, both when used as monotherapy as well as an adjunct therapy. Although enteral therapy is not considered a first-line therapy for UC at the present in the USA, its use as such is popular in Europe and Canada, and allows for steroid sparing/avoidance. The time to onset of remission using enteral therapy in Crohn's disease is much less with steroids than with enteral therapy, however, with the former occurring typically within 2 weeks, the latter taking usually 6–8 weeks to achieve similar clinical remission. Smaller studies have been conducted employing low-fat diets, lower processed sugar foods, and employing polyunsaturated fatty acids, such as in fish oils, which may play a role in the treatment of disease. The data regarding use of ω-3 fish oil supplements for the treatment of IBD currently do not support routine use as a main-line therapy, but may have an adjuvant role in maintaining disease remission.

Surgical treatment is indicated in UC when acute, fulminant disease does not respond to medical therapy, or persistent chronic disease which is refractory to medical (steroid) therapy and when the diagnosis has been confirmed, i.e., that Crohn's disease has been ruled out. Colectomy is curative in such instances.

Crohn's disease is more complex, and surgical intervention, limited to involved segments only, is not curative. Failure of medical therapy to reduce inflammation, critical stenosis of the involved segments with fibrosis leading to obstruction, perforation, fistulization, and abscess formation not amenable to medical therapy and frank gastrointestinal hemorrhage are indications for surgical intervention. Reactivation of disease can occur postoperatively, at the site of anastomoses or elsewhere.

The natural history of IBD is such that long-standing disease increases the risk of colonic dysplasia, particularly in the case of UC, besides being a curative intervention, making the case for colectomy more attractive in older patients. Ileoanal continuity can be achieved by means of surgical anastomoses. Pouchitis secondary to bacterial overgrowth, smoldering pockets of disease activity that may not have been resected or have become evident after resection, and loss of continence are common complications of these procedures.

Conclusion

The term "microbiome" refers to organisms and their genomes/functions, whereas "microbiota" refers to the organisms themselves. The difference is very subtle, and sometimes they are used interchangeably. The metabolome specifically refers to the collection of metabolites that organisms produce.

Dietary intake influences the intestinal luminal environment, and there is dynamic interplay between this intestinal substrate and the microbiota. These differences occur in the neonatal period, with differences in the metabolome developing between breastfed and formula-fed infants (Saavedra and Dattilo, 2012). Breastfed infants support a microbiome rich in Bifidobacterium, which has *anti*-inflammatory properties thought to against Crohn's disease in particular (Joossens et al., 2011; Imaoka et al., 2008; LoCascio et al., 2010). Diets high in animal fat and low in fiber have increased concentrations of Bacteroides species, whereas those diets rich in carbohydrates and low in fat have predominance of Prevotella species. This is suggestive of a bidirectional relationship between the microbiota and nutrition, further influenced by host genetic susceptibility from an immune regulation and/or intestinal barrier function perspective (Sartor, 2012; Goldsmith and Sartor, 2014; Ray and Dittel, 2015). Short-chain fatty acids, including butyric acid, are the preferred energy source for colonocytes and are produced by the bacteria clostridia, which ferments dietary fiber. This is the preferred substrate for Firmicutes. A lack of dietary fiber results in less fermentation of this substrate by gut

bacteria to short-chain fatty acids. Furthermore, lack of fiber ingestion results in intestinal bacteria targeting an alternative substrate in the absence of fiber, namely the unstirred mucus layer. The unstirred mucus layer serves a barrier between the lumen and enterocytes. When the unstirred mucus layer is targeted and permeated, this contributes to inflammation of the mucosa that is exposed underneath. Emulsifiers found in the diet also impact the intestinal unstirred mucus layer and permeability to *E. coli*, which may drive inflammation (Lock et al., 2018).

In patients with IBD, the composition of the gut microbiome is altered. There is decreased species diversity and richness in the IBD microbiota. There is a decrease in Firmicutes (including clostridia), *Faecalibacterium prausnitzii*, *Bilophila wadsworthia*, and *Bifidobacterium adolescentis* and an increase in Actinobacteria and enterobacteriae proteobacteria, including adherent invasive *Escherichia coli*. These microbiota profiles are different than unaffected individuals within the same geographic populations (Fujimoto et al., 2013; Tawfik et al., 2014) Furthermore, there are some differences in the microbiota in patients with UC vs. Crohn's (Schaffler et al., 2016). It remains unclear if these changes to the microbiota are causative or correlated with having IBD (Ni et al., 2017). Composition of dietary intake can influence risk for IBD (Amre et al., 2007). Additionally, the diet can be modified to effect/modulate intestinal microbiota in a manner that may reduce IBD related inflammation (Ananthakrishnan et al., 2013).

References

Amre, D.K., D'Souza, S., Morgan, K., et al., 2007. Imbalances in dietary consumption of fatty acids, vegetables, and fruits are associated with risk for Crohn's disease in children. Am. J. Gastroenterol. 102, 2016–e2025.

Ananthakrishnan, A.N., Khalili, H., Konijeti, G.G., et al., 2013. A prospective study of long-term intake of dietary fiber and risk of Crohn's disease and ulcerative colitis. Gastroenterology 145, 970–e977.

Fujimoto, T., Imaeda, H., Takahashi, K., et al., 2013. Decreased abundance of *Faecalibacterium prausnitzii* in the gut microbiota of Crohn's disease. J. Gastroenterol. Hepatol. 28, 613–e619.

Goldsmith, J.R., Sartor, R.B., 2014. The role of diet on intestinal microbiota metabolism: downstream impacts on host immune function and health, and therapeutic implications. J. Gastroenterol. 49, 785–e798.

Imaoka, A., Shima, T., Kato, K., et al., 2008. *Anti*-inflammatory activity of probiotic Bifidobacterium: enhancement of IL-10 production in peripheral blood mononuclear cells from ulcerative colitis patients and inhibition of IL-8 secretion in HT-29 cells. World J. Gastroenterol. 14, 2511–e2516.

Joossens, M., Huys, G., Cnockaert, M., et al., 2011. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. Gut 60, 631–e637.

LoCascio, R.G., Desai, P., Sela, D.A., Weimer, B., Mills, D.A., 2010. Broad conservation of milk utilization genes in *Bifidobacterium longum* subsp. infantis as revealed by comparative genomic hybridization. Appl. Environ. Microbiol. 76, e7373–e7381.

Lock, J.Y., Carlson, T.L., Wang, C.M., Chen, A., Carrier, R.L., 2018. Acute exposure to commonly ingested emulsifiers alters intestinal mucus structure and transport properties. Sci. Rep. 8, 10008.

Ni, J., Wu, G.D., Albenberg, L., Tomov, V.T., 2017. Gut microbiota and IBD: causation or correlation? Nat. Rev. Gastroenterol. Hepatol. 14, 573–e584.

Ray, A., Dittel, B.N., 2015. Interrelatedness between dysbiosis in the gut microbiota due to immunodeficiency and disease penetrance of colitis. Immunology 146, 359–e368.

Saavedra, J.M., Dattilo, A.M., 2012. Early development of intestinal microbiota: implications for future health. Gastroenterol. Clin. N. Am. 41, 717–e731.

Sartor, R.B., 2012. Gut microbiota: diet promotes dysbiosis and colitis in susceptible hosts. Nat. Rev. Gastroenterol. Hepatol. 9, 561–e562.

Schaffler, H., Kaschitzki, A., Alberts, C., et al., 2016. Alterations in the mucosa-associated bacterial composition in Crohn's disease: a pilot study. Int. J. Colorectal Dis. 31, 961–e971.

Tawfik, A., Flanagan, P.K., Campbell, B.J., 2014. Escherichia coli-host macrophage interactions in the pathogenesis of inflammatory bowel disease. World J. Gastroenterol. 20, 8751–e8763.

Further reading

Feldman, M., Sleisenger, M.H., Friedman, L.S. (Eds.), 2002. Sleisenger and Fordtran's Gastrointestinal and Liver Disease, seventh ed. WB Saunders Company, London.

Guyton, A.C., Hall, J.E. (Eds.), 2006. Textbook of Medical Physiology. Elsevier Saunders, London.

Kleinman, R.E., Goulet, O.J., Miele-Vergani, G., Sanderson, I.R., Sherman, P., Shneider, B.L. (Eds.), 2008. Walker's Pediatric Gastrointestinal Disease: Pathology, Diagnosis, Management, fourth ed. BC Decker, Toronto.

Shils, M.E., Shike, M., Ross, C.A., Caballero, B., Cousins, R.D. (Eds.), 2006. Modern Nutrition in Health and Disease, tenth ed. Lippincott, Williams and Wilkins, Philadelphia.

Cystic fibrosis

J. Dowsett and O. Tully, St Vincent's University Hospital, Dublin, Ireland

© 2012 Elsevier Ltd. All rights reserved.
This is a reproduction of J. Dowsett, O. Tully, Cystic Fibrosis, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2012, Pages 416–422, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00065-9.

Definition and Etiology 207
Prevalence 207
Prognosis 208
Clinical Features 208
Pathogenesis of Lung Disease 208
Gastrointestinal Complications 209
CF-Related Diabetes Mellitus 209
Liver Disease 209
Nutritional Management 210
Decreased Dietary Intake 210
Increased Energy Requirements 210
Increased Energy Losses 211
Dietary Management of CF 211
Dietary Supplements 211
Enteral Feeding 211
Specific Dietary Considerations 212
Liver Disease 212
CF-Related Diabetes Mellitus 212
Bone Disease in CF 212
Fertility Issues 212
Body Composition Studies in CF 212
Assessment of Nutritional Status 212
Vitamin Status in CF 213
Vitamin A 213
β-Carotene 213
Vitamin D 213
Vitamin E 214
Vitamin K 214
Water-Soluble Vitamins 214
Mineral Status in CF 214
The Oxidant/Antioxidant Imbalance in CF Patients 214
Conclusions 214
Further Reading 214

Definition and Etiology

Cystic fibrosis (CF) is a multisystem autosomal recessive disorder caused by the mutation of a single gene on the long arm of chromosome 7 that codes for the CF transmembrane regulator (CFTR). This protein regulates the passage of chloride through the membrane of secretory epithelia, the dysfunction of which results in an altered composition of epithelial secretions. Clinically, CF is characterized by chronic pulmonary infection with periods of acute exacerbation, pancreatic insufficiency, and excessive losses of sweat electrolytes. The latter forms the basis for the diagnostic test. The mutated gene was identified in 1989 and since then more than 800 CFTR mutations have been reported, the most common of these being ΔF508.

Prevalence

Approximately 5% of the Caucasian North European and North American populations are carriers of the gene defect causing CF, leading to an approximate incidence of 1 in 2500 live births. This inheritance is illustrated in **Figure 1**. The incidence of CF in non-Caucasians is much lower and estimated to be approximately 1 in 100 000 in Oriental populations.

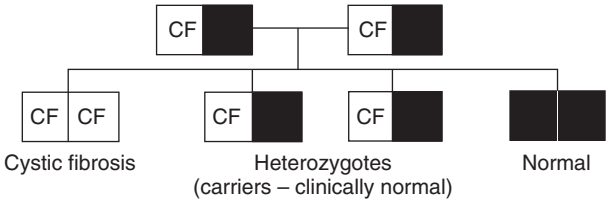


Figure 1 Mode of inheritance of CF: A Mendelian-inherited recessive characteristic. Reproduced from Figure 16.3 in Gibney MJ (ed.) (2005) Clinical Nutrition. *Nutriton Society Textbook Series*. Blackwell publishing. ISBN 0-632-05626-6, with permission from Wiley.

Prognosis

The median age of survival has dramatically risen from approximately 2 years in the 1940s to nearly 30 years in the 1990s. A current survival estimation following diagnosis is approximately 40 years. This improved prognosis can be attributed to a combination of factors, including aggressive management of infections, effective antibiotics, improved nutritional management, modern physiotherapy techniques, and the centralization of treatment in specialist centers. The survival age for females with CF would appear to be less than that for males. This may be related to poorer nutritional status among female CF patients. Expert management started immediately after an early diagnosis of CF by neonatal screening results in an important beneficial effect on outcome and may be critical to the clinical course of the condition and long-term prognosis. Although optimized nutrition, antibiotics, and chest physiotherapy remain the mainstay of CF management, new approaches to treatment are being developed that may add to the traditional medical therapy for CF. As prognosis and survival improves, nutrition-related issues become more prevalent, including the effective management of pregnancy, diabetes, osteoporosis, and transplantation.

Clinical Features

The clinical features of CF are listed in Table 1.

Pathogenesis of Lung Disease

Pulmonary disease can be demonstrated within the first few months of life. Bacterial infection is characterized by high levels of neutrophils and mediators of infection in the form of interleukin 1, interleukin 8, and elastases. Mucous glands become dilated

Table 1 Clinical features of respiratory features of CF

Respiratory features of CF	
Atelectasis	Incomplete expansion of a lung or part of a lung due to airlessness or collapse
Bronchiectasis	Chronic dilatation of the bronchi associated with coughing and expectoration of purulent mucus
Bronchitis	Inflammation of one or more bronchi
Pneumonia	Inflammation of the lungs with air spaces becoming filled with exudates
Pneumothorax	Accumulation of air in the pleural space
Gastrointestinal features of CF	
Cholelithiasis	The presence or formation of gallstones
Cirrhosis	Liver disease characterized by loss of normal liver tissue and fibrosis
Distal intestinal obstruction syndrome	Blockage of the bowel with feces, mucus, and undigested food
Gastroparesis	Paralysis of the stomach or delayed gastric emptying
Malabsorption	Impaired intestinal absorption of nutrients
Maldigestion	Impaired intestinal digestion of nutrients
Meconium ileus	Blockage of the bowel with meconium
Osteoporosis/osteopenia	Reduction in bone mass
Pancreatic insufficiency	Reduction of enzyme production from the pancreas
Portal hypertension	High pressure in the portahepatic artery
Rectal prolapse	Protrusion of the rectal mucous membrane through the anus
Splenomegaly	Enlargement of the spleen

leading to obstruction, secondary infection, and progressive lung damage. Frequent periods of respiratory infection and exacerbation are common in CF patients with increased cough, increased sputum production, and shortness of breath. The immune response appears to be of great significance. Chronic inflammation has been cited as the cause of considerable lung damage seen in CF. Steroidal anti-inflammatory drugs have been shown to be beneficial but have nutritional side effects, such as hyperglycemia and osteoporosis. Nonsteroidal anti-inflammatory drugs, such as ibuprofen, have been used in some centers with positive results, but their long-term effect on renal function is not yet known. The impact of malnutrition on lung disease and respiratory muscle function has been extensively studied in patients with CF. Malnutrition and deterioration of lung function are interdependent. Prevention of malnutrition from the time of diagnosis is associated with better lung function and improved survival.

Gastrointestinal Complications

Individuals with CF can develop a variety of gastrointestinal (GI) disorders related to the pathophysiological changes associated with CF. Pancreatic insufficiency, which is present in most CF patients, leads to many of the GI manifestations of CF, including steatorrhea, abdominal pain, distal intestinal obstruction syndrome (DIOS), and rectal prolapse. Gastroesophageal reflux occurs frequently in CF patients due to decreased lower esophageal sphincter pressure and is usually treated by proton pump inhibitors. In patients with advanced lung disease, vomiting is common after strenuous bouts of coughing and this over time may lead to decline in nutritional status. Peptic ulcer disease, pancreatitis, and intussusception also occur at varying degrees in patients with CF. Crohn's disease and celiac disease occur more frequently in the CF population than in controls; and GI tumors, although rare, have an increased incidence in CF.

Meconium ileus is the presenting complaint in up to 15% of infants with CF. This is a condition in which the small intestine is blocked with tenacious meconium and surgical intervention is required to correct it. Excessive mucus in the small bowel of patients with CF can provide a physical barrier to the absorptive surface. Undigested or unabsorbed food in association with this mucus, and possibly a reduced gut motility, can lead to a partial or complete obstruction of the GI tract in older children and adults known as meconium ileus equivalent, or more accurately DIOS. This is a condition specific to CF. The usual clinical presentation is one of abdominal pain, abdominal distension, and constipation. It can be precipitated by dehydration, change in eating habits, change in enzyme brand or dose, or immobility. DIOS is treated with a laxative regime and should have a diet and enzyme review.

CF-Related Diabetes Mellitus

Diabetes requiring insulin is the most common comorbidity in CF. The islets of Langerhans are the last cells to be damaged in the process of fibrosis of the pancreas. The incidence of diabetes in CF has been reported to be 8–15%, but this may be underestimated due to lack of screening. It is estimated that 50% of patients older than 30 years will have some degree of glucose intolerance. The primary cause of CF-related diabetes mellitus (CFRD) is insulin deficiency secondary to pancreatic fibrosis. Diagnostic criteria for CFRD are the same as for non-CFRD. Glucose metabolism is also affected by many factors, including infection, malabsorption, abnormal intestinal transit time, and steroid use, all features of CF. Although CFRD shares many of the characteristics of both type 1 and type 2 diabetes, it is itself a distinct clinical condition. Hyperglycemia may adversely influence weight and pulmonary function; and as the age of survival increases, it may lead to the development of microvascular complications. Retrospective studies have shown that in those individuals presenting with overt diabetes mellitus, deterioration in weight and respiratory status for 2 years before diagnosis are reversed once insulin therapy is instituted. A program of multiple daily insulin injections and self-monitoring of blood glucose with the aim of normoglycemia is the preferred treatment with regular follow-up with the Endocrinology team. All patients with CF should be screened annually for CFRD using the oral glucose tolerance test. Minimal dietary restrictions are imposed on this group of patients in an attempt to maximize nutritional intake. See the Section on **Dietary Management of CF**.

Liver Disease

Another complication associated with increased longevity in CF is liver disease, which affects between 2% and 37% of adults with CF. The development of liver disease in patients with CF has been attributed to the blockage of small bile ductules with thick secretions and the subsequent development of progressive cholestasis, biliary fibrosis, and eventually biliary cirrhosis and portal hypertension. The persisting acidic conditions in the upper small bowel lead to bile salt precipitation and defective lipid emulsification. Unhydrolyzed fat and other products of maldigestion may interfere with bile acid reabsorption in the terminal ileum, thereby reducing the total bile salt pool. Fecal losses of primary and secondary bile acids lead to an imbalance of bile salts, which further increases the viscosity of the already tenacious bile. Treatment with ursodeoxycholic acid has led to an improvement in bile excretion and liver function tests. Complications of liver disease, including ascites, gastric, and esophageal varices may further exacerbate a patient's nutritional status. In a small number of patients, liver failure may require liver transplantation. See the Section on **Dietary Management of CF**.

Nutritional Management

Aggressive nutritional management of patients with CF is key in their overall management. Nutritional management of CF involves maximizing dietary intake; minimizing malabsorption and maldigestion; monitoring vitamin intakes and serum levels; and adapting eating patterns in the event of diabetes, osteoporosis, DIOS, or liver disease. Nutritional support in the form of nocturnal gastrostomy feeding may be necessary if nutritional failure persists (body mass index (BMI) lower than 18.5 kg m^{-2}). It is well recognized that the malnutrition seen in CF patients is due to an energy imbalance caused by three main factors: decreased dietary intake, increased energy requirements, and increased energy losses. There appears to be a direct association between the degree of malnutrition and the severity of pulmonary disease, affecting overall prognosis. Many patients are capable of balancing these factors effectively and have a normal growth velocity and good nutritional status. However, as lung function deteriorates, energy requirement increases and appetite decreases leading to a loss of energy stores and lean tissue further contributing to progressive deterioration of lung function (see Figure 2).

Decreased Dietary Intake

People with CF are advised to consume a diet high in energy with no fat restriction. Before the development of enteric-coated enzymes in the mid 1980s, patients with CF were advised to follow a low-fat diet in an attempt to minimize fat malabsorption and steatorrhea. Unfortunately, older patients continue this practice as they have developed an aversion to fatty foods after many years of avoiding them. Decreased dietary intake secondary to anorexia is common in CF patients and can become more of a problem during recurrent chest infections. There has also been an increased number of reports of eating disorders and abnormal eating behavior in the CF population. In addition, polypharmacy, repeated exacerbations of CF, organomegaly, GI problems, food intolerance, and poor social circumstances can reduce oral intake.

Increased Energy Requirements

Energy requirements are increased during periods of infection by catabolism and fever and continue to increase with advanced pulmonary disease. It has been estimated that CF patients require 120–150% of the estimated average requirement of energy. As pulmonary function deteriorates, mobility also decreases and, as a result, overall energy expenditure is reduced. Owing to the heterogenicity of CF, the energy requirements of individuals will vary and should be assessed on an individual basis. Energy losses through sputum may also be significant in a patient with a marginal energy intake. Salbutamol, often used as a bronchodilator in CF, can increase basal metabolic rate.

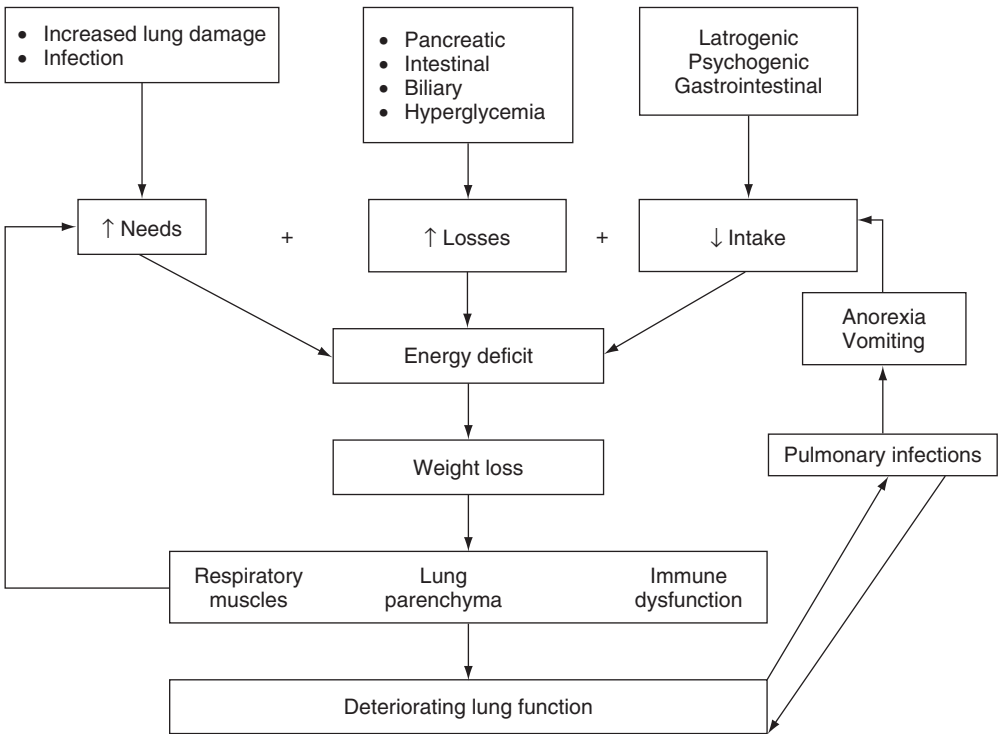


Figure 2 Interdependent factors that may give rise to progressive energy deficit as lung function deteriorates.

Increased Energy Losses

Pancreatic changes are caused by the obstruction of small ducts with thick secretions and cell debris. Functional tissue becomes replaced with fibrotic tissue leading to pancreatic exocrine insufficiency when more than 90% of the normal structure of the pancreas is lost. Pancreatic insufficiency is the most common GI manifestation in CF, occurring in at least 95% of patients. The production of pancreatic secretions, including enzymes and bicarbonate, is reduced, necessitating pancreatic enzyme replacement therapy (PERT). PERT is supplied in the form of gelatin capsules containing microspheres, which are swallowed whole with food. The capsule dissolves within the stomach and releases the microspheres, which are protected from the gastric acid by an enteric coating. Enzymes should be taken immediately before or during a meal to maximize their efficacy. The microspheres mix with the stomach contents and pass through the pylorus into the duodenum where they become activated. Microspheres should be less than 1.5 mm in diameter to ensure that they leave the stomach with food. Fibrosis of the pancreas tends to be a progressive process, so increasing amounts of oral enzyme supplements are often required as patients get older. All people with CF have some level of pancreatic dysfunction but requirements of enzymes are variable and must be assessed individually. Clinically, the aim of PERT is to correct symptomatic steatorrhea, relieve any abdominal pain, reduce the mass and frequency of stool passed, and achieve weight gain within normal limits.

The enteric coating on enzyme supplements is designed to dissolve at pH 6, the optimal pH for pancreatic enzymatic action. Owing to the reduced production of bicarbonate and the resulting lower pH of the duodenum in patients with CF, the enteric coating of the enzyme may fail to dissolve so that the enzyme does not become activated at the absorptive surface of the small bowel. Increasing the duodenal pH by taking proton pump inhibitors may improve absorption. Changing the brand of enzyme may also improve absorption as dissolution characteristics of the enteric coating and proportions of enzymes contained within the microspheres vary. Patients should be dissuaded from chewing enzymes as this breaks the enteric coating and leads to deactivation in the acid medium of the stomach. Even with maximal PERT, it has been estimated that between 10% and 20% of ingested fat will be malabsorbed. Colonic strictures known as fibrosing colonopathy (FC) in CF populations receiving high-potency enzymes with a more concentrated dose of lipase and protease per capsule have been reported. The etiology of this FC remains unclear. Recently, it has been suggested that FC may be related to the presence of methacrylic acid copolymer coating present in some preparations rather than actual enzyme strength. Some adult patients continue to take high-dose enzymes and are advised to do so within recommended levels. The working group on PERT use recommends that no more than 10 000 units of lipase per kilogram body weight should be taken per day.

Dietary Management of CF

Patients with CF are encouraged to consume a diet providing 150% of the recommended intake for age and sex. However, this is only a guideline, because in practice, the energy requirement for a patient with CF is that which maintains their ideal body weight when malabsorption has been controlled. Maximizing energy intake from everyday foods should be the initial step in the promotion of a high-energy diet. As fat is the most concentrated source of energy in the diet, liberal use of fat should be encouraged; this can best be achieved by recommending frequent consumption of high-fat meals and snacks, including confectionery, desserts, and cakes. PERT should be dosed accordingly.

Dietary Supplements

The energy intake of many patients with CF is commonly suboptimal. Many patients find it difficult to eat sufficient food daily to attain or maintain their ideal body weight. During a respiratory exacerbation of CF, energy requirements are at a maximum, but appetite is often reduced. Dietary supplements in the form of sip feeds can be a useful adjunct to a high-energy diet. Care should be taken to ensure that supplements are used in addition to a diet and not as a substitute for normal foods.

Enteral Feeding

When diet and oral dietary supplements are undesirable or ineffective and nutritional failure persists, i.e., BMI lower than 18.5 kg m^{-2}, enteral feeding should be considered. Research has demonstrated a sustained weight gain and a slowing decline in respiratory function associated with supplemental enteral feeding. Artificial nutritional support can be provided via nasogastric or gastrostomy tube depending on patient's preference. Gastrostomy feeding is becoming more popular, whether passed endoscopically or under fluoroscopic guidance. The introduction of low-profile gastrostomy feeding tubes or 'button' tubes have made this method of nutritional support more acceptable to patients. The type of feed used and the PERT given with it varies between centers. Feeds are usually administered overnight in an attempt to provide 30–50% of energy requirements and to allow for maximal oral intake during the day. Gastrostomy feeds can be used over longer periods during acute pulmonary infection, loss of appetite, or in a severely malnourished patient. Patients with a history of poor intake should be monitored for refeeding syndrome.

Specific Dietary Considerations

There are some medical complications of CF that warrant particular nutritional attention.

Liver Disease

Patients with liver disease as a complication of their CF may have ascites, gastric, or esophageal varices, all of which may affect nutritional status and options for nutritional support. Dietary management of the patient with CF and liver disease centers on maximizing energy intake and is best achieved by encouraging small, frequent, energy-dense meals, snacks, and drinks. Suboptimal oral intake can arise in patients with hepatomegaly or splenomegaly, who often have a feeling of fullness after eating, 'a condition' referred to as the 'small stomach syndrome.' The benefits of gastrostomy insertion should be carefully weighed in a patient with gastric varices or splenomegaly due to risk of bleeding. A moderate sodium restriction may alleviate ascites. If coagulation is impaired, supplementation with vitamin K may be indicated.

Treatment of liver disease in CF is with ursodeoxycholic acid, which has a positive effect on liver enzymes. Whether this improvement is associated with improvement in nutritional status is unknown.

CF-Related Diabetes Mellitus

The dietary treatment of CFRD varies from standard diabetic dietary advice. The principle of the diet centers on maintaining caloric intake while ensuring glycemic control. The treatment of CFRD should enhance rather than impair a patient's nutritional status. This is done by encouraging a high-fat diet and confining the intake of refined carbohydrate to mealtimes. Insulin doses should be increased so as to maximize the flexibility of the diet, particularly in those patients who are already nutritionally compromised. Patients taking oral nutritional supplements and/or overnight gastrostomy feeds need to have their insulin doses carefully monitored and adjusted accordingly.

Bone Disease in CF

Osteopenia and osteoporosis are now widely recognized in the CF population. There are a number of contributing factors to this early development of bone disease, including steroid usage; malabsorption of calcium and, more importantly, vitamin D; poor nutritional status; decreased levels of physical activity; and a reduced peak bone mass in CF patients compared with healthy individuals. Assessment of bone health is by dual-energy X-ray absorptiometry scanning and there are a variety of treatment options available depending on the severity of disease ranging from dietary calcium and vitamin D supplementation to the use of bisphosphonate drugs, which aim to halt the progression of bone loss and promote bone formation.

Fertility Issues

As the number of people with CF of a reproductive age increases, so does the incidence of pregnancy in this group. Although almost all males with CF are infertile owing to the absence of the vas deferens, most females are fertile. Pregnancy in women with CF requires special nutritional attention with regular monitoring, particularly with respect to adequate weight gain, and vitamin and mineral status.

Body Composition Studies in CF

Studies of body composition in CF patients have shown deficits in total body mass, lean body mass, and body fat, which affect body density. As skinfold thickness percentiles are derived from body density, it has been suggested that the assessment of the body fat content of children with CF using, or derived from, body density, such as skinfold thickness, is invalid. Muscle function indices have been shown to respond to refeeding in malnourished patients with CF before body composition or biochemical indices of protein status improved, and so appear to be sensitive markers of nutritional status.

Assessment of Nutritional Status

Malnutrition in CF remains a major clinical problem. Growth and nutritional status should be monitored at each clinic visit to ensure early detection of any deterioration and to prompt appropriate nutritional intervention. The many factors that complicate nutritional status in CF are shown in Table 2.

When weight falls to a BMI of less than 18.5 kg m^{-2}, nocturnal enteral feeding should be considered. At diagnosis and when the patient shows clinical deterioration, the following should be determined: electrolytes, serum albumin and other liver function tests, oral glucose tolerance test, full blood count, serum retinol, and α-tocopherol. If there is any evidence of iron deficiency, iron status

Table 2 Factors affecting nutritional status

- Variation in gene mutation
- Frequency of pulmonary exacerbations
- Gastroesophageal reflux
- Distal intestinal obstruction syndrome
- Pancreatitis
- Liver disease
- Diabetes mellitus
- Drug therapy
- Dietary dislikes and misconceptions
- Psychological problems/eating disorders
- Pregnancy
- Transplantation

should be assessed. Other medical disorders should be considered in the evaluation of nutritional failure. These include diabetes mellitus, liver disease, Crohn's disease, celiac disease, chronic abdominal pain, DIOS, and esophagitis.

Vitamin Status in CF

At least 85% of CF patients have some level of pancreatic insufficiency leading to a degree of fat malabsorption. For this reason, unless supplemented, most patients are at risk of developing either clinical or subclinical deficiencies of the fat-soluble vitamins, vitamin A, D, E, and K. Those most at risk appear to be individuals with poorly controlled malabsorption, poor adherence to treatment, liver disease, bowel resection, or following a late diagnosis.

Vitamin A

Vitamin A should be supplemented at a dose of 4000–10 000 IU day^{-1}. However, low-serum levels of retinol have been noted even at this dose. If retinol levels are persistently low despite adequate supplementation, an assessment of compliance, retinol-binding protein (RBP), and zinc levels should be checked. Special care should be given to vitamin A supplementation during pregnancy as high levels are reported to be teratogenic.

It is important to consider hepatotoxicity with large supplemental doses of vitamin A in a patient who may store vitamin A in the liver, yet shows low-serum levels of retinol, and who may display ocular signs of deficiency. The free alcohol retinol is almost entirely attached to RBP, which is synthesized in the liver. Decreased levels of RBP, which may occur in up to 25% of patients with CF, may be due to an abnormality in its production by the liver, zinc deficiency, or protein–energy malnutrition. Even with adequate vitamin supplementation and PERT, up to 20% of patients may have ocular signs of deficiency of retinol. Xerosis may improve by increasing the dose of vitamin A alone or combined with zinc. It has been suggested that there may exist a specific defect in the handling of retinol in the GI tract of people with CF unrelated to the level of fat malabsorption. A correlation has been demonstrated between low levels of vitamin A and poor lung function.

β-Carotene

β-carotene is one of the carotinoids present in plasma and a precursor of vitamin A. It is effective as an antioxidant at lower oxygen saturation states than vitamin E. It has a biological role as a lipid-soluble chain-breaking antioxidant in biomembranes. Routine supplementation with β-carotene could diminish lipid peroxidation and improve essential fatty acid status.

Vitamin D

Vitamin D deficiency may be caused by malabsorption, underexposure to sunlight, or defects in metabolism due to liver disease. Although skin exposure to sunlight is the major source of vitamin D, serum concentrations will vary between individuals depending on endogenous production in the skin. Rickets as a result of vitamin D deficiency is rare but has been described in CF patients. Osteopenia and retarded bone maturation have been reported in a number of CF patients, even with supplementation to recommended levels. Bone density has been shown to be significantly decreased in all sites compared with that of normal young adults. Other variables such as activity levels and nutritional status have not been adequately researched, although the incidence of osteoporosis was found to be higher in those patients with severe respiratory disease. To attain and maintain normal serum levels, a daily dose of 400–2000 IU is generally required in adults.

Vitamin E

Cholestasis and a reduced enterohepatic circulation of bile acids contribute to the malabsorption of fat-soluble vitamins from the small intestine. Vitamin E is highly lipophilic and its deficiency correlates with a degree of fat malabsorption. Subclinical neuroelectrophysiological abnormalities are already present in approximately 40% of patients by 2 months of age. Neurological signs of vitamin E deficiency are responsive to supplementation if initiated early, but are irreversible if treatment starts after the neurological lesions are present. As circulating α-tocopherol is transported in the blood attached to lipid, it is should be expressed as a ratio to total lipid to be correctly interpreted. Current recommendations are to monitor serum vitamin E levels annually and adjust supplementation accordingly. A daily dose of 400 IU day^{-1} should achieve normal serum levels in adults.

Vitamin K

A review of the literature providing conflicting opinions in the area of routine supplementation of vitamin K as the prevalence of vitamin K deficiency has not been established. Theoretically, the risk factors for patients developing vitamin K deficiency are pancreatic insufficiency, severe liver disease, extensive small bowel resection, and chronic broad-spectrum antibiotic use. Monitoring the coagulation system is advised, as vitamin K estimations are not generally routinely available. It seems prudent to prescribe vitamin K supplements to patients with the above-mentioned risk factors. Vitamin K has recently been shown to play an important role in bone health. There are no specific guidelines on supplementation, but doses of 5–10 mg appear to be a prudent guide. Annual monitoring of fat-soluble vitamin levels should be carried out and doses of vitamins altered as appropriate.

Water-Soluble Vitamins

Supplementation with water-soluble vitamins is, in general, thought to be unnecessary in CF. In cases where dietary intake is poor or unbalanced, supplementation of vitamin C is advised. Supplementation with other water-soluble vitamins is not routinely recommended.

Mineral Status in CF

Fat malabsorption can lead to the formation of insoluble fatty acid complexes with minerals in the gut, leading to a reduction in their absorption. CF may also be associated with intestinal mucosal defects, which may further retard the absorption of nutrients. Suboptimal levels of zinc, selenium, manganese, and iron have all been described in CF. Routine iron supplementation is not recommended as it has been suggested that *Pseudomonas aeruginosa* grows in tissues with a high concentration of iron. In addition, levels of iron may be suppressed as a normal body response in times of infection, and attempting to correct this is potentially harmful. Sodium and chloride do not need to be supplemented unless in very hot climates or during excessive exercise.

The Oxidant/Antioxidant Imbalance in CF Patients

Patients with CF frequently exhibit increased oxygen free radical generation from activated neutrophils due to chronic lung inflammation. This, coupled with antioxidant deficiencies due to exocrine pancreatic insufficiency, results in an oxidant/antioxidant imbalance. Consequently, free radical attack on unsaturated fatty acids of lipid structures occurs leading to lipid peroxidation. An efficient antioxidant supply is suggested to control tissue damage by restoring the oxidant/antioxidant balance.

Conclusions

There is a complex relationship between physiological, environmental, and genetic variables leading to a great variability in energy requirements among individuals with CF. Despite advances in the treatment of CF, the need for good nutritional strategies in CF will continue. Individually tailored nutritional advice for each patient with CF by a dietitian experienced in the area of CF is essential.

Further Reading

Amin, R., Ratjen, F., 2008. Cystic fibrosis: A review of pulmonary and nutritional therapies. Advances in Pediatrics 55, 99–121.
Chinuck, R.S., Fortnum, H., Baldwin, D.R., 2007. Appetite stimulants in cystic fibrosis: A systematic review. Journal of Human Nutrition and Dietetics 20 (6), 526–537.
Colombo, C., 2007. Liver disease in cystic fibrosis. Current Opinion in Pulmonary Medicine 13 (6), 529–536.
Erskine, J.M., Lingard, C., Sontag, M., 2007. Update on enteral nutrition support for cystic fibrosis. Nutrition in Clinical Practice 22 (2), 223–232.
Ferguson, J.H., Chang, A.B., 2009. Vitamin D supplementation for cystic fibrosis. Cochrane Database of Systematic Reviews. Issue 4. doi: 10.1002/14651858.CD007298.pub3.
Innis, S.M., Davidson, A.G., 2008. Cystic fibrosis and nutrition: Linking phospholipids and essential fatty acids with thiol metabolism. Annual Review of Nutrition 28, 55–72.
Laguna, T.A., Nathan, B.M., Moran, A., 2010. Managing diabetes in cystic fibrosis. Diabetes Obesity & Metabolism 12 (10), 858–864.

Messick, J., 2010. A 21st-century approach to cystic fibrosis: Optimizing outcomes across the disease spectrum. Journal of Pediatric Gastroenterology and Nutrition 51 (supplement 7), S1–S7.

Michel, S.H., Maqbool, A., Hanna, M.D., Mascarenhas, M., 2009. Nutrition management of pediatric patients who have cystic fibrosis. Pediatric Clinics of North America 56 (5), 1123–1141.

Robinson, K.A., Saldanha, I.J., McKoy, N.A., 2009. Management of infants with cystic fibrosis: A summary of the evidence for the cystic fibrosis foundation working group on care of infants with cystic fibrosis. Journal of Pediatrics 155 (supplement 6), S94–S105.

Rosenstein, B.J., 1998. The diagnosis of cystic fibrosis: A consensus statement. Journal of Pediatrics 132 (4), 589–595.

Shamseer, L., Adams, D., Brown, N., Johnson, J.A., Vohra, S., 2010. Antioxidant micronutrients for lung disease in cystic fibrosis. Cochrane Database of Systematic Reviews. Issue 12. doi: 10.1002/14651858.CD007020.pub2.

Stallings, V.A., Stark, L.J., Robinson, K.A., Feranchak, A.P., Quinton, H., 2008. Evidence-based practice recommendations for nutrition-related management of children and adults with cystic fibrosis and pancreatic insufficiency: Results of a systematic review. Journal of the American Dietetic Association 108 (5), 832–839.

Dehydration

Shashank S. Dhareshwar, Department of Nephrology, St John's Medical College and Hospital, Koramangala, Bengaluru, India

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A.W. Subudhi, E.W. Askew, M.J. Luetkemeier, Dehydration, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 1–9, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00068-4.

Introduction 216
Physiology of water balance 217
Distribution of water in the body spaces 217
Kidney and body fluid balance 217
Gastrointestinal tract and body fluid balance 217
Sweat and thermoregulation 218
Respiration related water loss 219
Pathophysiology of dehydration 219
Tonicity and osmolality 219
Blood volume 220
Dehydration and heat illness 220
Groups at risk for dehydration 221
Infants and children 221
Old age 221
People who work or exercise outside 221
Predisposing factors for dehydration and heat illness 221
Identifying types of dehydration 221
Clinical features and diagnosis 221
Management of dehydration 222
Goals of treatment 222
Treating different types of dehydration 224
Prevention of dehydration 224
Conclusion 225
References 226
Further reading 226

Key points

- Understand the physiology of body water balance and pathophysiology of dehydration.
- Recognize the different clinical and laboratory abnormalities in isotonic, hypotonic, and hypertonic dehydration.
- Know how to manage dehydration.

Glossary

Hyper/hypotonic Increased/decreased concentration of solutes (most typically sodium) in the blood
Hypo/hypernatremia Low/high plasma sodium concentration
Osmolality A measure of solute content in the blood. Dehydration tends to raise blood osmolality
Tonicity Measure of the osmotic pressure gradient between two solutions

Introduction

Water is essential for the maintenance of life. As part of regular homeostasis, body water is lost mainly through the kidneys as urine but also through skin, lungs and the gastrointestinal tract. Dehydration results when water losses from the body exceed water replacement. It implies loss of water from both extracellular (intravascular and interstitial) and intracellular spaces. Dehydration alters circulatory hemodynamics and may result in disruption of the delicate electrolyte and acid-base balance needed to maintain healthy cells and tissues. The most common causes of dehydration result in both salt and water loss from the extracellular space of

which diarrheal diseases are the primary example, in contrast, pure water loss occurs due to conditions resulting in excessive sweating or in kidney diseases like diabetes insipidus, resulting in intracellular water loss. Management of dehydration depends on the severity and the precipitating medical conditions. In mild-to-moderate cases, oral rehydration with adequate fluid solutions may be sufficient. In severe cases, or when gastrointestinal function is severely impaired, intravenous fluids may be required, along with appropriate electrolyte replacement.

Physiology of water balance

After oxygen, water is the most essential nutrient needed to sustain human life and it is responsible for connecting the diverse physiological functions of the body, some of which are shown in Table 1.

Distribution of water in the body spaces

Water is distributed across various fluid compartments in the body. For the average 70 kg man, 60% of the total body weight is comprised of water, equaling 42 L. The average adult woman is made up of 55% water because women naturally have more fatty tissue than men. The body's fluid separates into two main compartments: Intracellular fluid (ICF) and extracellular fluid (ECF). Using the average man as an example, of the 42 L of water found in the body, two-thirds of it is within the intracellular space, which equates to 28 L, forming the ICF. The ECF, which makes up one-third of the total body water, equivalent to 14 L is comprised of two components: the interstitial fluid (ISF), three-fourth of the ECF or 10.5 L and plasma volume (PV), one-fourth of the ECF or 3.5 L. Representational image is shown in Fig. 1.

Kidney and body fluid balance

The maintenance of the osmolality of body fluids within a narrow physiologic range is made possible by homeostatic mechanisms that control the intake and excretion of water. Significant fluctuations of solute and water intake fail to disrupt the total solute concentration of body fluids which remains virtually constant and the serum osmolality is maintained between the physiological range of 285–290 mOsm kg^{-1} H2O. In a normal adult, approximately 130 L–180 L of fluid are filtered across the glomerular capillaries each day. More than 98%–99% of the filtrate is then reabsorbed by the tubules, resulting in a urine output averaging 1–2 L day^{-1}. Urine volume can be as little as 0.5 L day^{-1} during water deprivation to 20 L day^{-1} during water loading (Nielsen et al., 2007).

Volume and electrolyte changes in response to decreased blood water content (increased osmolality) trigger the hypothalamus to stimulate arginine vasopressin (AVP) release from the posterior lobe of the pituitary gland. AVP acts on the kidney to increase tubular water resorption and maintain plasma volume. Decreased plasma volume also results in a complex series of events resulting in the release of renin from the kidneys and the subsequent formation of angiotensin II and the mineralocorticoid, aldosterone. Angiotensin II is a potent vasoconstrictor and stimulator of thirst. The osmotic threshold for thirst usually occurs at 290–295 mOsm kg^{-1} H2O and is above the threshold for AVP release. Aldosterone promotes sodium resorption, which allows the blood to retain more water. AVP also increases distal nephron reabsorption of urea and its recycling to improve the efficiency of water reabsorption. The net result of these regulatory mechanisms is concentrated urine and maintenance of the plasma volume, provided that exogenous fluid intake increases proportionally. If fluid intake is not increased, dehydration will result (Fig. 2).

Gastrointestinal tract and body fluid balance

In an adult, during a normal day, 8000 mL of fluid enters the small bowel; 500 mL as saliva, 1200 mL as gastric fluid, 600 mL as bile, 1200 mL as pancreatic juice, 2000 mL as small bowel secretion, and the remainder as ingested fluid. Almost 1.5 L is reabsorbed in the small intestine, and 1.4 L is reabsorbed in the colon. Only 100 mL is normally lost in the feces per day. With gastrointestinal diseases like vomiting or diarrhea, large volumes of water and electrolytes may be lost to the outside or, in the case of intestinal obstruction or ileus, pooled in the gut and lost to the functional compartments.

Table 1 Major physiological functions of water.

Function	*Example*
Waste product removal	Excretion by kidneys
Solvent for chemical reactions	Glycolysis in the cell cytosol
Transport medium	Blood
Lubrication	Synovial fluid of joints
Shock absorber	Disks between vertebrae of spinal column
Temperature regulation	Evaporative sweat loss

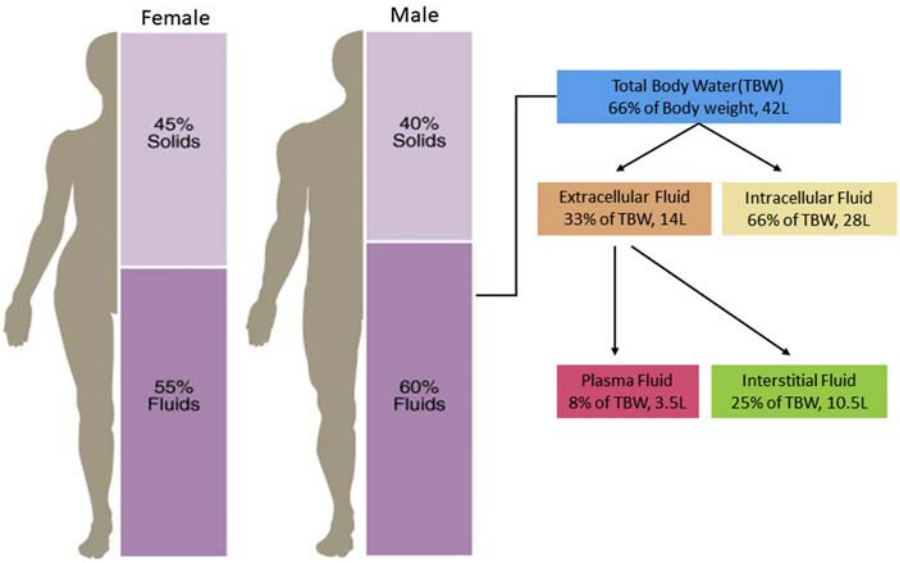


Fig. 1 Distribution of body fluids.

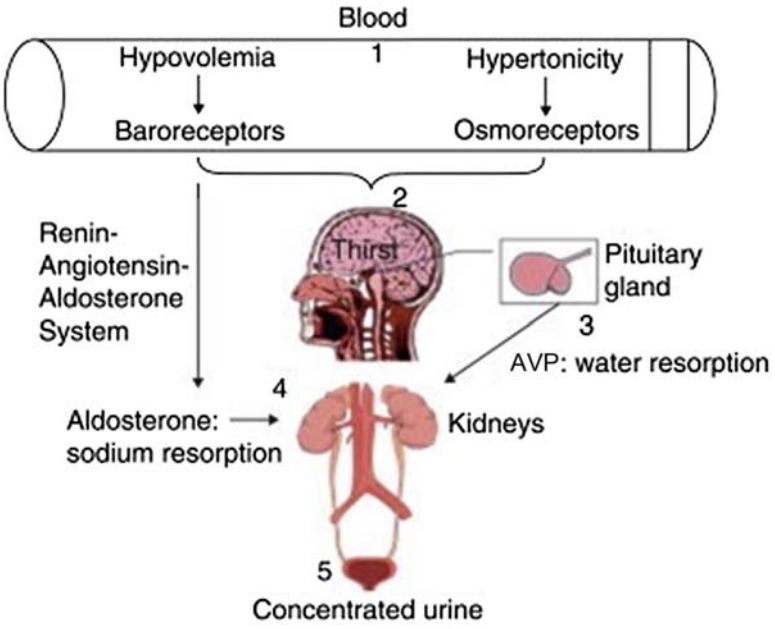


Fig. 2 Water and sodium physiology: mechanisms controlling body water gain and loss. As water is lost from the body via sweat, urine, respiration, and feces: (1) Plasma osmolality increases and plasma volume decreases with water loss. (2) The increase in osmolality acts on the "thirst center" in the hypothalamus to secrete AVP and stimulates the conscious desire for water. (3) The release of AVP from the pituitary gland increases tubular resorption of water by the kidney. (4) Aldosterone is formed via a series of reactions involving renin, which is released from the adrenal cortex in response to decreased blood pressure, and a plasma protein, angiotensinogen. Aldosterone promotes sodium resorption by the kidney to maintain plasma volume. (5) These events conserve water and result in the production of concentrated urine.

Sweat and thermoregulation

Water absorbs heat produced at the cellular level and transfers it to the surface of the skin, where it can be dissipated to the external environment. Approximately 500 mL of sweat is lost per day under average ambient environmental conditions. Such obligatory water loss occurs without visible or tactile sensations and is termed "insensible" sweat. However, given a sufficient thermal challenge, humans are capable of producing approximately 10 L of "sensible" sweat per day. Sweat is hypotonic (sodium concentration 15–65 mmol L^{-1}).

Respiration related water loss

Insensible loss from the respiratory tract is also about 400 mL day^{-1} in an unstressed adult. The water loss here is variable: it is increased if minute ventilation increases and can be decreased if inspired gas is fully humidified at a temperature of 37 °C (Newburgh and Johnston, 1942).

In summary, under normal physiological conditions, the body is able to regulate its water content tightly over a 24 h period (approximately ±200 mL) with the help of various feedback mechanisms (Fig. 3).

Pathophysiology of dehydration

Dehydration explicitly refers to the loss of total body water producing hypertonicity but it is often used interchangeably with volume depletion, which refers to a deficit in extracellular fluid volume. The distinction between these two conditions is important as it will determine the type of fluids used for treatment.

Tonicity and osmolality

Tonicity i.e., the measure of the osmotic pressure gradient between two solutions, is a physiologic term that refers to the volume of cells in a solution; cell volume tends to expand as body fluids become hypotonic or shrink as surrounding fluids become hypertonic.

Hypertonicity usually results from a disproportionate fall in total body water relative to total body sodium producing hypernatremia. With pure water loss, extracellular tonicity rises and draws fluid from the intracellular compartment. Hypertonicity thus results in intracellular volume contraction, while volume depletion results in blood volume contraction. Neurons are particularly sensitive to hypertonicity. Extreme pure water losses resulting in serum Na+ >170 mEq L^{-1} cause hemodynamic alterations but significant neurologic manifestations become apparent much before this level is reached (Bhave and Neilson, 2011).

The homeostatic response to hypertonicity is AVP-mediated urinary water conservation and stimulation of thirst seeking water. AVP release and thirst are much more sensitive to hypertonicity compared to hypovolemia (Stricker, 2004). Decreased urine output is evident early in hypertonicity (<5% increase over set point), while appearing late in hypovolemia as deficits of at least 10% of the blood volume are required to stimulate AVP release and raise urine osmolality.

Osmolality depends upon only the total solute concentration, while tonicity is determined by how it affects cell volume, which depends not only on the solute concentration but also on the solute permeability of cell membranes. Plasma osmolality can be easily determined by directly measuring or can be calculated from the formula: Plasma osmolality = 2 × Na + Glucose (mg dL^{-1})/18 + BUN (mg dL^{-1})/2.8. The effective plasma osmolality is determined by those solutes in the plasma which do not freely permeate the cell wall and act to hold water within the ECF. So, lipid soluble solutes such as urea which can cross the cell membrane do not contribute to the osmotic pressure gradient between ECF and ICF. Effective osmolality = 2 × Na (mEq L^{-1}) + Glucose (mg dL^{-1})/18 (mOsm Kg^{-1}).

Normal serum osmolality is 285–295 mOsm kg^{-1} H2O. Serum Osmolality >295 mOsm kg^{-1} H2O can be considered to suggest impending dehydration. Pure water losing dehydration results in an elevated serum osmolality.

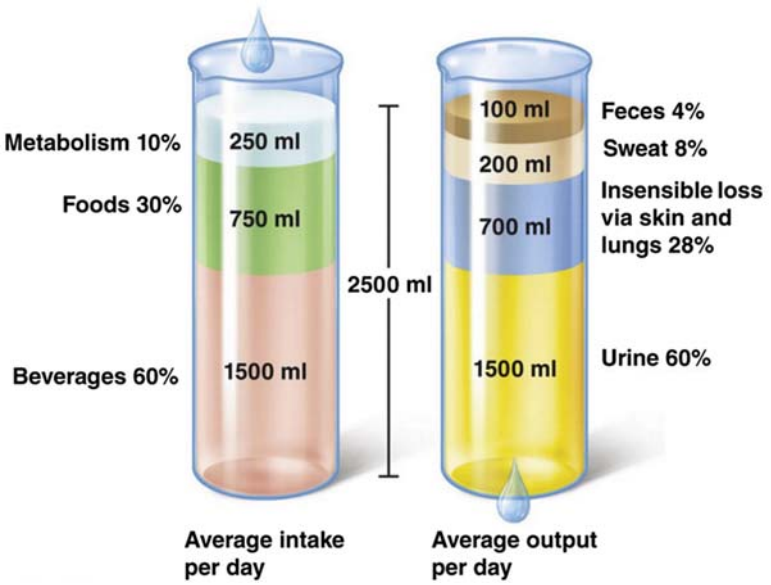


Fig. 3 Fluid intake and loss over 24 h.

Blood volume

Fluid lost via the gastrointestinal tract, renal or via third spacing into the interstitial space is usually slow enough to draw from the ECF compartment although initially the fluid is derived from the plasma volume. Acute hemorrhagic fluid loss is mostly derived from the plasma volume. When net fluid loss is isotonic, it draws completely from the ECF and thus the volume of fluid loss exactly equals the volume deficit. Changes in heart rate or blood pressure do not become evident in normal subjects until 15–20% of blood volume is removed acutely (Knopp et al., 1980). Assuming a 15% fall in blood volume as a minimal threshold for clinically detectable volume depletion, a non-hemorrhagic, isotonic loss of about 15% of ECF amounting to 7% of total body water is required. In contrast, a pure water deficit equivalent to 15% of total body water is needed to reach the same hemodynamic threshold. Consequently, isotonic losses are about 2-fold more potent than pure water losses at depleting blood volume. Isotonic losses alter systemic hemodynamics, reduce blood volume and GFR, and leave body tonicity unchanged. Conversely, an equivalent pure water deficit does not measurably alter blood volume or GFR, while hypernatremia and hypertonicity are prominent.

Dehydration and heat illness

Dehydration also adversely affects the body's ability to regulate temperature. If water loss due to sweating is not replaced during exercise, plasma volume and sweat rate will be decreased (**Fig. 4**). The combination of reduced peripheral blood flow for heat exchange and reduced sweat volume for evaporative cooling leads to an overall reduction in the ability to dissipate heat (Sawka et al., 2000).

The consequence of impaired heat dissipation is hyperthermia. Without evaporative cooling, human core temperatures can elevate 5 °C h^{-1} during moderate intensity work. The heat production is proportional to the intensity and duration of work, ranging from 75 kcal h^{-1} (314 kJ) at rest to 300 kcal h^{-1} (1256 kJ) during moderate exercise and 600 kcal h^{-1} (2512 kJ) for maximal sustained work. Brief periods of intense exercise can generate heat at the rate of 900 kcal h^{-1} (3768 kJ).

Hyperthermia can lead to serious or even life-threatening heat injury if left unchecked. Heat injury can result if the rate of heat production is greater than the rate of cooling. When fluid losses are not replaced during activity, heat dissipation mechanisms are compromised. The buildup of heat in blood and tissues adversely affects various physiological systems. Minor heat injury syndromes include prickly heat (skin rash resulting from plugged sweat glands), heat syncope (light headedness due to pooling of blood in the extremities), and heat cramps (muscle cramps related to electrolyte loss). These heat illnesses are of concern but not life-threatening. Major hyperthermia syndromes involving dehydration are heat exhaustion which can progress to life threatening heat stroke that requires immediate medical treatment (Cheshire, 2016).

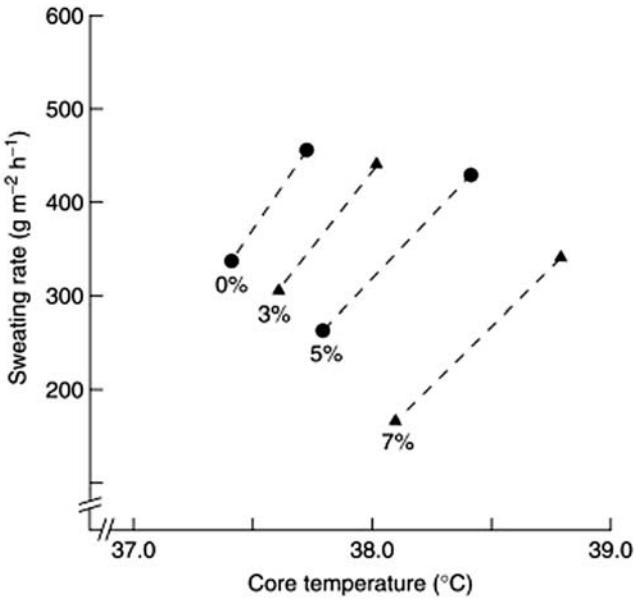


Fig. 4 The influence of water loss by dehydration (hypohydration) on the sweating response to exercise following normal hydration (0%) and dehydration equal to 3%, 5%, and 7% of the body weight. The primary stimulus for sweating is the increase in core temperature (thermal drive). Note that dehydration reduces the sweating rate at any given level of thermal drive. Hypohydration compromises exercise by reducing sweat rate and evaporative cooling and increasing body core temperature. Reproduced from Sawka, M.N., Young, A.J., Francesconi, R.P., et al., 1985. Thermoregulatory and blood responses during exercise at graded hypohydration levels. J. Appl. Physiol. 59: 1394–1401, with permission from APS.

Groups at risk for dehydration

Infants and children

According to the WHO, diarrheal disease is the second leading cause of death in children under 5 years old, and is responsible for killing around 525,000 children every year. Infants and young children are especially vulnerable because they lack the ability to convey their thirst to caregivers or to access fluids on their own. They also have increased insensible losses due to a higher body surface area.

Old age

With aging, there is a decline in total body water, in both the extracellular and intracellular fluid volume. There is also decreased fluid intake related to decreased thirst due to a higher osmotic operating point (the point at which the thirst sensation is triggered). Aging is also associated with a reduced glomerular filtration rate, increased proximal tubular renal absorption and a decline in maximal tubular fluid concentrating ability (Thomas et al., 2008).

The normal nocturnal rise that occurs with AVP is blunted with aging resulting in nocturia. These physiological changes that occur with aging place older persons at major risk for dehydration.

Certain behavioral factors may also influence drinking patterns in older adults who may wish to avoid the physical difficulty associated with trips to the bathroom. Dehydration also alters the effective dosage of medications through plasma volume changes, leading to further medical complications in the elderly. Dehydration in the elderly often accompanies or results from clinical conditions and/or medications.

People who work or exercise outside

When the weather is hot and humid, risk of dehydration and heat illness increases. When the air is humid, sweat can't evaporate and cool as quickly as it normally does, and this can lead to an increased body temperature and the need for more fluids.

Predisposing factors for dehydration and heat illness

Obesity (extra exertion, heat production, and sweating are required to move a larger mass), insufficient heat acclimation (associated with reduced sweating and evaporative cooling and increased cardiovascular and renal stress), socioeconomic barriers to cooling methods (fans, air conditioners, etc.), pyrexial illness (fever), drug and alcohol abuse (interferes with fluid balance and thermoregulation), physical work in environments that contribute to dehydration (heat: sweating; cold: respiratory water loss and diuresis; altitude: respiratory water loss and diuresis), and athletic competition and training (if athletes do not replace sweat loss).

Identifying types of dehydration

Dehydration usually occurs along a continuum of fluid and electrolyte loss. The ratio of water to electrolyte loss determines the type of dehydration present. A classification of the types of dehydration, their characteristics and likely causes is shown in **Table 2**.

Clinical features and diagnosis

Thirst is not a good short-term regulator of fluid balance. Humans frequently lose up to 2% of their body weight as water before the thirst mechanism is activated but one of the first symptoms of dehydration is thirst. Mild dehydration can also cause dry mouth, headaches, tiredness, lack of energy and feeling faint on standing up. Another early sign of dehydration is urine that is darker yellow in color than usual. Monitoring the urine color has been suggested as a noninvasive way to measure hydration status, where light, pale yellow urine generally indicates a favorable hydration status. Assessing urine color may be a simple method to assess hydration status, but it can also be artificially influenced by dietary intake (i.e., nutritional supplements). Moderate dehydration results in dizziness, muscle cramps, pale and dry skin, and sunken eyes. Severe dehydration results in confusion or disorientation.

In common with most diagnoses, that of dehydration depends mainly on the history and examination. A detailed history of fever, diarrhea (duration, frequency, consistency, presence or absence of mucus or blood), vomiting (duration, frequency, consistency), medication history (e.g., recent antibiotic use, diuretics, laxatives); exposure to heat and/or cold; weight loss; and potential ingestions (e.g., drugs) should be taken.

Physical examination findings, although frequently nonspecific, may help support a diagnosis of dehydration and include: altered mental status, decreased capillary refill, decreased skin turgor which is tested on inner aspect of thighs or the skin overlying the sternum, and is less reliable in older patients due to decreased skin elasticity with age, dry mucous membranes of the tongue and oral mucosa, orthostatic hypotension, which is determined by taking supine blood pressure after the patient is lying for 5–10 min, and then taking the blood pressure as soon as the patient sits or stands up, and again in this position after two to 3 min. A drop in systolic blood pressure $\geq$20 mm Hg or a drop in diastolic blood pressure $\geq$10 mm Hg from supine indicates orthostatic hypertension.

Table 2 Types of dehydration.

Hypertonic dehydration
Water loss exceeds salt loss e.g., inadequate water intake, fever, excessive sweating, osmotic diuresis and diuretic drugs
There is osmotic shift of water from the ICF to the ECF
Elevated blood osmolality and hypernatremia may be seen
Symptoms include thick, doughy texture to skin (tenting is uncommon), tachypnea, intense thirst
Isotonic dehydration
Isotonic loss of both water and solutes from the extracellular fluid (ECF) e.g., vomiting, diarrhea or inadequate intake
No osmotic water shift from the ICF to the ECF
Blood electrolytes usually normal
Hypotonic dehydration
Sodium loss is higher than water loss e.g., in some instances of high sweat or gastro-intestinal fluid losses or when fluid and electrolyte deficits are treated with water replacement only
Characterized by an osmotic shift of water from the ECF to the ICF
Lethargy and irritability are common and vascular collapse can occur early

Laboratory testing for mild dehydration is not usually necessary, especially when the underlying cause is apparent. For moderate to severe cases of dehydration, blood investigations including complete blood count, electrolytes, blood urea nitrogen, creatinine, acid-base status, glucose, and urinalysis (specific gravity, hematuria, glucosuria) may be considered. Measurements of plasma osmolality and urine-specific gravity and osmolality can be used to assess the relative dehydration if baseline (euhydrated) values are known. However, due to significant interindividual variation, the use of absolute specific gravity and osmolality values for the diagnosis of dehydration remains questionable. The plasma sodium concentration gives no clue to the total body sodium content, unless water balance is also known. It merely reflects the relative proportion of sodium and water in the extracellular space. Thus, with pure water deficit, the plasma sodium concentration increases. With a mixed deficit of sodium and water, the plasma sodium reflects the relative proportion of the two that are lost. Sodium levels in moderate-to-severe dehydration are useful to determine the type and speed of fluid repletion. An acute change in body weight is the most common and practical method used to assess hydration status. It is assumed that the short-term body weight loss is primarily the result of water loss.

The history and physical examination combined with a knowledge of the natural history of the presenting condition, provides the main clues to the presence of salt and/or water depletion, the likely evolution of the condition and the necessary treatment.

Table 3 gives the American Academy of Pediatrics guidelines for the assessment of dehydration in children (American Academy of Pediatrics Provisional Committee on Quality Improvement, 1996).

Management of dehydration

Goals of treatment

The primary goals of immediate treatment are to estimate the fluid and/or electrolyte deficit, rehydrate the patient at the earliest which involves, replacement of fluid and electrolyte deficits, maintenance needs and ongoing fluid losses.

The American Academy of Pediatrics has published guidelines for oral rehydration therapy of infants and children younger than 5 years old with acute gastroenteritis. An algorithm for the treatment and hydration of children with acute diarrhea is shown in Fig. 5.

Table 3 Guidelines for the assessment of dehydration.[a]

	Dehydration		
Variable	*Mild, 3–5%*	*Moderate, 6–9%*	*Severe, ≥10%*
Blood pressure	Normal	Normal	Normal to reduced
Quality of pulses	Normal	Normal or slightly decreased	Moderately decreased
Heart rate	Normal	Increased	Increased, severe cases bradycardia
Skin turgor	Normal	Decreased	Decreased
Fontanel	Normal	Sunken	Sunken
Mucous membranes	Slightly dry	Dry	Dry
Eyes	Normal	Sunken orbits	Deeply sunken orbits
Extremities	Warm, normal capillary refill	Delayed capillary refill	Cool, mottled
Mental status	Normal	Normal to listless	Normal to lethargic or comatose
Urine output	Slightly decreased	<1 mL kg^{-1} h^{-1}	$\ll 1$ mL kg^{-1} h^{-1}
Thirst	Slightly increased	Moderately increased	Very thirsty or too lethargic to indicate

[a]The percentages of body weight loss and their corresponding categorization sometimes vary depending on the author.

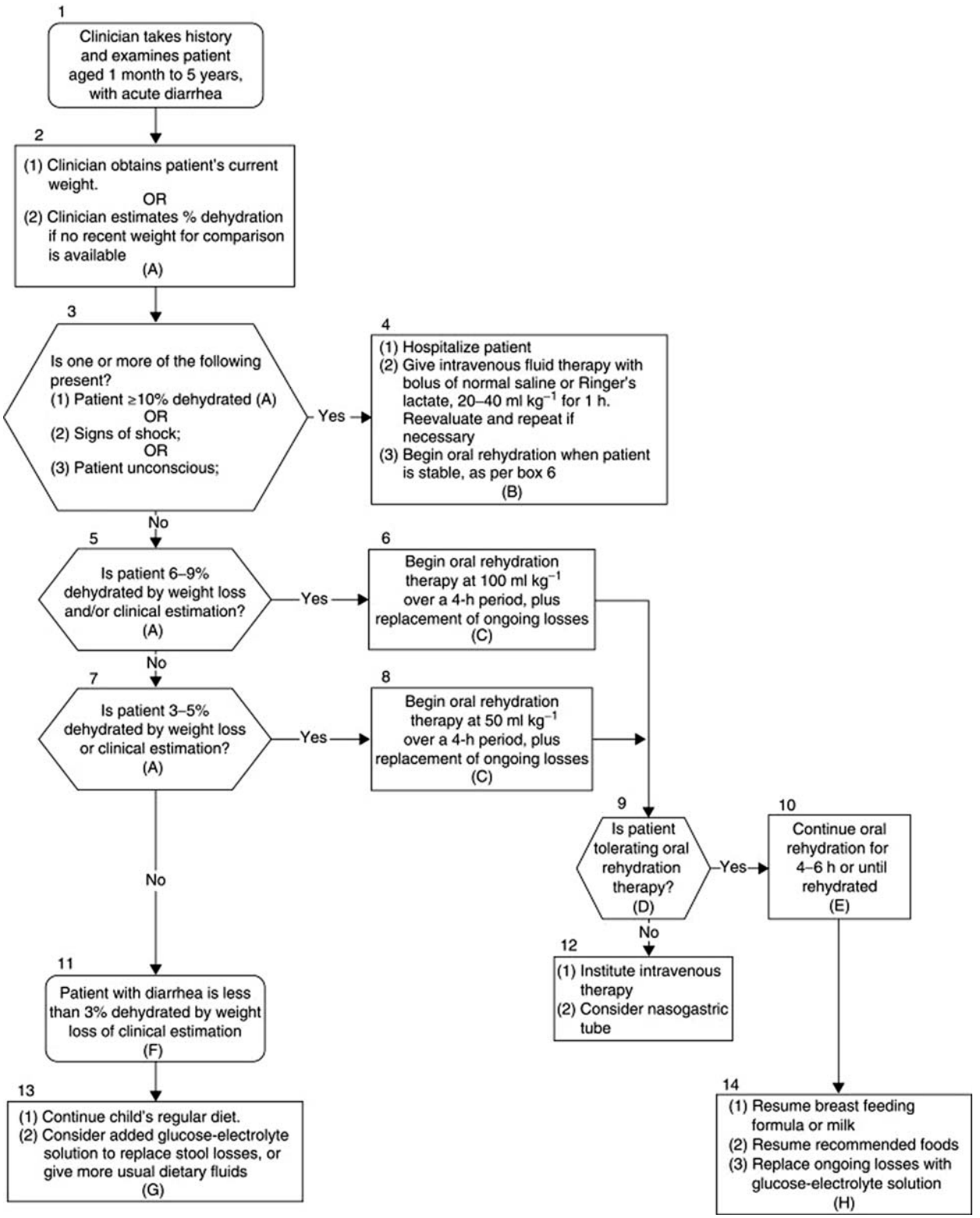


Fig. 5 Algorithm for children with dehydration from acute diarrheal disease. The letters at the bottom of the decision boxes refer to the following: (A) See **Table 3** for guidance in the assessment of the degree of dehydration. (B) Restoration of cardiovascular stability is critical and is accomplished by giving bolus i.v. therapy. In the patient who does not respond, consider the possibility of an underlying disorder. When the patient is in a stable condition and has achieved satisfactory mental status, oral rehydration therapy (ORT) can be implemented. (C) Solutions containing 45–90 mmol L^{-1} sodium should be given in a volume of 100 mL kg^{-1} for moderate dehydration and 50 mL kg^{-1} for mild dehydration. Giving the child these volumes requires patience and persistence, and progress must be monitored frequently. (D) Intractable, severe vomiting, unconsciousness, and ileus are contraindications to ORT. Persistent refusal to drink may require a trial of i.v. therapy. (E) The rehydration phase usually can be completed

Treating different types of dehydration

In the majority of simple, mild dehydration cases, plain water is an adequate rehydration solution. However, there are instances (e.g., children younger than 5 years of age dehydrated by vomiting and diarrhea) when water containing sodium and potassium is the proper hydrating agent. The most effective way of preventing and treating mild to moderate dehydration in infants and children with acute diarrhea is the oral administration of oral rehydration solutions (ORSs). An oral rehydration solution can be made at home with table salt and sugar using the following ingredients, 1/2 teaspoon of salt, 6 teaspoons of sugar, 4 cups (1 L) of water. There are also a number of commercially available ORSs. These solutions are designed to replace fluid and electrolytes when both the water and food intake have been restricted or compromised by diarrheal disease. The World Health Organization recommends the ORS shown in Table 4 for individuals afflicted with diarrheal disease and vomiting. Oral modes of fluid and electrolyte administration are always preferred in mild (3–5%) to moderate (6–9%) dehydration; however, intravenous fluids may be required in cases of severe dehydration ($\geq$10%) and vomiting or if the patient is in a comatose state. When i.v. fluids are administered, 0.45% saline with 5% dextrose is an effective hydrating agent.

The Holliday-Segar calculation approximates daily fluid loss, therefore the daily maintenance fluid requirements for 24 h is as follows: 100 mL kg^{-1} for the first 10 kg of wt, 50 mL kg^{-1} for the second 10 kg of wt, 20 mL kg^{-1} for the remaining wt. The 24 h number is often divided into approximate hourly rates for convenience, leading to the "4-2-1" formula. 100 mL kg^{-1} 24 h^{-1} = 4 mL kg^{-1} h^{-1} for the first 10 kg, 50 mL kg^{-1} 24 h^{-1} = 2 mL kg^{-1} h^{-1} for the second 10 kg, 20 mL^{-1} kg^{-1} 24 h^{-1} = 1 mL^{-1} kg^{-1} h^{-1} for the remainder. e.g., 70 kg patient would require 1500 mL plus 1000 mL for maintenance in a 24 h period). Add maintenance requirements plus replacement requirements for losses (vomiting, urine, etc.). A rough guide for adults is 200–400 mL for every loose stool, for children 200 mL per motion, and for infants 1–1.5 times the usual feed. If the patient is eating, calculate fluid replacement at 75% of total. If the patient is diagnosed with hypotonic or isotonic dehydration, calculate total fluids (maintenance plus replacement) for the first 24 h, and give half this amount over the next 8 h, and the other half over the next 16 h. In hypertonic dehydration, correct the fluid deficits slowly over 48 h (Holliday and Segar, 1957).

In most instances involving heavy sweating, plain water containing 1.25 g of NaCl L^{-1} is a suitable rehydration solution. Increasing the concentration of NaCl to 5 or 6 g L^{-1} may promote the rate of rehydration but may not be palatable for some individuals. Most commercial sports drinks contain 1.2–1.8 g NaCl L^{-1} and are also good rehydration solutions, especially when both fluid and electrolytes have been lost through sweating. Fruit juices can also provide fluid, energy, and electrolytes (e.g., fresh orange juice contains approximately 10 mg of sodium and 2000 mg of potassium L^{-}) but may be too concentrated and delay gastric emptying. Diluting fruit juices 1:3 with water may yield a more appropriate rehydration solution. The inclusion of carbohydrate in the rehydration solution provides energy for the intestinal sodium pump, which facilitates sodium transport across the intestinal cell wall into the blood, where it in turn exerts a positive osmotic effect on water absorption from the gut. Commonly used beverages, such as apple juice, tea, ginger ale, colas, and chicken broth, are inappropriate to use for rehydration because they do not contain the correct sodium and glucose ratio to promote salt and water reabsorption across the intestinal lumen. Glucose and electrolyte sports beverages are useful rehydration solutions for sporting activities but are not a good choice for children with diarrhea because these beverages have lower electrolyte and higher carbohydrate concentrations than recommended (Wilk et al., 1998).

Prevention of dehydration

Dehydration resulting from non-disease causes can be easily prevented provided that people are inclined to drink and have access to cool, safe sources of fluids. Drink flavoring, beverage temperature, and sodium chloride content are important promoters of fluid intake in active children. Education of athletic coaches, the general public, and health care providers is necessary to increase awareness of the importance of proper hydration. The American College of Sports Medicine has issued a set of guidelines for fluid replacement (Table 5).

Simple methods, such as recording body weight before and after exercise to determine fluid loss and observing the color of urine or the turgidity of skin, can be useful for monitoring hydration status. The simplest insurance against dehydration is to consume fluids before and during physical activity or heat exposure to match water loss. The amount of fluid needed to maintain a favorable hydration status is variable between individuals but often necessitates drinking in the absence of thirst. Excess fluid consumption is rarely a problem. However, caution should be used to avoid dilutional hyponatremia from overzealous hydration. Humans can acclimate to work in a hot environment and enhance their ability to thermoregulate and conserve fluid, but they cannot adapt

in 4 h; reevaluation should occur every 1 or 2 h. See referenced text for guidance to decide when rehydration has been achieved. (F) The type and intensity of therapy will vary with the individual clinical situation. (G) Often, a child has diarrhea but remains adequately hydrated. The parent can be reassured but should be taught to assess hydration and to identify a worsening condition. If the stool output remains modest, ORT may not be required if early, age-appropriate feeding is instituted and increased consumption of usual dietary fluids is encouraged. More significant stool losses can be replaced with an oral rehydrating solution at the rate of 10 mL kg^{-1} for each stool. (H) Breast-feeding should be resumed. Nonlactose formula, milk-based formula, or milk may be given, although a small percentage of children will not tolerate lactose-containing fluids. Lactose-containing solutions seem to be tolerated better when combined with complex carbohydrates in weaned children. Children who are eating foods may resume eating, although certain foods are tolerated better than others. Recommended foods include complex carbohydrates (rice, wheat, potatoes, bread, and cereals), lean meats, yogurt, fruit, and vegetables. Avoid fatty foods and foods high in simple sugars (including juices and soft drinks). Supplement feeding with an oral electrolyte solution, 10 mL kg^{-1} for each diarrheal stool and the estimated amount vomited for each emesis.

Table 4 Composition of recommended WHO/UNICEF oral rehydration solution.

Solute	*Content (mmol L^{-1})*
Glucose	75
Sodium	75
Chloride	65
Potassium	20
Citrate	10
Total osmolarity	245

Table 5 Fluid replacement: summary of recommendations of the American College of Sports Medicine.

It is recommended that individuals consume a nutritionally balanced diet and drink adequate fluids during the 24 h period before an event, especially during the period that includes the meal before exercise, to promote proper hydration before exercise or competition.

It is recommended that individuals drink approximately 500 mL of fluid approximately 2 h before exercise to promote adequate hydration and allow time for excretion of excess ingested water.

During exercise, athletes should start drinking early and at regular intervals in an attempt to consume fluids at a rate sufficient to replace all the water lost through sweating, or consume the maximal amount that can be tolerated.

During exercise lasting less than 1 h, there is little evidence of physiological or physical performance differences between consuming a carbohydrate–electrolyte drink and plain water.

Inclusion of sodium (0.5–0.7 g L^{-1} of water) in the rehydration solution ingested during exercise lasting longer than 1 h is recommended since it may be advantageous in enhancing palatability, promoting fluid retention, and possibly preventing hyponatremia in certain individuals who drink excessive quantities of fluid. There is little physiological basis for the presence of sodium in an oral rehydration solution for enhancing intestinal water absorption as long as sodium is sufficiently available from the previous meal.

Reproduced from the American College of Sports Medicine, 1996. Position stand on exercise and fluid replacement. Med. Sci. Sports Exerc. 28: i–vii, with permission from LWW.

to dehydration. Acute dehydration can decrease physical performance and thermoregulation ability and increase the risk of heat illness. Chronic dehydration can reduce the metabolic and thermoregulatory efficiency and increase predisposition to kidney disease. The deleterious effects of dehydration on physiological function are summarized in Fig. 6.

Conclusion

Dehydration is common in infants, children and elderly, especially following gastrointestinal illnesses. Oral rehydration can be safely and effectively accomplished in those with mild-to-moderate dehydration. Individuals with severe dehydration or with

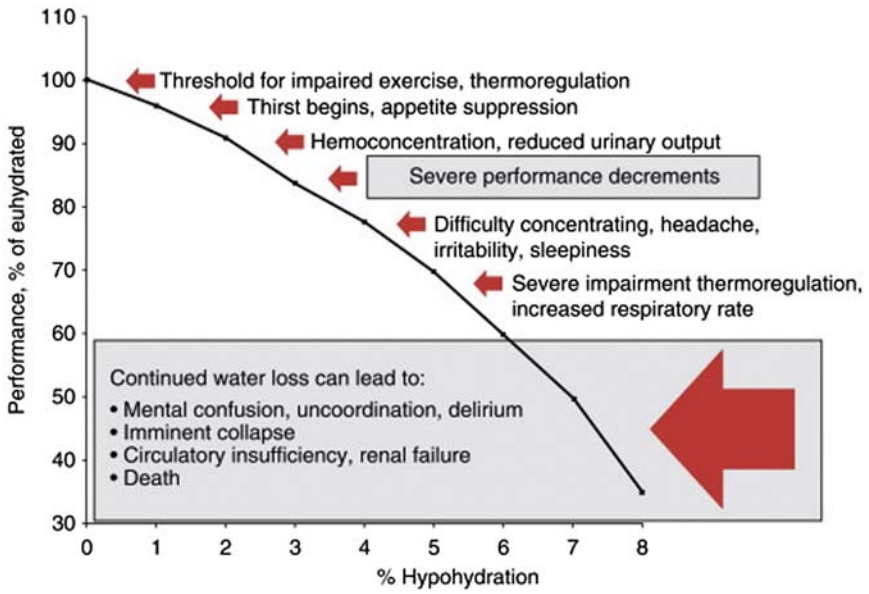


Fig. 6 Progressive physiological effects of dehydration on physical performance and pathophysiology of hypohydration. The onset, magnitude, and severity depend on the workload, level of physical fitness, ambient temperature, relative humidity, and degree of heat accumulation of the individual. Reproduced with permission from Askew, E.W., 1996. Water. In: Ziegler, E.E., Filer, L.J. (Eds.), Present Knowledge in Nutrition, pp. 98–108. Washington, DC: ILSI Press.

abnormal serum sodium values should be treated with intravenous infusions. Early identification of the precipitating cause and determining the correct fluid and electrolyte solutions to meet the individual's maintenance, deficit and ongoing losses will help prevent serious complications. Prevention by adequate education is especially important in non-disease related dehydration in individuals at risk.

References

American Academy of Pediatrics Provisional Committee on Quality Improvement, 1996. The management of acute gastroenteritis in young children. American Academy of Pediatrics, Provisional Committee on Quality Improvement, Subcommittee on Acute Gastroenteritis. Pediatrics 97, 424–436.

Bhave, G., Neilson, E.G., 2011. Volume depletion versus dehydration: how understanding the difference can guide therapy. Am. J. Kidney Dis. 58 (2), 302–309.

Cheshire, W.P., 2016. Thermoregulatory disorders and illness related to heat and cold stress. Auton. Neurosci. 196, 91–104.

Holliday, M.A., Segar, W.E., 1957. The maintenance need for water in parenteral fluid therapy. Pediatrics 19 (5), 823–832.

Knopp, R., Claypool, R., Leonardi, D., 1980. Use of the tilt test in measuring acute blood loss. Ann. Emerg. Med. 9 (2), 72–75.

Newburgh, L.H., Johnston, M.W., 1942. The insensible loss of water. Physiol. Rev. 22 (1), 1–18.

Nielsen, S., Knepper, M.A., Kwon, T.H., 2007. Regulation of water balance. Urine concentration and dilution. In: Schrier, R.W. (Ed.), Diseases of the Kidney and Urinary Tract, eighth ed. Lippincott Williams & Wilkins, Philadelphia, pp. 96–123.

Sawka, M.N., Latzka, W.A., Montain, S.J., 2000. Effects of dehydration and rehydration on performance. In: Nutrition in Sport, pp. 216–225.

Stricker, E., 2004. Thirst. In: Stricker, E., Woods, S. (Eds.), Neurobiology of Food and Fluid Intake. Plenum Publishers, New York, NY, pp. 507–545.

Thomas, D.R., Cote, T.R., Lawhorne, L., Levenson, S.A., Rubenstein, L.Z., Smith, D.A., Stefanacci, R.G., Tangalos, E.G., Morley, J.E., 2008. Understanding clinical dehydration and its treatment. J. Am. Med. Dir. Assoc. 9 (5), 292–301.

Wilk, B., Kriemler, S., Keller, H., Bar-Or, O., 1998. Consistency in preventing voluntary dehydration in boys who drink a flavored carbohydrate-NaCl beverage during exercise in the heat. Int. J. Sport Nutr. 8 (1), 1–9.

Further reading

American College of Sports Medicine, Sawka, M.N., Burke, L.M., Eichner, E.R., Maughan, R.J., Montain, S.J., Stachenfeld, N.S., 2007. American College of Sports Medicine position stand. Exercise and fluid replacement. Med. Sci. Sports Exerc. 39 (2), 377–390.

Arai, S., Stotts, N., Puntillo, K., 2013. Thirst in critically ill patients: from physiology to sensation. Am. J. Crit. Care 22 (4), 328–335.

Armstrong, L.E., Casa, D.J., Millard-Stafford, M., Moran, D.S., Pyne, S.W., Roberts, W.O., 2007. Exertional heat illness during training and competition. Med. Sci. Sports Exerc. 39 (3), 556–572.

Popkin, B.M., D'Anci, K.E., Rosenberg, I.H., 2010. Water, hydration, and health. Nutr. Rev. 68 (8), 439–458.

World Health Organization, 2017. Diarrhoeal Disease. Available at. https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease.

Diabetes mellitus: Diagnosis and heterogeneity

Ram Jagannathan[a], Kosuke Tamura[b], and Priyathama Vellanki[c], [a] Division of Hospital Medicine, Emory University School of Medicine, Atlanta, GA, United States; [b] Neighborhood Social & Geospatial Determinants of Health Disparities Laboratory, Population and Community Sciences Branch, Division of Intramural Research, National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD, United States; and [c] Division of Endocrinology, Metabolism and Lipids, Emory University School of Medicine, Atlanta, GA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of K.C. McCowen, R.J. Smith, Diabetes Mellitus: Classification and Chemical Pathology, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 17–24, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00072-6.

Introduction 227
Definition 227
Type 2 diabetes 227
Intermediate hyperglycemia/prediabetes 228
Diagnostic cut-points for diabetes diagnosis 228
Oral glucose tolerance test 228
Fasting plasma glucose 228
Non-agreement of FPG and 2 h PG 229
Glycated hemoglobin 229
Diagnostic criteria for diabetes in pregnant women 230
Diagnostic criteria for intermediate hyperglycemia/prediabetes 230
Classification 230
Heterogeneity in T1D 231
Heterogeneity in T2D 231
Conclusion 232
Funding acknowledgment 232
References 232
Further readings 232

Key points

- There is a global consensus for the diagnosis of diabetes.
- The definition/classification of prediabetes or intermediate hyperglycemia is not globally agreed.
- Newer data-driven, hypothesis free classification systems to discern diabetes heterogeneities are emerging.

Introduction

Diabetes Mellitus is a debilitating and costly disease. Although pathogenic pathways to diabetes include defects in insulin action, pancreatic β-cell function, or both, destruction of or dysfunctional β-cell function is the culminating pathway that can lead to a metabolic imbalance responsible for the development of the disease. Because type 2 diabetes mellitus (T2DM) may be prevented or delayed with various prevention modalities, early and accurate identification of individuals at high-risk of developing diabetes at the prediabetes or intermediate hyperglycemia stage is crucial (Haw et al., 2017). In this article, the definition and classification of diabetes will be discussed, and various diagnostic criteria will be outlined. Subsequently, we will also summarize the heterogeneity of diabetes and novel data-driven approaches to decipher different diabetes subtypes.

Definition

Type 2 diabetes

Accurate diagnosis of diabetes with reliable and inexpensive tests for screening and early identification of T2D has historically been challenging. Figure depicts the evolution of diabetes diagnosis. The present diagnostic cut points for diabetes (fasting plasma glucose (FPG) > 7.0 mmol/L (126 mg/dL), 2 h post oral glucose load (OGTT) plasma glucose (2 h PG > 11.1 mmol/L (200 mg/dL), or HbA1c > 6.5% (48.0 mmol/mol)) are derived based on glycemic levels above which a substantially increased risk of

diabetes-associated microvascular complications is apparent, particularly diabetic retinopathy. These criteria exploit the observation that there appears to be a linear association between glycemia and diabetic complications, and there was a clear glycemic threshold separating persons at high and low-risk diabetic-specific retinopathy. An individual-level pooled analysis of cross-sectional data from nine studies from five countries with 44,623 participants aged 20–79 years with diabetes-specific retinopathy from DETECT-2 collaborators suggests that the current FPG level for diagnosis of diabetes should be lowered to 6.5 mmol/L (117 mg/dL) and that an HbA1c of 6.5% (48.0 mmol/mol) is a suitable alternative diagnostic criterion. Each assessment provides pivotal information about glucose metabolism and reflects different physiological mechanisms. The FPG reflects post-absorptive glucose homeostasis, while the 2 h PG primarily reflects the disposal of an exogenous glucose load. The HbA1c level strongly correlates with overall glycemia, reflecting the average glucose over 2–3 months. The FPG strongly correlates with HbA1c in the nondiabetic range as elevations in the FPG concentration are present throughout the day. In contrast, postprandial hyperglycemic excursions are transient, occurring 3–4 h after each meal, while 2 h PG is more strongly associated with elevations in HbA1c with increasing overall glycemia. Therefore, it is not surprising that the HbA1c has a stronger correlation with the FPG than the 2 h PG.

Intermediate hyperglycemia/prediabetes

The development of T2DM is preceded by an intermediate prediabetic or hyperglycemia stage (impaired glucose tolerance (IGT) and impaired fasting glucose (IFG)) in which blood glucose level is elevated when compared to normal but is below diabetic thresholds. Prediabetes is a risk state that may last for several years and increases the risk of developing diabetes. However, it should be noted that only ~30–40% of those with prediabetes eventually develop T2DM. In a meta-analysis of prospective studies, when compared to normoglycemic individuals the relative risk for diabetes was: 5.52% in people with isolated IGT; 7.54% in people with isolated IFG and 12.13% in people with both IFG and IGT (combined glucose intolerance). In a separate study, prediabetes as identified by HbA1c criteria (5.7%–6.4%) identified lower proportion of individuals (20.4%) than FPG (30.2%) as high-risk for diabetes, the ability to predict progression to diabetes over 5-years was stronger for HbA1c than IFG criteria.

Identification of the prediabetic stage (IGT, IFG, or HbA1c) has been used to identify individuals with high risk for T2DM. Indeed, all the major landmark primary prevention studies that have evaluated intervention strategies for preventing T2DM have recruited individuals with IGT for assessing the benefits of intervention strategies (Haw et al., 2017). However, in large, prospective, epidemiological studies, only ~30–40% of the individuals with IGT/IFG develop overt T2DM eventually. Furthermore, ~40% of individuals who develop T2DM have normoglycemia at baseline. This limits the use of IGT/IFG as the sole means to identify individuals at high risk for T2DM.

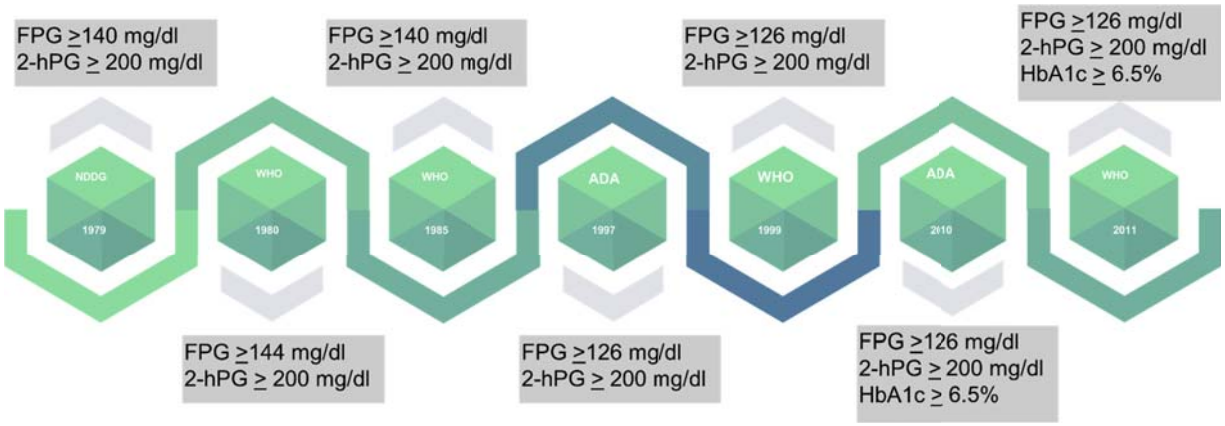
Diagnostic cut-points for diabetes diagnosis

Oral glucose tolerance test

For over a century, the oral glucose tolerance test (OGTT) has been the mainstay for the diagnosis of diabetes (Jagannathan et al., 2020). Over the years, there was substantial evolution in the conduct of OGTT with respect to the glucose loadings for testing (50, 75, or 100 g), using plasma glucose instead of whole/capillary blood, the timing and number of the sample (0, 30, 60, 90, 120 min) required for diagnosis, and the terminology for diagnosing dysglycaemia (e.g., prediabetes, intermediate hyperglycemia, chemical, borderline, subclinical, latent or overt diabetes). In 1980, the world health organization (WHO) endorsed global standardization of the OGTT with a 75 g glucose load which is still in practice (WHO Expert, 1980). Conversely, data from cross-sectional studies showed a strong association between FPG and 2 h PG values with diabetic-specific retinopathy, which led to a commonly agreed-upon protocol, endorsed by the US National Diabetes Data Group (NDDG) (National Diabetes Data Group, 1979) and the world health organization (WHO) (Report of the Expert, 1997) for the diagnosis of diabetes (FPG $\geq$ 7.8 mmol/L (140 mg/dL) and 2 h PG $\geq$ 11.1 mmol/L (200 mg/dL)) (**Fig. 1**).

Fasting plasma glucose

In the late 1960s, it was generally acknowledged that the FPG test was suboptimal for the diabetes diagnosis. Before 1997, diabetes was diagnosed based on FPG levels >7.8 mmol/L (140 mg/dL), arbitrarily determined to represent the upper threshold of normal FPG. In 1997, the ADA Expert Committee revised the criteria by lowering the FPG cut-point from 7.8 mmol/L (140 mg/dL) to 7.0 mmol/L (126 mg/dL) and retained the 2 h PG value for the diagnosis of diabetes (**Fig. 1**). The revision of the FPG cut-point was primarily owing to the poor agreement in diabetes diagnosis between the 2 h PG and FPG and/or an FPG < 7.8 mmol/L (140 mg/dL)) as well as to simplify the diagnostic procedure (FPG vs. OGTT). The ADA Expert Committee reasoned that as the association between FPG and diabetes-related complications follows a continuum, then at an FPG > 7.0 mmol/L (126 mg/dL), then diabetes must exist. Although the 2 h PG cut-point is justified based on diabetic-specific retinopathy studies, the optimal value for the diagnosis of FPG is inadequately standardized. Congruently, in 1999, the WHO also revised the cut-point of FPG > 7.0 mmol/L (126 mg/dL) and retained the 2 h PG threshold for diagnosing diabetes.



Legend. ADA: American Diabetes Association; FPG: Fasting plasma glucose; 2-h PG: 2-h post-load glucose. NDDG: National Diabetes Data Group; WHO: World Health Organization.

Notes. The figure depicts the evolution of type 2 diabetes over the past four decades. The OGTT based criteria were the mainstay for the diagnosis of diabetes between 1980-2010. The HbA1c as a diagnostic tool was recently added around a decade ago.

Fig. 1 Evolution of diagnosis of diabetes criteria over four decades. ADA: American Diabetes Association; FPG: Fasting plasma glucose; 2 h PG: 2 h post-load glucose. NDDG: National Diabetes Data Group; WHO: World Health Organization. Notes. The figure depicts the evolution of type 2 diabetes over the past four decades. The OGTT based criteria were the mainstay for the diagnosis of diabetes between 1980 and 2010. The HbA1c as a diagnostic tool was recently added around a decade ago.

Non-agreement of FPG and 2 h PG

The existing glucose-centric diagnostic tests, FPG, and the 2 h PG derived from the OGTT have well-known limitations with increased inter-individual variability (FPG (coefficient of variation (CV: 5.7%; 2 h PG (CV: 16.7%)). Due to these discrepancies and poor reproducibility, ADA and WHO agencies recommended repeated testing within a "short period of time", albeit without specifying the actual time frame to confirm the diagnosis, with the goal of reducing the probability of a false-positive diagnosis. As a result, in 1997, the ADA favored the FPG and precluded the use of 2 h PG from the OGTT (Selvin et al., 2018). Subsequently, the WHO adopted similar approach but was less restrictive about the use of the OGTT. Several observational studies showed that the association between FPG and 2 h PG with diabetes and/or diabetic-specific retinopathy was highly variable and affected by diabetogenic risk factors such as age, obesity status, sex, and ethnicity. For example, the optimal cut-points for identifying diabetic retinopathy were lower in Pima Indians, Egyptian and in US population, whereas the threshold was lower in Asian population. Therefore, revising the diagnostic criteria proposed by the latest ADA criteria resulted in the secular increase in the prevalence of diabetes and increased personal and healthcare costs. Recently, Selvin et al. proposed a clinical utility of diabetes diagnosis from a single-sample determinations of FPG and HbA1c to confirm the diabetes diagnosis (Agarwal, 2015). Performing both tests concurrently from a single sample may streamline T2DM diagnosis in resource-limited settings by eliminating the need for a second clinical visit for confirmatory testing.

Glycated hemoglobin

Glycated Hemoglobin (HbA1c) was first discovered in 1969, and in the 1980s, it was introduced in the clinical management for diabetes (as it reflects 60–90 days average of glycemia) and has become a cornerstone of clinical practice (Jagannathan et al., 2020). In 2009, the International Expert Committee that included representatives of the ADA, the international diabetes federation (IDF), and the European Association for the Study of Diabetes endorsed measurement of HbA1c to diagnose diabetes, with a cut-point of ≥6.5% (48 mmol/mol), and the ADA formally adopted this criteria in 2010. However, it must be noted that HbA1c is not a reliable biomarker for assessing undiagnosed diabetes. For instance, analyses of the US National Health and Nutrition Examination Survey (NHANES-3) data from 2005 to 2006 indicate that, assuming universal screening of the undiagnosed, the HbA1c test cut point of ≥6.5% (47.5 mmol/mol) identifies one-third fewer cases of undiagnosed diabetes than a fasting plasma glucose cut point of ≥7.0 mmol.

Several reports have shown that Blacks, Asians, and Latinos have ∼0.4% higher HbA1c levels than Whites at similar glycemic levels. Correspondingly, the HbA1c test has been associated with the imprecize classification of diabetes in these high-risk populations than in Whites. In the NHANES (year 2005–2014) study, the false-positive rates for prediabetes or diabetes were ∼3 times for Blacks (17.6%) than Whites (6.3%). Besides, HbA1c may not be reliable among children and adolescents in diagnosis of

prediabetes or diabetes, underestimating their true prevalence. Besides, HbA1c values can be affected by numerous factors such as decreased red cell turnover (e.g., iron deficiency, vitamin B12 deficiency, certain hemoglobinopathies), advanced kidney dysfunction stages, and individuals with severe immuno-compromized conditions such as HIV, recent organ transplant, or cystic fibrosis may have HbA1c discordant with blood glucose levels, resulting in underestimating the prevalence of dysglycemia.

On the contrary, the hemodialytic condition is typically associated with falsely decreased HbA1c levels due to the uremic environment, reduced erythrocyte lifespan, treatment with erythropoietin, or intravenous iron replacement treatment. Importantly, HbA1c is a relatively expensive test compared with glucose and not standardized in most of the low- and middle-income countries. In essence, although HbA1c is a valuable screening and diagnostic tool for identifying dysglycemia, it may not be sufficient to substitute for the information derived from OGTT.

Diagnostic criteria for diabetes in pregnant women

Identifying women at greater risk of developing gestational diabetes (GDM) is paramount since it is strongly associated with maternal (C-section, hypertension, cardiovascular diseases) and neonatal (macrosomia, neonatal hypoglycemia, and perinatal death) outcomes (Chung et al., 2020). A gold standard test for the screening and diagnosing of GDM is lacking, with the various testing options and thresholds are currently available (Jagannathan et al., 2020).

At present, there are two different two-step glucose thresholds criteria available to diagnose GDM, such as those defined by the Carpenter and Coustan (Agarwal, 2015), National Diabetes Data Group (NDDG) and Canadian Diabetes Association (National Diabetes Data Group, 1979; Report of the Expert, 1997). On the other hand, a one-step GDM screening approach is recommended by the International Association of Diabetes and Pregnancy Study Groups (IADPSG) (Harper et al., 2016). There has been a debate on the clinical applicability of the one-vs. two-step GDM screening test. One pragmatic randomized controlled study of 23,792 women assessed the efficacy of one-vs. two-step approach for GDM screening on perinatal and maternal outcomes (International Association of Diabetes et al., 2010). The study found that even though the one-step approach diagnosed 16.5% of women with GDM and the two-step approach diagnosed 8.5% of women with GDM, there were no differences in maternal or perinatal outcomes.

Although the HbA1c has preanalytical advantages as described in the previous section, its clinical utility in GDM has several inadequacies. A retrospective analysis from the US data showed that HbA1c in the range of 5.7%–6.4% (39–46 mmol/mol) measured at <20 weeks after gestation identified individuals at heightened risk of GDM. On the other hand, findings from New Zealand and Israel retrospective studies noted a lower HbA1c cut-off for predicting incident GDM. Therefore, more robust, prospective studies are warranted for the establishment of universal screening criteria for GDM.

Diagnostic criteria for intermediate hyperglycemia/prediabetes

Both WHO and ADA render guidance on screening for prediabetes or intermediate hyperglycemia. While the thresholds for IGT definition are the same to both guidelines, the ADA suggests a lower threshold for IFG compared to the WHO guidelines (IFG-ADA: 100–125; IFG-WHO: 110–125 mg/dL) in an effort to improve the agreement between IFG and IGT prevalence estimates.

In 2010 ADA also recommended assessing HbA1c to screen for prediabetes or intermediate hyperglycemia; however, it is to be noted that the WHO does not endorse this recommendation. These tests do not necessarily identify prediabetes or intermediate hyperglycemia in the same people, and according to ADA recommendation, abnormal results from any of the tests are adequate for prediabetes or intermediate hyperglycemia diagnosis. The most appropriate biochemical test for identifying individuals with a high risk of developing diabetes is currently unclear. The Atherosclerosis Risk in Communities (ARIC) study demonstrated that prediabetes or intermediate hyperglycemia definitions using HbA1c were more specific and provided modest improvements in risk discrimination for clinical complications. Congruently, IFG, as defined by ADA guidelines, was more sensitive diagnostic marker and HbA1c for identifying high-risk individuals. However, several other studies, disputed this finding and demonstrated the superiority of 2 h PG in predicting CVD events and mortality over other biochemical tests.

With a range of diagnostic criteria available, it is unsurprizing that individuals with prediabetes or intermediate hyperglycemia identified by each biochemical method vary widely in their phenotypes and have limited overlap. In the US Diabetes Prevention Program, Blacks had higher HbA1c levels than Whites, despite similar FPG and 2 h PG levels. Similar findings were also reported in other ethnicities. Besides, prediabetes or intermediate hyperglycemia as identified with the OGTT criteria have more severe metabolic abnormalities and diabetic-complications than when by the HbA1c criteria alone. These discrepancies in screening criteria for prediabetes or intermediate hyperglycemia may result in inaccurate diagnoses, leading to some being unnecessarily treated and others being left without treatment to prevent or delay the onset of overt diabetes. Similarly, it is arduous to estimate the global burden of prediabetes.

Classification

Around a century ago, Dr. Elliot Joslin's have made several observations such as "… It is well known that diabetic patients come too late for treatment. If the disease is detected early, it is far more susceptible to diet … Therefore, more energy must be exerted to discover this disease. One must hunt for it."

Table 1 Diabetes classification according to the ADA (American Diabetes Association, 2020).

Type	*Definition*
Type 1 diabetes	Autoimmune pancreatic β-cell destruction exacerbates to rapid loss of insulin deficiency.
Type 2 diabetes	Progressive loss of adequate insulin secretion, compounded by deteriorated insulin sensitivity.
Gestational diabetes	Diagnosed in the second or third trimesters, in mothers without overt diabetes prior to gestation.
Monogenic diabetes	There are two subgroups: MODY and neonatal diabetes
Diseases of the exocrine pancreas	Examples: pancreatitis and cystic fibrosis
Chemical or drug-induced	Caused by glucocorticoids, HIV/AIDS drugs, COVID-19, organ transplantation (immunosuppressive) drugs

The current classification of diabetes recognizes the significant forms of type 1 diabetes (T1D), T2D, GDM, and other uncommon types, such as monogenic diabetes (e.g., MODY), diseases of the exocrine pancreas (e.g., cystic fibrosis-related diabetes), or drug-induced diabetes (**Table 1**).

Distinguishing between T1DM and T2DM can be difficult. T1DM has classically been known to occur at a younger age and present with absolute insulin deficiency. T2DM, on the other hand, was classically described as occurring in older, obese individuals. However, T2DM is becoming more prevalent in younger age groups. Further, some people with diabetes present with diabetic ketoacidosis (DKA), also called ketosis-prone diabetes. People with ketosis-prone diabetes initially present with diabetic ketoacidosis at onset of diabetes (Vellanki and Umpierrez, 2017). Ketosis-prone diabetes is well described in people of African and Hispanic origin and obesity, has a high male to female prevalence and a low prevalence of pancreatic autoimmune antibodies. However, unlike patients with T1D, with intensive insulin treatment, ~70% of patients are able to discontinue insulin with glycemic control and maintained off all medications or on oral antidiabetic agents. Over the long-term, most people with ketosis-prone diabetes have a clinical course with decreasing beta-cell function similar to that of T2DM.

Thus the heterogeneity of diabetes has been the subject of considerable interest recently given the complexity of classifying diabetes and the availability of the extensive population-based databases. In the remaining chapter, we will describe the pathophysiology of diabetes "through the prism of heterogeneity" within and between diabetes types.

Heterogeneity in T1D

The temporal rise in the incidence of T1D suggests an interplay between gene-gene, gene-environment, and environment–environment interactions resulting in overt T1D. Several epidemiological and genome-wide association studies have identified over 60 risk regions across the human genome, marked by single nucleotide polymorphisms (e.g., PTEN, PTPN22, ERBB3, SH2B3, INS) and its potential interaction with environmental factors such as Enterovirus B, height, BMI, and waist circumference on the exacerbation of T1D onset. Currently, there is a debate on the definition of subtypes within T1D which may have implications for personalized treatment. Distinct histological phenotype associates with destructive insulitis on pancreatic β-cells, altered insulin processing, and the onset of T1D at a young age. Harnessing the systems biology approaches and multi-omics approach has significantly improved our understanding of various forms of T1D endotypes, and it will pave the way for a tailored therapy to the specific pathophysiological processes.

Heterogeneity in T2D

Dysregulation of many biological pathways leads to the development and progression of T2DM (Jagannathan et al., 2020). The measurement of glucose alone cannot reflect the complexity of metabolic disorders arising from a combination of excessive adipose tissue, peripheral insulin resistance, dysfunctional pancreatic β-cells, increased free fatty acids, impaired incretin effects, and obesity-induced inflammation. Hence, it is imperative to cluster individuals with T2D based on perceived phenotypic characteristics (Ahlqvist et al., 2020). Recent efforts have proposed phenotype-based approaches that take advantage of modern computational capabilities and clustering techniques. These strategies have the potential to highlight underlying disease pathophysiology, enhance prognostication, and refine treatment regimens.

In 2018, Ahlqvist et al. (Ahlqvist et al., 2020), in a Scandinavian population ($n = 8980$), studied six variables (glutamic acid decarboxylase [GAD] antibody, age, BMI, HbA1c, HOMA- β- (pancreatic β-cell function), and insulin resistance (HOMA2-IR)), assessed at the time of diagnosis with any form of new-onset diabetes. By employing k-means and hierarchical clustering algorithms, the study population was classified into five reproducible phenotypes: (1) a severe autoimmune (capturing T1D and latent autoimmune diabetes of adults (LADA)) group; (2) a severe insulin-deficient group; (3) a severe

insulin-resistant group; (4) a mild obesity-related form, and (5) a mild age-related form. The study investigators further investigated whether the subgroups differed regarding the intensification of therapy or diabetes-related complications; indeed, there were significant between-group differences, including diminished time to sustained insulin use in the severe autoimmune and severe insulin-deficient phenotypes increased risk of CKD in the severe insulin-resistant phenotype (Ahlqvist et al., 2020). Since then, several groups have replicated the findings of Ahlqvist et al. including in more ethnically diverse populations for identifying different heterogenous clusters to identify the subgroup of individuals at high risk of developing diabetes and its associated complications.

Conclusion

A shift to diagnosing high-risk individuals for diabetes or prediabetes even earlier than current screening modalities enables the opportunity for reducing progression to diabetes and its associated complications. However, the currently available diagnostic tests, as the FPG, 2 h PG, and HbA1c, have performance limitations and are inadequate to classify high-risk individuals owing to the multi-factorial nature of the diseases. Integrating our knowledge of genomics, immunology, and metabolic data will improve our understanding of the pathophysiology of diabetes heterogeneity and the ability to use precision medicine to tailor treatments and improve health outcomes (Chung et al., 2020).

Funding acknowledgment

PV is funded in part by NIH/NIDDK K23 DK 11324-01A1. K.T. was supported by the Division of Intramural Research (DIR) of the National Institute on Minority Health and Health Disparities at the National Institutes of Health (NIH) and the NIH Distinguished Scholars Program. The statements and views expressed of this work do not necessarily represent the official position of the NIH, the U.S. Department of Health and Human Services, and/or the US Government.

References

Agarwal, M.M., 2015. Gestational diabetes mellitus: an update on the current international diagnostic criteria. World J. Diabetes 6 (6), 782–791.
Ahlqvist, E., Prasad, R.B., Groop, L., 2020. Subtypes of type 2 diabetes determined from clinical parameters. Diabetes 69 (10), 2086–2093.
American Diabetes Association, 2020. Classification and diagnosis of diabetes: standards of medical care in diabetes—2020. Diabetes Care 43 (Suppl. 1), S14–S31.
Chung, W.K., Erion, K., Florez, J.C., et al., 2020. Precision medicine in diabetes: a consensus report from the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetologia 43 (7), 1617–1635.
Harper, L.M., Mele, L., Landon, M.B., et al., 2016. Carpenter-coustan compared with National Diabetes Data Group criteria for diagnosing gestational diabetes. Obstet. Gynecol. 127 (5), 893–898.
Haw, J.S., Galaviz, K.I., Straus, A.N., et al., 2017. Long-term sustainability of diabetes prevention approaches: a systematic review and meta-analysis of randomized clinical trials. JAMA Intern. Med. 177 (12), 1808–1817.
International Association of Diabetes, Pregnancy Study Groups Consensus Panel, Metzger, B.E., et al., 2010. International Association of Diabetes and Pregnancy Study Groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 33 (3), 676–682.
Jagannathan, R., Neves, J.S., Dorcely, B., et al., 2020. The oral glucose tolerance test: 100 years later. Diabetes Metab. Syndr. Obes. 13, 3787.
National Diabetes Data Group, 1979. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. National Diabetes Data Group. Diabetes 28 (12), 1039–1057.
Report of the Expert, 1997. Report of the Expert Committee on the diagnosis and classification of diabetes mellitus. Diabetes Care 20 (7), 1183–1197.
Selvin, E., Wang, D., Matsushita, K., Grams, M.E., Coresh, J., 2018. Prognostic implications of single-sample confirmatory testing for undiagnosed diabetes: a prospective cohort study. Ann. Intern. Med. 169 (3), 156–164.
Vellanki, P., Umpierrez, G.E., 2017. Diabetic ketoacidosis: a common debut of diabetes among African Americans with type 2 diabetes. Endocr. Pract. 23 (8), 971–978.
WHO Expert, 1980. WHO Expert Committee on diabetes mellitus: second report. World Health Organ. Tech. Rep. Ser. 646, 1–80.

Further readings

Ahlqvist, E., Storm, P., Käräjämäki, A., et al., 2018. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol. 6 (5), 361–369.
Anjana, R.M., Baskar, V., Nair, A.T.N., et al., 2020. Novel subgroups of type 2 diabetes and their association with microvascular outcomes in an Asian Indian population: a data-driven cluster analysis: the INSPIRED study. BMJ Open Diabetes Res. Care 8 (1), e001506.
Bergman, M., Abdul-Ghani, M., DeFronzo, R.A., et al., 2020. Review of methods for detecting glycemic disorders. Diabetes Res. Clin. Pract. 165, 108233.
Daryabor, G., Atashzar, M.R., Kabelitz, D., Meri, S., Kalantar, K., 2020. The effects of type 2 diabetes mellitus on organ metabolism and the immune system. Front. Immunol. 11 (1582).
Hillier, T.A., Pedula, K.L., Ogasawara, K.K., et al., 2021. A pragmatic, randomized clinical trial of gestational diabetes screening. N. Engl. J. Med. 384 (10), 895–904.
Kahkoska, A.R., Geybels, M.S., Klein, K.R., et al., 2020. Validation of distinct type 2 diabetes clusters and their association with diabetes complications in the DEVOTE, LEADER and SUSTAIN-6 cardiovascular outcomes trials. Diabetes Obes. Metab. 22 (9), 1537–1547.
Kramer, C.K., Campbell, S., Retnakaran, R., 2019. Gestational diabetes and the risk of cardiovascular disease in women: a systematic review and meta-analysis. Diabetologia 62 (6), 905–914.
Redondo, M.J., Hagopian, W.A., Oram, R., et al., 2020. The clinical consequences of heterogeneity within and between different diabetes types. Diabetologia 63 (10), 2040–2048.

Rose, S.M.S.F., Contrepois, K., Moneghetti, K.J., et al., 2019. A longitudinal big data approach for precision health. Nat. Med. 25 (5), 792–804.
Selvin, E., Crainiceanu, C.M., Brancati, F.L., Coresh, J.J.A., 2007. Short-term variability in measures of glycemia and implications for the classification of diabetes. Arch. Intern. Med. 167 (14), 1545–1551.
Warren, B., Pankow, J.S., Matsushita, K., et al., 2017. Comparative prognostic performance of definitions of prediabetes: a prospective cohort analysis of the Atherosclerosis Risk in Communities (ARIC) Study. Lancet Diabetes Endocrinol. 5 (1), 34–42.

Diabetes mellitus: Dietary management

M. Lu[a], S.J. Pilla[b,c], and S.H. Oh[d], [a] The Johns Hopkins University School of Medicine, Institute for Clinical and Translational Research, Johns Hopkins Hospital Bloomberg Children's Center, Baltimore, MD, United States; [b] School of Medicine, The Johns Hopkins University School of Medicine, Department of Medicine, Division of General Internal Medicine, Baltimore, MD, United States; [c] Welch Center for Prevention, Epidemiology & Clinical Research, Baltimore, MD, United States; and [d] Research Nutrition Core, The Johns Hopkins University School of Medicine, Institute for Clinical and Translational Research, Johns Hopkins Hospital, Baltimore, MD, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of S.H. Oh, R.R. Kalyani, A.S. Dobs, Diabetes Mellitus: Dietary Management, In Encyclopedia of Human Nutrition (Third Edition), edited by Benjamin Caballero, Academic Press, 2013, https://doi.org/10.1016/B978-0-12-375083-9.00070-2.

Overall objectives in the management of diabetes **235**
Control of blood glucose level 235
Prevention or control of comorbidities 236
Minimum intrusion on quality of life 236
Principles of dietary management of diabetes **236**
Medical nutrition therapy for diabetes 236
Nutrition assessment 236
Individualization 236
Nutritional instruction 237
Dietary approaches to diabetes **238**
Total energy intake 238
Distribution of energy intake 238
Different eating patterns for the management of diabetes 238
USDA Dietary Guidelines for Americans 239
Mediterranean-style diet 239
Dietary approaches to stop hypertension (DASH) 240
Low-carbohydrate and very low-carbohydrate 240
Low-fat and very low-fat 240
Vegetarian or vegan 241
Intermittent fasting 241
Supplements: herbal, trace elements, vitamins, and minerals 241
Effects of ingested nutrients on blood glucose **242**
Carbohydrate 242
Sucrose versus complex carbohydrate 242
Glycemic index 242
Non-nutritive sweeteners 244
Protein 244
Fat 245
Special considerations for diabetes subtypes **245**
Type 1 diabetes 245
Major objectives 245
Growth and development 247
Carbohydrate counting 247
Gastroparesis 247
Type 2 diabetes 247
Major objectives 247
Coexisting risk factors 248
Other types of diabetes 248
Diabetes medications and other non-nutrient factors that regulate blood glucose **248**
Insulin 248
Diabetes medications other than insulin 249
Physical activity 250
Stress 250
Dietary prevention and management of comorbidities **250**
Cardiovascular disease 250
Dyslipidemia 250

Hypertension 251
Chronic kidney disease 251
Conclusions 251
References 251
Further reading 251

Key points

- The successful treatment of diabetes mellitus starts with a healthful diet, although specifics will vary depending on the type of diabetes, medications, and individual circumstances and lifestyle.
- There is no one-size-fits-all diet for diabetes, rather approach should be individualized working with a registered dietitian.
- Individualization is a cardinal principle of medical nutrition therapy for diabetes, facilitating individual lifestyle and behavior changes that will lead to improved metabolic control
- Medical nutrition therapy goals include promoting and supporting healthful eating patterns, emphasizing a variety of nutrient-dense foods in appropriate portion sizes, to achieve individualized targets for HbA1c, blood pressure, and cholesterol, achieve body weight goals, and delay and prevent diabetes complications.
- There are many different eating patterns that support a healthy eating style for persons with diabetes, such as the Mediterranean meal pattern, the DASH diet, intermittent fasting, and the myplate method.
- There continues to be no clear evidence of benefit for trace element, herbal or vitamin and mineral supplementation for people with diabetes without underlying deficiencies.

  The successful treatment of diabetes mellitus starts with a healthful diet, although the specifics of the diet may vary depending on the type of diabetes being treated, medications taken and the individual circumstances and lifestyle. Since each person with diabetes will have unique needs and challenges, individualization is the cornerstone of medical nutrition therapy in diabetes treatment. Because approximately 90–95% of all people with diabetes have type 2 diabetes, and 80% of them also have overweight or obesity, weight reduction is often the main therapeutic goal. Many individuals will also require directed treatment of comorbidities such as hypertension or dyslipidemia. People with type 1 diabetes usually require far more attention to exact the quantity and type of carbohydrate they eat and how their food intake aligns with their insulin dosages and physical activity level. In all cases, education of the patient by a trained nutritionist is essential. Diabetes is rarely well controlled unless patients have at least a basic understanding of what they should eat and why.

Overall objectives in the management of diabetes

Control of blood glucose level

The goal of blood glucose control in diabetes is to achieve blood glucose levels that reduce the risk of diabetes complications while minimizing adverse effects of medications. Some individuals with type 2 diabetes can control their blood glucose without needing medications, or cure their diabetes, defined as having blood glucose levels below the diabetes range without needing diabetes medications or treatments for over one year. However, type 2 diabetes is a progressive disease and most patients will need medications to lower blood glucose. Therefore, for most patients with diabetes, the goal is not to normalize blood glucose but to lower it to a level where the harms of hyperglycemia are minimized, balanced against the potential harms of diabetes medications, such as hypoglycemia.

Elevated blood glucose is predominantly asymptomatic. To cause symptoms, hyperglycemia usually must be quite high, averaging >180–200 mg/dL. These symptoms include increased urination (polyuria), thirst (polydipsia), hunger (polyphagia), and blurry vision. In the most severe cases in which hyperglycemia becomes profound and remains unchecked for a long period, it can result in Hyperosmolar Hyperglycemic Syndrome characterized by dehydration, electrolyte imbalances, coma, or death. Most patients with diabetes are diagnosed while asymptomatic on the basis of hyperglycemia on laboratory testing.

Chronic hyperglycemia of any level increases the risk of long-term consequences of diabetes, also called diabetes complications. These complications are divided into microvascular and macrovascular complications: organ damage as a result of damage to small and large blood vessels, respectively. Microvascular complications are retinopathy (eye disease), nephropathy (kidney disease), and neuropathy (nerve damage). Macrovascular complications include heart disease, stroke, and peripheral vascular disease. A combination of neuropathy and vascular disease can result in damage to the extremities which may lead to amputation. Elevated blood glucose levels are more strongly associated with microvascular than macrovascular complications, the latter having a number of other strong metabolic risk factors such as hyperlipidemia, hypertension, and smoking. In studies assessing the benefits of reducing glucose levels to tight control (near-normal levels) versus more permissive levels, there are more significant benefits on microvascular complications, and only small macrovascular benefits. The benefits of tight glycemic control take years to accrue.

To determine the efficacy of treating glycemia, blood glucose must be monitored. There are three ways to assess glucose control: self-monitoring of blood glucose (SMBG), laboratory monitoring of hemoglobin A1c (HbA1c), and continuous glucose

monitoring (CGM). SMBG, done by obtaining a drop of blood and using a small handheld meter, measures the blood glucose at the time the measurement is taken. It is done only as often as necessary to inform decisions about diabetes medication adjustment, or before an insulin dose is administered for patients using insulin. CGM is a device with a small sensor worn on the body that measures blood glucose continuously throughout the day and night. HbA1c is a laboratory test that reflects glycemic control during the previous 60–90 days, and this value is what is most commonly followed for diabetes management in clinical practice. HbA1c less than 5.7% is the normal range. HbA1c of 5.7–6.4% is prediabetes which is mildly elevated glucose with an increased risk of developing diabetes. HbA1c of 6.5% and higher is diagnostic of diabetes. The treatment target for most people with diabetes will be HbA1c less than 7.0%, but this target should be individualized based on a patient's, age, comorbidities, health status, and other factors.

Prevention or control of comorbidities

Morbidity and mortality among people with diabetes are rarely due to acute hyperglycemia or diabetic ketoacidosis. Rather, the long-term complications are either specific to diabetes (e.g., diabetic retinopathy or nephropathy) or accelerated by diabetes (e.g., atherosclerosis). Diabetes significantly increases the risk of coronary artery, cerebrovascular, and peripheral vascular disease, with these cardiovascular complications being the most common cause of death among people with diabetes. Prudent dietary management of diabetes therefore often requires concurrent dietary management of cardiovascular disease risk factors such as hypertension and dyslipidemia. All people with diabetes should follow best practices for dietary control of their comorbidities, such as a low sodium DASH diet in patients with hypertension. In addition, weight loss of 5% or more of initial body weight is recommended for most patients with diabetes and overweight or obesity.

Minimum intrusion on quality of life

For individuals with diabetes, one of the most challenging areas of treatment is determining what to eat and how to incorporate healthy changes. Meals are an integral part of our lives and most patients will not totally abandon their lifetime dietary habits, such as forgoing favorite ethnic flavors and socially available foods. Meal patterns versus "strict diets" enable more flexibility in food choices and preferences. There is no "one size fits all" diet for persons with diabetes as every person has unique personal preferences and lifestyles. Rather, the prescribed diet that intrudes least on a person's quality of life and can be maintained is the most successful nutrition plan. Dietitians should work to identify exactly what changes are required and what favorite dishes, spices, or food groups can be built into a good nutrition prescription.

Principles of dietary management of diabetes

Medical nutrition therapy for diabetes

Medical nutrition therapy (MNT) is an evidence-based process aiming to treat or manage a disease through nutrition. Its components are comprehensive and include assessment of nutritional status and provision of nutritional diagnosis, diet modifications, counseling, and specialized nutrition therapies, provided by a Registered Dietitian Nutritionist (RDN) or a nutrition professional. Systematic reviews have found strong evidence that MNT can effectively improve HbA1c levels, medication use, and quality of life for people with type 1 and type 2 diabetes. Diabetes-focused MNT is usually reimbursed by insurance plans. It is recommended that people with type 1 or type 2 diabetes be referred to diabetes-focused MNT at diagnosis and then as needed throughout the life span in order to achieve optimal treatment goals. The American Diabetes Association recommends specific goals for MNT for persons with varying types of diabetes geared toward more flexibility and patient-centered (**Table 1**).

Nutrition assessment

The first step for planning an appropriate nutrition plan is a full assessment of the patient with diabetes. Every person has unique needs and circumstances, therefore a deep understanding and assessment is key to developing a more successful and individualized management plan. Recommended nutrition assessment components for patients with diabetes are included in **Table 2**.

Individualization

Individualization is a cardinal principle of medical nutrition therapy for diabetes, facilitating individual lifestyle and behavior changes that will lead to improved metabolic control. As no one diet fits all, a standard "diabetic diet" is neither evidence-based nor realistic for the prevention or management of diabetes. Rather, people with diabetes should be evaluated by a registered dietitian or other nutrition specialist who is able to take account of individual variations and provide personalized, diabetes-focused MNT. The nutrition intervention plan should be driven by the diagnosis, pharmacologic treatment, lifestyle, and treatment goals. Important consideration is given to dietary preferences, socioeconomic factors, and the patient's ability to understand and implement instructions. Some patients will need instruction on fine points such as carbohydrate counting and nutrition label reading; others will benefit from broader prescriptions, such as minimizing concentrated sweets and deep fried foods and avoiding

Table 1 Goals of medical nutrition therapy for persons with diabetes mellitus.

- To promote and support healthful eating patterns, emphasizing a variety of nutrient-dense foods in appropriate portion sizes, to improve overall health and specifically to:
  - Achieve individualized targets for HbA1c, blood pressure, and cholesterol
  - Achieve and maintain body weight goals
  - Delay or prevent diabetes complications
- To address individual nutrition needs based on personal and cultural preference, health literacy and numeracy, access to healthful food choices, willingness and ability to make behavioral change, barriers to change.
- To maintain the pleasure of eating by providing nonjudgmental messages about food choices while limiting food choices only when indicated by scientific evidence.
- To provide practical tools for daily meal planning rather than focusing on individual macronutrients, micronutrients or single foods.

Adapted from Lifestyle Management (2019).

Table 2 Recommendations for diabetes nutrition assessment.

Biomedical information
- Type of diabetes, glycemic control, lipid profile, blood pressure
- Presence of key comorbidities: Hypertension, hyperlipidemia, chronic kidney disease, cardiovascular disease, or other chronic illnesses significantly affecting health or function
- Prescriptions, use of over-the-counter medications, complementary or alternative medications
- Adherence to medications

Nutrition-focused physical findings
- Height, weight, BMI, waist circumference
- Weight management goals (weight loss, gain, or maintenance)
- Lower extremity edema
- Injection site for insulin or other injectable medications

History
- Demographic information, education, occupation
- Food insecurity and access
- Cultural dietary preferences
- Family nutrition-related medical history
- Health literacy and numeracy, knowledge, beliefs, attitudes
- Previous experiences with nutrition care services
- Motivation, readiness to behavior changes, self-efficacy
- Physical activity patterns

Food and nutrition-related history
- Food/beverage nutrient intake, including energy intake, portion sizes, meal-snack spacing and patterns, types and amounts of macronutrients, micronutrient intake, alcohol use
- Experience with cooking/preparing food, eating environment, access to healthy foods, frequency of eating out

Adapted from Franz et al. (2017).

sugar-sweetened beverages. **Table 3** presents several examples of nutritional goals that may need accommodation among different people with diabetes.

Nutritional instruction

To achieve successful nutritional outcomes, there are several key educational factors that must be communicated with patients. Basic survival skills for the patient include understanding the relationship between food intake, activity level, and insulin; healthy weight loss goals; healthy dietary patterns for blood glucose, blood pressure and lipids; and a healthy macronutrient balance. Meal planning education should include types and amounts of food, incorporating dietary fiber, understanding serving sizes and nutrition labels, management of eating out and special occasions, as well as incorporating favorite recipes. Patients should also be instructed on how to modify food intake during brief illnesses and changes in food intake based on activity level. For those requiring a more comprehensive approach, carbohydrate counting, nutrition label reading, and self-monitoring blood glucose levels are also vital factors. In addition, patients using high hypoglycemia risk medications (insulin, sulfonylureas, or meglitinides) should be instructed on managing hypoglycemic episodes including the definition of hypoglycemia (blood glucose <70 mg/dL), common symptoms, treatment (ingesting carbohydrates), and avoidance of driving with hypoglycemia.

Table 3 Examples illustrating the variable clinical factors affecting people with diabetes and the resulting diversity of their nutritional needs.

Patient	*1*	*2*	*3*	*4*
Type of diabetes	Type 1	Type 1	Type 2	Type 2
Age (years)	14	38	56	76
Duration of DM (years)	6	26	6	6
BMI	18	23	27	34
Physical activity	Vigorous	Moderate	Mild	Minimal
Prone to hypoglycemia	Yes	Yes	Yes	No
Blood lipids	Normal	Normal	High LDL cholesterol	High TG and low HDL
Blood pressure	Normal	High	Normal	High
Dietary preferences	Likes sweets and snacks	Healthy, little carbohydrate awareness	Spicy foods and irregular meals	Fried foods and sweets
Pharmacologic therapy	Multiple-dose insulin	Multiple-dose insulin	Oral agents plus long-acting insulin	Oral agents
Major nutritional considerations	Adequate caloric intake for growth	Stabilize carbohydrate intake, count carbohydrates	Mildly hypocaloric	Moderately hypocaloric
	Recognize carbohydrate portions, regularize carbohydrate intake, Minimize concentrated sweets	Low salt, high vegetable for hypertension (DASH diet)	Hypolipidemic diet (low saturated fat)	Hypolipemic diet (low saturated fat)
			Regularity of meals, consistency of carbohydrate and fat intake	Low salt, high vegetable for hypertension (DASH diet)
	Learn factors causing hypoglycemia			Control of dietary carbohydrate, especially high-energy concentrated sweets
	Healthy heart diet			

BMI, body mass index; DM, diabetes mellitus; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

Dietary approaches to diabetes

Total energy intake

It is important to start with a personalized total energy requirement calculation. This may be for overall weight management, whether weight loss, maintenance and in some cases, weight gain. As one starts to make changes to their diets, it essential to start with one's total energy requirement or total daily calories. The total energy requirement to maintain constant body weight may be calculated using an estimating equation, taking into consideration the patient's activity level. The weight-maintaining requirement is then adjusted according to the therapeutic objective—to accomplish weight loss, weight maintenance, or weight gain. Examples of how to calculate one's caloric requirements using the Mifflin–St Jeor equation are shown in Table 4. These calculations do not apply to childhood growth and development, pregnancy, malabsorption, or patients with nutritional deficiencies.

Distribution of energy intake

Macronutrients, including carbohydrate, protein, and fat, are the main sources of energy in one's diet. Current literature indicates that there is no ideal macronutrient distribution of an eating pattern for managing diabetes. Individuals with diabetes typically consume 45% of their daily calories from carbohydrate, 16–18% from protein, and 36–40% from fat. This intake pattern is similar to the Acceptable Macronutrient Distribution Ranges (AMDR) recommended by the National Academies for the general public to reduce risk of chronic diseases while getting enough intakes of essential nutrients (Table 5). While the AMDR or other Dietary Reference Intakes (DRIs) serve as general guidelines, there are many healthy eating patterns that fall outside of these specific macronutrient distributions; these are discussed later in this chapter. Any specific recommendations regarding distributions of macronutrients should be individualized based on the patient's therapeutic objectives, personal preferences, and food availability.

Besides macronutrient distribution, the distribution of energy intake throughout the day may also vary. Patients who require insulin, for example, may need a more evenly distributed energy intake, even including a bedtime snack to avoid hypoglycemia while sleeping. For those on fixed insulin doses, it is important to maintain consistent carbohydrate intake to avoid glucose excursions. Reduced or absent energy intake for prolonged periods is dangerous for patients using any glucose lowering medication because this may result in hypoglycemia, though it is most dangerous for patients taking insulin.

Different eating patterns for the management of diabetes

An eating/dietary pattern is a combination of different foods and food groups. There will always be varying expert opinions on which eating pattern is the best. With carbohydrate being the main dietary influence on postprandial blood glucose, some experts

Table 4 Sample calculations of energy requirement in differing circumstances using the Mifflin–St Jeor formula to determine caloric requirements for adults.

Caloric requirements = basal metabolic rate × activity factor
Basal metabolic rate (BMR)
For men: BMR = 10 × weight (kg) + 6.25 × height (cm) − 5 × age (years) + 5
For women: BMR = 10 × weight (kg) + 6.25 × height (cm) − 5 × age (years) − 161
Multiply BMR by one of the following activity factors:
Activity factors
1. Sedentary (little or no exercise): BMR × 1.2
2. Lightly active (light exercise/sports 1–3 days per week): BMR × 1.375
3. Moderately active (moderate exercise/sports 3–5 days per week): BMR × 1.55
4. Very active (hard exercise/sports 6–7 days per week): BMR × 1.725
5. Extra active (very hard daily exercise/sports and physical job or 2x day training): BMR × 1.9

To achieve caloric requirements for weight loss, use the above calculated formula and subtract by 500 calories (this is for a recommended 0.5–1 pound of weight loss per week):
Caloric requirements − 500 calories per day = modified caloric requirements for weight loss

Table 5 Acceptable Macronutrient Distribution Ranges (AMDR) for adults defined by the Institute of Medicine of the National Academies (Institute of Medicine of the National Academies, 2002/2005).

Macronutrient	*AMDR (percent of total daily energy intake)*
Fat	20–35%
Omega-6 polyunsaturated fatty acids	5–10%
Omega-3 polyunsaturated fatty acids	0.6–1.2%
Carbohydrate	45–65%
Protein	10–35%

recommend an eating pattern with a lower carbohydrate and higher fat intake, particularly of unsaturated fatty acids. Others advocate for a lower fat diet for its cardiovascular health benefits. The American Diabetes Association (ADA) and other expert groups support several different healthy eating patterns for managing diabetes. Healthy eating patterns for diabetes in general emphasize non-starchy vegetables and whole foods, and minimizing added sugars, refined grains, and ultra-processed foods. Current evidence does not support any eating pattern to be optimal for diabetes. However, there is evidence suggesting that reducing overall carbohydrate intake, when incorporated in different eating patterns based on individual preferences, is an effective approach to improving glycemia for people with diabetes. Regardless of which eating pattern to follow, one must take into consideration that each nutrient plays a unique role in the body metabolism, and limiting intake of any nutrient to an extreme may result in adverse health consequences. For example, carbohydrates are an important source of fiber, vitamins, minerals, and energy, and fat is an important source of fat-soluble vitamins and essential fatty acids. This is particularly true for omega-3 fatty acids, found in fatty fish such as salmon. Fat and protein intake also have satiating effects that can help with distributing meals during the day. Below, we discuss a list of eating patterns and summarize the evidence of their potential benefits on diabetes management.

USDA Dietary Guidelines for Americans

The Dietary Guidelines for Americans have provided science-based advice on what to eat and drink to promote health, reduce risk of chronic disease and meet the nutrient needs of the general population. The guidelines focus on dietary patterns with the understanding that nutrients and foods are not consumed in isolation. Rather, people consume them in various combinations over time—a dietary pattern. The USDA emphasizes a variety of vegetables as well as fruits, especially whole fruits; grains, at least half of which are whole intact grains; lower-fat diary; a variety of protein foods; and oils. This eating pattern limits saturated fats and *trans* fats, added sugars and sodium. These dietary recommendations can be illustrated by the use of the Plate method shown in Fig. 1. Half the plate should consist of fruits and vegetables, while the other half is divided between whole grains and lean sources of protein, with a serving of low- or non-fat dairy to complete the meal. These general guidelines for macronutrients are easy for patients to remember and are appropriate for adults with diabetes.

Mediterranean-style diet

The Mediterranean-Style diet is based on the traditional diet in the Mediterranean region and is rich in unsaturated fats from plants and seafoods, with olive oil being the primary source of dietary fat. It emphasizes fresh fruits, vegetables, nuts, whole grains, seafoods, with moderate amounts of dairy products (yogurt and cheese), eggs, and wine. It includes very small amounts of red meat and concentrated sugars. Large randomized controlled trials have demonstrated that a Mediterranean-Style diet results in better

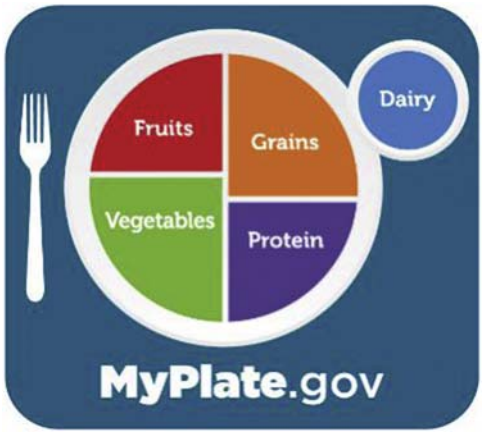


Fig. 1 The USDA Dietary Guidelines for Americans illustrated using the Plate method. Half the plate consist of fruits and vegetables, while the other half is divided between whole grains and lean sources of protein (Dietary Guidelines for Americans, 2020–2025).

glycemic control for people with type 2 diabetes, compared to a lower-fat eating pattern. The Mediterranean-Style diet is also considered a heart healthy diet and has also been shown to reduce the incidence of cardiovascular disease which is a major complication of diabetes.

Dietary approaches to stop hypertension (DASH)

The DASH diet was initially investigated as a dietary intervention to manage high blood pressure. It focuses on foods high in potassium, magnesium, calcium, fiber, and protein (fruits, vegetables, low-fat dairy) while limiting the amount of saturated fats, total fats, cholesterol, and added sugars (red meat, snacks, sweets). Combining the DASH diet with sodium restriction at the national recommended level (2300 mg/day) or at an even lower level (1150 mg/day) results in additional blood pressure lowering. Currently there is limited research on the effect of the DASH diet among people with diabetes. Some evidence suggests that the DASH diet may improve HbA1c, cholesterol levels, and weight loss in people with type 2 diabetes, but long-term, large-scale clinical trials among patients with diabetes are needed to clarify this.

Low-carbohydrate and very low-carbohydrate

The low-carbohydrate eating pattern is the most extensively studied eating pattern for type 2 diabetes due to the direct impact of dietary carbohydrate on postprandial blood glucose and insulin levels. A low-carbohydrate diet refers to an eating pattern with no more than 45% of calories from carbohydrate, typically no more than 100 g of carbohydrate per day. One meta-analysis shows that a low-carbohydrate diet, especially when carbohydrate is limited to 26% or less, can effectively reduce HgbA1c and the need for diabetes medications for people with type 2 diabetes in the short term (<12 months). A very low-carbohydrate diet, also known as a "ketogenic diet", contains even lower amount of carbohydrate, typically less than 50 g per day or 5–10% of total calories, with the majority of calories coming from fat (>70%) and the remaining from protein. Our central nervous system relies on glucose from dietary carbohydrates for fuel. When intake of carbohydrate is extremely low, the body will respond by increasing metabolization of fat and production of ketone bodies as an alternative fuel source (i.e., ketosis). Ketogenic diets were originally developed for the treatment of drug resistant epilepsy, but there is growing evidence suggesting their benefits in glycemic control. One systematic review concludes that a ketogenic diet can be even more effective in type 2 diabetes management in terms of improving blood glucose control, lowering body weight, and reducing medication needs when compared to a low-carbohydrate diet. Due to the carbohydrate-targeting nature of insulin and many antihyperglycemic medications, those who plan to adopt a low carbohydrate or ketogenic eating pattern should consult a provider to adjust insulin and medication dosing to prevent hypoglycemia. Furthermore, a very low-carbohydrate eating pattern may not be recommended to people with chronic kidney disease, disordered eating, and women who are pregnant due to theoretical concerns.

Low-fat and very low-fat

In contrast to a low-carbohydrate eating plan, the low-fat eating pattern focuses on limiting the amount of total fat in one's diet. There is no standard definition for a low-fat eating pattern. In most clinical trials utilizing a low-fat diet intervention, the amount of total fat is less than 30% of total calories, with the majority being lower than 25%. Some evidence suggests that lowering fat intake may help lower the risk of developing diabetes. However, for people with existing type 2 diabetes, most literature indicate that a low-fat eating pattern is not effective at improving glycemia or cardiovascular risk factors. It is also worth noting that most studies of low-fat diets are combined with calorie restriction or weight loss programs, making it difficult to disentangle the effect of dietary pattern from these other factors. A very low-fat diet (Ornish or Pritikin eating patterns) usually consists of only 10% of calories from fat and more than 70% from carbohydrate. Some small-scale, nonrandomized studies suggest that making comprehensive lifestyle changes including eating a very low-fat diet with no calorie restriction may improve fasting glucose levels, weight, blood pressure,

and need for antihyperglycemic medications for people with type 2 diabetes. From a safety standpoint, the World Health Organization and the current dietary reference intakes recommend the minimum fat intake of 20% of total calories to ensure adequate consumption of energy and essential fatty acids, facilitate absorption of fat-soluble vitamins, and preserve HDL cholesterols. Considering the restrictive nature of the eating plan and the lack of evidence, a low-fat or very low-fat eating pattern may not be recommended for the management of diabetes. Low fat and very low-fat diets tend to be very high in carbohydrates since protein intake tends to stay around 15–20% and from the authors' opinion, we would not recommend diets that emphasize greater than 60% of total calories from carbohydrates.

Vegetarian or vegan

A vegetarian eating pattern refers to a diet that does not include meat, fowl, seafood, or products containing these foods. There are several variations of the vegetarian diet, including lacto-ovo-vegetarian (eats dairy products and eggs), lacto-vegetarian (eats dairy but excludes eggs), and vegan vegetarian (excludes all animal-derived foods and food products). One systematic review concluded that vegetarian diets were related to significant reduction in HgbA1C levels in people with type 2 diabetes, compared to an omnivorous diet or a conventional diet for diabetes. Another more recent meta-analysis also suggests that vegetarian diets can effectively improve glycemic control, weight, and non-HDL cholesterol levels in people with type 2 diabetes, supporting the inclusion of such dietary patterns for diabetes management. However, for people who follow a vegetarian or vegan diet, it may be necessary to monitor and supplement certain nutrients including vitamin B-12, vitamin D, calcium, zinc, and long-chain fatty acids as these nutrients are mostly present in animal products.

Intermittent fasting

Intermittent fasting is technically not an eating plan but a way of eating that focuses on the timing rather than the type of foods. Fating means consuming very little to no calories for a period of time. There are different approaches to intermittent fasting, and most common regimens include time-restricted feeding (limiting the "eating window" to 6–8 h every day), alternate day fasting (one day of fasting and one day of ad lib feeding), and periodic fasting (fasting for 1–2 days a week). With the growing interest in intermittent fasting among the public, more researchers have been examining the effects of intermittent fasting on type 2 diabetes. Current literature shows consistent evidence that intermittent fasting can help achieve weight loss (by means of caloric restriction), but there have been mixed results for improvements in HbA1c, insulin sensitivity, and other glycemic control markers. While more research is needed on this topic, intermittent fasting can be recommended to people who wish to achieve desired body weight, as weight loss is also commonly recommended for type 2 diabetes management. While intermittent fasting does not provide guidelines on the type of foods to consume, people can incorporate a healthy eating pattern as discussed above into their intermittent fasting regimen to ensure the quality of their diets. Again, people who intend to follow this eating method should work closely with a professional to match insulin and medication dosing with their eating/fasting periods to achieve optimal glycemic control. Although water and non-caloric liquids are allowed during fasting periods, many people tend to drink less total fluid on their fasting days. Therefore, in order to prevent dehydrate and hypotension, patients should be advised to drink adequate liquids or reduce their diuretics and anti-hypertensive medications during fasting periods.

Supplements: herbal, trace elements, vitamins, and minerals

Herbal and micronutrient, such as vitamin and mineral supplementation are always of interest for the general population. However, there continues to be no clear evidence of benefit for people with diabetes without underlying deficiencies. There has been recurrent interest in whether the routine use of herbal and micronutrients, such as cinnamon, magnesium, vitamin D or chromium, improve glycemia in people with diabetes. There is insufficient evidence of efficacy and possibly concern over long-term safety. It would obviously be attractive if simple oral supplements could facilitate normoglycemia. In addition, since herbal and vitamin and mineral supplements are not regulated, the purity and advertised amounts of the active ingredients of many dietary supplements have been questioned. Further research is needed to adequately establish the role of herbal medicines in diabetes management. The evidence, though, is slim and unconvincing that supplementation of any of these herbal or micronutrient elements has a beneficial effect except when there is a true deficiency. Vitamin and mineral supplementation it is indicated when vitamin deficiency is suspected or likely. For example, populations such as the elderly with poor dietary intake, those pregnant or lactating, strict vegetarians, or those on a calorie-restricted diet may require vitamin supplements. Folate supplementation is well documented to improve the outcome of pregnancy, with or without diabetes. However, there is no clear evidence that supplementation is helpful for those eating an adequate diet. In addition, long-term use of metformin may cause vitamin B12 deficiency, and some experts recommend periodic testing of vitamin B12 levels, particularly in those with anemia or peripheral neuropathy.

In summary, evidence is weak that vitamin or trace element deficiencies occur due to diabetes. Supplementation in more normal circumstances has little or no role in the control of diabetes, and general nutritional guidelines for vitamins and trace elements should be followed. This may be particularly the case with vitamin D, because deficiencies have been reported in many populations due to decreased sun exposure, aging, and lactose intolerance as well as vitamin B12 for those taking metformin.

Effects of ingested nutrients on blood glucose

Carbohydrate

Carbohydrate ingestion causes blood glucose to increase and is one of the major determinants of postprandial glucose levels. In the intestine, dietary carbohydrates are rapidly hydrolyzed into sugars (approximately 75% glucose, 22% fructose and 2% galactose) by specific enzymes. The breakdown of dietary carbohydrates into sugars is so rapid that the rate-limiting step is not how fast these carbohydrates can be digested, but rather how quickly the resulting free sugars can be absorbed into the portal venous system. In people without diabetes, the normal increase in blood glucose in response to a meal is approximately 0.5–2.8 mmol/L (10–50 mg/dL) above baseline, returning to baseline within 1–3 h. The pancreatic hormonal response to dietary carbohydrate mediates the return to normal. Insulin is the central mediator of energy metabolism and is released from the pancreatic beta cells in response to the presence of carbohydrate in the intestine and after glucose passes through the liver into the peripheral circulation. The resulting level of circulating insulin is a signal for the liver to store glucose as glycogen and to reduce glucose production: it also stimulates skeletal muscle to rapidly remove glucose from the circulation, where the glucose is stored as muscle glycogen. The basics of insulin-dependent energy metabolism in the fed and the fasting states are depicted in Fig. 2.

Dietary fiber is a unique type of carbohydrate that is not digested in the small intestine. Rather, it enters the large intestine and gets fermented by the gut microbiota. Since it is not absorbed into the blood stream as glucose, ingestion of dietary fiber will not raise blood glucose levels like other carbohydrates do. There is also evidence suggesting that dietary fiber can slow down the absorption rate of other carbohydrates in the meal, lowering postprandial plasma glucose and plasma insulin in people with type 2 diabetes.

Although carbohydrate intake plays the major role in postprandial blood glucose, there are other factors to consider. Diet is not the only source of glucose in blood; hepatic gluconeogenesis maintains blood glucose in the absence of dietary intake such as when a person is not eating due to an illness. Sick-day instruction is essential for people with diabetes so that they do not simply stop their treatment if they are not eating well, although medication doses may need to be modified. Pharmacologic therapies (insulin or oral agents), of course, also affect blood glucose. A long-standing debate has surrounded the optimal proportion of intake from carbohydrate, fat, and protein. People with diabetes, especially when insulin is administered, will discover that if they hold back carbohydrate their blood glucose does not increase as much. However, holding back carbohydrate unless the diet is hypocaloric, inevitably leads to a high-fat diet, and carbohydrate restriction leaves insulin with no substrate to act on. In our experience, this can cause blood glucose levels to be more unstable, susceptible to swings of hypoglycemia and hyperglycemia. We recommend that for people with type 1 diabetes a lower limit of 130 g of carbohydrate per day unless they are following a low carbohydrate diet under supervision.

Sucrose versus complex carbohydrate

Careful metabolic studies suggest that, gram for gram, sucrose does not increase blood glucose more than complex carbohydrates, either acutely or over a matter of weeks. In these studies, sucrose was isoenergetically substituted for other carbohydrates, mostly under carefully defined research ward conditions in which precise substitutions can be made. Because complex carbohydrates and sucrose are both digested to monosaccharides before they are absorbed, it is not unexpected that each should cause the same glycemic excursion if administered in the same number of grams. It does run counter, however, to the traditional advice that people with diabetes should avoid concentrated sweets.

A number of organizations have cited these research studies in support of a recommendation that allows ingestion of concentrated sweets. The caveat, in the words of the American Diabetes Association, is that "sucrose should be substituted for other carbohydrate sources in the food/meal plan." In the authors' view, there is a practical fallacy in this recommendation: People are unlikely to substitute sucrose for complex carbohydrates in equal amounts. Owing simply to taste, concentrated sweets are likely to be taken in far greater quantity than the more filling and less sweet starches. Thus, in reality, people who routinely eat concentrated sweets are likely to have greater and less predictable glycemic excursions than those who stick to complex carbohydrates. There is also the significant risk that excess concentrated sweet intake will cause weight gain (as well as dental caries). However, if a person with diabetes can include a fixed amount of concentrated sweet in his or her diet and can demonstrate that his or her diabetes is well controlled and the postmeal glycemia is not excessive, there is no reason to deny the person the sweet. The other consideration is that complex carbohydrates are superior sources of fiber, vitamins, and minerals, and therefore, should be the larger portion of carbohydrate intake.

Glycemic index

The glycemic index (GI) is defined as the area under the 2 h curve of blood glucose after the ingestion of a set amount of carbohydrate compared with ingestion of the same amount of carbohydrate from a reference food (white bread or glucose). The glycemic load (GL) is an additional measure in which the amount of carbohydrate in a typical portion is taken into account. A high GI food is considered greater than 70 and a low value is less than 55. For glycemic load values, a high GL is considered >20 and low <10. However, it's important to note that there are a number of factors that can affect the glycemic index and glycemic load of foods. Among these are the fat and fiber content of the meal, cooking method and duration, the patient's absorptive rate, and micronutrient content. Further, it may be difficult to use reported GI and GL for single foods to determine the blood glucose response to meals which contain mixtures of different foods. Below are some basic guidelines to some low, moderate and high glycemic index examples.

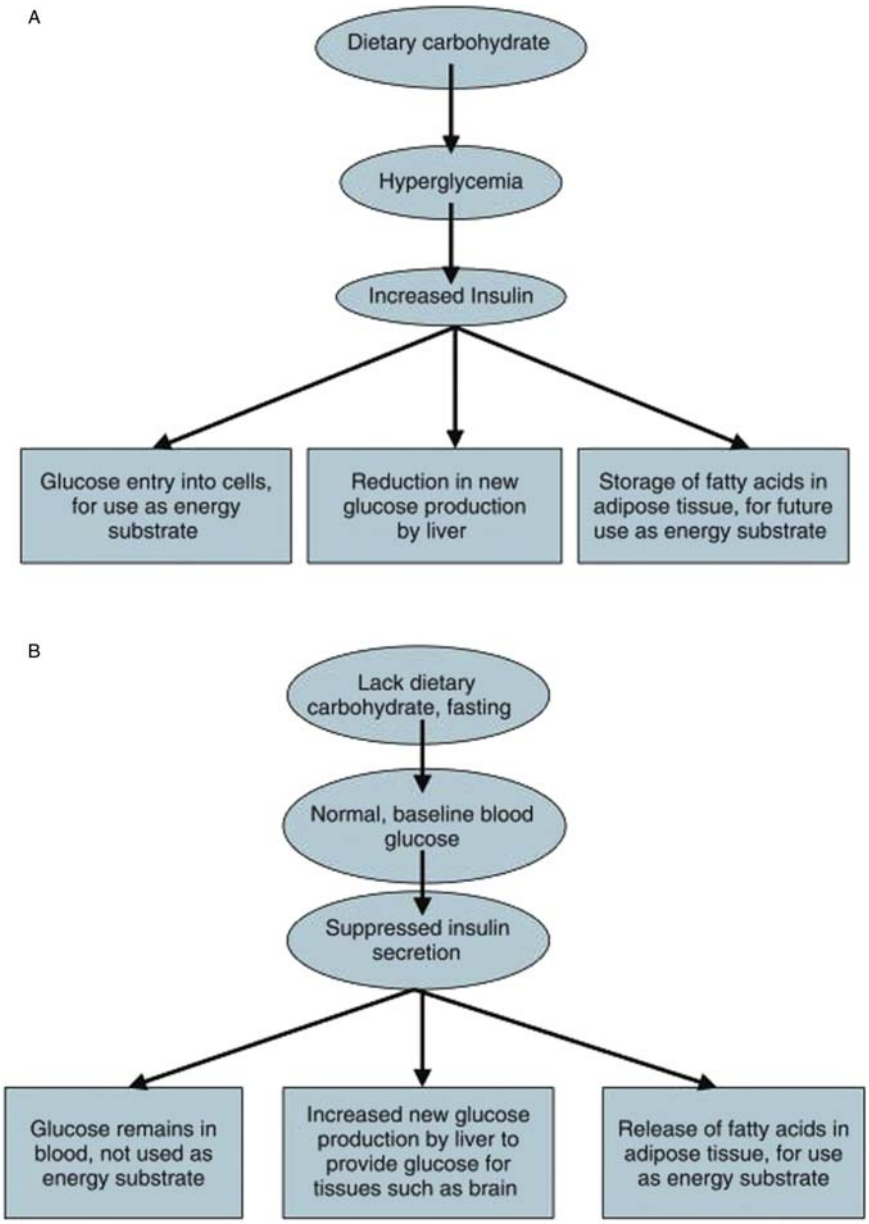


Fig. 2 Influence of insulin on basic energy metabolism. (A) With dietary carbohydrate intake, hyperglycemia induces insulin secretion that acts to enhance glucose entry into cells for utilization as metabolic fuel. Simultaneously, insulin decreases new glucose production in the liver, as dietary glucose is already available and excess stored in liver as glycogen, and stores the remaining excess caloric intake in adipose tissue as fat. (B) With lack of dietary carbohydrate, as in fasting, the reverse occurs: With lower blood glucose, insulin secretion is suppressed. This minimizes entry of glucose into cells but stimulates enough new glucose production from the liver to provide for obligate glucose-using tissues such as the brain. Meanwhile, low insulin concentration promotes fatty acid release from adipose tissue to serve as an alternate fuel for metabolism.

Low glycemic index (GI of 55 or less): Most fruits and vegetables, beans, minimally processed whole grains, whole grain pasta, low-fat dairy foods, and nuts.

Moderate glycemic index (GI 56 to 69): White and sweet potatoes, corn, white rice, couscous, breakfast cereals such as Cream of Wheat and Mini Wheats.

High glycemic index (GI of 70 or higher): White bread, rice cakes, most crackers, bagels, cakes, doughnuts, croissants, most packaged breakfast cereals.

Diets emphasizing low GI/GL have had inconsistent effects on diabetes and cardiovascular risk factors. In the authors' opinion, the concept of GI is valid in a research sense: certain carbohydrates, gram for gram, do raise blood glucose levels more, or with different glycemic patterns, than others. However, basing nutrition plans on the GI and the GL of foods is usually too much of a burden for people with diabetes, and may make it difficult to adhere to a healthy dietary pattern. It is more practical to encourage people to learn their own glycemic response to different foods from experience. Patients with type 1 diabetes may learn, for example, that far more insulin is needed before eating pizza or a bagel; they may learn to avoid certain "desserts." A general awareness of what preferred foods, in what amounts, raise blood glucose may be more practical than memorizing GI or GL.

Non-nutritive sweeteners

Sweeteners are important for the quality of life of people with diabetes. An essential distinction is to differentiate those with from those without significant energy content, i.e., nutritive and non-nutritive sweeteners, respectively. **Tables 6** and **7** provide many of the available non-nutritive and nutritive sweeteners. Non-nutritive sweeteners can be consumed without concern about their effect on blood glucose. They have been determined to be safe when consumed within the daily intake levels established by the Food and Drug Administration (FDA).

Many "diet" sweeteners, such as sorbitol or fructose-based snacks, do cause at least some degree of hyperglycemia. Sugar alcohols (polyols) such as sorbitol, mannitol, and xylitol are classified as hydrogenated monosaccharides, hydrogenated disaccharides, and oligosaccharides. They do contain calories, but because they are only partially absorbed in the small intestine, they have a reduced energy value per gram. A few randomized controlled trials studied gastrointestinal effects of polyols/sugar alcohols and consistently found that in moderate doses of up to 10–15 g/day, polyols/sugar alcohols are tolerated. At high doses (>30 g/day), consumption of some polyols/sugar alcohols (including lactitol, isomalt, and xylitol) may result in significant gastrointestinal discomfort such as increases in flatulence, defecation frequency and loose/watery stools. It is important for people with diabetes to understand these distinctions because many calories can be ingested with foods labeled as "diet" under the false assumption that they are without effect on blood glucose.

Protein

The average protein intake for people of all ages in the United States is between 16% and 18%. In most diets, approximately 50–100 g protein is ingested per day, compared with approximately 200–300 g carbohydrate. Therefore, protein is a calorically less significant part of the diet and far less important in regulating blood glucose. The effect of protein ingestion on blood glucose is far less pronounced than the effect of carbohydrate ingestion. Although protein ingestion affects postprandial blood glucose significantly less than carbohydrate, high-protein diets are not recommended particularly in anyone with diminishing renal function. Those with chronic kidney disease (with albuminuria and/or reduced estimated glomerular filtration rate) should target dietary protein intake at the recommended daily allowance of 0.8 g/kg body weight/day. Reducing the amount of dietary protein below the recommended daily allowance is not advised because it does not alter glycemic measures, cardiovascular risk measures, or the rate at which glomerular filtration rate declines. In people with type 2 diabetes, ingested protein can increase insulin response without increasing plasma glucose concentration. Thus, carbohydrate sources high in protein should not be used to treat or prevent hypoglycemia.

Table 6 Non-nutritive sweeteners.

Type	*US brand names*	*kcal g^{-1}*	*Description*
Saccharin	Sweet and Low, Sweet Twin, Sweet'N Low Brown and Necta Sweet	0	200–700 times sweeter than sucrose; noncarcinogenic and produces no glycemic response
Aspartame	Nutrasweet, Equal and Sugar Twin (blue box)	4	160–220 times sweeter than sucrose; noncarcinogenic and produces limited glycemic response
Acesulfame-K	Sunett, Sweet & Safe, and Sweet One	0	200 times sweeter than sucrose; noncarcinogenic and produces no glycemic response
Sucralose	Splenda	0	600 times sweeter than sucrose; noncarcinogenic and produces no glycemic response
Luo han guo extract	Monkfruit	0	150–300 times sweeter than sucrose; intended as a tabletop sweetener, a food ingredient, and a component of other sweetener blends
Stevia	Stevia	0	250 times sweeter than sucrose general use; heat stable for cooking and baking

Adapted with permission from Fitch et al. (2012).

Table 7 Nutritive sweeteners.

Type	*kcal g^{-1}*	*Description*
Monosaccharide polyols or novel sugars		
Sorbitol	2.6	50–70% as sweet as sucrose; some people experience a laxative effect from a load ≥50 g
Mannitol	1.6	50–70% as sweet as sucrose; some people experience a laxative effect from a load ≥20 g
Xylitol	2.4	As sweet as sucrose
Erythritol	0.2	60–80% as sweet as sucrose; also acts as a flavor enhancer, formulation aid, humectant, stabilizer and thickener, sequestrant, and texturizer
Disaccharide polyols or novel sugars		
Isomalt	2	45–65% as sweet as sucrose; used as a bulking agent
Lactitol	2	30–40% as sweet as sucrose; used as a bulking agent
Maltitol	2.1	90% as sweet as sucrose; used as a bulking agent
Polysaccharide polyols		
HSH	3	25–50% as sweet as sucrose; other names include hydrogenated starch hydrolyzates and maltitol syrup

Adapted with permission from Fitch et al. (2012).

Fat

Dietary fat alone has little, if any, immediate effect on blood glucose concentration because the constituent fatty acids do not produce new glucose and the glycerol moieties are insignificant in their contribution to blood glucose. However, there is considerable evidence that fat reduces early glucose response (first 2–3 h) and delays peak blood glucose due to delayed gastric emptying. The delayed delivery of carbohydrate to the circulation can cause a late, slow postprandial rise in blood glucose. In some cases, fat can lead to late postprandial (>3 h). Evidence suggests that meals containing carbohydrates and that are high in dietary fat cause sustained late postprandial hyperglycemia. This is important to understand, especially for those who consume higher fat diets and are also taking insulin. Insulin dosing may need to be adjusted based on meal composition rather than on carbohydrate content along. People who do not self-monitor frequently are unlikely to be aware of this effect of dietary fat. Although fat has little contribution to blood glucose, persons with diabetes must be aware that fat gram for gram contributes more than twice as many calories as carbohydrates or protein.

Special considerations for diabetes subtypes

Type 1 diabetes

Individuals with type 1 diabetes have essentially no endogenous insulin secretion due to autoimmune destruction of the insulin-producing beta cells of the pancreas. Therefore, patients with type 1 diabetes lose the body's finely tuned insulin secretory mechanism that continually provides insulin "on demand" to precisely match energy intake. This is replaced by administering exogenous insulin given in multiple daily injections, or using an insulin pump, a device that administers insulin subcutaneously on a continuous basis. However, it is difficult to precisely match exogenous insulin to glucose needs, so it is common for patients with type 1 diabetes to have episodes of hypoglycemia or hyperglycemia. Further, missing insulin for even a short period can result in significant hyperglycemia or diabetic ketoacidosis, a life-threatening condition requiring hospitalization. Advances in technology have improved the ability of patients with type 1 diabetes to keep their blood sugar within the normal range, including automated insulin delivery systems (also called artificial pancreas systems) that continuously monitor glucose levels and administer insulin automatically in response. These technologies continue to evolve rapidly and give individuals with type 1 diabetes more control over their disease.

Major objectives

Generally, the treatment objectives in type 1 diabetes are controlling blood glucose within an acceptable range, controlling other cardiovascular risk factors, and thus avoiding long-term complications. This requires close attention not only to diet but also to its interrelationships with insulin dose and timing, physical activity, stress, and other psychosocial factors. This also includes addressing obesity and related diseases which often coexist with type 1 diabetes. Individuals with type 1 diabetes need to learn how to match their caloric intake with their exogenous insulin needs. If carbohydrate, in particular, varies significantly from day to day and meal to meal, the person must learn to adjust insulin doses to match the changed intake. Therefore, comprehensive nutrition education, including how to monitor carbohydrate intake, is recommended for all individuals with type 1 diabetes.

It is useful for the nutritionist to understand the various insulin regimens that are used to treat people with type 1 diabetes. Typically, insulin for individuals with type 1 diabetes is administered as a combination of basal insulin which provides a continuous low level of insulin, and prandial (also called bolus) insulin, which provides a rapid peak in insulin level to cover meals. Insulins that can be used for basal coverage include long acting insulins (insulin glargine, detemir, or degludec) which last for approximately a full day or longer, and intermediate acting insulin (insulin NPH) which lasts for approximately half a day. Rapid acting insulins (insulin aspart, lispro, or glulisine) have a rapid onset of action and last for several hours. Sometimes, regular insulin is used as a rapid acting insulin, this lasts for about 5–8 h. Several different typical regimens, with comments on the dietary implications, are shown in Fig. 3.

In addition to carbohydrate awareness, dietary fat intake should be taken into consideration. Dietary fat is often the main determinant of serum lipids and contributes significantly to total energy intake and thus body weight. It also delays gastric emptying, prolonging the glycemic response to dietary carbohydrate.

Very few people continue to measure and weigh food, but weighing is a useful tool during the instruction phase. Ultimately, people with type 1 diabetes should become proficient in estimating the carbohydrate content of food so that their food selection becomes habitual.

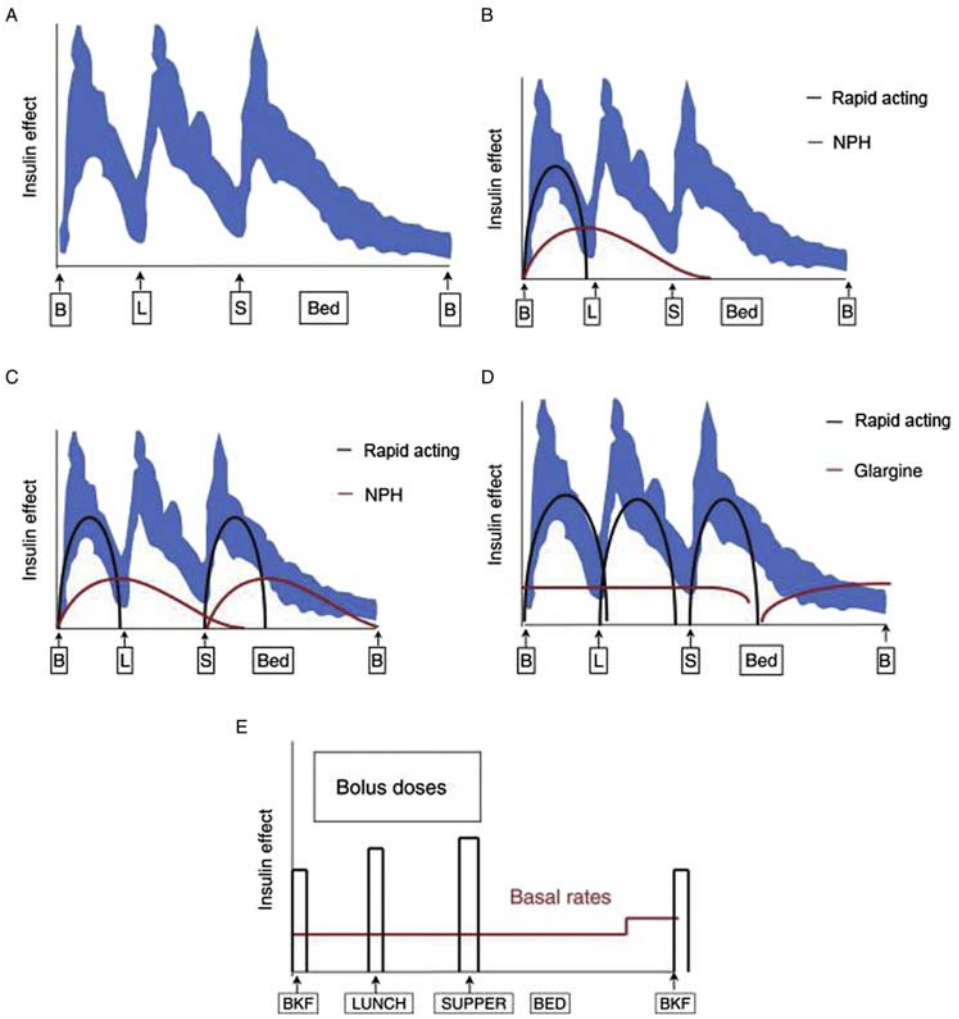


Fig. 3 Aligning insulin regimens with nutritional intake. (A) The normal insulin response to three meals: breakfast (B), lunch (L), and supper (S). Note that insulin increases sharply after ingestion of a carbohydrate-containing meal, declining to baseline within several hours. (B) When a combination of rapid-acting and intermediate-acting insulin given only at breakfast, the normal response to breakfast is reproduced and the intermediate-acting insulin "covers" lunch. (C) When an intermediate-acting insulin is given at breakfast and supper, there is better "coverage" of the supper meal, but the intermediate-acting insulin peaks near bedtime and the middle of the night, so a bedtime snack may be necessary. (D) Long-acting insulin given at bedtime provides "basal" coverage, lasting the full 24 h, while rapid-acting insulin provides mealtime coverage. (E) Use of an external insulin pump infuses insulin at a precise basal rate, and the patient signals the pump to deliver bolus doses of insulin with each meal.

Energy intake distribution will depend on the type of insulin, the number of injections, and the aggressiveness of glycemic targets. Often, small changes in food ingestion can make a significant difference. If, for example, a patient tends to develop hypoglycemia at approximately noon, the skillful dietitian can either emphasize the necessity of eating lunch regularly before noon or suggest the patient consume some of the lunch carbohydrates as an 11 a.m. snack. These changes may eliminate the need to change insulin dose. Alternatively, if weight loss and caloric restriction are desired, then the prebreakfast insulin dose can be decreased. Especially with intensive insulin therapy (three or four daily injections or an external insulin pump), there is some flexibility in the timing of the meals but also a need for more accurate assessment of meal content. Some patients will learn their own ratio of grams of carbohydrate to insulin dose necessary to maintain blood glucose in a good range.

Eating disorders pose a serious problem to the management of type 1 diabetes. Presumably because people with diabetes are often diet conscious, the prevalence of eating disorders is surprisingly high among teenagers with diabetes. The problem is especially dangerous because young people may skip insulin injections in order to induce glucosuria, a sort of "metabolic purging." These conditions clearly require prompt professional help.

Growth and development

The total daily energy intake of a person with type 1 diabetes should be calculated to maintain normal growth and development in a child and normal weight in an adult. Examples of these calculations are provided in **Table 4**. As most people with type 1 diabetes are not overweight, they do not need low-energy diets. In fact, underfeeding is a poor way to maintain blood glucose control. The energy needed to establish and maintain normal weight should be matched with the insulin needed to control glycemia. There is no need for a thin or normal-weight person with type 1 diabetes to be perpetually hungry.

Carbohydrate counting

Carbohydrate counting and calculations using food labels give the patient more freedom and flexibility with more choices. Food labels make the calculation of specific fat and carbohydrate content easier. The trend, therefore, is to emphasize the total carbohydrate in gram amounts or by carbohydrate "choices," where one choice is equal to 15 g of carbohydrate. Examples of 15 g carbohydrate choice include a slice of bread, one-third cup of pasta or rice, or a small apple. If using a food label, one can count the amount of "net carbohydrate" by subtracting fiber from total carbohydrate, as dietary fiber does not affect glycemia. For example, if a food has 22 g of total carbohydrate and 3 g of fiber per serving, this food contains 19 g net carbohydrate. Fat intake should also be addressed with more emphasis on the types of fats, saturated versus mono- and polyunsaturated. This shift in teaching allows for more emphasis on specific carbohydrate and fat awareness rather than lumping mixed foods together in exchanges.

Gastroparesis

An extremely difficult clinical challenge is posed by the patient with diabetic gastroparesis. This condition, a severe autonomic neuropathy reducing gastric motility and gastric emptying time, can sometimes be difficult to diagnose by standardized testing, such as gastric emptying studies. Gastroparesis causes symptoms of early satiety, nausea, vomiting, and abdominal pain, which can result in markedly variable food ingestion. Along with pharmacologic management and good glycemic control, the dietary prescription for diabetic gastroparesis should include small, frequent feeding as tolerated. Rarely, the condition can progress to the point that any oral intake is difficult, and tube feeding or a gastrostomy is required. Fortunately, diabetic gastroparesis tends to wax and wane in severity.

Type 2 diabetes

Approximately 90–95% of all people with diabetes have type 2 diabetes. The prevalence of type 2 diabetes has increased substantially over the last two decades and has paralleled rising rates of obesity. There are two pathophysiologic mechanisms underlying type 2 diabetes: the body's cells are resistant to the action of insulin (insulin resistance), and the pancreas is unable to secrete enough insulin to overcome that resistance (relative insulin insufficiency). There is a complex interaction between these two processes, and their relative contribution to diabetes pathogenesis and progression over time may vary from individual to individual. Insulin resistance is strongly associated with overweight or obesity, although only a minority of people who are overweight or obese have diabetes, so there are clearly other processes at play. Persistent insulin resistance in type 2 diabetes, together with deteriorating pancreatic insulin secretion over time, sometimes results in patients with longstanding type 2 diabetes requiring exogenous insulin therapy. This does not change the diagnosis to type 1 diabetes, which is a disease of entirely different pathogenesis.

Major objectives

The cornerstone of type 2 diabetes treatment is medical nutritional therapy which emphasizes a healthy diet, physical activity, and weight loss for individuals with overweight or obesity. For most patients with diabetes, a goal weight loss of $\geq$5% of their body weight is recommended. It should be noted that to achieve and maintain this weight loss goal, intensive interventions are needed, meaning that patients have frequent (at least monthly) contact over a period of at least several months; shorter and less intensive interventions are may be minimally effective. Intensive lifestyle interventions for patients with type 2 diabetes have demonstrated beneficial effects on weight, blood glucose, kidney function, long-term disability, and other benefits, although no benefits were found on cardiovascular outcomes. The best dietary strategy for accomplishing and maintaining weight loss is unclear and may vary from person to person depending on their individual eating patterns, preferences and nutritional needs. It should also be noted

that weight loss strategies for patients with type 2 diabetes should include consideration of pharmacologic therapy for weight loss and metabolic surgery, the latter being the most effective means of achieving weight loss and, for many individuals, remission of type 2 diabetes.

Coexisting risk factors

Obesity, dyslipidemia, and hypertension are especially prevalent in type 2 diabetes and contribute to cardiovascular risk. Given that the majority of people with type 2 diabetes die of cardiovascular causes, it is critical for medical nutrition therapy for type 2 diabetes to address these cardiovascular risk factors. In fact, most evidence suggests that the management of coexisting risk factors, particularly hypertension, dyslipidemia, and smoking cessation, is more important than the treatment of hyperglycemia in preventing cardiovascular morbidity and mortality. Fortunately, the same intensive lifestyle changes that benefit blood glucose also improve these comorbidities. Medical nutritional therapy should target these comorbidities, such as by using a low sodium DASH diet and moderation of alcohol intake in patients with hypertension. In addition, medical management with antihypertensives and cholesterol lowering medications is often needed to achieve treatment targets.

Other types of diabetes

Gestational diabetes, or elevated glucose that occurs during pregnancy, is relatively common, occurring in nearly 10% of pregnancies, and substantially increases the risk for developing subsequent type 2 diabetes. The management of gestational diabetes has many aspects unique to pregnancy and is not discussed here. Among non-pregnant adults, the vast majority of diabetes is due to type 2 diabetes and type 1 diabetes. Other types of diabetes are uncommon and include pancreatectomy-induced diabetes, diabetes due to chronic pancreatitis, cystic fibrosis, iron infiltration of the pancreas (hemochromatosis), or rare syndromes of insulin resistance. When there is widespread destruction of pancreatic cell mass, as with cystic fibrosis, pancreatectomy, or extensive cancer, the exocrine as well as endocrine functions are affected, leading to malabsorption and impaired glucagon secretion. Malabsorption causes steatorrhea and may require pancreatic enzyme replacement to avoid marked variability in carbohydrate as well as fat absorption.

Diabetes medications and other non-nutrient factors that regulate blood glucose

No element of diabetes management exists in a vacuum, so it is essential to consider how dietary therapy interacts with other elements of glucose regulation and diabetes treatment. Besides glucose-lowering medications, other non-nutrient factors that affect blood glucose include specific medications (e.g., glucocorticoids), acute illnesses, and changes in physical activity. Some of these non-nutrient factors are discussed below.

Insulin

As the two major types of diabetes (type 1 and 2) differ in pathogenesis, so the use of insulin differs for each. In type 1 diabetes the insulin doses must be closely matched to the meals ingested. As described earlier, insulin is administered for multiple purposes: basal insulin provides a continuous low level of insulin, and prandial insulin covers the increase in glucose with a meal. Prandial insulin may be given at prespecified doses (for example, 5 units with each meal) or be based on carbohydrate intake. In addition, patients may administer correctional insulin: rapid acting insulin given to cover high glucose levels, typically per a sliding scale (Table 8). Patients with type 1 diabetes typically will use all three of these types of insulin, for example 15 units of long-acting (basal) insulin at bedtime, and with each meal 1 unit of rapid acting (prandial) insulin per 15 g of carbohydrate plus an additional 1 unit of rapid acting (correctional) insulin for every 30 mg/dL blood glucose above 120 mg/dL. Alternatively, they may use an

Table 8 Example of an insulin regimen for a patient with type 1 diabetes based on blood glucose level before each meal.

		Units of insulin per blood glucose level (mg/dL)					
Meal	*Insulin type*	*<150*	*150–199*	*200–249*	*250–299*	*300–349*	*≥350*
Breakfast	Rapid acting prandial	4	4	4	4	4	4
	Rapid acting SS	0	1	2	3	4	5
Lunch	Rapid acting prandial	4	4	4	4	4	4
	Rapid acting SS	0	1	2	3	4	5
Dinner	Rapid acting prandial	6	6	6	6	6	6
	Rapid acting SS	0	1	2	3	4	5
Bedtime	Long acting	14	14	14	14	14	14

SS, sliding scale.

Table 9 Types of available insulin by onset, peak, and duration of action.

Category	*Insulin generic name*	*Approximate onset*	*Approximate peak (h)*	*Approximate duration (h)*
Rapid acting	Aspart	10–20 min	1	3–5
	Lispro	15–30 min	0.5–2.5	3–6.5
	Glulisine	25 min	1	4–5
Short acting	Regular	30–60 min	2–4	5–8
Intermediate acting	NPH	1–3 h	4–10	12–16
Long acting	Glargine	3–6 h	"Peakless"	11 – >24
	Detemir	3–4 h	3–9	6–23
	Degludec			

insulin pump or automated insulin delivery system. Many people with type 1 diabetes learn to adjust their insulin dose according to both the blood glucose at the start of a meal and the estimated amount that a unit of insulin will reduce their blood glucose—some people may be more sensitive to insulin than others, thus, the relationship between carbohydrate intake and insulin dose needs to be individualized.

For patients with type 2 diabetes, insulin is a medication of last resort. Of all classes of diabetes medications, it is the most dangerous with by far the highest risk for hypoglycemia. Hypoglycemia is a potentially serious complication that is associated with vascular events, falls, cognitive impairment, and mortality. Hypoglycemia also imposes a substantial burden on patients and their family and caregivers, and leads to worse medication adherence and glycemic control. Insulin also imposes significant practical burden on patients from injections, blood glucose monitoring, and sometimes also psychosocial burdens.

Therefore, insulin in type 2 diabetes is generally reserved for situations where blood glucose cannot be lowered by lifestyle change and other classes of diabetes medications, or there are contraindications or barriers to access other diabetes medications. Although nutritionists do not usually prescribe or adjust insulin, it is useful to know the various types of insulin available (Table 9) and patterns of insulin action (Fig. 3). In addition to those listed, there are also premixed insulin products, such as 70% NPH and 30% regular insulin, which provide a mixture of each insulins' action.

Diabetes medications other than insulin

Diabetes medications other than insulin are the primary treatment for type 2 diabetes, and may also be used by some patients with type 1 diabetes who have developed insulin resistance. The number of available non-insulin diabetes medications has expanded dramatically over the past two decades, now including 11 medication classes and over 25 compounds, not to mention combination products and all of the generic and brand names of these products. Some of these medications are oral and others injectable. Here we will give brief overview of each medication class, but for more details about the selection, administration, and monitoring of these medications, refer to a reference such as the American Diabetes Association Standards of Medical Care in Diabetes.

Metformin is the first line medication for all patients with type 2 diabetes, as it has the most evidence of benefiting long-term outcomes. It is an oral medication that acts by decreasing hepatic glucose production and increasing insulin sensitivity. It may cause some modest weight loss, but is mostly weight neutral. Its major contraindication is low renal function, though it may be used safely in many patients with chronic kidney disease. Common side effects are diarrhea and nausea that are often worst when first starting the medication, and can be decrease by using long-acting metformin formulations. Long-term use of metformin may cause vitamin B12 deficiency, so some experts recommend monitoring vitamin B12 levels routinely.

Sulfonylureas (glipizide, glyburide, glimepiride) are older diabetes medications that are commonly used because they are inexpensive, easy to administer, and have few contraindications. They are oral medications that act by stimulating pancreatic insulin secretion. They cause significant weight gain, and their major side effect is hypoglycemia, although the risk is substantially lower than insulin. Of note, they interact with a number of antibiotics that can increase the effective dose of the sulfonylurea and cause hypoglycemia, so it is important to caution patients using sulfonylureas about these interactions.

Thiazolidinediones (pioglitazone, rosiglitazone) are older, effective diabetes medications that are not commonly used due to concerns about rare but serious side effects which include bone fractures and bladder cancer. They are oral medications that act by increasing insulin sensitivity. They are weight neutral and generally well tolerated. They are not recommended in patients with chronic kidney disease or heart failure because they can cause fluid retention.

DPP-4 inhibitors (sitagliptin, linagliptin, other -gliptins) are newer diabetes medications that are very well tolerated but relatively weak in their blood glucose lowering effect. They are oral medications that act in gut hormone pathways to stimulate glucose-dependent insulin secretion and slow gastric emptying. They are weight neutral, have few sides effects, and some can be used without dose adjustment in chronic kidney disease. Unlike other newer classes of diabetes medications, they have not been demonstrated to have cardiovascular or renal protective benefits.

GLP-1 receptor agonists (liraglutide, dulaglutide, other -glutides) are newer diabetes medications that are being used increasingly due their cardiovascular and renal benefits and potential to cause weight loss. They are injectable medications administered daily or weekly that, like DPP-4 inhibitors, act on gut hormone pathways. Some may be used safely in chronic kidney disease. Common side effects include nausea and diarrhea, and they have been associated with increased risk of pancreatitis and gallbladder issues. Of note, liraglutide and semaglutide are also FDA approved for weight loss in patients with and without diabetes and are extremely effective, lowering weight by up to 15% of initial body weight. Therefore, these medications are being used increasingly for the dual purpose of glucose lowering and obesity treatment.

SGLT-2 inhibitors (empagliflozin, canagliflozin, other -flozins) are newer diabetes medications that are being used increasingly due to their cardiovascular and renal benefits. They are oral medications that act by increasing glucose loss in the urine. Calories lost this way are mostly countered by increased appetite but on average they cause modest weight loss. Dose adjustment for chronic kidney disease is required. They are also effective diuretics and have been shown to improve outcomes among patients with heart failure, although they are not approved solely for that purpose in patients without diabetes. Common side effects include genital mycotic infections, volume depletion and hypotension. In some studies, SGLT-2 inhibitors have been associated with rare but serious adverse events including bone fractures, amputations, diabetic ketoacidosis, and Fournier's gangrene (a life-threatening infection of the perineal area).

Other diabetes medication classes are used very infrequently. They are meglitinides, α-glucosidase inhibitors, bile acid sequestrants, dopamine-2 agonists, and amylin mimetics.

Physical activity

The effects of exercise on blood glucose levels are complex and sometimes unpredictable. Although moderate, extended aerobic exercise generally causes progressive lowering of blood glucose, intense exercise may transiently increase the blood glucose. Also, exercise can cause a drop in blood glucose 6–12 h later which may cause hypoglycemia for individuals who are treated with insulin. In general, it is recommended to modify diet to accommodate exercise, rather than changing the dose of diabetes medications, because the duration and intensity of exercise may be unpredictable. However, reducing insulin dosage in anticipation of exercise does work best for many individuals.

Stress

Stress in normal life is difficult to quantify or study, but the usual experience is that blood glucose control deteriorates under stress. This is most likely due to changes in their diet such as eating for stress relief, rather than physiologic changes. Especially for patients with type 1 diabetes, it is important to counsel patients to maintain consistent eating habits in times of stress. It is common for patients with diabetes to have clinical eating disorders, and counseling by behavioral specialists or therapists specializing in eating disorders may be helpful.

Dietary prevention and management of comorbidities

People with diabetes have a higher prevalence than the general population of important chronic conditions that require their own dietary considerations. These conditions include diabetes complications: cardiovascular disease and chronic kidney disease, as well hypertension and dyslipidemia that are risk factors for these complications. Dietary considerations for these comorbidities are discussed briefly in this section; for more detailed information refer to other chapters in this Encyclopedia of Human Nutrition, or other appropriate references mentioned below. In addition, smoking is a major contributor to all of these risk factors and complications, and for patient who use tobacco or e-cigarettes, smoking cessation counseling should be provided and the patient referred to smoking cessation resources.

Cardiovascular disease

For both primary and secondary prevention of cardiovascular disease, dietary approaches are aimed at addressing hypertension and dyslipidemia, the major cardiovascular risk factors discussed below. These risk factors contribute to the development of atherosclerotic cardiovascular disease which is the most common cause of death among people with diabetes. For individuals with failure and resultant propensity for fluid retention, it is important to follow a low sodium diet and avoid dietary indiscretion to sodium which can precipitate a heart failure exacerbation.

Dyslipidemia

Dietary control of dyslipidemia focuses on eating patterns that are low in saturated fats and *trans*-fat, and high in fiber and omega-3 fatty acids. Dietary patterns that achieve this include the Mediterranean diet and the DASH diet, discussed above. Increased physical activity is also important to reduce cholesterol and reduce cardiovascular risk: the American Heart Association recommends 150 min of physical activity of moderate intensity or greater per week for maximal cardiovascular benefit. It is important to note

that genetics explains approximately 40% of individual variation in lipid levels, so there is a limit to what changes in diet and lifestyle can accomplish. Patients should be encouraged to take and adhere to lipid lowering medications if they are indicated as these medications are well tolerated and effective at achieving dramatic lowering of LDL cholesterol and resultant prevention of cardiovascular disease and mortality.

Hypertension

Dietary changes can make a dramatic impact on blood pressure, especially among people with elevated blood pressure, such that diet can reduce or eliminate the need for blood pressure lowering medications for some people. The two major dietary factors that contribute to blood pressure are sodium intake and the overall dietary pattern. A low sodium diet of less than 2300 mg per day of sodium is recommended for people with diabetes and hypertension, and lower sodium intake may be even more effective. In terms of dietary pattern, the DASH diet has been shown to be effective at lowering blood pressure in multiple clinical trials, and is more effective when used in combination with a low sodium intake. Details of the DASH diet are described above.

Chronic kidney disease

Chronic kidney disease is a microvascular complication that occurs in 20%–40% of people with diabetes. Chronic kidney disease is defined and staged according to an individual's level of kidney function (glomerular filtration rate) and the presence and degree of protein loss in the urine. Dietary management of chronic kidney disease is complex, and further details can be found in guidelines specific to chronic kidney disease. Important factors to consider in dietary management are the stage of chronic kidney disease, use of renal replacement therapy (such as hemodialysis), degree of urinary protein loss, electrolyte balance, calcium and vitamin D levels, and use of medications with dietary implications such as phosphate binders or warfarin. For patients who have diabetes and chronic kidney disease stage 3 or greater, but are not on dialysis, a protein-restricted diet of approximately 0.8 g of protein per kg body weight per day is recommended as this has been demonstrated to slow the decline in kidney function. Patients who are on dialysis may need a higher level of protein intake as many patients who are on dialysis struggle with malnutrition. As in patients with heart failure, sodium restriction will be important for patients who have a tendency toward fluid accumulation. As chronic kidney disease progresses, electrolyte disturbances become more common, and in some patients a low potassium diet may need to be implemented to control potassium levels.

Conclusions

Medical nutrition therapy is essential to all people with diabetes, of whatever type or severity. A healthy diet should contain important components, including foods containing carbohydrates from whole grains, fruits, vegetables, vitamins, and low-fat dairy products. Although blood glucose control and management of coexisting risk factors are overall goals, the implementation of dietary management is a highly complex and individualized process. Principles of medical nutrition therapy that generally apply to all diabetes, and specific features that apply to the various types of diabetes, have been discussed; but it must be emphasized that good nutritional management requires the close interaction of each individual patient with a knowledgeable expert in dietetics.

References

Dietary Guidelines for Americans, ninth ed., 2020–2025 https://www.dietaryguidelines.gov/ https://www.myplate.gov/.

Fitch, C., Keim, K.S., Academy of Nutrition and Dietetics, 2012. Position of the Academy of Nutrition and Dietetics: use of nutritive and nonnutritive sweeteners. J. Acad. Nutr. Diet. 112 (5), 739–758. https://doi.org/10.1016/j.jand.2012.03.009. PMID: 22709780.

Franz, M.J., MacLeod, J., Evert, A., et al., 2017. Academy of Nutrition and Dietetics Nutrition practice guideline for type 1 and type 2 diabetes in adults: systematic review of evidence for medical nutrition therapy effectiveness and recommendations for integration into the nutrition care process. J. Acad. Nutr. Diet. 117, 1659–1679.

Institute of Medicine of the National Academies, 2002/2005. Dietary Reference Intakes for Energy, Carbohydrates, Fiber, Fat, Protein and Amino Acids (Macronutrients). The National Academies Press, Washington, DC.

Lifestyle Management, 2019. Lifestyle management: standards of medical care in diabetes—2019. Am. Diabetes Assoc. Diabetes Care 39 (Suppl. 1), S46–S60. https://doi.org/10.2337/dc19-S005.

Further reading

American Diabetes Association, Standards of Care in Diabetes—2021.

Resources for healthy eating plans and recipes

Academy of Nutrition and Dietetics (Diabetes Dietetic Practice Group), Recipes, https://www.dce.org/public-resources/recipes.

American Diabetes Association, Diabetes Food Hub, https://www.diabetesfoodhub.org/.

American Heart Association, Cookbooks, https://www.heart.org/en/healthy-living/healthy-eating/eat-smart/aha-cookbooks.

American Heart Association, Recipes, https://recipes.heart.org/en.

Diabetes mellitus: Etiology and epidemiology

Yara S. Beyh[a] **and K.M. Venkat Narayan**[a,b], [a] Laney Graduate School, Rollins School of Public Health, Department of Global Health, Emory University, Atlanta, GA, United States; and [b] Emory University School of Medicine, Emory University, Atlanta, GA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of J. Sudagani, G.A. Hitman, Diabetes Mellitus: Etiology and Epidemiology, In Encyclopedia of Human Nutrition (Third Edition), edited by Benjamin Caballero, Academic Press, 2013, https://doi.org/10.1016/B978-0-12-375083-9.00071-4.

Introduction 253
Type 1 diabetes mellitus (T1DM) 253
Prevalence 253
Etiology 253
Genetics 254
Environmental/infectious factors 255
Treatments 255
Current status and future 255
Type 2 diabetes mellitus (T2DM) 255
Prevalence 255
Etiology 256
Genetics 256
Heterogeneity 257
Other factors 257
Treatments 258
Future perspectives 259
Other types of diabetes 259
Gestational diabetes mellitus (GDM) 259
Maturity onset diabetes of the young (MODY) 260
Conclusions 261
Acknowledgment 261
References 261

Key points

- Diabetes Mellitus is a chronic multifactorial metabolic disorder characterized by a dysregulation in glucose metabolism leading to hyperglycemia and long-term complications, with a projection to affect 700 million people in 2045.
- Several types of diabetes have been classified by the WHO including T1DM, T2DM, GDM, and Maturity Onset Diabetes of the Young (MODY). Despite the heterogeneity of the disease, they all share hyperglycemia and defective insulin secretion or sensitivity.
- T1DM is an autoimmune disease characterized by the lack of insulin resulting from the destruction of pancreatic β cells on the islets of Langerhans. It is marked by its early onset and the presence of islets-specific autoantibodies (GADA, IA-2A, ZnT8). The key genetic risk factor is located on the Major Histocompatibility Complex (MHC or HLA) known as HLA-DR and HLA-DQ. Other non-MHC genes are involved in the development of T1DM, including *PTPN22*, *IFIH1*, *INS*, *CTLA4*, and *IL2RA*.
- T2DM is the most commonly diagnosed form of diabetes. T2DM is characterized by insulin resistance and/or insulin deficiency, making this disease heterogeneous in nature. The advancement in technologies has allowed recent genetic studies to discover five distinct subgroups of T2DM, which all have hyperglycemia in common, yet have different etiologies, and potentially different prognosis. This discovery will play a major role in advocating for a patient-centered treatment approach to prevent the possible complications.
- GDM is the third common type of diabetes diagnosed during the second half of pregnancy, and usually resolving with delivery. The main cause of GDM is the inability of pancreatic β cells to respond to increased glucose levels and insulin resistance. This form of diabetes is quite similar to T2DM, including the genes involved in its development.
- MODY are a cluster of heterogeneous autosomal dominant inherited non-autoimmune Diabetes Mellitus sharing similarities with both T1DM and T2DM. Given the early-onset of the disease, it is commonly misdiagnosed as T1DM; however, it has its own genetic etiology affecting the HNF gene family.

Introduction

Diabetes Mellitus is a chronic metabolic disorder characterized by a dysregulation of glucose metabolism leading to hyperglycemia and long-term vascular complications. Diabetes is a leading non-communicable disease globally, with over 463 million living with the disease currently and a projected 700 million people affected by 2045 (Saeedi et al., 2019). Despite the heterogeneous etiological and clinical nature of the disease, all existing types share hyperglycemia and defective insulin secretion or sensitivity. The most commonly diagnosed, Type 1 Diabetes Mellitus (T1DM) and Type 2 Diabetes Mellitus (T2DM), are quite distinct in both their epidemiology and etiology (Saeedi et al., 2019). T2DM accounts for 90% of diabetes cases worldwide and has been shown to be heterogeneous with etiologies ranging from insulin deficiency to insulin resistance (Ahlqvist et al., 2018; Saeedi et al., 2019). T1DM is an autoimmune disease characterized by a deficiency in insulin secretion from pancreatic β cells (Dimeglio et al., 2019; Ilonen et al., 2019). Other rarer types of diabetes are diagnosed regularly and have been classified by the World Health Organization (WHO). However, the scope of this review will mainly focus on T1DM and T2DM, with a brief overview on Gestational Diabetes Mellitus (GDM) and Maturity Onset Diabetes of the Young (MODY).

Type 1 diabetes mellitus (T1DM)

Prevalence

According to the International Diabetes Federation, T1DM makes up 10–15% of the global total diabetes prevalence (Katsarou et al., 2017). The prevalence of T1DM has increased at a rate of 2–3% per year (Dimeglio et al., 2019). This increase has been mainly observed in children under the age of 15 years, with a global incidence of 79,000 children per year in 2013 and 90,000 cases in 2015 (Katsarou et al., 2017; Patterson et al., 2014). About 25% of the cases were diagnosed in Europe, 20% in North America and Caribbean, and 6% in Western Pacific (Patterson et al., 2014). In adults aged 15–19 years, the highest incidence was in Estonia (39.9/100,000) and the lowest incidence in Mauritius (1.1/100,000) (Diaz-Valencia et al., 2015). Those differences can be attributed to factors including genetic susceptibility, environmental exposures, and socio-economic status (Dimeglio et al., 2019).

Etiology

T1DM is a chronic disease characterized by the lack of insulin resulting from the destruction of pancreatic β cells on the islets of Langerhans (Dimeglio et al., 2019; Ilonen et al., 2019). T1DM can appear at any age although it is commonly diagnosed in children or young adults before the age of 40 years (Ilonen et al., 2019).

T1DM is mainly characterized by a communication between the endocrine pancreas and the immune system (innate and adaptive) (Dimeglio et al., 2019). The main markers indicating destruction of β cells destruction are the appearance of islet-specific autoantibodies such as glutamic acid decarboxylase autoantibody (GADA), IA-2 antigen (IA-2A), and zinc transporter 8 (ZnT8), in addition to deficiency in insulin production (Ilonen et al., 2019; Katsarou et al., 2017). The subsequent progression to symptomatic T1DM depends on the number of expressed autoantibodies and the age at which the expression occurs. As such, those expressing more than two autoantibodies during the first three years of their lives have a higher risk of progressing into T1DM within 10–20 years (Ilonen et al., 2019; Katsarou et al., 2017). Notwithstanding, some people might express autoantibodies and yet not develop T1DM. Further, at diagnosis, people with T1DM show reduced β cell function, which may slightly improve with treatment over the course of the first year (known as the "honeymoon period") (Dimeglio et al., 2019). It is speculated that the autoimmune trigger is most likely to happen during the *in-utero* phase, when the person is first exposed to those autoantibodies, especially if the mother suffered from T1DM (Ilonen et al., 2019).

T1DM can also arise from insulitis, whereby immune cells infiltrate the pancreas and destroy β cells accordingly. The most commonly involved immune cells are $CD8^+$ T-cells, $CD4^+$ T-cells, B cells, and macrophages. Insulitis is also characterized by the hyperexpression of class I human leukocyte antigen (HLA) in islet cells, which stimulates Interferon synthesis, autoantigen presentation to $CD8^+$ T-cells, and accelerates the death of β cells (Dimeglio et al., 2019; Ilonen et al., 2019). A third mechanism involves the inflammation of the exocrine pancreas and is characterized by an increased expression of T-cells and $CD11c^+$ (Dimeglio et al., 2019; Ilonen et al., 2019). This mechanism, more common in T1DM without elevated levels of autoantibodies, changes the physiology of the pancreatic extracellular matrix and increases vascularization (Dimeglio et al., 2019; Ilonen et al., 2019). The pathogenesis of T1DM is described in **Fig. 1** below.

Another subgroup of T1DM diabetes is Latent Autoimmune Diabetes (LADA). This subgroup typically occurs in adulthood and does not require insulin treatment for the first 6 months following diagnosis. Accordingly, LADA is often regarded as a milder form of an autoimmune response with pathological features that closely resemble T2DM. With that said, LADA is a T1DM subgroup because of its resulting insulitis and expression of IA-2A (Dimeglio et al., 2019; Ilonen et al., 2019; Pozzilli and Pieralice, 2018). On the other end of the spectrum, the genetic susceptibility of LADA seems to be less expressed than it is for T1DM. The HLA, which are very common in the young and adult onset of T1DM, are less frequent in LADA. The prevalence of LADA is higher than originally thought. In fact, LADA cases constitute 2–12% of all adult-onset autoimmune diabetes in adults, and 4–14% of T2DM cases in Europe diagnosed in adulthood are positive for T1DM-associated autoantibodies. The global prevalence of LADA differs based on the inclusion criteria and diagnostic methods applied to identify the disease (Pozzilli and Pieralice, 2018). Hence, unlike T1DM, LADA may be prevented by a healthy lifestyle such as weight regulation, healthy eating, and physical activity.

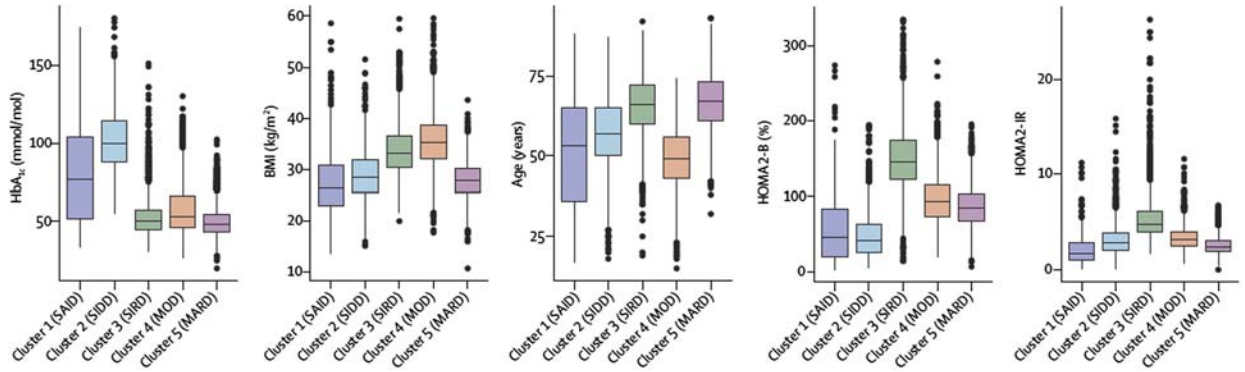


Fig. 1 Five clusters of T2DM showing its heterogeneity. Source: Ahlqvist et al. (2018, The Lancet).

Genetics

T1DM is a polygenic disease affected by both genetics and the environment with twin concordance varying between 30 and 70%, sibling concordance of 6–7%, and a 1–9% chance for children with at least one parent with diabetes (Dimeglio et al., 2019; Katsarou et al., 2017). The key genetic risk factor, accounting for 40–50% of T1DM, is located on the major histocompatibility complex (MHC) on the short arm of chromosome 6 (IDDM-HLA, 6p21), known as HLA-DR and HLA-DQ (Dimeglio et al., 2019; Katsarou et al., 2017, Pociot and Mcdermott, 2002; Redondo et al., 2018). Both DR and DQ consist of alpha and beta heterodimers; DRA and DRB loci encode for the alpha and beta of DR, while DQA1 and DQB1 loci encode for the alpha and beta chains of DQ (Pociot and Mcdermott, 2002). Several haplotypes of those alleles exist, varying by ethnicity, some of which are protective against T1DM (i.e., DR15-DQ6), while the commonly expressed DR3-DQ2 and DR4-DQ8 increase susceptibility to T1DM (Katsarou et al., 2017; Pociot and Mcdermott, 2002; Redondo et al., 2018). HLA alleles are associated with different autoantibodies: DR3-DQ2 was found to be associated with GADA, while DR4-DQ8 has been associated with IA-2A (Robertson and Rich, 2018). The strongest genetic association at the population level has been observed for HLA-DQ1, DQB1, and DRB1. Further, a heterozygotic inheritance of DR3-DQ2 and DR4-DQ8 has a much stronger association with T1DM than the homozygotic inheritance of either one of them (Katsarou et al., 2017; Pociot and Mcdermott, 2002; Redondo et al., 2018).

Studies suggest that the association between HLA and T1DM results from polymorphisms of genes encoding for various amino acids in the peptide-binding section of the HLA molecule. The latter impacts the binding affinity and variety of peptides that can reach T-cells, specifically, position 57 of HLA-DQB1 and position 13 of HLA-DRB1 (Redondo et al., 2018). Of the several classes of the HLA family, HLA class I and the peptide antigen complex play a major role in the development of T-cells selection in the thymus, as well as in the cytotoxicity mediated by antigen-specific T-cells (Pociot and Mcdermott, 2002).

There are also non-MHC genes associated with T1DM, though their effects are not as comparable to MHC genes (Pociot and Mcdermott, 2002; Robertson and Rich, 2018). The strongest T1DM-associated variants encoding candidate genes are *PTPN22*, *IFIH1*, *TYK2*, and *SIRPG* (Redondo et al., 2018; Robertson and Rich, 2018). They serve as important targets for direct variant-protein and function relationships. *PTPN22* on chromosome 1p13 encodes a major down-regulating protein, LYP, which in turn inhibits TCR signal transduction by the dephosphorylation of tyrosine residues on two main proteins, Lek and CD3 (Pociot and Mcdermott, 2002, Prasad and Groop, 2015; Redondo et al., 2018; Robertson and Rich, 2018). Further, functional studies on *PTPN22* have shown various novel frameshift mutations and splicing variants that change between generations. Similarly, functional studies on LYP protein showed that a new isoform of the protein reduced $CD4^+$ T-cells' response to antigens (Robertson and Rich, 2018). *IFIH1* encodes the protein melanoma differentiation-associated protein 5 (MDA5), which binds viruses and mediates the interferon response of the immune system to respond; hence, an insult to this protein reduces the capability to fight viruses and thus increases risk of T1DM (Redondo et al., 2018).

Another essential player is the variable number tandem repeat (VNTR) in the INS promoter on chromosome 11 (11p15.5) (Pociot and Mcdermott, 2002; Redondo et al., 2018; Robertson and Rich, 2018). VNTR INS regulates insulin expression in the thymus, hence modulating the negative selection of T-cells for the insulin-derived peptides. The relationship between the IRS gene and the immune system has been studied in nonobese diabetic mouse (NOD), where cloned diabetes-inducing $CD4^+$ T-cells created antigenic hybrid peptides through cross-linking proinsulin peptides to other peptides in secretory granules of pancreatic beta-cells (Robertson and Rich, 2018). In humans, it has been shown that $CD8^+$ T-cells, encoded by INS mRNA, have the capacity to attack and kill pancreatic beta cells, a common islet autoimmunity pathway leading to T1DM (Robertson and Rich, 2018). The other gene modifying T1DM risk is cytotoxic T-lymphocyte associated protein (*CTLA4*), essential for T-cell development and acts as a negative regulator of cytotoxic T-cells (Pociot and Mcdermott, 2002, Prasad and Groop, 2015; Redondo et al., 2018; Robertson and Rich, 2018). The established relationship is at the post-transcriptional level, where an immunoglobulin protein, Abatacept (CTLA4-Ig), blocks the interaction with CD28 by selectively binding CD80/CD86, which can reduce beta-cell loss in newly diagnosed people (Pociot and Mcdermott, 2002; Redondo et al., 2018; Robertson and Rich, 2018). Other genes and variants associated with T1DM have been discovered, such as Interleukin 2 receptor subunit alpha (*IL2RA*). The latter is essential for

T-regulatory cells' function, and hence, the abnormalities caused by genetic insult to the variant lead to an imbalance between T-regulatory cells and T-effector cells, thus increasing the risk of T1DM (Prasad and Groop, 2015; Redondo et al., 2018). Genetic variants are still being discovered on a daily basis, and their expression and the risk of increased susceptibility differs between populations and ethnicities.

Environmental/infectious factors

Over the years, T1DM diagnosis has overlapped with the seasonal spike in virus infections, mostly peaking during the colder months, as well as reported disease onset with the outbreaks of mumps and rubella (Ilonen et al., 2019). Recently, the focus has shifted to human enterovirus B (echovirus and coxsackie B), peaking late summer and early autumn. The most common explanation is the increased need for insulin, which accelerates the rate at which T1DM develops, especially in children. However, the time of birth may also be key, since exposure to maternal antibodies during fetal growth protects against viral infections and T1DM. Several studies comparing controls to T1DM people who died from the disease found that it is more common to find β-cell stain positive for enterovirus VP1 protein in those who had T1DM. The latter finding indicates that people with T1DM suffer from a low-grade infection in their pancreatic islets. For instance, coxsackie B virus infects monocyte-derived macrophages, which is a type of cells found abundantly in immune cells infiltrates of the pancreatic islets of T1DM patients. A controversial point in this issue is the timing of the viral infection compared to the diagnosis of T1DM. Some have found enterovirus infection to be present at diagnosis, while others, mainly in children, have found the infections to be present around the time of autoantibody seroconversion and not diagnosis (Ilonen et al., 2019). As for the mechanisms leading to β-cell destruction, there are several. The one with the least effect is cytolysis by productive infection, which due to its poor-replicative nature, would lead to functional deficiency in β-cells. Following infection, autoantigens are released from β-cells and transported to pancreatic lymph nodes by macrophages, where they activate autoantigen-specific T-cells. Upon this, $CD8^+$ cytotoxic T-cells are returned to the islets where the virus has increased interferon activity that presents antigens to the cytotoxic cells by increasing the expression of class I HLA, hence the start of the attack on β-cells (Ilonen et al., 2019).

Treatments

Treating T1DM is a complex process with an overall objective to promote glycemic control and a healthy lifestyle to prevent metabolic (i.e., hyperglycemia, hypoglycemia, and ketoacidosis) and longer-term micro- and macro-vascular complications of diabetes.

The most common treatment is insulin with analogues having various onsets and durations of action mimicking pancreatic insulin release (Dimeglio et al., 2019; Katsarou et al., 2017). The only non-insulin medication approved is pramlintide (Dimeglio et al., 2019). Yet, some patients use metformin, glucagon-like receptor agonists, and dipeptidyl peptidase-4 inhibitors. Continuous glucose monitoring (CGM) has been used to closely examine glucose levels throughout the day to titrate dosage and to prevent episodes of hypoglycemia or hyperglycemia. The American Diabetes Association (ADA) alongside other associations, have set targets for HbA1c to 7.5% for the pediatric population and 7% for adults. Those levels can be higher or lower based on the patient's health and overall condition (Dimeglio et al., 2019).

Current status and future

Several studies, such as the Diabetes Prevention Trial (DPP) and the European Nicotinamide Diabetes Intervention Trial (ENDIT) among others, have tried to prevent or delay the development of T1DM in predisposed individuals by administering an early regimen of insulin. Unfortunately, these trials have not yielded beneficial results. However, it is still possible that T1DM is one of the rare diseases that can benefit from primary prevention by enhancing the immune system, and thus preventing insulitis and the start of an autoimmune response. Further understanding of the etiology of T1DM and its treatment and prevention is needed.

Type 2 diabetes mellitus (T2DM)

Prevalence

T2DM makes up more than 90% of diabetes cases, which also happens to be one of the most common non-communicable diseases globally. In 2017, about 462 million people were living with T2DM and this number is expected to go over 600 million by 2040 (Kaiser et al., 2018; Khan et al., 2020; Zheng et al., 2018). The major contributors to this epidemic are the shifts from low to middle income socio-economic status and the accompanying epidemiological and nutritional transitions that countries are experiencing as a consequence of urbanization and industrialization (Zheng et al., 2018). The prevalence of the disease varies widely across ethnicity and geography, with the two main epicenters being China and India. The onset of the disease in these two countries has been linked to a lower BMI and a younger age than high-income Europid countries (Khan et al., 2020; Zheng et al., 2018). Currently, the countries with the highest numbers of people with T2DM are China with 88.5 million, India with 65.9 million, and the US with 28.9 million. However, the countries with the highest prevalence are in the Middle East with a prevalence of 12.8%, while Africa has the lowest prevalence of 4.7% (Khan et al., 2020; Saeedi et al., 2019; Zheng et al., 2018). This large disparity

in prevalence highlights differences in socio-economic development, and the presence of environmental and genetic factors associated with the disease. As such, the prevalence of the disease differs widely between Indigenous native populations and non-Indigenous populations, with a 3–5 times higher risk in indigenous populations, and between global populations (Khan et al., 2020; Ley et al., 2009; Prasad and Groop, 2015; Zheng et al., 2018). As such, Asians Indians have an earlier onset and etiology compared to Pima Indians (Khan et al., 2020; Narayan et al., 2021; Prasad and Groop, 2015; Staimez et al., 2019; Zheng et al., 2018).

Etiology

T2DM is a common multifactorial disease mainly caused by a defect in either insulin secretion or in the feedback loop between insulin secretion by pancreatic β cells and insulin action at the level of peripheral tissues, thus causing hyperglycemia. On one hand, reduced insulin secretion prevents the body from maintaining physiological glucose levels while on the other hand, insulin resistance leads to increased endogenous hepatic glucose synthesis and reduced peripheral (muscle and adipose tissues) glucose uptake. Those can be two separate events occurring at different times or simultaneously. However, their coexistence governs the progression to T2DM through a worsened hyperglycemia (Galicia-Garcia et al., 2020; Zheng et al., 2018). Certain conditions associated with T2DM, like obesity, are characterized by chronic low-grade inflammation, which infiltrates the pancreatic islets, causing ER stress and oxidative stress leading to the loss of integrity of the islets. The excessive presence of free fatty acids and glucose in circulation leads to lipotoxicity and glucotoxicity, which induce ER stress through the unfolded protein response pathways (UPRP), eventually causing β cell damage. The activation of UPRP causes accumulation of misfolded insulin, which increases the production of reactive oxygen species, thus worsening the inflammatory status. These pathological conditions can impair communications between islet cells of the pancreas and peripheral tissue, and contribute to poor regulation of insulin release, ultimately causing β cell impairment (Galicia-Garcia et al., 2020; Zheng et al., 2018). As for insulin resistance, it is characterized by a decreased physiologic response of insulin-sensitive tissues which progresses into a systemic insulin resistance. At the level of the muscles and adipose tissues, insulin resistance will lead to the reduced responsiveness of the glucose transporter GLUT4, hence hindering glucose uptake from circulation. This could also be caused by the defect in the activation of the insulin receptors (IRS-1 and IRS-2) at the surface of peripheral tissues, thus blocking the intracellular downstream signaling cascade leading to T2DM with time. The main role of insulin at the hepatic level is to stimulate glycogen storage and prevent the endogenous glucose synthesis. However, insulin resistance impairs the process, as well as increases lipogenesis and proinflammatory cytokines (leptin and adiponectin) release, which further increases inflammation and may contribute to the destruction of β cells in the long run (Galicia-Garcia et al., 2020). Hence, T2DM is a complex disease requiring the presence of either impaired β cell function or insulin resistance, or both at different or at the same time.

Genetics

Studies have shown that the concordance of T2DM in monozygotic twins (~70%) is higher compared to dizygotic twins (20–30%), highlighting the significant genetic contribution to the disease. Prediabetes traits have been also shown to be inherited, and ethnicity/race play a major role in determining susceptibility (Nair and Baier, 2015; Prasad and Groop, 2015). Nowadays, with the technological advancements and the reduced costs, scientists have been able to better understand the complexity of T2DM. Candidate-gene association studies (CGAS) and genome-wide association studies (GWAS) have shed light on the polygenic nature of the disease, showing its potential population-specific genetic etiology. The key candidate genes in T2DM reported are *KCNJ11*, peroxisome proliferator-activated receptor gamma (*PPARG*), *HNF1B*, *WFS1*, and insulin receptor substrate (*IRS1*). *KCNJ11* encodes Kir6.2, an essential subunit of the ATP-sensitive potassium channel involved in regulating glucose-stimulated insulin release from pancreatic islets. One variant, E23K, has been identified as the most potent and to be the target of the therapeutic drug sulfonylureas (Nair and Baier, 2015; Prasad and Groop, 2015). *PPARG* is mainly expressed in adipose tissues encoding for a type II nuclear receptor (PPAR-γ), involved in regulating fatty acid storage and glucose metabolism. Its variant P12A, causing a substitution of proline for alanine, is the target of the therapeutic drug family thiazolidinediones, aimed at improving insulin action and secretion (Nair and Baier, 2015; Prasad and Groop, 2015). Further, linkage studies have identified genes responsible for the Mendelian forms of T2DM such as MODY, which share a common gene with T2DM called hepatocyte nuclear factor subunit alpha 4 (*HNF4A*) (Nair and Baier, 2015; Prasad and Groop, 2015).

However, the first identified gene *via* linkage studies was cysteine protease calpain 10 (*CAPN10*) on chromosome 2q37, mainly reported in people from Mexican-American descent (Nair and Baier, 2015; Prasad and Groop, 2015). *CAPN10* is involved in glucose-stimulated secretion and action of insulin and β cell function through facilitating the necessary actin-reorganization; hence, its involvement in insulin resistance of T2DM (Meza-Espinoza et al., 2019). Further, an insertion/deletion variant in this gene known as *SNP-19* (*rs3842570*) has been linked to increased levels of glycated hemoglobin and obesity in T2DM people (Meza-Espinoza et al., 2019). Another gene, widely replicated for T2DM is *TCF7L2*, showing a strong linkage signal on chromosome 10q in the Icelandic population. With the exception of Pima Indians, variants of this gene have been found to be related to T2DM in most ethnic populations (mainly European), with an OR of 1.50 in non-Hispanic Whites (Nair and Baier, 2015; Prasad and Groop, 2015). *IRS1* gene encodes a protein responsible for mediating cellular processes by transmitting signals from insulin receptors to downstream signaling pathways. The IRS1 protein is responsible for stimulating autophosphorylation of its β chain, which in turn phosphorylates several downstream molecules to activate insulin action. On the one hand, a mutation in this

gene involving the C allele of rs2943641 has been shown to be associated with insulin resistance. On the other hand, the variants of *IRS1* gene reduce the basal levels of IRS1 protein, which diminishes the induction of IRS1-associated downstream PI3K activity in muscles. Additionally, the substitution of glycine for arginine at position 972 in the molecule, one of the strongest mutations involved in the progression of T2DM, has been linked to insulin resistance in the majority of the population (Yousef et al., 2018). Numerous genes have been discovered thus far *via* linkage studies, GWAS, CGAS, and whole genome sequencing (WGS) in different ethnic groups, most of which have been confirmed to play a role in T2DM predisposition in most ethnicities. With the advancement in techniques and the rise of the GWAS studies, novel approaches using Illumina sequencing uncovered the presence of over 150 single nucleotide polymorphisms (SNPs) mapping to more than 120 loci. Some of the variants discovered are trans-ethnic, and hence affect people all over the globe, while a good part of the discovered variants is ethnicity-specific or population-specific (Nair and Baier, 2015; Prasad and Groop, 2015). As such, minor variants in *GLIS3*, *ITM2-R3HDML-HNF4A*, *ZFAND*, *PSMD6*, *MAEA*, *KCNK16*, and *PEPD* among others, were found to be specific for the East Asian population; variant in *TMEM163* specific for the North Indian origin, and variant in *SLC16A11* in Mexicans (Prasad and Groop, 2015). Despite the novel genetic discoveries in the etiology of T2DM, it explains only less than 20% of the heritability of the disease and its prediction. In fact, T2DM is a multifactorial disease that has genes influenced by the environment, intrauterine environment, and other existing health conditions (Prasad and Groop, 2015). Also, diabetes inheritance relies on which parent is affected by the disease, as such, T2DM has a higher maternal inheritance, while T1DM has a higher paternal inheritance (Prasad and Groop, 2015).

Heterogeneity

Until a few years ago, diabetes was classified as either T1DM or T2DM (as major types). However, increased recognition of early-onset T2DM (younger than 30 years) in several populations (Fagot-Campagna et al., 2000), and growing recognition of T2DM in nonobese populations pointed to the need to explore heterogeneity of the disease (Gujral and Narayan, 2019; Gujral et al., 2018; Narayan, 2016). Thus far, the diabetes classification has relied on the presence or absence of islet-specific autoantibodies and the age of diagnosis, hence more than 80% of the cases are classified as T2DM (Ahlqvist et al., 2018; Ilonen et al., 2019). With additional variables (i.e., *PDX1*, *PAX4*, *INS*, *HNF1A*, *GCK*, etc.) (Lawlor and Stitzel, 2019), Ahlqvist et al. proposed five distinct diabetes subgroups possible (Ahlqvist et al., 2018). Two of the groups have variant traits associated with reduced β cell function and insulin deficiency, while the other three groups have variant traits related to insulin resistance (Ahlqvist et al., 2018; Lawlor and Stitzel, 2019). The first group, severe autoimmune diabetes (SAID), was characterized by an early-onset diabetes exhibiting relatively low BMI, poor metabolic control, insulin deficiency, and presence of GADA. The second group, severe insulin-deficient diabetes (SIID), was similar to SAID except it was GADA negative. The third group, severe insulin-resistant diabetes (SIRD), showed a severe insulin resistance and increased BMI. The fourth group, mild obesity-related diabetes (MOD), exhibited high BMI but not insulin resistance. Lastly, group five, mild age-related diabetes (MARD), was similar to group four, in addition to the late-onset of the disease (**Fig. 1**) (Ahlqvist et al., 2018). The replication of these clusters in certain European cohorts, such as ANDIS, SDR, ANDIU, and DIREVA, showed the same grouping pattern for newly diagnosed and existing diabetes cases (Ahlqvist et al., 2018). These results have been further replicated in several studies, and the results were in line with the initial findings (Bancks et al., 2021; Kanaya et al., 2011, 2014). It appears that the various diabetes phenotypes are present across populations, but their frequency might vary across population; for example, South Asians groups may have a higher frequency of T2DM phenotypes related to impaired insulin secretion. Furthermore, complications and disease progression were assessed for all five groups identified by Ahlqvist et al. This showed that SAID and SIID had higher A_{1c} levels, a stronger indicator for ketoacidosis, while SIRD had the highest risk of non-alcoholic fatty liver disease (NAFLD) and kidney disease (Dennis et al., 2019; Kanaya et al., 2011, 2014). The ADOPT and RECORD trials showed differences in glycemic progression and health outcomes per group, as to the ones found in the original study by Ahlqvist's team (Dennis et al., 2019). Other studies have also been conducted to highlight the differences in the subgroups of T2DM and their complications, such as the MASALA-MESA studies (Bancks et al., 2021; Kanaya et al., 2014). This discovery also allowed linking genetic variants and loci with their corresponding subgroups. As such, variant of gene *TCF7L2* was associated with SID, MOD, and MARD; another variant in gene *TM6SF2* has been associated with both SIRD and the risk of NAFLD; and variant in gene *HLA* has been associated with SAID (Ahlqvist et al., 2018). This is extremely helpful in moving diabetes treatment and diagnosis into a personalized or patient-centered approach to prevent any unwanted complications and delaying progression.

Other factors

There are strong data linking inter-generational factors and diseases—the Developmental Origin of Health and Diseases (DOHaD) theory. The latter postulates that the structure and physiology of organs influenced during fetal growth may adapt to adverse growth conditions predisposing to diseases in adulthood. As such, poor nutrition during pregnancy may lead to structural and functional changes in key organs including the pancreas (Barker, 2004; Vaiserman and Lushchak, 2019). In fact, an unhealthy pregnancy may cause *intra-uterine* growth restriction (IUGR) and dysfunction in pancreatic β cells. The natural response to this gestational insult is the reduced fetal insulin synthesis by building insulin resistance, which serves as a coping mechanism for unfavorable environmental conditions postnatally (Barker, 2004; Vaiserman and Lushchak, 2019). It may also be possible that *intra-uterine* growth restriction might impair beta cell function or beta cell mass. All this has led to the thrifty phenotype hypothesis, which conjectures that any *in utero* insult (i.e., placental dysfunction, stress, and inadequate nutrients etc) may lead to adaptations aiming at an

optimal nutrient storage to maximize metabolic functions. The fetal adaptations, be it reduced peripheral insulin sensitivity or impaired pancreatic development, may accrue into adulthood in the form of reduced insulin secretion or insulin resistance, which serve as the basis of T2DM (Barker, 2004; Vaiserman and Lushchak, 2019). Further, studies have shown a U-shaped association between birth weight, which is impacted by nutritional status during gestation, and risk of T2DM later in life. As such, both low birth weight (LBW) and high birth weight (HBW) increase the risk of T2DM in adulthood. The most likely explanation for the LBW offspring is the catch-up growth in the postnatal period which causes a disproportionate gain of fat to lean mass or impaired insulin secretion. Whereas, the explanation for HBW is the increased risk of overweight/obesity starting childhood, which might grow into adulthood if not addressed early in life (Vaiserman and Lushchak, 2019). Dietary factors, weight, and physical activity also play a role in the risk of developing T2DM. When it comes to weight, it was first thought that only overweight/obesity is associated with T2DM, while in fact the heterogeneity of the disease has shown that increased body fat mass despite a low/normal BMI may be a predisposing factor (Weisman et al., 2018). The changes in dietary patterns across populations into an industrialized diet has contributed tremendously to the increase in weight and fat mass. The industrialized diet pattern relies heavily on sugar sweetened beverages and energy-dense-nutrient-poor foods, mainly based on the ease of accessibility to those items away from home, as well as their reduced prices. Energy-dense foods are generally higher in their fat content, typically saturated fats compared to unsaturated fats. Further, consumption of high glycemic index foods has also been linked to increased T2DM risk, while a diet rich in low glycemic foods and rich in fiber is associated with a lower risk of T2DM (Weisman et al., 2018). Additionally, with economic development and modernization across the globe, people have started to rely more on cars and less on walking, mechanized labor, having more desk jobs, engaging in less physical activity due to their working schedules, and to higher use of screen time. Hence, physical inactivity and sedentary have grown over time, increasing the risk of obesity and insulin resistance and thus the risk of T2DM. In fact, both resistance and endurance exercises have been shown to improve glycemic control in people with diabetes, making it an essential component of diabetes treatment (Weisman et al., 2018). Finally, T2DM disproportionally affects people with lower socio-economic status, due to several factors—access to healthy foods and exercise, stress, etc. Lower education status and lower income increase food insecurity and reduce access to health care, making it more difficult for people to get the medical attention and healthy diet they need to prevent the development of T2DM (Weisman et al., 2018). **Fig. 2** summarizes some of the factors implicated in the risk of developing T2DM among different races and ethnicities.

Treatments

The main objective of treatment is either to prevent or delay T2DM in case of pre-diabetes or mitigate the progression of the condition and its potential complications. The first objective may be achieved through the implementation of proven interventions aimed at a healthy lifestyle, including diet and exercise, or pharmacology (e.g., metformin), while the second is attained by the further addition of pharmacologic treatment to lifestyle interventions, and also attention to blood pressure and lipid control. Diet and

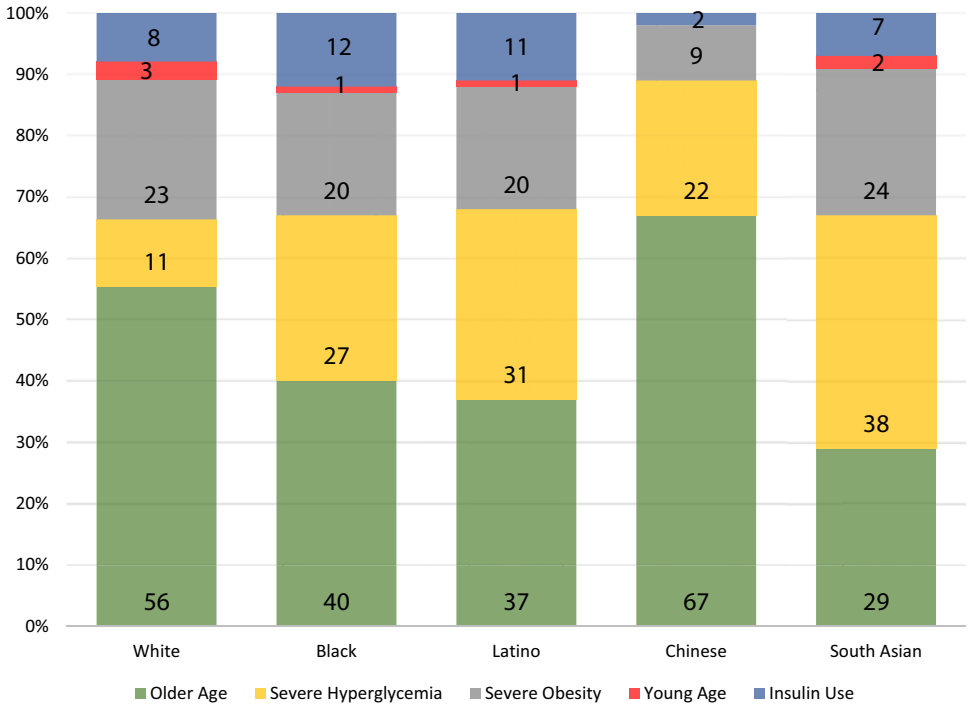


Fig. 2 Factors involved in risk of developing T2DM among different racial and ethnic groups. Source: Bancks et al. (2021, JCEM).

exercise have proven to be very efficient in the prevention and treatment of diabetes; in addition to good quality and duration of sleep (Diabetes Prevention Program (DPP) Research Group, 2002; Marín-Peñalver et al., 2016). In most cases, T2DM is accompanied by excessive weight and adiposity, which are key players in increasing insulin resistance and impairing insulin secretion from pancreatic cells, and in depositing ectopic fat in places like the liver, pancreas, and muscle. Hence, a moderate weight loss (mainly fat) in those who are overweight would be beneficial to achieve optimal levels of glycated hemoglobin and blood glucose (Marín-Peñalver et al., 2016; Pi-Sunyer, 2014). This is affected through adequate caloric and nutrients intake following a personalized plan. However, for those who do not suffer from excessive weight, an isocaloric and balanced diet is still crucial; results observed in several studies including the LookAHEAD trial (Marín-Peñalver et al., 2016; Pi-Sunyer, 2014). Both aerobic and resistance exercises help improve glycemic control, in addition to regulating blood lipids and blood pressure. The recommendation is to engage in at least 150 min of physical activity per week at moderate intensity. However, it is important to note that exercise is contraindicated in case of uncontrolled hyperglycemia or hypoglycemia, as this could damage the peripheries (Marín-Peñalver et al., 2016). One of the most successful lifestyle changes intervention addressing people with diabetes is that which was tested in the DPP trial. DPP was aimed at teaching people how to build a better lifestyle using nutrition, exercise, and behavioral changes. The main two findings of the trial were the superiority of lifestyle changes over oral pharmacological agents and a 58% reduction in the incidence of T2DM (Diabetes Prevention Program (DPP) Research Group, 2002). Several other efficacy and translation trials of interventions like the DPP have been reported and meta-analyzed, as well as other sustainable interventions such as HDC, Alliance, and REACH (Galaviz et al., 2018a,b; Haw et al., 2015, 2017). Pharmacologic treatments for glucose control in people with diabetes are numerous; but the two most commonly prescribed agents are: Biguanides and Sulfonylureas. Metformin is one the most commonly prescribed biguanides as the first treatment of choice for T2DM Metformin has the capacity of reducing intestinal glucose absorption and inhibiting hepatic gluconeogenesis through different mechanisms. Additionally, it increases insulin sensitivity, which opposes insulin resistance (Harrigan et al., 2001; Marín-Peñalver et al., 2016). Since their introduction in the 1950s, Sulfonylureas have been widely used given their stimulation of insulin release from pancreatic β cells and increased action. Their action is possible through the inhibition of an ATP potassium channel, which changes membrane potential by calcium influx to the cell, hence releasing insulin from secretory granules of the pancreas. In some cases, sulfonylureas and biguanides are prescribed in tandem (Harrigan et al., 2001; Marín-Peñalver et al., 2016). Other newer treatments have been introduced such as dipeptidyl peptidase-4 inhibitors (DPP4i) and sodium glucose co-transporter 2 inhibitor (SGLT2i). DPP4i targets the regulation of glucose through blocking incretins—agents that increase insulin secretion and inhibit glucagon. Generally, DPP4i are prescribed for patients with inadequate glucose control through lifestyle modifications, in combination with metformin (Harrigan et al., 2001; Marín-Peñalver et al., 2016). As for SGLT2i, they inhibit glucose renal reabsorption at the level of the proximal convoluted tubule and increase glucose excretion, hence modifying hyperglycemia. In fact, the increased glucosuria aids in weight loss and blood pressure reduction (Harrigan et al., 2001; Marín-Peñalver et al., 2016). It is important to note that despite insulin being the treatment of choice for T1DM, it is prescribed in many cases of T2DM based on onset and duration of action, as it can aid in regulating insulin release from the pancreas, especially for subgroups related to defect in insulin release from the pancreas (Harrigan et al., 2001; Marín-Peñalver et al., 2016). Finally, according to the ADA, the treatment of choice should follow a patient-centered approach based on the coexisting conditions and comorbidities.

Future perspectives

T2DM is a heterogeneous disease affecting populations at different rates and is likely connected to the developmental origin of health and disease together with contemporary lifestyle, reflecting the complexity of the disease. Genetic research, while still in infancy, has led us to better understand the different loci and variants involved in the prognosis and various etiologies of the disease. Despite the countless GWAS, CGAS, and WGS studies, more in-depth investigations in diverse ethnic populations are needed to better understand the differences and similarities, and possible gene-gene and gene-environment interactions. It is also important to note that T2DM is a preventable disease, as it has been shown that the risk can be substantially reduced with successful lifestyle interventions through diet and exercise. Yet, understanding the etiology, probably through genetic profiling, better grasp of phenotypes and their prognosis and treatments remains crucial for a personalized treatment.

Other types of diabetes

Gestational diabetes mellitus (GDM)

GDM is defined as glucose intolerance first diagnosed during the second half of pregnancy (second or third trimester) without previously known diabetes (T1DM or T2DM) (Mack and Tomich, 2017). Recently, the American College of Obstetricians and Gynecologists (ACOG), has released an updated clinical diabetes management guideline in which they included a rarely diagnosed pre-gestational diabetes (American College of Obstetricians and Gynecologists' Committee on Practice Bulletins—Obstetrics, 2018). The latter is defined as diabetes diagnosed in the first half of pregnancy (before week 20 of gestation) based on standard diagnostic criteria. Pre-gestational diabetes is classified as either Type 1 or Type 2, based on the similarities they share with T1DM and T2DM, respectively. Ethnic disparities, screening standards, and diagnostic criteria influence the prevalence of GDM, with the highest being in the Middle-East and North Africa region (15.2%) and the lowest was in European countries (6.1%).

Given the physiological adaptations to meet the requirements of both the mother and the growing fetus, the need for hepatic glucose production increases by 30%. In a normal pregnancy, the pancreatic compensation can match the increased needs to maintain euglycemia. However, in GDM pregnancy, the pancreas is unable to respond to the increased level of glucose, leading to an exhaustion of the pancreatic β cells and development of diabetes (Mcintyre et al., 2019). This is possibly due to the release of hormones opposing insulin action (human placental lactogen, progesterone, placental growth hormone, cortisol); and inflammation through the increased release of tumor necrosis factor (TNF) which interferes with insulin signaling cascade (Mcintyre et al., 2019). GDM shares similar genetic traits with T2DM, as it is seen in the variants on loci of the genes *TCF7L2*, *GCK*, *KCNJ11*, *MTNR1B*, and *CDKAL1*, all of which are major players in regulation of insulin sensitivity and glucose metabolism regulation (Mcintyre et al., 2019).

The main risk factors associated with GDM include maternal age, ethnicity, pre-pregnancy weight, GDM diagnosis in a previous pregnancy, and family history of GDM. It has been proposed that women older than 40 years are twice at risk of developing compared to women younger than 30 years of age (Mcintyre et al., 2019; Szmuilowicz et al., 2019). Overweight or obesity (BMI $\geq$ 25 kg/m^2) before pregnancy, in addition to a sedentary lifestyle during and prior to pregnancy are important risk factors for developing GDM. Furthermore, some environmental (endocrine disruptors and pollutants) and psychosocial (stress, depression, anxiety) factors may contribute to GDM development (Mcintyre et al., 2019; Szmuilowicz et al., 2019). The commonly observed consequences of GDM include increased risk of cesarean delivery, shoulder dystocia, birth injuries due to the large size of the infant, and hypoglycemia of the neonate. Further, in the short-term, the mother may suffer from pre-eclampsia, polyhydramnios, and birth canal lacerations. Whereas in the long-term, women suffering from GDM may be at increased risk of developing T2DM, metabolic syndrome, cardiovascular, kidney, retinal, and liver diseases (Mcintyre et al., 2019).

The main goal of GDM treatment is to maintain fasting glucose levels $\leq$95 mg/dL; 1 hr postprandial glucose $\leq$140 mg/dL; and 2 h post-prandial glucose $\leq$120 mg/dL, while maintaining adequate fetal growth (Mcintyre et al., 2019). This goal could be achieved through Medical nutrition therapy (MNT), which aims at organizing and diversifying food intake to include sufficient energy, macronutrients, micronutrients, and fibers needed during this phase, while keeping blood glucose level well maintained. Additionally, it is recommended to engage in at least 30 min of moderate-intensity aerobic exercise on most days of the week (Mcintyre et al., 2019). When glycemic control is not achieved *via* MNT, pharmacological treatment including insulin or an oral agent (i.e., metformin) is initiated. Insulin is the first treatment of choice, as it does not significantly cross the placental barrier; while the use of oral agents is still controversial with lacking long-term data on the offspring. Post-partum treatment depends on maternal glucose level after delivery, as in most cases, GDM resolves after birth (Mack and Tomich, 2017; Mcintyre et al., 2019).

Maturity onset diabetes of the young (MODY)

MODY are a cluster of heterogeneous autosomal dominant inherited non-autoimmune Diabetes Mellitus. The main features of MODY include its early onset (before the age of 35), and the presence of a family history of autosomal dominant inheritance of diabetes contrasting with both Type 1 and Type 2 Diabetes Mellitus (Naylor et al., 2018). Several genetic etiologies exist for MODY causing a dysfunction in pancreatic β cells starting at birth (Peixoto-Barbosa et al., 2020). The most common defects are GCK-MODY (MODY2) and HNF1A-MODY (MODY3), each accounting for 30–60% of all MODY cases, followed by HNF4A-MODY (MODY1) and HNF1B-MODY (MODY5) constituting 10% of all MODY cases (Naylor et al., 2018; Prasad and Groop, 2015) (**Table 1**). The defective insulin secretion in MODY2 is caused by a mutation to the gene encoding glucokinase (*GCK*), a key enzyme involved in the absorption and metabolism of glucose at the hepatic level, as well as stimulating insulin release from the pancreas (Sternisha and Miller, 2019). MODY1, MODY3, and MODY5, are caused by insults to the hepatocyte nuclear

Table 1 Major subgroups of maturity onset diabetes of the young.

Subgroup	*Gene name*	*Chromosome*	*Frequency*	*Clinical features*	*Frequency of microvascular complications*	*Treatment*
MODY1	*HNF4A*	20q12-q13.1	5–10%	• Large at birth. • Mild neonatal hyperinsulinemia/hyperglycemia. • Progressive insulin secretory defect.	Common	Insulin/sulfonylureas
MODY2	*GCK*	7p15-p13	30–60%	• Mild fasting hyperglycemia at birth. • Generally asymptomatic.	Rare	Exercise/diet
MODY3	*HNF1A*	12q24.2	30–60%	• Mild neonatal hyperinsulinemia/hyperglycemia. • Progressive insulin secretory defect. • Renal glucosuria.	Common	Insulin/sulfonylureas
MODY5	*HNF1B*	17q12	<5%	• Renal anomalies. • Pancreatic hypoplasia. • Intrauterine growth restriction.	Common	Insulin/sulfonylureas

factors (*HNF*) family (4α, 1α, and 1β, respectively), all of which are crucial in the development and function of the pancreatic β cells (Peixoto-Barbosa et al., 2020; Prasad and Groop, 2015). The clinical manifestations associated with MODY differ based on the type and genetic origin of the disease, which ultimately influences the treatment regimen. As such, some of them could be treated by dietary therapy alongside insulin or medication, while others require pure pharmacological treatment. For instance, people suffering from MODY3 can benefit from sulfonylureas, while those suffering from MODY1 are sensitive to this treatment, and those with MODY5 require insulin treatment. As for MODY2 patients, the treatment is not quite clear, and it may be only required during pregnancy depending on the fetus' genotype (Naylor et al., 2018). As for the microvascular complications, they are also dependent on the subgroup of MODY and its severity. Among the most common MODY types, MODY2 is the only one with very rare microvascular complications, while the others (MODY1, MODY3, and MODY5) exhibit them commonly. Those complications are mainly at the level of the kidneys in the form of renal cyst that may eventually lead to renal failure (Naylor et al., 2018).

Conclusions

Diabetes Mellitus is a chronic multifactorial metabolic disorder characterized by a dysregulation in glucose metabolism leading to hyperglycemia and long-term complications, projected to affect 700 million people in 2045. Despite the common traits shared by all major forms of diabetes, T1DM, T2DM, GDM, and MODY, the disease is known for its heterogeneity. T1DM is an autoimmune disease leading to the destruction of the pancreatic β cells of islets of Langerhans; while T2DM can be caused by either an insulin deficiency or resistance, or both, hence the different subgroups. On the other hand, GDM is generally a transient form of diabetes diagnosed during the second half of pregnancy and resolving with delivery; yet, if not well-controlled it could continue into T2DM. Finally, MODY is a cluster of non-autoimmune diabetes, yet, it could be misdiagnosed as T1MD given its early onset and the similarities shared with T1DM. Given the heterogenous nature of the disease and different etiologies, it is important to apply a patient-centered and personalized approach for treatment in the aim of preventing the disease or any major complications that may arise.

Acknowledgment

Research reported in this publication was partially supported by the National Institute Of Diabetes And Digestive And Kidney Diseases of the National Institutes of Health under Award Number P30DK111024 awarded to KMVN. The research reported in this publication was also partially supported by the Nalini and Ravi Saligram Scholarship awarded to YB. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health nor the Nalini and Ravi Saligram institute

References

Ahlqvist, E., Storm, P., Käräjämäki, A., Martinell, M., Dorkhan, M., Carlsson, A., Vikman, P., Prasad, R.B., Mansour Aly, D., Almgren, P., Wessman, Y., Shaat, N., Spégel, P., Mulder, H., Lindholm, E., Melander, O., Hansson, O., Malmqvist, U., Lernmark, Å., Lahti, K., Forsén, T., Tuomi, T., Rosengren, A.H., Groop, L., 2018. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol. 6, 361–369.

American College of Obstetricians and Gynecologists' Committee on Practice Bulletins—Obstetrics, 2018. ACOG Practice Bulletin No. 201: pregestational diabetes mellitus. Obstet. Gynecol. 132, e228–e248.

Bancks, M.P., Bertoni, A.G., Carnethon, M., Chen, H., Cotch, M.F., Gujral, U.P., Herrington, D., Kanaya, A.M., Szklo, M., Vaidya, D., Kandula, N.R., 2021. Association of diabetes subgroups with race/ethnicity, risk factor burden and complications: the MASALA and MESA studies. J. Clin. Endocrinol. Metabol. 106, e2106–e2115.

Barker, D., 2004. The developmental origins of adult disease. J. Am. Coll. Nutr. 23, 588S–595S.

Dennis, J.M., Shields, B.M., Henley, W.E., Jones, A.G., Hattersley, A.T., 2019. Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. Lancet Diabetes Endocrinol. 7, 442–451.

Diabetes Prevention Program (DPP) Research Group, 2002. The diabetes prevention program (DPP): description of lifestyle intervention. Diabetes Care 25, 2165–2171.

Diaz-Valencia, P.A., Bougnères, P., Valleron, A.J., 2015. Global epidemiology of type 1 diabetes in young adults and adults: a systematic review. BMC Public Health 15.

Dimeglio, L.A., Evans-Molina, C., Oram, R.A., 2019. Type 1 diabetes. Lancet 391, 2449–2462.

Fagot-Campagna, A., Pettitt, D.J., Engelgau, M.M., Burrows, N.R., Geiss, L.S., Rodolfo, V., Beckles, G.L.A., Saaddine, J., Gregg, E.W., Williamson, D.F., Narayan, K.M.V., 2000. Type 2 diabetes among north adolescents: an epidemiologic health perspective. J. Pediatr. 136, 664–672.

Galaviz, K.I., Narayan, K.M.V., Lobelo, F., Weber, M.B., 2018a. Lifestyle and the prevention of type 2 diabetes: a status report. Am. J. Lifestyle Med. 12, 4–20.

Galaviz, K.I., Weber, M.B., Straus, A., Haw, J.S., Narayan, K.M.V., Ali, M.K., 2018b. Global diabetes prevention interventions: a systematic review and network metaanalysis of the real-world impact on incidence, weight, and glucose. Diabetes Care 41, 1526–1534.

Galicia-Garcia, U., Benito-Vicente, A., Jebari, S., Larrea-Sebal, A., Siddiqi, H., Uribe, K.B., Ostolaza, H., Martín, C., 2020. Pathophysiology of type 2 diabetes mellitus. Int. J. Mol. Sci. 21.

Gujral, U.P., Narayan, K.M.V., 2019. Diabetes in normal-weight individuals: high susceptibility in nonwhite populations. Diabetes Care 42, 2164–2166.

Gujral, U.P., Weber, M.B., Staimez, L.R., Narayan, K.M.V., 2018. Diabetes among non-overweight individuals: an emerging public health challenge. Curr. Diabetes Rep. 18.

Harrigan, R., Nathan, M., Beattie, P., 2001. Oral agents for the treatment of type 2 diabetes mellitus: pharmacology, toxicity, and treatment. Ann. Emerg. Med. 38, 68–78.

Haw, J.S., Tantry, S., Vellanki, P., Pasquel, F.J., 2015. National strategies to decrease the burden of diabetes and its complications. Curr. Diabetes Rep. 15.

Haw, J.S., Galaviz, K.I., Straus, A.N., Kowalski, A.J., Magee, M.J., Weber, M.B., Wei, J., Narayan, K.M.V., Ali, M.K., 2017. Long-term sustainability of diabetes prevention approaches. J. Am. Med. Assoc. 177, 1808–1817.

Ilonen, J., Lempainen, J., Veijola, R., 2019. The heterogeneous pathogenesis of type 1 diabetes mellitus. Nat. Rev. Endocrinol. 15, 635–650.

Kaiser, A.B., Zhang, N., van der Pluijm, W., 2018. Global prevalence of type 2 diabetes over the next ten years (2018–2028). Diabetes 67.

Kanaya, A.M., Adler, N., Moffet, H.H., Liu, J., Schillinger, D., Adams, A., Ahmed, A.T., Karter, A.J., 2011. Heterogeneity of diabetes outcomes among Asians and Pacific Islanders in the U.S. Diabetes Care 39, 930–937.

Kanaya, A.M., Kandula, N.R., Ewing, S.K., Herrington, D., Liu, K., Blaha, M.J., Srivastava, S., Dave, S.S., Budoff, M.J., 2014. Comparing coronary artery calcium among U.S. South Asians with four racial/ethnic groups: the MASALA and MESA studies. Atherosclerosis 234, 102–107.
Katsarou, A., Gudbjörnsdottir, S., Rawshani, A., Dabelea, D., Bonifacio, E., Anderson, B.J., Jacobsen, L.M., Schatz, D.A., Lernmark, Å., 2017. Type 1 diabetes mellitus. Nat. Rev. Dis. Primers 3.
Khan, M.A.B., Hashim, M.J., King, J.K., Govender, R.D., Mustafa, H., Al Kaabi, J., 2020. Epidemiology of type 2 diabetes—global burden of disease and forecasted trends. J. Epidemiol. Glob. Health 10, 107–111.
Lawlor, N., Stitzel, M.L., 2019. (Epi)genomic heterogeneity of pancreatic islet function and failure in type 2 diabetes. Mol. Metabol. 27, S15–S27.
Ley, S.H., Harris, S.B., Mamakeesick, M., Noon, T., Fiddler, E., Gittelsohn, J., Wolever, T.M.S., Connelly, P.W., Hegele, R.A., Zinman, B., Hanley, A.J.G., 2009. Metabolic syndrome and its components as predictors of incident type 2 diabetes mellitus in an aboriginal community. Can. Med. Assoc. J. 180, 617–624.
Mack, L.R., Tomich, P.G., 2017. Gestational diabetes: diagnosis, classification, and clinical care. Obstet. Gynecol. Clin. N. Am. 44, 207–217.
Marín-Peñalver, J.J., Martín-Timón, I., Sevillano-Collantes, C., Del Cañizo-Gómez, F.J., 2016. Update on the treatment of type 2 diabetes mellitus. World J. Diabetes 7, 354–395.
Mcintyre, H.D., Catalano, P., Zhang, C., Desoye, G., Mathiesen, E.R., Damm, P., 2019. Gestational diabetes mellitus. Nat. Rev. Dis. Primers 5.
Meza-Espinoza, J.P., Leal-Ugarte, E., Peralta-Leal, V., Flores-Villarreal, L.E., Picos-Cardenas, V.J., Sainz-Gonzalez, E., Bocanegra-Alonso, A., Hernandez, J.L., 2019. Association of CAPN10 SNP-19 with metabolic traits in Mexican patients with type 2 diabetes. Int. J. Hum. Genet. 19, 48–53.
Nair, A.K., Baier, L.J., 2015. Complex genetics of type 2 diabetes and effect size: what have we learned from isolated populations? Rev. Diabet. Stud. 12.
Narayan, K.V., Kondal, D., Daya, N., Gujral, U.P., Mohan, D., Patel, S.A., Shivashankar, R., Anjana, R.M., Staimez, L.R., Ali, M.K., Chang, H.H., Kadir, M., Prabhakaran, D., Selvin, E., Mohan, V., Tandon, N., 2021. Incidence and pathophysiology of diabetes in South Asian adults living in India and Pakistan compared with US blacks and whites. BMJ Open Diabetes Res. Care 9.
Narayan, K.M.V., 2016. Type 2 diabetes: why we are winning the battle but losing the war? 2015 kelly west award lecture. Diabetes Care 39, 653–663.
Naylor, R., Knight, J.A., Del Gaudio, D., 2018. Maturity-Onset Diabetes of the Young Overview. University of Washington, Seattle (WA), pp. 1993–2021.
Patterson, C., Guariguata, L., Dahlquist, G., Solte'Sz, G., Ogle, G., Silink, M., 2014. Diabetes in the young—a global view and worldwide estimates of numbers of children with type 1 diabetes. Diabetes Res. Clin. Pract. 103, 161–175.
Peixoto-Barbosa, R., Reis, A.F., Giuffrida, F.M.A., 2020. Update on clinical screening of maturity-onset diabetes of the young (MODY). Diabetol. Metab. Syndrome 12, 1–14.
Pi-Sunyer, X., 2014. The look ahead trial: a review and discussion of its outcomes. Curr. Nutr. Rep. 3, 387–391.
Pociot, F., Mcdermott, M., 2002. Genetics of type 1 diabetes mellitus. Gene Immun. 2, 235–249.
Pozzilli, P., Pieralice, S., 2018. Latent autoimmune diabetes in adults: current status and new horizons. Endocrinol. Metabol. 33, 147–159.
Prasad, R.B., Groop, L., 2015. Genetics of type 2 diabetes—pitfalls and possibilities. Genes 6, 87–123.
Redondo, M.J., Steck, A.K., Pugliese, A., 2018. Genetics of type 1 diabetes. Pediatr. Diabetes 19, 346–353.
Robertson, C.C., Rich, S.S., 2018. Genetics of type 1 diabetes. Curr. Opin. Genet. Dev. 50, 7–16.
Saeedi, P., Petersohn, I., Salpea, P., Malanda, B., Karuranga, S., Unwin, N., Colagiuri, S., Guariguata, L., Motala, A.A., Ogurtsova, K., Shaw, J.E., Bright, D., Williams, R., 2019. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International diabetes federation diabetes atlas. Diabetes Res. Clin. Pract. 157 ninth ed.
Staimez, L.R., Deepa, M., Ali, M.K., Mohan, V., Hanson, R.L., Narayan, K.M.V., 2019. Tale of two Indians: heterogeneity in type 2 diabetes pathophysiology. Diabetes Metab. Res. Rev. 35.
Sternisha, S.M., Miller, B.G., 2019. Molecular and cellular regulation of human glucokinase. Arch. Biochem. Biophys. 663, 199–213.
Szmuilowicz, E.D., Josefson, J.L., Metzger, B.E., 2019. Gestational diabetes mellitus. Endocrinol. Metab. Clin. N. Am. 48, 479–493.
Vaiserman, A., Lushchak, O., 2019. Developmental origins of type 2 diabetes: focus on epigenetics. Ageing Res. Rev. 55.
Weisman, A., Fazli, G.S., Johns, A., Booth, G.L., 2018. Evolving trends in the epidemiology, risk factors, and prevention of type 2 diabetes: a review. Can. J. Cardiol. 34, 552–564.
Yousef, A.A., Behiry, E.G., Abd Allah, W.M., Hussien, A.M., Abdelmoneam, A.A., Imam, M.H., Hikal, D.M., 2018. IRS-1 genetic polymorphism (r.2963G>A) in type 2 diabetes mellitus patients associated with insulin resistance. Appl. Clin. Genet. 11, 99–106.
Zheng, Y., Ley, S.H., Hu, F.B., 2018. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat. Rev. Endocrinol. 14.

Diarrheal diseases

Phoebe Hodges[a] **and Paul Kelly**[a,b], [a] Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom; and [b] Tropical Gastroenterology and Nutrition Group, University of Zambia School of Medicine, Lusaka, Zambia

© 2023 Elsevier Ltd. All rights reserved.

This is an update of K. Zaman, A.H. Baqui, Diarrheal Diseases, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 47–49, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00073-8.

Introduction 263
Pathophysiology of diarrhea 264
Global overview 264
Contribution of diarrhea to malnutrition 264
Infectious causes of diarrheal disease 266
Bacterial 266
Cholera 267
Viral 267
Parasites 267
Human immunodeficiency virus (HIV) 267
Intestinal tuberculosis 268
Whipple's disease 268
Tropical sprue 268
Non-infectious causes 268
Inflammatory bowel disease 268
Celiac disease 268
Pancreatic exocrine insufficiency 269
Intestinal failure (often due to short bowel syndrome) 269
Diagnosis 269
Prevention and treatment 269
Conclusions 270
References 270
Further reading 270
Relevant website 271

Key points

- To give an overview of the global burden of diarrheal diseases
- To discuss the relationship between diarrheal disease and childhood malnutrition
- To give a brief summary of the causes of diarrhea, both infectious and non-infectious
- To discuss the treatment and prevention of diarrheal disease and provide a summary of the current evidence for public health interventions aimed at reducing the burden of diarrheal diseases in low-income countries

Introduction

Diarrhea is commonly defined as an increase in the frequency of bowel movement with increased fluidity of stool relative to usual patterns in an individual. There is a wide variation in the literature in the definition of a diarrhea episode, ranging from mothers' perceptions to operational definitions based on the frequency of loose stools in 24 h. The operational definitions vary but require at least three watery or loose stools in 24 h.

Diarrhea can be classified in a number of ways: duration of an episode (acute duration—less than 14 days; persistent—14 days or more), type of stool (watery or bloody), consistency of stool (loose, liquid, and watery), dehydration status (dehydrated, nondehydrated), pathophysiologic mechanism (inflammatory, secretory, osmotic), or based on causative agents (e.g., *Escherichia coli*, cholera, rotavirus, *Shigella*, etc.). Dysentery refers to acute bloody diarrhea resulting from infective colitis with bleeding from ulceration. In terms of overall burden of diarrhea, acute infectious diarrhea is by far the greatest contributor as almost every human being experiences it at some point in a lifetime.

Pathophysiology of diarrhea

During diarrhea there is an increased frequency of bowel movement and alteration in stool consistency. Normal fluid and solute movement across the intestinal membrane is altered resulting in increased loss of water and electrolytes (particularly sodium, potassium, chloride, and bicarbonate). Four mechanisms have been postulated to be responsible for these alterations: decreased fluid absorption, increased intestinal secretion, increased luminal osmolality, and altered intestinal motility. Inflammatory diarrhea probably results from a combination of all four. It has been suggested that a common mechanism exists in cases of bacterial, viral, and parasitic diarrhea where levels of cyclic nucleotide are enhanced. Loss of fluids and electrolytes may lead to dehydration, acidosis, hyponatremia, and hypokalaemia. The degree of dehydration depends on the net amount of fluid lost. World Health Organization (WHO) classifies the grades of dehydration as none, some, or severe dehydration. The estimated fluid deficits in the different types of dehydration are no dehydration (not enough signs to classify as some or severe dehydration), some dehydration (two or more of the following: restlessness, irritability; drinks eagerly, thirsty; sunken eyes), and severe dehydration (two or more of the following: lethargy/unconsciousness; sunken eyes; unable to drink/drinks poorly; skin pinch goes back very slowly). However, the most reliable measure of volume depletion in an adult is postural hypotension, and this important sign should be sought in all but the most trivial cases.

Global overview

Diarrhea remains the second leading cause of death in under-5s globally, with over 0.5 million children dying annually of diarrheal disease. The majority of these deaths are in children under the age of 2. Children in low-income countries (LICs) under the age of 3 have on average 3 episodes of diarrhea annually, which may have implications for nutritional deficits and growth faltering. The Global Enteric Multicenter Study (GEMS), which analyzed the burden and etiology of diarrheal disease in children across seven sites in Africa and Asia, found that the majority of cases of moderate-to-severe diarrhea were caused by just four pathogens: rotavirus; *Cryptosporidium*; enterotoxigenic *Escherichia coli* producing heat-stable toxin; and *Shigella*, and that children with moderate-to-severe diarrhea were significantly more likely to have linear growth faltering during the follow-up period (Kotloff et al., 2013). However, the carriage of enteropathogens does not necessarily lead to diarrhea, and it is likely that this is a reflection of the balance between the intensity of infection, often itself a consequence of size of inoculum, and host defense. A recent study in Zambia showed that children with stunting excreted up to 11 enteropathogens concurrently despite not having intercurrent diarrhea. These children also had a severe degree of enteropathy demonstrated on small bowel biopsies and this is likely to represent an adaptive response to constant enteropathogen exposure with the outcome of reduced microbial translocation (Amadi et al., 2021). The aforementioned enteropathy is known as environmental enteropathy or environmental enteric dysfunction and refers to abnormalities of small intestinal structure and function seen in populations living in unsanitary conditions. Changes include villous blunting, crypt hypertrophy, epithelial breaches, mucosal inflammation and increased mucosal permeability, resulting in systemic inflammation and nutrient malabsorption. This lesion in itself plays a central role in childhood growth failure. These data from Zambia are consistent with data from the Malnutrition and Enteric Disease Study (MAL-ED), which set out to assess the role of enteropathogen burden and diarrhea in childhood growth at eight sites across South America, Asia and sub-Saharan Africa. This study also showed high numbers of enteropathogens being carried by asymptomatic children in association with slower linear growth (MAL-ED Network Investigators, 2017).

Contribution of diarrhea to malnutrition

It has long been thought that diarrheal disease has a significant impact on child growth and development, however, recent evidence suggests that it is frequent exposure to enteropathogens as opposed to frequency of diarrheal disease that affects growth. Although the negative short-term association between diarrheal episodes and weight-for-age is widely accepted, several studies in recent years have examined the relationship between childhood diarrhea and nutrition in more detail, particularly with regards to longer-term impact on linear growth, with varied results. Data from multi-study analyses suggest that there is a multiplicative effect on growth with each diarrheal episode/day of diarrhea over the first two years of life. A pooled analysis of 1393 children from 9 studies across 5 countries (Peru, Brazil, Ghana, Guinea-Bissau and Bangladesh) found that the odds of stunting at 24 months of age increased by 1.13 when cumulative diarrheal incidence over the first 24 months of life increased by 5 episodes (Checkley et al., 2008). A more recent analysis of 7 longitudinal cohort studies from Peru, Brazil, Guinea-Bissau and Bangladesh, where data on diarrhea and anthropometric measures were available, found that a child with the average age-specific monthly diarrhea burden (equivalent to 23 days/year) was 0.38 cm shorter at 24 months than a child with no diarrhea. This deficit was reduced to 0.2 cm if the number of days of diarrhea/year was halved (Richard et al., 2013).

The MAL-ED study found no significant long-term relationship between diarrheal morbidity and linear growth, however, higher rates of enteropathogen detection in non-diarrheal stools were found to have a negative impact on linear growth. *Campylobacter* in particular was found to be negatively associated with growth rates for both length and weight (MAL-ED Network Investigators, 2017).

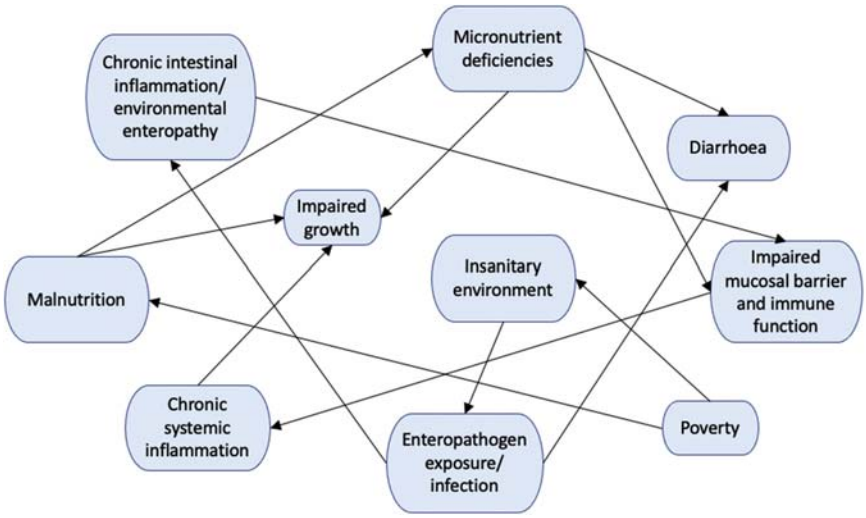


Fig. 1 Diagram demonstrating the complex interplay of factors linking diarrhea and malnutrition.

If it is indeed the case that luminal enteropathogen exposure contributes to malnutrition as opposed to diarrheal disease in itself, then interventions aimed at reducing enteropathogen burden in children might reasonably be expected to have a positive effect on childhood growth (Fig. 1). The WASH benefits trial was a cluster-randomized trial in which pregnant women in villages in rural Bangladesh and Kenya were randomized to one of seven interventions aimed at improving sanitation and nutrition. This showed that interventions aimed at improving sanitation alone did reduce childhood diarrhea in Bangladesh, however only interventions that included a nutrition element were seen to be associated with an improvement in linear growth (Luby et al., 2018). Sanitation interventions did not appear to have an impact on childhood diarrhea or on growth in Kenya (Null et al., 2018). The Sanitation Hygiene Infant Nutrition Efficacy (SHINE) trial, a cluster-randomized trial of water, sanitation and hygiene (WASH) and improved infant and young child feeding (IYCF) interventions in rural districts in Zimbabwe, showed that neither of these types of interventions reduced the prevalence of childhood diarrhea, and while IYCF interventions did increase linear growth, the combination of WASH interventions with IYCF did not provide any additional benefit (Humphrey et al., 2019). It is likely that the WASH interventions in these trials were not of sufficient intensity to result in improvement in child health outcomes, as enteropathogen burden remained high in the intervention groups. Interventions are needed that are more effective at reducing (and ideally excluding) fecal contamination of the domestic environment (Fig. 2).

Zinc deficiency is widespread among populations of LICs, and this may be both a consequence of and a contributor to diarrheal disease. Zinc is absorbed in the small intestine and decreased absorptive capacity of the gut may contribute to deficiency in LICs, where environmental enteric dysfunction is known to be widespread and reduced fractional absorption of zinc has been demonstrated in children. Fecal losses of zinc occurring in chronic or frequent diarrheal disease also contribute to deficiency. Zinc



Fig. 2 View of a high-density residential area in Lusaka, Zambia, typical of unplanned settlements in LICs where lack of proper drainage leads to regular flooding which contributes to insanitary conditions.

deficiency may lead to malabsorption through the impairment of digestive enzymes in the small intestine, hence perpetuating the cycle of enteropathy and zinc deficiency. The World Health Organization (WHO) recommends that zinc supplements at a dosage of 20 mg/day for 10–14 days are given to all children with acute diarrhea, and this has been shown to reduce the duration and severity of symptoms as well as the risk of subsequent episodes in the short-term.

The nutritional consequences of chronic non-infectious diarrhea should also be considered. Specific nutritional deficits will depend on the site of the gastrointestinal disease, for example, Crohn's disease often involves the terminal ileum hence vitamin B12 deficiency is common as this is the only site for absorption of this vitamin. Nutritional deficits relating to specific conditions will be discussed in the section "Non-infectious causes of diarrheal disease".

Infectious causes of diarrheal disease

Infectious causes of diarrhea can be subdivided according to whether the causative agent is bacterial, viral or parasitic, and we will discuss some of the commonest causes in each of these categories.

Bacterial

Common bacterial causes of diarrhea include *Campylobacter jejuni*, several pathotypes of *E. coli*, *Shigella*, *Vibrio cholerae*, nontyphoidal *Salmonella* species, *Clostridioides difficile* (previously *Clostridium difficile*), and *Yersinia enterocolitica*. *Vibrio cholerae* will be discussed below under "Cholera." All of these can present with bloody diarrhea although *Shigella* is the predominant bacterial cause of this.

C. jejuni is the most common cause of bacterial diarrhea worldwide and was found to be responsible for 7.5 million disability-adjusted life years in the 2010 Global Burden of Disease Study. This is a foodborne infection which in high-income countries is usually sporadic as opposed to causing outbreaks. In low- and middle-income countries (LMICs), *C. jejuni* is hyperendemic and spread via contaminated drinking water and food (e.g., poultry). Infection is usually only symptomatic in children, who can be infected multiple times.

There are 6 diarrheagenic *Escherichia coli* (DEC) pathotypes: enteropathogenic (EPEC); enterotoxigenic (ETEC); enteroaggregative (EAEC); *Shiga* toxin producing or enterohaemorrhagic (STEC/EHEC); enteroinvasive (EIEC); and diffusely adherent (DAEC), with ETEC being the principal pathotype in LMICs and in returning travelers. ETEC causes diarrhea through heat stable (ST) and heat-labile (LT) toxins which, like cholera toxin, open chloride channels after activation of secondary messenger signaling. EPEC causes diarrhea through complex processes initiated by intimate adherence. EIEC is invasive, with a similar pathogenesis to *Shigella*, while EHEC is usually akin to an EPEC which additionally carries a nephrotoxic cytotoxin gene. A study from Guatemala found that by the age of 36 months, cumulative seroprevalence of antibodies to ETEC was 83% (Steinberg et al., 2004). In addition to being one of the four most common pathogens causing moderate-to-severe diarrhea in children in LMICs, ETEC producing heat-stable toxin (ST-ETEC) as well as typical EPEC are associated with a higher risk of mortality in infants than other enteropathogens (Kotloff et al., 2013).

Nontyphoidal *Salmonella* species are a major cause of food-related outbreaks of diarrheal disease, usually related to undercooked poultry or eggs, and are also an important cause of diarrhea in returning travelers. In immunocompetent individuals, typical nontyphoidal salmonellosis causes self-limiting diarrheal disease and treatment with antimicrobial agents is not generally recommended as this can prolong bacterial shedding. Bacteremia secondary to invasive nontyphoidal salmonellosis can occur in high-risk groups (extremes of age, HIV- or malaria-infected, malnourished) with the majority of these cases occurring in Africa. *Salmonella enterica* serovars *typhimurium*, Dublin and Enteritidis are most commonly associated with invasive disease with the Typhimurium sequence type 313 being of particular concern as this demonstrates multidrug resistance to antimicrobial agents and may not require an animal reservoir. It is important to note that typhoid (due to *Salmonella enterica* serovar *typhi*) is not primarily a diarrheal disease and is not considered in this article.

Clostridioides difficile is one of the most common nosocomial infections worldwide and the most common iatrogenic cause of diarrhea. It is a spore-forming bacillus with infection usually occurring as a result of spore ingestion, as these are able to survive on fomites for several months. For this reason, measures including handwashing, glove wearing and decontamination of medical devices and patient environment as well as antimicrobial stewardship are important in the prevention of *C. difficile* infection (CDI) in healthcare settings. CDI is often associated with antibiotic use in hospitalized patients, particularly in the elderly or immunocompromized, as this depletes the intestinal microbiota and disrupts the protective mucous layer. There is a range of clinical presentations, from asymptomatic carrier status to pseudomembranous colitis with potential complications of toxic megacolon, bowel perforation, and death. Although antibiotics including metronidazole, oral vancomycin and fidaxomicin are often used as first line treatments for symptomatic CDI, there is increasing evidence that fecal microbiota transplantation is an extremely effective treatment.

Yersinia enterocolitica is a zoonotic infection, with pigs being the most common animal reservoir, and is usually transmitted to humans through ingestion of undercooked or raw pork products, although foodborne outbreaks have also been associated with contaminated water, milk, bean sprouts and tofu.

Cholera

Cholera typically causes epidemic outbreaks in LMICs and represents a major public health burden for some of these countries. There have been seven pandemic cholera outbreaks in the past, with the first six outbreaks originating on the Indian subcontinent during the nineteenth century and the seventh originating in the 1960s. Prior to the 19th century, epidemic cholera had largely been confined to Asia as it was in the Bengal river deltas that this formerly free-living estuarine bacterium acquired its virulence genes. *Vibrio cholerae* causes diarrhea through the cholera toxin, which is produced by *V. cholerae* serogroups 01 and 0139 and consists of an A subunit and 5 B subunits. The B subunit binds to monosialoganglioside GM1 receptors on the luminal surface of enterocytes and following endocytosis travels through retrograde transport from the plasma membrane to the trans-Golgi network and thence to the endoplasmic reticulum where the active portion of the A subunit (the CTA1 subunit) is unfolded. CTA1 stimulates adenylate cyclase activity through ADP-ribosylation, leading to increased cAMP production and efflux of chloride ions, and subsequently to secretory diarrhea with massive fluid and electrolyte losses. Treatment of cholera was revolutionized by the advent of oral rehydration solution (ORS) which came into widespread use in the 1970s. An inexpensive solution of glucose and electrolytes, ORS can in fact be used to treat acute watery diarrhea regardless of the etiology and has enabled reduction in mortality in cholera outbreaks from over 60% to less than 1%.

Viral

Rotavirus is the leading cause of severe acute diarrheal disease in children globally, with approximately half of deaths due to rotavirus occurring in just four countries: India; Nigeria; Pakistan; and the Democratic Republic of the Congo. Deaths have reduced significantly since the introduction of rotavirus vaccines in the early 21st century, however, unfortunately oral rotavirus vaccinations underperform in LICs where they are most needed, possibly due to the ubiquity of environmental enteropathy in these settings. Rotavirus causes an inflammatory diarrhea with a secretory component due to a viral enterotoxin encoded in the viral genome.

Adenovirus can cause a range of clinical presentations in addition to acute diarrheal disease including febrile upper and lower respiratory tract infections, hepatitis, keratoconjunctivitis, meningoencephalitis, cystitis and myocarditis. It usually causes a mild self-limiting clinical picture in the immunocompetent host however infection can be of particular significance in the context of immunosuppression, for example, in post-transplant patients.

Norovirus (previously known as Norwalk virus) usually presents with nausea and vomiting, diarrhea, and fever. It is highly transmissible due to the small infective dose (<100 viral particles) and was originally termed "winter vomiting sickness" before the causative organism was known as it tended to occur in seasonal outbreaks in healthcare settings such as hospitals and nursing homes in the Northern hemisphere. There is some evidence for a high prevalence of noroviruses in LMICs (a rate of 8% among healthy controls is cited in one review), albeit with a significant level of asymptomatic infection, and in tropical regions infections seem to correlate with the rainy season.

Sapoviruses cause a similar clinical presentation to norovirus albeit generally milder, with self-limiting diarrhea and vomiting, both in outbreaks and sporadic cases worldwide. Sapoviruses are thought to account for 5.9–22.6% outbreaks where samples are negative for norovirus and pathogenic bacteria.

Parasites

Parasitic causes of diarrhea include protozoa: *Entamoeba histolytica*, *Giardia intestinalis*, *Cryptosporidium*, *Cystoisospora belli*, *Cyclospora cayetanensis*, and some helminths including *Strongyloides stercoralis*, *Trichuris trichiuria* (whipworm), and *Schistosoma mansoni*. These are discussed further in the article on "Parasitism." Cryptosporidiosis in particular is a leading cause of childhood diarrhea in LMICs and is known to be associated with growth faltering even in the absence of symptoms of diarrhea. In the MAL-ED study, the presence of *Giardia* (although not associated with increased risk of acute diarrheal illness) was associated with increased intestinal permeability and worse growth outcomes at 24 months. Parasitic infections of the gastrointestinal tract can lead to specific nutritional deficits, for example, hookworm infection (*Necator americanus* and *Ancylostoma duodenale*) does not usually present with diarrheal disease but can lead to iron deficiency anemia as the adult worms attach to the mucosa of the small intestine leading to chronic blood loss. The question of whether worm infections (specifically the soil-transmitted helminths *Ascaris lumbricoides*, whipworm and hookworm) contribute to growth faltering in children remains controversial however a cluster-randomized trial of periodic deworming in 1 million pre-school children in India showed no significant effect on growth (Awasthi et al., 2013).

Human immunodeficiency virus (HIV)

Although HIV infection in itself does not cause diarrhea, immunocompromized individuals are more at risk of opportunistic infections causing diarrhea including *Cryptosporidium*, *Cystoisospora belli*, cytomegalovirus, *Mycobacterium avium* complex and *C. difficile*. Several antiretroviral drugs prescribed in HIV can also cause diarrhea.

Intestinal tuberculosis

Mycobacterium tuberculosis infection of the gastrointestinal tract may present in a non-specific way with chronic diarrhea with or without blood, abdominal pain, fever, and weight loss, and may be difficult to distinguish from inflammatory bowel disease (IBD), particularly Crohn's disease, as these are both inflammatory granulomatous conditions. Similarly, to Crohn's disease, it may affect any part of the gastrointestinal tract however has a predilection for the ileocecal region. Not all patients with intestinal TB will have concurrent pulmonary disease, with active pulmonary TB only being found to be present in a quarter of patients with intestinal infection.

Whipple's disease

Caused by infection with the bacteria *Tropheryma whipplei*, this rare multisystem disease can present with diarrhea and weight loss and can be diagnosed through histopathological analysis of duodenal biopsies or through PCR detection of *T. whipplei* in tissue. Classic Whipple's disease typically occurs in Caucasian individuals and, although carriage of *T. whipplei* seems to be common in native African and Asian populations, resulting disease is rare.

Tropical sprue

This condition, which manifests clinically as chronic diarrhea and weight loss, was previously commonly described in South Asia and the Caribbean and is now rarely diagnosed. It has been listed under infectious causes of diarrhea as it now seems likely that reports of this condition in the past were due to infections which were either not known about or which it was not possible to exclude with diagnostic techniques available at the time. For the traditional method of diagnosis of this disorder, it is necessary to exclude infectious causes, and to demonstrate evidence of enteropathy on small bowel biopsies (villous blunting, lymphocyte infiltration and increased crypt depth) as well as evidence of malabsorption of fat, vitamin B12 and xylose. However, in the present day it is difficult to arrange tests of fat and xylose absorption hence demonstration of enteropathy in the presence of folate and vitamin D deficiency in the absence of any other demonstrable cause should be accepted as sufficient for diagnosis of tropical sprue.

Non-infectious causes

There are many causes of non-infectious diarrhea, including the inflammatory bowel diseases ulcerative colitis and Crohn's disease (autoimmune conditions leading to intestinal inflammation and usually presenting with diarrhea) as well as malignant causes including colorectal cancer, neuroendocrine tumors and lymphoma. Drugs are also an important cause of non-infectious diarrhea, with common causative agents being antibiotics, proton pump inhibitors, H2 receptor antagonists, and metformin. Radiotherapy targeting cancers in the abdomen or pelvis can also result in enteropathy leading to chronic diarrheal disease. Celiac disease is another important cause of chronic diarrheal disease, as is malabsorption secondary to pancreatic exocrine insufficiency. Hyperthyroidism can also present with diarrhea. Specific nutritional deficiencies associated with several of these conditions are discussed below.

Inflammatory bowel disease

Although malnutrition can occur in both of the predominant forms of IBD, ulcerative colitis (UC) and Crohn's disease, it is more common in Crohn's as this condition can affect the small intestine and hence has greater capacity to disrupt the absorptive function of the gut. As previously discussed, vitamin B12 deficiency is common in Crohn's for this reason, as well as other micronutrient deficiencies including folate, iron, zinc, selenium, and vitamin D. These should be screened for routinely and supplementation prescribed as required. Low bone mineral density is also common in IBD due to both reduced absorption of calcium and vitamin D as well as corticosteroid use in these patients. Due to the inflammatory nature of IBD, patients with active disease are in an increased catabolic state and hence have greater protein requirements (1.2–1.5 g/kg/day in adults) than the general population where an intake of 1 g/kg/day is recommended. There is little evidence to support that energy expenditure in general is greater in IBD patients, with the possible exception of severe acute UC, in which state electrolyte losses (particularly K^+ and Mg^{2+}) are also common due to high volume diarrhea. Hypokalaemia may also be exacerbated by use of intravenous hydrocortisone in these patients.

Celiac disease

This is an immune-mediated gluten-sensitive enteropathy caused by antibodies against gluten peptides in genetically susceptible individuals. Treatment consists of lifelong dietary exclusion of gluten. In untreated celiac disease there is villous atrophy and inflammation with loss of small intestinal surface area as well as loss of disaccharidase activity, leading to carbohydrate maldigestion and malabsorption. Individuals with celiac may present with diarrhea, weight loss, abdominal bloating, iron deficiency, and poor

growth in children. Other nutritional deficiencies which are reported in untreated celiac disease include vitamin B12, folic acid, vitamin B6 (pyridoxine), vitamin D, calcium, zinc, and less commonly, magnesium.

Pancreatic exocrine insufficiency

Failure of the exocrine pancreas usually occurs as a result of chronic pancreatitis however can also be secondary to pancreatic malignancies, resection, cystic fibrosis, and less commonly diabetes mellitus. Maldigestion and malabsorption occur as a result of loss of functioning pancreatic parenchyma and hence reduced secretion of pancreatic enzymes, and affected individuals present with symptoms including diarrhea, steatorrhea, weight loss, abdominal pain and bloating. Nutritional deficiencies seen in pancreatic exocrine insufficiency include fat soluble vitamins (A, D, E and K), micronutrients including magnesium, trace elements and plasma proteins.

Intestinal failure (often due to short bowel syndrome)

Intestinal failure is the reduction of gut function below the minimum absorptive capacity required to maintain health and/or growth. Short bowel syndrome, which refers to the physical loss of or loss of function of a portion of the small bowel resulting in a reduced ability to absorb nutrients, is the most common cause of intestinal failure. This condition may be congenital or acquired as a result of the following: surgical resection, for example, following abdominal catastrophe such as infarction, trauma, or volvulus; extensive small bowel mucosal disease usually secondary to Crohn's disease; or radiotherapy for intraabdominal cancers. Fluid and electrolyte loss can be very high in this condition, particularly in the acute phase, when patients with jejunostomy may lose as much as 8 L/day. Depending on the length of remaining small bowel and whether or not the colon is in continuity, these patients may require intravenous supplementation with fluid, electrolytes, and in some cases parenteral nutrition.

Diagnosis

Diagnostic approach to the patient with diarrheal disease depends on the duration of symptoms, with acute diarrhea being more likely to be due to an infectious agent, and cases with a duration of greater than two weeks warranting investigation for non-infectious causes. All cases of persistent diarrhea and all cases of dysentery should have stool cultures and examination for ova, cysts and parasites on three separate stool samples. In most settings, investigation of acute watery diarrhea would only be indicated in specific circumstances such as outbreaks.

Prevention and treatment

General population-based preventive measures recommended by the WHO to reduce the burden of diarrheal disease include increasing access to clean drinking water, improved sanitation and handwashing with soap, improved food hygiene, rotavirus vaccination and exclusive breastfeeding for the first six months of life. The WHO recommends that oral cholera vaccines should be used as a reactive strategy during outbreaks and for prevention in high-risk areas. Currently available vaccinations for diarrheal diseases are summarized in **Table 1**.

The WASH Sustainable Development Goal set by the UN is for "universal access to safe and affordable drinking water and adequate and equitable sanitation and hygiene for all by 2030." Despite the seemingly obvious benefits inherent in this goal, it should be noted that three recent major trials of WASH interventions in LICs (WASH-Benefits Bangladesh, WASH-Benefits Kenya, and Sanitation Hygiene Infant Nutrition Efficacy (SHINE)) found no effect of these interventions on childhood growth and mixed effect on diarrheal disease. Interventions in these trials included increased chlorination of drinking water and increased access to improved pit-latrines and to handwashing stations with soap, and uptake of interventions was reported to be high. Although these

Table 1 Available vaccines for diarrheal diseases.

Type	*Disease*	*Vaccine constituents*	*Route*	*Protection*
Live, attenuated	Cholera	*V. cholerae* (CVD103-HgR, **Vaxchora**)	Oral	67% over 3 years
	Rotavirus	Attenuated virus (**Rotarix**; GSK)	Oral	85–100%
	Rotavirus	Human-bovine reassortant viruses (**RotaTeq**; Merck)	Oral	74% over one season, 94.5% vs. severe disease
Inactivated	Cholera	Heat-killed *V. cholerae* + CTB (**Dukoral**; Chiron)	Oral	80–90% over 6/12, 60% at 2 years
	Cholera	Vc O1 Inaba and Ogawa plus O139 (Euvichol, Shanchol)	Oral	40% over 6 months

interventions had no effect on diarrhea in children in Kenya and Zimbabwe, a 40% relative reduction in diarrheal disease was seen in Bangladesh, which may reflect differing baseline levels of contamination between environments. An expert meeting convened by the WHO and the Bill and Melinda Gates Foundation to discuss these results concluded that the household interventions in these trials were not sufficiently intensive to achieve clear benefits, and in fact community-level interventions such as providing a piped, microbially safe supply of drinking water to each household or a community-level sewered sanitation system may be required in order to impact childhood diarrheal disease and stunting.

Specific treatments for diarrheal disease will depend on the underlying cause, for example, targeted antimicrobial treatment in infectious diarrhea where the causative organism is known. However supportive treatments can be used regardless of the etiology of the diarrheal disease, including ORS, replacement of electrolytes, zinc supplements, and appropriate nutritional support.

Conclusions

Although the burden of diarrheal diseases remains significant particularly in LICs due to limited access to clean drinking water and sanitation, the advent of simple interventions such as ORS has reduced mortality particularly in the setting of outbreaks. It seems that the relationship between diarrhea and childhood malnutrition is indirect, with enteropathogen exposure as opposed to diarrhea itself responsible for limiting growth. Large scale community-level interventions to improve sanitation and water supply are required to address this problem in LICs.

References

Amadi, B., Zyambo, K., Chandwe, K., et al., 2021. Adaptation of the small intestine to microbial enteropathogens in Zambian children with stunting. Nat. Microbiol. 6 (4), 445–454.

Awasthi, S., Peto, R., Read, S., et al., 2013. Population deworming every 6 months with albendazole in 1 million pre-school children in North India: DEVTA, a cluster-randomized trial. Lancet 381, 1478–1486.

Checkley, W., Buckley, G., Gilman, R.H., et al., 2008. Multi-country analysis of the effects of diarrhea on childhood stunting. Int. J. Epidemiol. 37, 816–830.

Humphrey, J.H., Mbuya, M.N.N., Ntozini, R., et al., 2019. Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural Zimbabwe: a cluster-randomized trial. Lancet Glob. Health 7, e132–e147.

Kotloff, K.L., Nataro, J.P., Blackwelder, W.C., et al., 2013. Burden and aetiology of diarrheal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet 382, 209–222.

Luby, S.P., Rahman, M., Arnold, B.F., et al., 2018. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhea and child growth in rural Bangladesh: a cluster randomized controlled trial. Lancet Glob. Health 6, e302–e315.

MAL-ED Network Investigators, 2017. Relationship between growth and illness, enteropathogens and dietary intakes in the first 2 years of life: findings from the MAL-ED birth cohort study. BMJ Glob. Health 2, e000370.

Null, C., Stewart, C.P., Pickering, A.J., et al., 2018. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhea and child growth in rural Kenya: a cluster-randomized controlled trial. Lancet Glob. Health 6, e316–e329.

Richard, S.A., Black, R.E., Gilman, R.H., et al., 2013. Diarrhea in early childhood: short-term association with weight and long-term association with length. Am. J. Epidemiol. 178, 1129–1138.

Steinberg, E.B., Mendoza, C.E., Glass, R., et al., 2004. Prevalence of infection with waterborne pathogens: a seroepidemiologic study in children 6–36 months old in San Juan Sacatepequez, Guatemala. Am. J. Trop. Med. Hyg. 70, 83–88.

Further reading

Aksan, A., Farrag, K., Blumenstein, I., et al., 2021. Chronic intestinal failure and short bowel syndrome in Crohn's disease. World J. Gastroenterol. 27, 3440–3465.

Bischoff, S.C., Escher, J., Hébuterne, X., et al., 2020. ESPEN practical guideline: clinical nutrition in inflammatory bowel disease. Clin. Nutr. 39, 632–653.

Church, J.A., Parker, E.P.K., Kosek, M.N., et al., 2018. Exploring the relationship between environmental enteric dysfunction and oral vaccine responses. Future Microbiol. 13, 1055–1070.

Crawford, S.E., Ramani, S., Tate, J.E., et al., 2017. Rotavirus infection. Nat. Rev. Dis. Prim. 3, 17083.

Cumming, O., Arnold, B.F., Ban, R., et al., 2019. The implications of three major new trials for the effect of water, sanitation and hygiene on childhood diarrhea and stunting: a consensus statement. BMC Med. 17, 173.

Czepiel, J., Dróżdż, M., Pituch, H., et al., 2019. *Clostridium difficile* infection: review. Eur. J. Clin. Microbiol. Infect. Dis. 38, 1211–1221.

Dolmans, R.A., Boel, C.H., Lacle, M.M., Kusters, J.G., 2017. Clinical manifestations, treatment, and diagnosis of *Tropheryma whipplei* infections. Clin. Microbiol. Rev. 30, 529–555.

Haselbeck, A.H., Panzner, U., Im, J., et al., 2017. Current perspectives on invasive nontyphoidal *Salmonella* disease. Curr. Opin. Infect. Dis. 30, 498–503.

Jesser, K.J., Levy, K., 2020. Updates on defining and detecting diarrheagenic *Escherichia coli* pathotypes. Curr. Opin. Infect. Dis. 33, 372–380.

Kedia, S., Das, P., Madhusudhan, K.S., et al., 2019. Differentiating Crohn's disease from intestinal tuberculosis. World J. Gastroenterol. 25, 418–432.

Kreutz, J.M., Adriaanse, M.P.M., Van Der Ploeg, E.M.C., Vreugdenhil, A.C.E., 2020. Narrative review: nutrient deficiencies in adults and children with treated and untreated celiac disease. Nutrients 12.

Lindenmayer, G.W., Stoltzfus, R.J., Prendergast, A.J., 2014. Interactions between zinc deficiency and environmental enteropathy in developing countries. Adv. Nutr. 5, 1–6.

Lion, T., 2014. Adenovirus infections in immunocompetent and immunocompromized patients. Clin. Microbiol. Rev. 27, 441–462.

Louis-Auguste, J., Kelly, P., 2017. Tropical enteropathies. Curr. Gastroenterol. Rep. 19, 29.

Mans, J., 2019. Norovirus infections and disease in lower-Middle and Low-income countries, 1997–2018. Viruses 11.

Oka, T., Wang, Q., Katayama, K., Saif, L.J., 2015. Comprehensive review of human sapoviruses. Clin. Microbiol. Rev. 28, 32–53.

Phillips, M.E., Hopper, A.D., Leeds, J.S., et al., 2021. Consensus for the management of pancreatic exocrine insufficiency: UK practical guidelines. BMJ Open Gastroenterol. 8, e000643.

Pickering, A.J., Null, C., Winch, P.J., et al., 2019. The WASH Benefits and SHINE trials: interpretation of WASH intervention effects on linear growth and diarrhea. Lancet Glob. Health 7, e1139–e1146.
Platts-Mills, J.A., Babji, S., Bodhidatta, L., et al., 2015. Pathogen-specific burdens of community diarrhea in developing countries: a multisite birth cohort study (MAL-ED). Lancet Glob. Health 3, e564–e575.
Robilotti, E., Deresinski, S., Pinsky, B.A., 2015. Norovirus. Clin. Microbiol. Rev. 28, 134–164.
Rogawski, E.T., Bartelt, L.A., Platts-Mills, J.A., et al., 2017. Determinants and impact of *Giardia* infection in the first 2 years of life in the MAL-ED birth cohort. J. Pediatr. Infect. Dis. Soc. 6, 153–160.
Sabina, Y., Rahman, A., Ray, R.C., Montet, D., 2011. *Yersinia enterocolitica*: mode of transmission, molecular insights of virulence, and pathogenesis of infection. J. Pathog. 2011, 429069.
Steiner, K.L., Ahmed, S., Gilchrist, C.A., et al., 2018. Species of cryptosporidia causing subclinical infection associated with growth faltering in rural and urban Bangladesh: a birth cohort study. Clin. Infect. Dis. 67, 1347–1355.
Wernick, N.L., Chinnapen, D.J., Cho, J.A., Lencer, W.I., 2010. Cholera toxin: an intracellular journey into the cytosol by way of the endoplasmic reticulum. Toxins 2, 310–325.
Zyambo, K., Hodges, P., Chandwe, K., et al., 2021. Reduced fractional absorption of zinc in children with environmental enteropathy in Zambia. J. Pediatr. Gastroenterol. Nutr. https://doi.org/10.1097/MPG.0000000000003292.

Relevant website

https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease.

Diet and oral health

Anja Heilmann, Yanela Aravena Rivas, and Richard G. Watt, University College London, London, United Kingdom

© 2023 Published by Elsevier Ltd. All rights reserved.

This is an update of R. Cottrell, Dental Disease: Etiology and Epidemiology, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 10–16, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00069-6.

Introduction 272
Background 273
Impact of oral diseases 273
Epidemiology and etiology of diet-related oral diseases 273
Dental caries 273
Prevalence and trends 273
Etiology of dental caries 274
Periodontal disease 275
Prevalence and trends 275
Etiology of periodontal disease 275
Dental erosion 276
Prevalence 276
Etiology 276
Oral cancer 276
Prevalence 276
Etiology of oral cancers 276
Social and commercial determinants of oral diseases 276
Social determinants of oral diseases 276
Commercial determinants of sugar consumption 276
Policy strategies to reduce free sugar consumption 278
Upstream strategies 278
Regulation of industry influence on policy development 278
Sugar production and trade 279
Voluntary agreements to achieve reformulation 279
Taxation 279
Advertising restrictions 279
Labeling 279
Midstream strategies 279
Downstream strategies 280
Conclusions 280
References 280
Further reading 280

Key points

- Dental caries of the permanent teeth is the most common of all chronic diseases.
- Dental caries is caused by high intakes of free sugars. The availability of fluoride in the mouth helps to prevent dental caries, however the link between free sugar intake and caries development remains also when fluoride is present.
- WHO recommends to restrict free sugar intake to less than 10% of total energy intake.
- Dietary choices are influenced by the wider social and commercial environments people live in.
- Policy strategies to reduce free sugar intake need to create supportive environments through a combination of upstream, midstream, and downstream strategies, including the regulation of industry activities.

Introduction

Despite being largely preventable, oral diseases are the most common conditions of mankind, with a significant impact on the individuals affected and their wider communities. Oral diseases are highly prevalent across the entire life course from early life to old age. Globally, diseases of the teeth and mouth affect 3.5 billion people, nearly half of the world's population (G.B.D. Oral Disorders

Collaborators, 2020). In 2015 oral diseases accounted for an estimated US$ 357 billion in direct costs (such as treatment expenditures) and US$ 188 billion in indirect costs (such as productivity losses due to absence from work or school) (Righolt et al., 2018). Oral diseases disproportionally affect the poor and vulnerable in society with stark and persistent oral health inequalities evident globally.

In recognition of the global public health importance of oral diseases, in May 2022 the World Health Organization (WHO) agreed a new landmark Global Oral Health Strategy, which highlights the need for prevention and health promotion to combat oral diseases and reduce oral health inequalities (WHO, 2022). The WHO strategy also recognizes the importance of adopting an integrated approach to prevention through tackling the shared causes of oral diseases and other non-communicable conditions, including diet and nutrition.

This article will present an epidemiological overview of oral diseases with the main focus on dental caries, the most common oral condition, and will outline the need for a comprehensive range of policy measures to reduce free sugars consumption, the principal cause of dental caries.

Background

The World Health Organization defines oral health as "the state of the mouth, teeth and orofacial structures that enables individuals to perform essential functions, such as eating, breathing and speaking, and encompasses psychosocial dimensions, such as self-confidence, well-being and the ability to socialize and work without pain, discomfort and embarrassment" (WHO, 2022). Oral health is clearly a very important and integral element of overall health and wellbeing enabling essential daily functions. The mouth and teeth are also a fundamental feature of personal identity (Peres et al., 2019).

A wide range of diseases and disorders can affect the mouth and teeth, but the main oral diseases of global public health significance include dental caries (tooth decay), periodontal (gum) disease and oral cancers.

Dental caries is the localized destruction of dental hard tissues (enamel and dentine) by acidic by-products from the bacterial fermentation of free sugars (Pitts et al., 2017). Dental caries is a chronic condition that can affect individuals across the life course from early childhood to later life.

Periodontal diseases are chronic inflammatory conditions that affect the tissues surrounding and supporting the teeth. Initially, periodontal disease presents as gingivitis, a reversible inflammation of the periodontal soft tissues which causes gum bleeding and swelling. In a proportion of susceptible individuals, gingivitis progresses to periodontitis. Periodontitis involves the destruction of the periodontal tissue support including the bone surrounding the teeth and can ultimately lead to tooth loss. Severe periodontal disease is largely restricted to middle aged and older people.

Oral cancer is a broad term that is mostly used to refer to oral cavity cancer i.e., cancers of the lips, tongue, gums, floor of mouth, palate, and cheeks (ICD codes C00-C06). Some definitions include throat cancers, i.e., oropharyngeal cancer (ICD codes C07-C14). Squamous cell carcinoma is the most common type of oral cancer. The prevalence of oral cancers increases with age.

Impact of oral diseases

Oral diseases have a considerable impact on individuals and societies (**Fig. 1**). The mouth is the site of fundamental human functions, such as eating and communicating, and an integral part of the self (Peres et al., 2019). Dental caries, periodontal disease, and their endpoints of tooth loss affect quality of life of children and adults due to the associated pain and discomfort, by reducing the ability to eat and enjoy a variety of foods, and by affecting self-esteem and willingness to socialize. In children, severe levels of caries can impact sleep, concentration, performance at school, and even growth and development. Very young children may require extraction of carious teeth under general anesthesia. Because most dental treatments are provided by highly skilled dental professionals and require specialist equipment, the economic costs of oral diseases are also high. In many countries, out-of-pocket expenditure for dental treatment places a significant burden on individuals and families, with poorer households disproportionately affected. Societal costs are significant as well: due to the high prevalence of oral diseases, the costs of dental care in high-income countries exceed the treatment costs of other common non-communicable diseases (NCDs) such as cancer and respiratory diseases. In many low-income countries, the available healthcare budgets cannot accommodate these costs and oral diseases remain largely untreated.

Epidemiology and etiology of diet-related oral diseases

Dental caries

Prevalence and trends

Dental caries is the most common of all chronic NCDs. Despite an overall decline in caries levels between 1990 and 2017, the global prevalence remains unacceptably high. The Global Burden of Disease study estimated that in 2017, untreated dental caries of the permanent teeth affected 2.3 billion people worldwide, corresponding to an age-standardized prevalence of 29.4%, while dental caries in deciduous (milk) teeth affected 523 million children globally (G.B.D. Oral Disorders Collaborators, 2020). The distribution of dental caries is highly skewed toward more disadvantaged population groups, following clear social gradients among both



Fig. 1 Impact of oral diseases (Heilmann et al., 2015).

children and adults. At a global level, there is concern over the increasing burden of dental caries in many low- and middle-income countries (G.B.D. Oral Disorders Collaborators, 2020).

Etiology of dental caries

Dental caries is inextricably linked to diet—the main cause of caries is the consumption of free sugar. Caries develops through the interaction between acid-producing bacteria that are present in dental biofilms or plaque (these are mainly mutans streptococci, but also other groups of acidogenic and acid-tolerating species) and dietary free sugars (Pitts et al., 2017).

Not all dietary sugars are considered cariogenic—the sugars of concern are the so-called free sugars, defined by the WHO as "sugars added to foods or beverages by the manufacturer, cook or consumer; as well as sugars naturally present in honey, syrups, fruit juices and fruit juice concentrates" (WHO, 2015). Intrinsic sugars (those naturally present within the structure of intact fruits and vegetables) and milk sugars (naturally present in milk and milk products i.e., lactose and galactose) are not considered to increase caries risk. The WHO classification of sugars is shown in **Fig. 2** (WHO, 2015; Heilmann et al., 2021).

When free sugars are consumed, they are fermented by oral bacteria, a process which produces acid and causes the pH in the dental biofilm to fall, resulting in the demineralization of the dental hard tissues, i.e., the enamel and dentine. The critical pH at which demineralization occurs is about 5.5 for enamel and about 6.0 for dentine. The main acid involved in the caries process is lactic acid. During demineralization, mineral ions (calcium and phosphate) that are part of the enamel and dentine are released from the surface layer of enamel. This process is reversible—after the sugars are cleared from the mouth, the pH slowly rises again due to the buffering capacities of the saliva. Lost minerals are redeposited back into the enamel structure and remineralization occurs. As long as there is a balance between demineralization and remineralization, the tooth can remain intact. However, if free sugars are ingested too frequently and/or in high amounts, the repeated acid attacks can shift the balance toward demineralization and a carious lesion can develop (Pitts et al., 2017).

Given that acid-producing bacteria are ubiquitous in humans, free sugar consumption is considered the key causal factor in the development of dental caries. The associations between the amount of free sugars consumed and levels of dental caries were examined and confirmed by a systematic review of the evidence that informed the 2015 WHO guideline on sugars intake for adults and children (Moynihan and Kelly, 2014). In view of this evidence, and the body of evidence linking free sugar intake and unhealthy weight gain, the WHO guideline recommends that for both adults and children, sugar intake should be restricted to less than 10% of total energy intake (strong recommendation). A further reduction to less than 5% of total energy intake is given as a conditional recommendation (WHO, 2015).

Free sugar intake is in large parts driven by the consumption of ultra-processed foods and drinks. The term "ultra-processed" refers to manufactured foods and drinks that underwent a series of industrial processes to create products that are hyper-palatable, cheap, and ready to consume. They are made mostly from derived substances and have long lists of ingredients that

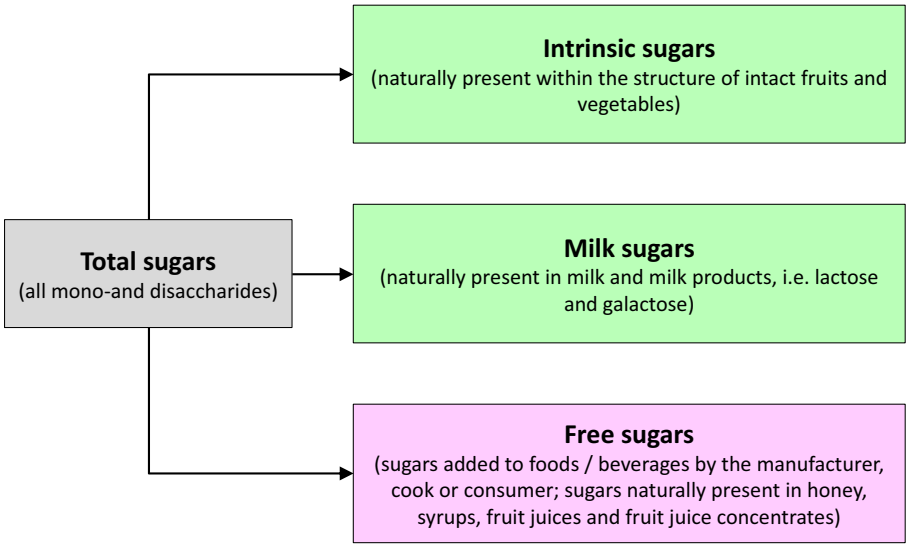


Fig. 2 WHO classification of sugars (Heilmann et al., 2021; WHO, 2015).

include additives such as flavors, colors, sweeteners, emulsifiers, and processing aids, while containing little or no intact (unprocessed) foods. Ultra-processed foods and drinks are energy-dense, often devoid of nutrients, and their free sugar content is generally much higher than that of unprocessed or minimally processed foods. Examples are sugar-sweetened beverages (SSB), breakfast cereals, sweets and biscuits, sugared milk products, instant sauces, and many baby products (Monteiro et al., 2018).

A question often asked is about the relative importance of frequency versus amount of free sugars consumption. Both factors play a role in the caries process; however, as they are highly correlated in human diets, the practicability and usefulness of attempting to disentangle them may be questionable. Given that the amount of sugar intake is important from a general health point of view, caries prevention strategies taking a Common Risk Factor Approach should aim to reduce the amount, while usefully highlighting that this might be achieved by reducing the frequency of intake.

The relationship between free sugar intake and dental caries is moderated by a number of other factors. These include salivary flow, the anatomy of the tooth, and most importantly, the availability of fluoride ions in the mouth (Pitts et al., 2017). The presence of fluoride in the mouth inhibits demineralization and enhances remineralization, moving the balance in favor of caries prevention. This means that the caries inhibiting effect of fluoride is mainly topical rather than through systemic uptake during the development of enamel, as was originally believed. Given its widespread use, accessibility and high acceptability, fluoridated toothpaste is considered the most important vehicle for getting topical fluoride into the mouth. The effectiveness of fluoridated toothpaste has been studied extensively in randomized controlled trials. The Cochrane review on the topic estimated that twice daily toothbrushing with fluoridated toothpaste prevents about 24% of caries lesions among children and adolescents. However, even with the widespread use of fluoride toothpastes and other forms of fluoride delivery, the link between free sugar intake and caries development remains (Moynihan and Kelly, 2014).

Periodontal disease

Prevalence and trends

Periodontal disease (gum disease) is another highly prevalent oral disease—according to the Global Burden of Disease study it is the sixth-most prevalent chronic condition in the world. Its prevalence increases with increasing age. The global age-standardized prevalence for severe periodontal disease was 9.8% in 2017, peaking at age 60–64 years. Like dental caries, periodontal disease is socially patterned, disproportionately affecting the more disadvantaged population groups. While the age-adjusted prevalence of periodontal disease has decreased in high-income countries since 1990, middle- and low-income countries have seen an increase over the same period (G.B.D. Oral Disorders Collaborators, 2020).

Etiology of periodontal disease

The most important risk factors for periodontal disease are poor oral hygiene and tobacco use. However, there are also associations with dietary factors, with evidence for a role of vitamin deficiencies (mainly B vitamins, vitamin C, and vitamin D), calcium, and proinflammatory saturated fatty acids. A diet high in free sugars has also been implicated in the development of periodontal disease, an inflammatory disease. Dietary sugars contribute to oxidative stress, which plays a role in inflammatory processes.

Dental erosion

Prevalence

Dental erosion, or tooth wear, is defined as non-carious tooth surface loss. Globally, there are large variations in the reported prevalence of dental erosion due to varying methods of measurement. Available data suggest that the prevalence of dental erosion in deciduous teeth ranges between 30% and 50%, and in permanent teeth between 20% and 45%.

Etiology

Dental erosion is caused by acids that are either intrinsic, i.e., produced in the body (most commonly through gastro-esophageal reflux disease or eating disorders that involve vomiting); or extrinsic, i.e., coming from the diet. It does not involve bacteria. Dietary sources of acid are mainly citric acid and phosphoric acid that are found in citrus fruits, fruit juices, soft drinks, vinegar and some alcoholic drinks—however, their potential to cause erosion depends not only on acidity but also on mineral content.

It should be stressed that while an excessive intake of acidic fruits has also been linked to dental erosion, the risk is far outweighed by the health benefits of a diet high in fruits and vegetables and no conflicting messages should be given to patients and the public.

Oral cancer

Prevalence

Oral cancers (cancers of the lip, tongue, and mouth) were the 16th most prevalent cancers worldwide in 2018, with the highest incidence occurring in South-Central Asia and parts of Oceania. Prevalence is socially graded with lower socioeconomic position associated with higher risk.

Etiology of oral cancers

Oral cancer is strongly associated with alcohol consumption, particularly in combination with smoking. It has been estimated that alcohol and tobacco together account for 80% of all oral cancers.

On the other hand, a diet high in fruits and vegetables is thought to have a protective effect, which is attributed to anti-inflammatory and antioxidant properties of micronutrients contained in fruits and vegetables. The evidence is based mainly on case-control studies with the potential for bias due to unmeasured confounding and selective reporting, and further research is needed.

Social and commercial determinants of oral diseases

Social determinants of oral diseases

The previous sections emphasized the causal role of dietary behaviors—in particular, free sugar consumption—for the development of oral diseases. It is important to recognize that health-related behaviors such as diet are not taking place in a vacuum but are influenced by the wider environment people live in—the social determinants of health. These include distal factors such as the political and economic context in a given society, the extent of structural inequality, and degree to which people's socio-economic position affects their living and working conditions and ability to access health services. These wider determinants shape individual behaviors through material and psychosocial pathways as shown in **Fig. 3** (Peres et al., 2019).

The importance of socio-economic factors for good oral health is reflected in the stark social gradients that are seen within and between countries. **Fig. 4** shows the stepwise increase in the prevalence of any caries experience with increasing levels of area-level deprivation among 5-year-olds in England in 2019 (Public Health England, 2020).

Moreover, oral diseases share common risk factors with general diseases. A diet high in free sugar increases the risk of dental caries, as well as obesity and type-2 diabetes. Alcohol consumption is not only linked to a higher risk of oral cancer but is also a leading cause of global morbidity and mortality from a range of other cancers and conditions. Addressing these common risk factors, which include the wider social determinants described above, to prevent oral diseases will therefore also help to prevent other chronic diseases. The implication is that oral and general diseases should be tackled together in an integrated manner.

More recently, the concept of the commercial determinants of health has been introduced. These refer to strategies used by the private sector to promote products and choices that are detrimental to health (Kickbusch et al., 2016). Sugar is a prime example for the way commercial determinants of health operate. The commercial determinants of sugar consumptions and implications for policy strategies aiming to improve oral and general health by reducing sugar intake are discussed in the following sections.

Commercial determinants of sugar consumption

Ultra-processed foods and drinks are the primary sources of free sugar intake in the population. They are cheap and readily available worldwide in both urban and rural settings. The sugar industry uses complex strategies to promote high sugar consumption.

The strategies used by the industry that have negative impacts on population health include advertising, lobbying (influencing legislators and policymakers to favor corporate goals), corporate social responsibility (social and environmental actions such as

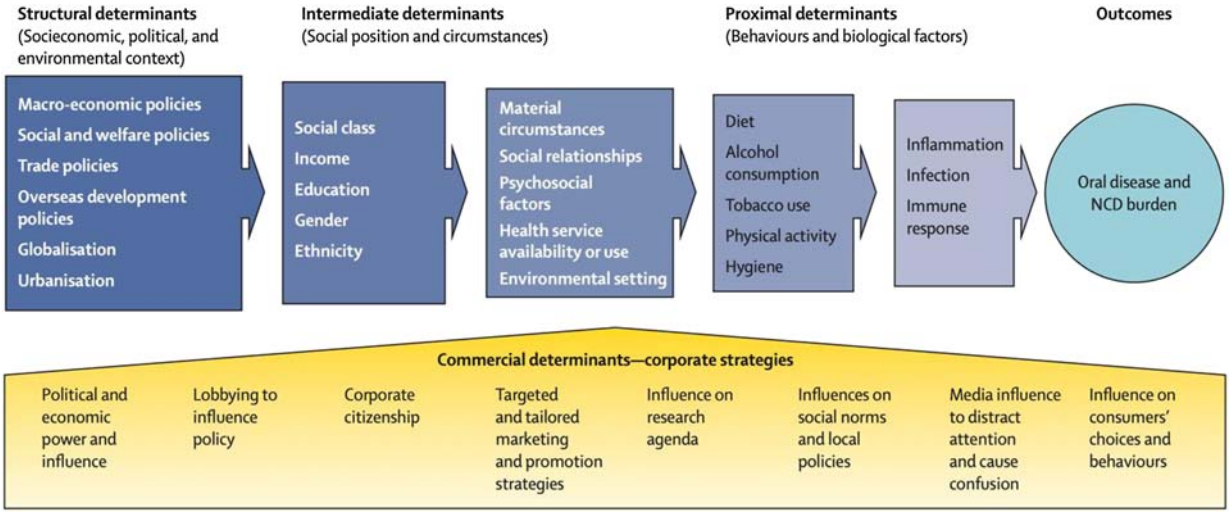


Fig. 3 Social and commercial determinants of oral diseases (Peres et al., 2019).

promoting sports or donating to charities) to increase social acceptability, and extending their supply chains, i.e., their manufacturing and distribution networks (Kickbusch et al., 2016). As a result, transnational sugary foods corporations have significant political and economic influence on the international landscape. Evidence shows how industry has used its power to protect their interests by frustrating public health initiatives. When the policy agenda does not suit industry interests, policies are less likely to be implemented, and if implemented, they are more likely to fail. One of the best-documented cases of industry interference with public health policymaking is when the Sugar Association (which brings together prominent sugar industry actors) pressured the US Congress to get the 2004 WHO Global Strategy on Diet, Physical Activity and Health withdrawn. The association demanded that the Congress withdraw US funding of WHO, if WHO did not remove from the guideline their recommendation to restrict added sugar intake to 10% of daily energy consumption.

The sugar industry has also opposed initiatives to reduce sugar consumption at the national level. As a result, policies to introduce the taxation of SSB and food labeling have faced substantial barriers in their implementation. The industry uses misleading arguments as well as their political and economic power to influence the discourse, including the setting up and funding of opposition groups.

The sugar industry spends billions of dollars on advertising and marketing with sophisticated strategies such as gamification, product placement in films, social media campaigns, and the use of celebrities to endorse their products. Industry spending on advertising to promote their unhealthy products far surpasses the funds available to WHO and national governments to promote healthy diets. Particularly concerning are the marketing campaigns directly targeting children, which often promote products that are touted as being "healthy" but are very high in free sugar.

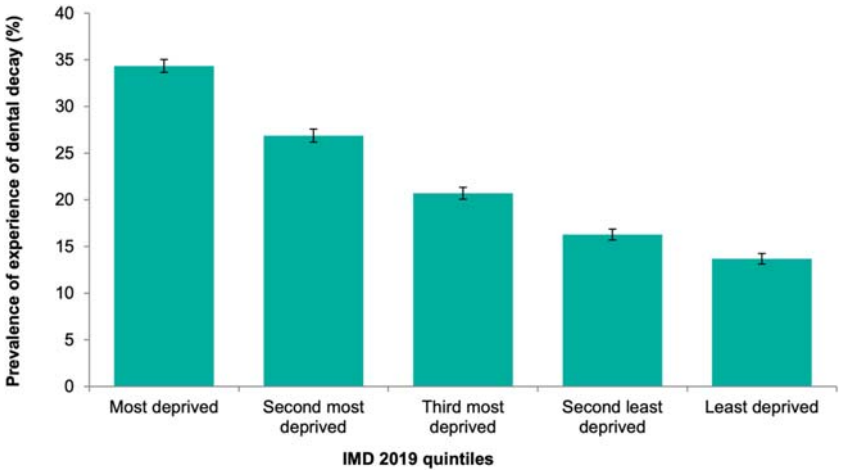


Fig. 4 Prevalence of caries experience in 5-year-olds in England, 2019 by national Index of Multiple Deprivation (IMD) 2019 quintiles (Public Health England, 2020). Error bars represent 95% confidence limits.

The WHO has acknowledged these commercial determinants and the need for public health actions to respond, and has recently initiated an action program to strengthen the evidence, develop tools, increase capacity, raise awareness, and help build partnerships to address the commercial determinants of health.

Policy strategies to reduce free sugar consumption

From a public health perspective, policy strategies that regulate the amount of free sugars present in foods and drinks are key to reduce sugar consumption at the population level. These strategies can range from upstream approaches aimed at structural changes to the food environment, to downstream approaches aiming to change individual dietary behaviors. Policy efforts need to be coordinated and implemented at international, national, and local levels. A summary of the most common policies currently used by different countries is shown in Fig. 5.

Upstream strategies

Upstream approaches aim to target the causes of disease at their source and can be highly effective. Key upstream strategies are regulation and fiscal policies, i.e., taxation.

Regulation of industry influence on policy development

The need to regulate the food industry has been highlighted as a critical upstream policy strategy to promote healthier eating, and governments worldwide have implemented a range of different policy measures. Examples include transparency mechanisms, managing conflicts of interest, monitoring corporate practices, and managing interactions with industry. However, there are currently no strategies directly regulating the sugar industry. Moreover, most of the strategies targeting different types of industries do not establish limits to curb their influence over policy-making processes, which is highly relevant in relation to the sugar industry. Thus, more robust regulation of industry policy influence is needed.

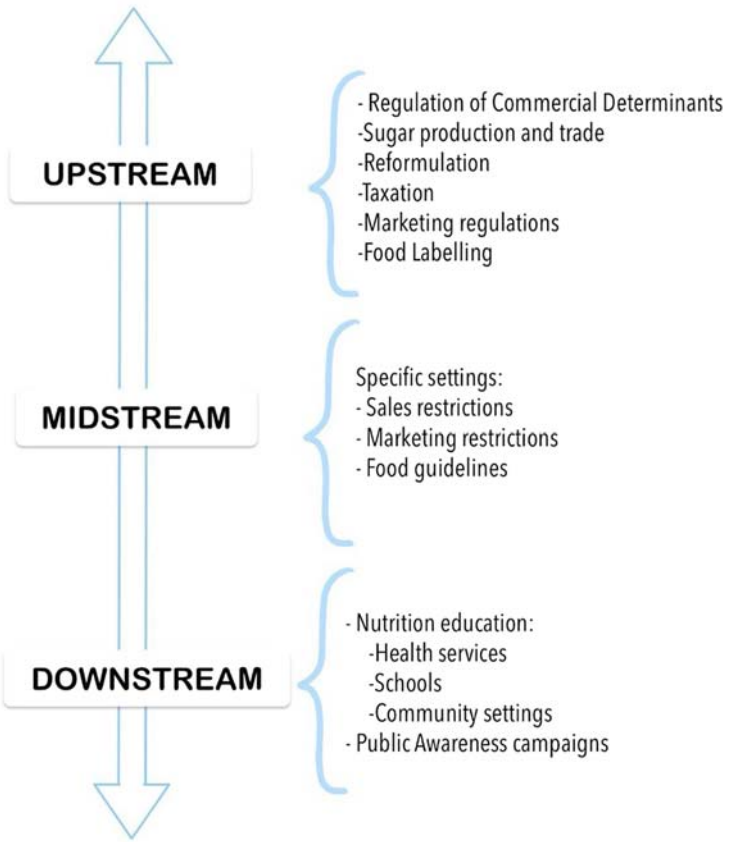


Fig. 5 Policy strategies to reduce free sugar consumption.

Sugar production and trade

Policies focusing on sugar production and its trade shape the availability of sugar at the population level and indirectly impact health outcomes. Agricultural diversification and alternative uses for sugar crops fall within this group of policies. For example, India has local programmes to promote the growing of traditional crops and use of sugarcane for ethanol production instead of refined sugar. Additionally, regulating the sugar trade can shape access to free sugars and their use by manufacturers of processed foods such as soft drinks and confectionery. Several countries have tariffs on importing sugars. However, agricultural and trade policies are usually based on economic arguments, particularly in countries where the sugar production industry and the sugar manufacturing industry generate many jobs and are a central part of the national economy.

Voluntary agreements to achieve reformulation

Several countries have used voluntary agreements to encourage manufacturers to reformulate their products and reduce their free sugar contents for better population health. Voluntary agreements with the industry exist in places such as Australia, Europe, and Southeast Asia. For example, Singapore's top seven soft drinks manufacturers pledged to reduce the sugar content of their products by 12% by 2020. However, voluntary efforts have failed to deliver significant progress due to their limited scope and lack of ambitious goals. An evaluation of the Public Health Responsibility Deal in England (a public–private partnership involving voluntary pledges between government, industry and other organizations to improve public health) concluded that it had largely failed. These failures of voluntary agreements with the food industry have highlighted the need for mandatory regulatory policies.

Taxation

Price is a major determinant of sugar consumption and fiscal measures such as taxation of sugary products can be highly effective. Taxation of SSB is increasingly popular because of their low nutritional value and high intake in many countries. According to the WHO Global Database on the Implementation of Nutrition Action, 84 member states have implemented SSB taxes. An important advantage of these strategies is that they do not require consumers to change their behavior. Evidence suggests that a 20% increase in price is needed for a significant reduction in SSB consumption. In some countries where taxation policies have been implemented, the industry responded with reformulation schemes to avoid price increases. For example, after announcement of the 2018 implementation of the soft drinks industry levy in the UK, the proportion of drinks with more than 5 g sugar per 100 mL fell by 34% points.

Advertising restrictions

The sugar industry spends billions on complex marketing strategies to promote its products. Because of this, national authorities have found themselves in need of stricter marketing regulations. Advertising restrictions can apply to media such as TV, radio, movies, or the internet, with additional rules to protect children. For example, several countries have introduced policies prohibiting TV adverts for sugary products before 9 p.m.

In Chile, cartoons, toys, or prizes as incentives to sell unhealthy foods to children are banned. Thus, cereal mascots cannot be used to promote breakfast cereals, and chocolate eggs with hidden toys cannot be sold in the country. Bans of industry sponsorship of community and sports events, and advertising bans on transport systems such as the Transport for London Network also fall within this category. However, these strategies are not widely implemented due to strong opposition from industry actors.

Labeling

Food labeling provides straightforward information about food composition to consumers. Food labeling systems such as traffic lights are simple and easy to understand by most people. However, the labeling of free sugars is inconsistent and often confusing. The industry's voluntary efforts have focused on using 'healthy stamps' systems to identify healthier products. Their effectiveness in changing purchase patterns is however low. More recently, countries such as Chile, Mexico, Peru, and Israel have implemented mandatory "high in sugars" warning signs for all processed foods. Early evaluations have shown these mandatory policies to be easily understood by consumers and more effective in reducing purchases of sugary products.

Midstream strategies

Midstream strategies aim to reduce free sugars consumption by creating healthy and supportive environments using a settings approach. For these strategies to be successful, joint action with actors such as businesses, public and private institutions, and communities is fundamental.

An important strategy in this group are mandatory food guidelines in institutions such as schools, nurseries, healthcare services and workspaces, to improve the quality of the food provided to students and staff. Examples are the replacement of fruit juices and soft drinks with water, or sugary desserts with fruit or sugar-reduced alternatives. In addition, policies banning the marketing and sale of sugary products in schools, workplaces, and healthcare settings have been proven effective in reducing sugar intake.

Sales restrictions in supermarkets and other retailers have also been implemented. For example, some supermarkets in the UK have replaced confectionery and sugary snacks usually sold at checkouts with healthier options. More recently, the UK has implemented restrictions on price promotions for foods high in sugars such as "buy one, get one free" offers.

Downstream strategies

The strategies that are most frequently implemented by local governments and/or health professionals are downstream policy strategies that target individual behavior change. They focus on the provision of dietary advice including individual support by health professionals and have been implemented with different degrees of success. Examples include nutrition education in health settings, schools, or community spaces. Evidence shows that these activities have modest results in reducing sugar intake. The types of nutritional education that are more likely to be effective are those focusing on healthcare workers, schoolteachers, or social care workers who can act as health champions in the community, and initiatives that teach skills such as cooking and growing foods.

Finally, there is a varied group of strategies aimed at providing information on healthy eating to consumers. Most countries have implemented public awareness campaigns at the national or regional level. However, research has shown that giving non-tailored, general advice on healthy eating can raise knowledge on the topic but does not lead to long-term behavior change. A disadvantage of downstream approaches is that they can widen oral health inequalities, because the messages are more likely to reach and benefit the more advantaged population groups.

Conclusions

Oral diseases, and dental caries in particular, pose significant global public health problems that affect a large proportion of the world's population. The main cause of dental caries is the consumption of free sugars. A range of upstream, midstream, and downstream actions are therefore urgently needed to reduce free sugars consumption to prevent dental caries. Upstream legislation, regulation and fiscal action should be implemented whenever possible to effectively and sustainably reduce free sugars consumption at a population level.

References

G.B.D. Oral Disorders Collaborators, 2020. Global, regional, and national levels and trends in burden of oral conditions from 1990 to 2017: a systematic analysis for the global burden of disease 2017 study. J. Dent. Res. 99 (4), 362–373.

Heilmann, A., Machuca Vargas, C., Watt, R.G., 2021. Sugar consumption and oral health. In: Peres, M.A., Antunes, J.L.F., Watt, R.G. (Eds.), Oral Epidemiology: A Textbook on Oral Health Conditions, Research Topics and Methods. Springer International Publishing, Cham, pp. 307–317.

Heilmann, A., Tsakos, G., Watt, R.G., 2015. Oral health over the life course. In: Burton-Jeangros, C., Cullati, S., Sacker, A., et al. (Eds.), A Life Course Perspective on Health Trajectories and Transitions. Springer International Publishing, pp. 39–59.

Kickbusch, I., Allen, L., Franz, C., 2016. The commercial determinants of health. Lancet Global Health 4 (12), e895–e896.

Monteiro, C.A., Cannon, G., Moubarac, J.C., et al., 2018. The UN Decade of Nutrition, the NOVA food classification and the trouble with ultra-processing. Publ. Health Nutr. 21 (1), 5–17.

Moynihan, P.J., Kelly, S.A.M., 2014. Effect on caries of restricting sugars intake: systematic review to inform WHO guidelines. J. Dent. Res. 93 (1), 8–18.

Peres, M.A., Macpherson, L.M.D., Weyant, R.J., et al., 2019. Oral diseases: a global public health challenge. Lancet 394 (10194), 249–260.

Pitts, N.B., Zero, D.T., Marsh, P.D., et al., 2017. Dental caries. Nat. Rev. Dis. Prim. 3 (1), 17030.

Public Health England, 2020. National Dental Epidemiology Programme for England: Oral Health Survey of 5-Year-Olds 2019. Public Health England, London. Available at. https://www.gov.uk/government/statistics/oral-health-survey-of-5-year-old-children-2019.

Righolt, A.J., Jevdjevic, M., Marcenes, W., Listl, S., 2018. Global-, regional-, and country-level economic impacts of dental diseases in 2015. J. Dent. Res. 97 (5), 501–507.

WHO, 2015. Guideline: Sugars Intake for Adults and Children. World Health Organization, Geneva. Available at. https://www.who.int/publications/i/item/9789241549028.

WHO, 2022. Draft Global Strategy on Oral Health. Available at. https://apps.who.int/gb/ebwha/pdf_files/WHA75/A75_10Add1-en.pdf. (Accessed 6 June 2022).

Further reading

Abanto, J., Tsakos, G., Paiva, S.M., et al., 2014. Impact of dental caries and trauma on quality of life among 5- to 6-year-old children: perceptions of parents and children. Community Dent. Oral Epidemiol. 42 (5), 385–394.

Barbour, M.E., Lussi, A., 2014. Erosion in relation to nutrition and the environment. Monogr. Oral Sci. 25, 143–154.

Bartold, P.M., Van Dyke, T.E., 2013. Periodontitis: a host-mediated disruption of microbial homeostasis. Unlearning learned concepts. Periodontol. 2000 62 (1), 203–217.

Bernabé, E., Vehkalahti, M.M., Sheiham, A., et al., 2016. The shape of the dose-response relationship between sugars and caries in adults. J. Dent. Res. 95 (2), 167–172.

Borrell, L.N., Crawford, N.D., 2012. Socioeconomic position indicators and periodontitis: examining the evidence. Periodontol. 2000 58 (1), 69–83.

Boulton, J., Hashem, K.M., Jenner, K.H., et al., 2016. How much sugar is hidden in drinks marketed to children? A survey of fruit juices, juice drinks and smoothies. BMJ Open 6 (3), e010330.

Bragg, M.A., Roberto, C.A., Harris, J.L., et al., 2018. Marketing food and beverages to youth through sports. J. Adolesc. Health 62 (1), 5–13.

Chapple, I.L.C., Bouchard, P., Cagetti, M.G., et al., 2017. Interaction of lifestyle, behaviour or systemic diseases with dental caries and periodontal diseases: consensus report of group 2 of the joint EFP/ORCA workshop on the boundaries between caries and periodontal diseases. J. Clin. Periodontol. 44 (S18), S39–S51.

Conway, D.I., Purkayastha, M., Chestnutt, I.G., 2018. The changing epidemiology of oral cancer: definitions, trends, and risk factors. Br. Dent. J. 225 (9), 867–873.

Corvalán, C., Reyes, M., Garmendia, M.L., Uauy, R., 2019. Structural responses to the obesity and non-communicable diseases epidemic: update on the Chilean law of food labelling and advertising. Obes. Rev. 20 (3), 367–374.

Costa, S.M., Martins, C.C., Bonfim Mde, L., et al., 2012. A systematic review of socioeconomic indicators and dental caries in adults. Int. J. Environ. Res. Publ. Health 9 (10), 3540–3574.

Croker, H., Lucas, R., Wardle, J., 2012. Cluster-randomised trial to evaluate the "Change for Life" mass media/social marketing campaign in the UK. BMC Publ. Health 12 (1), 404.

Ejlerskov, K.T., Sharp, S.J., Stead, M., et al., 2018. Supermarket policies on less-healthy food at checkouts: natural experimental evaluation using interrupted time series analyses of purchases. PLoS Med. 15 (12), e1002712.

Fejerskov, O., Escobar, G., Jøssing, M., Baelum, V., 2013. A functional natural dentition for all–and for life? The oral healthcare system needs revision. J. Oral Rehabil. 40 (9), 707–722.
Gil-Montoya, J.A., de Mello, A.L., Barrios, R., et al., 2015. Oral health in the elderly patient and its impact on general well-being: a nonsystematic review. Clin. Interv. Aging 10, 461–467.
Gilchrist, F., Marshman, Z., Deery, C., Rodd, H.D., 2015. The impact of dental caries on children and young people: what they have to say? Int. J. Paediatr. Dent. 25 (5), 327–338.
Government of India, 2021. Ministry of Petroleum and Natural Gas. Roadmap for Ethanol Blending in India 2020–25: Report of the Expert Committee. Available at. https://www.niti.gov.in/sites/default/files/2021-06/EthanolBlendingInIndia_compressed.pdf.
Griffiths, M.L., Powell, E., Usher, L., et al., 2020. The health benefits and cost-effectiveness of complete healthy vending. PLoS One 15 (9), e0239483.
Griswold, M.G., Fullman, N., Hawley, C., et al., 2018. Alcohol use and burden for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 392 (10152), 1015–1035.
Hagenaars, L.L., Jevdjevic, M., Jeurissen, P.P.T., Klazinga, N.S., 2020. Six lessons from introducing sweetened beverage taxes in Berkeley, Cook County, and Philadelphia: a case study comparison in agenda setting and decision making. Health Pol. 124 (9), 932–942.
Harris, R., Gamboa, A., Dailey, Y., Ashcroft, A., 2012. One-to-one dietary interventions undertaken in a dental setting to change dietary behaviour. Cochrane Database Syst. Rev. (3), CD006540. https://doi.org/10.1002/14651858.CD006540.pub2.
Heilmann, A., Watt, R.G., 2017. Dental health. In: Buttriss, J.L., Welch, A.A., Kearney, J.M., et al. (Eds.), Public Health Nutrition, second ed. Wiley-Blackwell, pp. 298–309.
Jankauskiene, B., Narbutaite, J., 2010. Changes in oral health-related quality of life among children following dental treatment under general anaesthesia. A systematic review. Stomatologija 12 (2), 60–64.
Jevdjevic, M., Trescher, A.L., Rovers, M., Listl, S., 2019. The caries-related cost and effects of a tax on sugar-sweetened beverages. Publ. Health 169, 125–132.
Jones, A., Thow, A.M., Ni Mhurchu, C., et al., 2019. The performance and potential of the Australasian Health Star Rating system: a four-year review using the RE-AIM framework. Aust. N. Z. J. Publ. Health 43 (4), 355–365.
Knai, C., Petticrew, M., Durand, M.A., et al., 2015. Has a public–private partnership resulted in action on healthier diets in England? An analysis of the Public Health Responsibility Deal food pledges. Food Pol. 54, 1–10.
Lauber, K., Rutter, H., Gilmore, A.B., 2021. Big food and the World Health Organization: a qualitative study of industry attempts to influence global-level non-communicable disease policy. BMJ Glob. Health 6 (6).
Lawrence, H.P., Thomson, W.M., Broadbent, J.M., Poulton, R., 2008. Oral health-related quality of life in a birth cohort of 32-year olds. Community Dent. Oral Epidemiol. 36 (4), 305–316.
Lean, M.E.J., Te Morenga, L., 2016. Sugar and type 2 diabetes. Br. Med. Bull. 120 (1), 43–53.
Leite, F.R.M., Nascimento, G.G., Scheutz, F., López, R., 2018. Effect of smoking on periodontitis: a systematic review and meta-regression. Am. J. Prev. Med. 54 (6), 831–841.
Listl, S., Galloway, J., Mossey, P.A., Marcenes, W., 2015. Global economic impact of dental diseases. J. Dent. Res. 94 (10), 1355–1361.
Lussi, A., Carvalho, T.S., 2014. Erosive tooth wear: a multifactorial condition of growing concern and increasing knowledge. Monogr. Oral Sci. 25, 1–15.
Mackenbach, J.P., Lingsma, H.F., van Ravesteyn, N.T., Kamphuis, C.B., 2013. The population and high-risk approaches to prevention: quantitative estimates of their contribution to population health in The Netherlands, 1970–2010. Eur. J. Publ. Health 23 (6), 909–915.
Malam, S., Clegg, S., Kirwan, S., McGinigal, S., 2009. Comprehension and Use of UK Nutrition Signpost Labelling Schemes. Food Standards Agency, London.
Mansfield, J.L., Savaiano, D.A., 2017. Effect of school wellness policies and the Healthy, Hunger-Free Kids Act on food-consumption behaviors of students, 2006–2016: a systematic review. Nutr. Rev. 75 (7), 533–552.
Marinho, V.C., 2009. Cochrane reviews of randomized trials of fluoride therapies for preventing dental caries. Eur. Arch. Paediatr. Dent. 10 (3), 183–191.
Marinho, V.C., Higgins, J.P., Logan, S., Sheiham, A., 2003. Fluoride toothpastes for preventing dental caries in children and adolescents. Cochrane Database Syst. Rev. (1).
Meier, T., Deumelandt, P., Christen, O., et al., 2017. Global burden of sugar-related dental diseases in 168 countries and corresponding health care costs. J. Dent. Res. 96 (8), 845–854.
Mialon, M., Vandevijvere, S., Carriedo-Lutzenkirchen, A., et al., 2020. Mechanisms for addressing and managing the influence of corporations on public health policy, research and practice: a scoping review. BMJ Open 10 (7), e034082.
Micha, R., Karageorgou, D., Bakogianni, I., et al., 2018. Effectiveness of school food environment policies on children's dietary behaviors: a systematic review and meta-analysis. PLoS One 13 (3), e0194555.
Miranda-Filho, A., Bray, F., 2020. Global patterns and trends in cancers of the lip, tongue and mouth. Oral Oncol. 102, 104551.
Mishu, M.P., Tsakos, G., Heilmann, A., Watt, R.G., 2018. Dental caries and anthropometric measures in a sample of 5- to 9-year-old children in Dhaka, Bangladesh. Community Dent. Oral Epidemiol. 46 (5), 449–456.
Monse, B., Duijster, D., Sheiham, A., et al., 2012. The effects of extraction of pulpally involved primary teeth on weight, height and BMI in underweight Filipino children. A cluster randomized clinical trial. BMC Publ. Health 12, 725.
Moynihan, P., Makino, Y., Petersen, P.E., Ogawa, H., 2018. Implications of WHO Guideline on Sugars for dental health professionals. Community Dent. Oral Epidemiol. 46 (1), 1–7.
Moynihan, P., Miller, C., 2020. Beyond the chair: public health and governmental measures to tackle sugar. J. Dent. Res. 99 (8), 871–876.
Najeeb, S., Zafar, M.S., Khurshid, Z., et al., 2016. The role of nutrition in periodontal health: an update. Nutrients 8 (9).
Nanchahal, K., Vasiljevic, M., Petticrew, M., 2022. A content analysis of the aims, strategies, and effects of food and nonalcoholic drink advertising based on advertising industry case studies. Obes. Sci. Pract. 8 (2), 208–218.
Ng, S.W., Ni Mhurchu, C., Jebb, S.A., Popkin, B.M., 2012. Patterns and trends of beverage consumption among children and adults in Great Britain, 1986–2009. Br. J. Nutr. 108 (3), 536–551.
Pavia, M., Pileggi, C., Nobile, C.G., Angelillo, I.F., 2006. Association between fruit and vegetable consumption and oral cancer: a meta-analysis of observational studies. Am. J. Clin. Nutr. 83 (5), 1126–1134.
Petersen, P.E., Ogawa, H., 2012. The global burden of periodontal disease: towards integration with chronic disease prevention and control. Periodontol. 2000 60 (1), 15–39.
Qadri, G., Alkilzy, M., Franze, M., et al., 2018. School-based oral health education increases caries inequalities. Community Dent. Health 35 (3), 153–159.
Rebelo, M.A.B., Rebelo Vieira, J.M., Pereira, J.V., et al., 2018. Does oral health influence school performance and school attendance? A systematic review and meta-analysis. Int. J. Paediatr. Dent. https://doi.org/10.1111/ipd.12441.
Rodríguez-Molinero, J., Migueláñez-Medrán, B.D.C., Puente-Gutiérrez, C., et al., 2021. Association between oral cancer and diet: an update. Nutrients 13 (4).
Santonocito, S., Polizzi, A., Palazzo, G., et al., 2021. Dietary factors affecting the prevalence and impact of periodontal disease. Clin. Cosmet. Invest. Dent. 13, 283–292.
Scarborough, P., Adhikari, V., Harrington, R.A., et al., 2020. Impact of the announcement and implementation of the UK Soft Drinks Industry Levy on sugar content, price, product size and number of available soft drinks in the UK, 2015–19: a controlled interrupted time series analysis. PLoS Med. 17 (2), e1003025.
Schlueter, N., Luka, B., 2018. Erosive tooth wear—a review on global prevalence and on its prevalence in risk groups. Br. Dent. J. 224 (5), 364–370.
Schwendicke, F., Dörfer, C.E., Schlattmann, P., et al., 2015. Socioeconomic inequality and caries: a systematic review and meta-analysis. J. Dent. Res. 94 (1), 10–18.
Sheiham, A., James, W.P., 2015. Diet and dental caries: the pivotal role of free sugars reemphasized. J. Dent. Res. 94 (10), 1341–1347.
Solar, O., Irwin, A., 2010. A Conceptual Framework for Action on the Social Determinants of Health. World Health Organization, Geneva. Available at. https://apps.who.int/iris/handle/10665/44489.
Taillie, L.S., Reyes, M., Colchero, M.A., et al., 2020. An evaluation of Chile's Law of Food Labeling and Advertising on sugar-sweetened beverage purchases from 2015 to 2017: a before-and-after study. PLoS Med. 17 (2), e1003015.
Takahashi, N., 2005. Microbial ecosystem in the oral cavity: metabolic diversity in an ecological niche and its relationship with oral diseases. Int. Congr. 1284, 103–112.

Te Morenga, L., Mallard, S., Mann, J., 2013. Dietary sugars and body weight: systematic review and meta-analyses of randomised controlled trials and cohort studies. BMJ 346.
Warnakulasuriya, S., 2009. Global epidemiology of oral and oropharyngeal cancer. Oral Oncol. 45 (4), 309–316.
Warnakulasuriya, S., Greenspan, J.S., 2020. Epidemiology of oral and oropharyngeal cancers. In: Warnakulasuriya, S., Greenspan, J.S. (Eds.), Textbook of Oral Cancer: Prevention, Diagnosis and Management. Springer International Publishing, Cham, pp. 5–21.
Watt, R.G., Daly, B., Allison, P., et al., 2019. Ending the neglect of global oral health: time for radical action. Lancet 394 (10194), 261–272.
Watt, R.G., Listl, S., Peres, M.A., Heilmann, A., 2015. Social Inequalities in Oral Health: From Evidence to Action. University College London: International Centre for Oral Health Inequalities, Research & Policy. Available at. http://www.icohirp.com/monograph.html.
Watt, R.G., Sheiham, A., 2012. Integrating the common risk factor approach into a social determinants framework. Community Dent. Oral Epidemiol. 40 (4), 289–296.
WHO, 2020. Healthy Diet. Available at. https://www.who.int/news-room/fact-sheets/detail/healthy-diet. (Accessed 6 June 2022).
WHO, 2022b. Commercial Determinants of Health. Available at. https://www.who.int/news-room/fact-sheets/detail/commercial-determinants-of-health. (Accessed 22 May 2022).
WHO, 2022c. Global Database on the Implementation of Nutrition Action (GINA). Available at. https://extranet.who.int/nutrition/gina/en/home. (Accessed 22 May 2022).
World Cancer Research Fund, 2019. International Nourishing Database. Available at. https://policydatabase.wcrf.org/level_one?page=nourishing-level-one. (Accessed 22 May 2022).

Down's syndrome: Nutritional aspects

A. Laverty, Learning Disability Team, Coleraine, Co. Londonderry, UK

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of A. Laverty, Down's Syndrome: Nutritional Aspects, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 84–89, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00079-9.

Nutritional Status 284
Dietary Assessment 284
Laboratory Assessment 284
Carbohydrate metabolism 284
Protein Metabolism 285
Lipid Metabolism 285
Vitamins 285
Minerals 285
Anthropometric Assessment 285
Nutritional Requirements 286
Nutritional Therapy 286
Dietary Management 286
Dietary Guidelines 286
Feeding Behavior 286
Dental Problems 287
Obesity 287
Celiac Disease 288
Ageing 288
Care in the Community 289
Further Reading 289
Relevant Websites 289

Glossary

Bruxism Grinding of the teeth which can cause the teeth to wear and damage the tooth enamel through time.
Chromosome One of the thread-like structures in a cell nucleus that carry the genetic information.
Cystathionine β-synthase Enzyme that catalyses the first step of the transulfuration pathway from homocysteine to cystathionine.
Lipid peroxidation Refers to the oxidative degradation of lipids. It is the process whereby free radicals oxidize lipids in cell membranes resulting in cell damage.
Nonstarch polysaccharide (NSP) Complex carbohydrates other than starches found in foods. Contribute to dietary fiber in the diet. Insoluble NSP found in wheat, maize and rice and have a stool bulking effect. Soluble NSP found in oats, barley, rye and beans, can help lower blood cholesterol.
Osteoporosis Condition resulting in loss of bone tissue resulting in bones becoming brittle and liable to fracture.
Purine A nitrogenous compound with a two-ring molecular structure. Examples are adenine and guanine which occur in DNA structure, and uric acid.
Superoxide dismutase An enzyme which catalyses the dismutation of superoxide into oxygen and hydrogen peroxide. It is an important antioxidant defense in nearly all cells exposed to oxygen.
Trisomy A condition where there is one extra chromosome present in each cell in addition to the normal pair.

Down's syndrome, named after John Langdon Down, is a most widely recognized chromosomal disorder found in humans and falls into a category of chromosomal disruptions known as trisomies, hence, the other term for the condition; Trisomy 21. People with Down's syndrome vary widely in their abilities, but the syndrome is the most common genetic cause of intellectual disability.

More than 90% of Down's syndrome individuals have a total of 47 chromosomes in cells instead of the usual 46. The remaining cases are mainly either translocations, where there is a rearrangement of fragments of chromosomes, or mosaics in whom there are both normal and trisomic cells, that is, mosaic trisomy 21. There is a relationship between the frequency of Down's syndrome births and age, with both very young mothers and older mothers having a higher incidence of affected infants. It has been suggested that

nutrition may be implicated in the nondisjunction of the chromosomes. The additional chromosomal material in Down's syndrome usually comes from the mother, but it can, on occasion, come from the father. This observation may be indicative of hormonal changes in the older mother that reduce the likelihood of spontaneous abortion in an abnormal pregnancy.

The incidence of Down's syndrome is approximately 1 in 600–1000 live births. Prevalence is rising as life expectancy has improved over recent years with advancing medical knowledge and higher standards of care. In addition, women are having children at the older age and the incidence of Down's syndrome is increasing with increasing maternal age, although there are significant differences between various racial and social groups.

Physical defects common in Down's syndrome include congenital anomalies of the gastrointestinal tract (e.g., duodenal atresia and intestinal aganglionosis), which occur in approximately 12% of infants with Down's syndrome. Most of these anomalies require the neonate to be operated on immediately to allow nutrition. Congenital heart disease occurs in approximately 40% of infants with Down's syndrome. Children with congenital heart disease may present with failure to thrive, but after surgical repair of heart defects these children usually improve. Immune dysfunction, increased susceptibility to leukemia, and premature ageing with Alzheimer-like changes in the brain are major features of the syndrome.

Thyroid dysfunction is more common in people with Down's syndrome with the incidence increasing with age. Hypothyroidism is most frequently reported but hyperthyroidism can also occur. Correction of the thyroid function is essential to allow normal learning processes to take place and to help weight control.

There are many biochemical anomalies associated with the syndrome, mainly quantitative rather than qualitative. It is presumed that the overexpression of genes on chromosome 21 contributes to both the structural and functional pathology. Overdose effects of the genes already mapped to chromosome 21 are thought to alter pathways controlling the production of monocarbons, purines, pyrimidines, tubulins, and myelin.

The nutritional complications associated with Down's syndrome are summarized in **Table 1**.

Nutritional Status

It is debatable how relevant reference data from normal groups are for people with Down's syndrome.

Dietary Assessment

In children with Down's syndrome, conflicting reports have shown energy intake to be less than, similar to or greater than age-matched comparison groups, with a small percentage of children exceeding the recommended daily intake by more than 50%. However, as children with this syndrome tend to be shorter than age-matched children, energy intake comparisons need to be calculated per unit of body height.

Lower than recommended intakes of nonstarch polysaccharide coupled with higher than recommended consumption of protein and fats have also been reported. Researchers have reported low intakes of calcium, particularly in preschool and school-age children who refuse or limit milk consumption. Iron intakes have been reported to be low, particularly nonhem iron. Vitamins A and C intakes are limited in those who have a poor intake of fruit and vegetables. Intake of vitamin B has also been reported as low.

Laboratory Assessment

Carbohydrate metabolism

Fasting blood glucose levels are usually in the normal range, but the glucose tolerance curve has been reported to be flatter and often with a double humped curve, suggestive of delayed absorption. There is an increased incidence of both Type 1 (insulin dependent) and Type 2 diabetes in Down's syndrome. Studies have shown increased insulin resistance in obese and overweight females and adults with Down's syndrome but further research is required.

Table 1 Nutritional complications of Down's syndrome

Physical	Problems with muscle tone, oral health and dentition, chewing and swallowing
Metabolic	Anomalies in carbohydrate protein and lipid metabolism
	Increased demands on antioxidant defence system and methylation pathways
	Increased incidence of diabetes, celiac disease, obesity and thyroid disorders and leukaemia
	Growth retardation
Behavioral	Food consumption and exercise choices

Protein Metabolism

Disturbances in protein metabolism are common in Down's syndrome. An increased level of immunoglobulin A and immunoglobulin G antibodies to food antigens has been reported, and several studies have reported an increased prevalence of celiac disease. Abnormal levels of fasting plasma and urinary amino acids have been reported.

Lipid Metabolism

One study reported no significant differences between study and control groups, drawn from within the same families, in levels of total cholesterol, low-density lipoprotein, apolipoprotein B, and the apolipoprotein B to apolipoprotein A–I ratio. Triacylglycerol levels were significantly increased and serum high-density lipoprotein cholesterol to total cholesterol ratio significantly decreased in Down's syndrome. This suggests increased risk for coronary heart disease. The results of this and other studies reporting no difference between Down's syndrome and comparison groups in atherosclerosis, contrast with early reports that suggested a decreased incidence of coronary artery disease in Down's syndrome. It is not clear whether the differences reflect nutritional variables or population variable changes reflecting the increased survival rate in infancy.

There is evidence of increased lipid peroxidation in Down's syndrome.

Vitamins

Some studies have reported biochemical evidence of deficiency of thiamin, nicotinic acid, pyridoxine, cobalamin, folate, ascorbic acid, retinol, β-carotene, and α-tocopherol in patients with trisomy 21. Vitamin D metabolites have been reported to be in the normal range in a Spanish study that demonstrated wide seasonal variation linked to intensity of solar radiation.

Minerals

Low iron, calcium, manganese, and zinc blood concentrations have been reported in patients, and the iron to copper ratio has been reported to be decreased. Recent studies reported that intracellular zinc in blood mononuclear cells was approximately 47% lower than normal controls and it is possible that this could play a role in thyroid dysfunction, immunodeficiency, retarded growth, and faulty DNA repair. Low zinc status may contribute to the chemical disturbances that usually appear with ageing in individuals with Down's syndrome. Further research is required to determine whether zinc supplements are beneficial and if so at what dose. Supplementation with selenium aimed at increasing levels of the selenium-dependent enzyme glutathione peroxidase is reported to have led to a decrease in initially high blood mononuclear cell levels of copper, but did not affect iron or zinc.

Vitamin and mineral levels have been held to reflect not just nutrient intake, but also abnormal metabolism. Assessments of antioxidants and of oxidation by-products are useful indicators of nutritional status in people with Down's syndrome. The overexpression of the superoxide dismutase system, the purine synthesis pathway and cystathionine β-synthase are thought to create extra demands for antioxidants and for folate, but despite gene dosage effects the many biochemical anomalies that have been reported in people with Down's syndrome show a great deal of individual variation.

Anthropometric Assessment

Growth delay is one of the main characteristics of Down's syndrome but impaired growth velocity is particularly evident at certain stages of development.

The fetal growth has usually been reported to be relatively normal and the length of the neonate is often within normal limits, allowing for gestation. Some studies have reported the prenatal growth delay and a major Italian study comparing neonatal length, weight, head circumference, and weight/length squared reported all percentiles of growth variables lower in Down's syndrome infants except for weight/length squared percentiles.

At approximately 6 months of age, when growth starts to become regulated by growth hormone, growth velocity usually begins to show a marked reduction from normal levels. Although for the Down's syndrome child the period between birth and 2 years and the period between 6 years and 10 years of age are times of accelerated growth, the deviation from normal levels remains significant. Slow growth velocity is also a particular feature of adolescence, although there is a pubertal growth spurt. The deviation of adult stature from the means of reference groups is greater than the deviations in early infancy.

The short stature in Down's syndrome seems to be mainly the result of the impaired growth of the long bones of the leg, as sitting height measurements show that the growth of the vertebral column is closer to normal.

Why there is growth delay in Down's syndrome is not entirely clear, and several hypotheses have been advanced. Both human growth hormone therapy and zinc sulfate supplementation of the diet have been reported to accelerate the growth.

Children with Down's syndrome tend to be not only shorter, but also heavier than reference children. Charting the height and weight of a child with Down's syndrome using reference norms from the general population, will show the abnormality of the growth pattern. However, it is more useful clinically, to compare the height and weight of an individual against syndrome-specific norms, as that will show up any deviation from the growth patterns of children with Down's syndrome.

Italian percentile charts have been drawn up for neonates with Down's syndrome based on a large sample of consecutively born infants. The specific growth charts for children with Down's syndrome have been constructed based on anthropometric assessments of US children, Sicilian children (thought to be representative of southern European children) and Dutch children (thought to be representative of northern European children). On average the Dutch children were taller than the US children and the US children were taller than the Sicilians. More recently growth charts have been developed for UK children and in the US work is currently being undertaken to update growth charts to encompass current research that shows children with Down's syndrome are growing better.

Nutritional Requirements

Children and adults with Down's syndrome need the same range of nutrients as the general population. Energy intake standards based on age groups are not appropriate for children with Down's syndrome. Energy intakes in both children and adults need to be tailored to height and weight and to the physical activity.

Nutritional Therapy

In the 1970s and 1980s hopes were raised that megadoses of vitamins and minerals would boost intelligence in children with Down's syndrome, but rigorous studies have shown these doses lead neither to higher intelligence nor to better health. In addition, there is anxiety about possible side effects, particularly of the fat-soluble vitamins.

As more has been learned about the genes on chromosome 21, interest shifted to targeted nutritional intervention aimed at correcting the metabolic anomalies that are common in Down's syndrome owing to genetic overexpression, with the emphasis on nutrients to maintain health and prevent disease. Targeted nutritional supplementation with vitamins, minerals, amino acids, digestive enzymes, and essential fatty acids remains controversial. Clinicians have reported differences between children treated and not treated in health, growth, and cognitive and speech functions but few well-designed studies have been performed.

Dietary Management

Dietary Guidelines

Dietary recommendations are as for the general population until research proves otherwise. There are, as yet, no specific dietary guidelines for the woman pregnant with a Down's syndrome child. Periconceptual folic acid supplements of 400 μg daily in addition to folate-rich foods may be beneficial in decreasing risk of Down's syndrome, as well as neural tube effects. There are also indications that antioxidant and essential fatty acid intake may be particularly important, but at present dietary advice is the same as for other pregnant women.

The situation is similar for infant feeding. Brain lipids in the human infant are known to change with changing intakes of fatty acids. The needs of a newborn with Down's syndrome for the long-chain polyunsaturated fatty acids docosahexenoic acid and arachidonic acid have not yet been determined. Because breast milk contains the preformed dietary very long-chain fatty acids that seem to be essential for the development of the brain and the retina, it seems prudent to encourage breast-feeding.

The antioxidant defense system has a particularly important role in Down's syndrome and parents and caregivers can be advised on providing a diet rich in antioxidants. Dietary intakes need to be considered for the sulfur amino acids (which are needed for glutathione synthesis), of fat-soluble vitamins A, C, and E, water-soluble vitamins B_6, B_{12}, and folic acid, and of the minerals selenium and zinc. In latitudes where no vitamin D is synthesized in the winter months, it is particularly important to ensure exposure to sunlight during summer months to maintain adequate stores of the vitamin throughout the year as recent studies indicate an increase in the incidence of osteoporosis in Down's syndrome. Those who are housebound or have poor mobility may benefit from a vitamin D supplement.

Feeding Behavior

Feeding skills tends to be delayed in the young child with Down's syndrome, but the sequence of the emergence of the skills is the same as with other children if appropriate learning opportunities are provided.

Infants with Down's syndrome have a smaller oral cavity, which makes it easier for liquids to spill from the sides of the mouth. If a child is hypotonic, the tongue is likely to flatten out when the child sucks instead of forming a groove round the nipple, so the child will have a weak suck, may gag, and milk will leak from the mouth. Feeding will be exhausting, and particularly where the child has a cardiac defect, the child may have difficulty taking in enough milk to meet energy requirements. Tube feeding may be necessary until the child develops better tongue control. Feeding will be easier if the infant is wide awake and extra support for the infant during feeding, in particular supporting the infant's chin to help steady the jaw, can all help encourage intake. Because of the benefits of breast-feeding, it is essential that nursing mothers are given help and advice when their infants have initial difficulties. Breathing during feeding may be helped if the mouth and nose are cleared of mucus with a syringe before feeding.

As with other children, it is important to introduce textured food when the child is developmentally ready, and information should be provided for parents and caregivers regarding appropriate expectations and helpful feeding techniques as well as dietary

advice. In children with Down's syndrome, poor neuromotor control of the tongue may result in the continued use of pureed food. There may be slow initiation of the swallow response, possibly because of hypotonic pharyngeal muscles, and oral sensitivity problems may also make the transition to textured foods difficult. Persistent feeding problems merit multidisciplinary assessment and therapy. Impaired swallow can result in food being aspirated and contribute to respiratory problems. The presence of the tongue protrusion reflex past the age of 12–18 months can result in delayed progression to solid food and can contribute to malocclusion of teeth. Also, dental abnormalities can exacerbate difficulties with chewing and can contribute to poor nutrition, because children who have problems chewing may be offered soft often high-energy food and be given little opportunity to accept meats, fresh fruits and vegetables, which are lower in energy.

Fresh fruit and vegetables provide the nonstarch polysaccharide that can help prevent the constipation common in Down's syndrome. Fruit juices and water between meals also help with constipation. Because the hypotonia in Down's syndrome also contributes to sluggish bowel habits, this is another reason for children and adults to be encouraged to take part in physical activity. If constipation does not respond to dietary management, there should be a medical assessment to exclude gastrointestinal and thyroid problems.

Dental Problems

Dental anomalies in Down's syndrome include changes in the tooth structure, reduced total number of teeth and delayed or abnormal eruption. Together with the physical abnormalities of the facial appearance and oral cavity, these can all impact on feeding. Dental disease is common in Down's syndrome as teeth are more at risk of wear through bruxism and decay due to fragile enamel. In addition, gum disease (gingivitis), and oral infections due to mouth breathing can lead to teeth becoming loose and falling out. A healthy balanced diet, including plenty of fruit and vegetables, low in sugar-containing fluids and fizzy drinks (including 'diet' varieties), and avoiding frequent snacks will help preserve teeth.

Obesity

Obesity is common in Down's syndrome having been reported from different cultures and ethnic backgrounds. From Australian and North American studies it has been reported that by 2–3 years of age more than 30% of children with trisomy 21 are overweight and that by 9 years of age the average child with Down's syndrome is obese.

High rates of overweight and obesity have been reported in adults with Down's syndrome, both living in the community and at home, and more commonly in females than males. Overweight and obesity are particularly associated with living in the family home compared to supervised community units or hospital, but they are not significantly associated with the degree of learning disability.

Because excessive weight gain in childhood often leads to adult obesity, it is important to encourage healthy choices in childhood. Why children with Down's syndrome have a tendency to become fat is not clear, but probably several factors influence the weight gain. Retardation of growth resulting in short stature may be of prime importance. Obesity in people with Down's syndrome has also been linked with several physiologic features of the syndrome (Table 2).

Prepubescent children with Down's syndrome have a decreased resting metabolic rate compared to control children matched for the body mass index. Children of approximately the same body composition, whether or not they have Down's syndrome, expend similar levels of energy in the physical activity. Since obesity is negatively correlated with the motor performance, it is likely to lead to a reduction in sporting and physical recreation activities, and thus obesity has social as well as health implications, in children with Down's syndrome as in other children. However, children and adolescents with Down's syndrome have been shown to have difficulty with sustained physical exercise in both laboratory and recreational situations, and this has been attributed to physiological impairments, notably cardiovascular, as well as to lack of motivation.

Children, adolescents and adults with Down's syndrome have a deficit in isokinetic strength, and by the age of 14 years adolescents with testosterone levels in the normal range fail to show the pubertal muscle strength increase. Progressive resistance exercise

Table 2 Factors predisposing to obesity in Down's syndrome

	Increased	*Decreased*
Poor eating behavior	↑	
Calorie intake	↑	
Resting metabolic rate		↓
Muscle tone		↓
Exercise		↓
Thyroid function		↓
Substrate fat oxidation		?↓
Leptin levels	?↑	

programs can help to build muscle strength, and regular aerobic exercise will improve exercise tolerance. Often individuals can attain high standards in competitive gymnastics and swimming. The overexpression of collagen genes on chromosome 21 affects both muscle and connective tissue, and it has been claimed that targeted nutritional treatment leads to rapid improvement in both muscle strength and joint stability.

In a cross-sectional study of men and women with Down's syndrome, the body mass index declined with increasing age. Further research is needed to clarify whether individuals lose weight as they age, or whether there is a shorter life expectancy for individuals with higher body mass indices.

Celiac Disease

There is an increased incidence of celiac disease in Down's syndrome compared with the general population. Symptoms such as gut dysfunction (both diarrhea and constipation), abdominal bloating, dyspepsia, mouth ulceration, mood change, arthritis, general fatigue, and mild anemia may be indicators of celiac disease, although many of these are common in Down's syndrome without celiac disease. Blood tests such as antiendomysial antibodies (AEA) and tissue transglutaminase antibody status can give an indication of celiac disease but are not conclusive. Definitive diagnosis remains the identification of villous atrophy on small bowel biopsy. Quality of life can be greatly improved on a gluten-free diet so implementation following diagnosis and good compliance should be encouraged.

Ageing

The rapid ageing that characterizes Down's syndrome is in line with the accumulating evidence that many degenerative diseases are associated with deleterious activated oxygen species reactions. Activated oxygen species can damage genetic material and inactivate membrane-bound enzymes as well as cause lipid peroxidation in cell membranes. Of particular relevance to Down's syndrome is the evidence relating to cancer, inflammatory joint disease, diabetes, degenerative vascular disorders, degenerative eye disease and senile dementia – all reported to have increased prevalence in Down's syndrome.

The gene for copper/zinc-superoxide dismutase is on chromosome 21, and copper/zinc-superoxide dismutase levels are elevated by 50% in a range of cells of people with Down's syndrome, including erythrocytes, blood platelets, leucocytes, and fibroblasts. The increase has also been reported in fetal cerebral cortical cells. Although copper/zinc-superoxide dismutase usually functions as an antioxidant it seems likely that in Down's syndrome the raised levels lead to oxidative stress. When the increased production of hydrogen peroxide through catalysis of superoxide free radicals is not matched by a sufficient increase in glutathione peroxidase to metabolize the additional hydrogen peroxide to water and oxygen, there is thought to be an increase in highly reactive hydroxyl radicals leading to increased lipid peroxidation.

Fibroblasts derived from people with Down's syndrome show elevated lipid peroxidation, and levels of thiobarbituric reaction products, which indicate extent of lipid peroxidation, have been reported to be raised in erythrocytes from Down's syndrome subjects compared to control subjects.

A reported increase in the activity of the hexose monophosphate pathway in Down's syndrome is thought to be a compensatory mechanism to deal with increased hydrogen peroxide, allowing greater production of the reduced form of nicotinamide-adenine-dinucleotide phosphate, thus improving the ability of cells to reduce oxidized glutathione. However, it has been suggested that this shift of glucose utilization from energy production to reducing power may compromise cellular cation pumps.

Among the genes so far identified on chromosome 21 is that for β-amyloid precursor protein. Amyloidosis is evident in the brain tissue of both patients with Alzheimer's disease and those with Down's syndrome. Studies are investigating the implications of the anomalies in the expression of the β-amyloid precursor protein and also of the effect on cobalamin/folate metabolism of the gene for the enzyme cystathionine β-synthase, also on chromosome 21. The overexpression of both these genes is believed to contribute substantially to the development of dementia. Although all people with Down's syndrome have evidence of brain pathology similar to Alzheimer's disease by their early thirties, not all show Alzheimer-like behavior changes as they age.

It may be that an increase in dietary antioxidants could delay the onset of Alzheimer's type symptoms but more research is required. However, standard dietary recommendations for healthier lifestyles (i.e., eating more fruit and vegetables and including more oily fish in the diet) may have the added potential benefits of increasing antioxidant intake. Unfortunately these are often the foods least favored by individuals with Down's syndrome.

Low vitamin E levels have been found to be associated with dementia, not only in the elderly but also in those with Down's syndrome. Vitamin E may have a potential therapeutic role in Alzheimer's like neurological changes by protecting the integrity of the muscarinc receptors. In addition, vitamin E is a strong antioxidant. Likewise, vitamin B_{12} deficiency has been reported in cases of Down's syndrome and Alzheimer's disease; however, replacement therapy does not change the evolution of the underlying disease. Continuing research into the etiology of the Down's syndrome phenotype is expected to lead to advances in the treatment of both Down's syndrome and Alzheimer's disease.

Care in the Community

Most people with Down's syndrome live in the community; some live with parents or caregivers, but adults often live independently or semi-independently. Many people with Down's syndrome can learn about healthy eating and manage their own diets. A dietitian's role in a community learning disability support team is likely to encompass not only individual assessment, but also teaching and educating people with Down's syndrome as well as parents, caregivers and other professionals.

Further Reading

Ani, C., Grantham-McGregor, S., Muller, D., 2000. Nutritional supplementation in Down's syndrome: Theoretical considerations and current status. Developmental Medicine and Child Neurology 42, 207–213.

Antila, E., Westermarck, T., Huovinen, K., Lehto, J., Johansson, E., 1996. Indications for nutritional supplementation in Down's syndrome, 10th World Congress of the International Association for the Scientific Study of Intellectual Disabilities. Trends in Biomedicine in Finland 7 (supplement), 50–53.

Bell, E.J., Haidonis, J., Townsend, G.C., 2002. Tooth wear in children with Down's syndrome. Australian Dental Journal 47 (1), 30–35.

Cremers, M.J.G., van der Tweel, I., Boersma, B., Wit, J.M., Zonderland, M., 1996. Growth curves of Dutch children with Down's syndrome. Journal of Intellectual Disability Research 40 (5), 412–420.

Hennequin, M., Faulks, D., Veyrune, J.L., Bourdiol, P., 1999. Significance of oral health in persons with Down syndrome: A literature review. Developmental Medicine and Child Neurology 41, 275–283.

Hewitt, P., Smith, D., 1983. Down's syndrome – nutritional aspects. In: Macrea, K., Robinson, R.K., Sadler, M.J. (Eds.), Encyclopaedia of Food Science, Food Technology and Nutrition, vol. 2. Academic Press, London.

Lejeune, J., 1990. Pathogenesis of mental deficiency in trisomy 21. American Journal of Medical Genetics 7 (supplement), 20–30.

Melville, C.A., Cooper, S.-A., McGrother, C.W., et al., 2005. Obesity in adults with Down syndrome: A case-control study. Journal of Intellectual Disability Research 49 (2), 125–133.

Pipes, P.L., 1988. Feeding management of children with Down syndrome. In: Dmitriev, V., Oelwein, P. (Eds.), Advances in Down's Syndrome, ch. 16. Special Child Publications, Seattle.

Piro, E., Pennino, C., Cammarata, M., et al., 1990. Growth charts of Down syndrome in Sicily: Evaluation of 382 children 0–14 years of age. American Journal of Medical Genetics 7 (supplement), 66–70.

Prasher, V.P., 1995. Overweight and obesity amongst Down's syndrome adults. Journal of Intellectual Disability Research 39 (5), 437–441.

The UK (2000) *Growth Charts – Growth Charts for Down's Syndrome*. Maxwell Street, South Shields, Tyne & Wear, NE33 4PU, UK: Harlow Printing Ltd.

Relevant Websites

www.dsmig.org.uk. – Down's Syndrome Association.
www.downs-syndrome.org.uk. – Down's Syndrome Medical Interest Group.
www.ndss.org. – National Down's Syndrome Society.

Eating disorders: Anorexia nervosa

Meredith Kells and Jennifer E. Wildes, Department of Psychiatry & Behavioral Neuroscience, University of Chicago, Chicago, IL, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A.R. Rolla, Eating Disorders: Anorexia Nervosa, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 113–119, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00083-0.

Introduction 290
Diagnostic criteria 290
Other clinical characteristics 291
Epidemiology 291
Pathophysiology 292
Differential diagnoses 292
Screening 292
Clinical presentation and physical examination 292
Common laboratory findings 293
Testing 293
Management 293
Conclusion 294
Funding information 294
References 294
Further reading 295

Key points

- Anorexia Nervosa (AN) is a serious psychiatric disorder associated with significant medical morbidity and higher mortality rates relative to the general population
- AN is characterized by energy restriction, fear of weight gain, behaviors to inhibit weight gain, and body image disturbance
- AN is associated with high rates of self-harm, suicidality, and co-morbid psychiatric diagnoses, in particular mood and anxiety disorders and substance abuse
- The effects of malnutrition are seen across body systems, with common findings including cardiovascular compromise (bradycardia, orthostatic hypotension, and less commonly, arrhythmias), amenorrhea, low bone mineral density (BMD), and gastrointestinal complaints.
- Treatment for AN may occur on a spectrum from outpatient to inpatient psychiatric interventions; all levels of care should include an interdisciplinary team of providers.

Introduction

Anorexia nervosa (AN) is a serious psychiatric disorder classified under the category of Feeding and Eating Disorders in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; American Psychological Association, 2013). The key characteristics of AN are centered on an individual's perception of their weight and subsequent behaviors to avoid weight gain resulting in low weight. Full diagnostic criteria are described below. Individuals with AN commonly present with co-morbid psychiatric disorders and physiologic manifestations of malnutrition across body systems. AN is associated with elevated risk of death compared to the general population consequent to effects of malnutrition and suicide (Keshaviah et al., 2014).

Diagnostic criteria

AN is a distinct diagnosis outlined under the category of Feeding and Eating Disorders in DSM-5 (APA, 2013). Although an individual may be diagnosed with more than one eating disorder during their lifetime, concurrent disorders are not possible within the framework of current diagnostic criteria. Diagnosis of AN per DSM-5 criteria is focused on the following features:

- Energy restriction (restriction of food intake) leading to body weight that is significantly low for that individual by age, sex, growth trajectory, health, or what is minimally normal or expected if a child or adolescent.

- Profound fear of gaining weight or becoming fat with behavior that prevents weight gain despite low weight.
- Body image disturbance, significant self-evaluation tied to weight or shape, or poor insight into the seriousness of weight loss/low body weight

The DSM-5 was published in 2013 with notable differences from the diagnostic criteria in the previously published Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition Text Revision. First, the discrete cutoff of 85% of ideal body weight (IBW) was removed as an exemplar of low body weight and replaced with language describing "significantly low" as "less than minimally normal, or for children and adolescents, less than that minimally expected" (APA, 2013, p. 338). Additionally, the amenorrhea criterion was removed in consideration of younger adolescents who are premenarchal, those who have irregular menses or body composition (body fat) that influences menstruation, and males. Further revisions included adding persistent behavior that inhibits weight gain as an index of fear of weight gain, and replacing "denial" of the seriousness of low body weight with "persistent lack of recognition" of the seriousness of low body weight (APA, 2013).

DSM-5 describes two subtypes of AN: restricting and binge-eating/purging. Individuals with the restricting AN subtype lose weight or maintain low weight through dietary restriction and exercise and have not engaged in binge-eating or purging in the past 3 months. Alternatively, those with the binge-eating/purging subtype have experienced recurrent binge eating or purging through laxatives, vomiting, diuretics, enemas, etc. within the 3 months prior to diagnosis.

The International Classification of Diseases 11th Revision (ICD-11) also includes restricting and binge-eating/purging AN subtypes along with weight-related diagnostic qualifiers of "significantly low weight" and "dangerously low weight." Notably, however, the ICD-11 does not provide set cutoffs for "significantly low" or "dangerously low" when evaluating weight status. Per the DSM-5, in adults severity of weight status is conceptualized using World Health Organization (WHO) classifications as follows: mild—body mass index (BMI) > 17 kg/m^2, moderate—BMI of 16–16.99 kg/m^2, severe—BMI 15–15.99 kg/m^2, and extreme—BMI < 15 kg/m^2. Importantly, clinicians should consider level of functioning and symptom profile, in addition to weight, when determining severity of AN. For children and adolescents, BMI percentile and percent median BMI are used, and should consider a child's previous growth trajectory when determining severity of illness (Society for Adolescent Health and Medicine [SAHM], 2015).

Likewise, atypical anorexia nervosa (AAN) is diagnosed in individuals who meet criteria for AN with regards to energy restriction, fear of weight gain, and body image disturbance yet do not exhibit low weight as per WHO classifications for BMI described above. Those with AAN may present with significant weight loss or, in the case of youth, fail to continue on the expected growth trajectory. This presentation would fall under the DSM-5 category of Other Specified Feeding or Eating Disorder (OSFED) with the specific designation of AAN.

Other clinical characteristics

Individuals with AN commonly present with comorbid psychiatric disorders. One survey reported that 56.2% of individuals with AN met criteria for at least one other DSM-IV psychiatric disorder, of which mood and anxiety disorders (including obsessive compulsive disorder), and substance abuse were most prevalent (Andrés-Pepiñá et al., 2020). Longitudinal studies show that over one-third of individuals (39%) who recover from AN continue to exhibit other psychiatric disorders at follow-up (17–29 years post presentation; Andrés-Pepiñá et al., 2020).

Additionally, individuals with AN experience high rates of self-harm. As many as 21% of individuals with AN engage in non-suicidal self-injury (NSSI) in their lifetime (Cucchi et al., 2016). High mortality rates in AN are due, in part, to death by suicide (Keshaviah et al., 2014), with greater risk of suicide attempt in individuals with the binge-eating/purging subtype (Mandelli et al., 2019). Moreover, suicide attempts in individuals with AN are reported to have a greater level of intent and potential lethality than those found in other eating disorder diagnoses (Guillaume et al., 2011). Suicide is the cause of death for 1 in 5 individuals who die as a result of AN; this population has been found to be 31 times more likely to die by suicide than age matched peers (Preti et al., 2011).

Epidemiology

Lifetime prevalence of AN is estimated to be between 0.3 and 2.2% (Bulik et al., 2006; Hudson et al., 2007). The peak incidence is during adolescence, specifically 15–19 years of age and is more common in females than males (Hoek and van Hoeken, 2003). Although eating disorders are perceived to affect only white individuals, they are found across all racial and ethnic groups (Marques et al., 2011). Non-white individuals with eating disorders may present differently than white women and are less likely to receive treatment (Marques at al., 2011). Additionally, AN may be more common in sexual and gender minorities than cisgender, heterosexual peers (Nagata et al., 2020).

There is no consensus definition of recovery from AN. It has been suggested that AN exists along a spectrum from subsyndromal (body image concerns and eating restriction without changes in weight) to full syndrome, as well as a spectrum from partial remission with "minimal" restriction through full remission with no behaviors present (Steinglass et al., 2020). Commonly, recovery from AN is defined as normalization of disordered behaviors such as eating patterns, resolution of dietary restriction or binge eating,

improvement in cognitive distortions (e.g. body image disturbance), and physiologic recovery (e.g. weight) and may be episodic (Steinglass et al., 2020). It has been suggested that 46% of individuals fully recover from AN and up to 20% develop chronic illness, although studies with adolescent-onset AN show higher rates of recovery (up to 65%; Andrés-Pepiñá et al., 2020). Relapse is most common in the first 2 years following treatment discharge. Prognosis is better for those who enter treatment at a young age or those with shorter duration of illness prior to entering treatment (Andrés-Pepiñá et al., 2020).

Pathophysiology

The development of AN is likely a complex bio-psycho-social phenomenon involving genetic predisposition in combination with environmental factors. Recent genome wide association studies of nearly 17,000 cases suggest a strong genetic influence and posit AN as a metabo-psychiatric disorder (Watson et al., 2019).

Within the environment, media-driven societal pressure to conform to the "thin ideal" (romanticized thin body type) likely contributes to the development of restrictive eating disorders. The notion that life would be better if one was thinner, or internalization of the thin ideal, and greater exposure to media sources promoting thin bodies are associated with increased eating pathology. Resultant dieting behavior in childhood is a key predictor of the onset of eating disorders (Culbert et al., 2015).

Additional environmental factors include parenting style, parental personality, and history of adverse childhood events (ACEs; (Culbert et al., 2015). Personality traits associated with eating disorder development include negative emotionality, perfectionism, drive for thinness, poor interoceptive awareness, and obsessive-compulsiveness, neuroticism, impulsivity, negative urgency, and inhibitory control have been implicated for those with binge-purge behaviors. Additionally, neurotransmitter disturbances and regulation have been suggested as contributing factors to eating disorder development (Culbert et al., 2015).

Differential diagnoses

When considering a diagnosis of AN, primary medical and psychiatric illnesses must be ruled out. Medication use, gastrointestinal diseases, food allergies and intolerances, thyroid abnormalities, new onset Type 1 diabetes mellitus, malignancy, and HIV/AIDs can cause weight loss. Psychiatric disorders that have overlapping symptoms also should be considered, including anxiety disorders and phobias that inhibit intake, depressive disorders, and schizophrenia-related erratic eating. Additionally, drug use that suppresses appetite (e.g. stimulants, cocaine) must be ruled out.

Screening

There are a number of instruments available to aid clinicians in screening for the presence of eating disorder pathology. These tools are not appropriate for diagnosing AN, but can be used to identify patients in need of further clinical assessment or referral to a specialist. Very brief, self- or clinician-administered screening tools, like the SCOFF Questionnaire (Morgan et al., 1999) and Eating Disorder Screening in Primary Care (ESP-PC; Cotton et al., 2003), often are used in medical settings. More thorough assessments include the Eating Attitude Test-26 (EAT-26; Garner et al., 1982), Eating Disorder Examination-Questionnaire (EDE-Q; Fairburn and Beglin, 1994), and the Eating Disorder Diagnostic Scale (EDDS; Stice et al., 2000), all of which are self-administered.

Clinical presentation and physical examination

Presentation may vary depending on clinical setting; however, it is imperative that clinicians across settings are aware of warning signs and can identify potential cases for early intervention. Clinicians should consider an eating disorder in any individual with new or rapid weight loss, growth stunting in adolescents or children, new dieting behavior or compulsive exercise, rigid or obsessive food habits, or negative thoughts about weight or shape.

Effects of malnutrition may be seen regardless of weight or BMI. Individuals may present to care based on symptoms such as dizziness or fainting, gastrointestinal distress, and changes in menstrual patterns (amenorrhea). Clinicians might see indications of purging behavior, such as Russell's sign (abrasion on the knuckles of the hands due to repeated self-induced vomiting), dental enamel erosion, or enlarged parotid glands. Hemodynamic instability, arrhythmias, dehydration, peripheral edema, bloating, gastric distention, xerosis, hair loss, lanugo, cold intolerance, brittle hair and nails, and subcutaneous fat loss also may be noted on exam (Westmoreland et al., 2016).

Initial assessment should include a complete history including medical, family, and social histories. Nutrition history including changes in eating patterns, weight loss, and 24 h dietary recall, and menstrual history, may elucidate warning signs as well. Growth charts for children and adolescents should be reviewed or weight history for adults. Complete physical examination including height, weight, BMI, vital signs, laboratory studies, urinalysis, and electrocardiogram (EKG) should be completed (Campbell and Peebles, 2014).

Common laboratory findings

The effects of malnutrition can lead to a number of laboratory abnormalities and may be related to a specific behavior (e.g. hypokalemia induced by purging). Patient-specific factors such as illness severity, weight status, medications, pre-morbid or co-morbid illness, and eating disorder behaviors may guide clinicians in type and frequency of laboratory monitoring.

Chemistries frequently are monitored closely to evaluate electrolytes tied to nutritional intake and starvation status. Refeeding syndrome, or shifting serum phosphorus, potassium, magnesium, and glucose in a starved person on restarting adequate nutrition, is a rare but potentially fatal complication of AN. Refeeding hypophosphatemia is a common finding during nutrition restoration and is a hallmark of impending refeeding syndrome (O'Connor and Nicholls, 2013). Additionally, clinicians should evaluate potassium levels in relation to purging through vomiting or diuretic use, with those who are underweight and using diuretics being most at risk. Low potassium and elevated bicarbonate seen in metabolic alkalosis are indicative of vomiting and can lead to serious cardiac arrhythmias. Low serum chloride may also be seen in those who vomit frequently. Dehydration may result in a relationally elevated blood urea nitrogen (BUN), creatinine, potassium, sodium, calcium, and albumin. Dangerously low sodium can be due to diuretic misuse, and rarely in those who water load (overhydrate) prior to appointments in an attempt to falsify weight.

The most common liver complications include elevation in aspartate aminotransferase (AST) and alanine aminotransferase (ALT), which show an inverse relationship with body weight (as body weight decreases, AST and ALT increase). Less commonly, elevated AST and ALT may occur during nutrition rehabilitation, causing fat and glucose deposition in the liver leading to hepatic steatosis. Bilirubin, alkaline phosphatase, and international normalized ratio (INR) may be implicated as starvation causes liver injury at low weight. Liver glycogen stores and impaired gluconeogenesis related to starvation-mediated liver injury also contribute to hypoglycemia. Despite these weight-induced changes, albumin most often remains normal. However, pre-albumin is a reliable marker of malnutrition and has been shown to be predictive of serious outcomes in severely malnourished individuals with AN.

The complete blood count may reveal a number of abnormalities related to malnutrition. Leukopenia (low white blood cell count), anemia, and thrombocytopenia (low platelets) may be related to insufficient intake. Elevated white blood cell count is seen with infection, inflammation, stress, or anemia. Elevated red blood cell count may reflect dehydration.

Malnutrition results in alterations in the hypothalamic-pituitary-thyroid axis with resultant amenorrhea, low androgens, and low estrogen levels. Additionally, laboratory findings may show an elevated growth hormone (GH) with low levels of insulin-like growth factor-1 (IGF1), and hypercortisolemia. Thyroid stimulating hormone (TSH) may be normal or low, as well as low free T4 and T3 related to adaptation to low energy availability and metabolic rates and normalizes with weight restoration (Schorr and Miller, 2017). Parathyroid hormone (PTH) levels are commonly low, and have importance when considering calcium, PTH, and vitamin-D mediated bone health implications for this population (Lenherr-Taube et al., 2020).

Testing

Cardiovascular complications due to nutrition-related structural changes (e.g. myocardial atrophy) as well as electrolyte imbalances may lead to serious, and potentially deadly consequences in individuals with AN. Routine electrocardiograms (EKG) are recommended to monitor for conduction and repolarization abnormalities; however, echocardiograms are not typically necessary unless EKG shows signs of pericardial or valve disease. EKGs often show bradycardia (abnormally slow heart rate), in part due to autonomic dysfunction (Sachs et al., 2016).

Low bone mineral density (BMD) is a result of malnutrition, vitamin D deficiency, low IGF1 and hypocortisolemia (Solmi et al., 2016). These effects may have long-term implications as peak bone mass is achieved during adolescence, coinciding with peak incidence of AN, and may not be recoverable later in life leading to higher rates of osteoporosis and fractures than in healthy peers. As amenorrhea and illness chronicity are associated with low BMD in this population (Solmi et al., 2016), dual energy x-ray absorptiometry measurements are recommended for those who are not menstruating or with long duration of illness.

Management

Management of AN is best achieved with an interdisciplinary team including, at minimum, mental health, medical, and nutrition services (Ozier et al., 2011). Ideally, the treatment team communicates regularly to monitor progress and set goals in conjunction with the individual and family. As outcomes are improved for those who receive early, comprehensive care (Andrés-Pepiñá et al., 2020), identification and timely treatment are key.

Level of care for AN is on a continuum from outpatient to inpatient psychiatric hospitalization. Achieving goals while maintaining safety in the least restrictive setting is preferable to higher levels of care; however, escalating care may be necessary for individuals who fail to meet outpatient goals (e.g. weight gain, behavioral improvement), have safety concerns, or are medically unstable (Buchman et al., 2019).

In the outpatient setting, family-based treatment (FBT) is the first-line intervention for children and adolescents with AN (Buchman et al., 2019). FBT is a manualized, three phase, stepwise approach in which primary caregivers (usually parents) intervene to guide nutritional and weight restoration, slowly restoring autonomy to the adolescent and ending with a focus on the adolescent's development. For older adolescents and adults, and younger patients for whom FBT is not possible, individual psychotherapy is the

preferred outpatient treatment modality. Promising psychotherapeutic approaches for AN include cognitive behavior therapy (CBT), focal psychodynamic psychotherapy (FPT), the Maudsley Model of Anorexia Nervosa Treatment for Adults (MANTRA), and Specialist Supportive Clinical Management (SSCM), though none of these interventions has demonstrated consistent superiority to the others (Brockmeyer et al., 2018).

Intensive outpatient (IOP) treatment involves components of mental health, meal supervision, nutrition, and family therapy in a treatment center three or more days per week for a portion of the day (usually 3–5 h). Partial hospitalization (PHP) is the next higher level of intensity in which individuals attend a day-based program, returning home in the evenings and weekends. Individuals commonly attend PHP for a few to several weeks at a time. Residential level of care is an unlocked facility where participants live and receive treatment. Residents may stay at residential care for weeks to months. Finally, inpatient psychiatric or inpatient eating disorder (less common) facilities are locked units for those with significant safety concerns or uncontrolled behaviors (e.g. purging) that cannot be managed in other settings.

Inpatient medical care is intended to be short-term stabilization with the intention of returning to the individual to eating disorder care for ongoing management. Indications for inpatient medical hospitalization include hemodynamic instability (e.g. bradycardia, orthostatic hypotension), rapid weight loss, food refusal, significant electrolyte derangements (e.g. hypokalemia), or end-organ damage (SAHM, 2015).

Medication management may be utilized in conjunction with psychotherapy in some cases. However, in those with low weight, psychopharmacologic agents may be less effective. Antidepressants (selective serotonin reuptake inhibitors, tricyclic antidepressants) are a commonly prescribed class of medications for those with AN, although they have not proven to significantly improve behaviors, promote weight gain, or maintain weight in this population. Olanzapine, a second-generation antipsychotic, may be used to facilitate short-term weight gain and may assist in those who experience obsessionality (Davis and Attia, 2017). Medications to treat nutritional deficiencies, electrolyte imbalances, gastrointestinal complaints (constipation) are used as indicated. Physiologic estrogen replacement therapy for low bone mineral density has been shown to improve bone density in DXA scans and may be utilized based on DXA z-scores. Hormonal contraceptives (e.g. birth control pills) are not recommended to induce menses as they may mask resumption of spontaneous menses and have not been shown to improve DXA scores. Nutrition therapy considers the age, life-stage, laboratory findings, and historical eating patterns of the individual (Ozier et al., 2011). Nutrition therapy focuses on resolving nutritionally-based deficiencies, weight restoration, and normalization of eating patterns.

Conclusion

AN is a serious psychiatric illness with high rates of mortality, co-morbid psychiatric illness, and medical consequences. Early identification and treatment are associated with improved outcomes, and clinicians should be aware of warning signs and familiar with appropriate screening tools. Treatment for AN occurs across a spectrum of settings dependent on individual psychiatric and medical needs.

Funding information

This work was supported by a grant from the National Institute of Mental Health (T32 MH082761).

References

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders. Washington, DC, fifth ed.

Andrés-Pepiñá, S., Plana, M.T., Flamarique, I., Romero, S., Borràs, R., Julià, L., Gárriz, M., Castro-Fornieles, J., 2020. Long-term outcome and psychiatric comorbidity of adolescent-onset anorexia nervosa. Clin. Child Psychol. Psychiatr. 25 (1), 33–44.

Brockmeyer, T., Friederich, H.C., Schmidt, U., 2018. Advances in the treatment of anorexia nervosa: a review of established and emerging interventions. Psychol. Med. 48 (8), 1228–1256.

Buchman, S., Attia, E., Dawson, L., Steinglass, J.E., 2019. Steps of care for adolescents with anorexia nervosa: a Delphi study. Int. J. Eat. Disord. 52 (7), 777–785.

Bulik, C.M., Sullivan, P.F., Tozzi, F., Furberg, H., Lichtenstein, P., Pedersen, N.L., 2006. Prevalence, heritability, and prospective risk factors for anorexia nervosa. Arch. Gen. Psychiatr. 63, 305–312.

Campbell, K., Peebles, R., 2014. Eating disorders in children and adolescents: state of the art review. Pediatrics 34, 582–592.

Cotton, M.A., Ball, C., Robinson, P., 2003. Four simple questions can help screen for eating disorders. J. Gen. Intern. Med. 18 (1), 53–56.

Cucchi, A., Ryan, D., Konstantakopoulos, G., Stroumpa, S., Kaçar, A.Ş., Renshaw, S., Landau, S., Kravariti, E., 2016. Lifetime prevalence of non-suicidal self-injury in patients with eating disorders: a systematic review and meta-analysis. Psychol. Med. 46 (7), 1345–1358.

Culbert, K.M., Racine, S.E., Klump, K.L., 2015. Research review: what we have learned about the causes of eating disorders – a synthesis of sociocultural, psychological, and biological research. J. Child Psychol. Psychiatry. 56 (11), 1141–1164.

Davis, H., Attia, E., 2017. Pharmacotherapy of eating disorders. Curr. Opin. Psychiatr. 30 (6), 452–457.

Fairburn, C.G., Beglin, S.J., 1994. Assessment of eating disorders: interview or self-report questionnaire? Int. J. Eat. Disord. 16 (4), 363–370.

Garner, D.M., Olmsted, M.P., Bohr, Y., Garfinkel, P.E., 1982. The eating attitudes test: psychometric features and clinical correlates. Psychol. Med. 12 (4), 871–878.

Guillaume, S., Jaussent, I., Olié, E., Genty, C., Bringer, J., Courtet, P., Schmidt, U., 2011. Characteristics of suicide attempts in anorexia and bulimia nervosa: a case-control study. PLoS One 6 (8), e23578.

Hoek, H.W., van Hoeken, D., 2003. Review of the prevalence and incidence of eating disorders. Int. J. Eat. Disord. 34, 383–396.

Hudson, J.I., Hiripi, E., Pope Jr., H.G., Kessler, R.C., 2007. The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. Biol. Psychiatr. 61 (3), 348–358.
Keshaviah, A., Edkins, K., Hastings, E.R., Krishna, M., Franko, D.L., Herzog, D.B., Thomas, J.J., Murray, H.B., Eddy, K.T., 2014. Re-examining premature mortality in anorexia nervosa: a meta-analysis redux. Compr. Psychiatr. 55 (8), 1773–1784.
Lenherr-Taube, N., Trajcevski, K., Sochett, E., Katzman, D.K., 2020. Low PTH levels in adolescents with anorexia nervosa. Front. Pediatr. 8, 99.
Mandelli, L., Arminio, A., Atti, A.R., De Ronchi, D., 2019. Suicide attempts in eating disorder subtypes: a meta-analysis of the literature employing DSM-IV, DSM-5, or ICD-10 diagnostic criteria. Psychol. Med. 49 (8), 1237–1249.
Marques, L., Alegria, M., Becker, A.E., Chen, C.N., Fang, A., Chosak, A., Diniz, J.B., 2011. Comparative prevalence, correlates of impairment, and service utilization for eating disorders across US ethnic groups: implications for reducing ethnic disparities in health care access for eating disorders. Int. J. Eat. Disord. 44 (5), 412–420.
Morgan, J.F., Reid, F., Lacey, J.H., 1999. The SCOFF questionnaire: assessment of a new screening tool for eating disorders. BMJ 319 (7223), 1467–1468.
Nagata, J.M., Ganson, K.T., Austin, S.B., 2020. Emerging trends in eating disorders among sexual and gender minorities. Curr. Opin. Psychiatr. 33 (6), 562–567.
O'Connor, G., Nicholls, D., 2013. Refeeding hypophosphatemia in adolescents with anorexia nervosa: a systematic review. Nutr. Clin. Pract. 28, 358–364.
Ozier, A.D., Henry, B.W., American Dietetic Association, 2011. Position of the American Dietetic Association: nutrition intervention in the treatment of eating disorders. J. Am. Diet Assoc. 111 (8), 1236–1241.
Preti, A., Rocchi, M.B., Sisti, D., Camboni, M.V., Miotto, P., 2011. A comprehensive meta- analysis of the risk of suicide in eating disorders. Acta Psychiatr. Scand. 124 (1), 6–17.
Sachs, K.V., Harnke, B., Mehler, P.S., Krantz, M.J., 2016. Cardiovascular complications of anorexia nervosa: a systematic review. Int. J. Eat. Disord. 49 (3), 238–248.
Schorr, M., Miller, K.K., 2017. The endocrine manifestations of anorexia nervosa: mechanisms and management. Nat. Rev. Endocrinol. 13 (3), 174–186.
Society for Adolescent Health and Medicine, 2015. Position paper of the Society for Adolescent Health and Medicine: medical management of restrictive eating disorders in adolescents and young adults. J. Adolesc. Health 56, 121–125.
Solmi, M., Veronese, N., Correll, C.U., Favaro, A., Santonastaso, P., Caregaro, L., Vancampfort, D., Luchini, C., De Hert, M., Stubbs, B., 2016. Bone mineral density, osteoporosis, and fractures among people with eating disorders: a systematic review and meta-analysis. Acta Psychiatr. Scand. 133 (5), 341–351.
Steinglass, J.E., Glasofer, D.R., Dalack, M., Attia, E., 2020. Between wellness, relapse, and remission: stages of illness in anorexia nervosa. Int. J. Eat. Disord. 53 (7), 1088–1096.
Stice, E., Telch, C.F., Rizvi, S.L., 2000. Development and validation of the Eating Disorder Diagnostic Scale: a brief self-report measure of anorexia, bulimia, and binge-eating disorder. Psychol. Assess. 12 (2), 123–131.
Watson, H.J., Yilmaz, Z., Thornton, L.M., Hübel, C., Coleman, J., Gaspar, H.A., Bryois, J., Hinney, A., Leppä, V.M., Mattheisen, M., Medland, S.E., Ripke, S., Yao, S., Giusti-Rodríguez, P., Anorexia Nervosa Genetics Initiative, Hanscombe, K.B., Purves, K.L., Eating Disorders Working Group of the Psychiatric Genomics Consortium, Adan, R., Alfredsson, L., et al., 2019. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. Nat. Genet. 51 (8), 1207–1214.
Westmoreland, P., Krantz, M.J., Mehler, P.S., 2016. Medical complications of anorexia nervosa and bulimia. Am. J. Med. 129 (1), 30–37.

Further reading

Bardone-Cone, A.M., Harney, M.B., Maldonado, C.R., Lawson, M.A., Robinson, D.P., Smith, R., Tosh, A., 2010. Defining recovery from an eating disorder: conceptualization, validation, and examination of psychosocial functioning and psychiatric comorbidity. Behav. Res. Ther. 48 (3), 194–202.
Berends, T., Boonstra, N., van Elburg, A., 2018. Relapse in anorexia nervosa: a systematic review and meta-analysis. Curr. Opin. Psychiatr. 31 (6), 445–455.
Cachelin, F.M., Rebeck, R., Veisel, C., Striegel-Moore, R.H., 2001. Barriers to treatment for eating disorders among ethnically diverse women. Int. J. Eat. Disord. 30 (3), 269–278.
Golden, N.H., Katzman, D.K., Kreipe, R.E., Stevens, S.L., Sawyer, S.M., Reese, J., Nicholls, D., Rome, E.S., 2003. Eating disorders in adolescents: position paper of the society for adolescent medicine. J. Adolesc. Health 33, 496–503.
Keel, P.K., Forney, K.J., 2013. Psychosocial risk factors for eating disorders. Int. J. Eat. Disord. 46, 433–439.
Keski-Rahkonen, A., Hoek, H.W., Susser, E.S., Linna, M.S., Sihvola, E., Raevuori, A., Bulik, C.M., Kaprio, J., Rissanen, A., 2007. Epidemiology and course of anorexia nervosa in the community. Am. J. Psychiatr. 164 (8), 1259–1265.
Misra, M., Katzman, D., Miller, K.K., Mendes, N., Snelgrove, D., Russell, M., Goldstein, M.A., Ebrahimi, S., Clauss, L., Weigel, T., Mickley, D., Schoenfeld, D.A., Herzog, D.B., Klibanski, A., 2011. Physiologic estrogen replacement increases bone density in adolescent girls with anorexia nervosa. J. Bone Miner. Res. 26 (10), 2430–2438.
Reiter, C.S., Graves, L., 2010. Nutrition therapy for eating disorders. Nutr. Clin. Pract. 25 (2), 122–136.
Rienecke, R.D., 2017. Family-based treatment of eating disorders in adolescents: current insights. Adolesc. Health Med. Therapeut. 8, 69–79.
Thompson, J.K., Stice, E., 2001. Thin-ideal internalization: mounting evidence for a new risk factor for body-image disturbance and eating pathology. Curr. Dir. Psychol. Sci. 10, 181–183.
Trace, S.E., Baker, J.H., Peñas-Lledó, E., Bulik, C.M., 2013. The genetics of eating disorders. Annu. Rev. Clin. Psychol. 9, 589–620.

Eating disorders: Binge eating

M.A. Kalarchian[a] **and M.D. Marcus**[b], [a] School of Nursing, Associate Provost for Interdisciplinary Research and Strategic Initiatives, Duquesne University, Pittsburgh, PA, United States; and [b] Departments of Psychiatry and Psychology, University of Pittsburgh School of Medicine, Western Psychiatric Hospital, Pittsburgh, PA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of M.D. Marcus, J.E. Wildes, Eating disorders: Binge Eating, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 120–125, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00085-4.

Introduction 297
Assessment of binge eating 297
Epidemiology and risk factors 298
Comorbidity 298
Treatment of binge eating 299
Psychosocial treatments 299
Cognitive behavior therapy 299
Interpersonal psychotherapy 299
Dialectical behavior therapy 300
Guided self-help 300
Behavioral weight management 300
Pharmacotherapy 300
Psychostimulant medication 301
Antidepressant medications 301
Anticonvulsant medication 301
Weight loss medications 301
Selection of treatment for specific patients 301
Eating disorder and obesity history 301
Psychiatric status 302
Available resources 302
Summary 302
References 302

Key points

- BED is often a chronic and fluctuating disorder characterized by persistent and recurrent episodes of binge eating without the regular use of inappropriate compensatory behaviors
- Several methods can be used to assess BED, including standardized clinical interview and self-report measures.
- Biological, psychological, and social risk factors may interact to determine the onset and maintenance of BED, and both psychiatric and medical comorbidities, including obesity, are common.
- Evidence-informed treatments include psychosocial treatments, behavioral weight management, and pharmacotherapy.
- Treatment selection involves weighing the advantages and disadvantages of available interventions for an individual client, but more research is needed to identify personalized approaches to treatment of BED.

Glossary

Abstinence from binge eating The absence of binge eating behavior for a specified period of time (usually at least one month)

Binge eating The ingestion of a large amount of food within a discrete period of time accompanied by a sense of loss of control over when, what, or the quantity of food that is eaten

Compensatory behavior Inappropriate behavior designed to counteract the effects of binge eating or prevent weight gain. Examples of compensatory behaviors include self-induced vomiting; misuse of laxatives, enemas, diuretics, or diet pills; excessive exercise; and fasting

Eating disorder A persistent pattern of aberrant eating or inappropriate dieting behaviors accompanied by distressing thoughts and emotions. Eating disorders are associated with significant medical morbidity, as well as physical, social, or psychological dysfunction

Loss of control The subjective feeling that one cannot stop eating or control when, what, or the amount of food that is eaten

Obesity Refers to an excess of body fat. At present, there is no clear division between normal and abnormal levels of fat. However, body mass index (BMI), a ratio of weight to height calculated by weight in kilograms divided by the square of height in meters, is widely utilized as a screening measure for obesity given its robust associations with adiposity and medical comorbidity (obesity is operationalized as a BMI $\geq$30)

Introduction

In 1959, Stunkard noted three patterns of eating behavior in patients with obesity: night eating, binge eating, and eating without satiation. However, it was not until the 1980s that binge eating received attention as a distinct clinical phenomenon. Spitzer and colleagues proposed diagnostic criteria for Binge Eating Disorder (BED) and subsequently evaluated them in two field trials in the 1990s. These initial investigations led to the inclusion of BED in the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition (DSM-IV) as an example of Eating Disorder Not Otherwise Specified (EDNOS) and as a proposed diagnostic category requiring further study. Subsequently, BED was included as a formal diagnosis in the Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition (DSM-5).

BED is characterized by persistent and recurrent episodes of binge eating without the regular use of inappropriate compensatory behaviors seen in Bulimia Nervosa (BN) and the binge eating/purging subtype of Anorexia Nervosa (AN). In this chapter, we address several findings related to BED including its assessment, epidemiology, risk factors, and comorbid conditions. Evidence-supported treatments also are reviewed, including guidelines for choice of treatment approach (Fig. 1).

Assessment of binge eating

A binge episode is defined as the consumption of a large amount of food within a discrete period of time, accompanied by a sense of loss of control over eating. Researchers and clinicians have agreed that loss of control involves the subjective feeling that one cannot stop eating, or control what or how much is being eaten. Indeed, many observers have concluded that loss of control, rather than the amount of food ingested, is the hallmark of binge eating. There has been much less agreement about the size and duration of binge eating episodes. Specifically, there is no definitive threshold for a "large amount" of food (other than agreement that the amount of food eaten is more than others typically would eat in a similar situation), and the duration of binge eating episodes can vary widely, sometimes continuing over a period of many hours.

To receive a clinical diagnosis of BED (DSM-5), binge episodes must occur on at least one day per week for three months or more. Moreover, the binge eating episodes must be associated with three or more associated clinical features: eating much more

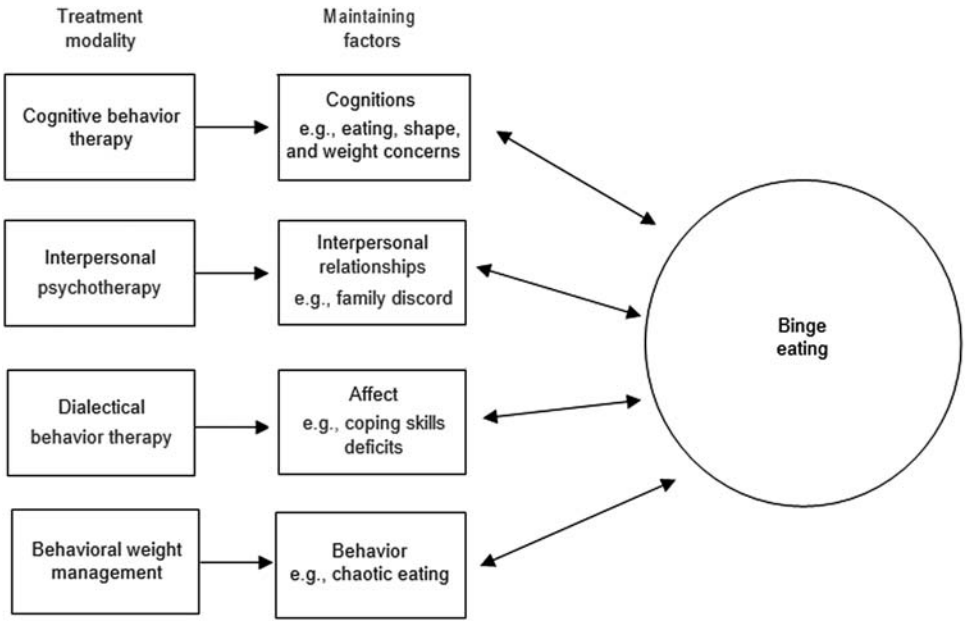


Fig. 1 Psychosocial and behavioral interventions and treatment targets for binge eating.

rapidly than normal; eating until uncomfortably full; eating large amounts of food when not hungry; eating alone because of embarrassment about how much one is eating, and feeling disgusted, depressed, or very guilty after overeating.

Several methods can be used to assess binge eating and BED, including standardized clinical interviews and self-report measures. A clinical interview by a trained professional is the preferred assessment method, as it provides the opportunity to standardize definitions of key concepts such as a "large amount of food" and "loss of control." Structured interviews include the Eating Disorder Examination (Fairburn et al., 2014) and the Eating Disorder Assessment for DSM-5 (Sysko et al., 2015). Questionnaires are easier to administer, and the use of psychometrically sound questionnaires is widespread for research and clinical screening. Self-report assessments include, but are not limited to, the Eating Disorder Diagnostic Scale (Stice et al., 2000) and the Questionnaire for Eating and Weight Patterns—5 (Yanovski et al., 2015). Interview-based assessments tend to yield ratings of binge eating that are lower, but more precise, than questionnaire-based surveys (Everett et al., 2021). Because binge eating and related psychopathology may occur across a range of behavioral phenotypes, readers may wish to consider the consensus measures for phenotypes and exposures website (www.phenx.org) for additional measures and protocols that may cut across multiple domains.

Epidemiology and risk factors

Findings from a nationally representative sample of 36,309 US adults assessed with lay-administered diagnostic interviews based on DSM-5 criteria documented lifetime prevalence estimates of 0.80%, 0.28% and 0.85% for AN, BN, and BED, respectively (Udo and Grilo, 2018). This study also found that the lifetime prevalence of BED was more common among women, and that the rates for BN and BED were roughly comparable across racial/ethnic groups. BED is also associated with obesity and extreme obesity. Findings from the Udo and Grilo (2018) study are in contrast with previous findings based on population data from the US National Comorbidity study (Hudson et al., 2007), which reported higher rates of BED (lifetime prevalence of 2.8%), and equivalent rates of BED in men and women. The Hudson et al. (2007) study utilized DSM-IV rather than DSM-5 criteria and utilized a much smaller sample than the Udo and Grilo (2018) study. Nevertheless, future population studies are needed to resolve differences in BED prevalence rates. All available studies have documented that binge eating is more prevalent among individuals with obesity in both community and clinical samples.

Multiple potential risk factors have been implicated in the etiology of BED. Family history and twin data have documented that BED is a moderately heritable illness that aggregates in families independently of obesity. Although work focusing on the genetics of BED is preliminary, there are indications that genetic polymorphisms related to dopamine functioning associated with poor impulse control are linked to BED. Other research has documented the involvement of serotonergic genes in BED. There have been investigations suggesting the involvement genes associated with numerous other factors (e.g., ghrelin, endocannabinoid system), but results are not definitive, and findings from genome-wide association studies have not yet been reported. Overall, research has indicated that BED is characterized by elevated response to food reward as well as increased impulsivity and compulsivity (Boswell et al., 2021).

Many studies have identified other putative risk factors for the development of eating disorders and binge eating disorder. These investigations have indicated that biological (e.g., childhood obesity, family history of obesity), psychological (e.g., body dissatisfaction, negative affect, perceived stress, depressive symptoms, dietary restraint), and social (e.g., repeated exposure to negative comments about shape, weight, or eating, perceived pressure to be thin) factors are implicated in the pathogenesis of binge eating.

Risk factors do not operate independently, and more recent work has investigated patterns of risk in longitudinal samples of adolescent women with body dissatisfaction (Stice and Van Ryzin, 2019). Results indicated that pursuit of the thin ideal with subsequent body dissatisfaction, dieting, and unhealthy weight control behaviors predicted the onset of binge eating disorders, and that impaired interpersonal functioning and negative affect preceded the onset of any eating disorder. Prospective risk factor studies including males and females of different racial and ethnic groups will improve our understanding of how multiple factors interact to determine the onset and maintenance of binge eating.

Comorbidity

BED has been associated with weight gain over time and higher risk of diabetes and other metabolic dysfunction Indeed, severity of binge eating is positively associated with degree of overweight, and there are important differences between overweight individuals with and without BED. Patients with BED report earlier onset of obesity along with a history of more severe obesity, dieting, and weight fluctuations. Moreover, when compared to equally overweight individuals without binge eating problems, BED patients report considerably less "restraint" or control over eating, lower self-esteem, more fear of weight gain, more preoccupation with food, and higher body dissatisfaction.

Although it is well documented that obesity is linked to adverse medical and psychosocial outcomes, research has indicated that BED is associated with poor health independent of the effects of comorbid obesity or psychopathology. Epidemiologic data suggest that BED is associated with type 2 diabetes, hypertension, dyslipidemias, sleep problems/disorders, and pain conditions. BED is also related to asthma and gastrointestinal symptoms and disorders, as well as menstrual dysfunction, pregnancy complications, intracranial hypertension, and polycystic ovary syndrome in women.

BED is also associated with increased risk of other psychiatric disorders and symptoms. Individuals with lifetime BED have higher rates of mood, anxiety, substance use, and other eating disorders compared to equally overweight individuals without binge eating problems. Data also suggest co-occurrence of binge eating with bipolar disorder and attention-deficit/hyperactivity disorder. Personality disorders also appear to pe prevalent among patients with BED. Finally, symptoms of BED, particularly binge eating, also are associated with other psychiatric symptoms, particularly depressive symptoms.

Treatment of binge eating

Among those who seek treatment, BED tends to be a chronic and fluctuating disorder. The clinical picture in BED often involves onset in late adolescence or the early 20s, with numerous periods of relative control over eating and weight loss alternating with periods characterized by binge eating and weight gain. Individuals with BED often seek obesity treatment rather than treatment of disordered eating *per se*.

A variety of psychosocial and pharmacological interventions can help individuals gain control over binge eating. There also is some evidence that behavioral weight management is associated with weight loss and short-term improvements in BED symptoms, although data suggest that eating disorder-focused treatments may be required to achieve longer-lasting remission from binge eating. Nevertheless, a substantial number of patients are not abstinent from binge eating after treatment, suggesting the need for clinical trials of novel therapeutic approaches as well as combinations and sequencing of treatments.

Psychosocial treatments

Treatments for BED have been adapted from those that have been shown to be effective in reducing binge eating among individuals with BN (Linardon et al., 2017). The majority of the research on psychosocial treatments has supported two structured, focused, short-term psychotherapies, Cognitive Behavior Therapy (CBT) and Interpersonal Psychotherapy (IPT), both of which have been shown to be more effective than no treatment in decreasing the frequency of binge episodes and improving the psychopathology associated with binge eating (da Luz et al., 2020). In addition, the use of Dialectical Behavior Therapy (DBT) (Ben-Porath et al., 2020) and interventions based on Acceptance and Commitment Therapy show promise as alternative treatments for BED. Finally, evidence supports the utility of self-help interventions in the treatment of binge eating.

Cognitive behavior therapy

Therapist-led CBT has been the most extensively studied treatment and has been shown to be associated with long-term reductions in binge eating and related psychopathology (Hilbert et al., 2020). It should be noted that many patients with binge eating do not meet full diagnostic criteria for BED or fluctuate between disorders, and thus one of the most widely-used formats of CBT is CBT enhanced for eating disorders (CBT-E), which adopts a transdiagnostic approach to binge eating. CBT-E has been found to be effective for treatment of individuals with a range of diagnoses (Atwood and Friedman, 2020).

CBT for BED is based on the premise that binge eating is maintained in the context of ongoing dietary restraint, weight concerns, negative emotions, and low self-esteem. Treatment focuses first on normalizing eating, and then on the identification and restructuring of maladaptive thoughts and beliefs, particularly those related to eating, shape, and weight.

CBT for BED has been adapted to reflect important differences between individuals with BN and BED (Fairburn, 2008). Specifically, cognitions relating to having a large body size are directly targeted in treatment. Overweight individuals with BED may be helped to accept a larger than average body size and to change unrealistic expectations for weight loss. For the majority of BED patients, a five or ten kilogram weight loss does not correspond with their desired weight, even though a modest weight loss may relate to improvements in binge eating and overall health. It is therefore important to help patients adopt realistic goals for the body weight and shape they are likely to achieve.

Another adaptation of CBT for BED relates to differences in the role of dieting between individuals with BED and those with BN. Although the treatment of BN stresses the role of dietary restraint in precipitating binge episodes, and treatment focuses on decreasing dietary restraint, patients with BED do not necessarily binge eat in response to restraint or hunger. Indeed, the preponderance of evidence suggests that increasing dietary restraint may help to ameliorate binge eating in obese individuals. Thus, CBT for BED does not stress decreased dietary restraint; rather, treatment encourages the development of a moderate, structured, healthy eating pattern.

Interpersonal psychotherapy

Klerman and Weissman's IPT originally developed for treatment of depression has been adapted for group treatment of individuals with BED (Wilfley et al., 1998). IPT for binge eating is based on the idea that dysfunctional eating behavior is maintained in the context of interpersonal difficulties. Treatment focuses on identifying and addressing specific, problematic interpersonal patterns in an effort to ameliorate binge eating. Treatment can focus on: (1) role disputes, such as marital or family discord; (2) role transitions, such as the adjustment to motherhood or a new job; (3) grief, such as the loss of a spouse or loved one; or, (4) interpersonal

deficits, such as loneliness and social isolation. IPT for BED does not directly target eating behaviors or attitudes about eating, shape, and weight. Although the ways in which CBT and IPT conceptualize and treat binge eating differ, both appear to be effective in reducing the frequency of binge eating. Evidence suggests that IPT may be particularly effective for individuals with low self-esteem and a high level of specific eating disorder psychopathology (Wilson et al., 2010).

Dialectical behavior therapy

Developed by Linehan for the treatment of individuals with borderline personality disorder, DBT has been adapted for treatment of BED (Ben-Porath et al., 2020). DBT is a comprehensive treatment program based on cognitive and behavioral principles and complemented by the use of mindfulness strategies derived primarily from Zen Buddhism. In addition to weekly individual outpatient treatment, traditional DBT prescribes a weekly group meeting in which the goal is to increase participants' behavioral skills. A group-only version of DBT for individuals with BED has been shown to decrease binge eating relative to active control eating, and wait list control; adaptations included changes to treatment format, content, and duration, and thus the effect of standard DBT cannot be determined from available studies. Nonetheless, despite its emphasis on emotion regulation, DBT has not shown superiority in reducing negative affect or enhancing adaptive affect regulation skills among individuals with BED. Thus, additional research is needed to determine the mechanisms by which DBT affects change in BED symptoms, and to evaluate the efficacy of DBT relative to CBT and IPT.

Guided self-help

There is growing evidence for the efficacy of guided self-help programs, in which a therapist or other professional provides support to individuals in completing a manualized self-help intervention, in the treatment of patients with eating disorders including binge eating (Traviss-Turner et al., 2017). Guided self-help programs have some advantages over more specialized psychotherapies for binge eating in that they can be implemented by therapists with little experience in the treatment of eating disorders and may offer a cost-effective approach to treating binge eating in community settings. For example, an eight-session guided self-help program resulted in significantly greater rates of abstinence from binge eating and reductions in the cognitive correlates of disordered eating relative to treatment as usual when provided by nonspecialist clinicians in a large health maintenance organization in the USA (DeBar et al., 2011). Another report found that guided self-help based on CBT principles was more effective than behavioral weight loss in producing remission from binge eating at two-year follow-up (Grilo and Masheb, 2005). Available findings suggest that CBT-based guided self-help may be a useful first stage treatment for mild-to-moderate BED, but more research is needed (Traviss-Turner et al., 2017). Other forms of self-help, such as unguided self-help and DBT-based self-help may also be effective for some patients.

Behavioral weight management

As most individuals with BED are overweight and want to lose weight, and because obesity is associated with significant medical and psychosocial consequences, weight loss is a potentially important outcome in the treatment of BED. Several reports have indicated that participation in a behavioral weight management program that focuses on moderate calorie restriction, provides education about sound nutritional principles, and promotes physical activity is associated with decreases in binge eating and improvements in mood among individuals with BED (Grilo, 2017). There is also evidence that behavioral weight management interventions produce greater weight loss among individuals with BED than do specialized psychotherapies for binge eating (e.g., CBT, IPT), which generally have been shown to have little effect on weight status. Although more widely available, behavioral weight management programs may be less effective than specialized psychotherapies and CBT-based guided self-help in producing lasting remission from binge eating. These findings indicate that psychosocial interventions that target disordered eating may be preferred in treating patients with BED, depending previous treatment history, current weight status, and personal goals for treatment.

Pharmacotherapy

A number of medications have been used to treat BED. Pharmacotherapy for BED tends to lead to moderate improvements in binge eating, but a modest impact on body weight. Medications may increase access to treatment and carry less stigma than psychotherapy. Lisdexamfetamine was the first drug approved for treatment of moderate-to-severe BED by the Food and Drug Administration (FDA) in 2015, and off-label treatments include antidepressants, weight loss medications, and anticonvulsants.

Emerging insights into the pathophysiological mechanisms involved in BED influence investigations into novel pharmacologic approaches. Animal models and small clinical trials have investigated a group of potential drug targets involving neurotransmitters/hormones like dopamine, glutamate, and endogenous opioids. However, larger and more robust investigations are required to determine the safety and efficacy of novel pharmacological agents.

Psychostimulant medication

Approved by the FDA for treatment of attention deficit hyperactivity disorder, the rationale for its use in BED was due in part to shared features of the disorders, including impulsivity, as well the drug's side effect of decreased appetite and weight loss. Current evidence documents utility of lisdexamfetamine in reducing binge eating days per week in adults. A review documented that an initial dose of 30 mg/day titrated to 50–70 mg/day is effective and generally well tolerated for treating moderate-to-severe BED in adults, with most adverse events being mild-to-moderate (e.g., dry mouth, headache, insomnia) (Heo and Duggan, 2017). Before starting treatment, potential for abuse and dependence, as well as serious cardiovascular-related reaction, and monitored throughout treatment.

Antidepressant medications

Because of their efficacy in ameliorating binge eating and purging behaviors in individuals with BN, antidepressants have been used widely in the treatment of BED (Davis and Attia, 2019). Early research comparing tricyclic antidepressants (e.g., desipramine, imipramine) to placebo showed greater reductions in binge eating among obese binge eaters treated with active medication than with a placebo. In subsequent research, several selective serotonin reuptake inhibitors (e.g., fluoxetine, fluvoxamine, sertraline, citalopram) have been shown to be associated with moderate reductions in binge eating in BED patients. Moreover, the effects of antidepressant treatment on binge eating are independent of any effects on mood.

Anticonvulsant medication

Anticonvulsant medications were initially investigated for BED because they are used to treat disorders that are often comorbid with BED, such as mood and substance use disorders, and some have been associated with weight loss (McElroy, 2017). Topiramate and zonisamide have both been effective for treatment of BED, although use of both may be limited due to side effect profiles. There also is preliminary evidence that augmenting group CBT for BED with topiramate is associated with greater reductions in body weight and higher rates of remission from binge eating compared to group CBT with placebo (Claudino et al., 2007). Topiramate has been the most studied, but its long-term on binge eating and weight loss remain unknown.

Weight loss medications

Anorectic agents used in the treatment of obesity, sibutramine and orlistat, have been investigated in BED patients with promising results, but sibutramine was withdrawn from the US market in 2010. Research has shown that adding orlistat to CBT-based guided self-help is associated with higher rates of remission from binge eating at post-treatment and greater weight loss at both post-treatment and 3-month follow-up relative to CBT-based guided self-help plus placebo (Grilo et al., 2005). Emerging data on newer anti-obesity agents (e.g., semaglutide) herald a more promising future for the use of anti-obesity agents, but the efficacy of these agents in the management of BED is not known. Nevertheless, findings suggest that augmenting psychosocial treatments for disordered eating with weight loss medications may be an effective approach to treating patients with BED and obesity.

Selection of treatment for specific patients

No single treatment approach is effective for all patients, and additional research is needed to guide personalized treatment for individual patients with BED. Until such information becomes available, clinicians and patients must decide on a course of treatment based on a careful assessment and thorough consideration of the pros and cons of available options, including cost and access. A review found that combining certain medications with cognitive behavioral therapy or behavioral weight management produces superior outcomes to pharmacotherapy only for BED, but does not substantially improve outcomes achieved with psychosocial treatment only (Grilo et al., 2016). Recent evidence suggests, however, that weight-loss medication enhanced outcomes for adults with BED with comorbid obesity in an adaptive stepped care approach (Grilo et al., 2020). More research is needed on comparative efficacy, mechanisms through which treatments work, and complex models of care.

Eating disorder and obesity history

A history of early onset of binge eating, binge eating in the absence of obesity, or obesity in combination with numerous bouts of weight loss and regain over time (i.e., "yo-yo" dieting), suggest a course of eating disorders treatment. Such patients can be reassured that significant improvements in the aberrant eating and eating disorder psychopathology associated with BED can be obtained without weight loss.

On the other hand, clinical experience suggests that patients who report adult onset of binge eating and obesity, and do not have a history of marked weight fluctuations, may benefit from a behavioral weight management approach. Behavioral weight management also may be indicated for patients who remain overweight after a trial of eating disorders treatment. However, weight loss

among individuals with BED is modest, and the re-emergence of binge eating can lead to weight regain, so it is important for each individual to evaluate the likelihood that he or she will be able to sustain changes in eating and exercise.

Psychiatric status

Given the high psychiatric comorbidity in BED, a thorough evaluation is important for all patients who seek treatment. Although mild to moderate depression or anxiety is likely to improve during treatment of binge eating, the presence of marked or severe current illness suggests primary treatment of the mood or anxiety disorder. In addition, the presence of personality disorders characterized by emotional, dramatic, or impulsive behavior appears to be related to severity of binge eating. Although data indicating that co-morbid personality disorder is associated with treatment outcome, it may be that an approach like DBT, which targets emotional regulation may benefit these individuals.

Available resources

Clinicians trained in the use of psychosocial treatments for eating disorders are likely to be found in most academic medical centers and urban areas but may not be available in rural or underserved areas. Insurance companies vary in coverage for treatment of eating disorders, and some insurance plans may pay for obesity treatment only if there is a clear medical indication (e.g., hypertension or other cardiovascular risk). Thus, clinicians may need to consider pragmatic factors such as the availability of specialty care, or patient insurance plan coverage. Guided self-help programs based on CBT principles may be a promising alternative for patients who do not have access to specialized psychotherapies. There is growing interest in the use of mHealth (e.g., apps and internet-based interventions) as a means of disseminating therapist-guided self-help for BED, and preliminary data have been encouraging. However, comorbid psychopathology and high frequency binge eating may require more intensive clinical intervention.

Summary

BED is a fluctuating disorder and frequently chronic disorder that is common among obese individuals who seek treatment and is associated with elevated rates of psychopathology. Although the exact causes of binge eating remain unknown, available data support a multidimensional model that includes biological, psychological, and social factors. Once established, binge eating is maintained by a complex interplay among eating behaviors, cognitions, affect, and interpersonal factors. Nevertheless, available research indicates that most people who binge-eat can be helped with specialized psychotherapy for eating disorders, therapist-guided self-help programs, or behavioral weight management. Pharmacotherapy also appears to reduce binge eating, and treatment with certain medications may result in larger and more sustained weight losses when combined with psychosocial interventions than psychosocial intervention alone. A careful assessment, review of the benefits and disadvantages of the different therapies, and consideration of the availability of trained clinicians should guide the choice of treatment for an individual with BED. More research is necessary to fully understand this problematic eating disorder and to improve strategies for management and treatment.

References

Atwood, M.E., Friedman, A., 2020. A systematic review of enhanced cognitive behavioral therapy (CBT-E) for eating disorders. Int. J. Eat. Disord. 53, 311–330.
Ben-Porath, D., Duthu, F., Luo, T., et al., 2020. Dialectical behavioral therapy: an update and review of the existing treatment models adapted for adults with eating disorders. Eat. Disord. 28, 101–121.
Boswell, R.G., Potenza, M.N., Grilo, C.M., 2021. The neurobiology of binge-eating disorder compared with obesity: implications for differential therapeutics. Clin. Therapeut. 43, 50–69.
Claudino, A.M., De Oliveira, I.R., Appolinario, J.C., et al., 2007. Double-blind, randomized, placebo-controlled trial of topiramate plus cognitive-behavior therapy in binge-eating disorder. J. Clin. Psychiatr. 68, 1324–1332.
Da Luz, F.Q., Hay, P., Wisniewski, L., Cordás, T., Sainsbury, A., 2020. The treatment of binge eating disorder with cognitive behavior therapy and other therapies: an overview and clinical considerations. Obes. Rev. 22, e13180.
Davis, L.E., Attia, E., 2019. Recent advances in therapies for eating disorders. F1000Res. 8, F1000. Faculty Rev-1693.
Debar, L.L., Striegel-Moore, R.H., Wilson, G.T., et al., 2011. Guided self-help treatment for recurrent binge eating: replication and extension. Psychiatr. Serv. 62, 367–373.
Everett, V.S., Crochiere, R.J., Dallal, D.H., et al., 2021. Self-report versus clinical interview: discordance among measures of binge eating in a weight-loss seeking sample. Eat. Weight Disord. 26, 1259–1263.
Fairburn, C., 2008. Cognitive Behavior Therapy and Eating Disorders. Guilford, New York.
Fairburn, C., Cooper, Z., O'Connor, M., 2014. Eating Disorder Examination, seventeenth ed. 0D; April, 2014.
Grilo, C.M., 2017. Psychological and behavioral treatments for binge-eating disorder. J. Clin. Psychiatr. 78 (Suppl. 1), 20–24.
Grilo, C.M., Masheb, R.M., 2005. A randomized controlled comparison of guided self-help cognitive behavioral therapy and behavioral weight loss for binge eating disorder. Behav. Res. Ther. 43, 1509–1525.
Grilo, C.M., Masheb, R.M., Salant, S.L., 2005. Cognitive behavioral therapy guided self-help and orlistat for the treatment of binge eating disorder: a randomized, double-blind, placebo-controlled trial. Biol. Psychiatr. 57, 1193–1201.
Grilo, C.M., Reas, D.L., Mitchell, J.E., 2016. Combining pharmacological and psychological treatments for binge eating disorder: current status, limitations, and future directions. Curr. Psychiatr. Rep. 18, 55.
Grilo, C.M., White, M.A., Masheb, R.M., et al., 2020. Randomized controlled trial testing the effectiveness of adaptive "SMART" stepped-care treatment for adults with binge-eating disorder comorbid with obesity. Am. Psychol. 75, 204–218.

Heo, Y.A., Duggan, S.T., 2017. Lisdexamfetamine: a review in binge eating disorder. CNS Drugs 31, 1015–1022.
Hilbert, A., Petroff, D., Herpertz, S., et al., 2020. Meta-analysis on the long-term effectiveness of psychological and medical treatments for binge-eating disorder. Int. J. Eat. Disord. 53, 1353–1376.
Hudson, J.I., Hiripi, E., Pope, H.G., Kessler, R.C., 2007. The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. Biol. Psychiatr. 61, 348–358.
Linardon, J., Wade, T.D., De La Piedad Garcia, X., Brennan, L., 2017. The efficacy of cognitive-behavioral therapy for eating disorders: a systematic review and meta-analysis. J. Consult. Clin. Psychol. 85, 1080–1094.
Mcelroy, S.L., 2017. Pharmacologic treatments for binge-eating disorder. J. Clin. Psychiatr. 78 (Suppl. 1), 14–19.
Spitzer, R.L., Yanovski, S.Z., Marcus, M.D., 1993. Questionnaire on Eating and Weight Patterns-Revised (QEWP-R). New York State Psychiatric Institute, New York, NY.
Stice, E., Telch, C.F., Rizvi, S.L., 2000. Development and validation of the Eating Disorder Diagnostic Scale: a brief self-report measure of anorexia, bulimia, and binge-eating disorder. Psychol. Assess. 12, 123–131.
Stice, E., Van Ryzin, M.J., 2019. A prospective test of the temporal sequencing of risk factor emergence in the dual pathway model of eating disorders. J. Abnorm. Psychol. 128, 119–128.
Sysko, R., Glasofer, D.R., Hildebrandt, T., et al., 2015. The eating disorder assessment for DSM-5 (EDA-5): development and validation of a structured interview for feeding and eating disorders. Int. J. Eat. Disord. 48, 452–463.
Traviss-Turner, G.D., West, R.M., Hill, A.J., 2017. Guided self-help for eating disorders: a systematic review and metaregression. Eur. Eat Disord. Rev. 25, 148–164.
Udo, T., Grilo, C.M., 2018. Prevalence and correlates of DSM-5-defined eating disorders in a nationally representative sample of US adults. Biol. Psychiatr. 84, 345–354.
Wilfley, D., Frank, M., Welch, R., Spurrell, E., Rounsaville, B., 1998. Adapting interpersonal psychotherapy to a group format (IPT-G) for binge eating disorder: toward a model for adapting empirically supported treatments. Psychother. Res. 8, 379–391.
Wilson, G.T., Wilfley, D.E., Agras, W.S., Bryson, S.W., 2010. Psychological treatments of binge eating disorder. Arch. Gen. Psychiatr. 67, 94–101.
Yanovski, S.Z., Marcus, M.D., Wadden, T.A., Walsh, B.T., 2015. The questionnaire on eating and weight patterns-5: an updated screening instrument for binge eating disorder. Int. J. Eat. Disord. 48, 259–261. https://doi.org/10.1002/eat.22372.

Eating disorders: Bulimia nervosa

Heather A. Davis and Jennifer E. Wildes, Department of Psychiatry & Behavioral Neuroscience, University of Chicago, Chicago, IL, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A.J. Hill, S. Heywood-Everett, U. Philpot, Eating Disorders: Bulimia Nervosa, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 126–131, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00084-2.

Introduction 305
Diagnostic criteria 305
Prevalence of BN 305
Gender differences 305
Racial/ethnic differences 306
Food insecurity 306
Associated psychopathology 306
Weight/shape overvaluation 306
Comorbidity 306
Etiology and maintenance 307
Genetic risk 307
Developmental risk 307
Personality risk 307
Theories of BN maintenance 307
Health consequences of BN 307
Nutritional consequences 307
Medical consequences 308
Treatment 308
Dietetic management 308
Screening 308
Professional responsibilities 308
Nutrition care process 309
Psychological treatment 309
Cognitive behavior therapy 309
Interpersonal psychotherapy 309
Dialectical behavior therapy 310
Integrative cognitive-affective therapy 310
Family-based treatment 310
Pharmacotherapy 310
Treatment of underrepresented populations 311
Long-term prognosis 311
Conclusion 311
Funding information 311
References 311
Further reading 312

Key points

- Bulimia nervosa is a serious eating disorder associated with significant impairment and distress, health complications, psychiatric comorbidity, and a chronic course.
- Bulimia nervosa occurs across genders, racial/ethnic groups, and socioeconomic status.
- Treatment for bulimia nervosa should include medical management and monitoring, empirically supported psychological intervention, and, if necessary for malnourished patients, nutritional support. Pharmacotherapy is a safe and effective adjunct to psychological treatment.

Introduction

Bulimia nervosa (BN) is a severe eating disorder characterized by recurrent episodes of binge eating, inappropriate compensatory behaviors, and overvaluation of weight and shape (APA, 2013). This article will begin by introducing the diagnostic criteria for BN. Next, we will describe the prevalence of BN, followed by a discussion of research regarding the psychopathology of BN. We will then describe the etiology of BN, followed by the nutritional and medical complications of the disorder. We will continue with a description of dietary management, psychological treatment, and pharmacotherapy, evidence for the effectiveness of each, and special considerations for future research. Finally, long-term prognosis will be discussed.

Diagnostic criteria

Diagnostic schedules, including the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; APA, 2013) and proposed guidelines for the International Classification of Diseases, 11th Edition (ICD-11), agree that the following features must be present for a diagnosis of BN: (1) recurrent episodes of binge eating; (2) recurrent inappropriate compensatory behavior to prevent weight gain, such as self-induced vomiting, laxative or diuretic misuse, fasting, or excessive exercise; and (3) self-evaluation that is unduly influence by body shape and weight (APA, 2013). Importantly, BN is distinct from anorexia nervosa (AN) because one cannot be underweight and meet criteria for BN (APA, 2013). Additionally, BN is distinct from binge-eating disorder (BED), as BED includes recurrent binge eating without the presence of recurrent inappropriate compensatory behaviors or the requirement of undue influence of body shape and weight on self-evaluation (APA, 2013).

Binge eating is distinct from everyday overeating or dietary indulgence. According to DSM-5, three features must be present for an eating episode to be considered an objective binge episode: consumption of an unusually large amount of food, over a short period of time, and a subjective sense of loss of control over eating (APA, 2013). The size of binge-eating episodes varies; in laboratory studies, binge-eating episodes in BN are around 3000 calories. Notably, the literature has identified that some binge-eating episodes are considered "subjective" rather than objective. Subjective binge eating involves loss of control as with objective binge eating, but the amount of food is not deemed excessive. Importantly, the amount of food consumed is not the primary criterion for defining a binge, as both subjective and objective binge-eating episodes are associated with loss of control and thus distress.

The DSM-5 diagnostic criteria and ICD-11 proposed guidelines differ in two important ways. First, ICD-11 includes subjective binge-eating episodes to satisfy the binge eating requirement. Another key difference is in duration of symptoms: the ICD-11 criteria for BN require a shorter duration of binge eating and compensatory behaviors, 1 month, compared to the required duration of 3 months in DSM-5.

One concern of diagnostic classification experts is the prevalence of residual, "catch all" diagnoses such as the DSM-5 category, "Other Specified Feeding or Eating Disorder" (OSFED), which encompasses a wide variety of subthreshold eating disorders, including BN of low frequency (e.g., binge-eating episodes or compensatory behaviors occurring less than once per week) or limited duration (e.g., behaviors that have not persisted for three consecutive months). In studies comparing the ICD-11 and DSM-5 criteria to diagnose patients, the broader proposed criteria of ICD-11 do appear to facilitate greater clinical utility. In an investigation of 107 adults with eating disorders, the proposed ICD-11 BN criteria appeared to capture more individuals with BN than the DSM-5 criteria, leaving fewer individuals in the residual category of OSFED BN (Amorim Palavras et al., 2018). Furthermore, individuals who met criteria for ICD-11 BN but not DSM-5 BN endorsed no significant differences in severity of eating disorder psychopathology, depression, anxiety, stress, or health-related quality of life (Amorim Palavras et al., 2018), indicating that the DSM-5 criteria do not better predict distress or impairment when compared with the ICD-11 criteria.

The proposed ICD-11 diagnostic guidelines also appear to be superior to ICD-10 criteria in classifying eating disorders. In a study of 2288 clinicians using ICD-10 and proposed ICD-11 criteria to diagnose case vignettes of individuals with feeding or eating disorders, the ICD-11 diagnostic guidelines better distinguished cases of eating disorders and demonstrated higher clinical utility ratings across clinicians, who were diverse in terms of language (including Chinese, English, French, Japanese, and Spanish), indicating that the ICD-11 criteria as written appeared valid and culturally sensitive.

Prevalence of BN

Gender differences

Recent research indicates that the lifetime prevalence of BN among female-identified individuals is 0.46%, and among male-identified individuals is 0.08%. However, the prevalence of specific bulimic symptoms is higher in both genders; a study of college students from 11 universities revealed that 49% of female students endorsed binge eating compared with 30% of male students, but rates of compensatory behaviors were similar: 31% of female students versus 29% of male students endorsed excessive exercise, fasting, self-induced vomiting, or laxative/diuretic use to control their weight or compensate for binge eating. The average age of onset for BN is 18 years, but binge eating and compensatory behaviors in children have been documented as early as 10 years old, in boys and girls.

Although point prevalence of BN decreases across the lifespan for women, it remains constant for men, suggesting gender differences diminish over time. Women from age 20 to age 40 demonstrate higher drive for thinness and bulimia symptoms across time compared to men, which may help explain greater prevalence of BN among adolescent girls and young women.

Research on the prevalence of BN in gender minority populations is limited. However, recent findings indicate that rates of self-reported BN are higher in transgender and genderqueer individuals relative to their cisgender peers, suggesting that gender minority groups may be at heightened risk for BN (Simone et al., 2020). Further research is needed to understand this disparity.

Racial/ethnic differences

Early research of racial/ethnic differences suggested that the prevalence of BN was higher in White individuals. However, in recent years, studies of eating disorder prevalence have found no significant racial/ethnic differences in rates of BN in population samples or high-risk college women, suggesting little to no difference in prevalence across racial/ethnic groups. Notably, BN samples in research studies historically have been largely White, limiting conclusions that can be drawn regarding prevalence. Research should continue to include samples diverse in terms of race/ethnicity to address the disparities of previous samples.

Food insecurity

Food insecurity is defined as "limited or uncertain availability of nutritionally adequate and safe foods or limited or uncertain ability to acquire acceptable foods in socially acceptable ways". Food insecurity is associated with high weight and overeating (Franklin et al., 2012), potentially through a pattern of dietary restriction followed by increased consumption of lower cost, low-nutrition, energy dense foods (Bruening et al., 2012). Recent studies have shown associations between food insecurity and binge eating. In an investigation of 873 adults, low and very low food security were associated with increased likelihood of BN diagnosis (Lydecker and Grilo, 2019). Individuals with food insecurity experience intermittent availability of food throughout the month, due to various factors including paycheck allocation or payments from food assistance programs, which is thought to contribute to a scarcity-induced dietary restriction and subsequent binge eating cycle. The relationship between dietary restriction and binge eating within food secure populations is well-established (Fairburn, 2008), but the relationship in the context of food insecurity may be unique due to the bidirectional process of restriction and binge eating that results from externally imposed food restriction due to financial difficulties, rather than internally motivated dietary restriction for the purpose of weight loss or maintenance. Because most studies of food insecurity and BN are correlational, further research is needed to elucidate the temporal development of BN in the context of food insecurity, and the mechanisms for the development and maintenance of bulimic behaviors in this population.

Associated psychopathology

Weight/shape overvaluation

In addition to being a core symptom of BN, weight/shape overvaluation is thought to play a central role in the onset and maintenance of dietary restriction, binge eating, purging (e.g., self-induced vomiting, laxative abuse, and diuretic abuse), excessive exercise, and other eating-related concerns, all of which are evident in BN (Fairburn, 2008). Patients with BN demonstrate significantly greater levels of weight/shape overvaluation and body image disturbance when compared with patients with AN and healthy controls. Related to weight/shape overvaluation, fear of weight gain also is a prominent feature of BN; a network analysis of adults with BN revealed that fear of weight gain was central to BN psychopathology.

Comorbidity

There is extensive evidence that the problems and distress associated with BN extend beyond disordered eating. As many as 74.4% of individuals with BN experience symptoms and diagnoses of concurrent mood disorders, anxiety disorders, post-traumatic stress disorder, substance use disorder, and personality disorders. BN also is associated with increased rates of suicide (Goldstein and Gvion, 2019). Substance use disorders demonstrate perhaps the most significant overlap, as 37% of individuals with BN meet criteria for a substance use disorder. Emphasizing the seriousness of this overlap, the presence of multiple comorbidities in BN is associated with inpatient hospitalization stays, a poorer prognosis, and increased mortality.

The mechanisms underlying the relationship between bulimic symptoms and comorbid psychopathology in BN warrant further investigation. Theorists have proposed that certain transdiagnostic risk factors may help to explain the overlap between symptoms. One such risk factor might be the experience of shame, an overarching negative evaluation of the self, which has been shown to predict a variety of different forms of psychopathology including bulimic behaviors, substance abuse, self-harm, depression, and anxiety.

Etiology and maintenance

A variety of psychological, biological, and social factors are involved in the emergence of BN. We review three categories of risk: genetic, developmental, and personality-based risk.

Genetic risk

Research on the genetic epidemiology of BN is limited when compared with AN. It appears that individuals with family members with BN are predisposed to developing BN themselves (Strober et al., 2000). The heritability of BN is estimated at 0.60 (Bulik et al., 2010). Genetic studies indicate BN is strongly correlated with AN and alcohol use disorder, supporting their phenotypic overlap.

Developmental risk

Engagement in and exposure to dieting behavior, through parents or other family members, and childhood obesity appear to increase risk for disordered eating behaviors consistent with BN. Adverse childhood experiences also appear to increase risk for BN; across studies of patients with BN, 35% report childhood sexual abuse, 33% report physical abuse, and 81% report emotional abuse (Molendijk et al., 2017). Moreover, 49% of patients with BN report multiple forms of childhood maltreatment (Molendijk et al., 2017). Experience of childhood maltreatment among patients with eating disorders is associated with comorbid psychiatric diagnosis, increased suicidality, earlier age of eating disorder onset, and a more severe illness course (Molendijk et al., 2017).

Personality risk

Although impulsivity historically was thought to predispose one to, and correlate with, BN, research has clarified that impulsivity is a broad term that encompasses several facets of rash behavior, and does not demonstrate reliable associations with, or prediction of, behaviors consistent with BN (Fischer et al., 2008). This said, the impulsigenic trait of negative urgency, a disposition that describes the tendency to act rashly when feeling distressed, demonstrates reliable and consistent prediction of behaviors central to BN (i.e., binge eating and purging), as well as prediction of other comorbid maladaptive behaviors and symptoms, including self-harm, substance use, and depression (Berg et al., 2015). Other personality traits that predispose one to the development of BN include heightened perfectionism, negative affect (a general dimension of negative mood), neuroticism (i.e., the tendency to experience anxiety, depression, impulsiveness, and vulnerability to stress), and avoidance motivation (the tendency to avoid harm or discomfort).

Theories of BN maintenance

There are several established theories that seek to explain the maintenance of disordered eating behaviors within BN. Binge-eating episodes often follow periods of extreme dietary restriction or fasting (Fairburn, 2008). Restraint theory posits that dietary restriction is a risk factor for binge eating due to chronic hunger and a reliance on cognitive control of food intake, rather than physiological cues of hunger and fullness. Fasting also is thought to increase physiological and cognitive hunger drives, and deplete self-regulatory resources. Other theories attempt to incorporate the experience of difficult emotions in explanatory models of binge eating; the affect regulation model posits that binge eating is used to cope with strong negative emotions, and escape theory posits that binge eating is used to relieve or distract from subjective distress. Empirical studies support both dietary restriction and heightened negative affect as preceding binge eating in BN samples (Haedt-Matt and Keel, 2011), but evidence for relief of negative affect following binge eating is mixed, with some studies reporting alleviated negative affect following binge eating (Berg et al., 2013), and others reporting elevated negative affect following binge eating (Haedt-Matt and Keel, 2011). Further research is needed to clarify the emotional consequences of binge eating and thus the maintenance process of BN.

Health consequences of BN

Nutritional consequences

Binge eating often follows caloric restriction and tends to occur on consecutive days. Individuals with BN frequently report caloric restriction following a binge-eating episode, to prevent further weight gain (Fairburn, 2008). Caloric restriction can result in restrictive eating behavior that may fail to achieve the recommended energy intake, leading to nutritional deficiencies and weight loss. Although a diagnosis of BN requires that a patient not be underweight, many patients with BN are weight suppressed (i.e., at a current weight that is several pounds below pre-morbid weight). Because weight suppression has been found to maintain BN in longitudinal research, it is essential that nutritional treatment of BN seek to help the patient decrease dietary restriction by eating a larger range and greater quantity of food consistently to allow for the resumption of a normative, healthy pattern of eating and restoration to premorbid weight.

Medical consequences

Eating disorders, including BN, are associated with increased risk of death (Arcelus et al., 2011) somewhat heightened by medical complications attributed to specific bulimic behaviors, such as purging. Importantly, purging behavior of all types can be associated with severe medical complications across all body systems (see Forney et al., 2016, for a review). Self-induced vomiting is the most frequently endorsed purging behavior in BN and research shows that the medical complications of self-induced vomiting can be extensive (Forney et al., 2016). Recurrent self-induced vomiting is associated with dehydration and electrolyte loss, including potassium. Regarding oral health, self-induced vomiting directly causes dental erosion due to the stomach acid that enters the mouth. Case studies indicate that self-induced vomiting causes parotid gland (a salivary gland between the mouth and ear) swelling, which can be painful to patients (Forney et al., 2016). Self-induced vomiting in BN also is associated occasionally with gastroesophageal reflux, retrosternal burning (heartburn), and acid regurgitation, potentially due to exposure of stomach acid to the esophagus. Of note, these symptoms are not associated with frequency of purging behavior, but instead are thought to be due to intraindividual sensitivity to esophagus damage. Rarely, cardiovascular complications occur due to purging behavior, generally when substances such as ipecac are used to induce vomiting. Given the potential severity of these symptoms, patients who vomit frequently should be educated on the medical consequences of such behavior.

Laxative and diuretic use are less common than self-induced vomiting among individuals with BN, but nevertheless associated with problems of various organ systems (Forney et al., 2016). Consequences of laxative and diuretic misuse include renal inflammation, calcium deposits, kidney stones, gastrointestinal symptoms such as bowel dysfunction and abdominal discomfort, cardiac arrest, seizures, and rhabdomyolysis (deterioration of muscle tissue) (Forney et al., 2016). Given the possibility of extensive complications, providers should be aware of laxative and diuretic abuse and intervene swiftly to avoid long-term damage to vital organ systems.

Treatment

Below, we review the recommended components of treatment including dietetic management, psychological intervention, and pharmacotherapy.

Dietetic management

Practice guidelines set by the Academy for Eating Disorders recommend dietetic involvement in eating disorder treatment, including for BN. However, the recommended, manualized psychological treatments for adults and youth do not typically mention the incorporation of dietetic treatment, despite the explicit mention of nutrition-focused content in 91% of treatment manuals. This is likely because of the distinction between malnourishment and undernourishment in BN. Many patients with BN are undernourished due to dietary restriction, but not malnourished. Because restriction not leading to low body weight still can be associated with serious health complications due to the avoidance of certain food groups, basing malnutrition only on current body weight runs the risk of missing treatable complications. Moreover, BN can result in the context of a history of AN; thus, it is recommended that providers take possible malnutrition seriously to avoid a host of life-threatening problems. If a patient with BN is undernourished but not malnourished, dietetic treatment may not be indicated, as the other providers as part of the treatment team (e.g., primary care physician, therapist, psychiatrist) may be equipped to help the patient by prescribing a regular meal pattern. At all levels of care, patients with BN should be screened routinely for medical instability and nutrition deficiency to ensure safety for treatment and determine if nutrition interventions are needed.

When dietetic support for BN is necessary (e.g., in cases of medical instability or malnourishment), dietitians should follow the practice and training standards set forth by the Academy for Eating Disorders. The components of care can be divided usefully into three categories: screening, professional responsibilities, and the nutrition care process, each described in further detail below.

Screening

The dietitian should be aware of the groups at high-risk for disordered eating, and screen those individuals appropriately. Screening tools may include symptom checklists or clinical interviews.

Professional responsibilities

The dietitian should prioritize the use of a reflective practice model and demonstrate a commitment to and use of professional clinical supervision, which improves patient outcomes and ensures safe practices, as well as maintains clinical resilience. Dietitians should seek out supervision or consultation services regularly, and initiate contact with the patient's treatment team (e.g., therapist, physician, psychiatrist) early in the treatment process.

Nutrition care process

In addition to the usual assessment procedures taken by a dietitian (e.g., thorough assessment of medical history, allergies and diseases, individual macronutrient requirements), there are several special considerations for patients with BN. The nutrition care process should include a thorough assessment of premorbid and current dietary intake, considering all aspects of intake (e.g., macro and micro nutrients, fluid, caffeine, alcohol). The dietitian should carefully assess the patient's sociocultural background and its influence on dietary preferences, the purchasing and preparing of food, flexibility of the patient's schedule, timing of the meal pattern, and the relation of these factors to eating disorder behaviors. It also is important to remain mindful of medication and supplement use, and their potential impact on weight, eating behavior, and appetite. The dietitian should gain some knowledge of the patient's beliefs regarding food, eating, and body image to allow for an understanding of the context of the eating problem. Special attention should be paid to the lived experience of the patient, including possible past harms related to weight stigma. It also is essential that the dietitian assess and take into account activity level and athlete status in prescribing recommendations for the meal pattern.

A nutrition diagnosis is formulated based on nutrition intake problems stemming from BN (e.g., over or under-hydration, avoidance of certain nutrients), and a statement is developed to inform the treatment team of the plan for treating specified problems. The dietitian should take special note of the nutrition problems related to body image disturbance, history of dieting, family history of an eating disorder, and fear of weight gain, all of which can inform the patient's approach to or avoidance of the specified nutrition rehabilitation plan. For adolescents, it is necessary to involve the patient's family in assessment, treatment planning, and treatment implementation.

The nutrition intervention should consider the patient's stage in the eating disorder treatment process and level of independence, which often is based on level of care (e.g., outpatient, intensive outpatient, hospital, residential). The recommended meal pattern for all patients is three nutritionally adequate meals per day (e.g., morning meal, mid-day meal, evening meal), and 2–3 snacks per day (e.g., snacks between the morning and mid-day meal, mid-day and evening meal, and after evening meal). A wide variety of food intake should be encouraged, as well as flexibility of food choice and an emphasis on eating for enjoyment rather than to lose or maintain weight. The dietitian should be mindful of providing education on common side effects of resuming a regular meal pattern, including gastrointestinal effects (e.g., bloating, discomfort, constipation) and topics such as gut function, appetite regulation, weight and body changes, and social eating, among others.

It is important that dietitians adapt nutrition counseling to the patient's individual level of motivation, developmental expectations, psychological treatment model, and psychiatric comorbidities. Using motivational interviewing, employing collaborative goal setting, and encouraging self-monitoring can aid in keeping the patient engaged in treatment. Importantly, the dietitian should communicate regularly with the patient's treatment team throughout the duration of nutrition counseling.

Psychological treatment

Randomized controlled trials (RCT) show that cognitive behavior therapy (CBT), interpersonal psychotherapy (IPT), and dialectical behavior therapy (DBT) demonstrate effectiveness in reducing symptoms of BN in adults in the short- and long-term (e.g., Agras et al., 2000; Safer et al., 2001). A newer treatment, integrative cognitive-affective therapy for BN (ICAT-BN) demonstrates similar effectiveness as CBT (Wonderlich et al., 2014). Research of treatment for BN in adolescents is limited, but RCTs show that family-based treatment (FBT) demonstrates effectiveness in reducing symptoms when compared with supportive psychotherapy, and CBT demonstrates some evidence in reducing binge eating and purging when compared with established treatments. We describe each treatment below and provide a brief summary of evidence for their effectiveness.

Cognitive behavior therapy

"Enhanced CBT" for eating disorders (CBT-E; Fairburn, 2008) is a "transdiagnostic" treatment that can be adapted for any type of eating disorder, including BN. CBT-E works by addressing the thoughts and behaviors that maintain bulimic behaviors, including overvaluation of weight and shape and unhealthy weight control behaviors (i.e., dietary restraint) that perpetuate the binge-eating/purging cycle. The conceptualization of these maintaining factors is individualized for each patient, so that the treatment is tailored to their specific eating problem. In CBT-E, the therapist and patient work together to understand the eating problem, regulate the patient's eating using a prescribed, regular meal pattern and self-monitoring, and address the maintaining factors for the individual patient. There is evidence to support CBT-E as an effective treatment for BN in adults and older adolescents. CBT-E appears to result in significantly greater post-treatment and 60-week follow-up outcomes compared to another established treatment for BN, interpersonal psychotherapy (IPT; described below) (Fairburn et al., 2015).

Interpersonal psychotherapy

Individuals with BN report difficulty and problems in interpersonal relationships, including hostility toward others, communication difficulties, trouble understanding others' perspectives, and overall distress about relationships. The IPT theory posits that interpersonal difficulties cause low self-esteem and negative affect which then predispose one to bulimic behaviors. IPT focuses on

addressing these interpersonal problem areas, and deficits in social skills, which are thought to maintain eating disorder behaviors. By improving interpersonal functioning (and consequently, binge eating and purging as means for coping with dysfunction), IPT can be helpful in resolving bulimic symptoms. Two RCTs demonstrate IPT's effectiveness in treating BN (Agras et al., 2000; Fairburn et al., 1991). In each trial, CBT appeared superior to IPT in terms of end-of-treatment symptom reports. However, symptom reports were similar at 1-year follow-up, suggesting IPT may have a delayed, but similarly effective, impact on bulimic symptoms.

Dialectical behavior therapy

DBT for BN seeks to improve a patient's skills in four domains: interpersonal effectiveness, distress tolerance, emotion regulation, and mindfulness. Skill-building in these areas is thought to reduce affective lability, which in turn decreases bulimic behaviors. DBT was first developed for the treatment of chronic suicidality (Linehan, 2014), but was adapted for use with individuals with BN given commonalities in the symptom profiles of both groups (e.g., emotion dysregulation that causes engagement in impulsive behaviors such as binge-eating or self-harm). DBT is considered a "third wave" cognitive-behavioral therapy. Results from two RCTs offer preliminary evidence for DBT's effectiveness in reducing bulimic behaviors in BN. In one RCT, 28.6% of women with BN who were randomized to receive DBT reported abstinence from bulimic behaviors compared to no women in the waitlist control group, which was a statistically significant difference (Safer et al., 2001). In the other RCT, 26.9% of women with BN in the DBT condition reported abstinence from bulimic behaviors, and 61.5% no longer met criteria for BN at end of treatment, which was statistically significant difference compared to the control condition (Hill et al., 2011).

Integrative cognitive-affective therapy

ICAT-BN is the most recently developed of all the BN treatments. This treatment focuses on reducing emotion dysregulation, improving interpersonal skills, and addressing intrapersonal factors (including self-discrepancy and nutrition). ICAT-BN is based on models of conditioning and learning, which guide its interventions of eating disorder behaviors. ICAT-BN theory posits that individuals with BN experience negative affect due to the discrepancy between their actual and ideal self, which then causes them to engage in bulimic behaviors and self-criticism. A RCT showed comparable effectiveness of ICAT-BN and CBT at end-of treatment and 4-month follow-up (Wonderlich et al., 2014). ICAT-BN may be an acceptable alternative treatment option for individuals who do not respond to CBT and present with emotion dysregulation, interpersonal distress, and self-criticism.

Family-based treatment

FBT-BN is used for adolescents with BN and emphasizes the role of the family in helping the adolescent normalize their eating patterns and resume healthy adolescent development. FBT was originally developed for adolescents with AN, but has been adapted for the treatment of BN, with many of the same core components of the original treatment. FBT-BN seeks to empower parents to restore their adolescent back to health, and separate the adolescent from the illness. Although FBT-BN is newer than its predecessor, FBT-AN, there is preliminary evidence for its effectiveness in reducing bulimic behaviors. FBT-BN demonstrated a statistically significant difference in rates of abstinence from eating disorder behaviors when compared with nondirective supportive psychotherapy in one RCT (Le Grange et al., 2007). In another RCT, FBT-BN demonstrated comparable end of treatment and 6-month follow-up outcomes to CBT for adolescents (Le Grange et al., 2015).

Pharmacotherapy

Pharmacologic medication has been tested for BN (see Bello and Yeomans, 2018 for a review). Notably, pharmacotherapy can be used with or without psychotherapy, and tends to offer increased accessibility to treatment, as well as more convenient long-term management due to the low time commitment for medication management and the need for less intensive training (compared with psychotherapy training) for the treating health care professional. Pharmacotherapy can be especially helpful in addressing symptoms comorbid to BN (e.g., depression, anxiety, impulsivity), as better management of these symptoms can help patients fully engage in eating disorder treatment.

Fluoxetine is a selective serotonin reuptake inhibitor that was approved by the Food and Drug Administration (FDA) in 1994 for acute treatment and maintenance of BN. Across randomized placebo-controlled studies, there is evidence for the effectiveness of fluoxetine in treating BN. In a multi-center study of 387 women with BN, an 8-week course of fluoxetine (60 mg/day) resulted in a 67% reduction of binge-eating episodes and a 56% reduction of vomiting episodes from pretreatment; in contrast, the placebo condition resulted in a 33% reduction in binge-eating episodes and 5% reduction in vomiting episodes (Fluoxetine Bulimia Nervosa Collaborative Study Group, 1992). Findings also support that fluoxetine aids in relapse reduction; a 52-week RCT of 232 individuals with BN demonstrated that there was a significant decrease in relapse to binge eating and purging behaviors among the fluoxetine group, compared to a placebo condition. In some studies, fluoxetine is associated with minor side effects compared with placebo conditions, including insomnia, nausea, anxiety, dizziness, and decreased libido. However, other studies demonstrate few side effects. In sum, fluoxetine demonstrates both effective and safe reduction of bulimic behaviors in BN (Bello and Yeomans, 2018).

Treatment of underrepresented populations

Most of the research on BN treatment has occurred in high SES, White, female-identified patients. This disparity may be due to a lack of diagnosis in other groups; one study showed that White individuals were more likely to be diagnosed with and receive treatment for BN, despite Black individuals and girls from low-income families being more likely to exhibit bulimic behavior than White individuals and girls from high income families (Ham et al., 2015). Consequently, it is unclear whether established assessments and treatments are effective among people of color, men, gender and sexual minority groups, and low SES groups such as individuals experiencing food insecurity. Special attention should be paid to researching treatment effectiveness in these groups, and in adapting current treatment approaches to be culturally sensitive and accessible to underrepresented and underserved populations. Of note, guided self-help versions of specialty eating disorder therapies, such as CBT (Linardon et al., 2017), have been somewhat helpful in addressing the treatment-accessibility gap, but further work is needed to improve dissemination, cost, and access to the specialty BN treatments described above, particularly in underserved populations that appear to be at comparable or heightened risk.

Long-term prognosis

Help-seeking rates for eating disorder symptoms are quite low (13–35.6%). One review found that across studies, the average duration of untreated BN was 53 months, and it appeared that shorter duration of untreated BN was associated with greater likelihood of remission (Austin et al., 2021). Of those who do obtain treatment, symptom remission is not guaranteed; over 60% of patients fail to fully abstain from core BN symptoms, despite receiving first-line, empirically supported psychological treatments (Linardon and Wade, 2018). In terms of pharmacotherapy, post-treatment effects across medications are moderate, indicating that taking medication also does not guarantee remission. Findings such as these highlight the urgent need for research to improve the effectiveness and accessibility of existing treatments, and the development of culturally sensitive guidelines for implementing these treatments within groups historically underrepresented in the literature.

Conclusion

BN is a serious eating disorder associated with a chronic course, psychiatric comorbidity, and impairment. BN impacts individuals of various backgrounds and identities. Treatment should be prompt, and comprehensive in terms of psychological, pharmacological, and nutrition focus, specifically for those demonstrating undernourishment. Delays in treatment may prolong illness course, emphasizing the need for improved access, particularly for marginalized groups who may be at higher risk.

Funding information

This work was supported by a grant from the National Institute of Mental Health (T32 MH082761).

References

Agras, W.S., Walsh, B.T., Fairburn, C.G., Wilson, G.T., Kraemer, H.C., 2000. A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa. Arch. Gen. Psychiatr. 57 (5), 459–466.

Amorim Palavras, M., Hay, P., Claudino, A., 2018. An investigation of the clinical utility of the proposed ICD-11 and DSM-5 diagnostic schemes for eating disorders characterized by recurrent binge eating in people with a high BMI. Nutrients 10 (11), 1751.

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders. Washington, DC, fifth ed.

Arcelus, J., Mitchell, A.J., Wales, J., Nielsen, S., 2011. Mortality rates in patients with anorexia nervosa and other eating disorders: a meta-analysis of 36 studies. Arch. Gen. Psychiatr. 68 (7), 724–731.

Austin, A., Flynn, M., Richards, K., et al., 2021. Duration of untreated eating disorder and relationship to outcomes: a systematic review of the literature. Eur. Eat Disord. Rev. 29 (3), 329–345.

Bello, N.T., Yeomans, B.L., 2018. Safety of pharmacotherapy options for bulimia nervosa and binge eating disorder. Expert Opin. Drug Saf. 17 (1), 17–23.

Berg, K.C., Crosby, R.D., Cao, L., et al., 2013. Facets of negative affect prior to and following binge-only, purge-only, and binge/purge events in women with bulimia nervosa. J. Abnorm. Psychol. 122 (1), 111–118.

Berg, J.M., Latzman, R.D., Bliwise, N.G., Lilienfeld, S.O., 2015. Parsing the heterogeneity of impulsivity: a meta-analytic review of the behavioral implications of the UPPS for psychopathology. Psychol. Assess. 27 (4), 1129–1146.

Bruening, M., MacLehose, R., Loth, K., Story, M., Neumark-Sztainer, D., 2012. Feeding a family in a recession: food insecurity among Minnesota parents. Am. J. Publ. Health 102 (3), 520–526.

Bulik, C.M., Thornton, L.M., Root, T.L., et al., 2010. Understanding the relation between anorexia nervosa and bulimia nervosa in a Swedish National Twin Sample. Biol. Psychiatr. 67 (1), 71–77.

Fairburn, C.G., Jones, R., Peveler, R.C., et al., 1991. Three psychological treatments for bulimia nervosa: a comparative trial. Arch. Gen. Psychiatr. 48 (5), 463–469.

Fairburn, C.G., Bailey-Straebler, S., Basden, S., et al., 2015. A transdiagnostic comparison of enhanced cognitive behaviour therapy (CBT-E) and interpersonal psychotherapy in the treatment of eating disorders. Behav. Res. Ther. 70, 64–71.

Fairburn, C.G., 2008. Cognitive Behavior Therapy and Eating Disorders. Guilford Press.
Fischer, S., Smith, G.T., Cyders, M.A., 2008. Another look at impulsivity: a meta-analytic review comparing specific dispositions to rash action in their relationship to bulimic symptoms. Clin. Psychol. Rev. 28 (8), 1413–1425.
Fluoxetine Bulimia Nervosa Collaborative Study Group, 1992. Fluoxetine in the treatment of bulimia nervosa. Arch. Gen. Psychiatry. 49 (2), 139–147.
Forney, K.J., Buchman-Schmitt, J.M., Keel, P.K., Frank, G.K., 2016. The medical complications associated with purging. Int. J. Eat. Disord. 49 (3), 249–259.
Franklin, B., Jones, A., Love, D., et al., 2012. Exploring mediators of food insecurity and obesity: a review of recent literature. J. Community Health 37 (1), 253–264.
Goldstein, A., Gvion, Y., 2019. Socio-demographic and psychological risk factors for suicidal behavior among individuals with anorexia and bulimia nervosa: a systematic review. J. Affect. Disord. 245, 1149–1167.
Haedt-Matt, A.A., Keel, P.K., 2011. Revisiting the affect regulation model of binge eating: a meta-analysis of studies using ecological momentary assessment. Psychol. Bull. 137 (4), 660–681.
Ham, J.C., Iorio, D., Sovinsky, M., 2015. Disparities in bulimia nervosa: who is left behind? Econ. Lett. 136, 147–150.
Hill, D.M., Craighead, L.W., Safer, D.L., 2011. Appetite-focused dialectical behavior therapy for the treatment of binge eating with purging: a preliminary trial. Int. J. Eat. Disord. 44 (3), 249–261.
Le Grange, D., Crosby, R.D., Rathouz, P.J., Leventhal, B.L., 2007. A randomized controlled comparison of family-based treatment and supportive psychotherapy for adolescent bulimia nervosa. Arch. Gen. Psychiatr. 64 (9), 1049–1056.
Le Grange, D., Lock, J., Agras, W.S., Bryson, S.W., Jo, B., 2015. Randomized clinical trial of family-based treatment and cognitive-behavioral therapy for adolescent bulimia nervosa. J. Am. Acad. Child Adolesc. Psychiatr. 54 (11), 886–894.
Linardon, J., Wade, T.D., 2018. How many individuals achieve symptom abstinence following psychological treatments for bulimia nervosa? A meta-analytic review. Int. J. Eat. Disord. 51 (4), 287–294.
Linardon, J., Wade, T.D., De la Piedad Garcia, X., Brennan, L., 2017. The efficacy of cognitive-behavioral therapy for eating disorders: a systematic review and meta-analysis. J. Consult. Clin. Psychol. 85 (11), 1080–1094.
Linehan, M., 2014. DBT Skills Training Manual. Guilford Publications.
Lydecker, J.A., Grilo, C.M., 2019. Food insecurity and bulimia nervosa in the United States. Int. J. Eat. Disord. 52 (6), 735–739.
Molendijk, M.L., Hoek, H.W., Brewerton, T.D., Elzinga, B.M., 2017. Childhood maltreatment and eating disorder pathology: a systematic review and dose-response meta-analysis. Psychol. Med. 47 (8), 1402–1416.
Safer, D.L., Telch, C.F., Agras, W.S., 2001. Dialectical behavior therapy for bulimia nervosa. Am. J. Psychiatr. 158, 632–634.
Simone, M., Askew, A., Lust, K., Eisenberg, M.E., Pisetsky, E.M., 2020. Disparities in self-reported eating disorders and academic impairment in sexual and gender minority college students relative to their heterosexual and cisgender peers. Int. J. Eat. Disord. 53 (4), 513–524.
Strober, M., Freeman, R., Lampert, C., Diamond, J., Kaye, W., 2000. Controlled family study of anorexia nervosa and bulimia nervosa: evidence of shared liability and transmission of partial syndromes. Am. J. Psychiatr. 157 (3), 393–401.
Wonderlich, S.A., Peterson, C.B., Crosby, R.D., et al., 2014. A randomized controlled comparison of integrative cognitive-affective therapy (ICAT) and enhanced cognitive-behavioral therapy (CBT-E) for bulimia nervosa. Psychol. Med. 44 (3), 543–553.

Further reading

Ali, K., Fassnacht, D.B., Farrer, L., et al., 2020. What prevents young adults from seeking help? Barriers toward help-seeking for eating disorder symptomatology. Int. J. Eat. Disord. 53 (6), 894–906.
Becker, C.B., Middlemass, K., Taylor, B., Johnson, C., Gomez, F., 2017. Food insecurity and eating disorder pathology. Int. J. Eat. Disord. 50 (9), 1031–1040.
Bennett, H.S., Spiro, A.J., Pollack, M.A., Zucker, P., 1982. Ipecac-induced myopathy simulating dermatomyositis. Neurology 32 (1), 91.
Bickel, G., Nord, M., Price, C., Hamilton, W., Cook, J., 2000. Guide to Measuring Household Food Security, Revised. U.S. Department of Agriculture, Food and Nutrition Service, Alexandria, VA.
Bodell, L.P., Keel, P.K., 2015. Weight suppression in bulimia nervosa: associations with biology and behavior. J. Abnorm. Psychol. 124 (4), 994–1002.
Bove, C.F., Olson, C.M., 2006. Obesity in low-income rural women: qualitative insights about physical activity and eating patterns. Women Health 44 (1), 57–78.
Brown, T.A., Forney, K.J., Klein, K.M., Grillot, C., Keel, P.K., 2020. A 30-year longitudinal study of body weight, dieting, and eating pathology across women and men from late adolescence to later midlife. J. Abnorm. Psychol. 129 (4), 376–386.
Cândea, D.M., Szentagotai-Tătar, A., 2018. Shame-proneness, guilt-proneness and anxiety symptoms: a meta-analysis. J. Anxiety Disord. 58, 78–106.
Claudino, A.M., Pike, K.M., Hay, P., et al., 2019. The classification of feeding and eating disorders in the ICD-11: results of a field study comparing proposed ICD-11 guidelines with existing ICD-10 guidelines. BMC Med. 17 (1), 1–17.
Davis, H.A., Holland, L.A., Keel, P.K., 2014. A preliminary examination of a nonpurging compensatory eating disorder. Int. J. Eat. Disord. 47 (3), 239–243.
Davis, H.A., Ortiz, A.M.L., Smith, G.T., 2018. The occurrence and covariation of binge eating and compensatory behaviors across early to mid-adolescence. J. Pediatr. Psychol. 43 (4), 402–412.
Farstad, S.M., McGeown, L.M., von Ranson, K.M., 2016. Eating disorders and personality, 2004–2016: a systematic review and meta-analysis. Clin. Psychol. Rev. 46, 91–105.
Goldschmidt, A.B., Accurso, E.C., Crosby, R.D., et al., 2016. Association between objective and subjective binge eating and psychopathology during a psychological treatment trial for bulimic symptoms. Appetite 107, 471–477.
Grilo, C.M., Crosby, R.D., Machado, P.P., 2019. Examining the distinctiveness of body image concerns in patients with anorexia nervosa and bulimia nervosa. Int. J. Eat. Disord. 52 (11), 1229–1236.
Hadassah Cheng, Z.H., Perko, V.L., Fuller-Marashi, L., Gau, J.M., Stice, E., 2019. Ethnic differences in eating disorder prevalence, risk factors, and predictive effects of risk factors among young women. Eat. Behav. 32, 23–30.
Herman, C.P., Mack, D., 1975. Restrained and unrestrained eating. J. Pers. 43 (4), 647–660.
Heruc, G., Hurst, K., Casey, A., et al., 2020. ANZAED eating disorder treatment principles and general clinical practice and training standards. J. Eat. Disord. 8 (1), 1–9.
Hilbert, A., Pike, K.M., Goldschmidt, A.B., et al., 2014. Risk factors across the eating disorders. Psychiatr. Res. 220 (1–2), 500–506.
Himmerich, H., Hotopf, M., Shetty, H., et al., 2019. Psychiatric comorbidity as a risk factor for the mortality of people with bulimia nervosa. Soc. Psychiatr. Psychiatr. Epidemiol. 54 (7), 813–821.
Hudson, J.I., Hiripi, E., Pope Jr., H.G., Kessler, R.C., 2007. The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. Biol. Psychiatr. 61 (3), 348–358.
Jebeile, H., Lister, N.B., Baur, L.A., Garnett, S.P., Paxton, S.J., 2021. Eating disorder risk in adolescents with obesity. Obes. Rev. 22 (5), e13173.
Karam, A.M., Fitzsimmons-Craft, E.E., Tanofsky-Kraff, M., Wilfley, D.E., 2019. Interpersonal psychotherapy and the treatment of eating disorders. Psychiatr. Clin. 42 (2), 205–218.
Keel, P.K., Baxter, M.G., Heatherton, T.F., Joiner Jr., T.E., 2007. A 20-year longitudinal study of body weight, dieting, and eating disorder symptoms. J. Abnorm. Psychol. 116 (2), 422–432.
Kim, S., Thibodeau, R., Jorgensen, R.S., 2011. Shame, guilt, and depressive symptoms: a meta-analytic review. Psychol. Bull. 137 (1), 68–96.

Kiss, A., Wiesnagrotzki, S., Abatzi, T.A., et al., 1989. Upper gastrointestinal endoscopy findings in patients with long-standing bulimia nervosa. Gastrointest. Endosc. 35 (6), 516–518.
Lipson, S.K., Sonneville, K.R., 2017. Eating disorder symptoms among undergraduate and graduate students at 12 US colleges and universities. Eat. Behav. 24, 81–88.
Luoma, J.B., Chwyl, C., Kaplan, J., 2019. Substance use and shame: a systematic and meta-analytic review. Clin. Psychol. Rev. 70, 1–12.
McMaster, C.M., Wade, T., Franklin, J., Hart, S., 2020. A review of treatment manuals for adults with an eating disorder: nutrition content and consistency with current dietetic evidence. Eat. Weight Disord. 1–14.
Milos, G.F., Spindler, A.M., Buddeberg, C., Crameri, A., 2003. Axes I and II comorbidity and treatment experiences in eating disorder subjects. Psychother. Psychosom. 72 (5), 276–285.
Munn-Chernoff, M.A., Baker, J.H., 2015. A primer on the genetics of comorbid eating disorders and substance use disorders. Eur. Eat Disord. Rev. 24 (2), 91–100.
Munn-Chernoff, M.A., Grant, J.D., Agrawal, A., et al., 2015. Genetic overlap between alcohol use disorder and bulimic behaviors in European American and African American women. Drug Alcohol Depend. 153, 335–340.
Patel, R.S., Olten, B., Patel, P., Shah, K., Mansuri, Z., 2018. Hospitalization outcomes and comorbidities of bulimia nervosa: a Nationwide Inpatient Study. Cureus 10 (5), e2583.
Pearson, C.M., Wonderlich, S.A., Smith, G.T., 2015. A risk and maintenance model for bulimia nervosa: from impulsive action to compulsive behavior. Psychol. Rev. 122 (3), 516–535.
Polivy, J., Herman, C.P., 1985. Dieting and binging: a causal analysis. Am. Psychol. 40 (2), 193–201.
Prefit, A.B., Cândea, D.M., Szentagotai-Tătar, A., 2019. Emotion regulation across eating pathology: a meta-analysis. Appetite 143, 104438.
Rasmusson, G., Lydecker, J.A., Coffino, J.A., White, M.A., Grilo, C.M., 2018. Household food insecurity is associated with binge-eating disorder and obesity. Int. J. Eat. Disord. 52 (1), 28–35.
Romano, S.J., Halmi, K.A., Sarkar, N.P., Koke, S.C., Lee, J.S., 2002. A placebo-controlled study of fluoxetine in continued treatment of bulimia nervosa after successful acute fluoxetine treatment. Am. J. Psychiatr. 159 (1), 96–102.
Rosten, A., Newton, T., 2017. The impact of bulimia nervosa on oral health: a review of the literature. Br. Dent. J. 223 (7), 533–539.
Sheehy, K., Noureen, A., Khaliq, A., 2019. An examination of the relationship between shame, guilt and self-harm: a systematic review and meta-analysis. Clin. Psychol. Rev. 73, 101779.
Silva, C.L.M., de Moraes, C.E.F., da Veiga, G.V., 2021. An evaluation of binge eating characteristics in individuals with eating disorders: a systematic review and meta-analysis. Appetite 105176.
Stice, E., 2002. Risk and maintenance factors for eating pathology: a meta-analytic review. Psychol. Bull. 128 (5), 825–848.
Striegel-Moore, R.H., Dohm, F.A., Kraemer, H.C., 2003. Eating disorders in white and black women. Am. J. Psychiatr. 160 (7), 1326–1331.
Tozzi, F., Thornton, L.M., Mitchell, J., et al., 2006. Features associated with laxative abuse in individuals with eating disorders. Psychosom. Med. 68 (3), 470–477.
Udo, T., Grilo, C.M., 2018. Prevalence and correlates of DSM-5–defined eating disorders in a nationally representative sample of US adults. Biol. Psychiatr. 84 (5), 345–354.
Ulfvebrand, S., Birgegård, A., Norring, C., Högdahl, L., von Hausswolff-Juhlin, Y., 2015. Psychiatric comorbidity in women and men with eating disorders results from a large clinical database. Psychiatr. Res. 230 (2), 294–299.
Volpe, U., Tortorella, A., Manchia, M., et al., 2016. Eating disorders: what age at onset? Psychiatr. Res. 238, 225–227.
Whitelaw, M., Lee, K.J., Gilbertson, H., Sawyer, S.M., 2018. Predictors of complications in anorexia nervosa and atypical anorexia nervosa: degree of underweight or extent and recency of weight loss? J. Adolesc. Health 63 (6), 717–723.
Yao, S., Kuja-Halkola, R., Thornton, L., et al., 2019. Overlap between anorexia and bulimia nervosa: understanding genetic and environmental effects in a Swedish population registry. Eur. Neuropsychopharmacol. 29, S817–S818.
Zunker, C., Peterson, C.B., Crosby, R.D., et al., 2011. Ecological momentary assessment of bulimia nervosa: does dietary restriction predict binge eating? Behav. Res. Ther. 49 (10), 714–717.

Energy balance

Yves Schutz, Department of Medicine, Physiology Unit, University of Fribourg, Fribourg, Switzerland

© 2023 Elsevier Ltd. All rights reserved.

This is an update of Y. Schutz, ENERGY: Balance, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 154–163, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00089-1.

Introduction 315
Energy balance: definition 315
Model of energy balance: a dynamic state 316
Gross and metabolizable energy 316
Total energy expenditure and its components 317
Basal metabolic rate and resting metabolic rate 317
Thermic effect of food or postprandial thermogenesis 318
Energy expenditure due to physical activity 318
Energy balance and macronutrient balance 318
The classical fat balance concept: needs to be refined under certain nutritional conditions 319
Energy imbalance and body weight relationship 319
Energy balance regulation 320
Dynamic and static evolution of energy balance: overfeeding and underfeeding conditions 321
Conclusions 323
References 323
Further reading 323

Key points

- Energy balance is the difference between metabolizable energy intake and total energy expenditure. It is strongly related to macronutrient balances, and the sum of the individual substrate balances, expressed as energy, are mathematically equivalent to the overall energy balance.
- Energy in foods is furnished by carbohydrate, proteins, fats, and alcohol; only 5–10% is lost through the feces and urine.
- The energy available to the body, called 'metabolizable energy', is on average 17 kJ g^{-1} of carbohydrate, 17 kJ g^{-1} of protein, 37 kJ g^{-1} of fat, and 29 kJ g^{-1} of alcohol. These figures vary slightly according to the types of carbohydrate, protein, or fat in the diet.
- The energy used in the body, or energy expenditure, is classically assessed by indirect calorimetry. It involves measuring non-invasively the oxygen consumption and carbon dioxide production by an individual.
- Short-term regulation of energy balance is poor, but (in most people) long-term regulation is relatively accurate. The exact mechanism is still unknown, but must include, in certain individuals, conscious alterations in lifestyle to correct unwanted changes in body weight.
- During long periods of energy imbalance, the weight gained (or lost) is initially glycogen plus water with an energy density of 1.0000 kcal/kg. If the imbalance continues, after a week the tissue gained (or lost) is a mixture of mostly fat, water, and proteins.
- Undernutrition leads to a decrease in energy expenditure. Part of the decrease in absolute metabolic rate is related to weight loss.
- During overfeeding, although much of the excess energy intake will be stored in adipose (fat) tissue, there are compensatory increases in thermogenesis and energy expenditure which will moderate the level of imbalance over time together with the parallel effect of body weight gain.

List of abbreviations

BMR Basal metabolic rate
REE Resting metabolic rate
TEF thermic effect of food
EE Energy expenditure
NEAT Non-exercise activity thermogenesis

Introduction

To maintain physiologic functions, the human body continuously expends energy by oxidative metabolism. This energy is used to maintain chemical and electrochemical gradients across cellular membranes for the biosynthesis of macromolecules such as proteins, glycogen, and triglycerides, and for muscular contraction. Another part of the energy is lost as heat because of the inefficiency of metabolic transformations. Ultimately all the energy produced by the organism is dissipated as heat.

The energy expended by an individual can be assessed by two different techniques: indirect and direct calorimetry. The term indirect calorimetry stems from the fact that the heat released by chemical processes within the body can be indirectly calculated from the rate of oxygen consumption. Indirect calorimetry measures the heat released by the oxidative metabolic processes and direct calorimetry assesses the heat dissipated by the body. The average values of the two over a period of 24 h or more are very close in steady state conditions. A relationship exists between these two variables although they are not in strict temporal synchronization since when an acute exercise is performed during the day, the total heat loss response lags behind that of heat production, thereby resulting in a small and transient increase in body central temperature. The difference between metabolic heat production and heat dissipation represents the body "heat balance", which, over a 24 period, will be generally in equilibrium in healthy subjects without developing fever.

The total body heat losses are equal to the sum of several components, namely the non-evaporative components (i.e., the radiant heat exchange, the convective and conductive heat transfers) plus the evaporative component (i.e., heat losses through the skin and the pulmonary losses consecutive to water vapor loss during expiration).

When average total heat production is not equal to average total heat loss, the heat imbalance will indicate either a net heat *storage* when the former is larger than the latter or a net heat *dissipation* when the latter is higher than the former. As a result, core (internal) body temperature will progressively rise or decrease respectively.

Energy balance: definition

Overall energy balance is given by the following equation:

$$\text{Energy balance} = \text{total energy intake} - \text{total energy expenditure}$$

Thus, if the total energy contained in the body (as total fat, protein and glycogen) is not altered, average energy expenditure must be equal to average energy intake. In a state of positive energy balance over a long period of time, an increase in body energy stores will occur, primarily as body fat, and vice versa in a state of negative energy balance. The differences between the concepts of energy balance, substrate balance and heat balances are presented in **Fig. 1**.

This very simple equation has plagued nutritionists and physiologists for many decades, since it masks a number of difficulties that are of practical, methodological and conceptual nature:

(1) The methodology to assess energy intake is much more uncertain than that assessing total energy expenditure by modern methods namely by stable non-radioactive isotopes (D2O18) or respiration (whole body) chamber.

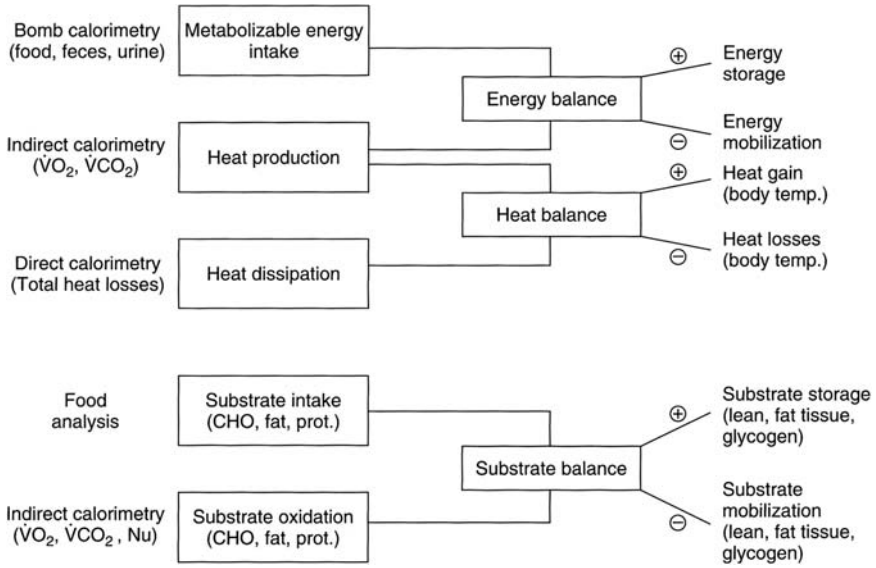


Fig. 1 Heat balance, energy balance, and substrate balance: three different complementary concepts measuring different things.

(2) The inherent bias of the subject under investigation can lead to large errors in the quantitative food intake assessment. In contrast, one cannot cheat on the total energy expenditure value as it is generally inconspicuously and objectively assessed by physiological measurements.

This issue of large errors inherent to food energy intake measurement in free-living conditions has become important for the medical/nutritional scientists. As a result, it was proposed to calculate "true" energy intake indirectly (using the formula above) from total energy expenditure (TEE). As a result, the energy balance is assumed to be in equilibrium (or close to equilibrium) over a week or so, namely Ein = TEE.

If we accept this simplified postulate, the relative error in food energy intake assessment by classical methodology can be then estimated. Incidentally, large (non-voluntarily) under-estimation of total energy intake has been systematically reported in obese subjects/patients.

(3) The natural variability of energy intake from day-to-day in a given subject is much greater than that of energy expenditure in free-living conditions. This is quite expected since just one day fasting (or feasting), in a time series of habitual food intakes, will drastically inflate its intra-individual coefficient of variation. In contrast, energy expenditure is never zero in a living individual but has a baseline value equal to resting post-absorptive metabolic rate and a lower minimal value during the night while sleeping.

(4) Eating meals constitutes a discontinuous process (generally 2–5 meals per day), whereas metabolic energy expenditure is of continuous nature. Thus, when calculated over day-time, energy balance will largely be positive whereas it is compensated by a negative value during night-time, a period during which the person normally sleeps and rarely eats meals.

This rhythmic day/night pattern of energy balance occurs normally in most individuals, who are not involved in work shifts. When the 2 phases are cumulated, the 24 h energy balance can be in equilibrium, the diurnal period of positive balance is compensated by the negative one during the nocturnal period.

This emphasizes that an energy balance should be calculated over a minimum of 24 h (for example in a respiration chamber) in order to be able to assess the potential energy storage or mobilization of body tissues. However, evidence of full continuous energy equilibrium may require a longer measurement duration such as several days and even weeks (see below).

(5) When the magnitude of energy imbalance (either positive or negative) has a large gap, this permits an appropriate interpretation of the results for comparing those values with the time course of body weight and body composition simultaneously observed.

Model of energy balance: a dynamic state

There are multiple reciprocal direct and indirect influences of energy intake on energy expenditure and vice versa: for example, energy intake influences resting energy expenditure by increasing postprandial and dietary-induced thermogenesis, whereas changes in energy expenditure via a modification of physical activity are susceptible to influence energy intake to re-equilibrate energy balance. To ensure an accurate regulation of body stores, a double servo-control is essential (Fig. 2).

Body weight and body composition are not invariant with time, but small corrections of both input and output from day-to-day, week-to-week or more, ensure energy homeostasis. When attempting to explain the actual responses in energy balance and weight regulation in real life, we need to recognize that several factors may be operating on both sides of the energy balance equation. Compensatory adjustments occur in both intake and output sides, but unraveling their respective importance, is not an easy task in humans. Much more advanced, sophisticated and complex models of energy balance have been developed over these last decades for predicting the evolution of body weight and body composition profile, but these are outside the scope of this paper.

Gross and metabolizable energy

The traditional way of measuring the energy content of foodstuffs is to use a "bomb calorimeter" in which the heat produced when a small representative sample of food is combusted (under high pressure of oxygen) is measured. The food is completely and instantaneously oxidized to water, carbon dioxide and small mineral incombustible residues. The total heat liberated (expressed in

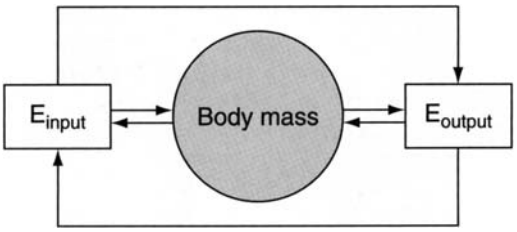


Fig. 2 Simplified conceptual model of energy balance showing the numerous feedback systems operating between energy input and energy output toward body weight regulation.

kilocalories or kilojoules) represents the gross energy value or heat of combustion of the food. The heat of combustion differs between carbohydrates, proteins, and fats. There are also important differences within each category of macronutrients. For example, sucrose has a gross energy of 16.5 kJ g^{-1}, whereas starch yields 17.5 kJ g^{-1}. The energy yield of butterfat is 38.5 kJ g^{-1} and of lard is 39.6 kJ g^{-1}. These values have been rounded off to give 17.3 kJ g^{-1} for carbohydrates rich in starch and poor in sugar, 39.3 kJ g^{-1} for average fat, and 23.6 kJ g^{-1} for mixtures of animal and vegetable proteins.

However, the gross energy value of foodstuffs does not represent the energy actually available to the body, since no potentially oxidizable substrate (macronutrients) can be considered available to the body until it is presented to the cell for oxidation. None of the foodstuffs are completely absorbed in the small intestine and some residual unabsorbed foods are excreted in the feces. Digestibility of the major macronutrients are high; on the average, 97% of ingested mixed carbohydrates, 95% of average fats and 92% of average exogenous proteins are absorbed from the intestinal lumen.

In the body, the tissues are able to oxidize carbohydrate and fat completely to carbon dioxide and water, but the oxidation of protein is incomplete since it results in the formation urea in the liver (and other nitrogenous compounds such ammonia), which are excreted in the urine with an energy equivalent of 5.2 kJ g^{-1} of protein ingested. With a mixed diet, rich in carbohydrates and fibers, the metabolizable energy of food is approximately 90% of the gross energy value (heat of combustion). The remaining 10% is mainly due to unabsorbed energy (fecal energy losses). The latter value increases with the amount of dietary (insoluble) fibers (for example, cellulose).

Total energy expenditure and its components

Classically we simply considered energy expenditure as being made up of 3 components: the energy spent for basal metabolism or basal metabolic rate, (BMR), the energy spent on physical activity, and the increase in resting energy expenditure in response to a variety of stimuli (in particular food, cold, stress, and drugs). Additional subcomponents have been included such as the partition of physical activity into structured exercise (such as sporting activities) and non-exercise activity thermogenesis (NEAT) such as fidgeting.

For a more detailed outline of the components of TEE, refer to the chapter "Energy adaptation".

Basal metabolic rate and resting metabolic rate

These are the largest component of energy expenditure accounting for between half to 3/4 of total daily energy expenditure. BMR is measured under standardized conditions, that is, in an awake subject lying in the supine position, in a state of physical and mental rest, in a comfortable warm environment, and in the morning in the postabsorptive state, usually 10–12 h after the last meal. Resting metabolic rate (RMR) is measured at rest, supine but the conditions of measurements are less strict, for example in inter-prandial conditions. The distinction between RMR and BMR in the literature is not always clear. RMR could be considered almost equivalent to BMR if the measurements are made in postabsorptive conditions; the terms RMR and BMR are used interchangeably in the sections below.

It seems difficult to partition RMR into various subcomponents because the metabolic rates of individual organs and tissues are hard to assess in humans under noninvasive experimental conditions. BMR can vary up to 10% between individuals of the same age, gender, body weight, and body composition (fat-free mass), suggesting that genetic factors are also important. Day-to-day (intraindividual) variability of BMR is low in men (coefficient of variation of 1–3%) but is slightly greater in women because the menstrual cycle affects BMR. In both women and men, sleeping metabolic rate is lower than BMR by 5–10%, the difference being explained by the effect of arousal.

The major part of the whole-body RMR (~60%) stems from organs with high-metabolic activity such as the liver, kidneys, brain, and heart, although these account for a small proportion of the total body weight (5%). Per unit body weight, the kidneys, and heart have a metabolic rate more than twice as high as the liver and the brain. In contrast, the metabolic rate of muscle per unit body weight is nearly 35 times lower than that of the heart and kidneys. Since the proportion of muscle to non-muscle changes with age from birth to adulthood, the RMR per unit body weight is not constant with age. The tissue with the lowest metabolic activity per unit body weight is adipose (fat) tissue, which accounts for only 4% of the whole-body RMR in nonobese subjects. Calculations show that this value can increase up to 10% or more in obese subjects with a large excess in body fat. Skin and intestines (which have a relatively large protein mass and protein turnover), as well as bones and lungs, also contribute significantly to RMR.

Numerous studies have demonstrated that a major factor explaining the variation in RMR between individuals is fat-free mass (FFM), a heterogeneous component that can be partitioned into muscle mass and non-muscle mass. Unfortunately, there is no simple and accurate way to assess these two subcomponents. Owing to the larger variation between individuals in fat mass, as compared to FFM, and because in grossly obese women fat mass can represent a nonnegligible component of total RMR, the prediction models for RMR that include both FFM and fat mass explain significantly more variance in RMR than FFM alone. In addition, age, sex, and family membership are additional factors that should be taken into account.

The effects of gender on RMR are explained by differences in body composition. Caution should be used when comparing RMR expressed per kilogram FFM in men and women, because the composition of FFM is influenced by gender. The muscle mass of men is greater than that of women and this tends to give a lower value of RMR per kilogram FFM in men when compared to that of

women. This is explained by a greater component of tissue with a low metabolic rate per unit mass (resting muscle) in men than in women.

In clinical work, where body composition is difficult to assess, body weight, gender, and age can be used to estimate BMR and RMR, bearing in mind that many important determinants of RMR, in addition to body size, have been identified under noninvasive experimental conditions. BMR can vary up to 10% or more between individuals of the same age, gender, body weight, and FFM, suggesting that genetic factors are also important.

Thermic effect of food or postprandial thermogenesis

The basal energy expenditure increases significantly after a meal. The so-called thermic effect of food is mainly due to the energy cost involved from the mouth to the endogenous tissues. The total thermic effect of food represents about 10% of the total energy intake over 24 h in sedentary subjects. The thermic effect of macronutrients mainly depends upon the energy costs of processing and/or storing the exogenous nutrient. Expressed as a percentage of the energy content of each macronutrient, values of about 8%, 2%, 20–30%, have been reported for glucose, fat, protein respectively, and as much as 22% for alcohol (ethanol).

For example, protein synthesis requires much energy due to the high cost of peptide bonds synthesis and biochemical processing. For the concept of altered thermic effect of food in adaptive thermogenesis, refer to the chapter "Energy adaptation".

Energy expenditure due to physical activity

The energy spent on physical activity depends on the type and intensity of the physical activity and on the time spent in different activities. Physical activity is often considered to be synonymous with "muscular work," which has a classical definition in physics (i.e., force *x* distance) when external work is performed on the environment (such as climbing).

During muscular work (muscle contractions), the muscle produces three to four times more heat than mechanical energy so that useful work costs more than muscle work. There is a wide variation in the energy cost of any activities both within and between individuals. The latter variation is due to differences in body size and in the speed and dexterity with which an activity is performed. To adjust for differences in body size, the energy cost of physical activities is expressed as multiples of BMR, the latter taken as baseline. The values generally range from 1 (= resting) to 5–6 for most daily habitual individual activities, but can reach values between 10 and 15 during intense exercise. In terms of daily energy expenditure, physical activity accounts for 15–40% of total energy expenditure (TEE) but it can represent up to 70% of daily energy expenditure in an individual involved in heavy manual work or competition sports. However, for most people in industrialized societies the contribution of physical activity to daily energy expenditure is relatively limited by occupation.

The energy expenditure associated with everyday life activity, which excludes structured exercise, is called non-exercise activity thermogenesis (NEAT). It plays a pivotal role in the regulation of human energy metabolism and body weight regulation. In practice, NEAT includes the energy expended while moving around (e.g., during household, occupational or leisure activities), walking back and forth (pacing) and fidgeting. It can be measured by the so-called factorial approach: each component of NEAT is quantified, and total NEAT is calculated by summing up these components. NEAT varies substantially between people of similar size in part because of the substantial variation in the amount of spontaneous physical activity they perform. Obesity is generally associated with low NEAT's, the deficit being not negligible: Obese individuals stand and ambulate significantly less and seat more than their lean counterparts.

Just as described above for a specific activity, it has been customary to express TEE relative to RMR (TEE/RMR or TEE/BMR), in order to (partially) offset the large variation in RMR among subjects of different body weight and body composition. This quotient is called physical activity level (PAL) and expressed as a the multiple of RMR measured over a 24 h period. For example, a PAL of 1.5 indicates that TEE is 50% greater than RMR (BMR) over 24 h.

Energy balance and macronutrient balance

Since macronutrients (carbohydrate, fat, protein) are the major sources of energy (if alcohol is excluded), it is logical to consider energy balance and macronutrient balance together as the opposite side of the same coin Schutz (1995).

Provided the respiratory quotient, i.e., the ratio of CO2 production to oxygen consumption is known, indirect calorimetry also allows an estimation of the macronutrient oxidation rates in the whole body in addition to energy expenditure.

The 3 macronutrients crude balance are defined as:

$$\text{Carbohydrate balance} = \text{carbohydrate intake} - \text{carbohydrate oxidation}$$
$$\text{Lipid(Fat) balance} = \text{lipid(fat) intake} - \text{lipid(fat) oxidation}$$
$$\text{Protein balance} = \text{protein intake} - \text{protein oxidation}$$

There is a direct mathematical relationship between total energy balance and total macronutrient balance, since the arithmetic sum of individual substrates balance (when macronutrients are expressed as energy) is equivalent to the overall energy balance. Note that the importance of lipid (fat) balance concept merits more details, given in a subsequent paragraph.

In a study composed of heterogeneous individuals (both lean and obese ones pooled), energy balance calculated over a period of weeks, was well correlated to total fat balance. This indicates that when there is an increase in body energy stores due to positive energy balance, this was mostly explained by body fat accumulation in adipose tissue rather than increase in muscles.

Total body energy stores (constituted mainly of fat stores in adipose tissue) represent a large proportion of the total daily energy ingested. For example, let's take a 60 kg non-obese woman (with 25% body fat), who is eating 2000 kcal/day), this would represent a total adipose tissue storage (at 8000 kcal/kg for fat) of 120,000 kcal. Hence, this represents theoretically 60 days of fat available stored, whereas carbohydrate storage (as glycogen) constitutes only 3–4 days of reserve.

The classical fat balance concept: needs to be refined under certain nutritional conditions

The definition of "crude" fat (lipid) balance described above is somehow more complex than a simple balance equation, since, in the intermediary metabolism, one substrate (or molecule) can be converted into another substrate. For example, carbohydrates can be transformed into fat in acute carbohydrate overfeeding conditions. This process is called *de novo lipogenesis*, meaning new synthesis of fat derived from another precursor that exogenous fat (Schutz, 2004).

Taken together the above considerations mean that we must distinguish 3 types of fat balance rather than only one (Fig. 3):

(1) *Classical* (crude) fat balance, typically used in nutrition/dietetics
(2) *Dynamic* fat balance, a very useful concept to use when energy dislocation (+ or − energy imbalance) leads to body weight gain and body weight loss respectively.
(3) Further *sub-partitioning* of fat balance, which can be split into an exogenous component (coming from food) and that mobilized from the body (endogenous one). This more complex type of fat balance is useful when interpreting acute overfeeding (or underfeeding) conditions. It primarily concerns situations such as a surfeit of foods rich in carbohydrates or sugar dissolved in liquids (such as soda, the so-called *soda syndrome*). Note that chronic excess alcohol consumption has also lipogenic capacities.
(4) In habitual nutritional conditions, the fat stored in adipose tissue mainly comes from dietary fat (i.e., exogenous). The best evidence of this, is that the relative composition of fatty acids in adipose tissue (expressed in % of total tissue fat) closely reflects that of fat ingested in particular the nutritionally essential poly-unsaturated fatty acids of w3 and w6 series.

The transformation of exogenous carbohydrates into endogenous fat occurs in the liver and, to a lesser extent, in adipose tissue. This has important consequences for this key organ, as the liver is a real metabolic "carrefour", which regulates and controls a vast number of functions for maintaining metabolic homeostasis and is responsible of many biosynthesis activities.

The mechanism of *de novo* lipogenesis leading to excess endogenous fat in the liver has negative functional consequences in particular if it is associated to physical inactivity. The placidity results in a blunted mobilization of carbohydrates stored as glycogen in the liver and muscles, precluding its subsequently oxidation by the active working muscles. The progressive accumulation in the liver is subsequently converted into fat as schematically presented in Fig. 4.

Energy imbalance and body weight relationship

Positive energy balance leads to body weight gain and negative energy balance leads to body weight loss. There is no fixed relationship between these two variables, since there could be considerable variations in the composition of the weight change in terms of fat, lean tissues and associated water stored or mobilized.

1) *Classical nutritional (static)* : Fat bal. = Fat_{in} - Fat_{ox}

2) *Dynamic fat balance (dynamic)*: Δ Fat bal. = $\Delta\,\text{Fat}_{in}$ - $\Delta\,\text{Fat}_{ox}$

3) *Partitioning of fat balance between exogenous/endogenous*

a) Exog. Fat bal. = Fat_{in} - $\text{Fat}_{ox\,exog.}$

b) Endog. Fat bal. = *de novo* Fat synthesis - $\text{Fat}_{ox\,endog.}$

Total: (a+b) = (Fat_{in} + *de novo* $\text{Fat}_{synth.}$) - ($\text{Fat}_{ox\,exo\,+\,endo}$)

Fig. 3 When macronutrients are considered, the classical fat balance is considered to be static. In dynamic conditions, during which fat storage or fat mobilization occur, *delta* fat balance will be more compatible with the independent measurement of serial body composition. In addition, the classical fat balance has to be adjusted in case of particular nutritional conditions, during which *net de novo* lipogenesis i.e., a transformation of exogenous carbohydrates into endogenous fat. The total equation considering these factors is given in the bottom line. Abbreviations: bal. = balance; in = intake; ox = oxidation; Δ = change in; exog. = exogenous; endo. = endogenous; synth = synthesized.

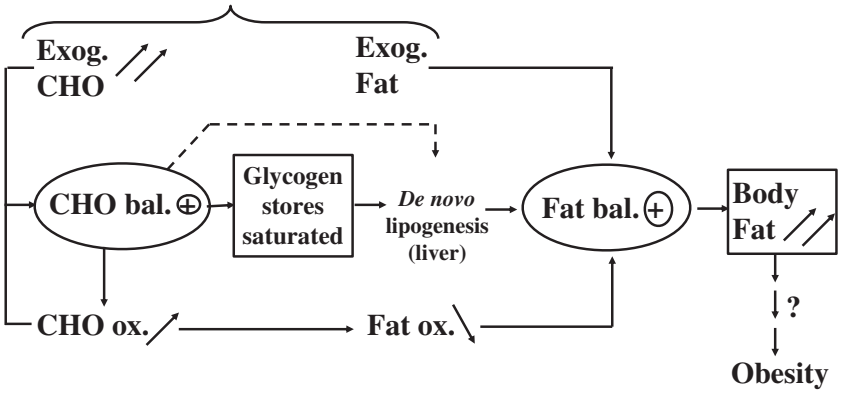


Fig. 4 The mechanism of *de novo* lipogenesis is schematically shown when surfeit carbohydrates are deliberately over consumed continuously for a period of approximately one week. In real-life situation of repeated high CHO (sugar) overfeeding on obesity risk development, the dual mechanism perpetuating a net positive fat balance accompanied by a rise in body fat stores, can be summarized: the total body fat gain is explained by (1) a drop in endogenous fat oxidation (antilipolytic effect of insulin) to which is superposed to (2) the effect of stimulation in *de novo* lipogenesis resulting from the net conversion of carbohydrates into fat notwithstanding the retention of exogenous fat of the meal (if any). Abbreviations: CHO = carbohydrates; Exog. = Exogenous; bal. = balance; ox. = oxidation; arrow up = increase; arrow down = decrease.

Over short term, relatively small energy retention or mobilization on a daily basis can cumulate over time and be accompanied by important body weight gain and vice versa. The confounding factor is the associated water retained in certain tissues in particular when glycogen storage or glycogen mobilization occur in the early phase of acute feeding or depriving carbohydrate, respectively. The calculated water dilution rate of "dry" glycogen by water in the tissue is approximately 3, indicating that each 100 g of glycogen stored/mobilized was associated to approximately 300 g of water in the glycogen-water pool displaced. Over longer term, this transient situation plays little role and cumulated energy retention (or mobilization) becomes compatible with fat retention (or mobilization), if we neglect the protein storage or mobilized.

In summary, day-to-day body weight fluctuations are generally accommodated by water retention due to changes in carbohydrate storage as well as acute sodium intake fluctuations. Body weight tend to be spontaneously maintained over middle term, despite large natural day-to-day fluctuations of macronutrients balance and body weight. The large fluctuations of CHO balance are explained by the variability of total CHO eaten from day-to-day vs that oxidized.

Energy balance regulation

Any regulated function varies within a certain window that is largely determined by the physiological homeostasis which ultimately will favor survival Galgani and Ravussin (2008).

Short-term day-to-day energy imbalance is mostly accommodated by rapid changes in carbohydrate balance, whereas over a prolonged period of time, positive energy balance is mostly expressed as fat storage since carbohydrate stores remain limited to a maximum of about 1 kg. By contrast, the fact that the body weight of certain individuals varies in a very narrow range over years/decades, in spite of large day-to-day fluctuations in both the amount of food (macronutrients) consumed and also in the level of physical activity, may suggest that body weight is accurately regulated in these individuals. A critical feature of any regulated system is the metabolic response to a voluntary disturbance: the regulated variable should result in a compensatory response that tend to attenuate the initial disturbance and to restore the system to its "set" or "preferred" value or initial state. It seems important to highlight key features of energy balance and weight regulation (refer to the companion chapter on Energy Adaptation).

First, the energy balance spontaneously varies from day-to-day because human beings cannot precisely match total energy intake and total energy expenditure on a day-to-day basis. The changes in daily energy intake and expenditure appear relatively random and not necessarily synchronized. Positive energy balance on one day may not be spontaneously and fully compensated by negative energy balance on the subsequent day, so that it is important to consider the overall energy balance regulation over a sufficiently prolonged period of time.

Near equality of energy intake and expenditure in non-obese individuals can most often appear over a period of 1–2 weeks. Longer term research measurements (over several weeks) are technically difficult and impractical to conduct for the subjects, and also because of cumulative systematic energy measurements errors on both sides of the energy balance.

Second, this matching of long-term energy intake and energy expenditure must be very accurate in certain individuals, because a theoretical error (imbalance) of only 1% between energy input and energy output, if constant and persistent over several years and if not compensated, will lead to a theoretical gain or loss of approximately 10 kg in the first decade of imbalance if the body would be invariant.

Dynamic and static evolution of energy balance: overfeeding and underfeeding conditions

To better understand the dynamic aspect of energy balance (i.e., in non-steady state or not) we need to stimulate the body by an acute and continuous perturbation such as overfeeding/underfeeding to demonstrate that the system is not inert and invariant; see Bray, and Bouchard (2020) for an extensive analysis of human overfeeding studies. As shown in Fig. 5A step (and sustained) increase in energy intake above energy requirement will lead to a gradual gain in body weight. The size of the energy imbalance will progressively diminish with time as weight is gained. The reason for this is that the expansion of FFM and fat mass (adipose tissue) will be accompanied by a rise in energy expenditure. A new equilibrium in weight is eventually reached after adaptation of each component of TEE, that is, RMR, diet-induced thermogenesis and body weight.

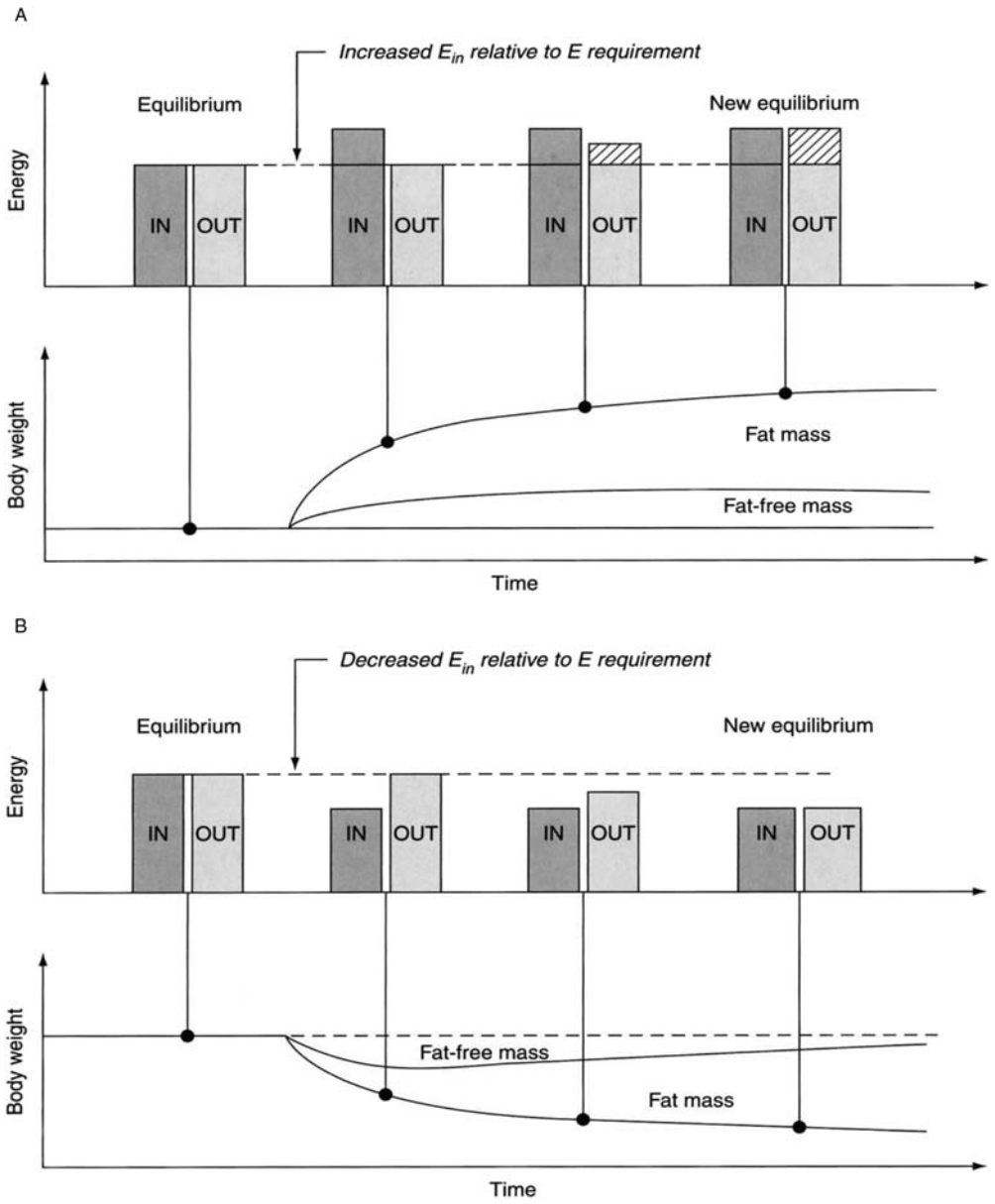


Fig. 5 Dynamic change in energy balance following a step (and sustained) increase (A) or decrease (B) in total energy intake. The time required to reach a new equilibrium in energy balance is very long (several years) and depends on the magnitude of initial energy imbalance, the magnitude of adaptation of energy expenditure in response to the step change in energy intake, and on the factors related to the subject (obesity vs. leanness). Let us take the following practical example: small increase in energy intake, for example, of 100–200 kcal/day, will induce small increases in body weight, associated to a rise in energy expenditure as the mass of lean tissue increases. If these changes occur month-by-month, and if after 3–5 years the adult is still eating an excess of 200 kcal/day over the baseline value, they will now be progressively heavier. Ultimately, after a long period of adaptations (years), they will reach a new energy balance equilibrium at a higher energy intake level. As a result, their weight will tend to reach a plateau after an expected (theoretical) 8–10 kg weight gain.

In a perfectly regulated system, any increase in energy intake should be offset by a change in energy expenditure of the same magnitude and direction. However, a 100% efficient adaptive process would obviously be counterproductive, since this would signify that an increase in energy storage (required during the anabolic state of nutritional rehabilitation) or an increase in energy mobilization (required for decreasing body weight) would be very limited. Briefly, adaptation to energy imbalance occurs at the cost of increasing (or decreasing) body weight.

In fact, excess energy intakes result in an increased metabolic turnover and energy flux through the mechanism of adaptive thermogenesis. The efficiency of energy storage is not constant and depends upon several factors including the magnitude of energy imbalance the composition of the surfeit energy fed, as well as endogenous factors.

Analysis of underfeeding experiments shows that the progressive decrease in total energy expenditure has 3 components. First, if energy intake is decreased the thermic effect of food (approximately 10% of energy intake) is similarly decreased. Second, there is a quick adaptive decrease in metabolic rate during the first week, related in part to a decrease in sympathetic activity. The magnitude of this decrease is only 5–8%. Third, there is a decrease in metabolic rate related to body weight lost: most investigators find a decrease in 10–12 kcal per day per kg weight loss.

The effect combined effects of all 3 processes are that a person who lost weight from, say, from 100 kg to 70 kg (a 30% reduction in weight) would experience approximately a 15% reduction in energy expenditure and hence in energy requirements for weight maintenance. Thus, a decrease in energy intake causes a reduction in body weight up to the time that a new equilibrium will be reached at which the reduced requirement will be again compatible with the reduced intake, and therefore body weight will ultimately stabilize.

Taken together, we can conclude that the efficiency of energy utilization is lower in overfeeding than in underfeeding conditions because substrate storage in tissues is energetically costly (ATP needs), whereas the process of energy mobilization requires little energy. Adaptive changes in thermogenesis attenuates the impact of excessive or insufficient food consumption on energy balance. The magnitude of adaptive thermogenesis, which shows high inter-individual variability, also varies as a function of the nature of excess substrates fed; protein has a greater effect than carbohydrate, followed by fat. As shown in a graphical form (Fig. 6) the relationship **in steady state conditions** between total energy intake below and above energy equilibrium (considered as energy needs) vs energy expenditure and energy balance is not quite linear, indicating an improved net efficiency of energy utilization below energy maintenance and a decreased one above energy equilibrium.

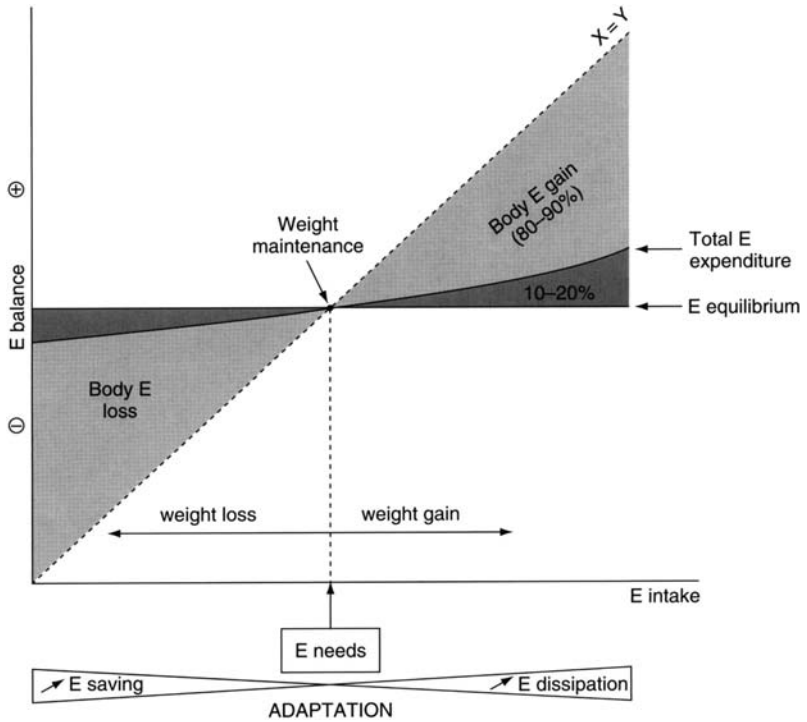


Fig. 6 Relationship between the level of energy intake on energy balance in steady state nutritional conditions, i.e., after adaptation: from underfeeding (below energy needs) to body weight maintenance when energy needs are fully covered, up to overfeeding situations (above weight maintenance) conditions. The metabolic adaptations occur in both sides of the energy imbalance, i.e., by increased energy saving in undernutrition and increased energy dissipation in overfeeding by stimulation of diet-induced thermogenesis. Note that the straight line (where x = y) would represent the unreal and unphysiological situations where E intake would be equal to Energy expenditure immediately no matter what is the level of E intake, without a long transient adaptation over the years, at the cost of body weight gain or body weight loss. E = energy.

Conclusions

The magnitude of day-to-day variability of energy expenditure in a given individual under free-living conditions is substantially lower (coefficient of variation CV of around 5–10%) than that of total energy intake (CV with much larger rang: 10–20%). This has consequences for the extent to which both variables can be controlled for regulating energy balance and substrate balance, and hence to maintain body weight.

Several studies have shown that the relationships between energy intake, energy expenditure vs body weight is not straightforward, since it depends upon several factors, the most important ones being the time frame under which it is considered (one day, one week, one month, or more) and the accuracy and precision with which energy intake and expenditure are measured in humans; energy intake being much more difficult to track than energy expenditure.

The metabolic energy used in the body, or energy expenditure, is classically assessed by indirect calorimetry, involving the measurement of oxygen consumption and carbon dioxide production. These validated systems, have been used as a golden standard for assessing energy needs/requirements of individuals and groups of well-defined subjects by FAO/WHO (1985), admitting implicitly that food energy intake failed to be accurate enough to assess energy needs in various populations.

References

Bray, G.A., Bouchard, C., 2020. The biology of human overfeeding: a systematic review. Obes. Rev. 21, e13040.

Food and Agriculture Organization of the United Nations, World Health Organization & United Nations University, 1985. Energy and Protein Requirements: Report of a Joint FAO/WHO/UNU Expert Consultation.

Galgani, J., Ravussin, E., 2008. Energy metabolism, fuel selection and body weight regulation. Int. J. Obes. 32 (Suppl. 7), S109–S119.

Schutz, Y., 1995. Macronutrients and energy balance in obesity. Metabolism 44 (Suppl. 3), 7–1.

Schutz, Y., 2004. Concept of fat balance in human obesity revisited with particular reference to de novo lipogenesis. Int. J. Obes. Relat. Metab. Disord. 28 (Suppl. 4), S3–S11.

Further reading

Dhurandhar, N.V., Schoeller, D., Brown, A.W., Heymsfield, S.B., et al., 2015. Energy balance measurement: when something is not better than nothing. Int. J. Obes. 39, 1109–1113.

Elia, M., Stratton, R., Stubbs, J., 2003. Techniques for the study of energy balance in man. Proc. Nutr. Soc. 62 (2), 529–537.

Fernández-Verdejo, R., Aguirre, C., Galgani, J.E., 2019. Issues in measuring and interpreting energy balance and its contribution to obesity. Curr. Obes. Rep. 8 (2), 88–97.

Schutz, Y., Byrne, N.M., Dulloo, A., Hills, A.P., 2014. Energy gap in the aetiology of body weight gain and obesity: a challenging concept with a complex evaluation and pitfalls. Obes. Facts 7 (1), 15–25.

Spiegelman, B.M., Flier, J.S., 2001. Obesity and the regulation of energy balance. Cell 104, 531–543.

Enteral and parenteral nutrition

Kent Aje, Sara Yacyshyn, Lindsay Dowhan, and Lisa Reitz, The Cleveland Clinic Foundation, Cleveland, OH, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of S. Devi Rampertab, A.K. Fischer, G.E. Mullin, Parenteral Nutrition, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 14–20, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00204-X.

Part I: Enteral nutrition **325**
Introduction 325
Indications and contraindications 325
Types 325
Access 326
Nasogastric route 326
Nasoduodenal or nasojejunal route 326
Percutaneous endoscopic gastrostomy (PEG), percutaneous jejunostomy (PEJ) or percutaneous endoscopic gastrostomy with jejunal extension tube (PEG-JET) 326
Formula selection 326
Modular formulas 327
Medication administration and tube occlusions 327
Methods of infusion 329
Complications and monitoring 329
Home enteral nutrition 329
Conclusion **329**
Part II: Parenteral nutrition **330**
Indications for parenteral nutrition 330
Perioperative support in severe malnutrition 330
Contraindications to parenteral nutrition 330
Vascular access 330
Nutrition components of parenteral nutrition 331
Amino acids 331
Dextrose 331
Lipid emulsions 331
Electrolytes 332
Vitamins 332
Trace elements 332
Titration of volume 332
Parenteral nutrition monitoring 333
Cyclic parenteral nutrition 333
Complications of parenteral nutrition 334
Catheter occlusion/thrombosis 334
Infection 334
Metabolic 334
Hepatic injury 334
Metabolic bone disease 335
Home parenteral nutrition 335
Conclusions **335**
References **335**
Further reading **336**

Key points

- Understand when Enteral Nutrition is appropriate
- Understand the benefits and complications of Enteral Nutrition
- Review the indications and components of parenteral nutrition
- Understand the complications associated with parenteral nutrition

Part I: Enteral nutrition

Introduction

Enteral nutrition (EN) is the modality of feeding via the gastrointestinal (GI) tract when one cannot sustain or progress their nutritional needs orally. Feeding is delivered directly into the GI tract for absorption of nutrients, fluid, electrolytes, and vitamins.

EN dates back a few thousand years, notably 100 BC when barley and wheat were provided via rectal feeds in patients who were not capable of eating orally (Vassilyadi et al., 2013; Dudrick and Palesty, 2011). It wasn't until 1598 that Capivacceus used a hollow tube placed in the esophagus to deliver liquid nutrients (Dudrick and Palesty, 2011) and Hunter in 1790 (Vassilyadi et al., 2013; Dudrick and Palesty, 2011) first used EN for therapeutic purposes to bypass the esophagus and feed directly into the stomach. The first documented case of gastrostomy tube placement was in 1876 by Verneuil (Vassilyadi et al., 2013) and early enteral feeding attempts and techniques kept evolving well into the twentieth century leading to the development and innovation of enteral nutrition that we know today. EN has been able to provide a vital avenue for patients to receive and meet their nutritional needs.

Indications and contraindications

EN is the preferred method of feeding patients who cannot eat, absorb, or use a normal diet in the presence of a useable GI tract. Indications for EN are listed in **Table 1** and should be initiated in the critically ill within 24–48 h.

EN can also be harmful if clinically used when the GI tract is not accessible, safe to use, functional or ethically don't align with a patient's goals of care. **Table 2** lists common contraindications to EN.

Types

EN requires liquid formulas to be administered through a tube in the upper GI tract. The type of feeding tube used for EN delivery can be divided into two categories: nasogastric or nasoenteral tubes, which enter the GI tract through the nose or gastrostomies and jejunostomies, which enter through the abdominal wall. In general, feedings administered through nasal tubes is indicated for

Table 1 Potential indications for enteral nutrition.

Inability to ingest food normally	
Prolonged NPO (nothing by mouth)	Multiple sclerosis
Persistent nausea	Amyotrophic lateral sclerosis
Early satiety	Guillain-Barre syndrome
Dysphagia, odynophagia	Huntington's chorea
Mucositis	Parkinson's disease
Chronic abdominal pain	Respiratory failure
Head and neck surgery	Fracture of mandible
Stroke	Irradiation of head and neck
Unconsciousness, stupor, coma	Chemotherapy
Brain tumor or injury	Oropharyngeal neoplasm
Gastrointestinal dysfunction	
Acute pancreatitis	
Chronic liver disease	
Crohn's disease	
Ulcerative colitis	
Diabetic gastroparesis	
Enteric fistula (if access is distal to fistula)	
Esophageal stricture or neoplasm	
Gastric outlet obstruction	
Psychiatric	
Anorexia nervosa	
Depression	
Medical	
Hypermetabolic state (trauma, sepsis, burns)	
Critical illness	
Renal failure	
Cancer or cardiac cachexia	
Chronic obstructive pulmonary disease	
Cystic fibrosis	
Hyperemesis gravidarum	

Adopted from Rath, M., McLaughlin, K., Murphy, R., 2020. Enteral nutrition. In: Nutrition Support Handbook. Cleveland Clinic, Cleveland OH, 60–86.

Table 2

Contraindications for enteral nutrition
Mechanical obstruction
Intractable vomiting or diarrhea that is refractory to medications
Extreme malabsorption
Short Bowel Syndrome (100 cm small bowel remaining)
Paralytic Ileus
High Output Fistula
Peritonitis
GI bleeding
Hemodynamic Instability
Inability to access GI tract
Terminal disease/Hospice

periods lasting less than 4 weeks whereas feedings that require a longer duration of time use more permanent tubes (gastrostomies and jejunostomies).

Access

Nasogastric route

Short-term feeding tubes should be used for patients who are expected to receive EN for less than 4 weeks or for whom a long-term feeding tube is not an option. Nasogastric tube placement, where a silicone or polyvinyl tube passes through the nose into the stomach, is one of the most common types of enteral access and can be used to safely administer short-term EN and or in patients who are at risk of aspiration. This procedure can be performed bedside. Nasogastric tubes vary in length from 30 to 43 in with diameters of 8–14 French. Bolus or continuous infusions can be used to administer feedings. Complications that may occur with short-term tubes include patient discomfort, sinusitis, or minor epistaxis. Both, nasogastric or nasoenteral (duodenal or jejunal), when placed with a nasal bridle to secure the tube in place decrease the risk of dislodgement and helps ensure adequate calorie delivery (Stabler et al., 2018; Seder et al., 2010).

Nasoduodenal or nasojejunal route

Nasojejunal tubes are longer tubes than their nasogastric counterparts and are also used for short-term feeding (4 weeks), in patients with gastric dysmotility, nausea or vomiting, gastroparesis, and high-risk patients. A nasojejunal tube is more difficult to insert and place, particularly if blind placement of the tube occurs, where complications of misplacement into the trachea or into the lung pleura could lead to a pneumothorax. An electromagnetic placement device allows for the visualization of the placement of small-bore feeding tubes using a transmitting stylet and a receiver placed on the patient's abdomen (Powers et al., 2018). Alternatively, tubes can be advanced via endoscopic or fluoroscopic guidance. Feedings via nasojejunal route require an infusion pump for administration.

Percutaneous endoscopic gastrostomy (PEG), percutaneous jejunostomy (PEJ) or percutaneous endoscopic gastrostomy with jejunal extension tube (PEG-JET)

PEG tubes are used in patients who will need long term, full, or supplemental feedings lasting more than 4 weeks. The PEG tube can be placed directly into the stomach through the abdominal wall using an endoscope. Tubes can then be advanced by endoscopic guidance into the duodenum and then brought out through the abdominal wall to provide the access route for enteral feedings. The advantages of PEG tube feeding are that they are easier to place than a jejunostomy, less expensive, and easier to maintain. Disadvantages are that it is contraindicated in patients with obstructions and there may be a moderate risk of aspiration in high-risk patients. The tube diameters commonly used range from 6 to 8 mm French units. In general, small diameter tubes should be avoided in patients with poor gastric emptying who require intragastric administration of medication. Percutaneous jejunostomy (PEJ) is also used in long-term feeding, in patients who require full or supplemental feedings for more than 4 weeks. PEJ placement requires higher skill and is more difficult to place than PEG, but can safely feed patients with gastric resection, gastric dysmotility, and patients with a higher risk for aspiration. Sometimes clinical scenarios arise when the need to both decompress the stomach while providing nutrition must occur. A combination of the two tubes via a percutaneous endoscopic gastrostomy with jejunal extension tube (PEG-JET or J-PEG) is performed either with fluoroscopy or endoscopy to achieve this (Ponsky et al., 1985; DeLegge et al., n.d.). A limitation in PEJ tubes is that the tubes are typically smaller in size and may clog more easily.

Formula selection

Although the composition of EN formulas vary widely, almost all formulas are nutritionally complete when administered in the appropriate quantities. Factors to consider when choosing a formula depend on the nutritional needs of the patient, the

functionality of the GI tract and desired metabolic response (Rath et al., 2020). Most formulas are gluten and lactose free, with some containing Halal and Kosher designations as well as non-GMO and vegan varieties. Water comprises the majority of EN with formulas ranging between 70 and 85% water depending on the caloric density. A wide selection of EN formulas are commercially available depending on the needs of the patient and can be classified into four separate groups:

- Standard or Polymeric formulas: These formulas contain intact proteins from milk and soy, oligosaccharides and lipid sources primarily from lecithin and oils which contain both long-chain triglycerides (LCT) and medium-chain triglycerides (MCT). Standard formulas vary in caloric density ranging 1.0 calorie (kCal) per milliliter (mL) to 2.0 kCal/mL. Protein provision can range between 15 and 25% and is most often the deciding factor on formula selection. Standard formulas are typically well tolerated across many patient populations, including the medical and surgical intensive care units (ICU's) (McClave et al., 2016).
- Elemental and Semi-Elemental formulas: These formulas require less digestion than standard formulas as their components have been enzymatically hydrolyzed. They contain peptides and free amino acids with lower osmolality. They are often used for patients that have malabsorptive conditions and impaired GI function, which may include short bowel syndrome, Crohn's and/or pancreatitis. The protein in these formulas is in the form of peptides and free amino acids, while fat is provided as a combination of LCT and MCT. Carbohydrates are partially hydrolyzed starch maltodextrins. Formulas may provide 1–1.5 kCal/mL and protein can range 18–38% of total kCal.
- Disease specific formulas: These formulas are designed to provide specific nutrient needs based on specific illnesses. They are intended for patients who have specific metabolic requirements and may be enriched with specialized nutrients to improve patient outcomes (See **Table 3**).
- Blenderized: These formulas are defined as blended foods and liquids provided via feeding tube (Escuro, 2014) and can be commercially made or blended in the hospital and/or home environment. Blenderized foods may date back to 1949 when purified amino acids, sucrose, corn oil, starch, vitamin, and minerals were fed to study participants (Dudrick and Palesty, 2011) and has gained popularity as of recent to a subset of patient/families interested in real-food ingredients. Risks associated with its use include bacterial contamination and inconsistencies with nutrient content (Milton et al., 2020). Hang time for administration should be limited to 2 h or less to prevent bacterial contamination (Boulatta et al., 2017).

Modular formulas

Modular formulas are used for supplemental use and are designed to provide additional calories, protein, or fiber to meet individual nutrient needs. Modulars used to provide additional kCal contain both fat and protein (typically from vegetable oils and calcium caseinate respectively) and can provide upward of 300 kCal per serving. Protein modulars come in both powder and liquid formulations and can provide between 6 and 10 g protein per serving from whey protein isolates or hydrolyzed collagen. Fiber modulars are also available and can be used to enhance GI function by providing 3 g soluble fiber per serving from partially hydrolyzed guar gum.

Soluble and Insoluble fibers: Tube feeding formulas can be fiber-free, contain a mix of both insoluble and soluble fiber, or contain soluble fibers alone depending on the medical needs of the patient. Patients may benefit from a mixed fiber formula or soluble fiber additive for management of diarrhea when other sources of diarrhea have been ruled out (McClave et al., 2016). Soluble fiber helps to decrease diarrhea through increased absorption of sodium and water (Rath et al., 2020) and has been shown to maintain healthy microbiota in the bowel (McClave et al., 2016). Insoluble fibers are used to provide bulk to stool and promote intestinal transit (Rath et al., 2020). All types of fiber should be avoided in the critically ill with severe dysmotility or those at increased risk for bowel ischemia (McClave et al., 2016).

Medication administration and tube occlusions

Enteral access devices can be used to provide medications to patients, but complications may occur because many medications are not formulated for feeding tube administration. The goal of medication administration is to both reduce drug toxicity and ensure therapeutic levels are being delivered to the patient (Boullata et al., 2017). Gastric administration of medication is ideal as post-pyloric administration can alter a drug's bioavailability (Boullata et al., 2017). The safest method of medication delivery is by mixing crushed compatible solid tablets with 30–60 mL of water and diluting liquid medications to reduce its viscosity (Boullata et al., 2017).

Enteral tube occlusion is a common problem that can sometimes be resolved but may require replacement of the tube. Medications are a common cause. Water flushes should be provided before and after medications are given, every 4 h during continuous feedings, and before and after each bolus feeding (Rath et al., 2020). A single-center survey in the UK found that there were significant practice variations among nurses when administering medications via enteral access devices. Theoretically, this may have contributed to blocked tubes and excessive fluid administration to some patients (Tillott et al., 2020). An interdisciplinary team of licensed independent practitioners, nurses and pharmacists are important in preventing errors and developing protocols for medication administration via enteral access devices (Boullata et al., 2017).

Table 3 Disease-specific formulas.

Type of formula	*Description of use*
Diabetic	Lower carbohydrate, higher fat formula with fiber to help with glycemic control Can use standard formula as there is limited evidence to support its use in adult patients (McMahon et al., 2013) Avoid overfeeding to prevent hyperglycemia
Renal	Fluid restricted, energy dense formula that contains less potassium, sodium, phosphorus and magnesium for patients not on dialysis. Protein content varies depending on need. If patient on frequent hemodialysis or CRRT, recommend increased protein up to 2.5 g/kg/day
Pulmonary/ARDS	High fat, low carbohydrate formula in order to reduce CO2 production and respiratory quotient, however there is limited data to support use. Fluid restriction, energy dense formulas are recommended for patients with acute respiratory failure Avoid overfeeding kCal especially when weaning from ventilator
Hepatic	Hepatic formulas no longer available. Standard formulas are recommended as there is no evidence for use of branched –chain amino acids in this population. Fluid restricted, lower protein, energy dense formula with high ratio of branched-chain amino acids to aromatic amino acids. High MCT to LCT ratio. Protein restriction has not been shown to reduce hepatic encephalopathy and may lead to malnutrition and decreased muscle mass
Immune enhancing	These formulas contain specialized nutrients to improve immune function, such as arginine, glutamine, nucleic acids and omega-3 fatty acids. There appears to be a synergistic effect with arginine and omega-3 fatty acids Recommended in pre- and post-operative phase for patients with malnutrition; or pre-operatively in well-nourished patients undergoing elective surgeries. May also be used with traumatic brain injury as it may mitigate infections and improve recovery (McClave et al., 2016). Limited studies have shown efficacy with its use in trauma. Not used in severe sepsis
Plant/food based	Contains whole foods for patients who prefer real food sources, vegan options, and Non-GMO ingredients. Can meet 100% nutritional needs with specific volume. Caution with food allergies, as some products contain allergenic foods (i.e., milk, fish, gluten, egg, soy, tree nuts, wheat)

kCal—Calorie CRRT—Continuous renal replacement therapy.

Adopted from: McClave, S.A., Taylor, B.E., Martindale, R.G., et al., 2016. Guidelines for the provision and assessment of nutrition support therapy in the adult critically Ill patient: Society of Critical Care Medicine (SCCM) and American Society for Parenteral and Enteral Nutrition (A.S.P.E.N.). J. Parenter. Enteral Nutr. 40(2), 159–211.; McMahon, M., Nystrom, E., Braunschweig, C., Miles, J., Comper, C., 2013. American Society for Parenteral and Enteral Nutrition (A.S.P.E.N.) clinical guidelines: nutrition support of adult patients with hyperglycemia. J. Parenter. Enteral Nutr. 37, 23–26.

Methods of infusion

Enteral feedings can be administered via pump assist, which includes a continuous drip, bolus feedings or intermittent infusions; or can be provided using gravity drip or manual bolus (syringe). The method selected depends on the stability of the patient, EN access route, GI anatomy and other medical factors.

- Pump assisted:
  - Continuous drip: EN is administered at a designed rate over a 24 h period. This method is used in hospitals when initiating tube feeding and is well tolerated. Initiation of tube feeds typically start at 10–15 mL/h and increases every 4–8 h depending on EN tolerance until patient's goal rate is met. This method is used with both gastric and small bowel tube feedings.
  - Bolus infusion: Deliver a higher volume of nutrients provided multiple times daily to mimic normal eating patterns. Bolus infusions should be used for gastric tube feeding only.
  - Intermittent infusions: Larger volumes of EN are infused over a designated time period. This type of feeding may result in more adequate volumes being administered and will also provide time off infusion. This method is used with cyclic and nocturnal feeding regimens.
- Gravity feed: Method of administration using a feeding bag with a roller clamp attached to the tubing to adjust the flow to allow more intermittent delivery of nutrients. It is often used in gastric tube feeding patients. This mode of infusion is less expensive than pump-assisted delivery and is also easier to administer (Rath et al., 2020).
- Manual bolus (syringe): Larger volume of EN is delivered multiple times daily via syringe bolus. This method allows the patient more freedom, is easy to use, and is the cheapest route of administration as it does not require a pump. This is a common method to administer tube feeding at home and long-term care facilities. Rapid installation of feeding into the stomach is given by syringe and most patients tolerate this method. Bolus feeding must be provided when the patient is in an upright position to avoid aspiration.

Rate based infusion is the method of providing an EN infusion over a constant hourly rate. Historically, this was a common approach to EN delivery, however volume-based infusion has shown efficacy in providing more nutrient delivery as a prescribed amount of EN is delivered over a 24 h period to help ensure the patient will receive adequate nutrition (Boulatta et al., 2017). This method helps to make up for any missed feedings that may occur during the day in a hospitalized patient undergoing various test or procedures. Studies have shown its safety and improvement in nutrient delivery when compared to rate-based infusion techniques (McClave et al., 2015).

Complications and monitoring

EN is generally safer than Parenteral Nutrition (PN) but it still has the potential to induce GI, metabolic and mechanical complications. Diarrhea is the most common adverse event associated with tube feedings, but it is rarely due to the formula of the tube feeding itself. Etiologies such as underlying disease, enteric pathogens or sorbitol-containing medications contribute to diarrhea symptoms (Rath et al., 2020).

An adverse event to monitor for while patients are on EN is aspiration. Traditionally, measurements of Gastric Residual Volumes (GRV) were used as an indicator for aspiration risk (Boullata et al., 2017). Elevated and increasing residual volumes may be a symptom of another underlying problem manifesting itself as delayed gastric emptying (Boullata et al., 2017). Alternative methods for monitoring for aspiration include careful daily physical examinations, review of abdominal radiologic studies, and evaluation of clinical risk factors for aspiration. Current ASPEN guidelines recommend against stopping EN for GRVs <500 mL in the absence of other signs of intolerance (Boullata et al., 2017).

Home enteral nutrition

Transitioning patients to home with EN is a common practice in the United States. To date, there is an estimated $>400{,}000$ patients receiving home enteral nutrition (Mundi et al., 2017). A multidisciplinary approach to patient management is key in setting up the patient for success. The team needs to ensure the patient is tolerating the EN formula and that adequate education to either the patient and/or other caregivers who will be providing support for the patient at home. Social work or Case management is involved in assessing the safety of the home environment prior to hospital discharge (Boulatta et al., 2017). Coverage for EN depends on individual insurance coverage, however Medicare and Medicaid combined make up the largest insurers to cover the cost of EN (Bonnes et al., 2017). Documentation for payment coverage must include specific parameters a patient must meet, starting with a permanent condition expecting to last >3 months, limiting the patient's ability to meet their estimated nutritional needs.

Conclusion

EN is viewed as the safest and most efficacious method to support nutritional status in patients who are unable to eat orally. If managed properly, enteral feeding is associated with improved clinical outcomes and reduced infectious complications and is the preferred form of nutritional support.

Part II: Parenteral nutrition

Indications for parenteral nutrition

The use of parenteral nutrition (PN) can be considered in those situations where the GI tract is inaccessible or unsafe for enteral nutrition, a non-functional GI tract or during a period of required bowel rest (ASPEN, 2009). PN is also appropriate when it is anticipated that enteral feedings cannot begin for at least 7 days. However, PN should not be initiated if it is expected to last less than 5–7 days because risk may outweigh potential benefit of this therapy. Enteral nutrition has many advantages over parenteral nutrition and should be considered whenever possible. Table 4 lists indications for parenteral nutrition. In general, the gastrointestinal tract should not be used in these conditions until the underlying problem is treated.

Perioperative support in severe malnutrition

Malnutrition has been found to be associated with an increased risk of morbidity and mortality in the postoperative period. The Veteran's Affairs Parenteral Nutrition Cooperative Trial evaluated the benefits of preoperative parenteral nutrition in patients with varying degrees of malnutrition (Veterans Affairs Total Parenteral Nutrition Cooperative Study Group, 1991). Of note, an increased rate of infectious complications was seen in mildly-to-moderately malnourished patients receiving parenteral nutrition. If enteral feeding is feasible and tolerated, preoperative enteral nutrition has been found to be equally effective when used for 7–14 days or longer in malnourished surgical patients. In those postoperative patients who have developed a paralytic ileus from various causes, parenteral nutrition may be attempted if there is no return of gastrointestinal function for 7–10 days.

Contraindications to parenteral nutrition

The main contraindication to PN includes a functional gastrointestinal tract. Other possible instances where PN would not be recommended would be when a patient is do not resuscitate (DNR) status, or in those cases where patients have unstable fluid, electrolyte, or cardiopulmonary status.

Vascular access

PN can be infused either into the venous system centrally or peripherally. Central or peripheral access is determined by the position of the distal catheter tip. When the tip is located outside a large-caliber central vein such as the superior vena cava (SVC) or the inferior vena cava (IVC), it is considered a peripheral access device. Lower concentration of dextrose and amino acids may be given through peripheral veins (peripheral parenteral nutrition or PPN) for a short duration of therapy (less than 10–14 days). Such formulas usually do not provide the patient's full nutritional needs and may require large volumes of fluid and therefore should not be used in patients where fluid restriction is necessary (i.e., cardiac patients and renal patients). Osmolality of peripheral

Table 4 Indications for parenteral nutrition.

GI tract is unsafe for enteral nutrition
• Preoperative nutrition optimization for severely malnourished patients unable to meet needs through enteral nutrition
• Severe persistent GI bleeding
• Inability to maintain or obtain enteral access
GI tract is non-functional
• Short bowel syndrome or severe malabsorption
• Bowel obstruction or prolonged ileus
• High output enterocutaneous fistula (>500 mL/d)
• Intestinal dysmotility
• Intestinal ischemia
• Intractable vomiting or diarrhea
• Radiation enteritis
Bowel rest needed
• Acute exacerbations of inflammatory bowel disease
• Graft-versus-host disease of the gut
• Severe acute pancreatitis (>7–10 days)
• Severe persistent enteritis
• Persistent chyle leak despite very low fat enteral nutrition
• Acute exacerbations of inflammatory bowel disease

Adopted from Ratliff, A., Jaroch, L., Nishnick, A., 2020. Parenteral nutrition. In: DeChicco, R.S., et al., Nutrition Support Handbook, fourth ed. The Cleveland Clinic Foundation, 88–100.

solutions needs to be less than 900 mOsm l^{-1}. Potential complications of PPN include phlebitis, infiltration, or fluid-overload issues. Frequent peripheral IV site rotations are required (usually every 48–72 h). Formulas that have higher concentrations of nutrients, thereby making them hyperosmolar, must be administered directly into the SVC or IVC to allow for rapid dilution. Some of the central catheters commonly used to deliver PN include subclavian vein catheters, peripherally inserted central catheters, subcutaneously tunneled percutaneous catheters, or implanted subcutaneous infusion ports. These delivery systems are preferred to peripheral options because they allow for maximal amounts of calories and protein per volume to be provided.

Nutrition components of parenteral nutrition

PN is a compounded formulation, which includes protein as amino acids, energy in the form of dextrose and fats, as well as electrolytes, vitamins, minerals, and trace elements. Sterile water is added to provide necessary volume to the formula. Parenteral solutions are divided into two main categories: 3-in-1 or 2-in-1 solutions. A 3-in-1 solution, referred to as Total Nutrient Admixture, is composed of amino acids, dextrose, and lipids all combined in one bag. A 2-in-1 solution contains the mixture of amino acids and dextrose in one bag and a separate intravenous fat emulsion infusion.

Amino acids

Amino acids yield 4 kcal g^{-1} when oxidized for energy. Amino acid solutions range from 3 to 20% concentrations. Standard amino acid solutions are a combination of essential and nonessential amino acids. Some standard amino acid solutions also contain an appreciable amount of phosphorus which should be considered when dosing electrolytes. Disease specific amino acid solutions for renal or hepatic failure are available commercially, however, evidence does not support improved outcomes with these products compared to standard solutions. Ascertaining goal protein requirements is vital in terms of maintaining lean body mass and promoting positive nitrogen balance. Degree of suspected catabolism, renal function, and hepatic function all play an important role in determining this. **Table 5** provides a listing of protein needs estimates based on clinical condition. Reductions in protein doses may be required in cases of hepatic encephalopathy complicating hepatic failure. Likewise, renal insufficiency may also be a situation where lower amounts of protein can be used depending on the severity of renal failure and whether dialysis is indicated. IBW should be utilized in morbidly obese individuals to estimate protein needs.

Dextrose

Carbohydrate is the primary energy source for the human body, especially the brain. Carbohydrate calories come in the form of dextrose, which provides 3.4 kcal g^{-1} of energy when metabolized. Excessive carbohydrate administration is associated with increased infections, hepatic steatosis, hyperglycemia and increased carbon dioxide production. To minimize these complications, the maximum parenteral carbohydrate infusion rate should be 7 mg/kg/minute in hospitalized patients and 4–5 mg/kg/minute in the critically ill patient (McClave et al., 2016) (**Table 6**).

Lipid emulsions

Parenteral nutrition formulas contain lipid emulsions as a source of fat calories that prevents essential fatty acid deficiency (EFAD). In US, the lipid formulations are mostly composed of n-6 long-chain fatty acids derived from vegetable oils. Intravenous fat emulsion provides 10 kcal g^{-1} of energy due to the addition of the glycerol molecule. Lipids are contraindicated in patients with significant hypertriglyceridemia. It is recommended that intravenous fat emulsion be held or limited from the parenteral nutrition regimen if serum triglyceride concentration exceeds 400 mg/dL. Withholding lipid emulsions greater than 2 weeks is not advised due to increased risk of essential fatty acid deficiency (Hamilton et al., 2006). Since ILE contains egg yolk phospholipids, it is contraindicated in the rare instance of documented egg allergy. In this case, safflower or sunflower oil can be administered topically or enterally (Miller et al., 1987). Two polyunsaturated fatty acids, linoleic and a-linolenic, cannot be synthesized by the body and are therefore considered essential in that they have to be brought into the body via diet. Thus, to prevent essential fatty acid deficiency, 1–2% of daily energy requirements should be derived from linoleic acid and approximately 0.5% of energy from linolenic

Table 5 Estimation of protein needs.

BMI	*g protein/kg ABW*
<25	1.5–2
25–29.9	1.3–1.7
30–34.9	1.2–1.6
≥35	1.1–1.5

Adopted from Nutrition requirements. In: Coughlin, K. (Ed.), Cleveland Clinic Nutrition Support Handbook. 35–38.

Table 6 Determination of energy needs.

Condition	*Need (kcal kg^{-1})*
Overnourished/obese	20 (upper end IBW)
Maintenance	25
Undernourished	30
Stressed/critically ill	25

acid. Clinical manifestations of essential fatty acid deficiency are important to recognize and include scaly dermatitis, dermatitis, alopecia, hepatomegaly, thrombocytopenia, fatty liver, and anemia.

Electrolytes

Electrolytes are added as salts to PN solutions depending on the patient's requirements for daily maintenance and to replace losses (Table 7). Sodium and potassium can be added as chloride or acetate salts, depending on acid–base needs. In general, acetate and chloride content of the solution should be adjusted carefully because they help to maintain the acid–base balance. Also, calcium and phosphorus concentrations must be watched closely to prevent precipitation. Many factors may influence the solubility of these electrolytes in the PN solution, including the concentration of electrolytes, the pH of the final formula, temperature, and the presence of other components. Fluid and electrolyte status can be maintained utilizing standard intravenous (IV) fluids along with additional oral or IV replacements.

Vitamins

Multivitamin preparations, including both water- and fat-soluble vitamins, are available for inclusion in the PN formulation. Available parenteral multivitamin products for adults contain 12 or 13 known vitamins (with or without vitamin K). Table 8 lists the composition of standard adult multivitamin products. Iron is not usually added to PN solutions because it can result in destabilization of the lipid emulsion. However, addition of iron in the form of iron dextran can be administered with 2-in-1 solution. A test dose of the iron should be given separately initially to ensure that there is no adverse reaction before it is incorporated routinely into a 2-in-1 bag.

Trace elements

Commonly used trace elements in PN formulas include zinc, selenium, manganese, copper, and chromium. Intravenous trace element preparations are available as single component products as well as a variety of combination products. Table 9 provides a list of the doses of trace elements in a common combination product. The commercially available multiple trace element combinations in the US contain three-to five-times the recommended doses for manganese, and this is important to keep in mind in those patients on long-term PN.

Titration of volume

Final volume can be titrated by the addition of sterile water for injection. Concentrated substrates may be used to minimize volume in those situations where patients need to be volume restricted. Typical PN volumes range from 800 to 2500 mL daily.

Table 7 Daily electrolyte requirements for patients receiving nutrition support.

Electrolyte	*Parenteral*
Sodium	1–2 mEq/kg
Potassium	1–2 mEq/kg
Chloride	As needed to maintain acid-base balance
Acetate	As needed to maintain acid-base balance
Calcium	10–15 mEq
Magnesium	8–20 mEq
Phosphorus	20–40 mmol

Note- General ranges for safe administration in healthy people with normal losses.
Adopted from Krishnan, K., Meyer, S., Wolford, M., 2020. Parenteral nutrition. In: DeChicco, R.S., et al., Nutrition Support Handbook, fourth ed. The Cleveland Clinic Foundation, 88–100.

Table 8 Contents of parenteral multivitamin preparations.

Vitamin component	*Current FDA requirement (2020)*
Vitamin A	900 mcg RAE
Vitamin D	20 mcg
Vitamin E	15 mg alpha- tocopherol
Vitamin K	120 mcg
Vitamin C	90 mg
Folate	400 mcg DFE
Niacin	16 mg NE
Vitamin B_2	1.3 mg
Vitamin B_1	1.2 mg
Vitamin B_6	1.7 mg
Vitamin B_{12}	2.4 mcg
Pantothenic acid	5 mg
Biotin	30 mcg

Note- General ranges for safe administration in healthy people with normal losses. Needs for pregnant or lactating women may be higher.

Table 9 Contents of a common parenteral trace element preparation (Tralement®).

Component	*Dose*
Zinc	3 mg
Copper	0.3 mg
Manganese	55 mcg
Chromium	0 mcg
Selenium	60 mcg

Parenteral nutrition monitoring

Careful evaluation and monitoring of patients is essential to provide optimal nutritional therapy while avoiding complications. Changes in patient clinical condition necessitate frequent reassessment. The possibility of enteral feeding should continue to be readdressed. Moreover, as parenteral nutrition is transitioned to enteral intake, tolerance should be closely monitored, allowing for weaning and eventual discontinuation. Initial laboratory measurements should include liver chemistries, electrolyte levels, as well as renal profile. Electrolytes including calcium, magnesium, and phosphorus should be monitored daily (Misra and Kirby, 2000). Volume status is also an important parameter and intake and output should be continuously assessed. Physical exam signs looking for evidence of volume overload include weight gain, increasing swelling of the extremities, ascites, and pulmonary congestion. Likewise, tachycardia, low blood pressure, poor skin turgor, and dry mucous membranes may indicate hypovolemia. Increased losses such as urine, diarrhea, nasogastric output, emesis, fistula, and drains should also be considered. Blood glucose levels should be monitored to follow any hyper or hypoglycemia. Coverage with subcutaneous sliding-scale insulin is frequently utilized to manage hyperglycemia; however, separate intravenous insulin infusion or the addition of insulin as a component of the total parenteral nutrition (TPN) could also be done.

Cyclic parenteral nutrition

When initiating PN in an acutely ill patient, continuous infusion over 24 h is started. Once the patient has demonstrated stability on a given formula by way of physical exam, clinical parameters such as weight, volume status, and ins/outs as well as labs such as electrolyte balance and nutritional parameters, the patient can then be cycled to have the same volume of solution run over a shorter time frame (for example 12 h at night). Infusion pumps can aid in programming desired volumes and administration times. Although cycling does allow the patient to have more flexibility and therefore lead to a more active lifestyle, limitations include fluid or glucose intolerance. Abrupt cessation of the infusion may be associated with rebound hypoglycemia due to circulating insulin; therefore, the rate of administration is often tapered down at the end of the cycle to allow for downregulation of pancreatic release of insulin.

Complications of parenteral nutrition

Catheter occlusion/thrombosis

Catheter occlusion is the most common noninfectious complication associated with long-term venous access. Symptoms of a catheter-related venous thrombosis include vein distension, edema, tingling, or pain over the ipsilateral arm and neck, and a prominent venous pattern over the anterior chest. It may require thrombolytic treatment or line replacement. Parenteral formulas with inappropriate calcium/phosphate ratios favoring precipitation as well as lipid residue collecting within the catheter lumen can also create occlusion. Catheter flushing protocols should be followed closely to prevent the risk of occlusion. Kinking of the catheter tubing or angulation of the catheter can also contribute. Pinch-Off Syndrome refers to the compression of a subclavian vein central venous catheter (CVC) between the clavicle and the first rib resulting in an intermittent or permanent catheter obstruction. Findings are confirmed by venogram and/or chest radiograph and requires catheter replacement.

Infection

Catheter-related bloodstream infections (CRBSI) are another serious, and common complication of PN. The CDC recently estimated 250,000 cases of CRBSI annually in the US. Moreover, there was an average of 5 cases per 1000 catheter days and mortality ranged between 12% and 25% for these infections. Infection with fungal pathogens has an extremely high rate of mortality, ranging from 30% to 60%. Fever and unexplained hyperglycemia can herald the presence of a catheter-related infection. When a CRBSI is suspected, blood cultures should be obtained from each lumen of the catheter as well as the peripheral blood before antibiotics are started. The most common infective micro-organisms are coagulase negative staphylococci, *Staphylococcus aureus*, Klebsiella, Candida, and Enterobacter (Mermel et al., 2009). Treatment of CRBSI typically include the removal of the catheter as well as antibiotic or antifungal therapy. In those patients on long-term home PN with limited venous access, catheter salvage is certainly important and can be considered with the help from an infectious disease specialist. PN should be held if an infection is suspected to prevent worsening bacteremia due to infusion of hypertonic dextrose and other nutrients. A 5% dextrose solution can be given instead through a peripheral line until the infection has been adequately treated.

Metabolic

Metabolic complications are seen in 5–10% of patients on PN. Hyperglycemia is the most common complication and is typically related to the proportion of dextrose in the formulation, but other contributing factors include stress due to acute illness or sepsis, steroid use, postoperative period, diabetes, and pancreatitis. Excess sugar through PN can lead to hepatic steatosis, increased carbon dioxide production, and adverse outcomes. Blood glucose levels should be monitored very closely and a sliding-scale coverage of regular insulin can be used for elevated levels. Two-thirds of the total amount of sliding-scale insulin required over 24 h can be added to the next day's parenteral nutrition formulation. However, hypoglycemia can result from the abrupt cessation of parenteral nutrition. To decrease the risk of rebound hypoglycemia, a 1–2 h taper down of the rate of parenteral nutrition may be required. Alternatively, if the infusion is abruptly stopped, a 5% or 10% dextrose infusion should be utilized for at least an hour once the parenteral nutrition infusion is discontinued to avoid a possible rebound hypoglycemia. To avert hypertriglyceridemia, intravenous fatty emulsion intake should be limited to less than 30% of total calories or 1 g/kg/d. Hypertriglyceridemia may increase the risk of pancreatitis, particularly in those with triglyceride levels greater than 1000 mg dL^{-1}. It is generally recommended that lipid infusion be discontinued when serum triglyceride concentration exceeds 400 mg $dl1^{-1}$. Electrolyte disturbances also commonly occur and warrant close clinical evaluation. Lower concentrations or even elimination of selected electrolytes from parenteral nutrition are indicated in patients with renal failure. Refeeding syndrome is another metabolic complication seen in malnourished patients (Solomon and Kirby, 1990). The rapid delivery of calories in the form of carbohydrates stimulate insulin secretion, which causes an intracellular shift of phosphorus, magnesium, and potassium. Clinically, the patient develops symptoms of generalized fatigue, lethargy, muscle weakness, edema, cardiac arrhythmia, and hemolysis. Prevention of this phenomenon is achieved by identifying individuals at risk, repleting electrolytes before PN initiation, and slow caloric advancement of PN with daily monitoring of electrolytes.

Hepatic injury

There are three types of hepatobiliary conditions that are associated with PN. Steatosis is a benign disorder that is seen in the setting of overfeeding and characterized by fat accumulation in the liver. Most patients are asymptomatic and usually transaminases are mildly elevated. Progression to fibrosis and subsequently cirrhosis may rarely occur in long-term PN. Parenteral-associated cholestasis is characterized by impaired bile secretion or biliary obstruction and presents as an elevation of alkaline phosphatase, g-glutamyl transpeptidase, and conjugated (direct) hyperbilirubinemia (Guglielmi et al., 2008). Lastly, lack of enteral feeding that results in decreased cholecystokinin release, which leads to impaired bile flow and gallbladder contractility, causes the development of gallstones and gallbladder sludge with subsequent cholecystitis. The duration of PN therapy appears to correlate with the development of biliary sludge. When a patient on PN develops hepatic complications, a trial of decreasing calories to avoid overfeeding should be considered. A balanced carbohydrate:fat ratio needs to be obtained with approximately 70–85% of non-protein calories supplied as carbohydrates and 15–30% as fat. In addition, carbohydrate content should not exceed 7 g/kg/day in adults. Infusing cyclic PN rather than as a 24 h infusion may also reduce the risk of liver injury. Early transition to enteral or oral feedings should be optimized, as small amounts of intake would promote enterohepatic circulation of bile acids.

Metabolic bone disease

There is a high prevalence of metabolic bone disease in patients on long-term PN which contributes to an increase in morbidity due to fracture risk. Metabolic bone disease may develop due to underlying disease, inadequate intake of calcium particularly because the amount that can be included into the parenteral nutrition formulation is physically limited due to risk of precipitation and hypercalciuria. Higher protein levels in parenteral formula further enhance hypercalciuria, therefore reducing protein intake to maintenance doses (0.8–1 g/kg/d) is recommended (Seidner, 2002). Both vitamin D deficiency and excess can be contributing factors. Calcium balance and phosphorus levels also play a critical role in the development of metabolic bone disease. For selected long-term PN patients, treatment with bisphosphonates along with calcium and vitamin D supplementation should be considered to prevent complications. An increase in weight-bearing activity and avoiding corticosteroids may also be useful. Careful attention to magnesium, phosphorus, and calcium concentrations in the PN solution and establishing a good acid–base balance is also critical to prevent metabolic bone disease.

Home parenteral nutrition

Home PN is a viable option when it has been determined that a hospitalized patient on PN is not able to advance to enteral or oral feedings and is otherwise stable. The patient and/or caregiver should be taught to administer PN safely utilizing sterile technique, in addition to actively watching for signs of infection, fluid imbalances, or other complications. Cycling PN to 12 h/day provides increased mobility and flexibility and helps minimize hepatobiliary complications.

Conclusions

Parenteral nutrition can be a lifesaving therapy for individuals who are unable to utilize their gastrointestinal tract for a prolonged time. Patients and caregivers should be appropriately evaluated and educated to help prevent complications associated with PN. Management of PN is most effective when done through a multidisciplinary approach, utilizing the expertise of physicians, pharmacists, dietitians, nurses, case managers, and social workers.

References

American Society for Parenteral and Enteral Nutrition Board of Directors, 2009. Clinical guidelines for the use of parenteral and enteral nutrition in adult and pediatric patients. J. Parenter. Enteral Nutr. 33, 255-25.

Bonnes, S.L., Salonen, B.R., Hurt, R.T., McMahon, M.T., Mundi, M.S., December 2017. Parenteral and enteral nutrition—from hospital to home: will it be covered? Nutr. Clin. Pract. 32 (6), 730–738.

Boulatta, J.I., Carrera, A.L., Harvey, L., et al., 2017. ASPEN safe practices for enteral nutrition therapy. J. Parenter. Enteral Nutr. 41 (1), 15–103.

Dudrick, S.J., Palesty, J.A., 2011. Historical highlights of the development of enteral nutrition. Surg. Clin. N. Am. 91, 945–964.

Escuro, A.A., 2014. Blenderized tube feedings: suggested guidelines to clinicians. Practical Gastroenterol. 38, 58–66.

Guglielmi, F.W., Regano, N., Mazzuoli, S., et al., 2008. Cholestasis induced by total parenteral nutrition. Clin. Liver Dis. 12 (1), 97–110 viii.

Hamilton, C., Austin, T., Seidner, D., 2006. Essential fatty acid deficiency in human adults during parenteral nutrition. Nutr. Clin. Pract. 21, 387–394.

McClave, S.A., Saad, M.A., Esterle, M., et al., 2015. Volume-based feeding in the critically ill patient. J. Parenter. Enteral Nutr. 39 (6), 707–712.

McClave, S.A., Taylor, B.E., Martindale, R.G., et al., 2016. Guidelines for the provision and assessment of nutrition support therapy in the adult critically ill patient: Society of Critical Care Medicine (SCCM) and American Society for Parenteral and Enteral Nutrition (A.S.P.E.N.). J. Parenter. Enteral Nutr. 40 (2), 159–211.

McMahon, M., Nystrom, E., Braunschweig, C., Miles, J., Comper, C., 2013. American Society for Parenteral and Enteral Nutrition (A.S.P.E.N.) clinical guidelines: nutrition support of adult patients with hyperglycemia. J. Parenter. Enteral Nutr. 37, 23–26.

Mermel, L.A., Allon, M., Bouza, E., et al., 2009. Clinical practice guidelines for the diagnosis and management of intravascular catheter-related infection: 2009 Update by the Infectious Diseases Society of America. Clin. Infect. Dis. 49 (1), 1–45.

Miller, D.G., Williams, S.K., Palombo, J.D., et al., 1987. Cutaneous application of safflower oil in preventing essential fatty acid deficiency in patients on home parenteral nutrition. Am. J. Clin. Nutr. 46, 419–423.

Milton, D.L., Johnson, T.W., Johnson, K., Murphy, B., Carter, H., Hurt, R.T., et al., 2020. Accepted safe food handling procedures minimizes microbial contamination of home-prepared blenderized tube feeding. Nutr. Clin. Pract. 35 (3), 479–486.

Misra, S., Kirby, D.F., 2000. Micronutrient and trace element monitoring in adult nutrition support. Nutr. Clin. Pract. 15, 120–126.

Mundi, M.S., Pattison, A., McMahon, M.T., Davidson, J., Hurt, R.T., 2017. Prevalence of home parenteral and enteral nutrition in the United States. Nutr. Clin. Pract. 32 (6), 799–805.

Ponsky, J.L., Gauderer, M.W., Stellato, T.A., Aszodi, A., 1985. Percutaneous approaches to enteral alimentation. Am. J. Surg. 149 (1), 102–105. https://doi.org/10.1016/s0002-9610(85)80017-9.

Powers, J., Luebbehusen, M., Aguirre, L., Cluff, J., David, M.A., Holly, V., Linford, L., Park, N., Brunelle, R., 2018. Improved safety and efficacy of small-bore feeding tube confirmation using an electromagnetic placement device. Nutr. Clin. Pract. 33 (2), 268–273. https://doi.org/10.1002/ncp.10062.

Rath, M., McLaughlin, K., Murphy, R., 2020. Enteral nutrition. In: Nutrition Support Handbook. Cleveland Clinic, Cleveland OH, pp. 60–86.

Seder, C.W., Stockdale, W., Hale, L., Janczyk, R.J., 2010. Nasal bridling decreases feeding tube dislodgment and may increase caloric intake in the surgical intensive care unit: a randomized, controlled trial. Crit. Care Med. 38 (3), 797–801. https://doi.org/10.1097/CCM.0b013e3181c311f8.

Seidner, D.L., 2002. Parenteral nutrition-associated metabolic bone disease. J. Parenter. Enteral Nutr. 26 (Suppl. 5), S37–S42.

Solomon, S.M., Kirby, D.F., 1990. The refeeding syndrome: a review. J. Parenter. Enteral Nutr. 14, 90–97.

Stabler, S.N., Ku, J., Brooks, L., Gellatly, R., Haljan, G., 2018. Implementation of a nasogastric tube securement device in a tertiary care intensive care unit. Can. J. Crit. Care Nurs. 29 (1).

Tillott, H., Barrett, D., Ruan, J., Li, V., Merrick, S., Steed, H., Morrissey, H., Anthony Ball, P., 2020. Survey of nurses' knowledge and practice regarding medication administration using enteral tubes. J. Clin. Nurs. 29 (23–24), 4614–4622. https://doi.org/10.1111/jocn.15498.

Vassilyadi, F., Panteliadou, A.K., Panteliadis, C., April 2013. Hallmarks in the history of enteral and parenteral nutrition: from antiquity to the 20th century. Nutr. Clin. Pract. 28 (2), 209–217.
Veterans Affairs Total Parenteral Nutrition Cooperative Study Group, 1991. Perioperative total parenteral nutrition in surgical patients. N. Engl. J. Med. 325, 525–532.

Further reading

DeChicco, R.S., et al., 2004. Nutrition Support Handbook, fourth ed. The Cleveland Clinic Foundation, Cleveland, OH.
Institute of Medicine, 2005. Food and Nutrition Board: Dietary Reference Intakes for Macronutrients: Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein and Amino Acids. National Academy of Press, Washington, DC, pp. 265–645.
Kelly, D.A., 2006. Intestinal failure-associated liver disease: what do we know today? Gastroenterology 130 (2 Suppl. 1), S70–S77.
Klein, S., Kinney, J., Jeejeebhoy, K., et al., 1997. Nutrition support in clinical practice: review of published data and recommendations for future research directions. National Institutes of health, American society for parenteral and enteral nutrition, and American society for clinical nutrition. J. Parenter. Enteral Nutr. 21 (3), 133–156.
McClave, S.A., 2007. Nutrition support in acute pancreatitis. Gastroenterol. Clin. N. Am. 36 (1), 65–74.
Mueller, C. (Ed.), 2017. The A.S.P.E.N. Adult Nutrition Support Core Curriculum, third ed. American Society for Parenteral and Enteral Nutrition, Silver Spring, MD, pp. 473–488.
Raman, M., Aghdassi, E., Baun, M., et al., 2006. Metabolic bone disease in patients receiving home parenteral nutrition: a Canadian study and review. J. Parenter. Enteral Nutr. 30 (6), 492–496.
Ratliff, A., Jaroch, L., Nishnick, A., 2020. Parenteral nutrition. In: DeChicco, R.S., et al. (Eds.), Nutrition Support Handbook, fourth ed. The Cleveland Clinic Foundation, pp. 88–100.

Additional resources

Community Resources (i.e., ASPEN, Oley Foundation, Feeding Tube Awareness Foundation).

Food allergies: Diagnosis and management

T.J. David, University of Manchester, Manchester, UK

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of T.J. David, Food Allergies: Diagnosis and Management, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 270–276, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00114-8.

Documenting Possible Food Allergies 337
Documenting a Diagnosis of Food Allergy 338
Practical Diagnostic Difficulties 338
Multiple Mechanisms 338
Inability to Predict Outcome 338
Diagnostic Tests 338
Skin Prick Tests 338
Intradermal Testing 339
Skin Application of Food before Food Challenges 339
Tests for Circulating IgE Antibodies: The Radioallergosorbent (RAST) Test 339
Provocation Tests 340
Open and Blind Challenges 340
The Purpose of Provocation Tests 340
Special Types of Provocation Testing 341
Management 341
Dietary Elimination 341
Malnutrition 341
Calcium 341
Protein and Energy 342
Iodine 342
High-Risk Factors 342
The Role of the Dietitian 342
Cows' Milk Protein Avoidance 342
Egg Avoidance 342
Soy Avoidance 343
Wheat-Free and Gluten-Free 343
Peanut Avoidance 343
Drug Treatment in the Management of Food Allergy 343
Desensitization 343
Further Reading 344

Documenting Possible Food Allergies

The diagnosis of food allergy is made from the history, supported by investigations and by responses to avoidance of specific food triggers. Because the value of investigations is limited, it is especially important to obtain a clear history. There are a number of practical points to be made:

- *Speed of onset*: In general, the quicker the onset of the allergic reaction, the more reliable is the history. If a child develops a violent allergic reaction within a minute or two after ingesting a food, it is much easier to link the reaction to a specific food than if a reaction only occurs hours or days after eating a food.
- *Coincidences need to be excluded*: If a child becomes unwell (e.g., starts wheezing) an hour after eating a specific food, the wheezing could be caused by the food, or it could just be a coincidence. The more times that such a sequence has been observed, the more likely it is that there is a cause and effect relationship.
- *Observations need to be tested for internal consistency*: Someone may believe that he or she is allergic to a food if a symptom (e.g., urticaria) occurs on (say) three occasions after eating a specific food. It is important to find out the following:

1. Whether the subject has had the same symptoms on other occasions when the suspect food trigger was not taken.
2. Whether the subject has taken the suspect food on one or more other occasions without any adverse effects.

Failure to seek inconsistencies such as these is one factor that is responsible for the overdiagnosis of food allergy.

Documenting a Diagnosis of Food Allergy

If it is reported that someone is allergic to an item, it is important to probe further and find out on what basis the person has been deemed allergic. It is common to find children and adults who are believed to be allergic to a food solely on the basis of tests such as skin tests or blood tests, which are in fact almost wholly unreliable (see below). It is also common for people to believe that they are allergic to something because a health professional said so one day, which on further enquiry turns out to be on flimsy or nonexistent grounds.

Another common problem is the misinterpretation of a sequence of events. For example, a child with an ear infection is given an antibiotic, and 3 days later gets diarrhea, so the parents come to believe the child is allergic to the antibiotic. In fact the cause of the diarrhea is far more likely to be either an underlying viral infection, or a disturbance of the gut flora. Another example is the report of a child who is believed to be allergic to sesame seeds because of the reactions occurring after eating buns coated with sesame seeds; many such children are in fact not allergic to sesame seeds but are reacting to the egg glaze that has been used as an adhesive for the seed coating. Another common example is the child with asthma who coughs and wheezes after drinking a diluted orange squash drink, with the result that it is believed that the child is reacting to the yellow-orange coloring agent tartrazine. In fact such reactions are more likely to be due to sulfite preservatives in the squash; sulfites trigger symptoms in 60% or more of children with asthma.

Practical Diagnostic Difficulties

Multiple Mechanisms

Reactions to foods are a heterogeneous group of disorders caused by a variety of different immunological and pharmacological mechanisms. In any individual case, the precise mechanism is often not known. No single type of laboratory test could possibly cover all the different types of possible mechanisms of reactions to foods. Even if one focuses on food allergy, there are a number of different possible immunological mechanisms, including IgE antibody mediated, cell mediated, and circulating immune complexes.

Inability to Predict Outcome

In many situations (e.g., atopic disease), the subject wants to know whether there will be any benefit from food avoidance (e.g., not drinking cows' milk or not eating apples). Even if there were valid tests for the diagnosis of food intolerance, the outcome of avoidance measures depends on a number of other variables. Allergen avoidance may succeed for the following reasons:

(1) the patient was intolerant to the item; (2) coincidental improvement; (3) placebo response.

The reasons why a trial of food avoidance may fail to help can be summarized as follows:

(1) The subject is not allergic to the food. (2) The period of elimination was too short. For example, where a child has an enteropathy (damage to the small intestine) due to food allergy, it may take a week or more for improvement in symptoms to occur. (3) The food has been incompletely avoided. This may happen in a subject supposedly on a cows' milk protein-free diet who still continues to receive food that contains cows' milk proteins such as casein or whey. (4) The subject is allergic to other items, which have not been avoided. For example, a child with an allergy to cows' milk protein who fails to improve when given a soy-based milk to which they also have an allergy. (5) Coexisting or intercurrent disease, for example, gastroenteritis in a child with loose stools who is trying a cows' milk-free diet. (6) The patient's symptoms are trivial and have been exaggerated or do not exist at all and have either been imagined or made up by the parents. It is unrealistic to expect there to be a simple test that can overcome all these problems.

Diagnostic Tests

Skin Prick Tests

The principle of skin prick tests is that the skin weal and flare reaction to an allergen demonstrates the presence of mast-cell-fixed antibody, which is mainly IgE antibody. IgE antibody is produced in plasma cells, and is distributed in the circulation to all parts of the body, so that sensitization is generalized and therefore can be demonstrated by skin testing. In the presence of specific IgE antibody, mast cells in the skin release histamine, which in turn causes a visible weal and flare reaction in the skin.

The procedure involves a drop of allergen solution being placed on the skin, which is then pricked with a hypodermic needle. Two control solutions should also be used: the diluent, in order to detect false-positive reactions; a positive control (e.g., a histamine solution), to enable comparison with a positive result of an allergen solution. The skin prick test induces a response that reaches a peak in 8–9 min for histamine and 12–15 min for allergens. The size of the weal reaction (and not the larger red flare) is measured.

There are numerous problems with skin prick tests, including the following:

1. There is no agreed definition about what constitutes a positive reaction.
2. The size of the weal depends to some extent on the potency of the extract.

3. Antihistamines and tricyclic antidepressants suppress the histamine-induced weal and flare response of a skin test. The suppressive effect of antihistamines may last from a week up to several months for some of the more recently introduced nonsedating antihistamines.
4. *False-positive tests*: Skin prick test reactivity may be present in subjects with no clinical evidence of allergy or intolerance. This is sometimes described as 'asymptomatic hypersensitivity' or 'subclinical sensitization.' Whilst many with positive skin prick tests will never develop the allergy, some subjects with positive skin prick tests do develop symptoms later. However, since the test cannot identify those who are going to develop symptoms, the skin test information is of no practical value.
5. *False-positive results*: Skin prick test reactivity may persist after clinical evidence of intolerance has subsided. For example, in a study of children with egg allergy, it was noted that five out of 11 who grew out of egg allergy had persistently positive skin prick tests after the allergy had disappeared.
6. *False-negative tests*: Skin prick tests are negative in some subjects with genuine food allergies.
7. Skin prick tests mainly detect IgE antibody. However, many adverse reactions to food are not IgE mediated, in which case skin prick tests can be expected to be negative. Taking cows' milk protein intolerance as an example, patients with quick reactions often have positive skin prick tests to cows' milk protein, but those with delayed reactions usually have negative skin prick tests.
8. False-negative results are a problem in infants and toddlers, when the weal size is much smaller than later in life.
9. There is a poor correlation between the results of provocation tests (e.g., double-blind food challenges) and skin prick tests. For example, in one study of 31 children with a strongly positive (weal >3 mm in diameter) skin prick test to peanut, only 16 (56%) had symptoms when peanuts were administered.
10. Commercial food extracts (sometimes heat treated) and fresh or frozen raw extracts may give different results (more positives with raw foods), reflecting the fact that some patients are allergic to certain foods only when taken in a raw state. In others the reverse is the case.

Skin prick tests are mainly used in research studies. The results of skin tests cannot be taken alone, and standard textbooks on allergy acknowledge that "the proper interpretation of results requires a thorough knowledge of the history and physical findings." The problems in clinical practice are, for example, whether or not a subject with atopic disease (eczema, asthma, or hay fever) or symptoms suggestive of food intolerance will benefit from attempts to avoid certain foods or food additives. However, skin prick test results are unreliable predictors of response to such measures.

Skin test results are known to be misleading in cases of inhalant allergy (e.g., allergy to dust mites or grass pollen) but skin prick tests for food allergy are especially unreliable because of the large number of false-positive and false-negative reactions.

Intradermal Testing

Intradermal testing comprises the intradermal injection of 0.01–0.05 ml of an allergen extract. It can cause fatal generalized allergic reaction (anaphylaxis), and is only performed if a preliminary skin prick test is negative. Intradermal tests are more sensitive than skin prick testing, and hence also produce even more false-positive reactions, making the interpretation of the results of intradermal testing even more difficult than that for skin prick testing. The difficulty in interpretation of the results, the pain of intradermal injections, and the risk of anaphylaxis mean that intradermal testing has no place in the routine investigation of food allergy.

Skin Application of Food before Food Challenges

There is one situation where direct application of food to the skin may be of practical value, and that is before a food challenge in a child in whom one fears an anaphylactic reaction. An example might be a 6-month-old infant with a history of a severe allergic reaction to egg. If the parents wish to see if the child has outgrown the allergy without directly administering egg and risking a violent reaction, a simple approach is to rub some raw egg white into the skin and observe the skin for a few minutes. If the skin application of egg in this way causes an urticarial reaction, then a gradual diminution and disappearance of this response during the succeeding months and years can probably be taken to indicate the development of tolerance, and a continuing brisk response to skin contact would constitute a deterrent to an oral challenge. However, this is only an approximate guide, and there are a number of possible reasons why such testing may give false-positive (e.g., using a raw food when the food is usually eaten cooked, such as egg or potato) or false-negative (e.g., the child is receiving antihistamine drug) results.

Tests for Circulating IgE Antibodies: The Radioallergosorbent (RAST) Test

The RAST test is the best known of a number of laboratory procedures for the detection and measurement of circulating IgE antibody. Unfortunately, the clinical interpretation of RAST test results is subject to most of the same pitfalls as that for skin prick testing. Additional problems with RAST tests are the cost, and the fact that a very high level of total circulating IgE (e.g., in children with severe atopic eczema) may cause a false-positive result. Depending on the criteria used for positivity, there is a fair degree of correlation between the RAST test and skin prick test results.

Provocation Tests

A provocation test may be useful to confirm a history of allergy. An example might be a child who developed wheezing and urticaria minutes after eating a rusk that contained, as its main ingredients, wheat and cows' milk protein. To determine which component, if any, caused the reaction, oral challenges with individual components can be conducted.

However, the results of provocation tests cannot prove that improvement in a disease has been caused by food avoidance. For example, a child with atopic eczema is put on a diet avoiding many foods, and the eczema improves. This improvement could be a coincidence, a placebo effect, or due to the diet. Just because the child is shown to react to a single food does not prove that avoidance of that food was the cause of the improvement.

Open and Blind Challenges

Where the subject and the observer knows the identity of the administered material at the time of the challenge, the procedure is said to be an 'open' challenge. In a 'single-blind challenge' the observer but not the patient or family know the identity of the test material. To avoid bias on the part of the observer, a double-blind challenge is required. A 'double blind' challenge involves exposing the subject to a challenge substance, which is either the item under investigation or an indistinguishable inactive (placebo) substance. Neither the subject nor the observer knows the identity of the administered material at the time of the challenge or during the subsequent period of observation.

The Purpose of Provocation Tests

The aim of a food challenge is to study the consequences of food or food additive ingestion. Provocation tests are helpful in the following ways:

(1) to confirm a history (parents' observations of alleged food allergy are notoriously unreliable, as are adults' beliefs about their own allergies); (2) to confirm the diagnosis, for example, of cows' milk protein allergy in infancy, where the diagnostic criteria include improvement on elimination diet and relapse on reintroduction; (3) to see if a subject has grown out of a food intolerance; (4) as a research procedure. The food challenge should replicate normal food consumption in terms of dose, route, and state of food. It should also be performed in such a way that the history can be verified. Thus, for example, there is no point solely looking for an immediate reaction if the parents report a delayed reaction.

Open food challenges are the simplest approach, but open food challenges run the risk of bias influencing the parents' (or doctors') observations. Often this is unimportant. But in some cases belief in food intolerance may be disproportionate, and where this is suspected there is no substitute for a double-blind placebo-controlled challenge. An open challenge may be an open invitation to the overdiagnosis of food intolerance. For example, in the UK parents widely believe that there is an association between food additives and bad behavior, but in one series, double-blind challenges with tartrazine and benzoic acid were negative in all cases in a study of 24 children with a clear parental description of adverse reaction.

The double-blind placebo-controlled challenge is regarded as the state-of-the-art technique to confirm or refute histories of adverse reactions to foods. The ability to unravel food-related problems is said to be limited only by the imagination of the physician and a clever dietitian. In fact, the technique is subject to a number of potential limitations, not all of which can be overcome.

Effect of Dose

In some cases of food intolerance, minute quantities of food (e.g., traces of cows' milk protein) are sufficient to provoke florid and immediate symptoms. In other cases, much larger quantities of food are required to provoke a response. Hill *et al.* demonstrated that whereas 8–10 g of cows' milk powder (corresponding to 60–70 ml of milk) was adequate to provoke an adverse reaction in some patients with cows' milk protein allergy, others (with late onset symptoms and particularly atopic eczema) required up to 10 times this volume of milk daily for more than 48 h before symptoms developed.

Concealing Large Doses is Difficult

Standard capsules that contain up to 500 mg of food are suitable for validation of immediate reactions to tiny quantities of food, but concealing much larger quantities of certain foods (especially those with a strong smell, flavor, or color) can be very difficult.

Route of Administration

Reactions to food occurring within the mouth are likely to be missed if the challenge by-passes the oral route, e.g., administration of foods in a capsule or via a nasogastric tube. In practice, patients whose symptoms are exclusively confined to the mouth are unusual, and where there is a history of purely oral reactions an alternative challenge procedure can be employed. In subjects who are intolerant to sulfites, it is well recognized that the administration of sulfites in capsules or directly into the stomach via a nasogastric tube usually fails to provoke an adverse reaction, whereas the oral administration of solution will succeed in doing so.

Problems with Capsules

Capsules are unsuitable for use in children who cannot swallow large capsules, and this is a major limitation as most cases of suspected food allergy are in infants and toddlers. Furthermore, it is unsatisfactory to allow patients or parents to break open capsules and mix the contents with food or drink, as the color (e.g., tartrazine) or smell (e.g., fish) will be difficult or impossible to conceal and the challenge will no longer be blind.

Anaphylactic Shock Danger

There is a danger of producing anaphylactic shock, even if it had not occurred on previous exposure to the food. For example, in Goldman's classic study of cows' milk protein intolerance, anaphylactic shock had been noted before cows' milk challenge in five out of 89 children, but another three developed anaphylactic shock as a new symptom after cows' milk challenge. In a study of 80 children with atopic eczema treated with elimination diets, anaphylactic shock occurred in four out of 1862 food challenges. The risk appears to be greatest for those who have received elemental diets.

Effect of Disease Activity

A food challenge performed during a quiescent phase of the disease (e.g., urticaria, eczema, or asthma) may fail to provoke an adverse reaction.

Additive Effect of Triggers

Although some patients react repeatedly to challenges with single foods, it is possible (but unproven) that some patients only react adversely when multiple allergens are given together. There certainly are some subjects who only react in the presence of a nonfood trigger, such as exercise or taking aspirin.

Special Types of Provocation Testing

Other than giving a suspect food by mouth, and asking the subject to swallow it, there are some alternative approaches, which are outlined below.

Oral Mucosal Challenge

A small portion of food is applied to the mucosa inside the mouth, and one looks for reactions such as swelling of the lips, and tingling or irritation of the mouth or tongue, possibly followed by other more generalized symptoms such as urticaria, asthma, vomiting, abdominal pain, or anaphylactic shock. Patients with food intolerance commonly make use of these oral symptoms, spitting out and avoiding further consumption of a food that provokes the symptom.

Gastric Mucosal Challenge

In this procedure, an allergen is applied directly to the gastric mucosa via an endoscope, and the mucosa is then observed for signs of a reaction. In addition, it is possible to take biopsies of the gastric mucosa to study the histological changes and measure the tissue concentration of mediators of inflammation such as histamine.

Rectal Challenges

The standard test to confirm a diagnosis of celiac disease is the jejunal biopsy, in which a small portion of jejunal mucosa is obtained with the aid of a special capsule that is swallowed, and which passes into the small intestine. When in the correct location, the capsule is triggered and withdrawn; it contains a portion of intestinal mucosa, which can then be examined under the microscope. Alternatively, gluten can be instilled into the rectum, in order to look for a reaction that would signify celiac disease. This procedure requires multiple biopsies from the rectum, and it is uncertain whether the results are reliable.

Management

Dietary Elimination

The management of food allergy consists largely of elimination from the diet of the trigger food or foods. Elimination diets are used either for the diagnosis or the treatment of food intolerance, or for both. A diet may be associated with an improvement in symptoms because of intolerance to the food, a placebo effect, or the improvement may have been a coincidence. The degree of avoidance that is necessary to prevent symptoms is highly variable. Some patients are intolerant to minute traces of food, but others may be able to tolerate varying amounts. Strict avoidance and prevention of symptoms are the aims in certain instances, but in many cases it is unknown whether allowing small amounts of a food trigger could lead to either enhanced sensitivity or to the reverse, increasing tolerance. The duration required for dietary avoidance varies. For example, intolerance to food additives may last only a few years, whereas intolerance to peanuts is usually lifelong. Although food allergy is common in children, most have grown out of the problem by the age of 5 years; an important exception is those with nut allergy.

Malnutrition

Malnutrition is a major risk of unsupervised diets.

Calcium

Cows' milk is an important source of calcium, and avoidance of cows' milk and its products carries the risk of an inadequate intake of calcium. Unfortunately, it is far from clear what constitutes an adequate intake for various different age groups.

Protein and Energy

Milk, eggs, fish, meat, wheat, and their respective manufactured food products are important sources of protein and energy. Avoidance of these without the provision of alternative sources of protein and energy runs the risk of an inadequate intake, and growth failure, serious malnutrition, and weight loss are well documented sequelae of unsupervised and inappropriate dietary elimination.

Iodine

Cows' milk and dairy products are important sources of dietary iodine. Exclusion of cows' milk products and a number of other items from the diet, coupled with the consumption of large amounts of soy milk, which has been reported to cause hypothyroidism by increasing fecal loss of thyroxine, have resulted in hypothyroidism and growth failure due to dietary iodine deficiency.

High-Risk Factors

The risk of malnutrition from an elimination diet is particularly high in the following situations:

(1) The diet is not supervised by a dietitian. (2) There is chronic disease before diagnosis, or concurrent chronic disease such as severe atopic eczema. The subject's nutrient requirements may be increased. (3) Malabsorption or enteropathy increases the risk of malabsorption of nutrients. (4) The subject is avoiding sunlight. The risk of vitamin D deficiency may compound the effects of a low calcium intake. (5) The subject is already on a diet that excludes multiple foods, e.g., vegan or macrobiotic diet.

The Role of the Dietitian

The dietitian has three roles in the management of elimination diets. One is to ensure that the resulting diet is nutritionally adequate, and to prevent potential deficiency states by recommending (in an infant) appropriate amounts of infant milk formula, and (in older children or adults) supplements of calcium, vitamins, and so on. Another role is to advise how to avoid specific foods, particularly those contained in manufactured foods. Third, the dietitian makes suggestions as to how to make the diet practical and palatable, and suggests recipes for use with a limited range of foods (e.g., how to make biscuits with potato flour).

Cows' Milk Protein Avoidance

Any form of cows' milk, whether fresh, skimmed, condensed, or evaporated, needs to be avoided. Also forbidden are milk products that contain casein, whey, and nonfat milk solids. Where milk substitutes are required, the choice lies between formulas based on soy protein, casein hydrolysate, or whey hydrolysate. Soya formulas are cheaper, but unsuitable for those who are also intolerant to soya.

Butter, margarine, cream, cheese, ice cream, and yogurt all need to be avoided. Fats that can be used instead include margarines made from pure vegetable fat (e.g., Tomor) and lard. Caution is required with baby foods, as a large number of manufactured products, e.g., rusks, contain milk protein. A common trap is the so-called 'vegetarian' cheese, often wrongly believed to be safe for subjects with cows' milk allergy. In fact, it differs from ordinary cheese only in the use of nonanimal rennet and is unsuitable for people with cows' milk allergy. Meat, game, and poultry are all allowed, but sausages and pies should be avoided unless it is known that they are milk free. Intolerance to cows' milk protein is not a reason to avoid beef. Eggs are allowed, but not custard or scrambled egg, which may contain milk. Fish is permitted, unless it is cooked in batter (which unless otherwise stated should be assumed to contain milk) or milk. Lemon curd, chocolate spread, chocolate (unless stated to be milk-free), toffee, fudge, caramels, and butterscotch are all unsuitable. All ordinary cereals (e.g., oats) are allowed, but caution is required with manufactured breakfast cereals, some of which contain milk powder.

It is essential to check the list of ingredients on the label of any manufactured foods. There is a special problem with unwrapped foods, because there is no label of ingredients. Examples include bread, sausages, or confectionery.

Egg Avoidance

Eggs (both the white and the yolk) and all products that contain egg or albumen must be avoided. As well as hen's eggs, eggs of other birds such as geese, turkeys, and quails must be avoided. Eggs are widely used to make cakes and are sometimes used in the manufacture of bread. Egg wash or glaze is commonly brushed on to the surface of rolls, buns, or baps, and also bread, cakes, and pastry used in puddings (e.g., apple pie). Sweets can be a hazard because they are usually sold without information about ingredients, and egg is included in several products.

Mayonnaise normally contains egg; custard usually does not, with the exception of egg custard and egg custard tarts. Eggs are an essential ingredient of souffles and certain sauces, such as Bearnaise or Hollandaise sauce.

Egg allergy is not a reason to avoid eating chicken.

Soy Avoidance

The major difficulty is mass-produced bread, because in the UK soy is often included as an ingredient in flour. Soy is also found in manufactured products that contain hydrolyzed or textured vegetable protein, and minced beef, which unless described as 'pure beef' has been known to include quantities of soy protein.

Wheat-Free and Gluten-Free

These terms cause confusion; they are not interchangeable. Subjects who are allergic to wheat cannot tolerate foods that contain any type of wheat. Subjects with celiac disease can tolerate all wheat proteins other than the gluten fraction.

Peanut Avoidance

Peanut is also known as groundnut or arachis, so these three names need to be sought on labels of manufactured foods as well as some pharmaceutical products. The difficulty comes with 'vegetable oil,' which may include peanut oil; only by writing to the manufacturer of individual products can the composition of the vegetable oil be determined. It is not known to what extent subjects with peanut allergy should avoid peanut oil. Most peanut oil used in food manufacture is highly refined, and contains only very minute quantities of peanut protein. In a number of small-scale studies, subjects with peanut allergy were found not to react when given highly refined peanut oil. However, it remains possible that such oil contains traces of protein sufficient to result in enhanced reactivity, such that when the subject does ingest peanut accidentally the reaction is worse than previously. On this basis, subjects with peanut allergy should really be advised to avoid peanut oil.

Drug Treatment in the Management of Food Allergy

At present, drug treatment has little part to play in the management of food allergies. There are two exceptions. First, there are a very small number of cases in which the reaction to a food is exclusively gastrointestinal, and in whom the reaction can be blocked by taking the drug sodium cromoglycate by mouth 20 min before the trigger food is swallowed. Second, there are a small number of individuals who develop the life-threatening reaction, of anaphylactic shock when exposed to a trigger food. There are three ways in which anaphylactic shock may prove fatal. First, rapid swelling of the soft tissues in the pharynx may completely obstruct the airway; the treatment is to bypass the obstruction, either by passing an endotracheal tube, or by performing a tracheostomy. Another mechanism is severe shock, with a profound drop in blood pressure; the life-saving treatment is to restore the circulating volume with intravenous fluids and to give oxygen. The third mechanism is severe bronchoconstriction (asthma); here, the life-saving treatment is with bronchodilator drugs and artificial ventilation. If patients with life-threatening anaphylactic shock are to be saved, they must be given urgent (within minutes) medical attention. For individuals who have already experienced a life-threatening allergic reaction to a food, it is a common practice to provide them with a syringe preloaded with adrenaline (epinephrine), with the aim that this should be administered while waiting for medical help. Unfortunately, self-administered adrenaline is not without its hazards (e.g., inadvertent intravenous administration causing fatal cardiac arrest), and there is no proof that it is life saving; indeed, there are many cases in which the subject died despite the use of epinephrine. Nevertheless, it is the best one can do when faced with someone who is experiencing a life-threatening allergic reaction to a food. The need for urgent medical help cannot be overemphasized.

There is little evidence that antihistamine drugs are of any value. It would be reasonable to take a nonsedating fast-acting antihistamine such as terfenadine if experiencing an allergic reaction to a food, but it is questionable whether it will have much effect.

A number of new approaches to the treatment of IgE-mediated food allergy are being examined. In a double-blind placebo-controlled study of monthly injections of a preparation of anti-IgE antibodies, treated patients with peanut allergy required significantly greater amounts of peanut protein to elicit allergic symptoms compared with control subjects. Another anti-IgE preparation has been used in the treatment of asthma but has not been evaluated in peanut allergy. Theoretically, anti-IgE antibody treatment should be protective against multiple food allergens, although it would have to be administered indefinitely. Other experimental approaches include a concoction of traditional Chinese herbs, injection of heat-killed *Escherichia coli* containing mutated recombinant peanut proteins Ara h 1 to Ara h 3, the use of immunostimulatory sequences, and the use of chimeric protein that could form complexes with allergen-specific IgE bound to mast cells and basophils.

Desensitization

In theory it ought to be possible to desensitize subjects with food allergy by giving injections of gradually increasing quantities of an appropriate extract of the food trigger. In practice, such treatment is not available. One at present insurmountable difficulty is that desensitization (also known as hyposensitization) treatment carries a small risk of death from the treatment itself. A subject has a series of injections without any major problem, but then without warning drops dead from anaphylaxis after the next injection. There is some data to show that desensitization performed in this way can work, but such subjects would probably require

maintenance injections on a permanent basis, and the very subjects most at risk of fatal anaphylaxis from accidental injection are quite probably also the ones most at risk from fatal anaphylaxis resulting from desensitization treatment.

Further Reading

Bock, S.A., Sampson, H.A., Atkins, F.M., et al., 1988. Double-blind, placebo-controlled food challenge as an office procedure: A manual. Journal of Allergy and Clinical Immunology 82, 986–997.

Burks, W., Kulis, M., Pons, L., 2008. Food allergies and hypersensitivity: A review of pharmacotherapy and therapeutic strategies. Expert Opinion on Pharmacotherapy 9 (7), 1145–1152.

Buttriss J (ed.) (2002) Adverse Reactions to Food. *The Report of a British Nutrition Foundation Task Force*. Oxford: Blackwell.

Chafen, J.J., Newberry, S.J., Riedl, M.A., et al., 2010. Diagnosing and managing common food allergies: A systematic review. Journal of the American Medical Association 303 (18), 1848–1856.

David, T.J., 1987. Reactions to dietary tartrazine. Archives of Disease in Childhood 62, 119–122.

David, T.J., 1989. Hazards of challenge tests in atopic dermatitis. Allergy 44 (Supplement 9), 101–107.

David, T.J., 1993. Food and Food Additive Intolerance in Childhood. Blackwell Scientific Publications, Oxford.

Demoly, P., Piette, V., Bousquet, J., 1998. In vivo methods for study of allergy. Skin tests, techniques, and interpretation. In: Adkinson, N.F., Yunginger, J.W., Bisse, W.W., et al. (Eds.), Middleton's Allergy Principles & Practice, 6th edn. Mosby, St. Louis, pp. 631–643.

Fries, J.H., Glazer, I., 1950. Studies on the antigenicity of banana, raw and dehydrated. Journal of Allergy 21, 169–175.

Gonsalves, N., 2007. Food allergies and eosinophilic gastrointestinal illness. Gastroenterology Clinics of North America 36 (1), 75–91.

Houle, C.R., Leo, H.L., Clark, N.M., 2010. A developmental, community, and psychosocial approach to food allergies in children. Current Allergy and Asthma Reports 10 (5), 381–386.

Lee, L.A., Burks, A.W., 2006. Food allergies: Prevalence, molecular characterization, and treatment/prevention strategies. Annual Review of Nutrition 26, 539–565.

Li, J.T., 2007. Allergic diseases: Diagnosis and treatment. Journal of the American Medical Association 298, 2205–2219.

Metcalfe, D.D., Sampson, H.A., Simon, R.A. (Eds.), 1997. Food Allergy: Adverse Reactions to Foods and Food Additives, 2nd edn. Blackwell, Oxford.

Montalto, M., Santoro, L., D'Onofrio, F., et al., 2008. Adverse reactions to food: Allergies and intolerances. Digestive Diseases 26 (2), 96–103.

Patterson, R., Grammer, L.C., Greenberger, P.A. (Eds.), 1997. Allergic Diseases. Diagnosis and Management, 5th edn. Lippincott-Raven, Philadelphia.

Simons, F.E.R., 2004. First-aid treatment of anaphylaxis to food: Focus on epinephrine. Journal of Allergy and Clinical Immunology 113, 837–844.

Young, M.C., Munoz-Furlong, A., Sicherer, S.H., 2009. Management of food allergies in schools: A perspective for allergists. Journal of Allergy and Clinical Immunology 124 (2), 175–182, 182 e171–174; quiz 183–174.

Food intolerance

B. Caballero, Johns Hopkins University, Baltimore, MD, USA

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of B. Caballero, Food Intolerance, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 315–321, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00121-5.

Definition of Food Intolerance **345**
Any Food taken in Excess may be Harmful 345
Principal Mechanisms and Pathophysiology of Food Intolerance **346**
Food Allergy 346
Enzyme Defects 346
Lactase Deficiency 346
Hereditary Fructose Intolerance 346
Pharmacological Mechanisms 346
Caffeine 346
Sodium Nitrite 347
Tyramine, Histamine, and Other Vasoactive Amines 347
Tyramine and Migraine 348
11 β-Hydroxysteroid Dehydrogenase and Liquorice 348
Irritant Mechanisms 348
Specific Drug–Food Combinations 348
Toxic Mechanisms 349
Protease Inhibitors 349
Lectins 349
Lathyrogens 349
Mimosine 349
Djenkolic Acid 349
Goitrogens 349
Cyanogens 349
Vicine and Convicine 349
Cycasin 350
Pyrrolizidine Derivatives 350
Lupin Alkaloid 350
Other Examples 350
Food Storage **350**
Practical Applications **351**
Further Reading **351**

Definition of Food Intolerance

Food intolerance can be defined as a reproducible adverse reaction to a specific food or food ingredient, and it is not psychologically based. Although this appears straightforward, a key limitation of this definition is the lack of consistent definition on what constitutes an 'adverse reaction.' The phrase has a strong subjective component, from the part of both the patient as well as the physician. The problem is compounded by the wide variation in tolerance to the different 'reactions' associated with eating in the general population.

Any Food taken in Excess may be Harmful

The definition above does not take into account the dosage. Large quantities of certain foods may result in disease in certain individuals, although such disorders are not usually included in the category of food intolerance. Any food, however harmless, can be harmful if taken in excess. Notable examples of this are:

1. Apples, pears, and honey are rich sources of fructose – a sugar which in early childhood is not well absorbed if taken in large quantities. Thus, if a child takes a quantity of fructose in excess of that which can be absorbed in the gastrointestinal tract, the result will be loose stools (diarrhea) due to the osmotic effect of unabsorbed fructose. It should be noted that whereas this

applies to normal children, there are, in addition, a few children who are especially poor at handling ingested fructose, and in these children even small quantities of fructose-containing foods will cause florid diarrhea.
2. Chicken liver is a rich source of vitamin A. There are reported cases of infants who were fed large quantities of chicken liver, and who developed raised intracranial pressure as a consequence of vitamin A toxicity.
3. In those who are genetically predisposed, ingestion of an excess of purine-rich foods contributes to hyperuricemia, leading to gout, a disorder which is not usually regarded as a form of food intolerance.

Principal Mechanisms and Pathophysiology of Food Intolerance

Food Allergy

The term 'allergy' implies a definite immunological mechanism. This could be antibody mediated, cell mediated, or due to circulating immune complexes. The clinical features of an allergic reaction include urticaria (nettle rash), angioedema, rhinitis (sneezing, nasal discharge, blocked nose), worsening of preexisting atopic eczema, asthma (wheezing, coughing, tightness of the chest, shortness of breath), vomiting, abdominal pain, diarrhea, and anaphylactic shock.

Enzyme Defects

Inborn errors of metabolism may affect the digestion and absorption of carbohydrate, fat, or protein. In some subjects the enzyme defect is primarily gastrointestinal, causing defects in digestion or absorption. An example is lactase deficiency (see below). In other subjects, the enzyme defect is systemic. An example is the rare disorder of hereditary fructose intolerance, described below.

Lactase Deficiency

An example of an enzyme defect causing food intolerance is lactase deficiency. In this condition, which is primarily a disorder that affects infants and young children, there is a reduced or absent concentration of the enzyme lactase in the small intestinal mucosa. Affected individuals are unable to break down ingested lactose, the main sugar found in milk, and which if unabsorbed passes into the large intestine, where there are two consequences. One is an osmotic diarrhea. The other is that some of the unabsorbed lactose is broken down by intestinal bacteria, accompanied by the production of gas (hydrogen) leading to abdominal distension and flatus and the production of organic acids that cause perianal soreness or excoriation. The elimination in exhaled air of hydrogen produced in the colon from unabsorbed carbohydrate can be used as a simple yet reliable test for malabsorption (breath hydrogen test). Most commonly used to assess lactose malabsorption, the test provides quantitative estimates of the amount of carbohydrate that was not absorbed in the small intestine and reached the colon. Because hydrogen is produced by colonic bacteria, the test requires a normal gut flora and is therefore affected by antibiotic use, diarrhea, or other large intestine disorders.

The management of lactose intolerance is to avoid foods that contain lactose (mainly cows' milk and its products). For infants it is worth noting that the soya-based infant formulas are lactose free. In theory, an alternative is to add microbial β-galactosidase to cows' milk, which can produce a lactose-free milk, with the inconvenience that it has a sweeter flavor and requires a 24-h incubation period at 4 °C.

In infants and young children, lactase deficiency is usually a transient problem occurring after an episode of gastroenteritis, but it is commonly a feature of any disease that causes damage to the intestinal mucosa (e.g., celiac disease). Levels of lactase tend to fall during mid- to later childhood, and in a number of populations (e.g., African, Mexican, and Greenland Eskimo) a high proportion of adults have very little lactase activity. This adult deficiency is believed to have a genetic basis. Man is the only animal apart from the domestic cat that drinks milk after weaning, and deficiency of lactase in adults could in certain populations be considered the normal state.

Hereditary Fructose Intolerance

In this condition, which has autosomal recessive inheritance, there is deficiency of the liver enzyme fructose 1,6-bisphosphate aldolase. As a result, fructose-1-phosphate accumulates in liver cells, and acts as a competitive inhibitor for phosphorylase. The resulting transient inhibition of the conversion of glycogen to glucose leads to severe hypoglycemia (low blood glucose concentration). Affected infants are symptom free as long as their diet is limited to human milk. If they receive milk formulas or any food that contains fructose they develop attacks of hypoglycemia, shock, coma, and convulsions. There may be jaundice, an enlarged liver, and sometimes progressive liver disease. The treatment requires the complete elimination of fructose from the diet, which may be difficult as fructose is a widely used food additive and a sweetener. A trivial but interesting feature of the condition in survivors is a notable reduction in dental caries, a beneficial result from the need to avoid many types of confectionery.

Pharmacological Mechanisms

Caffeine

A good example of a pharmacological agent found in food with the ability to cause adverse reactions is caffeine. The stimulant effect, which may be welcome at times but unwelcome at others, of 60 mg caffeine in a cup of tea or 100 mg caffeine in a cup of coffee are

well recognized. What is less well recognized is that heavy coffee or tea drinkers can suffer a number of other side effects of caffeine, which stimulates gastric secretion and can cause heartburn, nausea, vomiting, diarrhea, and intestinal colic. Also common are irregular heartbeats, episodes of rapid pulse, sweating, tremor, anxiety, and sleeplessness. Caffeine also has a diuretic effect.

Sodium Nitrite

Another pharmacological effect occurs when unusually large quantities of sodium nitrite are ingested. Sodium nitrite is an antioxidant used as an antibacterial agent, and in quantities of 20 mg or more it can cause dilatation of blood vessels causing flushing and headache, and urticaria.

Tyramine, Histamine, and Other Vasoactive Amines

A further example of a pharmacological mechanism is the adverse effect of various vasoactive amines such as tyramine, serotonin, tryptamine, phenylethylamine, and histamine, which are found in a range of foods such as tuna, pickled herring, sardines, anchovy fillets, bananas, cheese, yeast extracts (such as Marmite), chocolate, wine, spinach, tomato, and sausages. There appear to be three main mechanisms in operation:

1. An abnormally high intake of vasoactive amines, such as histamine or tyramine, either because of a high content in food or because of synthesis of these chemicals in the gut by bacteria.
2. An abnormal effect whereby drugs or chemicals in food interfere with the enzymes that break down vasoactive amines.
3. An abnormal release from mast cells of histamine and other mediators of inflammation, triggered by eating certain foods such as strawberries, shellfish, and alcohol.

Vasoactive amines are the normal constituents of many foods. They arise mainly from the decarboxylation of amino acids, but they may also develop during normal food cooking and during the storage of food. **Table 1** shows the histamine level of various sausages. The term 'semidry' when applied to sausages (**Table 1**) means sausages that are fermented for varying periods. During this sausage-ripening process, the histamine concentration increases, depending on the length of the ripening process. It is estimated that 70–1000 mg of histamine ingested in a single meal is necessary for the onset of toxicity, depending on individual sensitivity. Thus, 130 g of the pepperoni sample that contained 55.0 mg histamine per 100 g (see **Table 1**) would be necessary to cause symptoms in the most sensitive individuals.

The largest amounts of histamine and tyramine are found in fermented foods such as cheese, alcoholic drinks, sausage, sauerkraut, and tinned fish. Badly stored food (see below) such as mackerel and tuna can also contain large amounts of histamine.

The effects of large doses of tyramine, histamine, and other vasoactive amines are extremely variable. Histamine causes flushing (by dilatation of blood vessels), constriction of smooth muscle in the intestine and the bronchi, increased heart rate, headache, fall in blood pressure, and asthma. Tyramine causes constriction of blood vessels, and it stimulates the release of noradrenaline from nerve endings. It can also cause the release of histamine and prostaglandins from mast cells. Dietary tyramine is known to induce hypertension and headache in patients who are taking monoamine oxidase inhibitor drugs. This effect has been shown to be due to inhibition, by these drugs, of intestinal and hepatic metabolism of tyramine, so that the amine accumulates.

The variable effect of histamine taken by mouth is in part due to the varying degree of inactivation in the gastrointestinal tract. Histamine is inactivated by mucoproteins that are produced in the gastrointestinal tract mucosa, but this inactivation can be

Table 1 Histamine levels in sausages

Type of sausage	*Histamine level (mg/100 g)*	
	Mean	*Range*
Cooked sausages[a]		
Bologna	0.55	0.19–0.84
Cooked salami	0.83	0.47–5.86
Kosher salami	0.50	0.33–0.97
Semidry sausages[a]		
Thuringer cervelat	2.35	1.03–3.63
Thuringer	1.19	0.31–2.56
Dry sausages[a]		
Italian dry salami	2.14–24.5[b]	0.42–36.4[b]
Pepperoni	1.03–38.1[b]	0.72–55.0[b]
Chorizo	2.29	0.60–8.08

Source: Reproduced from Taylor SL, Leatherwood M, and Lieber ER (1978) A survey of histamine levels in sausages. *Journal of Food Protection* 41: 634–637, with permission from International Association For Food Protection

[a]The sausages were obtained from retail markets in the San Francisco Bay area.

[b]Depending on the brand tested.

blocked by other amines such as cadaverine and putresceine, which also bind strongly to mucoproteins. Thus, when food that contains cadaverine and putresceine is ingested, more histamine can be absorbed. In fact, most of the histamine that is absorbed is degraded as it is transported across the mucosa by the intestinal enzyme diamine oxidase. Cadaverine and putresceine also have a high affinity for diamine oxidase and can also interfere with the inactivation of histamine by this enzyme. Another barrier to the absorption of histamine is provided by the liver enzyme methyl transferase.

Thus, the effect of histamine and other vasoactive amines on an individual will depend on a number of factors, which include:

1. The amount of vasoactive amine that is present in food.
2. The amount of histamine released (as a result of an allergic process).
3. The permeability of the gastrointestinal tract, including inactivation by mucus and by mechanisms in the gut mucosa.
4. Interference with the synthesis or release of enzymes involved in amine breakdown (e.g., liver damage causing reduced activity of methyl transferase).

Tyramine and Migraine

There has been interest in a possible relationship between dietary tyramine and migraine. One hypothesis is that some patients with migraine have defective metabolism of ingested tyramine in the intestinal wall, which leads to increased absorption, apparently explaining why foods that contain tyramine can provoke attacks in susceptible individuals. However, there is no evidence that the activity of monoamine oxidase, the main tyramine-metabolizing enzyme, is lower in patients with food-induced migraine than in other individuals prone to migraine, although levels of monoamine oxidase in platelets are generally lower in patients with migraine.

Set against these theoretical arguments, in fact, most attempts to induce migraine by tyramine challenge in children and adults have been unsuccessful. Furthermore, a controlled study of exclusion of dietary vasoactive amines in children with migraine failed to demonstrate benefit. In the latter study, patients were randomly allocated to either a high-fiber diet low in dietary amines or a high-fiber diet alone. Although there was no significant difference in the results for the two groups, both groups showed a highly significant decrease in the number of headaches, emphasizing the need for a control diet in studies designed to show that dietary manipulation improves disease.

Of the foods reported to be common triggers of attacks of migraine, only cheese is rich in tyramine. Chocolate is low in this and other vasoactive amines, and red wine usually contains no more tyramine than white wine. Alcoholic drinks, particularly red wine, are commonly reported to provoke attacks of migraine. Whether these attacks are due to the alcohol itself or some other compound is a matter of debate. The major chemical difference between red and white wines is the former's high concentration of phenolic flavonoids such as anthocyanins and catechins, which as well as having direct effects on blood vessels may also inhibit the enzyme phenolsulfotransferase. Patients with food-induced migraine were shown to have significantly lower levels of platelet phenolsulfotransferase activity, and it has been hypothesized (but not proven) that low activity of this enzyme could lead to an accumulation of phenolic or monoamine substrates, which in turn might directly or indirectly provoke attacks of migraine.

Regardless of the possible mechanism, there are a number of subjects with migraine who are made worse by specific dietary triggers such as cheese or wine, for whatever reason, and avoidance of specific food triggers in susceptible subjects may prove helpful in reducing the frequency of attacks.

11 β-Hydroxysteroid Dehydrogenase and Liquorice

Liquorice contains an enzyme that inhibits 11 β-hydroxysteroid dehydrogenase, resulting in sodium and water retention, hypertension, hypokalemia, and suppression of the renin–aldosterone system.

Irritant Mechanisms

Certain foods have a direct irritant effect on the mucous membranes of the mouth or gut, such as the irritant effect of coffee or curry. In certain individuals, food intolerance only occurs in the presence of a coexisting medical disorder. For example, the ingestion of spicy food, coffee, or orange juice provoke esophageal pain in some patients with reflux esophagitis. This effect is unconnected to the temperature or acidity of the food, or to any effect on the lower esophageal sphincter. The treatment in susceptible individuals is to avoid the trigger food item.

Specific Drug–Food Combinations

One example of drug-induced food intolerance is potentiation of the pressor effects of tyramine-containing foods (e.g., cheese, yeast extracts, and fermented soya bean products) by monoamine oxidase inhibitor drugs. Another is the effect of taking alcohol in patients with alcohol dependence during treatment with disulfiram (Antabuse). The reaction, which can occur within 10 min of alcohol and may last for several hours, consists of flushing and nausea.

Toxic Mechanisms

Nature has endowed plants with the capacity to synthesize substances that are toxic, and thus serve to protect them from predators whether they be fungi, insects, animals, or humans. Thus, many plant foods contain naturally occurring toxins. On a worldwide scale, reactions to naturally occurring toxins may outnumber allergic reactions, although it is currently fashionable to pay more attention to the latter.

Protease Inhibitors

Soya beans were originally introduced into the US as a source of oil, the extracted meal being used as a by-product that could provide animals with a source of protein. However, it was recognized that it was necessary to subject soya beans to heat treatment if they were to support the growth of animals. It was later found that the substance responsible for growth inhibition in raw soya beans was a protease (trypsin) inhibitor, and it is now known that protease inhibitors are widely distributed throughout the plant kingdom, particularly in legumes, and to a lesser extent in cereal grains and tubers. In addition to inhibition of growth, one of the most characteristic responses of most animals to trypsin inhibitor is enlargement of the pancreas. The depression of growth is believed to result from endogenous loss of protein (i.e., loss into the gastrointestinal tract) due to hypersecretion by the pancreas. Soya bean products that have been adequately heat treated to inactivate trypsin inhibitor are safe for consumption.

Lectins

There is a protein present in most legumes and cereals that has the property of being able to agglutinate the red blood cells of various species of animals: The so-called phytohemagglutinins or lectins. Some of these lectins, such as ricin from the castor bean, are extremely toxic. Others, such as those in the soya bean, are nontoxic. Lectins appear to be responsible for the fact that many other legumes, unless properly cooked, not only fail to support the growth of animals but can lead to death. Lectins are found in many food items commonly consumed in the human diet including tomatoes, bean sprouts, raw vegetables, fruits, spices, dry cereals, and nuts, and it is not known whether these are harmful in any way. However, it is well recognized that inadequate cooking of red kidney beans can cause severe gastrointestinal upset, with vomiting and diarrhea. It is for this reason that it is recommended that raw red kidney beans should be cooked by initially boiling hard for 10 min.

Lathyrogens

Lathyrism is a paralytic disease that is associated with the consumption of chickling pea or vetch, *Lathyrus sativus*. The causative factor is believed to be an amino acid derivative, β-*N*-oxalyl-, -diaminopropionic acid; this is a metabolic antagonist of glutamic acid, a substance that is involved in the transmission of nerve impulses in the brain.

Mimosine

Mimosine is an amino acid that comprises 1–4% of the dry weight of the legume *Leucaena leucocephala*, and consumption of its leaves, pods, and seeds leads to hair loss in animals. Mimosine is also a goitrogen (see below).

Djenkolic Acid

In parts of Sumatra the djenkol bean is a popular food item. The bean is a seed of the leguminous tree, *Pithecolobium lobatum*, and resembles the horse chestnut in size and color. Consumption of this seed leads to kidney failure that is accompanied by blood and needle-like clusters in the urine, which have been identified as containing the amino acid djenkolic acid.

Goitrogens

Substances capable of producing goiter are present in plants belonging to the cabbage family, including cabbage, turnip, broccoli, cauliflower, brussel sprouts, kale, rape seed, and mustard seed. Cows' milk is a vector for the transmission of goitrogens from animals fed kale and turnips, and may have been responsible for endemic goiter in countries such as Australia and Finland.

Cyanogens

A number of plants are potentially toxic because they contain glycosides from which hydrogen cyanide may be released by enzymatic hydrolysis. The most common plants eaten by humans, in order of their potential cyanide content, are: Lima beans (*Phaseolus lunatus*), sorghum, cassava, linseed meal, black-eyed pea (*Vigna sinensis*), garden pea (*Pisum sativum*), kidney bean (*Phaseolus vulgaris*), Bengal gram (*Cicer arietinum*), and red gram (*Cajanus cajans*).

Vicine and Convicine

These are β-glucosides that are present in broad beans (*Vicia faba*). When consumed by individuals with deficiency of the enzyme glucose-6-phosphate dehydrogenase, these substances precipitate the condition of favism, which is characterized by anemia caused by hemolysis of red blood cells. The enzyme deficiency is a genetic disorder that is confined largely to inhabitants of countries surrounding the Mediterranean basin (Italy, Sicily, Lebanon, Israel, and North Africa) although individuals of the same ethnic background residing in other countries may also suffer from favism.

Cycasin

Cycad seeds or nuts are obtained from *Cycad circinalis*, a palm-like tree that grows throughout the tropics and subtropics. The seeds, unless thoroughly washed, are extremely toxic, causing poisoning in humans and tumors in experimental animals. The toxic ingredient methyl-azoxymethanol, the aglycone of cycasin, is released on hydrolysis of cycasin by intestinal bacteria.

Pyrrolizidine Derivatives

Pyrrolizidine alkaloids are found in a wide variety of plant species. The toxic ingredient belongs to a class of compounds that are derivatives of pyrrolizidine. Large numbers of people have been poisoned through consumption of cereal and grain crops contaminated with pyrrolizidine-containing plants. It is also possible that milk from cows grazing on pastures that contain such plants could act as a vector for the transmission of pyrrolizidine to humans. In one part of western USA one such plant, the tansy ragwort (*Senecio jacobea*) is readily consumed by cows and goats, and the milk from such animals has been shown to contain significant amounts of a pyrrolizidine derivative, jacoline.

Lupin Alkaloid

Milk from animals that have eaten plants from the lupin family, notably *Lupinus latifolius*, may contain quinolizidine alkaloids such as anagyrine. There is strong evidence that these alkaloids are teratogenic in animals, causing severe bony deformities, and there is some evidence that similar defects may occur in the offspring of human mothers who drink alkaloid-containing milk in pregnancy.

Other Examples

There are numerous other examples of toxic substances present in foodstuffs. These include solanidine in potatoes, cyanide in tapioca, mycotoxins in mushrooms and cereal grains, and phototoxic furocoumarins in angelica, parsley, dill, and celeriac, which in sufficient quantities can give rise to a wide variety of toxic reactions (**Tables 2 and 3**).

Food Storage

Chemical changes in food during storage can produce substances that cause food intolerance. An example is intolerance to ripe or stored tomatoes in subjects who can safely eat green tomatoes, where ripening of the fruit produces a new active glycoprotein. Some adverse reactions resulting from food storage come into the category of toxic reactions, such as the rise in levels of histamine and tyramine in certain foods during storage as a result of bacterial decarboxylation. An example of this is the production of histamine in badly stored mackerel and other fish: Scombroid fish poisoning. Contamination of food by antigens such as storage mites or microbial spores may give rise to adverse effects, particularly asthma and eczema. Contamination of food by microorganisms may result in adverse effects. For example, celery, parsnip, and parsley may become infected with the fungus *Sclerotinia scleriotiorum* ('pink rot'), resulting in the production of the photosensitizing chemicals psoralen, 5-methoxypsoralen, and 8-methoxypsoralen.

Table 2 Examples of toxic constituents of plant foodstuffs and their role in plant physiology

Toxic constituent	*Type of food containing toxic constituent*	*Physiological role of toxic constituent*	*Role in plant defense: mechanism of toxic constituent*
Protease inhibitors	Legumes, cereals, potatoes, pineapple	?Prevents degradation of storage protein during seed maturation	Part of defense against invading microbes following mechanical damage to leaves
Hemagglutinins	Legumes, cereals, potatoes	(a) Attach glycoprotein enzymes (b) Role in embryonic development/ differentiation (c) Role in sugar transport or store (d) ?Involved in root nodule nitrogen-fixing bacteria symbiosis	(a) Counteract soil bacteria (b) Antifungal (c) Protect against seed predators
Glucosinolates	Radish, horseradish, turnip, cabbage, rape seed	?Disease & insect resistance role	
Cyanogens	Almonds, cassava, corn, peas, butter beans, bamboo shoots		
Saponins	Alfalfa, French beans, soya beans		

Source: Adapted from Leiner IE (ed.) (1980) *Toxic Constituents of Plant Foodstuffs*, 2nd edn. New York: Academic Press.

Table 3 Examples of foodborne toxins or toxin-producing organisms, excluding plant foodstuffs

Pathogen or toxin	*Principal symptoms*	*Common food source*
Bacillus cereus	(a) Diarrhea	Proteinaceous food, vegetables, sauces, puddings
	(b) Vomiting	Fried rice
Bacillus subtilis	Vomiting, diarrhea	Meat and pastry
	Flushing, sweating	Meat/seafood with rice
Bacillus licheniformis	Diarrhea	Cooked meat and vegetables
Clostridium botulinum	Neuroparalytic disease (botulism)	Meat, fish, vegetables, hazelnut conserve
Clostridium perfringens	Diarrhea, abdominal pain	Meat, poultry
Salmonella enteridis	Diarrhea, abdominal pain, fever, vomiting	Poultry, eggs
Staphylococcus aureus	Vomiting, abdominal pain, diarrhea	Numerous but specially cooked high-protein foods
Verotoxin-producing *Escherichia coli*	Hemorrhagic colitis	Ground beef
Listeria monocytogenes	Listeriosis	Unpasteurized cheese, undercooked meat
Dioxins and dibenzofurans	Adverse effects uncertain when consumed in quantities found in food	Fish
Cantharidin	Sensitivity to urethra and genitalia; priapism	Frogs that have Meloidae (blister beetles)
Methyl mercury	Brain damage	Fish, bread
Toxic alkaloid (saxitoxin) in dinoflagellates and plankton	Diverse neurological disorders (paralytic shellfish poisoning)	Clams, oysters, scallops, and mussels
Brevetoxins	Paresthesia, abdominal pain, diarrhea, transient blindness, paralysis, death (neurotoxic shellfish poisoning)	Clams, oysters, scallops, and mussels
Ciguatera toxin	Diverse gastrointestinal and neurological disorders	Fish (especially reef predators)
Tetrodotoxin	Diverse gastrointestinal and neurological disorders	Puffer fish, certain newts
Domoic acid	Vomiting, diarrhea, hyperexcitation, seizures, memory loss (amnesic shellfish poisoning)	Mussels
Okadaic acid, dinophysis toxins, yessotoxin, pectenotoxins	Diarrhea, vomiting, abdominal pain (diarrhetic shellfish poisoning)	Mussels, scallops
Scombrotoxin (usually histamine)	Headache, palpitations, gastrointestinal disturbance	Mackerel, tuna, and related species
Tetramine (red whelk poisoning)	Diplopia, dizziness, leg pains	Whelks
Grayanotoxins (in honey from areas of Turkey where Rhododendrons are grown)	Hypotension, bradycardia, vomiting, sweating	Honey
Unknown (? in algae) (turtle flesh poisoning)	Cardiorespiratory failure, death	Turtles

Practical Applications

Food arouses not only the appetite but also the emotions. The passion for food that is natural (i.e., free from extraneous ingredients) is not new; in 1857, a survey of adulterants in food showed that childrens' sweets were commonly colored by red lead (lead oxide), lead chromate, mercuric sulfide, and copper arsenite. By the late 1850s, 'pure and unadulterated' had become the stock advertising slogan of those anxious to cash in on the then newly awakened fears of the public. The current scale of the use of additives in food comes as a surprise to most people, and it is understandable that many should find these substances vaguely menacing. Nonetheless, the current phobia of food additives and food processing, and the obsession for the so-called natural or health food arises largely out of misinformation and ignorance. Obsession with the so-called natural or health food ignores the wide range of naturally occurring toxins in foods. The concept of health food is wholly misleading. For example, a survey of 'crunchy' peanut butter showed that 11 out of 59 samples from health food producers contained over 100 $\mu g\,kg^{-1}$ of aflatoxins, over 10 times the proposed maximum permitted level for total aflatoxins. Only one of the 26 samples from other producers contained aflatoxins in excess of 10 $\mu g\,kg^{-1}$, and none contained more than 50 $\mu g\,kg^{-1}$.

Further Reading

Asero, R., 2002. Multiple intolerance to food additives. The Journal of Allergy and Clinical Immunology 110 (3), 531–532.
Assa'ad, A.H., 2006. Gastrointestinal food allergy and intolerance. Pediatric Annals 35 (10), 718–726.
Campbell, A.K., Matthews, S.B., Vassel, N., et al., 2010. Bacterial metabolic 'toxins': A new mechanism for lactose and food intolerance, and irritable bowel syndrome. Toxicology 278 (3), 268–276.
Farese, R.V., Bigieri, E.G., Shackleton, C.H.L., Irony, I., Gomez-Fontes, R., 1991. Licorice induced hypermineralocorticoidism. New England Journal of Medicine 325, 1223–1227.
Forsythe, W.I., Redmond, A., 1974. Two controlled trials of tyramine in children with migraine. Developmental Medicine and Child Neurology 16, 794–799.
Horwitz, D., Lovenberg, W., Engelman, K., Sjoerdsma, A., 1964. Monoamine oxidase inhibitors, tyramine, and cheese. Journal of the American Medical Association 188, 90–92.

Jacobsen, M.B., Aukrust, P., Kittang, E., et al., 2000. Relation between food provocation and systemic immune activation in patients with food intolerance. Lancet 356 (9227), 400–401.
Kaufman, H.S., 1986. The red wine headache: A pilot study of a specific syndrome. Immunology and Allergy Practice 8, 279–284.
Lessof, M.H., 1992. Food Intolerance. Chapman & Hall, London.
Littlewood, J.T., Glover, V., Davies, P.T.G., Gibb, C., Sandler, M., Rose, F.C., 1988. Red wine as a cause of migraine. Lancet 1, 558–559.
Moffett, A., Swash, M., Scott, D.F., 1972. Effect of tyramine in migraine: A double-blind study. Journal of Neurology, Neurosurgery and Psychiatry 35, 496–499.
Moffett, A.M., Swash, M., Scott, D.F., 1974. Effect of chocolate in migraine: A double-blind study. Journal of Neurology, Neurosurgery and Psychiatry 37, 445–448.
Nelson, M., Ogden, J., 2008. An exploration of food intolerance in the primary care setting: The general practitioner's experience. Social science & medicine 67 (6), 1038–1045.
Noah, N.D., Bender, A.E., Reaidi, G.B., Gilbert, R.J., 1980. Food poisoning from raw red kidney beans. British Medical Journal 281, 236–237.
Osborn, D.A., Sinn, J., 2006. Formulas containing hydrolysed protein for prevention of allergy and food intolerance in infants. Cochrane Database of Systematic Reviews (4): CD003664.
Pascual, C.Y., Crespo, J.F., Perez, P.G., Esteban, M.M., 2000. Food allergy and intolerance in children and adolescents, an update. European Journal of Clinical Nutrition 54 (Supplement 1), S75–S78.
Patriarca, G., Schiavino, D., Pecora, V., et al., 2009. Food allergy and food intolerance: Diagnosis and treatment. Internal and Emergency Medicine 4 (1), 11–24.
Salfield, S.A.W., Wardley, B.L., Houlsby, W.T., et al., 1987. Controlled study of exclusion of dietary vasoactive amines in migraine. Archives of Disease in Childhood 62, 458–460.
Schafer, T., Bohler, E., Ruhdorfer, S., et al., 2001. Epidemiology of food allergy/food intolerance in adults: associations with other manifestations of atopy. Allergy 56 (12), 1172–1179.
Taylor, S.L., 1986. Histamine food poisoning: Toxicology and clinical aspects. CRC Critical Reviews in Toxicology 17, 90–128.
Wuthrich, B., Ortolani, C. (Eds.), 1996. Highlights in Food Allergy. Karger, Basel.

Hyperactivity: Nutritional aspects

Klaus W. Lange, Institute of Psychology, University of Regensburg, Regensburg, Germany

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A.B. Bax, M.L. Wolraich, Hyperactivity: Nutritional Aspects, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 436–441, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00146-X.

Introduction 353
Methodological issues 354
Dietary patterns in ADHD 355
Elimination of food elements 355
Sucrose restriction 355
Food additives 356
Few-foods or oligoallergenic diet 356
Supplementation of food elements 357
Vitamins 357
Minerals 357
Micronutrient combinations 357
Amino acids 357
Polyunsaturated fatty acids 358
Probiotics 358
Problems of research on nutrition in ADHD 359
Conclusion 359
References 360

Key points

- The current pharmacological treatment of individuals diagnosed with hyperactivity or attention-deficit hyperactivity disorder (ADHD) is unsatisfactory.
- Alternative treatment options include nutritional interventions, some of which attempt to improve ADHD symptoms either through the avoidance of certain food compounds or the supplementation of specific food elements.
- Deficiencies of nutrients such as vitamins, minerals or polyunsaturated fatty acids have not convincingly been shown to be involved in the etiology of ADHD.
- Evidence of a causative link between ADHD and dietary surpluses of nutritional items such as sucrose or artificial food additives or other foods causing behavioral reactions in sensitive children is lacking.
- At group level, there is no conclusive evidence supporting dietary interventions for the treatment of ADHD.
- The few-foods diet allows the design of individually tailored diets and may offer treatment opportunities in subgroups of children with ADHD.
- Further research is required to show the utility of nutrition-related interventions alone or in combination with other treatment approaches in ADHD.

Introduction

While nutrition and diet are an established mainstay of physical and metabolic health, their role in mental health and disease is less commonly recognized. However, considerable evidence points to an important role of nutrition in brain development and functioning. Observational studies have consistently shown that the risk for mental disorders is associated with diet quality. A range of macronutrients and micronutrients have been linked to behavioral, affective and cognitive functions as well as to the incidence and prevalence of mental disorders. Nutritional deficiencies and other dietary factors have been suggested to be involved in the etiology, pathophysiology, prevention, management, and modification of common mental disorders, such as depression, schizophrenia, autism spectrum disorder and attention-deficit hyperactivity disorder (ADHD). Research on dietary and nutritional approaches in the therapy of ADHD and other mental disorders has therefore attracted increasing scientific interest.

The concept of ADHD has developed over the past 50 years, from its early description as a hyperkinetic reaction of childhood to one of the most common diagnoses in child and adolescent psychiatry. ADHD is increasingly diagnosed in adults and is regarded as a lifespan condition. The disorder is complex, multifactorial and heterogeneous, with a wide diversity of features in affected

individuals, including behavioral, cognitive, emotional and motivational profiles, as well as environmental exposures, social influences and genetic susceptibilities. The core symptoms of ADHD, which are age-inappropriate levels of hyperactivity, inattention, and impulsivity, are considered to be continuous traits across the general population, and the disorder is frequently comorbid with other psychiatric conditions. An interaction between genetic susceptibility and environmental and social factors is believed to contribute to the etiology of ADHD. Biological or neuropsychological markers underpinning the validity of the concept of ADHD are currently lacking, and no structural or functional brain alterations can be consistently found in those diagnosed with ADHD. As there are no viable biomarkers or biological diagnostic tools, the diagnosis of ADHD relies solely on clinical symptoms and behavioral assessment. ADHD is diagnosed above a certain degree of symptom severity combined with significant problems in daily life, including social, academic, and occupational dysfunctioning.

Pharmacotherapy with stimulant medications, behavior therapy and parent management training have consistently been demonstrated to yield significant short-term symptom reduction in children and adolescents with ADHD. However, the extent of efficacy of these therapeutic approaches is a matter of debate (Lange, 2017). Many studies have found that ADHD medications administered for several weeks produce statistically significant improvement of core symptoms compared to placebo, as assessed using rating scales. However, small though statistically significant effects on symptom scores may not be noticeable by patients or physicians and do not therefore necessarily equate to any clinically relevant improvements in subjective wellbeing or difficulties in daily life. Furthermore, the minimum difference in rating scale scores indicating an ecologically relevant effect remains unknown. In particular, the impact of medications on academic performance and quality of life in children and adolescents with ADHD has been found to be low.

Cochrane reviews of randomized controlled trials investigating benefits and harms of methylphenidate for children and adolescents with ADHD concluded, in view of low quality of outcome measures and possible bias of studies, that the extent of therapeutic efficacy is unclear (Storebø et al., 2018). Randomized controlled trials supporting the long-term administration of psychostimulants for more than a few months are lacking. As the results of studies investigating efficacy and safety of medications over short-time periods cannot be extrapolated to long-term outcomes, conclusive evidence of longer lasting benefits of ADHD medications on social, educational, and vocational outcomes remains elusive. Furthermore, potential effects are compounded by low adherence to pharmacological treatment.

Short-term administration of stimulant medications in ADHD seems to be relatively safe, but little is known about long-term safety. Long-term observations have suggested that children with ADHD who received psychostimulants into adulthood may present with reduced height as adults with no ongoing improvement in symptomatology. Current information may therefore indicate that pharmacological treatment for ADHD should be used with caution when administered for more than a few months. Given the unproven long-term benefits of medications in the treatment of ADHD, their value may be outweighed by their risks. Concerns regarding the quality and interpretation of the available evidence as well as uncertainties as to the benefit-versus-harm profile of methylphenidate have resulted in the rejection by the World Health Organization Model List of Essential Medicines of applications for the inclusion of this drug in the treatment of ADHD.

In summary, while the pharmacotherapy of ADHD using drugs such as methylphenidate almost certainly provides some symptom reduction in children and adolescents with ADHD, the magnitude and relevance of these effects is unknown. Future large-scale, sufficiently powered and methodologically rigorous randomized controlled trials focusing on both beneficial and harmful effects are required to investigate the effectiveness of pharmacological treatment in alleviating the real-life problems faced by those diagnosed with ADHD.

The lack of evidence demonstrating the ability of either stimulant or non-stimulant agents to address the broader clinical needs of many individuals with ADHD has brought the effectiveness of these treatments into question. The unknown long-term efficacy of ADHD medications as well as concerns regarding unwanted side effects of pharmacotherapy have led to a search for improved or alternative therapeutic options (Lange, 2020a). Lifestyle factors, including diet and nutrition, have been suggested to play a role in the pathogenesis and management of ADHD (Lange, 2018a). As early as the 1920s, a possible relationship between food and hyperkinetic behavior was proposed. More recently, children diagnosed with ADHD have been found to have different dietary patterns from normally developing children. Nutritional interventions in ADHD are mainly based on the assumption that dietary compounds may either produce adverse effects on behavior and should therefore be eliminated or are deficient and need to be supplemented. Dietary agents thought to induce hyperactivity and other ADHD symptoms include sugar as well as food dyes and preservatives. Major dietary components suggested to be involved in the pathogenesis of ADHD and to be effective in assisting in the treatment of ADHD include micronutrients, such as vitamins and minerals, amino acids, and polyunsaturated fatty acids (Lange et al., 2022).

Methodological issues

Various scientific approaches may be employed in establishing the efficacy of therapies and other interventions, including the potential health benefits of foods and nutrients. The primary methods used are observational (epidemiological or population-based) studies and interventional studies (clinical trials).

Observational studies include (1) cross-sectional studies with data collection from an entire population or a representative subgroup at a certain point of time, (2) case-control studies with two existing groups differing in a health outcome and being compared based on a supposed causal factor and (3) longitudinal studies performing correlational analyses of repeated

observations of the same variables over extended periods of time. All types of observational studies can establish associations between the consumption of certain diets or food components and the risk of certain diseases. However, the major limitation of observational studies is their correlational character and their inability to establish causal relationships, since both dietary habits and health risks may be caused by other factors, which were not recorded but were, in fact, the causal influences.

The highest level of evidence capable of establishing a cause-and-effect relationship between an intervention and a health outcome of nutrition can be derived only from the gold standard of randomized controlled trials (RCTs). The design of these studies involves the random assignment of individuals to two different groups. The experimental group receives the dietary intervention, while the placebo control group receives a diet that resembles the intervention diet but does not restrict or add the dietary elements to be studied. It is important that both study participants and individuals implementing the diet and measuring the outcomes are blinded to which group is receiving the experimental diet. This is a major challenge when, for example, parents are involved in the delivery of the diet and the assessment of its outcome.

Dietary patterns in ADHD

The role of dietary patterns and whole diets in ADHD has been explored in several cross-sectional and case-control studies. For example, low adherence to a Mediterranean diet was associated with elevated prevalence of ADHD diagnosis. Hyperactivity in children correlated positively with the intake of processed meat and salty snacks, while a negative correlation was observed for the consumption of vegetables, coarse cereals, aquatic foods, beef, mutton, and milk. Furthermore, the diagnosis of ADHD in adolescents was found to be associated with a higher score for a Western dietary pattern, containing heavily processed foods rich in total fat, saturated fat, sodium, and refined sugars. In contrast, ADHD diagnosis was not associated with a dietary pattern characterized by the intake of omega-3 fatty acids, fiber and folate and a relatively low content of total fat, saturated fat, and refined sugars. In Korea, a traditional-healthy dietary pattern, high in kimchi, grains, bonefish and minerals and low in fat, fast foods and sugary beverages, was associated with reduced likelihood of childhood ADHD. A snacking dietary pattern, with high intake of snacks and processed meat and low intake of noodles, was positively associated with the risk of ADHD. A dietary pattern with relatively large amounts of high-fat, processed foods, and another with high intake of sweets, chocolate and other high-sugar foods, correlated positively and significantly with ADHD symptoms in preschool children in China; while a vegetarian pattern, with coarse food grain, wheat foods, beans, all kinds of vegetables and fresh fruit or vegetable juice, was negatively correlated with ADHD symptoms. A positive association of sweet and fast-food dietary patterns with the prevalence of ADHD was also found in Iranian children. A fish-white meat dietary pattern and mineral-protein nutrient diet have been suggested to have beneficial effects in children with ADHD in mainland China, where high intakes of deep-water fish, shellfish, white meat, freshwater fish, algae and fungi were associated with less ADHD in children. Furthermore, a pattern rich in protein, zinc, phosphorus, selenium, calcium and riboflavin also showed an inverse association with ADHD.

While the findings of observational studies emphasize a potential role of dietary patterns in ADHD, a caveat is the observational study designs, which are unable to establish a causal relationship. Reverse causation, with ADHD behaviors leading to a preference for certain diets, may also explain the above associations. Furthermore, various other lifestyle factors, such as physical activity, may correlate with dietary patterns and may be more important factors in ADHD symptomatology. Risk factors for ADHD, including family history, emotional abuse, prenatal problems, and parental education, frequently differed between individuals with ADHD and controls. Even when statistical adjustment for potential confounding variables was performed, residual confounding was still unavoidable. Commonly used food frequency questionnaires are known to contain some degree of measurement error. Moreover, in many studies, dietary patterns derived from food component analysis explained less than 50% of total variance, suggesting the existence of other patterns.

Elimination of food elements

The elimination of certain elements from the diet is based on the hypothesis that some children may display behavioral changes following the consumption of certain types of food. Elimination diets aim to remove sugar or artificial food colorants and other synthetic food additives. The so-called few-foods diet excludes various foods and additives from children's diets.

Sucrose restriction

When hyperactive children were reported to show greater hyperactivity following the consumption of large amounts of sucrose, several observational studies examined this issue and found that sucrose intake was associated with ADHD symptoms in children and adolescents. However, whether hyperactive behavior was caused by sucrose or by the circumstances associated with high sucrose consumption remained unclear. Several subsequent studies provided convincing evidence that sucrose consumption is not related to ADHD symptoms (Wolraich et al., 1995). For example, the administration of sucrose versus aspartame resulted in similar behavior in hyperactive boys. Furthermore, no differences in signs of ADHD in children believed to be sensitive to sucrose were found between those receiving sucrose for three weeks and those receiving aspartame or saccharine. A meta-analysis of relevant trials, which were mainly challenge studies using sucrose, glucose, or fructose, concluded that sucrose appeared to have no effect on

the behavior or cognitive performance of children (Wolraich et al., 1995). Definitive conclusions could not be drawn due to the limited number of available studies. Insufficient statistical power did not allow the detection of small effects and a small effect on a subgroup of children cannot be ruled out. The absence of effects of added sugars on hyperactivity have largely discredited the sucrose hypothesis of ADHD.

In summary, sucrose does not appear to be the cause of hyperactivity or ADHD, even though children with ADHD may tend to consume sugar-rich foods and beverages. Current evidence is insufficient to recommend a restriction of sucrose consumption to improve behavior or cognitive functioning.

Food additives

In the 1970s, Feingold proposed that food additives, such as artificial food colorants and flavors, can cause adverse behavioral effects, with increased sensitivity to these additives underlying hyperactive behavior in some children. A diet free of dyes, colorings, additives and naturally occurring salicylates (Feingold diet), hypothesized to improve symptoms in children with hyperactivity, became a popular approach. While the findings of initial studies suggested that symptoms of hyperactivity could be reduced through a diet free of artificial food colors and other additives, these early studies lacked control groups. More rigorous, controlled research was less conclusive, with only a small fraction of children with ADHD responding to this kind of intervention. More recent evidence, based on a large, well-designed randomized controlled study, has rekindled interest in the theory that food additives may trigger hyperactivity, especially in younger children. Artificial colors and/or a sodium benzoate preservative were shown to cause increased hyperactivity in three-year-old children, as well as children aged eight and nine, in the general population. The findings of this research led public policy groups to petition food manufacturers and governments in several countries to ban the use of artificial food colors. In response to these demands, the British government mandated the removal of most synthetic dyes from food products, and the European Food Safety Authority introduced regulations requiring that foods containing synthetic colors carry warnings stating that their intake may be associated with adverse effects on activity and attention in children. However, the elimination of food colors appears to be unable to provide a general solution to the multifaceted behavioral problems of hyperactivity and ADHD.

The available studies on the effects of artificial food colors assessed the questions of whether hyperactivity or other symptoms of ADHD improved following a diet excluding artificial colorants and whether the symptoms worsened after exposure to these compounds. Several meta-analyses of these studies concluded that artificial food colors have small, but statistically significant, adverse effects on ADHD symptoms in some children (Nigg et al., 2012; Sonuga-Barke et al., 2013). However, the exclusion of artificial food colorants may be limited to children selected for food sensitivities.

In short, although food additives are not the primary cause of ADHD, they have been considered in recent years to be a minor causative factor in hyperactive behavior in at least some children (Nigg et al., 2012). The elimination of food colorants from the diet would not seem to offer a general remedy for the varied behavioral problems of hyperactivity and ADHD. However, the available evidence carries sufficient weight to warrant further investigation, which should attempt to identify subgroups of those with ADHD who may benefit from the exclusion of food additives.

Few-foods or oligoallergenic diet

As well as artificial food additives and sucrose, various other food items have been hypothesized to elicit allergic reactions or sensitivity in some children diagnosed with ADHD. Foods causing reactions in children, such as subtle behavioral effects or symptoms of ADHD, may vary significantly between children. The hypothesis of a link between specific foods and ADHD has led to the development of the few-foods or oligoallergenic diet.

A strict elimination diet excluding many food items has been found to be a valuable tool in assessing whether symptoms of ADHD are induced by individual foods. The general approach of the few-foods, or oligoallergenic, diet in the identification of foods triggering ADHD is the introduction of a restricted diet, eliminating most food items for a limited period, with the subsequent reintroduction of single foods one at a time. Children responding to the few-foods diet show improvements in behavior or cognitive performance following several weeks on this diet. In those showing a response, food items are added consecutively in a controlled way to determine which foods are causative of adverse reactions or specific behavioral signs and symptoms. This process can be complemented by blinded challenges using disguised forms of foods to which the child had reacted. Finally, an individualized diet is created, eliminating only the foods causing adverse effects.

Significant effects of the few-foods diet, as examined using a RCT design, were found in an unselected sample of children with ADHD. Furthermore, several double-blind placebo-controlled studies assessing the effects of the few-foods diet have consistently demonstrated the role of foods in triggering ADHD, suggesting the existence of a food-related subtype of ADHD (Pelsser et al., 2017). The primary objective of the few-foods approach is the identification of food items provoking symptoms in each individual case and the design of an individually tailored diet for future use.

In summary, the few-foods, or oligoallergenic, diet may offer treatment opportunities in subgroups of children with ADHD. In the Netherlands, the few-foods diet is applied in general practice (Pelsser et al., 2020). Available evidence suggests that this personalized dietary treatment approach deserves further systematic investigation.

Supplementation of food elements

Brain development requires a sufficient supply of nutrients. Diets deficient in essential nutrients may contribute to dysfunction of brain regions thought to be involved in the pathogenesis of ADHD (such as the prefrontal cortex) and to occurrence of symptoms of the disorder. Dietary supplementation of certain macronutrients or micronutrients rests on the assumption that children with ADHD may be deficient in these nutrients, due to low intake or increased requirement. Research in this area has focused mainly on vitamins, minerals, amino acids, and polyunsaturated fatty acids. As various minerals and vitamins serve as cofactors of enzymes required in the synthesis, uptake and metabolism of neurotransmitters in the brain, metabolic dysfunction resulting from a decreased availability of mineral and vitamin cofactors may underlie the symptoms of mental disorders. The potential value of probiotics in the treatment of ADHD has also attracted increasing interest.

Vitamins

The findings of various cross-sectional and longitudinal studies suggest a role of vitamin D in the pathogenesis of mental disorders in childhood and adolescence. Mean 25-hydroxyvitamin D concentration was found to differ significantly between children and adolescents with ADHD, relative to controls. While a meta-analysis of RCTs using vitamin D as adjunctive treatment to methylphenidate in children with ADHD, showed small, statistically significant improvements in behavior, hyperactivity and inattention scores, these effects were limited by the low to very low quality of evidence provided by the available studies.

As B vitamins (B6, B9 and B12) are involved in the synthesis of monoaminergic neurotransmitters and metabolism of fatty acids, deficiencies may impact cognitive function. The anti-oxidative effects of vitamin C improve iron absorption and may reduce oxidation in the brain. Several studies have examined the potential benefits of multivitamin supplementation, with vitamins B3, B5, B6 and C administered in doses many times above the commonly recommended daily allowance, in children with ADHD. While some of these studies found beneficial effects, severe methodological limitations, such as short intervention periods, lack of randomization, and poor statistical analysis, undermine their scientific value. Other well-designed RCTs found no effects. Taken together, available findings provide no evidence that the administration of vitamin mega-doses can ameliorate ADHD symptoms in children. Furthermore, excessive intakes of vitamins B3, B5, B6 and C carry risk of liver toxicity.

Minerals

Possible associations between zinc, magnesium, iron, copper and selenium status and the occurrence of ADHD have been suggested, as has the potential value of supplementation of these minerals in the therapy of ADHD (Robberecht et al., 2020). Zinc is involved in the synthesis of monoaminergic neurotransmitters, and deficiencies in zinc and iron may cause cognitive impairment in children. Observational studies have reported trends toward lower blood concentrations of zinc, magnesium and ferritin in children with ADHD. However, available studies have yielded inconsistent results, potentially resulting from different diagnostic parameters, heterogeneity of studies or variation in daily intake of the minerals. The studies also differed significantly in respect of demographic variables. Moreover, whether differences in mineral status are a cause or a consequence of ADHD is unknown.

Several studies have examined the effects of supplementation with zinc, magnesium and iron. Two double-blind RCTs showed statistically significant beneficial effects of zinc supplementation on ADHD symptoms, with children demonstrating the lowest blood concentrations of zinc and essential fatty acids showing the greatest benefit. A significant effect of iron supplementation was found in a small uncontrolled intervention, while another small RCT showed no effect on ADHD symptoms. Magnesium supplementation showed an improvement in ADHD symptoms, but the evidence was undermined by severe methodological flaws, such as lack of randomization and problematic ADHD symptom assessment.

In summary, the evidence for therapeutic supplementation of minerals in ADHD is insufficient. Zinc supplementation may be considered in individuals with zinc deficiency.

Micronutrient combinations

Several studies have used combinations of various vitamins and minerals in children with ADHD and found improvements in behavioral, cognitive and emotional symptoms (Rucklidge et al., 2019). However, most studies had small sample sizes and were highly heterogeneous in regard to individuals included and treatments.

Amino acids

Amino acids have attracted interest in the context of ADHD management, as the monoaminergic neurotransmitters dopamine, norepinephrine, and serotonin, which are believed to be involved in the pathogenesis of ADHD, are synthesized from the amino acid precursors, phenylalanine, tyrosine, and tryptophan. Supplementation of these amino acids has therefore been proposed as a potentially useful therapy in children with ADHD. A few randomized controlled trials with small sample sizes have examined the effects of amino acid supplementation. However, results showed only small effects of tryptophan, and no effects of tyrosine or phenylalanine on ADHD symptoms. Thus, there is currently no evidence of beneficial effects of amino acid supplementation in children with ADHD.

Polyunsaturated fatty acids

Brain structure and function are critically dependent on an adequate intake of polyunsaturated fatty acids (Lange, 2020b). Both omega-3 fatty acid deficiencies and supplementation with fish oil affect monoaminergic neurotransmitter concentrations in the brain. The role of free fatty acids in the etiology and treatment of ADHD is controversial (Lange, 2018a,b). Early studies using omega-6 fatty acid supplements from vegetable oils gave rise to an interest in fatty acids in ADHD. Since then, the focus of interest in this respect has been the long-chain omega-3 fatty acids, docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA) from fish oil. Maternal fish oil administration during lactation has been associated with better attention and physical activity in the offspring in both animal and human studies. Children with hyperactivity have been found to present with clinical signs of fatty acid deficiency, such as dry skin and hair, as well as polydipsia and polyuria. Furthermore, plasma levels of DHA, EPA and arachidonic acid were found to be significantly decreased in children and adolescents with ADHD. Children with ADHD symptoms have been shown to have increased ratios of omega-6 to omega-3 fatty acids, compared to those without symptoms. This could be due to dietary patterns, alterations in gut microbiota or abnormal fatty acid metabolism. The findings in a population-based birth cohort suggest that maternal diet during pregnancy with a high ratio of omega-6 to omega-3 fatty acids may increase the risk of subclinical ADHD symptoms in the offspring in childhood.

Numerous interventional studies have assessed the effects of omega-3 fatty acid supplementation on ADHD symptoms. Available RCTs using fish oil in children with ADHD have several limitations. Sample sizes were usually small, the dosages of omega-3 fatty acids were low, the intervention periods were short, and multiple testing using many evaluation scales without adequate statistical correction was performed. Meta-analyses of RCTs assessing effects of polyunsaturated fatty acids on ADHD symptoms found statistically significant but small reductions, the clinical significance of which remains to be determined (Sonuga-Barke et al., 2013). A systematic review of meta-analyses of double-blind placebo-controlled trials in children with ADHD found small effect sizes for the supplementation with omega-3 fatty acids, according to ratings of symptoms by parents and teachers (Pelsser et al., 2017). A more recent randomized placebo-controlled trial investigating the effects of omega-3 fatty acids, administered for three months to children and adolescents with ADHD, found that individuals receiving placebo showed significantly greater total reduction in an ADHD rating score than those on omega-3 fatty acids (Cornu et al., 2018). Pooling of these findings with previous results suggested no overall effect of omega-3 fatty acids on ADHD symptoms (Cornu et al., 2018).

In summary, the evidence of efficacy of omega-3 fatty acids in the treatment of the core symptoms of ADHD in children and adolescents is marginal or non-existent. Pre-treatment status regarding polyunsaturated fatty acids may influence supplementation effects, and therapeutically relevant effects may be confined to children with a deficiency of free fatty acids.

Probiotics

Based on the hypothesis that gut microbiota and probiotics may influence brain activity, behavior and mental health, the search for factors involved in the pathogenesis of ADHD has recently included the intestinal microbiota and the gut-brain axis. The findings of animal studies suggest that the gut microbiota play a role in brain development and function and may be involved in neurotransmission and neuronal plasticity. Neuronal, metabolic, endocrine, and immune pathways have been proposed to be involved in the modulation by intestinal microbiota of brain development and function as well as behavior. Disturbances of the developing gut microbiota in early life have been shown to influence neurodevelopment and to be associated with mental health issues later in life. Furthermore, early probiotic intervention may influence the risk of mental disorders later in childhood. Based on the findings of animal studies, the gut microbiome has also been suggested as a potential therapeutic target in ADHD. Individuals diagnosed with ADHD have been found to have a different composition of intestinal microbiota compared to healthy controls, with certain bacteria being more and others being less abundant (Bull-Larsen and Mohajeri, 2019).

Bifidobacteria have protective effects on the intestinal barrier function in the gut and influence the dopamine system by elevating phenylalanine and dopamine concentrations through increased production of cyclohexadienyl dehydratase. Intestinal Bifidobacterium levels have been suggested to play a role in the pathogenesis of ADHD and are influenced by several factors related to an elevated risk of ADHD. These factors include delivery by caesarean section, preterm delivery, and the administration of antibiotics during the first months of life. Breastfeeding has been found to be associated with increased Bifidobacterium and lower prevalence of ADHD. However, varying results concerning levels of Bifidobacterium in people with ADHD preclude its use as a potential biomarker for ADHD (Bull-Larsen and Mohajeri, 2019). Furthermore, the role of the gut microbiota in ADHD is still ill-defined due to various factors, such as small sample sizes, differences in age, sex, and medication of children with ADHD included, as well as heterogeneous microbiome sequence analyses.

Few intervention studies have examined probiotics in ADHD. The effects of a 6-month supplementation of the probiotic strain *Lactobacillus rhamnosus* GG soon after childbirth were assessed after 13 years. In children diagnosed with ADHD, the quantity of Bifidobacteria in the feces was lower than in healthy children. At age 13 years, of 35 children in the placebo group, three were diagnosed with ADHD and two with both ADHD and Asperger syndrome, while none of 40 children in the probiotic group were found to have ADHD. This preliminary evidence suggests a preventive role of probiotics in ADHD. A randomized placebo-controlled 3-month pilot trial, investigating the effect of *Lactobacillus rhamnosus* GG on ADHD symptoms in children and adolescents, found better health-related quality of life, but no clear effects on psychometric parameters following probiotic treatment, compared to placebo controls. The potential role of the microbiome in the pathophysiology of ADHD and as a biomarker of the disorder, as well as of probiotics in the prevention and treatment of ADHD requires further large-scale studies.

In summary, it remains unclear whether vitamin (e.g. B vitamins, vitamin D) or mineral (e.g. zinc, magnesium, iron) deficiencies are involved in the pathophysiology of ADHD. At present, the evidence regarding micronutrient supplementation in the treatment of ADHD is inconclusive. Modest effects on hyperactive behavior may be yielded by the supplementation of free fatty acids, but these may be confined to subgroups of children with ADHD. Potential effects of probiotics in ADHD require further investigation.

Problems of research on nutrition in ADHD

Scientists and laypeople interested in nutrients and their effects on hyperactivity and ADHD usually accept the validity of the concept of ADHD and are seldom acquainted with the problems surrounding the diagnosis and treatment of this disorder. The identification of a role of nutrients in ADHD is hindered by the complex, heterogeneous and ill-defined nature of ADHD as well as the lack of biological markers underpinning the validity of the diagnosis. The validity of ADHD as a diagnostic entity in children and adolescents has been disputed based on inconsistent symptom clustering, lack of biomarkers, high prevalence of comorbid childhood-onset mental disorders and differing cultural perceptions. Particularly in older studies, there is uncertainty regarding the diagnostic classification of ADHD and the psychometric qualities of the assessment tools used. Furthermore, it is often unclear whether those rating behavioral changes were truly blinded (e.g. parents involved in delivering the dietary intervention). The assessment of outcomes of therapeutic interventions in ADHD is commonly based on symptom rating and does not account for measures of adaptive skills, daily functioning in the context of social relationships, education and employment or quality of life. However, these aspects need to be considered when outcome measures with ecological validity beyond the ADHD core symptoms are defined.

Evidence supporting the therapeutic efficacy of single nutrients in ADHD is at present unavailable. Multi-nutrient supplementation may be necessary to achieve beneficial effects on ADHD symptoms. Furthermore, in view of the controversy surrounding the concept of ADHD and the heterogeneity of the disorder in terms of etiology, pathophysiology and clinical presentation, potential therapeutic benefits of nutritional elimination and supplementation could be confined to subpopulations of children with ADHD yet unidentified.

A sufficient supply of certain nutrients is essential during critical phases of brain development, especially during gestation, and dietary supplementation of nutrients after these time windows may be of limited benefit. Controlled interventional studies investigating the effects of single nutrients during childhood and adolescence may fail to account for the requirement of the brain for adequate quantities of various nutritional components at critical times. Nutrients needed during late fetal and early neonatal life include protein, choline, various minerals, and polyunsaturated fatty acids. Prenatal exposure to diets deficient in these essential nutrients has been found to be associated with symptoms of ADHD in childhood. Moreover, low maternal serum vitamin D in the first trimester of pregnancy appear to be associated with ADHD-like symptoms at preschool age.

The effects of unhealthy dietary patterns, such as snack or Western patterns, in children with ADHD are not fully understood. The cumulative benefits of the range of ingredients comprising healthy diets, such as the Mediterranean diet, may result in better outcomes compared to a supplementation of individual nutrients. However, establishing and maintaining adherence to longer term changes in whole diets beyond nutrient supplementation is likely to be a major challenge. Furthermore, associations between adherence to healthy diets and low prevalence of ADHD do not necessarily imply protective effects of healthy foods consumed during childhood. The mothers of children consuming a healthy diet may also have adhered to healthy diets during pregnancy and have provided their children with essential nutritional compounds during critical phases of brain development.

Diet and nutrition interact and overlap with other lifestyle factors, such as physical activity. Children with ADHD may benefit from generally improved lifestyle choices, and the interrelationship between nutrition and lifestyle should play a more prominent role in research on treatment approaches to ADHD (Lange, 2018a).

Conclusion

The current pharmacological treatment of individuals diagnosed with ADHD is unsatisfactory. Insufficient evidence of the effectiveness of commonly used ADHD medications on clinically and ecologically relevant outcome measures as well as the unproven long-term efficacy and potentially problematic adverse effects of drug treatment have led to a search for alternative treatment options, including dietary interventions, some of which attempt to improve ADHD symptoms either through the avoidance of certain food compounds or the supplementation of specific food elements.

Low plasma concentrations of minerals such as zinc, magnesium and iron have been found in children with ADHD at the group level, and mineral supplementation may reduce ADHD symptoms in children with deficiencies. However, evidence in support of this is lacking. The question of whether vitamin deficiencies contribute to the pathogenesis of ADHD is also unclear, and, consequently, the potential value of therapeutic compensation through multivitamin supplementation is undetermined. The role of polyunsaturated fatty acids in the treatment of ADHD is controversial. Furthermore, potential adverse effects of long-term administration of food supplements should be considered. These may become apparent many years following supplementation and therefore elude detection (Lange et al., 2019).

In summary, deficiencies of nutrients including vitamins, minerals and polyunsaturated fatty acids and surpluses of food items such as sucrose and artificial food additives have not convincingly been shown to be involved in the etiology of hyperactivity or ADHD. Furthermore, there is currently no conclusive evidence supporting the efficacy of dietary interventions eliminating or

supplementing single nutrients in reducing hyperactivity or other symptoms of ADHD. However, the few-foods diet provides a diagnostic tool in the identification of children with sensitivity to certain foods. This diet also appears to hold some promise in improving symptoms of ADHD and providing a novel individualized dietary treatment approach. Further large-scale, well-designed studies are needed to demonstrate the utility of dietary interventions alone or in combination with other treatment approaches in ADHD.

References

Bull-Larsen, S., Mohajeri, M.H., 2019. The potential influence of the bacterial microbiome on the development and progression of ADHD. Nutrients 11, 2805.

Cornu, C., Mercier, C., Ginhoux, T., Masson, S., Mouchet, J., Nony, P., Kassai, B., Laudy, V., Berquin, P., Franc, N., Le Heuzey, M.-F., Desombre, H., Revol, O., 2018. A double-blind placebo-controlled randomized trial of omega-3 supplementation in children with moderate ADHD symptoms. Eur. Child Adolesc. Psychiatr. 27, 377–384.

Lange, K.W., Nakamura, Y., Gosslau, A.M., Li, S., 2019. Are there serious adverse effects of omega-3 polyunsaturated fatty acid supplements? J. Food Bioact. 7, 1–6.

Lange, K.W., Nakamura, Y., Reissmann, A., 2022. Diet and food in attention-deficit hyperactivity disorder. J. Future Foods 2.

Lange, K.W., 2017. The treatment of attention deficit hyperactivity disorder has no proven long-term benefits but possible adverse effects. Mov. Nutr. Health Dis. 1, 11–25.

Lange, K.W., 2018a. Lifestyle and attention deficit/hyperactivity disorder. Mov. Nutr. Health Dis. 2, 22–30.

Lange, K.W., 2018b. Do food bioactives play a role in attention-deficit/hyperactivity disorder? J. Food Bioact. 4, 1–7.

Lange, K.W., 2020a. The need for alternative treatments for attention-deficit/hyperactivity disorder. Mov. Nutr. Health Dis. 4, 1–9.

Lange, K.W., 2020b. Lipids in the treatment of mental disorders. In: Shahidi, F. (Ed.), Bailey's Industrial Oil and Fat Products, seventh ed. Wiley, pp. 371–390.

Nigg, J.T., Lewis, K., Edinger, T., Falk, M., 2012. Meta-analysis of attention-deficit/hyperactivity disorder or attention-deficit/hyperactivity disorder symptoms, restriction diet, and synthetic food color additives. J. Am. Acad. Child Adolesc. Psychiatr. 51, 86–97.

Pelsser, L.M., Frankena, K., Toorman, J., Rodrigues Pereira, R., 2017. Diet and ADHD, reviewing the evidence: a systematic review of meta-analyses of double-blind placebo-controlled trials evaluating the efficacy of diet interventions on the behavior of children with ADHD. PLoS One 12, e0169277.

Pelsser, L., Frankena, K., Toorman, J., Rodrigues Pereira, R., 2020. Retrospective outcome monitoring of ADHD and nutrition (ROMAN): the effectiveness of the few-foods diet in general practice. Front. Psychiatr. 11, 96.

Robberecht, H., Verlaet, A.A.J., Breynaert, A., De Bruyne, T., Hermans, N., 2020. Magnesium, iron, zinc, copper and selenium status in Attention-Deficit/Hyperactivity Disorder (ADHD). Molecules 25, 4440.

Rucklidge, J.J., Eggleston, M.J.F., Darling, K.A., Stevens, A.J., Kennedy, M.A., Frampton, C.M., 2019. Can we predict treatment response in children with ADHD to a vitamin-mineral supplement? An investigation into pre-treatment nutrient serum levels, MTHFR status, clinical correlates and demographic variables. Prog. Neuropsychopharmacol. Biol. Psychiatr. 89, 181–192.

Sonuga-Barke, E.J.S., Brandeis, D., Cortese, S., Daley, D., Ferrin, M., Holtmann, M., Stevenson, J., Danckaerts, M., van der Oord, S., Döpfner, M., Dittmann, R.W., Simonoff, E., Zuddas, A., Banaschewski, T., Buitelaar, J., Coghill, D., Hollis, C., Konofal, E., Lecendreux, M., Wong, I.C.K., Sergeant, J., 2013. Nonpharmacological interventions for ADHD: systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments. Am. J. Psychiatr. 170, 275–289.

Storebø, O.J., Faltinsen, E., Zwi, M., Simonsen, E., Gluud, C., 2018. The jury is still out on the benefits and harms of methylphenidate for children and adolescents with attention deficit/hyperactivity disorder. Clin. Pharmacol. Ther. 104, 606–609.

Wolraich, M.L., Wilson, D.B., White, J.W., 1995. The effect of sugar on behavior or cognition in children. A meta-analysis. J. Am. Med. Assoc. 274, 1617–1621.

Hyperlipidemia

Penny M. Kris-Etherton[a], Lisa Sanders[b], Olivia Lawler[a], Terrence Riley[a], and Kevin Maki[b,c,d], [a] Department of Nutritional Sciences, Penn State University, University Park, PA, United States; [b] Midwest Biomedical Research: Center for Metabolic and Cardiovascular Health, Addison, IL, United States; [c] MB Clinical Research, Boca Raton, FL, United States; and [d] Department of Applied Health Science, School of Public Health, Indiana University, Bloomington, IN, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of T.R. Trinick, E.B. Duly, Hyperlipidemia: Overview, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 442–452, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00148-3.

Overview 362
Cholesterol 363
Triacylglycerol 363
Fatty acids 364
Phospholipids 364
Apolipoproteins 364
Apolipoprotein A-I 364
Apolipoprotein A-II 366
Apolipoprotein A-IV 366
Apolipoprotein A-V 366
Apolipoprotein B-48 366
Apolipoprotein B-100 366
Apolipoproteins C-I, C-II and C-III 366
Apolipoprotein E 366
Apolipoprotein (a) 367
Lipoproteins 367
Chylomicrons 367
Chylomicron remnants 367
Very low-density lipoproteins 367
Intermediate-density lipoproteins 368
Low-density lipoproteins 368
High-density lipoproteins 368
Lipoprotein(a) 368
Enzymes and transfer proteins 368
Acyl-coenzyme A:cholesterol O-acyltransferase (ACAT) 368
ATP binding cassette transporter A1 and G1 (ABCA1, ABCG1) 369
Cholesterol ester transfer protein (CETP) 369
Fatty acid binding protein (FABP) 369
Hepatic lipase (HL) 369
HMG CoA reductase 369
Lecithin:cholesterol acyltransferase (LCAT) 369
Lipoprotein lipase (LPL) 370
Microsomal triglyceride transfer protein (MTP) 370
Phospholipid transfer protein (PLTP) 370
Sterol regulatory element binding proteins (SREBPs) 370
Lipoprotein receptors and transporters 370
LDL receptor (LDLR) 370
LDLR-related protein (LRP) 370
Scavenger receptor (SR) family 370
Exogenous (dietary) lipid pathways 371
Endogenous lipid pathways 372
Reverse cholesterol transport 373
Risks associated with hypercholesterolemia and hypertriglyceridemia 373
Benefits of lipid-lowering on ASCVD risk 373
Categories and classifications of hyperlipidemia 374
Causes of hyperlipidemia 375
Genetics 375
Lifestyle 375
Secondary causes 376
Dietary management of hyperlipidemia 377

Healthy dietary patterns 377
Dietary recommendations for reduction of atherogenic cholesterol 377
Dietary management of hypertriglyceridemia 377
Summary **378**
References **378**

Key points

- Lipid and lipoprotein metabolism are exquisitely regulated and there many metabolic pathways that are subject to dysregulation resulting in hypercholesterolemia and hypertriglyceridemia
- Dyslipidemia can be caused by genetics or unhealthy lifestyle practices
- There is a long-established classification system for the familial dyslipidemic phenotypes
- Hyperlipidemia is a major risk factor for CAD
- A healthy lifestyle including a healthy dietary pattern is recommended for the treatment of hypercholesterolemia and hypertriglyceridemia

Overview

Hyperlipidemia is a disorder of lipid metabolism that results in an elevated level of circulating cholesterol and/or triglycerides, also known as triacylglycerols (TG). Because of their insolubility in water, these lipids are transported in the circulation packaged in lipoproteins. Lipoprotein particles are comprised of a hydrophobic lipid core (primarily cholesteryl esters and triacylglycerol) and a hydrophilic surface membrane of polar lipids (unesterified cholesterol and phospholipids) and apolipoproteins (Fig. 1). There seven major lipoprotein classes are differentiated on the basis of size, lipid, and apolipoprotein (Apo) composition:

- Chylomicrons
- Chylomicron remnants
- Very low-density lipoproteins (VLDL)
- Immediate-density lipoproteins (IDL)
- Low-density lipoproteins (LDL)
- High-density lipoproteins (HDL)
- Lipoprotein (a) [Lp(a)]

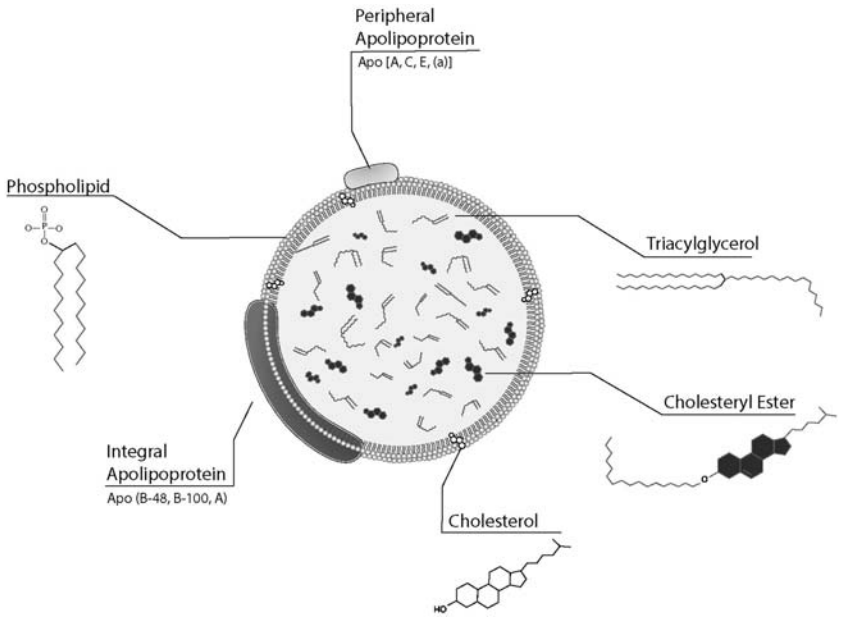


Fig. 1 Lipoprotein particle. Apolipoprotein (Apo).

Lipoproteins are synthesized in the intestine, liver and modified in the plasma, the latter by enzyme modification and removed from the plasma by a receptor-mediated process that is largely regulated by apolipoproteins. The physicochemical characteristics of the main lipoprotein classes are shown in **Table 1**. The different lipoprotein classes have important and distinct physiological functions. Disruptions in their normal metabolic fates result in disorders of lipid and lipoprotein metabolism.

These disorders cause hyperlipidemia that play a central role in the development of atherosclerosis, particularly coronary heart disease (CHD), stroke, peripheral vascular disease, which are the major cardiovascular diseases (CVD). This has been reviewed recently by Feingold (2021).

Cholesterol

Cholesterol is a sterol that is present in all animal tissues. It emanates from diet and endogenous synthesis. Cholesterol is found in the body in two metabolically active pools, free and esterified to fatty acids, largely as free cholesterol in membranes, but in the plasma, it is two-thirds esterified, mainly as cholesteryl linoleate and cholesteryl oleate. Body cholesterol homeostasis is tightly regulated by the interplay between absorption, synthesis, and excretion or conversion of cholesterol into bile acids (Alphonse and Jones, 2016). The rate-limiting step in synthesis is sensitive to cellular levels of cholesterol, which, in turn, are sensitive to plasma cholesterol levels. The feedback regulation occurs through changes in the amount and activity of the enzyme, 3-hydroxy-3-methylglutaryl CoA reductase (HMG-CoA reductase), which catalyzes the formation of mevalonate, the rate-limiting step in cholesterol biosynthesis. The rate of synthesis of HMG-CoA reductase mRNA is controlled by the sterol regulatory element binding protein (SREBP). SREBP in its inactive state is attached to the endoplasmic reticulum or nuclear membrane, but when cholesterol levels decline the amino-terminal domain is released from its association with the membrane by proteolytic cleavage, it migrates to the nucleus and binds to the sterol regulatory element of the reductase gene to enhance transcription. As cholesterol levels increase, the proteolytic release of SREBP is blocked, SREBP in the nucleus is rapidly degraded, and cholesterol synthesis is switched off. Hepatic cholesterol homeostasis is important for the regulation of plasma LDL-cholesterol (LDL-C).

The Niemann–Pick C1-like 1 (NPC1L1) transporter that facilitates the uptake of cholesterol is located in the brush border membrane of enterocytes in the proximal jejunum of the small intestine. Two other transporters, adenosine triphosphate protein binding cassette transporters G5/G8 (ABCG5 and ABCG8), are found in the proximal small intestine. These transporters promote the efflux of unesterified cholesterol from the enterocyte back into the intestinal lumen. The non-effluxed intracellular cholesterol translocates to the endoplasmatic reticulum and subsequently gets esterified by acetyl-coA cholesterol acyltransferase-2 (ACAT2) and then is incorporated into chylomicrons along with TGs, phospholipids, and Apolipoprotein (Apo) B-48 and delivered to the lymph.

Triacylglycerol

Triacylglycerols are glycerol molecules esterified with three fatty acid molecules (**Fig. 1**). Diacylglycerols and monoacylglycerols have two and one fatty acid molecules, respectively. Triacylglycerols constitute the main energy storage form in mammals and are the primary storage form of fatty acids.

Dietary TGs are hydrolyzed by pancreatic lipases in the intestinal lumen to monoglycerides and free fatty acids, which are taken up by enterocytes where they are re-synthesized into TGs and packaged into chylomicrons for secretion into the lymph. The chylomicron surface is covered with a phospholipid monolayer, contains free cholesterol and is enveloped by Apo B-48. In addition to Apo B-48, which is synthesized in the small intestine, there are several other apolipoproteins on the surface of the chylomicron, including Apo AI, Apo AIV, and Apo Cs. Chylomicron fatty acid composition is correlated with dietary fatty acids.

Table 1 Physiochemical characteristics of lipoprotein classes.

Lipoprotein	*Density (g/mL)*	*Size (nm)*	*Major lipids*	*Lipid composition*	*Major apolipoproteins*
Chylomicrons	<0.930	75–1200	TG	98–99	Apo B-48, Apo C, Apo E, Apo A-I, Apo A-II, Apo A-IV
Chylomicron remnants	0.930–1.006	30–80	TG and Cholesterol	92–94	Apo B-48, Apo E
VLDL	0.930–1.006	30–80	TG	90-93[a]	Apo B-100, Apo E, Apo C
IDL	1.006–1.019	25–35	TG and Cholesterol	89	Apo B-100, Apo E, Apo C
LDL	1.019–1.063	18–25	Cholesterol	79	Apo B-100
Lp(a)	1.055–1.100[b]	30	Cholesterol	65	Apo B-100, Apo (a)
HDL	1.063–1.210	5–20	Cholesterol and Phospholipids	43–68	Apo A-I, Apo A-II, Apo C, Apo E

[a]Ratio of TG to cholesterol is 5:1.
[b]Also found in small amounts between 1.100 and 1.21 g/mL.

Fatty acids

Fatty acids are present in TGs, as part of lipoprotein particles, and also as free fatty acids (bound to albumin) in the circulation. Common dietary fatty acids and major food sources are listed in **Table 2**. Fatty acids are straight-chain compounds of differing lengths connecting a hydrocarbon group to a hydroxyl group. Saturated fatty acids (SFA, the most common saturated fatty acid is palmitic acid, C16:0) do not have double bonds; with one or more additional double bonds, the fatty acid is unsaturated. Monounsaturated fatty acids (MUFA) have one double bond (e.g., oleic acid, C18:1, is the major MUFA in the diet). Polyunsaturated fatty acids (PUFA) have two or more double bonds and are classified as either an omega-6 or omega-3 fatty acid based on the position of the first double bond from the methyl terminus. For an omega-6 PUFA the first double bond is on the sixth carbon atom (e.g., linoleic acid, C18:2, is the major PUFA in the diet) whereas for omega-3 PUFA, the first double bond occurs on the third carbon atom from the methyl end (e.g., linolenic acid, C18:3). The presence of a double bond allows for two isomers, depending on whether the hydrogen atoms attached to the carbon atoms on either side of the double bond lie on the same side (cis) or opposing sides (trans). In liquid food oils, the unsaturated fatty acids are cis isomers. Trans isomers are produced by microbial biohydrogenation of unsaturated fatty acids in ruminant animals and also by partial hydrogenation of liquid oils, during food processing. Naturally occurring trans fatty acids are found in foods that come from ruminant animals (e.g., milk, cheese, beef, lamb). Industrially produced trans fatty acids have largely been removed from the food supply since they lost GRAS (Generally Recognized as Safe) status by FDA.

Dietary unsaturated fatty acids decrease risk of CVD whereas SFA increase risk. The biological effects of dietary fatty acids are due to changes in membrane lipid composition, cellular metabolism, intracellular signal transduction, and the regulation of gene expression. Recent evidence has emerged that suggests that some fatty acids may act at the level of the microbiome, which could have important health implications.

Phospholipids

The common phospholipids in plasma are derived from glycerol and consist of TG containing phosphate and a nitrogenous base (glycerophospholipids). The phosphate group is usually attached at position 3 of the glycerol molecule, and the nitrogenous base is usually an amino acid or an alcohol. The phosphatidyl cholines (lecithins) are the most common phospholipid and are found in plasma and in cell membranes. Lecithin–cholesterol acyl transferase (LCAT) catalyzes the transfer of a fatty acyl group at position 2 on glycerol to cholesterol to produce cholesteryl ester and leaves monoacyl glycerophosphate (lysolecithin). Another class of phospholipids, the cephalins, includes phosphatidyl ethanolamine, phosphatidyl serine, and phosphatidyl inositol.

Phospholipids are essential for the structural integrity of the membrane that surrounds lipoprotein particles. Phospholipids form a bridge between nonpolar lipids and water and act to allow the particle to be transported in an aqueous environment (i.e., the plasma). In lipoproteins, the hydrophobic core of TG and cholesteryl esters is surrounded by phospholipids and free cholesterol. The nonpolar hydrocarbon end of the phospholipid is attracted to lipid, whereas the polar phosphate group is attracted to water.

Apolipoproteins

There is a diverse array of apolipoproteins located on the surfaces of the lipoprotein particles, which play crucial roles in the formation, transport and metabolism of lipoproteins. Their major functions are to: (1) play a structural role—they interact with phospholipids to solubilize cholesterol esters and TG; (2) act as ligands for lipoprotein receptors; (3) help with the assembly of lipoproteins; and (4) serve as activators or inhibitors of enzymes involved in lipoprotein metabolism. For example, they are involved in the regulation of enzymes (e.g., lecithin cholesterol acyltransferase (LCAT), lipoprotein lipase (LPL), and hepatic lipase (HL)), and bind with cell surface receptors to initiate the cellular metabolism of lipoproteins or their constituents. Of note, levels of some are associated with CVD risk (e.g., Apo B and Apo C-III).

Apolipoprotein A-I

This is the main structural protein of HDL and accounts for 70% of HDL protein. It is synthesized in both the liver and intestine. Apo A-I has two important roles: (1) it is involved in the regulation of the interaction of HDL with ATP-binding cassette protein A1 (ABCA1), ABCG1, and class B, type I scavenger receptor (SR-B1); and (2) activates LCAT, which is responsible for esterification of free cholesterol in plasma and allows the binding of HDL to many cell surfaces. The production and catabolism of ApoA-I determine the plasma concentration of HDL cholesterol. Elevated levels of Apo A-I are associated with lower CVD risk.

Table 2 Average triacylglycerol fatty acid composition of common foods, fats and oils.

		Saturated (% TG)				*Monounsaturated (% TG)*		*Polyunsaturated (% TG)*				
Fat/oil	*Average TG per 100 g (%)*	*Total*	*Lauric (12:0)*	*Myristic (14:0)*	*Palmitic (16:0)*	*Stearic acid (18:0)*	*Total*	*Oleic (18:1)*	*Total*	*Linoleic (18:2 n−6)*	*α-Linolenic (18:3 n−3)*	*EPA (22:5 n−3)*	*DHA (22:6 n−3)*
Canola oil	100	7	5		2		66	64	27	19	7	0	0
Almond oil	100	9	7		2		73	73	18	18	0	0	0
Coconut oil	100	91	74		3		7	7	2	2	0	0	0
Corn oil	100	14	12		2		30	29	56	55	1	0	0
Butter (dairy fat)	82	70	47		11		26	23	4	3	TRACE	TRACE	TRACE
Flaxseed oil	100	9	6		3		19	19	71	15	56	0	0
Lard (pork fat)	100	41	27		14		47	43	12	11	2	0	0
Olive oil	100	16	13		3		74	72	10	9	1	0	0
Palm oil	100	51	46		5		39	38	10	10	0	0	0
Palm kernel oil	100	86	76		3		12	12	2	2	0	0	0
Cottonseed oil	100	27	25		2		19	18	54	54	0	0	0
Peanut oil	100	17	9		3		61	60	21	21	0	0	0
Safflower oil	100	8	5		2		77	76	15	14	TRACE	TRACE	0
Sesame oil	100	15	9		5		41	41	43	42	TRACE	0	0
Soybean oil	100	16	12		4		23	23	61	54	7	0	0
Grapeseed oil	100	10	7		3		17	17	73	73	TRACE	0	0
Sunflower oil	100	10	5		3		68	68	22	22	TRACE	TRACE	0
Tallow (beef fat)	100	52	31		20		44	38	4	3	1	0	0
Walnuts	65	10	7		3		14	14	76	61	15	0	0
Salmon (Atlantic)	11	22	19		3		37	19	41	6	1	6	13
Herring (Atlantic)	10	26	24		TRACE		47	19	27	2	1	9	11

Percentages are given as approximations rounded to the nearest whole number. (TRACE) <1%. (0) = undetectable amount.
Data from US Department of Agriculture food composition tables: https://fdc.nal.usda.gov/.

Apolipoprotein A-II

Apolipoprotein A-II, which is synthesized in the liver, accounts for approximately 20% of HDL protein. In contrast to Apo A-I, the biological and physiological functions of Apo A-II remain unclear. Clinical and epidemiological studies report conflicting results about the role of Apo A-II and CVD risk.

Apolipoprotein A-IV

Apo A-IV, which is synthesized in the intestine during the absorption of fat, is associated with both chylomicrons and high-density lipoproteins. Many questions remain about the role of Apo A-IV in lipoprotein metabolism.

Apolipoprotein A-V

Apo A-V is synthesized in the liver and is located on triglyceride rich lipoprotein particles (chylomicrons and VLDL), and also on HDL. Apo A-V activates LPL mediated lipolysis and, consequently, affects the metabolism of triglyceride rich lipoproteins, which modulates plasma triglyceride concentrations.

Apolipoprotein B-48

There are two circulating forms of apolipoprotein B, Apo B-48 (which is synthesized in the small intestine) and Apo B-100 (which is produced in the liver). Apo B-48 is the amino-terminus of Apo B-100 and has a molecular mass of 48% of the parent protein, which was the basis for its name. It is the predominant structural protein on chylomicrons and chylomicron remnants. There is a single molecule of Apo B-48 on each chylomicron particle. Importantly, the LDL receptor (LDL-R) does not recognize Apo B-48.

Apolipoprotein B-100

Apo B-100 is the main protein component of VLDL, IDL and LDL, and is synthesized in the liver. Each of these lipoproteins has one molecule of Apo B-100. Apo B-100 is the primary ligand for the LDL-R and is involved in the clearance cholesterol-rich lipoproteins. Elevated Apo B-100 levels increase risk of atherosclerosis.

Apolipoproteins C-I, C-II and C-III

These Apo C particles are synthesized in the liver and all are found in chylomicrons, VLDL and HDL. These C apolipoprotein exchange freely between lipoprotein particles. Notably, they have different biological functions. Apo C-I activates LCAT, which catalyzes the transfer of cholesterol from the surface of HDL to the core of HDL. Apo C-II is a co-factor for lipoprotein lipase (LPL), which hydrolyzes TG in TG-rich lipoproteins to fatty acids that are taken up by cells and the conversion of chylomicrons to chylomicron remnants and VLDL into IDL. Apo C-III is the most abundant form of Apo C and is mainly produced in the liver and to a lesser extent in the intestine. Importantly, plasma Apo C-III levels are positively correlated with plasma TG levels and CVD risk. Clinical and preclinical studies have shown that Apo C-III raises plasma TG levels by inhibiting hepatic TRL clearance, which occurs primarily as a result of its effect to reduce the rate of TG hydrolysis in plasma by both LPL and HL. In addition, Apo C-III inhibits the binding of TG-rich lipoproteins with their receptors. In the Framingham Heart Study Apo C-III was an independent CVD risk factor.

Apolipoprotein E

Apo E is a glycoprotein with three common isoforms Apo E-2, E-3, and E-4. Apo E-3 is the most common isoform. The Apo E isoforms are synthesized in many tissues, but liver and intestine are the major sources. Apo E exchanges between lipoprotein particles and is associated with chylomicrons, chylomicron remnants, VLDL, IDL, and a subgroup of HDL particles. Apo E-3 and E-4 are ligands for the LDL-R, whereas Apo E-2 is poorly bound. Patients who are homozygous for Apo E-2 can develop dysbetalipoproteinemia (type III hyperlipoproteinemia) in which chylomicron remnants and IDL are cleared slowly, which increases risk of CVD. Apo E-4 is a risk factor for the development of CVD and Alzheimer's disease.

Apolipoprotein (a)

Apo(a), when bound to Apo B-100 on an LDL particle results in the creation of a unique lipoprotein referred to as Lp(a). Apo(a) is synthesized in the liver, shares close sequence homology with plasminogen and has been shown to inhibit fibrinolysis. In addition, it also increases the uptake of lipoproteins by macrophages, which initiates and promotes atherosclerosis.

Lipoproteins

Lipoprotein transport and metabolism are highly complex and exquisitely regulated. Each of the seven lipoprotein classes has unique biological functions. An important function is to transport lipids from one organ to another. In addition, there is a transfer of lipids and apolipoproteins, as well as between lipoproteins and lipoprotein classes which affects their functionality (Feingold, 2021). Lipoproteins also function in transporting toxic foreign hydrophobic and amphipathic compounds so that they are cleared from the body.

Chylomicrons

These are the largest lipoproteins, consisting mainly of TG derived principally from the diet via lipid digestion and absorption. These particles contain apolipoproteins A-I, A-II, A-IV, A-V, B-48, C-II, C-III, and E. Apo B-48 is the core structural protein and each chylomicron particle contains one Apo B- 48 molecule. Their major role is transport of dietary TG and cholesterol to peripheral tissues and the liver. Chylomicron size varies depending on the amount of fat consumed. Higher fat intake results in both the formation of larger chylomicron particles and higher postprandial TG levels. Peak chylomicronemia occurs 3–6 h after a meal, with a half-life of less than 1 h, and is cleared from the circulation after a 12 h fast. The lipolytic processes result in smaller particles known as chylomicron remnants (reviewed by Ginsberg et al., 2021).

Chylomicron remnants

Lipolysis of chylomicrons results in the formation of chylomicron remnants that are enriched in cholesterol, have an altered apolipoprotein composition, as well as physicochemical properties. The resulting remnants have retained virtually all the cholesterol contained in their precursors and acquire several molecules of apo E via exchange with HDL. Because of these changes, remnant lipoproteins are rendered highly atherogenic (reviewed by Ginsberg et al., 2021).

Because of their size, chylomicrons and large VLDLs are not taken up by the artery wall. In contrast, smaller chylomicron and VLDL remnants can and do penetrate the artery wall. Unlike chylomicrons and large VLDL, remnant particles have a 2-fold increased affinity for crossing the endothelium and accumulating into the subendothelial space, where they become trapped by subendothelial proteoglycans and may ultimately be engulfed by vascular macrophages. This fuels the atherogenic process.

Very low-density lipoproteins

These TG-rich lipoproteins are produced and secreted mainly by the liver (and the intestine to a limited extent) and contain apolipoprotein B-100, C-I, CII, C-III, and E. VLDL size varies as a function of the amount of TG that is packaged in the particle during hepatic synthesis. Once released in the circulation, the nascent VLDL are transformed into mature VLDL by accumulating cholesterol ester, apo C, and apo E from HDL. Subsequently they undergo delipidation by the actions of lipoprotein lipase and are converted to IDL. Small VLDL is converted into LDL, via IDL, to a greater extent than large VLDL, which is converted to a form of IDL that appears to be removed from the plasma before conversion to LDL (Ginsberg et al., 2021).

Table 3 Classification of HDL.

Method of classification	*Types of HDL*
Density gradient ultracentrifugation	HDL2, HDL3, very high-density HDL
Nuclear magnetic resonance	Large, medium, and small
Gradient gel electrophoresis	HDL 2a, 2b, 3a, 3b, 3c
2-dimensional gel electrophoresis	Pre-beta 1 and 2, alpha 1, 2, 3, 4
Apolipoprotein composition	A-I particles, A-I: A-II particles, A-I: E particles

Source: Feingold (2021).

Intermediate-density lipoproteins

IDLs are intermediate particles formed from the conversion of VLDL to LDL. Also known as VLDL remnants, some are removed directly from plasma, whereas some are converted to LDL.

Low-density lipoproteins

LDL is the major cholesterol-carrying particle in the plasma with apo B-100 being the predominant apolipoprotein. LDLs are derived from VLDL and IDL due to the delipidation process and are further enriched in cholesterol. LDLs consist of a spectrum of particles that range in size. All LDL particles are atherogenic (Ference et al., 2017). Small dense LDL particles are particularly pro-atherogenic because of prolonged retention time in the circulation, which reflects a decreased affinity for the LDL receptor-mediated uptake and removal. In addition, they more easily enter the arterial wall and readily bind to intra-arterial proteoglycans, which trap them in the arterial wall thereby promoting the atherogenic process. Finally, small dense LDLs are very susceptible to oxidation, which is thought to enhance uptake by macrophages followed by foam cell formation, which accelerates atherogenesis. Oxidized LDL in the cell wall stimulates inflammation and the production of cytokines and growth factors, which promote monocyte recruitment and smooth muscle cell proliferation. Moreover, the modified LDL particle is also a ligand for certain receptors (the scavenger receptor and perhaps a specific receptor for oxidized LDL), which contributes to the formation of cholesterol-laden foam cells.

High-density lipoproteins

Nascent HDL is secreted by the liver and gut. There is a family of HDL particles that vary in size and density (see **Table 3**). HDL particles are heterogeneous due to differences in size, lipid, and protein content. Based on ultracentrifugation, there are two main subfractions, HDL2 and HDL3. These can be further sub-fractionated into at least five subpopulations (HDL2b, HDL2a, HDL3a, HDL3b, and HDL3c).

The long-held view that HDL-C is inversely associated with the risk of cardiovascular events stems from several epidemiological studies. There is evidence that HDL particles have an important role in reverse cholesterol transport (RCT) from peripheral tissues to the liver, which may explain its anti-atherogenic property, along with having antioxidant, anti-inflammatory, anti-thrombotic, and anti-apoptotic properties. However, recent epidemiologic studies found that individuals not only with the lowest HDL-C but also those with the highest HDL-C levels have an increased risk of CVD mortality. This "U-shaped" association raises the question of whether a "sweet spot "exists for the HDL-C–CVD relationship, and whether the reasons for the increased association with CVD (and mortality) at both ends of the distribution are based on different characteristics of the HDL particle profile. The idea is evolving that the diversity of HDL subclasses, each perhaps endowed with some specific biological effects, rather than being based solely on their cholesterol content. Also, involvement in cholesterol efflux capacity may account, at least in part, for the U-shaped risk profile. Thus, estimating CVD risk on the basis of HDL-C concentration is limited by the complexity of the HDL family (Casula et al., 2021).

Lipoprotein(a)

Lp(a) is an LDL-like particle containing apolipoprotein (a) linked to Apo B-100 via a disulfide bond. High concentrations of Lp(a) are associated with increased risk of CVD-related events, including myocardial infarction, stroke, and heart failure, as well as valvular aortic stenosis. Lp(a) also has atherothrombotic traits and thrombo-inflammatory traits (Wilson et al., 2019).

Enzymes and transfer proteins

The enzymes presented below all play important roles in lipid and lipoprotein metabolism. Related to dyslipidemia, HMG-CoA reductase and LPL are essential for modulating plasma cholesterol and TG levels, respectively. Many of the perturbations in plasma cholesterol and triglyceride levels occur because of a complex genetic etiology with gene variants that occur in key regulatory steps in the pathways that modulate cholesterol and TG levels. These genetic variants can occur in enzymes, apoproteins and other factors (binding proteins, receptors) that affect cholesterol and TG metabolism.

Acyl-coenzyme A:cholesterol O-acyltransferase (ACAT)

Acyl-coenzyme A:cholesterol O-acyltransferase is an intracellular enzyme that catalyzes the acylation of cholesterol with long-chain fatty acids to form cholesteryl esters (CE). The CE in nascent VLDL are generated by ACAT. There are two isoforms of ACAT. ACAT-1 is expressed predominantly in liver (Kupffer cells), macrophage, adrenal gland and the brain. ACAT-2 is synthesized in the

hepatocytes and enterocytes. ACAT-1 has been associated with adverse effects on the arterial wall and macrophage-derived foam cell formation. An important role of ACAT-2 is for the synthesis of CE used for chylomicron and VLDL production.

ATP binding cassette transporter A1 and G1 (ABCA1, ABCG1)

Both ABC transporters function as a lipid transporter and are essential for RCT. ABCA1 and ABCG1 are critical receptors for the initial step of RCT in atherosclerotic plaques, i.e., cholesterol efflux out of foam cells. The activity ABCA1 in hepatocytes and enterocytes plays a central in plasma HDL production. The transporter mediates the release of phospholipid and free cholesterol from the plasma membrane of these cells to apo A-I in the extracellular medium, thereby producing nascent HDL particles. Thus, ABCA1 is important in the efflux of cholesterol to lipid poor pre-beta Apo A-1 particles. ABCG1 also mediates the efflux of cholesterol from the cell to mature HDL particles and is present in a diverse array of cells.

Cholesterol ester transfer protein (CETP)

CETP facilitates the transfer of esterified cholesterol from anti-atherogenic HDL to pro-atherogenic LDL and VLDL in exchange for triglycerides. The hypothesis that inhibition of CETP activity might prevent CHD was based on the evidence that it both reduces plasma LDL-C concentration, and raises HDL-C. Several large trials evaluating different CETP inhibitors have failed to prevent cardiovascular events even though there was a marked increase in HDL-C. Subsequent studies showed that HDL functionality and reverse cholesterol transport were compromised in response to these inhibitors (Yamashita and Matsuzawa, 2016).

Fatty acid binding protein (FABP)

Fatty acid binding proteins (FABPs) are a family of lipid chaperones that are involved in several lipid signal transduction pathways that regulate lipid trafficking and metabolic and inflammatory pathways. At least nine different FABP isoforms have been identified. FABP4 is primarily expressed in adipocytes and macrophages and plays an important role in insulin resistance and atherosclerosis. FABPs are involved in the solubilization of long-chain fatty acids and their CoA-esters to various intracellular organelles. FABPs serve as intracellular receptors of LCFAs and are involved in ligand-dependent transactivation of peroxisome proliferator-activated receptors in trafficking LCFAs to the nucleus.

Hepatic lipase (HL)

Hepatic lipase (HL) is an endothelial-bound enzyme synthesized and secreted from hepatocytes that catalyzes the hydrolysis of TG and phospholipids in several lipoproteins. HL is involved in the remodeling of apo B-containing lipoproteins and HDL, and in the production of small, dense LDL, as well as small HDL remnants. HL hydrolyzes HDL triacylglycerol and phospholipids to form HDL3 from HDL2, contributing to the process of HDL regeneration in the reverse cholesterol transfer process. Thus, HL plays a role in determining the size, density, and consequent metabolic fate of these lipoprotein particles.

HMG CoA reductase

HMG-CoA reductase is the key (and rate limiting) enzyme in the synthesis of cholesterol. It catalyzes the NADPH-dependent reduction of HMG-CoA to mevalonic acid in the cholesterol synthetic pathway. Statin drugs are HMG-CoA reductase inhibitors, the first line drugs for treating hypercholesterolemia. Clinical studies have shown that they effectively reduce CVD morbidity and mortality in primary and secondary prevention. Maximal reduction of LDL-C is approximately 60% with high dose statin therapy.

Lecithin:cholesterol acyltransferase (LCAT)

Lecithin:cholesterol acyltransferase (LCAT), the only enzyme capable of esterifying cholesterol in plasma, mediates the esterification of cholesterol by transferring a fatty acid from lecithin to cholesterol to form cholesteryl ester. Thus, LCAT is centrally involved the maturation of HDL. Because it maintains an unesterified cholesterol gradient between peripheral cells and extracellular acceptors, LCAT is considered to be a key enzyme in reverse cholesterol transport. The preferred lipoprotein substrate is pre β-HDL, a newly assembled small discoidal HDL. Following interaction with LCAT, pre β-HDL is converted to mature, spherical and α-migrating HDL, which comprises most of plasma HDL.

Lipoprotein lipase (LPL)

Lipoprotein lipase (LPL) is an endothelial-bound enzyme that has long been known to be the rate-limiting step in the removal of TG from TG-rich lipoproteins. Lipoprotein lipase is activated by apoC-II and catalyzes the hydrolysis of TG in chylomicrons and VLDL. It is a biomarker for diagnosing Type I hyperlipidemia.

Microsomal triglyceride transfer protein (MTP)

Microsomal triglyceride transfer protein (MTP) is an endoplasmic reticulum (ER) resident protein in enterocytes and hepatocytes that facilitates the transfer of neutral lipids to nascent apo B. It plays a key role in the assembly and secretion of apo B-containing lipoproteins.

Inhibition of MTP reduces apo B synthesis and, hence, lipoprotein assembly and secretion, and decreases plasma cholesterol levels (LDL-C), apo B and TG by over 50% (Hussain, 2014).

Phospholipid transfer protein (PLTP)

Plasma PLTP catalyzes the transfer of phospholipids from apoB-containing TG-rich lipoproteins to HDL, and also exchanges phospholipids between lipoproteins. In addition, PLTP has been shown to act like a fusion factor to enlarge HDL particles. PLTP is required for efficient PLTP-mediated HDL enlargement (Jiang, 2018).

Sterol regulatory element binding proteins (SREBPs)

Sterol regulatory-element binding proteins (SREBPs) are transcription factors that regulate the promoters of genes involved in cholesterol biosynthesis and the LDL receptor pathway that regulates sterol balance. SREP is activated when cellular cholesterol levels are decreased and LDL receptor gene expression is increased, thereby increasing LDL-C uptake from the plasma. When cellular cholesterol levels are high, SREBP remains inactive.

Lipoprotein receptors and transporters

Lipoprotein metabolism is importantly regulated by different lipoprotein receptors and transporters that, consequently, affect atherogenesis and CVD risk (Mineo, 2020).

LDL receptor (LDLR)

The LDLR is a transmembrane glycoprotein present on the plasma membrane of most cell surfaces. The LDLR binds to Apo B-100 and Apo E, which is the first step in receptor-mediated uptake of LDL, chylomicron remnant and IDL. Following internalization, the particle is degraded in lysosomes and the free cholesterol is released. The accumulation of free cholesterol reduces both cell synthesis of cholesterol and cell uptake of more LDL-C via regulatory steps involving HMG CoA reductase and the SREBPs. LDLR gene expression is regulated by changes in intracellular cholesterol.

LDLR-related protein (LRP)

In addition to LDLR, there are many other structurally and functionally related receptors that have been identified, which include low-density lipoprotein receptor-related protein (LRP)1, LRP5, LRP6, very low-density lipoprotein receptor, and apolipoprotein E receptor 2. In the liver, LRP1 partners with the LDLR in the endocytosis and clearance of cholesterol-rich chylomicron remnants from the circulation. The LRP family is a large and complex array of different proteins that bind to dozens of different ligands that not only affect lipoprotein metabolism but other cellular processes.

Scavenger receptor (SR) family

The scavenger receptors belong to a large family of pattern recognition receptors, which can interact with a wide range of ligands. Interacting with the circulating native or modified lipoproteins, such as oxLDL, these receptors modulate vascular inflammation, lipid accumulation, and plaque formation. Two important SR, CD36 and SR-BI, play a role in atherogenesis in different ways.

CD36 promotes endothelial dysfunction, macrophage foam cell formation, vascular inflammation, and thrombosis. In contrast, SR-B1, a high-affinity HDL receptor mediates the uptake of HDL cholesterol ester into the liver. SR-BI also facilitates the bidirectional flux of free cholesterol (FC) between cells and HDL. Also, hepatic SR-BI has been linked to clearance of proatherogenic lipoproteins VLDL and Lp(a). Thus, hepatic SR-BI is thought to protect against atherosclerosis by promoting RCT and maintaining anti-atherogenic characteristics of HDL and decreasing atherogenic lipoproteins. Lastly, there is compelling evidence in support of atheroprotective functions of macrophage SR-BI.

Exogenous (dietary) lipid pathways

In the small intestine, dietary lipids undergo pancreatic lipase-mediated lipolysis and are emulsified by bile acids. Lipolysis of TG produces 2-monoacyglycerol (2-MG) and fatty acids that are taken up by intestinal enterocytes. Cholesterol ester is hydrolyzed by carboxyl ester hydrolase (cholesterol esterase) to free cholesterol and fatty acids. The free cholesterol is incorporated into micelles, which are primarily comprised of bile acids and lesser amounts of phospholipids, fatty acids and 2-MG. These micelles are transported across the intestinal unstirred water layer to the brush border of the enterocyte where cholesterol, fatty acids and 2-MG are taken up.

In the enterocyte, re-esterification of fatty acids into TG and cholesterol into cholesteryl ester occurs to form chylomicrons, to which is added a surface layer of A-I, A-II, A-IV, A-V, B-48, C-II, C-III, and E, phospholipid, and free cholesterol. This allows secretion of the chylomicron into the intestinal lymphatics. Apo B-48 is required for secretion of the chylomicron.

Chylomicrons in the circulation take up Apo C from HDL (releasing it back to HDL later) and acquire Apo E. Apo C-II allows the chylomicron to activate lipoprotein lipase on capillary endothelial cells of muscle and fat. This promotes hydrolysis of TG, releasing glycerol and fatty acids to be taken up by local tissue. Surface phospholipids, free cholesterol, and Apo C transfer to HDL as the

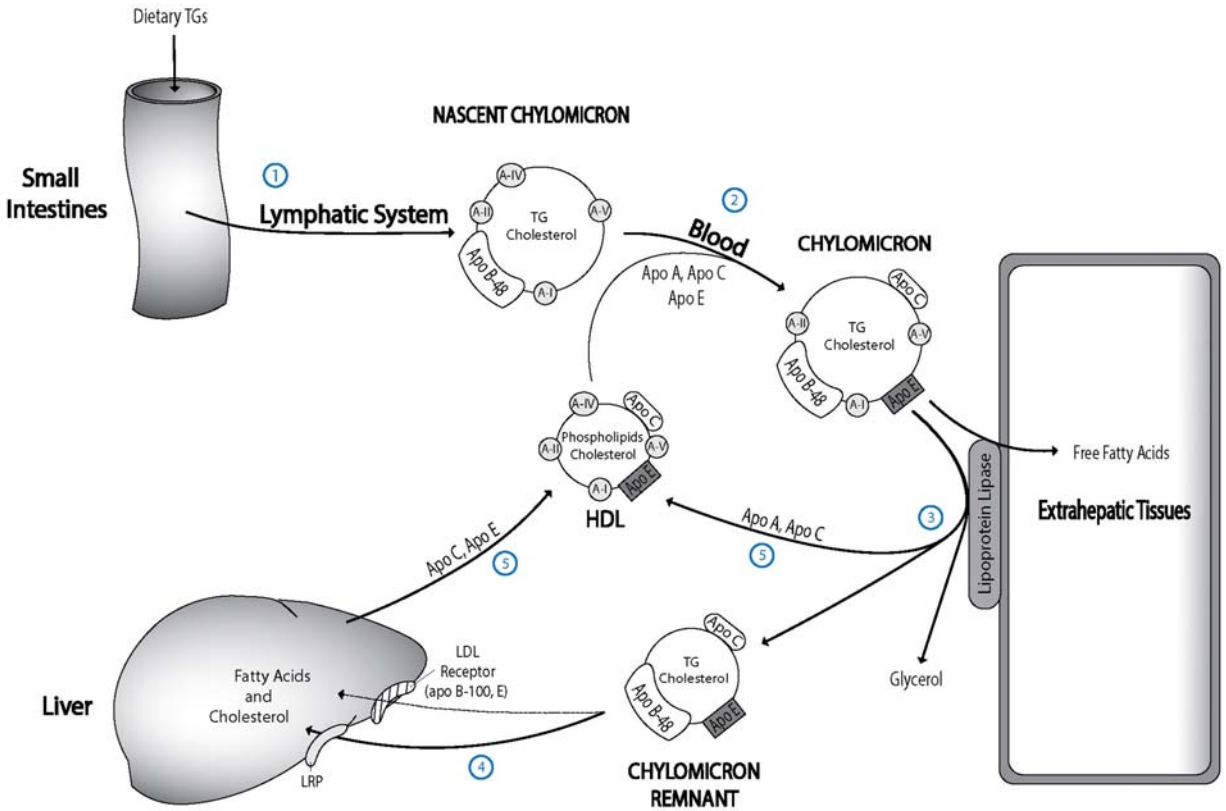


Fig. 2 Exogenous (dietary) lipid pathway. Chylomicrons transport triglycerides from the intestines to peripheral tissues. (1) Lipids are digested and absorbed in the small intestines where chylomicron synthesis occurs on the template of apo B-48. Nascent chylomicrons are released into the lymphatic system before entering circulation. (2) Mature chylomicrons are formed after receiving the full complement of apolipoproteins from HDL. (3) Lipids are made available to peripheral tissues by lipoprotein lipase mediated lipolysis ultimately forming a chylomicron remnant. (4) Removal of chylomicron remnants occurs by LRP and LDL receptor in the liver. (5) Apolipoproteins are recycled through chylomicron remnant—HDL interaction and by the liver. Triacylglycerol (TG). Apolipoprotein (Apo). LDL Receptor-related protein (LRP). Low-density lipoprotein (LDL). High-density lipioprotein (HDL).

particle shrinks thereby becoming a chylomicron remnant. Thereafter the apo E on the chylomicron remnant binds to the hepatic LDLR which catalyzes remnant uptake. This transport of dietary lipid from the intestinal to the peripheral tissues is shown in Fig. 2.

Endogenous lipid pathways

The endogenous lipoprotein pathway is dynamic and complex and involves the synthesis and transport of TG and cholesterol from sites of origin to tissues/organs (Ko et al., 2020). The liver is the main source of endogenous lipid (Fig. 3). In particular, the liver secretes the TG-rich VLDL. Triacylglycerol, which is formed from fatty acids either newly synthesized or taken up from plasma, together with free cholesterol, which is delivered to the liver in chylomicron remnants, join with Apo B and phospholipids to form VLDL. Apo C and Apo E are added in the circulation. Triacylglycerol is progressively hydrolyzed in VLDL (same way as chylomicrons) resulting in the formation of VLDL remnants (IDL). The IDL particles are either removed from the circulation by the liver (via the LDLR) or converted to LDL in the plasma. IDL is then modified through interaction with HL leading to a further decrease in TG which occurs in conjunction with transferring apoproteins to other lipoproteins that results in the formation of LDL (see Fig. 3). The production rate of LDL is inversely related to hepatic LDLR activity (i.e., high receptor activity decreases LDL production and low receptor activity increases production). Circulating LDL is cleared by LDLR-mediated uptake in the liver.

HDL synthesis begins with the synthesis of Apo A-1 by the liver and intestine (Fig. 4). Once Apo A-1 is secreted, it is the nascent HDL and matures by accumulation of cholesterol and phospholipids that come from the liver and intestine and also from chylomicrons and VLDL (during lipolysis which also promotes of transfer of other apolipoproteins to HDL).

Free cholesterol transfers to HDL and is esterified by LCAT (which is activated by Apo A-1). Cholesteryl ester in the core of HDL can be transferred to Apo B-containing particles, via CETP, in exchange for TG transfer to HDL. In this way, VLDL becomes smaller and transforms to become IDL, although some small VLDLs may be removed directly. HL catabolizes TG in HDL resulting in the formation of small HDL particles with the release and degradation of Apo A-1. HDL delivers cholesterol to the liver. Thereafter a small HDL particle is released into the circulation.

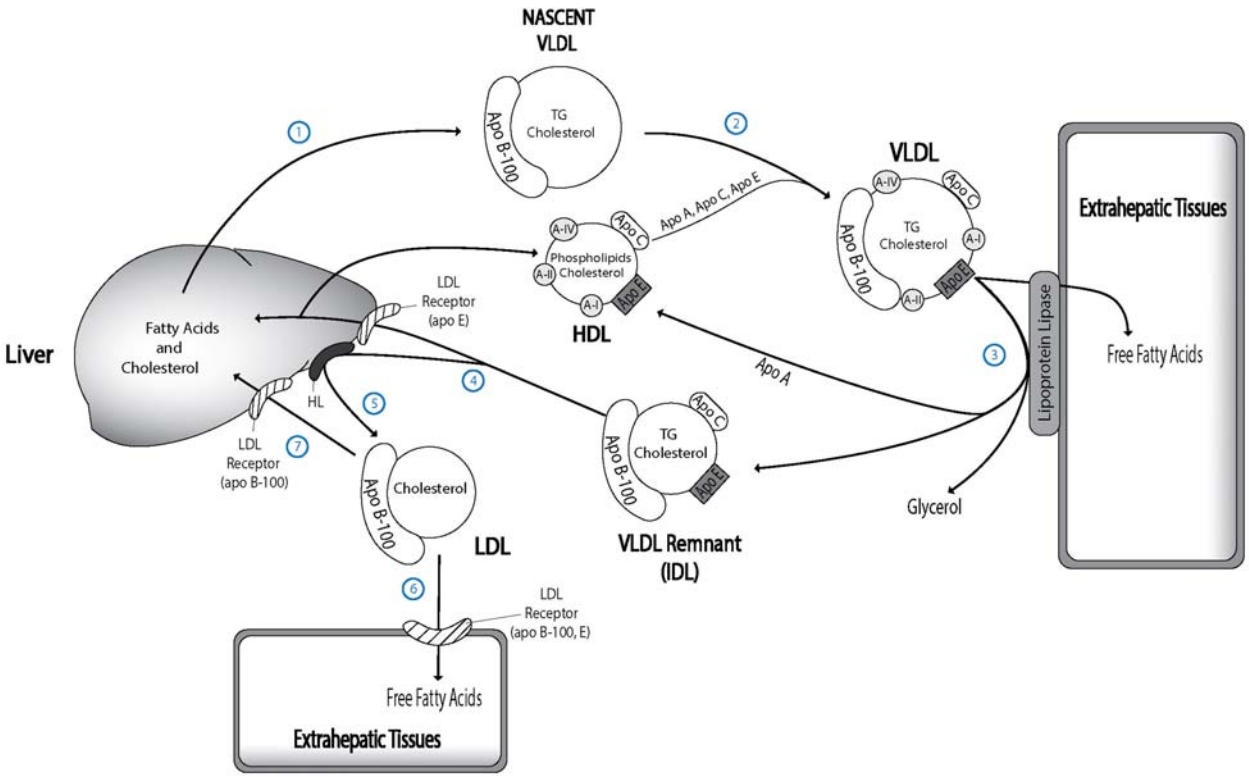


Fig. 3 Endogenous lipid pathway. Transport of TGs and cholesterol from sites of origin (liver) to peripheral tissues/organs. (1) The liver secretes TG-rich, apo B-100 containing, nascent VLDL. (2) Apolipoproteins from HDL are transferred to nascent VLDL to form VLDL. (3) TGs and cholesterol are delivered to peripheral tissues by VLDL decreasing the size which forms VLDL remnants. (4) Further lipolysis of VLDL remnants forms IDL. IDL is either taken up by the liver or (5) is converted to LDL after removal of apo E and apo C. (6) LDL delivers TGs to peripheral tissues or (7) is removed from circulation by LDL receptor in the liver. Triacylglycerol (TG). Apolipoprotein (Apo). Very Low-density lipoprotein (VLDL). High-density lipioprotein (HDL). Intermediate-density lipoprotein (IDL). Hepatic lipase (HL).

Reverse cholesterol transport

RCT is defined as the process by which cholesterol moves out of cells in peripheral tissues (including foam cells in atherosclerotic plaques), enters the circulation, and is excreted in the feces. The CV protective effect of HDL has been attributed to its ability to function as an acceptor of cholesterol from cells and as the cholesterol carrier in the RCT pathway, including delivery to the liver (Fig. 4).

ABCA1 preferentially lipidates small HDL particles, specifically apo A-I to form nascent HDL, ABCG1 stimulates net cholesterol efflux to larger HDL, but not to lipid-poor apo A-I. Efflux of cholesterol to HDL involves both passive diffusion and also active cholesterol transfer; ABCA1, ABCG1 as well as unrelated scavenger receptor class B type 1 (SR-B1) mediate lipid transfer to HDL. After cholesterol transfer to HDL particles, LCAT esterifies the acquired cholesterol to form cholestoryl esters (CE), and, hence, mature HDL. Remodeling of HDL particles occurs via hydrolysis of HDL triglycerides and phospholipids, mediated by hepatic lipase and endothelial lipase, respectively. CE in the HDL core can be transferred to TG-rich lipoproteins by CETP for elimination via hepatic clearance by the LDLR, or taken up by SR-B1, which acts as a hepatic receptor for CE on HDL. Therefore, RCT to the liver of cholesterol derived from peripheral cells involves two routes: (1) direct (HDL-SR-B1); and (2) indirect (HDL-LDL/VLDL-liver LDLR). In the liver, the CE is hydrolyzed and the free cholesterol is either converted to bile acids or transported into the bile for excretion into the feces (Ouimet et al., 2019).

Risks associated with hypercholesterolemia and hypertriglyceridemia

Hyperlipidemia can have multiple clinical manifestations, including atherosclerosis, xanthomas, xanthelasmas, corneal arcus, and pancreatitis. Visible manifestations, such as xanthomas and corneal arcus, are due to lipid buildup in the skin, tendons, and the eye. Acute pancreatitis is a serious risk for individuals with elevated cholesterol and triglycerides carried in chylomicrons and/or VLDL. This is a particular concern for individuals with genetic conditions that impair chylomicron clearance, such as familial chylomicronemia syndrome (FCS).

For most individuals with hyperlipidemia characterized by elevations in LDL-C alone or in combination with hypertriglyceridemia, the primary concern is increased risk of atherosclerotic cardiovascular disease (ASCVD). Atherosclerosis begins when lipoproteins, such as LDL and VLDL remnants, infiltrate artery walls and are modified (e.g., oxidized), triggering an inflammatory response. Macrophages internalize the lipoproteins leading to foam cell formation and fatty lesions. Continued inflammatory signaling promotes a greater immune response and smooth muscle cells are recruited to proliferate and create a barrier between prothrombotic factors in the lesion and procoagulant factors in the blood. This leads to plaque formation and partial occlusion of the artery lumen. If inflammation within the plaque remains unresolved, the plaque may rupture, leading to thrombus formation and possible ischemic events, such as myocardial infarction or stroke. Individual risk for ASCVD events must also consider other major risk factors beyond hyperlipidemia, including age, family history of CHD, smoking, hypertension, and low HDL-C (Jacobson et al., 2015).

Evidence demonstrates that LDL-C and TG (which is highly correlated with VLDL-C) independently and additively increase the risk of CHD and cardiovascular events. A recent study of individuals on statins for cholesterol-lowering reported participants with elevated TG ($\geq$3.9 mmol/L or 150 mg/dL) had a 26% increased risk of cardiovascular events compared to individuals with TG $<$ 3.96 mmol/L (150 mg/dL) (Toth et al., 2021). The relative risk was further increased for individuals with TG 5.2–12.9 mmol/L (200–499 mg/dL). For this reason, measures including both LDL-C and TG-rich VLDL-C, such as Apo B and non-HDL-C, are considered better indicators of ASCVD risk than LDL-C alone, particularly in individuals with elevated TG. In fact, individuals with desirable LDL-C $<$2.6 mmol/L (100 mg/dL) but elevated non-HDL-C ($>$3.4 mmol/L or 130 mg/dL) are at increased risk for cardiovascular events compared to individuals with desirable LDL-C and desirable non-HDL-C (Boekholdt et al., 2012).

Benefits of lipid-lowering on ASCVD risk

There is considerable evidence of reductions in ASCVD risk with reductions in LDL-C; thus, the consensus of experts is that lower is better for LDL-C (Grundy et al., 2019). Further, there is a strong indication that lower LDL-C for a longer period is better for risk reduction of ASCVD. LDL-C-lowering therapies in studies where subjects were followed for $\sim$5 years, resulted in a 22% reduction in CHD risk for every 1 mmol/L (38.7 mg/dL) reduction in LDL-C. In observational studies with longer follow up, from approximately a decade to $>$50 years, the reduction in CHD risk for every 1 mmol/L (38.7 mg/dL) reduction LDL-C increased to 32% (mean follow-up of $\sim$12 years) and 52% (mean follow-up of multiple decades) (Ference et al., 2017).

Lowering of VLDL-C, which can be approximated by TG/5 in those without marked hypertriglyceridemia, can also reduce the risk of ASCVD events. Ference et al. (2019) demonstrated a 0.56 mmol/L (50 mg/dL) reduction in TG, equivalent to a 0.26 mmol/L (10 mg/dL) reduction in VLDL-C, was associated with a similar reduction in CHD risk as a 0.26 mmol/L (10 mg/dL) reduction in LDL-C. These results suggest that Apo B concentration, or non-HDL-C concentration (cholesterol carried by all Apo B particles) is likely to be a better predictor of ASCVD risk than LDL-C or TG levels individually. In fact, after adjustment for Apo B concentration, both LDL-C and TG concentration lose statistical significance as predictors of CHD risk (Ference et al.,

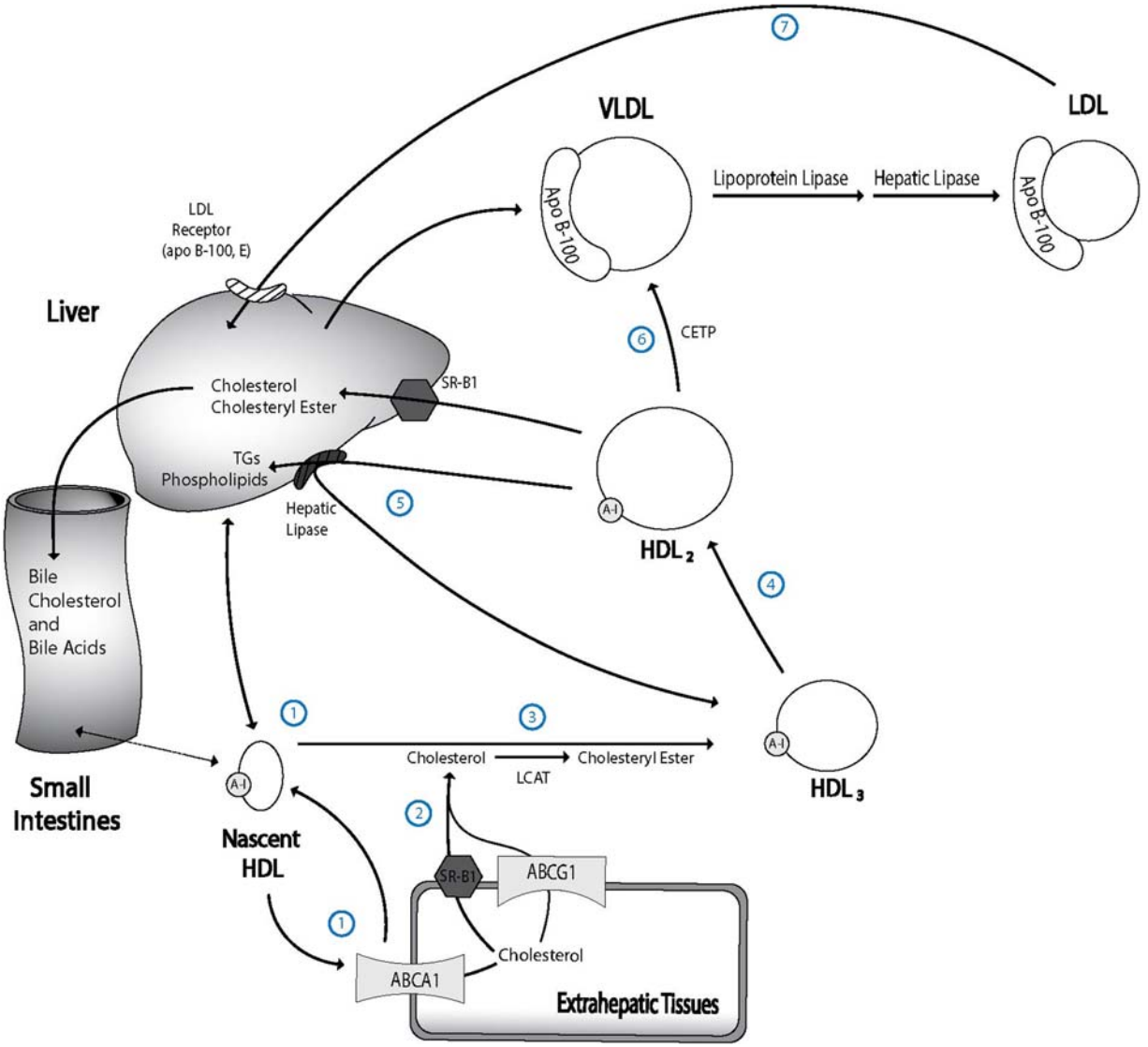


Fig. 4 Reverse cholesterol transport. Removal of peripheral cholesterol from peripheral tissues to the liver and transfer of cholesteryl ester toTG-rich lipoproteins. (1) Nascent HDL formed on Apo A-I derived from the liver and small intestines develop by aggregating cholesterol and phospholipids released by ABCA1 in the intestines, liver, and peripheral tissues. (2) Peripheral SR-B1and ABCG1 facilitate cholesterol transfer from cells into nascent HDL. (3) LCAT esterifies cholesterol in nascent HDL to cholesteryl ester to prevent cholesterol release. (4) The nascent HDL matures to HDL_3 with the accumulation of cholesteryl esters and eventually to HDL_2. (5) Mature HDL_2 is converted back to HDL_3 by depositing cholesteryl esters, phospholipids, and TG in the liver for excretion or lipoprotein synthesis. (6) CETP facilitates exchange of cholesteryl esters in $HDL_2$for TG in TG-rich lipoproteins (VLDL, LDL) which is then removed from circulation by the liver. ATP binding cassette transporter A1/G1 (ABCA1/G1). Scavenger receptor B-1 (SR-B1). Lecithin:Cholesterol acyltransferase (LCAT). Cholesteryl ester transport protein (CETP). Triacylglycerol (TG).

2019). Over periods of ~5 year, each mmol/L reduction in non-HDL-C (LDL-C + VLDL-C) is associated with a reduction in major ASCVD event risk of ~20–25%.

Categories and classifications of hyperlipidemia

The different phenotypes of dyslipidemia were first classified in 1965 by Fredrickson, Levy and Lees (Fredrickson et al., 1965) and are still commonly used today (**Table 4**). This classification was also adopted by the World Health Organization in 1970 (Beaumont et al., 1970). Classifications for LDL-C, triglycerides, non-HDL-C and HDL-C for healthy adults without known ASCVD or other high-risk conditions are included in **Table 5** (Miller et al., 2011; Jacobson et al., 2015).

Causes of hyperlipidemia

Genetics

Primary hyperlipidemia can result from monogenic (one gene) or polygenic (multiple genes) causes. Monogenic causes of hypertriglyceridemia are rare but include FCS or Type I hyperlipoproteinemia, which most often results from loss-of-function mutations in both alleles of the lipoprotein lipase gene, responsible for chylomicron clearance, although defects in apolipoprotein CII and CIII can also manifest as FCS. FCS is usually diagnosed in childhood and risk of acute pancreatitis is of primary concern for these individuals. Acute pancreatitis can lead to systemic inflammation and multiple organ failure; thus, strict dietary fat restriction, which reduces chylomicron formation, is necessary to treat FCS. Interestingly, individuals with FCS do not typically have markedly increased risk for ASCVD, perhaps due to the large size of chylomicron particles which prevent translocation into the arterial wall and subsequent accumulation of atherosclerotic plaques. Type IIa hyperlipoproteinemia, most commonly results from multiple genetic polymorphisms that predispose individuals to elevated LDL-C. More than 50 genetic loci have been identified as possible contributors to polygenic hypercholesterolemia (Warden et al., 2018). However, environmental factors, such as obesity or a high saturated fat diet, are often necessary for hypercholesterolemia to manifest in these individuals.

There are also monogenic causes for inherited or familial hypercholesterolemia (FH) (Warden et al., 2018). These dominantly inherited conditions involve genetic mutations that result in defects in the LDL receptor, apolipoprotein B (apo B), or proprotein convertase subtilisin/kexin type 9 (PCSK9) which is involved in the degradation of LDL receptors. Individuals with heterozygous familial hypercholesterolemia experience high levels of LDL-C, typically from a young age, and are at risk of early ASCVD. Homozygous FH results when two copies of a defective gene are inherited for the LDL receptor, apo B or PCSK9 are inherited, resulting in a severe elevation of LDL-C. Early diagnosis is critical to prevent clinical ASCVD events as well as tendon xanthomas that are often observed in patients with FH.

Type IIb hyperlipoproteinemia, sometimes referred to as familial combined hyperlipidemia (FCH) or mixed dyslipidemia, is characterized by elevated LDL-C and TG (as well as VLDL). The prevalence is estimated to be around 1 in 40 individuals (Brahm et al., 2013). However, the genetics are not completely understood, and it is likely that multiple genetic polymorphisms combined with environmental factors, such as obesity, sedentary lifestyle, and a diet high in saturated fats and added sugars, contribute to the development of FCH. Individuals with FCH have an elevated Apo B level, which is the primary protein in LDL and VLDL. Since each LDL and VLDL particle contains a single Apo B molecule, the circulating Apo B concentration is a direct measure of the concentration of potentially atherogenic lipoprotein particles. It should be noted that chylomicron remnants make a quantitatively small contribution to the circulating Apo B concentration. Each chylomicron particle contains a single molecule of a truncated form of Apo B that contains 48 amino acids, whereas there are 100 amino acids in the Apo B contained in LDL and VLDL particles. Individuals with FCH are at increased risk of ASCVD, thus early diagnosis and management is important.

Familial dysbetalipoproteinemia, or Type III dyslipidemia, is a rare form of hyperlipidemia occurring in individuals homozygous for the Apo E2 genotype. The prevalence is estimated at 1 in 5000 to 1 in 10,000 people, although the prevalence of homozygosity for Apo E2 is roughly 1%. In Type III dyslipidemia, there is a decrease in clearance of VLDL and chylomicron remnant lipoprotein particles because the Apo E2 variant does not interact efficiently with hepatic receptors. The result is elevated cholesterol and TG because TG-rich particles are not efficiently converted to smaller LDL particles. Most individuals with the Apo E2/E2 genotype do not develop marked dyslipidemia in the absence of a secondary condition, such as diabetes, obesity, or hypothyroidism. Palmar xanthomas, identified as orange lipid deposits in the plantar creases of the hands, may occur as a clinical manifestation of this condition. Tuberoeruptive xanthomas may also be found in joint areas, such as elbows and knees. These individuals are also at high risk of ASCVD.

Type IV dyslipidemia is the most common hypertriglyceridemia and is most often polygenic (Cruz-Bautista et al., 2021). Typically, there is a defect in the catabolism of VLDL particles leading to elevated blood TG. Unlike FCH, individuals with Type IV dyslipidemia do not have elevations in LDL-C. When individuals with Type IV dyslipidemia experience other environmental factors, such as obesity or poorly controlled type 2 diabetes, TG levels can exceed 5.6 mmol/L (500 mg/dL) and increase the risk of acute pancreatitis, particularly when above 8.5 mmol/L (750 mg/dL), which is the TG concentration above which fasting chylomicronemia generally manifests.

Type V hyperlipoproteinemia seems to result from a similar genetic predisposition as individuals with Type IV, but with additional polygenic and environmental factors that contribute to an elevation in chylomicron particles in addition to VLDL elevation. Unlike FCS, Type V hyperlipoproteinemia does not typically manifest until adulthood and often in response to secondary conditions, such as poorly controlled diabetes or excess weight gain. Extremely high TG levels increase the risk of pancreatitis in these individuals. In both Type IV and Type V dyslipidemia, ASCVD risk may be elevated, but this is generally true for the subset of such individuals who have elevated levels of Apo B.

Lifestyle

Lifestyle factors, particularly an unhealthy diet, elevated body weight, physical inactivity and smoking/tobacco use, as well as poor sleep are contributors to hyperlipidemia and numerous randomized controlled trials have shown the benefit of lifestyle therapies in reducing hyperlipidemia and risk of ASCVD (Jacobson et al., 2015; Lloyd-Jones, 2022).

Dietary components, particularly fats and carbohydrates have been implicated in hyperlipidemia. Saturated and *trans* fatty acids, are contributors to elevations in LDL-C and dietary cholesterol can also raise LDL-C. Unsaturated fats and viscous dietary fibers

lower LDL-C. Excess body weight can also contribute to dyslipidemia (higher LDL-C and TG, lower HDL-C) and weight loss improves the lipid profile (Jensen et al., 2014).

Physical activity also has important effects on the lipid profile, generally lowering the TG level and raising the HDL-C concentration, particularly if accompanied by weight loss (Kraus et al., 2002). Smoking is associated with elevations in TG and reductions HDL-C (Maeda et al., 2003). Smoking is a strong risk factor for ASCVD, in part because of its association with other risk factors associated such as endothelial dysfunction and arterial stiffness, as well as its effects on the lipid profile. It has been estimated that 40% of all smoking-related deaths are due to CVD (Athyros et al., 2013).

Secondary causes

Hyperlipidemia can result from various conditions, such as metabolic diseases, excess alcohol consumption, or from certain medications. In these situations, often treating the primary condition or altering medications can restore normal lipid levels. For example, individuals with hypothyroidism often have an elevated LDL-C due to a reduction in LDL catabolism. Treating the hypothyroidism may reduce LDL-C. Similarly, individuals with uncontrolled diabetes mellitus or obesity can experience hypertriglyceridemia which may resolve when glycemia and body weight are controlled. Excessive alcohol consumption can also increase TG which can be decreased by reducing or abstaining from alcohol intake.

Several medications can alter lipoprotein lipid levels. For example, postmenopausal women on estrogen therapy may experience hypertriglyceridemia that improves when estrogen is stopped. Medications used to treat hypertension, such as diuretics and beta-blockers, may also increase LDL-C and/or TG in some individuals. Anti-viral therapies, immunosuppressants, and retinoid medications used to treat skin disorders have all been reported to increase blood lipids in some individuals. For individuals with these

Table 4 Fredrickson, Levy and Lees classification of hyperlipoproteinemias.

Type	*Elevated blood lipids*	*Elevated lipoproteins*	*Clinical manifestations*
I	Triglycerides	Chylomicrons	Eruptive xanthomas, pancreatitis
II-a	Cholesterol	LDL	Tuberous and/or tendon xanthomas, coronary atherosclerosis
II-b	Cholesterol and possibly triglycerides	LDL and VLDL	Coronary atherosclerosis, xanthelasmas, corneal arcus
III	Cholesterol and triglycerides	Chylomicron remnants and VLDL	Palmar, tuberoeruptive xanthomas, coronary atherosclerosis
IV	Triglycerides	VLDL	Coronary atherosclerosis (if also elevated Apo B)
V	Cholesterol and triglycerides	Chylomicrons, VLDL	Eruptive xanthomas, pancreatitis

Data sources: Fredrickson et al. (1965), Beaumont et al. (1970).

Table 5 Classifications for normal and hyperlipidemia.

LDL-C	*mmol/L*	*mg/dL*
Desirable	<2.6	<100
Above desirable	2.6–3.3	100–129
Borderline high	3.4–4.1	130–159
High	4.2–4.9	160–189
Very high (severe)	≥5.0	≥190
Triglycerides		
Optimal	<1.1	<100
Normal	<1.7	<150
Borderline high	1.7–2.2	150–199
High	2.3–5.5	200–499
Very high (severe)	≥5.6	≥500
Non-HDL-C		
Desirable	<3.4	<130
Above desirable	3.4–4.1	130–159
Borderline high	4.2–4.9	160–189
High	5.0–5.6	190–219
Very high (severe)	≥5.7	≥220
HDL-C		
Low	<1.3 (women) <1.0 (men)	<50 (women) <40 (men)

Data sources: Miller et al. (2011), Jacobson et al. (2015).

situations, a risk-benefit analysis should be undertaken to determine if medication use should continue or if there may be alternative medications that could treat the condition without adversely altering blood lipids.

Dietary management of hyperlipidemia

Healthy dietary patterns

Clinical guidelines consistently recommend a healthy dietary pattern for the management of hyperlipidemia and prevention of CVD (Eckel et al., 2014; Jacobson et al., 2015; Arnett et al., 2019; Visseren et al., 2021; Lichtenstein et al., 2021). A healthy dietary pattern emphasizes consumption of fruits, vegetables, whole grains, nuts, seeds, legumes and non-tropical oils, and minimizes intake of saturated fat, *trans* fat, sodium, and added sugars, consistent with the Dietary Guidelines for Americans (US Department of Agriculture and US Department of Health and Human Services, 2020). The recommendations also emphasize incorporating more plant foods and lean sources of protein in the diet and reducing intake of processed or high-fat animal products. Consumption of fatty fish (containing omega-3 fatty acids) at least once or twice a week is often recommended. It is important to ensure the diet meets caloric needs, but in individuals with overweight or obesity, caloric restriction may be necessary to achieve weight loss and maintenance. These recommendations are reflective of dietary patterns such as Dietary Approaches to Stop Hypertension (DASH) and Mediterranean diet, which are associated with reduced risk of CVD (Eckel et al., 2014; Jacobson et al., 2015; Arnett et al., 2019; Visseren et al., 2021; Lichtenstein et al., 2021).

Dietary recommendations for reduction of atherogenic cholesterol

In addition to a healthy dietary pattern, there are additional dietary modifications that have been shown to improve LDL-C and non-HDL-C levels. While healthy dietary patterns recommended by the Dietary Guidelines for Americans suggest <10% of calories from saturated fat, for individuals with elevated LDL-C it is recommended to target <7% of calories from saturated fat (Eckel et al., 2014; Jacobson et al., 2015). Replacing calories from saturated fat with carbohydrate, protein, monounsaturated fatty acids, or polyunsaturated fatty acids all contribute to a reduction in LDL-C. However, the effect on overall blood lipids is more favorable when saturated fat is replaced with polyunsaturated fatty acids (Sacks et al., 2017).

Dietary cholesterol has been shown to raise LDL-C, but the effect is modest with every 100 mg of dietary cholesterol raising LDL-C by approximately 0.05 mmol/L (2 mg/dL) (Jacobson et al., 2015). For this reason, some recommendations do not include specific dietary restrictions of cholesterol since following a healthy dietary pattern low in saturated fat should also achieve a reduction in dietary cholesterol (Eckel et al., 2014; Visseren et al., 2021; Lichtenstein et al., 2021).

For individuals who have overweight or obesity, caloric restriction combined with physical activity to achieve weight loss is recommended to reduce LDL-C. A loss of 5% of body weight can reduce LDL-C by 3%–5% (Jacobson et al., 2015).

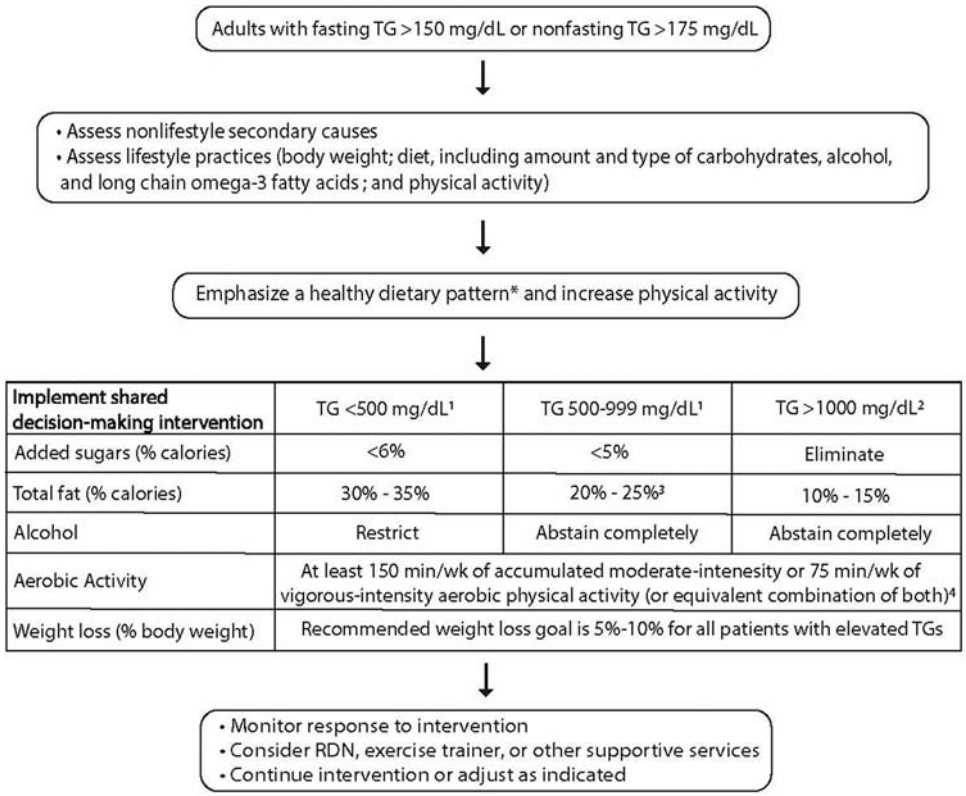
Dietary adjuncts may be useful when a healthy dietary pattern low in saturated fat, trans fat and cholesterol and weight loss, if indicated, are insufficient. These include 5–10 g/day of viscous dietary fibers and 2 g/day plant sterols/stanols. Randomized controlled trials have shown these dietary components can each lower LDL-C by 4%–10%.

A combination of dietary approaches can achieve substantial reductions in LDL-C possibly eliminating the need for pharmacological interventions. For example, achieving a diet low in saturated fat in combination with 5% weight loss could lower LDL-C by 8%–15%. The portfolio diet approach which combines a diet low in saturated fat with viscous fiber, plant sterols/stanols, soy protein and nuts has been shown to reduce LDL-C up ~30% in controlled settings and by 15% in free-living subjects (Jacobson et al., 2015).

Dietary management of hypertriglyceridemia

There are 4 main categories of hypertriglyceridemia (HTG), Borderline: between 150 and 199 mg/dL, High: 200–499 mg/dL, Very High: 500–999 mg/dL and Severe: >1000 mg/dL with chylomicronemia (Virani et al., 2021). Diet is the first line of treatment for elevated triglycerides. Recommendations are based on the TG level (Virani et al., 2021). For all individuals with elevated triglycerides, a weight loss of 5–10% is recommended, if indicated. This reduction in weight can lead to a 10–20% decrease in triglyceride levels in most patients, with some patients experiencing a reduction in triglycerides up to 70% (Virani et al., 2021).

Fig. 5 describes dietary recommendations for individuals with elevated triglycerides. If triglyceride levels are <500 mg/dL added sugars (jams, jellies, and desserts) should be less than 6% of total calories, total fat is ideally between 30 and 35% of calories, and alcohol intake should be restricted. An emphasis on vegetables, legumes, fatty fish at least twice a week, and whole grains is encouraged. For individuals with triglyceride levels between 500 and 999 mg/dL added sugars should be less than 5% of energy, total fat should be between 20 and 25%, and alcohol should be eliminated. Vegetables, legumes, fatty or lean fish twice a week, and whole grains are emphasized. For individuals with triglyceride levels >1000 mg/dL added sugars and alcohol should be eliminated, and total fat should comprise only 10–15% of calories. Vegetables, legumes, lean fish, and whole grains are encouraged (Virani et al., 2021). Dietary modifications (that include alcohol restriction or abstinence) can lower triglycerides by over 70%, although, responses may vary depending on the initial triglyceride level, as well as dietary adherence. Physical activity is important for all patients HTG. Exercise can lower triglycerides by up to 30%, however, the response depends on the type, duration, and intensity of the activity (Virani et al., 2021).



Implement shared decision-making intervention	TG <500 mg/dL[1]	TG 500-999 mg/dL[1]	TG >1000 mg/dL[2]
Added sugars (% calories)	<6%	<5%	Eliminate
Total fat (% calories)	30% - 35%	20% - 25%[3]	10% - 15%
Alcohol	Restrict	Abstain completely	Abstain completely
Aerobic Activity	At least 150 min/wk of accumulated moderate-intenesity or 75 min/wk of vigorous-intensity aerobic physical activity (or equivalent combination of both)[4]		
Weight loss (% body weight)	Recommended weight loss goal is 5%-10% for all patients with elevated TGs		

Fig. 5 Recommendations for lifestyle interventions in patients with increasing levels of weight loss and effects on triglycerides. Registered Dietitian Nutritionist (RDN). Triglyceride (TG).

*Recommendations for a healthy dietary pattern: vegetables; fruits; legumes; nuts; whole grains; and fish/seafood (other healthy proteins such as low-fat dairy, low-fat poultry); liquid plant-based oils; and replacing saturated fatty acids with monounsaturated fatty acids and polyunsaturated fatty acids. Recommendations also emphasize limiting: red and processed meats; refined carbohydrates; added sugars (sweets and sugar-sweetened beverages); sodium and dietary cholesterol; and avoiding trans fats.

[1]RDN referral advised.

[2]RDN referral necessary.

[3]Clinicians may opt to reduce total fat as percent of calories in some of these patients to 10%–15% (examples include those with a history of pancreatitis or those at the higher end of this range).

[4]Although clinicians should aim for their patients to meet the guideline-recommended goals for physical activity, any amount of physical activity is likely beneficial in sedentary individuals and should therefore be encouraged to reduce cardiometabolic risk. Source: Virani et al. (2021).

Summary

Dyslipidemia is a major risk factor for CAD. Metabolism of lipoproteins is a complex biological system that can be disrupted at many regulatory steps by genetics or lifestyle leading to hypercholesterolemia and/or hypertriglyceridemia, thereby increasing CAD risk. The resulting dyslipidemias present in a variety of clinical phenotypes that have been classified by Frederickson, Levy and Lees. A healthy lifestyle, including a recommended dietary pattern, is the primary therapy for managing dyslipidemia and CAD risk.

References

Alphonse, P.A., Jones, P.J., 2016. Revisiting human cholesterol synthesis and absorption: the reciprocity paradigm and its key regulators. Lipids 51 (5), 519–536. https://doi.org/10.1007/s11745-015-4096-7.

Arnett, D.K., et al., 2019. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines. Circulation 140 (11), e596–e646.

Athyros, V.G., et al., 2013. Effect of tobacco smoking and smoking cessation on plasma lipoproteins and associated major cardiovascular risk factors: a narrative review. Curr. Med. Res. Opin. 29 (10), 1263–1274.

Beaumont, J.L., et al., 1970. Classification of hyperlipidaemias and hyperlipoproteinaemias. Bull. World Health Organ. 43 (6), 891–915.

Boekholdt, S.M., et al., 2012. Association of LDL cholesterol, non–HDL cholesterol, and apolipoprotein B levels with risk of cardiovascular events among patients treated with statins: a meta-analysis. JAMA 307 (12), 1302–1309.

Brahm, A., et al., 2013. Hypertriglyceridemia. Nutrients 5 (3), 981–1001.

Casula, M., Colpani, O., Xie, S., Catapano, A.L., Baragetti, A., 2021. HDL in atherosclerotic cardiovascular disease: in search of a role. Cells 10 (8), 1869. https://doi.org/10.3390/cells10081869.

Cruz-Bautista, I., et al., 2021. Familial hypertriglyceridemia: an entity with distinguishable features from other causes of hypertriglyceridemia. Lipids Health Dis. 20 (1), 14.

Eckel, R.H., et al., 2014. 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. J. Am. Coll. Cardiol. 63 (25, Part B), 2960–2984.

Feingold, K.R., January 19, 2021. Introduction to lipids and lipoproteins. In: Feingold, K.R., Anawalt, B., Boyce, A., Chrousos, G., de Herder, W.W., Dhatariya, K., Dungan, K., Grossman, A., Hershman, J.M., Hofland, J., Kalra, S., Kaltsas, G., Koch, C., Kopp, P., Korbonits, M., Kovacs, C.S., Kuohung, W., Laferrère, B., McGee, E.A., McLachlan, R., Morley, J.E., New, M., Purnell, J., Sahay, R., Singer, F., Stratakis, C.A., Trence, D.L., Wilson, D.P. (Eds.), Endotext. MDText.com, Inc, South Dartmouth (MA).

Ference, B.A., Ginsberg, H.N., Graham, I., Ray, K.K., Packard, C.J., Bruckert, E., Hegele, R.A., Krauss, R.M., Raal, F.J., Schunkert, H., Watts, G.F., Borén, J., Fazio, S., Horton, J.D., Masana, L., Nicholls, S.J., Nordestgaard, B.G., van de Sluis, B., Taskinen, M.R., Tokgözoglu, L., Landmesser, U., Laufs, U., Wiklund, O., Stock, J.K., Chapman, M.J., Catapano, A.L., 2017. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. Eur. Heart J. 38 (32), 2459–2472. https://doi.org/10.1093/eurheartj/ehx144.

Ference, B.A., et al., 2019. Association of triglyceride-lowering LPL variants and LDL-C-lowering LDLR variants with risk of coronary heart disease. J. Am. Med. Assoc. 321 (4), 364–373.

Fredrickson, D.S., et al., 1965. A system for phenotyping hyperlipoproteinemia. Circulation 31, 321–327.

Ginsberg, H.N., Packard, C.J., Chapman, M.J., Borén, J., Aguilar-Salinas, C.A., Averna, M., Ference, B.A., Gaudet, D., Hegele, R.A., Kersten, S., Lewis, G.F., Lichtenstein, A.H., Moulin, P., Nordestgaard, B.G., Remaley, A.T., Staels, B., Stroes, E.S.G., Taskinen, M.R., Tokgözoğlu, L.S., Tybjaerg-Hansen, A., Stock, J.K., Catapano, A.L., 2021. Triglyceride-rich lipoproteins and their remnants: metabolic insights, role in atherosclerotic cardiovascular disease, and emerging therapeutic strategies-a consensus statement from the European Atherosclerosis Society. Eur. Heart J. https://doi.org/10.1093/eurheartj/ehab551 ehab551.

Grundy, S.M., et al., 2019. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines. Circulation 139 (25), e1082–e1143.

Hussain, M.M., 2014. Intestinal lipid absorption and lipoprotein formation. Curr. Opin. Lipidol. 25 (3), 200–206. https://doi.org/10.1097/MOL.0000000000000084.

Jacobson, T.A., et al., 2015. National lipid association recommendations for patient-centered management of dyslipidemia: part 1—full report. J. Clin. Lipidol. 9 (2), 129–169.

Jensen, M.D., et al., 2014. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the obesity society. J. Am. Coll. Cardiol. 63 (25, Part B), 2985–3023.

Jiang, X.C., May 2018. Phospholipid transfer protein: its impact on lipoprotein homeostasis and atherosclerosis. J. Lipid Res. 59 (5), 764–771. https://doi.org/10.1194/jlr.R082503.

Ko, C.W., Qu, J., Black, D.D., Tso, P., March 2020. Regulation of intestinal lipid metabolism: current concepts and relevance to disease. Nat. Rev. Gastroenterol. Hepatol. 17 (3), 169–183. https://doi.org/10.1038/s41575-019-0250-7.

Kraus, W.E., et al., 2002. Effects of the amount and intensity of exercise on plasma lipoproteins. N. Engl. J. Med. 347 (19), 1483–1492.

Lichtenstein, A.H., et al., 2021. 2021 Dietary guidance to improve cardiovascular health: a scientific statement from the American Heart Assoication. Circulation 144 (23), e472–e487.

Lloyd-Jones, D.M., et al., 2022. LIfe's essential 8: updating and enhancing the American Heart Association's construct of cardiovascular health: a presidential advisory from the American Heart Association. Circulation 146 (5), e18–e43.

Maeda, K., et al., 2003. The effects of cessation from cigarette smoking on the lipid and lipoprotein profiles: a meta-analysis. Prev. Med. 37 (4), 283–290.

Miller, M., et al., 2011. Triglycerides and cardiovascular disease. Circulation 123 (20), 2292–2333.

Mineo, C., June 1, 2020. Lipoprotein receptor signalling in atherosclerosis. Cardiovasc. Res. 116 (7), 1254–1274. https://doi.org/10.1093/cvr/cvz338.

Ouimet, M., Barrett, T.J., Fisher, E.A., May 10, 2019. HDL and reverse cholesterol transport. Circ. Res. 124 (10), 1505–1518. https://doi.org/10.1161/CIRCRESAHA.119.312617.

Sacks, F.M., et al., 2017. Dietary fats and cardiovascular disease: a presidential advisory from the American Heart Association. Circulation 136 (3), e1–e23.

Toth, P.P., et al., 2021. Real-world analyses of patients with elevated atherosclerotic cardiovascular disease risk from the Optum Research Database. Future Cardiol. 17 (4), 743–755.

US Department of Agriculture and US Department of Health and Human Services, 2020. Dietary Guidelines for Americans, 2020–2025.

Virani, S.S., Morris, P.B., Kris-Etherton, P.M., Ladden-Stirling, A.B., Miller, M., et al., 2021. 2021 ACC expert consensus decision pathway on the management of ASCVD risk reduction in patients with persistent hypertriglyceridemia. J. Am. Coll. Cardiol. 78 (9), 960–993.

Visseren, F.L.J., et al., 2021. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice: developed by the Task Force for cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies with the special contribution of the European Association of Preventive Cardiology (EAPC). Eur. Heart J. 42 (34), 3227–3337.

Warden, B.A., et al., 2018. Familial hypercholesterolemia: genes and beyond. In: Feingold, K.R., Anawalt, B., Boyce, A. (Eds.), Endotext. South Dartmouth. MDtext.com, Inc, MA.

Wilson, D.P., Jacobson, T.A., Jones, P.H., Koschinsky, M.L., McNeal, C.J., Nordestgaard, B.G., Orringer, C.E., 2019 May–Jun. Use of Lipoprotein(a) in clinical practice: a biomarker whose time has come. A scientific statement from the National Lipid Association. J. Clin. Lipidol. 13 (3), 374–392. https://doi.org/10.1016/j.jacl.2019.04.010.

Yamashita, S., Matsuzawa, Y., 2016. Re-evaluation of cholesteryl ester transfer protein function in atherosclerosis based upon genetics and pharmacological manipulation. Curr. Opin. Lipidol. 27 (5), 459–472. https://doi.org/10.1097/MOL.0000000000000332.

Hyperlipidemia: Nutritional prevention and management

A.H. Lichtenstein, Tufts University, Boston, MA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A.H. Lichtenstein, Hyperlipidemia: Prevention and Management, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 453–461, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00147-1.

Introduction 380
Assessing the impact of diet on CVD risk 381
Dietary lipids: approaches for the prevention and management of CVD 381
Amount of dietary fat 381
Type of dietary fat 381
Saturated fatty acids 383
Unsaturated fatty acids 383
MUFA 384
PUFA 385
N-6 fatty acids 385
N-3 fatty acids 385
Trans fatty acids 385
Dietary cholesterol 386
Other dietary considerations for the prevention and management of CVD 387
Fiber 387
Plant sterols (phytosterols) 388
Nutrient supplements 388
Dietary patterns 388
Conclusions 389
References 389
Further reading 390

Key points

- Summarize key issues related to the relation between the amount and type of dietary fat and CVD.
- Discuss the differences between the major classes of fatty acids and CVD prevention and treatment.
- Review the controversies related to dietary cholesterol, and plasma lipids concentrations and lipoprotein profiles.
- Discuss the importance of adequate dietary fiber for CVD prevention and treatment.
- Discuss the role of plant sterols (phytosterols) for CVD prevention and treatment.
- Summarize the available data related to nutrient supplements and CVD prevention and treatment.

Introduction

A wide range of diet related factors alter risk of developing cardiovascular disease (CVD). Some were identified early in the 20th century and have stood the test of time. Others have emerged more recently. In some cases, they too have stood the test of time (e.g., moderate rather than low fat diets, unmodified plant oils rather than hydrogenated plant oils), yet, others have not (e.g., vitamin E and vitamin D supplements). Although emerging data have resulted in shifts in certain dietary guidance over time, none of the changes are without controversy, due to inconsistent results. Some inconsistencies are likely attributable to differences across study cohorts in biological factors (e.g., genetics, sex), physiological characteristics (e.g., body weight, body composition), and long-term lifestyle practices (e.g., physical activity, exposure to tobacco products). Also, the approach used for data analyses may contribute to differences in reported responses to similar dietary interventions; for example, expression of the data in relative (e.g., energy percent) or absolute (e.g., grams, servings) terms. This article will present evidence-based current trends in dietary approaches for the prevention and management of CVD.

Assessing the impact of diet on CVD risk

It is logistically difficult, if not impossible, to directly assess the effect of dietary interventions on hard CVD outcomes, given that the natural course of the disease is measured in years or decades rather than days or months (Virani et al., 2021). Hence, most dietary interventions aimed at reducing CVD risk are evaluated based on intermediate biomarkers. As the number of dietary variables potentially associated with CVD risk has increased (and occasionally decreased) over the past few decades, so have the number of potential biomarkers of CVD used to assess the efficacy of dietary interventions. The independent and relative importance of each biomarker has yet to be fully adjudicated. With the widespread advent of 'omics technology the field is expected to continue to evolve.

Historically, total cholesterol, individual lipoproteins (low density lipoprotein-cholesterol [LDL-C] and high-density lipoprotein-cholesterol [HDL-C]) and triacylglycerol (triglyceride) concentrations were used to evaluate the efficacy of a dietary intervention for CVD risk. Total cholesterol, HDL-C and triglyceride concentrations were measured directly, whereas LDL-C concentration was calculated using the Friedewald formula (LDL-C = total cholesterol—HDL-C—[triglyceride/5]) unless triglyceride concentrations exceeded 400 mg/dL. Reliable automated direct assays for LDL-C are now available and obviate the need for this calculation. Circulating factors, such as lipoprotein(a) (Lp[a]), C-reactive protein (CRP) and other markers of inflammation, LDL particle size, hematologic factors, apoprotein concentrations and genotypes, measures of glucose homeostasis, plant sterol and intermediates of cholesterol biosynthesis, HDL subspecies, cholesterol efflux capacity, trimethylamine N-oxide concentration, and remnant-like particle (triglyceride-rich lipoprotein particles) concentration are also considered potentially informative in estimating CVD risk and the effect of dietary interventions on CVD risk. As new data emerge, the relative importance of each is elucidated and additional factors are identified. Currently, much dietary guidance aimed at decreasing CVD risk is made on the basis of plasma lipid and lipoprotein concentrations. In some cases, additional factors, such as CRP and Lp(a), are taken into consideration for individuals classified as at intermediate risk to inform risk management.

Dietary lipids: approaches for the prevention and management of CVD

Amount of dietary fat

For most individuals consuming Western-type diets, dietary fat serves as a major energy source. One gram of fat contributes 9 kcal, a little more than twice that contributed by protein or carbohydrate (4 kcal/g), and somewhat more than that contributed by alcohol (7 kcal/g). Focusing on the importance of total dietary fat on CVD prevention and management, there are two independent factors for consideration; plasma lipoprotein profiles and body weight. The former is used to determine diet and pharmacotherapy treatment. The latter is of importance due to secondary effects of excess body weight on plasma lipid profiles and factors associated with the metabolic syndrome (elevated blood pressure, waist circumference and blood glucose concentration, and dyslipidemia [low HDL cholesterol and high triglyceride concentrations]).

The effects of total dietary fat (as a percent of total energy intake) on plasma lipoprotein profiles, are measured by triglyceride and HDL-C concentrations, and the total cholesterol to HDL-C ratio. Relatively consistent evidence indicates that, when body weight is maintained at a constant level, a decrease in the percent of energy contributed by dietary fat, accompanied by a concomitant increase in dietary carbohydrate, results in increased triglyceride and decreased HDL-C concentrations, and less favorable (higher) total cholesterol to HDL-C ratio. The underlying mechanism is carbohydrate induced hypertriglyceridemia. The influx of glucose exceeds the capacity of peripheral tissues to oxidize it or store as glycogen, forcing the liver to use the excess glucose to synthesize fatty acids, which are released into circulation as triglyceride-rich particles termed very low density lipoprotein (VLDL). This situation can be exacerbated when the source of carbohydrate is refined, such as added sugar, which results in a rapid raise in blood glucose contractions. Both low HDL-C and high triglyceride concentrations are associated with increased risk for CVD, and frequently coexist. Very low fat diets are not recommended, and should be particularly avoided by individuals with insulin resistance, because such a dietary pattern is more likely to result in dyslipidemia (low HDL-C and high triglycerides) than higher fat diets.

Available historical data suggest a null or weak association between the amount of dietary fat and body weight. Data from two relatively long-term weight loss intervention trials confirm this finding, indicating that the percent of energy from dietary fat is not a major determinant of weight loss (**Figs. 1** and **2**) (Gardner et al., 2018; Sacks et al., 2009). In one study, a major determinate of weight loss was adherence to the diet intervention, as assessed by number of group sessions attended during the study period, rather than the proportions of dietary fat, carbohydrate or protein. Hence, despite the theoretical logic of the assertion that high fat diets lead to weight gain, the data do not support the premise.

Type of dietary fat

The relation between dietary fat and atherosclerosis, the major form of CVD, was first identified at the turn of the 20th century. Subsequently, a series of meticulously conducted studies in humans during the 1960s demonstrated that changes in individual dietary fatty acids altered plasma total cholesterol concentration. As knowledge expanded and analytical techniques became more sophisticated, the focus on biomarkers for CVD risk shifted from total cholesterol to lipoprotein concentrations; LDL-C and HDL-C. Although many studies have confirmed early observations, inconsistencies among more recent work are not rare. These

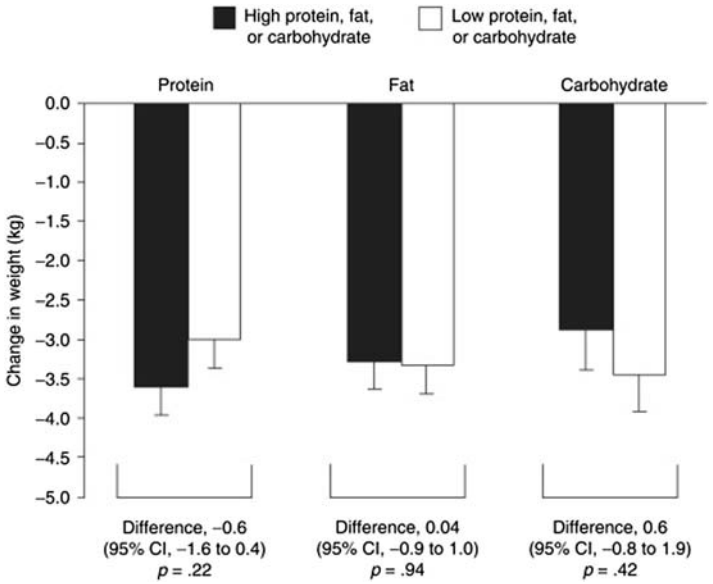


Fig. 1 Mean change in body weight and waist circumference from baseline to 2 years according to dietary macronutrient content. Low protein, fat or carbohydrate. High protein, fat or carbohydrate. From Sacks FM, Bray GA, Carey VJ, Smith SR, Ryan DH, Anton SD, McManus K, Champagne CM, Bishop LM, et al. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. NEJM. 2009;360:859-873.

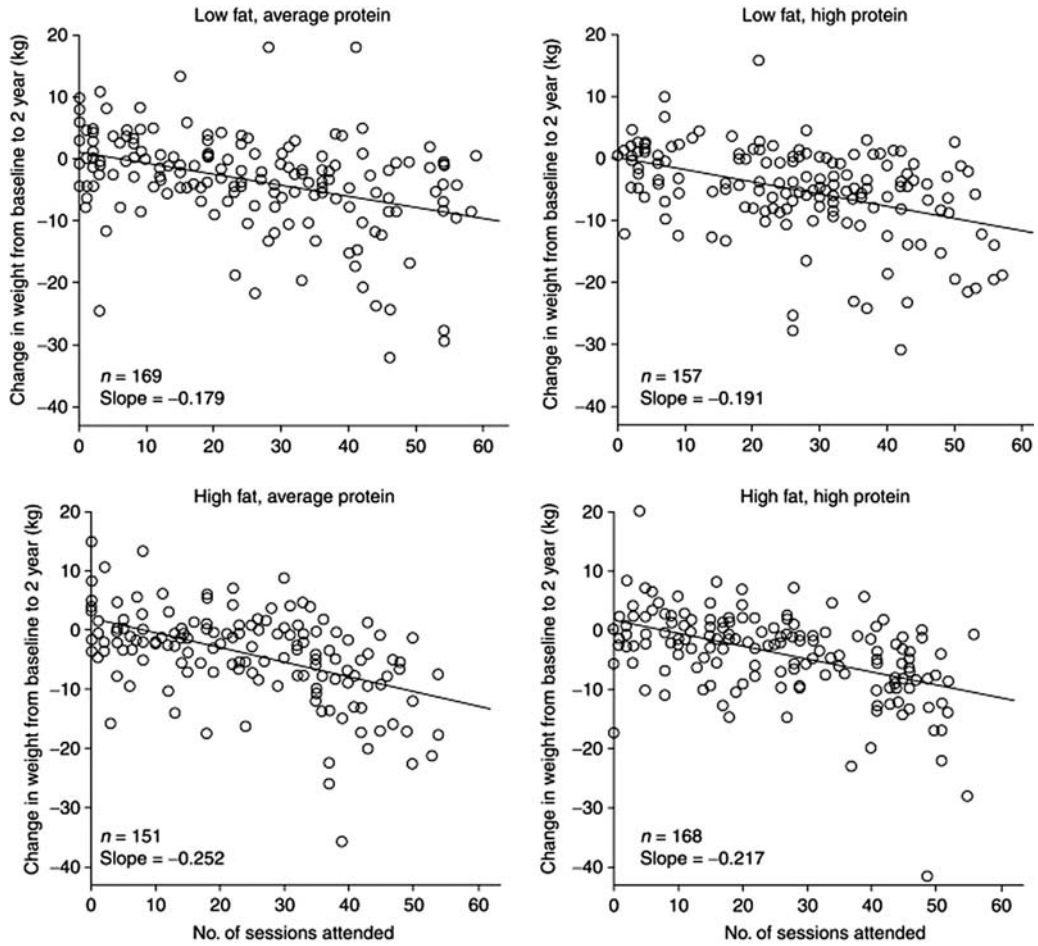


Fig. 2 Waterfall plot of weight loss by diet group (A) healthy low fat diet; (B) healthy low carbohydrate (carb) diet. From Sacks FM, Bray GA, Carey VJ, Smith SR, Ryan DH, Anton SD, McManus K, Champagne CM, Bishop LM, et al. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. NEJM. 2009;360:859-73.

inconsistencies are likely due to differences across diet interventions, such as absolute vs relative amounts of dietary fat, length of intervention period, and diet prior to the start of the study period (background diet), onto which the dietary variable(s) was superimposed. Additional differences among studies may include characteristics of participants such as age, sex, genetics, efficiency of cholesterol absorption and baseline plasma lipid concentrations.

Saturated fatty acids

Early evidence showed that diets relatively high in total saturated fatty acids (SFA) increased plasma total cholesterol concentration, compared to diets relatively high in a comparable amount of unsaturated fatty acids. It was also noted that not all SFA had identical effects. Subsequent work confirmed the hypercholesterolaemic effect of SFA, demonstrating that increasing SFA intake results in an increase in both LDL-C and HDL-C concentrations, although LDL-C more so that HDL-C, and reaffirmed that individual SFA had different effects on plasma lipoprotein particles. Short-chain fatty acids (6:0 to 10:0) and stearic acid (18:0) appear to have little or no effect on LDL-C or HDL-C concentrations, whereas intermediate chain SFA (lauric [12:0], myristic [14:0] and palmitic [16:0] acids) appear to be the most potent in increasing LDL-C concentration (**Table 1**). Subsequent work indicated that a high proportion of stearic acid (18:0) is converted to the monounsaturated fatty acid oleic acid (18:1) and, therefore, has a relatively neutral effect on plasma cholesterol concentration. The underlying mechanism by which fatty acids with 10 or fewer carbon atoms have different effects from those with 12–16 carbons on plasma cholesterol may be related to the mode of absorption (portal vein rather than lymphatic system) and preferential hepatic oxidation. An analysis that pooled data from multiple studies demonstrated that when dietary SFA was replaced by polyunsaturated fatty acids (PUFA), CVD risk was lowered (**Fig. 3**) (Jakobsen et al., 2009). When dietary SFA was replaced by carbohydrate (resulting in lower fat diets), this advantage was lost.

The major contributor to dietary SFA are fats from animal origin, meat and dairy fat. Approaches to reduce SFA intake from meat include replacing fatty cuts with leaner cuts, trimming visible fat prior to cooking, and reducing portion size. Reducing dietary SFA from dairy products can be achieved by substituting non-fat and low-fat (1%) for full-fat and reduced-fat (2%) versions and reducing portions of full fat diary foods such as cheese. Foods high in SFAs tend to be solid at room temperature. Notable exceptions include the tropical oils, palm, palm kernel and coconut. Tropic oils are liquid at room temperature because they have a relatively high concentration of short chain SFA. Reducing dietary SFA from tropical oils can be achieved by using the information in the ingredient list to identify sources and replace them with similar foods formulated with unsaturated fats, such as soybean and canola oils.

Unsaturated fatty acids

Unsaturated fatty acids are those that contain one or more double bonds in the acyl chain. As their name implies, monounsaturated fatty acids (MUFA) have one double bond and PUFA have multiple double bonds (two or more). Most fatty acids containing

Table 1 Major dietary fatty acids.

Dietary fatty acids		
Code	*Common name*	*Formula*
Saturated		
12: 0	Lauric acid	$CH_3(CH_2)_{10}COOH$
14: 0	Myristic acid	$CH_3(CH_2)_{12}COOH$
16: 0	Palmitic acid	$CH_3(CH_2)_{14}COOH$
18: 0	Stearic acid	$CH_3(CH_2)_{16}COOH$
Monounsaturated		
16: 1n–7 cis	Palmitoleic acid	$CH_3(CH_2)_6CH = (C)CH(CH_2)_7COOH$
18: 1n–9 *cis*	Oleic acid	$CH_3(CH_2)_7CH = (C)CH(CH_2)_7COOH$
18: 1n–9 *trans*	Elaidic acid	$CH_3(CH_2)_7CH = (t)CH(CH_2)_7COOH$
Polyunsaturated		
18: 2n–6, 9 all cis	Linoleci acid	$CH_3(CH_2)_4CH = (c)CHCH_2CH = (c)CH(CH_2)_7COOH$
18: 3n–3, 6, 9 all *cis*	α-Linoleci acid	$CH_3CH_2CH = (c)CHCH_2CH = (c)CHCH_2CH = (c)CH(CH_2)_7COOH$
18: 3n–6, 9,12 all *cis*	Υ-linoleci acid	$CH_3(CH_2)_4CH = (c)CHCH_2CH = (c)CHCH_2CH = (c)CH(CH_2)_4COOH$
20: 4n–6, 9,12,15 all *cis*	Ara childoleci acid	$CH_3(CH_2)_4CH = (c)CHCH_2CH = (c)CHCH_2CH = (c)CHCH_2CH = (c)CH(CH_2)_3COOH$
20: 5n–3, 6, 9,12,15 all *cis*	Eicosapentaenoic acid	$CH_3(CH_2CH = (c)CH)_5(CH_2)_3COOH$
20: 6n–3, 6, 9,12,15,18 all *cis*	Docosahexaenoic acid	$CH_3(CH_2CH = (c)CH)_6(CH_2)_2COOH$

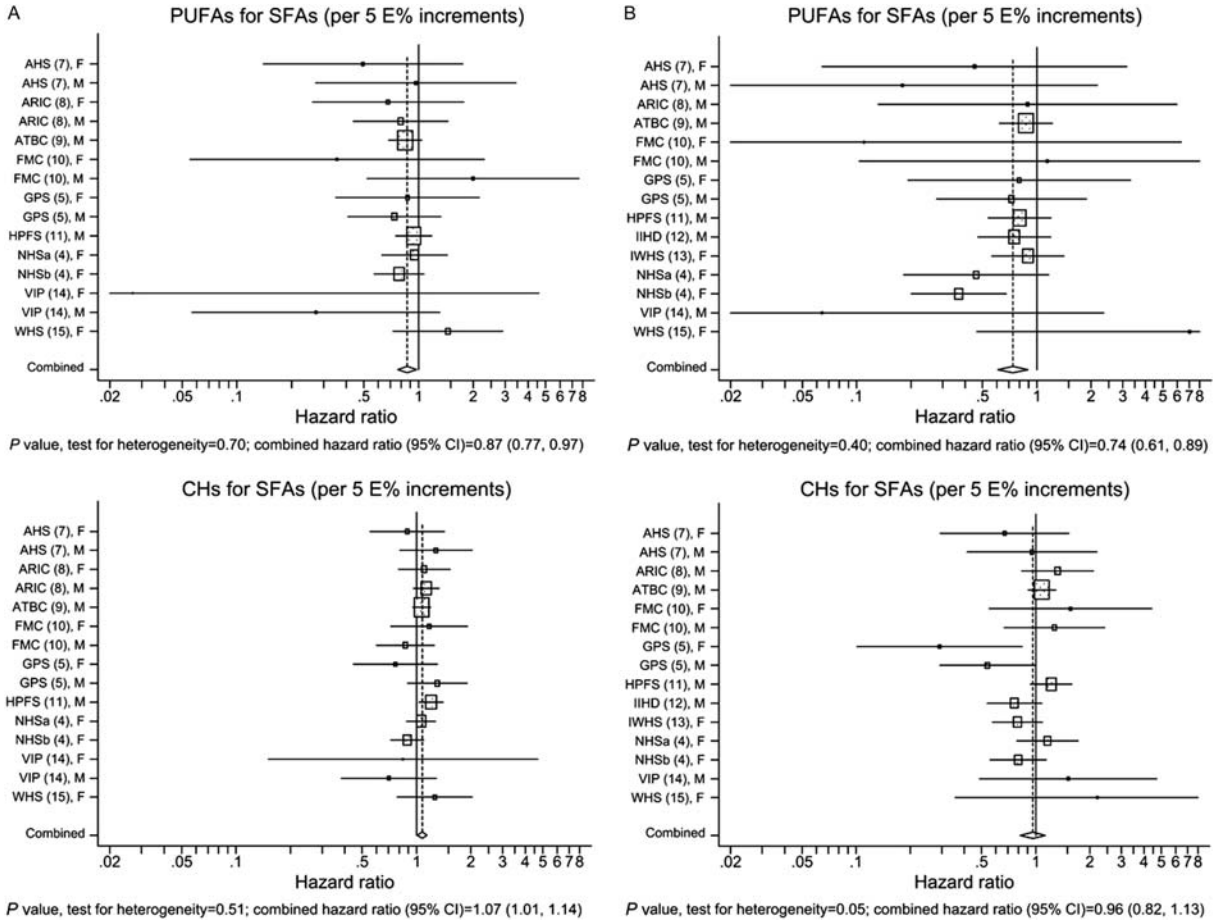


Fig. 3 Study-specific and combined hazard ratios and 95% CIs for coronary events (A) and coronary deaths (B) in the pooling project of cohort studies on diet and coronary disease. The squares and horizontal lines represent the study-specific hazard ratios and 95% CIs, respectively. The area of the squares reflects the study-specific weight (inverse of the variance). The diamonds represent the combined hazard ratios and 95% CI. AHS, Adventis Health Study; ARIC, Atherosclerosis Risk in Communities Study; ATBC, Alpha-Tocopherol and Beta-Carotene Cancer Prevention Study; FMC, Finnish Mobile Clinic Health Study; GPS, Glostrup Population Study; HPFS, Health Professionals Follow-Up Study; IIHD, Israeli Ischemic Heart Disease Study; IWHS, Iowa Women's Health Study; NHSa, Nurses' Health Study 1980; NHSb, Nurses' Health Study 1986; VIP, Västerbotten Intervention Program; WHS, Women's Health Study. SFA, saturated fatty acids; PUFA, polyunsaturated fatty acids; CH, carbohydrate. Jakobsen MU, O'Reilly EJ, Heitmann BL, Pereira MA, Bälter K, Fraser GE, Goldbourt U, Hallmans G, Knekt P, Liu S, Pietinen P, Spiegelman D, Stevens J, Virtamo J, Willett WC, Ascherio A. Major types of dietary fat and risk of coronary heart disease: a pooled analysis of 11 cohort studies. Am J Clin Nutr. 2009;89:1425-32.

double bonds in food are in the *cis* configuration, that is, the hydrogen atoms associated with the carbons forming the double bond are oriented on the same side of the acyl chain. Alternatively, some double bonds occur in the *trans* configuration, and the hydrogen atoms associated with the carbons forming the double bond are on the opposite side of the acyl chain. This difference in double bond conformation has implications for bond angle (see section on ***Trans* fatty acids**). This part of the discussion of unsaturated fatty acids will be restricted to those containing *cis* double bonds.

Relative to SFA, both MUFA and PUFA lower both LDL-C and HDL-C concentrations. The absolute magnitude of the change is greater for LDL-C than HDL-C, resulting in an improvement in the total cholesterol/HDL-C ratio. Most data suggest that MUFA has a somewhat smaller effect than PUFA in lowering both LDL-C and HDL-C concentrations (**Fig. 4**). Regardless, a shift from food sources rich in SFA to those rich in MUFA and PUFA should be encouraged for the prevention and management of CVD.

MUFA

The major MUFA in the diet is oleic acid (18:1) (**Table 1**). Plant oils high in MUFA include canola (rapeseed) and olive. Fats from meats are relatively high in MUFA, but unlike plant oils, they also contain relatively high levels of SFA, and therefore are not recommended as good source of dietary MUFA.

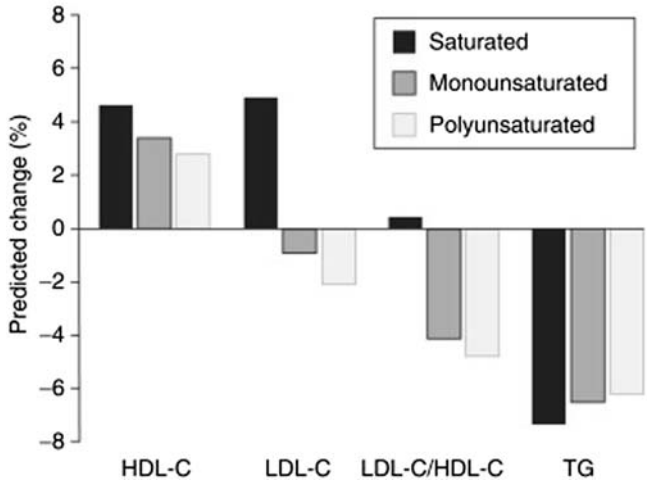


Fig. 4 Predicted changes based on replacement of 5% of energy as carbohydrates with specific fatty acids under isocaloric conditions. From Hu FB, Willett WC. Optimal diets for prevention of coronary heart disease. JAMA. 2002;288:2569-78.

PUFA

There is a wider range of PUFA than MUFA in the diet. Dietary PUFA vary based on chain length, degree of unsaturation (number of double bonds) and position of the double bond(s) (positional isomers). Two positional isomers of interest with respect to diet and CVD risk are omega (n)-6 and n-3 fatty acids (**Table 1**). The distinction is made based on the location of the first double bond from the methyl end of the fatty acyl chain (as opposed to the carboxyl end). If the first double bond is six carbons from the methyl end, the fatty acid is classified as n-6 fatty acid. If the first double bond is three carbons from the methyl end, the fatty acid is classified as n-3 fatty acid.

N-6 fatty acids

Diets rich in n-6 PUFA, relative to SFA, are associated with optimal cardiovascular outcomes (see prior section for SFA). Plant oils high in PUFA include soybean, corn, and unmodified sunflower and safflower. The major n-6 PUFA in the diet is linoleic acid (18:2n-6). Other n-6 PUFA, such as γ-linolenic acid (18:3n-6) and arachidonic acid (20:4n-6), occur in smaller amounts. These fatty acids are important for a wide range of metabolic functions.

N-3 fatty acids

Quantitatively, the major n-3 PUFA in the diet is α-linolenic acid (18:3n-3). Major dietary sources include soybean and canola oils. Two metabolically important n-3 PUFA that occur in smaller amounts are eicosapentaenoic acid (EPA, 20:5n-3) and docosahexaenoic acid (DHA, 22:6n-3). These are very long chain n-3 fatty acids (**Table 1**). Although humans have the capacity to convert α-linolenic acid to EPA and DHA, efficiency is low, estimated at about 3%–5%. The major dietary sources are fish and seafood. Dietary patterns rich in fish (very long chain n-3 fatty acids) have been associated with lower risk of coronary heart disease (**Fig. 5**). The beneficial effects of fish and seafood have been attributed, in part, to their EPA and DHA content (Wang et al., 2006) A variety of factors may contribute to the beneficial effect of dietary EPA and DHA, including lower likelihood of ventricular fibrillation, resulting in lower risk of sudden death, as well as lower triglyceride concentration, platelet aggregation and blood pressure. An additional benefit of dietary patterns that routinely include fish and seafood is their use as a substitute for other protein sources such as meat and cheese, rich in SFA. For plant-based diets, EPA and DHA rich algae is now available.

Trans fatty acids

Trans fatty acids, by definition, contain at least one double bond in the *trans* configuration (**Fig. 6**) (Mensink and Katan, 1990). In contrast to the more common *cis* double bonds, *trans* double bonds have a larger bond angle, resulting in a fatty acid with a straighter acyl chain, resulting in a conformation similar to saturated fatty acids. Fatty acids with *trans* double bond(s) can either be MUFA or PUFA. The intake of *trans* fatty acid containing foods has been positively linked to CVD risk (Lichtenstein et al., 1999). Concern about the use of partially-hydrogenated fat emerged in the early 1990s when it was reported that fatty acids with a *trans* double bond resulted in unfavorable lipoprotein profiles (**Fig. 7**) and later, with increased CVD risk (Ascherio et al., 1999).

Dietary *trans* fatty acids occur naturally in meat and dairy products because of anaerobic bacterial fermentation in ruminant animals and subsequent absorption and deposition in milk and muscle. *Trans* fatty acids are also introduced into foods when they are prepared with oils that have been partially hydrogenated. Partial hydrogenation results in changes in the fatty acyl chains;

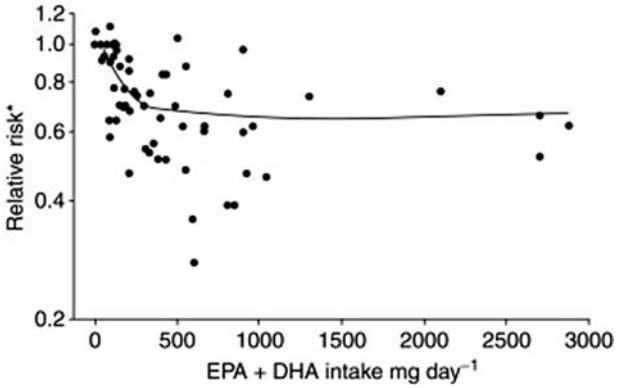


Fig. 5 Forest plot of fish consumption with CHD incidence. The diamond denotes summary risk estimate, and horizontal lines represent 95% CI. Abbreviations: RR, relative risk; CI, Confidence interval. Zhang B, Xiong K, Cai J, Ma A. Fish Consumption and coronary heart disease: a meta-analysis. Nutrients. 2020;12:2278.

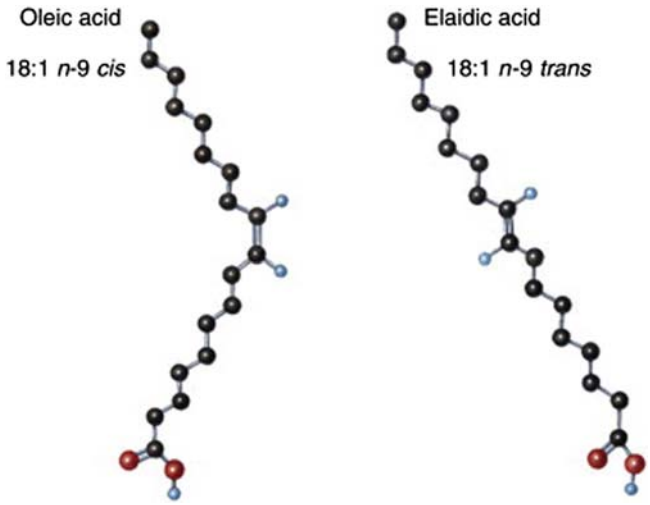


Fig. 6 Geometric isomers of 18:1n-9 (oleic and elaidic acids). Effect of dietary trans fatty acids on high-density and low-density lipoprotein cholesterol levels in healthy subjects. Mensink RP, Katan MB.N Engl J Med. 1990 Aug 16;323(7):439-45.

conversion of *cis* to *trans* double bonds, saturation of double bonds and migration of double bonds along the acyl chain. Plant oils are partially hydrogenated to increase viscosity (change a liquid oil into a semi-liquid or solid) and extend shelf life (decrease susceptibility to oxidation). It has been suggested that there is a difference in response to *trans* fatty acids of ruminant fat compared to partially hydrogenated fat origin. However, most of the evidence suggests that, if present, differences are small and all sources of *trans* fatty acids should be limited.

The major source of dietary *trans* fatty acids worldwide is from partially hydrogenated fat, primarily in commercially fried foods and baked goods. However, over the past decades there has been a shift away from the use of partially hydrogenated fats in some countries. In some cases, this change was spurred by consumer demand, and in some cases, by legislation. In the US, changes in legislation include mandatory inclusion of *trans* fat on the Nutrient Facts panel and removal of partially hydrogenated fat from the FDA's Generally Recognized as Safe (GRAS) list. It is expected that this trend will continue globally.

Dietary cholesterol

The observation that dietary cholesterol increased plasma total cholesterol concentration leading to an acceleration of arteriosclerotic lesion progression was originally made early in the 20th century, in rabbits. In humans, positive associations have been observed between dietary cholesterol and plasma LDL-C concentration when dietary cholesterol was provided at high intake levels (800–900 mg/d; ~4–5 eggs/d) (Carson et al., 2020). Within the context of amounts habitually consumed, the effect was modest and not significant. Hence, the most recent editions of the Dietary Guidelines for Americans and the American Heart Association do not include a numerical target for dietary cholesterol. Dietary cholesterol is present only in foods of animal origin. Major sources

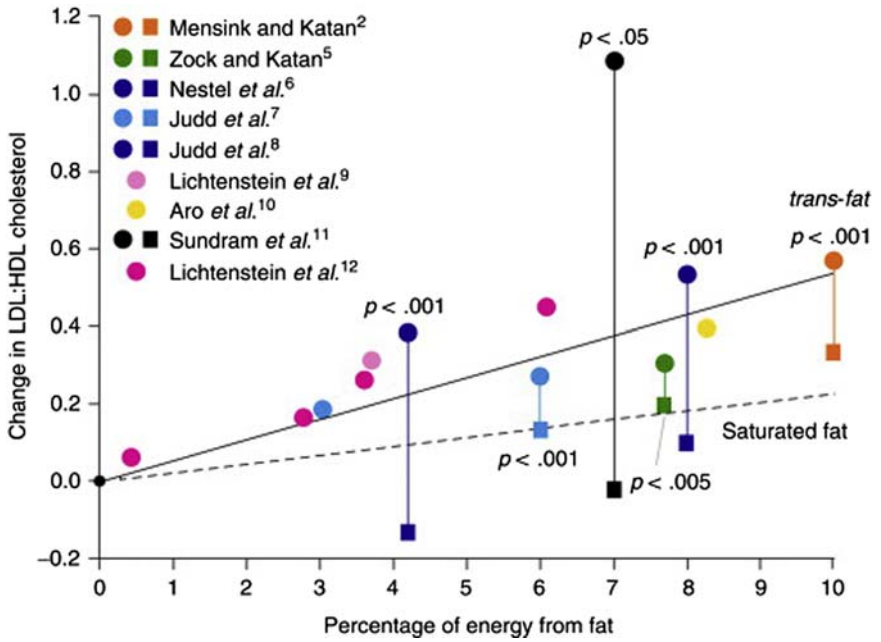


Fig. 7 Change in LDL:HDL cholesterol ratio in response to *trans* fatty acids (solid line) and saturated fatty acids (dashed line). From Ascherio, A., Katan, M. B., Zock, P. L., Stampfer, M. J., Willett, W. C. *Trans* fatty acids and coronary heart disease. N Engl J Med 1999;340: 1994-1998.

include eggs, fatty cuts of meat and full fat dairy products. Current dietary patterns recommended to lower chronic disease risk, particularly, cardiometablic risk, are relatively low in dietary cholesterol.

Other dietary considerations for the prevention and management of CVD

Fiber

Dietary soluble fiber, primarily β-glucan, has a modest independent effect of lowering LDL-C concentration (Armet et al., 2020; Hui et al., 2019). Approximately 50% of human controlled intervention trials report a significant benefit (Pereira et al., 2003; Doucette et al., 2019). Potential LDL-C lowering mechanisms of fiber include binding bile acids and cholesterol in the intestine, leading to increased fecal loss, and providing substrate for the production of short-chain fatty acids by the gut microbiota. In addition to whole grains, other fiber rich foods include fruit and vegetables, nuts and seeds, and legumes. Dietary patterns abundant in fiber rich foods, usually those that are predominately plant based, are associated with lower CVD risk (Fig. 8).

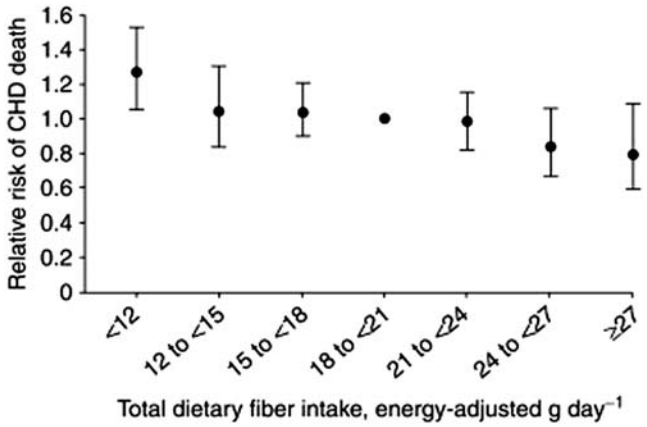


Fig. 8 Relative risk of death from coronary heart disease (CHD) by category of total dietary fiber intake. Pereira MA, O'Reilly E, Augustsson K, Fraser GE, Goldbourt U, Heitmann BL, Hallmans G, Knekt P, Liu S, Pietinen P, Spiegelman D, Stevens J, Virtamo J, Willett WC, Ascherio A. Dietary fiber and risk of coronary heart disease: a pooled analysis of cohort studies. Arch Intern Med. 2004;164:370-6.

Plant sterols (phytosterols)

The term sterol represents a group of compounds that are essential constituents of cell membranes in animals and plants (Katan et al., 2003). Cholesterol is the major sterol of mammals. Phytosterols, such as beta-sitosterol, campesterol, and stigmasterol, are the major sterols of plants (Fig. 9). Humans cannot synthesize plant sterols and they are poorly absorbed. When consumed they replace cholesterol in intestinal micelles, thereby interfering with cholesterol absorption. Maximal LDL-C lowering attributable to plant sterols occurs at a dose of about 2 g/d (Grundy et al., 2003). Although a relatively wide range of responses has been reported, most work suggests an expected LDL-C lowering of 5%–10% in individuals with high cholesterol. Given that the mechanism is at the point of absorption, this response can be additive to other LDL-C lowering approaches, such as stain therapy. In addition to plant sterol supplements, foods enriched with plant sterols, referred to as functional foods, are available. In some cases, the saturated form of the plant sterol, stanols, are added to these products. With respect to LDL-C lowering, the efficacy of plant sterols and stanols is similar. It is important to note that if an individual with hypercholesterolemia is encouraged to use plant sterols or stanols to lower their LDL-C and they choose food forms, such as margarine, orange juice or yogurts, these foods should replace, rather than be added to their habitual diet. No benefit will be derived from exceeding the maximal recommended dose or using plant sterols as a prophylactic measure to prevent the development of hypercholesterolemia.

Nutrient supplements

In the past, considerable interest had been generated for the potential benefit of supplementation with vitamin C, beta-carotene, vitamin E, folate, or vitamin D for reducing CVD risk. Support came from observational studies, and in some cases biological plausibility was derived from experimental animal models and *in vitro* work. Despite these links, randomized controlled intervention trials have failed to demonstrate a benefit of nutrient supplementation (Manson et al., 2019) and, in some cases, have identified risks at high intake levels (Miller et al., 2005; Banach and Penson, 2021). Currently, the data do not support a use of nutrient supplements for the prevention or management of CVD.

Dietary patterns

Because multiple components of a typical day's food and beverage intake, such as type of fat and carbohydrate and amount of fiber or added sugar, affect blood lipid levels and CVD risk, to facilitate adoption, current prevention recommendations have shifted from individual foods and nutrients to dietary patterns. This includes the most recent guidance from the American Heart Association (Lichtenstein et al., 2021). The advantages include adaptability to personal preferences, ethnic and religious practices and life stages, uniformity with guidance across multiple chronic risk reduction strategies (e.g., diabetes, cancer), and consistency with best practices for sustainability and environmental health. The guidance is summarized into 10 features (Table 2) and pictorially (Fig. 10). The features include adjusting energy intake and expenditure to achieve and maintain a healthy body weight; eating plenty and a variety of fruits and vegetables; choosing whole grain foods and products; choosing healthy sources of protein (mostly from plants, regularly from fish and seafood, primarily low-fat or fat-free dairy products, and if meat or poultry is desired, mostly lean cuts and unprocessed forms); using liquid plant oils rather than tropical oils and partially hydrogenated fats; choosing minimally processed foods instead of ultra-processed foods; minimizing the intake of beverages and foods with added sugars; choosing and preparing foods with little or no salt; if not consuming alcohol, do not start, if you choosing to consume drink alcohol, limiting intake; and adhering to this guidance regardless of where food is procured prepared or consumed.

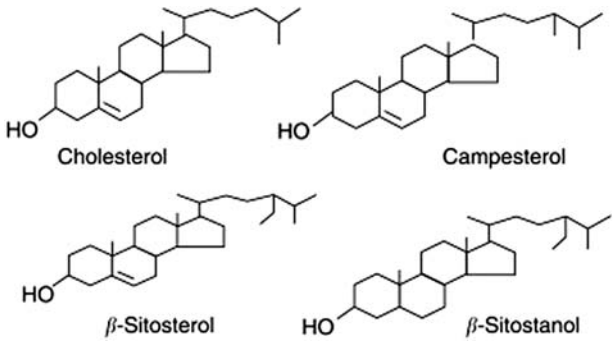


Fig. 9 Structures of cholesterol, and the phytosterols campesterol, β-sitosterol and sitostanol

Table 2 Evidence-based dietary guidance to promote cardiometabolic health.

1. Adjust energy intake and expenditure to achieve and maintain a healthy body weight
2. Eat plenty of fruits and vegetables, choose a wide variety
3. Choose whole grain foods and products made mostly with whole grains rather than refined grains
4. Choose healthy sources of protein a. Mostly protein from plants (legumes and nuts)
   b. Regular intake of fish and seafood
   c. Low-fat or fat-free dairy products instead of full-fat dairy products
   d. If meat or poultry are desired, choose lean cuts and avoid processed forms
5. Use liquid plant oils rather than tropical oils (coconut, palm and palm kernel), animal fats (e.g., butter and lard) and partially-hydrogenated fats
6. Choose minimally processed foods instead of ultra-processed foods[a]
7. Minimize intake of beverages and foods with added sugars
8. Choose and prepare foods with little or no salt
9. If you don't drink alcohol, do not start; if you choose to drink alcohol, limit intake.
10. Adhere to this guidance regardless of where food is prepared or consumed

[a]There is no commonly accepted definition for ultra-processed foods, and some healthy foods may exist within the ultra-processed food category.

From Lichtenstein, A.H., Appel, L.J., Vadiveloo, M., Hu, F.B., Kris-Etherton, P., Rebholz, C.M., Sacks, F.M., Thorndike, A.N., Van Horn, L., 2021. Dietary guidance to improve cardiovascular health: Scientific Statement from the American Heart Association. Circulation CIR0000000000001031. PMID: 34724806.

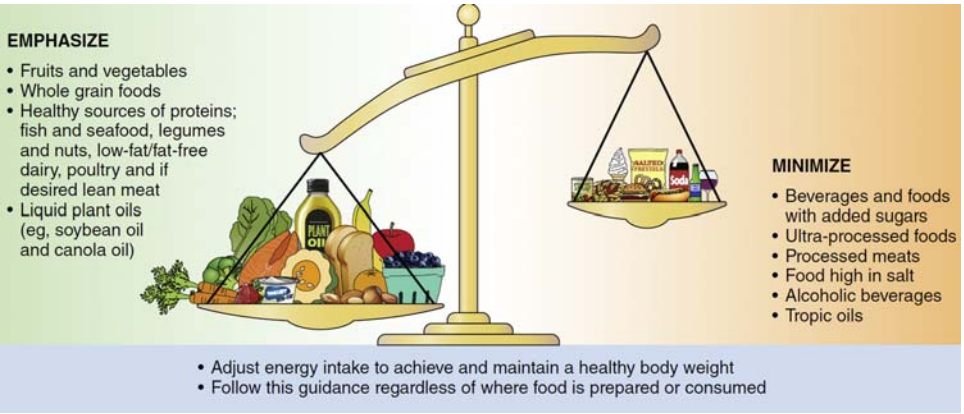


Fig. 10 Dietary patterns to promote cardiovascular health. From Lichtenstein, A.H., Appel, L.J., Vadiveloo, M., Hu, F.B., Kris-Etherton, P., Rebholz, C.M., Sacks, F.M., Thorndike, A.N., Van Horn, L., 2021. Dietary guidance to improve cardiovascular health: Scientific Statement from the American Heart Association. Circulation CIR0000000000001031. PMID: 34724806.

Conclusions

Relationships between diet and CVD risk have been clearly established, particularly with lipid and lipoprotein profiles, in addition to other risk factors. The current recommendation is to focus on the whole dietary pattern rather than not individuals foods or nutrients. Specific focus is on type rather than amount of dietary fat, and to replace SFA and *trans* fatty acids with unsaturated fatty acids, particularly PUFA, consuming at least two fish meals per week and fiber rich diets. Daily intake of plant sterols can result in LDL-C concentration lowering, in addition to that achieved with statins (Han et al., 2016). Currently, data do not support the use of nutrient supplements for the prevention or treatment of CVD. Attainment or maintenance of a healthy body weight should be emphasized along with engagement in regular physical activity. These recommendations are the culmination of more than a century of work. They have evolved slowly and continue to be fine-tuned as new findings emerge, leading to subsequent modifications.

References

Armet, A.M., Deehan, E.C., Hewko, S.J., Thöne, J.V., Walter, J., 2020. The effect of isolated and synthetic dietary fibers on markers of metabolic diseases in human intervention studies: a systematic review. Adv. Nutr. 11, 420–438.

Ascherio, A., Katan, M.B., Zock, P.L., Stampfer, M.J., Willett, W.C., 1999. *Trans* fatty acids and coronary heart disease. NEJM 340, 1994–1998.

Banach, M., Penson, P.E., 2021. Natural compounds as anti-atherogenic agents: clinical evidence for improved cardiovascular outcomes. Atherosclerosis 316, 58–65.

Carson, J.A.S., Lichtenstein, A.H., Anderson, C.A.M., Appel, L.J., Kris-Etherton, P.M., Meyer, K.A., Petersen, K., Polonsky, T., Van Horn, L., 2020. Dietary cholesterol and cardiovascular risk: a science advisory from the American Heart Association. Circulation 141, e39–e53.

Doucette, S., Hui, S., Lang, H., Liu, K., Liu, Y., Mi, M., Wang, X., Yi, L., Zhu, X., 2019. Comparative effects of different whole grains and brans on blood lipid: a network meta-analysis. Eur. J. Nutr. 58, 2779–2787.

Gardner, C.D., Trepanowski, J.F., Del Gobbo, L.C., Hauser, M.E., Rigdon, J., Ioannidis, J.P.A., Desai, M., King, A.C., 2018. Effect of low-fat vs low-carbohydrate diet on 12-month weight loss in overweight adults and the association with genotype pattern or insulin secretion: the DIETFITS randomized clinical trial. JAMA 319, 667–679.

Grundy, S.M., Jones, P., Katan, M.B., Law, M., Miettinen, T., Paoletti, R., 2003. Participants Stresa Workshop (2003) efficacy and safety of plant stanols and sterols in the management of blood cholesterol levels. Mayo Clin. Proc. 788, 965–978.

Han, S., Jiao, J., Xu, J., Zimmermann, D., Actis-Goretta, L., Guan, L., Zhao, Y., Qin, L., 2016. Effects of plant stanol or sterol-enriched diets on lipid profiles in patients treated with statins: systematic review and meta-analysis. Sci. Rep. 19, 31337.

Hui, S., Liu, K., Lang, H., Liu, Y., Wang, X., Zhu, X., Doucette, S., Yi, L., Mi, M., 2019. Comparative effects of different whole grains and brans on blood lipid: a network meta-analysis. Eur. J. Nutr. 58, 2779–2787.

Jakobsen, M.U., O'Reilly, E.J., Heitmann, B.L., Pereira, M.A., Balter, K., Fraser, G.E., Goldbourt, U., Hallmans, G., Knekt, P., Liu, S., Pietinen, P., Spiegelman, D., StevensJ, V.J., Willett, W.C., Ascherio, A., 2009. Major types of dietary fat and risk of coronary heart disease: a pooled analysis of 11 cohort studies. Am. J. Clin. Nutr. 89, 1425–1432.

Katan, M.B., Grundy, S.M., Jones, P., Law, M.T., Paoletti, R., Participants Stresa Workshop, 2003. Efficacy and safety of plant stanols and sterols in the management of blood cholesterol levels. Mayo Clin. Proc. 78, 965–978.

Lichtenstein, A.H., Ausman, I, Jalbert, S.M., Schaefer, E.J., 1999. Comparison of different forms of hydrogenated fats on serum lipid levels in moderately hypercholesterolemic female and male subjects. NEJM 340, 1933–1940.

Lichtenstein, A.H., Appel, L.J., Vadiveloo, M., Hu, F.B., Kris-Etherton, P., Rebholz, C.M., Sacks, F.M., Thorndike, A.N., Van Horn, L., 2021. Dietary guidance to improve cardiovascular health: Scientific Statement from the American Heart Association. Circulation. CIR0000000000001031. PMID: 34724806.

Manson, J.E., Cook, N.R., Lee, I.M., Christen, W., Bassuk, S.S., Mora, S., Gibson, H., Gordon, D., Copeland, T., D'Agostino, D., Friedenberg, G., Ridge, C., Bubes, V., Giovannucci, E.L., Willett, W.C., Buring, J.E., VITAL Research Group, 2019. Vitamin D supplements and prevention of cancer and cardiovascular disease. NEJM 380, 33–44.

Mensink, R.P., Katan, M.B., 1990. Effect of dietary *trans* fatty acids on high-density and low-density lipoprotein cholesterol levels in healthy subjects. NEJM 323, 439–445.

Miller III, E.R., Pastor-Barriuso, R., Dalal, D., Riemersma, R.A., Appel, L.J., Guallar, E., 2005. Meta-analysis: high-dosage vitamin E supplementation may increase all-cause mortality. Ann. Intern. Med. 142, 37–46.

Pereira, M.A., O'Reilly, E., Augustsson, K., Fraser, G.E., Goldbourt, U., Heitmann, B.L., Hallmans, G., Knekt, P., Liu, S., Pietinen, P., Spiegelman, D., Stevens, J., Virtamo, J., Willett, W.C., Ascherio, A., 2003. Dietary fiber and risk of coronary heart disease: a pooled analysis of cohort studies. Arch. Intern. Med. 164, 370–376.

Sacks, F.M., Bray, G.A., Carey, V.J., Smith, S.R., Ryan, D.H., Anton, S.D., McManus, K., Champagne, C.M., Bishop, L.M., Laranjo, N., Leboff, M.S., Rood, J.C., de Jonge, L., Greenway, F.L., Loria, C.M., Obarzanek, E., Williamson, D.A., 2009. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. NEJM 360, 859–873.

Virani, S.S., Alonso, A., Aparicio, H.J., Benjamin, E.J., Bittencourt, M.S., Callaway, C.W., Carson, A.P., Chamberlain, A.M., Cheng, S., Delling, F.N., Elkind, M.S.V., Evenson, K.R., Ferguson, J.F., Gupta, D.K., Khan, S.S., Kissela, B.M., Knutson, K.L., Lee, C.D., Lewis, T.T., Liu, J., Loop, M.S., Lutsey, P.L., Ma, J., Mackey, J., Martin, S.S., Matchar, D.B., Mussolino, M.E., Navaneethan, S.D., Perak, A.M., Roth, G.A., Samad, Z., Satou, G.M., Schroeder, E.B., Shah, S.H., Shay, C.M., Stokes, A., VanWagner, L.B., Wang, N.Y., Tsao, C.W., 2021. Heart Disease and Stroke Statistics-2021 update: a report from the American Heart Association. Circulation 143, e254–e743.

Wang, C., Harris, W.S., Chung, M., Lichtenstein, A.H., Balk, E.M., Kupelnick, B., Jordan, H.S., Lau, J., 2006. N-3 fatty acids from fish or fish-oil supplements, but not alpha-linolenic acid, benefit cardiovascular disease outcomes in primary- and secondary-prevention studies: a systematic review. Am. J. Clin. Nutr. 84, 5–17.

Further reading

Agrawal, R., Ballantyne, C.M., Barter, P.J., Bash, D., Davidson, M.H., Garcia, M., Himmelmann, A., Kastelein, J.J.P., Katona, B.G., Koenig, W., Lincoff, A.M., Loss, L.E., Lundström, T., McGuire, D.K., Menon, V., Mozaffarian, D., Nicholls, S.J., Nissen, S.E., Ray, K.K., Rensfeldt, M., Ridker, P.M., Wolski, K., 2020. Effect of high-dose omega-3 fatty acids vs corn oil on major adverse cardiovascular events in patients at high cardiovascular risk: the STRENGTH randomized clinical trial. JAMA 324, 2268–2280.

Kalstad, A.A., Myhre, P.L., Laake, K., Tveit, S.H., Schmidt, E.B., Smith, P., Nilsen, D.W.T., Tveit, A., Fagerland, M.W., Solheim, S., Seljeflot, I., Arnesen, H., OMEMI Investigators, 2021. Effects of n-3 fatty acid supplements in elderly patients after myocardial infarction: a randomized, controlled trial. Circulation 143, 528–539.

Lichtenstein, A.H., Appel, L.J., Brands, M., Carnethon, M., Daniels, S., Franch, H.A., Franklin, B., Kris-Etherton, P., Harris, W.S., Howard, B., Karanja, N., Lefevre, M., Rudel, L., Sacks, F.B., Van Horn, L., Winston, M., Wylie-Rosett, J., 2006. Diet and lifestyle recommendations revision 2006: a scientific statement from the American Heart Association Nutrition Committee. Circulation 114, 82–96.

Manson, J.E., Cook, N.R., Lee, I.M., Christen, W., Bassuk, S.S., Mora, S., Gibson, H., Albert, C.M., Gordon, D., Copeland, T., D'Agostino, D., Friedenberg, G., Ridge, C., Bubes, V., Giovannucci, E.L., Willett, W.C., Buring, J.E., VITAL Research Group, 2019. Marine n-3 fatty acids and prevention of cardiovascular disease and cancer. NEJM 380, 23–32.

Nicholls, S.J., Lincoff, A.M., Garcia, M., Bash, D., Ballantyne, C.M., Barter, P.J., Davidson, M.H., Kastelein, J.J.P., Koenig, W., McGuire, D.K., Mozaffarian, D., Ridker, P.M., Ray, K.K., Katona, B.G., Himmelmann, A., Loss, L.E., Rensfeldt, M., Lundström, T., Agrawal, R., Menon, V., Wolski, K., Nissen, S.E., 2020. Effect of high-dose omega-3 fatty acids vs corn oil on major adverse cardiovascular events in patients at high cardiovascular risk: the STRENGTH randomized clinical trial. JAMA 2324, 2268–2280.

Penson, P.E., Banach, M., 2020. Natural compounds as anti-atherogenic agents: clinical evidence for improved cardiovascular outcomes. Atherosclerosis 316, 58–65.

Stone, N.J., Robinson, J.G., Lichtenstein, A.H., Goff Jr., D.C., Lloyd-Jones, D.M., Smith Jr., S.C., Blum, C., Schwartz, J.S., 2013. ACC/AHA cholesterol guideline panel. Ann. Intern. Med. 160, 339–343.

Willett, W.C., 2012. Dietary fats and coronary heart disease. J. Intern. Med. 272, 13–24.

Hypertension: Dietary factors

L.J. Appel, Johns Hopkins University, Baltimore, MD, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of L.J. Appel, Hypertension: Dietary Factors, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 462–468, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00150-1.

Introduction 391
Dietary factors that lower BP 393
Weight loss 393
Reduced salt intake 393
Increased potassium intake 395
Moderation of alcohol intake 396
Whole dietary patterns 396
Vegetarian diets 396
DASH-style dietary patterns 396
Protein intake 396
Dietary factors with limited or uncertain effect on BP 397
Fiber 397
Calcium and magnesium 397
Fish oil supplementation 397
Fats (other than fish oil) and cholesterol 398
Vitamin C 398
Gene–diet interactions 398
Special populations 398
Children 398
Pregnant women 398
Older people 398
Populations defined by race/ethnicity or geography 398
Conclusion 399
References 399

Key points

- High blood pressure, also termed hypertension, affects over 1.4 billion individuals worldwide and is the leading preventable cause of mortality.
- High blood pressure results from dietary factors, other environmental factors, genetic factors, and their interactions; of these factors, diet likely has a predominant role.
- Dietary factors that effectively lower blood pressure are weight loss, reduced dietary sodium intake, increased dietary potassium intake, moderation of alcohol intake (among those who drink), and adoption of a DASH-style or vegetarian dietary pattern.
- The current challenge is to develop and implement clinical and public health strategies that lower blood pressure though sustained dietary changes among individuals and more broadly among populations.

Introduction

Worldwide, high blood pressure (BP) is the leading cause of preventable death, not just in high income countries, but also in low- and middle-income countries (GBD Collaborators, 2018). The high global burden from high BP reflects its causal relationship with multiple diseases and its enormous prevalence—an estimated 1.4 billion person have hypertension (Mills et al., 2016) High BP is causally related to cardiovascular (CV) and kidney diseases, including stroke, myocardial infarction, heart failure, kidney failure, and likely cognitive decline. As blood pressure (BP) rises, so does the risk of these diseases (**Fig. 1**) (Lewington et al., 2002). The relationship is strong, consistent, continuous, independent, and etiologically relevant. Accordingly, the adverse consequences of high BP are not just restricted to individuals with hypertension (classically defined as a systolic BP $\geq$ 140 mm Hg and/or a diastolic BP $\geq$ 90 mm Hg, or more recently revised to a systolic BP $\geq$ 130 mm Hg and/or a diastolic BP $\geq$ 80 mm Hg) (Whelton et al., 2018). A normal BP is considered a systolic BP $<$ 120 mm Hg and diastolic BP $<$ 80 mm Hg. Those with intermediate levels

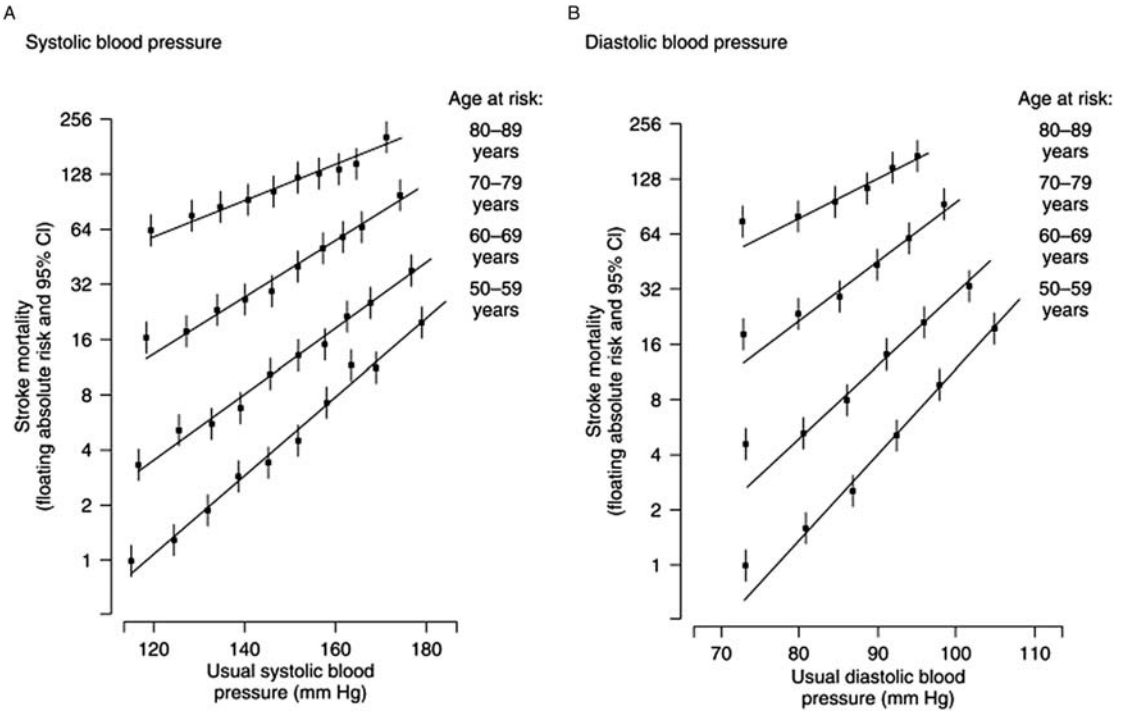
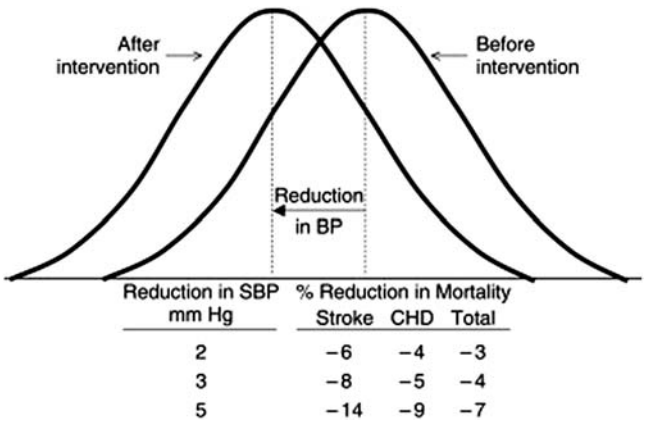


Fig. 1 Stroke mortality rate by decade of age versus systolic BP (A) and diastolic BP (B): meta-analysis of 61 prospective studies with 2.7 million person-years. Reproduced from Lewington, S., Clarke, R., Qizilbash, N., Peto, R., Collins, R., 2002. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet 360, 1903–1913.

between normal BP and hypertensive BP have a high probability of developing hypertension and carry an excess risk of cardiovascular disease (Vasan et al., 2001). It has been estimated that approximately 20% of all deaths and >50% of deaths from ischemic heart disease and stroke can be attributed to high BP (Forouzanfar et al., 2017).

In Western countries and most economically developing countries, systolic BP rises with age in both children and adults. As a consequence, the lifetime risk of developing hypertension is extremely high, approximately 90% among US adults older than the age of 50 years (Vasan et al., 2002). However, the rise in BP with age is not inevitable. There are numerous isolated populations in which the rise in BP is blunted or even flat (Rose et al., 1988). These populations are typically characterized by extremely low intakes of salt, relatively high intakes of potassium, and a lean body habitus.

Lifestyle modification, which includes dietary changes and increased physical activity, has important roles in both non-hypertensive and hypertensive individuals (Appel et al., 2006b). In non-hypertensive individuals, lifestyle modifications have the potential to prevent hypertension, reduce BP, and thereby lower the risk of BP-related cardiovascular disease. Even an apparently small reduction in BP, if applied to an entire population, could have an enormous beneficial impact (Stamler, 1991). Specifically, it



Reduction in SBP mm Hg	% Reduction in Mortality		
	Stroke	CHD	Total
2	−6	−4	−3
3	−8	−5	−4
5	−14	−9	−7

Fig. 2 Estimated effects of population-wide shifts in systolic BP on mortality. Reproduced from Stamler, R., 1991. Implication of the intersalt study. Hypertension 17, I-16–I-20, with permission from LWW.

has been estimated that a 3 mm Hg reduction in systolic BP could lead to an 8% reduction in stroke mortality and a 5% reduction in mortality from coronary heart disease (Fig. 2). In hypertensive individuals, lifestyle modifications can serve as initial treatment before the start of drug therapy and as an adjunct to medication in people already on antihypertensive drug therapy. In hypertensive individuals with medication-controlled BP, lifestyle therapies can facilitate drug step-down and potentially drug withdrawal in individuals who sustain lifestyle changes (Whelton et al., 1998).

High BP results from dietary factors, other environmental factors, genetic factors, and their interactions (Appel et al., 2006a; Levy et al., 2009). Of these factors, diet likely has a predominant role in BP homeostasis. Dietary factors that effectively lower BP are weight loss, reduced dietary sodium intake, increased dietary potassium intake, moderation of alcohol intake (among those who drink), and adoption of a DASH-style or vegetarian dietary pattern. Other aspects of diet might also affect BP, but the effects are small and/or the evidence uncertain.

Dietary factors that lower BP

Weight loss

On average, as weight increases, so does BP. The importance of this relationship is reinforced by the high and increasing prevalence of overweight and obesity throughout the world. With rare exception, clinical trials have documented that weight loss lowers BP. Importantly, reductions in BP occur before and without attainment of a desirable body weight. In one meta-analysis that aggregated results across 25 trials, mean systolic and diastolic BP reductions from an average weight loss of 5.1 kg were 4.4 and 3.6 mm Hg, respectively (Neter et al., 2003). Greater weight loss leads to greater BP reduction. Still, the long-term effects of weight loss on BP are unclear, with some studies suggesting that BP reductions attenuate over time (Stevens et al., 2001).

Additional trials have documented that modest weight loss can prevent hypertension by approximately 20% among overweight, individuals without hypertension, and can facilitate medication step-down and drug withdrawal (1997). Lifestyle intervention trials have uniformly achieved short-term weight loss, primarily through a reduction in total calorie intake. In some instances, substantial weight loss has also been sustained over 3 or more years (TOHP Collaborative Research Group, 1997; Svetkey et al., 2008).

In aggregate, available evidence strongly supports weight reduction, ideally attainment of a body mass index less than 25 kg m^{-2}, as an effective approach to prevent and treat hypertension. Weight reduction can also prevent diabetes and control lipids. Hence, the beneficial effects of weight reduction in preventing cardiovascular–renal disease should be substantial. Finally, in view of the well-recognized challenges of maintaining weight loss, efforts to prevent weight gain among those with a normal body weight are critical.

Reduced salt intake

On average, as dietary salt (sodium chloride) intake rises, so does BP. (In view of the format of published data and of most dietary recommendations, data are presented as g/day (mmol/day) of sodium rather than g/day of salt.) To date, over 100 randomized trials have tested the effects of salt on BP, including dose–response trials; several meta-analyses have aggregated data across trials. In a recent meta-analysis (Mozaffarian et al., 2014), a reduction in sodium intake of 2.3 g/d lowered systolic BP by 3.8 in adults; larger BP reductions occurred in older than younger persons, blacks compared to whites, and hypertensive individuals compared to normotensive individuals.

One of the most important dose–response trials is the Dietary Approaches to Stop Hypertension (DASH)-Sodium trial, which tested the effects of three different salt intakes separately in two distinct diets—the DASH diet and a control diet more typical of what Americans eat (Sacks et al., 2001). As displayed in Fig. 3, the rise in BP with higher salt intake was evident in both diets. The BP response to salt intake was nonlinear. Specifically, decreasing salt intake caused a greater lowering of BP when the starting sodium intake was less than 2.3 g day^{-1} (100 mmol day^{-1}) than when it was above this level.

The BP response to changes in salt intake is heterogeneous. Despite the use of the terms “salt sensitive” and “salt resistant” to classify individuals in research studies, the change in BP in response to a change in salt intake is not binary. Rather, the change in BP from a reduced salt intake has a continuous distribution, with individuals having greater or lesser degrees of BP reduction. Genetic factors influence the response to salt reduction. Concomitant diet also modifies the effects of salt on BP. The rise in BP for a given increase in salt intake is blunted in the setting of either the DASH diet or a high potassium intake (Fig. 3). In general, the effects of salt on BP tend to be greater in blacks, middle-aged and older people, and individuals with hypertension, or chronic kidney disease (McMahon et al., 2015). Importantly, although it is possible to identify groups that tend to be salt sensitive, it is impossible, given currently available diagnostic tools, to identify individuals who are salt sensitive.

In addition to lowering BP, clinical trials have documented that a reduced salt intake can prevent hypertension by approximately 20% (with or without concomitant weight loss) (1997) and can lower BP in the setting of antihypertensive medication (Appel et al., 2001). Evidence from observational studies suggests that a reduced salt intake can blunt the age-related rise in systolic BP (Fig. 4) (Rose et al., 1988). A reduced salt intake may also reduce the risk of left ventricular hypertrophy, osteoporosis, and gastric cancer (Institute of Medicine, 2005). Still, the evidence base does have limitations, largely because of the difficulty in measuring sodium intake and the potential for reverse causality. For these reasons, results from observational studies have been inconsistent and occasionally paradoxical with some studies suggesting that a low sodium intake might be harmful. However, such analyses are fraught

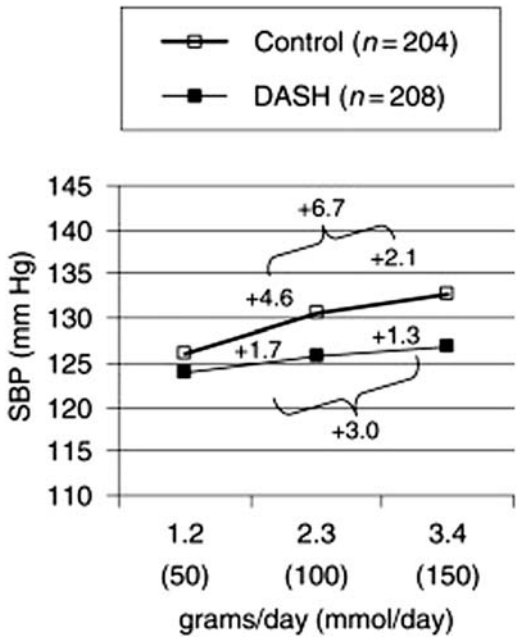


Fig. 3 Mean systolic BP change in the DASH-Sodium Trial from salt reduction in two diets and from the DASH diet at three salt levels. Reproduced with permission from Sacks, F.M., Svetkey, L.P., Vollmer, W.M., et al., 2001. A clinical trial of the effects on BP of reduced dietary sodium and the DASH dietary pattern (The DASH-Sodium Trial). N. Engl. J. Med. 344, 3–10.

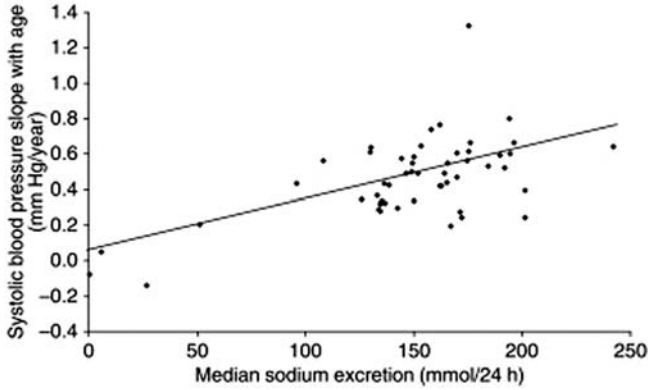


Fig. 4 Slope of systolic BP increase with age plotted by median sodium excretion in 52 communities worldwide: results from intersalt. Reproduced with permission from Rose, G., Stamler, J., Stamler, R., et al., 1988. Intersalt: an International Study of electrolyte excretion and BP. Results for 24 h urinary sodium and potassium excretion. Br. Med. J. 297, 319–328.

with methodologic issues (Cobb et al., 2014) and have been largely dismissed in policy-making (National Academies of Sciences, 2019).

Recently, two major studies have provided strong evidence is support of population-wide sodium reduction. First, a major trial documented that replacement of regular salt with a potassium-enriched salt substitute significantly reduced the risk of stroke (14%) and total mortality (12%) in rural China (Neal et al., 2021). Second, a meta-analysis of observational studies in generally healthy adults documented that higher sodium and lower potassium intakes, as measured in multiple 24 h urine samples, were associated in a dose–response manner with a higher cardiovascular risk, without evidence of harm at lower levels of sodium excretion (Ma et al., 2021).

Still, the effects of salt on health have been debated. Some have argued that the increases in plasma renin activity, uric acid and perhaps insulin resistance that occur as a result of a reduced salt intake mitigate the beneficial effects of salt reduction on BP (Graudal et al., 2012). However, in contrast to BP, the clinical relevance of increased plasma renin activity and uric acid is uncertain, especially because antihypertensive medications that raise plasma renin levels and uric acid actually lower cardiovascular disease risk. It has also been argued that a reduced salt intake has little or no effect on BP in many individuals.

Available evidence strongly supports population-wide sodium reduction, as recommended by the World Health Organization (WHO), most governments, and numerous organizations. The WHO has recommended that adults consume less than 5 g of salt per

day, corresponding to less than 2000 g of sodium per day. A National Academy of Medicine (NAM) committee recommended an upper limit of 2300 g of sodium per day (National Academies of Sciences, 2019). In view of the high burden of BP-related disease in middle and older-aged persons, and individuals with hypertension and chronic kidney disease, several authoritative bodies have recommended an even lower intake of sodium (Lloyd-Jones et al., 2010). The NAM committee set a lower limit, termed Adequate Intake (AI) of 1500 g of sodium per day. This level of salt intake allows for excess sweat salt loss among unacclimatized individuals who become physically active or who become exposed to high temperatures.

In most Western counties, average intake of sodium is high, greatly exceeding 2.3 g day^{-1} (100 mmol day^{-1}). In the United States, the mean intake of sodium from foods, not including salt added at the table, is ~4200 g/day in adult men and ~3000 in adult women (Bailey et al., 2015). In addition to gender, sodium intake varies considerably by age. Worldwide, there is substantial variation in sodium intake, ranging from an estimated mean intake of 0.02 g day^{-1} (1.0 mmol day^{-1}) in Yanomamo Indians to more than 10.3 g day^{-1} (450 mmol day^{-1}) in northern Japanese (Rose et al., 1988).

In aggregate, available data strongly support current population-wide recommendations to lower salt intake. In some regions of the world, e.g., rural China, most dietary sodium comes from the addition of salt during cooking (Anderson et al., 2010). In this context, counseling and replacement of regular salt with potassium-enriched, reduced-sodium salt are the preferred, sodium reduction strategies. However, in the vast majority of countries, the primary source of sodium is processed and restaurant foods, which account for more than 70% of sodium intake (**Fig. 5**) (Harnack et al., 2017). Hence, any meaningful strategy to reduce sodium intake must involve the efforts of food manufacturers, who should reduce the amount of salt added during food processing.

Increased potassium intake

High levels of potassium intake are associated with reduced BP. Observational data have been reasonably consistent in documenting this inverse relationship, whereas data from individual trials, which typically achieve a higher potassium intake through pill supplements, have been less consistent. Meta-analyses of these trials have documented a significant inverse relationship between potassium intake and BP in studies of hypertensive individuals, and non-significant effects in studies of non-hypertensive persons. In one meta-analysis, average net systolic/diastolic BP reductions from increased potassium intake were 3.5/2.0 mm Hg (Aburto et al., 2013). Available studies have documented greater BP reductions from potassium in blacks compared to non-blacks and in hypertensive compared to non-hypertensive individuals. A high potassium intake appears to blunt the rise in BP in response to increased salt intake (Morris et al., 1999). Potassium appears to have greater BP lowering in the context of a higher salt intake and lesser BP reduction in the setting of a lower salt intake. Conversely, the BP reduction from a reduced salt intake is greatest when potassium intake is low. These data are consistent with subadditive effects of reduced salt intake and increased potassium intake on BP.

Most trials that tested the effects of potassium on BP used pill supplements, typically potassium chloride. However, in foods, the conjugate anions associated with potassium are mainly citrate and other bicarbonate precursors. The latter is important because other potential benefits of foods rich in potassium (i.e., reduced risk of kidney stones and reduced bone turnover) likely result from effects of the conjugate anion. Because a high dietary intake of potassium can be achieved through diet rather than pills and because potassium derived from foods also comes with a variety of other nutrients, the preferred strategy to increase potassium intake is to consume foods, such as fruits and vegetables, rather than potassium supplements.

On the basis of available data, a National Academy of Medicine committee set an Adequate Intake for potassium of 3400 mg per day for healthy men and 2600 mg per day for healthy women (National Academies of Sciences, 2019). This level of dietary intake should lower BP levels and have the potential benefits of reducing the adverse effects of salt on BP, reducing the risk of kidney stones, and possibly decreasing bone loss (Institute of Medicine (U.S.) Panel on Dietary Reference Intakes for Electrolytes and Water, 2005). Currently, dietary intake of potassium is lower than these levels. In recent surveys, the mean intake of potassium in the United States was ~3200 mg/day in adult men and ~2400 mg/day in adult women (Bailey et al., 2015). Because African Americans have a relatively low intake of potassium and a high prevalence of high BP and salt sensitivity, this subgroup of the population would especially benefit from an increased potassium intake.

In the generally healthy population with normal kidney function, a potassium intake from foods higher than 3400 mg/day poses no potential for increased risk because excess potassium is readily excreted in the urine. However, in individuals whose

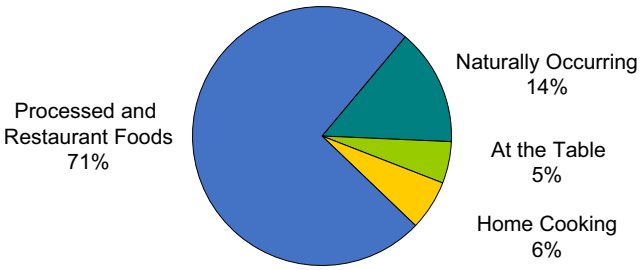


Fig. 5 Sources of dietary sodium. Data from Harnack, L.J., et al., 2017. Sources of sodium in US adults from 3 Geographic Regions. Circulation 135, 1775–83.

urinary potassium excretion is impaired, a potassium intake of less than 3400 mg/day might be appropriate because of the potential for adverse cardiac effects (arrhythmias) from hyperkalemia. Common drugs that impair potassium excretion are angiotensin converting enzyme inhibitors, angiotensin receptor blockers, and potassium-sparing diuretics. Medical conditions associated with impaired potassium excretion include diabetes, chronic kidney disease, end stage kidney disease, severe heart failure, and adrenal insufficiency. Elderly individuals are at increased risk of hyperkalemia because they often have one or more of these conditions or take one or more of the medications that impair potassium excretion.

Moderation of alcohol intake

Alcohol intake raises BP, particularly at an alcohol intake above approximately two drinks per day (~1 oz. or ~28 g of ethanol per day). In a meta-analysis of 32 trials with 767 participants, the effects of alcohol differed by dose and time after ingestion. Low-dose alcohol (<14 g/d) had no effect on BP. Medium-dose alcohol (14–28 g/d) lowered BP soon after ingestion, within 6 h, but did not affect BP after 6 h. However, high-dose alcohol (>30 g/d) lowered BP within 6 h and up to 12 h later. However, after 12 h, high-dose alcohol increased systolic BP by a mean of 3.7 mmHg and diastolic BP by a mean of 2.4 mmHg (Tasnim et al., 2020). In non-hypertensives and hypertensives, BP reductions did not differ, potentially because of small sample sizes. In aggregate, evidence supports recommendations to lower alcohol intake (among those who drink) as an approach to lower BP. Alcohol consumption be limited to no more than 1 oz. (30 mL) of ethanol (e.g., 24 oz. (720 mL) beer, 10 oz. (300 mL) wine, or 2 oz. (60 mL) 100-proof whiskey) per day in most men and to no more than 0.5 oz. (15 mL) ethanol per day in women and lighter weight people.

Whole dietary patterns

Vegetarian diets

Vegetarian diets have been associated with low BP. In observational studies, vegetarians also experience a markedly lower, age-related rise in BP (Yokoyama et al., 2014). Aspects of a vegetarian lifestyle that might affect BP include nondietary factors (e.g., physical activity), established dietary risk factors (e.g., salt, potassium, weight, and alcohol), and other aspects of a vegetarian diet (e.g., high fiber and no meat). To a very limited extent, observational studies have controlled for the well-established determinants of BP. In a meta-analysis of trials of dietary patterns, consumption of a vegan dietary pattern had no significant effect on BP, while a lacto-ovo vegetarian patterns significantly reduced systolic and diastolic BP by 5.5 and 2.5 mmHg, respectively (Gibbs et al., 2021).

DASH-style dietary patterns

The DASH trial tested whether modification of whole dietary patterns affect BP (Appel et al., 1997). In this trial, participants were randomized to eat one of three diets: (1) a control diet; (2) a diet rich in "fruits and vegetables" but otherwise similar to control; or (3) the DASH diet. The DASH diet emphasizes fruits, vegetables, and low-fat dairy products; includes whole grains, poultry, fish, and nuts; and is reduced in fats, red meat, sweets, and sugar-containing beverages. Accordingly, it is rich in potassium, magnesium, calcium, and fiber and reduced in total fat, fat, and cholesterol; it is also slightly increased in protein.

Among all participants, the DASH diet significantly lowered mean systolic BP by 5.5 mm Hg and mean diastolic BP by 3.0 mm Hg. The fruits and vegetables diet also reduced BP but to a lesser extent—approximately half of the effect of the DASH diet. The effect was relatively rapid; the full effect was apparent after 2 weeks (**Fig. 6**). In subgroup analyses, the DASH diet significantly lowered BP in all major subgroups (men, women, Blacks, non-Blacks, hypertensives, and non-hypertensives) (Svetkey et al., 1999). However, the effects of the DASH diet were especially prominent in Blacks, who experienced net systolic/diastolic BP reductions of 6.9/3.7 mm Hg, and hypertensive individuals, who experienced net BP reductions of 11.6/5.3 mm Hg.

Subsequently, the OmniHeart trial compared three variants of the DASH diet (a diet rich in carbohydrate (58% of calories), a second version rich in protein (approximately half from plant sources), and a third diet rich in unsaturated fat (predominantly monounsaturated fat)) (Appel et al., 2005). In several respects, each diet was similar to the DASH diet—each was reduced in saturated fat, cholesterol, and sodium, and rich in fruit, vegetables, fiber, and potassium at recommended levels. Although each diet lowered systolic BP, substituting some of the carbohydrate (approximately 10% of total kcal) with either protein (approximately half from plant sources) or with unsaturated fat (mostly monounsaturated fat) further lowered BP.

Results from the DASH and OmniHeart trials have important clinical and public health implications. The effect of the DASH diet in hypertensive individuals was similar in magnitude to that of drug monotherapy. From a public health perspective, the DASH diet could potentially shift the population distribution of BP downward, thereby reducing the risk of BP-related cardiovascular disease (**Fig. 2**). In a meta-analysis of 11 trials (n = 1400 participants) that tested the effects of the DASH diet on BP, adoption of the DASH diet lowered mean systolic and diastolic BP by 5.5 and 3.8 mmHg, respectively (Gibbs et al., 2021). In the same meta-analysis, adoption of a Mediterranean style diet lowered mean systolic BP by ~1 mmHg and had no significant effect on diastolic BP.

Protein intake

A large and generally consistent body of evidence from observational studies and clinical trials has documented that higher protein intake, particularly protein from plant-based sources, is associated with lower BP (Rebholz et al., 2012; Elliott et al., 2006). Several trials have tested the effects of soy-based interventions on BP. In several but not all of these trials, soy supplementation reduced BP; however, it is unclear whether the BP reductions are due to soy protein or isoflavones (He et al., 2011). As previously discussed, the

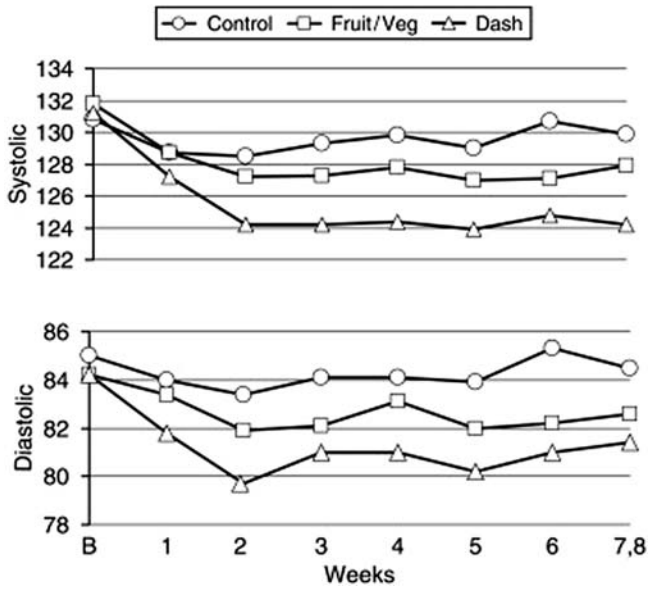


Fig. 6 BP by week during the DASH feeding study in three diets (control diet, fruits and vegetables diet, and DASH diet). Reproduced with permission from Appel, L.J., Moore, T.J., Obarzanek, E., et al., 1997. The effect of dietary patterns on BP: results from the Dietary Approaches to Stop Hypertension (DASH) clinical trial. N. Engl. J. Med. 336, 1117–1124.

OmniHeart trial documented that a dietary pattern with increased protein intake from mixed sources, predominantly plants, lowered BP (Appel et al., 2005).

Dietary factors with limited or uncertain effect on BP

Fiber

Evidence from observational studies and several clinical trials suggests that increased fiber intake may reduce BP. A meta-analysis documented that supplemental fiber (average net increase of 11 g day^{-1}) was associated with net systolic/diastolic reductions of 1.2/1.7 mm Hg, respectively (Whelton et al., 2005). Still, high-quality epidemiological studies and clinical trials are needed before one can recommend increased fiber intake as a means to lower BP.

Calcium and magnesium

Evidence that increased calcium intake might lower BP comes from a variety of sources, including animal studies, observational studies, clinical trials, and meta-analyses. Meta-analyses of trials documented small reductions in systolic and diastolic BP with calcium supplementation (400–2000 mg/day) (Allender et al., 1996). There is also evidence that calcium intake may affect the BP response to salt. Overall, data are insufficient to recommend supplemental calcium alone as a means to lower BP.

The body of evidence implicating magnesium as a major determinant of BP is likewise inconsistent. In observational studies, often cross-sectional in design, a common finding is an inverse association of dietary magnesium with BP (Jee et al., 2002). However, in pooled analyses of clinical trials, there was no clear effect of magnesium intake on BP. Hence, data are insufficient to recommend increased magnesium intake alone as a means to lower BP.

Fish oil supplementation

High-dose, ω-3 polyunsaturated fatty acid (commonly termed "fish oil") supplements can lower BP in hypertensive individuals. In a meta-analysis of 17 trials with a total of 1524 participants, high dose fish oil reduced systolic and diastolic BP by a mean of 2.6 and 1.5 mmHg, respectively, in persons with hypertension, but had no significant effect in non-hypertensive persons. It is noteworthy that the dose of fish oil was relatively high in many trials, i.e., 3 g or more per day (Campbell et al., 2013). Side effects from fish oil supplements, particularly belching and a fishy taste, are common. In view of the side effect profile and the high dose required to lower BP, fish oil supplements are not routinely recommended as a means to lower BP.

Fats (other than fish oil) and cholesterol

Numerous studies, including both observational studies and clinical trials, have examined the effects of fat intake on BP. Overall, there is no apparent effect of saturated fat and *n*-6 polyunsaturated fat intake on BP. Although a few trials suggest that an increased intake of monounsaturated fat may lower BP, evidence is insufficient to make recommendations. Likewise, few studies have examined the effect of dietary cholesterol intake on BP. Hence, although modification of dietary fat and cholesterol intake can be recommended as a means to prevent and treat hyperlipidemia and dyslipidemia, evidence is insufficient to recommend these changes alone as a means to lower BP.

Vitamin C

Laboratory studies, depletion–repletion studies, and epidemiological studies suggest that increased vitamin C intake or status is associated with lower BP. Several trials, many with methodological limitations, have also addressed this issue (Juraschek et al., 2012). Overall, it remains unclear whether an increased intake of vitamin C lowers BP.

Gene–diet interactions

A rapidly increasing body of evidence indicates that genetic factors affect BP levels and the BP response to dietary changes. Most of the evidence relates to genetic factors that affect the BP response to salt (Elijovich et al., 2016). Several genotypes that influence BP have been identified. Most of these genotypes influence the renin–angiotensin–aldosterone axis or renal salt handling.

Special populations

Children

High BP begins well before adulthood, during the first two decades of life and likely during gestation. In additional to the age-related rise in BP observed in children, numerous studies have documented that BP tracks with age from childhood into the adult years (Dekkers et al., 2002). Hence, efforts to reduce BP and to prevent the age-related rise in BP in childhood are prudent.

With the exception of sodium, few trials have tested the effects of dietary factors as a mean to lower BP in children and adolescents. In a meta-analysis of 10 trials conducted in children, sodium reduction significantly lowered BP (He and MacGregor, 2006). Because few trials have tests dietary interventions in children, the effect of diet on BP in children and adolescents is, in large part, extrapolated from trials conducted in adults. Such extrapolations are reasonable because high BP is a chronic condition resulting from the insidious rise in BP throughout childhood and adulthood.

Pregnant women

Hypertension during pregnancy is a constellation of diverse clinical conditions, some of which can be extremely serious. Of substantial concern are preeclampsia and eclampsia. Both are multisystem disorders that are manifest by the onset of hypertension and proteinuria during the second half of pregnancy. Convulsions occur in the setting of eclampsia but not preeclampsia. Many studies have explored the possibility that dietary factors cause hypertensive disorders in pregnancy. In clinical trials, several dietary interventions, including sodium reduction, fish oil supplementation, and calcium supplementation, have been tested as a means to prevent preeclampsia, but none were effective. Currently, the focus of research on the pathogenesis of pre-eclampsia is abnormal placental development and subsequent placental ischemia.

Older people

Because of the age-related rise in systolic BP and because of the high prevalence of BP-related cardiovascular disease in middle-aged and older people, dietary strategies should be especially beneficial as adults age. It is well documented that older people can make and sustain dietary changes, specifically weight loss and dietary salt reduction (Whelton et al., 1998). Furthermore, salt sensitivity increases as individuals age. Lastly, because of the high attributable risk associated with high BP in older people, the beneficial effects of dietary changes on BP should translate into substantial reductions in cardiovascular risk in this age group.

Populations defined by race/ethnicity or geography

The worldwide prevalence of hypertension is enormous—an estimated 1.4 billion have hypertension, of whom over 1 billion live in low- and middle-income countries (Mills et al., 2016). Yet, such staggering estimates mask the fact that there is substantial variation in BP among populations. In certain primitive societies, such as the Yanomamo Indians in Brazil, BP does not rise with age, and hypertension is absent (Mueller and Appel, 2019). Until recently, rural populations tended to have a lower population than corresponding urban populations. However, this pattern is now less apparent with rural populations having a similar, or even higher,

prevalence of hypertension. Among urbanized populations, the prevalence of hypertension is high, especially among African Americans, a population in which the prevalence of hypertension approaches 40%. Other groups, such as Australian Aborigines, Eastern Europeans, and Russians, also have a high prevalence of hypertension.

Understanding the causes of geographic variation is difficult. However, migration studies provide strong evidence that modifiable environmental factors (e.g., diet and physical activity) rather than genetic factors or geographic factors account for this variation (He et al., 1991; Poulter et al., 1990). Furthermore, as noted previously, trials have documented that compared to non-blacks, blacks achieve greater BP reduction from several nonpharmacological therapies, specifically a reduced salt intake, increased potassium intake, and the DASH diet. The potential benefits of these dietary therapies is amplified because US survey data indicate that Blacks consume less potassium than non-Blacks, while salt intake is high and similar in Blacks and non-Blacks. In this context, changes in diet should provide a means to reduce racial and perhaps geographic disparities in BP.

Conclusion

In view of the continuing epidemic of BP-related cardiovascular disease, efforts to reduce BP in both non-hypertensive and hypertensive individuals are warranted. Such efforts will require individuals to change behavior and society to make substantial environmental changes. The current challenge to the general public, health care providers, researchers, and policy makers is to develop and implement effective clinical and public health strategies that lead to sustained dietary changes among individuals and more broadly among populations.

References

Aburto, N.J., Hanson, S., Gutierrez, H., Hooper, L., Elliott, P., Cappuccio, F.P., 2013. Effect of increased potassium intake on cardiovascular risk factors and disease: systematic review and meta-analyses. BMJ 346, f1378.

Allender, P.S., Cutler, J.A., Follmann, D., Cappuccio, F.P., Pryer, J., Elliott, P., 1996. Dietary calcium and blood pressure: a meta-analysis of randomized clinical trials. Ann. Intern. Med. 124, 825–831.

Anderson, C.A., Appel, L.J., Okuda, N., Brown, I.J., Chan, Q., Zhao, L., Ueshima, H., Kesteloot, H., Miura, K., Curb, J.D., Yoshita, K., Elliott, P., Yamamoto, M.E., Stamler, J., 2010. Dietary sources of sodium in China, Japan, the United Kingdom, and the United States, women and men aged 40 to 59 years: the INTERMAP Study. J. Am. Diet Assoc. 110, 736–745.

Appel, L.J., Moore, T.J., Obarzanek, E., Vollmer, W.M., Svetkey, L.P., Sacks, F.M., Bray, G.A., Vogt, T.M., Cutler, J.A., Windhauser, M.M., Lin, P.H., Karanja, N., 1997. A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group. N. Engl. J. Med. 336, 1117–1124.

Appel, L.J., Espeland, M.A., Easter, L., Wilson, A.C., Folmar, S., Lacy, C.R., 2001. Effects of reduced sodium intake on hypertension control in older individuals: results from the Trial of Nonpharmacologic Interventions in the Elderly (TONE). Arch. Intern. Med. 161, 685–693.

Appel, L.J., Sacks, F.M., Carey, V.J., Obarzanek, E., Swain, J.F., Miller III, E.R., Conlin, P.R., Erlinger, T.P., Rosner, B.A., Laranjo, N.M., Charleston, J., Mccarron, P., Bishop, L.M., 2005. Effects of protein, monounsaturated fat, and carbohydrate intake on blood pressure and serum lipids: results of the OmniHeart randomized trial. JAMA 294, 2455–2464.

Appel, L.J., Brands, M.W., Daniels, S.R., Karanja, N., Elmer, P.J., Sacks, F.M., American Heart Association, 2006. Dietary approaches to prevent and treat hypertension: a scientific statement from the American Heart Association. Hypertension 47, 296–308.

Bailey, R.L., Parker, E.A., Rhodes, D.G., Goldman, J.D., Clemens, J.C., Moshfegh, A.J., Thuppal, S.V., Weaver, C.M., 2015. Estimating sodium and potassium intakes and their ratio in the American diet: data from the 2011–2012 NHANES. J. Nutr. 146, 745–750.

Campbell, F., Dickinson, H.O., Critchley, J.A., Ford, G.A., Bradburn, M., 2013. A systematic review of fish-oil supplements for the prevention and treatment of hypertension. Eur. J. Prev. Cardiol. 20, 107–120.

Cobb, L.K., Anderson, C.A., Elliott, P., Hu, F.B., Liu, K., Neaton, J.D., Whelton, P.K., Woodward, M., Appel, L.J., American Heart Association Council on Lifestyle and Metabolic Health, 2014. Methodological issues in cohort studies that relate sodium intake to cardiovascular disease outcomes: a science advisory from the American Heart Association. Circulation 129, 1173–1186.

Dekkers, J.C., Snieder, H., Van Den Oord, E.J., Treiber, F.A., 2002. Moderators of blood pressure development from childhood to adulthood: a 10-year longitudinal study. J. Pediatr. 141, 770–779.

Elijovich, F., Weinberger, M.H., Anderson, C.A., Appel, L.J., Bursztyn, M., Cook, N.R., Dart, R.A., Newton-Cheh, C.H., Sacks, F.M., Laffer, C.L., American Heart Association Professional and Public Education Committee of the Council on Hypertension; Council on Functional Genomics and Translational Biology; and Stroke Council, 2016. Salt sensitivity of blood pressure: a scientific statement from the American Heart Association. Hypertension 68, e7–e46.

Elliott, P., Stamler, J., Dyer, A.R., Appel, L., Dennis, B., Kesteloot, H., Ueshima, H., Okayama, A., Chan, Q., Garside, D.B., Zhou, B., 2006. Association between protein intake and blood pressure: the INTERMAP Study. Arch. Intern. Med. 166, 79–87.

Forouzanfar, M.H., Liu, P., Roth, G.A., NG, M., Biryukov, S., Marczak, L., Alexander, L., Estep, K., Hassen Abate, K., Akinyemiju, T.F., Ali, R., Alvis-Guzman, N., Azzopardi, P., Banerjee, A., Barnighausen, T., Basu, A., Bekele, T., Bennett, D.A., Biadgilign, S., Catala-Lopez, F., Feigin, V.L., Fernandes, J.C., Fischer, F., Gebru, A.A., Gona, P., Gupta, R., Hankey, G.J., Jonas, J.B., Judd, S.E., Khang, Y.H., Khosravi, A., Kim, Y.J., Kimokoti, R.W., Kokubo, Y., Kolte, D., Lopez, A., Lotufo, P.A., Malekzadeh, R., Melaku, Y.A., Mensah, G.A., Misganaw, A., Mokdad, A.H., Moran, A.E., Nawaz, H., Neal, B., Ngalesoni, F.N., Ohkubo, T., Pourmalek, F., Rafay, A., Rai, R.K., Rojas-Rueda, D., Sampson, U.K., Santos, I.S., Sawhney, M., Schutte, A.E., Sepanlou, S.G., Shifa, G.T., Shiue, I., Tedla, B.A., Thrift, A.G., Tonelli, M., Truelsen, T., Tsilimparis, N., Ukwaja, K.N., Uthman, O.A., Vasankari, T., Venketasubramanian, N., Vlassov, V.V., Vos, T., Westerman, R., Yan, L.L., Yano, Y., Yonemoto, N., Zaki, M.E., Murray, C.J., 2017. Global burden of hypertension and systolic blood pressure of at least 110 to 115 mm Hg, 1990–2015. JAMA 317, 165–182.

GBD Collaborators, 2018. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 392, 1923–1994.

Gibbs, J., Gaskin, E., Ji, C., Miller, M.A., Cappuccio, F.P., 2021. The effect of plant-based dietary patterns on blood pressure: a systematic review and meta-analysis of controlled intervention trials. J. Hypertens. 39, 23–37.

Graudal, N.A., Hubeck-Graudal, T., Jurgens, G., 2012. Effects of low-sodium diet vs. high-sodium diet on blood pressure, renin, aldosterone, catecholamines, cholesterol, and triglyceride (Cochrane Review). Am. J. Hypertens. 25, 1–15.

Harnack, L.J., Cogswell, M.E., Shikany, J.M., Gardner, C.D., Gillespie, C., Loria, C.M., Zhou, X., Yuan, K., Steffen, L.M., 2017. Sources of sodium in US adults from 3 geographic regions. Circulation 135, 1775–1783.

He, F.J., Macgregor, G.A., 2006. Importance of salt in determining blood pressure in children: meta-analysis of controlled trials. Hypertension 48, 861–869.

He, J., Klag, M.J., Whelton, P.K., Chen, J.Y., Mo, J.P., Qian, M.C., Mo, P.S., He, G.Q., 1991. Migration, blood pressure pattern, and hypertension: the Yi Migrant Study. Am. J. Epidemiol. 134, 1085–1101.

He, J., Wofford, M.R., Reynolds, K., Chen, J., Chen, C.S., Myers, L., Minor, D.L., Elmer, P.J., Jones, D.W., Whelton, P.K., 2011. Effect of dietary protein supplementation on blood pressure: a randomized, controlled trial. Circulation 124, 589–595.

Institute Of Medicine (U.S.) Panel On Dietary Reference Intakes For Electrolytes And Water, 2005. DRI, Dietary Reference Intakes for Water, Potassium, Sodium, Chloride, and Sulfate. National Academies Press, Washington, D.C.

Jee, S.H., Miller III, E.R., Guallar, E., Singh, V.K., Appel, L.J., Klag, M.J., 2002. The effect of magnesium supplementation on blood pressure: a meta-analysis of randomized clinical trials. Am. J. Hypertens. 15, 691–696.

Juraschek, S.P., Guallar, E., Appel, L.J., Miller III, E.R., 2012. Effects of vitamin C supplementation on blood pressure: a meta-analysis of randomized controlled trials. Am. J. Clin. Nutr. 95, 1079–1088.

Levy, D., Ehret, G.B., Rice, K., Verwoert, G.C., Launer, L.J., Dehghan, A., Glazer, N.L., Morrison, A.C., Johnson, A.D., Aspelund, T., Aulchenko, Y., Lumley, T., Köttgen, A., Vasan, R.S., Rivadeneira, F., Eiriksdottir, G., Guo, X., Arking, D.E., Mitchell, G.F., Mattace-Raso, F.U.S., Smith, A.V., Taylor, K., Scharpf, R.B., Hwang, S.-J., Sijbrands, E.J.G., Bis, J., Harris, T.B., Ganesh, S.K., O'donnell, C.J., Hofman, A., Rotter, J.I., Coresh, J., Benjamin, E.J., Uitterlinden, A.G., Heiss, G., Fox, C.S., Witteman, J.C.M., Boerwinkle, E., Wang, T.J., Gudnason, V., Larson, M.G., Chakravarti, A., Psaty, B.M., Van Duijn, C.M., 2009. Genome-wide association study of blood pressure and hypertension. Nat. Genet. 41, 677–687.

Lewington, S., Clarke, R., Qizilbash, N., Peto, R., Collins, R., 2002. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet 360, 1903–1913.

Lloyd-Jones, D.M., Hong, Y., Labarthe, D., Mozaffarian, D., Appel, L.J., Van Horn, L., Greenlund, K., Daniels, S., Nichol, G., Tomaselli, G.F., Arnett, D.K., Fonarow, G.C., Ho, P.M., Lauer, M.S., Masoudi, F.A., Robertson, R.M., Roger, V., Schwamm, L.H., Sorlie, P., Yancy, C.W., Rosamond, W.D., American Heart Association Strategic Planning Task Force & Statistics Committee, 2010. Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic impact goal through 2020 and beyond. Circulation 121, 586–613.

Ma, Y., He, F.J., Sun, Q., Yuan, C., Kieneker, L.M., Curhan, G.C., Macgregor, G.A., Bakker, S.J.L., Campbell, N.R.C., Wang, M., Rimm, E.B., Manson, J.E., Willett, W.C., Hofman, A., Gansevoort, R.T., Cook, N.R., Hu, F.B., 2021. 24-Hour urinary sodium and potassium excretion and cardiovascular risk. N. Engl. J. Med. 386 (3).

Mcmahon, E.J., Campbell, K.L., Bauer, J.D., Mudge, D.W., 2015. Altered dietary salt intake for people with chronic kidney disease. Cochrane Database Syst. Rev. CD010070.

Mills, K.T., Bundy, J.D., Kelly, T.N., Reed, J.E., Kearney, P.M., Reynolds, K., Chen, J., He, J., 2016. Global disparities of hypertension prevalence and control: a systematic analysis of population-based studies from 90 countries. Circulation 134, 441–450.

Morris Jr., R.C., Sebastian, A., Forman, A., Tanaka, M., Schmidlin, O., 1999. Normotensive salt sensitivity: effects of race and dietary potassium. Hypertension 33, 18–23.

Mozaffarian, D., Fahimi, S., Singh, G.M., Micha, R., Khatibzadeh, S., Engell, R.E., Lim, S., Danaei, G., Ezzati, M., Powles, J., 2014. Global sodium consumption and death from cardiovascular causes. N. Engl. J. Med. 371, 624–634.

Mueller, N.T., Appel, L.J., 2019. Absence of the association of blood pressure with age in a remote Venezuelan population renews the call for population-wide interventions-reply. JAMA Cardiol. 4, 497.

National Academies of Sciences, Engineering, and Medicine, 2019. Dietary Reference Intakes for Sodium and Potassium. The National Academies Press, Washington (DC).

Neal, B., Wu, Y., Feng, X., Zhang, R., Zhang, Y., Shi, J., Zhang, J., Tian, M., Huang, L., Li, Z., Yu, Y., Zhao, Y., Zhou, B., Sun, J., Liu, Y., Yin, X., Hao, Z., Yu, J., Li, K.C., Zhang, X., Duan, P., Wang, F., Ma, B., Shi, W., Di Tanna, G.L., Stepien, S., Shan, S., Pearson, S.A., Li, N., Yan, L.L., Labarthe, D., Elliott, P., 2021. Effect of salt substitution on cardiovascular events and death. N. Engl. J. Med. 385, 1067–1077.

Neter, J.E., Stam, B.E., Kok, F.J., Grobbee, D.E., Geleijnse, J.M., 2003. Influence of weight reduction on blood pressure: a meta-analysis of randomized controlled trials. Hypertension 42, 878–884.

Poulter, N.R., Khaw, K.T., Hopwood, B.E., Mugambi, M., Peart, W.S., Rose, G., Sever, P.S., 1990. The Kenyan Luo Migration Study: observations on the initiation of a rise in blood pressure. BMJ 300, 967–972.

Rebholz, C.M., Friedman, E.E., Powers, L.J., Arroyave, W.D., He, J., Kelly, T.N., 2012. Dietary protein intake and blood pressure: a meta-analysis of randomized controlled trials. Am. J. Epidemiol. 176 (Suppl. 7), S27–S43.

Rose, G., Stamler, J., Stamler, R., et al., 1988. Intersalt: an international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion. Intersalt Cooperative Research Group. BMJ 297, 319–328.

Sacks, F.M., Svetkey, L.P., Vollmer, W.M., Appel, L.J., Bray, G.A., Harsha, D., Obarzanek, E., Conlin, P.R., Miller III, E.R., Simons-Morton, D.G., Karanja, N., Lin, P.H., 2001. Effects on blood pressure of reduced dietary sodium and the Dietary Approaches to Stop Hypertension (DASH) diet. DASH-Sodium Collaborative Research Group. N. Engl. J. Med. 344, 3–10.

Stamler, R., 1991. Implications of the INTERSALT Study. Hypertension 17, I16–I20.

Stevens, V.J., Obarzanek, E., Cook, N.R., Lee, I.M., Appel, L.J., Smith West, D., Milas, N.C., Mattfeldt-Beman, M., Belden, L., Bragg, C., Millstone, M., Raczynski, J., Brewer, A., Singh, B., Cohen, J., 2001. Long-term weight loss and changes in blood pressure: results of the Trials of Hypertension Prevention, phase II. Ann. Intern. Med. 134, 1–11.

Svetkey, L.P., Simons-Morton, D., Vollmer, W.M., Appel, L.J., Conlin, P.R., Ryan, D.H., Ard, J., Kennedy, B.M., 1999. Effects of dietary patterns on blood pressure: subgroup analysis of the Dietary Approaches to Stop Hypertension (DASH) randomized clinical trial. Arch. Intern. Med. 159, 285–293.

Svetkey, L.P., Stevens, V.J., Brantley, P.J., Appel, L.J., Hollis, J.F., Loria, C.M., Vollmer, W.M., Gullion, C.M., Funk, K., Smith, P., Samuel-Hodge, C., Myers, V., Lien, L.F., Laferriere, D., Kennedy, B., Jerome, G.J., Heinith, F., Harsha, D.W., Evans, P., Erlinger, T.P., Dalcin, A.T., Coughlin, J., Charleston, J., Champagne, C.M., Bauck, A., Ard, J.D., Aicher, K., Weight Loss Maintenance Collaborative Research Group, 2008. Comparison of strategies for sustaining weight loss: the weight loss maintenance randomized controlled trial. JAMA 299, 1139–1148.

Tasnim, S., Tang, C., Musini, V.M., Wright, J.M., 2020. Effect of alcohol on blood pressure. Cochrane Database Syst. Rev. 7, CD012787.

Tohp Collaborative Research Group, 1997. Effects of weight loss and sodium reduction intervention on blood pressure and hypertension incidence in overweight people with high-normal blood pressure. The Trials of Hypertension Prevention, phase II. The Trials of Hypertension Prevention Collaborative Research Group. Arch. Intern. Med. 157, 657–667.

Vasan, R.S., Larson, M.G., Leip, E.P., Evans, J.C., O'donnell, C.J., Kannel, W.B., Levy, D., 2001. Impact of high-normal blood pressure on the risk of cardiovascular disease. N. Engl. J. Med. 345, 1291–1297.

Vasan, R.S., Beiser, A., Seshadri, S., Larson, M.G., Kannel, W.B., D'agostino, R.B., Levy, D., 2002. Residual lifetime risk for developing hypertension in middle-aged women and men: the Framingham Heart Study. JAMA 287, 1003–1010.

Whelton, P.K., Appel, L.J., Espeland, M.A., Applegate, W.B., Ettinger Jr., W.H., Kostis, J.B., Kumanyika, S., Lacy, C.R., Johnson, K.C., Folmar, S., Cutler, J.A., 1998. Sodium reduction and weight loss in the treatment of hypertension in older persons: a randomized controlled trial of nonpharmacologic interventions in the elderly (TONE). TONE Collaborative Research Group. JAMA 279, 839–846.

Whelton, S.P., Hyre, A.D., Pedersen, B., Yi, Y., Whelton, P.K., He, J., 2005. Effect of dietary fiber intake on blood pressure: a meta-analysis of randomized, controlled clinical trials. J. Hypertens. 23, 475–481.

Whelton, P.K., Carey, R.M., Aronow, W.S., Casey Jr., D.E., Collins, K.J., Dennison Himmelfarb, C., Depalma, S.M., Gidding, S., Jamerson, K.A., Jones, D.W., Maclaughlin, E.J., Muntner, P., Ovbiagele, B., Smith Jr., S.C., Spencer, C.C., Stafford, R.S., Taler, S.J., Thomas, R.J., Williams Sr., K.A., Williamson, J.D., Wright Jr., J.T., 2018. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 71, e13–e115.

Yokoyama, Y., Nishimura, K., Barnard, N.D., Takegami, M., Watanabe, M., Sekikawa, A., Okamura, T., Miyamoto, Y., 2014. Vegetarian diets and blood pressure: a meta-analysis. JAMA Intern. Med. 174, 577–587.

Hypoglycemia

R.A. Sulaiman[a,b], **R. Gama**[c], **and V. Marks**[d], [a] Department of Medical Genomics, Center for Genomic Medicine, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; [b] Alfaisal University, Riyadh, Saudi Arabia; [c] School of Medicine and Clinical Practice, University of Wolverhampton, Black Country Pathology Services, Royal Wolverhampton NHS Trust, Clinical Chemistry, Black Country Pathology Services, Wolverhampton, United Kingdom; and [d] Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of V. Marks, Hypoglycemia, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 469–478, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00152-5.

Introduction 402
Brain function and hypoglycemia 402
The blood glucose concentration 402
Mechanism of hypoglycemia 402
The glucose pool in fasting subjects 402
Insulin release in response to eating and fasting 402
The role of the liver in glucose homeostasis 403
Hypoglycemic syndromes 404
Brain malfunction from hypoglycemia 404
Neuroglycopenic syndromes 404
Acute neuroglycopenia (adrenergic symptoms) 404
Sub-acute neuroglycopenia or hypoglycemic unawareness 404
Chronic neuroglycopenia 405
Diagnosis 405
Whipple's triad 405
Causes of hypoglycemia 405
Postprandial (reactive) hypoglycemia 405
Prolonged glucose load test 405
Definition 407
The postprandial syndrome 407
Differential diagnosis 407
Management of reactive (postprandial) hypoglycemia 408
Treatment of attacks 408
Prevention 408
Non-hypoglycemia 409
Exercise-induced hypoglycemia 409
Drug induced hypoglycemia 409
Hepatic and renal failure 409
Sepsis 410
Inborn errors of metabolism 410
Starvation 410
Hypoglycemia in the elderly sick 410
Conclusion 410
References 411
Further reading 411
Relevant websites 411

Key points

- Spontaneous hypoglycemia is not a diagnosis but a manifestation of an underlying disorder.
- Hypoglycemic symptoms may be non-specific. Acute and subacute neuroglycopenia may only be confidently confirmed when Whipple's triad is fulfilled; namely neuroglycopenic symptoms, a low blood glucose and symptoms relieved by raising blood glucose to or above normal.
- Although uncommon, it is important to recognize spontaneous hypoglycemia and its etiology because preventative or curative therapy is often available.

- Appropriate laboratory tests in hypoglycaemic samples and subsequent radiological investigations identify most of the causes of hypoglycemia.

Introduction

Hypoglycemia is defined as a blood glucose concentration of 2.5 mmol L^{-1} (plasma glucose concentration of 3.0 mmol L^{-1}) or less. Its definition is necessarily arbitrary and owes its importance to the fact that hypoglycemia (low blood glucose) of this severity produces brain dysfunction by depriving its neurons of glucose.

Hypoglycemia is not a disease but a manifestation of it. It has, however, come to have a totally different meaning, among certain sections of the population, that has nothing to do with blood glucose concentration but much to do with feelings of well-being, discomfort, attitudes to life and, above all, with the role of diet in the achievement and maintenance of good health. No discussion of the dietary treatment of hypoglycemia can be meaningful without reference to this concept—referred to, for want of a better term, as non-hypoglycemia. Hypoglycemia will, throughout this article, be used only to describe a condition associated with a measured low blood glucose concentration.

Brain function and hypoglycemia

The brain malfunction and symptoms to which hypoglycemia gives rise will be referred to as neuroglycopenia to distinguish them from a low measured blood glucose concentration.

Although glucose is the major source of energy for brain, the brain is able, under certain circumstances including prolonged fasting, to utilize the "ketone bodies" (β-hydroxybutyrate and acetoacetate) produced exclusively in the liver from partial oxidation of fatty acids. Under these circumstances the need for glucose and its supply through gluconeogenesis is drastically reduced. The survival value of this adaptation is immense as it permits fat stores rather than structural muscle and other tissue proteins to be utilized for maintenance of vital processes under stressful conditions. Only when fat stores have become completely exhausted and plasma ketone levels fail to rise does the brain's demand for glucose exceed the ability of gluconeogenesis to provide it. At this point hypoglycemia intervenes and portends death from starvation or inanition (see **Starvation**).

The blood glucose concentration

Failure to appreciate the differences between arterial and venous blood glucose is a major cause of the confusion that has surrounded the recognition and diagnosis of hypoglycemia and been responsible for non-hypoglycemia becoming a common diagnosis among those whom Singer and coworkers refer to as the folk sector, which includes many health writers.

In the fasting subject the concentration of glucose in arterial and venous blood is virtually identical but following ingestion of a carbohydrate rich meal it may differ by as much as 2.5 mmol L^{-1}. It is arterial blood glucose that determines glucose supply to the brain, regulates the secretion of insulin and other hormones, and is itself homeostatically controlled and therefore, it is necessary to define hypoglycemia in terms of glucose in arterial (or more practically free flowing capillary) than in venous blood.

Mechanism of hypoglycemia

The glucose pool in fasting subjects

Glucose is confined within the body to the extracellular or interstitial fluid where it is referred to as the glucose pool; detailed discussion of its regulation is outside the scope of this article except to stress that its size is reflected by the concentration of glucose in the blood. This remains remarkably constant despite huge changes in the rates of delivery and utilization of glucose, by meals and exercise (and fasting), respectively, and is described as glucose homeostasis (**Fig. 1**). The main but far from sole regulator is insulin.

Insulin release in response to eating and fasting

After a carbohydrate-containing meal, glucose derived from food enters the portal vein and transported to the liver where about one-third of it is extracted and converted to glycogen. The un-extracted glucose passes into the systemic circulation, producing small and variable rises in arterial, capillary, and, initially, venous blood glucose concentrations. The modest rise in arterial blood glucose concentration perfusing the pancreas, augmented by nervous stimuli and insulinotrophic hormones stimulate insulin secretion. The most notable insulinotrophic hormones are glucagon-like polypeptide 1 (GLP-1) and gastric inhibitory hormone, sometimes called glucose dependent insulinotrophic peptide but best referred to simply as GIP (Marks, 2020) and collectively called incretins (Marks, 2018). These are released from the gut in response to meals containing carbohydrate and fat and lead to the secretion of insulin in greater amounts than simply due to the rise in blood glucose concentration, a phenomenon known as the "incretin effect." The

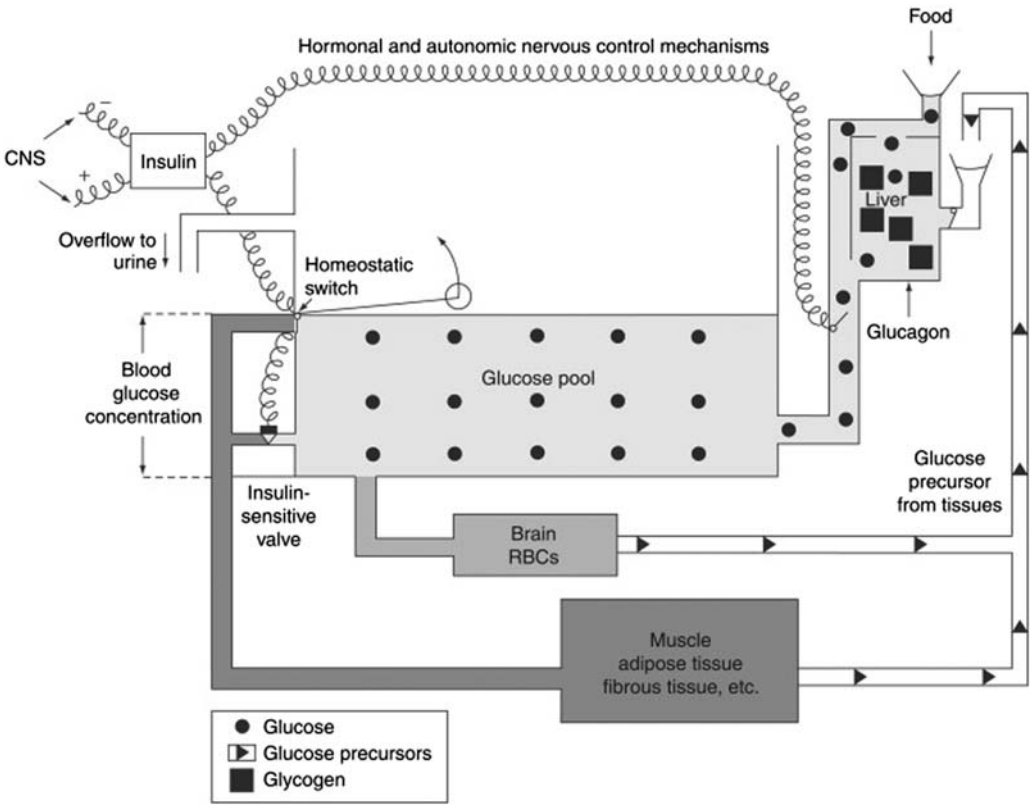


Fig. 1 Schematic representation of homeostatic control of blood glucose level and mechanism of hypoglycemia. Hypoglycemia results whenever inflow of glucose from the gut or liver fails to meet the outflow of glucose from the glucose pool, which consists of glucose dissolved in the extracellular water only. Imbalance arises from: (1) excessive outflow into the tissues owing to insulin (or very rarely IGF-II) overproduction or activity; or (2) in the fasting state, an inability of the liver to liberate or produce glucose at a rate sufficient to meet the noninsulin-dependent, and obligatory, requirements of the brain and red blood cells for glucose.

incretin effect minimizes blood glucose excursions very efficiently following a meal and plays an important role in maintaining postprandial glucose homeostasis.

Evidence for a "cephalic phase" of insulin secretion in humans is scanty and conflicting. Most observers have found a minimal, if any, response to the prospect of eating, or the reality of drinking, a non-calorigenic sweet drink. Others have reported insulin secretion especially in the 10 min or so after eating an appetizing meal and before the blood glucose concentration has risen. They attribute it to vagal activation.

In the postprandial period, as the blood glucose concentration falls toward its homeostatically controlled level, insulin secretion declines to a level that is just sufficient to suppress unbridled lipolysis. Absence of this constitutive insulin secretion in patients with type 1 diabetes is the cause of diabetic ketoacidosis.

The role of the liver in glucose homeostasis

The liver, under the influence of insulin reaching it in high concentration in the portal vein after ingestion of a meal, switches from being a net exporter to net importer of glucose from the glucose pool. Any insulin not extracted and degraded by the liver passes through the heart and lungs to reach peripheral tissues, notably muscle, adipose tissue, and skin, where, providing the concentration of insulin in blood is sufficiently high, it promotes glucose uptake.

Except in disease, the glucose pool, amounting to just 5–15 g, rarely expands by more than 100% even after ingestion of a meal providing up to 300 g of carbohydrate as starch or glucose; nor does it shrink to less than 4 g, corresponding to a blood glucose concentration of about 3.5 mmol L^{-1}, even after many days of fasting.

Entry of glucose into the glucose pool is limited by the rate at which it can be absorbed from the intestine. This is normally in the region of 25–50 g h^{-1}. In people with normal glucose tolerance, venous blood glucose levels generally return to overnight fasting values within 2–3 h of eating a meal regardless of how much carbohydrate it contains. Arterial blood glucose levels take longer to return to pre-ingestion levels but they too are always within the normal fasting range by 3–4 h, even though the evidence provided by measurement of the gut hormone GIP, indicates that absorption of large meals continues for much longer. Absorption of a 200 g liquid glucose meal by normal healthy subjects, for example, is still incomplete 5 h later even though both their venous and arterial blood glucose levels have long since returned to normal.

The outflow of glucose into the tissues depends upon many factors; the two most important are the plasma insulin and blood glucose concentrations. Under maximum insulin stimulation—and at "normal" blood glucose levels—glucose disappears from the glucose pool at a rate of up to 40–50 g h^{-1} but these conditions are rarely encountered except experimentally or in cases of gross insulin overdose.

Onset of insulin action is almost instantaneous and persists for as long as insulin remains bound to insulin receptors. Glucose continues to enter insulin dependent cells for up to 30 min after plasma insulin levels have returned to "fasting" levels. During this time the glucose pool may shrink sufficiently to produce hypoglycemia unless replenished by glucose continuing to enter from the intestine (or experimentally/therapeutically by intravenous infusion) or from the liver, once it has switched from the glycogenic to glycogenolytic mode.

Small, and always temporary, imbalances between the rate at which insulin action declines and glucose enters the glucose pool can occur in healthy subjects after ingestion of a large dose of glucose in solution on an empty stomach, but is rare following the ingestion of an ordinary mixed meal.

A slight delay in stimulating insulin release in response to a meal is the earliest and most characteristic abnormality observed in patients with typical type 2 diabetes mellitus who may secrete more insulin in total than people of comparable age and body mass index. They are, however, generally insulin resistant, which explains why, despite the larger amounts of insulin secreted in response to meals in the early stages of glucose intolerance, they do not suffer from meal-induced hypoglycemia.

Hypoglycemic syndromes

Brain malfunction from hypoglycemia

The brain ordinarily receives a regular and plentiful supply of glucose from the blood by active transport across the blood brain barrier utilizing the glucose transporter protein GLUT1. Reduction of glucose supply to below critical limits causes the brain to malfunction and this manifests itself clinically as activation of the autonomic nervous system and cognitive neural deficit. The blood glucose level at which autonomic activation and cognitive impairment occurs varies. Symptoms are unusual at blood glucose levels above 3.0 mmol L^{-1} except in diabetic and elderly subjects in whom they may occur at higher levels. Objective evidence of cerebral impairment can however often be discerned by an investigator at blood glucose levels around 3.5–4 mmol L^{-1} especially in people with diabetes.

Causes of neuroglycopenia other than hypoglycemia, i.e., normoglycemic neuroglycopenia, are rare but include reduced GLUT1 activity and poorly controlled diabetes. GLUT1 deficiency, due to a *SLC2A1* gene mutation, results in reduced blood glucose delivery to the brain in the presence of normal blood glucose levels (Klepper and Voit, 2002). Patients present with epilepsy, movement disorders and impaired neurological growth. The possibility that less severe forms are more common than currently supposed and may be responsible for some cases of "non-hypoglycemia" cannot be completely dismissed. It also explains why, under research conditions, some people diagnosed with this condition develop symptoms at higher blood glucose levels than control subjects.

Neuroglycopenic syndromes

Acute neuroglycopenia (adrenergic symptoms)

This syndrome comprises a collection of vague symptoms such as feelings of alternating hot and cold, feeling unwell, anxiety, panic, inner trembling, unnatural feelings, blurring of vision, and palpitations, any or all of which may be accompanied by objective signs of facial flushing, sweating, tachycardia, and unsteadiness of gait. There is no particular order in which these features occur, nor are they constant. Nevertheless, patients on insulin or sulfonylurea therapy for diabetes, in whom they are common, rely upon them to warn of more severe neuroglycopenic impairment culminating in loss of consciousness. These patients can be taught to abort progression of symptoms by eating carbohydrate hence they are often referred to as warning symptoms.

Many of the features of acute neuroglycopenia resemble those produced by adrenaline which is released in large amounts and consequently are often referred to as adrenergic.

Sub-acute neuroglycopenia or hypoglycemic unawareness

This syndrome is more insidious and may go completely unrecognized unless or until the patient loses consciousness. Often, however, there is loss of spontaneous activity, impairment of cognitive function, and the onset of somnolence that is more discernible to the bystander than to the patient and which, when it occurs *de novo* in an insulin-treated diabetic, is often referred to as hypoglycemia unawareness. Although suffering from sub-acute neuroglycopenia patients may perform quite complicated actions, such as driving a motorcar, of which they are total unaware when restored to consciousness. In other words they may behave as an automaton.

Acute can proceed to sub-acute neuroglycopenia and both can progress to stupor or coma unless relieved by food or injection of glucagon. Even if this is not done, full recovery under the influence of endogenous counter-regulatory mechanisms, is almost invariable and is the reason why treatment with insulin is relatively safe despite the dangers of hypoglycemia.

Chronic neuroglycopenia

This neuroglycopenic syndrome is exceedingly rare. It occurs only when the blood glucose concentration remains low, either owing to the presence of an insulin-secreting tumor of the pancreas or overzealous treatment of diabetes with insulin for weeks or months. It is characterized by mental dysfunction resembling clinical depression, schizophrenia, or dementia, the symptoms of which are not relieved by restoring the blood glucose level to normal. Partial recovery may, however, take place over the ensuing months or years if the cause of the hypoglycemia is remedied.

This condition might be confused with "non-hypoglycemia" were it not for the fact that the blood glucose concentration is invariably low (<3.0 mmol L^{-1}) while the patient is fasting, does not rise normally in response to food, and evidence of underlying disease can always be found.

Diagnosis

Whipple's triad

Whipple's triad of occurrence of symptoms, a simultaneous measured low blood glucose concentration, and rapid symptomatic relief with correction of hypoglycemia following glucose administration is a *sine qua non* for ascertaining neuroglycopenia due to hypoglycemia (Whipple, 1938).

Causes of hypoglycemia

There are in the region of 100 causes of hypoglycemia but all, apart from exogenous (or iatrogenic) insulin overdose, are uncommon. The most important causes of recurrent hypoglycemia are listed and briefly described in **Table 1**. Differentiation is seldom simple and always rests heavily upon the results of laboratory data of which measurements of plasma insulin, proinsulin, C-peptide and beta-hydroxybutyrate (BHB) are the most important.

Endocrinological, anatomico-pathological, iatrogenic and toxic causes of hypoglycemia will not be considered further. Instead, attention will be given to those conditions (including "nonhypoglycemia") that have a mainly or exclusively dietary etiology and which respond partially or completely to dietary measures.

Postprandial (reactive) hypoglycemia

Within a year of the discovery of insulin, and the symptoms to which hypoglycemia can give rise, Seale Harris, an American physician, had proposed that spontaneous overproduction of endogenous insulin might produce a similar condition. Confirmation of this hypothesis soon followed. The seminal work of Whipple on the diagnosis of insulinoma and of Conn on diet-induced postprandial reactive hypoglycemia, both in 1936, distinguished between fast-induced (fasting) hypoglycemia and that which occurred only in response to feeding. The latter, reactive or postprandial hypoglycemia, could be reproduced by oral administration of large doses of glucose in solution and this became the standard criterion for its diagnosis—the 5 h glucose tolerance or load test.

Prolonged glucose load test

The observation that in a substantial percentage of normal healthy people glucose taken in solution on an empty stomach produces a rebound fall in venous blood glucose levels to below fasting levels was made very soon after blood glucose measurements became possible and before the discovery of insulin. It attracted little attention at the time being considered to have only curiosity value and little pathological significance.

The situation changed dramatically during the early 1950s and, subsequently, particularly in the US, with the appearance of books written for lay consumption attributing a vast array of common symptoms to hypoglycemia, whether the blood glucose concentration was low at the time or not. Belief in the importance and prevalence of hypoglycemia grew among fashionable medical practitioners and the general public alike to such an extent that, by the early 1970s, alarm bells began to ring among consumer action groups and the scientific medical community.

With the passage of time the original, well-defined syndrome of postprandial reactive hypoglycemia had become so distorted and the criteria for its diagnosis so blurred, that anyone with vague symptoms could be, and often was, described as suffering from hypoglycemia, without bothering to measure their blood glucose concentration.

Not until a consensus Statement on Postprandial or Reactive Hypoglycemia was issued by the Third International Symposium on Hypoglycemia and generally recognized by medical practitioners throughout the world did scientific criteria for the diagnosis of reactive hypoglycemia gain universal acceptance and its purported incidence declined dramatically. This has been emphasized by endocrine society clinical practice guidelines which define hypoglycemic symptoms without Whipple's triad as a functional disorder.

Table 1 Diagnostic criteria and treatment of the main types of hypoglycemia.

Description	*Mechanism*	*Diagnostic criteria*	*Management*
Insulin-secreting tumor: insulinoma	Abnormal β-cells with failure to suppress insulin secretion in response to hypoglycemia	Fasting hypoglycaemia. Inappropriate high plasma insulin ($>$18 pmol L^{-1}) and C-peptide ($>$200 pmol L^{-1}) concentrations in presence of hypoglycemia (BG $<$ 2.5 mmol L^{-1}). Suppressed β-hydroxybutyrate levels ($<$0.6 mmol L^{-1})	High carbohydrate intake until curative surgical ablation, or effective hyperglycemic therapy with diazoxide plus chlorothiazide or with octreotide, can be instituted.
Non insulinoma pancreatogenous hypoglycemia (NIPHS)	Exact cause unknown; no h/o surgery; excessive insulin secretion after meals	Postprandial hypoglycemia 3–5 h after eating. High insulin, C-peptide. Negative imaging, diffusely positive intra-arterial calcium stimulation test	Frequent small mixed meals low in absorbed carbohydrates, rich in soluble dietary fiber. May benefit from treatment with acarbose or miglitol (α-glucosidase inhibitors)
Post bariatric surgery and idiopathic rapid gastric emptying	Accelerated deposition of nutrients in duodenum and increased release of insulinotrophic hormones, for example, GIP, GLP-1.	Postprandial hypoglycemia only follows 1–3 h after eating. History of gastrectomy or objective evidence of rapid gastric emptying. Exaggerated insulinaemic response to food.	Frequent small mixed meals, low in rapidly absorbed carbohydrate; rich in dietary fiber. May benefit from treatment with acarbose (α-glucosidase-inhibitors), diazoxide or octreotide
Autoimmune insulin syndrome	Delayed release of insulin from antibody binding after all of meal has been absorbed	Postprandial hypoglycemia 3–12 h after last eating: total plasma insulin high; C-peptide high, normal or low; proinsulin normal or high. Antibodies to insulin present. Common in Japan, infrequent elsewhere	Frequent small mixed meals, low in rapidly absorbed carbohydrate; rich in dietary fiber
Non-islet cell tumor hypoglycemia (NICTH) Mesenchymal, epithelial and hematopoietic tumors	Tumor cells secrete big IGF-II which binds to insulin receptors and IGF receptors, reducing hepatic glucose production, increasing glucose uptake by peripheral tissues	Fasting hypoglycemia. Low plasma insulin and C-peptide levels: low plasma IGF-I, normal or raised IGF-II levels; high IGF-II: IGF1 ratio. Suppressed β-hydroxybutyrate levels ($<$0.6 mmol L^{-1})	High carbohydrate intake orally or intravenously until curative treatment of tumor or debulking is instituted. Glucagon and somatostatin analogs may be helpful
Fasting alcohol-induced hypoglycemia	Alcohol impaired hepatic gluconeogenesis	Low blood glucose, raised blood alcohol, lactate and usually β-hydroxybutyrate: low plasma insulin and C-peptide	Avoid drinking alcohol while fasting or on a low energy diet.
Endocrine disease for example, hypopituitarism, addison's disease	Reduced availability of diabetogenic or hypoglycemia counter-regulatory hormones	Clinical features of primary disease with subnormal levels of appropriate counter-regulatory hormones, for example, cortisol, growth hormone. Appropriately raised β-hydroxybutyrate levels ($>$500 μmol L^{-1}) during hypoglycemia. low plasma insulin and C-peptide	High carbohydrate intake orally or intravenously until effective hormone replacement therapy has been established
Drugs Insulin, sulfonylurea Others-quinine, beta blocker, etc.	Drugs other than insulin and SU may cause hypoglycemia mostly in patients with restricted food intake, age, liver or kidney disease	High insulin, low C-peptide in insulin overdose High insulin, high C-peptide with SU and quinines Low insulin, low C-peptide with others	Treat hypoglycemia, adjust insulin or SU dose. Remove other potential offenders
Critical illness Sepsis, renal failure hepatic failure	Excess glucose utilization and impaired glucose formation	Clinical features of primary disease Low insulin, low C-peptide	Treat hypoglycemia and primary disease
Inherited metabolic disorders			
a. Glycogen storage diseases (GSDs)	Inability to release glucose from liver during fasting	Usually present in childhood: low blood glucose, high β-hydroxybutyrate levels, low insulin and C-peptide: high lactate: impaired or absent glucose response to glucagon	Avoid fasting: a constant intake of slowly absorbed carbohydrate (uncooked corn starch) may be required day and night in infants

Table 1 Diagnostic criteria and treatment of the main types of hypoglycemia.—cont'd

Description	*Mechanism*	*Diagnostic criteria*	*Management*
b. Disorders of mitochondrial β-oxidation of fatty acid	Defective utilization of fat as fuel in tissues: compensatory increase in glucose utilization	Presents mostly in infancy: low glucose, low insulin and C-peptide, high FFA, low β-hydroxybutyrate, increased urinary organic acids. Low carnitine in some cases	Avoid fasting: frequent high carbohydrate low fat feeding.
c. Gluconeogenesis disorders—fructose 1,6 biphosphatase deficiency	Fructose biphosphatase enzyme deficiency	Hypoglycemia, lactic acidosis following fasting and ingestion of fructose	Avoid prolonged fasting and fructose containing food
d. Hereditary fructose intolerance	Aldolase B enzyme deficiency causing impaired release of glucose from liver in response to ingestion of fructose or sucrose.	Hypoglycemia evoked by food containing fructose or sucrose	Avoid fruit and fructose/sucrose containing food.
e. Congenital endogenous hyperinsulinemia due to genetic mutations	Dysregulated insulin secretion	Mostly seen in infants and children, low glucose, high insulin and C-peptide, low free fatty acids, low β-hydroxybutyrate levels	Frequent feeding of low carbohydrate (avoiding short-acting carbohydrates), high-protein diet, and addition of fiber and fat.

Definition

It is now accepted that a few people exhibit, in the course of their everyday life, symptoms similar to those caused by acute neuroglycopenia and may, if accompanied by a capillary or arterialized venous blood glucose concentration of 2.5–3.0 mmol L^{-1} or less, justify description as being of postprandial reactive hypoglycemic origin. Reactive hypoglycemia may itself be a consequence of anyone of a large number of well-recognized but generally uncommon organic conditions, such as an islet cell tumor, that also produce fast-induced hypoglycemia. It is, therefore, only after all of these have been excluded by appropriate laboratory investigations that a diagnosis of functional or dietary reactive hypoglycemia is justified.

Specifically, the prolonged oral glucose load (tolerance) test is now deemed inappropriate for the diagnosis of postprandial or reactive hypoglycemia because the high incidence of false-positive results with this test makes it meaningless, especially if, as is so often the case, venous rather than arterial blood is sampled. A mixed meal test *containing protein, carbohydrates, and fat, however, may be used to identify postprandial hypoglycemia.*

The postprandial syndrome

Typically, the patient is a woman of 20–50 years whose main complaint is of vague feelings of distress occurring predominantly mid-morning, but occasionally mid-afternoon or evening and never before breakfast. In between attacks, characterized by feeling of faintness, anxiety, nervousness, irritability, inner trembling, rapid heartbeat, headache, and sweatiness, either alone or in combination, they may be completely well. Often they describe themselves as suffering from increased tiredness, lacking in zest for life, and apathetic: symptoms often associated with depression or chronic alcohol abuse.

Patients often have diagnosed themselves, on the basis of articles they may have read, as suffering from hypoglycemia. Almost without exception they reject the possibility that their symptoms might have a contributory, or even large, psychogenic element.

Symptoms wax and wane during middle life but often remit completely for years or may never recur. They are not progressive and never cause severe neurological dysfunction such as coma, psychosis, or dementia. Hypoglycemia cannot be demonstrated during spontaneous symptomatic episodes in people with the postprandial syndrome and some other explanation should be sought for them. Psychotherapy is often helpful.

Differential diagnosis

Studies using continuous interstitial glucose monitoring confirm that only a very small proportion of those with postprandial syndrome have symptomatic hypoglycemia. Of those who do, a substantial proportion have an identifiable cause for it. The commonest, in the West, is partial gastrectomy, pyloroplasty and bariatric surgery (gastric sleeve or gastric bypass) for the treatment of obesity (Lee et al., 2015): in the Far East, for example, Japan, it is the autoimmune insulin syndrome (AIS) which also occurs in Europe and the US but is rare (Lupsa et al., 2009). Other rare causes include insulinoma and noninsulinoma pancreatogenic hypoglycemia.

In some people postprandial hypoglycemia occurs only in response to a specific dietary indiscretion, for example, ingestion of moderate amounts of alcohol (eg gin) and glucose (eg tonic) on an empty stomach or fructose in people with a rare inborn errors of

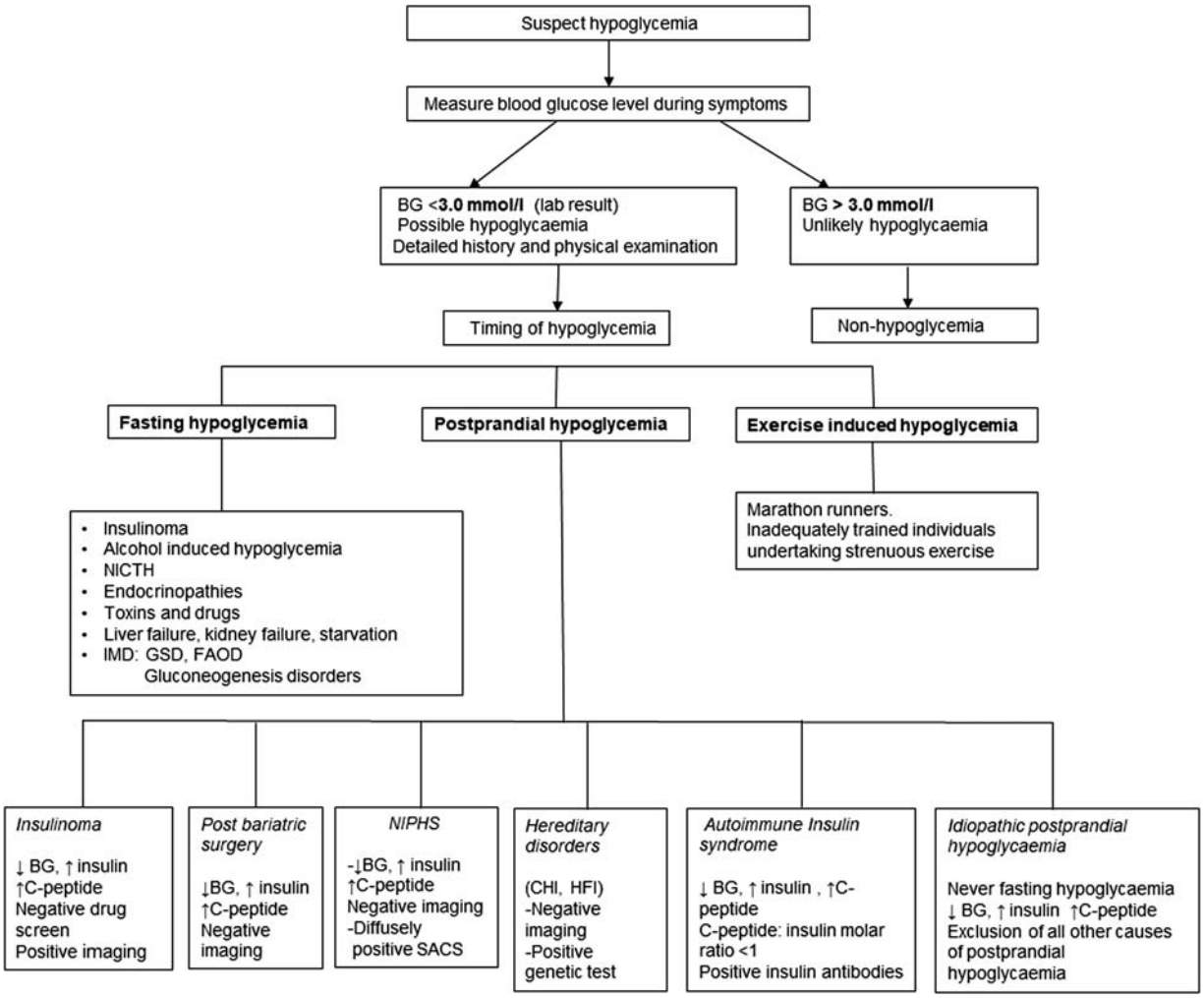


Fig. 2 Investigation of hypoglycemia. NICTH: Non islet cell tumor hypoglycemia, IMD: inherited metabolic disorders; GSD: glycogen storage disease; FAOD: fatty acid oxidation defect; NIPHS: non insulinoma pancreatogenous hypoglycemia; CHI: congenital hyperinsulinism; HFI: hereditary fructose intolerance; SACS: selective intra arterial calcium stimulation test.

metabolism-hereditary fructose intolerance (Douillard et al., 2012). In a small number of subjects for whom no satisfactory pathogenic mechanism can be identified (Fig. 2), it is justified to describe them as suffering from idiopathic reactive (postprandial) hypoglycemia.

Management of reactive (postprandial) hypoglycemia

Treatment of attacks

Because of their short duration and modest severity, acute spontaneous neuroglycopenic episodes due to reactive hypoglycemia require no specific treatment beyond ingestion of a rapidly assimilable form of carbohydrate (e.g., a lump of sugar). There is no evidence that this ever produces rebound hypoglycemia and should it do so the grounds for making the diagnosis should be reviewed.

Postprandial hypoglycemia secondary to disease is similarly treated with oral carbohydrate but if severe may require parenteral glucagon followed by carbohydrate ingestion.

Prevention

Dietary prevention of postprandial hypoglycemia is based on the premise that it is caused by imbalance between the timing and amount of insulin secreted in response to the ingestion of a meal and disposal of the glucose derived from it. Frequent small meals containing only modest amounts of sugars (glucose and sucrose) and refined starches but rich in poorly absorbed complex carbohydrates and containing dietary fiber have replaced the diets rich in proteins (and fats) previously advocated, but evidence of their unique efficacy is lacking. Avoidance of drinks rich in sucrose or glucose, especially with alcohol, may be helpful in subjects who are

highly susceptible to this combination. There is no evidence that confectionery eaten in moderation e.g., 25–50 g, is uniquely detrimental, though excessive use should be discouraged, if only, on general health grounds.

Dietary modification alone may be insufficient and supplementary pharmacological intervention may be required to achieve a satisfactory therapeutic outcome. Guar, acarbose and miglitol, which slow glucose absorption and decrease the insulinemic response to food, may be effective. When these medications fail to improve the symptoms, somatostatin analogs which reduce synthesis and secretion of insulin may also be effective. Interestingly, diazoxide which inhibits insulin secretion has not been found effective except in patients with proven endogenous hyperinsulinism due to insulinoma or pancreatic islet hyperplasia.

Non-hypoglycemia

No account of dietetic treatment of hypoglycemia would be complete without a brief description of non-hypoglycemia, which has been described as a controversial illness and epidemic in the US. Clinically, the illness is indistinguishable from (idiopathic) postprandial hypoglycemia, except that the blood glucose level is never pathologically low during symptomatic episodes. Moreover, although transient turns are often a major feature of the illness, only rarely if ever, does the patient consider their health, between turns, as normal.

The attribution of these patients' illness to hypoglycemia had its origins in the early 1950s with the appearance, in the US, of a book by Drs. Abrahams and Pezet entitled "Body, Mind and Sugar." Other American practitioners, notably John Tintera, founder of the Hypoglycemia Foundation Inc., Stephen Gyland, Harry Saltzer and, others, including the medical writer Carlton Fredericks, publicized the concept. This led to hypoglycemia being held, by a large section of the public, responsible for such diverse diseases as coronary artery disease, allergy, asthma, rheumatic fever, susceptibility to viral infections, epilepsy, gastric ulcer, alcoholism, suicide, and even homicide, as well as for a whole galaxy of symptoms in their own right. Hypoglycemia was treated as though it was a disease entity and asserted by its advocates to be "one of the most common illnesses in the United States" and that because of it "thousands of Americans have forgotten, or perhaps never known, what it is like to feel completely healthy." Diagnosis of non-hypoglycemia generally depended upon the results of the now discredited 6 h oral glucose tolerance test, using venous blood, although some have dispensed even with this discredited formality in favor of just purely clinical criteria.

The appearance in the New England Journal of Medicine of an article in 1974 entitled "Non-hypoglycemia is an epidemic condition" first drew international attention to its existence (Yager and Young, 1974). Many patients with non-hypoglycemia undoubtedly derive some benefit, probably through a powerful placebo effect, from severely restrictive dietary regimes. Although differing in details most of the diets emphasize the purported specifically detrimental effects of sugar (sucrose), salt, alcohol, and caffeine.

Although the cause of illness in people with non-hypoglycemia remains unknown, it is unlikely to be the same in all cases. In some it is chronic alcoholism and in a tiny number of others it is owing to caffeine intoxication, which can be confirmed by a dietary history and measurement of plasma caffeine levels. Such patients benefit specifically from reducing their intake of caffeinated beverages, though not necessarily from avoiding them completely. Ironically, and probably significantly, caffeine restores hypoglycemia awareness to diabetic patients on insulin who have become insensitive to it. The possibility exists, therefore, that a combination of reasonable or normal caffeine intake occurring in combination with the normal rebound in arterial blood glucose to just below fasting levels that sometimes occurs 3–5 h after a meal in someone with an unusually low threshold for neuroglycopenia, might precipitate symptoms. This explanation must, however, be considered no more than speculative (Kerr et al., 1993).

On the other hand such diverse illnesses as hyperventilation, panic attacks, drug abuse, and genuine food intolerances are all established as capable of producing the non-hypoglycemia syndrome and should always be considered in the differential diagnosis.

Exercise-induced hypoglycemia

Previously only associated with marathon running, hypoglycemia is now recognized to be comparatively common in inadequately trained individuals undertaking strenuous exercise (Brun et al., 2001). Consumption of rapidly absorbed carbohydrate before taking exercise may encourage its appearance although consumption of slowly absorbed, low glycemic index foods may prevent it as does appropriate training.

Exercise, however, may provoke hypoglycemia in those who are fasting with an underlying organic cause for hypoglycaemia, such as insulinoma.

Drug induced hypoglycemia

In addition to insulin or insulin secretagogues, a number of medications which are not used to treat hyperglycemia, have been implicated in causing hypoglycemia such as quinine, disopyramide, beta blockers, salicylates and pentamidine, quinolones, angiotensin-converting enzyme agents and IGF. Liver or kidney disease, old age and polypharmacy are the predisposing factors for drug induced hypoglycemia.

Hepatic and renal failure

Considering the importance of the liver and kidneys in the maintenance of blood glucose levels, hypoglycemia is rare in both hepatic and renal disease. In liver disease hypoglycemia is virtually confined to patients with acute toxic hepatic necrosis, whether owing to overwhelming viral infection or specific hepatotoxins such as poisonous mushrooms, unripe akee fruit, or paracetamol in excess. Its appearance always portends an extremely poor prognosis as it does in all critically ill patients regardless of their pathology.

The association of hypoglycemia with primary cancer of the liver is comparatively common and owing to over-expression and secretion of big IGF-II, and is not due to nonspecific destruction of hepatic tissue as previously supposed.

Renal failure is one of the most common causes of hypoglycemia in nondiabetic hospital inpatients but does not carry quite as grave a prognostic significance as in patients with liver disease. It generally responds to appropriate dietary and other supportive treatments for end-stage kidney disease (Arem, 1989).

Sepsis

Infections, usually bacterial, not necessarily severe enough to produce a full blown Systemic Inflammatory Response Syndrome (SIRS) may cause hypoglycemia. The hypoglycemia is usually uncovered serendipitously by a blood glucose measurement, and generally portends a fatal outcome (wang et al., 2021). It results from increased glucose utilization by the peripheral tissues, reduced glycogen reserves and impaired gluconeogenesis in liver. Treatment is that of the primary condition as well as of the hypoglycemia.

Inborn errors of metabolism

Hypoglycemia is a manifestation and often a presenting feature of many inborn errors of metabolism (see **Table 1**) usually in children but occasionally in adults (Douillard et al., 2012).

In glycogen storage diseases (GSD), various enzymes deficiencies prevent breakdown of glycogen into glucose, resulting in severe hypoglycemia. GSDs usually present in childhood with fasting hypoglycemia, hyperketonemia and lactic acidosis which may persist in adulthood particularly in GSD type I and type III. These are managed with slowly absorbed uncooked cornstarch, which is broken down slowly to provide a continuous supply of glucose, and avoidance of prolonged periods of fasting. In infants and children continuous feeding through a nasal or gastrostomy tube may be necessary to prevent hypoglycemia, especially at night when glycogenolysis normally maintains normoglycemia.

Abnormalities of fatty acid metabolism, on the other hand, are characterized by hypoglycemia, increased non-esterified fatty acids and hypoketonemia. As with children with liver GSD, treatment is to ensure that they are constantly supplied with carbohydrates and are never fasting for more than a very short period. The fasting tolerance, however, increases with age. Milder forms of fatty acid oxidation defects may remain undiagnosed till adulthood.

Hereditary fructose intolerance (HFI) results from a deficiency of the enzyme aldolase-B which breaks down fructose. Ingestion of fructose (fruit sugar) or sucrose (cane or beet sugar) leads to the accumulation of fructose-1-phosphate in the liver, inhibiting glycogenolysis and gluconeogenesis. This results in hypoglycemia and lactic acidosis. The diagnosis may be delayed till adult life. Similarly, deficiency of the enzyme fructose 1,6 biphosphatase prevents gluconeogenesis resulting in hypoglycemia, lactic acidosis and ketosis triggered during prolonged fasting, inter-current illness or ingestion of large amount of fructose containing food. The mainstay of treatment in these conditions is a fructose-free diet.

Endogenous hyperinsulinemic hypoglycemia, a genetically heterogeneous condition, is emerging as a common cause of severe persistent hypoglycemia in neonates and children. Hypoglycemia may occur fasting, postprandial or related to exercise. Normoglycemia is maintained with oral and intravenous glucose. Some resolve spontaneously or require treatment with diazoxide or if severe partial pancreatectomy.

Starvation

Although average fasting blood glucose levels are lower in victims of famine than in well-fed populations, hypoglycemia is rare. Even in patients suffering from Kwashiorkor, hypoglycemia is uncommon and is usually associated with infection, hypothermia, and coma. Patients with anorexia nervosa develop hypoglycemia only as an agonal phenomenon and its appearance generally portends imminent death. The characteristic clinical biochemistry findings are of low plasma insulin, proinsulin, C-peptide, IGF-1, nonesterified fatty acids (NEFA) and β-hydroxybutyrate levels and elevated growth hormone and cortisol levels. Relief of hypoglycemia by re-feeding is the only measure carrying any chance of preventing death, but it is rarely successful.

Hypoglycemia in the elderly sick

The high incidence of hypoglycemia in sick elderly patients—especially those with infections—has become apparent from the use of routine blood glucose measurements. The cause is seldom attributable to any of the well-recognized causes of hypoglycemia found in younger fitter people. It is probably owing to chronic malnutrition that is so common in the elderly sick, compounded by coincident disease but which is not of itself sufficiently severe to produce hypoglycemia.

Conclusion

Spontaneous (non-diabetic) hypoglycemia needs laboratory confirmation followed by full evaluation to establish and treat the underlying cause. Prompt treatment of acute hypoglycemia is important to prevent irreversible neurological sequelae.

References

Arem, R., 1989. Hypoglycemia associated with renal failure. Endocrinol. Metab. Clin. North Am. 103–121.
Brun, J.F., Dumortier, M., Fedou, C., Mercier, J., 2001. Exercise hypoglycemia in nondiabetic subjects. Diabetes Metab. 27, 92–106.
Douillard, C., Mention, K., Dobbelaere, D., et al., 2012. Hypoglycaemia related to inherited metabolic diseases in adults. Orphanet J. Rare Dis.
Kerr, D., Sherwin, R.S., Pavalkis, F., et al., 1993. Effect of caffeine on the recognition of and the responses to hypoglycemia in humans. Ann. Intern. Med. 119, 799–804.
Klepper, J., Voit, T., 2002. Facilitated glucose transporter protein type 1 (GLUT1) deficiency syndrome: impaired glucose transport into brain—a review. Eur. J. Pediatr. 161, 295–304.
Lee, C.J., Clark, J.M., Schweitzer, M., et al., 2015. Prevalence of and risk factors for hypoglycemic symptoms after gastric bypass and sleeve gastrectomy. Obesity 23, 1079–1084.
Lupsa, B., Chong, A.Y., Cochran, E.K., et al., 2009. Autoimmune forms of hypoglycemia. Medicine (Baltimore) 88, 141–153.
Marks, V., 2018. Rebirth of the incretin concept: its conception and early development. Peptides 100.
Marks, V., 2020. The early history of GIP 1969-2000: from enterogastrone to major metabolic hormone. Peptides 125. https://doi.org/10.1016/j.peptides.2020.170276.
Wang, J., Zhu, C.K., Yu, J.Q., et al., 2021. Hypoglycemia and mortality in sepsis patients: a systematic review and meta-analysis. Heart Lung 50, 933–940.
Whipple, A.O., 1938. The surgical therapy of hyperinsulinism. J Int Chir 3, 237–276.
Yager, J., Young, R.T., 1974. Non-hypoglycemia is an epidemic condition. NEJM.

Further reading

Cryer, P.E., Axelrod, L., Grossman, A.B., et al., 2009. Evaluation and management of adult hypoglycemic disorders: an endocrine society clinical practice guideline. J. Clin. Endocrinol. Metab. 94, 709–728.
Gama, R., Teale, J.D., Marks, V., 2003. Clinical and laboratory investigation of adult spontaneous hypoglycaemia. Best Practice No. 173. J. Clin. Pathol. 56, 641–646.
Honkaa, H., Salehi, M., 2019. Postprandial hypoglycemia after gastric bypass surgery: from pathogenesis to diagnosis and treatment. Curr. Opin. Clin. Nutr. Metab. Care 22 (4), 295–302.
Hughes, L.E., Fenn, J.S., Ford, C., Gama, R., 2021. Autoimmune hypoglycaemia: a narrative review from the laboratory perspective. J. Lab. Precis. Med. 6, 20. https://doi.org/10.21037/jlpm-20-111.
Lefebvre, P.J., Andreani, D., Marks, V., Creutzfeldt, W., 1988. Statement on postprandial hypoglycemia. Diabetes Care 11, 439–440.
Marks, V., 1976. The measurement of blood glucose and the definition of hypoglycemia. Hormone and Metabolic Research, supplement 6 Stuttgart. In: Andreani, D., Lefebvre, P.J., Marks, V. (Eds.), Hypoglycemia. Proceedings of the European Symposium, Rome. George Thieme, pp. 1–6.
Marks, V., 1987. Functional hypoglycaemia: fact or fancy. In: Andreani, D., Marks, V., Lefebvre, P.J. (Eds.), Hypoglycaemia: Serono Symposia Publications, vol. 38. Raven Press, New York, pp. 1–17.
Mon, S., Ito, H., 1988. Hypoglycemia in the elderly. Jpn. J. Med. 27, 160–166.
Palardy, J., Havrankova, J., Lepage, R., et al., 1989. Blood glucose measurements during symptomatic episodes in patients with suspected postprandial hypoglycemia. N. Engl. J. Med. 321, 1421–1425.
Patti, M.E., Goldfine, A.B., 2010. Hypoglycaemia following gastric bypass surgery—diabetes remission in the extreme? Diabetologia 53, 2276–2279.
Service, F.J., Natt, N., Thompson, G.B., et al., 1999. Non-insulinoma pancreatogenous hypoglycemia: a novel syndrome of hyperinsulinemic hypoglycemia in adults independent of mutations in Kir6.2 and SUR1 genes. J. Clin. Endocrinol. Metab. 84, 1582–1589.
Service, F.J., 1989. Hypoglycemia and the postprandial syndrome. N. Engl. J. Med. 321, 1472–1473.
Shah, P., Rahman, S.A., Demirbilek, H., Güemes, M., Hussain, K., 2017. Hyperinsulinaemic hypoglycaemia in children and adults. Lancet Diabetes Endocrinol. 5 (9), 729–742.
Shilo, S., Berezovsky, S., Friedlander, Y., Sonnenblick, M., 1998. Hypoglycemia in hospitalized nondiabetic older patients. J. Am. Geriatr. Soc. 46, 978–982.
Simpson, E.J., Holdsworth, M., Macdonald, I.A., 2008. Interstitial glucose profile associated with symptoms attributed to hypoglycemia by otherwise healthy women. Am. J. Clin. Nutr. 87, 354–361.
Singer, M., Arnold, C., Fitzgerald, M., Madden, L., von Legat, C.V., 1984. Hypoglycemia: a controversial illness in US society. Med. Anthropol. 8, 1–35.
Snorgaard, O., Lassen, L.H., Rosenfalck, A.M., Binder, C., 1991. Glycaemic thresholds for hypoglycaemic symptoms, impairment of cognitive function, and release of counterregulatory hormones in subjects with functional hypoglycaemia. J. Intern. Med. 229 (343), 350.
Sulaiman, R.A., 2021. Postprandial hypoglycaemia in adults: pathogenesis, diagnosis and management. J. Lab. Precis. Med. 6, 13. https://doi.org/10.21037/jlpm-20-102.

Relevant websites

http://www.hypodiab.com/, A site devoted to iatrogenic hypoglycemia.
http://www.insulinoma.net/english%20homepage/Eindex.html, A site devoted to hypoglycemia mainly due to non-iatrogenic causes.

Inborn errors of metabolism: Classification and biochemical aspects

D.L. Marsden, Boston Children's Hospital, Boston, MA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of D.L. Marsden, Inborn Errors of Metabolism: Classification and Biochemical Aspects, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 1–10, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00155-0.

Introduction 413
Newborn screening 413
Disorders of protein metabolism 414
Amino acid disorders 414
Maple syrup urine disease (MSUD) 415
Cofactor deficiencies 417
Disorders of fatty acid oxidation 418
Disorders of carbohydrate metabolism 419
Galactosemia 419
Glycogen storage disorders 420
Disorders of fructose metabolism 420
Disorders of micronutrient metabolism 421
Disorders of copper metabolism 421
Disorders of iron metabolism 422
Hemochromatosis 422
Conclusions 422
References 422
Further reading 423

Key points

- Understand the basis of inborn errors of metabolism
- The key clinical features of disorders due to an enzyme deficiency in the catabolic pathways of:
  - carbohydrates
  - proteins
  - fatty acids
  - essential vitamins and cofactors
- Understand the pathology of disorders
  - accumulation of toxic intermediates
  - depletion of end products/energy

List of abbreviations

ASL argininosuccinate lyase
ASS argininosuccinate synthetase
BCAA branched chain amino acids
BH4 tetrahydrobiopterin
BIA bacterial inhibition assay
Cbl cobalamin
CPS carbamoyl synthetase
CPT I, II carnitine palmityl transferase
GALE UDP galactose 4-epimerase
GALK galactokinase
GALT galactose-1-phosphate uridyl transferase
HCS holocarboxylase synthetase
LCHAD long chain 3-hydroxy acyl CoA dehydrogenase

MCAD medium chain acyl CoA dehydrogenase
MS/MS tandem mass spectrometry
MSUD maple syrup urine disease
NAG N-acetylglutamine
NH3 ammonia
OTC ornithine transcarbamylase
PAH phenylalanine hydroxylase
PKU phenylketonuria
SCAD short chain acyl CoA dehydrogenase
TCA tricarboxylic acid cycle
VLCAD very long chain acyl CoA dehydrogenase

Introduction

Garrod described the first inborn error of metabolism in his Croonian Lectures (Garrod, 1908) when he associated the symptoms that had been observed in patients with Alkaptonuria as being due to an inherited enzyme deficiency. Since that time more than 400 disorders have been described that are due to an enzyme deficiency in the catabolic pathways of protein, fatty acids, and carbohydrates. The resulting accumulation of the toxic intermediates, and in some cases, the depletion of a necessary end product, causes a variety of metabolic derangements, often with significant neurological sequelae. The severity and the age of onset of symptoms usually, although not always, depend on the amount of residual enzyme activity.

Genetic disorders are inherited by various modes. The most common pattern is autosomal recessive, where each parent carries a recessive in a specific gene in one of the autosomal chromosomes, i.e., not one of the sex-determining chromosomes, X or Y. As there are four possible combinations for the chromosomes carrying the abnormal gene., the offspring has 25% chance of inheriting two mutations (one from each parent) and therefore being affected (also referred to as being homozygous for the disorder) with the disorder, a 50% chance of inheriting one mutation from one or other parent, thus being an unaffected carrier (heterozygous) and a 25% chance of inheriting two normal genes.

In autosomal dominant inheritance, a dominant mutation in one of the autosomal chromosomes is inherited in 50% of the offspring, irrespective of sex. The affected parent is usually symptomatic, although the onset of symptoms may not be apparent until later in life (such as in Huntington's disease), or the spectrum of severity may be variable in different generations, so the disease may not always be recognized. Approximately 30% of dominant mutations arise spontaneously and are not inherited.

In X-linked inheritance, the mutation occurs on the X chromosome. Females have two X chromosomes (one of which is always inactivated, usually randomly), and males an X and a Y. At fertilization, the mother contributes one X and the father either an X or a Y chromosome. If there is a mutation on the inherited X chromosome and the offspring receives a Y chromosome from the father, the mutation will be expressed. If the offspring receives a normal X chromosome from the mother, and an X from the father, the offspring will be a normal female. If the abnormal X is received with a Y from the father, the son will be affected. There is thus a 50% chance of sons from a carrier mother being affected and a 50% chance of daughters being carriers. In rare circumstances, X inactivation is nonrandom or skewed, so that the female carrying the abnormal gene can be symptomatic. X inactivation is also often referred to as lyonization, named after Dr. Mary Lyon who first described the phenomenon. In X-linked disorders, approximately 30% of mutations arise spontaneously, so not inherited from the mother.

The vast majority of genetic disorders are inherited in an autosomal recessive fashion.

While the individual inborn errors of metabolism are rare, based on recent results of expanded newborn screening programs (in which more than 30 disorders can be detected), the overall incidence is approximately 1/5000 live births worldwide (Waters et al., 2018). The incidence of disorders may vary in different populations because of the "founder effect," where a specific mutation arises and is maintained in subsequent generations, or where there is a higher incidence of consanguinity.

With a few exceptions, infants are normal at birth because the placenta efficiently eliminates the toxic metabolites, but symptoms can become apparent in some cases, within days.

Newborn screening

Mass population screening of newborns was introduced in the 1960s, initially for phenylketonuria (PKU), after the development of the bacterial inhibition assay (BIA) for phenylalanine by Guthrie (1961). This simple method, popularly referred to as the Guthrie Test, is still the mainstay of screening for PKU in much of the world. Essentially, the method entails the addition of a solution of *B. subtilis* into an agar well, into which is added a standardized punched sample from the newborn screening filter paper from which the blood is then eluted. High levels of phenylalanine inhibit growth of the bacteria, and the laboratory technician can easily

visually identify this "no-growth" zone as abnormal. Quantification is necessary, using a follow-up method such as high performance liquid chromatography (HPLC). BIA has been adapted for screening for elevated levels of leucine (for maple syrup urine disease (MSUD)) and for methionine (for homocystinuria).

The most significant advance in newborn screening since its inception has been the adaptation of tandem mass spectrometry (MS/MS), (Millington et al., 1990). With this technology, multiple compounds can be identified (both amino acids and acylcarnitine species) from the same dried blood filter paper sample after a simple preparation. More than 30 different inborn errors of metabolism can now be identified. The major drawback, however, is the relative expense of the equipment and lack of long-term data on the outcome of infants detected and treated presymptomatically. Further modification of MS/MS will enable future screening for many more inborn errors of metabolism, as well as increased efficiency by adding already screened disorders to the same platform. Methods have been developed for screening for several lysosomal storage disorders and introduced in several countries. New screening methods are also under development.

Disorders of protein metabolism

Amino acid disorders

Amino acidopathies are due to an enzyme deficiency early in the catabolic pathway of one or more amino acids that results in the accumulation of the amino acid(s), detected by plasma amino acid analysis of serum or plasma. Symptoms may be due to the chronic accumulation of toxic amino acid(s) or due to acute metabolic decompensation, for which aggressive intervention is necessary to prevent death or severe morbidity. Treatment is dietary restriction of the toxic amino acid by limiting the intake of whole protein and supplementing with special modular amino acid formulas to provide the appropriate nutrients for normal growth and development. All disorders are inherited in an autosomal recessive fashion.

The classic example is PKU. In PKU, a deficiency of the phenylalanine hydroxylase (PAH) enzyme (Fig. 1) results in a high level of phenylalanine, which, if not treated with dietary restriction of phenylalanine in the early newborn period, causes severe, irreversible mental retardation. The diagnosis is confirmed by finding a phenylalanine level 1200 mmol/L^{-1} in an infant on unrestricted protein intake. The incidence of PKU is approximately 1/20,000 in caucasians (Hillert et al., 2020). Although PKU is pan ethnic, the incidence varies in certain populations. The level of phenylalanine can vary in individual patients because of the amount of residual enzyme activity, which in turn depends on the specific variants. There are currently more than 400 known variants. Most have biallelic variants previously described as compound heterozygote) i.e., have one copy each of two different variants). Before the introduction of newborn screening, PKU was the commonest cause for inherited mental retardation. Early recognition of the presymptomatic infant allows for initiation of dietary treatment with a phenylalanine-restricted diet, with the best outcomes when recommended phenylalanine levels are attained by 2 weeks of age.

Untreated patients develop progressive severe mental retardation, often with seizures and Parkinson disease-like neurological symptoms. The primary pathogenesis is due to the toxic effect of phenylalanine on the central nervous system; secondary symptoms

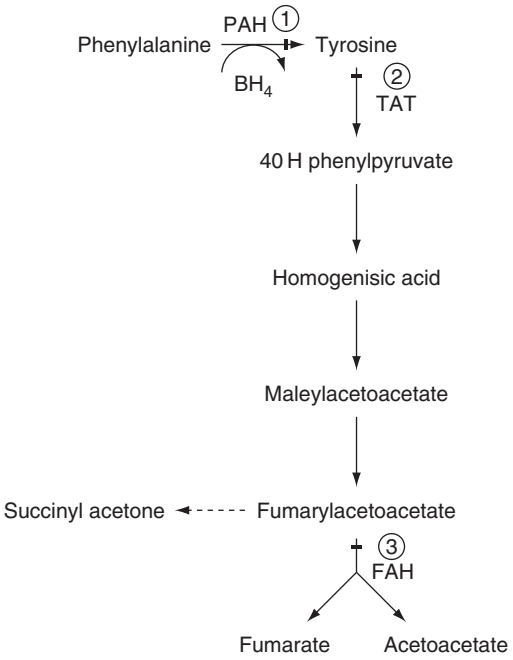


Fig. 1

may be due to a deficiency of tyrosine, which is an important precursor for the synthesis of some neurotransmitters. These symptoms include anxiety and depression.

Treatment is discussed in detail in the following article.

Benign or mild hyperphenylalaninemias is due to allelic variants of PAH, which result in greater residual enzyme activity. On an unrestricted diet, levels are typically in the range 120–360 mmol/L^{-1}, and no dietary treatment is necessary.

Moderate elevation of phenylalanine is also present with defects of tetrahydrobiopterin (BH4), the cofactor for PAH. BH4 is also the cofactor for other enzymes, tryptophan hydroxylase and tyrosine hydroxylase. These amino acids are important precursors for the neurotransmitters 5-hydro-xytryptophan and dopamine. A deficiency causes a neurological syndrome characterized by hypotonia, seizures, and movement disorder (dystonia).

Maple syrup urine disease (MSUD)

MSUD has an incidence approximately 1/185,000 births It is due to a deficiency of the branched chain ketoacid dehy-drogenase enzyme, resulting in the accumulation of the branched chain amino acids (BCAAs), leucine, isoleucine, and valine, which are detected on plasma amino acid analysis (Chuang, 1998). Elevation of alloisoleucine (a derivative of isoleucine) is pathognomonic. In classic MSUD, symptoms typically occur in the first week of life and, if untreated, rapidly progress to cerebral edema, coma, and death. Toxicity is due primarily to high levels of leucine. The characteristic maple syrup (or burnt sugar) odor is due to presence of sotolone, a metabolite of isoleucine or alloisoleucine. It is only detectable when the BCAAs are significantly elevated; the ester is concentrated in the urine and in the earwax of affected patients. Ingestion of the spice, fenugreek, produces a similar odor, but the branched chain amino acids are normal.

Variant forms of MSUD occur. Intermediate MSUD typically presents later in infancy with developmental delay; seizures may occur. Moderate levels of the BCAAs (including alloisoleucine) are present.

Intermittent MSUD is associated with intermittent symptoms during acute infections or periods of prolonged fasting. Typical symptoms include ataxia, vomiting, and seizures. Acute severe decompensation may occur, similar to the classic form of MSUD. The BCAAs are only elevated during the episode of acute symptoms. Other disorders are listed in **Table 1**.

Urea cycle disorders

Urea cycle defects are due to enzyme deficiencies associated with the elimination of waste nitrogen produced by the normal catabolism of protein (AhMew et al., 2003). There are six enzymatic steps involved in this process (**Fig. 2**): a deficiency in any of the first five enzymes causes accumulation of nitrogen, in the form of ammonia (NH3) and increased levels of the amino acids glutamine, alanine and glycine. Glutamine and alanine are the major carriers of nitrogen.

Another very rare disorder has recently been characterized, due to a defect of N-acetylglutamate synthase (NAGS). N-acetylglutamate is the essential cofactor for the first enzyme in the urea cycle, carbamoylphosphate synthase (CPS).

Symptoms occur typically in the newborn period, except for arginase deficiency, but milder late-onset variants have been well described. Symptoms include lethargy, poor feeding, vomiting, tachypnea, and progressive encephalopathy. Routine biochemical testing shows respiratory alkalosis and hyperammonemia. The liver transaminases are usually elevated. Hypoglycemia is not typical.

Plasma amino acid and urine organic acid analysis is necessary to make a presumptive diagnosis. In ASS, citrulline is elevated, in ASL argininosuccinic acid and citrulline are elevated and in arginase deficiency, arginine is elevated, and ornithine low.

Table 1

Disorder (deficient enzyme)	*Elevated analyte*	*Clinical features*	*Treatment*
Tyrosinemia type I (fumarylacetoacetase)	Tyrosine SA	Cirrhosis Liver failure Failure to thrive Renal tubular acidosis Rickets Hepatocellular carcinoma (late)	NTBC (inhibits SA production) Tyrosine restriction
Tyrosinemia type II (tyrosine aminotransferase)	Tyrosine (↑↑)	Keratoconjunctivitis Palmar keratosis Mental retardation	Tyrosine restriction
Homocystinuria (cystathionine β synthase)	Methionine Total homocysteine Free homocystine+ Mixed disulfides	Mental retardation Thromboembolism Lens dislocation Osteoporosis	Vitamin B_6 (50% respond) Methionine restriction
Nonketotic hyperglycinemia (glycine cleavage enzyme deficiency)	Glycine (↑↑) (plasma and CSF)	Seizures developmental delay	Sodium benzoate (decreases glycine)

CSF, cerebrospinal fluid; SA, succinylacetone; NTBC, 2-(2-nitro-4-trifluoro-methylbenzoyl)-1,3-cyclohexanedione.

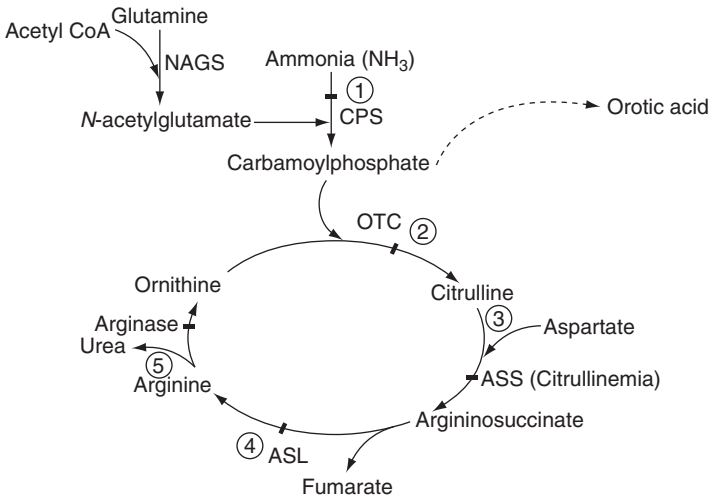


Fig. 2

In OTC, however, the citrulline is very low or absent. In each of these disorders, orotic acid is also present (found on urine organic acid analysis), produced via the pyrimidine cycle from excessive carbamoylphosphate that accumulates due to the deficiency of enzymes distal to CPS. In CPS and NAGS, however, the amino acids and orotic acid levels are normal, so the diagnosis is essentially one of exclusion in a patient who presents with the typical symptoms and severe hyperammonemia in which no other cause is determined. Confirmation of the diagnosis has previously required liver biopsy for enzyme analysis for CPS, NAGS, and OTC, however molecular genetic detection of pathogenic variants is more available, and necessary for confirmation of an OTC carrier mother. Skin fibroblasts can be assayed for ASS and ASL and red blood cells for arginase deficiency. .

OTC is the most common urea cycle defect. It is inherited in an X-linked disorder (all other disorders are autosomal recessive). Male infants usually present in the first few days of life with severe life-threatening hyperammonemia. Newborn screening may be abnormal because of very low or absent citrulline, although severe symptoms and irreversible brain damage may occur before results are available. Females carrying the X-chromosome variant may be asymptomatic, but can present with variable symptoms ranging from acute hyperammonemia to recurrent episodes of nausea, vomiting, and headache. The severity of these symptoms depends on the degree of lyonization (the inactivation of one of the X chromosomes) and the resultant residual enzyme activity. Some women may remain asymptomatic and a diagnosis is made after the birth of a symptomatic son.

Patients with ASL deficiency also have progressive liver cirrhosis, possibly due to the direct toxic effect of the argininosuccinic acid. In arginase deficiency, hyperammonemia is rare (most of the urea has already been eliminated) but arginine itself is toxic to the central nervous system causing progressive spastic quadriplegia and developmental delay; seizures are common.

The toxicity of these disorders is primarily due to the accumulation of ammonia (NH3) and glutamine, which is increased because of the transfer of excess ammonium ions (transamination). Acute, severe hyperammonemia in the newborn period is catastrophic and often fatal. Survivors have variable neurological deficits. Increased awareness and prompt diagnosis and treatment can improve outcomes.

Acute treatment of hyperammonemia due to a urea cycle defect is elimination of dietary protein, elimination of ammonia (by hemodialysis or peritoneal dialysis), high concentration of intravenous dextrose to reverse catabolism, arginine (except in arginase deficiency) to regenerate the cycle, and the nitrogen scavenging drugs, sodium benzoate (which conjugates with glycine to form hippurate) and sodium phenylacetate (which conjugates with glutamine to form sodium phenylbutyrate), glycerol phenylbutyrate is a recently developed drug that is more tolerable for oral administration. An intravenous solution of sodium phenylbutyrate and sodium phenylacetate is used acutely, until the patient can transition to an oral scavenger medication.

Early reintroduction of limited dietary protein is necessary to provide substrate for anabolism and to prevent further catabolism. This should consist of whole protein and a special formula to provide enough essential amino acids to ensure normal weight gain, without producing excessive amounts of nitrogen for ammonia production. Chronic treatment includes similar dietary protein restriction, citrulline and an oral form of nitrogen scavenging medication.

A derivative of the amino acid, L-glutamic acid, N-carbamylglutamate (NCG), has recently been approved for treatment of NAGS, by providing an alternate activator of CPS. For arginase deficiency, dietary protein restriction and formula are usually adequate.

Liver transplant can be an option for treatments. It has been most widely used for infants with OTC, or in patients with argininosuccinic aciduria (ASL), which also often has liver disease.

Gene therapy, where the abnormal gene may be replaced by functional gene, is currently being evaluated in clinical trials for OTC; an enzyme replacement therapy is also in clinical trial for arginase deficiency.

Organic acidemias

Organic academia are due to enzyme deficiencies further along the catabolic pathway, resulting in the accumulation of the toxic products of intermediary metabolism (organic acids). In some cases, there is a functional defect of the enzyme due to a deficiency of the enzyme cofactor, rather than of the enzyme itself. Examples of this are biotinidase deficiency and defects of cobalamin (vitamin $_{B12)}$ metabolism.

Accumulation of large amounts of organic acids causes severe metabolic acidosis and ketosis. Hyperammonemia is often present, due to secondary inhibition of the urea cycle. Hypoglycemia may be variably present, due to secondary inhibition of fatty acid oxidation. Symptoms are often present in the newborn period; recurrent episodes of metabolic decom-pensation can occur because of excessive protein intake or because of catabolism (and therefore an increased load of amino acids endogenously released from muscle) associated with acute infections or prolonged periods of fasting. Morbidity and mortality are due to acute acidosis and the associated neurologic sequelae.

The diagnosis is made by finding high levels of the characteristic organic acids in urine. Newer analytic methods, such as MS/MS can detect even small elevations of characteristic plasma acylcarnitine and urine acylglycine conjugates of the intermediary metabolites. Confirmation is by enzyme analysis, usually in skin fibroblasts. Molecular genetic analysis is available for many disorders. Treatment typically entails limitation of dietary protein and supplementation with a specific metabolic formula that is depleted in the amino acid precursors in the specific catabolic pathway.

Propionic acidemia is a typical organic acidemia. It is due to an isolated defect of the enzyme, propionyl CoA carboxylase in the catabolic pathways of the amino acids isoleucine, valine, methionine, and threonine as well as cholesterol and odd chain fatty acids (**Fig. 3**). The resulting accumulation of the intermediary metabolites, 3-hydroxypropionic acid, methylcitric acid, propionylglycines, and tiglyglycine, can cause severe metabolic acidosis, ketosis, coma, and death. Other associated symptoms can be hyperammonemia, hypoglycemia, and pancytopenia, due to bone marrow suppression by the accumulated toxic organic acids, cardiomyopathy and renal dysfunction, can occur, usually later in life.

Symptoms can occur within days of birth in the classic disease or later in infancy or childhood in the milder variant forms. The later-onset form may be associated with persistent vomiting, failure-to-thrive, and developmental delay, but often without severe episodes of metabolic acidosis. Dystonia may occur due to infarction of the basal ganglia.

Liver and/or kidney transplant may improve outcomes in some patients.

The potential for gene therapy is being evaluated in animal studies.

Cofactor deficiencies

Biotin is an essential cofactor for the four carboxylase enzymes, propionyl CoA carboxylase, methylcrotonyl CoA carboxylase, pyruvate CoA carboxylase, and acetyl CoA car-boxylase. It is endogenously derived from lysine and also present in its protein-bound form in small amounts in many foods. Holocarboxylase synthetase (HCS), which forms the inactive parent apoenzyme, is also biotin dependent. Enzyme activation requires free biotin, which is released by the action of biotinidase; this enzyme also plays an essential role in the recycling of biotin for further use. A deficiency of biotinidase, therefore, results in depletion of biotin and a functional defect of the carboxylases. Symptoms include hypotonia, lethargy, vomiting, and ataxia. Recurrent metabolic acidosis may occur. Alopecia and a generalized erythematous rash are common. The symptoms are more severe in HCS deficiency. The characteristic pattern of organic acids is present in both disorders. The diagnosis is made by measuring biotinidase activity in plasma or carboxylase enzyme activity in leukocytes or fibroblasts. Treatment with pharmacologic doses of biotin is effective.

Multiple defects of cobalamin (vitamin $_{B12)}$ metabolism can occur, starting with the transport of vitamin $_{B12}$ into the cell (defects of the transporter proteins, Transcobalamin I and II) or subsequent intracellular utilization of the different biologically active forms. These disorders are classified as complementation groups, depending on whether the defect is in adenosylcobalamin (Cbl A and B), methylcobalamin (Cbl G and E), or both (Cbl C and D).

Adenosylcobalamin is the cofactor for methylmalonyl CoA mutase; a defect results in a milder form of methylmalonic acidemia than found with a defect of the enzyme itself. Methylcobalamin is the cofactor for methionine synthase, which results in low methionine and homocystinuria (distinct from classic homocystinuria due to a defect of cystathionine b synthase). A defect of both adenosyl and methylcobalamin causes both methylmalonic acidemia and homocystinuria (**Table 2**).

Symptoms vary with the complementation group, but can include metabolic acidosis, hypotonia, developmental delay, macular degeneration, and megaloblastic anemia.

Treatment with hydroxocobalamin corrects some of the biochemical derangements, especially in Cbl A and B. Addition of betaine, a remethylating agent, may decrease homo-cysteine levels in CblC. Treatment is less successful in the other groups.

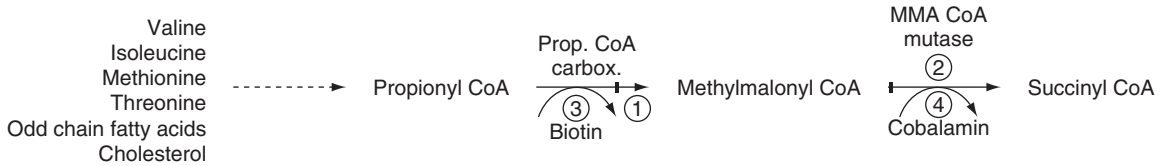


Fig. 3

Table 2

Disorder (deficient enzyme)	*Elevated analyte(s)*	*Clinical features*	*Treatment*
Methylmalonic acidemia (methylmalonyl CoA mutase)	Methylmalonic acid	Metabolic acidosis Hyperammonemia Failure to thrive Vomiting	Protein restriction Carnitine
Isovaleric acidemia (isovaleryl CoA dehydrogenase)	Isovalerylglycine	Metabolic acidosis Vomiting	Protein restriction Glycine
Glutaric aciduria type I (glutaryl CoA dehydrogenase)	3-Hydroxyglutaric acid Glutaric acid	Metabolic acidosis Vomiting Macrocephaly Developmental delay	Protein restriction Carnitine
3-Methylcrotonyl glycinuria (3-methylcrotonyl CoA carboxylase)	3-Hydroxyisovaleric acid 3-Methylcrotonylglycine	Metabolic acidosis Hypoglycemia Hyperammonemia Seizures (Some patients asymptomatic)	Protein restriction Carnitine
Mitochondrial acetoacetyl CoA thiolase deficiency	2-Methyl-3-hydroxybutyrate acid 2-Methylacetoacetic acid Tiglylglycine	Metabolic acidosis Vomiting	Protein restriction Carnitine

A syndrome similar to Cbl C has been described in the breastfeeding infants of strict vegetarian (vegan) mothers and in mothers with pernicious anemia, who are vitamin B12 deficient.

Disorders of fatty acid oxidation

These disorders have only been recognized since the early 1980s, but as a group are the most common inborn errors of metabolism. Fat provides a source of energy as ATP and ketone bodies during times of metabolic stress (such as febrile illness) or with prolonged fasting. Free fatty acids, released from the adipose tissue, are transported into the mitochondria via the carnitine shuttle system, where they then undergo beta-oxidation (Fig. 4), the progressive cleavage from an 18-carbon length very long-chain fatty acid to the 2-carbon acetoacetyl CoA, the substrate for the TCA cycle, ketones and ATP through the mitochondrial respiratory chain. A deficiency of any of the enzymes in this pathway can cause symptoms, primarily hypoketotic hypoglycemia and hepatic encephalopathy, with hyperammonemia (due to secondary inhibition of the urea cycle) and sudden death. Many cases of what would previously have been diagnosed as Reye syndrome are now known to be due to fatty acid oxidation defects. Symptoms can occur at any time, from the newborn period to adulthood.

Carnitine has a dual role: as well as its critical role in the transport of free fatty acids into mitochondria, it also conjugates with the fatty acyl CoA intermediates that accumulate proximal to an enzyme block, forming acylcarnitine species that can be excreted by the kidneys. They can also be measured in plasma for diagnostic purposes and in the newborn screening dried blood spot. Increased utilization of carnitine due to an enzyme defect causes a secondary depletion, further impairing fatty acid oxidation.

Defects of the carnitine shuttle (CPTI, CACT, CPT II) and beta-oxidation (VLCAD, TFP, and LCHAD) may present in the newborn period or later in infancy with severe hypoketotic hypoglycemia, cardiomyopathy and hepatic encephalopathy, due to deposition of fat in the heart and liver. Rhabdomyolysis (lysis of muscle cells) is common. Pigmentary degeneration of the retina is almost universal in patients with LCHAD, thought to be due to impaired endogenous production of DHA, which is necessary for normal retinal function. Peripheral neuropathy is common in TFP and retinopathy has been found in older patients Variant forms of CPT II and VLCAD may present in adolescence or adulthood with exercise intolerance, muscle cramping and rhabdomyolysis, which may be severe enough to cause acute renal failure due to the deposition of the muscle pigment, myoglobin, in the renal tubules, causing acute necrosis.

Treatment for these disorders includes frequent feeding and avoidance of fasting and limitation of long-chain fat, which is the most abundant fat in human diets. Supplementation with medium-chain triglycerides), which bypass the long-chain enzyme deficiencies due to independent diffusion into the mitochondria, provide an alternate substrate for energy metabolism. MCT refers even chain triglycerides, of 6, 8 and 10 carbon length.

Special formulas (medical foods) can provide the medium chain triglycerides, but some are deficient in essential fatty acids, such as linoleic and linolenic acids and DHA. Addition of oil, such as canola, provides most of the essential fatty acids. DHA can be added to the formula as walnut oil. Uncooked cornstarch has been used to provide an alternate source of complex carbohydrate (especially for overnight fasting) after the age of about 9 months. Normal pancreatic amylase activity is necessary, and may not be active e

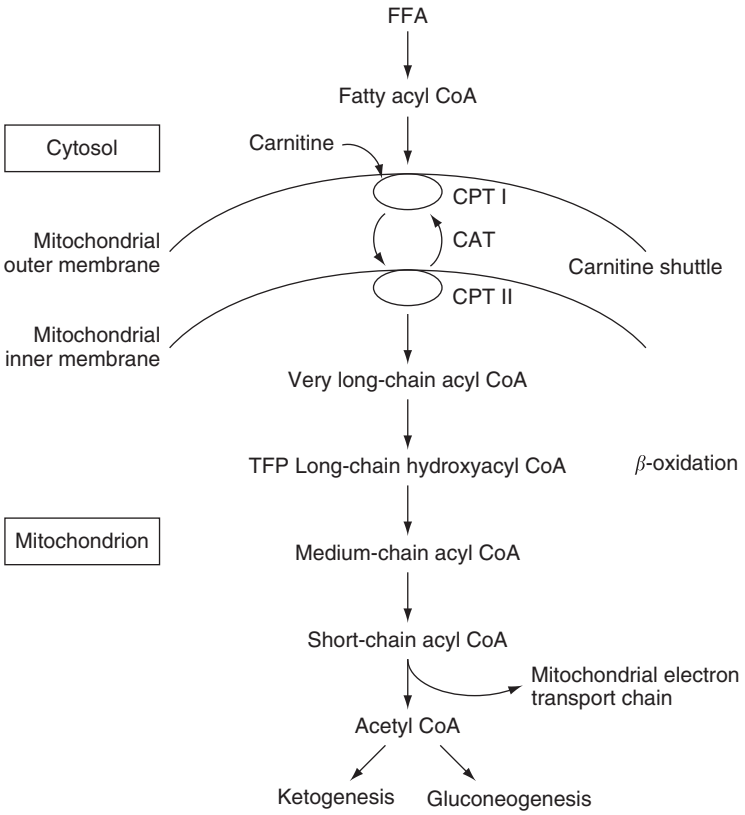


Fig. 4

before this age. A dietary guideline for VLCAD has recently been published (Van Calcar et al., 2020). Nutrition management guideline for very-long chain acyl-CoA dehydrogenase deficiency (VLCAD): an evidence-and-consensus-based approach).

In 2020, FDA (Food and Drug Administration) approved triheptanoin (Dojolvi), a 7 carbon chain triglyceride, in the US, as a source of calories in children and adults with long-chain fatty acid oxidation disorders. FDA based its approval on a 4-month double-blind randomized controlled study in which triheptanoin and trioctanoin (8-carbon demonstrated similar efficacy) (Gillingham et al., 2017).

Treatment for the medium- and short-chain defects (MCAD and SCAD) is simpler, involving avoidance of fasting and early intervention during acute illness to prevent hypoglycemia. Carnitine supplementation is frequently used to prevent secondary depletion. Dietary fat recommendations are approximately 30% of total calories, or a "heart healthy" diet for MCAD. Most cases of SCAD (detected on newborn screening) are due to two common variants in the general population, and considered to be non-pathogenic (Pedersen et al., 2008).

Newborn screening, where available, may detect the fatty acid oxidation disorders and allow early confirmation and intervention. This has resulted in significant improvement in morbidity and mortality.

Disorders of carbohydrate metabolism

Galactosemia

Galactose is derived primarily from dietary lactose, which is the major disaccharide in dairy products, human breast milk and many fruits and vegetables. There is also a small contribution from endogenous production. There are three known enzyme deficiencies in the pathway that oxidizes galactose to glucose (Fig. 5); all are autosomal recessive genetic disorders.

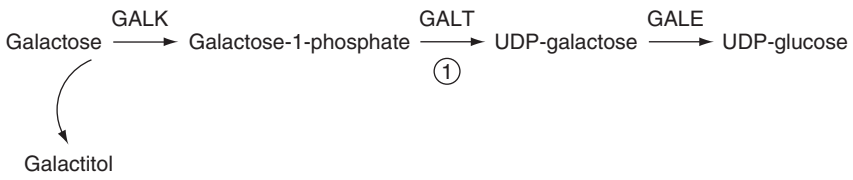


Fig. 5

Classic galactosemia is due to almost complete absence of galactose-1-phosphate uridyl transferase enzyme (GALT) activity. Symptoms generally occur in the first few weeks of life, with poor weight gain, lethargy, hypotonia, and liver disease (hyperbilirubinemia, coagulopathy, and hepatomegaly) and renal tubular acidosis. Hypoglycemia can occur, but is rare. *E. coli* sepsis may also be a complication; elevated galactose is thought to impair leukocyte bactericidal activity, allowing the bacteria to more easily invade the red blood cells with subsequent dissemination. Cognitive impairment may be a long-term complication.

The underlying pathogenesis of galactosemia is not fully understood; despite compliance with the lactose-restricted diet, speech delay is almost universal in children; some patients have learning disorders; many female patients develop ovarian failure, though successful pregnancies have been reported. Early recognition and hormone replacement therapy may be used.

Treatment is restriction of lactose in the diet, primarily by elimination of dairy products and other foods known to be high in galactose (Welling et al., 2017).

Variant forms of galactosemia occur due to mutations in the GALT gene that result in greater residual enzyme activity. The commonest variant is the Duarte Variant, in which there is usually one copy of a classic galactosemia mutation (e.g., Q188R) and one copy of the variant N314D. This combination results in approximately 25% residual enzyme activity. There is varying opinion on whether or not dietary treatment is necessary; some programs consider that the residual enzyme activity is adequate to prevent the pathologic sequelae, others elect to treat with lactose restriction for the first year of life. There are no long-term outcome data to support either approach (Fridovich-Keil et al., 2020).

Galactokinase (GALK) deficiency causes excessive accumulation of galactitol, which is oxidized from galactose by an alternative pathway. High levels of galactitol cause cataract formation, which is the only symptom of this disorder. Lactose restriction is necessary.

Epimerase deficiency (GALE) is very rare. There are two isoforms of the enzyme, one in red blood cells, and one in the liver. The most common disorder is due to an isolated deficiency of the RBC isoform, which will be detected incidentally by newborn screening programs that measure total galactose. There are no clinical symptoms and no treatment is necessary. A defect of both isoforms will cause symptoms similar to classic galactosemia and should be treated similarly.

Glycogen storage disorders

Glycogen is a complex carbohydrate stored primarily in liver and muscle. Liver glycogen provides glucose to maintain blood sugar levels in between normal feedings; defects of the liver enzymes for glycogen degradation lead to hypoglycemia and/or liver disease because of excessive accumulation of glycogen. Muscle glycogen is an important substrate for energy production and normal muscle function; symptoms are typically exercise induced muscle pain or rhabdomyolysis (Ellingwood and Cheng, 2018).

Glycogen Storage Disease Type I (GSD Ia, (von Gierke disease) (**Fig. 6**), the most common disorder, is due to a deficiency of glucose-1-phosphatase in liver, kidney, and intestinal mucosa. Symptoms typically occur in infancy when frequency of feeding decreases. Profound hypoglycemia can occur; progressive hepatomegaly and liver dysfunction is due to storage of glycogen. Other metabolic derangements include hypertriglyceridemia and hyper-cholesterolemia (causing xanthomas); hyperuricemia (causing gout and renal calculi) is due to decreased renal excretion) as well as increased uric acid production due to phosphate depletion. Other long-term complications include progressive renal disease (proteinuria) and hepatocellular carcinoma. Treatment is frequent meals and continuous nocturnal feeding (in infants) and supplemental uncooked cornstarch to provide exogenous glucose from the age of about 9 months when pancreatic amylase is mature and can metabolize the corn starch (Kishnani et al., 2014).

Gene therapy is being evaluated in clinical trials for GSD 1a and III (www.Clinicaltrials.gov) (gene therapy, GSD1a, GSD 111).

Other disorders are summarized in **Table 3**.

Disorders of fructose metabolism

There are three disorders of fructose metabolism, all inherited in an autosomal recessive fashion. Fructose is widely distributed in the diet as the primary sugar in fruits, vegetables, and honey. It is also derived from sucrose and sorbitol, which are also found in a large variety of products, including infant formulas and intravenous fluids. The toxic effect of fructose is due to inhibition of gluconeogenesis by high levels of fructose-1-phosphate and subsequent depletion of inorganic phosphate and thus, ATP.

Essential fructosuria is a benign disorder due to a defect of the enzyme, fructose kinase. Patients have increased urinary excretion of fructose, which is usually an incidental finding on routine testing for reducing substances.

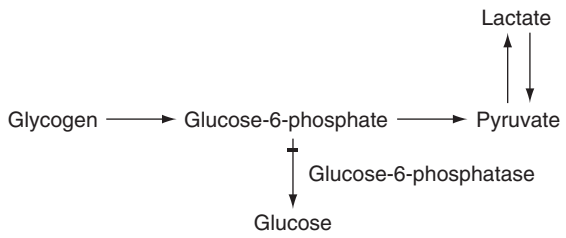


Fig. 6

Table 3

Disorder	*Deficient enzyme*	*Primary affected tissue*	*Symptoms*	*Treatment*
GSD 0	Glycogen synthase	Liver	Hypoglycemia	Uncooked cornstarch, frequent feeds
GSD I	Glucose-6-phosphatase	Liver, muscle	Hypoglycemia, hepatomegaly, growth retardation, proteinuria, lactic acidemia, hyperlipidemia, hyperuricemia (gout), hepatocellular carcinoma	Uncooked cornstarch, frequent feeds
GSD II (Pompe disease)	Acid maltase (α glucosidase)	Lysosomes of muscle (skeletal and cardiac)	Cardiomyopathy, skeletal myopathy, cardiorespiratory failure	Enzyme replacement (in clinical trial)
GSD III	Debranching enzyme (amylo-1, 6-glucosidase)	Liver, muscle	Hypoglycemia (mild), hepatomegaly, myopathy, hyperlipidemia	Uncooked cornstarch, frequent feeds
GSD IV (amylopectinosis)	Branching enzyme	Liver	Hepatomegaly, cirrhosis, liver failure, myopathy	Liver transplant, uncooked cornstarch
GSD V (McArdle disease)	Myophosphorylase	Muscle	Muscle cramping (with exercise)	Oral glucose, high-protein diet
GSD VI (Hers disease)	Liver phosphorylase	Liver	Hepatomegaly, hypoglycemia, myopathy	Frequent feeds
GSD VII (Tarui disease)	Phosphofructokinase	Muscle	Fatigue exercise intolerance, cramping	Avoidance of strenuous exercise
GSD IX	Phosphorylase kinase	Liver, muscle	Hepatomegaly, growth retardation	Frequent feeds

GSD, glycogen storage disorder.

Hereditary fructose intolerance (HFI) is due to a deficiency in aldolase B, which splits fructose-1-phosphate into glycer-aldehyde and dihydroxyacetone. Symptoms only occur after exposure to fructose, usually from dietary ingestion, although they are more severe after intravenous infusion. These symptoms include gastrointestinal dyscomfort, vomiting, and hypoglycemia. Chronic exposure causes failure to thrive, liver disease, and renal tubular acidosis. Affected patients are often misdiagnosed as having behavioral problems or an eating disorder. Treatment is elimination of fructose from the diet.

Fructose-1,6-bisphosphatase is a defect of gluconeogenesis and is not dependent on exposure to fructose. Symptoms of recurrent episodes of vomiting, lactic acidosis, tachypnea, seizures and apnea, occur when dietary glucose and glycogen stores are depleted, such as during periods of fasting or with febrile illnesses. Approximately 50% of patients are symptomatic in the newborn period. Treatment is prevention of fasting and supplementation with uncooked cornstarch to provide a source of complex carbohydrate. Acute episodes respond to intravenous infusions of dextrose.

Disorders of micronutrient metabolism

Disorders of copper metabolism

Copper plays an essential role in normal cell metabolism as a cofactor for several critical enzyme pathways. Normal homeostasis is regulated through a balance of gut absorption and biliary excretion (Hasan and Lutsenko, 2021). There are two major inherited disorders of copper metabolism: Wilson disease and Menkes disease.

Wilson disease (Mulligan and Bronstein, 2020) is an autosomal recessive disorder with an incidence of approximately 1/30,000 (higher in China, Japan, and Sardinia), due to a deficiency of hepatic ATPase(ATP 7B gene). Symptoms usually start in childhood, but there is a wide range of clinical features. Copper that cannot be excreted in the bile is initially deposited in the liver, with reduced incorporation into the carrier protein ceruloplasmin, causing chronic liver dysfunction (cirrhosis), and in some cases, acute liver failure. Increasing levels of copper are also deposited in other tissues, especially the cornea (Kayser-Fleischer rings) and the central nervous system, which is usually later in life, manifesting as a movement disorder and poor coordination. Difficulties with speech and swallowing can also occur. Psychiatric symptoms are also common, often beginning in adolescence.

The diagnosis is made by finding decreased serum ceruloplasmin with increased urine excretion of copper; copper deposition may also be seen on liver biopsy. Confirmation is by mutation analysis.

Treatment is aimed at increasing the excretion of the stored copper with chelating agents, such as D-penicillamine, and a copper-depleted diet. Liver transplantation may be an option. Clinical trials for gene therapy are underway (www.clinicaltrials.gov) Wilson Disease, gene therapy.

Menkes disease is an X-linked disorder due to a deficiency of the membrane transporter, ATP 7A, resulting in intracellular copper accumulation and lack of transport to critical tissues, such as the CNS, kidneys, and also to connective tissue. It usually presents in the first few months of life with typical facial features and unusual kinky hair (pili torti), seizures, poor weight gain, and loss of developmental milestones. Death usually occurs by approximately 3 years of age. Diagnosis is made by finding decreased serum

copper and ceruloplasmin and mutation analysis. Treatment with copper chloride injections before 10 days of age may slow the progression of the disease in some patients (Vairo et al., 2018).

A milder form of the disease (due to higher residual enzyme activity) is Occipital Horn Syndrome. Patients have characteristic occipital calcifications (occipital horns), borderline cognitive delay, a distinctive facial appearance, and connective tissue abnormalities (e.g., joint laxity). Treatment with copper-histidine may be beneficial if started in the neonatal period.

Disorders of iron metabolism

Hemochromatosis

There are four different subtypes of hereditary hemochromatosis, due to excessive absorption of iron from the gastrointestinal system, leading to toxic accumulation of iron in various tissues.

Hereditary hemochromatosis is the most common disorder, inherited in autosomal recessive fashion, due to pathogenic variants in the HFE gene (Gan et al., 2011). The carrier frequency is approximately 1:10 in Caucasians with a disease frequency of approximately 1/400. The typical presentation is in adults by age 40–60 years, with increased skin pigmentation, chronic liver disease (cirrhosis), portal hypertension, and diabetes mellitus (also called 'bronzed diabetes'), due to excessive iron deposition in skin, liver, and pancreas, respectively. Deposition also occurs in the heart and other endocrine tissues, including the pituitary gland, leading to hypogonadism in men. Hepatocellular carcinoma is a complication in approximately 30% of patients. There appears to be a male preponderance for end-stage organ failure. Homozygosity for the common mutation, C282Y appears to be associated with greater iron overload, but there is no clear correlation with disease severity. The diagnosis is suggested by increased serum iron and ferritin studies, with confirmation by mutation analysis. Current treatment is removal of excess iron by regular phlebotomy.

Other rare forms of inherited hemochromatosis are juvenile hemochromatosis (Piperno et al., 2020). Which is due to bi-allelic pathogenic variants in the HJV or HAMP genes. It presents in the first to third decades, with hypogonadotropic hypogonadism, cardiomyopathy, liver fibrosis, cirrhosis. Hemochromatosis type 3, due to a defect in the transferring receptor-2 gene, is similar to HFE-related hemochromatosis.

Type 4 is an autosomal dominant disorder (only one copy of the pathogenic variant is required. It transports iron from the diet.

Neonatal hemochromatosis (Feldman and Whitington, 2013) also known as Gestational Alloimmune Liver Disease (GALD) is very severe and presents as neonatal liver failure or cirrhosis. Fetal demise and/or still-birth can occur. When a mother is exposed to a fetal antigen that is not recognized as "self", IgG antibodies, which activates a complement cascade, causing liver injury. Mortality is high. Early intervention with IVIG may be beneficial.

Conclusions

Inborn errors of metabolism, while individually rare, are the cause of a significant number of inherited disorders. With the development and continued evolution of newborn screening, many more disorders are detected early and may benefit from early treatment.

The classic management with dietary treatment will likely continue to be the mainstay for many disorders, however new approaches, such as gene therapy, and newer medications, may provide much greater tools for managing many disorders and improving outcomes and quality of life.

References

AhMew, N., Simpson, K., Gropman, A., 2003. GeneReviews®. University of Washington, Seattle, Seattle (WA), pp. 1993–2021.
Chuang, D.T., 1998. Maple syrup urine disease: it has come a long way. J. Pediatr. 132 (3Pt2), 517–523.
Ellingwood, S., Cheng, A., 2018. Biochemical and clinical aspects of glycogen storage diseases. J. Endocrinol. 238 (3), 131–141.
Feldman, A., Whitington, F., 2013. Neonatal hemochromatosis. J. Clin. Exp. Hepatol. 3 (4), 313–320.
Fridovich-Keil, J.L., Gambello, M.J., Singh, R.H., Sharer, J.D., 2020. Duarte variant galactosemia. In: GeneReviews®. University of Washington, Seattle, Seattle (WA), pp. 1993–2021.
Gan, E.K., Powell, L.W., Olynyk, J.K., 2011. Natural history and management of HFE-hemochromatosis. Semin. Liver. Dis. 31 (3), 293–301.
Garrod, A.E., 1908. Inborn errors of metabolism. Lancet 4427, 1–7.
Gillingham, M.B., Heitner, S.B., Martin, J., et al., 2017. Triheptanoin versus trioctanoin for long-chain fatty acid oxidation disorders: a double blinded, randomized clinical trial. J. Inherit. Metab. Dis. 40 (6), 832–843.
Guthrie, R., 1961. Blood screening for phenylketonuria. JAMA 178, 86.
Hasan, N.M., Lutsenko, L., 2021. Regulation of copper transporters in human cells. Curr. Top Membr. 69, 137–161.
Hillert, F., Anikster, Y., Belanger-Quintana, A., 2020. The genetic landscape and epidemiology of phenylketonuria. Am. J. Hum. Genet. 6 (107), 234–250.
Kishnani, P.S., Austin, S.L., Abdenur, J.E., 2014. Diagnosis and management of glycogen storage disease type 1: a practice guideline of the American College of Genetics and Genomics. Genet. Med. 16 (11).
Millington, D.S., Kodo, N., Norwood, D.L., Roe, C.R., 1990. Tandem mass spectrometry: a new method for acylcarnitine profiling with potential for newborn screening for inborn errors of metabolism. J. Inherit. Metab. Dis. 13 (3), 321–324.
Mulligan, C., Bronstein, J.M., 2020. Wilson disease: an overview and approach to management. Neurol. Clin. 38 (2), 417–432.
Pedersen, C.B., Kolvraa, S., Kolvraa, A., et al., 2008. The ACADS gene variation spectrum in 114 patients with short-chain acyl CoA dehydrogenase (SCAD) is dominated by missense variations leading to misfolding at the cellular level. Hum. Genet. 124 (1), 43–56.

Piperno, A., Bertola, F., Ventivegna, A., 2020. Juvenile hemochromatosis. In: Adam, M.P., Ardinger, H.H., Pagon, R.A., Wallace, S.E., Bean, L.J.H., Gripp, K.W., Mirzaa, G.M., Amemiya, A. (Eds.), GeneReviews®. University of Washington, Seattle, Seattle, WA, pp. 1993–2022.
Vairo, F.P.E., Chwal, B.C., Perini, S., et al., 2018. A systematic review and evidence- based guideline for diagnosis and treatment of Menkes disease. Mol. Genet. Metab. 126 (1), 6–13.
Van Calcar, S.C., Sowa, M., Rohr, F., Beazer, J., Setlock, T., Weihe, T.U., Pendyal, S., Wallace, L.S., Hansen, J.G., Stembridge, A., Splett, P., Singh, R.H., 2020. Nutrition management guideline for very-long chain acyl-CoA dehydrogenase deficiency (VLCAD): an evidence- and consensus-based approach. Mol Genet Metab 131 (1-2), 23–37. https://doi.org/10.1016/j.ymgme.2020.10.001. Epub 2020 Oct 6. PMID: 33093005.
Waters, D., Davies, A., Woolham, D., et al., 2018. Global birth prevalence and mortality from inborn errors of metabolism: a systematic review of evidence. J. Glob. Health 8 (2), 021102.
Welling, L., Bernstein, L.E., Berry, G.T., et al., 2017. International clinical guideline for management of classical galactosemia: diagnosis, treatment and follow up. J. Inherit. Metab. Dis. 40 (2), 171–176.

Further reading

Burton, B., 1998. Inborn errors of metabolism in infancy; a guide to diagnosis. Pediatrics 102 (6), E69.
Caldovic, L., Ah Mew, N., Shi, O., et al., 2010. N-acetylglutamate synthase: structure, function and defects. Mol. Genet. Metabol. 100 (Suppl. 1), S13–S19.
Chace, O.H., Kalas, T.A., Naylor, E.W., 2003. Use of tandem mass spectrometry for multianalyte screening of dried blood specimens from newborns. Clin. Chem. 49 (11), 1797–1817.
Fernandes, J., Saudubray, J.-M., van den Berghe, G. (Eds.), 2000. Inborn Metabolic Diseases, third ed. Springer-Verlag, Berlin.
Gillingham, M.B., Connor, W.E., Matern, O., et al., 2003. Optimal dietary therapy of long-chain-3-hydroxyacyl-CoA dehydrogenase deficiency. Mol. Genet. Metabol. 79 (2), 114–123.
Holme, E., Lindstedt, S., 1998. Tyrosinemia type I and NTBC (2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione). J. Inherit. Metab. Dis. 21 (5), 507–517.
Kahler, S.G., Fahey, M.C., 2003. Metabolic disease and mental retardation. Am. J. Med. Genet. 117C (1), 31–41.
Levy, H.L., Sepe, S.J., Walton, D.S., et al., 1978. Galactose-1-phosphate uridyl transferase deficiency due to the duarte/galactosemia combined variation: clinical and biochemical studies. J. Pediatr. 92 (3), 390–393.
Lyon, M.F., 2002. X-chromosome inactivation and human genetic disease. Acta Pediatr. Suppl. 91 (439), 107–111.
Podebrad, F., Heil, M., Reichert, S., et al., 1999. 4,5-dimethyl-3-hydroxy-2[5H]-furanone (sotolone)—the odour of maple syrup urine disease. J. Inherit. Metab. Dis. 22, 107–114.
Vockley, J., Singh, R.H., Whiteman, D.A., 2002. Diagnosis and management of defects of mitochondrial beta-oxidation. Curr. Opin. Clin. Nutr. Metab. Care 5 (6), 601–609.
Walter, J.H., Collins, J.E., Leonard, J.V., 2000. Recommendations for the management of galactosemia. UK Galactosemia Steering Group. Arch. Dis. Child. 82 (3), 266.
Zytkovicz, T.H., Fitzgerald, E.F., Marsden, D., et al., 2001. Tandem mass spectrometric analysis for amino, organic and fatty acid disorders in newborn dried blood pots: a two-year summary from the new England Newborn Screening Program. Clin. Chem. 47 (11), 1945–1955.

Inborn errors of metabolism: Nutrition management of phenylketonuria

D.L. Marsden[a], **B. Reineking**[b], **and F.J. Rohr**[c], [a] Boston Children's Hospital, Boston, MA, United States; [b] Ultragenyx Pharmaceutical Inc., Novato, CA, United States; and [c] Met Ed Co, Boulder, CO, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of D.L. Marsden, F.J. Rohr, K.C. Costas, Inborn errors of metabolism: Nutritional Management of Phenylketonuria, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 11–15, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00156-2.

Introduction 424
Introduction to nutrition management 425
Growth 427
Fatty acids 427
Vitamins and minerals for consideration 428
Bone mineral density 428
Maternal PKU 429
Pharmacological therapies 429
Tetrahydrobiopterin 429
Pegvaliase 429
Conclusion 429
References 430

Key points

- Phenylketonuria was the first inherited metabolic disorder to be identified by newborn screening and managed by nutrition to prevent adverse intellectual outcomes.
- A phenylalanine-restricted diet requires the use of medical food to provide protein without phenylalanine; the semi-synthetic nature of the diet and restriction of whole protein result in increased risk for inadequate growth, reduced bone mineral density, as well as essential fatty acid and vitamin and mineral deficiencies.
- While the phenylalanine-restricted diet prevents developmental delay, it is a difficult diet to sustain chronically, and blood phenylalanine concentrations often increase as children reach adolescence and adulthood resulting in suboptimal control and associated neurocognitive deficits.
- Strict control of blood phenylalanine is necessary in pregnant women with PKU to prevent the teratogenic effects of high blood phenylalanine.
- Pharmacological therapies for PKU include sapropterin, a tetrahydrobiopterin supplement that can increase phenylalanine hydroxylase activity in some individuals with PKU and pegvaliase, an injectable enzyme substitution therapy.

List of abbreviations

BH4 Tetrahydrobiopterin
LNAA Large neutral amino acid
PKU Phenylketonuria
RDA Recommended daily allowance
DRI Dietary reference intake

Introduction

Phenylketonuria (PKU) is a disorder of amino acid metabolism caused by a deficiency in the enzyme phenylalanine hydroxylase, which converts the essential amino acid phenylalanine to tyrosine. High concentrations of phenylalanine are toxic to the central nervous system (CNS), resulting in severe irreversible developmental delay. Details of the biochemistry are discussed elsewhere in the encyclopedia.

PKU is often considered a paradigm for the nutritional therapy for metabolic disorders. It was the first inborn error of metabolism identified by newborn screening, thus allowing for early dietary treatment. Early treatment was successful in preventing the

developmental delay associated with untreated PKU. Since the advent of successful nutrition management of PKU over 5 decades ago, the field has expanded greatly, but the principle of treating PKU with diet remains the same—to control the intake of the amino acid that is not metabolized normally. This principle applies to all amino acidopathies, but PKU is used here as an example.

Dietary treatment is started as soon as the diagnosis is confirmed in a newborn. Outcomes are best when the diet is implemented and the phenylalanine concentrations are within the recommended range by 2 weeks of age (Camp et al., 2014). Lifelong treatment is recommended as individuals with PKU who have resumed an unrestricted diet, although intellectually normal, have been shown to have an increased incidence of neuropsychiatric deficits as well as increased anxiety and depression (Bilder et al., 2016).

The pathophysiology of PKU is not well understood, and may be related to the disturbance of normal amino acid concentrations in the brain. Phenylalanine competes with other large neutral amino acids for transport across the blood–brain barrier, and it is theorized that high levels of brain phenylalanine and low levels of other amino acids, specifically tyrosine and tryptophan, may impede neurotransmitter synthesis in the brain and be responsible for the symptoms associated with untreated PKU (Schuck et al., 2015). Although the ideal brain concentration of phenylalanine has not been established, treatment guidelines have been established for blood phenylalanine at various ages, as these guidelines differ slightly in different countries. In the United States, the American College of Medical Genetics (ACMG) recommends that blood phenylalanine be maintained between 120 and 360 μmol/L throughout the lifespan (Vockley et al., 2014).

The goal of nutritional therapy is to keep blood phenylalanine controlled while providing a nutritionally sound diet. This necessitates the use of a special medical food that provides amino acids other than phenylalanine because the phenylalanine restriction required to maintain blood concentration within the desired range is so severe that the amount of natural protein allowed in the diet would not support normal growth and development.

Medical foods vary in amino acid content and most also provide carbohydrates, fats, vitamins, and minerals, but others do not. The amount of medical food prescribed is intended to meet protein needs at various ages in the life cycle, which is shown in **Table 1**.

Introduction to nutrition management

The phenylalanine-restricted diet is introduced after the diagnosis is confirmed. Typically, a phenylalanine-free medical food is given for a few days so that blood phenylalanine will quickly decrease to an acceptable concentration. Then, a limited amount of breast milk or standard infant formula is added to the diet to provide 130–430 mg/d of phenylalanine (Singh et al., 2014), depending on the severity of the PAH mutation. This prescribed amount of intact or natural protein is needed to meet the infant's phenylalanine requirements and prevent phenylalanine deficiency, which leads to muscle protein catabolism and inadequate weight gain.

If parents prefer to formula-feed their infant, both standard infant formula and the PKU medical food are mixed in prescribed amounts and are bottle-fed. Breast milk as the protein source is encouraged. Mature breast milk contains approximately 46 mg per 100 cc of phenylalanine compared to approximately 57 mg per 100 cc in cow's milk protein-based formula and approximately 90 mg per 100 cc in soy-based formulas (USDA, 2022). Breast-fed infants, therefore, may initially have slightly lower blood phenylalanine concentrations. If the parents wish to continue breast-feeding, they are taught the proper ratio of breast milk to PKU medical food to feed the infant. The infant can feed the breast for some feedings and is given the PKU medical food from a bottle at others. The key to success with either feeding method is frequent monitoring of blood phenylalanine and adjusting the diet based on phenylalanine intake, weight gain, and blood phenylalanine. In the newborn period, weekly or twice weekly monitoring is recommended (Singh et al., 2014). The method used for monitoring varies depending on the resources available at individual PKU clinics—either frequent visits to the clinic for blood drawing or collecting filter paper specimens at home and mailing them to the lab for analysis. Because of the time delay in the latter method, it is more suitable for use after blood phenylalanine concentrations have stabilized.

When an infant with PKU is 4–6 months old, solid food is introduced. Because the total amount of phenylalanine tolerated remains relatively constant, as phenylalanine from solid food is added to the diet, the amount of regular infant formula or breast milk decreases. Since nearly all foods contain some phenylalanine, it must be counted with the goal of reaching the prescribed

Table 1 Recommended daily intakes of PHE, TYR and protein for individuals with PKU (Singh et al., 2014).

Age	*PHE (mg)*	*TYR (mg)*	*Protein (g/kg)*
0 to <3 months	130–430	1000–1300	2.5–3.0
3 to <6 months	135–400	1400–2100	2.0–3.0
9 to <12 months	145–370	2500–3000	2.0–2.5
7 to <9 months	135–330	2500–3000	2.0–2.5
1–4 years	200–320	2800–3500	1.5–2.1
>4 years	200–1100	4000–6000	120–140% DRI for age

PHE, phenylalanine; TYR, tyrosine.

amount of phenylalanine each day. Several methods of counting are used. Historically, all food was measured or weighed, and every milligram of phenylalanine was counted. Caregivers were given lists of the phenylalanine content of foods and the daily phenylalanine intake goal. Now, most clinics teach parents the "simplified diet" in which foods that are relatively low in phenylalanine (less than 75 mg phe/100 g of food) are not counted, and instead parents only count the foods that have higher phenylalanine content. Since the uncounted foods are fruits and some vegetables, the simplified diet is intended to encourage healthy eating and reduce parental stress associated with weighing measuring and recording (Hansen et al., 2020). While not counted by the caregiver, dietitian accounts for the amount of for the phenylalanine content of the uncounted foods, usually about 30% of the total phe intake (Hansen et al., 2020). Typically, pureed fruits and vegetables are offered first, and then as the infant's appetite increases, "counted" foods, such as vegetables with >75 mg/100 g, for example corn, peas and potatoes, as well as bread and cereal products. For most individuals with PKU, intake regular grain products is either not allowed or severely limited. Instead, specialty low-protein foods are available, often through mail order. A whole array of low-protein breads, cereals, crackers, bagels, pasta, cakes, cookies, and low-protein cheeses, "meats" and "peanut butter" are helpful to include in the diet. These foods provide much-needed variety and calories to the diet. High-protein foods such as meat, fish, poultry, dairy, nuts, eggs, and legumes are not allowed on a PKU diet. Thus, the medical food continues to be the main source of protein for life. A variety of medical foods are available for older children and adults with PKU in order to meet different tastes and energy needs. Many of the medical foods for children, teens, and adults are packaged in ready to drink containers for convenience and come in several flavors. Nevertheless, many individuals with PKU struggle with taking the full amount of the prescribed medical food, perhaps because off the taste of amino-acid based formulas, taste fatigue, or inconvenience. If the full amount of medical food is not taken, nutritional intake is inadequate and may lead to catabolism of lean body mass, which, in turn, leads to poor control of blood phenylalanine.

Many medical foods contain L-amino acids as the nitrogen source, but others contain glycomacropeptide (GMP), a derivative of whey protein that is naturally low in phenylalanine, although not phenylalanine-free. There is some evidence that GMP-based medical foods improve satiety and are better-tasting than amino acid-based medical foods (Ney et al., 2016).

Other medical foods contain primarily the large neutral amino acids (LNAAs), phenylalanine, tyrosine, tryptophan, and branched-chain amino acids which share the same L-amino acid transport system across the blood–brain barrier. High concentrations of phenylalanine in the blood, therefore, impede the transport of these other amino acids into the CNS. Tyrosine and tryptophan are important neurotransmitter precursors; relative deficiency or imbalance of which may contribute to the neuropsychiatric symptoms seen in some adult PKU patients who have resumed an unrestricted diet. Treatment with supplemental LNAAs theoretically will increase the competition with phenylalanine for transport into the CNS. A net reduction in phenylalanine and an increase in CNS tyrosine and tryptophan may result in improvement in symptoms. This treatment is not suitable for children or women in the childbearing years that might be contemplating pregnancy (Vockley et al., 2014).

Once established, the amount of dietary phenylalanine an individual is allowed remains the same, except for periods of rapid growth, when more phenylalanine may be necessary. A typical phenylalanine intake for a child with severe PKU is approximately 250 mg day^{-1} and that for a child with moderate PKU is 400 mg day^{-1}. Thus, in addition to getting the proper amount of medical food, the crux of the diet is to provide the prescribed amount of phenylalanine while making the diet taste and appear as appetizing and socially acceptable as possible. Families require a good deal of support in doing this. Parent/patient organizations such as the National PKU Alliance, family events, and camps for children with PKU provide a connection for families and a forum for exchange of practical information and emotional support. PKU clinic personnel are another source of support and reliable information on medical advances in treating PKU.

All patients with PKU should have blood phenylalanine and other amino acids monitored regularly as long as they remain on diet; they should also be seen on a regular basis for a physical examination, especially for assessment of growth parameters in children and adolescents, and review of the dietary intake since the previous visit. Extensive nutrition counseling is an ongoing process. It is also recommended that adult patients, who are not following phenylalanine-restricted diets with prescribed medical foods, should be seen at least yearly for nutritional assessment, as they often tend to self-limit their protein intake and may have inadequate diets (Singh et al., 2014).

Nutritional therapy is the mainstay of treatment for individuals living with PKU (van Wegberg et al., 2017) (Vockley et al., 2014). Although nutritional therapy has been shown to prevent the severe neurological manifestations of the disease (Jameson and Remmington, 2020; Acosta and Matalon, 2010), a profoundly modified nutritional intake creates risk to growth, development, and nutritional adequacy in the PKU population.

Restriction of whole protein sources can create deficiencies in essential fatty acids, vitamin B_6, vitamin B_{12}, selenium, zinc, calcium, and iron (Blau et al., 2010). The use of medical foods for supplementation of the required macro and micronutrients mitigates the risk of deficiency, however, adherence to the consumption of medical foods varies between age groups and decreases throughout adolescence (Strisciuglio and Concolino, 2014). Additionally, individuals that are non-adherent to the consumption of medical food tend to consume high-carbohydrate diets that lead to obesity, vitamin and mineral deficiencies, and secondary disease sequela (McBurnie et al., 1991).

Nutritional therapy, including nutritional assessments, nutritional interventions, and nutritional prescriptions, followed by monitoring and evaluation are essential for maintaining the health of individuals with PKU.

Growth

Adequate growth in individuals with PKU is both a historical and current consideration in the manifestation of the disease. Studies examining the growth of individuals with PKU have demonstrated concern for growth retardation, even in the presence of nutritional therapy providing adequate micronutrient and macronutrient needs (Dobbelaere et al., 2003). There is, however, evidence of improved linear growth in the presence of increased intakes above the Recommended Daily Allowance (RDI) (ranging from 113% to 129% RDI) for protein (Singh et al., 2014) suggesting that the protein needs of individuals with PKU may be higher than the standard population. Growth must be carefully monitored and evaluated at, a minimum, of every clinic visit (Singh et al., 2014).

Some evidence suggests the rate of overweight and obesity within the PKU population is higher than in the general population (Acosta et al., 2003; Walkowiak et al., 2019) however, data is inconsistent (Shakiba et al., 2020). An analysis from the Collaborative Study of Children Treated for Phenylketonuria (PKU) was compared to growth data from the National Center for Health Statistics (NCHS) to determine if growth patterns were affected in the PKU population. This investigation found that growth in children treated for PKU differs from national standards for weight by age and weight by height, but not for height by age, with poor diet adherence being a consideration for an increased risk of obesity (McBurnie et al., 1991). A positive correlation between phenylalanine with body weight and age has been indicated in other investigations (de Almeida et al., 2020) demonstrating the importance of metabolic control on overweight and obesity.

The first multi-center European and Turkish cross-sectional study of 947 patients with PKU found a higher risk for obesity and overweight in the female cohort than the male cohort (Ozel et al., 2014). An observational study of 30 pediatric (ages 5–18 years) with PKU, however, demonstrated no difference in lean body mass (LBM) or fat free mass (FM) percentage in female subjects compared to control, but did see less LBM and higher FM percentage in males compared to controls. Additionally, 30% of the subjects with PKU and 6.67% of the controls were overweight (BMI 85-95th percentile, $n = 9$ vs. $n = 2$); and 10% of the subjects with PKU and 13.33% of the controls were obese (Sailer et al., 2020). Additionally, male subjects with PKU demonstrated significantly lower height z-scores compared to controls, however, pubertal maturation and/or genetic potential were not included in the analysis (Sailer et al., 2020).

Further investigation into the risk of overweight and obesity in the PKU population is warranted, however, metabolic control of the disease and gender may play a role in the risk of overweight and obesity (Walkowiak et al., 2019). Weight must be carefully monitored and evaluated at, a minimum, of every clinic visit (Singh et al., 2014).

Fatty acids

Individuals with PKU following a protein restriction consume limited to no animal products, including eggs and dairy, or nuts. These products are primary sources of long-chain polyunsaturated fatty acids (LC-PUFAs). Main n-3 LC-PUFA in food sources are α-linolenic acid (ALA), docosahexaenoic acid (DHA), and eicosapentaenoic acid (EPA). Main n-6 LC-PUFA in food sources include linoleic acid (LA) and arachidonic acid (AA or ARA) (Abedi and Sahari, 2014). These LC-PUFAs are essential fatty acids that cannot be produced in adequate amounts endogenously and must be consumed from foods to prevent essential fatty acid deficiency. As natural sources of essential fatty acids are restricted in the nutritional treatment of PKU and there is evidence of changes in concentrations of plasma/serum lipoprotein lipids, as well as in the fatty acid profile of plasma and red blood cells in the PKU population (Guerra et al., 2020), essential fatty acid deficiency is of concern.

Evidence demonstrates aberrations in lipoprotein components in individuals with PKU, even with metabolic stability (Guerra et al., 2020). The majority of individuals receiving adequate amounts of the essential fatty acid precursors, linoleic acid and alpha linolenic acid, are able to synthesize LC-PUFAs via elongation and desaturation reactions. When LC-PUFAs are not supplemented in treated infants with PKU, however, plasma and erythrocyte levels contain lower concentrations of ARA and DHA than controls (Giovannini et al., 1995). Additionally, the essential fatty acid profiles of individuals with PKU vary from their peers and essential fatty acid concentrations in the blood appear to be dependent on the consumption of essential fatty acids in this population (van Gool et al., 2000). A systematic review of literature of 65 publications from 1990 to July 2016 demonstrated significantly lower levels of DHA, EPA and cholesterol in individuals with PKU compared to healthy controls (Parra et al., 2018). Supplementation with EFA precursors, or with preformed docosahexaenoic acid (DHA) and/or ARA, may be necessary in individuals with PKU at any age and may be conditionally essential nutrients in this population (Giovannini et al., 1995).

Several studies suggest improved clinical outcomes in individuals with PKU when supplemented with DHA and/or AHA. A study of 36 children supplemented with 15 mg DHA per kilogram body weight showed improvement in visual processing, motor function, and coordination though they had been receiving adequate alpha linolenic acid (Giovannini et al., 1995) and had adequate metabolic control, however, further investigation is needed on the effect of DHA supplementation on neurocognitive function. Additionally, supplementation with DHA has been shown to correct and prevent essential fatty acid deficiency (Agostoni et al., 2001) and has been shown to decrease visual evoked potential P100 wave latency in children (Couce Pico et al., 2019). Further research is necessary into the full benefits of supplementation of LC-PUFAs in the PKU population.

In infancy, including breast milk or standard formulas containing DHA and arachidonic acid, as well as medical foods that contain DHA/arachidonic acid, can help ensure adequacy (Singh et al., 2014). Medical foods, however, are not consistently supplemented and, therefore, may not provide adequate sources of essential fatty acids. Essential fatty acid supplementation should be

considered for individuals with PKU of all ages and yearly monitoring, or sooner if deficiency is suspected, of essential fatty acid profiles is recommended (Singh et al., 2014).

Vitamins and minerals for consideration

The standard of care for PKU is restriction of dietary protein. This restricted intake leads to deficient vitamins B_2, Niacin (B_3), B_{12}, D, iron, zinc, and selenium (Bakaloudi et al., 2021). B-group vitamins, niacin, vitamin D, iron, zinc, and selenium all play a critical role in inflammatory responses and may affect cognition (Godos et al., 2020), possibly by decreasing reactive intermediates in the brain.

Several studies within the PKU population reveal micronutrient deficiencies in vitamins B_2, Niacin (B_3), B_{12}, D, zinc, and selenium, among others. Current data suggests that supplementation with micronutrients, both individually and within medical foods, has improved the nutritional status in those living with PKU, however, supplementation from medical foods may be insufficient (Lammardo et al., 2013) and concentrations in the blood may remain low (Nazari et al., 2016). The mechanisms for utilization of micronutrients in the PKU population and the bioavailability of the micronutrients added to medical foods are not sufficiently researched (Nazari et al., 2016).

Individuals living with PKU may be at significantly higher risk of deficiency of vitamins B_2, Niacin (B_3), B_{12} due to the lack of vitamin B rich foods in the diet and other theorized mechanisms (Brantley et al., 2018). The risk of B vitamin deficiencies is greater in those individuals with PKU not under medical supervision (Schoen et al., 2019). Vitamin B supplementation is warranted in the PKU population when not obtaining adequate amounts of these vitamins from medical food (Kose and Arslan, 2019) or when serum concentrations are low. Careful monitoring of vitamin B status is recommended (Schoen et al., 2019).

A higher rate of vitamin D deficiency exists in the PKU population (Kose and Arslan, 2019), however, the risk of deficiency is lowered in those individuals consuming medical foods (Geiger et al., 2016). The rate of vitamin D deficiency in the PKU population is as high as 30–53.6% depending on age and access to sunlight. Zinc and iron status in individuals with phenylketonuria has long been researched due to potential deficiency in the protein restricted diet. Low concentrations of zinc and iron in individuals with PKU have been evidenced in multiple studies (Acosta et al., 2004), even despite intake above the recommended daily allowances (Evans et al., 2014). A systematic review of literature from 1990 to July 2016 exhibited lowered zinc concentrations in those living with PKU compared to controls (Parra et al., 2018). Although the meta-analysis did not show statistical significance, the observation of lowered zinc concentrations in the PKU cohort may be of clinical consideration (Parra et al., 2018). Selenium deficiency has been demonstrated in the PKU population (Evans et al., 2014), even in the presence of supplementation, however, the rate of deficiency is lower with selenium supplementation (Darling et al., 1992). Until recently, selenium was not routinely added to PKU formulas, and reliance on the selenium content of soil may not be sufficient.

Active monitoring of B vitamins, zinc, and selenium concentrations is recommended. Monitoring may be necessary with inadequate consumption or supplementation of these micronutrients in medical foods, clinical signs/symptoms of nutritional inadequacy including poor growth, or serious intercurrent illness (van Wegberg et al., 2017). Vitamin D and iron statuses should be assessed yearly (Singh et al., 2014). Monitoring the impact of supplementation or nutrition intervention after one month is recommended to ensure improvement (Singh et al., 2014).

Bone mineral density

Abnormal bone mineral density (BMD) has long been observed (Feinberg and Fisch, 1962) in the PKU population. An elevation in the prevalence of low BMD-for-age in PKU, defined as a Z-score ≤ -2.0, has been demonstrated in numerous studies and has been shown to affect approximately 20% of individuals with PKU compared to 2% in the general population (de Castro et al., 2020). Although the exact mechanisms for decreased BMD in this population are unknown, there are numerous factors that may be of influence. It has been theorized that calcium and vitamin D status, the quality of bone proteins, physical activity, endocrine status, and genetic and environmental factors impact BMD in individuals with PKU (van Wegberg et al., 2017). Gender and intake of medical foods may also play a role in the risk for decreased bone mineral density, although the data is confounding (Stroup et al., 2018). A systematic review of cross-sectional and cohort studies on bone status in phenylketonuria demonstrated that mean BMD is reduced in PKU patients compared with healthy controls, however, they did not observe an increased prevalence of bone mass below the expected range for age as defined by Z-scores of −2.0 SD (de Castro et al., 2020). The authors also observed lower significant differences in bone mineral content (BMC) between PKU patients and healthy controls, however, the concentration of minerals was heterogeneous (de Castro et al., 2020). Additionally, the authors found a trend toward an imbalance between bone formation and bone resorption, favoring bone removal in the PKU population (de Castro et al., 2020). An alternative systematic review and meta-analysis further demonstrated individual study findings of low BMD in the PKU population, however, available data suggested the reduction in BMD is not clinically relevant when compared to the standard definitions of low BMD (Demirdas et al., 2015). Additionally, they concluded that phenylalanine concentrations, vitamin D, PTH, and nutrient intake do not correlate with BMD or BTM (Demirdas et al., 2015).

Numerous methods of evaluation have been implemented in the PKU population to study bone mineral density. These methods have included dual-energy X-ray absorptiometry (DEXA), peripheral quantitative computed tomography (pQCT), and single-photon absorptiometry (SPA). The various methods used could be leading to conflicting findings within the PKU population on the risk factors for decreased BMD. A European survey study of seventeen centers specialized in treating adults with PKU was unable

to identify any risk factors for low BMD in adults living with PKU and found that low BMD occurred only in a small subset of PKU patients (Lubout et al., 2020).

Further investigation into the prevalence and mechanism of decreased BMD in this population is needed, however, the need for frequent monitoring of BMD in the PKU population is well understood. Measurement of BMD through dual-energy X-ray absorptiometry (DEXA or DXA) is indicated in PKU (Singh et al., 2014; van Wegberg et al., 2017) medical intervention to improve BMD in this population should include all considerations for the healthy population.

Maternal PKU

For women with PKU who intend to become pregnant, following a strict phenylalanine-restricted diet and controlling blood phenylalanine to 120–360 μmol/L is critical to offspring health. Women with PKU who have high blood phenylalanine levels are at high risk of having children with microcephaly, developmental delay, low birth weight and, congenital heart anomalies. In an International Study of Maternal PKU, women who were in good metabolic control by 10 weeks' gestation had babies with good birth outcomes and development. In women in poor control, the degree of microcephaly and mental retardation was proportional to the concentration of blood phenylalanine (Koch et al., 2003). Congenital heart disease (CHD), however, was not directly related to the degree of metabolic control, suggesting that etiology is multifactorial. In the international study, a significantly higher risk for CHD was found in the women with lower natural protein and medical food intake, regardless of blood Phe levels. In addition, these women had lower red blood cell folate concentrations (Matalon et al., 1991).

The recommendation is for women to maintain blood phenylalanine concentrations in the recommended range before conceiving in order to prevent damage to the fetus (Vockley et al., 2014). Nevertheless, many women come to medical attention during pregnancy, indicating the need for better strategies for keeping women on treatment for life, as well as determining the safety of pegvaliase and other emerging therapies during pregnancy.

Pharmacological therapies

Tetrahydrobiopterin

Tetrahydrobiopterin (BH4) is the cofactor for phenylalanine hydroxylase. A synthetic form of BH4, sapropterin dihydrochloride, has been used in conjunction with dietary therapy for over a decade and lowers blood phenylalanine in some individuals with PKU. Responsiveness to BH4 therapy depends on the severity of PKU; patients with mild or moderate degrees of PKU are more likely to respond and those with 2 null mutations do not. Data from a registry of 504 patients continuously treated with sapropterin showed an average 34% decrease in blood phenylalanine μmol/L ($p = 0.0009$) after 5 years. Dietary phenylalanine tolerance increased an average of 54% above baseline intake after 6 years, however most patients remain on some degree of protein restriction and medical food intake (Longo et al., 2015).

Pegvaliase

Pegvaliase, an injectable form of phenylalanine ammonia lyase, converts phenylalanine to ammonia and cinnamic acid and is an approved enzyme substitution therapy for adults with PKU who have blood phe >600 μmol/L. In clinical trials, patients treated with pegvaliase showed a sustained reduction in blood phenylalanine concentrations while phenylalanine intake increased, and medical food was discontinued. By 24 months of therapy, approximately 60% of patients achieved blood phenylalanine concentrations below 360 μmol/L (the recommended upper limit for blood phe by ACMG). Common adverse events include injection site reactions, arthralgia, hypersensitivity reactions, and headaches. Most patients on pegvaliase can ultimately consume an unrestricted diet, but because of the likelihood of side effects, the drug is titrated slowly, and it often takes 6 months or more before blood phenylalanine concentrations decrease. In addition, because of the possible risk of anaphylaxis, patients must carry auto injectable epinephrine (Thomas et al., 2018).

Conclusion

Phenylketonuria was the first inherited amino acid disorder be identified by newborn screening and for which nutrition management prevented the serious developmental delay associated with untreated PKU. The diet for PKU became a paradigm for other amino acid disorders and lead to the development of medical foods based on L-amino acids without the offending amino acids. While nutrition management has been a huge public health success in preventing a generation of individuals with PKU from developmental delay, the diet has considerable limitations, including the difficulty of maintaining blood phenylalanine in the recommended treatment range due to the severe limitation of protein-containing foods, the neurocognitive deficits that are associated with suboptimal control and the nutritional risks the diet imposes. Pharmacological therapies offer an adjunct to or a substitution for the phenylalanine-restricted diet.

References

Abedi, E., Sahari, M.A., 2014. Long-chain polyunsaturated fatty acid sources and evaluation of their nutritional and functional properties. Food Sci. Nutr. 2 (5), 443–463.

Acosta, P.B., Matalon, K.M., 2010. Nutrition management of patients with inherited disorders of aromatic amino acid metabolism. In: Nutrition Management of Patients With Inherited Metabolic Disorders. Jones and Bartlett Publishers, Boston, pp. 119–174.

Acosta, P.B., Yannicelli, S., Singh, R., Mofidi, S., Steiner, R., DeVincentis, E., Jurecki, E., Bernstein, L., Gleason, S., Chetty, M., 2003. Nutrient intakes and physical growth of children with phenylketonuria undergoing nutrition therapy. J. Am. Diet. Assoc. 103 (9), 1167–1173.

Acosta, P.B., Yannicelli, S., Singh, R.H., Elsas, L.J., Mofidi, S., Steiner, R.D., 2004. Iron status of children with phenylketonuria undergoing nutrition therapy assessed by transferrin receptors. Genet. Med. 6 (2), 96–101.

Agostoni, C., Scaglioni, S., Bonvissuto, M., Bruzzese, M.G., Giovannini, M., Riva, E., 2001. Biochemical effects of supplemented long-chain polyunsaturated fatty acids in hyperphenylalaninemia. Prostaglandins Leukot. Essent. Fatty Acids 64 (2), 111–115.

Bakaloudi, D.R., Halloran, A., Rippin, H.L., Oikonomidou, A.C., Dardavesis, T.I., Williams, J., Wickramasinghe, K., Breda, J., Chourdakis, M., 2021. Intake and adequacy of the vegan diet. A systematic review of the evidence. Clin. Nutr. 40 (5), 3503–3521.

Bilder, D.A., Noel, J.K., Baker, E.R., Irish, W., Chen, Y., Merilainen, M.J., Prasad, S., Winslow, B.J., 2016. Systematic review and meta-analysis of neuropsychiatric symptoms and executive functioning in adults with phenylketonuria. Dev. Neuropsychol. 41 (4), 245–260.

Blau, N., van Spronsen, F.J., Levy, H.L., 2010. Phenylketonuria. Lancet 376 (9750), 1417–1427.

Brantley, K.D., Douglas, T.D., Singh, R.H., 2018. One-year follow-up of B vitamin and iron status in patients with phenylketonuria provided tetrahydrobiopterin (BH4). Orphanet J. Rare Dis. 13 (1), 1–10.

Camp, K.M., Parisi, M.A., Acosta, P.B., Berry, G.T., Bilder, D.A., Blau, N., Bodamer, O.A., Brosco, J.P., Brown, C.S., Burlina, A.B., 2014. Phenylketonuria Scientific Review Conference: state of the science and future research needs. Mol. Genet. Metab. 112 (2), 87–122.

Couce Pico, M. de la L., Castro López, M. J. de, Lamas Pérez, C. de, Leis Trabazo, M.R., 2019. Effects of LC-PUFA supplementation in patients with phenylketonuria: a systematic review of controlled trials. Nutrients 11 (7).

Darling, G., Mathias, P., O'regan, M., Naughten, E., 1992. Serum selenium levels in individuals on PKU diets. J. Inherit. Metab. Dis. 15 (5), 769–773.

de Almeida, B.N. de F., Laufer, J.A., Mezzomo, T.R., Shimada, N.C., Furtado, I.H.F., Dias, M.R.M.G., Pereira, R.M., 2020. Nutritional and metabolic parameters of children and adolescents with phenylketonuria. Clin. Nutr. ESPEN 37, 44–49.

de Castro, M.J., de Lamas, C., Sánchez-Pintos, P., González-Lamuño, D., Couce, M.L., 2020. Bone status in patients with phenylketonuria: a systematic review. Nutrients 12 (7), 2154.

Demirdas, S., Coakley, K.E., Bisschop, P.H., Hollak, C.E.M., Bosch, A.M., Singh, R.H., 2015. Bone health in phenylketonuria: a systematic review and meta-analysis. Orphanet J. Rare Dis. 10 (1), 1–17.

Dobbelaere, D., Michaud, L., Debrabander, A., Vanderbecken, S., Gottrand, F., Turck, D., Farriaux, J.P., 2003. Evaluation of nutritional status and pathophysiology of growth retardation in patients with phenylketonuria. J. Inherit. Metab. Dis. 26 (1), 1–11.

Evans, S., Daly, A., MacDonald, J., Preece, M.A., Santra, S., Vijay, S., Chakrapani, A., MacDonald, A., 2014. The micronutrient status of patients with phenylketonuria on dietary treatment: an ongoing challenge. Ann. Nutr. Metab. 65 (1), 42–48.

Feinberg, S.B., Fisch, R.O., 1962. Roentgenologic findings in growing long bones in phenylketonuria: preliminary study. Radiology 78 (3), 394–398.

Geiger, K.E., Koeller, D.M., Harding, C.O., Huntington, K.L., Gillingham, M.B., 2016. Normal vitamin D levels and bone mineral density among children with inborn errors of metabolism consuming medical food–based diets. Nutr. Res. 36 (1), 101–108.

Giovannini, M., Biasucci, G., Agostoni, C., Luotti, D., Riva, E., 1995. Lipid status and fatty acid metabolism in phenylketonuria. J. Inherit. Metab. Dis. 18 (3), 265–272.

Godos, J., Currenti, W., Angelino, D., Mena, P., Castellano, S., Caraci, F., Galvano, F., del Rio, D., Ferri, R., Grosso, G., 2020. Diet and mental health: review of the recent updates on molecular mechanisms. Antioxidants 9 (4), 346.

Guerra, I.M.S., Ferreira, H.B., Neves, B., Melo, T., Diogo, L.M., Domingues, M.R., Moreira, A.S.P., 2020. Lipids and phenylketonuria: current evidences pointed the need for lipidomics studies. Arch. Biochem. Biophys. 688, 108431.

Hansen, J., Hollander, S., Drilias, N., van Calcar, S., Rohr, F., Bernstein, L., 2020. Simplified diet for nutrition management of phenylketonuria: a survey of US metabolic dietitians. JIMD Rep. 53 (1), 83–89.

Jameson, E., Remmington, T., 2020. Dietary interventions for phenylketonuria. Cochrane Database Syst. Rev. 7.

Koch, R., Hanley, W., Levy, H., Matalon, K., Matalon, R., Rouse, B., Trefz, F., Güttler, F., Azen, C., Platt, L., 2003. The maternal phenylketonuria international study: 1984–2002. Pediatrics 112 (Suppl. 4), 1523–1529.

Kose, E., Arslan, N., 2019. Vitamin/mineral and micronutrient status in patients with classical phenylketonuria. Clin. Nutr. 38 (1), 197–203.

Lammardo, A.M., Robert, M., Rocha, J.C., van Rijn, M., Ahring, K., Belanger-Quintana, A., MacDonald, A., Dokoupil, K., Ozel, H.G., Goyens, P., 2013. Main issues in micronutrient supplementation in phenylketonuria. Mol. Genet. Metab. 110, S1–S5.

Longo, N., Arnold, G.L., Pridjian, G., Enns, G.M., Ficicioglu, C., Parker, S., Cohen-Pfeffer, J.L., 2015. Long-term safety and efficacy of sapropterin: the PKUDOS registry experience. Mol. Genet. Metab. 114 (4), 557–563.

Lubout, C.M.A., Arrieta Blanco, F., Bartosiewicz, K., Feillet, F., Gizewska, M., Hollak, C., van der Lee, J.H., Maillot, F., Stepien, K.M., Wagenmakers, M.A.E.M., 2020. Bone mineral density is within normal range in most adult phenylketonuria patients. J. Inherit. Metab. Dis. 43 (2), 251–258.

Matalon, R., Michals, K., Azen, C., Friedman, E.G., Koch, R., Wenz, E., Levy, H., Rohr, F., Rouse, B., Castiglioni, L., 1991. Maternal PKU collaborative study: the effect of nutrient intake on pregnancy outcome. J. Inherit. Metab. Dis. 14 (3), 371–374.

McBurnie, M.A., Kronmal, R.A., Schuett, V.E., Koch, R., Azeng, C.G., 1991. Physical growth of children treated for phenylketonuria. Ann. Hum. Biol. 18 (4), 357–368.

Nazari, F., Ghadiri Anari, A., Mozaffari-Khosravi, H., 2016. Comparison of zinc status between patients with phenylketonuria undergoing therapy metabolic formula and healthy people. Tolooebehdasht 15 (3), 13–21.

Ney, D.M., Stroup, B.M., Clayton, M.K., Murali, S.G., Rice, G.M., Rohr, F., Levy, H.L., 2016. Glycomacropeptide for nutritional management of phenylketonuria: a randomized, controlled, crossover trial. Am. J. Clin. Nutr. 104 (2), 334–345.

Ozel, H.G., Ahring, K., Bélanger-Quintana, A., Dokoupil, K., Lammardo, A.M., Robert, M., Rocha, J.C., Almeida, M.F., van Rijn, M., MacDonald, A., 2014. Overweight and obesity in PKU: the results from 8 centres in Europe and Turkey. Mol. Genet. Metab. Rep. 1, 483–486.

Parra, G.A.M., Singh, R.H., Cetinyurek-Yavuz, A., Kuhn, M., MacDonald, A., 2018. Status of nutrients important in brain function in phenylketonuria: a systematic review and meta-analysis. Orphanet J. Rare Dis. 13 (1), 1–16.

Sailer, M., Elizondo, G., Martin, J., Harding, C.O., Gillingham, M.B., 2020. Nutrient intake, body composition, and blood phenylalanine control in children with phenylketonuria compared to healthy controls. Mol. Genet. Metab. Rep. 23, 100599.

Schoen, M., Coakley, K., Douglas, T., Singh, R., 2019. Functional biomarkers of vitamin B12 status, diet, and metabolic control in women with phenylketonuria (P24-005-19). Curr. Dev. Nutr. 3 (Suppl. 1) nzz044-P24.

Schuck, P.F., Malgarin, F., Cararo, J.H., Cardoso, F., Streck, E.L., Ferreira, G.C., 2015. Phenylketonuria pathophysiology: on the role of metabolic alterations. Aging Dis. 6 (5), 390.

Shakiba, M., Alaei, M., Saneifard, H., Mosallanejad, A., 2020. Assessment of anthropometric indices in patients with phenylketonuria. Iran. J. Child Neurol. 14 (2), 27.

Singh, R.H., Rohr, F., Frazier, D., Cunningham, A., Mofidi, S., Ogata, B., Splett, P.L., Moseley, K., Huntington, K., Acosta, P.B., 2014. Recommendations for the nutrition management of phenylalanine hydroxylase deficiency. Genet. Med. 16 (2), 121–131.

Strisciuglio, P., Concolino, D., 2014. New strategies for the treatment of phenylketonuria (PKU). Metabolites 4 (4), 1007–1017.
Stroup, B.M., Hansen, K.E., Krueger, D., Binkley, N., Ney, D.M., 2018. Sex differences in body composition and bone mineral density in phenylketonuria: a cross-sectional study. Mol. Genet. Metab. Rep. 15, 30–35.
Thomas, J., Levy, H., Amato, S., Vockley, J., Zori, R., Dimmock, D., Harding, C.O., Bilder, D.A., Weng, H.H., Olbertz, J., 2018. Pegvaliase for the treatment of phenylketonuria: results of a long-term phase 3 clinical trial program (PRISM). Mol. Genet. Metab. 124 (1), 27–38.
USDA, 2022. USDA FoodData Central. https://fdc.nal.usda.gov/.
van Gool, C.J.A.W., van Houwelingen, A.C., Hornstra, G., 2000. The essential fatty acid status in phenylketonuria patients under treatment. J. Nutr. Biochem. 11 (11–12), 543–547.
van Wegberg, A.M.J., MacDonald, A., Ahring, K., Bélanger-Quintana, A., Blau, N., Bosch, A.M., Burlina, A., Campistol, J., Feillet, F., Giżewska, M., 2017. The complete European guidelines on phenylketonuria: diagnosis and treatment. Orphanet J. Rare Dis. 12 (1), 1–56.
Vockley, J., Andersson, H.C., Antshel, K.M., Braverman, N.E., Burton, B.K., Frazier, D.M., Mitchell, J., Smith, W.E., Thompson, B.H., Berry, S.A., 2014. Phenylalanine hydroxylase deficiency: diagnosis and management guideline. Genet. Med. 16 (2), 188–200.
Walkowiak, D., Kaluzny, L., Bukowska-Posadzy, A., Oltarzewski, M., Staszewski, R., Moczko, J.A., Musielak, M., Walkowiak, J., 2019. Overweight in classical phenylketonuria children: a retrospective cohort study. Adv. Med. Sci. 64 (2), 409–414.

Ketosis

D.H. Williamson[†], Radcliffe Infirmary, Oxford, UK

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of D.H. Williamson, Ketosis, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 47–53, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00166-5.

Introduction 432
Formation of Ketone Bodies 432
Extrahepatic Regulation 433
Intrahepatic Regulation 433
Role of Malonyl-CoA 434
Intramitochondrial Regulation 435
Function of Ketone Bodies 436
Pathways of Ketone Body Utilization 436
Mitochondrial Pathway 436
Cytosolic Pathway 437
Ketosis 437
Physiological Ketosis 437
Pathological Ketosis 438
Metabolic Acidosis 439
Further Reading 440

Introduction

The two ketone bodies acetoacetate ($CH_3COCH_2COO^-$) and D-3-hydroxybutyrate ($CH_3CHOHCH_2COO^-$) are the only freely soluble lipids in the circulation.

The name ketone bodies originates from the German *Ketonkörper* (literally, ketones excreted from the body) and refers to their discovery in the urine of diabetic patients in the latter half of the nineteenth century. In reality, the term is a misnomer because 3-hydroxybutyrate is not a ketone. It arose because the reagent originally used reacted positively with ketones in diabetic urine. Acetone (CH_3COCH_3), the product of the spontaneous decarboxylation of acetoacetate, is also a ketone and is present in blood and urine when the plasma concentration of acetoacetate is elevated. It is excreted via the kidneys and lungs and is responsible for the sweet smell on the breath in ketotic states.

The association of ketone bodies with the pathology of diabetes resulted in the view that they were toxic waste products. It is only in the past 30 years that this view has been convincingly reversed. Two factors led to this change, namely the development of an enzymatic method for the determination of acetoacetate and 3-hydroxybutyrate, which in turn allowed the dramatic finding of Cahill and colleagues in 1967 that adult human brain removed appreciable amounts of ketone bodies from the circulation in prolonged starvation.

The aim in this contribution is to review (a) the formation of ketone bodies in physiological and pathological situations and (b) the function of ketone bodies as physiological substrates and signals.

Formation of Ketone Bodies

It is well established that in humans and other mammals the only organ that contributes significant amounts of ketone bodies to the blood is the liver; this organ, unlike peripheral tissues, is unable to utilize ketone bodies to any appreciable extent. More recently it has been found that during the suckling period (high-fat diet) the intestine also has the capacity (approximately 10% of that of the liver) to produce ketone bodies. Whether ketone bodies are used *in situ* or are transported via the portal blood to supplement the existing hyperketonemia is an open question.

The main blood-borne substrates for the synthesis of ketone bodies (ketogenesis) are the nonesterified fatty acids; others of lesser importance are the branched-chain amino acids, leucine and isoleucine. In addition, acetate (sources: intestinal fermentation, in vinegar, or an oxidation product of ethanol) is a ketogenic substrate.

Long-chain fatty acids contained in dietary lipids do not enter the portal blood directly but are esterified in the intestinal cells, packaged with proteins and phospholipids to form chylomicrons (large lipoproteins), and transported via the lymphatic system to

[†]Deceased.

the thoracic duct, where they enter the blood. In contrast, the short- and medium-chain fatty acids (below C_{14}) contained in dairy products or in clinical medium-chain triacylglycerol preparations are directly absorbed as the respective fatty acids and are transported to the liver via the portal blood (Figure 1). The long-chain fatty acids in the plasma are bound to albumin and are released from adipose tissue triacylglycerol stores by the process of lipolysis.

Extrahepatic Regulation

A key factor in the regulation of ketogenesis is the availability of nonesterified long-chain fatty acids to the liver, which in turn is controlled by their release from adipose tissue. The enzyme responsible for the initiation of the hydrolysis of stored triacylglycerols to fatty acids is hormone-sensitive lipase. As its name implies, this enzyme is exquisitely sensitive to hormones: adrenaline (in the plasma) and noradrenaline (released from sympathetic nerve endings) are activators, whereas insulin inhibits the activity. In small mammals glucagon is also an activator of the enzyme, but this does not seem to be the case in the human.

Insulin has an additional effect on the net release of long-chain fatty acids from adipose tissue in that it stimulates their re-esterification to triacylglycerols. Thus after a high-carbohydrate meal, when insulin secretion and its concentration in the plasma is high, the release of fatty acids from adipose tissue is suppressed and their concentration in the plasma is low (Figure 2). In contrast, during stress, when adrenaline and noradrenaline are elevated, the release of fatty acids is increased and their plasma concentration is high.

In experimental animals increased plasma ketone body concentrations (hyperketonemia) can inhibit adipose tissue lipolysis (a) indirectly by increasing the secretion of insulin or (b) by a direct effect on the tissue (Figure 3). This can be viewed as a feedback mechanism for controlling the rate of ketogenesis via fatty acid supply to the liver, but whether this is important in the human is not known. In contrast, the supply of short- and medium-chain fatty acids to the liver is mainly dependent on the dietary intake and on the proportion that escapes further metabolism in the intestinal tract; there is no known involvement of hormones in the process.

Intrahepatic Regulation

There are situations (e.g., stress) where the supply of fatty acids to the liver may be increased, but there is no necessity to increase the availability of ketone bodies to the peripheral tissues. Consequently, there is a requirement that the rate of hepatic ketogenesis should be controlled independently of the supply of fatty acids. However, it must be stressed that without an increase in the supply of fatty acids the rate of ketogenesis cannot increase.

Much of the current interest is concerned with how the intrahepatic metabolism of fatty acids (Figure 4) is regulated. Long-chain fatty acids entering the liver have three main fates:

1. They can be re-esterified to phospholipids and triacylglycerols and then be secreted as very-low-density lipoproteins (VLDL).
2. They can be oxidized via the mitochondrial β-oxidation complex to acetyl-CoA. The latter can combine with another molecule of acetyl-CoA in the reaction catalyzed by acetoacetyl-CoA thiolase and then enter the hydroxymethylglutaryl-CoA pathway to form acetoacetate.
3. The acetyl-CoA derived from the fatty acids can be completely oxidized in the tricarboxylate cycle.

The short- and medium-chain fatty acids cannot be re-esterified to any appreciable extent in mammalian liver and therefore they are either metabolized to ketone bodies or completely oxidized. In addition, unlike the long-chain fatty acids, they are transported directly into the mitochondrial matrix without the need to be converted first to the corresponding acyl-CoA derivatives.

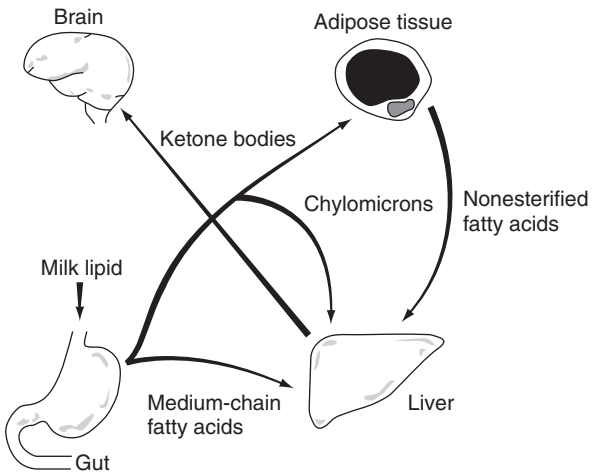


Figure 1 Intertissue fluxes of substrates in the suckling neonate. Thickness of line denotes rate of flux.

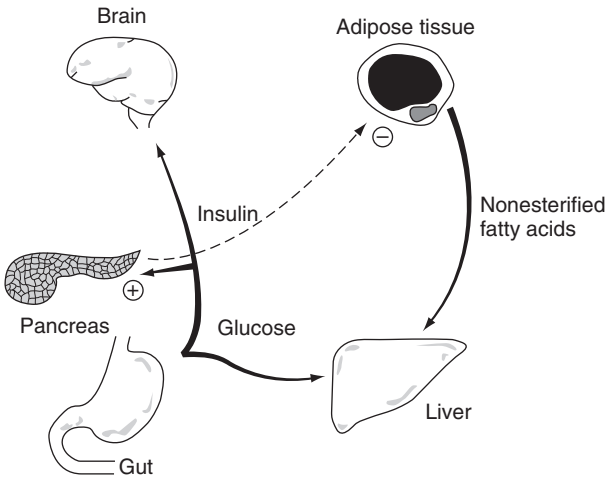


Figure 2 Intertissue fluxes of substrates in the fed state. Thickness of line denotes rate of flux.

Role of Malonyl-CoA

The entry of free long-chain fatty acids into the hepatocyte is via a specific carrier on the plasma membrane. Once inside the cytosol the long-chain fatty acids are bound to binding proteins, converted to the acyl-CoA derivatives, and then can either be esterified or enter the mitochondria via a complex transport system, the carnitine–acyl-CoA transferase (CAT) system. This consists of two proteins: CAT I located on the outer mitochondrial membrane and CAT II on the inner mitochondrial membrane (Figure 5). The overall action of the two enzymes results in the transfer of a long-chain fatty acyl-CoA to the mitochondrial matrix and the return of free carnitine to the cytosol via an exchange mechanism. Although carnitine is not consumed in the reaction, the available concentration can be critical. In nutritional carnitine deficiency there is impairment of long-chain fatty acid oxidation and ketogenesis.

The activity of CAT I is the key to the intrahepatic regulation of fatty acid metabolism in most situations. Its activity increases in ketogenic situations. More importantly, CAT I is inhibited by malonyl-CoA and the sensitivity of CAT I to this inhibitor changes in various pathophysiological situations such as fasting or diabetes.

As malonyl-CoA is a key intermediate in the synthesis of fatty acids (lipogenesis) from products (pyruvate and lactate) of glucose metabolism, this interaction provides a regulatory link between lipid and carbohydrate metabolism (Figure 5). Thus on high-carbohydrate diets, when the rate of hepatic lipogenesis, and consequently the cytosolic concentration of malonyl-CoA, is high, the activity of CAT I will be inhibited and fatty acids will be diverted to esterified products and secretion as VLDL rather than oxidation and conversion to ketone bodies. Conversely, on high-fat diets or in starvation, when lipogenesis is inhibited, malonyl-CoA concentration is low and CAT I is active. The sensitivity of CAT I to malonyl-CoA generally correlates with the prevailing concentration of the latter.

The short- and medium-chain fatty acids do not utilize the CAT I and II system to enter the mitochondrial matrix and therefore their oxidation is not greatly influenced by the prevailing 'carbohydrate status' (amount of glycogen, direction of carbohydrate flux, glycolysis, or gluconeogenesis) of the liver (Figure 5).

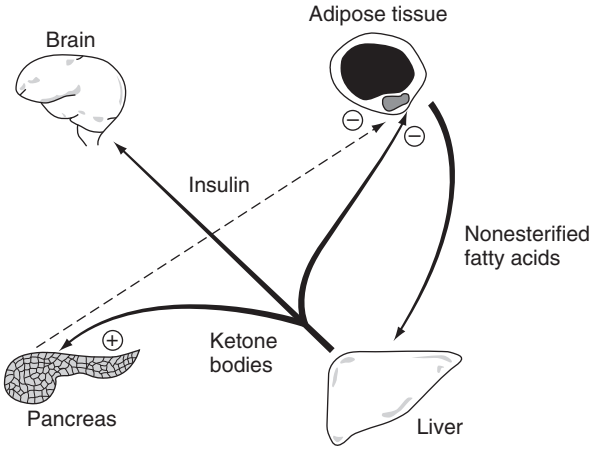


Figure 3 Role of ketone bodies as feedback regulators.

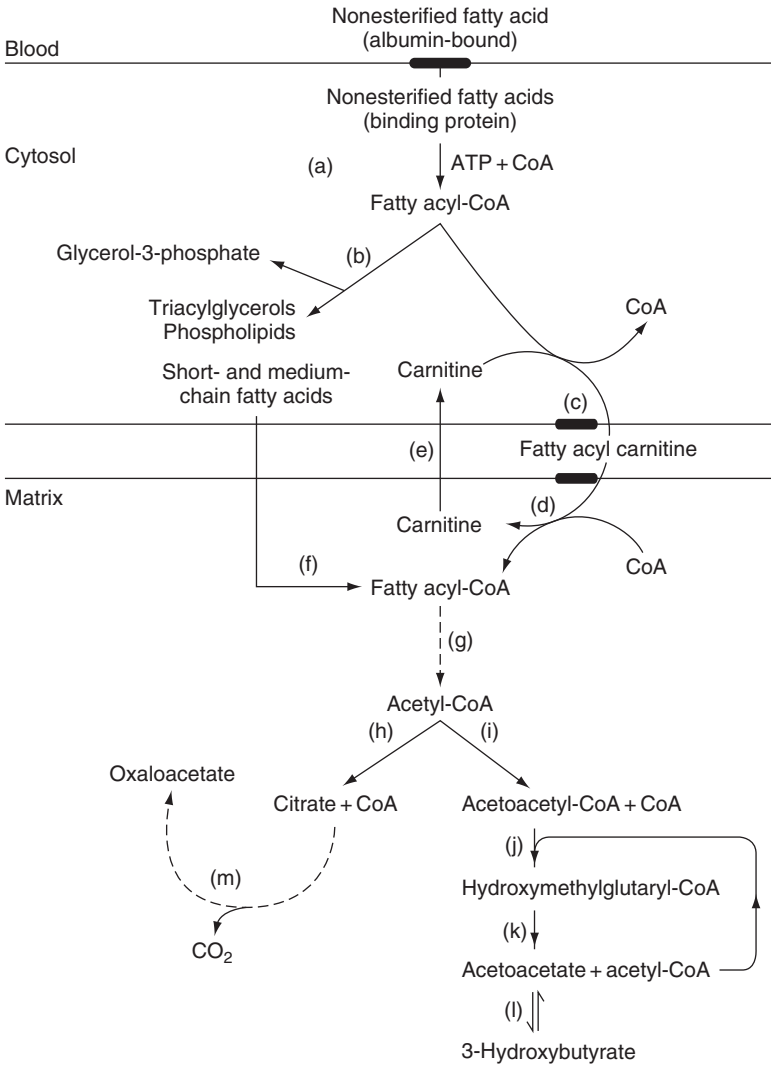


Figure 4 Pathway of fatty acid catabolism in liver. Enzymes involved: (a) long-chain fatty acyl-CoA synthetase; (b) glycerol-3-phosphate acyl-CoA transferase; (c) CAT I; (d) CAT II; (e) carnitine exchange; (f) short- and medium-chain fatty acyl-CoA synthetase; (g) fatty acid oxidation complex; (h) citrate synthase; (i) acetoacetyl-CoA thiolase; (j) hydroxymethylglutaryl-CoA synthase; (k) hydroxymethylglutaryl-CoA lyase; (l) hydroxybutyrate dehydrogenase; and (m) tricarboxylate cycle.

Insulin can rapidly depress the rate of ketogenesis *in vitro*. This effect is thought to result mainly from its stimulatory action on a key enzyme of lipogenesis, acetyl-CoA carboxylase, which in turn increases the concentration of malonyl-CoA. Glucagon and catecholamines have the opposite effect. Thus hormonal effects can be exerted at both the extrahepatic (lipolysis) and intrahepatic (modulation of lipogenesis) levels.

Intramitochondrial Regulation

Once the fatty acyl-CoA molecule is attached to the mitochondrial β-oxidation complex there appears to be little regulation exerted until release of the acetyl-CoA fragments. As indicated above, the acetyl-CoA can enter the tricarboxylate cycle and be oxidized to CO_2 or can be converted to ketone bodies via the hydroxymethylglutaryl-CoA pathway.

It appears that in most experimental situations the complete oxidation of fatty acids proceeds at a low, but relatively similar, rate and it is the activity of the hydroxymethylglutaryl-CoA pathway that shows larger changes. This has led to the view that the pathway might be regulated by mechanisms other than substrate supply.

Studies on the expression of 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase have shown that both the mRNA coding for the protein and the amount of protein increase during the onset of ketogenic states (fasting, diabetes) and that these changes are rapidly reversed (refeeding, insulin treatment). However, the finding that rates of ketogenesis from medium-chain fatty acids (CAT I and II) do not alter greatly with change in physiological state, if the rate of fatty acid supply is held constant, would seem to rule out appreciable regulation within the hydroxymethylglutaryl-CoA pathway. Indeed, current thinking suggests that the activity of CAT I is

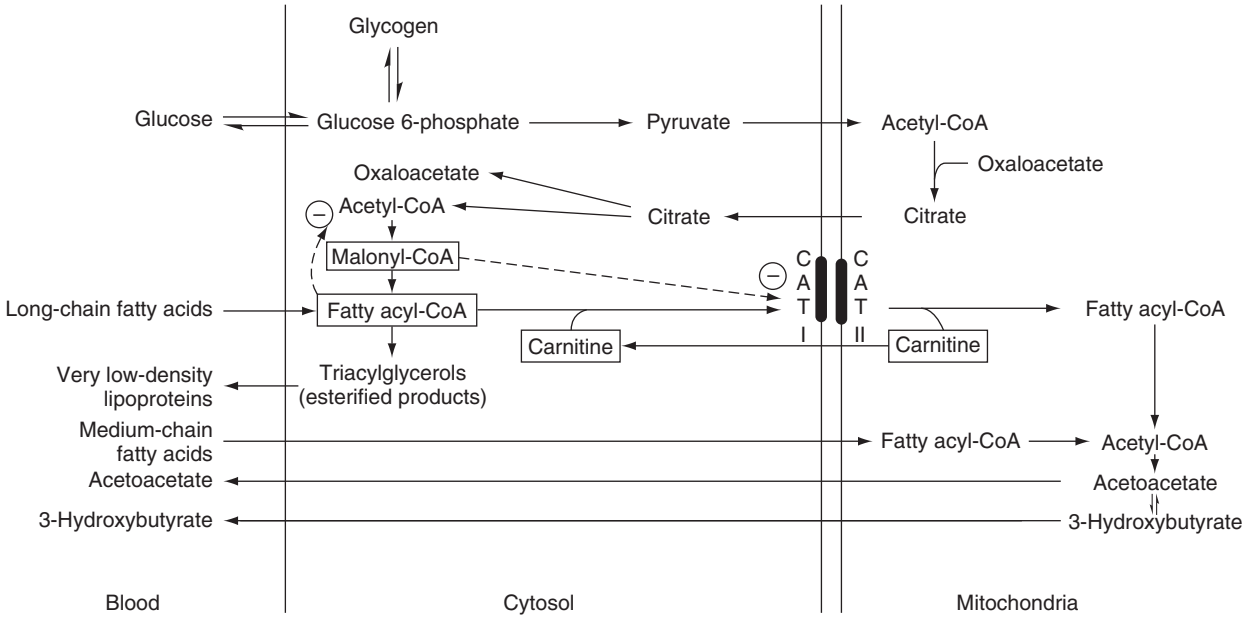


Figure 5 Inter-relationship between hepatic carbohydrate metabolism, lipogenesis, and ketogenesis. Circled minus signs indicate inhibition by the metabolite.

the primary intrahepatic site for the regulation of fatty acid oxidation and ketogenesis. If there is another important site, particularly during situations associated with the reversal of ketogenesis, it is likely to be proximal to the step catalyzed by this protein (e.g., the supply of fatty acids to the liver). Thus *in vivo* there is little doubt that the primary step that controls ketogenic flux is the rate of long-chain fatty acid release from adipose tissue.

Function of Ketone Bodies

The major role of ketone bodies is to supply an alternative oxidizable substrate to glucose for the brain in situations where the availability of the latter is impaired (e.g., starvation). In addition, ketone bodies can act as precursors for the acetyl-CoA required in neural lipid synthesis (myelin). Other mammalian tissues, including heart, skeletal muscle, kidney, and lactating mammary gland, can utilize ketone bodies but, in contrast to glucose utilization, no energy can be obtained in the absence of oxygen. In these tissues metabolism of ketone bodies results in the inhibition of glucose utilization and inhibition of the oxidation of pyruvate. The net result is a sparing of carbohydrate for the brain and the strictly glycolytic tissues (erythrocytes, retina).

Pathways of Ketone Body Utilization

Mitochondrial Pathway

The major site of ketone body utilization in peripheral tissues is the mitochondria (Figure 6). Although transporters for ketone bodies have been described on the plasma and inner mitochondrial membranes of some tissues, these do not appear to limit the flux. The initiating enzyme for acetoacetate metabolism is 3-oxoacid-CoA transferase:

$$\text{acetoacetate} + \text{succinyl} - \text{CoA} \rightleftharpoons \text{acetoacetyl} - \text{CoA} + \text{succinate}$$

The resulting acetoacetyl-CoA is cleaved to two molecules of acetyl-CoA by acetoacetyl-CoA thiolase; they are then oxidized in the tricarboxylate cycle.

3-Hydroxybutyrate is converted to acetoacetate by 3-hydroxybutyrate dehydrogenase:

$$3 - \text{hydroxybutyrate} + \text{NAD}^+ \rightleftharpoons \text{acetoacetate} + \text{NADH} + \text{H}^+$$

The ready reversibility of the three enzymes of the mitochondrial pathway (Figure 6) means that if the overall system is near equilibrium within the cell *in vivo*, the utilization of the ketone bodies will be dependent on their respective concentrations and on the rate of removal of the products. Thus acetoacetate utilization will be promoted when mitochondrial acetyl-CoA is decreased, whereas an increase in the latter will have the opposite effect. Similarly, oxidation of hydroxybutyrate will increase if the concentrations of $NADH_2$ and acetoacetate fall. Unlike the hepatic hydroxymethylglutaryl-CoA pathway for ketogenesis, which is essentially irreversible, the free reversibility of this pathway in peripheral tissues can be viewed as means of buffering the

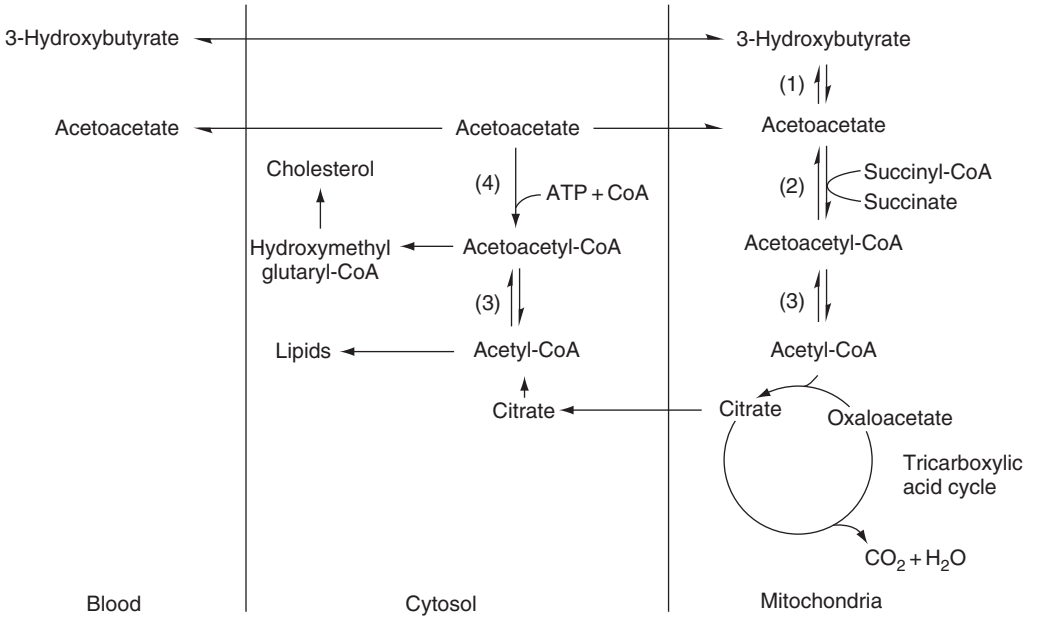


Figure 6 Pathways of ketone body utilization in peripheral tissues: (1) hydroxybutyrate dehydrogenase; (2) 3-oxoacid-CoA transferase; (3) acetoacetyl-CoA thiolase; and (4) acetoacetyl-CoA synthetase.

mitochondrial acetyl-CoA pool and hence energy production. Some of the acetyl-CoA can be transported to the cytosol in the form of citrate to act as a precursor for lipogenesis (Figure 6).

Cytosolic Pathway

The cytosol of tissues where active lipogenesis occurs (adipose tissue, developing brain, lactating mammary gland, and liver) contains an enzyme, acetoacetyl-CoA synthetase, which converts acetoacetate to acetoacetyl-CoA (Figure 6):

$$\text{acetoacetate} + \text{ATP} + \text{CoA} \rightarrow \text{acetoacetyl} - \text{CoA} + \text{AMP} + \text{pyrophosphate}$$

Its activity is at least an order of magnitude lower than that of the mitochondrial 3-oxoacid-CoA transferase, whereas its affinity for acetoacetate is appreciably higher. The presence of acetoacetyl-CoA thiolase in the cytosol allows the conversion of acetoacetate to acetyl-CoA and then to lipids without the involvement of the mitochondria.

Brain cytosol also contains 3-hydroxy-3-methylglutaryl-CoA synthase, allowing acetoacetate to act as a direct precursor for sterol synthesis. Evidence from *in vivo* experiments with ^{14}C-labeled acetoacetate has confirmed the existence of this pathway in developing brain and liver. The cytosolic route for acetoacetate utilization can be seen as a mechanism for directing this substrate to lipid or sterol synthesis rather than to oxidation.

Ketosis

The concentration of ketone bodies in the blood at any time represents a balance between the rate of hepatic ketogenesis and the rate of utilization by peripheral tissues. It is generally assumed that an increase in ketogenesis leads to a rise in blood ketone bodies, which in turn results in their increased utilization. In rare situations, such as congenital absence of key enzymes involved in ketone body utilization (e.g., 3-oxoacid-CoA transferase) or inhibition of these enzymes by pharmacological agents, blood ketone bodies may increase without any concomitant increase in ketogenesis.

The concentration of ketone bodies in the blood is exquisitely sensitive to changes in pathophysiological state. It is therefore useful to define *normoketonemia* in mammals as a concentration of total ketone bodies in blood below 0.2 mmol l^{-1}, *hyperketonemia* as above this level, and *ketoacidosis* (ketosis; by analogy to the definition of lactic acidosis) as above 7 mmol l^{-1}. In adult mammals there are small but characteristic diurnal variations in ketone body concentrations. Larger increases in concentration occur in humans in response to change in pathophysiological state (Table 1). The concentrations span a 200-fold range and it is this which underlines the important role of ketone bodies as substrates and signals.

Physiological Ketosis

Physiological hyperketonemia is found in the suckling neonate (high-fat diet of the milk; Figure 1), postexercise (depletion of hepatic glycogen reserves), and after prolonged fasting (more than 24 h; Figure 7). All these situations have in common a low

Table 1 Range of blood ketone body concentrations in humans

Situation	*Ketone body concentration (mmol l^{-1})*
Fed normal diet	About 0.1
Fed high-fat diet	Up to 3
Fasted: 12–24 h	Up to 0.3
Fasted: 48–72 h	2.0–3.0
Postexercise	Up to 2
Late pregnancy	Up to 1
Late pregnancy: fasted 48 h	4.0–6.0
Neonate: 0–1 day	0.2–0.5
Neonate: 5–10 days	0.7–1.0
Hypoglycemia	1.0–5.0
Untreated diabetes mellitus	Up to 25

hepatic carbohydrate status (depletion of glycogen or activation of gluconeogenesis) and therefore from a physiological standpoint one would expect an increased rate of ketogenesis. Comparison of the factors that can influence ketogenesis in suckling and fasting (Table 2) shows the expected broad agreement.

More detailed information on the hierarchy of the regulatory factors during onset and reversal of ketogenesis has been obtained for the fasting state by measurements at short time intervals. The first event after withdrawal of food is a lowering of plasma insulin accompanied by an increase in plasma fatty acids (stimulation of lipolysis). However, for an appreciable period (8–10 h) there is no increase in blood ketone bodies or in the *in vitro* rates of hepatic ketogenesis (measured with saturating fatty acid concentrations). The major increment in ketogenic rate occurs at the nadir of the hepatic malonyl-CoA concentrations and when the sensitivity of CAT I to malonyl-CoA starts to increase rapidly. This long time lag before a change in sensitivity of the protein to malonyl-CoA inhibition is thought to be due to the time required to bring about alterations to the lipid environment of the outer mitochondrial membrane.

Confirmation of this view is that on refeeding, when insulin rapidly increases and plasma fatty acids decrease with a parallel decrease in blood ketone bodies, there is again a time lag before malonyl-CoA concentrations rise and a longer one before sensitivity returns. In physiological and nutritional terms this delay of return to the normal feeding settings of intrahepatic regulation makes excellent sense. It is only when the refeeding consists primarily of large amounts of carbohydrate that the starved liver needs to inhibit the activity of CAT I to prevent the oxidation of newly synthesized fatty acids. If the meal consists mainly of lipid with little carbohydrate the activity of CAT I needs to remain high to allow oxidation of the excess fatty acids. Thus the liver must sense a prolonged increase in plasma insulin before the high activity of CAT I is suppressed.

Pathological Ketosis

The major example of pathological ketosis is of course insulin-dependent or type 1 diabetes. Essentially the changes in this condition are similar to those that occur during fasting, but they are more pronounced. Insulin is absent or very low in the plasma and therefore there is no antagonistic action to restrain the opposing hormones, adrenaline, noradrenaline, and glucagon. Consequently, lipolysis in adipose tissue is greatly stimulated and plasma fatty acids increase to high levels.

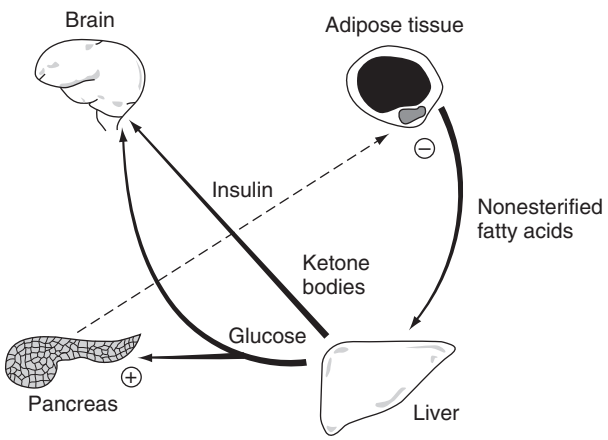


Figure 7 Intertissue fluxes of substrates in the starved state. Thickness of line denotes rate of flux.

Table 2 Comparison of factors influencing ketogenesis in suckling and fasted states

Factor	*Suckling*	*Fasted*
Plasma nonesterified fatty acids	Increased	Increased
Plasma insulin	Decreased	Decreased
Plasma glucagon	Increased	Increased
Hepatic carnitine	Increased	Increased
Hepatic lipogenesis	Decreased	Decreased
Hepatic malonyl-CoA	Decreased	Decreased
Hepatic CAT I activity	Increased	Increased
Sensitivity to malonyl-CoA	Decreased	Decreased

The lack of insulin and the large flux of fatty acids to the liver means that lipognesis is inhibited at the level of acetyl-CoA carboxylase and there is the expected decrease in malonyl-CoA concentration. In addition, the sensitivity of CAT I to inhibition by malonyl-CoA is considerably decreased. The level of expression of hepatic CAT I and II proteins also increases several-fold in diabetes. Thus the liver is in the ideal mode for producing excessive amounts of ketone bodies.

It has been suggested that diversion of oxaloacetate to hepatic glucose synthesis (which is also increased in insulin deficiency) may also play a role in the increased rate of ketogenesis by diverting acetyl-CoA from the tricarboxylate cycle. However, the present evidence suggests that this makes a minor contribution. Although the excessive output of ketone bodies by the liver undoubtedly makes the major contribution to their high levels in the blood, it is likely that there is also a degree of underutilization by peripheral tissues. The net result is ketoacidosis and excretion of large amounts of energy as ketone bodies in the urine.

A rare, but intriguing, example of pathological ketosis (ketone bodies up to 10 mmol l^{-1}) is the inborn error of hepatic glycogen synthase deficiency (Figure 8). Here glycogen is virtually absent from the liver so that after short-term fasting (5–10 h) the glucose falls to hypoglycemic levels, plasma insulin is decreased, plasma fatty acids increase, and ketogenesis is switched on. On consuming a meal the pattern is reversed until the blood glucose falls again. This case illustrates the importance of hepatic glycogen (and its mobilization) in the smooth transition of substrate supply from the fed to the fasted state. Treatment in this case was to recommend the consumption of more frequent high-carbohydrate snacks. It is of interest that this particular child suffered no ill effects from the daily exposure to high concentrations of ketone bodies, underlining their role as normal substrates for the brain when available.

Metabolic Acidosis

The great disadvantage of ketone bodies is that both acetoacetate and hydroxybutyrate are relatively strong acids. When they increase to high concentration there is the expected decrease in the blood pH, the plasma hydrogen carbonate concentration, and the partial pressure of carbon dioxide in blood and body fluids. The symptoms of acidosis include malaise, weakness, anorexia, and vomiting and these may eventually lead to coma. Treatment of diabetic ketoacidosis is to give insulin as soon as possible, usually as a continuous intravenous infusion. This rapidly decreases the raised plasma fatty acids and more slowly lowers the blood glucose and ketone bodies. Prolonged starvation, where the blood ketone bodies may reach 8–10 mmol l^{-1}, does not usually cause a serious disturbance of the acid–base balance. Loss of ketone bodies via the urine occurs but is not excessive. The nonenzymic decarboxylation of acetoacetate to acetone and carbon dioxide can be seen as a primitive mechanism for removing the potential acidotic effects of ketone bodies. The fact that acetone can be converted to glucose by the liver at low rates is an extra bonus.

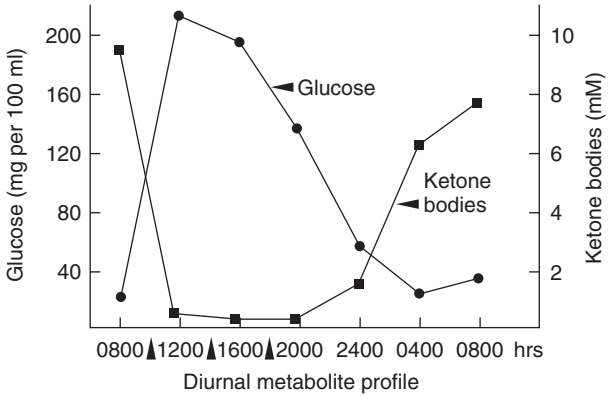


Figure 8 Diurnal blood metabolite profile of a child with glycogen synthetase deficiency.

The other common form of metabolic acidosis is lactic acidosis. This can arise because of infection, tissue hypoxia (anaerobic glycolysis), can be drug induced (ethanol, hypoglycemic biguanides), or can arise because of a congenital defect (pyruvate dehydrogenase or pyruvate carboxylase deficiency). In addition to the acidosis caused by lactic acid or ketone bodies there is a group of organic acidurias (some 25–30 different types) in which an inborn error results in the accumulation of an organic acid in the blood and urine. However, frank acidosis is not always associated with these conditions. The key investigation is chromatographic identification of the organic acid.

Further Reading

Bach, A.C., Ingenbleek, Y., Frey, A., 1996. The usefulness of dietary medium-chain triglycerides in body weight control: Fact or fancy? Journal of Lipid Research 37, 708–726.
Girard, J.R., Ferré, P., Pégorier, J.P., Duée, P.H., 1992. Adaptations of glucose and fatty acid metabolism during perinatal period and suckling–weanling transition. Physiological Reviews 72, 507–562.
Krebs, H.A., Woods, H.F., Alberti, K.G.M.M., 1975. Hyperlactataemia and lactic acidosis. Essays in Medical Biochemistry 1, 81–103.
McGarry, J.D., Foster, D.W., 1980. Regulation of hepatic fatty acid oxidation and ketone body production. Annual Review of Biochemistry 49, 395–420.
Nehlig, A., de Vasconcelos, A.P., 1993. Glucose and ketone body utilization by the brain of neonatal rats. Progress in Neurobiology 40, 163–221.
Owen, O.E., Morgan, A.P., Kemp, H.G., et al., 1967. Brain metabolism during fasting. Journal of Clinical Investigation 46, 1589–1595.
Page, M.A., Williamson, D.H., 1971. Enzymes of ketone body utilisation in human brain. Lancet 2, 66–68.
Porter, R., Lawrenson, G. (Eds.), 1982. Metabolic acidosis. Proceedings of the Ciba Foundation Symposium. Pitman, London, 87.
Robinson, A.M., Williamson, D.H., 1980. Physiological roles of ketone bodies as substrates and signals in mammalian tissues. Physiological Reviews 60, 143–187.
Williamson, D.H., 1982. The production and utilization of ketone bodies in the neonate. In: Jones, C.T. (Ed.), The Biochemical Development of the Fetus. Elsevier Biomedical, Amsterdam, pp. 621–650.
Williamson, D.H., 1987. Brain substrates and the effects of nutrition. Proceedings of the Nutrition Society 46, 81–87.
Zammit, V.A., 1996. Role of insulin in hepatic fatty acid partitioning: Emerging concepts. Biochemical Journal 314, 1–14.

Lactose intolerance

D.M. Paige, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of D.M. Paige, Lactose Intolerance, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 67–73, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00169-0.

Overview 441
Historical and Geographic Perspective 441
Lactase Nonpersistence 442
Lactose Digestion and Gastrointestinal Function 442
Prevalence 443
Children 443
Adults 444
Pregnant Women 445
Secondary Lactase Deficiency 445
Lactose Digestion and Diet 446
Calcium 446
Osteoporosis 446
Nutrition policy 446
Summary 447
Further Reading 448

Overview

Low lactase levels due to genetic nonpersistence is reported in approximately 70% of the world's adult population. The prevalence is lowest in individuals of northern European descent (15%) and highest in many Asian populations with reports approaching 100%. The prevalence of lactase nonpersistence in individuals of African descent is approximately 70% to 80%. Similar levels are reported for Latinos and those of Eastern European and South American ancestry. Not all individuals with a reduced level of the enzyme lactase experience symptoms with the ingestion of dietary lactose. The presence or absence of symptoms varies with the amount and type of food consumed, intestinal transit time and level of residual intestinal lactase. Individuals with low lactase levels may tolerate a moderate intake of lactose.

Uniform agreement regarding the application and definition of terms identifying lactose nonpersistence has been lacking and has led to confusion and controversy. The 2010 National Institutes of Health Consensus conference on lactose intolerance has underscored the use of an agreed terminology. Hypolactasia is defined as a relative diminution, or very low levels of lactase enzyme activity. Lactose malabsorption, identifies a lactose test result following a lactose challenge resulting in an abnormal rise in breath hydrogen resulting from undigested lactose reaching the colon. The undigested lactose may result in one or more of the following symptoms: bloating, abdominal cramps, flatulence, and diarrhea. Lactose intolerance, is the term used to identify individuals with any of the above clinical symptoms resulting from unhydrolyzed lactose. It is a reliable indicator of unhydrolyzed lactose when properly used and interpreted. Milk intolerance, self reported, may be due to lactose maldigestion but may result entirely or partly for other reasons.

The most reliable method for diagnosing lactase deficiency is determining lactase activity in the small bowel. The test is invasive and expensive. Lactose maldigestion can generally be identified by a breath hydrogen test, the most commonly used test to measure the level of undigested lactose reaching the colon. Bacterial fermentation of the undigested lactose is responsible for the volume of breath hydrogen production. A lactose tolerance test measuring blood sugar rise is also used. Genetic testing is available. Lactose elimination trials represent a noninvasive, no cost alternative, albeit often difficult to carry out and interpret. Small bowel bacterial overgrowth can confound results. Individuals experiencing discomfort with lactose ingestion can elect to consume commercially hydrolyzed milk that is readily available, milk substitutes or alternative food sources equally rich in calcium.

Historical and Geographic Perspective

The first herd animals, sheep, are reported to have been domesticated approximately 10 000 BC. Herd animals were primarily used for meat and perhaps certain other purposes. The historical record suggests that herd animals during this period were not milked. Evidence that humans milked domesticated animals dates to approximately 4000 to 3000 BC, in northern Africa and southwest Asia. Following that time, dairying spread across Eurasia and into sub-Saharan Africa. Dairying was not, however, adopted by all groups in Asia and Africa who had suitable herd animals. Even as late as AD 1500, the beginning of the great European overseas

expansion, there were sizable areas occupied by nonmilking groups. In Africa the zone of nonmilking centered on the Congo Basin but extended beyond to cover approximately one-third of the continent. In Asia the zone of nonmilking covered the bulk of the eastern and southeastern portions of the continent, including Thailand, Vietnam, China, and Korea as well as the islands to the east. Dairying remained unknown in the Pacific region and in the Americas in pre-European times. Animal milk was not part of their diet. At that time the nonmilking people of Asia, Africa, and the Americas consumed mother's milk as infants, but normally ingested no milk after weaning.

It was striking that adults of all groups whose origins lay in the traditional zone of nonmilking were predominantly maldigesters, usually from 75% to 100% of the individuals tested. Also striking was the fact that the people with low prevalences of lactose maldigestion (northwest Europeans and certain East African pastoral groups) came from a long tradition of consuming milk, much of it in lactose-rich forms. This suggests the geographic or culture–historical hypothesis. The hypothesis is based on the assumption that in the hunting and gathering stage, human groups everywhere were like most other land mammals in their patterns of lactase activity. That is, in the normal individual lactase activity would drop at weaning to low levels, which prevailed throughout life. With the beginning of dairying, however, significant changes occurred in the diets of many human groups. As a result, there may have been a selective advantage for those aberrant individuals who experienced high levels of intestinal lactase throughout life. That advantage would have occurred only in certain situations: Where milk was an especially critical part of the diet, where the group was under dietary stress, and where people did not process all their milk into low lactose products such as aged cheese. Under those conditions, most likely to occur among pastoral groups, such aberrant individuals would drink more milk, would benefit more nutritionally as a result, and would enjoy increased prospects of survival, wellbeing, and of bearing progeny and supporting them. In classic evolutionary terms, the condition of high intestinal lactase activity throughout life, or lactase persistence, would come to be typical of such a group.

Lactase Nonpersistence

In its pure form, lactose can not be transported across the mucosa of the small intestine. To be absorbed, it must be hydrolyzed by lactase to yield glucose and galactose. These two simple sugars are rapidly and completely absorbed in the normal small intestine. The rate of lactase synthesis is high from birth until the age of 3–5 years. However, between ages 5 and 14 years, many people undergo a genetically programmed reduction in lactase synthesis resulting in only 5–10% of the lactase levels in infancy. This reduction, known as lactase nonpersistence or primary lactase deficiency, is not related to the continued intake of milk or lactose. As noted, less than one-third of the world's adult population is genetically predisposed to maintain a high degree of lactase activity or lactase persistence throughout adulthood.

Lactase persistence in the human population is inherited as a dominant genetic trait. It has been observed that low lactase level is "ancient and globally distributed" predating the appearance of a persistent lactase variant that was naturally selected in dairying regions. Hollox *et al*. report, "the continued adult production of lactase results from the persistent expression of the protein lactase-phlorizin hydrolase which is encoded by the lactase gene (LCT) on chromosome 2". Swallow notes, "the distribution of different lactase phenotypes in human populations is highly variable and is controlled by a polymorphic element cis-acting to the lactase gene. A putative causal nucleotide change has been identified and occurs on the background of a very extended haplotype that is frequent in northern Europeans, where lactase persistence is frequent".

Lactase persistence is a likely result of the advent of dairying and the result of natural selection. Samples of ancient human mitochondrial deoxyribonucleic acid (DNA) sequences from ancient skeletons in the early Neolithic Europeans support the hypothesis. Investigators did not observe the allele most often identified with lactase persistence in Europeans suggesting lactase persistence was uncommon in early European farmers thereby reinforcing the cultural–historical hypothesis.

Lactose Digestion and Gastrointestinal Function

Lactose is hydrolyzed at the intestinal jejunal brush border by the enzyme lactase into its absorbable monosaccharides glucose and galactose. Lactase activity is robust during infancy and as is the case in humans along with most mammals declines after weaning. Accordingly, the general pattern of lactase nonpersistence is a continuous decline in genetically programmed populations. A shifting pattern of lactose digestion and gastrointestinal function are the result of lactase nonpersistence. The pattern can be described and monitored during three distinct clinical phases.

First, there is a decreasing ability to digest the large lactose load consumed during the screening test. It is important to recognize that this is not an all or nothing phenomenon but rather a slowly progressive decline in available lactase activity, and that this decline, as earlier noted, can be influenced by transit time, the vehicle in which the lactose is consumed, and the intake of additional foods along with lactose.

Next, with the continued decline of lactase activity, a point is reached when available lactase activity is no longer sufficient to hydrolyze more modest levels of lactose. Therefore, the consumption of a glass of milk or another product containing the equivalent level of lactose will result in incomplete hydrolysis of the lactose consumed. The individuals so tested frequently do not recognize signs or symptoms associated with the incomplete digestion of lactose.

Finally, with the continued decline of lactase activity with increasing age, individuals become symptomatic as a result of the undigested lactose. The decline in available lactase activity reaches a recognizable clinical threshold with increasing age.

Initially, many reports had treated the population studied as a single unit and had paid incomplete attention to age-specific considerations. Distinctions between secondary lactose malabsorption due to short-term intestinal injury, and primary lactose malabsorption that has a genetic basis, were not always made. This introduced additional confounding variables. Differences in an individual's capacity to hydrolyze and tolerate a lactose challenge dose compared to his or her ability to utilize lesser amounts of lactose found in usually consumed amounts of milk created additional areas of confusion.

When attention is paid to the many factors associated with lactose digestion from infancy to old age, it is possible to place many of the seeming contradictions into perspective. What may have appeared to be incongruities in reported data appear to merge into a relatively predictable pattern of lactose digestion.

Lactose maldigestion and intolerance are influenced by age, infection, size of the lactose bolus, gastric emptying time, intestinal transit time, individual sensitivities, eating habits, genetics, environment, food ideologies, and cultural patterns. Further, symptoms of lactose malabsorption may also be the result of bacterial fermentation of undigested carbohydrate in the colon. The type and extent of the colonic bacterial profile and the absorption of hydrogen and the volatile fatty acids will influence individual reports of symptoms associated with lactose intolerance. Clearly, lactose malabsorption is not a homogeneous event. Neither is it an all or nothing phenomenon having its origins in a single etiology. Clinical expressions of lactose malabsorption, lactose intolerance, and milk rejection find their origins in one or more of the causes outlined above (**Figure 1**).

Prevalence

Children

A review of reported data on diverse populations support the conclusion that in later childhood and adolescence an important transition in lactose digestion occurs. Below 3 years of age there is lactase persistence. Between 3 and 11 years of age the beginning of a genetically controlled lactose nonpersistence is recognized. Older children and young adults are increasingly unable to digest even modest amounts of lactose. This results in increased symptom production, recognition of discomfort, and avoidance of lactose-containing products that provoke symptoms (**Table 1**).

A progressive decrease in lactase is noted from approximately 1–5 years of age through adolescence. Reported rates in United States African-American children ranged from 27% lactose maldigestion following lactose testing using a lactose load equivalent to two 8-ounce glasses of milk at 1–2 years to 74% in 11–12-year-old children. The progressive decrease in the ability to hydrolyze a lactose challenge was observed in children of both high and low socioeconomic status. Studies in white children 1–12 years of age identified only 17% of children maldigesting a lactose challenge. Signs and symptom production associated with a reduction in lactose digestion in a child population is difficult to measure due to the nature of the symptoms being reported and the signs observed and the subjective nature of the reports. This is reinforced by a report of 21 African-American girls of 11–15 years of

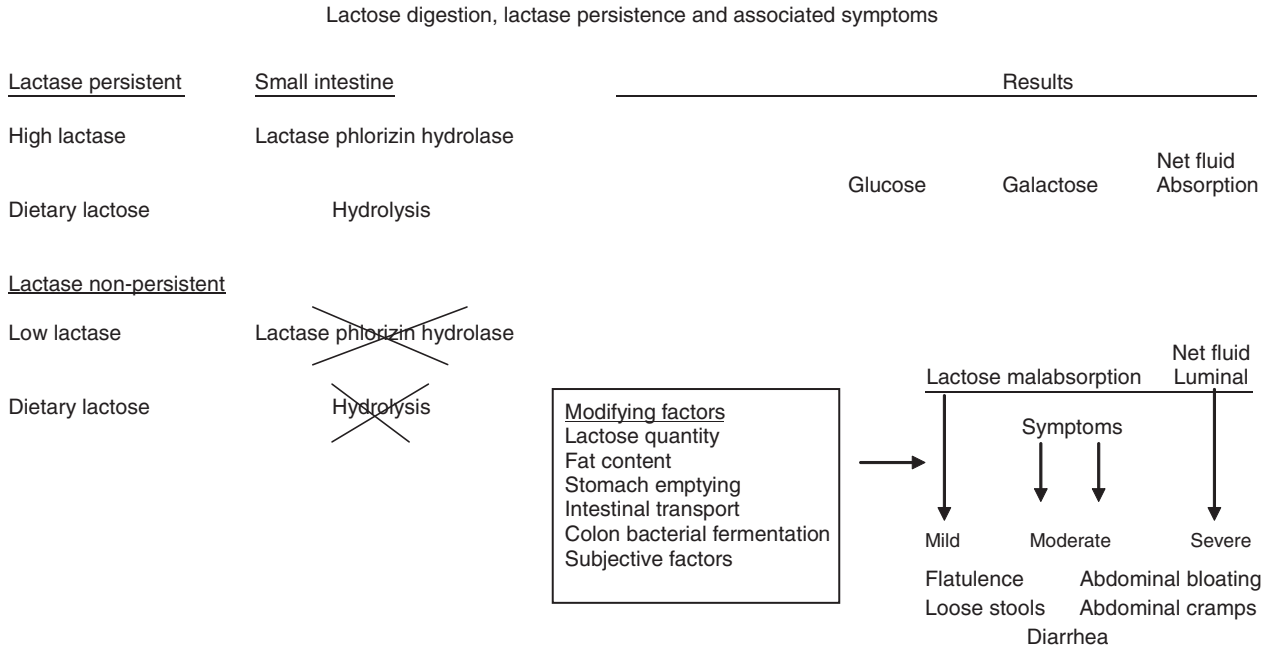


Figure 1 Lactose digestion, lactase persistence, and associated symptoms.

Table 1 Genetically determined lactase levels in healthy individuals by age and lactase persistence

Age	*Lactase level*	
	Low lactase nonpersistent individual	*High lactase persistent individual*
Fetal period	Low	Low
Birth	High	High
Weaning	Decline	High
3–12-year-old child	Reduced	High
Adolescent	Low	High
Adult	Lower	Average
Elderly	Lowest	Decline

age indicating 82% had evidence of lactose maldigestion with reports of gastrointestinal symptoms being negligible and breath hydrogen excretion, while remaining high, varied between two time periods. Consistent with the above data, milk consumption studies, both observed and reported, suggest a progressive decline in milk intake with increasing age in the African-American population of children and parallel reports in children from other populations with a high prevalence of lactose maldigestion (Table 2).

Phenotypic lactase nonpersistence needs to be distinguished from secondary lactose maldigestion and intolerance as a result of a variety of other conditions. Secondary lactose maldigestion can be observed with diarrheal disease and infection, celiac disease, allergic enteropathy, Crohns' disease, chemotherapy, radiation, and small bowel resection.

Adults

The progressive increase in prevalence of lactose maldigestion increases with age reaching reported adult levels of approximately 70% of the world's adult population. The exceptions are populations of northern and central Europeans and some Middle Eastern populations as well as groups of primarily European descent in Australia, New Zealand, and North America. Thus, minority populations in North America and Europe, as well as adult populations in most developing countries are lactose maldigesters (Table 3).

Reported milk drinking patterns of individuals classified as maldigesters vary considerably in adults. Data range from 50% reporting symptoms with one 8-ounce glass of milk, to 75% reporting symptoms with two 8-ounce glasses of milk and 30% reporting not drinking any milk. Nevertheless, caution must be exercised in interpreting reported symptoms and making the diagnosis of lactose intolerance. There can be considerable crossover between individuals who self-identify as intolerant to lactose and are not diagnosed as lactose maldigesters *versus* those in whom the diagnosis was carefully established. More attention to identifying and categorizing symptoms more precisely may help. A recent Finnish study notes flatulence as the most severe symptom in maldigesters whereas abdominal bloating is most frequently reported by individuals self-identifying as lactose maldigesters. Moreover, microbiota may play a role in the presence and intensity of lactose-related symptoms. Data suggest that increased levels of colonic bacteria, as well as their diversity, may play a role, as a result of increased fermentative capacity in reducing the symptoms associated with lactose intolerance.

Table 2 Patterns of lactose digestion by lactase status

Lactase status	*Test results*	*Symptoms*	*Lactose intolerance*	*Milk consumption*
Adequate	Normal (−)	0	0	Average (+)
Marginal lactase Deficiency	+ −	0/+	0/+ −	+
Moderate lactase Deficiency	+	0/+	0/+ −	+
Severe lactase Deficiency	+++	++	++	−

Sidney, Phillips, Paige & Bayless.

Table 3 Prevalence of lactose maldigestion in selected populations

Population	*Country*	*% Lactose maldigestion*	*Population*	*Country*	*% Lactose maldigestion*
African-American 18–54 years	USA	75	General 21–65 years	Finland	15
Asian 23–39 years	USA	100	General 20.3 years	Germany	70
Native American 18–54 years	USA	81	General 16–54 years	Chile	80
African-American 13–19 years	USA	69	Non Caucasian	Peru	94
Mexican 18–94 years	USA	53	General 38–49 years	Brazil	80
Vietnamese 22–63 years	USA	100	Arab adult	Israel	81
Sicilian 25 years average	Italy	71	General male 14–34 years	Egypt	73
Northern 28.7 years average	Italy	52	General 15–78 years	Greece	45
Central 36 years average	Italy	18	Bantu 13–43 years	Uganda	100
Romai	Hungary	56	Yoruba 13–70 years	Nigeria	83
Austrian 22 years average	Austria	20	General adult	India	61
General 20.3 years average	Finland	17	General 17–83 years	Korea	75
Aboriginal	Australia	84	General 15–64 years	Japan	100

Pregnant Women

The role of lactose digestion in pregnant women is of special interest. Despite the nutritional value of milk during pregnancy, the lactase levels in some individuals in a number of racial and ethnic groups may be insufficient to hydrolyze commonly consumed amounts of lactose resulting in lactose maldigestion and possibly milk intolerance. The Institute of Medicine report notes, that "lactose intolerance among pregnant African-American women may result in their subsequent avoidance of milk". Other populations may also experience lactose maldigestion and intolerance to milk during pregnancy.

Lactose maldigestion, in pregnant women in our studies as measured by breath hydrogen response to 240 ml of low fat (1%) milk, reinforces the Institute of Medicine's concern with lactose digestion among pregnant African-American women. We report the prevalence of lactose maldigestion in early (13–16 weeks), late (30–35 weeks) and 8 weeks postpartum as 66%, 69%, and 75% respectively. The prevalence in nonpregnant control women was 80% (**Table 4**).

Accordingly, healthcare providers instructing African-American women on the optimal dietary pattern during pregnancy need to be mindful of a high rate of lactose maldigestion. Implications for fetal growth and development remain to be answered by further study. Furthermore, health providers need to be aware that the presence or absence of symptoms may be unevenly reported by pregnant African-American women; and symptoms do not represent a reliable guide to lactose digestion. Less than 25% of pregnant lactose maldigesting women reported any symptoms with 240 ml of low fat (1%) milk. Symptoms may be further reduced when milk is consumed with other foods. Unanswered is the level of digestion and absorption of a range of nutrients in the consumed milk. Health care providers should discuss with the pregnant woman, her ability to tolerate milk, and where and when appropriate, should educate her as to other food options. In this regard, Kingfisher and Millard report that "Euroamerican staff tended to give advice that was biologically appropriate for them but not for many of their patients, a process reflecting biocentrism".

Secondary Lactase Deficiency

Secondary lactase deficiency is distinct from genetically determined loss of lactase with age. Secondary lactase deficiency is frequently associated with diseases of the small intestine. Enteric viruses such as rotavirus and Norwalk agent can induce lactase deficiency by penetration of the enterocyte in the small intestine Rotaviruses are a principal cause of diarrhea and lactose intolerance

Table 4 Lactose maldigestion[a] in pregnant and nonpregnant African-American women

African-American women	*% Lactose maldigestion*
Early pregnancy (13–16 weeks)	66
Late pregnancy (30–35 weeks)	69
Postpartum (8 weeks)	75
Nonpregnant	80

[a]Breath hydrogen rise >20ppm following the consumption of 240 ml of low fat (1%) milk containing 12 g of lactose following an overnight fast.

in infancy. Denudation of the brush border of the jejunal mucosa associated with diarrhea can lead to the loss of the other two disaccharides, maltase and sucrose. Continued diarrhea may also lead to severe complications such as monosaccharide intolerance. Giardiasis and *Ascaris lumbricoides* have also been implicated as contributing to lactose maldigestion. Severe protein malnutrition is frequently associated with lactose maldigestion. Other disease conditions that give rise to secondary lactose maldigestion are celiac disease, gluten-induced enteropathy, and tropical and non–tropical sprue. The mucosal brush border of the small intestine is severely damaged in each case.

Lactose Digestion and Diet

Calcium

Dietary calcium is an important element in skeletal development. Dairy products can account for up to three-quarters of dietary calcium in some populations. Milk is a rich source of calcium. Nevertheless, many minorities in the United States and population groups throughout the world drink decreasing amounts of milk after early childhood and little milk as adults. Given the high prevalence of lactose intolerance, alternatives to cow's milk should be identified for those who cannot tolerate lactose and desire a milk alternative. Lactose-intolerant individuals ultimately attribute their discomfort to lactose-containing foods and voluntarily reduce or eliminate their milk intake. Data from National Studies in the United States indicate African-American and Hispanic women have lower intakes of calcium compared with non Hispanic women. An Institute of Medicine Report concludes that the disparity in calcium intake "may be explained in part by the much higher prevalence of lactose intolerance among African-Americans and Hispanics, sometimes resulting in their subsequent avoidance of milk". In general, populations at risk for lactose intolerance report a lower calcium intake as a result of the decline in the intake of milk and milk products. One solution to this problem is to educate lactose-intolerant groups as to alternative calcium-containing foods, reinforce appropriate cultural patterns and dietary practices that include alternatives to milk and identify other culturally acceptable calcium-containing foods. Meeting the calcium requirement with an alternative diet is a challenge but nevertheless is required for many in the community. Although milk may serve as a primary source of calcium, appreciable quantities of calcium can be found in nondairy foods (**Table 5**).

Clearly it is more difficult to meet the published calcium recommendation with a diet low in whole cow's milk. A review of the tables of food composition reveals a variety of foods that contain acceptable levels of calcium per 100-g portion or other standard portions. Other lactose-modified dairy products including hard cheeses, yogurts, and lactose-modified milk are good calcium sources.

In addition, lactose digestive aids are available and are increasingly used. The digestive aids commercially available include lactase tablets, lactase preparations, lactose-free milk, and prehydrolyzed milk. Live culture yogurt is another alternative to milk. Lactose in yogurt is better digested than lactose in milk. Tolerance to yogurt is thought to be due to the microbial beta-galactosidase activity that digests the lactose.

Osteoporosis

The role of lactose maldigestion, calcium intake and osteoporosis has been studied. Osteoporosis and osteoporitic fractures are major public health problems. The role of lactose maldigestion and osteoporosis remains unsettled. For example, minority populations consuming small amounts of milk should be at greater risk for osteoporosis. Nevertheless, African-American and Hispanic populations in the United States appear to have a lower risk of developing osteoporosis. Caucasian and Asian women were found to have the highest risk for osteoporosis, with fracture rates of 140.7/1 00 000 and 85.4/1 00 000, respectively. Hispanic and African-American females had lower age-adjusted rates, at 49.7/1 00 000 and 57.3/1 00 000, respectively. A study of gene-identified lactose intolerance in a Dutch Caucasian elderly population is associated with lower dietary Calcium intake and serum Calcium levels but not associated with bone mineral density or fractures. The paradox reinforces the complexity of the disease and the importance of biologic, genetic and as yet undetermined factors in the etiology of osteoporosis.

Nutrition policy

Apart from the nutritional implications outlined above, there are policy considerations that require attention. Clearly milk has important economic, nutritional, and emotional significance in Western culture, a culture strongly committed to the concept that milk is an ideal food. Yet, lactose digestion should be an important consideration in developing a suitable policy regarding the use of milk and dairy products by the lactose malabsorber and by ethnic or racial groups, among whom high rates of malabsorption prevail. Accordingly, a balance must be struck between dietary guidance and the interests of a diverse population with a large number of lactose maldigesters. For many the continued use of a limited amount of milk may be appropriate and comfortable. For others dietary modification and lactose reduction or elimination may be warranted. The substitution of low-lactose products or alternative foods may be successful. Traditional diets among lactose-maldigesting populations, using little or no milk or dairy products should be respected.

Table 5 Calcium content in milligrams per 100 g portion or as noted[a]

Canned sardines (3 oz.)	372	Brewer's yeast (2 tbs)	66
Buckwheat pancakes	249	Lobster	65
Kale (raw)	225	Green beans	63
Mustard greens	220	Flounder	61
Muffins[b]	206	Bran flakes	61
Waffles[b]	192	Canned apricots (1 cup)	57
Figs (dry)	186	Gingerbread (1 piece)	57
Canned salmon (3 oz. with bones)	167	Plain rolls	55
Collard greens	162	Toaster pastry (1 piece)[b]	54
Oat breakfast cereal[b]	160	Prunes (dry)	54
Wheat pancakes	158	Orange	54
Almonds	152	Whole egg	54
Tofu (8 oz.)	143	Peanuts	54
Egg yolk	147	Artichokes	51
Corn bread[b]	139	Cod	50
Kale (frozen)	134	Brussels sprouts	50
Filberts	120	Clams (3 oz.)	47
Beet greens	118	Lima beans	47
Oysters (½ cup)	113	Puffed wheat[b]	46
Whole cow's milk (100 g)	113	Whole wheat bread (2 slices)	46
Swiss chard	105	Sweet potato	46
Rhubarb (cooked ½ cup)	105	Fruit cocktail (1 cup)	46
Canned shrimp (3 oz.)	98	Raisins (1/2 cup)	45
Okra	92	Apricots	44
Soy beans (1 cup)	90	Farina (1 cup)	44
Sunflower seeds	88	Fig bars (4 cookies)	44
Broccoli	88	Pecans	43
Sauerkraut (1 cup)	85	White bread (2 slices)	42
Potato salad (1 cup)	80	Pecans	43
Peanut butter	74	White bread (2 slices)	42
Spinach	73	Tangerine	40
Dates (dry)	72	Raspberries (raw)	40
		Apple sauce	21

[a]Oski and Paige modified from Krause and Mahan & Burton.
[b]Enriched, fortified, or restored to legal standard when one exists.

Summary

In summary, the principles of genetics and evolution help to explain the emergence of continued lactase activity beyond weaning. Darwin referred to food as a major factor in selective pressures. Lactose digestion illustrates how a certain food, by indirectly favoring the survival of those able to digest that substance, can influence the evolutionary process.

Clinical and nutritional consequences of lactose digestion in adults must be examined in relation to digestion, intolerance, milk rejection, symptoms, and their recognition. Estimates of how frequently milk intolerance is a clinically significant problem in adults vary. The protocol of individual scientific studies can influence interpretation. A balance of factors tend to prevent or minimize symptoms when the stomach, small intestine, and colon can compensate for an increased solute load, but abdominal discomfort or diarrhea occur when these small intestinal and colonic physiologic mechanisms are loaded beyond their capacity. The role of the colonic flora in metabolizing unabsorbed sugar and the importance of colonic salvage of unabsorbed carbohydrate is an important variable in the symptom complex. Secondary lactase deficiency due to infectious gastroenteritis and malnutrition represents a distinct clinical syndrome and must be distinguished from lactose intolerance.

Dietary recommendations must take account of lactose maldigestion. Milk and dairy product consumption will vary among lactose-maldigesting and milk-intolerant individuals. Lactose-reduced or lactose-free products are available to lactose-maldigesting and – intolerant individuals who wish to drink milk and milk-based products. Nevertheless, dietary recommendations must be modified and respectful of those who do not drink milk. Accordingly, appropriate alternatives to milk and other lactose-containing foods must be identified and guidance provided in developing nutritionally equivalent diets.

Further Reading

Bayless, TM, Paige, DM, 1978. Lactose tolerance by lactose-malabsorbing Indians. Gastroenterology 74, 153.

Bayless, TM, Rosensweig, NS, 1966. A racial difference in the incidence of lactase deficiency: A survey of milk tolerance and lactase deficiency in healthy males. The Journal of the American Medical Association 197, 968–972.

Burger, J, Kirchner, M, Bramanti, B, Haak, W, Thomas, MG, 2007. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. Proceedings of the Nationall Academy of Sciences of the United States of America 6 104 (10), 3736–3741.

Enattah, NS, Sahi, T, Savilahti, E, Terwilliger, JD, Peltonen, L, Jarvela, I, 2002. Identification of a variant associated with adult-type hypolactasia. Nature Genetics 30 (2), 233–237.

Goldberg, JP, Folta, SC, 2002. Milk: Can a "good" food be so bad? Pediatrics 110 (4), 826–832.

Hollox, EJ, Poulter, M, Zvarik, M, et al., 2001. Lactase haplotype diversity in the Old World. The American Journal of Human Genetics 68 (1), 160–172.

Institute of Medicine (U.S.) (1990) Subcommittee on Nutritional Status and Weight Gain During Pregnancy. Nutrition During Pregnancy, National Academy of Sciences.

Itan, Y, Jones, BL, Ingram, CJ, Swallow, DM, Thomas, MG, 2010. A worldwide correlation of lactase persistence phenotype and genotypes. BMC Evolutionary Biology 10, 36.

Itan, Y, Powell, A, Beaumont, MA, Burger, J, Thomas, MG., 2009. The origins of lactase persistence in Europe. PLoS Computational Biology 5 (8).

Jackson, RT, Latham, MC, 1979. Lactose malabsorption among Masai children of East Africa. The American Journal of Clinical Nutrition 32, 779–782.

Kingfisher, CP, Millard, AV., 1998. Milk makes me sick but my body needs it: Conflict and contradiction in the establishment of authoritative knowledge. Medical Anthropology Quarterly 12 (4), 447–466.

Koek, WN, van Meurs, JB, van der Erden, BC, et al., 2010. The T-13910C polymorphism in the lactase phlorizin hydrolase gene is associated with differences in serum calcium levels and calcium intake. Journal of Bone and Mineral Research 9, 1980–1987.

Kretchmer, N, 1972. Lactose and lactase. Scientific American 227, 70–78.

Labayen, I, Forga, L, Gonzalez, A, Lenoir-Wijnkoop, I, Nutr, R, Martinez, JA., 2001. Relationship between lactose digestion, gastrointestinal transit time and symptoms in lactose malabsorbers after dairy consumption. Alimentary Pharmacology and Therapeutics 15 (4), 543–549.

Law, D, Conklin, J, Pimentel, M., 2010. Lactose Intolerance and the Role of the Lactose Breath Test. The American Journal of Gastroenterology 105 (8), 1726–1728.

Mendel, LB, Mitchell, PH, 1907. Chemical studies on growth: I The inverting enzymes of the alimentary tract, especially in the embryo. American Journal of Physiology 20, 81–96.

Novillo, A, Peralta, D, Dima, G, Besasso, H, Soifer, L., 2010. Frequency of bacterial overgrowth in patients with clinical lactose intolerance. Acta Gastroenterol Latinoam 40 (3), 221–224.

Nutrition Reviews, 1969. Lactase deficiency in Thailand. Nutrition Reviews 27, 278–280.

Omenn, GS., 2010. Evolution in health and medicine Sackler colloquium: Evolution and public health. Proceedings of the National Academy of Sciences of the United States of America Jan 26 107 (supplement 1), 1702–1709.

Oski, FA, Paige, DM., 1994. Cow's milk is a good food for some and a poor choice for others: Eliminating the hyperbole. Archives of Pediatrics and Adolscent Medicine 148, 104–107.

Paige, DM, 1981. Lactose malabsorption in children: Prevalence, symptoms, and nutritional considerations. In: Paige, DM, Bayless, TM (Eds.), Lactose Digestion: Clinical and Nutritional Implications. The Johns Hopkins University Press, Baltimore, MD, pp. 151–161.

Paige, D.M., 1982. Milk intolerance in human populations. In Animal Products in Human Nutrition. In: Beitz, D.C. (Ed.), Nutrition Foundation Monograph Series. Academic Press, New York.

Paige, DM, Bayless, TM (Eds.), 1981. Lactose Digestion: Clinical and Nutritional Implications. Johns Hopkins University Press, Baltimore, MD.

Paige, DM, Bayless, TM, Ferry, GD, Graham, GG, 1971. Lactose malabsorption and milk rejection in Negro children. Johns Hopkins Medical Journal 129, 163–169.

Paige, D.M., Bayless, T.M., Graham, G.G., 1973. Pregnancy and lactose intolerance. The American Journal of Clinical Nutrition 26, 238–240.

Paige, D.M., Bayless, T.M., Huang, S.S., Wexler, R., 1975. Lactose hydrolyzed milk. The American Journal of Clinical Nutrition 28, 818–822.

Paige, DM, Bayless, TM, Mellits, ED, Davis, L, 1977. Lactose malabsorption in preschool black children. The American Journal of Clinical Nutrition 30, 1018–1022.

Paige, D.M., Davis, L., Bayless, T.M., Mellits, E.D., Dellinger Jr, W.S., Kreitner, M., 1979. The effects of age and lactose tolerance on blood glucose rise with whole cow and lactose hydrolyzed milk. Journal of Agriculture and Food Chemistry 27 (4), 667.

Paige, D.M., Davis, L.R., 1986. Nutritional significance of lactose: Nutritional aspects of lactose digestion. In: Fox, P.R. (Ed.), In Dairy Chemistry. Elsevier Press, London, pp. 111–132.

Paige, DM, Leonardo, E, Corando, A, Nakashima, J, Adrianzen, TB, Graham, GG, 1972. Lactose intolerance in Peruvian children: Effect of age and early nutrition. The American Journal of Clinical Nutrition 25, 297–301.

Paige, DM, Witter, F, Bronner, YL, Kessler, L, Perman, J, Paige, T., 2003. Lactose intolerance in pregnant African-Americans. Public Health Nutrition 6 (8), 801–807.

Perman, JA, Barr, RB, Watkins, JB., 1978. Sucrose malabsorption in children: non-invasive diagnosis by interval breath hydrogen determination. Journal of Pediatrics 93, 17–22.

Perman, JA, Dudley, B, 1998. Dairy products: Try them – you'll like them? The American Journal of Clinical Nutrition 68, 995–996.

Pribila, BA, Hertzler, SR, Martin, BR, Weaver, CM, Savaiano, DA, 2000. Improved lactose digestion and intolerance among African-American adolescent girls fed a dairy-rich diet. Journal of the American Dietetic Association 100 (5), 524–528.

Rao, DR, Bello, H, Warren, AP, Brown, GE., 1994. Prevalence of lactose maldigestion. Digestive Diseases and Sciences 39, 1519–1524.

Romagnuolo, J, Schiller, D, Bailey, RJ., 2002. Using breath tests wisely in a gastroenterology practice: An evidence-based review of indications and pitfalls in interpretation. The American Journal of Gastroenterology 97 (5), 1113–1126.

Rush, D, Sloan, N, Leighton, J, et al., 1988. The national WIC evaluation: Evaluation of the special supplemental food program for Women, Infants, and Children (WIC). The American Journal of Clinical Nutrition 48, 439–483.

Scrimshaw, NS, Murray, EB., 1988. The acceptability of milk and milk products in populations with a high prevalence of lactose intolerance. The American Journal of Clinical Nutrition 48, 1083–1159.

Simoons, FJ, 1970. Primary adult lactose intolerance and the milking habit: A problem in biological and cultural interrelations: II. A culture historical hypothesis. The American Journal of Digestive Diseases 15, 695–710.

Simoons, FJ, 1973. The determinants of dairying and milk use in the Old World: Ecological, physiological and cultural. Ecology of Food and Nutrition 2, 83–90.

Suarez, F, Savaiano, D, Levitt, M., 1995. A comparison of symptoms after the consumption of milk or lactose-hydrolyzed milk by people with self-reported severe lactose intolerance. New England Journal of Medicine 333, 1–334.

Suchy, FJ, Brannon, PM, Carpenter, TO, et al., 2010. National Institutes of Health Consensus Development Conference: Lactose intolerance and health. Annals of Internal Medicine 152 (12), 792–796.

Swallow, DM, 2003. Genetics of lactase persistence and lactose intolerance. Annual Review of Genetics 37, 197–219.

Tursi, A, 2004. Factors influencing lactose intolerance. European Journal of Clinical Investigation 34, 314–315.

Voelker, R, 2010. NIH Panel tackles lactose intolerance. The Journal of the American Medical Association 303 (13), 1240–1242.

Vonk, RJ, Priebe, MG, Koetse, HA, et al., 2003. Lactose intolerance: Analysis of underlying factors. European Journal of Clinical Investigation 33, 70–75.

Vonk, RJ, Stellaard, F, Priebe, MG, et al., 2001. The $^{13}C/^{2}H$-glucose test for determination of small intestinal lactase activity. The European Journal of Clinical Investigation 31, 226–233.

Weiss, KM., 2004. The unkindest cup. The Lancet 363.

Zhong, Y, Priebe, MG, Vonk, RJ, et al., 2004. The role of calonic microbiota in lactose intolerance. Digestive Diseases and Sciences 49 (1), 78–83.

Liver disorders: Nutritional management

Jenifer Thompson[a] **and Wikrom Karnsakul**[b], [a] Department of Pediatric Gastroenterology, Hepatology and Nutrition, Johns Hopkins Hospital, Baltimore, MD, United States; and [b] Department of Pediatric Gastroenterology, Hepatology and Nutrition, Johns Hopkins University School of Medicine, Baltimore, MD, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of J. Hampsey, W. Karnsakul, Liver disorders: Nutritional Management, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 87–99, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00172-0.

Introduction 450
Nutritional aspects and liver physiology in normal liver and liver diseases 450
Bile salts 450
Macronutrient metabolism 451
Carbohydrates 451
Proteins 451
Lipids 451
Fat-soluble vitamins 451
Trace elements 452
Liver in specific hepatobiliary disorders and nutritional management 452
Hepatocellular diseases 452
Nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH) 452
Alcohol-associated liver disease 452
Viral hepatitis 454
Autoimmune hepatitis 454
Hepatobiliary disorders 454
Neonatal and infantile cholestatic disorders 454
Parenteral nutrition-associated liver disease 456
Cholestatic diseases in adults 457
Metabolic liver disorders 457
Pregnancy and liver disease 459
Acute liver failure 460
Chronic liver disease 460
End stage liver disease in pre- and post-liver transplantation 460
Conclusion 462
References 462
Further reading 462

Key points

- Understand basic physiology in normal liver and changes in physiology or pathology in liver with disorders.
- Employ general approach to nutritional assessment care of patients with various kinds of liver disorders.
- Translate recent medical literatures, practice-based knowledge, and guidelines to clinical practice to improve the quality of patient care.

Glossary

Cholestasis A condition where bile cannot flow from the liver to the duodenum. It can occur because of an obstruction of the intrahepatic or extrahepatic biliary system, genetic defects, or be acquired as a side effect of several medications

Dysgeusia The distortion of the sense of taste

Gluconeogenesis A metabolic pathway that results in the formation of glucose from noncarbohydrate carbon substrates such as lactate, glycerol, and glucogenic amino acids

Glycogenesis The pathway of glycogen synthesis, in which glucose molecules are combined with chains of glycogen for energy storage

Glycolysis The metabolic process that converts glucose into pyruvate

HELLP A life-threatening obstetric complication including hemolysis, elevated liver enzymes, and low platelets syndrome

Hepatoportoenterostomy A surgical procedure performed on young infants with extrahepatic biliary atresia to allow bile drainage from intrahepatic bile ducts to the small intestine
Steatohepatitis A form of liver disease, characterized by fat accumulation and inflammation in the liver

Introduction

The liver is the largest solid organ and gland in the body, and it plays a major role in complex metabolism and numerous functions, including glycogen storage, lipid and protein synthesis, bile salt production, hormone production, and detoxification. Impairment of the liver can result in numerous consequences throughout the body including nutrition and growth. This article will cover the role of the liver in normal nutrition, including the important functions of bile salt production, macronutrient metabolism, and fat-soluble vitamin absorption, metabolism, and storage. Next the pathogenesis of malnutrition in liver disease will be discussed including the role of the microbiota, starting off with the mechanisms of malnutrition in both acute and chronic liver failure. Specific nutritional issues in liver failure will be addressed, including disturbances in carbohydrates, proteins, and fats metabolism. Nutritional disturbances in the major types of specific liver diseases will be reviewed: hepatocellular, metabolic liver disease, and biliary tract disorders. Nutritional assessment and management of patients with acute, chronic liver disease, and end stage liver disease will be discussed.

Nutritional aspects and liver physiology in normal liver and liver diseases

Bile salts

A normal functioning liver will secrete 600–1200 mL of bile to the gall bladder on a daily basis. Bile is composed of bile salts, lecithin, conjugated bilirubin, phospholipids, cholesterol, electrolytes, and water. Bile salts, the predominant component of bile, are synthesized from cholesterol in hepatocytes. The primary function of bile salts lies in its interaction with lipid digestion. Bile salts bind to large fat particles, which alone are insoluble in water, and act on them as an emulsifier, breaking them down into smaller particles called micelles (Fig. 1). Micelles, the product of the fat particle and bile salt structure, aid in the transport of fat to the mucosal membrane for absorption. Cholesterol and fat-soluble vitamins are also incorporated into mixed micelles for proper absorption.

Micellar solubilization is only required for long chain fatty acids. Short and medium chain fatty acids (12 carbons or less) do not require micelle formation for absorption; they enter the portal circulation directly, bound to albumin, and are transferred to the liver

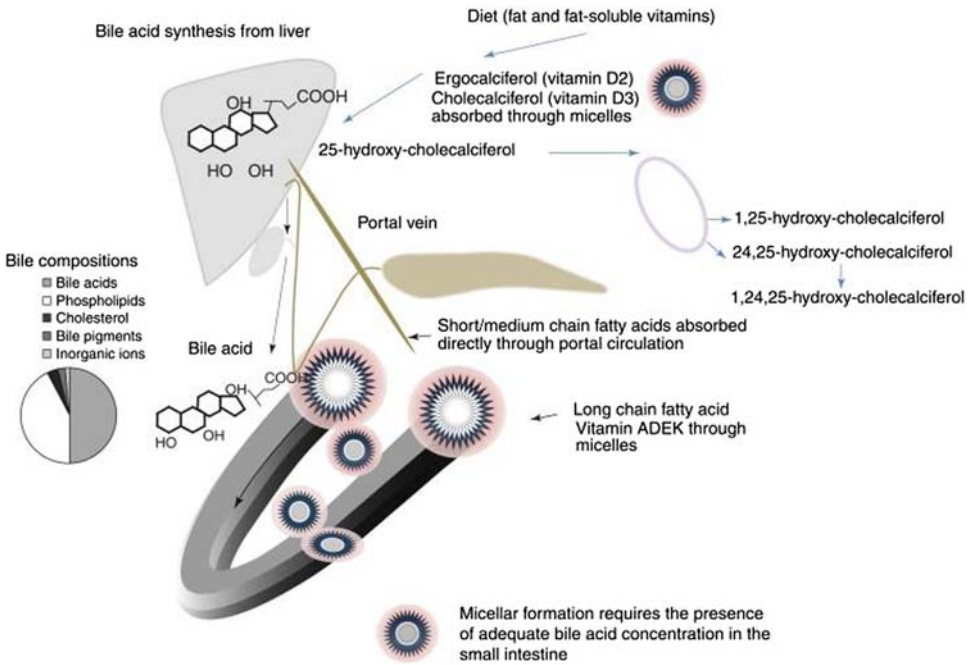


Fig. 1 Compositions of bile, bile acid excretion, micellar formation in the small intestine, absorption pathways of short, medium, and long chain fatty acids and fat-soluble vitamin, as well as vitamin D metabolism.

for oxidation. Approximately 94% of the micelle forming bile acids are reabsorbed in the ileum and shuttled via the hepatic portal vein bound to albumin back to the liver for re-use.

Macronutrient metabolism

Carbohydrates

The liver is responsible for maintaining normal blood glucose concentrations under various metabolic conditions. Among the several metabolic processes that allow this fine regulation are glycogenesis, gluconeogenesis, and glycolysis. The end product of carbohydrate digestion is largely glucose (80%), with the remaining 20% being fructose and galactose; the latter two are quickly converted into glucose in the liver. Once transported into hepatocytes, the glucose molecules are phosphorylated (via glucokinase) and cannot leave the cell unless dephosphorylated with glucose phosphatase. Glucose is either used for immediate energy release or stored as glycogen. Not surprisingly, in a patient with liver disease, glucose intolerance and insulin resistance are common. Cirrhotic patients are prone to developing diabetes. Energy from carbohydrates plays an important role in protein sparing mechanisms, preventing the use of protein as energy.

Proteins

Protein metabolism occurs in liver, specifically, the deamination of amino acids, urea formation for removal of ammonia, plasma protein synthesis, and in the interconversions between amino acids. Ingested protein is the sole source of the ten essential amino acids, and the primary source of nitrogen necessary for the synthesis of other amino acids. Protein is digested and broken down to amino acids which are absorbed into the circulation and taken to cells throughout the body, primarily the liver where they quickly become combined by peptide linkages. The plasma level of amino acids is tightly controlled and maintained near a constant level. Once the cellular limit of protein storage is met, excess amino acids are degraded and used for energy or stored as fat or glycogen. The liver is the primary site of all amino acid catabolism with the exception of branch-chained amino acid catabolism which occurs in muscle cells. The urea cycle, in which the toxic compound ammonia is converted to urea, occurs solely in the liver. The synthesis of the major plasma proteins albumin, fibrinogen, and globulin also occur in the liver.

Plasma proteins such as albumin and coagulation factors constitute approximately 50% of the proteins synthesized in the liver. In liver disease, decreased synthesis of these proteins has important clinical consequences including ascites from hypoalbuminemia and coagulopathy from decreased synthesis of coagulation factors. In end stage liver disease, hypoglycemia can result from decreased hepatic gluconeogenesis from amino acids. Decreased activity of the urea cycle enzymes result in hyperammonemia and hepatic encephalopathy, the ultimate expression of which can be cerebral edema.

Lipids

The liver is responsible for the metabolism of lipids via four key processes: fatty acid oxidation for energy, lipoprotein synthesis, the synthesis of cholesterol and phospholipids, and the conversion of carbohydrate to fat for storage. Digested fat is a key source of energy in which after splitting into fatty acids and glycerol, the fatty acid components further split via beta-oxidation into acetyl-CoA. Two molecules of acetyl-CoA become paired together to form acetoacetic acid and are transported to other cells to provide energy in the citric acid cycle.

In cholestatic liver disease there is malabsorption of dietary lipid, and consequent malnutrition. There are experimental data in primates showing that chronic alcohol consumption results in a decrease in liver phospholipids and of phosphatidylcholine. Consequently, the total phospholipid content in the mitochondrial membranes is decreased; mitochondria are altered both structurally and functionally. There is diminished mitochondrial oxidation because of decreased cytochrome oxidase activity, which can be restored by administration of phosphatidylcholine. The extent to which chronic liver disease of etiologies other than chronic alcohol consumption results in similar perturbations is unknown.

Fat-soluble vitamins

The liver plays a key role in the absorption of the fat-soluble vitamins, A, D, E, and K as they are only successfully absorbed in association with fat and sufficient quantities of bile salts (Fig. 1). The liver is also the primary storage site for several vitamins including vitamin A, E, K, and B_{12}. Vitamin A is stored in the largest quantity with amounts sufficient to prevent deficiency for 5–10 months. Stored vitamin D can last for 2–4 months and stored vitamin B_{12} can last at least 1 year after cessation of new supply. The liver is responsible for the first hydroxylation of vitamin D to the circulation form, 25-OH vitamin D.

Deficiencies of fat-soluble vitamins are common in liver disease associated with steatorrhea due to the concomitant malabsorption of fat. Vitamin A deficiency can result in anorexia, growth failure, decreased resistance to infections, and night blindness. Vitamin D deficiency results in osteopenia or osteoporosis, as well as rickets in children. The prevalence of fractures is increased

in women with a history of alcohol abuse; deficiencies of vitamin D as well as calcium, phosphorus, and fluoride may all play a role. The deficiency of vitamin E results in neuroaxonal dystrophy, clinically manifesting as peripheral neuropathy and cerebellar disturbances. Vitamin K deficiency results in hemorrhage because of reduced synthesis in clotting factors.

Trace elements

Zinc deficiency in individuals with cirrhosis may contribute to hypoalbuminemia and dermatitis as well as anorexia from dysgeusia. Deficiency of selenium can lead to decreased synthesis of important antioxidant selenoproteins such as glutathione peroxidase. Little is known about the effect of acute or chronic liver disease on other trace elements.

Liver in specific hepatobiliary disorders and nutritional management

Hepatocellular diseases

Nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH)

With the unrelenting rise in obesity worldwide, NAFLD and NASH have become very important causes of liver disease in both children and adults. Children with NAFLD may present before their fifth birthday, with a higher prevalence in male children. Hepatic fibrosis can occur and may even evolve into cirrhosis during childhood. Treatment consists of weight reduction and aerobic exercise. Practice guidance from the American Association for the Study of Liver diseases (2018) discussed though management of NAFLD/NASH (Chalasani et al., 2018).

In adults, NASH and NAFLD have been recognized for at least half a century as chronic liver diseases associated with obesity (with or without noninsulin-dependent diabetes mellitus and with or without hyperlipidemia). NAFLD may account for as much as 80% of cases of elevated liver enzymes in the US. Most adults with the disorders are 110–130% above ideal body weight. The prognosis of NAFLD is good if weight reduction is achieved. NASH is usually slowly progressive but can lead to cirrhosis and the need for liver transplantation in the minority of individuals affected. Progression from non-alcoholic fatty liver to NASH can be influenced by diet and gut microbiota, but there is also genetic predisposition, particularly the presence of the PNPLA3 which is associated with NASH development.

In many patients, NAFLD is a component of the insulin-resistance syndrome known as metabolic syndrome, which is characterized by central obesity, hypertension, hypertriglyceridemia, low levels of high-density lipoprotein-cholesterol, and hyperglycemia. In patients with this syndrome, it is hypothesized that there is greater insulin resistance in muscles and adipose tissue than liver. Those with the BMI class $>30\ \mathrm{kg\,m^{-2}}$ have an increased prevalence of each of the five components of the metabolic syndrome. Patients with NASH tend to exhibit a higher intake of saturated fatty acids, total fat, and cholesterol, and a lower intake of polyunsaturated fat, fiber, and vitamins E as an antioxidant. These findings provide a strong rationale for specific dietary modifications in NASH patients. A daily deficit of 500–1000 calories and approximately 150 min per week of aerobic exercise is recommended. Approximately 10% weight reduction from a baseline at a rate of approximately 1–2 pounds a week is generally the successful principle of weight management. Certain foods associated with NAFLD such as sodas and red meats should be avoided, and nuts which are associated with reduced risk of NAFLD should be encouraged (**Fig. 2**). Primary care physicians should screen for NAFLD in high-risk populations and implement the above therapy along with behavior modification programs as first line treatment. Other therapies include weight-loss medications, protein-sparing modified fasting, and weight-loss surgeries in those who have morbid obesity.

Probiotics are an emerging therapy for NAFLD. Patients with NAFLD have abnormal gut microbiota. They have much fewer Bacteroidetes and greater amounts of *Prevotella* and *Porphyromas* compared to healthy subjects. Additionally, they have elevated concentrations of *Lactobacillus*, *Escherichia*, and *Streptococcus* and diminished levels of *Ruminococcaceae* and *Faecalibacterium prausnitzii*. These changes may lead to increased activity of toll-like receptors and nucleotide-binding oligomerization domain pathways, resulting in an alteration of the tight junction multiprotein complexes and leading to increased gut permeability. In NASH, high-fat diet can change the composition of the gut microbiota, reducing the prevalence of protective intestinal bacteria and favoring opportunistic pathogenic Gram-negative species that produce lipopolysaccharides that can enter the portal circulation. There, the lipopolysaccharides can bind to toll-like receptor 4 and other co-receptors in the liver associated with inflammation, resulting in NASH development. In rodent models, ingestion of fructose has been shown to increase bacterial endotoxin levels in the plasma. In addition to probiotics, prebiotics such as dietary fiber are considered to have positive effects on health status due to its fermentation into short chain fatty acids by intestinal bacteria.

Alcohol-associated liver disease

The term alcohol-associated liver disease refers to a spectrum of types of hepatic injury associated with continuous alcohol consumption, ranging from alcoholic fatty liver, alcoholic steatohepatitis, cirrhosis, and hepatocellular carcinoma, typically progressing in that order. Nutritional disturbances in alcoholics are an important cause of morbidity and mortality; all classes of nutrients are affected. Anorexia leads to decreased food intake and subsequent protein-calorie malnutrition. Maldigestion and

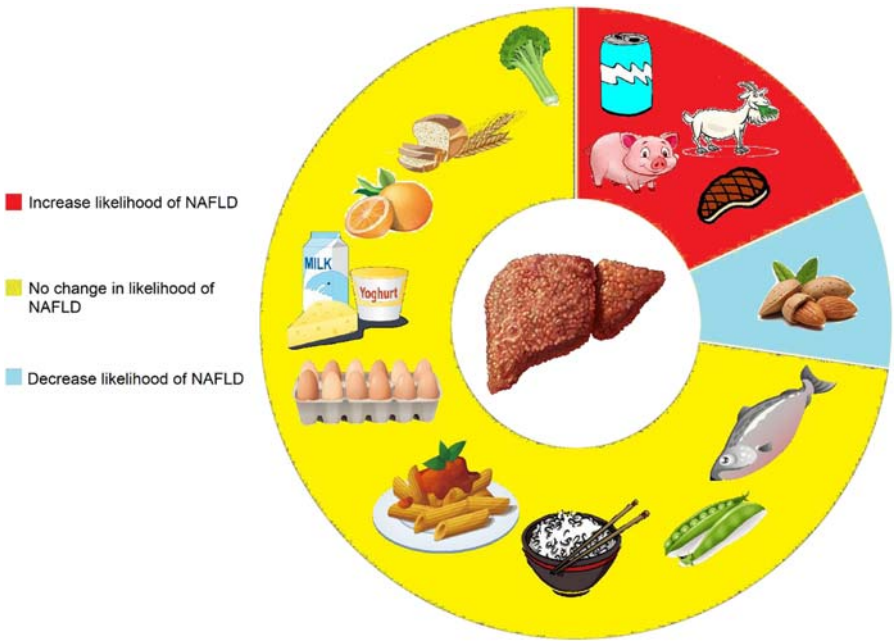


Fig. 2 Influence of various foods on the development of nonalcoholic fatty liver disease (NAFLD). Red meat and soft drinks increase likelihood of NAFLD, nuts decrease likelihood of NAFLD, and fish, legumes, refined grains, whole grains, eggs, dairy, fruits, and vegetables do not change likelihood of NAFLD.

malabsorption can occur secondary to chronic alcohol injury to the small intestinal mucosa. Alcohol consumption is often associated with chronic pancreatic insufficiency which results in steatorrhea and decreased absorption of dietary protein, fat, and fat-soluble vitamins. Chronic alcohol consumption also results in impaired hepatic amino acid uptake and protein synthesis.

In alcoholics, utilization of lipids and carbohydrates is markedly compromised due to an excess of reductive equivalents and impaired oxidation of triglycerides. Alcoholics are often resistant to insulin and exhibit impaired uptake of glucose into muscle cells. Insulin-dependent diabetes is common. Heavy alcohol consumption is frequently associated with deficiencies of a wide variety of micronutrients including the fat- and water-soluble vitamins, particularly folate, pyridoxal-5′-phosphate, thiamine, and vitamin A.

Table 1 summarizes the five published controlled trials of the effect of oral or enteral nutritional supplements on patients with alcoholic hepatitis. In most, nitrogen balance and protein synthesis improved, although no effect on mortality was shown, perhaps because of the small number of patients studied or the duration of follow-up. In the largest study, at 1 year of follow-up, the experimental group had a significantly better survival: 2/24 or 8% died as compared to 10/27 or 37% of the controls. In general, the effects of parenteral nutrition in alcoholic liver disease are similar to enteral nutritional supplements.

Many studies have examined the effect of oral or enteral nutritional supplementation in patients with alcoholic cirrhosis. Results are summarized in **Table 2**. Many studies are small and of short duration, so it is not surprising that results are inconclusive. Most studies demonstrated an improvement in nitrogen balance and protein synthesis; only one showed increased survival in the treated group. Taken together, these studies suggest that there are benefits to nutritional supplementation in this population. In addition, meta-analyses of nutritional support indicated an improvement in hepatic encephalopathy and fewer episodes of infections. However, more recent trial which compared intensive enteral nutrition to conventional nutrition did not demonstrate additional survival benefit of intensive nutrition. There was an association of increased rates of infection and mortality at 6 months when daily calorie intake was less than 21.5 kcal kg^{-1} day^{-1} (Crabb et al., 2020).

A variety of international associations have made nutritional recommendations for patients with various types of alcoholic liver disease. The primary and most important recommendation is abstinence, which may be all that is needed in patients with fatty liver. There is no safe amount of alcohol that can be consumed without risk. Weight loss is also an important lifestyle change that patients should implement. Alcohol in combination with obesity is known to synergistically induce liver injury, cirrhosis, and hepatocellular carcinoma. Patients with alcoholic hepatitis should take 40 kcal kg^{-1}, 1.3–1.5 g protein kg^{-1}, 4–5 g kg^{-1} of carbohydrates, and 1–2 g kg^{-1} of lipids per day. Those with cirrhosis without malnutrition should take 35 kcal kg^{-1}, 1.3–1.5 g protein $kg^{-1}$and carbohydrates and lipids as recommended for patients with alcoholic hepatitis. Those with cirrhosis and malnutrition should take higher amounts of protein (1.5–2.0 g kg^{-1}) and lipids (2.0–2 g kg^{-1}) and lower amounts of carbohydrates (3–4 g kg^{-1}). Fluid should be restricted to 2–2.5 L day^{-1}; B-vitamins, folate, thiamine, vitamins C, and K should be routinely supplemented. In addition, patients with cholestasis should take 50% of their dietary lipids as medium chain triglycerides and should be supplemented with the fat-soluble vitamins: A, D, E, and K. The major strategy in the management of alcoholic cirrhosis with ascites and edema is to restrict fluids to 1–1.5 L day^{-1} and to restrict sodium as well. Emerging therapies for alcoholic liver disease include probiotics, farnesoid X

Table 1 Studies on therapy of alcoholic hepatitis with oral or enteral nutritional supplements.

Design	*Patients (n)*	*Duration (days)*	*Experimental treatment (EXP)*	*Control Treatment (CTR)*	*Mortality*	*Secondary endpoints*
Open label	16	16–42	Oral (standard hospital diet) or intravenous supplement (51.6–77.4 g protein)	None	Not assessed	Nitrogen balance + albumin improved in EXP, CTR not assessed improvement of albumin. Transferrin. RBP
Historical controls	57	30	Standard hospital diet (2500 kcal day^{-1}) + 2200 kcal day^{-1} BCAA	Standard hospital diet	NS	Positive nitrogen balance in EXP, delayed hypersensitivity improved
Randomized, controlled	64	21	Standard diet (~2000 kcal day^{-1}) + 65 g standard AA or BCAA	Standard diet, 80 g protein day^{-1}	NS	Positive nitrogen balance in EXP, delayed hypersensitivty improved
Crossover	14	6	Nasoduodenal tube, 35 kcal kg^{-1} day^{-1}, fat/carbohydrate/protein 45/40/15%	3 days standard hospital diet (35 kcal kg^{-1} day^{-1})	0/6 contrs. 3/8 treatm	Nitrogen balance improved fivefold at 2 weeks
Randomized, controlled	71	28	Nasogastric tube, 2000 kcal day^{-1}, 72 g protein day^{-1}, 31% BCAA	Standard diet (1 g protein per kg) + 40 mg day^{-1} prednisolone	11/35 TEN 9/36 PRED NS FU: 2/24 TEN 10/27 ($P = 0.04$)	No dropouts in PRED, 8 dropouts in TEN; equal improvements of albumin, Child score, Maddrey score; equal rate of infections

Abbreviations are as follows: AA, amino acids; BCAA, branched-chain amino acids; FU, follow-up; NS, not significant; PRED, prednisolone group; TEN, total enteral nutrition group.

receptor (FXR) agonists, and sodium/glucose cotransporter 2 (SGLT2) inhibitors. Practice guidance from the American Association for the Study of Liver diseases (2019) discussed though management of alcohol-associated liver disease (Crabb et al., 2020).

Viral hepatitis

Patients with acute viral hepatitis are generally not at risk for nutritional deficiencies but they do experience anorexia and cholestasis resulting in a brief period of malabsorption. Patients chronically infected with hepatitis B and C viruses (HBV, HCV) may develop cirrhosis over 10–20 years with of the potential to develop hepatocellular carcinoma. Patients with chronic viral infection with significant alcohol intake, insulin resistance, obesity, cholestasis, cirrhosis, and cancer require additional nutritional assessment as mentioned in other sections of this article. Hepatic steatosis, obesity, and alcohol consumption are risk factors for antiviral treatment failure to achieve virologic response. Antiviral treatment for HBV and HCV with interferons could cause anorexia and further nutritional deficiency. HIV patients who have chronic HBV and HCV co-infection are expected to acquire more nutrient deficits.

Autoimmune hepatitis

Autoimmune hepatitides most frequently presents itself in both children and adults with a female preponderance. The nutritional assessment and therapy are not different from those of other types of hepatitis. The disease can be accompanied by intestinal diseases such as inflammatory bowel disease or celiac disease, and the nutritional management should take both organ systems into account. Although mild liver function abnormalities are common in celiac disease, there are recent reports of celiac disease in patients with severe liver disease, all of whom demonstrated an improvement in their liver disease with introduction of a gluten-free diet. Recent knowledge has appreciated the influence of the gut microbiota on liver via the gut-liver axis (**Fig. 3**). Disruption in normal microbiota composition may increase risk of autoimmune hepatitis.

Hepatobiliary disorders

Neonatal and infantile cholestatic disorders

Cholestasis is a condition that affects one in 2500 infants; it caused by poor bile flow resulting in increased concentrations of serum bile acids and conjugated hyperbilirubinemia. The major differential diagnoses of conjugated hyperbilirubinemia in the first 30 days of life are extrahepatic biliary atresia, infectious neonatal hepatitis from viruses, bacteria, and parasites as well as the neonatal hepatitis syndrome, for which a large number of specific genetic disorders have now been identified. These include alpha-1-antitrypsin deficiency, progressive familial intrahepatic cholestasis (PFIC I–III), inborn errors of bile salt synthesis, cystic fibrosis-associated liver disease, Alagille syndrome, hypothyroidism, panhypopituitarism, and other neonatal cholestatic

Table 2 Studies on treating alcoholic cirrhosis with oral and enteral nutritional therapy.

Design	*Patients (n)*	*Duration (days)*	*Experimental treatment (EXP)*	*Control treatment (CTR)*	*Mortality*	*Secondary endpoints*
Randomized, double-blind, crossover	8	12	BCAA (50%) formula	Standard diet (18% BCAA)	1 death (infection)	EXP equal to CTR positive nitrogen balance
Open label	10	4	40 g protein + 40 g BCAA formula day^{-1}	2100 kcal day^{-1} 80 g protein day^{-1}	None	EXP equal to CTR positive nitrogen balance
Randomized, controlled	36	28	50 kcal kg^{-1}, 1.5 g protein day^{-1}	Standard diet	EXP 2/17 CTR 5/19 (NS)	No differences child score
Randomized, controlled	35 (23 alc.)	23–35	211 kcal day^{-1} including 71 g BCAA formula	Standard diet	Improved ($P = 0.02$)	Improved, albumin improved
Randomized	64	90	Standard diet + BCAA supplement (0.24 g kg^{-1})	Standard diet + casein supplement	None	Nitrogen balance improved in both. BCAA better than standard diet
Randomized	31	28	Casein supplement (1.5 g protein day^{-1}, 40 kcal day^{-1} kg^{-1} day^{-1})	Standard diet	NS	Both groups improved nitrogen balance and albumin
Randomized, controlled	51	12 (months)	Standard diet + casein supplement (1000 kcal day^{-1}, 34 g protein day^{-1})	Standard diet	EXP 3/26 CTR 6/25 (NS)	Fewer hospitalizations, improved albumin and visceral protein
Open label	15	38	Increasing amounts of protein via standard diet (1.0–1.8 g kg^{-1} day^{-1})	None	None	Increased protein retention through gradual or protein intake
Open label	26	30	Standard diet	None	None	Anthropometric ratios improved
Open label	31	6 (months)	Standard diet + casein supplem. (1000 kcal day^{-1}, 34 g protein day^{-1})	None	6 deaths/31	Increased albumin, improved cellular immunity

Abbreviations used are as follows: BCAA, branched-chain amino acid; CTR, control group; EXP, experimental group; HRF, hepatorenal failure; NS, not significant.

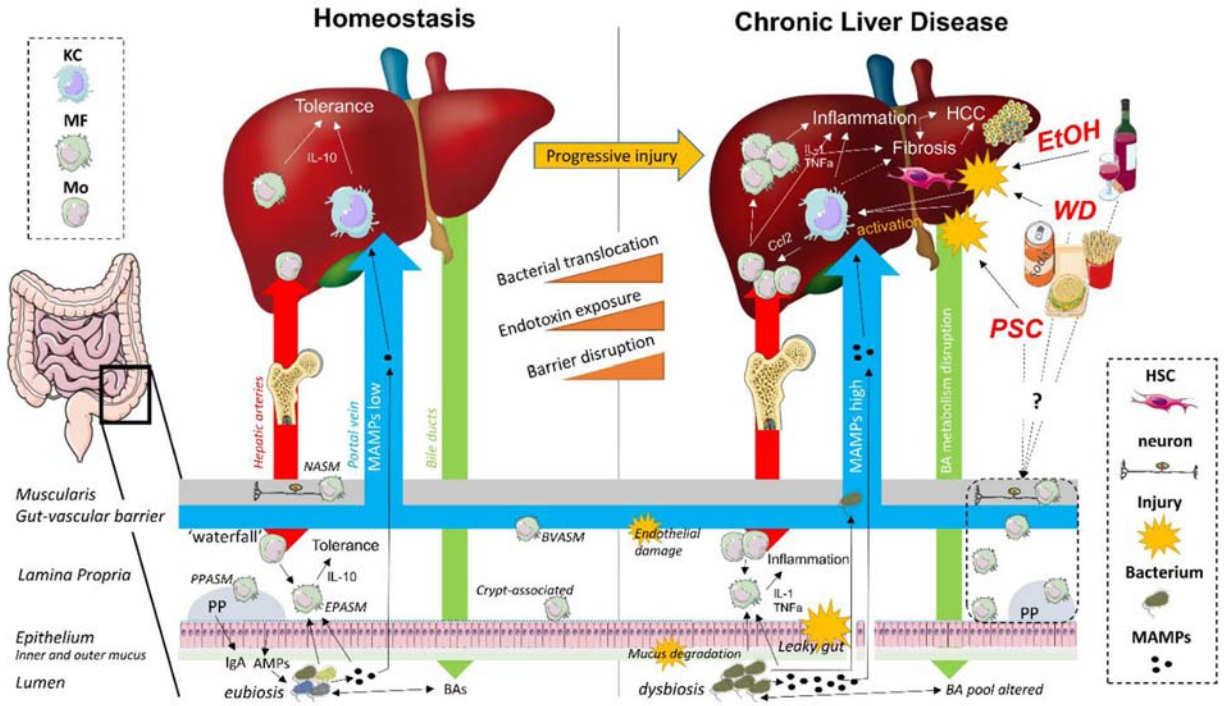


Fig. 3 Gut–liver axis in homeostasis and chronic liver disease. AMPs: anti-microbial peptides; BAs: bile acids; BVASM: blood vessel-associated macrophage; EPASM: epithelium-associated macrophage; EtOH: ethanol; HCC: hepatocellular carcinoma; HSC: hepatic stellate cell; KC: Kupffer cell; MF: macrophage; Mo: monocyte; NASM: neuron-associated macrophage; PP: Peyer's patch; PPASM: Peyer's patch-associated macrophage; PSC: primary sclerosing cholangitis; WD: Western diet. This figure is reproduced under the Creative Common CC BY license from De Muynck, K., Vanderborght, B., Van Vlierberghe, H., Devisscher, L., 2021. The gut–liver axis in chronic liver disease: a macrophage perspective. Cells. 10(11), 2959. https://doi.org/10.3390/cells10112959.

syndromes. Cholestasis and malabsorptive complications usually resolve when the specific treatment is applied in a timely manner; however, children with PFIC and Alagille syndrome tend to suffer from cholestasis and malnutrition without specific therapy such as biliary diversion (in some children with PFIC I and Alagille syndrome) and liver transplantation. Despite hepatoportoenterostomy, which if performed before 60 days of age may at least delay disease progression, these infants often develop malabsorption of fat and fat-soluble vitamins secondary to decreased bile flow and fat emulsification (Fig. 4). Aggressive formula feeding in early infancy period may be needed via nasogastric tube feeding with a high medium chain triglyceride containing infant formula. In other biliary obstructive conditions such as choledochal cyst, bile duct stricture, choledocholithiasis, and tumors or masses (intrinsic and extrinsic) in hepatobiliary regions, a surgical intervention may completely reverse cholestasis and nutrient malabsorption (Tessitore et al., 2021).

Malnutrition in cholestasis, particularly chronic cholestasis, are influenced by four major factors: poor nutrient intake, malabsorption/maldigestion, increased energy needs, and endocrine dysfunction (Fig. 5). Poor nutrient intake may be due to anorexia which could be caused by a change in amino acid metabolism, resulting in increased tryptophan levels and consequently increased brain serotonergic activity, potentially affecting regulation eating behavior. Malabsorption is due to that lack of bile acids, decreasing the absorption of fats and fat-soluble vitamins. Increased energy need in these patients is thought to be due to a hypermetabolic state induced by thyroid hormone activation by bile acids. Endocrine dysfunction can occur due to decreased production of insulin-like growth factor 1 (IGF-1) and IGF binding protein 3 (IGF-BP3), both produced in the liver, decreasing growth hormone activity and resulting in growth failure in these patients. The nutritional consequences are much the same for all—steatorrhea and malabsorption of the fat-soluble vitamins. Nutritional management is also much the same for all: use of a hydrolyzate or elemental formula rich in medium chain triglyceride and supplementation with vitamins A, D, E, and K, and possibly with additional micronutrients such as zinc, iron, magnesium, and calcium. Medium chain triglyceride should be administered with long chain triglycerides in a 1:1 ratio; although long chain triglycerides are poorly absorbed in these patients, some do ultimately get absorbed. Infants and children with cholestasis require higher protein intake than healthy individuals to ensure positive protein balance; diets enriched in branched chain amino acids may be beneficial. Carbohydrates are a major source of energy in these patients. Short chain polymers such as maltodextrin are typically given, but starch can also be administered although bloating and diarrhea may occur in very young patients due to their immature amylase. Water-miscible vitamin E is poorly absorbed; administration of vitamin E solubilized in polyethylene glycol succinate is a more effective way to administer vitamin E to cholestatic infants. Simultaneous administration of other fat-soluble vitamins with this form of vitamin E is expected to improve them as well. To better tailor supplementation, routine monitoring of vitamin concentration every 3–6 months is recommended for infants with cholestasis, or monthly for those with severe cholestasis. Infants with biliary atresia are more likely to have vitamin deficiency, particularly vitamin D deficiency (~82%), despite supplementation relative to other cholestasis etiologies. Ursodeoxycholic acid (UDCA) is the major bile acid from the black bear and has been used in traditional Asian medicine for centuries for the treatment of hepatobiliary diseases. Synthetic UDCA has been used effectively in children and adults with cholestasis to improve bile flow and nutrient malabsorption. After resolution of cholestasis, vitamin supplementation may be discontinued if serum vitamin concentrations are normal with routine monitoring early on.

Parenteral nutrition-associated liver disease

Premature infants and children with short bowel syndrome are particularly prone to develop this disorder and in the pediatric age group, parenteral nutrition -associated liver disease is usually cholestatic. The cholestasis can be solely intrahepatic or can be

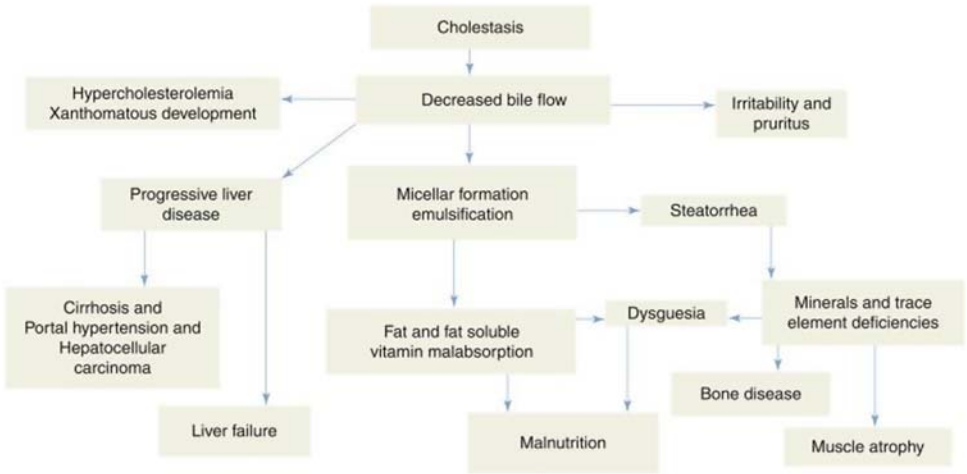


Fig. 4 Altered physiology of liver and bile flow (cholestasis) leads to nutrient malabsorption, malnutrition, and progressive liver disease.

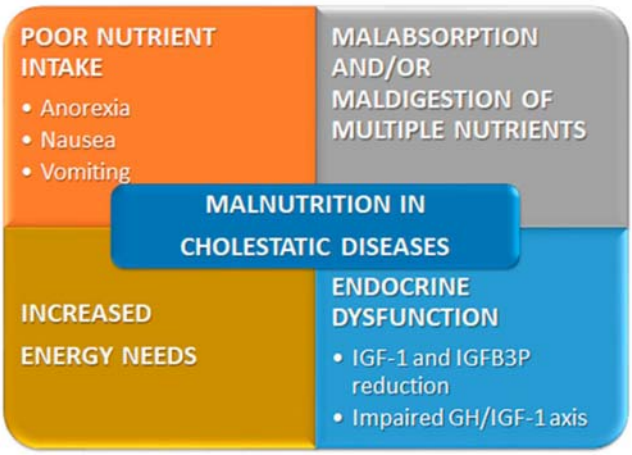


Fig. 5 Major factors that determine malnutrition in pediatric chronic cholestatic liver disease. GH: growth hormone; IGFB3P: insulin-like growth factor binding protein 3; IGF-1: insulin-like growth factor-1. This figure is reproduced under the Creative Common CC BY license from Tessitore, M., Sorrentino, E., Schiano Di Cola, G., Colucci, A., Vajro, P., Mandato, C., 2021. Malnutrition in pediatric chronic cholestatic disease: an up-to-date overview. Nutrients. 13(8), 2785. https://doi.org/10.3390/nu13082785.

associated with cholelithiasis. Parenteral nutrition-associated liver disease can be seen at any age and with any disease etiology resulting in long-term dependence on parenteral nutrition; in adults and older children, steatosis is more common as an initial presentation rather than cholestasis. Potential pathogenetic mechanisms include the gastrointestinal dysfunction associated with the lack of enteral nutrients as well as components of the parenteral nutritional solutions as potential hepatotoxins including amino acids, glucose, lipids (particularly peroxidizable lipids), and photo-exposed multi-vitamins. The most effective management is aggressive administration of enteral nutrients and decrease or discontinuation of parenteral nutrition as early as possible. Providing intravenous fat as fish oil, as well as limiting the overall amount of intravenous lipid may also help in minimizing liver damage seen in association with parenteral nutrition.

Cholestatic diseases in adults

Primary sclerosing cholangitis most commonly presents itself in association with ulcerative colitis, less commonly with Crohn's disease. The nutritional management of the disorder is essentially like that of other cholestatic disorders; in patients with Crohn's disease of the small bowel, aggressive administration of an elemental diet rich in medium chain triglycerides may be beneficial. Primary biliary cirrhosis, a disease generally presenting itself in young female adults, results in steatorrhea and malabsorption of the fat-soluble vitamins (Fig. 4). Osteoporosis and osteopenia are common. Other cholestatic diseases include common bile duct obstruction secondary to stone, stricture, parasitic infestation, and pancreatobiliary tumors or cancers. It is accepted however that endoscopic interventions should be used as needed in the case of significant biliary obstruction. For prevention of severe osteoporosis supplementation with vitamin D and calcium are needed. Vitamin K and alendronate may be beneficial in increasing bone mineral density. Serum levels of the fat-soluble vitamins should be monitored in high-risk patients and vitamins replaced as appropriate.

Metabolic liver disorders

Metabolic diseases or inborn error diseases of metabolism mostly are an autosomal recessive disorder resulting in an enzyme deficiency (Table 3). In general, a restriction of the responsible substances which cause abnormal metabolic pathways and noxious intermediates will diminish organ damages such as in galactosemia, hereditary fructose intolerance, tyrosinemia type 1, and urea cycle defects. Liver function improves by galactose and lactose elimination in individuals with galactosemia. The international guidelines for galactosemia treatment, published in 2017, recommend elimination of sources of lactose and galactose from dairy products but allowing small amounts of galactose from fruit, legume, vegetable, or mature cheese (<25 mg/100 g). Despite restriction, long-term complications such as mental disability, speech defects, ovarian failure, and neurologic syndromes can still occur as a result of endogenous production of galactose or minimal exposure to nondairy galactose-containing dietary sources. Some patients with galactosemia could tolerate galactose later in life, and reduction in diet restrictions should be strictly discussed with their physicians and dieticians. Adequate calcium supplements should be considered in this population. Asymptomatic heterozygous mothers should be recommended to avoid dietary galactose during subsequent pregnancies.

Table 3 Summary of metabolic liver diseases, management, clinical and laboratory monitoring.

Metabolic disorders	*Metabolic defects*	*Management*	*Laboratory and clinical monitoring*
		Elimination of galactose and lactose (mainly dairy products and breast milk).	Serum glucose for hypoglycemia as a complication in acute phase. Urine reducing
Galactosemia	Galactose-1- phosphate, toxic metabolite accumulates in liver and other organs	Appropriate commercial infant formulas (soy formula, lactose- free formulas) for infants. For complete elimination grains, fruits, and vegetables contain galactose such as American persimmon, papaya, tomato, watermelon, etc. Calcium and vitamin D supplementation possible.	Substance, serum galactose for compliant issues IQ test. Infertility from ovarian failure. Bone mineral density. Eye exam for cataracts.
Hereditary fructose intolerance	Accumulation of fructose-1-phosphate in the liver caused by aldolase B deficiency. This substance is a competitive inhibitor of phosphorylase which regulates the conversion of glycogen to glucose	Breast milk and sucrose-free formulas for infants. Restriction of fructose, sucrose, and sorbitol. Only certain vegetables are permitted. Commercial products, medication, and toothpaste may contain small amounts of sucrose.	Monitor growth, liver span, and intellectual development. Renal tubular function needs to be monitored in long-term fructose exposure. No biochemical method for monitoring fructose restriction. Monitor growth, liver span, and intellectual development.
Glycogen storage disease (GSD) (I, III, IV, VI, and IX)	Correct hypoglycemia (less in GSD IV), lactic acidosis, hyperuricemia (in type I), and hyperlipidemia	Continuous nocturnal nasogastric/ gastrostomy tube feeding with formula in infants. Uncooked corn starch (1–2 g kg^{-1} every 4 h) in older children and adults. High protein diet (3 mg kg^{-1} day^{-1}) in GSD type III only. NTBC in association with dietary treatment.	To maintain normal serum glucose, lipid profile, and especially lactic acid and uric acid in GSD I. Benign and malignant hepatic adenoma in GSD I and III.
Hepatorenal tyrosinemia I	Defects of tyrosine metabolism	Restriction of phenylalanine and tyrosine, and methionine (only if prolonged hypermethioninemia is observed). Low phenylalanine and tyrosine containing formula. Optimal protein diet.	To monitor plasma amino acid profile. To normalize plasma tyrosine and ensure normal growth.
Urea cycle disorders	Deficiencies of urea cycle enzymes	Restriction of upstream essential nutrients to prevent intoxication and supplementation of downstream nutrients.	Serum ammonia.
		Arginine substitution in severe ornithine transcarbamylase (OTC) deficiency and carbamyl-phosphate-synthetase (CPS)-I deficiency. Avoid catabolism. Ammonia removal with sodium benzoate. Vitamin supplementation (folate and trace elements). Carnitine supplementation if low. Sufficient fluid intake. Avoid hidden nitrogen, e.g., licorice. Liver transplantation in severe OTC and CPS deficiency. Gene therapy (future). Routine phlebotomy if serum ferritin $>$200g L^{-1} in female, $>$300g L^{-1} in male (limited data in children).	Urine orotate (defect beyond CPS). Plasma amino acid. Serum electrolytes (hypokalemia from sodium benzoate). Hemoglobin, serum iron, ferritin, and transferrin saturation.
Hemochromatosis	HFE mutation	Low-iron diet. Avoid vitamin C and medications or vitamin supplements containing iron.	To maintain ferritin level $<$50 μg L^{-1}. To document hepatic iron depletion by

Table 3 Summary of metabolic liver diseases, management, clinical and laboratory monitoring.—cont'd

Metabolic disorders	*Metabolic defects*	*Management*	*Laboratory and clinical monitoring*
		Copper-restricted diet (copper enriched diet includes chocolate, nuts, legumes, mushrooms, shellfish, and liver).	Liver biopsy or Magnetic Resonance Imaging (MRI).
	Copper accumulation	Copper chelating agents (trientine, D-penicillamine, and zinc).	Total serum copper, ceruloplasmin, urinary copper excretion.
Wilson's disease	Secondary to mutations in ATP7B, a copper- binding ATPase	Vitamin E as adjuvant therapy. Vitamin B_6 (pyridoxine) 25–50 mg day^{-1} for individuals using D-penicillamine as a chelator. Liver transplantation in severe or fulminant liver disease. Carbohydrate loading with intravenous glucose.	Serial eye exam in individuals with Kayser- Fleisher rings before therapy. To monitor peripheral neuropathy.
Hepatic porphyria (acute intermittent porphyria)	HMB-synthase deficiency	Avoid certain drugs (enzyme inducers), stress, fasting, menstruation, and alcohol. Hematin.	Long-term risk for hypertension, renal insufficiency, and hepatocellular carcinoma.

Infants with tyrosinemia type 1 not only require instant restriction of phenylalanine and tyrosine but also titration of protein intake to allow normal growth and development and to avoid tissue catabolism. The current recommendation is the specific treatment with 2-(2 nitro-4-3 trifluoro-methylbenzoyl)-1,3-cyclohexanedione (NTBC) coupled with a tyrosine- and phenylalanine-restricted diet including several commercial products containing tyrosine and phenylalanine-free amino acid mixtures supplemented with vitamins and minerals.

Urea cycle disorders present varying degrees of hyperammonemia. In the neonatal period, these disorders present dramatically, with somnolence, poor feeding, vomiting, lethargy, seizures, and even coma from hyperammonemia. In adults and older children, the presentation may be subtler and begin with chronic vomiting, developmental delay, seizures, psychiatric illness, postpartum decompensation, and hyperammonemia associated with valproate therapy, protein overconsumption, or increased catabolism.

Citrin deficiency, or neonatal-onset type II citrullinemia, is a relatively new genetic disorder caused by a homozygous or compound heterozygous mutation in the SLC25A13 gene. With several metabolic abnormalities, different clinical manifestations range from cholestasis, fatty liver, and growth retardation in infancy, to liver dysfunction and neuropsychiatric features in childhood and adulthood. Citrin deficiency is a self-limiting condition in almost all infants, and improving with age. Severe liver dysfunction requiring liver transplantation has been reported. Excessive intake of carbohydrate leads to various symptoms such as fatigue, anorexia, weight loss, neuropsychiatric symptoms, and liver failure, particularly in adult-onset type II citrullinemia. A low carbohydrate intake and a protein- and fat-rich diet with a protein–fat–carbohydrate ratio of 15–25%:40–50%:30–40% are recommended.

As a result of the impairment of NADH shuttling and glucose metabolism, patients with citrin deficiency have an aversion to carbohydrates and a unique preference for high-protein and high-fat foods, in contrast to patients with other urea cycle defects.

In hepatic porphyria, patients usually present neurologic, cutaneous, and gastrointestinal symptoms (especially acute intermittent porphyria) and mild elevation of transaminases.

Pregnancy and liver disease

Maternal health during pregnancy great affects fetal growth and development as well as long term outcomes for both the fetus and the mother. In general, pregnant women lack adequate knowledge on nutrition and would mostly benefit from consultation with trained nutrition practitioners. In pregnant women, due to estrogen effects, serum concentration of triglycerides, low-density and very-low-density lipoproteins, and cholesterol may be twice as high as that of nonpregnant women of the same age.

During pregnancy liver size and histology do not alter. The development of esophageal varices may occur in pregnant women with chronic liver disease as a result of an enlarging uterus creating an increase in venous return from the inferior vena cava to the azygous system. Liver diseases which can evolve as a consequence of pregnancy include intrahepatic cholestasis of pregnancy, acute fatty liver of pregnancy, and hemolysis, elevated liver enzymes, and low platelets syndrome (HELLP). Pregnant women with cirrhosis have a greatly increased likelihood of having protein-calorie malnutrition. NAFLD is associated with increased risk of gestational diabetes mellitus, preeclampsia, preterm birth, C-section delivery, and low birth weight. Management of NAFLD in pregnant women involves lifestyle modification. Weight loss and a Mediterranean diet have been shown to reduce risk of gestational diabetes mellitus.

Breastfeeding is known to be highly beneficial for the infant and can be done despite liver disease for most cases; women with hepatitis B or C infection can breastfeed their infants as long as the infant has received the appropriate prophylaxis and there is no

bleeding or cracks in the nipples. Additionally, breastfeeding has been shown to reduce the risk of the infant developing NASH or hepatic fibrosis from NAFLD later in life.

Acute liver failure

The nutritional status of someone with acute liver failure versus chronic liver failure can differ greatly. The primary goal of the nutritional management in acute liver failure is supportive. An increase in nausea, vomiting, and anorexia may be associated with acute liver disease which may result in a decreased intake of foods in general. If normal nutritional status before the insult is assumed, the patient will have a much higher nutritional reserve than that of a patient in chronic liver failure. Energy needs can be maintained by providing the Dietary Reference Intake for infants and children, and approximately 30 kcal kg^{-1} for adults. Fluid restriction to 70–90% of maintenance may be required if encephalopathy or cerebral edema is present. Hypoglycemia results from impaired gluconeogenesis and depleted glycogen stores. If fluids are provided intravenously, the glucose infusion rate may need to be increased with close monitoring of blood sugar levels.

The provision of adequate protein becomes crucial in fulminant hepatic failure and encephalopathy. Adequate protein must be provided to minimize catabolism which could exacerbate any hyperammonemia present. Excessive protein intake should be avoided as it could increase ammonia levels. Protein recommendations for adults and teenagers are 0.5–1.0 g kg^{-1} day^{-1}; for infants and children 1.2–1.5 g kg^{-1} day^{-1}. Additional protein restrictions or an increase in the intake of branched-chain amino acids intake may be beneficial. In health, the ratio of branched-chain amino acids/aromatic amino acids (leucine + isoleucine + valine/phenylalanine + tyrosine) is ~3:1 and in liver failure the ratio may decline to ~1:1, often in association with some degree of hepatic encephalopathy. There is some data to say that normalization of this ratio by administration of branched-chain amino acid formulae can improve hepatic encephalopathy.

Chronic liver disease

Chronic liver disease is often accompanied by nutritional deficiencies. The goals of nutritional management are to provide adequate energy and protein to prevent energy deficits and protein catabolism and to promote hepatic cell growth. Recommendations for nutritional management of children with chronic liver disease are presented in Fig. 6. Energy needs for adults with chronic liver disease are 30–35 kcal kg^{-1} day^{-1}. Energy requirements are increased to compensate for the weight loss that often occurs in cirrhosis. Protein should be provided as 0.8–1.0 g kg^{-1} for adults and the Dietary Reference Intake for protein should be provided for infants and children; unnecessary protein restriction should be avoided as it may only worsen total body protein losses. Energy from fat is best delivered as medium chain triglycerides due to the frequency of long chain fat malabsorption. Several infant, pediatric, and adult formulas are available with a large percentage of fat in the form of medium chain triglycerides.

Supplementation with fat-soluble vitamins (i.e., A, D, E, and K) in water-miscible solutions is necessary due to the potential for deficiencies associated with fat malabsorption. Serum levels should be monitored regularly to ensure appropriate levels and prevent toxicity. Supplementation with zinc, selenium, iron, and calcium should be given as needed. Copper and manganese should not be supplemented as they are excreted with bile and may build up to toxic levels. Sodium and fluid restrictions may be necessary in cirrhosis characterized with ascites and edema. This can impose difficulty as this restriction decreases the palatability of the diet, further decreasing oral intake.

End stage liver disease in pre- and post-liver transplantation

The prevalence of malnutrition in cirrhosis is as high as 65–90%, and malnutrition is predictive of survival in patients with liver cirrhosis. The severity of malnutrition is correlated with severity of liver disease and the development of serious complications such as hepatic encephalopathy, ascites, and hepatorenal syndrome. A variety of mechanisms are considered to contribute to malnutrition in cirrhosis (Table 4). Maintaining optimal nutritional status is important in the patient with end stage liver disease both pre- and post-transplant. However, nutritional assessment in end stage liver disease is particularly problematic. In the pre-transplant setting, fluid retention, ascites, and hepatosplenomegaly make body weight an unreliable nutritional index. True decreases in body weight, due to loss of fat stores and lean body mass may not be fully appreciated in solely following weight trends. In the pediatric population, linear growth is often a better indicator of nutritional status. Chronic malnutrition is often present as is reflected in a decrease in linear growth velocity.

Although anthropometric measurements, 24 h creatinine, bioelectric impedance analysis, and indirect calorimetry have all been used to assess nutrition, they are all affected by ascites and peripheral edema. *In vivo* neutron activation analysis and isotope dilution techniques are more accurate ways of assessing body composition but are time-consuming and costly. For practical purposes the indirect assessments of 24 h urinary creatinine excretion to assess body muscle mass and mid-arm muscle area can be used for patients without high volumes of extracellular fluid; in those with ascites the creatinine-height index is a better way of assessing body muscle mass.

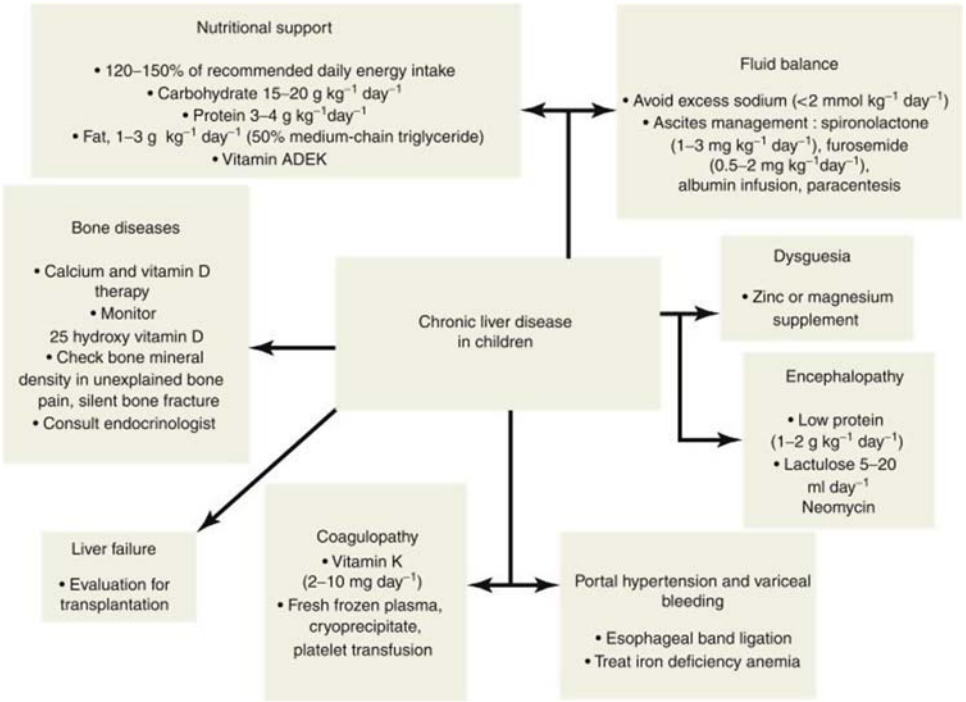


Fig. 6 Management in children with chronic liver disease.

Table 4 Factors and mechanisms that results in malnutrition in end stage liver diseases and treatment of the causes of malnutrition.

Factors that result in malnutrition in end stage liver diseases	*Causes*	*Management*
	Starvation.	Aggressive nutritional therapy.
Decreased dietary intake	Nil per os (NPO) status with frequent admissions to hospital.	Avoid NPO status unnecessary.
	Gastroesophageal reflux disease.	Appetite stimulant.
	Gastroparesis.	Prokinetic use.
Poor dietary intake from nausea and early satiety	Tense ascites.	Appropriate diuretic dosages.
	Small bowel dysmotility.	
	Small bowel bacterial overgrowth (SBBO).	SBBO treatment and prophylaxis with antibiotic (s).
	Sodium restriction to control ascites.	
A distortion or decrease in taste sensation (dysgeusia)	Trace element deficiencies (zinc, magnesium).	Supplement of trace elements.
Malabsorption	Atrophy of intestinal villi from starvation, prolonged NPO, gastroenteritis, SBBO, neomycin use. Reduced bile acid pool from cholestyramine for pruritus.	Treat causes.
Increased intestinal protein losses	Protein losing enteropathy from portal hypertension or hidden intestinal diseases (Celiac disease, intestinal lymphangiectasia).	Treat causes.
		Measuring the nitrogen balance.
Hypermetabolism	Increased sympathetic nervous system activity.	Adequate glucose or caloric intake to prevent muscle breakdown.
	Reduced glycogen storage hyperinsulinism sepsis.	Frequent meals. Treat infection.

Visceral proteins, including albumin, transferrin, prealbumin, and retinol binding protein are typically used in monitoring nutritional status due to the decrease seen in inadequate dietary protein intake, but should be used with caution in liver disease as the synthesis of these proteins also decreased in end stage liver disease. Serum levels of fat-soluble vitamins should be monitored closely as well.

Improving nutritional status before a transplant is imperative because malnutrition affects morbidity and mortality post-transplant. Although the degree of malnutrition may not be able to be reversed, aggressive nutritional support should be implemented to prevent further worsening of the nutritional state and possibly reduce pre- and post-transplant infection and complications. Reduced protein intakes, lactulose use, vegetable protein diets, zinc supplementation, and branched chain amino acid (BCAA)-enriched enteral supplements have been reported to reduce subclinical hepatic encephalopathy confirmed by psychometric test. Although treatment with bisphosphonates is recommended in those with osteoporosis, oral alendronate (not risedronate) may cause esophageal ulcer and could precipitate variceal bleeding.

Sarcopenia is defined by a reduction in muscle mass or muscle wasting and caused by decreased protein synthesis and increased protein degradation, especially in end-stage liver disease. It has been commonly neglected and associated with adverse clinical outcomes. This can be confirmed with dual-energy X-ray absorptiometry (DEXA), CT, MRI, and bioelectric impedance analysis (BIA) and requires rehabilitation strategies (Lai et al., 2021).

Post-transplant nutritional support should not be overlooked as the nutritional deficit is not cured merely by the transplant. Additionally, the surgery itself poses increased nutritional demand for post-surgery healing and support. Nutritional repletion may occur at a more rapid rate than pre-transplant as the patient now has a functional liver in which metabolism and digestion of macro and micronutrients will be improved. Long-chained fat and fat-soluble vitamins therefore should be normally absorbed. A side effect of corticosteroid use after liver transplant is growth hormone deficiency which should be suspected in children with persistent growth failure despite aggressive nutritional therapy. Post-transplant feeding should be started as soon as possible, ideally within 72 h. Muscle mass and fat storage rapidly improve within 3–6 months post-transplant. In children, marked catch-up growth is observed in those with stunted growth before transplantation. Most children catch up in weight to their peers by 1-year post-transplant; however, it may require up to 5 years to catch up in height. Of note, some conditions such as Alagille syndrome have underlying genetic abnormalities that limit growth regardless of liver function. Recombinant growth hormone therapy has successfully treated some of these children without deleterious effect to graft function. Post-transplant metabolic syndrome, fatty liver, and obesity are common complications after transplantation that may occur and should be monitored.

Conclusion

While the pathophysiology of the liver diseases discussed in this article are quite diverse, a key concept that unify them is the importance of proper nutritional management in the overall treatment of these diseases. Malnutrition in liver disease is associated with increased risks of morbidity and mortality, affects the outcomes of liver transplantation, and long-term survival, and quality of life. Careful assessment of the nutritional status by primary care takers or multidisciplinary team will initiate appropriate and early interventions. Therefore, It is essential to identify any nutrient deficiencies and specific nutritional needs which allow prompt preventive or corrective treatment.

References

Chalasani, N., Younossi, Z., Lavine, J.E., et al., 2018. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. Hepatology 67, 328–357.

Crabb, D.W., Im, G.Y., Szabo, G., Mellinger, J.L., Lucey, M.R., 2020. Diagnosis and treatment of alcohol-associated liver diseases: 2019 practice guidance from the American association for the study of liver diseases. Hepatology 71, 306–333.

Lai, J.C., Tandon, P., Bernal, W., et al., 2021. Malnutrition, frailty, and sarcopenia in patients with cirrhosis: 2021 practice guidance by the American association for the study of liver diseases. Hepatology 74, 1611–1644.

Tessitore, M., Sorrentino, E., Schiano Di Cola, G., et al., 2021. Malnutrition in pediatric chronic cholestatic disease: an up-to-date Overview. Nutrients 13, 2785.

Further reading

Bischoff, S.C., Bernal, W., Dasarathy, S., et al., 2020. ESPEN practical guideline: clinical nutrition in liver disease. Clin. Nutr. 39, 3533–3562.

De Meijer, V.E., Gura, K.M., Le, H.D., 2009. Fish oil-based lipid emulsions prevent and reverse parenteral nutrition-associated liver disease: the Boston experience. J. Parenter. Enteral Nutr. 33, 541–547.

European Association for the Study of the Liver, 2019. EASL Clinical Practice Guidelines on nutrition in chronic liver disease. J. Hepatol. 70, 172–193.

Plaza-Díaz, J., Solís-Urra, P., Rodríguez-Rodríguez, F., et al., 2020. The gut barrier, intestinal microbiota, and liver disease: molecular mechanisms and strategies to manage. Int. J. Mol. Sci. 21, 8351.

Sandhu, B.S., Sanyal, A.J., 2003. Pregnancy and liver disease. Gastroenterol. Clin. N. Am. 32, 407–436.

Stickel, F., Hoehn, B., Schuppan, D., Seitz, H.K., 2003. Review article: nutritional therapy in alcoholic liver disease. Aliment. Pharmacol. Ther. 18, 357–373.

Lung diseases

S.A. Unger, MRC International Nutrition Group, Medical Research Council Laboratories, The Gambia

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of S.A. Unger, Lung Diseases, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 111–123, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00175-6.

Introduction **464**
Chronic Obstructive Pulmonary Disease **464**
Epidemiology 464
Etiology 464
Clinical Features 464
Differential Diagnoses 464
Pathology 465
Nutrition and COPD 465
Reasons for Malnutrition in COPD 466
Obesity in COPD 466
Nutritional Support 466
Type of Nutritional Support 467
Tube Feeding 467
Cystic Fibrosis **468**
Definition 468
Epidemiology 468
Pathogenesis 468
Clinical Features 468
Malnutrition and CF 469
Decreased Intake 470
Increased Energy Expenditure 470
Nutritional Support 470
High-Energy/High-Protein Diet 470
Dietary Supplements 470
Enteral Feeding 470
Parenteral Nutrition 471
Vitamin, Mineral, and Pancreatic Enzyme Supplementation 471
Appetite Stimulants, Growth Hormones, and Omega-3 472
Asthma **473**
Definition 473
Epidemiology 473
Clinical Features 473
Pathogenesis 474
Nutrition and Asthma 474
Possible Preventative Effects of Nutrition 474
Secondary Prophylactic Effects of Nutrition in Asthma 475
Other Lung Diseases **475**
Obstructive Sleep Apnea Syndrome 475
Bronchiectasis 475
Chronic Lung Disease of Infancy 475
Respiratory Tract Infections and Micronutrients 476
Further Reading **476**

Glossary

Fat-free mass index Index of muscle mass in relation to height.
Forced expiratory volume Volume exhaled with force usually measured after 1 s.
Resting energy expenditure (REE) Amount of energy needed to maintain normal body functions at rest, excluding the thermal effect of food and physical activity.

Spirometry A pulmonary function test measuring the amount (volume) and speed (flow) of air that can be inhaled and exhaled.

Introduction

Adequate nutrition is vital for the development and continued health of our lungs. Evidence suggests that prenatal and early-life nutrition affect lung function in later life. Malnutrition is observed in several chronic lung conditions, such as chronic obstructive pulmonary disease (COPD) and cystic fibrosis (CF), where there is both an increased energy expenditure and a reduced intake leading to reduced lung function, increased morbidity, and decreased quality of life. Malnutrition with associated multiple micronutrient deficiencies increases the risk of acute respiratory infections. Conversely, obesity may compromise lung function in conditions such as asthma. This article gives a taste of the complex interaction between nutrition and lung disease and presents some of the evidence base for nutritional support.

Chronic Obstructive Pulmonary Disease

Chronic obstructive pulmonary disease is defined as a preventable and treatable disease characterized by chronic airflow obstruction that is not fully reversible. The impairment of lung function is usually progressive and associated with an abnormal inflammatory response of the lung to noxious particles or gases. The recent inclusion of 'preventable and treatable' in the definition represents a shift toward a more positive attitude.

COPD includes 'chronic bronchitis' defined by chronic bronchial secretions sufficient to cause expectoration occurring on most days for a minimum of 3 months for 2 consecutive years, and 'emphysema,' the pathological process of permanent destructive enlargement of the airspace distal to the terminal bronchioles without obvious fibrosis. Although pure forms of these two conditions exist, there is considerable overlap in the majority of patients, hence grouping them under COPD.

Epidemiology

The Global Burden of Disease Study has projected that COPD, ranking sixth as the cause of death in 1990, will become the third leading cause of death worldwide by 2020. Studies around the world estimate that the prevalence of COPD ranges from 7% to 19% and the disease is increasing in women. In developed countries, exacerbations of COPD present the greatest burden on the healthcare systems, accounting for 10% of all hospital medical admissions in the United Kingdom and direct costs of COPD are estimated to be 38.6 billion of the European and $18 billion of the United States health care budgets.

Etiology

The single most important risk factor for COPD is cigarette smoking with a direct correlation between number of cigarettes smoked and the likelihood of developing the disease. However, nonsmokers also develop COPD (never-smokers comprise approximately 23.3% of those classified as stage II+) and the individual susceptibility to smoking is very wide with only 15% of smokers likely to develop clinically significant COPD. Other risk factors include occupational exposure to dust and chemicals, air pollution, particularly in women in developing countries, and severe hereditary α_1-antitrypsin deficiency (**Figure 1**). Severe childhood respiratory infections have been associated with reduced lung function and increased respiratory symptoms in adulthood, which may, however, in turn be related to factors like low birth weight.

Clinical Features

Clinical features include repeated attacks of productive cough, progressive dyspnea, exertional breathlessness, recurrent infections, wheeze, and occasional chest tightness. Clinical respiratory examination may be normal in mild-to-moderate cases. In severe cases, clinical signs reflect pulmonary hyperinflation, hypoxemia, and the development of pulmonary hypertension and right heart failure (*cor pulmonale*) and polycythemia. Clinical classification according to spirometry results is widely used (**Table 1**).

Differential Diagnoses

Asthma can usually be distinguished from COPD based on reversible rather than irreversible airflow limitation. In the developing world, pulmonary tuberculosis and COPD are common and making the correct diagnosis is essential for management.

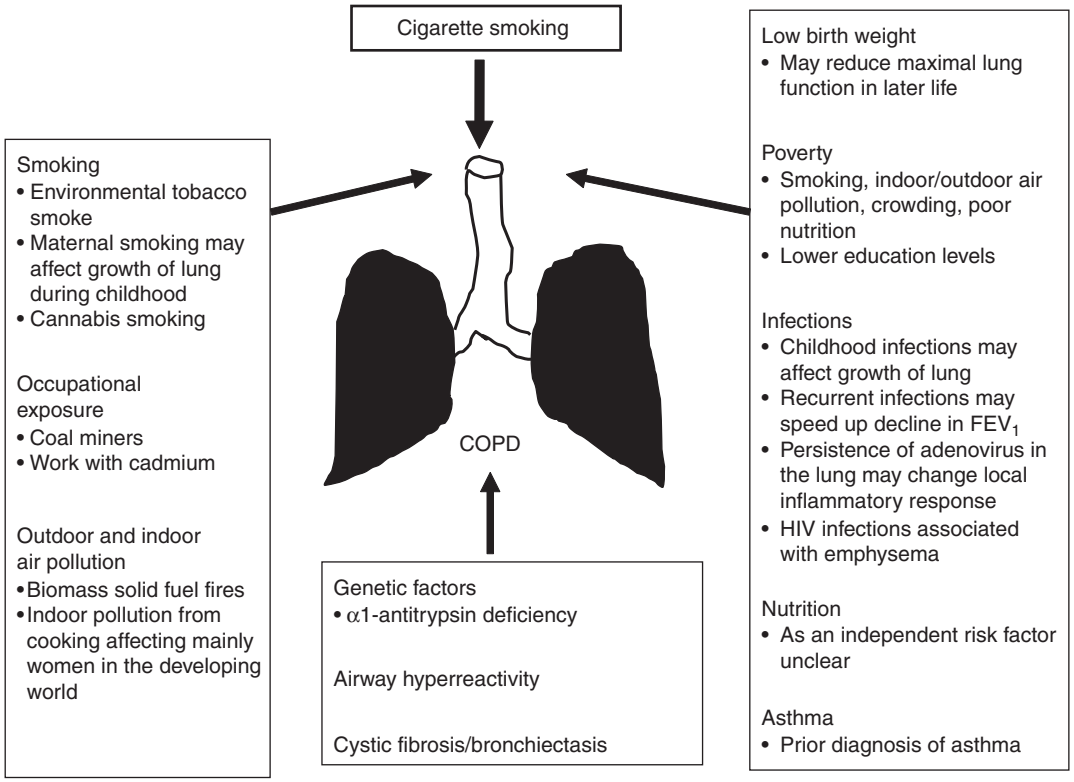


Figure 1 Risk factors for COPD. FEV_1, forced expiratory volume in 1 s.

Table 1 Spirometric classification of COPD

Stage	*Description*	*Definition*
I	Mild	$FEV_1 \geq 80\%$ predicted
II	Moderate	$50\% \leq FEV_1 < 80\%$ predicted
III	Severe	$30\% \leq FEV_1 < 50\%$ predicted
IV	Very severe	$FEV_1 < 30\%$ predicted or $FEV_1 < 50\%$ predicted plus chronic respiratory failure

Pathology

Pathological features include chronic air wall inflammation with airway smooth muscle hyperplasia and bronchial wall thickening, hypertrophy of mucus secreting glands, and a decrease in ciliated cells causing a less efficient transport of mucus in the airways. Small airways become obstructed; alveolar attachments and pulmonary elastic recoil are lost causing restriction of airflow. Emphysema is usually centriacinar with distension and damage affecting the respiratory bronchioles, alveolar ducts, and centrally located alveoli. More rarely, panacinar emphysema or paraseptal emphysema develops in the distal airway structures causing later blebs on the lung surface, giant bullae, or both.

Nutrition and COPD

Management of COPD needs a multicomponent approach with particular attention to existing comorbidities. Indices of airflow obstruction only poorly predict prognosis although used in classification. Body mass index (BMI), airflow obstruction, dyspnea, and exercise capacity (BODE) index has been proposed to account for the multiple components of COPD. Malnutrition represents an important clinical problem in a subpopulation with COPD and poor indices of nutrition such as low BMI and low fat-free mass index (FFMI) independently confer a poorer prognosis in patients with COPD in terms of mortality, risk of hospitalization, length of hospitalization, and health-related quality of life.

Between 24% and 71% of all COPD patients show some evidence of malnutrition. However, the percentage is higher in those with more severe disease and those in need of hospital admission. BMI is an indicator of poor prognosis, but even patients with a normal BMI may be undernourished. FFMI is a better marker of lean body mass compared to BMI, as it is associated with other

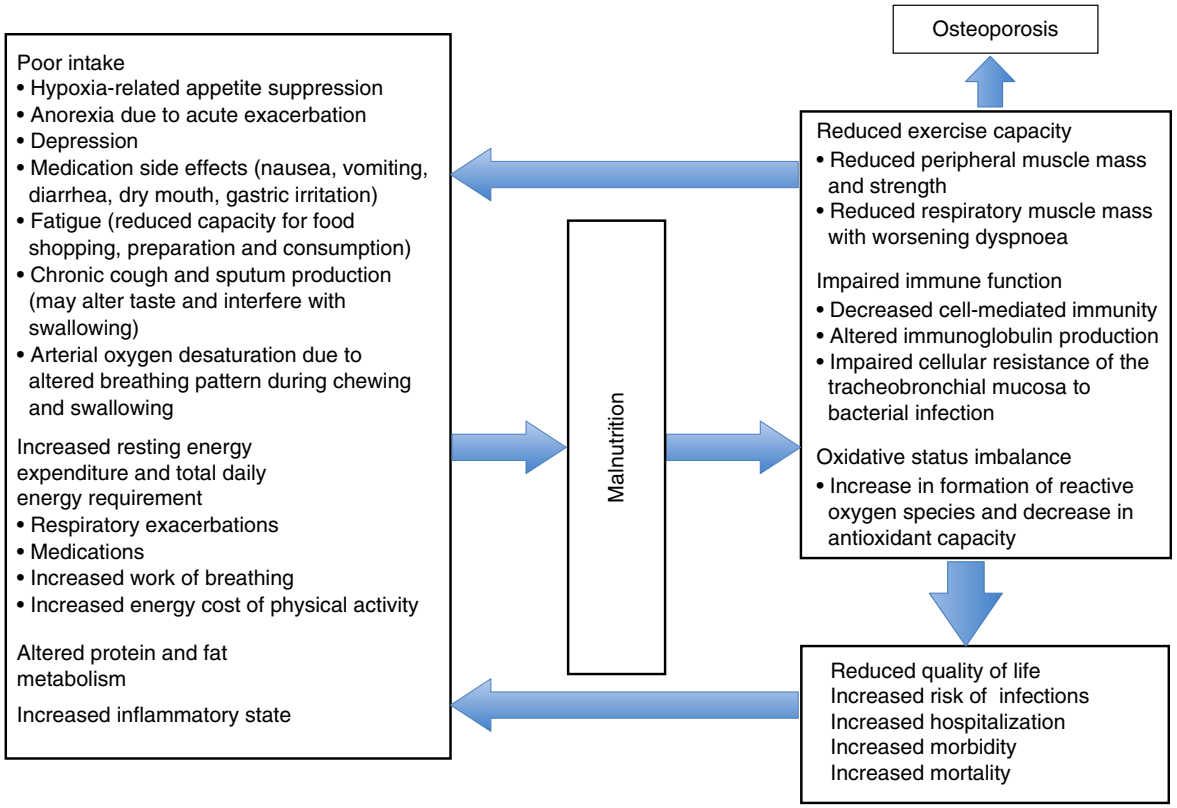


Figure 2 Malnutrition and COPD.

prognostic indices such as exercise capacity, dyspnea, and percentage of predicted FEV_1 (FEV_1=forced expiratory volume in 1 s). In up to 25% of COPD patients with normal weight, depletion of FFM can be noted. In malnourished patients, there may be loss of respiratory muscle and diaphragm mass as well as altered regulation of the oxidative phenotype and mitochondrial dysfunction affecting respiratory muscle strength and endurance. However, this does not imply a causal relationship; dyspnea may in turn prevent patients from exercising, contributing to muscle atrophy.

Reasons for Malnutrition in COPD

The mechanisms of weight loss in COPD are not fully understood. It may be the result of an imbalance between an elevated resting energy expenditure (REE), an elevated total daily energy requirement, and inadequate dietary intake (**Figure 2**). This imbalance may be further affected by altered protein metabolism suggested by the selective wasting of FFM. Decreased protein intake, particularly during the first days of acute exacerbation, decreased protein synthesis, and increased protein balance turnover have been reported. An increased systemic inflammatory state with elevated proinflammatory markers such as TNF-α may also adversely affect protein metabolism.

Obesity in COPD

Some COPD patients are obese, resulting in an increased energy cost for physical activity, and extreme obesity decreases lung function. A recent study showed higher survival rates among obese COPD patients (BMI>25) who lost weight. Another study showed that in those patients with a normal BMI or above, a high-carbohydrate diet may be of benefit to promote strength and function.

Nutritional Support

A 2008 Cochrane meta-analysis of 14 studies on simple nutritional supplementation in stable COPD patients did not identify improvements in anthropometric measures or functional exercise capacity in treated subjects versus control subjects. However, most studies did not use lean body mass as a primary outcome or health-related quality of life, which may be of more importance in COPD. In those studies showing any improvement in functional exercise, the improvements were not maintained when supplementation was discontinued. Nutritional supplementation as part of a multicomponent intervention was not reviewed.

More recently, it has been shown that a combination of nutritional supplementation with low-intensity exercise training increased weight and energy intake as well as exercise capacity and decreased REE and inflammatory cytokines in moderate to severe clinically stable malnourished COPD patients. However, further larger studies are required to examine the potential role of the combination of nutritional supplementation and exercise in the management of malnourished patients with COPD.

Initial examination of the combination of nutritional support and substances like creatine, L-carnitine, and growth hormone releasing factor has shown an increase in lean body mass but has not shown a consistent improvement in endurance or health-related quality of life. Grehlin stimulates growth hormone secretion, food intake, and weight gain without increasing adiposity unlike other appetite stimulants. It decreases muscle wasting via inhibition of production of anorectic proinflammatory cytokines. Appetite stimulants may, however, exacerbated by steroid medication, cause hyperglycemia.

A higher antioxidant food intake may improve lung function in COPD patients and may also have protective effects on the development of COPD. Studies have also shown a high prevalence of vitamin D deficiency among COPD patients, adversely affecting lung function. Supplementation trials of antioxidants and vitamin D are required before recommendations for routine supplementation can be made.

Type of Nutritional Support

Nutritional Advice, Exercise, and Supplementation

Despite the findings of the 2008 Cochrane review, the general logical advice is that nutritional support should be considered in all patients with COPD and it should be included in multidisciplinary pulmonary rehabilitation programs, which have shown some benefits. Individual dietary advice based on a dietary history and an analysis of nutritional status should be arranged for those who are underweight (BMI<18.5) or obese (BMI>30), and those whose weight is changing over time (weight loss greater than 10% in the past 6 months or more than 5% in the last month) (**Figure 3**). Furthermore, if the BMI is low, high-energy, high-protein nutritional supplements should be given in frequent, small amounts (to avoid postprandial dyspnea and satiety and to improve compliance) to increase the total calorific intake. Exercise should be encouraged to augment the effects of nutritional supplementation.

Tube Feeding

Enteral tube feeding should not be used in COPD patients unless oral intake is unsafe or oral methods of maintaining nutritional status have failed. One should be aware of the risks (**Table 4**). Although a high-carbohydrate diet produces more carbon dioxide (VCO_2) and therefore requires increased ventilation to expel the excess CO_2, there is no additional advantage in stable COPD patients of formulating low-carbohydrate, high-fat enteral feeds compared to standard high-protein, high-energy supplements.

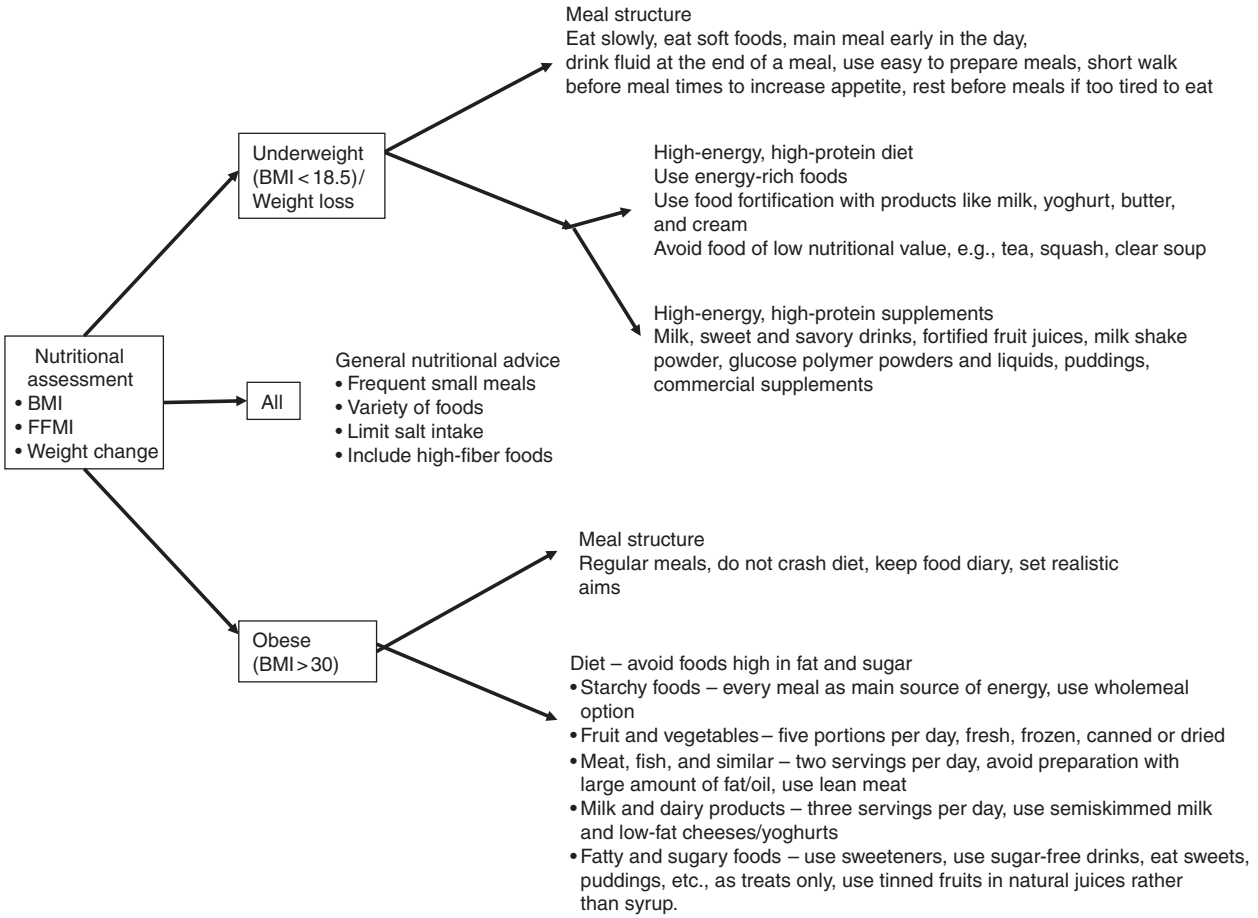


Figure 3 Nutritional advice in COPD.

However, in patients with acute respiratory problems requiring artificial ventilation, the composition of feeds has a profound effect on gas exchanges. Feeding formulas during ventilation should have low-carbohydrate, high-fat content to reduce VCO_2 and ventilation requirements. Overfeeding negates any beneficial response to high-fat feeds because the conversion of energy into fat involves disproportionately large production of CO_2. Bolus feeds during ventilation are as effective as continuous feeds.

Cystic Fibrosis

Definition

Cystic fibrosis is an autosomal recessive genetic disorder resulting from a mutation of a gene located on chromosome 7q31.3, which codes for a cyclic-AMP-activated chloride channel known as the cystic fibrosis transmembrane conductance regulator (CFTR) expressed in several epithelia. Over 1200 different mutations and 200 polymorphisms have been identified. Mutation ΔF508 accounts for 77% of CF chromosomes in the white population, whereas 3210+1G>A accounts for 46% of mutations in the black population.

Epidemiology

CF is the most common fatal genetic condition in Caucasians with a carrier rate of 1 in 25 and an incidence of 1 in 2500 live births. It is less common in the non-white races (1 in 20 000 and 1 in 1 million live births in the black and Oriental populations, respectively), but little is known about its prevalence in developing countries as CF often remains underdiagnosed. Early diagnosis with neonatal screening, aggressive treatment of lung infections, and nutritional support has led to a dramatic improvement in life expectancy. Now, there are more adults than children with CF in many developed countries.

Pathogenesis

The CFTR mutations lead to dysfunction of the exocrine glands in multiple systems including the skin, gut epithelium, pancreas, liver, and reproductive tract; in the respiratory epithelium, it results in increased resorption of sodium and water. Relative dehydration of the airway lining may predispose to chronic bacterial infection and ciliary dysfunction, leading to bronchiectasis and reduction of lung function.

Clinical Features

Most children, if not already identified by neonatal screening, present with malabsorption and failure to thrive as well as recurrent chest infections. Lung disease is the primary cause of morbidity and mortality in CF patients. At birth, the lungs appear normal macroscopically, although studies have shown the presence of an active inflammatory process. Progressive cycles of infection

Table 2 Complications of cystic fibrosis

Respiratory	
• Bronchiectasis	• Respiratory failure
• Pneumothorax	• *Cor pulmonale*
• Wheeze	• Lobar collapse due to secretions
• Hemoptysis	• Allergic bronchopulmonary aspergillosis
• Nasal polyps	
Gastrointestinal	
• Meconium ileus	• Malabsorption and steatorrhea
• Rectal prolapse	• Distal intestinal obstruction syndrome
• Intussusception	• Gastroesophageal reflux
• Abdominal distension	• Biliary cirrhosis
• Colonica strictures	• Hepatomegaly
• Cholelithiasis	• Portal hypertension
• Obstructive jaundice	• Cholecystitis
• Pancreatitis	
Others	
• Diabetes	• Salt depletion
• Male infertility	• Growth failure/weight loss/failure to thrive
• Amyloidosis	• Psychosocial problems
• Arthropathy	• Osteopenia/osteoporosis
• Delayed puberty	• Stress incontinence (repeated cough)
• Cutaneous vasculitis	

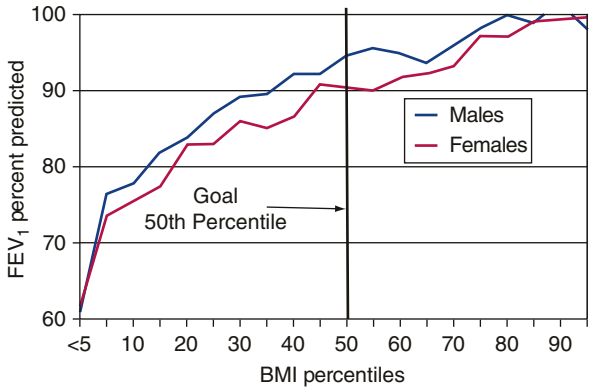


Figure 4 FEV_1 percent predicted vs. BMI percentile in children 6–20 years of age. Reproduced with permission from Cystic Fibrosis Foundation Patient Registry (2009) *2008 Annual Data Report.* Bethesda, Maryland: Cystic Fibrosis Foundation Patient Registry.

and inflammation lead to bronchiectasis in childhood, and progressive lung damage eventually leads to respiratory failure and death. Initially, the lungs are infected with *Staphylococcus aureus* and nontypable *Haemophilus influenzae*, but later approximately 80% are infected with *Pseudomonas aeruginosa*. Approximately 90% of CF patients have pancreatic insufficiency requiring pancreatic enzyme replacement therapy (PERT). Other features are summarized in **Table 2**.

Malnutrition and CF

Malnutrition in CF patients is still one of the main clinical manifestations despite increased knowledge and improved management. Of adults with CF, 60.8% have a BMI below the recommended levels and 15.7% of patients under 20 are below the fifth centile weight-for-age and gender.

A clear link has been shown between good nutrition and prognosis of CF. BMI percentiles in children and BMI values in adults are also directly and strongly correlated with pulmonary function in terms of FEV_1 (**Figure 4**).

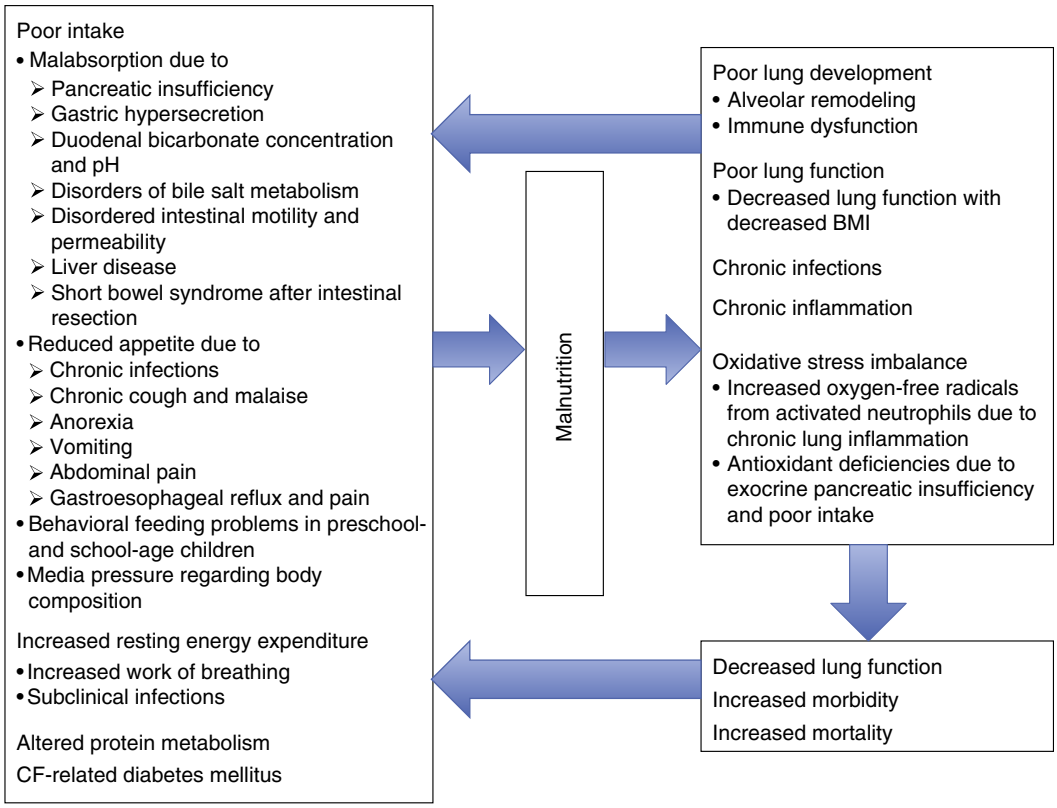


Figure 5 Cystic fibrosis and malnutrition.

Furthermore, based on multiple longitudinal and cross-sectional studies, there is an assertion that growth, particularly linear growth, is connected to the evolution of lung health in children who have CF. However, the causal relationship between malnutrition and pulmonary dysfunction remains unclear in CF. There is a suggestion that nutritional status early in life is a determinant for the progression of the lung disease. Studies in infants looking at later lung disease are obviously difficult in terms of ethics but retrospective studies have shown that appropriate nutritional status early in life has a positive effect on lung function later on, suggesting a causal relationship. Malnutrition may affect mechanical lung properties including alveolar remodeling in addition to factors including the activation of cytotoxic T lymphocytes and natural killer cells.

A variety of complex organic and psychosocial factors contribute to malnutrition in CF, which in turn worsens pulmonary dysfunction (Figure 5). Similar to COPD, there is an imbalance of decreased nutrient intake and increased energy expenditure in CF.

Decreased Intake

Pancreatic insufficiency and other complications may contribute to moderate and severe malabsorption affecting protein and fat-soluble vitamins despite adequate use of enzyme supplements (Figure 5). Reduced appetite due to chronic respiratory infections and other complications, as well as psychosocial factors such as eating disorders in teenagers, further lead to inadequate energy intake in CF patients.

Increased Energy Expenditure

The REE in severe obstructive lung disease such as CF and COPD is increased partly because of the increased work of breathing. REE is 10–20% greater in CF patients than in healthy controls and this increase appears to be closely associated with declining pulmonary function and subclinical infection. Bronchial sepsis leads to local release of leukotrienes, free oxygen radical, and cytokines, including TNF-α. Interestingly, antibiotics reduce energy requirements of moderately ill patients with chronic *P. aeruginosa*. Furthermore, the presence of even mild lung disease is associated with elevations in REE, indicating that REE could be a sensitive marker of clinical status before lung disease becomes clinically overt. The relationship between protein metabolism and pulmonary function in CF remains unclear.

Nutritional Support

Nutritional support should be based on complete individual nutritional assessment rather than a generalized approach. Current guidelines advise using weight-for-length in less than 2-year-olds, BMI percentiles for 2- to 18-year-olds, and BMI in the adult population (Table 3). The age of greatest risk of malnutrition appears to be during the first 2 years of life and during early adolescent years. Treatment for nutritional failure may include nutritional advice and behavioral interventions, oral supplements, enteral feeds, and parenteral nutrition (Table 3).

High-Energy/High-Protein Diet

The encouragement of a high-energy, high-protein diet aiming at 110–200% of recommended daily allowance will produce growth in the majority of children and adults with CF (Figure 6). Malnourished children achieve higher energy intake when more frequent meals are offered. Attention should be given to psychological, social, behavioral, and developmental aspects of feeding. There is evidence that particularly for children 1–12 years of age, behavioral intervention and nutrition counseling should be implemented when a risk of poor growth is present (Table 3).

Dietary Supplements

For children with growth deficits, adults with inappropriate weight status, and during acute chest infections, oral and enteral supplements should be used to improve weight. Dietary calorie supplements should complement normal food intake and not replace food, following recommended age-dependent quantities (Table 3). A Cochrane review concluded that oral calorie supplements, on top of standard dietary advice and monitoring, do not achieve any additional benefit in the nutritional management of moderately malnourished children with CF in addition to dietary advice and monitoring.

Enteral Feeding

When oral high-calorie diets and supplements are ineffective, enteral tube feeding may be considered, which is better tolerated by gastrostomy than by nasogastric tube in the long term (Table 4). Reported use of enteral tube feeding suggests that it results in nutritional and respiratory improvements. However, a Cochrane review concluded that its efficacy has not been fully evaluated and data are limited and randomized trials are needed to investigate efficacy as well as when to start enteral tube feeds. Furthermore, these interventions are not free from complications and should be balanced against the possible gains (Table 4). Enteral feeding may lead to gastroesophageal reflux, formula intolerance, and hyperglycemia. Patients receiving supplemental feeds who demonstrate repeated blood sugar levels higher than 11.1 mmol l^{-1} during the feed may benefit from insulin given before the feeds. Enteral feeds are usually given for 8–10 h overnight, with at least 40–50% of the estimated energy requirement given via the feed. Most patients tolerate an energy-dense polymeric feed providing at least 1.5 kcal ml^{-1} with additional pancreatic enzymes.

Table 3 Indications for nutritional interventions in cystic fibrosis

Age group	*0–2 years*	*2–18 years*	*>18 years*
Nutritional assessment indicator	Weight-for-length percentiles	BMI percentiles	BMI
Care categories			
Normal/routine nutritional care	Normal weight-for-length percentiles and within 2 centile bands of each other	BMI 25–95th percentile	BMI 20–27
At risk/noninvasive nutritional intervention	Weight and height percentiles decreasing with time or no weight gain	BMI 10–25th percentile OR Weight loss over 1–3 months OR Plateau in weight gain over 2–4 months	BMI 18.5–20 OR Weight<45 kg regardless of BMI OR 5% weight loss over 2 months
Malnutrition/aggressive nutritional support	Weight 2 or more centile bands below length OR Failure of noninvasive interventions to improve nutritional status	BMI<10th percentile OR Weight falling 2 or more percentile positions OR Plateau in weight gain for 6 months OR Failure of noninvasive interventions to improve nutritional status	BMI <18.5 OR Weight<40 kg regardless of BMI OR 5% weight loss over 2 months despite noninvasive nutritional interventions
Overweight/nutritional advice	N/A	BMI>95th percentile	BMI>27
Nutritional interventions			
Advice for routine energy intake as indicated for age and sex	Demand breast-feeding or whey-based artificial formulae. Intake of >200 ml kg^{-1} is not unusual. Early weaning not advised	110–200% standards for healthy population	110–200% standards for healthy population
Combined behavioral and nutritional intervention indicated for weight gain	Recommended for children aged 1–2 years	Recommended for age 2–12 years; insufficient evidence for age 13–18 years	Insufficient evidence for effectiveness for adults
Nutritional supplementation (oral and enteral) intervention if indicated for weight gain	Unrestricted demand feeds. Nutritional supplementation if poor growth or surgery for meconium ileus	Recommended	Recommended
Caloric supplementation (per day)	1–2 years 200 kcal	3–5 years 400 kcal 6–11 years 600 kcal >12 years 800 kcal	800 kcal
Pancreatic enzyme preparations (based on pancreatic status assessment)	Recommended Upper limit: 10 000 IU lipase kg^{-1} body weight day^{-1}	Recommended Upper limit: 10 000 IU lipase kg^{-1} body weight day^{-1}	Recommended Upper limit: 10 000 IU lipase kg^{-1} body weight day^{-1}
Vitamin Supplementation (fat-soluble vitamins A, D, and E)	Recommended in pancreatic insufficient patients	Recommended in pancreatic insufficient patients	Recommended in pancreatic insufficient patients
Vitamin K	Recommended in liver disease and with prolonged prothrombin time	Recommended in liver disease and with prolonged prothrombin time	Recommended in liver disease and with prolonged prothrombin time
Water-soluble vitamins, iron, and calcium supplements	Recommended if deficient or evidence of low intake	Recommended if deficient or evidence of low intake	Recommended if deficient or evidence of low intake
Sodium supplements (only in hot conditions and evidence of sodium deficiency)	Recommended <1 year 500 mg day^{-1}	Recommended 1–7 years 1 g day^{-1}, >7 years 2–4 g day^{-1}	Recommended 2–6 g day^{-1}

Parenteral Nutrition

Parenteral nutrition should be reserved for those with a nonfunctioning gastrointestinal tract (e.g., due to prolonged bowel obstruction or gastrointestinal surgery) and the critically ill as it is costly and associated with several risks, including catheter sepsis and metabolic complications such as hyperglycemia.

Vitamin, Mineral, and Pancreatic Enzyme Supplementation

Malabsorption and pancreatic insufficiency are treated with oral PERT and vitamins (for details, see Table 3). Low fat-soluble vitamin concentrations are associated with poorer clinical status and reduced lung function. In hot conditions and in proven deficiency, sodium supplements are recommended. Anorexia and poor growth may result from chronic salt depletion, and significant

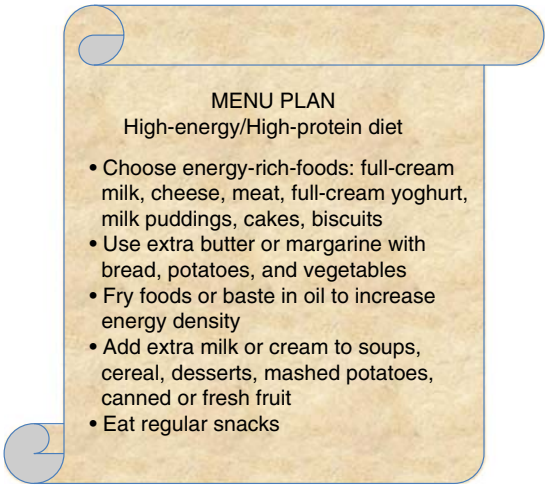


Figure 6 Menu plan for a high-calorie/high-protein diet.

Table 4 Advantages and disadvantages of different enteral feeding routes

Method	*Advantages*	*Disadvantages*
Nasogastric	• Short-term feeding	• Tube reinsertion may be ○ distressing to patient/caregiver/nurse ○ easily removed • Risk of aspiration • Psychosocial implications
Nasojejunal	• Less risk of aspiration • Short-term feeding	• Difficulty of insertion • Radiographic check of position • Easily removed • Risk of perforation • Abdominal pain and diarrhea unless continuous infusion of feed • Discomfort in nasopharynx • Reflux of bile is facilitated
Gastrostomy	• Cosmetically more acceptable • Long-term feeding	• Increase reflux if present • Local skin irritation • Stoma infection • Granulation tissue • Leakage • Gastric distension • Stoma closes within a few hours if accidentally removed
Jejunostomy	• Reduced risk of aspiration • Long-term feeding	• Surgical/radiology procedure • Risk of perforation • Must be constant infusion of feed • Bacterial overgrowth • Dumping syndrome can occur

hyponatremia may be accompanied by vomiting. Pancreatic enzymes should be given at the smallest dose possible to control steatorrhea and achieve normal patterns of growth and weight gain. Total fat excretion should be quantified in order to assess efficacy of PERT.

Appetite Stimulants, Growth Hormones, and Omega-3

Small randomized controlled trials of the appetite stimulant megestrol acetate noted significant improvement in weight and pulmonary function during the treatment period but gains seem to be short-lived. A Cochrane review is under way to examine the use of appetite stimulants in the management of CF. Important side effects include glucose intolerance and adrenal suppression. Cyproheptadine, another appetite stimulant, may be more promising in terms of side effect profile and long-term effect on weight gain.

Small trials with growth hormone given for 1 year showed improvements in respiratory status and exercise capacity and gains in lean body mass, but larger trials are required to make recommendations. Similarly, although small studies suggested an

improvement in lung function and clinical status, there is not enough evidence on the safety and effectiveness of fish oil supplements in the routine care of CF.

Asthma

Definition

Asthma (ancient Greek for 'panting') has no universally agreed definition. It is a heterogeneous disease described in terms of clinical, physiological, and pathological characteristics. Chronic airway inflammation and increased airway hyperresponsiveness to various stimuli lead to paroxysmal airway narrowing that may be reversible with or without treatment.

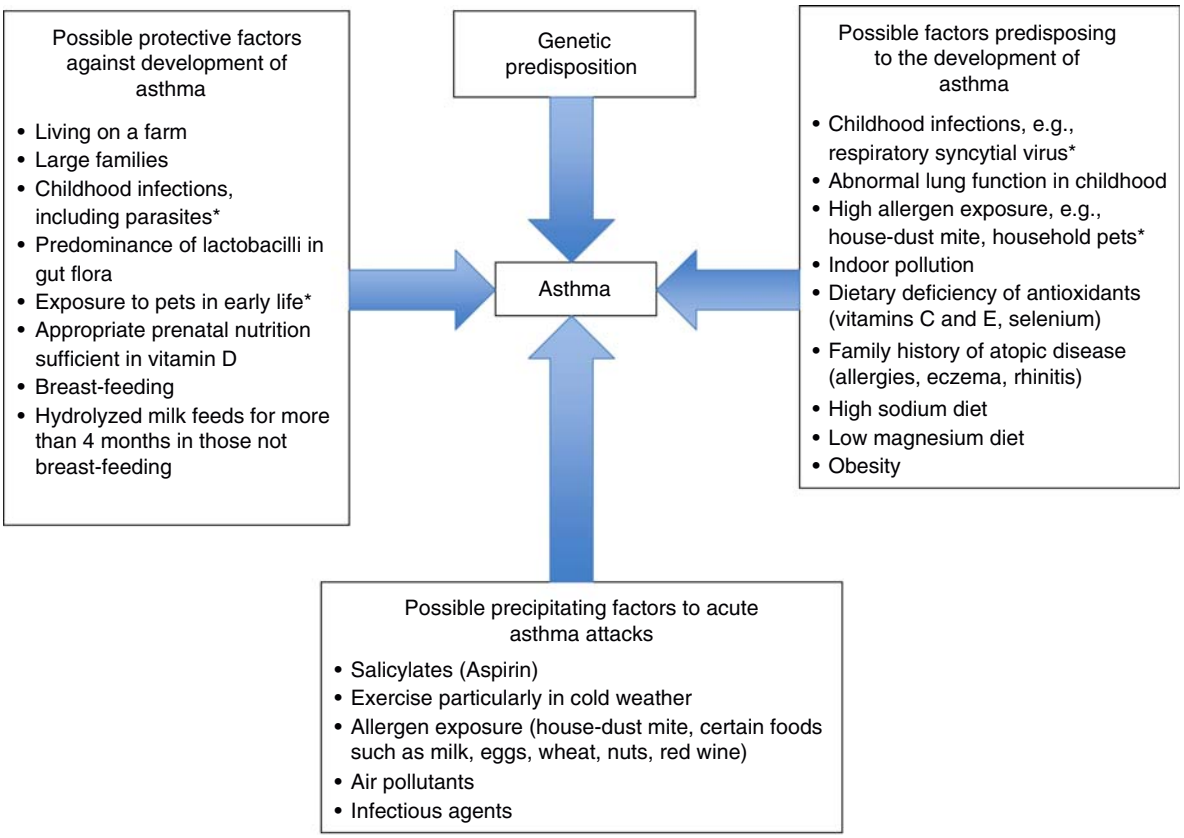
Epidemiology

Since the 1980s, there has been a worldwide increase in the prevalence of asthma in both children and adults with associated increases in morbidity and mortality. It currently affects 300 million people worldwide, and an additional 100 million persons will be diagnosed by 2025, with huge socioeconomic implications in terms of days lost from school or work, health care visits, etc. For most of the developing world, little or no standardized data are available.

A series of environmental factors and genetic predisposition have been implicated in the development of, or protection from, asthma (Figure 7). Of people with asthma, 50% develop the condition before age 10, and the majority before age 30. The earlier the onset of wheeze, the better the prognosis for disease progression. Boys are more likely to 'grow out' of their asthma.

Clinical Features

Central to all definitions of asthma is the presence of more than one of the following symptoms: recurrent episodes of wheeze, breathlessness, chest tightness, and cough. Precipitants include exercise in cold weather and viral upper respiratory infections (Figure 7). There is huge variability in symptom presentation between patients, ranging from those with periodic wheezing attacks with asymptomatic periods between exacerbations and patients with persistent, but fluctuating symptoms of breathlessness and wheeze. There is usually a diurnal pattern, with symptoms and lung function being worse early in the morning.



*There is controversy and conflictin evidence re ardin prenatal and early-life exposures.

Figure 7 Factors implicated in the development of, or protection from, asthma and trigger factors for acute asthma attacks.

Table 5 Nutritional effects on the development of asthma (primary preventative effects) and symptom control (secondary prophylactic effects)

Nutritional intervention	*Evidence*	*Comments*
Primary preventative effects		
Maternal food allergen avoidance during pregnancy and lactation	• Not related to subsequent development of asthma	• May adversely affect maternal and fetal nutrition • High-dose exposure may reduce subsequent sensitization
Early introduction of allergenic foods during weaning	• Not related to subsequent development of asthma	• Small study showed a nonsignificant increase in preschool wheezing with late introduction of egg
Fish oil supplementation during pregnancy	• Marginal effects on the reduction in wheeze and cough at 1 year	• Reduction in cytokine response to allergens in cord blood • No adverse effects observed
Fish oil supplementation in early infancy	• Reduces wheeze at 18 months but no effect on asthma by 5 years of age	• No adverse effects observed
Breast-feeding	• Potential protective effect in relation to early asthma	• In line with WHO recommendation
Hydrolyzed formula	• Suggested reduction of risk of asthma or wheeze in the first year of life	• Formula recommended when breast-feeding is not possible and the infant is at high risk
Low intake of antioxidants	• Possible association with higher prevalence of asthma	• Disputed by recent meta-analysis
Higher intake of fresh fruit and vegetable	• Improved pulmonary function and lower prevalence of asthma	• In line with healthy diet promotion for the prevention of cancer and cardiovascular disease
Secondary prophylactic effects		
Supplementation with antioxidants (vitamins C, E, selenium)	• Limited evidence of clinical benefits	• Selenium supplementation showed improvements in terms of 'clinical evaluation' but not in objective parameters of lung function and airway hyperresponsiveness
Supplementation with or diet high in fish oil	• No significant improvement in asthma control	• No known adverse effects
Probiotics	• No effect on clinical parameters	• Decrease in eosinophilia
Low sodium intake	• Minimal clinical improvements	• High sodium intake associated with increased bronchial hyperresponsiveness
Low magnesium intake	• Associated with higher prevalence of asthma	• Magnesium used as muscle relaxant in acute asthma attacks
Weight reduction in obese asthma patients	• Improved asthma control	• Limited evidence of a strict calorie-controlled diet
Food additive avoidance	• No effect on asthma control	• Restrictive diet seldom used in asthma

Note: The evidence is largely based on observational studies that are often multifaceted making it difficult to differentiate the effects of one exposure or intervention from another.

Pathogenesis

Asthma represents a spectrum of pathophysiologic processes involving different degrees of airway inflammation and remodeling. At the mild end of the spectrum, there is allergic asthma mainly driven by Th2-mediated inflammatory responses and associated with other allergic comorbidities such as rhinitis and atopic dermatitis. In moderate disease, altered epithelial–mesenchymal communication is observed, leading to the generation of growth factors and cytokines with subsequent sustained inflammation. When the epithelial–mesenchymal component becomes dominant in more severe disease, tissue damage and remodeling can be observed as well as increased mucus secretion with obstruction of peripheral airways.

Nutrition and Asthma

The relationship between nutrition and asthma has had increased interest over the past 10 years. Some studies have looked at nutritional changes influencing the establishment of asthma, whereas others have concentrated on possible secondary prophylactic measures once the diagnosis of asthma has been made (Table 5).

Possible Preventative Effects of Nutrition

The evidence for possible preventative effects of nutritional changes is based mainly on observational studies. Most studies are multifaceted and it is often difficult to differentiate the effects of one exposure or intervention from another. Allergenic food avoidance during pregnancy, lactation, and weaning has no effect on the subsequent development of asthma. Supplementation with, or diet high in, fish oil during pregnancy and infancy has only limited effect on asthma development. Low vitamin D status in pregnant

mothers is associated with a higher occurrence of wheeze in their children but further studies are required. Similarly, a higher prevalence of asthma has been reported with lower intake of antioxidants, although a recent meta-analysis does not support the hypothesis of linking vitamin C and E intake to a risk of asthma.

Breast-feeding is recommended over other types of feeding, and in infants at high risk of developing asthma (first-degree relative with atopy) where breast-feeding is not possible, hydrolyzed formula for a minimum of 4 months together with other preventative measures may reduce the risk of developing asthma or wheeze in the first year of life. Inconclusive evidence exists that growth rate during infancy may be related to the risk of development of asthma.

Secondary Prophylactic Effects of Nutrition in Asthma

Intervention studies could not confirm that supplementation with antioxidants produce clinical benefits in asthma. Better pulmonary function and less asthma have been observed, though, when there is higher intake of fresh fruit and vegetables and it seems to be reasonable to advise a healthy diet. Supplementation with, or a diet high in, fish oil as well as probiotics in asthmatic patients did not show any significant clinical improvement but equally showed no adverse effects with the latter showing a decrease in eosinophilia.

High sodium intake is associated with increased bronchial hyperresponsiveness; however, reducing salt intake results in only minimal clinical improvements. Low magnesium intakes may be associated with higher prevalence of asthma. Magnesium is used in the treatment of acute attacks for smooth muscle relaxation. Studies of long-term oral supplementation trials are limited and further larger randomized trials are required.

Obese asthma patients are advised to reduce weight based on evidence of an association between increased BMI and increased incidence and symptoms of asthma. Not enough evidence exists to recommend a strict calorie-controlled diet as a concomitant intervention with drug-based therapy.

According to a recent Cochrane review, there is no evidence that tartrazine, a commonly used food additive, makes asthma worse or avoiding it improves asthma control. Rarely, intolerance to certain foods may act as a trigger for asthma attacks and they should be avoided like other trigger allergens (Figure 7). A simple exclusion diet may be the most useful diagnostic diet, but strict food diets are seldom used in asthma.

Other Lung Diseases

Obstructive Sleep Apnea Syndrome

Obstructive sleep apnea syndrome (OSAS) is strongly associated with obesity and visceral fat mass and linked to serum leptin, insulin, and IL-6 and TNF-α levels. Early weight reduction with a very low-calorie diet and supervised lifestyle counseling is effective in improving symptoms in obese OSAS patients.

Bronchiectasis

Bronchiectasis unassociated with other conditions like CF is now uncommon in most developed countries, but remains a problem in developing countries and in certain indigenous populations mainly due to tuberculosis and other lung infections. The nutritional management of bronchiectasis is similar to that of CF and COPD.

Chronic Lung Disease of Infancy

Chronic lung disease of infancy (CLDI) describes a heterogeneous group of pulmonary disorders that originate from an acute respiratory disorder during the neonatal period, which may lead to chronic lung disease in childhood and adult life. The majority of cases are attributable to bronchopulmonary dysplasia (BPD), which is common in very preterm infants with respiratory distress syndrome. BPD has recently been defined as the requirement of supplemental oxygen for 21 of the first 28 days of life and is further defined according to severity (Table 6).

Table 6 Definition of bronchopulmonary dysplasia (BPD) according to severity (severity depends on the duration and level of supplemental oxygen and mechanical ventilatory support at 36 weeks postmenstrual age (PMA))

Gestational age	*<32 Weeks*	*≥32 Weeks*
Mild BPD	Breathing room air at 36 weeks PMA	Breathing room air by 56 days postnatal age
Moderate BPD	Requirement for <30% oxygen at 36 weeks PMA	Requirement for <30% oxygen at 56 days postnatal age
Severe BPD	Requirement for ≥30% oxygen and/or positive pressure at 36 weeks PMA	Requirement for ≥30% oxygen and/or positive pressure at 56 days postnatal age

Source: Data extracted from Table 1 on Jobe AH and Bancalari E (2001) Bronchopulmonary dysplasia. *American Journal of Respiratory and Critical Care Medicine* 163: 1723–1729, with permission from American Thoracic Society.

Table 7 Micronutrients: evidence for the effect of deficiency and supplementation on acute respiratory infections

Micronutrient	*Deficiency*	*Supplementation*
Zinc	• Increased prevalence and severity of LRTI	• Possible beneficial effects as adjunct therapy on length of symptoms and outcome in acute pneumonia • Reduction in LRTI prevalence
Vitamin A	• Increased risk of developing ARI	• No evidence as adjunct therapy during ARI • No evidence as preventative supplementation[a]
Vitamin D	• Increased risk and severity of LRTI	• No evidence yet as adjunct therapy or preventative supplementation
Iron	• Increased incidence of LRTI	• No overall effect on risk of LRTI
Selenium	• Found in premature babies with RDS and associated with CLDI in low birth weight infants	• Possible benefits as adjunct therapy in ARI
Multiple micronutrients	• Increased risk of LRTI	• Limited evidence. Suggested reduction in incident of LRTI similar to zinc supplementation alone

[a]Apart from in those who are malnourished.

Short-term vitamin A supplementation has been shown to reduce BPD at 36 postmenstrual age (postmenstrual age (weeks): gestational age plus chronological age) in extremely low birth weight infants (<1000 g), but the effects are similar to caffeine that is currently used, and the long-term risk–benefit ratio of vitamin A for neurodevelopment has not been fully established.

Growth failure is common in infants with BPD. They have increased metabolic demands and a relative state of protein-calorie malnutrition. Early growth restriction may increase long-term cardiovascular risks and have adverse effects on respiratory function. Currently, guidelines promote 'aggressive' nutrition for extremely preterm infants (born <28 weeks' gestation), which includes introducing parenteral amino acids from day one and lipids from day two, and using milk fortifiers to increase daily protein and caloric intakes to 4.4 g kg^{-1} and 130–150 kcal kg^{-1}, respectively.

Respiratory Tract Infections and Micronutrients

Studies have shown that undernourished children have a relative risk of 1.2 for an increased incidence of any acute respiratory illness (ARI) and 1.9 for lower respiratory tract infections (LRTIs). Single and multiple micronutrient deficiencies as part of malnutrition may be associated with increased prevalence of respiratory disease (Table 7). However, supplementation trials with single and multiple micronutrients in the prevention and as treatment adjunctive of ARIs have so far mainly shown contradictory results and further research is required to make recommendations. There are suggestions that adjunct therapy with zinc during the treatment of acute pneumonia may reduce the length of symptoms and improve outcome. Similar suggestions have been made with regard to selenium supplementations.

For detailed discussion of nutrition and tuberculosis, *see article* Nutrition and Susceptibility to Tuberculosis (00268).

Further Reading

Anker, S.D., John, M., Pedersen, P.U., et al., 2006. ESPEN guidelines on enteral nutrition: Cardiology and pulmonology. Clinical Nutrition 25, 311–318.

British Thoracic Society and Scottish Intercollegiate Guidelines Network (2009) *British Guideline on the Management of Asthma. A National Clinical Guideline.* London: British Thoracic Society; Edinburgh: Scottish Intercollegiate Guidelines Network.

Cystic Fibrosis in Australia (2006) *Australasian Clinical Practice Guidelines for Nutrition in Cystic Fibrosis.* Cystic Fibrosis in Australia.

Cystic Fibrosis Trust, 2002. Nutritional Management of Cystic Fibrosis. Cystic Fibrosis Trust, Kent.

Dine, J., Williams, J., DeLisser, H.M., 2009. Pulmonary disease. In: Hark, L., Morrison, G. (Eds.), Medical Nutrition and Disease, 4th edn. Wiley-Blackwell, Chichester, UK, pp. 355–393.

Ferreira IM, Brooks D, Lacasse Y, Goldstein RS, and White J (2008) Nutritional supplementation for stable chronic obstructive pulmonary disease. *Cochrane Database Systematic Reviews.* 2005 3, p. CD000998.

Kim, J.H., Ellwood, P.E., Asher, M.I., 2009. Diet and asthma: Looking back, moving forward. Respiratory Research 10, 49.

Lange, N.E., Litonjua, A., Hawrylowicz, C.M., Weiss, S., 2009. Vitamin D, the immune system and asthma. Expert Review of Clinical Immunology 5, 693–702.

Litonjua, A.A., Gold, D.R., 2008. Asthma and obesity: Common early-life influences in the inception of disease. Journal of Allergy and Clinical Immunology 121, 1075–1084.

Milla, C.E., 2007. Nutrition and lung disease in cystic fibrosis. Clinics in Chest Medicine 28, 319–330.

Rabe, K.F., Hurd, S., Anzueto, A., et al., 2007. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. American Journal of Respiratory and Critical Care Medicine 176, 532–555.

Reid PT and Innes JA (2010) Respiratory disease. In: Colledge NR, Walker BR, Ralston SHA (eds.) *Davidson's Principle and Practice of Medicine*, 21st edn., pp. 641–730.

Roth, D.E., Caulfield, L.E., Ezzati, M., Black, R.E., 2008. Acute lower respiratory infections in childhood: Opportunities for reducing the global burden through nutritional interventions. Bulletin of the World Health Organisation 86, 356–364.

Schulzke, S.M., Pillow, J.J., 2010. The management of evolving bronchopulmonary dysplasia. Paediatric Respiratory Reviews 11, 143–148.

Stallings, V.A., Stark, L.J., Robinson, K.A., Feranchak, A.P., Quinton, H., 2008. Evidence-based practice recommendations for nutrition-related management of children and adults with cystic fibrosis and pancreatic insufficiency: Results of a systematic review. Journal of the American Dietetics Association 108, 832–839.

Sugawara, K., Takahashi, H., Kasai, C., et al., 2010. Effects of nutritional supplementation combined with low-intensity exercise in malnourished patients with COPD. Respiratory Medicine 104 (12), 1883–1889.

Malabsorption syndromes: Nutritional management

P.M. Tsai, University of California, San Francisco, CA, USA
C. Duggan, Harvard Medical School, Boston, MA, USA

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of P.M. Tsai, C. Duggan, Malabsorption Syndromes: Nutritional Management, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 136–142, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00178-1.

Introduction 477
Pathophysiology and Symptoms 477
General Nutritional Management of Malabsorption 479
Specific Nutritional Management of Malabsorption 479
Fluids and Electrolytes 479
Carbohydrate Malabsorption 479
Lactose Intolerance 479
Sucrose 480
Fat Malabsorption: Fat and Fat-Soluble Nutrients 480
Protein Malabsorption 481
Route of Nutrition in Malabsorption 481
Selection of Enteral Formulas for Malabsorption 481
Clinical Management of Malabsorption 482
Inflammatory Bowel Disease 482
Short Bowel Syndrome 482
Summary 483
Further Reading 483
Relevant Websites 483

Glossary

Anthropometrics Body measurements used to determine body fat composition.
Bioelectrical impedance A method that uses a small electrical current to measure body composition.
DEXA scan A dual energy X-ray absorptiometry scan that uses X-ray beams to measure bone density and/or body composition.
Endoscopy A procedure that uses an instrument to look inside the body for diagnostic or therapeutic purposes.
Protein hydrolysate A type of formula in which proteins are broken down into smaller peptides by hydrolysis.

Introduction

The human gastrointestinal tract has an impressive capacitance for water, electrolyte, and nutrient absorption. In some disease states, however, this excess capacity is outpaced by either intestinal secretion or inadequate absorption. Malabsorption is defined as the inability of the gastrointestinal tract to adequately absorb nutrients. Although strictly speaking, malabsorption is distinct and contrasted with maldigestion (inadequate breakdown of nutrients in the intestinal lumen); the therapeutic implications of these two conditions are often similar. Multiple causes of malabsorption exist and reviews of these individual diagnoses can be found in separate sections of this text (e.g., inflammatory bowel disease, cystic fibrosis, short bowel syndrome, etc.). The pathophysiology, symptoms, and nutritional therapies for common malabsorption syndromes have been reviewed.

Pathophysiology and Symptoms

Malabsorption can occur when any of the several steps in nutrient digestion, absorption, and/or assimilation are interrupted; see **Table 1** for a list of congenital defects in nutrient assimilation. Carbohydrate malabsorption can occur, for instance, when intestinal disaccharidases are reduced in concentration at the enterocyte. The brush border membrane produces four disaccharidases that are important in carbohydrate digestion. These enzymes are sucrase-isomaltase, maltase-glucoamylase, trehalase, and lactase-phlorizin hydrolase. Worldwide, lactase deficiency is the most common type of acquired disaccharidase deficiency. With lactase deficiency, malabsorbed carbohydrate remains in the intestinal lumen and exerts an osmotic pull on fluids and electrolytes, leading to abdominal cramping and loose stools. Malabsorbed carbohydrate can be metabolized by gastrointestinal tract bacteria, and the fermented

Table 1 Congenital defects in nutrient assimilation. Included are congenital defects that are associated with gastrointestinal symptoms and/or nutritional deficiencies. Congenital defects not included here include multiple defects in amino acid absorption

Disorder	*Gene/protein affected*	*Symptoms*
Carbohydrate digestion		
Congenital lactase deficiency	Lactase	Lactose-induced diarrhea
Hypolactasia	Lactase	Lactose-induced diarrhea
Congenital sucrase-isomaltase deficiency	Sucrase-isomaltase	Sucrose-induced diarrhea
Carbohydrate absorption		
Glucose–galactose malabsorption	Sodium-glucose-co-transport (SGLT1); SLC5A1	Glucose-induced diarrhea
Fructose malabsorption	Facilitative fructose transport (GLUT5); SLC2A5	Fructose-induced diarrhea
Fanconi-Bickel syndrome	Facilitative glucose transport (GLUT2); SLC2A2	Diarrhea and nephropathy
Protein digestion		
Enterokinase deficiency	Serine protease 7	Diarrhea and edema
Trypsinogen deficiency	Trypsinogen	Diarrhea and edema
Fat digestion		
Pancreatic lipase deficiency	Pancreatic lipase	Steatorrhea
Fat assimilation		
Abetalipoproteinemia	Microsomal triglyceride transfer protein	Steatorrhea
Hypobetalipoproteinemia	Apolipoprotein B	Steatorrhea
Chylomicron retention disease	Sar1-ADP-ribosylation factor family GTPases	Steatorrhea
Primary bile acid malabsorption	Sodium-bile acid transporter; SLC10A2	Steatorrhea, bile acid diarrhea
Tangier disease	ATP-binding cassette transporter 1	Hepatosplenomegaly
Sitosterolemia	ATP-binding cassette, subfamily G, member 8; ABCG8	Atherosclerosis
Ion and metal absorption		
Congenital sodium diarrhea	Defective Na^+/H^+ exchange	Secretory diarrhea
Congenital chloride diarrhea	Defective Cl^-/HCO^{3-} exchange	Secretory diarrhea
Cystic fibrosis	CFTR	Pancreatic insufficiency, meconium ileus
Acrodermatitis enteropathica	Zinc and iron-regulated transport proteins; SLC39A4	Diarrhea and dermatitis
Menkes disease	Copper transporter	Developmental delay
Primary hypomagnesemia	Paracellin 1; claudin 16	Seizures, deafness and polyuria
Hemachromatosis	Hepcidin, others	Cirrhosis, cardiomyopathy, diabetes
Vitamin absorption		
Folate malabsorption	?	Macrocytic anemia, diarrhea, developmental delay
Congenital pernicious anemia	Intrinsic factor	Macrocytic anemia, developmental delay
Imerslund-Graesbeck syndrome	Cubilin, amnionless	Anemia, proteinuria
Congenital deficit of transcobalamin II	Transcobalamin II	Anemia, diarrhea, developmental delay
Thiamine-responsive megaloblastic anemia	Thiamine transport protein; SLC19A2	Anemia, diabetes, cranial nerve defects
Familial retinol binding protein (RBP) deficiency	Retinol-binding protein 4	Ophthalmologic problems
Selective vitamin E deficiency	Alpha-tocopherol transport protein	Vitamin E malabsorption

Source: Adapted from Martin M and Wright EM (2008) Congenital intestinal transport defects. In: Kleinman RE, Goulet O, Mieli-Vergani G, Sanderson IR, Sherman P, and Shneider B (eds.) *Walker's Pediatric Gastrointestinal Disease: Pathophysiology, Diagnosis, Management*, 5th edn., p. 290. Hamilton, Ontario: BC Decker.

gas produced is associated with flatulence and bloating. Bacterial overgrowth of the small intestine, as seen with short bowel syndrome, can also be associated with carbohydrate malabsorption.

Steatorrhea, excessive fat in the stools, results from fat malabsorption or maldigestion and can have several causes, most notably pancreatic insufficiency due to cystic fibrosis, chronic pancreatitis, Shwachman-Diamond syndrome, and Johanson-Blizzard syndrome. Failure of pancreatic secretion of lipase, amylase, and other digestive enzymes leads to persistence of dietary fat in the intestinal lumen, causing bloating, abdominal pain, and bulky, foul-smelling, oily stools. The stools often float due to a high gas content and test positive for fat. Patients also complain of blunted appetite and nausea. Other causes of fat malabsorption include hepatobiliary disease with inadequate bile salt circulation, severe mucosal disease, and short bowel syndrome.

The most common cause of protein malabsorption is so-called protein-losing enteropathy. Etiologies include diffuse mucosal disease such as celiac disease or Crohn's disease, elevated right heart pressure with resultant dilatation of lymphatics and leakage of lymph into the lumen, and invasive enteropathies as seen with *Shigella* or *Salmonella* infections. Because protein is a relatively minor component of dietary energy compared with carbohydrate and fat, symptoms of protein malabsorption can sometimes be minimal. However, infectious colitis or exacerbations of inflammatory bowel disease often present with frequent loose stools, which may be bloody. Rarely, congenital etiologies of protein malabsorption include enterokinase and trypsinogen deficiencies (**Table 1**).

Finally, the malabsorption of various micronutrients can occur in conjunction with or separate from the macronutrient malabsorption syndromes noted above. For instance, steatorrhea can be accompanied by excessive fecal losses of the fat soluble vitamins

A, D, E, and K, as well as calcium and other minerals. Alternatively, atrophic gastritis or surgical resection of the terminal ileum can lead to Vitamin B_{12} malabsorption in the absence of any symptoms of diarrhea. Proximal bowel resection can result in iron, zinc and calcium malabsorption. A rare cause of micronutrient inadequacy is abetalipoproteinemia in which fat soluble nutrients are normally digested and absorbed by the intestine but are not delivered to the circulation due to defective transepithelial transport. Other rare causes of micronutrient malabsorption are noted in **Table 1**.

General Nutritional Management of Malabsorption

As with all nutritional disorders, a thorough nutritional assessment is needed to plan rational therapy of malabsorption. Important historical points to review include duration of symptoms, underlying etiology of malabsorption, ability to meet nutritional needs by mouth, presence of food allergies, and concurrent medical and surgical problems. The patient's nutritional status (weight, height, body mass index, and their respective percentiles) should be determined. Tests of body composition such as arm anthropometrics, bioelectrical impedance, or DEXA scan should be considered. If the underlying cause of malabsorption is not known, diagnostic gastrointestinal endoscopy, laboratory studies, and/or imaging studies are indicated.

Specific Nutritional Management of Malabsorption

Fluids and Electrolytes

Diarrhea is usually the most distressing problem for patients with malabsorption and may cause dehydration. Care should be taken to correct fluid losses with appropriately designed oral rehydration solutions. Even in the setting of massive secretory diarrhea such as seen with cholera infections, oral rehydration solutions are effective at treating dehydration. Recent data have supported the safety and efficacy of oral rehydration solutions of reduced osmolarity in children with dehydration from acute diarrhea. An oral rehydration solution with the following composition: glucose 75 mmol l^{-1}, sodium 75 mmol l^{-1}, potassium 20 mmol l^{-1}, base 30 mEq l^{-1}, and osmolality 245 mOsm l^{-1} is well-suited for the rehydration and maintenance therapy during dehydration due to diarrhea.

In some cases of severe diarrhea, parenteral hydration is the mainstay of therapy. Examples include glucose–galactose malabsorption, congenital chloride diarrhea, microvillous inclusion disease, and tufting enteropathy. These cases, as well as other severe causes of more common malabsorptive syndromes, also frequently require the use of parenteral nutrition therapy.

Carbohydrate Malabsorption

Lactose Intolerance

Lactose intolerance is defined by the occurrence of symptoms after ingestion of lactose, the main carbohydrate in milk. These symptoms may include abdominal pain, bloating, diarrhea, or flatulence. Lactose intolerance is usually secondary to lactose malabsorption caused by a relative deficiency of the disaccharidase lactase, which reduces the ability to digest lactose. Primary lactase deficiency is a condition in which lactase activity falls after weaning around 2 years of age. Secondary lactose intolerance may be temporary and is usually due to mucosal injury associated with a condition or disease such as infectious diarrhea, Crohn's disease, or short bowel syndrome.

In addition to the presence or absence of the lactase enzyme, other factors determine whether a person will have symptoms of lactose malabsorption, including the amount of lactose in the diet, the mixture of lactose with other foods, gastric emptying rate, colonic scavenge of malabsorbed carbohydrate, ethnic origin and age. Although persons of Northern European ancestry commonly maintain the ability to digest lactose into adulthood, primary lactose intolerance is prevalent in African–American, Hispanic, Native American, and Asian populations.

Nutritional management of lactose intolerance consists largely of the removal of lactose from the diet. Lactose is a common ingredient in many foods, including breads, crackers, soups, cereals, cookies, and baked goods. Eliminating or reducing lactose-containing ingredients from one's diet is usually adequate to relieve symptoms. Individuals with primary lactose intolerance may require a permanent dietary change. Individuals with secondary lactose intolerance should eliminate all lactose from their diets for a short period of time ranging from 2 to 6 weeks. If symptoms resolve, lactose may be reintroduced slowly as tolerated by the individual. The amount of lactose that an individual can tolerate is highly variable. Many children can tolerate small amounts of lactose, particularly yogurt or hard cheese, without discomfort. Many adults who consider themselves lactose intolerant can actually tolerate moderate amounts of milk. Lactose intolerant individuals may also tolerate small amounts of lactose consumed over the course of the day better than a large dose all at once.

For individuals who choose to restrict lactose in their diets, a variety of lactose-free and low-lactose food choices are available. Lactose-reduced products, containing 70–100% less lactose than standard foods, are available commercially. Individuals may also choose to consume dairy products with concomitant administration of lactase enzyme tablets or drops.

Frequent consumption of milk and other dairy foods has been associated with better bone health in some studies, and a strict lactose-free diet may not contain adequate amounts of calcium and vitamin D. **Table 2** provides a list of some commercially available lactose-free and lactose-reduced calcium supplements.

Table 2 Commercial calcium supplements

Product	*Manufacturer*	*Mg Calcium/ tablet*	*IU vitamin D*
Citracal Regular	Bayer	500	400
OsCal 500+D	GlaxoSmithKline	500	200
Tums	GlaxoSmithKline	500	0
Calcium Milk Free	Nature's Plus	250	50
Cal-citrate+D	Freeda	250	100
Caltrate 600+D	Pfizer	600	400
Viactiv[a]	McNeil Nutritionals	500	500

Source: Adapted from DiSanto C and Duggan C (2005) Gastrointestinal diseases. In: Hendricks KM and Duggan C (eds.) *Manual of Pediatric Nutrition*, 4th edn., p. 212. Hamilton, Ontario: BC Decker.

[a]This product contains less than 0.5 mg lactose.

Sucrose

Congenital sucrase-isomaltase deficiency (SID) is the most common congenital disaccharidase deficiency. Patients with this disorder lack functional sucrase, although isomaltase deficiency may be normal or absent. Symptoms of SID can include diarrhea, abdominal pain, and poor weight gain. Dietary avoidance of sucrose or table sugar helps relieve symptoms, and can sometimes help with the diagnosis. Sucraid®, a sacrosidase produced from *Saccharomyces cerevisiae*, is an enzyme that can be given with meals and allows increased tolerance to sucrose.

Fat Malabsorption: Fat and Fat-Soluble Nutrients

Patients with pancreatic insufficiency are unable to produce and secrete enough enzymes to aid with the breakdown of fats in the intestinal lumen. In studies of normal adults and those with pancreatic insufficiency, pancreatic enzyme secretion needs to be lower than 15% of normal levels before significant steatorrhea is seen (**Figure 1**). Once clinically significant steatorrhea is determined, recovery of pancreatic function is therefore unlikely.

Historically, patients with pancreatic insufficiency due to cystic fibrosis (CF) were told to minimize symptoms of steatorrhea by limiting dietary fat. However, epidemiologic studies confirmed that this advice lead to negative energy balance, undernutrition, and higher mortality rates, compared to communities in which CF patients were treated with high-energy and high-fat diets. The introduction of effective pancreatic replacement therapy has been heralded as one of the most significant breakthroughs in the nutritional management of CF, responsible partly for the substantial increase in lifespan enjoyed by more recent generations of CF patients. In fact, the finding of a lower incidence of growth failure in CF patients diagnosed and treated with aggressive nutritional therapy early in infancy has been used as justification for neonatal screening of this condition.

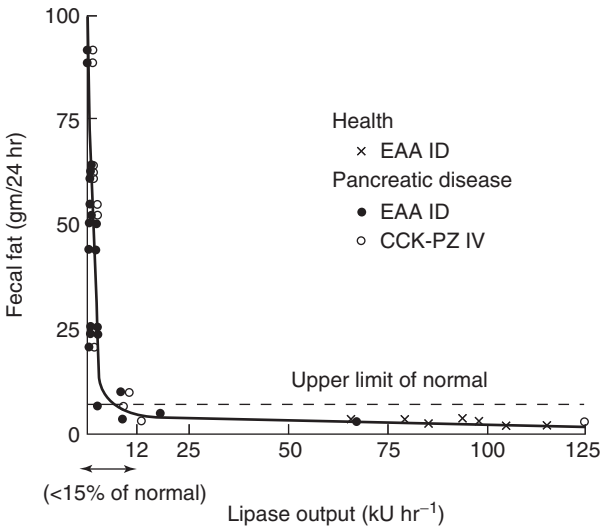


Figure 1 Pancreatic enzyme secretion and steatorrhea. Significant steatorrhea ensues when pancreatic function is less than 15% of normal. Reproduced with permission from DiMagno EP, Go VLW, and Sumerskill WHJ (1973) Relations between pancreatic enzyme output and malabsorption in severe pancreatic insufficiency. *New England Journal of Medicine* 288: 814. Copyright © 1973 Massachusetts Medical Society. All rights reserved.

Judicious use of pancreatic replacement enzymes is the hallmark of nutritional therapy of CF and other disorders of pancreatic insufficiency. Multiple commercial preparations of porcine pancreatic enzymes are available, most of which contain lipase, amylase, and protease enzymes. A nonporcine pancreatic enzyme is currently under development. The dose is usually titrated to the amount of steatorrhea. If meals take more than 30 min, the dose may be divided with half given before the meal and half given mid-way through the meal. Patients who cannot swallow pills may open the capsules and sprinkle the enzymes into acidic foods.

Another critical aspect of the nutritional management of fat malabsorption is routine supplementation with the fat-soluble vitamins A, D, E, and K. Multiple studies have confirmed that patients with CF, Crohn's disease, and other malabsorptive disorders are prone to micronutrient deficiencies, and some literature suggests that dietary needs for these and other antioxidant nutrients may be increased in settings of infectious and catabolic stress often suffered by these patients. The contribution of fat malabsorption contributing to other important mineral malabsorption, as in the case of calcium or zinc, should also be recognized.

Routine supplementation of fat-soluble vitamins is indicated in patients with fat malabsorption. In addition, serial measurement of fat-soluble vitamin biochemical status is recommended. Because blood nutrient concentrations of these and other nutrients can vary with the concentration of transport proteins, correction for these can aid the interpretation of these lab findings. For instance, vitamin A toxicity should be suspected if the molar ratio of vitamin A: Retinol-binding protein exceeds 1. Vitamin E concentrations, for example, should be corrected for circulating lipids.

Some patients with pancreatic malabsorption may benefit from a diet enriched in medium chain triglycerides (MCTs). MCTs are absorbed directly into the portal circulation and therefore bypass the steps of intraluminal digestion, reesterification, and enterocyte uptake. Therefore, these fats may be a dietary source of fats more easily absorbed in settings of fat malabsorption due to either pancreatic insufficiency or mucosal disease. However, MCT oils are less energy dense than long-chain fats, are more expensive, and do not contain the essential fatty acids alpha-linoleic and linolenic acid.

Protein Malabsorption

Protein-losing enteropathy (PLE) can also be treated with a variety of nutritional interventions. PLE due to dilated lymphatics as with right heart failure results in leakage of lymphocytes, proteins, and fats into the intestinal lumen. As with fat malabsorption, MCT-supplemented foods and formulas are therefore indicated to allow improved fat absorption in PLE. Fat-soluble vitamin supplementation is indicated. In congenital protein malabsorption syndromes, peptide- or amino acid-based formulas are often helpful.

Mucosal disorders including inflammatory bowel disease, allergic diseases, and celiac disease are other examples of disorders causing protein malabsorption. Once intestinal inflammation is reduced with appropriate medical or nutritional therapy, absorption of protein is usually improved. In *Shigella* infections, some studies have suggested improved nutritional outcomes with a high-protein diet during recovery from the acute symptoms of diarrhea.

Route of Nutrition in Malabsorption

Several factors need to be considered when recommending whether oral, enteral, or parenteral nutrition should be used in providing nutrition to the patient with malabsorption. These factors include etiology of malabsorption, severity of gastrointestinal disease, and underlying nutritional and medical conditions. Oral nutrition using modified diets as noted above is, of course, the most customary and desirable by physician and patient alike. In cases of mild lactose malabsorption, for instance, modification of a regular, healthy diet to avoid foods high in lactose should be sufficient. In cases where widespread gastrointestinal disease is leading to severe malabsorption, enteral or 'tube' feeding is helpful for two main reasons: (1) use of proprietary formulas specially designed for malabsorption are often indicated, and these formulas may be unpalatable, and (2) enteral feedings, especially with slow continuous 'drip' feedings make efficient use of nutrient transport kinetics, thereby maximizing residual gastrointestinal absorptive function. In severe cases of malabsorption in which tube feedings are unable to achieve adequate nutritional intake, parenteral nutrition may be indicated. Emerging data suggest that serum citrulline, an amino acid synthesized principally by the enterocyte, is a reliable biomarker of mucosal mass and may help distinguish among patients who require parenteral versus enteral nutrition.

Selection of Enteral Formulas for Malabsorption

A number of commercially available formulas are designed for patients with malabsorption, and these differ with regard to energy density, macronutrient composition, and indicated age. Because infant formulas are often handled in a separate regulatory fashion by governments, infant formulas are usually considered separately from formulas designed for older children and adults. In addition, formulas are also conventionally categorized by the extent of the hydrolysis of their protein source. Categories include intact protein formulas, protein hydrolysate formulas, and amino acid-based formulas. Protein hydrolysate formulas are also sometimes referred to as 'semielemental' formulas, and amino-acid formulas are sometimes called 'elemental' formulas. However, these terms suffer from vagueness and inaccuracies because not all of their macronutrients are semi or completely elemental. Marketing strategies often compound the confusion with misleading formula names. These terms should be discouraged, and the terms that refer to the composition and/or biochemical processing should be used instead.

Patients who have carbohydrate malabsorption from lactose intolerance should use lactose-free formula. Fat malabsorption may call for MCT enriched formula. In cases of protein malabsorption or severe enteropathy, a formula that is a protein hydrolysate or amino acid based would be most appropriate. Because many malabsorption syndromes overlap in terms of the macronutrient affected, as in cases of severe mucosal disease, some formulas are designed for fat, protein, and carbohydrate malabsorption. For example, all formulas designed for use in adults are lactose free, and several formulas contain both hydrolyzed proteins and MCT oils.

Clinical Management of Malabsorption

Two of the most clinically challenging scenarios for the management of malabsorption are inflammatory bowel disease (especially Crohn's disease) and short bowel syndrome. Both are discussed in separate articles of the text, but are considered briefly below.

Inflammatory Bowel Disease

Patients with Crohn's disease have widespread and intermittent gastrointestinal inflammation. Some patients with inflammatory bowel disease may require complete bowel rest for several days or even a few weeks to allow time for mucosal healing. To provide nutrition during this period of time, parenteral nutrition may be needed.

Numerous studies have shown that patients with Crohn's disease may safely and effectively achieve clinical remission with primary nutritional therapy. Early literature in the field highlighted the use of protein hydrolysate formulas, which, due to unpalatability, often required administration via a nasogastric or gastrostomy tube. More recent data have confirmed that intact protein formulas, termed 'polymeric' formulas when describing formulas designed for adults, may work as well as protein hydrolysates, and these formulas can be feasibly given by mouth.

As patients are recovering from an exacerbation and begin advancing their diet, patients should temporarily minimize the amount of fiber ingested to decrease trauma to healing mucosa. Patients whose disease affects the small intestine often benefit from temporary avoidance of lactose products as the mucosa heals and brush border membrane enzyme production is restored.

Micronutrients are also needed in the nutritional management of inflammatory bowel disease. Iron supplementation is recommended for anemia due to acute or chronic blood loss. Treatment of inflammatory bowel disease frequently requires the use of steroids, which affects bone density. Calcium and vitamin D supplementation is commonly needed to minimize the osteopenic effects of steroid therapy and/or the effects of malabsorption and chronic inflammation.

Short Bowel Syndrome

Patients who have suffered acquired or congenital loss of small intestinal surface area that makes them dependent on specialized enteral or parenteral support are said to have short bowel syndrome (SBS). Patients with SBS often malabsorb carbohydrates, proteins, fat, as well as numerous micronutrients, depending on the extent and location of bowel resection, as well as the presence of mucosal disease in the nonresected bowel.

Special attention should be given to exactly what part of the intestine remains as well as the length of the remaining intestine. Some patients may have the terminal ileum removed and are unable to absorb vitamin B_{12} and bile acids. Removal of the ileocecal valve increases the risk of bacterial overgrowth. Reduced length also means reduced surface area for the absorption of nutrients and decreased intestinal transit time.

In the immediate postoperative period, parenteral nutrition and gut rest should be used because significant stool output is the norm. Output should be quantified, and electrolytes must be carefully monitored in order to determine appropriate replacement fluids to make up for excess urine, stool, and ostomy losses. Replacement fluids should generally be given separately from standard parenteral nutrition so that they can be adjusted as needed to rapid shifts in fluid and electrolyte status.

As patients recover from surgery, every attempt should be made to feed them enterally as soon as is feasible. Enteral feeds facilitate growth and adaptation of the remaining bowel to allow partial compensation for the missing portion, and several studies have correlated early feeding with better long-term outcome. Attaining independence from parenteral nutrition may take weeks to months to years. **Table 3** outlines an approach to determine feeding advancement in SBS. Although some patients are able to grow well or maintain their body weight with only enteral feeds, many are dependent on parenteral nutrition. Some patients with SBS also have oral feeding aversion due to prematurity, prolonged mechanical ventilation, and/or prolonged orogastric or nasogastric feeding. Gastrostomy tubes are particularly helpful in this regard.

In infants, breast milk should be used if available. The breast milk may need to be fortified to increase calories, protein, or fat. For older patients or infants who are not receiving breast milk, protein hydrolysates or amino acid-based formulas may be better tolerated because the residual bowel more easily absorbs these nutrients. Lactose-free and MCT-containing formulas are often used, as well. Formulas may need to be supplemented with oral rehydration solutions if electrolyte abnormalities persist, particularly with sodium losses through persistent high stool or ostomy output.

Because many patients with SBS are dependent on parenteral nutrition for prolonged periods of time, selenium, carnitine, copper, and zinc blood concentrations should be checked periodically and supplemented if needed. Parenteral nutrition should be cycled off for a few hours each day to help simulate more natural cyclic fluctuations of gastrointestinal hormones. These patients also often have poor absorption of calcium and need calcium supplements to prevent osteopenia, which increases the risk of

Table 3 Feeding advancement in short bowel syndrome

1. Stool output	
If<10 g kg^{-1} d^{-1} or <10 stools d^{-1}	Advance rate by 10–20 ml kg d^{-1}
If 10–20 g kg^{-1} d^{-1} or 10–12 stools d^{-1}	no change
If>20 g kg^{-1} d^{-1} or>12 stools d^{-1}	Reduce rate or hold feeds[a]
or 2. Ileostomy output	
If<2 g kg^{-1} h^{-1}	Advance rate by 10–20 ml kg^{-1} d^{-1}
If 2–3 g kg^{-1} h^{-1}	no change
If>3 g kg^{-1} h^{-1}	Reduce rate or hold feeds[a]
3. Stool reducing substances	
If<1%	Advance feeds per stool or ostomy output
If=1%	no change
If>1%	Reduce rate or hold feeds[a]
4. Signs of dehydration	
If absent	Advance feeds per stool or ostomy output
If present	Reduce rate or hold feeds[a]
5. Gastric aspirates	
< four times previous hour's infusion	Advance feeds
> four times previous hour's infusion	Reduce rate or hold feeds[a]

Source: Adapted from Utter SL and Duggan C (2005) Short bowel syndrome. In: Hendricks KM and Duggan C (eds.) *Manual of Pediatric Nutrition*, 4th edn., pp. 728–729. Hamilton, Ontario: BC Decker.

[a]Feeds should generally be held for 8 h, then restarted at 3/4 the previous rate.

fractures. Iron may also be needed in patients with anemia from decreased absorption secondary to resection of the duodenum or jejunum. Ultimately, weaning from parenteral and enteral nutrition remain the goals of treatment, though lifelong dietary therapy is often needed.

Summary

Congenital or acquired diseases of the gastrointestinal tract can lead to life-threatening malabsorption of numerous macronutrients and micronutrients. Determining the type and etiology of malabsorption is essential to provide appropriate nutritional and medical therapy. Multiple formulas, supplements, and dietary regimens exist to target specific defects in the digestion, absorption, and assimilation of nutrients. In addition, many new nutrients are undergoing investigation that may become a standard part of care in the future, including probiotics, prebiotics, and various amino acids.

Further Reading

Cézard, J.P., 2008. Normal physiology of intestinal digestion and absorption. In: Kleinman, R.E., Goulet, O., Mieli-Vergani, G., Sanderson, I.R., Sherman, P., Shneider, B. (Eds.), Walker's Pediatric Gastrointestinal Disease: Pathophysiology, Diagnosis, Management, 5th edn. BC Decker, Hamilton, Ontario, pp. 245–252.

Crenn, P., De Truchis, P., Neveux, N., 2009. Plasma citrulline is a biomarker of enterocyte mass and an indicator of parenteral nutrition in HIV-infected patients. American Journal of Clinical Nutrition 90 (3), 587–594.

Holt, P.R., 2007. Intestinal malabsorption in the elderly. Digestive Diseases 25, 144–150.

Iannitti, T., Palmieri, B., 2010. Therapeutical use of probiotic formulations in clinical practice. Journal of Clinical Nutrition. https://doi.org/10.1016/j.clnu.2010.05.004. In press.

Olieman, J.F., Penning, C., Ijsselstijn, H., et al., 2010. Enteral nutrition in children with short-bowel syndrome: Current evidence and recommendations for the clinician. Journal of the American Dietetic Association 110, 420–426.

Rubio-Tapia, A., Murray, J.A., 2010. Celiac disease. Current Opinion in Gastroenterology 26, 116–122.

Shikhare, G., Kugathasan, S., 2010. Inflammatory bowel disease in children: Current trends. Journal of Gastroenterology 45, 673–682.

Stallings, V.A., Stark, L.J., Robinson, K.A., et al., 2008. Evidence-based practice recommendations for nutrition-related management of children and adults with cystic fibrosis and pancreatic insufficiency: Results of a systematic review. Journal of the American Dietetic Association 108, 832–839.

Treem, W.R., McAdams, L., Stanford, L., et al., 1999. Sacrosidase therapy for congenital sucrase-isomaltase deficiency. Journal of Pediatric Gastroenterology and Nutrition 28, 137–142.

Relevant Websites

http://www.nutrition.org/. – American Society for Nutrition.
http://www.celiac.org. – Celiac Disease Foundation.
http://www.cdhnf.org. – Children's Digestive Health and Nutrition Foundation.
http://www.ccfa.org. – Crohn's and Colitis Foundation of America.
http://www.cff.org. – Cystic Fibrosis Foundation.

Malnutrition

Joy Ngo[a,b] **and Lluis Serra-Majem**[a,c,d], [a] NGO Nutrition Without Borders, Barcelona, Spain; [b] Nutrition Research Foundation (FIN), University of Barcelona Science Park, Barcelona, Spain; [c] Institute of Biomedical and Health Sciences (IUIBS), University of Las Palmas de Gran Canaria (ULPGC), Las Palmas de Gran Canaria, Spain; and [d] CIBER Fisiopatología de la Obesidad y Nutrición (CIBER OBN), Instituto de Salud Carlos III, Madrid, Spain

© 2023 Elsevier Ltd. All rights reserved.

Introduction **485**
Definitions for the discussion of malnutrition 486
Malnutrition 486
Overnourishment 486
Overnutrition 486
Overweight and obesity 486
Protein energy malnutrition 486
Stunting 486
Undernourishment 486
Undernutrition 486
Underweight 486
Wasting 487
Malnutrition and mortality **487**
Nutrition and sustainable development **487**
Human and economic costs 489
Sustainable food systems 489
Types of malnutrition **490**
Undernutrition 490
Causes of undernutrition **492**
Determinants **493**
Biological and dietary factors 493
Socioeconomic factors and inequity 494
Environmental factors 494
Global nutrition targets **494**
Progress toward achieving global nutrition targets 496
Severe Acute Malnutrition (SAM) **496**
Immunity and SAM 498
Clinical manifestations and classification **498**
Mild and Moderate Acute Malnutrition (MAM) 498
Severe Acute Malnutrition (SAM) 498
Marasmus 498
Kwashiorkor 498
Marasmic-kwashiorkor 500
Long term consequences **500**
Chronic undernutrition **500**
Stunting 500
The role of fetal growth restriction 500
Anthropometric indicators for measuring malnutrition **501**
Malnutrition in children 501
Underweight, stunting and wasting 501
Malnutrition in adults 502
Moderate and severe thinness, underweight, overweight, obesity 502
Micronutrient malnutrition **503**
Micronutrient deficiencies 503
Conclusions **503**
References **506**
Further reading **506**
Relevant websites **506**

Key points

- Malnutrition has multiple drivers and a variety of manifestations which are complex and often times interrelated.
- Malnutrition encompasses various conditions/diseases, ranging from wasting and micronutrient deficiencies to obesity.
- Core factors leading to malnutrition include excessive intakes, insufficient food consumption or inadequate utilization of nutrients provided by foods ingested.
- A systems approach incorporating food, health, water and sanitation, education, and social protection is necessary to improve nutrition outcomes.
- Globally many countries are experiencing both under and overnutrition, in essence, the dilemma of overlapping and coexisting forms of malnutrition.

List of abbreviations

BMI Body mass index
FAO Food and Agriculture Organization of the United Nations
MDGs Millennium Development Goals
MAM Moderate Acute Malnutrition
NCD Non-communicable disease
PEM Protein Energy Malnutrition
SAM Severe acute malnutrition
SGA Small for gestational age
SDG Sustainable Development Goals
UNICEF United Nations Children's Fund
WHA World Health Assembly
WHO World Health Organization

Introduction

Throughout history, hunger has remained one of the most persistent and devasting problems facing humankind. Although statistics on food insecurity and malnutrition reveal noteworthy improvement in these areas over the last five decades, the current scenario involves factors that undermine the nutrition situation of populations—urbanization, inequities, globalization, environmental calamities, armed conflicts, humanitarian emergencies and health epidemics. In 2020 an estimated 720–811 million people in the world faced hunger, which has been further exacerbated by the COVID-19 pandemic, the effects of climate change, ongoing armed conflicts and the outbreak of war in the Ukraine. This has led to overwhelmed health systems, increased household food insecurity, and reversals in economic growth, all of which could hinder the progress made in addressing undernutrition, particularly among low-income and middle-income countries (LMIC).

Malnutrition doesn't always imply hunger or famine, which is the first association that typically comes to mind. Malnutrition is manifest in many forms and refers to deficiencies, excesses or imbalances in a person's intake of energy and/or nutrients. The resulting consequences cover two broad groups of conditions dealing with deficits and excesses, ranging from emaciation to obesity, which can be defined in terms of undernutrition, overnutrition and micronutrient deficiencies. Simply put, malnutrition is caused by excessive intakes, insufficient food consumption or inadequate utilization of nutrients provided by foods ingested. Insufficient intakes often result from precarious circumstances and inequalities existing at the level of economic resources, education, access to food and safe drinking water, adequate health care and sanitary living conditions. On the other hand, excessive intake of foods high in saturated fats, sugars and industrially processed, refined foods accompanied by decreased physical activity have led to increased obesity and diet-related non-communicable diseases (NCDs).

How malnutrition manifests itself is changing as it is present in various forms in nearly all countries. Currently, there is a global pandemic of overweight and obesity. Simultaneously, the burden of disease in low- and middle-income countries is rapidly changing from communicable to non-communicable diseases. Yet deficiencies in micronutrients and growth are still an issue as seen in rates for anemia as well as wasting, which haven't changed substantially for more than 20 years. The World Health Organization (WHO) estimates that at least one-third of the world's population is affected by micronutrient malnutrition. Moreover, according to the most recent Global Nutrition Report over 2.2 billion adults aged 18 and older are overweight (40% of the global population), of which an estimated 772 million are obese. In effect, no country is free of malnutrition and many households, communities, and regions struggle with all of these forms of malnutrition at the same time. Countries are confronted with the

burden of complex, overlapping and interrelated malnutrition problems, with the combination of under-five stunting, anemia in women of reproductive age, and adult overweight and obesity being the most common. The well documented "double burden of malnutrition" (child undernutrition and maternal overweight/obesity) has evolved into the "triple burden of malnutrition": undernutrition, in the form of stunting and wasting, an increasing prevalence of overweight and obesity accompanied by widespread micronutrient deficiencies. These different conditions are connected not only at a physiological level, but also at the level of resources and political will. As obesity and overweight is an extensive subject in and of itself, this article will focus solely on malnutrition in terms of undernutrition and micronutrient deficiencies.

Definitions for the discussion of malnutrition

For clarity, several definitions are presented so as to better understand the context of malnutrition and how to distinguish its various forms.

Malnutrition

An abnormal physiological condition caused by inadequate, unbalanced or excessive consumption of macronutrients and/or micronutrients. Malnutrition includes undernutrition and overnutrition as well as micronutrient deficiencies.

Overnourishment

Food and beverage intake that is continuously in excess of dietary energy requirements.

Overnutrition

A result of excessive food intake relative to dietary nutrient requirements.

Overweight and obesity

Body weight that is above normal for height as a result of excessive fat accumulation. It is usually a manifestation of overnourishment and/or sedentary lifestyles and a lack of physical activity. Overweight is defined as a Body Mass Index (BMI) of more than 25 but less than 30. Obesity is defined as a BMI of 30 or more.

Protein energy malnutrition

Classic medical terminology referring to a group of related disorders that include marasmus, kwashiorkor and intermediate states of marasmus-kwashiorkor. This designation implied that malnutrition was a deficiency of protein or energy and, as for other specific deficiencies, suggested that treatment should consist of proteins or calories (energy). Textbooks usually went on to say that Marasmus (extreme thinness) was due to insufficient calories, and Kwashiorkor (edematous malnutrition) was the result of a lack of proteins. This has now been proven incorrect due to the role of deficient intakes of essential micronutrients needed for growth and tissue renovation. Marasmus can occur with excess energy if no nutrients are ingested and Kwashiorkor can become manifest even with excess protein in the diet.

Stunting

Also known as linear growth, stunting is defined as height/length (cm) for age (months) <-2 SD of the WHO Child Growth Standards median. Low height-for-age is an indicator that reflects the cumulative effects of undernutrition and infections since and even prior to birth. It may be the result of long-term nutritional deprivation, recurrent infections and lack of water and sanitation infrastructures.

Undernourishment

The state in which an individual's usual food consumption is insufficient to provide the amount of dietary energy required to maintain a normal, active, healthy life.

Undernutrition

The outcome of poor nutritional intake in terms of quantity and/or quality, and/or poor absorption and/or poor biological use of nutrients consumed as a result of repeated instances of disease. It includes being underweight for one's age, too short for one's age (stunted), precariously thin for one's height (suffering from wasting) or deficient in vitamins and minerals (micronutrient deficiency).

Underweight

Low weight-for- age in children, and BMI of less than 18.5 in adults, reflecting a current condition resulting from inadequate food intake, past episodes of undernutrition or poor health conditions.

Wasting

Defined as low weight-for height/length (Weight for height/length <−2 SD of the WHO Child Growth Standards median). Low weight-for-height/length is an indicator of acute weight loss or a failure to gain weight and can be a consequence of insufficient food intake and/or an incidence of infectious diseases, particularly diarrhea.

Malnutrition and mortality

The 2013 Lancet Series on Maternal and Child Nutrition highlighted that undernutrition, including fetal growth restriction, suboptimum breastfeeding, stunting, wasting, and deficiencies of vitamin A, iodine and zinc, caused 3.1 million deaths annually in children younger than 5 years. This represented a staggering 45% of total child deaths in 2011. Included in this total was the combined distribution of suboptimum breastfeeding and fetal growth restriction in the neonatal period, which resulted in 1.3 million deaths or 19% of all deaths of children younger than 5 years. They provided clear evidence about the importance of undernutrition as a predisposing factor of an increased risk for illness and death and as an antecedent condition of structural and functional limitation. During the acute stage of severe undernutrition, the greater susceptibility of the child to infectious disease has major clinical significance and important consequences for child mortality.

Fig. 1 shows the contribution of different childhood illnesses on child mortality, with pneumonia (30%) and diarrhea (27%) comprising over 50% of childhood deaths. The area within the red circle represents the percentage of deaths that were caused due to the child being undernourished. When compared to a well-nourished child, children with severe, moderate or even mild undernutrition have a higher risk of mortality from these childhood illnesses. As such, undernutrition constitutes an underlying cause for most childhood deaths associated with severe infections. Recent Joint Child Malnutrition Estimates highlight that around 45% of deaths among children under 5 years of age are linked to undernutrition, most of which occur in low- and middle-income countries (LMIC).

Recent data from the United Nations Interagency Group for Child Mortality Estimation shows that despite a global decline in mortality over the past three decades, 5.2 million children died before the age of 5 years in 2019 alone, primarily of preventable or treatable causes (Fig. 2). Almost half of these deaths occurred in the first month of life. Another analysis evaluating the impacts of COVID-19 on childhood malnutrition and nutrition-related mortality in 118 LMICs suggests there could be a 14.3% increase in the prevalence of moderate or severe wasting among children younger than 5 years due to COVID-19-related economic declines. Currently, the interagency joint malnutrition estimates state that there are 45.4 million children under 5 years of age who suffer from moderate or severe wasting. When the projected increase in wasting in each country was combined with a projected yearly average reduction of 25% in nutrition and health services, an estimated 128,605 (ranging from 111,193 to 178,510 for best and worst case scenarios) additional deaths in children younger than 5 years were projected during 2020, with an estimated 52% of these deaths in sub-Saharan Africa. Sub- Saharan Africa remains the region with the highest under-5 mortality rate in the world, an average of 74 deaths per 1000 live births in 2020, 14 times higher than the risk for children living in Europe and North America.

Nutrition and sustainable development

In 1992, the first International Conference on Nutrition held in Rome called for a global commitment to address the problem of malnutrition. Since then several initiatives have continued to address this challenge. On January 1, 2016, 17 Sustainable

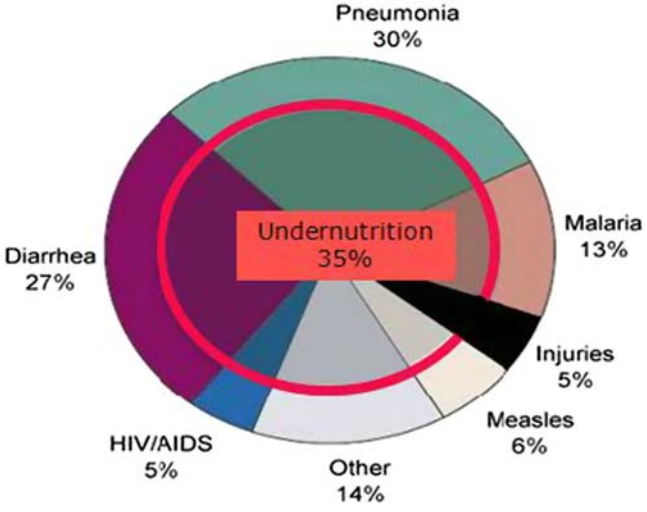


Fig. 1 Undernutrition-related risk of death associated with specific diseases.

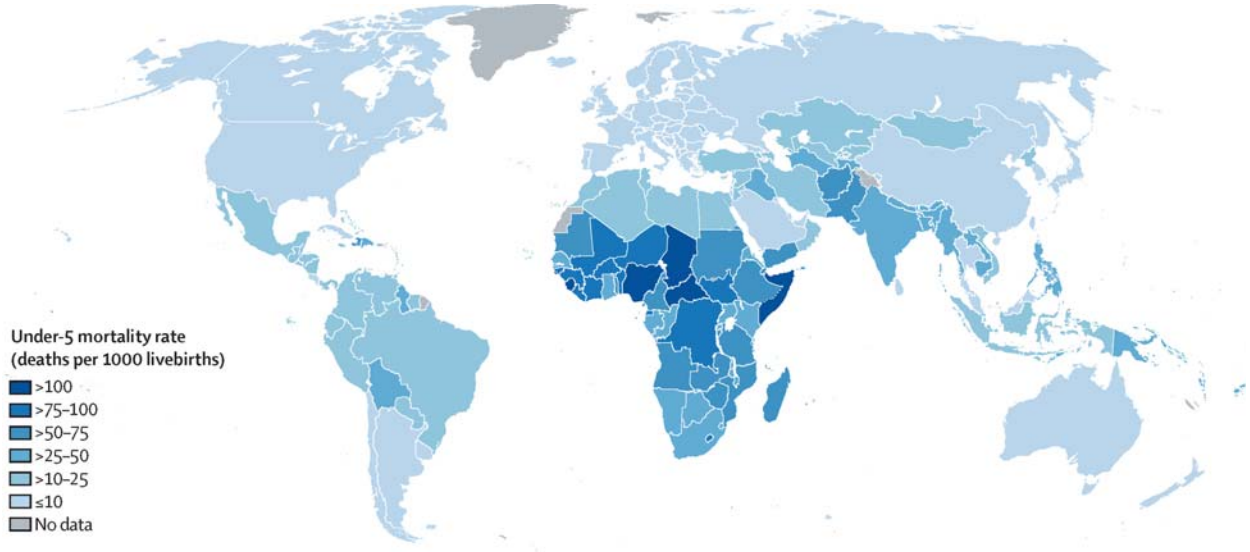


Fig. 2 Under-5 mortality rate in 2019, by country. Source: Sharrow et al. (2022).



Fig. 3 17 Sustainable development goals of the 2030 agenda for sustainable development. Source: United Nations (2022).

Development Goals (SDG), also known as the Global Goals, of the 2030 Agenda for Sustainable Development were formally recognized, which had been adopted by world leaders in September 2015 at a momentous UN Summit (see Fig. 3). They built on the success of the Millennium Development Goals (MDGs) with the objective of going further to end all forms of poverty. Moreover, at least 12 of the 17 Sustainable Development Goals contain indicators that are extremely relevant for nutrition, reflecting nutrition's central role in sustainable development. As such, the SDGs are broader in scope with 169 targets which recognize that ending poverty requires fomenting strategies that target economic growth and address a range of social needs including health, education, social welfare, and job opportunities, while taking on climate change and environmental protection. SDG 2, "End hunger, achieve food security and improved nutrition and promote sustainable agriculture", aims to put an end to hunger and ensure access by all people, particularly those in vulnerable situations, including infants, to safe, nutritious and sufficient food all year round by 2030. Also, by 2030, another target aims to end all forms of malnutrition, including achieving, by 2025, the internationally agreed upon targets on stunting and wasting in children under 5 years of age, and to address the nutritional needs of adolescent girls, pregnant and lactating women and older persons (Fig. 4).

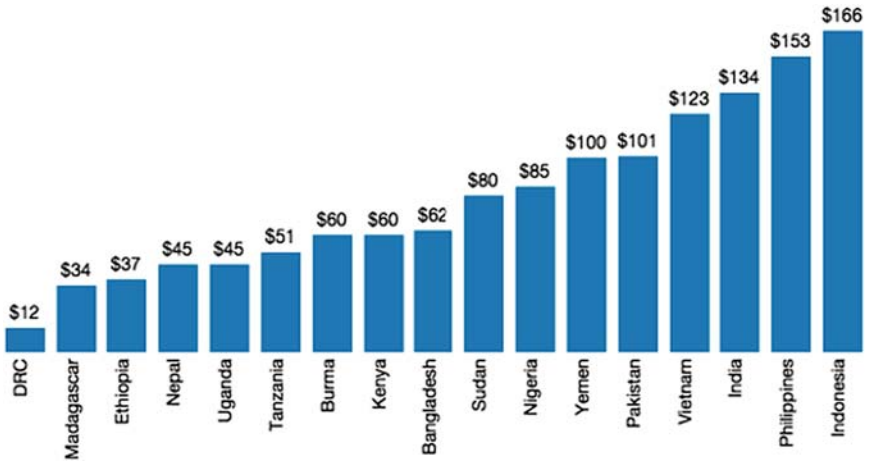


Fig. 4 Reducing stunting: benefit for every dollar spent by country, (discount rate = 3%, final working age = 50). Source: Horton and Hoddinott (2014).

Human and economic costs

Improving nutrition contributes to productivity, economic development, and poverty reduction by improving physical work capacity, cognitive development, school performance, and health by reducing disease and mortality. As the Copenhagen Consensus Center has shown, an entity that collaborates with the world's top economists to research the best solutions for global challenges, investing in nutrition is critical as poor nutrition perpetuates the cycle of poverty and malnutrition through three main pathways: (1) direct losses in productivity from poor physical status and losses caused by disease linked with malnutrition; (2) indirect losses from poor cognitive development and losses in schooling; and (3) losses caused by augmented health care costs. The economic costs of malnutrition are staggering—several billion dollars a year in terms of lost gross domestic product (GDP).

Since early childhood offers a special window of opportunity to improve nutrition, the bulk of the investment needs to be targeted between pre-pregnancy and 2 years of age. Economic studies show that every dollar spent on nutrition in the first 1000 days of a child's life (fetus to age two) can yield, on average, a savings of $45 and in some cases as much as $166.

Sustainable food systems

The UN's Zero Hunger Challenge (ZHC), launched in 2012, aims for all food systems to be sustainable. This approach is critical as what we eat affects both human and planetary health. Unsustainable food consumption and production constitute the primary causes of the triple planetary crises: climate change, biodiversity loss and pollution. Food systems should take into account their environmental impact, and in so doing, create a vital link between the nutrition agenda and the sustainability agenda of the SDGs. However, in many regions throughout the world, dietary patterns and their associated food systems are linked to child malnutrition, environmental degradation and climate change. Simultaneously, climate change is undermining and reversing the advances made in reducing child malnutrition. In fact, lower income countries bear a greater burden of the negative impacts of climate change. This is due to their economic reliance on climate-sensitive sectors (agriculture and fishing, for example), and to their limited financial, human and institutional capacity for forecasting and responding to the direct and indirect effects of climate change.

Data from the 2021 Global Nutrition Report shows that lower-income countries continue having the lowest consumption of health-promoting foods and the highest numbers of underweight, whereas higher-income countries have the highest consumption of foods with greater environmental and health impacts, and the highest levels of overweight and obesity. Economic and policy support measures for agricultural production primarily target staple foods, dairy and other protein-rich foods of animal origin, especially in high- and upper-middle-income countries. The most incentivized foods at a global scale include various categories of meat, rice and sugar while fruits and vegetables have less overall support, or may even be penalized in some low-income countries.

Moreover, in order to comply with the targets of the SDG 2 by 2030, increasing the supply of healthy foods should be accompanied by lower costs of these foodstuffs so that the population has access to them. This implies both an expansion in the supply of the nutritious foods that constitute a healthy diet and a shift in consumption toward them. As such, corresponding policies that promote changes in food environments and consumer behavior toward healthy eating patterns are also required. Agriculture policies and programs have concentrated so far on commodities and value chains. However, policy-makers are moving beyond the current focus on commodity goods and are beginning to implement actions centered on whole diets and food systems. This holistic approach can be seen in recent initiatives by key international agencies such as the United Nations Food Systems Summit and the Nutrition for Growth Summit and their focus on placing healthy diets and food systems at the forefront of their work. Among

others, a primary goal consists of "Repurposing food and agricultural policy support to make healthy diets more affordable" by reducing the costs of nutritious foods relative to other foods and people's income.

Dietary habits and the quantities and types of foods consumers demand are also associated with substantial and increasing levels of environmental pollution and resource use. Inadequate diets are both a cause and a symptom of climate change. Current food production and consumption patterns contribute to biodiversity loss, put pressure on natural resources, and provoke Greenhouse Gas (GHG) Emissions and excessive agro-chemical inputs, particularly in developed countries. Industrialized regions are living beyond the planet's capacity, yet their food systems are too often considered as role models for developing countries. Food systems—both in developed and developing countries—need to be restructured and take into account social, environmental and public health impacts. Biodiversity offers a wealth of untapped potential for livelihoods, health, nutrition and environments. Interdisciplinary collaboration is needed if global efforts are to tackle malnutrition, NCDs and the impact of food on climate change, water and biodiversity. Food production and consumption that is sustainable can safeguard the planet from degradation and lessen the effects of climate change and extreme weather events. In order to be sustainable, it's essential that diets be healthy as well as being compatible with the sustainable management of natural resources and ensuring social equity. As has been done with nutrition, advocacy for making sustainable food production central to development is essential.

Types of malnutrition

Undernutrition

Undernutrition is a physical condition characterized by a nutritional disorder resulting from insufficient food intake (undernourishment) and/or poor absorption and/or poor biological use of nutrients consumed as a result of repeated infectious disease. It encompasses being underweight for one's age, too short for one's age (stunted), dangerously thin for one's height (wasted) and deficient (status and function) in vitamins and minerals (micronutrient malnutrition). Although the terms malnutrition and undernutrition are often used interchangeably, the first is more extensive in scope. In this article both are used interchangeably. Indeed, malnutrition is a nutritional disorder that can manifest itself in distinct forms:

(a) *Overnutrition:* a result of excessive food intake relative to dietary nutrient requirements. Previously more prevalent in developed countries, but now also appearing in countries with emerging economies.
(b) *Undernutrition:* as previously defined above. It's worth noting that different organizations have distinct definitions for this term. Some refer to it as an insufficient intake of nutrients to sustain a state in which the physical function of an individual is impaired to the point where s/he can no longer maintain adequate bodily performance process such as growth, pregnancy, lactation, physical work and resisting and recovering from disease.
(c) *Secondary undernutrition:* not caused by diet but rather by diseases or pathologies (diarrhea, infections, measles, intestinal parasites, etc.) that prevent the body from absorbing ingested nutrients.
(d) *Micronutrient deficiency:* when essential micronutrients such as vitamins and/or minerals and trace elements are not present in adequate amounts in the diet.

The last three are common in developing countries, with distinctions being made between acute undernutrition, also known as Moderate Acute Malnutrition (MAM) and Severe Acute Malnutrition (SAM), and chronic undernutrition (stunting). These three situations, and especially undernutrition, are often referred to colloquially as hunger. When in a given place and time, it undergoes a process of escalation that is accompanied by other factors (impoverishment, epidemics, often an increase in mortality), the situation turns into a famine. Undernutrition affects all age groups, but some are particularly vulnerable, mainly children (especially those between 6 months and 5 years, after weaning), the elderly and pregnant and lactating women.

Currently undernutrition is a problem that affects developing as well as industrialized countries, and in both cases, the most economically disadvantaged are those that are most affected. In developed countries, obesity is rapidly extending, especially among the poorest segments of the population, bringing with it an epidemic of nutrition-related NCDs such as diabetes and cardiovascular disease, which increase health care costs and reduce productivity. In developing countries, although undernutrition and micronutrient deficiency diseases continue to be particularly pervasive, the swiftly spreading problem of obesity is also a burden. It is common to find underweight children and obese adults within the same household, both in developed and developing countries. Moreover, UNICEF affirms that in 2020, the nutritional state of the world's children was characterized by a triple burden of malnutrition-undernutrition, micronutrient deficiencies, and overweight. This triple burden is manifest in multiple forms and often coexists in the same population, household, and even in the same child. As shown in **Fig. 5**, 88% of countries analyzed faced overlapping burdens of malnutrition.

Manipulating market dynamics and prices fail to address the problem of undernutrition when families have no money to buy enough food or pay for health care expenses. Advocacy linked to human rights and equity should be considered, as well as other initiatives related to income generation, thus justifying government intervention to support those families in need. However, undernutrition also occurs in many families with adequate economic resources, as people do not always know which foods or feeding practices are best for themselves or for their children, in which inadequate marketing practices from industry play a part. It may also be difficult for parents to know whether their children are starting to suffer from undernutrition as growth deficits and micronutrient deficiencies are not easily detected by those without previous training or education on the topic.

The following table shows the main problems of undernutrition throughout the lifecycle (**Table 1**).

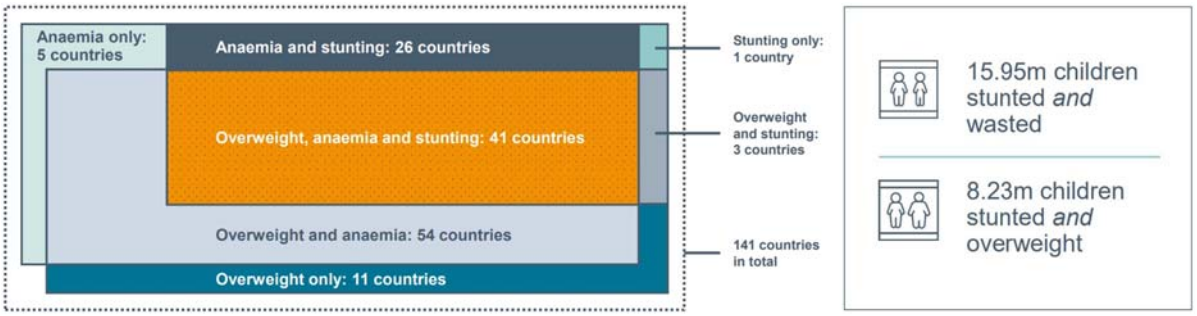


Fig. 5 Number of countries with overlapping forms of childhood stunting, anemia and overweight, in adult women, 2017 and 2018. Source: Global Nutrition Report 2018.

Table 1 Undernutrition throughout the lifecycle, according to disorder and main consequences.

Lifecycle stage	*Nutrition disorder*	*Main consequences*
Embryo/fetus	• Intrauterine growth retardation • Iodine deficiency disorders • Lack of folate	• Low birth weight • Prematurity • Incomplete nervous system development • Stillbirth • Increased risk for stunting, overweight/obesity, metabolic and cardiovascular diseases
Neonate (newborn)	• Low birth weight • Iodine deficiency disorders	• Growth retardation/Failure to thrive/Deafness • Delayed cognitive development • Anemia • Increased risk for stunting, overweight/obesity, metabolic and cardiovascular diseases
Infants and children	• Insufficient food intake • Iodine deficiency disorders • Vitamin A deficiency • Iron deficiency anemia	• Chronic undernutrition • Delayed physical and cognitive development • Increased risk of infections • Blindness • Anemia
Adolescents	• Insufficient food intake • Iodine deficiency disorders • Vitamin A deficiency • Iron deficiency anemia	• Delayed growth spurt • Stunting • Delayed intellectual development • Goiter • Increased risk of infections • Blindness • Anemia • Inadequate bone mineralization
Pregnant and breastfeeding women	• Insufficient food intake • Iodine deficiency disorders • Vitamin A deficiency • Iron deficiency anemia • Folate deficiency • Calcium deficiency	• Inadequate prenatal weight gain • Maternal anemia • Increased maternal mortality risk • Increased risk of infection • Night blindness • Increased risk of fetal mortality • Low birth weight
Adults	• Insufficient food intake • Iodine deficiency disorders • Obesity • Nutrition related diseases	• Thinness • Permanent fatigue • Obesity • Cardiac disease • Diabetes • Cancer • Arterial hypertension/Myocardial infarct • Anemia
Elderly	• Insufficient food intake • Iodine deficiency disorders • Obesity • Osteoporosis • Nutrition related diseases	• Obesity • Vertebral and hip fractures • Cardiovascular disease • Diabetes • Cancer • Cognitive impairment

Source: Adapted from WFP and CDC (2005).

Causes of undernutrition

The fundamental drivers of undernutrition and mortality quite often fall outside the nutrition sector and are associated with each country's levels of potential resources (i.e. energy, geography, climate, etc.), as well as with all the factors that condition their use. These include the management of population density in relation to available resources, poverty, social inequalities, the side effects of growth-oriented macroeconomic policies on structural adjustment and the migration to urban areas.

The classification of these causal chains can be simplified into three broad categories:

(a) *food insecurity:* includes problems of food production or supply at the national, regional or household level, as well as problems of families and communities in accessing sufficient amounts of safe and nutritious food for normal growth and development and an active and healthy life. It may be caused by the unavailability of food, insufficient purchasing power, inappropriate distribution or inadequate use of food at the household level. Food insecurity may be chronic, seasonal or transitory and may be induced by a myriad of factors including agricultural and trade policies, wars, famine and climate change, among others.

(b) *inadequate health and sanitation conditions:* The aspects of environmental hygiene include the supply of safe drinking water and food products, sanitation of the environment in all its forms and hygienic practices of the population. Health aspects address the areas of infectious and parasitic diseases on the one hand, and the systems and utilization of health care, on the other.

(c) *inappropriate care and feeding practices:* the concept of the provision of care and feeding practices refers both to care at the household level as well as to the broader aspects of social cohesion and protection at the community or national level. Thus, it encompasses the full range of maternal and child care, with mothers and children constituting the principal at-risk groups. In general, it refers to the attitudes and practices of the members of the household or community toward the most vulnerable members in their social sphere (i.e. available time, food distribution, emotional and material support) as well as the educational level of the caregivers.

The conceptual framework presented in Fig. 6 identifies three levels of causal factors leading to malnutrition in all its forms: immediate, underlying and enabling. It schematically illustrates the multifactorial and interconnected drivers that usually form the conceptual basis of "nutrition security". In other words, adequate nutritional status may be achieved by addressing the different

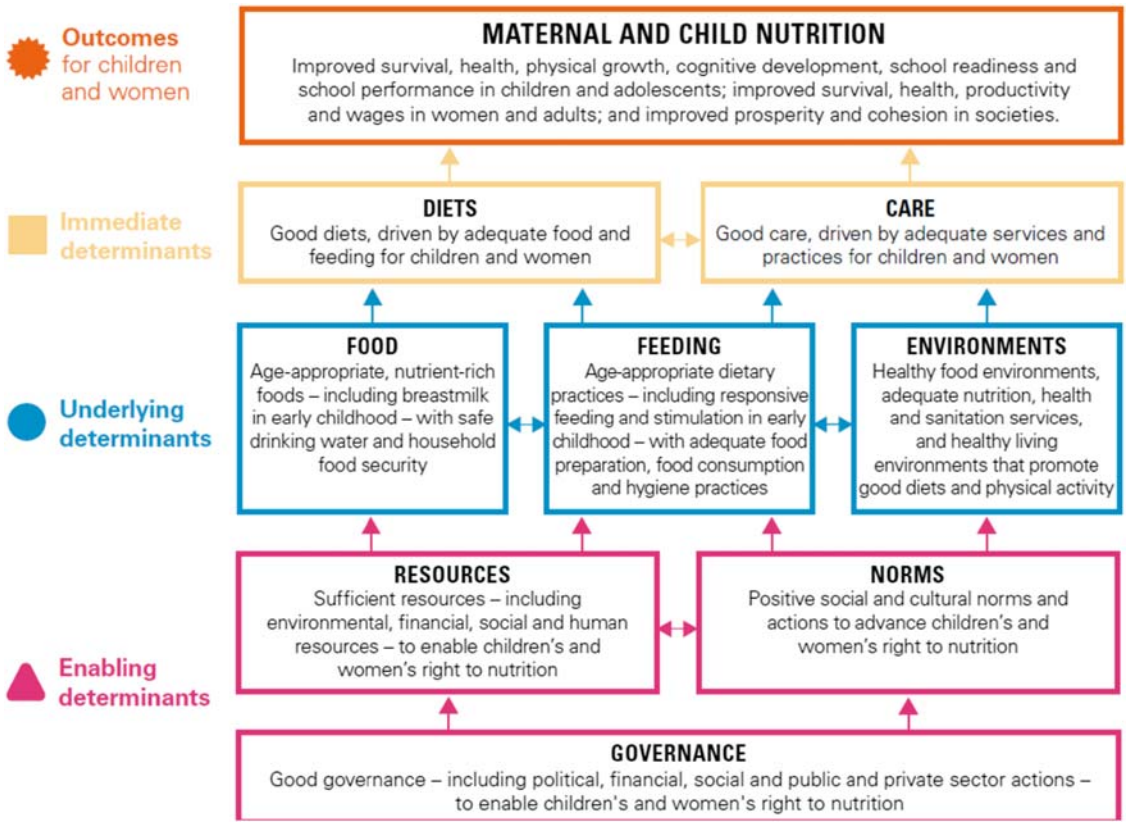


Fig. 6 UNICEF Conceptual Framework on the Determinants of Maternal and Child Nutrition, 2020. A framework for the prevention of malnutrition in all its forms. Source: UNICEF (2020).

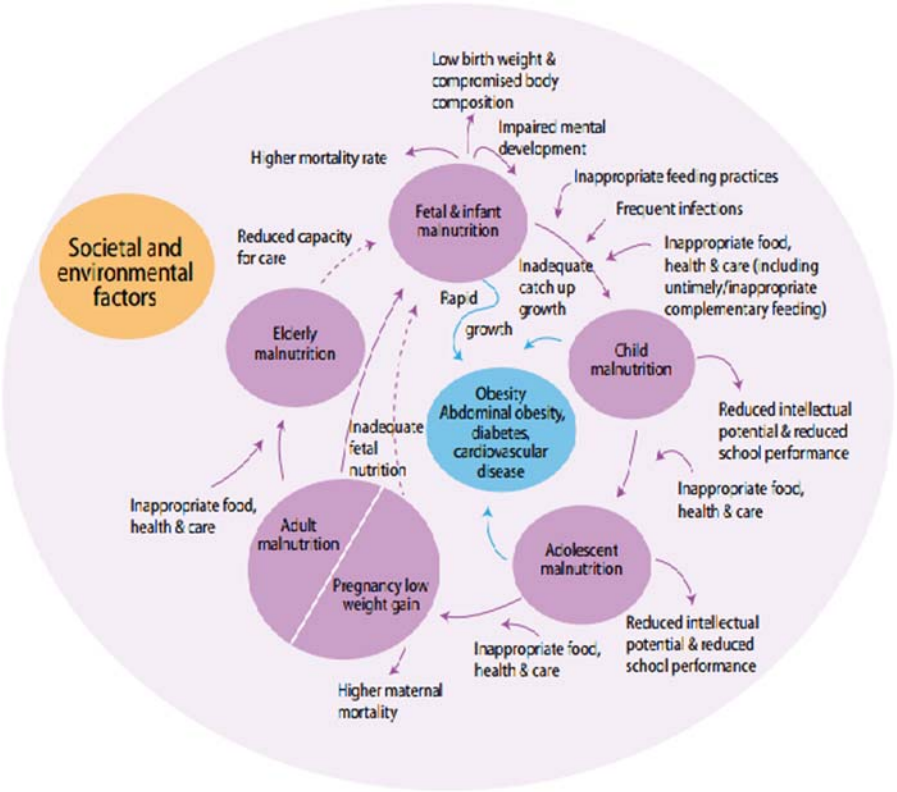


Fig. 7 Life-course: the proposed causal links of malnutrition. Source: Darnton-Hill et al. (2004).

determinants of malnutrition, with particular attention to the underlying causes. This analysis must be, however, more specifically tailored to each local situation, as well as the selection of appropriate indicators corresponding to the distinct causal levels and to each determinant (Fig. 7).

As previously mentioned and contrary to popular belief, undernutrition/malnutrition is not simply a result of food insecurity. Many children from environments that are food secure and from families with economic resources are below normal weight or have short stature (stunting), due to inappropriate infant feeding and care, poor access to health care services or inadequate sanitation.

In many countries where undernutrition is widespread, food production is not the limiting factor, with the exception of famine cases. The most important factors are, firstly, insufficient knowledge about the benefits of exclusive breastfeeding, complementary feeding practices and the function of micronutrients, and secondly, the lack of time women have to adequately care for their babies or for themselves during pregnancy. The most pernicious effects of undernutrition occur during pregnancy and the first 2 years of life, and the consequences of such early damage to health, brain development, intelligence, and to possibilities of benefiting from education and productivity are largely irreversible. There is evidence suggesting that obesity and NCDs such as cardiovascular disease and diabetes may also have their origins in early childhood. Moreover, the increase in overweight in conjunction with persisting undernutrition in young children has led to the aforementioned double burden of malnutrition in many countries, and studies have shown that both undernutrition and overweight have similar causes. As such, some have criticized the classification of malnutrition into under and overnutrition as being simplistic due to the interrelated nature between the two.

Determinants

Biological and dietary factors

Maternal undernutrition can result in intrauterine undernutrition and low birth weight, among others. Insufficient food intake to allow for compensatory growth that could offset this situation brings about the onset of Severe Acute Malnutrition (SAM) in early life stages.

The following figure shows the vicious circle of undernutrition. When a woman who is undernourished during pregnancy does not improve her nutritional status, the newborn presents with low birthweight. If that child is inadequately fed, it grows up being undernourished; and if the poor diet continues, s/he will become an undernourished adult. If female, when this undernourished woman becomes pregnant, she will give birth to an undernourished infant, and the cycle is perpetuated.

The American Academy of Pediatrics issued a 2018 policy statement emphasizing the importance and irreversibility of the 1000-day window of opportunity. It highlights the relationship between maternal prenatal and child nutrition in the first 1000 days (~2 years) of life and a child's neurodevelopment and mental health. Health risks both in childhood and as an adult, including obesity, hypertension and diabetes, may be determined by nutritional status during this time period. Adequate calorie intake, essential for growth of both the fetus and child, is not sufficient in and of itself for normal brain evolution and growth. The lack of key nutrients during this critical period of brain development may result in lifelong deficits in brain function even if there is subsequent nutrient repletion. Vital nutrients that support neurodevelopment include protein, zinc, choline, folate, iodine, vitamins A, D, B6, and B12.

Socioeconomic factors and inequity

Poverty is often associated with undernutrition due to reduced access to food, unsanitary living conditions, overcrowding and inadequate child care. Lack of knowledge leads to poor childcare practices, misconceptions about the use of various foods, inadequate feeding practices during illness and inadequate distribution of food among family members. The reduction in the initiation and duration of breast feeding, combined with inadequate weaning practices, are also associated with high rates of childhood undernutrition.

Social problems, such as the physical and emotional abuse of children, maternal deprivation, neglect of the elderly, alcoholism and drug addiction, can lead to the onset of undernutrition. Moreover, cultural and social practices that impose taboos and prohibit given foods, certain diet fads and the precarious conditions that characterize the migration from traditional rural areas to urban poverty zones, also contribute to the development of undernutrition. The following figure depicts highlights from the 2022 State of Food Security and Nutrition in the World, showing the relationship between undernutrition, wealth, maternal education and gender. Their analyses using the most recent data available per country (2015–2021) suggest that worldwide, under-five stunted children are more likely to be male, living in rural settings, poorer households, and whose mothers don't receive formal education. In contrast, obesity among women is most common in urban settings and wealthier households (**Fig. 8**).

Environmental factors

Overcrowded housing and poor sanitation favor the emergence of frequent infections and infectious disease outbreaks such as Covid-19 and Ebola. Natural disasters, as in droughts or floods, and human-induced disasters consisting of wars and forced migration, produce sudden, prolonged or cyclical food shortages, and can produce undernutrition in large population groups. For example, the war in Ukraine will have various impacts on global agricultural markets, affecting trade, production and prices and undermining food security as well as diet quality. Ukraine and the Russian Federation are major producers and exporters of important food commodities, fertilizer, minerals and energy. Both countries are jointly considered as the world's breadbasket, providing 30% and 20% of global wheat and corn exports, respectively, and 80% of worldwide exports of sunflower seed products. A minimum of 50 countries import 30% or more of their wheat from them, with many African and other LMIC importing more than 50%.

New studies have brought to light the impact of global climate change, which is projected to reduce future crop yields and thus threaten food security and eliminate much of the improvement in child undernutrition levels that would occur with no climate change. It has been estimated that climate change would lead to a relative increase in moderate stunting of 1–29% in 2050 compared to a future without climate change. Climate change would have a greater impact on severe stunting rates, projected to increase by 23% (central sub-Saharan Africa) to 62% (South Asia). Moreover, losses would be exacerbated by the additional loss of food post-harvest due to bad storage conditions as well as inadequate systems of food distribution, thus also contributing to the development of undernutrition, even after periods of agricultural abundance.

Global nutrition targets

Improving nutrition can be considered as one of the key developmental opportunities in the world today (**Table 1**).

In 2012 the World Health Assembly (WHA) Resolution 65.6 endorsed a comprehensive implementation plan on maternal, infant and young child nutrition (MIYCN), which specified a set of six global interrelated nutrition targets for the year 2025 with the aim to: (1) achieve a 40% reduction in the number of children under-5 who are stunted; (2) achieve a 50% reduction of anemia in women of reproductive age; (3)achieve a 30% reduction in low birth weight; (4) ensure that there is no increase in childhood overweight; (5) increase the rate of exclusive breastfeeding in the first 6 months up to at least 50%; and (6) reduce and maintain childhood wasting to less than 5%. These were extended to 2030 following the same groundwork and rationale as the 2025 nutrition targets established in 2012, and adjusting the targeted levels to newly available data (see **Table 2**). To a large degree they are aligned with the Sustainable Development Goals target levels to end all forms of malnutrition by 2030, although they aim to maintain the balance between achievability and established objectives, so as to uphold momentum for improving nutrition for health and development.

In addition, global nutrition targets not only refer to the WHA targets for maternal, infant and young child nutrition (MIYCN), but also those focusing on diet-related NCDs. The latter were adopted in 2013 by the WHA and were also to be reached by 2025. They include targets addressing four diet-related NCD indicators in adults: salt/sodium intake, raised blood pressure, diabetes and

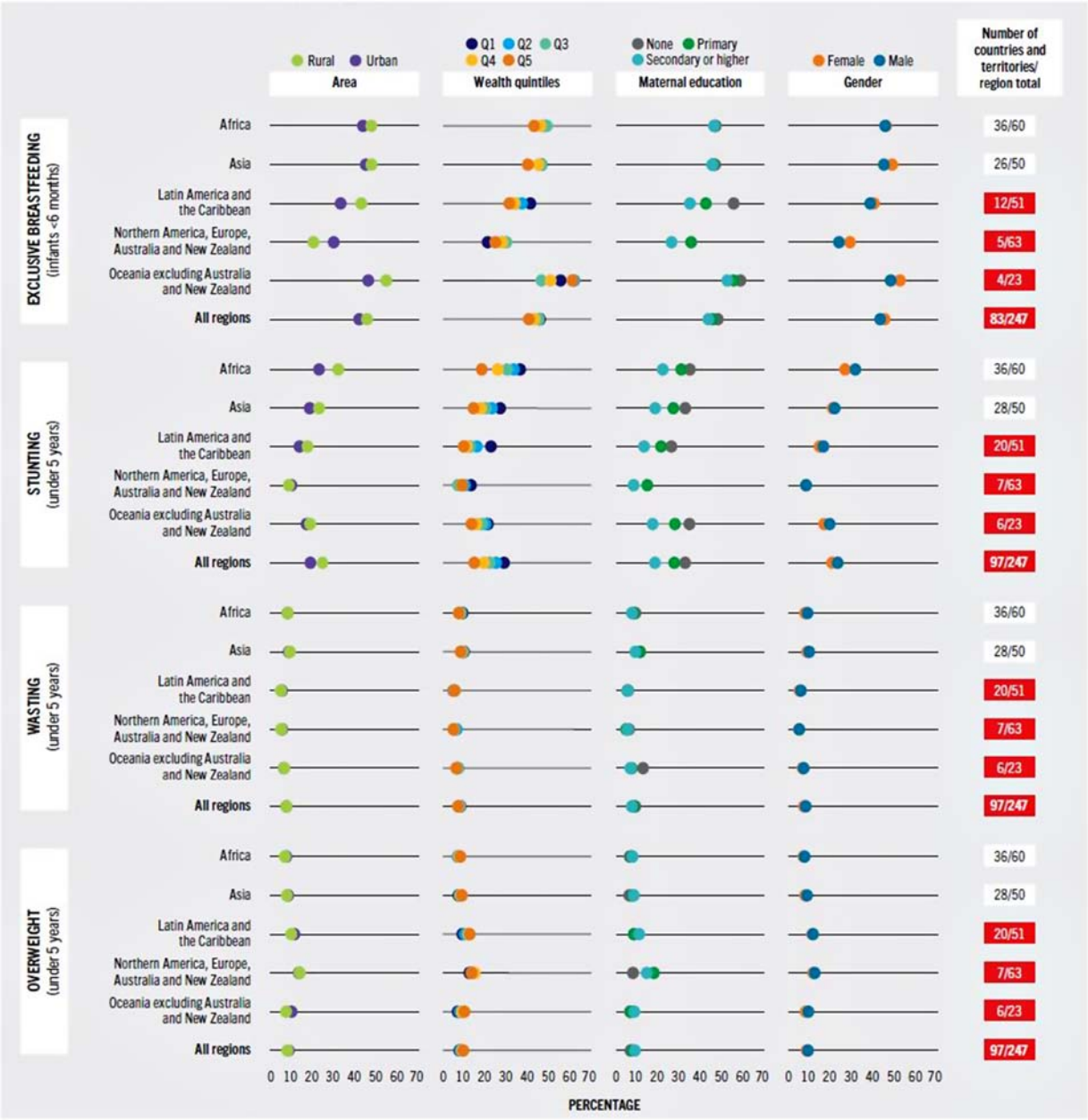


Fig. 8 Inequalities in exclusive breastfeeding and child malnutrition. Source: FAO et al. (2022).

Table 2 Global nutrition targets endorsed by the World Health assembly and their extension to 2030.

	2025 target	*2030 target*
Stunting (SDG)	40% reduction in the number of children under five who are stunted.	50% reduction in the number of children under five who are stunted.
Anemia (SDG)	50% reduction in anemia in women of reproductive age.	50% reduction in anemia in women of reproductive age.
Low birthweight	30% reduction in low birthweight.	30% reduction in low birthweight.
Childhood overweight (SDG)	No increase in childhood overweight.	Reduce and maintain childhood overweight to less than 3%.
Breastfeeding	Increase the rate of exclusive breastfeeding in the first 6 months up to at least 50%.	Increase the rate of exclusive breastfeeding in the first 6 months up to at least 70%.
Wasting (SDG)	Reduce and maintain childhood wasting to less than 5%.	Reduce and maintain childhood wasting to less than 3%.

Source: WHO and UNICEF (2017).

obesity. As such, the WHA's focus on these two sets of targets has pooled insights from studies evaluating the first 1000 days to prevent or cure child undernutrition, the connection between child undernutrition and the onset of overweight/obesity and NCDs later in life, and guidance on actions that can be applied throughout the life cycle. The NCD targets focus on preventing malnutrition related to unhealthy dietary and physical activity habits in the general population, whereas the principal focus of WHA targets is to decrease malnutrition in mothers and their young children. Thus, the WHA targets are complementary in nature and can enhance efforts to improve nutritional status and the population's wellbeing as a combined set. Both are well aligned with the SDGs, specifically SDG2 (food security and nutrition) and SDG3 (health).

Progress toward achieving global nutrition targets

As seen in the following figure, the burden of disease varies widely. Among the countries analyzed, the low- and lower-middle-income countries bore the greatest burden of stunting, wasting, low birthweight, and anemia cases whereas the upper-middle and high-income countries had the greatest burden of obesity cases.

According to the 2022 State of Food Security and Nutrition in the World, for the year 2020 in children less than 5 years of age, an estimated 149 million (22%) were stunted, 45 million (6.7 %) were wasted, and 39 million (5.7 %) were overweight. Improvements were observed for 2030 targets on stunting, while childhood overweight was worsening. Children who were stunted were more likely to live in low- or lower-middle-income countries (89% of the 2020 global burden), come from rural areas and have mothers without formal education. Nearly 30% of countries with representation in each subregion of North Africa, Oceania and the Caribbean demonstrated an increase in stunting prevalence. As such, these countries are not moving toward the 2030 target to decrease the number of stunted children by 50%.

Wasted children were also more likely to reside in low- or lower-middle-income countries (93% of the global burden) and come from households with lower economic resources. The levels of wasting remain above the 2030 target of less than 3% in many countries, particularly those in South and Southeast Asia.

It was more common to find overweight children residing in lower-middle- or upper-middle-income countries (77% of the 2020 global burden), coming from wealthier households and having mothers who had completed a minimum of secondary school education. As for the 2030 overweight target of less than 3%, more than half of the countries analyzed in West Africa and South Asia had achieved at least 75% progress, while overweight prevalence was increasing in the majority of countries analyzed in Southern Africa, Oceania, Southeast Asia, South America and the Caribbean.

Low birthweight decreased from 17.5% in 2000 to 14.6% in 2015 at a worldwide level, with most regions showing improvements. However, caution is needed in interpreting these results. Global monitoring of this indicator presents an important challenge with consequential data gaps, as nearly one in three newborns are not weighed at birth.

Exclusive breastfeeding has been progressing steadily. Globally, 43.8% of infants less than 6 months of age were exclusively breastfed in 2020, an increase from 37.1% in 2012. Infants who were exclusively breastfed were likely to come from low- or lower-middle-income countries (84% of the 2020 global number of exclusively breastfed infants), reside in rural areas, live in households with less economic resources, with mothers not receiving any formal education, and were more likely to be female. The majority of regions attained between 25 and 50% of the improvements needed to reach the 2030 exclusive breastfeeding target of at least 70%.

Global data showed that nearly one in three women aged 15–49 years (29.9%) in 2019 were affected by anemia. The trend demonstrated a leveling off, if not a slight reversal, in progress made since 2012 (28.5%). This means that globally, there were 571 million anemic women, whose characteristics included being more likely to come from rural settings, poorer households and not having any formal education. Movement toward the 2030 target of a 30% reduction in anemia was deteriorating for the vast majority of countries representing almost all the regions, especially in North America, Europe, Australia and New Zealand, Oceania and Southeast Asia.

An increase was observed for adult obesity in all regions, augmenting worldwide from 11.8% in 2012 to 13.1% in 2016. Obese adults were more likely to come from upper-middle- or high-income countries (73% of the 2016 global burden, the last year for which data was available), and a higher prevalence was seen among women. Obese women were more likely to come from urban areas and wealthier households (**Fig. 9**).

The following figure summarizes the progress made toward reaching the 2030 Global Nutrition Targets. It depicts that since the year 2000, overweight in children under five, anemia in women, and adult obesity have been increasing, whereas low birthweight, under 5 child-stunting and exclusive breastfeeding have steadily improved (**Fig. 10**).

Severe Acute Malnutrition (SAM)

Severe Acute Malnutrition (SAM) was previously known as Protein energy malnutrition (PEM), or protein calorie malnutrition, and is an important nutrition condition in developing countries due to its high prevalence and its association with high rates of infant mortality, changes in growth and development, decreased ability to work, altogether leading to inadequate social and economic development.

SAM occurs when the diet does not provide sufficient protein, energy substrates (calories), or both, to satisfy the body's nutritional needs. It is usually also associated with a deficiency of vitamins and minerals, essential micronutrients needed for growth and

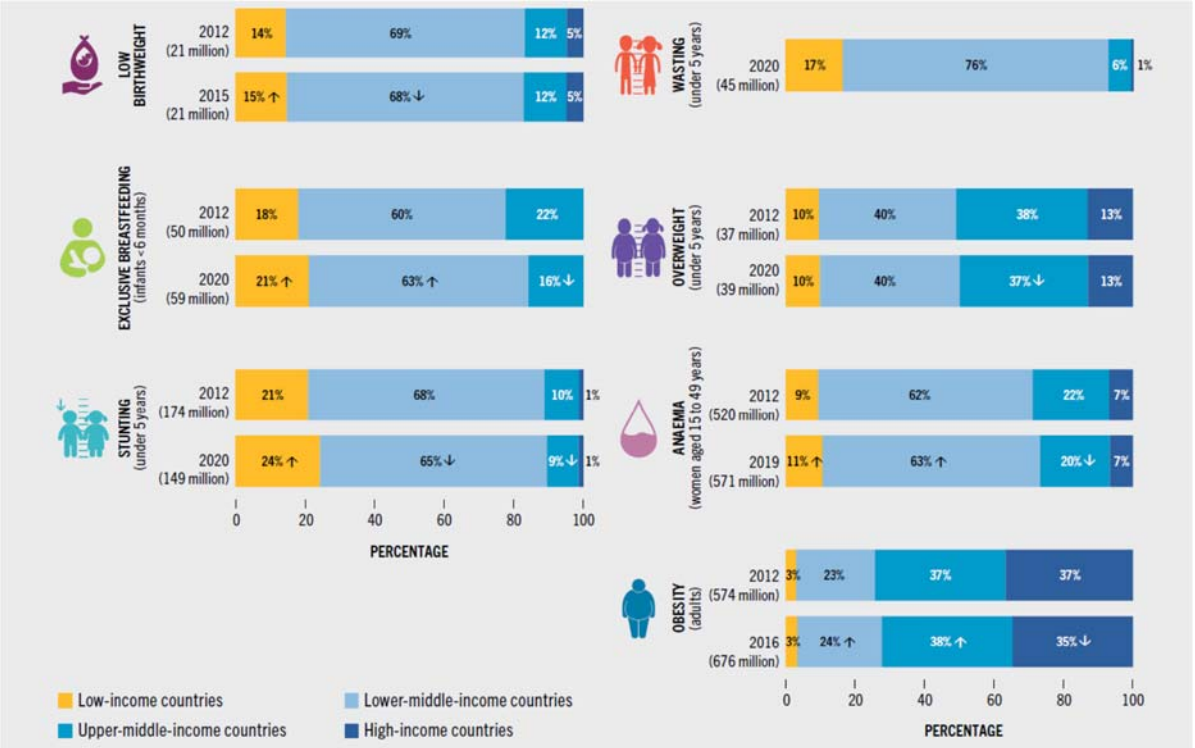


Fig. 9 Distribution of disease burden by country income level. Source: FAO et al. (2022).

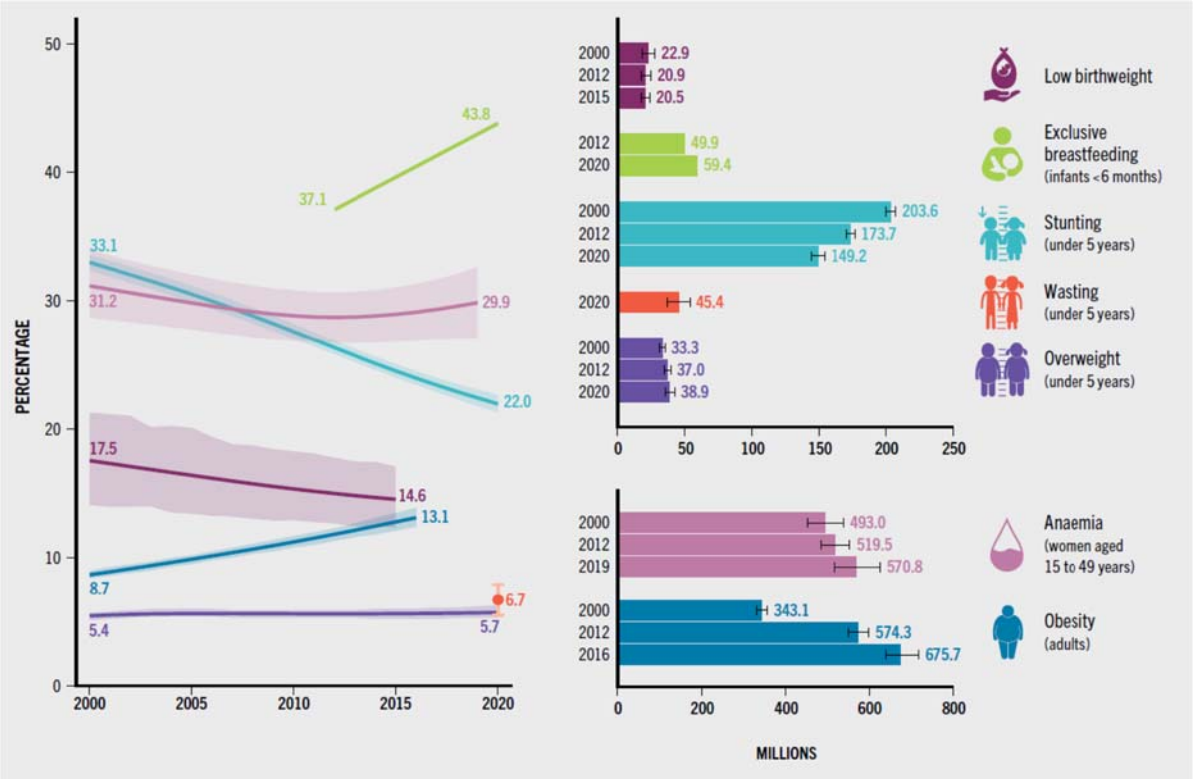


Fig. 10 Global Nutrition Targets. Trends in prevalence (%) and absolute numbers (millions). Source: FAO et al. (2022).

tissue renovation. The term SAM includes severe clinical syndromes of kwashiorkor (edema), marasmus (without edema), marasmic kwashiorkor (edema and the combination of chronic energy deficiency and chronic or acute protein deficiency), as well as mild and moderate cases that are much more numerous than severe forms. The term "malnutrition" is generally used in common vernacular to refer to SAM.

SAM may be of primary or secondary origin. Primary SAM is caused by inadequate nutrient intake. Secondary SAM results from disorders or drugs that interfere with nutrient use, such as cases of diseases accompanied by low food intake, inadequate nutrient absorption or utilization, increased nutritional requirements and/or increased nutrient losses. Although SAM is mainly an issue in early childhood, it may also occur in older children and adults, particularly the elderly.

Immunity and SAM

The main effects observed in SAM occur in the T-lymphocytes and the complement system. The number of lymphocytes originating in the thymus gland drastically decreases and the gland atrophies. The depletion of T-lymphocytes in the spleen and lymph nodes is observed as well. In SAM the production of certain complement components decreases. These deficiencies can account for the great propensity to sepsis by gram-negative bacteria. The resultant changes lead to increased susceptibility to infections and to severe complications.

Clinical manifestations and classification

Mild and Moderate Acute Malnutrition (MAM)

Its main clinical feature is weight loss with notable thinning accompanied by a reduction in subcutaneous adipose tissue. WHO/UNICEF defines Moderate acute malnutrition (MAM) as a weight-for-age between -3 and -2 z-scores below the median of the WHO child growth standards (see section on Anthropometric measures in this article). Criteria for Mid Upper Arm Circumference (MUAC) measurements is having a MUAC from 115 mm to <125 mm. MAM can be due to a low weight-for-height (wasting) or a low height-for-age (stunting) or to a combination of both. Similarly, moderate wasting and stunting are defined as a weight-for-height and height-for-age, respectively, between -3 and -2 z-scores.

MAM affects many children in poor countries. Children with moderate malnutrition are at increased risk of mortality and MAM is associated with a high number of nutrition-related deaths. If some of these moderately malnourished children do not receive adequate support, they may progress toward severe acute malnutrition (severe wasting and/or edema) or severe stunting (height-for-age less than -3 z-scores), which are both life-threatening conditions.

Children with mild to moderate malnutrition tend to be more sedentary and participate less frequently in games and activities that require some degree of physical exertion. Older children and adolescents may show a reduction in their ability to perform strenuous physical work for prolonged periods. Other non-specific manifestations include pallor, apathy, shortened attention span (particularly in school settings) and frequent episodes of diarrhea. Clinical laboratory tests do not contribute to the diagnosis of mild or moderate malnutrition.

Severe Acute Malnutrition (SAM)

In addition to clinical manifestations, the dietary and clinical history of the patient are important for diagnosis. Marasmus is usually associated with food shortages, prolonged semi-starvation, premature weaning, and extended time intervals between the administration of foods to infants and young children. Kwashiorkor is often associated with low protein intake, late weaning with inadequate complementary feeding practices and current or recent episodes of diarrhea or measles. Various bouts of infectious diseases and chronic or recurrent diarrhea occur frequently. WHO/UNICEF classification of SAM is based on presenting with weight-for-height (W/H) < -3 SD or MUAC <115 mm or edema, of which any are accompanied by complications such as lack of appetite, fever $\geq 38.5\,^{\circ}C$, hypothermia $\leq 35.5\,^{\circ}C$, vomiting or severe dehydration. Wasting refers to children that are too thin for their height and is a common manifestation of SAM.

Marasmus

Marked muscle wasting and the extreme reduction of subcutaneous fat in cases with severe non-edematous SAM present with a "skin and bones" appearance. Children with marasmus frequently have 60% or less of the expected weight-for- height and suffer from short stature. Their hair tends to be dry, thin, dull and at times lacking in quantity (Fig. 11).

Children with marasmus often show complications such as acute gastroenteritis, dehydration, respiratory infections and eye lesions due to hypovitaminosis A. Systemic infections can lead to septic shock or disseminated intravascular coagulation syndrome with high mortality rates. Obtaining a medical history plays an important role for the differential diagnosis of SAM secondary to AIDS, tuberculosis, cancer, and other debilitating diseases.

Kwashiorkor

The main feature is mild edema, being palpable without pain and usually presenting in the feet, ankles and legs, but in severe cases possibly extending to the perineum, upper extremities and the face. Many patients have pellagroid skin lesions at the edema sites



Fig. 11 Nutritional marasmus with severe wasting. Note: Original photo, no need to cite source.

and in areas of pressure or frequent irritation. The skin can be erythematous and shiny in the regions of edema, alternating with areas showing dry, hyperkeratosis and hyperpigmentation that tend to converge. Where the epidermis becomes dark and dry, it splits open when stretched to reveal pale areas between the cracks ("lacquered flaky paint", "crazy pavement dermatosis"), leaving exposed tissues that become infected easily. Subcutaneous fat tends to be preserved, and there may be some degree of muscle wasting. Weight loss, after correcting for edema, is usually not as severe as in marasmus. Children may have short or normal stature, depending on the length of the current episode and their nutritional history (Fig. 12).

The hair is dry, brittle, dull, and can be pulled out easily without causing pain. Curly hair becomes straight and pigmentation changes usually acquiring a brownish, reddish or yellow-white tone. Patients who alternately undergo periods of very poor and relatively good intakes of protein, may present with alternating bands of depigmented and normal hair.

Patients have the same complications as described for marasmus cases, but diarrhea and respiratory and skin infections are more frequent and severe. Apart from diarrhea, they frequently present with anorexia and postprandial vomiting. These general conditions improve as nutrition therapy progresses, without specific treatment for gastrointestinal disorders. Kwashiorkor cases may have infections without fever, tachycardia, shortness of breath or leukocytosis, making diagnosis difficult. The most frequent causes of death are pulmonary edema with bronchopneumonia, gastroenteritis, septicemia and electrolyte disturbances. A differential diagnosis of other conditions that cause hypoproteinemia and edema should be carried out, as well as secondary undernutrition due to alterations in metabolism and protein absorption.



Fig. 12 Kwashiorkor. Note: As this is public domain, does it need to be cited Source: NARA, Public domain, via Wikimedia Commons.

Marasmic-kwashiorkor

This form of edematous SAM combines the clinical features of marasmus and kwashiorkor. The major manifestations are the edema of kwashiorkor with or without skin lesions accompanied by the muscle wasting and subcutaneous fat reduction characteristic of marasmus. In the initial stages of treatment once the edema is resolved, the patient acquires an appearance that is very similar to the child with marasmus. The patient's biochemical profile is characteristic of kwashiorkor and marasmus but the alterations due to severe protein deficiency are those that predominate.

Long term consequences

Cognitive development studies indicate a close interrelationship between undernutrition and intellectual development. In recent years there has been progress in improved recovery of neurological development in severely undernourished infants who are systematically stimulated as part of their treatment. Augmented deterioration manifesting in severely undernourished children is usually due to electrolyte imbalance, gastrointestinal disorders, cardiovascular or kidney insufficiency, and compromised immune function to fight off infections. The psychological consequences can also be serious. Severe anorexia, apathy and irritability make it difficult to feed and handle the child, making prognosis more precarious.

Generally speaking, death is secondary to severe infections, bronchopneumonia, gram-negative septicemia, and acute cardiovascular, hepatic and renal failure. The clinical picture may be aggravated by severely insufficient folate, niacin and thiamin levels and/or acute deficiency of potassium, sodium and magnesium as well as chronic deficiency of iron, copper, zinc and chromium. The signs and symptoms of severe vitamin and mineral deficiencies can also play a critical role and may appear during early rehabilitation if an inadequate supply of these nutrients is not provided.

Chronic undernutrition

Chronic undernutrition is distinguished from the acute forms (SAM, MAM). A child who suffers from chronic undernutrition experiences delay in his/her growth. This growth retardation is measured by comparing the height of the child with the standard recommended for their age. It indicates a lack of the necessary nutrients over an extended period of time, and thus the increased risk of becoming ill and the subsequent effects of these events on the physical and intellectual development of the child. Chronic undernutrition, although a problem of greater magnitude in terms of the number of children affected, has up until recently received less attention than more acute forms of undernutrition.

Stunting

Chronic undernutrition is reflected in stunting, the indicator measuring length/height-for -age and is defined as having a length/height-for-age Z-score that is more than two standard deviations below the age-sex median for a well-nourished reference population. Stunting is a largely irreversible outcome of chronic inadequate nutrition and repeated bouts of infection during the first 1000 days of a child's life.

Specifically, factors that impact on stunted growth and development include poor maternal nutrition and health, inadequate infant and young child feeding practices (non-exclusive breastfeeding and complementary feeding that is limited in quantity, quality and variety), and infection. Special importance is placed on maternal nutritional and health status before, during and after pregnancy, which determines a child's early growth and development, beginning in the womb. Maternal contributors to stunting include inadequate nutrition during pregnancy leading to low birth weight, maternal short stature, short birth spacing, and adolescent pregnancy. The latter impairs nutrient availability to the fetus due to the competing demands of the adolescent mother's own growth.

Stunting has long-term effects on individuals and societies, including: diminished cognitive and physical development, reduced productive capacity and poor health, and an increased risk of degenerative diseases such as diabetes. In fact, stunted children who experience rapid weight gain after 2 years of age show an increased risk of becoming overweight or obese later in life, thus increasing risk of coronary heart disease, stroke, hypertension and type 2 diabetes onset. As for the economic impact, the World Bank estimates show that a 1% loss in adult height due to childhood stunting is associated with a 1.4% loss in economic productivity.

The role of fetal growth restriction

The most common definition of intrauterine growth restriction (IUGR) is a fetal weight that is below the 10th percentile for gestational age as determined through an ultrasound. This can also be called small-for gestational age (SGA) or fetal growth restriction. Findings in the Lancet 2013 series highlight the impact of SGA on attributable neonatal deaths. 32 million babies were born SGA, and about 800,000 neonatal deaths and 400,000 post-neonatal infant deaths could be attributed to the increased risk associated with having fetal growth restriction. This relatively recent data contradicts the widespread assumption that SGA infants, as compared to preterm babies, are not at a substantially increased risk of mortality. Additionally, infants who are SGA have an increased risk of growth faltering in the first 2 years of life. Causes of childhood growth faltering are multifactorial, but fetal growth restriction might

be an important contributor to stunting and wasting in children. The Lancet 2013 Maternal and Child Nutrition series demonstrated that about 20% of childhood stunting could have its origins in the fetal period, as shown by being born SGA. As such, there is now impetus to focus attention on preventing SGA by scaling up prenatal nutrition interventions. In 2010, 32.4 million babies were born SGA, with 27% of all births occurring in Low and Middle Income Countries (LMIC).

Anthropometric indicators for measuring malnutrition

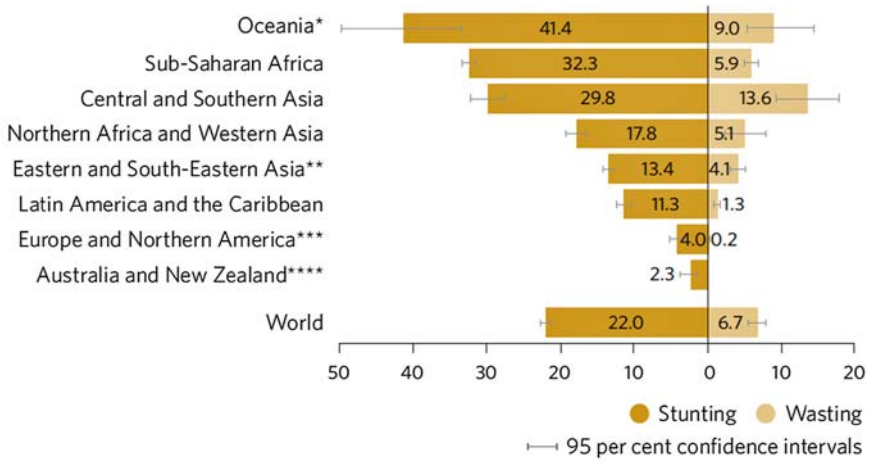
Several measures and anthropometric indices are used to diagnose and classify malnutrition. The choice of a measure or index of choice will depend on its simplicity, accuracy and sensitivity, as well as the availability of instruments to make measurements and the existence of reference standards for interpreting the results. In clinical practice the utilization of reference growth charts for weight, height and BMI for age and weight-for-height is common in the pediatric population. Growth charts consist of a series of percentile curves that illustrate the distribution of selected body measurements in children. The reference standards for anthropometric measurements presented in this article are those used in WHO's Nutrition Landscape Information System (NLIS), a comprehensive data collection system that compiles nutrition-related indicators in standardized form.

Malnutrition in children

The measurement of weight and height is used for the calculation of indicators such as "weight-for-age", "height-for-age", "weight for height" and the body mass index (BMI = weight kg/height 2 in meters). Other anthropometric measures employed are the measurements of arm circumference and head circumference. In statistical terms, it is more appropriate to use standard deviations ("Z-scores") or percentiles to evaluate differences between the patient and the reference standard. This is the preferred expression for anthropometric indicators in surveys. It is the difference between the value for an individual and the median value of the reference population for the same age or height, divided by the standard deviation of the reference population. In other words, by applying Z-scores, one can describe how far a child's weight is from the median weight of a child at the same height in the reference value (Fig. 13).

Underweight, stunting and wasting

These indicators are used to measure nutritional disorders resulting in undernutrition (assessed from underweight, wasting and stunting) and overweight. Child growth is globally acknowledged as an important indicator of nutritional status and health in populations. The percentage of children with a low height -for- age (stunting) reflects the accumulated effects of undernutrition and infections since and even preceding birth. Thus it can be interpreted as an indication of poor environmental conditions or long-term restriction of a child's potential to grow. The percentage of children who have low weight-for- age (underweight) can reflect "wasting" (i.e. low weight-for-height), indicating acute weight loss, "stunting", or both. As such, "underweight" is a composite indicator and may therefore be difficult to interpret.



* Excluding Australia and New Zealand.
** Excluding Japan for the wasting estimate.
*** The wasting estimate represents Northern America since population coverage for Europe was insufficient (<50 per cent) .
**** Wasting estimates are not available due to insufficient (< 50 per cent) population coverage.

Fig. 13 Proportion of children under age 5 who are affected by stunting and wasting, 2020 (percentage). Source: United Nations (2022).

Definitions used

- Underweight (thinness): weight for age <−2 standard deviations (SD) of the WHO Child Growth Standards median (Severe underweight: <−3 SD)
- Stunting (short for age): height for age <−2 SD of the WHO Child Growth Standards median (severely stunted: <−3 SD for length/height for age). Children under 2 are usually measured lying down (recumbent length) and those over 2, standing height is taken.
- Wasting: weight for length/height <−2 SD of the WHO Child Growth Standards median (Severe wasting <−3 SD for weight for length/height or BMI-for-age)
- Overweight: weight for height >+2 SD of the WHO Child Growth Standards median

Interpretation

Underweight

As weight is easy to assess, this has been the most widely collected indicator. Evidence has shown that the mortality risk of children who are even mildly underweight is increased, and severely underweight children are at even greater risk.

Stunting

Children who suffer from growth retardation as a result of poor diets or recurrent infections tend to be at greater risk for illness and death. Stunting is the result of long-term nutritional deprivation and often results in delayed mental development, poor school performance and reduced intellectual capacity. This in turn affects economic productivity at the national level. Women of short stature are at greater risk for obstetric complications because of a smaller pelvis. Small women are at greater risk of delivering an infant with low birth weight, contributing to the intergenerational cycle of malnutrition, as infants of low birth weight or retarded intrauterine growth tend be smaller as adults.

Wasting

Wasting in children reflects acute undernutrition, usually as a consequence of insufficient food intake or a high incidence of infectious and parasitic diseases, especially diarrhea. In addition, it impairs immune system functioning and can lead to increased susceptibility, severity and duration of infectious diseases as well as increased mortality risk. In the WHO protocols for Severe Acute Malnutrition management, Mid Upper Arm Circumference is also assessed, along with the presence/absence of edema.

Overweight

Childhood obesity is linked with a higher probability of obesity as an adult, which can result in a variety of disabilities and pathologies, such as diabetes and cardiovascular diseases.

Malnutrition in adults

The values for body mass index (BMI) are age-independent for adult populations and are similar for both males and females. However, differences in BMI may not correspond to the same degree of fatness in different populations as there may be variations in body proportions.

Moderate and severe thinness, underweight, overweight, obesity

- BMI < 17.0 indicates moderate and severe thinness
- BMI < 18.5 indicates underweight
- BMI 18.5–24.9 indicates normal weight
- BMI ≥ 25.0 indicates overweight
- BMI ≥ 30.0 indicates obesity

Moderate and severe thinness

A BMI < 17.0 indicates moderate and severe thinness in adult populations. This has been associated with increases in illness in adults studied in three continents and this is considered a reasonable further limit to choose as a cut-off point for moderate risk. A BMI < 16.0 is associated with a markedly increased risk for ill health, poor physical performance, lethargy and even death; as such this cut-off point is considered a valid extreme limit.

Underweight

The cut-off point of 18.5 for underweight in both genders has less scientific validity as a cut-off point for moderate and severe thinness but is considered a reasonable value for use pending further, comprehensive studies. The proportion of the population with a low BMI that is considered a public health problem is closely linked to the resources available for correcting the problem, the stability of the environment and government priorities. Approximately 3–5% of a healthy adult population have BMIs < 18.5.

Micronutrient malnutrition

Micronutrient deficiencies

More than 2 billion people are deficient in micronutrients such as vitamin A, iodine, iron, zinc and folic acid. Malnutrition due to the lack of vitamins and minerals (micronutrients) can make its appearance in a variety of ways. Fatigue, reduced learning capacity or impaired immunity are just a few examples. These nutrients are called micronutrients since the body requires them in small amounts for growth, development and to maintain health. The combination of different ecological factors such as climate, low production, availability and consumption of certain micronutrient rich foods, cultural patterns that limit the consumption of given foods and poverty induced food restrictions contribute to mapping a specific region's particular micronutrient and health profile. In regions with a high prevalence of undernutrition, associated micronutrient deficiencies occur, which contribute to the symptoms that appear. Or they may remain in an asymptomatic subclinical phase, without apparent clinical expression, thus giving rise to the expression "hidden hunger". This represents the phenomenon of a chronic lack of vitamins and minerals that often has no visible warning signs, so that those affected may not even be aware of it. Nevertheless, micronutrient deficiencies inevitably lead to significant negative health consequences.

Recent data from the 2021 Lancet series on maternal and child undernutrition show a high prevalence of vitamin A deficiency in Africa and South Asia. Data for zinc is limited but the few countries with this information suggest that almost 50% of all children are affected by zinc deficiency. Likewise, an estimated 60% of children under 5 years of age in LMIC have anemia (with higher rates for those aged 6–24 months), and very few improvements have been observed over the past decade. This review suggests that the high prevalence of micronutrient deficiencies in LMICs can be attributed to inadequate consumption of nutritive complementary foods as well as excessive losses caused by infectious conditions such as diarrhea.

Unlike many other impediments to social and economic development, micronutrient deficiencies can be reduced with relatively inexpensive investments in public health, agriculture and education. The technology exists to solve many of these deficiency states. However, they persist for various reasons, including the lack of awareness among policy makers on the importance of resolving the issues, as well as the lack of knowledge by program planners about their consequences and the strategies available to combat them.

In recent decades the research community, governments, development agencies and non-governmental organizations (NGOs) have made considerable progress in identifying the population groups that have the greatest risk of micronutrient deficiencies and in developing programs for short term improvement of given deficiencies. Food-based strategies have been emphasized, focusing on the promotion of producing and consuming micronutrient-rich foods as a sustainable solution to the problem of micronutrient deficits. Actions include improving access, availability and the consumption of foods rich in vitamins and minerals. The benefits of these food based strategies are that they not only address improving the intake of certain nutrients, but also enhance the overall situation in terms of food and health.

Globally, the three micronutrients having the greatest repercussions for public health are vitamin A, iron and iodine as their deficiency poses a major threat to the health and development of populations worldwide, particularly children and pregnant women in low-income countries. The magnitude of their impact can be seen in the following figures. Other micronutrients of relevance also exist, such as folate, zinc and calcium and the reader is referred to the supplementary materials for more information on the extensive topic of micronutrient deficiencies. The international community has emphasized these micronutrients since it is very difficult to meet their recommended intake levels without having access to an adequate variety of foods that ensure diet diversity (**Figs. 14–16**).

Conclusions

Decreasing wasting, stunting, underweight, micronutrient deficiencies, overweight, obesity and diet-related NCDs is critical for achieving several of the Sustainable Development Goals. However, no single country is on track to meet global nutrition targets. Although progress has been made since 2012, improvements have been modest and advances in reducing undernutrition in women and children have been fluctuating. Maternal and child undernutrition yield inadmissibly high prevalence rates, especially for low-income countries and for the poorest population segments in middle-income countries. The most recent evaluations show improvements in reducing stunting in children and underweight in women, as well as reductions in iodine deficiency-primarily through actions involving food fortification. Wasting has improved in certain countries, but still remains an important challenge in others. Although there have been small gains in exclusive breastfeeding rates, partial breastfeeding after 1 year of age shows no improvement. Low birthweight, anemia in children and women and zinc deficiency have also remained largely unchanged, whereas obesity is increasing globally. Socioeconomic inequalities persist and continue to curb the rate of progress worldwide. Moreover, the COVID-19 pandemic has disrupted health systems and, together with the conflict in Ukraine, has worsened household food insecurity, reversed economic growth, and could set back advances made in undernutrition across LMICs.

Achieving the 2030 global nutrition targets will require massive and coordinated efforts among key stakeholders. Fortunately, we have evidence based data to guide which actions will be effective that can be scaled up and maximize the utilization of resources to address the challenges presented. The following figure summarizes the recent review on recommended interventions published in the Lancet 2021 Series on maternal and child undernutrition. In addition, the Series also highlights the importance of providing critical nutrition interventions within the first 1000 days of life (early pregnancy up until the first 2 years of child life). It is necessary

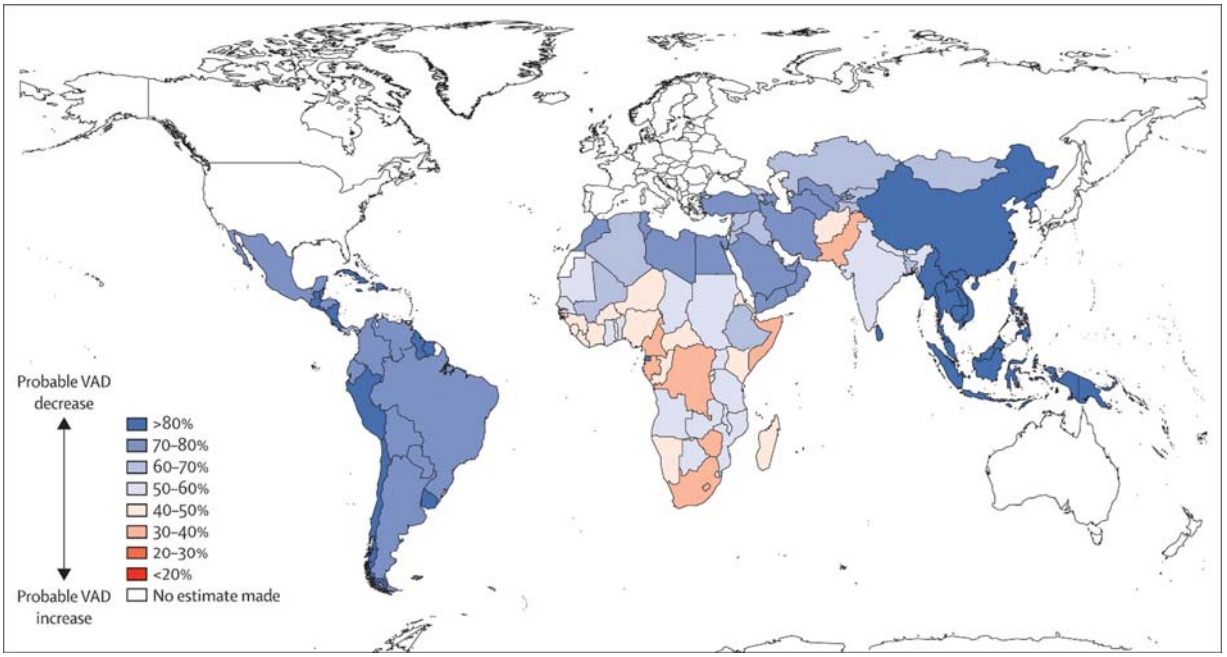


Fig. 14 Estimated global prevalence in of children under five with biochemical Vitamin A Deficiency (VAD). Source: Stevens et al. (2015).

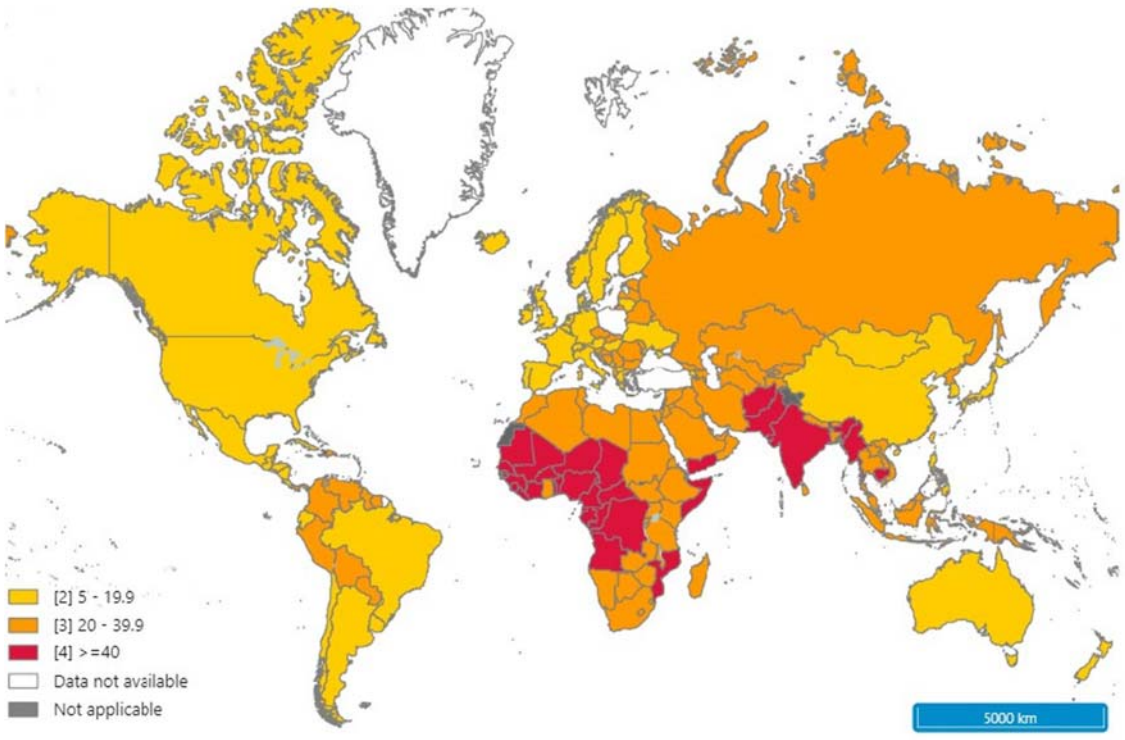


Fig. 15 Prevalence of anemia among women aged 15–49 years by country, 2019. Source: WHO (2019).

to strive for continued political commitment and cooperation among the wide range of relevant stakeholders, both in the public and private sector, and to implicate those in the area of nutrition as well as in other fields. Governments, international partners, civil society, non-governmental organizations and the private sector have a key role in the creation of healthy environments and conditions providing affordable, accessible and sustainable healthy dietary choices for communities and especially for those most at risk of malnutrition in all its forms (Fig. 17).

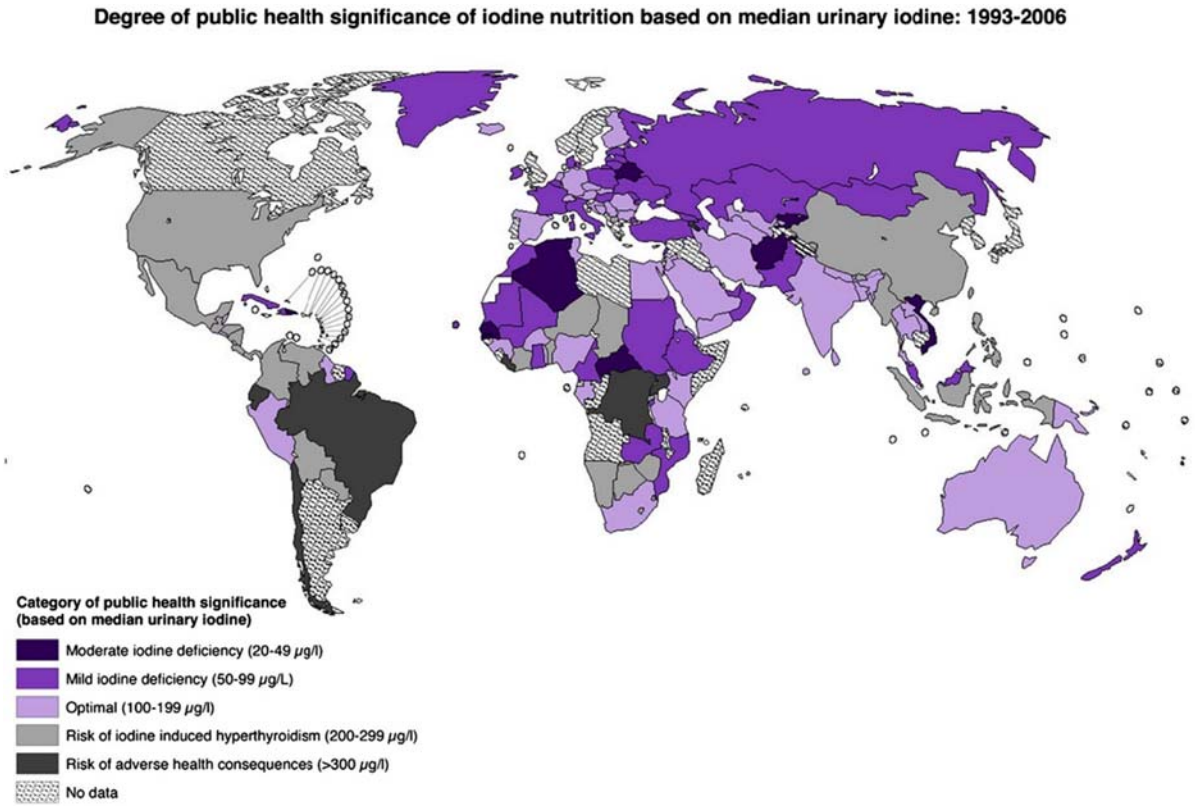


Fig. 16 Global iodine public health significance. Source: de Benoist et al. (2008).

Strong evidence for implementation	• Multiple micronutrient supplementation in pregnancy • Kangaroo mother care for preterm and low birthweight newborn babies • Delayed cord clamping for preterm newborn babies • Breastfeeding promotion and counselling • Complementary feeding education with food provision in food insecure populations • Vitamin A supplementation for children in vitamin A-deficient contexts • Therapeutic zinc supplementation for diarrhoea management • Small-quantity lipid-based nutrient supplements for growth among children • Ready-to-use supplementary food for management of acute malnutrition • Family planning and birth spacing* • Insecticide-treated bednets for the control of malaria* • Large-scale food fortification for the prevention of micronutrient deficiencies†
Moderate evidence for implementation	• Water, sanitation, and hygiene interventions‡ • Calcium supplementation in pregnancy in low intake populations • Balanced-energy protein supplementation in pregnancy for women who are undernourished • Complementary feeding education without food provision in food secure populations • Preventive zinc supplementation to reduce diarrhoea incidence • Micronutrient powders to reduce iron deficiency and anaemia among children
Weak evidence for implementation	• Food distribution programmes during pregnancy • Kangaroo mother care for term newborn babies
Emerging evidence	• Probiotics for preterm and low birthweight newborns • Emollient use (ie, coconut oil) for preterm and low birthweight newborns

Fig. 17 Recommended interventions to address malnutrition, according to strength of evidence.*Indirect health sector nutritional interventions. †Direct nutritional interventions outside of the health-care sector. ‡Indirect nutritional interventions outside of the health-care sector. All other points are direct health sector nutritional interventions. Source: Keats et al. (2021).

References

Darnton-Hill, I., Nishida, C., James, W.P.T., 2004. A life course approach to diet, nutrition and the prevention of chronic diseases. Publ. Health Nutr. 7 (1A), 101–121.
de Benoist, B., McLean, E., Andersson, M., Rogers, L., 2008. Iodine deficiency in 2007: global progress since 1993. Food Nutr. Bull. 29 (3), 195–202.
Development Initiatives, 2018. 2018 Global Nutrition Report: Shining a Light to Spur Action on Nutrition. Development Initiatives, Bristol, UK.
Development Initiatives, 2021. 2021 Global Nutrition Report: The State of Global Nutrition. Development Initiatives, Bristol, UK.
FAO, IFAD, UNICEF, WFP, WHO, 2022. The State of Food Security and Nutrition in the World 2022. Repurposing Food and Agricultural Policies to Make Healthy Diets More Affordable. FAO, Rome.
Horton, S., Hoddinott, J., 2014. Benefits and Costs of the Food and Nutrition Targets for the Post-2015 Development Agenda. Copenhagen Consensus Center Working Paper. Copenhagen Consensus Center, Copenhagen.
Keats, E.C., Das, J.K., Salam, R.A., et al., 2021. Effective interventions to address maternal and child malnutrition: an update of the evidence. Lancet Child Adolesc. health 5 (5), 367–384.
Sharrow, D., Hug, L., You, D., et al., 2022. Global, regional, and national trends in under-5 mortality between 1990 and 2019 with scenario-based projections until 2030: a systematic analysis by the UN Inter-agency Group for Child Mortality Estimation. Lancet Glob. Health 10 (2), e195–e206. https://doi.org/10.1016/S2214-109X(21)00515-5.
Stevens, G.A., Bennett, J.E., Hennocq, Q., et al., 2015. Trends and mortality effects of vitamin A deficiency in children in 138 low-income and middle-income countries between 1991 and 2013: a pooled analysis of population-based surveys. Lancet Glob. Health 3 (9), e528–e536.
United Nations, 2022. The Sustainable Development Goals Report 2022. United Nations, New York.
United Nations Children's Fund (UNICEF), 2020. Nutrition, for Every Child: UNICEF Nutrition Strategy 2020–2030. UNICEF, New York.
WHO, 2019. WHO Global Targets 2025 Tracking Tool: Indicator Mapping—Anemia. Available at. https://apps.who.int/nutrition/trackingtool/en/. (Accessed 15 June 2022).
WHO, UNICEF, 2017. The Extension of the 2025 Maternal, Infant and Young Child Nutrition Targets to 2030. Discussion Paper. WHO, Geneva, Switzerland.
World Food Program (WFP), Centers for Disease Control and Prevention (CDC), 2005. A Manual: Measuring and Interpreting Malnutrition and Mortality. World Food Program, Rome.

Further reading

Atinmo, T., Mirmiran, P., Oyewole, O.E., Belahsen, R., Serra-Majem, L., 2009. Breaking the poverty/malnutrition cycle in Africa and the Middle East. Nutr. Rev. 67 (Suppl. 1), S40–S46.
Bailey, R.L., West Jr., K.P., Black, R.E., 2015. The epidemiology of global micronutrient deficiencies. Ann. Nutr. Metab. 66 (Suppl. 2), S22–S33.
Headey, D., Heidkamp, R., Osendarp, S., et al., 2020. Standing together for nutrition consortium. Impacts of COVID-19 on childhood malnutrition and nutrition-related mortality. Lancet 396 (10250), 519–521. https://doi.org/10.1016/S0140-6736(20)31647-0.
HLPE, 2017. Nutrition and Food Systems. A Report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security. Rome.
Katz, J., Lee, A.C.C., Kozuki, N., et al., 2013. Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: a pooled country analysis. Lancet 382, 417–425.
Lancet, 2008. Series on Maternal and Child Nutrition. Available at. http://www.thelancet.com/series/maternal-and-child-undernutrition.
Lancet, 2013. Series on Maternal and Child Nutrition. Available at. http://www.thelancet.com/series/maternal-and-child-nutrition.
Lancet, 2021. Series on Maternal and Child Undernutrition Progress. Available at. https://www.thelancet.com/series/maternal-child-undernutrition-progress.
Sáez-Almendros, S., Obrador, B., Bach-Faig, A., Serra-Majem, L., 2013. Environmental footprints of Mediterranean versus Western dietary patterns: beyond the health benefits of the Mediterranean diet. Environ. Health 12, 118. https://doi.org/10.1186/1476-069X-12-118.
Schwarzenberg, S.J., Georgieff, M.K., Committee on Nutrition, 2018. Advocacy for improving nutrition in the first 1000 days to support childhood development and adult health. Pediatrics 141 (2), e20173716. https://doi.org/10.1542/peds.2017-3716.
United Nations Children's Fund (UNICEF), World Health Organization, International Bank for Reconstruction and Development/The World Bank, 2021. In: Levels and Trends in Child Malnutrition: Key Findings of the 2021 Edition of the Joint Child Malnutrition Estimates. World Health Organization. License: CC BY-NC-SA 3.0 IGO, Geneva.
United Nations Sustainable Development Goals. Available at https://www.un.org/sustainabledevelopment/.
United Nations System Standing Committee on Nutrition, 2017. By 2030, End All Forms of Malnutrition and Leave No One behind. Discussion Paper. United Nations System Standing Committee on Nutrition, Rome.
Victora, C.G., Christian, P., Vidaletti, L.P., et al., 2021. Revisiting maternal and child undernutrition in low-income and middle-income countries: variable progress towards an unfinished agenda. Lancet 397 (10282), 1388–1399.
WHA Global Nutrition Targets 2025. https://www.who.int/teams/nutrition-and-food-safety/global-targets-2025.
WHO Severe Acute Malnutrition Protocols. Available at https://apps.who.int/nutrition/topics/severe_malnutrition/en/index.html.

Relevant websites

Copenhagen Consensus on Hunger and Malnutrition. http://www.copenhagenconsensus.com/research-topic/hunger-and-malnutrition.
FANTA (Food and Nutrition Technical Assistance) Project. http://www.fantaproject.org/publications.
FAO The State of Food Insecurity in the World (SOFI) 2022. http://www.fao.org/state-of-food-security-nutrition.
FAO/WHO Second International Conference on Nutrition (ICN2). https://www.fao.org/documents/card/en/c/2c14f1da-ba0b-44b4-adc8-3c6697589106/.
Global Nutrition Report. http://globalnutritionreport.org/.
Nutrition Without Borders. https://www.nutricionsinfronteras.org/.
Scaling Up Nutrition (SUN). http://scalingupnutrition.org/.
Sustainable Development Goals. https://www.un.org/sustainabledevelopment/.
UNICEF-WHO-The World Bank: Joint child malnutrition estimates. https://data.unicef.org/resources/jme-report-2020/.
United Nations Standing Committee on Nutrition. https://www.unscn.org/.
WHO Nutrition Landscape Information System (NLiS). https://www.who.int/teams/nutrition-and-food-safety/databases/nutrition-landscape-information-system.
WHO Training Course on Child Growth Assessment. https://www.who.int/publications/i/item/9789241595070.
WHO Vitamin and Mineral Nutrition Information System (VMNIS). https://www.who.int/teams/nutrition-and-food-safety/databases/vitamin-and-mineral-nutrition-information-system.

Malnutrition: Secondary, diagnosis and management

L.H. Allen, University of California, Davis, CA, USA

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of L.H. Allen, Malnutrition: Secondary, Diagnosis and Management, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 143–147, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00181-1.

Introduction **507**
Causes of Secondary Under- and Overnutrition 507
Coexistence of Primary and Secondary Malnutrition **509**
The Reverse Paradigm: Underlying Pathology Revealed by Detection of Abnormal Nutrition **509**
Diagnosis of Secondary Malnutrition **509**
Caveats for the Diagnosis of Secondary Excess Nutriture 509
Caveats for the Diagnosis of Secondary Undernutrition 510
Management of Secondary Malnutrition **510**
Principles of Management 510
Public Health Approaches 511
Dietary Management of Secondary Overnutrition 511
Dietary and Nutritional Management of Secondary Undernutrition 511
Conclusions **512**
Further Reading **512**

Glossary

Cachexia Loss of weight, and fat and muscle mass, caused by disease and loss of appetite in people who are not trying to lose weight.

Enteral feeding Delivery of nutrients, in the form of special solutions, to the stomach or small intestine through a tube, i.e. 'tube-feeding'.

Nosocomial pneumonia Pneumonia contracted while in hospital, 2–3 days after admission.

Parenteral nutrition Delivery of nutrients through a tube inserted into a vein, because the intestine is nonfunctional.

Pathognomonic Characteristic or diagnostic of a specific disease.

Percutaneous Delivery of nutrients into the stomach or intestine by a tube passed through the abdominal wall.

Primary malnutrition Malnutrition due to inadequate or excessive food intake.

Secondary malnutrition Malnutrition caused primarily by illness, infections, or disease.

Introduction

In its broadest context, malnutrition is a state of having an inappropriate nutritional status with respect to one or more macronutrients (water, electrolyte, protein, or fat) or micronutrient (vitamin or mineral) constituent of the body. This imbalance can be a deficit, leading to an insufficient supply or content of the nutrient (undernutrition), or an excess, leading to an excessive content or overloading of the organism with a nutrient (overnutrition). The six possible causes for all nutrient deficiencies have been summarized as: inadequate intake, impaired absorption, increased wastage, impaired utilization, increased destruction, and elevated requirements. Correspondingly, overnutrition and excesses can result from reciprocal defects, that is: Hyperphagia, hyperabsorption, increased retention; decreased destruction, and decreased requirements.

As discussed in the previous article, the term 'primary' malnutrition relates almost exclusively to the first of these mechanisms, that of inadequate or excessive ingestion of nutrients from the diet. It is concerned with food consumption and intake. Secondary malnutrition, in contrast, concerns the disturbed and disordered handling of nutrients. When diseases or abnormal physiological conditions interfere with the normal disposition of nutrients ingested from the diet, this is the basis of a situation of 'secondary' malnutrition.

Causes of Secondary Under- and Overnutrition

A representative, but not exhaustive, list of diseases and conditions producing secondary undernutrition is provided in **Table 1**, and of secondary overnutrition in **Table 2**.

Table 1 Diseases and conditions associated with secondary macronutrient or micronutrient undernutrition

Inadequate nutrient absorption
Gastric abnormalities: Gastric atrophy
Pernicious anemia
Intestinal abnormalities: Celiac disease
Inflammatory bowel disease
Intestinal cryptosporidiasis
Radiation enteritis
Chronic intestinal pseudo-obstruction HIV/AIDS
Hepatobiliary abnormalities: Cystic fibrosis
Biliary obstruction
Pancreatic insufficiency
Increased nutrient excretion
Gastric disorders: Gastric adenoma
Intestinal abnormalities: Laxative abuse
Peptic ulcer
Gastrointestinal fistula
Colonic adenoma
Amebiasis
Hookworm
HIV/AIDS
Schistosomiasis
Hepatic disorders: Hepatic cirrhosis
Endocrine disorders: Diabetes mellitus
Hypoaldosteronism
Renal Disorders: Fanconi syndrome
Hemodialysis; peritoneal dialysis
Increased destruction or use of nutrients
Endocrine disorders: Hyperthyroidism
Chronic disease: Cardiac cachexia
Cancer cachexia
Cystic fibrosis
Bone marrow transplants
Infections: Pulmonary tuberculosis
HIV/AIDS
Decreased utilization of nutrients
Lead poisoning
Menkes' copper storage disease

The basis for suspecting secondary malnutrition is that there is evidence of deficiency or excess, but foods and nutrients are presumably being consumed in adequate amounts. Once the suspicion emerges, three distinct diagnostic principles need to be addressed: (1) the confirmation of dietary intake, and estimation of its adequacy; (2) the diagnosis and classification of abnormal nutritional status; and (3) the diagnosis of the functional, physiological, or pathological origins of disordered nutrient disposition.

Table 2 Diseases and conditions associated with secondary macronutrient or micronutrient excess (overnutrition)

Increased nutrition absorption
Wilson's disease
Hemochromatosis
Increased nutrition retention
Prader–Willi syndrome
Hypercorticosteroidism
Hyperpituitarism
Acute tubular necrosis
Chronic renal failure
Decreased destruction of nutrients
Hypothyroidism

The evaluator must remain attuned to the nutritional status of patients, clients, or populations, and sensitive to the possibility of a nonprimary origin of any under- or overnutrition.

Coexistence of Primary and Secondary Malnutrition

It is important to recognize the potential for the simultaneous coexistence of primary and secondary malnutrition in the same individual. Primary malnutrition in the free-living populations can be associated with famine conditions (crop failure, conflict, natural disaster, and refugee crisis), in which sufficient food is simply not available. Alternatively, it can arise from the poverty of landlessness or urban margination, where food is not accessible given the household income. A large number of communicable diseases with consequences of poor nutrient absorption, retention or utilization, such as parasitoses, tuberculosis or HIV/AIDS, are common in these situations of deprivation and misery. To the extent that a disease process produces anorexia or dysphagia, or even psychic depression, the net effect is to reduce total intake of dietary energy and nutrients. Whatever, malabsorptive or nutrient-wasting components of the underlying disorders will further compromise the nutritional state.

The Reverse Paradigm: Underlying Pathology Revealed by Detection of Abnormal Nutrition

In clinical medicine, a type of 'reversal of roles' often occurs. Rather than primarily recognizing the presentation of the underlying pathology, recognition of an abnormal nutritional status without a dietary explanation leads to the diagnosis of the underlying disorder before any specific (pathognomonic) sign or symptom has yet occurred. For instance, the Prader–Willi syndrome of pathological obesity would initially present as common obesity. Similarly, in hypercorticosteroidism (Cushing's syndrome), abnormal fat deposition and weight gain can be the changes that lead to the recognition of the underlying pituitary or adrenal dysfunction.

Classically, in type 2 diabetes, unexplained weight loss is a presenting complaint when polyuria is mild or absent. Moreover, with common forms of childhood gastrointestinal disorders, such as celiac sprue or Crohn's disease, arrested linear growth is often the first clue that something is clinically awry. It provokes the diagnostic inquiry that leads to the recognition of the bowel pathology. In milder presentations of cystic fibrosis, a similar growth failure occurring in infancy can indicate an underlying pathological disorder.

In fact, the entire roster of conditions listed in **Tables 1** and **2**, as well as others of a similar nature, are subject to being diagnosed as the result of a secondary change in nutritional status. The practical message is that the nutritional specialist, physician or nonphysician, may be the first person to whom the secondarily malnourished patient is referred, and the acumen of recognizing a secondary causation will guide the case to an appropriate clinical diagnostic program to uncover (and hopefully address and remedy) the underlying medical or surgical problems. Overarching guideline principles for uncovering secondary malnutrition states are provided in **Table 3**.

Diagnosis of Secondary Malnutrition

In general terms, a common set of principles applies for assessment of nutrient status whether the bases are primary, secondary, or a combination of both. These principles include: body composition measures, hematological and biochemical (biomarker) values, functional variables, and clinical signs and symptoms. It is more productive to focus here on the nuances, caveats, and distinctions for the detection of altered nutrition due to background conditions beyond spontaneous food intake.

Caveats for the Diagnosis of Secondary Excess Nutriture

The conditions that cause increased retention of energy and hypometabolism are listed in **Table 2**. When it comes to overweight and obesity, the absence of clear-cut overeating (which can be difficult to detect) combined with other characteristic signs of the different

Table 3 Three diagnostic principles related to secondary malnutrition

Assessment of dietary and nutrient intake: A quantitative and qualitative evaluation of usual dietary intake by a nutritionally trained practitioner or clinical dieticians serves to exclude the possibility that the situation is not primary (low or excessive intake) in nature and suggests a secondary basis for the nutritional problem. Caveat: In certain situations, a combination of reduced intakes and nutritional stress due to poor absorption, retention, or utilization may coexist.
Assessment of nutritional status: This includes measures of anthropometry and body composition, hematological status, biomarkers, and functional indicators, as well as clinical (physical) evaluation.
Diagnosis of underlying cause(s) of secondary nutritional imbalance: It is important, where possible, to identify the underlying entities that are causing the nutritional problem, such as absorptive or hormonal problems, to enable (where possible) a direct remedial approach to the cause of malnutrition and to orient management based on any pathophysiological knowledge about the underlying disease.

entities should raise suspicion. Excesses of vitamins and minerals may not easily be detected because the homeostatic control of circulating concentrations confounds biochemical diagnosis. Excessive urinary excretion rates of the nutrients or their metabolites often provide better indications than blood levels when micronutrient overload is the issue.

Caveats for the Diagnosis of Secondary Undernutrition

Undernutrition due to disease and dysfunction obviously requires establishment of the following: (1) the existence of deficiencies; and (2) the factors other than underconsumption that are influencing the deficiency states. The cut-point for diagnosing macronutrient undernutrition is a body mass index (BMI) of $<18.5\ kgm^{-2}$. However, with the worldwide pandemic of overweight, recent weight loss of 10% or more of usual body weight may be a more sensitive and reliable indicator of an incipient undernutrition problem. Weight problems diagnosed in this manner would certainly be detectable well before the BMI will have fallen to the aforementioned criterion.

Ill patients with adequate or excessive body mass indices can manifest metabolic substrate metabolism reminiscent of the severe malnutrition syndromes of adult kwashiorkor or marasmus (inanition). Moreover, fluctuations in weight under acute or semiacute situations often reflect changes in fluid balance. This is also the situation in patients with end-stage renal failure undergoing chronic dialysis. Methods such as bioelectrical impedance, dual X-ray absorptiometry, or isotope dilution in association with indirect calorimetry can assess true lean- and fat-mass status of patients with apparently normal body mass.

Hematological evaluation is important in nutritional assessment. A low hemoglobin, hematocrit, or red cell count signifies anemia, but in individuals with associated diseases, anemia can have a series of origins (hemolytic, hypoproliferative) that are nonnutritional and will not respond to nutritional therapy.

Biochemical evaluation for nutrient deficiency status in patients with associated disease is fraught with caveats and limitations. Two classes of nutritional deficiency are sometimes defined: in type 1 deficiencies, nutritional desaturation of tissue stores occurs, circulating levels of nutrients reflect the total body nutrient status and specific nutrient deficiency syndromes manifest (e.g., iron deficiency and anemia); in type 2 deficiency, there is homeostatic conservation of tissue, circulating concentrations of nutrients, such that blood concentrations remain virtually unaltered in the face of depletion. Typically, type 2 deficiencies manifest with growth failure or general signs of undernutrition. Deficiencies of zinc and magnesium, among others, fall into this second category. Inflammation and infection are stimuli that directly alter the circulating concentrations of nutrient indicators. Ferritin and circulating copper are elevated, whereas zinc, iron, and vitamin A concentrations are depressed with activation of the acute-phase response to injury or inflammation. In liver disease, depressed production of binding proteins can alter the usual indicators of nutritional status as a consequence of hepatic pathophysiology itself, rather than preexisting secondary malnutrition. Finally, it almost goes without saying that attempting biochemical nutrient evaluations from blood samples taken during concurrent infusion of micronutrient solutions in parenteral nutrition regimens – especially without a period of distribution and equilibration – will not reflect the tissue stores and total body reserves of the respective nutrients of interest.

Functional indicators of nutritional status have been applied to the assessment of secondary malnutrition and have been plagued by pitfalls. This applies to tests of nitrogen status, immune function, and hepatic protein secretion. Tests such as creatinine excretion, white blood cell counts, and cutaneous delayed hypersensitivity anergy, as well as decreased serum albumin, transferrin, transthyretin (prealbumin), and retinol-binding protein concentrations are sensitive to alteration by stress and injury. Failure to recognize distortion from stress underlies an early fallacy in surgical nutrition, in which low values for albumin, lymphocyte counts, and prealbumin, together with anergy, predicted poor postoperative outcomes. This misconception justified aggressive preoperative parenteral nutrition and albumin infusions, with little impact on predicted outcomes. In these situations, it was the stress and injury of the underlying disease, rather than nutritional status that was producing the abnormal values of the biomarkers. Recently, insulin-like growth factor has been advanced as a sensitive indicator of protein status in older patients, but whether it is confounded by nonnutritional features of disease remains to be clarified.

Management of Secondary Malnutrition

Secondary malnutrition has many faces and facets. It may have to be addressed both in a public health sense for communicable diseases such as parasitoses or HIV/AIDS, and in a medical care context for disorders that are particular and clinical in nature, such as hereditary or degenerative diseases.

Principles of Management

The first principle is to identify the underlying functional, physiological, or pathological cause of the malnourished state. If the condition is curable, then the management issues are simplified. For instance, if a person is dehydrated because of hyperglycemic diuresis in uncontrolled diabetes mellitus, the short-term management involves administration of exogenous intravenous fluids to restore normal hydration; however, restoring adequate diabetic control to the patient would be the long-term and definitive solution. The undernutrition and growth failure due to undetected celiac disease is easily eliminated by institution of a gluten-free diet. With inadequate nutrition in cystic fibrosis, adequate management of pulmonary problems and digestive-enzyme deficiencies

should allow patients to recover and maintain normal nutrition on a balanced oral diet. Thus, medical or surgical address of the underlying disorder, where possible, is the primary tool for management of secondary undernutrition.

Public Health Approaches

The management of the secondary iron deficiency attributable to hookworm or schistosomiasis can be achieved both by anthelminthic medications or supplemental iron to compensate for parasite-induced losses. In countries where HIV/AIDS is rampant, efforts for its prevention are fundamental. A food-security crisis grips the whole society in AIDS endemic areas, and this must be relieved with food and economic assistance. The wasting syndromes produced by tuberculosis are best addressed proactively by prevention of transmission and early detection. However, when primary prevention fails, as in the aforementioned infections, efforts to enhance the enteral intake of infected members of the community are particularly essential for their comfort and well being.

Dietary Management of Secondary Overnutrition

The dietary management of secondary overnutrition would logically include restricted intake of the nutrients accruing in excess. This is not always facile or feasible, however, due to the intrinsic complexity of foods and beverages, where most are sources of multiple essential micronutrients. Marked reduction in total energy intake can also jeopardize the intake of proteins and essential fats. For the metal-storage afflictions such as Wilson's disease and hemochromatosis, removing copper and iron from the diet, respectively, are the fundamental elements of management. Some additional benefits can be gained by blocking the metals' absorption, as with high doses of zinc in Wilson's disease or with strong black tea (tannins) in hemochromatosis. Fundamentally, however, the management of metal-storage diseases requires some interventions to selectively remove the overload – by chelating agents in Wilson's disease and recurrent phlebotomy in hemochromatosis. In a related variant condition, African hemosiderosis, common among Bantu in southern Africa, removing concentrated iron sources from the diet, specifically the iron-loaded traditional beers, provides effective long-term control.

Dietary and Nutritional Management of Secondary Undernutrition

The syllogism for dietary and nutritional management is to get enough nutrients into the body to restore nutritional adequacy and balance, taking any chronic barriers to uptake and retention into consideration. The blend of nutrients must be tailored to the specific absorptive or utilization problems, for example, compensatory fat-soluble vitamins in water-miscible forms with severe fat malabsorption, and extra doses of highly available iron with chronic blood loss. These can be delivered within a dietary context with supplements and fortified vehicles, in nonacute conditions. Even nondietary routes have been devised as in the treatment of vitamin D deficiency due to Crohn's disease with tanning bed ultraviolet B radiation.

When accumulated undernutrition is dangerously advanced, absorptive barriers are especially severe or nutrient losses are excessive, more concerted nutritional intervention is required. Intensive therapy can be delivered by three routes: orally, with special diets supplemented by liquid formulas; enterally, with liquid formulas perfused by intragastric or intraintestinal feeding tubes; and parenterally, with intravenous formulas infused into peripheral or central veins. Up to 50% of patients on dialysis have protein-energy malnutrition, which may continue undetected. For end-stage renal patients, intradialytic alimentation (adding nutrients to the dialysis fluids) has been used to reduce nutrient loss. Each approach has its distinct costs, special potential, and limitations and risks, and has been explored and refined in the context of age, physiological status, and specific disease states or surgical indications.

Tailoring of nutrient delivery is required with both enteral and parenteral nutrition, depending on the pathophysiology of the underlying conditions. Both hypo- and hypermetabolic states can occur; indirect calorimetry with metabolic carts is in vogue for prescribing energy delivery in intensive care. When pulmonary compromise is present, the balance among macronutrients is important to minimize carbon dioxide formation in metabolism.

Maintaining abundant amino acid supply promotes protein-sparing and prevents loss of lean tissue in catabolic states. Enrichment of enteral or parenteral regimens with branched-chain amino acids or keto-analog amino acids has been devised to compensate for the metabolic defects of nitrogen handling in hepatic or renal failure states. The objective of nutritional support in patients with liver failure is to provide adequate macronutrients to ensure the specific substrates for energy and protein synthesis and integrity of normal hepatic tissue function, without inducing or accentuating encephalopathy or otherwise aggravating hepatic insufficiency.

In juvenile cholestasis, large amounts of fat-soluble vitamin supplements and medium-chain triglycerides are usually required for optimum growth. With protracted secretory diarrheal diatheses, fluid and electrolyte balance may be the primary concern, followed by macro- and micronutrient nutriture, invoking the institution of parenteral feeding. Cancer cachexia is a major secondary consequence of disseminated neoplasms. It is tempting to prescribe aggressive nutritional support, but a caveat is that certain nutrients acting with certain neoplasms favor the tumor's growth and dissemination. To the extent that various forms of cachexia are partly driven by catabolic responses mediated by proinflammatory cytokines, antagonists directed at counteracting their action hold promise for retarding the nutrient-wasting in various forms of cachexia.

With intensive nutrition, there are risks and adverse consequences intertwined with the benefits. A variation of the refeeding syndrome, that is, hyperalimentation complications from excessive energy substrate perfusion or infusion, can produce hypophosphatemic and hypokalemic episodes. Improper formulation of fluids or liquids with micronutrients can cause deficiency or toxicity states in chronic nutritional support. The hazards of indwelling catheters are multiple, from phlebitis of the veins to sudden dislocation or migration. Fluid overload and sepsis are the most troubling complications of intravenous parenteral nutrition.

For tube-feeding enteral alimentation, tube placement is the crucial element. With nasal placement of the tube, there is a finite risk of respiratory tract inflammation and infection from aspiration of formula and secretions. In hospital, enteral nutrition is a risk factor for nosocomial pneumonia. An alternative site for long-term administration of tube-feeds is percutaneous placement of an intragastric feeding tube under endoscopic control.

Aggressive nutritional support, with its attendant expense and potential morbidity, in critically ill patients remains controversial. In terms of cost–benefit analysis, the use of the intensive formats of enteral artificial nutrition seems to be cost effective to reduce post-hip-fracture hospital stay in underweight women and for preoperative nutritional support, if carried out at home. Preoperative parenteral nutrition has been judged as prohibitively expensive for the small reduction in postoperative morbidity that it produces.

Conclusions

Dietary intake is the most important determinant of over- or undernutrition, but it is not the only influence on an individual's nutritional status. A series of extrinsic environmental factors or intrinsic clinical or physiological disorders can alter the absorption, retention, utilization, and integrity of nutrients. These give origin to secondary malnutrition states. Primary (dietary origin) and secondary (environmental, pathological) factors often combine within the same individuals. From a public health perspective, the goal is to implement broad policies and programs that increase the availability of specific nutrients imperiled by the local environmental problems, for example, iron in hookworm infested areas, while addressing the primary diseases. In the clinical setting, management requires diagnosing and managing the underlying pathological states interfering with nutritional health while providing compensatory measures to correct secondary nutritional imbalances.

Further Reading

Brooks, M.J., Melnik, G., 1995. The refeeding syndrome: An approach to understanding its complications and preventing its occurrence. Pharmacotherapy 15, 713–726.
Calder, P.C., Jackson, A.A., 2000. Undernutrition, infection and immune function. Nutrition Research Reviews 13, 3–29.
Dudrick, S.J., Maharaj, A.R., McKelvey, A.A., 1999. Artificial nutritional support in patients with gastrointestinal fistulas. World Journal of Surgery 23, 570–576.
Herbert, V., 1973. The five possible causes of all nutrient deficiencies: Illustrated by deficiencies of vitamin B_{12} and folic acid. American Journal of Clinical Nutrition 26, 77–88.
Higgins, P.A., Daly, B.J., Lipson, A.R., Guo, S.E., 2006. Assessing nutritional status in chronically critically ill adult patients. American Journal of Critical Care 15, 166–176.
McKenzie, C., Vicca, N., Ward, J.E., Coles, S.J., 2000. Prevalence of malnutrition on admission to four hospitals in England. The Malnutrition Prevalence Group. Clinical Nutrition 19, 191–195.
Ofman, J., Koretz, R.L., 1997. Clinical economics review: Nutritional support. Alimentation and Pharmacological Therapy 11, 453–471.
Paccagnella, A., Calo, M.A., Caenaro, G., et al., 1994. Cardiac cachexia: Preoperative and postoperative nutrition management. Journal of Parenteral and Enteral Nutrition 18, 409–416.
Phang, P.T., Aeberhardt, L.E., 1996. Effect of nutritional support on routine nutrition assessment parameters and body composition in intensive care unit patients. Canadian Journal of Surgery 39, 212–219.
Schaibe, U.E., Kaufmann, S.H.E., 2007. Malnutrition and infection: Complex mechanisms and global impacts. PLoS Medicine 4. https://doi.org/10.1371/journal.pmed.0040115.
Solomons, N.W., 1993. Pathways to impairment of nutritional status by gastrointestinal pathogens, with emphasis on protozoal and helminthic parasites. Parasitology 107 (supplement), S19–S35.
Solomons, N.W., Keusch, G.T., 1999. Clinical issues: Childhood illnesses, vaccinations and nutritional status. In: Gershwin, M.E., German, J.B., Keen, C.L. (Eds.), Nutrition and Immunology: Principles and Practice. Humana Press, Totowa, NJ, pp. 469–474.
Thapa, B.R., 1994. Intractable diarrhoea of infancy and its management: Modified cost effective treatment. Journal of Tropical Pediatrics 40, 157–161.
Von Roenn, J.H., Roth, E.L., Craig, R., 1992. HIV-related cachexia: Potential mechanisms and treatment. Oncology 49 (supplement 2), 50–54.
Zipf, W.B., 2004. Prader–Willi syndrome: The care and treatment of infants, children, and adults. Advances in Pediatrics 51, 409–434.

Nutrition and asthma

Charles B. Stephensen[a,b] **and Nicholas J. Kenyon**[c,d], [a] USDA Western Human Nutrition Research Center, Davis, CA, United States; [b] Department of Nutrition, University of California, Davis, CA, United States; [c] Division of Pulmonary, Critical Care, Sleep Medicine, Department of Medicine, School of Medicine, University of California, Davis, Davis, CA, United States; and [d] Veteran's Administration Northern California Health Care System, Mather, CA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of G.U. Schuster, N.J. Kenyon, C.B. Stephensen, Asthma, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 122–128, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00021-0.

Introduction 513
Overview of asthma 514
Epidemiology 514
Definition of asthma 514
Development of asthma 514
Management of chronic asthma 515
The role of nutrition in prevention and management of asthma 515
Breastfeeding 515
Probiotic supplements 516
Obesity and asthma 516
Vitamin D 517
Long-chain polyunsaturated fatty acids (LCPUFA) 517
Amino acid supplementation and asthma 518
Antioxidants 518
Fruits and vegetables 519
Magnesium 519
Conclusion 519
References 519

Key points

- Epidemiology and Development of Asthma
- Management of Asthma
- Maternal and infant nutrition and asthma risk
- Breastfeeding and infant intestinal microbiota and asthma risk
- Probiotics and asthma
- Obesity, weight loss and asthma
- Vitamin D and asthma
- Dietary n-3 PUFA and asthma
- Amino acid, vitamin and antioxidant supplements as asthma treatments
- Fruit and vegetable intake and asthma
- Oral magnesium therapy and asthma

Introduction

Asthma is a chronic inflammatory disease that develops in early childhood, though onset later in life is not atypical. Modern nutrition research has focused on the health benefits of specific nutrients, as well as dietary patterns, in preventing and treating such chronic conditions. The increased prevalence of asthma in recent decades has paralleled changes in environmental exposures and dietary patterns that likely are causing the increased prevalence of asthma, though specific cause-effect associations are hard to discern. In this article we will first examine the epidemiology, etiology, and management of asthma. We will then describe the dietary factors that are associated with a higher or lower risk of asthma or of asthma severity. Our approach in examining nutrition and asthma associations has been to identify specific nutrients and dietary practices that have been shown to be consistently associated with increasing or decreasing risk or severity of asthma as shown thorough repeated publication of human studies in the scientific literature, often reported in systematic reviews and meta-analyses. Some nutrition interventions can be readily examined in randomized, controlled trials (RCTs), but observational data is also important particularly for examination of factors such as

breastfeeding and dietary patterns. In summary, data published in the last decade has shown a substantial increase in the body of knowledge supporting diet (e.g., related to the increased prevalence of obesity) as a risk factor for developing asthma, though other practices, including breast feeding, fruit and vegetable consumption and intake of specific nutrients including vitamin D and n-3 LCPUFA, are protective and may be useful in decreasing the severity of symptoms in patients with asthma.

Overview of asthma

Epidemiology

Asthma affects an estimated 26 million people in the US, 70% of them adults. Worldwide approximately 250 million people are believed to have a diagnosis of asthma. In the US, asthma accounts for 13.9 million office visits for adults, 500,000 hospitalizations and two million emergency department (ED) visits per year. Of the estimated 25 million asthmatics in the US, most of their annual costs incurred are from the five to ten percent of patients with severe asthma. Approximately 3000 deaths occur annually, and most of these deaths occur outside the hospital. A recent study of the effect of age on asthma mortality showed that age >55 y conferred a 5-fold increased risk of death from asthma compared to younger adults and children (Tsai et al., 2012). Predictors of fatal asthma include three or more ED visits for asthma in the past year, an asthma hospitalization or ED visit in the past month, overuse of short-acting beta 2 agonists, a history of intubation or ICU stay for asthma, and patient characteristics, such as low socioeconomic status, female sex, and non-white ethnicity. In 2007, the incremental cost due to asthma was $56 billion in the United States, and the costs are much higher worldwide (Barnett and Nurmagambetov, 2011). It is estimated that up to 25% of all "adult-onset" asthmatics have a workplace trigger for their disease. 5–20% of all asthmatics suffer partial disability that affects their ability to work, and 40%–80% lose considerable income because of their disease.

Definition of asthma

In the 2021 Global Initiative for Asthma (GINA) guidelines, the definition of asthma highlights the facts that asthma is a "heterogeneous disease characterized by chronic airway inflammation" involving a history of episodic respiratory symptoms (shortness of breath, cough, wheezing) with "variable expiratory flow limitation" (GINA, 2021). Lung function testing such as spirometry is needed to help determine both the expiratory airflow abnormalities and airway hyperreactivity. Characterization of asthma also includes measuring fraction of nitric oxide in exhaled breath (FeNO) and peripheral blood eosinophil counts. While FeNO is not a standardized test for the diagnosis of asthma, recent evidence-based summaries suggest that a high FeNO level (e.g., >44 ppb) in breath has a very good positive predictive value for the diagnosis of asthma (Wang, 2017). Furthermore, skin prick testing and *in vitro* IgE specific assays can determine if a person with asthma has an antibody response to antigens in the environment. Sputum eosinophilia is another asthma marker, but is rarely used clinically, and is not part of the standard definition or work-up.

Development of asthma

Asthma is a complex syndrome and not a singular disease. Multiple sub-types of asthma exist in adulthood, though this is less true in early childhood. Asthma most commonly develops at a young age, and it remains difficult to diagnose before age three. Upper respiratory viral illnesses, particularly rhinovirus and respiratory syncytial virus, and environmental allergens combine to trigger an upper and lower airway inflammatory response in susceptible individuals. The contribution of air pollution and environmental events, including wildfires, to the development of asthma and to asthma attacks in children is becoming well-recognized. Several theories describe asthma pathogenesis broadly. One is the "atopic march" that states that allergic inflammation of the airways naturally follows eczema or atopic dermatitis in young children. Secondly, the "hygiene hypothesis" suggests that allergic inflammation results from changes in the patterns of exposure—less infections, more contact with indoor allergens like house dust mite—that many children now experience at a very young age in recent decades. Dietary changes characterized as adoption of the "Western diet" have also been associated with the increased prevalence of asthma seen over the past 50 years. Current research focuses on the hypothesis that such environmental and dietary changes may increase the risk of asthma development in early childhood by multiple mechanisms, including changes to the intestinal microbiome, which can then alter the development of immune function, increasing the risk of developing a type 2 pattern of inflammation which can lead to allergic sensitization to some environmental antigens (i.e., "allergens"). These factors may act on the infant *in utero* via maternal exposures (e.g., obesity related to Western diet, smoking, air pollution) or after birth (e.g., lack of breastfeeding, Western diet, decreased microbial exposure in urbanized settings). Delivery by caesarian section increases the risk of asthma development in early childhood via effects on immune development perhaps resulting from a decreased perinatal stress response related to delivery, altered development of the infant microbiome, or perhaps because of prophylactic use of antibiotics in such cases. Finally, genetics also plays a role, with infants of parents with atopic disease having a higher risk of developing asthma (Renz and Skevaki, 2021).

Inflammation of the airways is a defining feature of asthma and selective recruitment by leukotrienes, cytokines, and chemokines help direct trafficking of key effector cells to the airways. The key effector cells in asthma include eosinophils, basophils, T-helper Type 2 cells (Th2), and mast cells. The Th2 cytokines, interleukin (IL)-4, -5, and -13 or their receptors are all targets of novel therapies to dampen type 2 inflammation. Mast cells in the bronchial smooth muscle are important in the persistence of abnormal

airway hyper-responsiveness and chronic airway inflammation in difficult-to-control asthma. In addition to this pathway, lung group-2 innate lymphoid cells (ILC-2) in the lung drive an eosinophilic airway inflammatory response. Alarmins such as IL-33 and thymic stromal lymphopoietin (TSLP) are secreted by airway epithelial cells in response to cigarette smoke, pollutants and microbes and trigger activation of the ILC2 cells, which in turn increase IL-5 levels and eosinophilic recruitment.

Asthma has long been considered an eosinophil-predominant inflammatory airway disease; there is an established correlation between peripheral blood eosinophilia and worse outcomes from asthma and increased eosinophil numbers in induced sputum samples are found in most severe asthmatics (Louis et al., 2000). Eosinophils are bone marrow derived granulocytes that have long been recognized as the main mediator cells in allergic asthma. In addition to releasing varying cytokines and chemokines, they have been found to also release toxic granular proteins all of which promoting Th2 inflammation as well as airway epithelial damage (Dunn and Wechsler, 2015).

These findings support previous evidence that link airway inflammation and abnormal airway physiology indicating that reducing airway inflammation with corticosteroids improves airway function. While the classic eosinophilic pathogenesis of asthma does not adequately explain the subgroups of asthma, this pathway has driven novel drug development over the past decade.

Management of chronic asthma

The 2021 GINA guidelines (GINA, 2021) have simplified prior asthma guidelines that detailed often complex, progressive management actions within different steps of care. The medications have been repositioned to deemphasize "controller" inhaled corticosteroids (ICS) and "rescue" short acting beta2 agonist bronchodilators. Instead, the preferred regimen is now a combination of ICS and the long-acting b2-agonist (LABA) formoterol to be used as needed in mild asthma cases or scheduled and as needed for moderate asthma. Patient adherence with prescribed asthma therapy in the US is very poor and it is hoped that this simplified prescription will improve this. The important aspect of patient education and open communication between physician and patient must be emphasized. Emphasis placed on the two aspects of the written asthma action plan—(1) daily management and (2) how to recognize and handle worsening asthma. Discussions should also include physical, emotional, and environmental stressors that contribute to asthma symptoms, including diet and nutrition.

The past 7 years have brought a transformation in asthma care for the most severe asthmatics. There is now a panel of home-based injection monoclonal antibody therapeutics that can be deployed based on severe asthma patient's eosinophil count (>150/μL) and associated comorbidities. Omalizumab was the first monoclonal antibody approved in 2003. It binds to IgE and has been approved for patients with refractory allergic asthma that has been shown to decrease exacerbations, inhaled corticosteroids, and improved asthma-related quality of life measures in refractory asthmatics (Busse et al., 2011; Strunk and Bloomberg, 2006). In 2015, the approval of IL-5–neutralizing monoclonal antibodies (mepolizumab, reslizumab) for severe asthma led to remarkable clinical benefit in severe, corticosteroid-requiring asthma associated with sputum eosinophilia. That has been followed by benralizumab, a novel monoclonal antibody (mAb) that binds to the α chain of the interleukin 5 (IL-5) receptor (IL-5Rα), a receptor expressed by mature eosinophils, eosinophil-lineage progenitor cells and basophils (Rothenberg and Hogan, 2006), and dupilumab, a dual IL4/IL-13 receptor blocker approved for asthma and chronic sinusitis with nasal polyposis. Future therapeutics targeting the alarmins are expected soon. Together, these advanced therapeutics greatly expand the armamentarium to more severe asthma patients and reduce the risk and harm that results from asthma exacerbations.

The role of nutrition in prevention and management of asthma

Breastfeeding

Breastfeeding has multiple immediate health benefits for infants, providing optimal nutritional support and decreased risk of infection in infancy, but also provides many long-term health benefits, including a possible decrease in the risk of asthma (Victora et al., 2016). In addition to being rich in nutrients, breastmilk provides many immunologic factors that protect the infant against infection and can shape the development of the immune system (Dawod et al., 2021). Breastfeeding is also an important determinant of intestinal microbiota composition during infancy for many reasons, including the presence of human milk oligosaccharides that increase the growth of beneficial microbiota including bifidobacteria (Ho et al., 2018). The composition of an infant's microbiota directly affects early immune development and is associated with risk of allergic disease, including asthma (Renz and Skevaki, 2021). For many years it has been assumed that breastfeeding protects against the development of asthma through a variety of mechanisms, including effects on immune development and decreased risk of early lower respiratory tract infections which can increase the risk of asthma. The principal evidence for this comes from epidemiologic studies associating breastfeeding "level" (i.e., exclusive vs. supplemented) and duration with risk of asthma and allergic disease in infancy and later in childhood. As pointed out in a recent review, the evidence is mixed from these studies with some showing decreased risk of asthma and others showing no benefit (Miliku and Azad, 2018). Possible reasons for this lack of consistency include the difficulty of defining the extent (full, predominant, or partial) and duration of breastfeeding. It may also be important to determine whether infants feed from the breast or from expressed, stored milk (which could affect immune factors in milk). The risk of asthma development during the first few years of life (typically by 3–6 years of age) is relatively low, which can also make it difficult to define associations in epidemiologic studies that are not quite large. Fortunately, it is useful to examine the risk of wheezing as an endpoint because wheezing in infancy is itself predictive of later asthma development (Miliku and Azad, 2018). Many studies have examined the association of

breastfeeding with a risk of wheezing as a proxy for risk of asthma. A recent meta-analysis examined the association of breastfeeding with wheezing during infancy in infants with an elevated risk of asthma due to family history. The study found that breastfed infants had 32% lower odds of wheezing by 1 year of age compared to infants never breastfed (Harvey et al., 2021). In summary, the current evidence indicates breastfeeding can reduce the risk of wheezing in high-risk infants and may also reduce the risk of asthma in all infants, but the evidence for this latter point is not conclusive.

Probiotic supplements

A limited number of studies have examined the efficiency of probiotic supplementation in the treatment and the prevention of asthma during infancy and childhood. This strategy is based on the benefit of some commensal microbiota in regulating infant immunity and possibly decreasing the risk of type 2 inflammation developing in infancy and contributing to the risk of asthma (Renz and Skevaki, 2021). Probiotic supplements used in infancy and childhood typically include *Lactobacillus* or *Bifidobacterium* species. A recent meta-analysis of nineteen randomized trials using such probiotics in infancy did not show an overall benefit in preventing the development of asthma or wheeze during infancy or early childhood (Wei et al., 2020). A subgroup analysis showed a decreased risk of wheeze in infants with atopic disease, suggesting a benefit in high-risk infants. Probiotics have also been evaluated in randomized trials among asthma patients during childhood to determine if severity of disease could be diminished. A meta-analysis of such studies found a shift away from type 2 cytokine production (IL-4) in favor of type 1 cytokine production (IFN-γ) and a decrease in asthma episodes but not in other measures of asthma severity or pulmonary function (Lin et al., 2018). On balance, the use of probiotics to decrease the development or severity of asthma is not strongly supported by these findings. The underlying mechanistic rationale for such studies has merit, however, and future research in this area is warranted, including examination of the role of probiotics in treating asthma symptoms in adults.

Obesity and asthma

Asthma prevalence has increased in the US in recent decades at a time when the prevalence of obesity has also increased. This association is not a coincidence as there is ample evidence that obesity increases the risk of asthma in both adults and children and is also associated with greater asthma severity (Ali and Ulrik, 2013). Multiple mechanisms appear to be involved including chronic inflammation, mitochondrial dysfunction, Th17-induced neutrophilia, macrophage dysregulation, hormonal changes, lipid metabolism, insulin resistance, and body mechanics (Baffi et al., 2016; Shore and Cho, 2016). Other well-described abnormalities in obesity and metabolic syndrome, accelerated formation of advanced glycation end products (AGEs) and alterations in arginine metabolism that may be modulated by the anti-inflammatory incretin, glucagon-like peptide-1 (GLP-1) (Milutinovic et al., 2012; Ojima et al., 2013; Singh et al., 2015). Obesity, which is defined in adults as a body mass index (BMI; kg body weight/[height in meters]^2) >30, has a negative effect on the mechanics of breathing, and these effects are more pronounced as BMI increases. Thus, individuals with or at risk of asthma may have more severe diseases at higher BMIs because of the mechanical constraints on breathing imposed by obesity.

In addition, obesity causes local and systemic inflammation by multiple mechanisms (Ali and Ulrik, 2013). Local inflammation in adipose tissue deposits can result from local tissue damage related to adipocyte size, oxygenation of the local cells and tissues, and cell death related to these stresses. Markers of oxidative damage are increased in obesity, partly because of this tissue damage. As a result, inflammatory cells are attracted to and accumulate in adipose tissue, particularly abdominal adipose tissue. These cells include macrophages and T cells, particularly Th1 cells which drive type 1 inflammation by production of cytokines such as IFN-γ. These local effects can eventually produce systemic inflammation as indicated by elevated plasma cytokines, such as IL-6, and acute phase proteins, including C-reactive protein (CRP). While this phenomenon is well-characterized in obesity, the effect of such inflammation on asthma development is not clear. In addition, hormonal changes that occur in obesity can have direct effects on local and systemic inflammation. In particular, the hormone leptin, which is increased in obesity, enhances development of Th1 cells and thus promotes type 1 inflammation. Additionally, the hormone adiponectin has anti-inflammatory effects but is decreased in obesity. However, this bias toward type 1 inflammation, which could dampen type 2 inflammation that is common in allergic asthma, does not decrease the risk of asthma development, as far as is known, but obese individuals with asthma do have a phenotypically different type of inflammation with less pronounced eosinophilic inflammation than lean individuals (Ali and Ulrik, 2013). In children with asthma, a predominance of systemic type 1 cytokines is seen rather than the type 2 predominance seen in lean asthmatics (Nyambuya et al., 2020).

Information on the association of obesity with asthma in children has increased substantially in recent years. In case-control studies performed in children and adolescents, BMI is consistently associated with an increased risk of having asthma (Azizpour et al., 2018). In addition to the factors mentioned above, it is thought that common genetic factors may be associated with both obesity and asthma, and early life events may lead to epigenetic effects on immune function that could predispose to an increased risk of either or both conditions (Azizpour et al., 2018; Ali and Ulrik, 2013). Fortunately, weight loss in children can have some beneficial effects on asthma-related inflammation and quality-of-life indicators though improvements in pulmonary function have not been consistently shown in children, though improved pulmonary function has been shown in weight loss for adults (Okoniewski et al., 2019). Bariatric surgery in adults often results in greater weight loss than other interventions and has also been shown to improve asthma severity based on clinical indicators including medication usage and the occurrence of severe asthma exacerbations requiring emergency medical interventions (Hossain et al., 2021).

Vitamin D

Vitamin D affects immune function and has been hypothesized to have a benefit in treating asthma due to its ability to dampen inflammatory T-cell immunity (e.g., Th17 cell activity), enhance regulatory immunity (e.g., Treg cell activity), and enhance antiviral innate immunity, thus decreasing the risk or severity of viral infection-induced exacerbations in asthma patients (Shabana et al., 2019). A meta-analysis using individual participant data examined the effect of vitamin D supplementation of asthmatic children and adults on the occurrence of asthma exacerbations from seven published studies (Jolliffe et al., 2017). Vitamin D dosing regimens were varied, ranging from infrequent (every 2 m) high doses (e.g., 120,000 IU) to daily RDA-level doses (e.g., 500 2000 IU), to a combination of the two. The main outcome examined was the relative risk of severe exacerbations (needing either systemic corticosteroid treatment or ER attendance). The analysis found that severe exacerbations were reduced in the vitamin D treatment group relative to the control group using either a need for systemic corticosteroid treatment (showing a 31% reduction in incidence) or ER attendance (odds ratio, 0.46) as the case definition. Patients with poor vitamin D status (<10 ng/mL) tended to benefit more than those with better status. A subsequent RCT published in 2021 in 192 children with asthma with low plasma vitamin D (Lu et al., 2021) showed no benefit using time to a severe asthma exacerbation as the primary endpoint. In summary, vitamin D supplementation of asthma patients can be beneficial in decreasing severity of asthma exacerbations, presumably by regulating immune activity.

Asthma is often diagnosed in early childhood and its development can be influenced by prenatal exposures, perhaps including maternal vitamin D status. Three RCTs of prenatal vitamin D supplements (800 IU daily, or 400 IU daily with a higher loading dose of 2400 or 4000 IU), when analyzed individually, did not show any reduced risk of asthma or recurrent wheezing when analyzed individually (Brustad et al., 2019; Goldring et al., 2013; Litonjua et al., 2020), but a combined analysis of two of these (Wolsk et al., 2017) and a meta-analysis of all three (Li et al., 2019) did show a beneficial effect of prenatal supplementation on reducing risk in the vitamin D treatment group by approximately 25%. Wheezing itself may result from causes other than asthma (e.g., a respiratory tract infection) so some caution is warranted in concluding that wheezing is always specific to asthma. Observational data also show, as analyzed in the same meta-analysis (Li et al., 2019), a reduced risk of recurrent wheezing in association with higher maternal vitamin D intake. Taken together, these studies are consistent with higher maternal vitamin D intake reducing the risk of asthma in their offspring during early childhood. However, some caution is warranted as the meta-analysis (Li et al., 2019) showed a U-shaped relationship between maternal intake during pregnancy and wheezing/asthma risk in children, with risk decreasing to a nadir at approximately 800 IU/d, but with increased risk seen at higher levels.

Long-chain polyunsaturated fatty acids (LCPUFA)

The Western dietary pattern is typically rich in the n-6 PUFA linoleic acid (18:2n-6) from a variety of sources including corn oil. After consumption, linoleic acid can be efficiently elongated to the LCPUFA (a 20 or 22 carbon PUFA) arachidonic acid (AA; 20:4, n-6), which has many biological activities. AA can also be ingested directly from meat and eggs. AA is abundant in cell membranes and can affect immune cell responses as it is a precursor for the production of immune mediators including prostaglandins (PG) such as PGE2 produced by the cyclooxygenase (COX) pathway, and leukotrienes (LT) such as LTB4 produced by the 5-lipoxygenase (5-LOX) pathway. PGE2 promotes type 2 inflammation via effects on both T and B lymphocytes such as promotion of IgE production which is important in allergic responses. PGE2 also induces fever, vascular permeability to facilitate inflammation, and can induce the pain response. Not surprisingly, suppression of the COX pathway is one mechanism employed in pharmacology to produce common, over-the-counter non-steroidal anti-inflammatory drugs (NSAIDS). LTB4 causes bronchoconstriction, increases vascular permeability, and attracts neutrophils and eosinophils to sites of inflammation. LTB4 can also enhance inflammation by stimulation production of pro-inflammatory cytokines by innate immune cells, and by promoting T cell function. Dietary n-3 PUFA, such as α-linolenic acid (18:3, n-3), are found in flaxseed and canola oils and are less abundant in the Western dietary pattern than linoleic acid and can also act as precursors for LCPUFA production including eicosapentaenoic acid (EPA; 20:5, n-3) and docosahexaenoic acid (DHA; 22:6, n-3), though this elongation is very inefficient. However, EPA and DHA are abundant in some foods, primarily cold-water fish such as salmon, and are also commonly available now in supplements. Intake of n-3-containing plant oils, fish containing EPA and DHA, and supplements containing fish oil or purified EPA and DHA, is increasing because of the well-characterized anti-inflammatory properties of EPA and DHA. As with n-6 PUFA, these n-3 PUFA are incorporated in cell membranes, though EPA and DHA are always found at much lower levels than AA and are also available as precursors for production of immune mediators. The n-3-derived mediators tend to have anti-inflammatory properties while the n-6 mediators can promote inflammation. The n-3 LCPUFA and their mediators act via several mechanisms. First, EPA is a competitor of AA in the COX and 5-LOX pathways and thus can reduce the production of PGE2 and LTB4 to a limited degree, but the PGE3 and LTB5 mediators produced from EPA are also less inflammatory than PGE2 and LTB4 produced from AA. In addition, EPA and DHA serve as precursors for pro-resolving mediators termed resolvins that dampen and resolve inflammation by a variety of mechanisms. DHA, in addition, serves as a precursor for two additional classes of pro-resolving mediators, maresins and protectins. EPA and DHA can thus act via these mechanisms, as well as via other mechanisms not described here, to decrease inflammation. As a result, EPA and DHA are being evaluated as interventions (via diet or supplements) to reduce inflammation for many chronic inflammatory conditions, including asthma (Miles et al., 2021; Panigrahy et al., 2021; Djuricic and Calder, 2021).

As with vitamin D, prenatal exposure to n-3 LCPUFA has been examined as a protective factor for development of asthma in early childhood. Several cohort studies have examined usual dietary intake of fish containing n-3 LCPUFA by pregnant women

and by infants and children in relation to risk of developing asthma as reported in a meta-analysis (Yang et al., 2013). Fish consumption during pregnancy was not associated with lower risk of asthma in offspring but intake of fish during infancy and early childhood was associated with a 25% lower risk of developing asthma during follow-up periods ranging from 1 to 6 years. Similarly, infants ingesting higher levels of LCPUFA from breastmilk had a 29% lower risk of asthma over a period of 5 years. In summary, observational data suggest that exposure to n-3 LCPUFA may prevent development of asthma in children. These studies have the shortcoming of all observational studies in that other factors related to n-3 LCPUFA intake may be the cause of this association. To address such concerns, several placebo-controlled, randomized trials have examined the effect of prenatal n-3 LCPUFA supplements on risk of developing wheeze or asthma in offspring. Seven trials involving EPA plus DHA supplementation in the third trimester of pregnancy with children being followed for periods ranging from 6 m to 16 y (Lin et al., 2020). Since asthma and wheeze are difficult to distinguish in infants these endpoints were combined. The pooled analysis found that the relative risk of asthma/wheeze decreased by 19% in the LCPUFA group relative to the "placebo" oil (olive, soy, or a vegetable oil) group. The risk of asthma alone was not reduced when examined separately. Subgroup analysis showed that LCPUFA supplementation of high-risk children (those with a family history) had a 35% reduction in asthma/wheeze and in studies with higher doses of n-3 LCPUFA ($\geq$2 g/d) these supplements produced a 39% reduction in asthma/wheeze. The quality of the evidence from these trials was judged to be relatively low, which reduces confidence in the conclusions and suggests that more studies addressing study design issues should be undertaken. Also, long-term use of high-dose n-3 LCPUFA can decrease blood clotting and thus increases the risk of bleeding which is a risk that also requires further evaluation. In contrast to using high-dose supplements, a recent controlled intervention trial in asthmatic children that increased fish intake over that of the control group did show decreased pulmonary inflammation, though not improved pulmonary function (Papamichael et al., 2019). This partial benefit suggests that food interventions may be useful and would avoid risks of high-dose supplements.

Fewer studies have been done with n-3 LCPUFA in adults to prevent or treat asthma than have been done in childhood. Adult intake of fish and n-3 LCPUFA in observational studies has not associated with a lower of risk of developing asthma (Yang et al., 2013). Several therapeutic interventions have been done in adults with asthma but, as discussed in a recent review (Willemsen, 2016), and reconfirmed in a more recently published intervention trial (Lang et al., 2019), there is not consistent evidence from these studies that n-3 LCPUFA is an effective intervention for diminishing severity of asthma, though some indicators of inflammation were diminished in these studies.

Amino acid supplementation and asthma

Dietary supplementation of various amino acid formulations is common practice, including in patients with chronic illnesses. Several recent studies have added significantly to our understanding of the impact of supplemental amino acids on asthma. Creatine may be the most popular amino acid supplement in the general population, but data suggests that creatinine may be the most disadvantageous amino acid in asthma. In an ovalbumin exposure mouse model of asthma, it was shown that creatine at a dose 0.5 mg kg^{-1} day^{-1} worsened eosinophilic lung inflammation and airway hyper-responsiveness (Vieira et al., 2007). On the contrary, other amino acids appear to have some beneficial effects in asthma. For example, undenatured whey protein, a cysteine source, appears to increase antioxidant capacity in the lungs and improve airway hyperresponsiveness in asthmatics with exercise-induced symptoms. Also, L-arginine and L-citrulline could potentially act as bronchodilators and anti-inflammatory treatments. L-arginine is a semi-essential amino acid that can be enzymatically converted into nitric oxide by the nitric oxide synthase (NOS) enzymes. Dysregulation of arginine metabolism is a paradigm linked to multiple diseases, including obesity, metabolic syndrome, as well as asthma. This paradigm has led to the testable hypothesis that L-arginine supplementation is beneficial in cardiovascular disease (Bahadoran et al., 2016; Pahlavani et al., 2014) and sickle cell disease (Morris et al., 2013) among others. From these studies, we have learned that there are other factors to consider, such as asymmetric dimethylarginine (ADMA), which can accumulate and lead to impaired NO production (Bratt et al., 2011; Linderholm et al., 2014; Sibal et al., 2010). A recent clinical trial of L-citrulline supplementation in individuals with asthma found that L-citrulline increased fractional exhaled NO and modestly improved forced expiratory volume in 1 s (FEV1), especially in females with late-onset asthma (Holguin et al., 2019). We and others have found in murine models that manipulation of L-arginine content in the airway compartment via L-arginine treatment or inhibition of the arginase enzyme decreases airway inflammatory cell counts, lung lavage cytokine levels, airways hyperresponsiveness, and arginase activity (Mabalirajan et al., 2010). L-arginine, L-citrulline, and L-glutamine supplementation and metabolism will continue to be a focus of research in asthma, particularly given the low cost and widespread availability of these potential treatments.

Antioxidants

Oxidative stress during asthma can contribute to the severity of asthma symptoms and it is thus logical to consider how use of antioxidant nutrients might be useful in treating, or perhaps preventing, asthma (Sahiner et al., 2018). Vitamins C and E are the two principal vitamin antioxidants, though minerals like iron, zinc and selenium can also contribute to the activity of antioxidant proteins. Observational and small intervention studies have been conducted to examine the relationship between vitamins C and E and asthma. The results are quite heterogeneous and do not support use of these vitamins to treat asthma, though a possible benefit for vitamin C in exercise induced asthma has been reported (Wilkinson et al., 2014; Moreno-Macias and Romieu, 2014; Milan et al., 2013). The role of selenium in asthma has also been considered but supplement use has not been examined, though

selenium deficiency does occur in some regions of the world where supplements may be useful for improving status and treating underlying deficiency symptoms (Norton and Hoffmann, 2012).

Fruits and vegetables

Fruits and vegetables contain many antioxidants including vitamins C and E, carotenoids and other phytochemicals such as flavonoids, isoflavonoids and polyphenolic compounds that have antioxidant properties. For these reasons, many studies of varied design have examined the association of fruit and/or vegetable intake with asthma and a meta-analysis has examined 58 articles reporting such studies (Hosseini et al., 2017). Most studies were in children or adolescents with only 17 done in adults. Most studies were cross-sectional, 8 were case-control studies, 13 were cohort studies and only 7 were clinical trials. Studies were conducted in over 20 different countries. The authors of the meta-analysis focused on primary prevention studies reporting the risk of asthma or wheeze, and on secondary prevention studies reporting changes in the severity of asthma or wheeze. Overall, higher fruit intake was not associated with the odds of having asthma but was associated with 39% lower odds of having more severe asthma, but only two studies examined asthma severity. Higher fruit intake was also associated with 6% lower odds of having wheeze. Higher vegetable intake was associated with 6% lower odds of having asthma. Fruit and vegetable intake together were not associated with odds of having asthma. While combining studies with different design carried out in different ages and in different regions is challenging, these results support the role of a healthy dietary pattern including fruit and vegetable intake as being protective, to a limited degree, against development of asthma. The mechanism of protection is also uncertain when examining the intake of whole foods but does allow for consideration of dietary recommendations directly.

Magnesium

Magnesium deficiency has been associated with asthma but the mechanism underlying this association is not clearly understood. Magnesium may be anti-inflammatory and can inhibit the effect of calcium on causing smooth muscle contraction, for example in airways. Magnesium is given intravenously as an acute treatment for asthma exacerbations because it seems to have several biological effects attenuating asthma, including relaxation of smooth muscle cells, stabilization of mast cells, and counteracting bronchoconstriction. Oral magnesium supplementation has been examined in several RCTs with asthma patients and, as reported in a recent meta-analysis, a beneficial effect was seen on pulmonary function (increased FEV1 after 8 weeks of supplementation, but not at other timepoints) but other pulmonary function and clinical endpoints did not improve leading the authors to conclude that oral magnesium cannot be recommended at this time as an asthma treatment (Abuabat et al., 2019).

Conclusion

Several lines of evidence demonstrate the benefits of good nutrition, including maintaining a healthy body weight, in decreasing the risk and severity of asthma. This is true during pregnancy, infancy, and early childhood for the prevention of asthma in early life. Factors that have shown benefit include prevention of overweight and obesity, consuming higher levels of vitamin D and n-3 LCPUFA than are currently common in Western countries, providing breastfeeding in infancy for nutritional benefit and likely also for optimizing development of the intestinal microbiota. Studies in adults have been less frequent but there is a demonstrated benefit of weight loss for adults with obesity. Data suggest that a "healthy diet" including abundant fruits and vegetables is also of benefit in decreasing the risk and severity of asthma, though much of this data is observational. Use of RCTs is easier in single nutrient interventions and is providing evidence that vitamin D and n-3 LCPUFA intake are beneficial for preventing and perhaps decreasing severity of asthma. These findings are generally consistent with findings for other chronic inflammatory conditions, suggesting that maintaining a healthy body weight, optimizing intake of specific immune-modulating nutrients as well as of fruits and vegetables, and optimizing diet for development of health-promoting commensal microbiota (at least during infancy with breastmilk), protect against chronic inflammatory diseases, including asthma.

References

Abuabat, F., Alalwan, A., Masuadi, E., Murad, M.H., Jahdali, H.A., Ferwana, M.S., 2019. The role of oral magnesium supplements for the management of stable bronchial asthma: a systematic review and meta-analysis. NPJ Prim Care Respir Med 29, 4.

Ali, Z., Ulrik, C.S., 2013. Obesity and asthma: a coincidence or a causal relationship? A systematic review. Respir. Med. 107, 1287–1300.

Azizpour, Y., Delpisheh, A., Montazeri, Z., Sayehmiri, K., Darabi, B., 2018. Effect of childhood BMI on asthma: a systematic review and meta-analysis of case-control studies. BMC Pediatr. 18, 143.

Baffi, C.W., Wood, L., Winnica, D., Strollo Jr., P.J., Gladwin, M.T., Que, L.G., Holguin, F., 2016. Metabolic syndrome and the lung. Chest 149, 1525–1534.

Bahadoran, Z., Mirmiran, P., Tahmasebinejad, Z., Azizi, F., 2016. Dietary L-arginine intake and the incidence of coronary heart disease: tehran lipid and glucose study. Nutr. Metab. 13, 23.

Barnett, S.B., Nurmagambetov, T.A., 2011. Costs of asthma in the United States: 2002–2007. J. Allergy Clin. Immunol. 127, 145–152.

Bratt, J.M., Zeki, A.A., Last, J.A., Kenyon, N.J., 2011. Competitive metabolism of L-arginine: arginase as a therapeutic target in asthma. J. Biomed. Res. 25, 299–308.

Brustad, N., Eliasen, A.U., Stokholm, J., Bønnelykke, K., Bisgaard, H., Chawes, B.L., 2019. High-dose vitamin D supplementation during pregnancy and asthma in offspring at the age of 6 Years. JAMA 321, 1003–1005.

Busse, W.W., Morgan, W.J., Gergen, P.J., Mitchell, H.E., Gern, J.E., Liu, A.H., Gruchalla, R.S., Kattan, M., Teach, S.J., Pongracic, J.A., Chmiel, J.F., Steinbach, S.F., Calatroni, A., Togias, A., Thompson, K.M., Szefler, S.J., Sorkness, C.A., 2011. Randomized trial of omalizumab (anti-IgE) for asthma in inner-city children. N. Engl. J. Med. 364, 1005–1015.
Dawod, B., Marshall, J.S., Azad, M.B., 2021. Breastfeeding and the developmental origins of mucosal immunity: how human milk shapes the innate and adaptive mucosal immune systems. Curr. Opin. Gastroenterol. 37, 547–556.
Djuricic, I., Calder, P.C., 2021. Beneficial outcomes of omega-6 and omega-3 polyunsaturated fatty acids on human health: an update for 2021. Nutrients 13.
Dunn, R.M., Wechsler, M.E., 2015. Anti-interleukin therapy in asthma. Clin. Pharmacol. Ther. 97, 55–65.
GINA, 2021. Global Strategy for Asthma Management and Prevention, 2021. http://www.ginasthma.org/: Global Initiative for Asthma.
Goldring, S.T., Griffiths, C.J., Martineau, A.R., Robinson, S., Yu, C., Poulton, S., Kirkby, J.C., Stocks, J., Hooper, R., Shaheen, S.O., Warner, J.O., Boyle, R.J., 2013. Prenatal vitamin D supplementation and child respiratory health: a randomised controlled trial. PLoS One 8, e66627.
Harvey, S.M., Murphy, V.E., Whalen, O.M., Gibson, P.G., Jensen, M.E., 2021. Breastfeeding and wheeze-related outcomes in high-risk infants: a systematic review and meta-analysis. Am. J. Clin. Nutr. 113, 1609–1618.
Ho, N.T., Li, F., Lee-Sarwar, K.A., Tun, H.M., Brown, B.P., Pannaraj, P.S., Bender, J.M., Azad, M.B., Thompson, A.L., Weiss, S.T., Azcarate-Peril, M.A., Litonjua, A.A., Kozyrskyj, A.L., Jaspan, H.B., Aldrovandi, G.M., Kuhn, L., 2018. Meta-analysis of effects of exclusive breastfeeding on infant gut microbiota across populations. Nat. Commun. 9, 4169.
Holguin, F., Grasemann, H., Sharma, S., Winnica, D., Wasil, K., Smith, V., Cruse, M.H., Perez, N., Coleman, E., Scialla, T.J., Que, L.G., 2019. L-Citrulline increases nitric oxide and improves control in obese asthmatics. JCI Insight 4.
Hossain, N., Arhi, C., Borg, C.M., 2021. Is Bariatric surgery better than nonsurgical weight loss for improving asthma control? A systematic review. Obes. Surg. 31, 1810–1832.
Hosseini, B., Berthon, B.S., Wark, P., Wood, L.G., 2017. Effects of fruit and vegetable consumption on risk of asthma, wheezing and immune responses: a systematic review and meta-analysis. Nutrients 9.
Jolliffe, D.A., Greenberg, L., Hooper, R.L., Griffiths, C.J., Camargo Jr., C.A., Kerley, C.P., Jensen, M.E., Mauger, D., Stelmach, I., Urashima, M., Martineau, A.R., 2017. Vitamin D supplementation to prevent asthma exacerbations: a systematic review and meta-analysis of individual participant data. Lancet Respir. Med. 5, 881–890.
Lang, J.E., Mougey, E.B., Hossain, M.J., Livingston, F., Balagopal, P.B., Langdon, S., Lima, J.J., 2019. Fish oil supplementation in overweight/obese patients with uncontrolled asthma. A randomized trial. Ann. Am. Thorac. Soc. 16, 554–562.
Li, W., Qin, Z., Gao, J., Jiang, Z., Chai, Y., Guan, L., Ge, Y., Chen, Y., 2019. Vitamin D supplementation during pregnancy and the risk of wheezing in offspring: a systematic review and dose-response meta-analysis. J. Asthma 56, 1266–1273.
Lin, J., Zhang, Y., He, C., Dai, J., 2018. Probiotics supplementation in children with asthma: a systematic review and meta-analysis. J. Paediatr. Child Health 54, 953–961.
Lin, J., Zhang, Y., Zhu, X., Wang, D., Dai, J., 2020. Effects of supplementation with omega-3 fatty acids during pregnancy on asthma or wheeze of children: a systematic review and meta-analysis. J. Matern. Fetal Neonatal Med. 33, 1792–1801.
Linderholm, A.L., Bratt, J.M., Schuster, G.U., Zeki, A.A., Kenyon, N.J., 2014. Novel therapeutic strategies for adult obese asthmatics. Immunol. Allergy Clin. N. Am. 34, 809–823.
Litonjua, A.A., Carey, V.J., Laranjo, N., Stubbs, B.J., Mirzakhani, H., O'connor, G.T., Sandel, M., Beigelman, A., Bacharier, L.B., Zeiger, R.S., Schatz, M., Hollis, B.W., Weiss, S.T., 2020. Six-year follow-up of a trial of antenatal vitamin D for asthma reduction. N. Engl. J. Med. 382, 525–533.
Louis, R., Lau, L.C., Bron, A.O., Roldaan, A.C., Radermecker, M., Djukanović, R., 2000. The relationship between airways inflammation and asthma severity. Am. J. Respir. Crit. Care Med. 161, 9–16.
Lu, M., Litonjua, A.A., O'connor, G.T., Zeiger, R.S., Bacharier, L., Schatz, M., Carey, V.J., Weiss, S.T., Mirzakhani, H., 2021. Effect of early and late prenatal vitamin D and maternal asthma status on offspring asthma or recurrent wheeze. J. Allergy Clin. Immunol. 147, 1234-1241.e3.
Mabalirajan, U., Ahmad, T., Leishangthem, G.D., Joseph, D.A., Dinda, A.K., Agrawal, A., Ghosh, B., 2010. Beneficial effects of high dose of L-arginine on airway hyper-responsiveness and airway inflammation in a murine model of asthma. J. Allergy Clin. Immunol. 125, 626–635.
Milan, S.J., Hart, A., Wilkinson, M., 2013. Vitamin C for asthma and exercise-induced bronchoconstriction. Cochrane Database Syst. Rev. 2013, Cd010391.
Miles, E.A., Childs, C.E., Calder, P.C., 2021. Long-chain polyunsaturated fatty acids (LCPUFAs) and the developing immune system: a narrative review. Nutrients 13.
Miliku, K., Azad, M.B., 2018. Breastfeeding and the developmental origins of asthma: current evidence, possible mechanisms, and future research priorities. Nutrients 10.
Milutinovic, P.S., Alcorn, J.F., Englert, J.M., Crum, L.T., Oury, T.D., 2012. The receptor for advanced glycation end products is a central mediator of asthma pathogenesis. Am. J. Pathol. 181, 1215–1225.
Moreno-Macias, H., Romieu, I., 2014. Effects of antioxidant supplements and nutrients on patients with asthma and allergies. J. Allergy Clin. Immunol. 133, 1237–1244 quiz 1245.
Morris, C.R., Kuypers, F.A., Lavrisha, L., Ansari, M., Sweeters, N., Stewart, M., Gildengorin, G., Neumayr, L., Vichinsky, E.P., 2013. A randomized, placebo-controlled trial of arginine therapy for the treatment of children with sickle cell disease hospitalized with vaso-occlusive pain episodes. Haematologica 98, 1375–1382.
Norton, R.L., Hoffmann, P.R., 2012. Selenium and asthma. Mol. Aspect. Med. 33, 98–106.
Nyambuya, T.M., Dludla, P.V., Mxinwa, V., Nkambule, B.B., 2020. Obesity-related asthma in children is characterized by T-helper 1 rather than T-helper 2 immune response: a meta-analysis. Ann. Allergy Asthma Immunol. 125, 425-432.e4.
Ojima, A., Ishibashi, Y., Matsui, T., Maeda, S., Nishino, Y., Takeuchi, M., Fukami, K., Yamagishi, S., 2013. Glucagon-like peptide-1 receptor agonist inhibits asymmetric dimethylarginine generation in the kidney of streptozotocin-induced diabetic rats by blocking advanced glycation end product-induced protein arginine methyltranferase-1 expression. Am. J. Pathol. 182, 132–141.
Okoniewski, W., Lu, K.D., Forno, E., 2019. Weight loss for children and adults with obesity and asthma. A systematic review of randomized controlled trials. Ann. Am. Thorac. Soc. 16, 613–625.
Pahlavani, N., Jafari, M., Sadeghi, O., Rezaei, M., Rasad, H., Rahdar, H.A., Entezari, M.H., 2014. L-arginine supplementation and risk factors of cardiovascular diseases in healthy men: a double-blind randomized clinical trial. F1000Res. 3, 306.
Panigrahy, D., Gilligan, M.M., Serhan, C.N., Kashfi, K., 2021. Resolution of inflammation: an organizing principle in biology and medicine. Pharmacol. Ther. 227, 107879.
Papamichael, M.M., Katsardis, C., Lambert, K., Tsoukalas, D., Koutsilieris, M., Erbas, B., Itsiopoulos, C., 2019. Efficacy of a Mediterranean diet supplemented with fatty fish in ameliorating inflammation in paediatric asthma: a randomised controlled trial. J. Hum. Nutr. Diet. 32, 185–197.
Renz, H., Skevaki, C., 2021. Early life microbial exposures and allergy risks: opportunities for prevention. Nat. Rev. Immunol. 21, 177–191.
Rothenberg, M.E., Hogan, S.P., 2006. The eosinophil. Annu. Rev. Immunol. 24, 147–174.
Sahiner, U.M., Birben, E., Erzurum, S., Sackesen, C., Kalayci, Ö., 2018. Oxidative stress in asthma: Part of the puzzle. Pediatr. Allergy Immunol. 29, 789–800.
Shabana, M.A., Esawy, M.M., Ismail, N.A., Said, A.M., 2019. Predictive role of IL-17A/IL-10 ratio in persistent asthmatic patients on vitamin D supplement. Immunobiology 224, 721–727.
Shore, S.A., Cho, Y., 2016. Obesity and asthma: microbiome-metabolome interactions. Am. J. Respir. Cell Mol. Biol. 54, 609–617.
Sibal, L., Agarwal, S.C., Home, P.D., Boger, R.H., 2010. The role of asymmetric dimethylarginine (ADMA) in endothelial dysfunction and cardiovascular disease. Curr. Cardiol. Rev. 6, 82–90.
Singh, V.P., Aggarwal, R., Singh, S., Banik, A., Ahmad, T., Patnaik, B.R., Nappanveettil, G., Singh, K.P., Aggarwal, M.L., Ghosh, B., Agrawal, A., 2015. Metabolic syndrome is associated with increased oxo-nitrative stress and asthma-like changes in lungs. PLoS One 10, e0129850.
Strunk, R.C., Bloomberg, G.R., 2006. Omalizumab for asthma. N. Engl. J. Med. 354, 2689–2695.
Tsai, C.L., Lee, W.Y., Hanania, N.A., Camargo Jr., C.A., 2012. Age-related differences in clinical outcomes for acute asthma in the United States, 2006-2008. J. Allergy Clin. Immunol. 129, 1252-1258.e1.
Victora, C.G., Bahl, R., Barros, A.J., França, G.V., Horton, S., Krasevec, J., Murch, S., Sankar, M.J., Walker, N., Rollins, N.C., 2016. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. Lancet 387, 475–490.

Vieira, R.P., Duarte, A.C., Claudino, R.C., Perini, A., Santos, A.B., Moriya, H.T., Arantes-Costa, F.M., Martins, M.A., Carvalho, C.R., Dolhnikoff, M., 2007. Creatine supplementation exacerbates allergic lung inflammation and airway remodeling in mice. Am. J. Respir. Cell Mol. Biol. 37, 660–667.
Wang, Z., 2017. The Clinical Utility of Fractional Exhaled Nitric Oxide (FeNO) in Asthma Management. Agency for Healthcare Research and Quality, Rockville, MD.
Wei, X., Jiang, P., Liu, J., Sun, R., Zhu, L., 2020. Association between probiotic supplementation and asthma incidence in infants: a meta-analysis of randomized controlled trials. J. Asthma 57, 167–178.
Wilkinson, M., Hart, A., Milan, S.J., Sugumar, K., 2014. Vitamins C and E for asthma and exercise-induced bronchoconstriction. Cochrane Database Syst. Rev. 2014, Cd010749.
Willemsen, L.E.M., 2016. Dietary n-3 long chain polyunsaturated fatty acids in allergy prevention and asthma treatment. Eur. J. Pharmacol. 785, 174–186.
Wolsk, H.M., Chawes, B.L., Litonjua, A.A., Hollis, B.W., Waage, J., Stokholm, J., Bønnelykke, K., Bisgaard, H., Weiss, S.T., 2017. Prenatal vitamin D supplementation reduces risk of asthma/recurrent wheeze in early childhood: a combined analysis of two randomized controlled trials. PLoS One 12, e0186657.
Yang, H., Xun, P., He, K., 2013. Fish and fish oil intake in relation to risk of asthma: a systematic review and meta-analysis. PLoS One 8, e80048.

Nutrition and burn injury – manuscript

Marc G. Jeschke[a,b,c,d,f], **Sophie Brosset**[a,e], **and Anna Van Osch**[d], [a] Sunnybrook Research Institute, Toronto, ON, Canada; [b] Department of Surgery, Division of Plastic Surgery, University of Toronto, Toronto, ON, Canada; [c] Department of Immunology, University of Toronto, Toronto, ON, Canada; [d] Ross Tilley Burn Centre, Sunnybrook Health Sciences Centre, Toronto, ON, Canada; [e] Claude Bernard Lyon 1 University, Lyon, France; and [f] Hamilton Health Sciences, Deaprtment of Surgery Mc Master University, Hamilton, ON, Canada

© 2023 Elsevier Ltd. All rights reserved.

This is an update of S.A. Hill, Burns Patients, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 213–220, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00032-5.

Introduction: pathophysiology of burn injury and metabolism **523**
Nutrition after major burn **524**
Nutrition during the acute phase 524
Timing and route of nutrition support **525**
Increased energy expenditure **525**
Pathophysiology 525
Requirements 526
Substrate-specific requirements **526**
Protein 526
Pathophysiology 526
Requirements 526
Fat **527**
Pathophysiology 527
Requirements 527
Carbohydrate **528**
Pathophysiology 528
Requirements 529
Complementary treatment 529
Micronutrients and specific amino acids **529**
Glutamine **529**
Pathophysiology 529
Requirements 529
Vitamins **530**
Trace elements **530**
Monitoring of nutrition support **530**
Nutrition during recovery **531**
Metabolic disturbance during recovery and long-term effects 531
Nutrition requirements 533
Nutrition support after minor burn **533**
Conclusion **534**
References **534**
Further Readings **534**

Key points/objectives

Recommendations for nutritional management of major burns:

- The nutrition support should be adapted to REE, approximated by indirect calorimetry, multiplied by a factor of 1.2–1.4, and metabolic capacity.
- Enteral nutrition should be initiate early using a high-carbohydrate, high-protein, and low-fat formula.
  - Carbohydrate should be the primary energy source (>50–60% energy) but should not exceed the body's ability to assimilate substrate (7 g/kg/day).
  - Fat should provide < 15% of energy.
  - Protein intake of 1.5–2.0 g/kg/day should be provided to limit lean body mass (LBM) loss.
  - Vitamins C, D, E, and K and trace elements, copper (Cu), selenium (Se) and zinc (Zn), should be supplemented. Further studies are required to determine if glutamine supplementation is beneficial and to recommended micronutrient supplementation doses.

- Hyperglycemia should be managed with intensive insulin therapy.Recommendations for nutritional management of minor burns:
  - In robust patients, sufficient oral intake of a high-protein, high-carbohydrate diet should be encouraged along with supplementation of micronutrients to meet daily recommended intake (DRI) to facilitate wound closure.
  - Selected patients with pre-existing malnutrition or at risk for malnutrition such as children and elderly, may require nutrition support to achieve complete wound healing.

Glossary

Anabolism net accumulation of protein stores in response to hormones. This may be endogenous in recovery from injury or exogenous in response to administered steroids

Catabolism the net breakdown of protein and fat stores as a response to injury or stress

Hypermetabolism reflects an increase in whole body energy consumption above normal values. Patients are considered hypermetabolic when their resting energy expenditure (REE) is 10% or more above normal

Major burn burns covering greater than >10% total body surface area (TBSA) in elderly, >20% TBSA in adults, and >30% TBSA in children, initiating a prolonged, systemic inflammatory response

Minor burn burn covering less than 20% TBSA with inflammation localized to the injury site.

Nutrition support providing enteral and/or parenteral nutrition to treat or prevent malnutrition

Introduction: pathophysiology of burn injury and metabolism

The majority of burns sustained are thermal injuries caused by scald or flame; however, tissue damage can be induced by other agents, including friction, cold, heat, radiation, chemical, or electrical. Treatment options vary based on the mechanism of injury as well as the depth and size of burn. Superficial (affecting only the epidermis) and superficial-partial thickness (deeper, but still contained to the epidermis) burns are anticipated to heal spontaneously and do not require intervention beyond aseptic dressing changes. Deep partial thickness (extending into the dermis), full thickness (extend through the full dermis), and fourth-degree burns (involving deeper tissue, such as muscle or bone) require careful management including antimicrobial dressings and surgery. Surgery involves excising the burn until a viable wound bed is reached followed by restoration of the epidermal layer by autologous skin graft taken from a donor site on the patient.

Major burns, commonly categorized as >10% TBSA in elderly, >20% TBSA in adults, and >30% TBSA in children, are unique as they set off an immediate systemic and local stress response that, unlike sepsis and trauma, persists for a prolonged period. This persistent inflammatory reaction will lead to major metabolic disorder and organs dysfunction (Fig. 1). Extensive destruction of the skin induces the release of stress hormones and pro-inflammatory mediators. The immediate response comprises endothelial and acute heart dysfunction leading a state of shock and a hypometabolic response that lasts for ~72–96 h. The gut develops mucosal atrophy due to hypoperfusion and edema, enabling bacterial translocation. The systemic response rapidly evolves into a hypermetabolic phase fueled by the release of stress hormones (cathecholamines) and cytokines, that can persist for years following the initial injury The heart goes into hyperdynamic overdrive, increasing circulation and blood flow to increase oxygen and nutrient delivery. Increased stress signaling causes changes in metabolic demand. Protein is degraded to deliver energy for hepatic function. White adipose tissue changes to brown adipose tissue allowing for increased lipolysis with the accompanying expression of lipotoxic intermediates, such as triglycerides, free fatty acids, and diacylglycerols (DAG), all of which are transferred to the liver. Despite an increase of gluconeogenesis, the liver is unable to metabolize all the accumulating substances and hepatomegaly develops. Hyperlipidemia and hyperglycemia with insulin resistance are present, which worsens the hypermetabolic and inflammatory state. Furthermore, intercurrent complication and surgical intervention trigger inflammatory reactions. If hypermetabolism cannot be diminished or decreased, holistic catabolism ensues with decrease of the lean body mass, and subsequent immunocompromized, multiple organ failure and death.

It is not feasible to halt the hypermetabolic response, but complementary strategies are available to diminish its magnitude. Thermoregulation by increasing ambient room temperature and using occlusive dressings reduces REE by limiting the need for heat production. Early excision, within 72 h of injury, and closure of the burn wounds leads to substantially reduced REE and subsequent improvement in mortality rates. Anabolic treatment such as oxandrolone, insulin, metformin, growth hormones and propranolol, can also be used to promote wound healing. At present, beta-adrenergic blockade with propranolol is probably the most efficacious anti-catabolic therapy in the treatment of burns. However, early and adequate nutrition support remains the cornerstone in addressing this hypermetabolic response, as prolonged catabolism is associated with delayed wound healing and increased incidence of organ failure, infections and death (Porter et al., 2016). The inflammatory response, triggered almost immediately after injury, is intended to create an environment that fosters wound healing. During the brief hypometabolic period (ebb

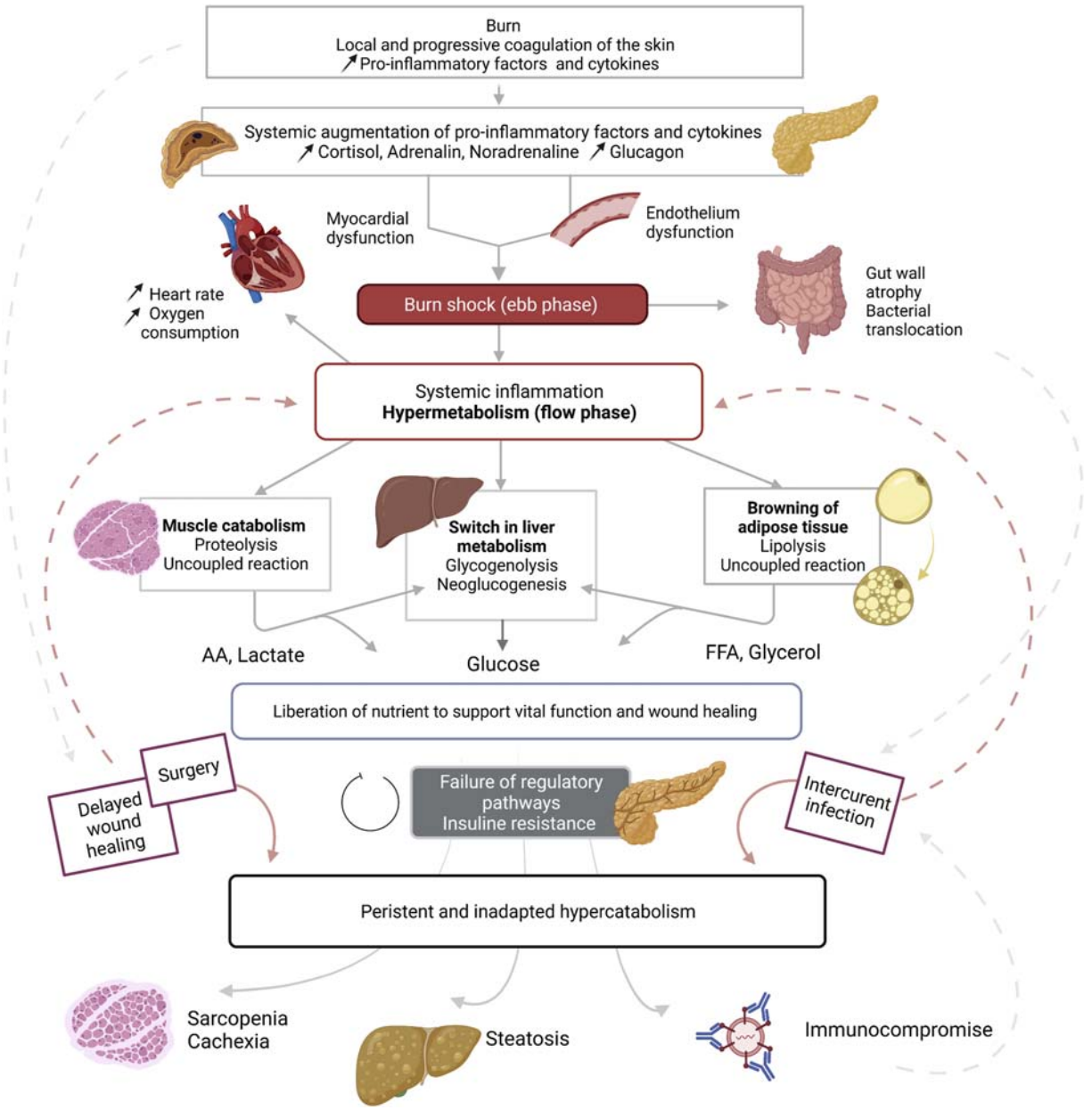


Fig. 1 Simplified physiopathology of systemic effect of major burn lesions (created with Biorender).

phase) following initial injury the body draws primarily on endogenous energy sources to maintain vital functions; In major burns this extensive inflammation can become uncontrolled resulting in an hypermetabolic state (flow phase) with persistent hypercatabolism to support it. (Jeschke et al., 2020). Aggressive nutrition support is required to counteract the potentially devastating effects of persistent hypercatabolism.

Nutrition after major burn

Nutrition during the acute phase

The purpose of nutrition is to provide adequate macro- and micronutrients for tissue regeneration, to limit the deleterious effects of hypermetabolism, and to support immune function. A nutrition focused assessment should be conducted by a qualified clinician before initiating nutrition support. A review of patient's pre-injury nutrition status which can include anthropometric data, biochemical data, nutrition intake, physical examination, and functional status will help to guide the nutrition intervention (Table 1). Mechanism, extent, site and time since injury should also be considered during assessment. In addition to mechanically

Table 1 Clinical assessment of factors influencing nutrition status of burn patients

*Biochemical measure should be interpreted within context of clinical picture.

Domain	*Indicators*
Anthropometric	Height
	Weight
	Body mass index (BMI) or ideal body weight
	Weight history
Biochemical *	Electrolytes
	Blood glucose
	Renal function (creatinine, urea)
	Inflammatory markers (C-reactive protein, white blood cell count)
Nutrition intake	Diet history (diet restrictions, allergies, changes in diet intake, use of oral supplements or nutrition support, etc.)
Nutrition focused physical exam	Evidence of muscle or subcutaneous fat loss (including sarcopenia and cachexia)
	Evidence of adiposity or obesity
	Fluid accumulation or edema
Functional status	Ability to perform usual activities

ventilated patients, those with major burns, inhalation injuries, failure to maintain LBM with oral intake or selected patients with less extensive burns but at risk for malnutrition, should be considered for initiation of nutrition support. The elderly, children, and obese patients with evidence of sarcopenia are groups at risk for under nutrition. Burn patients require frequent reassessment of their nutrition care plan as wound healing, physical therapy and medical complications can all affect nutrition needs.

Timing and route of nutrition support

In the first days following injury, energy production from endogenous sources in reaction to the inflammatory syndrome is increased due to a catabolic state and is sufficient to meet energy requirements. Full nutrition support in this phase may lead to overfeeding as exogenous nutrition cannot abolish endogenous energy production, and organs are unable to process the excess substrate. However, early enteral nutrition (EN), is recommended to maintain gastrointestinal (GI) trophicity (Wise et al., 2019). The American Society of Parenteral and Enteral Nutrition (ASPEN), European Society of Parenteral and Enteral Nutrition (ESPEN) and the Society of Critical Care Medicine (SCCM) recommend initiating EN within 12 h of injury, and as early as 4–6 h after injury to achieve maximal benefit (Wise et al., 2019; McClave et al., 2016; Rousseau et al., 2013). Early EN results in improved gut mucosal integrity which decreases incidence of bacterial translocation, enhances wound healing and LBM maintenance, modulates stress hormone levels and immunoglobulin production, and may improve EN tolerance for the duration of hospitalization. Once the patient is adequately resuscitated and hemodynamically stable, EN can be advanced to goal as quickly as tolerated.

EN is the preferred form of nutrition support; however, parenteral nutrition (PN) may be required for patients with non-functioning GI tracts, EN intolerance or unable to achieve goal EN in a reasonable time. There is no consensus on the appropriate timing of PN initiation. PN is avoided in many burn units due a historical reporting of increased risk for complications including hyperglycemia, infection, liver malfunction and mortality. The role of PN may require review, as improved infection control techniques and judicious prescribing to limit energy overfeeding and hyperglycemia have been demonstrated to be effective at increasing protein intake in pediatric patients without increasing incidence of blood or respiratory infections (Carson, 2018).

Increased energy expenditure

Pathophysiology

Thermal injury elevates resting energy expenditure (REE) via increased basal metabolic rate, tissue regeneration and thermoregulation (Fig. 2). Acute events, such as sepsis, fever and surgery, may punctuate the initial phase causing physiological stress and significant variation in energy requirements. In the presence of oxygen all macronutrients, including carbohydrate, fat, and protein, will be used to produce energy, either in the form of ATP or heat via oxidation, resulting in carbon dioxide and water formation. Degradation of amino acids will lead to the formation of nitrogen derivatives which are eliminated in the urine as urea.

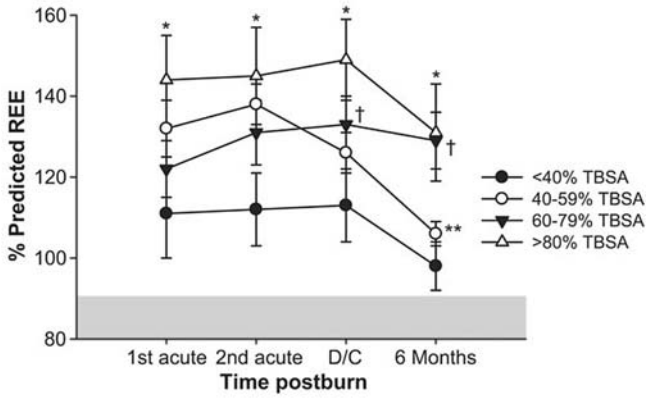


Fig. 2 Variation in resting energy expenditure from 1 week to 6 months following burn injury. Increase in predicted resting energy expenditure is correlated with the extent of burn injury. Reproduced from Jeschke, M.G., Finnerty, C.C., Herndon, D.N. et al. (2007) *Crit. Care* **11**, R90.

Requirements

REE can be approximated via indirect calorimetry by measuring the amount of inspired oxygen (VO2) and the amount of expired carbon dioxide (VCO2) using the following Haldane transformation of Weir's equation:

$$\text{REE (Kcal / day)} = 1.44 \times [(\text{VO2 (mL / min)} \times 3.94) + (\text{VCO2 (mL / min)} \times 1.11)]$$

In this formula the nitrogen derivatives are not considered leading to an error of 2–4%. Indirect calorimetry is the accepted gold standard for determining measured REE in the clinical setting (McClave et al., 2016). Minimum weekly repeated measures during acute hospitalization, or more frequently if there are changes in clinical status, are recommended to guide provision of nutrition support (Rodriguez et al., 2011). Generally, a brief measurement is taken during a steady state, when a patient is resting with no additional stresses, then extrapolated to predict energy expenditure for a 24-h period. Expensive equipment, clinician time and agitated patients with unstable oxygen requirements are common barriers to the use of indirect calorimetry.

Predictive equations are a popular substitute to estimate REE when indirect calorimetry is not available. The adapted Toronto formula correlates best with measured REE in the adult population, where Curreri, Harris-Benedict, Schofield-HW, and World Health Organization equations overpredict REE (Rodriguez et al., 2011). For pediatric patients, the Galveston equation is commonly recommended to predict energy requirements in the absence of indirect calorimetry (see *Monitoring* section, **Table 3**) (Rodriguez et al., 2011; Clark et al., 2017). Predictive equations should be used with caution as energy expenditure fluctuates during treatment and recovery, and fixed formulas often lead to underfeeding during periods of highest energy utilization and overfeeding late in the treatment course. Energy goals should be adapted based on clinical events and clinician expertise.

When using indirect calorimetry or predictive equations, a correcting factor of 30%, obtained by multiplying the REE by a factor of 1.2–1.4, results in predicted total energy expenditure (TEE) (Rodriguez et al., 2011). TEE considers energy expended on clinical activities, such as dressing changes, surgery, passive physiotherapy, and targets patient weight maintenance.

Substrate-specific requirements

Protein

Pathophysiology

Increase in skeletal muscle protein synthesis accompanies burn injury; however, it is surpassed by the concurrent hypercatabolic response, resulting in net losses of muscle proteins and systemic release of amino-acids. Amino acid efflux from skeletal muscle facilitates other vital metabolic processes in burn patients, such as gluconeogenesis, wound healing, immune response and hepatic protein synthesis. Major predictors of the magnitude of protein catabolism are age, weight and delay in definitive surgical treatment. Intercurrent sepsis episodes can also increase protein catabolism. Metabolic alterations interplay with protein metabolism, with hyperglycemia linked to increased muscle protein catabolism. Unabated, protein losses can exceed 150 g/day in severe burns leading loss of LBM in excess of half a pound of skeletal muscle daily. This can delay healing and significantly contribute to the long-term morbidity of burn survivors.

Requirements

Nutrition should supply sufficient exogenous protein to be used as a substrate for wound healing in order to spare the LBM protein pool. Nitrogen balance studies compare the difference between nitrogen intake and nitrogen losses and are a common clinical tool to assess for adequate protein provision. When using the Lee and Hatley formula a positive nitrogen balance of 2–4 g/day creates a desirable environment for anabolism. Two factors can affect the accuracy of these results in burns patients. Total urinary nitrogen

(TUN) is approximated by adding 4 g/dL to urinary urea nitrogen (UUN); however, this may be inappropriate during the acute burn phase of inconsistent ureagenesis (Clark et al., 2017). Secondly, major burns can loss an additional 20–25 g/m2 TBSA per day from wound exudate. A proposed solution is to add a correcting factor of 1.25 to TUN losses to account for wound exudate losses (Clark et al., 2017). Caution should be applied when using this formula as actual nitrogen losses can vary significantly from the proposed constants depending on the extent of injury.

$$\text{Nitrogen Balance} = 24 - \text{hour protein intake} - [1.25 \times (\text{UUN} + 4)]$$

As well documented in the critical care literature, visceral proteins should not be used as a nutrition marker in the acute or post-acute setting as they are subject to fluid shifts and hepatic reprioritization.

At present, protein recommendations are generally based on expert consensus and established institution practices. Supplying supranormal doses of protein does not reduce the catabolism of endogenous protein stores, but it does facilitate protein synthesis, particularly in skin, and reduces negative nitrogen balance. Accepted clinical practice guidelines recommend protein doses of 1.5–2.0 g/kg/day in thermally injured adults (McClave et al., 2016; Rousseau et al., 2013). Burn specific studies suggest that doses of greater than 2.2 g/kg/day did not have effects on net protein synthesis and should be carefully monitored as there is an increased risk of developing azotemia (McClave et al., 2016; Rousseau et al., 2013). In the pediatric population recommended protein doses range from 2.5 to 4.0 g/kg/day with the International Society of Burn Injury recommending 3 g/kg/day (Wise et al., 2019; Rodriguez et al., 2011).

Fat

Pathophysiology

Under prolonged adrenergic stress, major modifications of fatty tissue are observed. The energy storing white adipocytes switch their morphology into energy releasing "beige" adipocytes by adopting brown fat characteristics, a phenomenon called "browning". Heat production by uncoupled reaction and lipolysis are increased. Triglyceride (TG) reserves in adipocytes are broken down into glycerol and free fatty acids (FFA) and released in the blood stream to be metabolized by essential organs as a substrate for ATP production. Lipid profiling analysis over time revealed a prominent elevation of saturated fatty acid, mono-unsaturated fatty acid and omega-6 (w-6) polyunsaturated acid (PUFA). In cases of major burn injury, the rate of lipolysis significantly exceeds the capacity of metabolism. At most 30% of FFA are oxidized for energy, compared to up to 90% in starvation. The remaining 70% of FFA are recycled and undergo re-esterification in adipose and non-adipose tissue resulting in a futile cycling of energy substrates. Circulating FFA, particularly palmitic and oleic are pro-inflammatory and cause endoplasmic reticulum stress and apoptosis, leading to reduction in fat β-oxidation and worsening of the imbalance between FFA release and metabolism. The heart, muscles, kidneys and particularly liver are infiltrated by ectopic fat stores (i.e., steatosis) contributing to organ dysfunction. The regulatory pathway which is mediated by insulin, is inefficient due to a high peripheral insulin resistance. In most patients, FFA and TG concentrations normalize over a four-week period, but enhanced lipolysis can persist years after the initial injury resulting in adipose tissue wasting and systemic lipotoxicity. The modulation of the adipose tissue response to burn injury is now a major therapeutic target.

Requirements

The current understanding of altered fat metabolism after burn injury favors a high-carbohydrate, low-fat diet. While high fat intake doesn't diminished proteolysis, low-fat nutrition support has been shown to reduce infectious morbidity and shorten length of stay (LOS) in burn patients. A minimum of 2–4% of energy should be supplied from fat to avoid essential fatty acid deficiency. General expert consensus is that fat should make up 3–15% of energy requirements. While trying to minimize fat intake, titration of fat calories may be required to meet energy requirements while respecting the body's upper limit of glucose utilization (see *Carbohydrate* section). These low-fat recommendations may be difficult to achieve in the clinical setting as commonly available commercial, liquid formulas tend to provide 29–45% of energy as fat. This has led some burn centers to fortify existing formulas to achieve a more desirable macronutrient distribution, but this is not always feasible given available resources.

It is also prudent to account for non-dietary fat calories. Commonly used for sedation, 1 or 2% propofol infusion delivers 1.1 kcal/mL of fat and can significantly contribute to metabolic alterations if the additional fat calories are not accounted for in the nutrition care plan.

Composition of administered fat should also be considered. Omega-3 (w-3) and w-6 PUFAs, and their derivatives are predominantly derived from exogenous sources and considered essential in promoting and resolving inflammation following trauma by providing a source of resolvins, eicosenoids, and prostaglandins. Omega-6 fatty acids, the most common fat source in enteral and parenteral formulas in the United States, have been shown to encourage inflammation which may exacerbate the already heightened inflammatory response of burn patients. Lipids that contain a high percentage of w-3 fatty acids are metabolized without promoting proinflammatory molecules and have been linked to enhanced immune response, reduced hyperglycemia, and improved outcomes. An omega 6:3 ratio closer to 1:1 could enhance the immune response. As more research becomes available, the role of different fat source and provision of total fat calories may require reassessment.

Carbohydrate

Pathophysiology

A state of hyperglycemia beginning within the first hours of injury is associated with major burns (Fig. 3). Release of large quantities of catecholamine, glucagon and corticoids immediately following injury induces breakdown of glycogen in the liver. Increased proteolysis and lipolysis release the substrates necessary for gluconeogenesis, pyruvate, lactate, glycerol, and amino acids, in large quantities. As a result the rate of glucose release from the liver is increased two-fold. These physiological reactions liberate abundant endogenous energy reserves to allow for continued maintenance of vital functions and cell regeneration.

Two phenomena contribute to prolonged imbalance of carbohydrate metabolism and concurrent hyperglycemia in burn patients. First the glucose metabolism pathway is deeply altered. Anaerobic glycolysis prevails, allowing for limited production of ATP. Pyruvate and lactate formation are promoted, with subsequent gluconeogenesis via an extremely ATP-consuming reaction. Secondly, regulatory pathways are inefficient. Exogenous glucose infusion does not fully inhibit hepatic glucose production and despite gradual elevation of fasting insulin levels, hyperglycemia persists (Fig. 3). The mechanisms underlying this insulin resistance are poorly understood, and could result from direct hyperglycemic action of catecholamine, glucagon and cortisol, by alteration of insulin-receptor signaling by cytokine as monocyte chemoattractant protein-1 and interleukin-6 and/or from organelle dysfunction in the peripheral cell secondary to cellular stress responses, resulting in reduced glucose assimilation. The persistence of insulin resistance during the flow phase could also be due in part to the catabolic decrease in muscle mass seen with severe injury, since skeletal muscle is responsible for 70–80% of whole body insulin-stimulated glucose uptake.

Burn patients with poor glucose control have shown worse outcomes, including a significant risk of infectious complications, especially wound infections, pneumonia and bacteremia, which are potentially explained by altered immunity, impaired wound healing, reduction in graft take, enhanced muscle protein catabolism, and overall increase in mortality. Retrospective studies and a prospective randomized trial have further confirmed that good glucose control is beneficial in terms of post-burn morbidity and organ function.

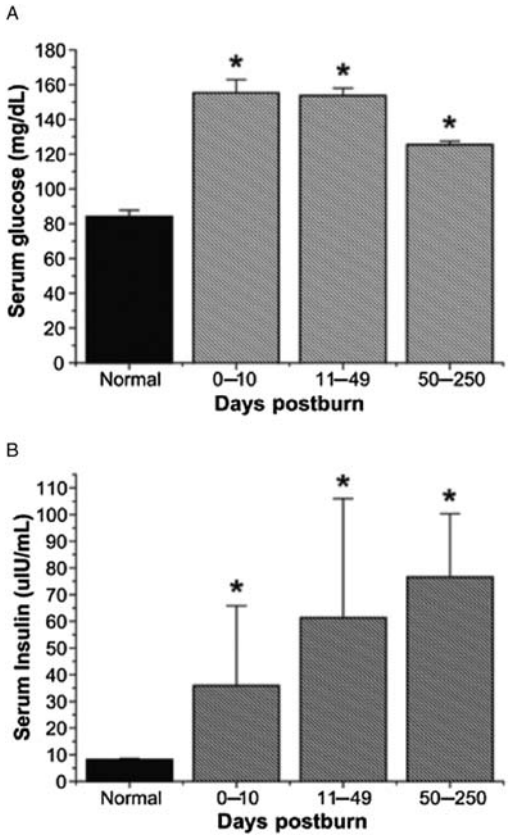


Fig. 3 Major burn injury leads to vast inflammatory stress responses, hyperglycemia, and hyperinsulinemia. Levels of fasting serum glucose (A) and insulin (B) were measured at various time points postburn and compared with non-burned children. Bars represent mean; error bars correspond to SEM. *Statistical difference between burned children versus non-burned children, $P < 0.05$. Reproduced from Jeschke, M.G., Finnerty, C.C., Herndon, D.N. et al. (2012) *Ann Surg* **255, *370–378***.

Requirements

Carbohydrates remain the preferred energy source for burn patients as high-carbohydrate diets impart a protein-sparing effect and promote wound healing. Ideally >50–60% of energy in the diet should come from carbohydrate (Rousseau et al., 2013; Carson, 2018). At minimum of 2 g/kg/day of carbohydrate should be provided to meet baseline adult glucose requirements, but carbohydrate intake should not exceed 5–7 g/kg/day or 5 mg/kg/min, the maximum rate the body is able to oxidize glucose. This is important to note as energy requirements may exceed the body's metabolic capacity to assimilate substrate. In this case, providing a limited amount of dietary fat can diminishes the need for carbohydrates and ameliorates glucose tolerance. If glucose is given in excess of what can be utilized, it leads to hyperglycemia, the conversion of glucose to stored fat, glucosuria and dehydration (Rodriguez et al., 2011). Although a large prospective randomized controlled trial (RCT) is lacking, a recent review was consistent with current recommendation for high carbohydrate diet by demonstrating possible clinical benefits including less pneumonia, less wound infections, shorter hospital LOS, less acute respiratory distress syndrome, shorter healing time, and lower rates of sepsis. Use of fiber containing formulas is encouraged, as high opioids doses for pain management can lead to significant constipation.

Complementary treatment

Exogenous insulin is often required to manage hyperglycemia in the acute phase. Exogenous insulin can help achieve normoglycemia while conferring the additional benefit of stimulating muscle protein synthesis and improving wound healing (Rousseau et al., 2013; Rodriguez et al., 2011). While the exact cut off for benefit is not defined, maintaining blood glucose levels below 130 mg/dL using intensive insulin therapy has been shown to reduce mortality and morbidity in burn patients (Rousseau et al., 2013). Severely burned patients who received insulin infusions, in conjunction with a high-carbohydrate, high-protein diet, have improved donor site healing, LBM, bone mineral density, and decreased LOS. Intensive insulin therapy should be closely monitored given the risk of hypoglycemia which can quickly increase risks of morbidity and mortality (Rodriguez et al., 2011; Clark et al., 2017). Due to the concern that cellular response to exogenous insulin could be limited by organelle dysfunction, alternative strategies to improve glucose assimilation in peripheral cells and to decrease gluconeogenesis, including the use of metformin and the PPAR-agonist fenofibrate, are being investigated.

Micronutrients and specific amino acids

Micronutrients, consisting of vitamin and mineral, are nutrients required in very small amount to enable proper metabolism, as they function to maintain the structure of enzymes or as enzyme cofactors. A major burn sets off a metabolic sequela that increases the body's need for specific micronutrients and amino acids above normal levels to enable wound healing.

Glutamine

Pathophysiology

Supplementation of specific amino acids, particularly glutamine (GLN) and arginine, is the ongoing subject of research. GLN is the main energic substrate for enterocytes and lymphocytes. GLN deficiency can lead to intestinal epithelial barrier dysfunction, with in vitro and in vivo studies demonstrating that GLN is an essential nutrient for lymphocyte proliferation, cytokine production, macrophage phagocytic plus secretory activities, and neutrophil bacterial killing. Significant decrease in GLN plasmatic concentration is observed in major burns, and is postulated be due to a major increase in hepatic consumption for gluconeogenesis. A small, randomized study reported a positive effect of GLN supplementation in the burn population with significant mortality and LOS reductions. A reduction in bacterial translocation and therefore gram-negative infections is also reported by a small double-blinded, RCT.

Requirements

For units choosing to supplement GLN, the current recommendation is to administer 0.3–0.5 g/kg/day enteral GLN for 10–15 days starting with the initiation of EN for patients with burns >20% TBSA and adjusting doses for patients with a body mass index >35 kg/m2 (Rousseau et al., 2013). Caution should be used when supplementing GLN in patients with hepatic failure, renal failure, septic shock, or high dose vasopressors as it can lead to accumulation of urea and ammonia. A study and systemic review using high dose GLN supplementation and lower dose parenteral GLN supplementation showed no benefit and a trend toward increased 28 day and 6 month mortality in critically ill patients with multi-organ failure. However, the systemic review did show a signal toward reduced hospital mortality in burn patients, prompting the RE-ENERGIZE multicenter RCT. The RE-ENERGIZE trial recently closed, seek to evaluate the effect of enteral GLN supplementation in 1200 burn patients and should clarify the potential role for GLN supplementation. This large multicentre trial however indicates that glutamine has no benefical effect in terms of mortlaity and incidence of infections (Heyland et al., 2022 NEJM in press).

Vitamins

Ascorbic acid (vitamin C), cholecalciferol (vitamin D), α-tocopherol (vitamin E), and menaquinone (vitamin K) have all been reported as depleted in burned patients and deficiencies are suspected to contribute to delayed wound healing and immune dysfunction after major burn injury. Supplementation of these vitamins at pharmacological doses has been proposed to improve outcomes.

Vitamin C acts to reduce damage to capillaries from reactive oxygen species, stabilize endothelial membranes and limits their permeability during initial shock (Blaauw et al., 2013). Two clinical trials were able to demonstrate high dose vitamin C infusion during resuscitation were effective at deceasing fluid volume needed for stabilization, resulting in less edema, and in one case less respiratory impairment, without inducing renal failure. While pharmacological doses of vitamin C during resuscitation appear safe, more research is needed to determine optimal doses. During wound healing, vitamin C is paramount for the synthesis and cross-linking of collagen to restore basement membrane and dermal collagen matrix.

Vitamin C and E work in tandem to scavenge free radicals and dispose of toxic oxidants, with vitamin C acting as a regenerator of oxidized vitamin E. By limiting oxidative damage to cell membrane structure there is the potential to improve end organ function and wound healing. Burned children show tissue storage of vitamin E decrease by half as early as three weeks after injury. Burned children supplemented with a combination of vitamins C, E, and zinc showed decreased malondialdehyde, which is a marker of oxidative stress, and needed significantly fewer days to heal compared to un-supplemented children.

Systemic inflammatory response syndrome, bleeding, decreased intake and absorption of vitamin D, organ and metabolic dysfunction, and decreased sun exposure during hospitalization and recovery as part of scar management leads to a high prevalence of vitamin D deficiency following large burns. Long-lasting impairment of burn scar tissue and adjacent tissue to effectively activate vitamin D has been shown, with impairment observed for up to seven years after injury creating an environment for chronic deficiency. While more work needs to be done to elucidate the preferred form and dose of vitamin D supplementation, there is agreement DRI doses are insufficient for repletion and maintenance (Blaauw et al., 2013).

Malabsorption, limited enteral intake, antibiotic therapy, and multiple surgical procedures puts burn patients at risk of vitamin K deficiency. Ninety-one percent of burned children showed below expected values of vitamin K for the first four weeks post-burn, however, clinical relevance remains unclear as no relationship was found between serum values and prothrombin time or activated partial thromboplastin time.

While clinical benefits of pharmacological doses of vitamins are being investigated, clinicians should ensure complimentary vitamins are provide to at minimum meet DRIs (Blaauw et al., 2013).

Trace elements

Trace elements (TE), Cu, iron (Fe), manganese (Mn), Se, and Zn are essential to catalysis reactions in enzyme systems and enabling electron transfer; particularly for reduction of reactive nitrogen and oxygen species, innate and adaptative immune defenses, energy metabolism and tissue repair (**Table 2**) (Berger et al., 2007). Inflammation immediately following burn injury decreases circulating TEs, while rapid uptake by the liver, spleen and kidney induces a relative depletion of Fe, Se, Zn while Cu and Mn levels increase. Without replacement, TE losses from wound exudate and peri-operative blood loss can lead to biochemical deficiencies of Cu, Fe, Mn, Se and Zn by the end of the first week. Clinical signs of deficiency have only been reported at late stages of recovery, when increased requirements for skin healing have completely depleted the body's stores (Berger et al., 2007). TE deficiencies are linked to poorer outcomes, specifically Cu deficiency which has been linked to fatal arrhythmias.

Evidence of clinical benefit of TE supplementation remains sparse. In a small, randomized trial, TE supplementation was associated with higher circulating plasma and skin tissue content of selenium and zinc with improved wound healing and fewer pulmonary infections. A meta-analysis confirmed decreased infectious episodes but failed to prove significant effect on wound healing, LOS or mortality.

Antioxidant protocols, which combine vitamin and TE supplementation, aim to reduce inflammation following burn or traumatic injury through their interplay against free radicals. Reduced energy expenditure and pro-inflammatory cytokine were observed after introduction of antioxidant protocol in a cohort of patients. A large study of critically ill patients failed to confirm any benefit with similar supplementation. A large, RCT in burn patients is needed to assess efficacy.

Monitoring of nutrition support

Careful clinical monitoring of abdominal pain and distention, vomiting, and bowel function are prudent to ensure safe EN administration. There is conflicting data about the validity of gastric residual volumes as an accurate marker of EN tolerance; however, some centers continue to use them as part of routine monitoring. Paralytic ileus of the stomach and colon is a common consequence of large volume fluid resuscitation, and insertion of a post-pyloric feeding tube can help enhance EN tolerance in the early acute phase despite gastroduodenal dysfunction (Rousseau et al., 2013; Rodriguez et al., 2011).

Under delivery of EN is an established problem in burn units with frequent interruptions to EN infusions, such as for surgeries, dressing changes and tests, resulting in formula equivalent of up to 930 kcal/day not being infused. A variety of strategies can be implemented to minimize feed interruptions or compensate for lost infusion time. Burn centers have demonstrated success at

Table 2 Roles of trace elements in burn metabolism (not exhaustive)
(Berger and Shenkin, 2007). DRI doses of micronutrients may be sufficient for minor burns; however, strong consideration should be given to providing therapeutic doses of intravenous (IV) micronutrients for burns that exceed 20% TBSA. Following a review of available burn literature, Blaauw et al. propose that daily IV doses of 500–700 μg Se, 4 mg Cu, and 40 mg Zinc for 2–3 weeks following major burn is safe and does not require specific blood concentration monitoring (Blaauw et al., 2019). Mn and Fe should be provided to meet DRI values. IV supplementation is recommended as absorption by enteral route may be variable. Patients receiving continuous renal replacement therapy (CRRT) may also require higher TE doses (Pantet et al., 2019).

Trace element	*Effect*
Copper	Humoral and cellular immunity
	Wound healing: collagen synthesis, lysyl oxidase
	Fatal arrhythmias
	Oxidative stress reduction: superoxide dismutase enzymes
Iron	Humoral and cellular immunity
	Cofactor in oxygen-carrying proteins
	Wound healing: elastin and collagen synthesis
Selenium	Humoral and cellular immunity
	Modulation of inflammatory response: activation of the transcription factor NFkB
	Oxidative stress reduction: component of the active site of glutathione peroxidase (in neutrophilis)
Zinc	Humoral and cellular immunity: lymphocyte function
	Wound healing: protein synthesis, DNA replication
	Modulation of inflammatory response: activation of the transcription factor NFkB
	Oxidative stress reduction: superoxide dismutase enzymes
	Metabolic pathways
Manganese	Wound healing: mucopolysaccharide and glycoprotein synthesis
	Oxidative stress reduction: superoxide dismutase enzymes

achieving prescribed EN goals after implementing volume-based feeding protocol, where hourly infuse rates are adjusted to meet 24-h formula volume targets. Intra-operative feeding of patients with protected airways is gaining popularity as adult and pediatric studies have demonstrated increased total calories and protein delivered compared to traditionally fasted patients, with no increased risk of mortality, wound infections, pneumonia or aspiration using both gastric and post-pyloric feeding tubes.

Nutrition support should continue until the patient demonstrates their oral intake is adequate to meet nutrition demands. Oral nutrition supplements and foods fortified to increase calorie and protein density can aid in achieving oral energy and protein goals.

Nutrition during recovery

Metabolic disturbance during recovery and long-term effects

The hypermetabolic response to burn is essential to maintain organ function and whole-body homeostasis under demanding trauma conditions; however, persistent metabolic alteration can be disruptive to recovery. Children recovering from 30% TBSA or greater burns were shown to have a significant elevation in cortisol, catecholamines, cytokines and acute phase proteins compared to their non-injured counterparts. Indirect calorimetry revealed measured REE exceeding predicted REE. Persistent negative skeletal muscle protein balance has been demonstrated for the first year following burn injury. Significant hepatomegaly and cardiac dysfunction, including increased heart rate and cardiac output, also persisted into the recovery phase. Epidemiological studies report an increased risk of morbidity and mortality in the long term due to cardio-circulatory complication after extensive burns in pediatric and adult populations, although the precise pathophysiology has not been elucidated. Significant growth delays are seen in children, with data showing it take 1–2 years for pediatric burn patient to grown again and improve on their height and weight percentile. Lasting effects of burn injury are likely underestimated, and long-term follow-up and treatment of severely burned patients is therefore mandatory.

Table 3 Process for nutritional assessment and monitoring of patients with burn injury

LBM = lean body mass, EN = enteral nutrition, PN = parenteral nutrition, TEE = total energy expenditure, MREE = measured resting energy expenditure, REE = resting energy expenditure, TBSA = (burn size) total body surface area, CI = caloric intake over previous 24 h, HBE = Harris Benedict Equation, T = average temperature over previous 24 h (°C), PBD = post burn day, DRI = daily recommended intake.

Nutrition Assessment

- Identify patients requiring initiation of nutrition support:
  - mechanically ventilated
  - >20% TBSA burns
  - inhalation injury
  - failure to maintain LBM with oral intake
  - malnourished or high risk for becoming malnourished (elderly, children, obese with sarcopenia)
- Conduct nutrition focused exam (see Table 1)

Initiation of nutrition support

- EN is strongly preferred over PN
- Choose high protein, high carbohydrate, low fat formula
- Consider PN if:
  - non-functioning GI tract
  - intolerant to EN
  - unable to achieve goal EN in a reasonable time
- Initiate trophic EN within 4–6 h of injury for maximal benefit
- Once patient is adequately resuscitated and hemodynamically stable advance to goal feeds as quickly as tolerated

Determining Nutrition Requirements

Nutrient	Requirements	Monitoring
Energy	Indirect calorimetry: TEE = MREE × 1.2−1.4 Predictive Equations Adults: Toronto Equation: REE = −4343 + (10.5 × %TBSA) + (0.23 × CI) + (0.84 × HBE) + (114 × T) − (4.5 × PBD) TEE = REE × 1.2−1.4 Predictive Equations Children: Galveston Infant (0–1 year) 2100 kcal/m^2 + 1000 kcal/m^2 burn Galveston Revised (1–11 years) 1800 kcal/m^2 + 1300 kcal/m^2 burn Galveston Adolescent (12–16 years) 1500 kcal/m^2 + 1500 kcal/m^2 burn	Minimum weekly repeated measurements or reassessment with clinical changes Clinical evidence of underfeeding: • LBM loss • Poor wound healing Clinical evidence of overfeeding: • Hyperglycemia • Fatty liver infiltration • Increased infections Monitor for long term weight changes
Protein	Adults: 1.5–2.0 g/kg/day Children: 3.0 g/kg/day	Nitrogen Balance = 24-h protein intake – [1.25 x (UUN + 4)] Wound closure Evidence of LBM loss/muscle wasting
Carbohydrate	>50–60% of total energy intake Adults: Minimum: 2 g/kg/day Maximum: 5–7 g/kg/day or 5 mg/kg/min	Hyperglycemia following implementation of appropriate insulin management
Fat	3–15% of total energy intake Adults: Minimum: 2–4% of energy Maximum: 25–35% of energy Account for non-dietary fat calories (i.e. Propofol) in nutrition care plan	Evidence of inadequate fat administration: • Essential fatty acid deficiency Evidence of excess fat administration: • Hypertriglyceridemia • Steatosis • Deposition of adipose tissue
Micronutrients	2–3 weeks IV micronutrients for >20% TBSA burns • Selenium 500–700 μg/day • Copper 4 mg/day • Zinc 40 mg/day Minimum micronutrient doses to meet DRI Consider additional vitamin C, D, E and K supplementation	Suggested micronutrient doses are considered safe if routine blood concentration monitoring is not available.

Table 3 Process for nutritional assessment and monitoring of patients with burn injury
LBM = lean body mass, EN = enteral nutrition, PN = parenteral nutrition, TEE = total energy expenditure, MREE = measured resting energy expenditure, REE = resting energy expenditure, TBSA = (burn size) total body surface area, CI = caloric intake over previous 24 h, HBE = Harris Benedict Equation, T = average temperature over previous 24 h (°C), PBD = post burn day, DRI = daily recommended intake.—cont'd

Monitoring of nutrition support
Ongoing monitoring of EN tolerance
- abdominal pain and distention
- vomiting
- constipation or diarrhea

Implementation of strategies to optimize EN delivery
- insertion of post-pyloric feeding tubes
- volume-based feeding protocols
- peri-operative feeding protocols

Cessation of nutrition support
Patient can maintain nutritional status via oral diet. Clinical evidence includes:
- caloric intake equal to predicted energy requirements
- stable weight
- wound closure
- improving functional status

Nutrition requirements

Data on the optimal diet as patients progress from the acute post-burn phase to recovery is virtually non-existent. As the hypermetabolic state can persist for over a year post-injury, a high calorie and protein diet may be required to meet elevated energy requirements and replete muscle mass lost because of prolonged critical illness. Weight maintenance or targeted weight gain can be used a marker of overall nutritional adequacy. Resistance exercise is recommended to improve LBM and prevent burn-wound contracture. Level of physical activity and persistent open wounds will impact nutrition requirements. Improvements in functional status are anticipated in the setting of adequate nutrition and physical therapy. Targeted intervention by a dietitian may be warranted in instances of weight loss, failure to close wounds or functional decline.

Nutrition support after minor burn

Minor burns do not incite the same hypermetabolic response as there is less tissue to regenerate. Energy and protein requirements may rise slightly above baseline, but the inflammatory response remains local so systemic hypercatabolism is not expected. In the absence of preexisting comorbidities or malnutrition, a healthy adult should have adequate nutritional stores to fully recover a few weeks after excision and grafting surgery. Pediatric, elderly and patients with preexisting comorbidities may require nutrition support to facilitate wound closure, even for relatively minor burns.

Initial trauma, pain and change in eating environment due to hospitalization, makes pediatric patients, especially those under three years of age, prone to anorexia. Close monitoring of oral intake and weight is vital in both hospitalized and ambulatory patients. Hospitalization and initiation of EN may be required if a child is unable to maintain their weight. Adequate pain control is a priority in restoring normal appetite.

Elderly patient are at higher risk of developing malnutrition following a minor burn or having delayed healing of superficial burns or donor skin sites due to higher incidence of preexisting undernutrition and decreased LBM. Beginning as early as age 40, humans lose an average of 0.6% of LBM per year, producing a "sarcopenia of aging". Elderly patients also require more exogenous protein than younger patients since they have a blunted anabolic response to dietary protein that requires a greater concentration of serum amino acids to initiate muscle protein synthesis. Adequate nutrition and physical activity are vital to maintaining LBM, functional status and independence during hospitalization and recovery.

Finally, burns occur disproportionately in low socioeconomic status population, even in high-income countries. This population is particularly at risk for nutritional deficiencies. There are several validated tools available to help screen for and diagnose malnutrition. Screening on admission then weekly thereafter can help identify patients that would benefit from nutrition support.

There are no specific nutrition recommendation for the management of minor burns; however, general recommendations to promote wound healing in cases of chronic wounds or during the peri-operative period may be applicable. Transient oral supplementation of Vitamin A, B12, C, D, essential amino-acids, w-3 fatty acids, protein, Zn, and Cu could be proposed but supplementary studies are needed to evaluate the benefit. Adequate oral intake to promote rapid wound healing and maintain functional status should be encouraged in all patients, with nutrition support and micronutrient supplementation prescribed on a patient-by-patient basis.

Conclusion

Nutrition support is a vital component in addressing the hypermetabolic response to burn injury and to support wound healing and maintain essential organ functions. Profound alterations of fat, protein and carbohydrate metabolism and increased energy requirements makes the nutrition management of burn patients challenging. Strong evidence favors early EN with a high carbohydrate and protein and minimal fat formula. Supplementation of specific vitamins, TE and GLN at pharmacological doses are attempted to enhance outcomes, but further studies are required to confirm benefits. While most patients with minor burns will not require nutrition intervention, patients at risk of undernutrition may require nutrition assessment and targeted nutrition to facilitate wound healing.-

Harris Benedict Equation:

$$\text{Men}: \ \text{BMR} = 66.5 + (13.76 \times \text{Weight in Kg}) + (5.003 \times \text{height in cm}) - (6.755 \times \text{age in years})$$

$$\text{Women}: \ \text{BMR} = 655 + (9.563 \times \text{Weight in Kg}) + (1.850 \times \text{height in cm}) - (4.676 \times \text{age in years})$$

References

Berger, M.M., Binnert, C., Chiolero, R.L., et al., 2007. Am. J. Clin. Nutr. 85, 1301–1306.
Blaauw, R., Osland, E., Sriram, K., et al., 2013. J. Parenter. Enteral Nutr. 43 (Suppl. 1), S5–S23.
Carson, J.S., 2018. Needs and nutritional support for the burned patient. In: Herndon DN, Total Burn Care, fifth ed. Elsevier, pp. 287–300.
Clark, A., Imran, J., Madni, T., et al., 2017. Burns Trauma 5, 11.
Jeschke, M.G., van Baar, M.E., Choudhry, M.A., et al., 2020. Nat. Rev. Dis. Prim. 6, 1–25.
McClave, S.A., Taylor, B.E., Martindale, R.G., et al., 2016. J. Parenter. Enteral Nutr. 40, 159–211.
Porter, C., Tompkins, R.G., Finnerty, C.C., et al., 2016. Lancet Lond Engl 388, 1417–1426.
Rodriguez, N.A., Jeschke, M.G., Williams, F.N., et al., 2011. J. Parenter. Enteral Nutr. 35, 704–714.
Rousseau, A.F., Losser, M.R., Ichai, C., Berger, M.M., 2013. Nutr Edinb Scotl 34, 497–502.
Wise, A.K., Hromatka, K.A., Miller, K.R., 2019. Nutr Clin Pract Off Publ Am Soc Parenter Enter Nutr 34, 673–680.

Further Readings

Casaer, M.P., Greet, V. den B., 2014. N. Engl. J. Med. 370, 1227–1236.
Demling, R.H., Seigne, P., 2000. World J. Surg. 24, 673–680.
Gus, E.I., Shahrokhi, S., Jeschke, M.G., 2020. Burns J. Int. Soc. Burn Inj. 46, 19–32.
Herndon, D.N., Tompkins, R.G., 2004. Lancet 363, 1895–1902.
Heyland, D., Muscedere, J., Wischmeyer, P.E., 2013. N. Engl. J. Med. 368, 1489–1497.
Jeschke, M.G., Phelan, H.A., et al., 2020. J. Burn Care Res. Off. Publ. Am. Burn Assoc. 41, 65–83.
Masch, J.L., Bhutiani, N., Bozeman, M.C., 2019. Nutr. Clin. Pract. Off. Publ. Am. Soc. Parenter. Enter. Nutr. 34, 666–671.
Mecott, G.A., Al-Mousawi, A.M., Gauglitz, et al., 2010. Shock Augusta Ga 33, 10.1097.
Shields, B.A., VanFosson, C.A., Pruskowski, K.A., et al., 2019. Nutr. Clin. Pract. Off. Publ. Am. Soc. Parenter. Enter. Nutr. 34, 688–694.
Tancheva, D., Arabadziev, J., Gergov, G., et al., 2005. Ann. Burns Fire Disasters 18, 16–18.

Nutrition and HIV/AIDS

S. Filteau and D. Manno, London School of Hygiene and Tropical Medicine, London, UK

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of S. Filteau, D. Manno, Nutrition and HIV/AIDS, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 303–308, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00143-4.

Introduction 535
HIV/AIDS 535
Effects of HIV/AIDS on Nutritional Status 536
Calories and Macronutrients 536
Micronutrients 537
Gastrointestinal Function 537
Nutritional Intervention Studies 537
Macronutrient Intervention Studies 537
Micronutrient Intervention Studies 537
Interactions Between Antiretroviral Drugs and Nutrition 538
Infant Feeding and HIV/AIDS 539
Social and Economic Interactions with HIV/AIDS 539
Conclusions 540
Further Reading 540

Glossary

AIDS Acquired immunodeficiency syndrome – a severe disease of the human immune system caused by the human immunodeficiency virus (HIV).

Antiretroviral drugs (ARVs) Drugs used to treat HIV.

Antiretroviral therapy (ART) Treatment with combinations of several drugs against HIV.

Body mass index (BMI) Calculated as weight in kg per height in meters squared; often used as an indicator of overall nutritional status.

CD4 cells A subgroup of lymphocytes which are targets of HIV infection. Destruction of CD4 cells is both a cause and a marker of HIV/AIDS disease progression; therefore blood CD4 count is often used to measure the degree of immunodeficiency.

HAART Highly active antiretroviral therapy; treatment for HIV/AIDS combining at least three antiretroviral drugs.

HIV Human immunodeficiency virus – a family of closely related viruses which infect and destroy the human immune system.

Lipodystrophy Altered proportion and distribution of body fat, often as a consequence of antiretroviral therapy and associated with blood lipid abnormalities.

NFκB A gene transcription factor with a central role in inflammatory responses.

Introduction

HIV infection and nutrition are intimately linked at many levels: biological, clinical, social, and economic. Nutritional problems have been a defining characteristic of human immunodeficiency virus (HIV)/acquired immuno deficiency syndrome (AIDS) since the early years of the epidemic. In Africa, AIDS was first recognized as 'slim disease' because sufferers experienced severe weight loss. HIV infection can lead to malnutrition, although poor diet can in turn speed disease progression. In contrast, in high income countries, HIV/AIDS and antiretroviral therapy (ART) for the disease have been associated with abnormal fat metabolism and distribution and with chronic diseases associated with overnutrition. Thus HIV/AIDS is a challenge to nutritionists working at both the individual patient level and the public health level globally (Figure 1).

HIV/AIDS

AIDS is a disease of the human immune system caused by HIV. HIV is transmitted through direct contact of a mucous membrane or the bloodstream with a body fluid, such as blood, genital secretions, or breast milk of someone who is infected with the virus. The transmission most commonly occurs during sexual contacts and from mothers to infants during pregnancy, at delivery or through breastfeeding. HIV can also be transmitted through transfusion of blood products or injection with contaminated needles.

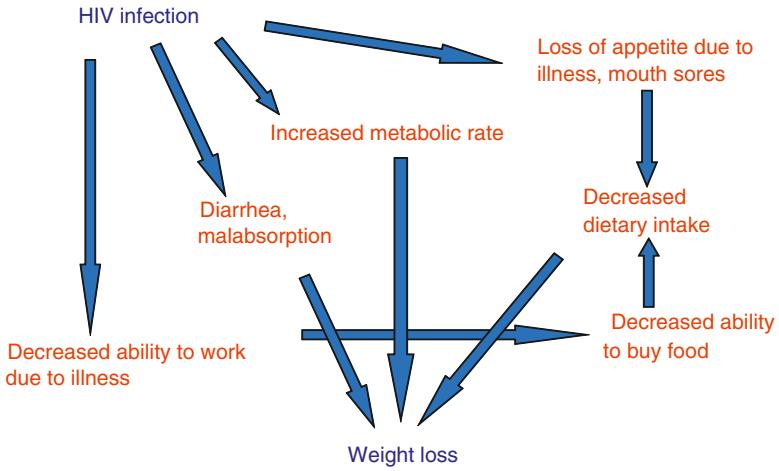


Figure 1 Clinical and socioeconomic pathways whereby HIV/AIDS can cause weight loss.

HIV infects and destroys cells of the immune system, resulting in immunodeficiency. Key cells involved in this process are the CD4 subclass of lymphocytes. The number of CD4 cells in the blood is often used as an indicator of the degree of HIV-induced immunodeficiency, that is, disease progression. A low number of CD4 cells is associated with an increased risk of opportunistic infections which are infections which do not normally cause disease in otherwise healthy people but which do in people who are immunosuppressed. Common opportunistic infections include *Pneumocystis jirovecii* pneumonia, disseminated cryptococcosis, and cerebral toxoplasmosis. In addition, certain infections such as tuberculosis can occur more frequently and cause more serious disease among HIV-infected people than among HIV-uninfected people.

HIV-immunosuppressed individuals are also at higher risk of developing cancers such as Kaposi sarcoma, invasive cervical carcinoma, Burkitt, immunoblastic, or primary central nervous system lymphomas.

There is currently no cure for HIV infection or AIDS. Vaccines have been investigated but are not publicly available at the moment. Current treatment for HIV infection is based on highly active antiretroviral therapy (HAART). This consists in a combination of at least three drugs belonging to at least two types or 'classes' of antiretroviral agents (ARVs). HAART has significantly improved prognosis and quality of life for people living with HIV. However, access to treatment remains a challenge throughout the world, especially, but not only, in resources-poor countries.

Effects of HIV/AIDS on Nutritional Status

Calories and Macronutrients

HIV, like many infections, raises metabolic rate. Among asymptomatic patients at an early stage of HIV infection, resting metabolic rate is increased by about 10%. Acutely ill patients may have higher resting metabolic rates but this is offset by their lower level of activity so that total energy expenditure may not be increased. They may also be anorexic so that acute illness is not the best time to try to improve nutritional status. There is evidence that weight loss during HIV infection is more a result of decreased food intake owing to anorexia than it is to raised metabolic rate owing to infection. When patients receive therapy for HIV or opportunistic infections and begin to feel better, their energy needs to support regain of weight and increased activity can become very high, up to 50% greater than normal healthy requirements. Provision of adequate food to cover the extra energy needs can be difficult for many in low income countries.

Although energy needs change during HIV infection, there is no evidence that protein needs, as a percent of food energy, are changed. Therefore, it is usually best to meet energy needs through increased intake of a balanced diet which will then provide adequate protein.

Tissue lost during illness and that regained during recovery may differ in proportions of fat and lean. There is a tendency for increased central fat deposition and a loss of lean tissue as a result of this weight cycling. This can have important implications for health since loss of lean tissue is associated with decreased ability to work or manage other aspects of daily living as well as with increased risk of mortality. On the other hand, given the role of central fat in regulation of the immune system and in increasing systemic inflammation, it is possible that increased central fat has short-term health benefits for HIV-infected people at high risk of infectious disease, even though it may increase their long-term risk of chronic diseases. This is an area requiring considerable further research.

Micronutrients

Anorexia and increased wastage of some micronutrients during illness, for example, fecal zinc losses during diarrhea, may result in deficient micronutrient status. However, infections change plasma levels of many markers of micronutrient status as part of the normal acute phase response and this does not necessarily imply changed micronutrient status itself. Therefore, it is not always easy to determine the effects of HIV/AIDS on micronutrient status. The best indication of inadequate micronutrient status has come from intervention trials which in some cases have shown benefits of improving micronutrient status in HIV/AIDS (*see* Micronutrient Intervention Studies).

Antioxidant micronutrients may be especially important because infection is an oxidative stress: the immune response to combat infection involves generation of activated oxygen metabolites and increased energy expenditure. High cellular oxidant levels can lead to increased inflammation through regulation of the transcription factor, NFκB. Although inflammation is an essential part of the response to infection, both inflammation and activated oxygen metabolites can cause tissue damage. Therefore inflammation and production of oxygen metabolites must be controlled and limited to the site of the immune response. Adequate status of antioxidant micronutrients – vitamin E, vitamin C, and selenium – is needed for this.

Gastrointestinal Function

Nutrition and gastrointestinal function are obviously intimately linked. HIV infection can disrupt these links. HIV and opportunistic infections can lead to diarrhea, malabsorption, and nutrient wastage with consequent macro- and micro-nutrient deficiencies. Nutritional deficiencies can further impair immune functions and the ability to repair damaged tissues.

Nutritional Intervention Studies

Macronutrient Intervention Studies

Many governments, nongovernmental organizations, religious, and other community groups have responded to the crisis of malnutrition in HIV/AIDS patients by various types of food and nutrition interventions. Many projects are small scale and few have been formally evaluated. Where evaluations have been conducted, the effectiveness of the program for improving nutritional status, health, and survival has generally been disappointing.

A Cochrane systematic review of macronutrient interventions for people with HIV/AIDS found evidence that interventions increased protein and energy intake but no evidence that this had any benefits for weight, body composition, or health. However, trials were few, fairly small, and conducted mainly in high income countries where undernutrition is fairly uncommon.

Since that review there have been a few macronutrient intervention studies among malnourished or food insecure HIV-infected people in low income countries. Provision of macronutrient supplements, as extra food rations or specific composite foods such as fortified blends or lipid-based supplements, can in some cases and in the short term increase body mass index (BMI) or improve compliance with ART. However, there has been little or no documented benefit for health or survival. Reasons for this are currently unclear. It may be that macronutrient deficiencies, even among malnourished HIV-infected people, are not limiting factors for health. It may be that diarrhea and anorexia limit intake and thus effectiveness of the foods provided. It is also possible that the foods provided are not consumed in sufficient amounts, because of sharing or selling, to have detectable benefits for health.

One area where macronutrient intervention trials have been conducted and programmes evaluated is management of severe malnutrition among HIV-infected young children. In HIV-endemic countries, a high proportion of children referred to malnutrition wards are HIV-infected. A recent meta-analysis has shown that mortality rates are higher for HIV-infected than uninfected malnourished children and are above target levels even in facilities with high quality care as evidenced by low mortality rates among HIV-uninfected children. For those HIV-infected children who do survive, it seems that weight gains are similar to gains of uninfected children. It is important to test HIV status of children admitted for severe malnutrition in high HIV prevalence areas in order to permit simultaneous management of HIV infection with ARVs and opportunistic infections (prophylaxis and treatment), and, hopefully, improve survival. It is notable that nutritional care of HIV-infected malnourished children has usually been initiated on malnutrition wards, not in clinics managing HIV. Fortunately some clinics managing HIV services for prevention of mother-to-child HIV transmission are beginning to incorporate into their services interventions to prevent severe malnutrition rather than to worry later about treating it.

Micronutrient Intervention Studies

Micronutrient interventions appear to have somewhat greater impact on health of HIV-infected people than do macronutrient interventions. This may be because, compared with macronutrients, micronutrient deficiencies are more important limiting factors for health of people with HIV or simply because micronutrient supplements are easier to deliver in the face of anorexia and are less likely to be shared or sold. There is evidence that vitamin A supplements have similar benefits for reducing mortality and diarrhea among HIV-infected children as among other children. Multiple micronutrient supplements have decreased mortality of HIV-infected adults. Multiple micronutrients given to HIV-infected pregnant Tanzanian women improved many aspects of maternal and child health. However, these studies were conducted before the widespread availability of ARVs in the communities and it

is unclear whether micronutrient interventions would have added benefits when given with ARVs. Furthermore, there is very limited information regarding the optimal micronutrient supplements for efficacy and safety among different patient groups.

The micronutrient intervention data is from research trials and, although micronutrient supplements are often taken by people with HIV/AIDS in high income countries and micronutrient interventions have been implemented in some parts of Africa, there is little programmatic evidence for the effectiveness of these interventions, mainly because of lack of evaluations. It seems to be generally assumed that providing nutrients to people with nutritional deficiencies must be beneficial but, in fact, the experience with macronutrient interventions for people with HIV/AIDS indicates that this is not necessarily true. Demonstration of program effectiveness is essential for program sustainability.

Interactions Between Antiretroviral Drugs and Nutrition

ARVs to decrease viral load and improve health, although they do not cure the infection, have been available in high income countries for many years and in Africa during the past decade. Treatments are constantly improving but continue to have side effects, such as nausea or taste abnormalities, which can affect dietary intake. Many ARVs need to be taken with food and, as mentioned above (*see* Effects of HIV/AIDS on Nutritional Status), treating the infection can lead to regain of weight which requires additional food. Therefore, increasing availability of ART may actually be increasing rather than decreasing the need for nutritional support of people with HIV/AIDS (**Figure 2**).

There is evidence that provision of ART can improve growth in children. It may not be possible to completely restore adequate growth in HIV-infected children since stunting, that is, poor length growth, is generally irreversible after age about 2 years. There is some danger that additional weight will then be deposited as fat. Children and their carers should be provided information about healthy eating and exercise, similar to that provided for HIV-uninfected children but possibly even more important.

One side effect of some ARVs is altered proportion and distribution of body fat referred to as lipodystrophy. Partly not only in association with cycles of weight loss and gain (*see* Effects of HIV/AIDS on Nutritional Status) but also as a result of some ARVs themselves, there is a tendency toward decreased peripheral fat and increased central fat. This has cosmetic and social implications which in the past may have decreased compliance with ART. There are more important metabolic effects as well such as increased plasma cholesterol and triglycerides and increased insulin resistance. As for HIV-uninfected people, these abnormalities are associated with increased risk of chronic diseases such as cardiovascular disease and diabetes. Newer drugs do not have such strong lipodystrophic effects so changing regimen may ameliorate the problems. Other interventions which could potentially promote deposition of lean rather than fat tissue are provision of androgenic drugs, supplementation with micronutrients, or exercise. Although historically it may have been odd to worry about excess fat deposition in Africa, even poor African countries are now undergoing the nutrition transition as evidenced by rising prevalence of overweight and obesity. It will require sensitive and intelligent nutritional counseling to promote good nutrition and exercise for preventing overweight and chronic diseases among HIV-infected Africans who have been more usually worried about weight loss and malnutrition.

There is some evidence that ARVs are metabolized differently, that is, more slowly, among malnourished people and this may lead to increased drug toxicity. Differences in metabolism and toxicity, although supported by only limited evidence, are not surprising given that drug metabolism and excretion involve enzymes which often have micronutrients as cofactors and may use oxidative metabolism. In Africa with the ART roll-out, it has now become apparent that, unlike in high income countries, starting ART is associated with a high mortality in the first few months of taking the drugs. Mechanisms for the high early mortality are still

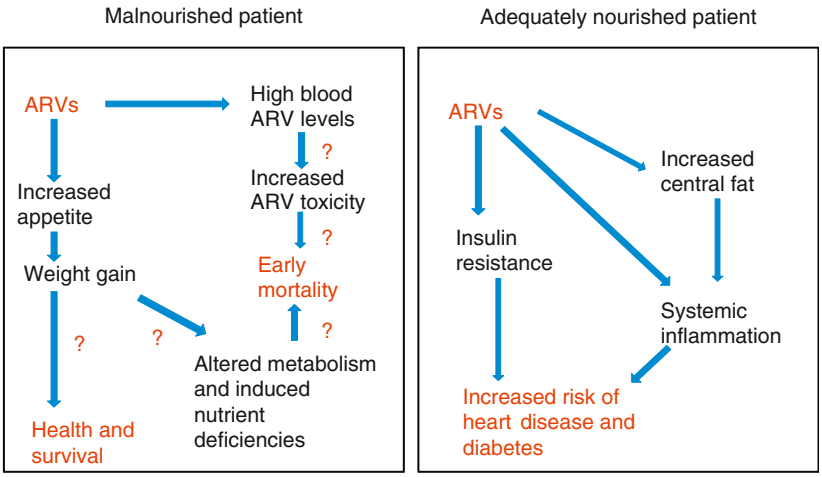


Figure 2 Interactions between antiretroviral drugs (ARVs) and nutrition.

under review but one risk factor which consistently remains significant even in multivariable analyses controlling for CD4 count and co-infections such as tuberculosis, is malnutrition, as represented by low BMI. Low BMI, in addition to being associated with loss of lean tissue, may be a marker for metabolic derangements such as low plasma phosphate. There is concern that when ARVs are first provided and recovery begins, increased metabolic rate could further lower the phosphate problem causing a refeeding syndrome with increased risk of death.

Infant Feeding and HIV/AIDS

One topic in the area of nutrition and HIV which has attracted huge amounts of research, commentary, and concern is HIV and infant feeding. Early in the epidemic it became apparent that HIV can be found in breast milk and that breastfed infants can acquire the infection from their mothers (about 15% of infants of HIV-infected mothers). Mother-to-child transmission can also occur *in utero* (to about 7% of infants of infected mothers) and at delivery (to about 15% of infants of infected mothers) (**Figure 3**). With the advent in Africa around 2000 of single dose nevirapine, an ARV, to mothers and infants at delivery, the proportion of infants infected at delivery decreased so that breastfeeding became the major mode of mother-to-child HIV transmission.

In high income countries HIV-infected mothers have access to ART during pregnancy and at delivery and are advised not to breastfeed. Therefore few infants in these countries become HIV-infected. However, in low income countries, notably in Africa where HIV prevalence is highest, not breastfeeding is associated with a high risk of morbidity and mortality from other infectious diseases. Furthermore, safe and nutritious alternatives to breast milk are not available or affordable for many African women. HIV-infected African women and health services supporting them have been faced with very difficult decisions regarding the best infant feeding mode. Not only do women need to make a feeding decision in the absence of specific information about their own child's risk of infection, but also they need to consider the social implications of not breastfeeding in societies where breastfeeding is almost universal and not breastfeeding advertises a positive HIV status.

World health organization (WHO) has regularly updated its recommendations based on research on HIV and infant feeding. Whereas this was done for the best of reasons – to save children's lives – the rapidly changing recommendations have resulted in some confusion among women and health care workers. Current recommendations are to provide ART to women during late pregnancy and at delivery and to the mother or infant throughout the period of breastfeeding. Although this has been shown to reduce mother-to-child transmission to less than 10% in research studies, this is an expensive recommendation to implement – not primarily in terms of drug costs but more because of costs of health workers' time needed – and so is not yet widely implemented in Africa.

It has recently become apparent that even those children of HIV-infected mothers who escape HIV infection themselves are at increased risk of nutritional and health problems compared to HIV-unexposed children. These HIV-exposed, uninfected children tend to be smaller than HIV-unexposed children and to suffer more episodes of serious morbidity and higher mortality. It is as yet unclear whether nutritional or other interventions can improve these children's growth and health.

Social and Economic Interactions with HIV/AIDS

In HIV-endemic Africa, HIV/AIDS is not only a disease of individuals but also of the whole society. HIV/AIDS strikes mainly young and middle-aged adults who are at their most economically productive ages. These adults are normally also the main carers of young

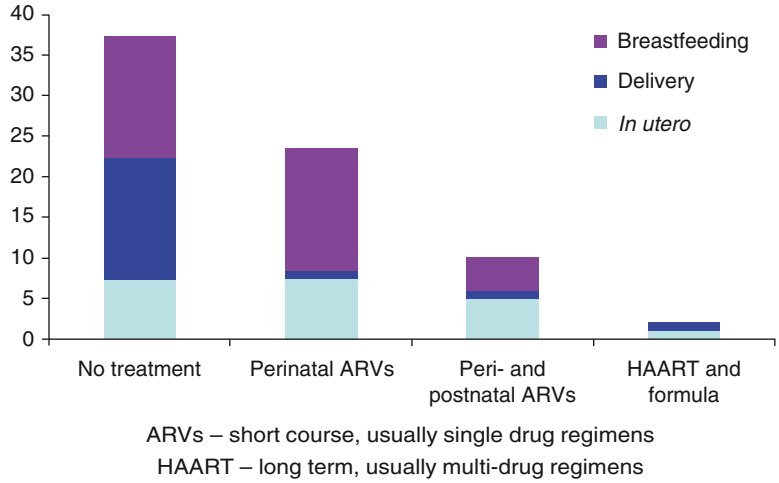


Figure 3 Estimated rate and timing of mother-to-child HIV transmission under different treatment conditions.

children. Loss of caring capacity for children can cause them long term health and emotional problems. Additional financial losses occur because other family members are needed as carers, because of overt medical costs or less obvious costs such as for transport to clinic visits, or because of funeral expenses. Therefore HIV infection of one member of a household can affect the entire household and can be devastating to families, many of whom are near the poverty line already in Africa.

The best approaches for improving nutrition of HIV-infected individuals and HIV-affected families are not clear. Should individuals be targeted on the basis of low BMI or on the basis of food insecurity? Which types of food support and which type of targeting are most cost-effective for outcomes in the patient? Should we also be considering family outcomes such as family members in employment and children in schools? What are the ethics of providing food support to HIV-affected food insecure families in areas where large proportions of the whole population are food-insecure? There is an urgent need for research addressing these questions.

Conclusions

Since early in the HIV/AIDS epidemic it has been clear that nutrition plays an important role. Paradoxically, nutrition may be of even greater importance with the increased availability of ART. The problems are complex and multifactorial and there remain many questions as to how best these can be solved, particularly in Africa where both the HIV epidemic and nutritional problems are most widespread and acute. Evaluation of nutrition intervention programmes in HIV care is essential since nutrition interventions may be expensive but can be justified if they improve patient and family outcomes.

Further Reading

Academy of Science of South Africa (2007) HIV/AIDS, TB and Nutrition. Pretoria: Academy of science of South Africa. http://www.assaf.co.za/wp-content/uploads/reports/evidence_based/3060%20ASSAf%20HIV%20TB%20and%20Nutrition.pdf

Fawzi, WW, Msamanga, GI, Spiegelman, D, et al., 1998. A randomized trial of multivitamin supplements and HIV disease progression and mortality. New England Journal of Medicine 351, 23–32.

Fergusson, P, Tomkins, A, 2009. HIV prevalence and mortality among children undergoing treatment for severe acute malnutrition in sub-Saharan Africa: A systematic review and meta-analysis. Transactions of the Royal Society of Tropical Medicine and Hygiene 103, 541–548.

Filteau, S, 2009. The HIV-exposed, uninfected African child. Tropical Medicine and International Health 14, 276–287.

Grinspoon, S, Carr, A, 2005. Cardiovascular risk and body-fat abnormalities in HIV-infected adults. New England Journal of Medicine 352, 48–62.

Koethe, JR, Chi, BH, Megazzini, KM, Heimburger, DC, Stringer, JS, 2009. Macronutrient supplementation for malnourished HIV-infected adults: A review of the evidence in resource-adequate and resource-constrained settings. Clinical Infectious Diseases 49, 787–798.

Lawn, SD, Harries, AD, Anglaret, X, Myer, L, Wood, R, 2008. Early mortality among adults accessing antiretroviral treatment programmes in sub-Saharan Africa. AIDS 22, 1897–1908.

Mahlungulu, SSN, Grobler, L, Visser, MME, Volmink, J, 2007. Nutritional interventions for reducing morbidity and mortality in people with HIV. Cochrane Database of Systematic Reviews Issue 3. https://doi.org/10.1002/14651858.CD004536.pub2. Art. No.: CD004536.

Ndekha, MJ, van Oosterhout, JJ, Zijlstra, EE, Manary, M, Saloojee, H, Manary, MJ, 2009. Supplementary feeding with either ready-to-use fortified spread or corn-soy blend in wasted adults starting antiretroviral therapy in Malawi: Randomised, investigator blinded, controlled trial. British Medical Journal 338, 1867.

Sztam, KA, Fawzi, WW, Duggan, C., 2010. Macronutrient supplementation and food prices in HIV treatment. Journal of Nutrition 140, 213S–223S.

Serwadda, D, Mugerwa, RD, Sewankambo, NK, et al., 1985. Slim disease: A new disease in Uganda and its association with HTLV-III infection. Lancet 2, 849–852.

World Health Organization HIV/AIDS http://www.who.int/topics/hiv_aids/en/

World Health Organization (2005) Consultation on nutrition and HIV/AIDS in Africa. http://www.who.int/nutrition/topics/consultation_nutrition_and_hivaids/en/index.html

World Health Organization (2009) New WHO recommendations on infant feeding in the context of HIV. ww.who.int/child_adolescent_health/news/archive/2009/30_11_09/en/index.htm

Nutrition and parasitism

Olga P. García[a], Gerardo A. Zavala[b], Maiza Campos-Ponce[c], and Katja Polman[d], [a] School of Natural Sciences, Universidad Autónoma de Querétaro, Querétaro, Mexico; [b] Department of Health Sciences, University of York, Heslington, United Kingdom; [c] Faculty of Earth and Life Sciences, VU Amsterdam University, Amsterdam, the Netherlands; and [d] Department of Biomedical Sciences, Institute of Tropical Medicine Antwerp, Antwerp, Belgium

© 2023 Elsevier Ltd. All rights reserved.

This is an update of P.G. Lunn, Parasitism, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 6–13, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00223-3.

Introduction **541**
Parasitology **542**
Ascaris lumbricoides 543
Hookworms 543
Schistosomes 543
Trichuris trichiura 543
Strongyloides stercoralis 543
Giardia lamblia 543
Cryptosporidium spp 543
Entamoeba histolytica 544
Parasites and nutrition **544**
Gut microbiota, inflammation and gut permeability 545
Loss of appetite and anorexia 545
Maldigestion and malabsorption 546
Nutrient losses 546
Treatment 546
Prevention **547**
Conclusion **547**
References **547**

Key points

- Despite technological and scientific advances in the last years, intestinal parasitic infections remain a leading cause of morbidity and mortality worldwide, affecting mostly under privileged populations, living in areas with tropical weather and limited access to health services and sanitation.
- Both soil-transmitted helminths and protozoa are the main causes of intestinal parasitic infections.
- Diarrhea caused by intestinal parasitic infections affects millions of people and is one of the major causes of mortality in children under 5 years of age.
- Intestinal parasites affect the nutritional status of the host, and malnutrition is the primary cause of immunodeficiency, which in turn, increases the risk of parasitic infection.
- Infection with intestinal parasites cause dysbiosis, local and systemic inflammation and gut permeability.
- The main nutritional consequences of parasitic infections include malabsorption, nutrient loss, anorexia, which in turn cause stunting, micronutrient deficiencies and anemia in the host.
- Treatment for parasitic infection have been available for years. Prevention needs to be a priority particularly in endemic populations.

Introduction

Intestinal parasitic infections (IPI) and malnutrition are very common in children of low and middle income countries (Zavala Gomez, 2017). IPI, obesity and micronutrient deficiencies account for more than 10% of the global disability-adjusted-life-years (DALYs) (Ng et al., 2014; Black, 2014).

Intestinal parasites are a broad number of organisms that require a host for their own survival and that live in the gastrointestinal (GI) tract. IPI continue to be public health concerns worldwide, affecting mainly low-income countries. They occur in regions that have poor living conditions, poor sanitation, limited access to safe drinking water, and with tropical and subtropical weather (Harhay et al., 2010). IPI are considered endemic in some regions of the world such as Africa, Central and South America, Eastern Europe

and South East Asia, and they can also be found, to a lesser extent, in some high income countries (World Health Organization, 2013, 2018; Brooker et al., 2006; Kantzanou et al., 2021). Children from rural communities are at highest risk of infection, due to the lack of health, sanitation and water supply services and due to behavioral risks, such as frequent outdoor exposure and poor personal hygiene (Strunz et al., 2014).

The IPI are mainly caused by soil transmitted helminths (STHs) and by intestinal protozoan, and infections by both are highly prevalent worldwide (**Table 1**). The majority of the STHs infections are caused by *Ascaris lumbricodies* (roundworm), *Necator americanus* and *Ancylostoma* spp (hookworm), and *Trichuris Trichuria* (whipworm) (Brooker et al., 2006; World Health Organization, 2013). It is estimated that approximately 1.5 billion people worldwide are infected with STHs, and according to the World Health Organization (WHO), more than 260 million preschool-age children and over 560 million school-age children live in areas where infections are prevalent (World Health Organization, 2013). Amoebiasis, caused by *Entomoeba hystolica*, and giardiasis, caused by *Giardia* are among the most common protozoan infections. In developing countries, the prevalence of reported amoebiasis ranges from 10 to 40%, and in developed countries is mainly observed in travelers or immigrants from countries with high prevalence of infection (Shirley et al., 2018). *Giardia duodenalis* is considered to be the most prevalent intestinal parasite present in humans, ranging from 20 to 30% in developing countries, and 2–7% in developed countries (Dixon, 2021). Both, amoebiasis and giardiasis are among the top causes of severe diarrhea across ages worldwide, and are considered one of the leading causes of diarrhea in children from developing countries (Lanata et al., 2013; GBD 2013 Mortality and Causes of Death Collaborators, 2015).

Despite the efforts and improvements in health programs and initiatives, IPI are still one of the major causes of global morbidity and mortality, and rate reductions have been slow in many developing countries. Mortality associated to diarrhea due to IPI account for approximately 800,000 reported deaths of children under 5 years of age and approximately 1.3 million deaths in the general population (GBD 2013 Mortality and Causes of Death Collaborators, 2015). The IPI change the immune response of the host both systemically and locally, thus, impairing its function, and leading to complications both acute and chronic (Brosschot and Reynolds, 2018). Overall, IPI cause inflammation, dysbiosis, diarrhea, malabsorption, acute and chronic blood loss, reduced work capacity in adults, and stunting and impaired cognitive development in children (Koruk et al., 2010; Jourdan et al., 2018; Pullan et al., 2014). IPI may also affect the nutritional status of the host leading to protein-energy malnutrition and micronutrient deficiencies, and recently, it has been observed that it may also be associated with obesity (Jourdan et al., 2018; Zavala et al., 2016). The parasitology, the effect on the nutritional status of the host and its mechanisms as well as IPI treatment and prevention will be discussed in the present article.

Parasitology

Intestinal protozoa have a direct life cycle where transmission typically occurs via the fecal-oral route (Price, 2017). They can be classified according to their effect on human health as "pathogenic" or "non-pathogenic". Pathogenic protozoa such as *Giardia lamblia*, *Entamoeba histolytica* and *Cryptosporidium* can cause acute diarrhea and dysentery, malabsorption, blood loss and reduced growth (Zavala Gomez, 2017). "Non-pathogenic" protozoa such as *Entamoeba coli* (*E.coli*) and *Endolimax nana*, are usually not associated with illness or clinical symptoms, though recently it has been observed that *E. coli* may be associated with obesity in children (Zavala et al., 2016).

STHs refer to the intestinal worms that are transmitted through contaminated soil. Infection occurs by ingestion of eggs (e.g., *A. lumbricoides* and *T. trichiura*) or by penetration of the skin (e.g., by hookworm larvae). The Global Burden of Disease Study 2010 estimated that STHs rank first among all neglected tropical diseases in DALYs, with 5.1 years lived with disability (Pullan et al., 2014). These DALYs were estimated using multiple studies associating STHs with moderate and severe anemia, micronutrient deficiencies and reduced growth (Pullan et al., 2014).

Table 1 Estimated population infected with intestinal parasites worldwide.

Parasite	*Approximate number of infected population (millions)*
Soil transmitted helminths	
Ascaris lumbricoides	800–1100
Trichuris trichuria	600–800
Necator americanus and *Ancycostoma duodenale*	600–700
Protozoa	
Giardia	200
Entamoeba histolytica/dispar	500

Source: CDC (2021), WHO (2021).

Ascaris lumbricoides

Ascariasis is the most common human helminthic infection globally. *Ascaris lumbricoides* reside in the lumen of the small intestine and are very large: adult females are 20–35 cm, and adult males are 15–30 cm (Dold and Holland, 2011). Infection occurs when infective eggs are swallowed, then reach the small intestine, where the eggs hatch and larvae invade the intestinal mucosa. The larvae then reach the lungs via systemic circulation, where they mature, are coughed and then swallowed. When they reach the small intestine, they develop into adult worms that produce eggs that are excreted in the feces. The fertilized eggs develop infective larvae after 18 days or more. Adult worms can live up to 2 years and usually do not cause acute symptoms. However, heavy infections may cause intestinal obstruction and potentially perforation (Centers for Disease Control and Prevention, 2018).

Hookworms

The two main hookworms that cause IPI are *Ancylostoma duodenale* and *Necator americanus*, though recently, it has been suggested that *A. ceylanicum* could be an emerging parasite that may also infect humans (Centers for Disease Control and Prevention, 2018). Hookworms reside in the lumen of the small intestine, typically the distal jejunum where they can live up to 5 years. While in the small intestine, the female worms can produce up to 10,000 *(N. americanus)* or 25,000 *(A. duodenale)* eggs per day which are passed out in the feces. Eggs hatch within 48 h and the free larvae can survive for 3–4 weeks in favorable conditions. The larvae penetrate the skin of the human host, and travel to the heart, from there to the lungs, where they are coughed and then swallowed. Once in the small intestine, the larvae become adults and they attach to the intestinal epithelial producing blood loss. Intestinal hookworm infections can also be asymptomatic (Centers for Disease Control and Prevention, 2018).

Schistosomes

Various species of *Schistosoma* cause Schistosomiasis, and they require specific snail intermediate hosts as part of their life cycle. In the snail, infective cercariae is produced and released, and then penetrate the human through the skin. In the human host, the cercariae becomes schistosomulae and migrates to the lung, then the heart and finally to the liver, where it develops (Basch, 1991). The mature worms, both male and female, leave the liver, copulate and reside in different locations (depending on the species), such as the mesenteric veins in the small or large intestine. The eggs are eliminated via feces or urine (Centers for Disease Control and Prevention, 2018).

Trichuris trichiura

The adult worms of *Trichuris trichiura* live in the colon for approximately 1 year. The infective stage of the *T. trichiura* begins with the intake of embryonated eggs present in soil-contaminated hands or food. When the eggs reach the small intestine, they hatch and the larvae transit to the colon where they mature into adult worms. The adult worms stay in the colon attached to the intestinal mucosa, where the female worms shed their eggs, which are later excrete in feces. Once in the soil, the excreted unembroyenated eggs become embryonated after several stages becoming infective (Centers for Disease Control and Prevention, 2018; Harhay et al., 2010).

Strongyloides stercoralis

Even though there are several species of *Strongyloides*, the major contributor of infection in humans is *Strongyloides stercoralis*. *S. stercolaris* has two life-cycles: the free-living cycle and the parasitic cycle. Also, autoinfection may occur. In the parasitic cycle, the filariform larvae from contaminated soil penetrates the skin of a human host at contact. Then, the larvae migrate to the small intestine, where they become adult female worms. The females produce the rhabditiform larvae in the submucosa of the small intestine, then transit through the colon and, are either excreted in the stool, or cause autoinfection. Autoinfection, when left untreated, may be responsible for persistent infection (Centers for Disease Control and Prevention, 2018; Jourdan et al., 2018).

Giardia lamblia

Infection with *G. lamblia* can occur when a person swallows its cysts from contaminated water, food, hands, surfaces, or objects, or by fecal-oral route. Cysts are immediately infectious when passed in the stool or shortly afterward, and can survive several months in cold water or soil. After swallowing, *Giardia* cysts get to the small intestine where each cyst releases two trophozoites (excystation), that feed from the nutrients of the human host. In the lumen of the small intestine, the trophozoites multiply and, either attach to the mucosa by the sucking disk of the ventral surface, or stay free. The cysts are found in the stool of infected humans after encystation when transiting to the colon (Centers for Disease Control and Prevention, 2018; Dixon, 2021).

Cryptosporidium spp

Cryptosporidium parvum and *C. hominis* are the leading causes of human cryptosporidiosis. Its transmission is mainly through contaminated water or food with feces, or by direct contact with infected animals or humans. After ingestion, the sporozoites

are released and get to the GI tract where they undergo multiplication (asexual and sexual), then oocysts develop and sporulate. The oocyts can be either excreted in feces (thick-walled oocyts) or produce autoinfection (thin-walled oocyts). The excreted oocysts are infectious. Infection with *Cryptosporidium* spp. and genotypes results in a wide range of signs and symptoms (Centers for Disease Control and Prevention, 2018; Fayer et al., 2000).

Entamoeba histolytica

Entamoeba hystolitica is as a pathogenic amoeba responsible for the infections of millions of people in the world due to fecal contaminated hands, food or water. Following ingestion, trophozoites are released in the small intestine after excystation, and then travel to the large intestine. Once there, three kind of infections may happen. The non-invasive infection is when the trophozoites remain in the large intestine, and the host is a carrier, passing cysts in the stool. While cysts can be infectious outside the human body, trophozoites cannot. The majority of infections restricted to the lumen of the intestine are asymptomatic. The intestinal infection is when the trophozoites invade the intestinal mucosa causing abdominal pain and diarrhea. Lastly, the extra-intestinal infection happens when the trophozoites reach blood vessels and travel to the liver, the brain and the lungs (Centers for Disease Control and Prevention, 2018; Ravdin and Petri, 1990).

Parasites and nutrition

IPI and malnutrition are bidirectionally linked. As seen in **Fig. 1**, intestinal parasites can cause malnutrition and malnutrition is the primary cause of immunodeficiency, which in turn, increases the risk of infection (Katona and Katona-Apte, 2008).

The interactions between host nutrition and GI parasites follow mechanisms that include alterations of gut microbiota, inflammation, and gut permeability. The most common nutritional adverse effects of intestinal parasitic infection are stunting, anorexia, malabsorption of fat and carbohydrates, reduced lactase activity, decreased transit time, reduced nitrogen retention, anemia and micronutrient deficiencies (**Table 2**). It is important to clarify that in most cases parasitic infections remain asymptomatic. The manifestation of adverse effects and the severity of these effects depends on the host immune and inflammatory response, initial nutritional status of the individual or the population, genetics, comorbidities and most importantly the intensity of the parasite infection (Quihui-Cota et al., 2004; Sackey et al., 2003; Zavala et al. 2018).

The type of nutritional adverse effects is dependent on the morphology, feeding and reproductive mechanism of each intestinal parasite. For example, parasites which feed on blood are associated with iron deficiency and anemia, while larger parasites (i.e., *Ascaris lumbricoides* and tapeworms) tend to be associated with abdominal pain, malabsorption and stunting.

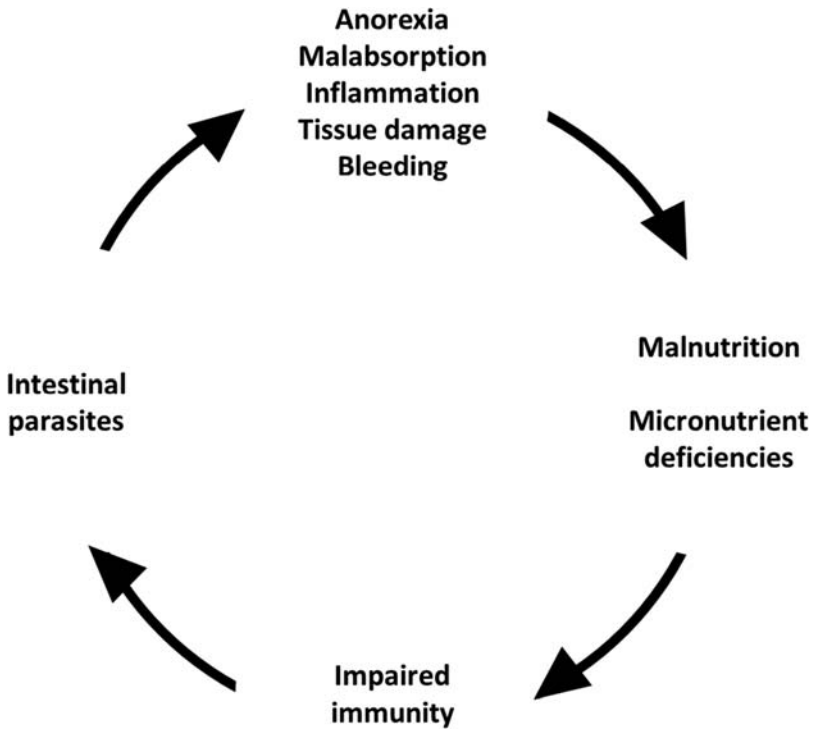


Fig. 1 Intestinal parasites and malnutrition vicious cycle. Original figure with information from Holmes and Coop (2011), Katona and Katona-Apte (2008).

Table 2 Symptoms and nutritional effects of intestinal parasites.

	Ascaris lumbricoides	*Hookworm*	*Schistosoma spp.*	Trichuris trichiura	*Strongyloides spp.*	Giardia intestinalis	*Cryptosporidium spp.*	Entamoeba histolytica
Symptoms								
Anorexia	X	X	X	X	X	X	X	
Abdominal pain	X	X	X	X	X	X	X	X
Diarrhea		X	X	X		X	X	X
Vomiting				X	X	X	X	
Acute-phase response	X		X	X	X	X	X	X
Mucosal disruption						X	X	X
Blood loss		X	X	X				X
Dysentery								X
Nutritional effects								
Lower food intake (macro and micronutrients)	X	X	X	X	X	X		X
Growth retardation	X	X	X	X	X	X	X	X
Weight loss	X	X	X	X	X	X	X	X
Malabsorption syndrome	X				X	X		
Lower vitamin A concentration	X							
Lower iron concentration and anemia		X	X	X				
Lower zinc concentrations				X		X		X
Fluid and electrolyte loss								X

The effects that IPI have on the microbiota and immune response are discussed as follows, including the main nutritional adverse effects of IPI, which are anorexia, malabsorption and nutrient losses.

Gut microbiota, inflammation and gut permeability

GI parasites have been shown to increase gut permeability and intestinal inflammation (Allain and Buret, 2020). One of the mechanisms is through the effect they have on the composition of the microbiota. The gut microbiota is defined as the array of living microorganisms (bacteria, viruses and fungi) in the human gut that are important for maintaining and developing of the intestinal barrier, and also for stimulating immune homeostasis and immune response (Berg et al., 2020). Infections with intestinal parasites have been shown to cause microbiota dysbiosis which results in gut permeability and translocation of bacterial metabolites that lead to low-grade systemic inflammation (Oriá et al., 2016). For instance, STH infection alters the composition of the microbiota directly by the secretion of products with antimicrobial activity such as glycoproteins, immunomodulatory proteins, and miRNAs, and also by altering the intestinal mucus of the host (Brosschot and Reynolds, 2018). In addition, infection with STH is known to promote a type 2 immune response both systemic and locally in the intestine, characterized by the production of T-helper type 2 cells and anti-inflammatory cytokines, such as interleukin 10 (IL-10) (Zavala et al., 2018; Brosschot and Reynolds, 2018). Though an anti-inflammatory immune response is necessary to eradicate the parasite, the response is not efficient enough due to the overall effect STH have on the immune response, both direct and indirect (Brosschot and Reynolds, 2018).

The inflammation caused by IPI can affect the intestinal barrier, where the innate immune response contributes to the lesions caused in the epithelial cell (Garzón et al., 2017). Also, IPI can affect the intestinal barrier by causing cell damage through binding directly to cell surfaces or disrupting the epithelial cells (Garzón et al., 2017). A disturbed intestinal barrier limits its ability to repair itself, and may lead to gut permeability. The inflammation associated to increased gut permeability as a consequence of parasitic infections may compromise growth and cognitive development in children and increase the risk of chronic diseases in the adult (Garzón et al., 2017; Goto et al., 2009). The severity of the inflammation, both systemic and intestinal, the dysbiosis and the gut permeability caused by the IPI, will depend on the nutritional and immune status of the host.

Loss of appetite and anorexia

Appetite loss and anorexia are common in people infected with intestinal parasites, which has been explained by the symptoms associated with infection, changes in the gut microbiome and inflammatory molecules released to face the parasitic infection (Holmes and Coop, 2011). In many cases parasitic infection remains asymptomatic, but in other cases (heavy infections) parasitic infection (helminths and protozoa) can cause nausea, abdominal pain, flatulence, and distension and discomfort, while protozoan intestinal parasites can also cause vomiting, diarrhea, or dysentery. For instance, hookworms and *Strongyloides stercoralis* stimulate abdominal pain, nausea, and anorexia. As mentioned before, intestinal parasites can also disrupt the gut microbiome composition,

and start inflammatory responses involving molecules as leptin and adiponectin, which are well known appetite regulators (Fata et al., 2019).

Maldigestion and malabsorption

Intestinal parasites can physically and metabolically disrupt macro- and micronutrient absorption. Heavy infections can cause damage and obstruction of the mucosa of the intestine, or cause flattening or thickening of villi, which damages the absorptive surface area and reduces absorption of macro- and micro nutrients. In addition, parasites such as hookworm and *Entamoeba histolytica* can cause damage to the cells, diminishing their absorptive properties and limiting active transport processes (Mohammadi et al., 2015).

Maldigestion and malabsorption main symptom is diarrhea, which is the leading cause of childhood malnutrition and second cause of mortality among children under 5 years of age (Parashar et al., 2003). For example, chronic strongyloidiasis is generally asymptomatic, but a variety of GI manifestations may occur, including diarrhea. In the case of *Entamoeba histolytica*, when the mucosa is invaded, severe dysentery and associated complications can occur; severe chronic infections may lead to further complications such as peritonitis, perforations, and the formation of amebic granulomas. Infected persons with *Giardia* can also experience severe diarrhea and malabsorption. Acute giardiasis develops after an incubation period of 1–14 days (average of 7 days) and usually lasts 1–3 weeks. Symptoms include diarrhea, abdominal pain, bloating, nausea, and vomiting. In chronic giardiasis the symptoms are recurrent and malabsorption and debilitation may occur. Immunocompetent patients infected with *Cryptosporidium* spp. may present diarrheal illness that is self-limiting, typically resolving itself within 2–3 weeks. In immunocompromised patients' infection can lead to more severe complications, such as life-threatening malabsorption and wasting. Diarrheal illness in all cases may be accompanied by fever or fatigue.

Nutrient losses

Direct and indirect loss of nutrients is the most important mechanism by which intestinal parasitic infections compromise the nutritional status of the host (Hesham et al., 2004). Direct losses occur during the feeding and burrowing of blood-sucking and tissue invading parasites, such hookworm, *Schistosoma* spp. and *Trichuris trichiura*, or due to vomiting and diarrhea. Blood and tissue ingested by the parasites contributes to one part of the nutrient loss. Hookworms, for instance, cause iron deficiency anemia due to blood loss at the site of intestinal attachment of adult worms, especially in heavy infections. However, the highest contribution to nutrient loss is due to lesions caused by feeding and burrowing activity; these lesions usually continue bleeding after the parasites have moved to another feeding/burrowing place (Crompton, 2000).

The passage of schistosome eggs though the intestine often causes tissue damage and blood loss, and thus, has been associated with iron-deficiency anemia (Friedman et al., 2005). Another mechanism is the accelerated disposal of erythrocytes damaged by the intestinal parasites to the lumen (Kapczuk et al., 2020).

Similarly to hookworms, loss of blood and anemia are the most common nutritional effects of *Trichuris trichiura* infection. *Trichuris trichiura* feeds from tissue secretions rather than blood, but there is blood loss due to the parasite-induced lesions (Stephenson et al., 2000). Anorexia, vomiting and diarrhea are associated with light and moderate infections, while more severe symptoms such as intestinal inflammation, abdominal pain and rectal prolapse are associated with heavy infections (Stephenson et al., 2000).

Indirect losses are caused by inflammatory and immunological mechanisms which are stimulated to fight the parasitic infection and repair tissue damage (Rivera et al., 2003). Local inflammation at the site of the parasite often accompanied by lymphocytic infiltration of tissues cause further damage to the mucosa, augmenting maldigestion, malabsorption, and nutrient losses as more damaged cells are shed. Activation of the systemic inflammatory system, that is, the acute-phase response, is a general reaction of the body to pathogen invasion or tissue damage. It results in a widespread cytokine-mediated catabolic response, which increases the catabolism of macro and micronutrient reserves (Grimble, 1990). This process is also associated with anorexia, which decreases nutrient intake, increases losses of amino acids, minerals, and vitamins in the urine and feces, and finally leads to protein-energy malnutrition, micronutrient deficiencies, stunting and impaired cognitive development, especially among children.

Another indirect way of losing nutrients is through competition of nutritional sources between the host and the GI parasites. For instance, there is evidence concerning competition for nutrients from tapeworm (Alroy and Gilman, 2020). *Diphyllobothrium latum* concentrates large amounts of vitamin B12 in its tissue. Thus, B12 deficiency and megaloblastic anemia has been reported in children and adults infected with tapeworm (Jimenez et al., 2012). Also, it has been suggested that zinc competition in the gut during *Giardia* infection is one of the causes of zinc deficiency (Astiazarán-García et al., 2015).

Treatment

In **Table 3**, the recommended drugs against intestinal parasites are shown. For the treatment of the STHs, the WHO recommends albendazole (400 mg) and mebendazole (500 mg). These are effective, inexpensive and easy to administer by non-medical personnel. In addition, they have been tested extensively and have been used in millions of people with few and minor side-effects (Anon, n.d.).

Table 3 Recommended drug according to parasitic infection.

Treatment	Ascaris lumbricoides	*Hookworm*	Schistosoma *spp.*	Trichuris trichiura	Strongyloides *spp.*	Giardia intestinalis	Cryptosporidium *spp.*	Entamoeba histolytica
Mebendazole	X	X		X				
Albendazole	X	X		X		x		
Praziquantel			X					
Tricabendazole			x					
Thiobendazole					x			
Ivermectine					x			
Metronidazole						x		X
Tinidazole						x		X
Secnidazole						x		
Nitazoxanide							x	
Spiramycin							x	
Clindamycin							x	
Nitazoxanide								
Paromomycin							x	

Even though the pathology of schistosomiasis is caused mainly by the eggs trapped in tissue, treatment with praziquantel has been shown to improve periportal fibrosis and splenomegaly, as well as physical fitness, appetite and school performance. Also, iron supplementation can improve iron status after anti-helminth treatment (Gryseels et al., 2006).

For giardiasis and amoebiasis, treatment has been available for several decades. There are no effective remedies for infection with *Cryptosporidium* spp. In chronically infected HIV-positive patients, treatment with multidrug regimens usually results in rapid resolution of the diarrhea and, in many instances, eradication of the parasite (Gilles and Hoffman, 2002; Rossignol, 2010).

Prevention

In 2012, the WHO published a roadmap to overcome the global impact of neglected tropical diseases. For intestinal parasites, WHO and partners have targeted to treat 50% of pre-school and school-aged children by 2015, and reach 75% coverage by 2020, which was not achieved (World Health Organization, 2010).

Mass drug administration (MDA) is currently being used to control STH infections. However, there are several obstacles for these programs to be successful, including the lack of access to the anthelminthic drugs, inadequate funding to expand existing MDA programs or to implement new ones in areas that do not currently have them, deficient geospatial and population prevalence data to guide focused targeting of MDA, potential development of drug resistance to existing anthelminthic, among others. In addition, for these programs to be sustainable, MDA programs require reinforcement with improved water quality, sanitation, and hygiene (WASH). Water and sanitation are critical to preventing reinfections or new infections of all intestinal parasites.

Conclusion

Despite the advances in science and technology in the past years, IPI are still a major cause of mortality and morbidity worldwide, affecting mainly children living in developing countries. The effects the IPI have on the nutritional status are varied and may last for years, affecting the development and work capacity of the hosts. Even though some progress has been made, progress is slow, and there is still a need for effective programs and treatments that are accessible to the population, particularly targeting children.

References

Allain, T., Buret, A.G., 2020. Pathogenesis and post-infectious complications in giardiasis. Adv. Parasitol. 107, 173–199.
Alroy, K.A., Gilman, R.H., 2020. Tapeworm Infections. Hunter's Tropical Medicine and Emerging Infectious Diseases, pp. 932–940. https://doi.org/10.1016/b978-0-323-55512-8.00130-7. Available at.
Anon, Soil-Transmitted Helminth Infections. Available at http://www.who.int/news-room/fact-sheets/detail/soil-transmitted-helminth-infections (Accessed January 10, 2022).
Astiazarán-García, H., et al., 2015. Crosstalk between zinc status and Giardia infection: a new approach. Nutrients 7 (6), 4438–4452.
Basch, P.F., 1991. Schistosomes: Development, Reproduction, and Host Relations.
Berg, G., et al., 2020. Microbiome definition re-visited: old concepts and new challenges. Microbiome 8 (1). https://doi.org/10.1186/s40168-020-00875-0. Available at.
Black, R.E., 2014. Global distribution and disease burden related to micronutrient deficiencies. Nestle Nutr. Inst. Workshop Ser. 78, 21–28.
Brooker, S., Clements, A.C.A., Bundy, D.A.P., 2006. Global epidemiology, ecology and control of soil-transmitted helminth infections. Adv. Parasitol. 62, 221–261.
Brosschot, T.P., Reynolds, L.A., 2018. The impact of a helminth-modified microbiome on host immunity. Mucosal Immunol. 11 (4), 1039–1046.

Centers for Disease Control and Prevention, 2018. Centers for Disease Control and Prevention (CDC) Webpage. CDC. Available at. https://www.cdc.gov/parasites/index.html.
Crompton, D.W., 2000. The public health importance of hookworm disease. Parasitology 121 (Suppl. I), S39–S50.
Dixon, B.R., 2021. *Giardia duodenalis* in humans and animals—transmission and disease. Res. Vet. Sci. 135, 283–289.
Dold, C., Holland, C.V., 2011. Ascaris and ascariasis. Microb. Infect. 13 (7), 632–637.
Fata, C., Naeem, F., Barthel, E.R., 2019. Small bowel obstruction secondary to *Ascaris lumbricoides* in the setting of prior exploratory laparotomy. J. Pediatr. Surg. Case Rep. 47, 101254.
Fayer, R., Morgan, U., Upton, S.J., 2000. Epidemiology of *Cryptosporidium*: transmission, detection and identification. Int. J. Parasitol. 30 (12–13), 1305–1322.
Friedman, J.F., Kanzaria, H.K., McGarvey, S.T., 2005. Human schistosomiasis and anemia: the relationship and potential mechanisms. Trends Parasitol. 21 (8), 386–392.
Garzón, M., et al., 2017. Association of enteric parasitic infections with intestinal inflammation and permeability in asymptomatic infants of São Tomé Island. Pathog. Glob. Health 111 (3), 116–127.
GBD 2013 Mortality and Causes of Death Collaborators, 2015. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 385 (9963), 117–171.
Gilles, H.M., Hoffman, P.S., 2002. Treatment of intestinal parasitic infections: a review of nitazoxanide. Trends Parasitol. 18 (3), 95–97.
Goto, R., Mascie-Taylor, C.G.N., Lunn, P.G., 2009. Impact of intestinal permeability, inflammation status and parasitic infections on infant growth faltering in rural Bangladesh. Br. J. Nutr. 101 (10), 1509–1516.
Grimble, R.F., 1990. Nutrition and cytokine action. Nutr. Res. Rev. 3 (1), 193–210.
Gryseels, B., et al., 2006. Human schistosomiasis. Lancet 368 (9541), 1106–1118.
Harhay, M.O., Horton, J., Olliaro, P.L., 2010. Epidemiology and control of human gastrointestinal parasites in children. Expert Rev. Anti Infect. Ther. 8 (2), 219–234.
Hesham, M.S., Edariah, A.B., Norhayati, M., 2004. Intestinal parasitic infections and micronutrient deficiency: a review. Med. J. Malaysia 59 (2), 284–293.
Holmes, P.H., Coop, R.L., 2011. Parasite and nutrition interaction: parasite infection effects. In: Encyclopedia of Animal Science, second ed., pp. 855–857. https://doi.org/10.1081/e-eas2-120041150 Available at.
Jimenez, J.A., et al., 2012. *Diphyllobothrium pacificum* infection is seldom associated with megaloblastic anemia. Am. J. Trop. Med. Hyg. 87 (5), 897–901.
Jourdan, P.M., et al., 2018. Soil-transmitted helminth infections. Lancet 391 (10117), 252–265.
Kantzanou, M., et al., 2021. Prevalence of intestinal parasitic infections among children in Europe over the last five years. Trop. Med. Infect. Dis. 6 (3). https://doi.org/10.3390/tropicalmed6030160. Available at.
Kapczuk, P., et al., 2020. The influence of selected gastrointestinal parasites on apoptosis in intestinal epithelial cells. Biomolecules 10 (5). https://doi.org/10.3390/biom10050674. Available at.
Katona, P., Katona-Apte, J., 2008. The interaction between nutrition and infection. Clin. Infect. Dis. 46 (10), 1582–1588.
Koruk, I., et al., 2010. Intestinal parasites, nutritional status and physchomotor development delay in migratory farm worker's children. Child Care Health Dev. 36 (6), 888–894.
Lanata, C.F., et al., 2013. Global causes of diarrheal disease mortality in children <5 years of age: a systematic review. PLoS One 8 (9), e72788.
Mohammadi, R., et al., 2015. The relationship between intestinal parasites and some immune-mediated intestinal conditions. Gastroenterol. Hepatol. Bed Bench 8 (2), 123–131.
Ng, M., et al., 2014. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 384 (9945), 766–781.
Oriá, R.B., et al., 2016. Early-life enteric infections: relation between chronic systemic inflammation and poor cognition in children. Nutr. Rev. 74 (6), 374–386.
Parashar, U.D., Bresee, J.S., Glass, R.I., 2003. The global burden of diarrhoeal disease in children. Bull. World Health Organ. 81 (4), 236.
Price, D.L., 2017. Procedure Manual for the Diagnosis of Intestinal Parasites. CRC Press.
Pullan, R.L., et al., 2014. Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. Parasites Vectors 7, 37.
Quihui-Cota, L., et al., 2004. Prevalence and intensity of intestinal parasitic infections in relation to nutritional status in Mexican schoolchildren. Trans. R. Soc. Trop. Med. Hyg. 98 (11), 653–659.
Ravdin, J.I., Petri, W.A., 1990. *Entamoeba histolytica* (amebiasis). In: Principles and Practice of Infectious Diseases, pp. 2036–2049.
Rivera, M.T., et al., 2003. Trace elements, innate immune response and parasites. Clin. Chem. Lab. Med. 41 (8), 1020–1025.
Rossignol, J.-F., 2010. *Cryptosporidium* and *Giardia*: treatment options and prospects for new drugs. Exp. Parasitol. 124 (1), 45–53.
Sackey, M.-E., Weigel, M.M., Armijos, R.X., 2003. Predictors and nutritional consequences of intestinal parasitic infections in rural Ecuadorian children. J. Trop. Pediatr. 49 (1), 17–23.
Shirley, D.-A.T., et al., 2018. A review of the global burden, new diagnostics, and current therapeutics for amebiasis. Open Forum Infect. Dis. 5 (7), ofy161.
Stephenson, L.S., Holland, C.V., Cooper, E.S., 2000. The public health significance of *Trichuris trichiura*. Parasitology 121 (Suppl. I), S73–S95.
Strunz, E.C., et al., 2014. Water, sanitation, hygiene, and soil-transmitted helminth infection: a systematic review and meta-analysis. PLoS Med. 11 (3), e1001620.
World Health Organization, 2018. Guideline: Preventive Chemotherapy to Control Soil-Transmitted Helminth Infections in At-Risk Population Groups. World Health Organization.
World Health Organization, 2013. Soil-Transmitted Helminths Infections Factsheet. Available at. https://www.paho.org/en/topics/soil-transmitted-helminthiasis.
World Health Organization, 2010. Working to Overcome the Global Impact of Neglected Tropical Diseases: First WHO Report on Neglected Tropical Diseases. World Health Organization.
Zavala Gomez, G.A., 2017. Intestinal Parasites, Inflammation and Nutritional Status in Mexican Children. Ph.D. VU University Press. Available at. https://eprints.whiterose.ac.uk/155772/. (Accessed 10 January 2022).
Zavala, G.A., et al., 2016. Children with moderate-high infection with *Entamoeba coli* have higher percentage of body and abdominal fat than non-infected children. Pediatr. Obes. 11 (6), 443–449.
Zavala, G.A., et al., 2018. Intestinal parasites: associations with intestinal and systemic inflammation. Parasite Immunol. 40 (4), e12518. https://doi.org/10.1111/pim.12518. Available at.

Nutrition and susceptibility to tuberculosis

Pranay Sinha[a] **and J. Peter Cegielski**[b], [a] Boston University School of Medicine, Harvard University School of Public Health, Boston, MA, United States; and [b] Emory University Rollins School of Public Health, Atlanta, GA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of J. Peter Cegielski, D.N. McMurray, Nutrition and Susceptibility to Tuberculosis, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 309–314, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00268-3.

References 554
Further reading 554

Glossary

Cell-mediated immunity (CMI) The principle host defense against TB mediated primarily by T lymphocytes rather than antibodies

Droplet nuclei Microscopic particles (1–5 mm in diameter) that can be expelled when a person coughs, sneezes, shouts or sings. The droplets produced by an infectious TB patient can carry tubercle bacilli and can remain suspended in the air for prolonged periods of time.

Undernutrition A type of malnutrition caused by the lack of food or failure of the body to absorb or assimilate nutrients properly

Hematogenous re-seeding The process by which bacilli make their way back to the lung through the bloodstream where they infect all parts of the lung

Bacillus of Calmette and Guerin (BCG) vaccine A tuberculosis vaccine used in many parts of the world containing an attenuated strain of tubercle bacillus developed by repeated culture on medium containing bile.

Undernutrition is an important risk factor for the development of tuberculosis (TB) both at the individual level and at the population level. Undernutrition profoundly affects cell-mediated immunity (CMI), and CMI is the principal host defense against TB. Although these concepts may be widely accepted, the relative and attributable risks of TB due to undernutrition are not well known. Moreover, the effects of specific nutrients have not been established. Recent evidence suggests that overweight and obesity may actually decrease the risk of developing TB. This article will summarize available evidence on the relationship between nutritional status and the risk of TB.

Understanding the link between undernutrition and susceptibility to TB is based on a conceptual model for the transmission and pathogenesis of TB (McMurray, 2007). *Mycobacterium tuberculosis* is transmitted by the aerosol route when an individual who has pulmonary or tracheobronchial TB produces droplets containing viable *M. tuberculosis* by coughing. The moisture content of smaller droplets evaporates quickly and the resulting "droplet nuclei" are the main vehicle for airborne transmission of TB. In most people, the innate and adaptive immune responses control the infection without overt symptoms or other clinical manifestations of TB disease. In a small minority of individuals, estimated to be ~5%, the immune response will not control the infection, and the infection progresses to active TB disease known as "progressive primary TB." The risk of progression, however, is much greater in persons with immune systems weakened by undernutrition, HIV infection, immunosuppressive medications, and extremes of age. In persons whose immune systems successfully control the primary infection, mycobacteria may remain viable within granulomas for years or decades. One in four humans are believed to be latently infected with TB in this way. Approximately 10% of individuals with latent TB infections develop active TB disease in their lifetime (called secondary TB disease). Annually approximately ten million develop active TB disease globally. The risk of an infection progressing to active TB disease depends on the host's immune system and the microbe's infectious dose and pathogenicity. The influence of nutritional status on host defenses is a central theme of this article.

Once the inhaled mycobacteria reach the alveoli, alveolar macrophages bind and internalize the bacilli activating these macrophages that then produce pro-inflammatory cytokines such as tumor necrosis factor-alpha (TNF-α) and other chemokines (McMurray, 2007). These chemical messengers recruit macrophages and other immune response cells to the site of infection. In some individuals, the bacilli are destroyed at this stage, but in others, as the inflammatory response progresses, the mycobacteria and infected macrophages are taken in the draining lymph to the hilar and mediastinal lymph nodes, leading to activation of the acquired or adaptive immune system. From these lymph nodes, bacilli and infected phagocytes flow into the thoracic duct and the blood stream, disseminating to potentially every organ or tissue in the body. Bacilli make their way back to the lung through the bloodstream where they infect all parts of the lung in a process known as hematogenous re-seeding which apparently occurs in all infected individuals. Blood-borne organisms establish "secondary" granulomas, in which the organisms may persist despite an active immune response, and which are often the sites of secondary TB disease described above.

Mycobacterial antigens are presented to $CD4^+$ (helper) and $CD8^+$ (cytotoxic) T lymphocytes in the lymph nodes initiating the CMI (McMurray, 2007). The population of TB-specific T lymphocytes expands and circulates to sites of active infection where they produce macrophage-activating cytokines such as interferon-gamma (IFN-γ). The combination of IFN-γ and TNF-α from phagocytes further activates macrophages to inhibit mycobacterial growth. During this period, the host begins to express evidence of a TB-specific T-lymphocyte-mediated immune response as manifested classically by a positive tuberculin (PPD) skin test. The process by which macrophages inhibit or destroy mycobacteria may also cause local tissue destruction and necrosis. Immunological feedback mechanisms modulate the intense inflammatory response to limit tissue damage while controlling the infection. Nevertheless, local tissue destruction in the lungs can lead to formation of cavities, characteristic of contagious, reactivation TB.

While there is little evidence of increased risk of primary *infection*, the risk of progression from infection to *disease* increases substantially in undernourished individuals (Cegielski and McMurray, 2004). This is likely due to the adverse effects of undernutrition on CMI which is essential for control of mycobacterial growth. Protein undernutrition clearly compromises CMI. Experimental animal studies have demonstrated that even modest protein deprivation impairs host responses to BCG vaccination and tuberculin skin test responsiveness after aerosol exposure to *M. tuberculosis*. Protein repletion rapidly restores these responses. Protein from the diet and from somatic stores is crucial for many aspects of host defense against TB as described above. Essential amino acids play important physiological roles apart from serving as building blocks for protein synthesis. Other macronutrients may also influence the immune system, especially fatty acids, including *n*-6 and *n*-3 polyunsaturated fatty acids (PUFA) (McFarland and et al., 2008). *n*-3 PUFA have a direct influence on immune cell membrane composition and function, as well as on the production of eicosanoids and other inflammatory mediators. High intakes of long chain *n*-3 PUFA dampen inflammatory responses and ameliorate chronic inflammatory conditions such as rheumatoid arthritis, potentially at the cost of increasing susceptibility to infection including TB (see below). The balance *n*-3 and *n*-6 PUFA may be a key determinant of their effects on the immune system and disease resistance (McFarland et al., 2008).

Recent findings suggest that chronic excess caloric intake may actually decrease the risk of TB (Leung et al., 2007). A 2009 study from China demonstrated that persons who were overweight and obese had ~1/4th to ~1/5th the risk of developing TB, respectively, compared with persons having normal body mass index. These findings have been corroborated by a contemporaneous cohort study from Peru which found that those with premorbid BMI > 25 kg/m^2 had approximately 1/3rd the risk as compared to those with BMIs in the normal range (18.5–25 kg/m^2) (Aibana et al., 2017). These findings are consistent with other studies which are described below. The mechanisms by which excessive body mass decreases TB incidence are ripe for investigation. Recent research demonstrates that adipose tissue may be a reservoir for non-replicating *M. tuberculosis*, and that such bacilli accumulate triacylglycerol in intracytoplasmic lipid bodies visible with appropriate staining by regular light microscopy (Garton et al., 2008). A triacylglycerol synthase-encoding gene (*tgs1)* in *M. tuberculosis is* a member of the DosR regulon that may control the development of a non-replicating "persister" phenotype in the mycobacteria. Induction of the isocitrate lyase gene, *icl1*, in these mycobacteria allows a shift to utilization of lipids as a source of carbon and energy (Garton et al., 2008). Indeed, Roth speculated that obesity may have provided an evolutionary advantage during TB epidemics in the past, predisposing to obesity the modern day descendants of the survivors (Roth, 2009). Taken together, these recent data suggest the presence of a non-replicating bacterial population which persists especially in adipose tissue. Such mycobacteria may stop replicating or replicate more slowly, decreasing the probability of active TB disease.

Vitamin A may affect resistance to TB by altering the innate and the adaptive immune response to *M. tuberculosis* (Aibana et al., 2017). Vitamin A is necessary for the function of the respiratory mucosa and the integrity of pulmonary epithelial tissues. Retinol modulates the maturation of dendritic cells which are needed for antigen presentation and T-cell priming. Retinoic acid fosters the T-helper (Th)-1 response against M. tuberculosis. Moreover, it also modulates T-regulatory cells and downregulates Th-17 cells. Lastly, all-trans-retinoic-acid may have direct antimycobacterial properties as it disrupts lipid metabolism.

Vitamins of the B complex and vitamin C play important roles in B-cell mediated humoral immune responses, but at present there is less evidence for their role in CMI. Vitamin B_{12} may be an exception to this generalization; B_{12} repletion in patients with pernicious anemia has been shown to reverse skin test anergy. The antioxidant functions of vitamin C, vitamin E, selenium, and glutathione have critical roles in protection against oxidative stresses, including reactive oxygen intermediates that play effector roles in cellular immunity as described below. Vitamin E may have other effects on both cellular and humoral limbs of the immune system as well especially in relation to *n*-3 PUFA. Vitamin E also stimulates the Th-1 response which is critical for suppressing *M. tuberculosis*.

Vitamin D may have a role in *anti*-TB immunity through activation of macrophages (Cegielski and McMurray, 2004; Cegielski and Vernon, 2015). Induction of a cell-surface receptor for the vitamin D metabolite, 25-hydroxycholecalciferol, as well as the hydroxylase which forms 1,25-dihydroxycholecalciferol, appear to be part of the mechanism that helps these cells destroy the pathogen. These vitamin D effects are linked with production of the endogenous antibiotic peptide, cathelicidin. Two cross-sectional studies on vitamin D metabolism in relation to TB have focused on the molecular and cellular mechanisms of the interaction rather than on the direction of causality (reviewed in Cegielski and McMurray (2004)). Cells recovered by bronchoalveolar lavage (BAL) and peripheral blood mononuclear cells from TB patients both produced $1,25(OH)_2D3$, the amount correlating with the number of $CD8^+$ T lymphocytes but not with other cell types. $CD4^+$ T lymphocytes in BAL fluid from TB patients expressed specific receptors for $1,25(OH)_2$ vitamin D3 but not 25(OH)D3. Since $1,25(OH)_2D3$ can improve the capacity of macrophages to kill mycobacteria, these results support the conclusion that cellular interactions mediated in part by $1,25(OH)_2D3$

may be important in resistance to TB. In a prospective study, 1103 persons with HIV had their vitamin D levels measured at the time of antiretroviral therapy initiation. Subjects were followed monthly. This study found that, after adjusting for confounders, compared to participants with sufficient vitamin D levels, those with vitamin D deficiency (serum 25(OH)D3 levels <20 ng/mL) were significantly more likely to develop TB disease (Odds Ratio 2.89; 95% confidence interval: 1.13–7.41). Those with vitamin D insufficiency (serum 25(OH)D3 levels <20 ng/mL) did not have a significantly elevated risk of TB disease (Odds Ratio 1.14; 95% confidence interval: 0.60–2.18). However, a randomized double-blind trial that provided vitamin D supplementation to individuals with vitamin D insufficiency did not find a difference in the incidence of pulmonary tuberculosis between those receiving the supplementation and the control group (Ganmaa et al., 2020).

Dietary mineral deficiencies also affect CMI. Zinc deficiency interferes with T cell replication and maturation leading to lymphopenia. Zinc deficiency is also associated with elevated glucocorticoid levels that suppress cell-mediated immunity. Iron plays a critical role in support of CMI, however, iron is also critical to the replication of mycobacteria and other pathogens. Mycobacteria have evolved intense iron-scavenging mechanisms, and iron deficiency may have worse consequences for the microbe than the host. Conversely, iron overload in the reticuloendothelial system may increase susceptibility to TB. Indeed, a meta-analysis has found an association between SLC11A1 mutations in Asian and African populations and increased TB disease incidence. Thus, the impact of iron on susceptibility to TB is difficult to predict.

Despite the seemingly clear pathways through which nutrition affects CMI and resistance to TB, the evidence in humans linking undernutrition to TB risk is indirect and surprisingly weak from the perspective of scientific rigor. Early research on the interaction of nutrition and TB was exhaustively reviewed in the classic text by Scrimshaw, Taylor, and Gordon (Scrimshaw et al., 1968). The studies in humans were either ecological or uncontrolled observational studies. For example, studies of the sharp increases in TB incidence and mortality in France and Germany during the two World Wars, in the Netherlands during the famine of 1944–1945 (the hunger winter), or in the Warsaw ghetto during World War II, do not isolate the effects of undernutrition from the impact of extreme crowding, social and environmental degradation, lack of medical services, and catastrophic social circumstances. Another notable study of this era includes the Papworth experiment which studied the impact of a suite of socioeconomic interventions on TB. TB patients were ensured a job and were given adequate nutrition and lodging after their discharge from sanatoria. There was a dramatic decrease in TB disease incidence among children under 5 years of age who were born within the Papworth settlement as compared to the rate prevailing at the time in England: 0 vs. 1217 cases/100,000 person-years (Bhargava et al., 2012). While highly suggestive, again, these studies are unable to isolate the effect of undernutrition from the effects of poor housing, overcrowding, lack of medical care, poor hygiene, social disruption, and poverty. Although much of this observational evidence is scientifically weak, it constitutes a large body of repeated observations supporting a relationship between undernutrition and TB. In addition, the decline in TB in the global North prior to the mid-20th Century advent of modern *anti*-TB chemotherapy and widespread BCG vaccination is often ascribed to improvements in living conditions, especially nutrition, a concept championed in seminal work by Thomas McKeown (McKeown and Record, 1962). The remainder of this article summarizes the evidence from observations in human populations and from experimental animal models with relevance to human TB.

Three ecological studies present compelling evidence that undernutrition, isolated to some extent from other confounding circumstances, plays a direct role in TB morbidity and mortality (Cegielski and McMurray, 2004; McMcMurray and Cegielski, 2007). During World War I, neutral Denmark exported the bulk of its meat, fish, poultry, and dairy products to support the war effort elsewhere so the local diet lacked these protein-, vitamin-, and mineral-rich foods. During that period, TB rates increased similar to the warring countries. In 1918, however, Germany blockaded Denmark making exports impossible, creating a local surplus of these foods. TB rates in Denmark plummeted while rates in the neighboring countries continued to increase unabated. The second study involves the Trondheim, Norway, Naval Training School, where an extremely high rate of TB among recruits in the early 20th century was ascribed to crowding, poor housing, and unhygienic conditions. TB rates did not decrease after improved housing and hygiene were provided. On the other hand, when the diet was fortified with milk, margarine, cod liver oil, whole wheat bread, and fresh fruits and vegetables, TB morbidity promptly declined to the prevailing level for young adults of that area. After WWII, Leyton reported on British and Russian prisoners of war (POW) held in the same German POW camps, sharing the same living conditions and diet, except the British received Red Cross food supplements amounting to 30 g protein and 1000 kcal per day. In a subsequent radiographic survey, the TB rate among the British POW was only 1.2% while the rate in Russian POW was 15%–19%. In the malnourished prisoners, TB was more severe, the onset was more rapid, and patients died rapidly with large pulmonary cavities and massive tissue breakdown. Granuloma formation was poor in the malnourished prisoners, supporting the idea there was a deficit of CMI in this group.

Protein–energy undernutrition compromises CMI and may exacerbate TB, however, TB itself can adversely affect nutritional status. Understanding the temporal relationship between the onset of undernutrition and the development of TB is crucial to correctly assess any possible cause–effect relationship. Cross-sectional and case–control studies generally suffer from the same flaw: Patients with and without TB disease are compared in terms of their concurrent nutritional status. While these studies demonstrate substantial macro- and micronutrient deficits in TB patients, they are not useful in determining the effect of undernutrition on susceptibility to TB because TB causes undernutrition. The intrinsic uncertainty over the chronological sequence of cause and effect in case–control and cross–sectional studies become intractable.

The effects of undernutrition on the immune response to mycobacterial proteins closely related to the CMI required for resistance to infection with *M. tuberculosis*, namely delayed-type hypersensitivity (DTH) responses, have been studied following BCG vaccination. Satyanarayana et al. showed that milder grades of undernutrition did not affect the tuberculin skin test response 6 months after immunization with BCG, but children with kwashiorkor were skin test negative (Satyanarayana et al., 1980).

Chandra and Newberne demonstrated that the DTH skin test response to tuberculin was reduced in protein–energy undernourished children and adults (Chandra and Newberne, 1977). Among TB patients, PPD skin test reactivity was directly proportional to serum transferrin level, a sensitive indicator of protein undernutrition. Similarly, malnourished children did not develop skin test responses to tuberculin as often or as large after BCG vaccination as did well-nourished children. Importantly, this defect has been demonstrated even in modest protein–energy undernutrition. A study of interferon gamma release assays (IGRA) among children in Bangladesh found that undernourished children (BMI z-score < -2) were more likely to have lower levels of interferon gamma response to *M. tuberculosis* antigens and were also more likely to have indeterminate IGRA results.

After intestinal bypass surgery for morbid obesity, patients experience rapid weight loss and malabsorption due to their short-circuited bowels (Snider, 1982). In several case series, the postoperative incidence of TB was 10- to 100-fold higher than historical or population comparison groups. Similarly, partial gastrectomy for ulcer disease was shown to predispose men to TB, especially among men whose weight was <85% of ideal.

The unique strength of cohort studies is nutritional status is measured prior to the onset of TB. Only a handful of cohort studies have examined the relationship between micronutrients and TB incidence. In the 1940s, Getz and coworkers enrolled a cohort of 1100 men who were free of TB at baseline, and followed them for up to 5 years with serial clinical, radiographic, and laboratory examinations (Getz et al., 1951). Plasma vitamin A levels were low in 13 of 16 men (81%) who developed TB compared to 318 of 1058 (30%) of those who did not. Similarly, plasma vitamin C levels were low in 100% of the subjects who developed active TB compared to 117 of 1013 (11%) of those who did not. Exposure to TB did not differ between the groups. In a Finnish study on cancer prevention, investigators randomized 26,975 healthy male smokers aged 50–69 years to supplementation with tocopherol, beta-carotene, both, or neither (Hemilä et al., 1999). The subjects were followed for a mean of 6.7 years. In >173,000 person-years of follow-up, 167 cases of TB were detected. Higher intakes of fruits and vegetables were associated with lower risk of TB (the adjusted relative risk of TB was 0.4; 95% confidence interval, 0.24–0.69). This study is noteworthy for its size and data quality, however, detecting TB through hospital discharges selects TB patients who were sick enough to require hospital admission. Lower dietary intakes of key nutrients may have been associated with higher rates of hospitalization rather than (or in addition to) higher rates of TB. More recently, a case-control study nested in a prospective cohort of 6751 HIV-negative household contacts of persons with TB in Peru, studied the impact of premorbid vitamin A and E deficiency on incident TB (Aibana et al., 2017). In this study, 180 individuals developed active TB disease. Vitamin A deficiency was associated with a ten-fold higher risk of incident TB disease (Odds Ratio: 10.53; 95% confidence interval, 3.73–29.70) and baseline alpha-tocopherol deficiency was associated with higher risk of incident TB disease (Odds Ratio: 1.59; 95% confidence interval, 1.02–2.50).

As part of the long-term follow-up of participants in large-scale BCG vaccine trials in Georgia and Alabama, Comstock and Palmer reported the incidence of TB was 2.2 times higher in children with 0–4 mm subcutaneous fat than in children with 10 mm subcutaneous fat (Comstock and Palmer, 1966). Cegielski examined the relationship between undernutrition, as determined in the first National Health and Nutrition Examination Survey (NHANES-1), and TB incidence as ascertained in the NHANES-1 Epidemiological Follow-up Study (Cegielski et al., 2012). NHANES-1 was a cross-sectional survey based on a representative sample of the US population from 1971 to 1975. In the Follow-up Study, the adult subjects of NHANES-1, aged 25–74 years at baseline, were followed up until 1992. Having body mass index (BMI), average skin-fold thickness, or mid-upper arm cross-sectional muscle area in the lowest decile of the population increased the adjusted hazard of TB from 6- to 10-fold, controlling for other known risk factors for TB.

In a related vein, three massive studies focused on "body build" as a risk factor for TB incidence (Palmer et al., 1957; Edwards et al., 1971; Tverdal, 1986; Snider, 1987). Palmer et al. reported on the relationship between TB incidence and naturally acquired delayed-type tuberculin sensitivity among nearly all US Navy recruits from 1949 to 1951. Of 68,754 subjects with follow up data, 8704 (12.7%) had tuberculin sensitivity recorded as >0 mm induration. During 4 years of follow-up, 109 developed TB: 28 among those with 0 mm skin test reactions, 29 among those with 1–9 mm reactions, and 57 among those with 10 mm or greater reactions. Later, these investigators related the risk of TB to the recruits' height and weight data on a stratified random sample of 1138 subjects. TB incidence was 75/105 for those 15% or more below the median weight for height, decreasing to 19/105 for those at least 5% overweight for height. Edwards et al. extended this study to more than 823,000 Navy recruits and found that TB developed 3-fold more often in young men 10% or more below their ideal body weight compared with those 10% or more above their ideal body weight. Rather than attribute these results to nutritional status, the authors concluded that there was an association between "body build" and risk of TB disease. The relationship between body build and TB was reviewed by Snider in 1987. One study stands out. Norway screened 42%–85% of the population older than age 14 years for TB with radiography from 1963 to 1975. Height and weight were measured accurately for approximately 80% of those screened. As reported by Tverdal et al. more than 1.7 million Norwegians were followed up via the national notification system through 1982 (i.e., 8–19 years of follow-up; mean, 12.1 years). A total of 2531 incident cases of TB were identified. The incidence of pulmonary TB declined logarithmically with increasing BMI for both sexes, all age groups, and at all durations of follow-up. The age-adjusted incidence of pulmonary TB was five times higher in the lowest BMI category than in the highest. Even though the study was based on BMI, Tverdal argued that the observed relationship was a function of body build, not nutritional status. Comstock suggested body build may influence susceptibility to TB because of differences in pulmonary mechanics, but no data supporting this hypothesis exist. Interpreting the findings in terms of unknown factors associated with body build rather than the most obvious explanation, i.e., nutritional status, discounts the relationship between body weight, caloric intake, and energy expenditure. A more nuanced view may be that body habitus is a function of genetic endowment and of nutrient intake plus physical activity, each of which affects the incidence of TB in complex ways.

A unique study of the effect of micronutrient supplementation on TB incidence was reported by Downes in 1949 (Downes, 1950). In a controlled trial among the families of black TB patients in the Harlem ghetto of New York City, 194 of 218 families under public health supervision were examined and divided into two groups matched for family size. The families were allocated alternately to receive vitamin and mineral supplements versus no supplements. The two groups were similar in terms of prior attack rates and mortality from TB, prevalence of TB at the start of the study, sputum smear positivity among the index cases, and relation of the index case to the rest of the family. In addition, the groups were similar in terms of their economic status, crowding, and eating habits. After 5 years of follow-up, the risk of TB in the control group was 2.8-fold higher than the supplemented group. However, there was substantial nonadherence with the supplements. Compared with those who actually took the vitamin supplements throughout the observation period, the risk of TB in the control group was 5.9-fold higher. Therefore, vitamin supplementation substantially reduced the risk of TB among family contacts of active TB cases.

Experimental animal models allow investigators to elucidate the causal links between nutritional deficiencies, immune system function, and TB in ways not possible in human studies. The link between diet, antimycobacterial immunity, and resistance to TB has been investigated in a highly relevant guinea pig model of low-dose pulmonary TB that mimics the pathogenesis of TB in humans (Smith and Harding, 1977; McMurray et al., 1981; McMurray and Bartow, 1992; McMurray, 1998; Chan et al., 1996). Moderate, chronic deficiencies of protein and other nutrients (e.g., zinc, vitamin D) induced in guinea pigs had many of the metabolic hallmarks of human dietary deficiencies. Groups of BCG-vaccinated and nonvaccinated guinea pigs were given different diet treatments and then challenged with an aerosol containing a low dose of virulent *M. tuberculosis*. Antigen-specific immune responses in vitro and in vivo were assessed several weeks later and the ability of the guinea pigs to control the infection was determined quantitatively by culture of viable mycobacteria from the lungs and spleens.

Moderate, chronic protein deficiency (modeled by a 10% ovalbumin-based diet) over several weeks resulted in a dramatic loss of CMI in infected guinea pigs. Protein deprived animals had much smaller DTH reactions, and their T lymphocytes proliferated poorly to PPD in vitro and produced significantly less interleukin (IL)-2, IFNγ. Macrophages from protein malnourished guinea pigs produced less TNFα in response to infection with virulent *M. tuberculosis*. Protein-deficient guinea pigs were unable to form mature, well-circumscribed granulomas in the lung. BCG-induced protection was diminished by protein deficiency. Protein undernutrition altered the numbers of $CD4^+$ and $CD8^+$ T cells in the spleen and bronchotracheal lymph nodes draining the infected lung. Thus, protein deficiency was accompanied by impairment of the normal trafficking of T lymphocytes that would be required for the formation of protective granulomas due, perhaps, to diet-induced changes in the production or function of chemokines, or by perturbations in the expression of adhesion molecules on T cells or endothelial cells.

Macrophages from TB patients are known to produce suppressive factors for T cells, including transforming growth factor-beta (TGF-β). Alveolar macrophages effectively down regulate T cell activation to mitigate potentially damaging pulmonary inflammation in response to inhaled antigens. Alveolar macrophages from protein-deficient guinea pigs exerted a 10-fold greater suppression of T cells compared to cells from normally nourished animals, perhaps due to the greater levels of TGF-β produced by these cells. Recombinant human TGF-β injected daily into guinea pigs infected with virulent *M. tuberculosis* suppressed T cell functions and impaired bacillary control in a manner similar to dietary protein deficiency. Thus, macrophages from protein-deprived guinea pigs appear to be more suppressive for T lymphocyte functions, and this suppression may be mediated, in part, by overproduction of TGF-β.

One of the most important findings from this model is that the profound loss of T cell-mediated resistance that accompanies chronic dietary protein deprivation was substantially and rapidly reversible. Protein-deficient, BCG-vaccinated guinea pigs given a normal diet beginning on the day of pulmonary challenge with *M. tuberculosis* displayed DTH reactivity and control of bacillary growth within a few weeks that were indistinguishable from those in BCG-vaccinated animals that had never been protein deficient. Similar results were observed in protein-malnourished mice. Using a high-dose, intravenous challenge model, Chan and colleagues observed many of the same T cell defects that have been reported in low protein guinea pigs, including inability to control the virulent infection, impaired granuloma formation, and recovery of resistance following refeeding with an adequate diet (Chan et al., 1996). These studies confirm the fundamental nature of the effects of protein deprivation on susceptibility to TB even when host species, and infection dose and route are altered.

Guinea pig and mouse models of low-dose, pulmonary TB have been used to demonstrate the significant effects of dietary n-3 PUFA on TB resistance (Garton et al., 2008). Guinea pigs fed a diet enriched in fish oil or transgenic *fat*-1 mice producing n-3 PUFA endogenously were more susceptible to infection with virulent *M. tuberculosis*. Immune cells from *fat-1* mice or cells from wild-type mice cultured in medium containing *n*-3 PUFA produced less TNFα, IL-6, and IL-1β, and exhibited reduced oxidative burst and impaired phagosome-lysosome fusion. These *n*-3 PUFA-enriched macrophages were significantly impaired in their ability to control *M. tuberculosis* over several days of culture.

This article has reviewed and critiqued published studies in human populations and in relevant animal models covering the in vivo evidence relating the risk of TB due to nutritional status and nutritional factors. Although TB is clearly related to undernutrition, the risk relative to specific levels and types of protein-energy deficiency and micronutrient deficiencies remain to be defined. Analysis of the NHANES-1 Epidemiological Follow-up Study provides plausible estimates of a 6- to 10-fold increase in relative risk among undernourished adults as well as a substantially decreased risk in overweight and obese individuals. Severe protein-energy deficiencies may increase the relative risk more than mild or moderate undernutrition, but severe undernutrition affects fewer people, even in low-income countries, except during famines, war, natural disasters, pandemics, etc. Mild to moderate protein-energy or micronutrient deficiencies affect more people at risk for TB, so prevention efforts should target those groups as well. The population attributable risk of TB due to undernutrition may be substantial, especially in populations where both TB

infection and undernutrition are prevalent. Undernourished individuals have an increased likelihood of primary or latent infection progressing to active disease. In populations with substantial latent TB infection, the occurrence of undernutrition may be an important determinant of the incidence of reactivation TB. In many developing countries, the risk of becoming infected with TB is as high as 1%–2% per year of life. The United Nations Food and Agriculture Organization estimates that one billion people face food insecurity. When combined with an estimate of two billion people latently infected with *M. tuberculosis*, even modest decreases in resistance affecting such large numbers of people may result in substantial increases in TB incidence. The potential public health impact of undernutrition on the global incidence of TB was summarized in a US Surgeon General's Report on Nutrition and Health, which emphasized that undernutrition was the leading cause of acquired, correctable immune system dysfunction throughout the world (US Public Health Service, 1988). Population groups at highest risk for poor nutrition are also at high risk for TB; poverty is the common denominator.

References

Aibana, O., Franke, M.F., Huang, C.C., Galea, J.T., Calderon, R., Zhang, Z., et al., 2017. Impact of vitamin A and carotenoids on the risk of tuberculosis progression. Clin. Infect. Dis. 65 (6), 900–909.

Bhargava, A., Pai, M., Bhargava, M., Marais, B.J., Menzies, D., 2012. Can social interventions prevent tuberculosis?: the Papworth experiment (1918–1943) revisited. Am. J. Respir. Crit. Care Med. 186 (5), 442–449.

Cegielski, J.P., McMurray, D.N., 2004. The relationship between malnutrition and tuberculosis: evidence from studies in humans and experimental animals. Int. J. Tubercul. Lung Dis. 8, 286–298.

Cegielski, J.P., Vernon, A., 2015. Tuberculosis and vitamin D: what's the rest of the story? Lancet Infect. Dis. 15, 489–490.

Cegielski, J.P., Arab, L., Cornoni-Huntley, J., 2012. Nutritional risk factors for tuberculosis among adults in the United States, 1971–1992. Am. J. Epidemiol. 176 (5), 409–422.

Chan, J., Tian, Y., Tanaka, K.E., et al., 1996. Effects of protein calorie malnutrition on tuberculosis in mice. Proc. Natl. Acad. Sci. U. S. A. 93, 14857–14861.

Chandra, R.K., Newberne, P.M., 1977. Nutrition, Immunity, and Infection: Mechanisms of Interactions. Plenum Press, New York, NY.

Comstock, G.W., Palmer, C.E., 1966. Long-term results of BCG vaccination in the Southern United States. Am. Rev. Respir. Dis. 93, 171–183.

Downes, J., 1950. An experiment in the control of tuberculosis among negroes. Milbank Mem. Fund. Q. 28, 180–220.

Edwards, L.B., Livesay, V.T., Acquaviva, F.A., Palmer, C.E., 1971. Height, weight, tuberculous infection, and tuberculous disease. Arch. Environ. Health 22, 106–112.

Ganmaa, D., Uyanga, B., Zhou, X., Gantsetseg, G., Delgerekh, B., Enkhmaa, D., et al., 2020. Vitamin D supplements for prevention of tuberculosis infection and disease. N. Engl. J. Med. 383 (4), 359–368.

Garton, N.J., Waddell, S.J., Sherratt, A.L., Lee, S.M., Smith, R.J., Senner, C., et al., 2008. Cytological and transcript analyses reveal fat and lazy persister-like bacilli in tuberculous sputum. PLoS Med. 5 (4), e75.

Getz, H.R., Long, E.R., Henderson, H.J., 1951. A study of the relation of nutrition to the development of tuberculosis: influence of ascorbic acid and vitamin A. Am. Rev. Tubercul. 64, 381–393.

Hemilä, H., Kaprio, J., Pietinen, P., Albanes, D., Heinonen, O.P., 1999. Vitamin C and other compounds in vitamin C rich food in relation to risk of tuberculosis in male smokers. Am. J. Epidemiol. 150, 632–641.

Leung, C.C., Lam, T.H., Chan, W.M., Yew, W.W., Ho, K.S., Leung, G., et al., 2007. Lower risk of tuberculosis in obesity. Arch. Intern. Med. 167 (12), 1297–1304.

McFarland, C.T., et al., 2008. Dietary polyunsaturated fatty acids modulate resistance to *Mycobacterium tuberculosis* in Guinea pigs. J. Nutr. 138, 2123–2128.

McKeown, T., Record, R.G., 1962. Reasons for the decline of mortality in England and Wales during the nineteenth century. Popul. Stud. xvi, 94–122.

McMurray, D.N., 1998. Impact of nutritional deficiencies on resistance to experimental pulmonary tuberculosis. Nutr. Rev. 56, S147–S152.

McMurray, D.N., 2007. Human immunity (ch. 5) and pathogenesis of *Mycobacterium tuberculosis* infections in humans (ch. 6). In: Academy of Sciences of South Africa Consensus Panel on Nutrition, HIV/AIDS, TB (Eds.), HIV/AIDS, TB, and Nutrition: Scientific Inquiry Into the Nutritional Influences on Human Immunity With Special Reference to HIV Infection and Active TB in South Africa. Academy of Sciences of South Africa, Pretoria, South Africa, pp. 85–100.

McMcMurray, D.N., Cegielski, J.P., 2007. The influence of nutrition on the risk and outcomes of tuberculosis. Ch. 11. In: Academy of Sciences of South Africa Consensus Panel on Nutrition, HIV/AIDS, TB (Eds.), HIV/AIDS, TB, and Nutrition: Scientific Inquiry Into the Nutritional Influences on Human Immunity With Special Reference to HIV Infection and Active TB in South Africa. Academy of Sciences of South Africa, Pretoria, South Africa.

McMurray, D.N., Bartow, R.A., 1992. Immunosuppression and alteration of resistance to pulmonary tuberculosis in Guinea pigs by protein undernutrition. J. Nutr. 122, 738–743.

McMurray, D.N., Loomis, S.A., Casazza, L.J., Rey, H., Miranda, R., 1981. Development of impaired cell mediated immunity in mild and moderate malnutrition. Am. J. Clin. Nutr. 34, 68–77.

Palmer, C.E., Jablon, S., Edwards, P.Q., 1957. Tuberculosis morbidity of young men in relation to tuberculin sensitivity and body build. Am. Rev. Tuberc. Pulm. Dis. 76, 517–539.

Roth, J., 2009. Evolution speculation about tuberculosis and the metabolic and inflammatory processes of obesity. JAMA 301, 2586–2588.

Satyanarayana, K., Bhaskaran, P., Seshu, V.C., Reddy, V., 1980. Influence of nutrition on post-vaccinial tuberculin sensitivity. Am. J. Clin. Nutr. 33, 2334–2337.

Scrimshaw, N.S., Taylor, C.E., Gordon, J.E., 1968. Effect of malnutrition on resistance to infection. In: Scrimshaw, N.S., Taylor, C.E., Gordon, J.E. (Eds.), Interactions of Nutrition and Infection. World Health Organization, Geneva, pp. 60–142.

Smith, D.W., Harding, G.E., 1977. Animal model: experimental airborne tuberculosis in the Guinea pig. Am. J. Pathol. 89, 273–276.

Snider, D.E., 1982. Jejunoileal bypass for obesity: a risk factor for tuberculosis. Chest 81, 531–532.

Snider, D.E., 1987. Tuberculosis and body build. JAMA 258, 3299.

Tverdal, A., 1986. Body mass index and incidence of tuberculosis. Eur. J. Respir. Dis. 69, 355–362.

US Public Health Service, 1988. The Surgeon General's Report on Nutrition and Health. DHHS Publ No (PHS) 88-50210, Washington, DC, pp. 427–463.

Further reading

Aibana, O., Franke, M.F., Huang, C.C., Galea, J.T., Calderon, R., Zhang, Z., et al., 2018. Vitamin E status is inversely associated with risk of incident tuberculosis disease among household contacts. J. Nutr. 148, 56–62.

Bonilla, D.L., Fan, Y.Y., Chapkin, R.S., McMurray, D.N., 2010. Transgenic mice enriched in omega-3 fatty acids are more susceptible to pulmonary tuberculosis: impaired resistance to tuberculosis in *fat-1* mice. J. Infect. Dis. 201, 399–408.

Cooper, A.M., 2009. Cell-mediated immune responses in tuberculosis. Annu. Rev. Immunol. 27, 393–422.

Ganmaa, D., Giovannucci, E., Bloom, B.R., Fawzi, W., Burr, W., Batbaatar, D., et al., 2012. Vitamin D, tuberculin skin test conversion, and latent tuberculosis in Mongolian school-age children: a randomized, double-blind, placebo-controlled feasibility trial. Am. J. Clin. Nutr. 96 (2), 391–396. https://doi.org/10.3945/ajcn.112.034967.

Harland, P.S.E.G., Brown, R.E., 1965. Tuberculin sensitivity following BCG vaccination in undernourished children. East Afr. Med. J. 42, 233–238.

Kielmann, A.A., Uberol, I.S., Chandra, R.K., Mehra, V.L., 1976. The effect of nutritional status on immune capacity and immune responses in preschool children in a rural community in India. Bull. World Health Organ. 54, 477–483.

Liu, P.T., Modlin, R.L., 2008. Human macrophage host defense against *Mycobacterium tuberculosis*. Curr. Opin. Immunol. 20, 371–376.

Lonnroth, K., Williams, B.G., Cegielski, P., Dye, C., 2010. A consistent log-linear relationship between tuberculosis incidence and body mass index. Int. J. Epidemiol. 39, 149–155.

McFarland, C.T., Fan, Y.Y., Chapkin, R.S., Weeks, B.R., McMurray, D.N., 2008. Dietary polyunsaturated fatty acids modulate resistance to *Mycobacterium tuberculosis* in Guinea pigs. J. Nutr. 138, 2123–2128.

McMurray, D.N., Bartow, R.A., Mintzer, C.L., Hernandez-Frontera, E., 1990. Micronutrient status and immune function in tuberculosis. Ann. N. Y. Acad. Sci. 587, 59–69.

Muttucumaru, D.G., Roberts, G., Hinds, J., Stabler, R.A., Parish, T., 2004. Gene expression profile of *Mycobacterium tuberculosis* in a nonreplicating state. Tuberculosis 84, 239–246.

Neyrolles, O., Hernandez-Pando, R., Pietri-Rouxel, F., Fornes, P., Tailleux, L., et al., 2006. Is adipose tissue a place for *Mycobacterium tuberculosis* persistence? PLoS One 1, e43. https://doi.org/10.1371/journal.pone.0000043.

Sirakova, T.D., Dubey, V.S., Deb, C., Daniel, J., Korotkova, T.A., et al., 2006. Identification of a diacylglycerol acyltransferase gene involved in accumulation of triacylglycerol in *Mycobacterium tuberculosis* under stress. Microbiology l152, 2717–2725.

Steinberg, B.E., Grinstein, S., 2008. Pathogen destruction versus intracellular survival: the role of lipids as phagosomal fate determinants. J. Clin. Invest. 118, 2002–2011.

Ulrichs, T., Lefmann, M., Reich, M., et al., 2005. Modified immunohistological staining allows detection of Ziehl-Neelsen-negative *Mycobacterium tuberculosis* organisms and their precise localization in human tissue. J. Pathol. 205, 633–640.

Nutritional support: Adults, enteral

Berkeley Limketkai[a], **Octavia Pickett Blakely**[b], **and Gerard E. Mullin**[c], [a] Division of Digestive Diseases, UCLA School of Medicine, Los Angeles, CA, United States; [b] Division of Gastroenterology, University of Pennsylvania Perlman School of Medicine, Philadelphia, PA, United States; and [c] Division of Gastroenterology & Hepatology, Johns Hopkins University School of Medicine, Baltimore, MD, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A.K. Fischer, S.D. Rampertab, G.E. Mullin, Nutritional Support: Adults, Enteral, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 258–263, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00203-8.

Introduction 556
Types 557
Access 557
Orogastric or nasogastric route 557
Oroenteric, nasoduodenal or nasojejunal route 558
Percutaneous endoscopic gastrostomy (PEG) or percutaneous endoscopic jejunostomy (PEJ) 558
Feeding formulas 558
Feeding formula classification 558
Compositions 559
Polymeric formulas 559
Elemental or semielemental formulas 559
Disease-specific formulas 559
Diabetic 559
Renal 559
Hepatic 560
Pulmonary 560
Immune enhancing 560
Arginine 560
Glutamine 560
Omega-3 fatty acids 560
Probiotics 560
Modular 561
Indications and contradictions 561
Methods of infusion 561
Potential challenges 562
Summary/conclusions 562
References 562
Further reading 563

Key points

- To define enteral nutrition support.
- To describe the indications for enteral nutrition.
- To provide a summary of the types of enteral nutrition and disease-specific formulations.
- To provide an overview of the types of access devices and infusions available for the provision of enteral nutrition.
- To list the common complications of enteral nutrition.

Introduction

Enteral nutrition (EN) (also known as enteral feeding or tube feeding) is a method of providing nutrients directly into the stomach or intestines through a thin flexible tube when a person cannot receive food orally. It is used in patients who have a functional gastrointestinal (GI) tract that can adequately digest and absorb food, but in whom oral intake is inadequate to maintain or restore optimal nutritional status. Inadequate oral intake could arise from conditions such as difficulty swallowing, altered mental status, or loss of appetite. There are several routes of delivery for EN (Tariq et al., 2020). Depending on the individual patient's needs, feeding

tubes can be placed via the mouth or nose and advanced into the stomach (orogastric, nasogastric tube) or duodenum (oroenteric, nasoenteric tube) (Boullata, 2021). Tubes can also be placed percutaneously via the abdominal wall into the stomach (gastrostomy) or the small intestine (jejunostomy). EN is generally considered the safer and most physiologic method of delivering nutritional support and therefore, is preferred over parenteral or intravenous nutrition support (Pironi et al., 2020). In this article, the use of enteral feeding is reviewed.

Types

EN requires liquid formulas to be administered through a tube into the GI tract. Once it has been decided that an individual is a candidate for EN, the tube type and appropriate route of access for tube placement can be selected. In general, feedings administered through oral or nasal tubes is indicated for periods lasting less than 2 weeks, whereas, when enteral feeding is required for a longer duration of time, more permanent tubes (gastrostomies and jejunostomies) are utilized (Boullata, 2021). The route of access that is selected for enteral feeding depends on multiple factors:

1. Length of time that enteral feeding will be administered.
2. Degree of risk for aspiration or device displacement.
3. Patency and function of the upper GI tract.
4. Anticipated/planned or history of prior GI surgery with altered upper GI tract anatomy.
5. Nutrition administration issues such as volume and formula viscosity (**Fig. 1**).

Access

Orogastric or nasogastric route

Short-term feeding tubes should be used for patients who are expected to receive EN for less than 2 weeks or for whom a percutaneous feeding tube is not an option (Bischoff et al., 2020). Complications of long-term use can include sinus irritation and infection. Orogastric and Nasogastric tube placement are among the most common types of enteral access and can be used to safely

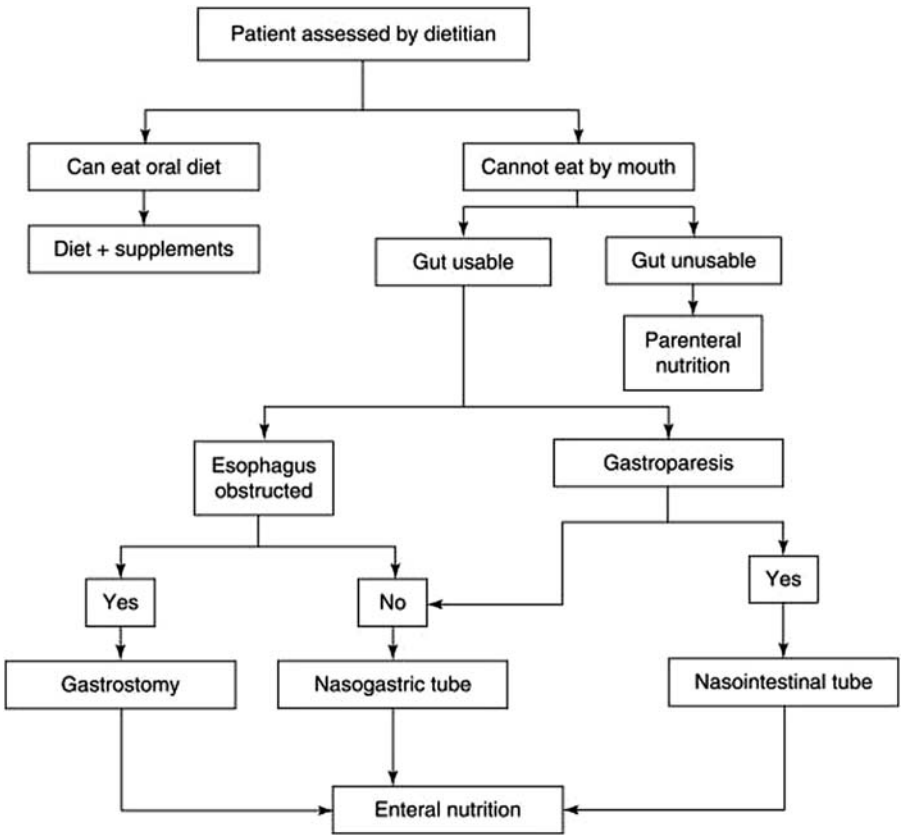


Fig. 1 Algorithm for nutritional support. Reproduced from Encyclopedia of Human Nutrition, Enteral Nutrition, Adults 2005.

administer short-term EN (Pash, 2018). Orogastric and nasogastric tubes passes through mouth and nose, respectively into the stomach and are used when there is adequate GI function. They do not require surgery and can be placed by a nurse or physician at the patient's bedside. Orogastric and nasogastric tubes are made from silicone or polyvinyl, and vary in length from 30 to 43 inches (76–109 cm) with diameters of 8–14 French (~2.5–4.5 mm). Bolus or continuous infusions can be used to administer feedings.

Oroenteric, nasoduodenal or nasojejunal route

Similar to orogastric and nasogastric tubes, although longer, these tubes are also used for short-term feeding (2 weeks) in patients with conditions where the stomach needs to be bypassed for feeding. Such conditions include gastric dysmotility, nausea or vomiting. Feedings via this route require an infusion pump for administration. Tubes can be placed at the bedside or with endoscopic or fluoroscopic guidance (Pash, 2018). If placed at bedside, abdominal radiographs can help confirm proper placement of the tube.

Percutaneous endoscopic gastrostomy (PEG) or percutaneous endoscopic jejunostomy (PEJ)

PEG tubes are used in patients who will need long-term feedings lasting longer than 4 weeks (Bechtold et al., 2016). Requiring moderate to deep sedation, the PEG tube can be placed directly into the stomach through the abdominal wall using an endoscope. Some PEG tubes are designed to allow attachment of a jejunal extension to permit feeding or administration of medications either in the stomach or small bowel. Advantages of PEG tubes include options for long-term use, reasonable cost, and relative ease to maintain. PEG tubes should not be used for feeding patients with small bowel obstructions or moderate risk of aspiration. In patients with small bowel obstruction, PEG tubes are sometimes used as palliative measure to decompress (empty) the stomach, but not for feeding. The tube diameters that are commonly used range from 6 to 8 mm (18–24 French). In general, small-diameter tubes should be avoided in patients who require administration of medication to avoid frequent clogging of the tubes (group SEss et al., 2011).

Percutaneous endoscopic jejunostomy (PEJ) tubes are also used in long-term feeding, in patients who require enteral feedings for more than 4 weeks (Zhu et al., 2012). PEJs can safely feed patients with gastric resection, gastric dysmotility, and patients with a higher risk for aspiration. The PEJ tube is more difficult to insert, requiring a higher skill level, and can be placed percutaneously under endoscopic guidance or surgically. Additionally, because the tubes are typically smaller in size, they may clog more easily. Overall, PEG and PEJ tubes, if maintained properly, can function well with few complications. It should be noted that both feeding gastrostomy and jejunostomy tubes can be placed under radiographic guidance or surgically when the endoscopic method is not clinically appropriate or a physician with endoscopic expertise is not available.

Feeding formulas

A wide selection of enteral feeding formulas are commercially available. Although the composition of formulas varies widely, almost all formulas are nutritionally complete when administered in the appropriate quantities (Hassan-Ghomi et al., 2017). Since 1989, the US Food and Drug Administration has used the term "Medical Foods" to define EN products, meaning they must be used under the supervision of a physician. Formulas are selected based on an assessment of the patient's age, medical condition, nutritional status, and gastrointestinal tract functional ability to absorb and digest nutrients. Nutrition related factors that are also incorporated in the formula selection process include (Tariq et al., 2020): the patient's energy, protein, and fluid requirements (Boullata, 2021), the patient's need for fiber modifications, and (Pironi et al., 2020) the patient's food allergies and intolerances such as gluten sensitivity. The types of formulas are classified below:

Feeding formula classification

1. Polymeric formulas: The most commonly used of the enteral feeding formulas, polymeric formulas can be used orally or through a feeding tube. These formulas contain intact proteins, oligosaccharides, and with fat partly as long-chain triglycerides (LCT) and partly as medium-chain triglycerides (MCT). Lactose and gluten-free, polymeric formulas are also low in osmolality (Muscaritoli and Pradelli, 2021).
2. Elemental formulas: These formulas are used for patients that have malabsorptive conditions, which may include short gut syndrome, Crohn's disease (Sharma et al., 2021), or pancreatitis (Petrov et al., 2009) among other conditions. These formulas require less digestion than polymeric formulas, so they are often used in patients with impaired digestion or absorption. Elemental formulas are amino acid based and are hyperosmolar. Semielemental diets have a mix of peptides and free amino acids, with a lower osmolarity than the elemental diets. They are considerably more expensive than standard formulas.
3. Disease-specific formulas: These formulas are designed to provide specific nutrient needs based on the underlying illnesses. They are intended for patients who have specific metabolic requirements in the context of liver, renal and pulmonary diseases, diabetes, and metabolic disorders (Anand, 2017). They may be enriched with branched-chain amino acids (BCAA), glutamine, omega-3 fatty acids, and arginine whereas others might have a high-fat content. Effectiveness of disease-specific formulas is controversial due to mixed research results.

4. Modular formulas: These formulas are designed for supplemental use. They may add calorie or protein density and can be used to tailor tube feedings to individual nutritional needs as a supplement to commercial enteral formulas. Specific nutrient composition of the patient can be customized. Disadvantages include the risk of bacterial contamination from excessive handling of formulas (Perry et al., 2015).
5. Blenderized foods: Blenderized enteral preparations are whole foods that are mechanically hydrolyzed into a semi-liquid/liquid form and delivered via feeding tube. Although there has been a surge of consumer enthusiasm for commercially available blenderized formulas, historically, some patients have utilized this method of delivering "homemade" formula. Advantages of this type of enteral support is that it more closely mimics food ingestion in the normal state. Disadvantages include greater risk of enteral tube clogging due to the viscosity.

Compositions

Polymeric formulas

These formulas are the standard formulas that are most commonly used for enteral feedings. They contain whole proteins, complex carbohydrates, and mostly LCTs for fat with varying percentage of free water (Hojsak et al., 2020).

Standard formulas provide 1.0 and 1.2–2.0 kcal/mL of formula. This affects the free water provided per milliliter, as a 1.0 formula provides ~85% free water and a 2.0 formula provides ~70% free water.

On average, carbohydrate percentage ranges from 51% to 57% and formulas may include or omit fiber. Polymeric formulas contain a moderate amount of fat with percentages ranging from 29% to 33% to meet essential fatty acid needs and contain a mix of LCT and MCT, with the majority coming from LCT. Sources of fat are of vegetable origin and include corn, canola, soybean, sunflower, and safflower oils.

Standard formulas contain intact protein including casein, egg albumin, and whey protein among others sources. Protein may have the most variability in standard formulas and is one of the deciding factors of which formula will be chosen. Protein percentages often range from 14% to 19%.

Elemental or semielemental formulas

These formulas require less digestion than polymeric formulas because they contain protein and carbohydrates that have been partially or fully hydolyzed and require minimal digestion. They do not contain lactose or fiber and are used for patients that have impaired digestion (Tiengou et al., 2006).

The protein in these formulas are in the form of peptides, fat as LCT or MCT or a combination of both, and carbohydrates as partially hydrolyzed starch maltodextrins to facilitate digestion and absorption.

Elemental formulas provide 1–1.5 kcal/mL. Fat from plant sources such as corn oil, soybean oil and canola oil provide 1–5% of calories and protein from hydrolyzed casein, whey, or soy protein, among other sources, provides 12–20% of total calories.

Disease-specific formulas

Diabetic

Diabetic formulas often contain a mix of soluble and insoluble fibers, with a composition of 31–40% carbohydrate, 42–49% fat, and 17–20% protein content. Lower in carbohydrates and higher in fat, diabetic formulas also contain fiber to help in maintaining glycemic control by slowing gut transit and the absorption of glucose. Consequently the high-fiber and high-fat content of diabetic EN formulations, may cause slow gastric emptying which could be problematic for diabetic patients with gastroparesis (Vallumsetla et al., 2019). It is important to avoid overfeeding diabetic patients to prevent hyperglycemia. In patients with diabetes that is difficult to manage, the initial use of standard formulas with insulin, as needed for glycemic control, is appropriate with the use of diabetic formulas.

Renal

Renal formulas are designed to provide optimal nutrition whilst minimizing blood urea nitrogen, maintaining water and electrolyte balance, and reducing accumulation of toxic products in the body. These formulas also contain a lower electrolyte content as well as lower potassium, phosphorus, and magnesium, and are concentrated to reduce volume for fluid restriction (Rees et al., 2021). They provide 1.8–2 kcal/mL, have less protein at 7–18%, 34–58% carbohydrates, and 35–48% fat.

Renal patients receiving dialysis require an increased amount of protein due to protein loss during dialysis. If a patient has not begun dialysis treatment and a protein restriction is indicated, renal formulas may have too much protein. There are specialized renal formulas available with a very low protein profile, providing only approximately 10% of calories from protein (Rees et al., 2021).

The need for a decreased electrolyte formula will depend on dialysis and subsequent serum electrolyte levels and urinary output. If a patient is receiving continuous renal replacement therapy, there may be no need for a renal formula. Volume status and electrolyte labs will dictate whether this formula is required. These formulas will meet the recommended dietary intake (RDI) for renal patients if the patient receives on average one liter of enteral formula daily.

Hepatic

The disease-specific formulas for liver disease contain higher proportions (40–50%) of the BCAAs (valine, leucine, and isoleucine) and lower levels of aromatic amino acids (AAAs) (tryptophan, tyrosine, and phenylalanine) to counteract a BCAA:AAA imbalance. Patients with hepatic encephalopathy tend to have decreased levels of BCAAs and increased levels of AAAs. An altered ratio may increase AAA transport through the blood–brain-barrier, causing a "false neurotransmitter effect" that could precipitate encephalopathy (Yirui et al., 2021).

Energy density is slightly increased in these enteral formulas containing 1.5–2 kcal/mL of fluid, allowing for full caloric provision in smaller volumes which can help circumvent delayed gastric emptying and complaints of fullness, while meeting fluid restrictions. Recent clinical trials examining the benefit of BCAAs in these patients have been inconclusive.

Pulmonary

The metabolism of macronutrients produces carbon dioxide, with the breakdown of carbohydrate producing the greatest amount. Based on these concepts, an enteral formula was developed with the intent to minimize carbon dioxide production by containing reduced carbohydrate and higher fat content. The fat content in these formulas comprises 50% and is derived from soy or corn oil versus approximately 30% in standard formulas; protein comprises 7–18%, carbohydrates 34–58%, and 1.8–2 kcal/mL (Yirui et al., 2021).

Immune enhancing

Specialized formulas have been designed to improve immune function in the immunocompromised critically ill patient. Specifically patients admitted with traumatic injuries, burns or requiring major abdominal surgeries and at high risk for development of infections and other ICU and postsurgical complications. These formulas may include arginine, glutamine, omega-3 fatty acids, probiotics, nucleic acids, and fiber and often are peptide-based to help maximize absorption (Ding et al., 2020).

Arginine

Arginine is not a traditional "essential" amino acid, but becomes a conditionally essential amino acid during periods of stress, including burns and trauma. It is important for T-cell functioning, collagen synthesis and production of growth hormone, prolactin, somatostatin, insulin, and glucagon (Breuillard et al., 2012). Following trauma and surgery, there is a drop in arginine synthesis. Multiple studies have demonstrated a decrease in postoperative infections and hospital length of stay in patients when using an arginine-containing formula in conjunction with other immune-enhancing ingredients, particularly, omega-3 fatty acids.

Glutamine

Glutamine is the most abundant nonessential amino acid in the body and, like arginine, it becomes a conditionally essential amino acid in states of stress. It is the preferred fuel source for the small bowel enterocyte, which is thought to help maintain its structure and function during times of stress. In septic and malnourished patients, muscle glutamine is depleted, and it is hypothesized that in these patients the availability of glutamine lymphocytes and the gut is reduced, resulting in increased risk of sepsis. Although enteral formulas designed to improve immunity have given mixed results, glutamine supplementation has not been shown to be harmful and has reduced complications in patients with bone marrow transplantation, after surgery, and in those with critical illness and burns (Ma et al., 2018). Studies using parenteral glutamine have generally been more positive than those employing enteral glutamine, although the data are still limited.

Omega-3 fatty acids

The fatty acids found in fish oil, called eicosapentaenoic (EPA) and docosahexaenoic (DHA), are precursors of prostaglandins and thromboxanes that alter the prothrombotic effects of similar compounds derived from linoleic acid. They promote anti-inflammatory cascade under stress conditions and may be beneficial in patients with acute respiratory distress syndrome and severe acute lung injury.

Probiotics

The administration of a probiotic with EN may have a role in reducing septic complications in patients with transplantation, major abdominal surgery, and severe trauma. Differences in species of probiotics may have different effects of variable impact on patient outcomes. One report of adverse outcomes in a randomized trial in patients with severe pancreatitis has generated some concern about the broad usage of probiotics in critically ill patients (Xie et al., 2018).

Modular

These formulas can provide macronutrients and micronutrients separately and can be used to provide additional calories or protein. Typically modular formulas contain only one or two macronutrients and are used to enhance other formulas. Protein modular products are powders and must be mixed with water before administration. Carbohydrates come in the form of glucose polymers and fat as triglycerides of long-chain polyunsaturated or medium-chain fatty acids. Some modular formulas can be combined with liquid vitamins and minerals for individualization for patients with unique needs. Normally, modulars are added to meet nutritional needs in patients who have disproportionate requirements.

Indications and contradictions

EN is the preferred method of feeding patients who cannot eat, absorb, or use a normal diet in the presence of a useable GI tract. The following are indications for EN (Bischoff et al., 2020; Saarnio et al., 2014):

1. Critical care patients, including those with trauma and burns, and also after major surgery.
2. Anorexia in patients with malignant disease, sepsis, liver and renal failure, and inflammatory bowel disease (IBD).
3. Upper GI obstruction or ulceration of the pharynx, esophagus, stomach, and duodenum may prevent the ingestion of normal food. Examples of these conditions are cancer, central nervous system disorders, and stenosis following ulceration.
4. Short bowel and severe malabsorption: In controlled trials, enteral diets are not better absorbed than normal food. Therefore, the presence of a short bowel *per se* is not an indication for enteral feeding. However, some patients with severe malabsorption may benefit from the use of elemental diets.
5. IBD: In IBD, enteral feeding is used in the following situations:
   i. Profound anorexia preventing the ingestion of a normal diet.
   ii. Abdominal discomfort due to partial bowel obstruction or intestinal inflammation.
   iii. In Crohn's disease, controlled trials have suggested that enteral feeding can induce a remission comparable to that seen with steroids.
6. Dementia: Patients unable to feed themselves because of mental changes.
7. Fistulas of the distal small bowel or colon.
8. Nausea or vomiting. Patients who suffer from nausea or vomiting due to a gastric disorder can be fed into the jejunum.
9. Recurrent aspiration. Formula should be delivered through a jejunostomy.
10. High nutritional requirements that are not being met by oral intake (mostly applying to patients with burn injury who require increased nutrient intake).
11. Cancers of the head and neck.
12. Dysphagia associated with disabling neurologic conditions including amyotrophic lateral sclerosis (ALS) and multiple sclerosis (MS).
13. Neoplasms: Advanced primary and secondary intracranial tumors.
14. Coma.
15. Severe head injuries.
16. Cerebrovascular accidents.
17. Transition from parenteral nutrition.

The following are contradictions for EN (Bischoff et al., 2020):

1. Mechanical obstruction that cannot be bypassed.
2. Intractable vomiting or diarrhea refractory to medical management.
3. Extreme malabsorption.
4. Short bowel syndrome (100 cm small bowel remaining).
5. Paralytic ileus.
6. High output fistula.
7. Peritonitis.
8. Mild GI bleeding.
9. Hemodynamic instability.
10. Inability to access GI tract.
11. Imminently terminal disease.
12. Aggressive interventions not desired.

Methods of infusion

Enteral feedings can be administered by continuous drip, bolus feedings, or intermittent infusions. The method selected depends on the stability of the patient, gastric emptying rate, caloric and protein needs, patient mobility and central access route (Deane and Chapman, 2021).

Continuous drip – Tube feedings are administered for 24 h with minimal interruptions. An advantage of this type of feeding is that it may be tolerated well in critically ill patients. A disadvantage is that an infusion pump is usually required, to ensure accuracy of volume is delivered. Initiation typically begins at 25–50 mL/h and is increased by 50 mL/h every 4–8 h (McNelly et al., 2020).

Intermittent infusions – Larger volumes (240–480 mL) are administered 3–6 times per day infused over longer periods (30–60 min). This type of feeding may result in more adequate volumes being administered. Another advantage is that it allows some time off of feedings and the feedings are usually given by a gravity drip over period of 30 min to 1 h. The gravity feeding system is less expensive than a pump. A disadvantage may be intolerance to the higher rates of administration. Initiation of feeding begins with 1/2–1 can per feeding increased by 1/2 can per feeding per day until the goal is met (McNelly et al., 2020).

Bolus feed –Bolus feeding is the preferred method because it allows the patient more freedom, is easier to use, takes less time, and is cheaper. The bolus method does not require a pump and can be given in larger volumes (240–480 mL) administered 3–6 times per day over short periods of time (10–15 min). Rapid installation of feeding into GI tract can be given by syringe or funnel. The majority of patients tolerate this method. To avoid aspiration the bolus feeding must be given when the patient is sitting or reclining at 45°. Initiation of feeding begins with 1/2–1 can per feeding and can be increased by 1/2 can per feeding per day until the goal is met (Hubbard et al., 2019).

EN is viewed as the safest and most efficacious method to support nutritional status in patients who are unable to eat orally. When administered properly, enteral feeding is associated with improved clinical outcomes and reduced infectious complications and is the preferred form of nutritional support.

Potential challenges

Enteral nutrition has several favorable indications, although there are several challenges that may limit its use. For individuals who consume enteral nutrition by mouth, its long-term use may be limited by poor palatability. On a broader scale, there would be concerns of availability and costs, particularly for specialty formulas and supplies, in several regions of the world. Additional contemporary challenges include supply chain issues brought on by the coronavirus disease 19 (COVID-19) pandemic and increased costs related to inflation. These practical barriers would need to be considered when recommending the use of enteral nutrition therapy.

Summary/conclusions

Enteral nutrition provides essential nutrients and can be a potentially lifesaving therapy for individuals who are unable to sustain caloric intake per oral route. Although minor complications may result from this form of therapy, if patients are chosen appropriately and close attention to components of the enteral formulation are taken, these patients would benefit greatly. Management of enteral nutrition is most effective when done through a multidisciplinary approach, utilizing the expertise of physicians, pharmacists, dieticians, nurses, case managers, and social workers.

References

Anand, A.C., 2017. Nutrition and muscle in Cirrhosis. J. Clin. Exp. Hepatol. 7, 340–357.
Bechtold, M.L., Mir, F.A., Boumitri, C., et al., 2016. Long-term nutrition: a Clinician's guide to successful long-term enteral access in Adults. Nutr. Clin. Pract. 31, 737–747.
Bischoff, S.C., Austin, P., Boeykens, K., et al., 2020. ESPEN guideline on home enteral nutrition. Clin. Nutr. 39, 5–22.
Boullata, J.I., 2021. Enteral medication for the tube-fed patient: making this route safe and effective. Nutr. Clin. Pract. 36, 111–132.
Breuillard, C., Darquy, S., Curis, E., et al., 2012. Effects of a diabetes-specific enteral nutrition on nutritional and immune status of diabetic, obese, and endotoxemic rats: interest of a graded arginine supply. Crit. Care Med. 40, 2423–2430.
Deane, A.M., Chapman, M.J., et al., 2. Technology to inform the delivery of enteral nutrition in the intensive care unit. JPEN J. Parenter. Enteral. Nutr. 42 (5), 949–959. https://doi.org/10.1002/jpen.1038.
Ding, H., Xu, J., You, J., et al., 2020. Effects of enteral nutrition support combined with enhanced recovery after surgery on the nutritional status, immune function, and prognosis of patients with esophageal cancer after Ivor-Lewis operation. J. Thorac. Dis. 12, 7337–7345.
group SEss, Pedron Giner, C., Martinez-Costa, C., et al., 2011. Consensus on paediatric enteral nutrition access: a document approved by SENPE/SEGHNP/ANECIPN/SECP. Nutr. Hosp. 26, 1–15.
Hassan-Ghomi, M., Nikooyeh, B., Motamed, S., et al., 2017. Efficacy of commercial formulas in comparison with home-made formulas for enteral feeding: a critical review. Med. J. Islam. Repub. Iran 31, 55.
Hojsak, I., Matic, K., Sila, S., et al., 2020. Characteristics of polymeric formula and route of delivery of exclusive enteral nutrition have no effect on disease outcome and weight gain in pediatric Crohn's disease. Clin Nutr 39, 1108–1111.
Hubbard, G.P., Andrews, S., White, S., et al., 2019. A survey of bolus tube feeding prevalence and practice in adult patients requiring home enteral tube feeding. Br. J. Nutr. 122, 1271–1278.
Ma, C., Tsai, H., Su, W., et al., 2018. Combination of arginine, glutamine, and omega-3 fatty acid supplements for perioperative enteral nutrition in surgical patients with gastric adenocarcinoma or gastrointestinal stromal tumor (GIST): a prospective, randomized, double-blind study. J. Postgrad. Med. 64, 155–163.
McNelly, A.S., Bear, D.E., Connolly, B.A., et al., 2020. Effect of intermittent or continuous feed on muscle wasting in critical illness: a Phase 2 clinical trial. Chest 158, 183–194.
Muscaritoli, M., Pradelli, L., 2021. Medium-chain triglyceride (MCT) content of adult enteral tube feeding formulas and clinical outcomes. A systematic review. Front. Nutr. 8, 697529.
Pash, E., 2018. Enteral nutrition: options for short-term access. Nutr. Clin. Pract. 33, 170–176.

Perry, J., Stankorb, S.M., Salgueiro, M., 2015. Microbial contamination of enteral feeding products in thermoneutral and hyperthermal ICU environments. Nutr. Clin. Pract. 30, 128–133.
Petrov, M.S., Loveday, B.P., Pylypchuk, R.D., et al., 2009. Systematic review and meta-analysis of enteral nutrition formulations in acute pancreatitis. Br. J. Surg. 96, 1243–1252.
Pironi, L., Boeykens, K., Bozzetti, F., et al., 2020. ESPEN guideline on home parenteral nutrition. Clin. Nutr. 39, 1645–1666.
Rees, L., Shaw, V., Qizalbash, L., et al., 2021. Delivery of a nutritional prescription by enteral tube feeding in children with chronic kidney disease stages 2-5 and on dialysis-clinical practice recommendations from the Pediatric Renal Nutrition Taskforce. Pediatr. Nephrol. 36, 187–204.
Saarnio, J., Pohju, A., Ahtola, H., 2014. Indications and execution of enteral nutrition. Duodecim 130, 2239–2244.
Sharma, S., Gupta, A., Kedia, S., et al., 2021. Efficacy and tolerability of exclusive enteral nutrition in adult patients with complicated Crohn's disease. Intest. Res. 19, 291–300.
Tariq, N., Ali, A., Chen, C., 2020. Endoscopic enteral access. Surg. Clin. 100, 1091–1113.
Tiengou, L.E., Gloro, R., Pouzoulet, J., et al., 2006. Semi-elemental formula or polymeric formula: is there a better choice for enteral nutrition in acute pancreatitis? Randomized comparative study. JPEN J Parenter Enteral Nutr 30, 1–5.
Vallumsetla, N., Epp, L., Hurt, R.T., et al., 2019. Effect of home enteral nutrition on diabetes and its management. Nutr. Clin. Pract. 34, 250–256.
Xie, H., Lu, Q., Wang, H., et al., 2018. Effects of probiotics combined with enteral nutrition on immune function and inflammatory response in postoperative patients with gastric cancer. J BUON 23, 678–683.
Yirui, L., Yin, W., Juan, L., et al., 2021. The clinical effect of early enteral nutrition in liver-transplanted patients: a systematic review and meta-analysis. Clin. Res. Hepatol Gastroenterol 45, 101594.
Zhu, Y., Shi, L., Tang, H., et al., 2012. Current considerations of direct percutaneous endoscopic jejunostomy. Can. J. Gastroenterol. 26, 92–96.

Further reading

DeBruyne, L.K., Pinna, K., Whitney, E., 2011. Nutrition and Diet Therapy: Principles and Practice. Thomson Wadsworth, Belmont, CA.
Garcia-de-Lorenzo, A., Zarazaga, A., Garcia-Luna, P.P., et al., 2003. Clinical evidence for enteral nutritional support with glutamine. Syst. Rev. Nutr. 19 (9), 805–811.
Lochs, H., Allison, S.P., Meier, R., et al., 2006. Introductory to the ESPEN guidelines on enteral nutrition: terminology, definitions and general topics. Clin. Nutr. 25 (2), 180–186.
Mahadeva, S., Malik, A., Hilmi, I., Qua, C.S., Wong, C.H., Goh, K.L., 2008. Transnasal endoscopic placement of nasoenteric feeding tubes: outcomes and limitations in non-critically ill patients. Nutr. Clin. Pract. 23 (2), 176–181.
Mahan, L.M., Raymond, J., 2017. Krause's Food, Nutrition and Diet Therapy, fourteenth ed. WB Saunders Co, Philadelphia, PA.
Markides, G.A., Al-Khaffaf, B., Vickers, J., 2011. Nutritional access routes following oesophagectomy – a systemic review. Eur. J. Clin. Nutr. 65, 565–573.
Meier, R., Ockenga, J., Pertkiewicz, M., Pap, A., Milinic, N., MacFie, J., Arvanitakis, M., et al., 2020. ESPEN guidelines on clinical nutrition in acute and chonic pancreatitis. Clin. Nutr. 39 (3), 612–631.
Enteral nutrition: options for short-term access. Pash E. Nutr. Clin. Pract. 33 (2), 2018 Apr, 170–176. https://doi.org/10.1002/ncp.10007. Epub 2018 Feb 10.PMID: 29427560 Review.
Rafferty, G.P., Tham, T.C., 2010. Endoscopic placement of enteral feeding tubes. World J. Gastrointest. Endosc. 2 (5), 155–164.

Nutritional support: In the home setting

M. Elia and T.R. Smith, University of Southampton, Southampton, UK

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of M. Elia, T.R. Smith, Nutritional Support: In the Home Setting, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 269–277, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00201-4.

Oral Nutritional Support **565**
Dietary Counseling and Fortification 565
Oral Nutritional Supplements 566
Artificial Nutritional Support: Home Parenteral Nutrition and Home Enteral Tube Feeding **566**
Origins and Development 566
Indications 567
Home Enteral Tube Feeding 567
Home Parenteral Nutrition 568
Organization 568
Standards of Care 568
Monitoring 568
Outcome 570
Intestinal Transplantation 572
Ethical Issues 572
Conclusions **572**
Further Reading **573**

The prevalence of nutritional problems in developed societies is a cause of growing concern. At one end of the nutritional spectrum, the obesity 'epidemic' is spreading at an alarming rate. At the other end of the spectrum, protein–energy malnutrition and nutrient deficiencies are also common, especially in the elderly and in those with disease. Table 1 shows the frequency of specific vitamin deficiencies and underweight (body mass index<20 kg m^{-2}) in people aged 65 years or older, residing in the United Kingdom. Complimentary information on protein–energy status can be obtained by considering simple criteria, such as those used by the 'Malnutrition Universal Screening Tool' (MUST) (Figure 1). This tool, which depends on weight loss and body mass index (and an acute disease effect, which does not normally apply to community patients), has been used to help estimate the distribution of malnutrition in the United Kingdom. Approximately 2% of individuals with medium and high risk of malnutrition are in hospital, 5% in care homes, and the remaining 93% in the community, of which 2–3% live in sheltered housing. Similar distributions are thought to exist in other countries although in low-income countries, the proportion of malnourished people in institutions is expected to be even less than in more developed countries. National surveys in various countries have demonstrated that malnutrition is common on admission to institutions, for example, approximately 28% among those admitted to hospitals in the United Kingdom (MUST criteria). This means that national strategies to combat malnutrition must consider its origins and causes in the community and attempt to prevent them at an early stage. However, because there are so many people who are discharged from the hospital in a malnourished state, often more malnourished than on admission, there should be an opportunity to identify them before discharge from the hospital and initiate treatment there, which can continue in the community. The same principles apply to individuals attending hospital outpatient clinics. The use of a consistent framework for identifying and treating malnutrition within

Table 1 Proportion of subjects aged 65 years or older with selected vitamin deficiencies and body mass index <20 kg m^{-2}

Free living (%)		*Institutions (%)*[a]	*Criteria*	
Vitamin deficiencies				
Folate deficiency	29	35	Red blood cell concentration	<345 mmol l^{-1}
Severe deficiency	8	16		<230 mmol l^{-1}
Thiamine deficiency	9	14	Erythrocyte transketolase activation coefficient (ratio)	>1.25
Vitamin B_{12} deficiency	6	9	Plasma concentration	<118 pmol l^{-1}
Vitamin D deficiency	1–2	1–5		<12 mmol l^{-1}
Vitamin C deficiency	14	40	Plasma concentration	<11 mmol l^{-1}
Severe deficiency	5	16		<5 mmol l^{-1}
Underweight	3	16	Body mass index	<20 kg m^{-2}

Source: Based on the National Dietary and Nutrition Survey (1998) in the United Kingdom.
[a]Registered residential homes (57%), nursing homes (30%), dual-registration homes (9%), and other facilities (4%).

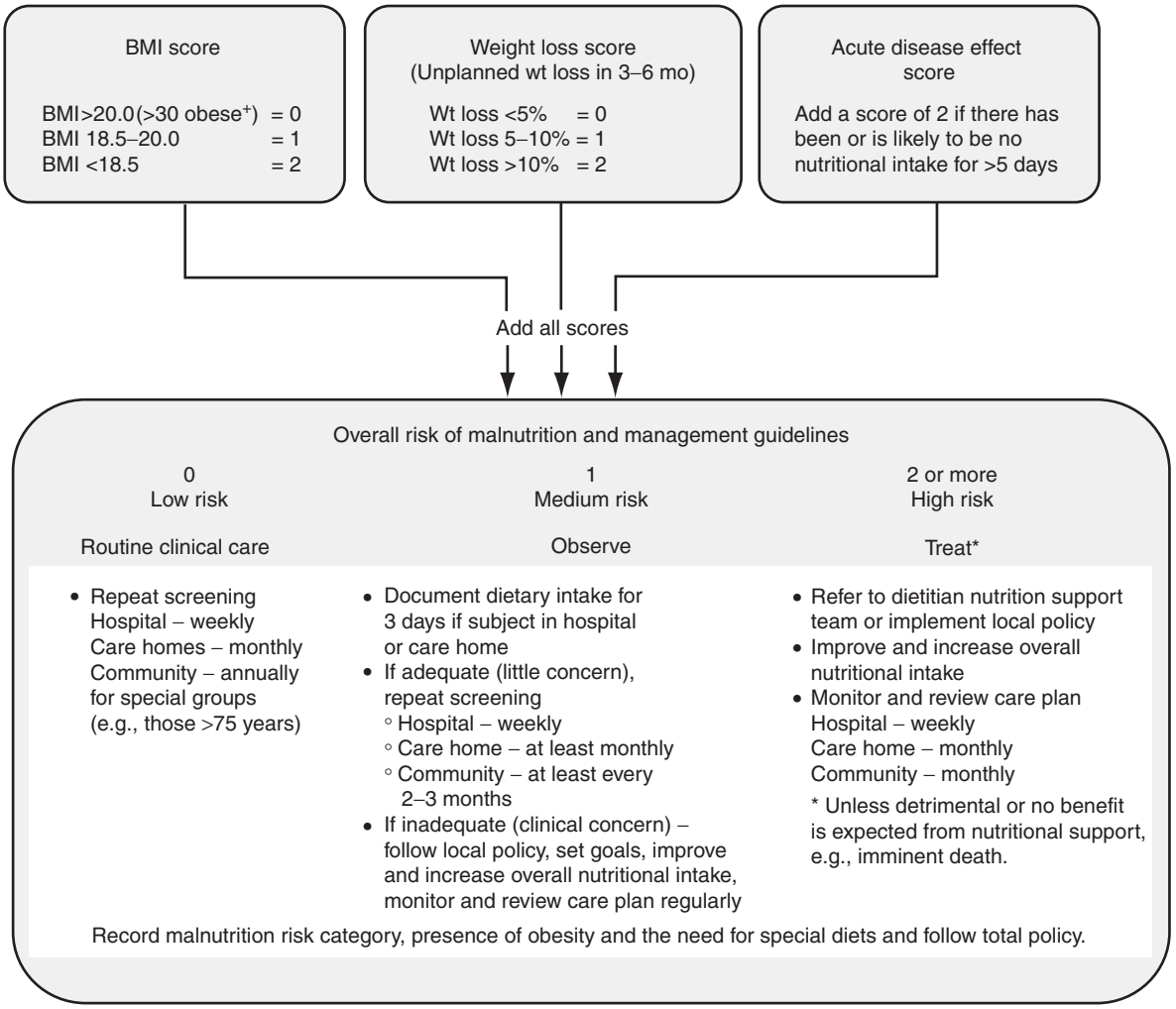


Figure 1 Malnutrition Universal Screening Tool (MUST). A copy of MUST and further details on taking alternative measurements, special circumstances, and subjective criteria can be downloaded at www.bapen.org.uk

and between care settings is important in facilitating continuity of care during the patient journey from one setting to another. Unlike other screening tools, MUST was developed for this specific purpose and adopted nationally in a number of countries with this in mind.

This article focuses on the treatment of malnutrition (rather than obesity) in the home setting. This treatment includes dietary counseling and fortification, oral nutritional supplementation (mixed macro- and micronutrient supplements), and artificial nutritional support (enteral tube feeding (ETF) and parenteral nutrition (PN)). The simplest and most commonly used treatment involves oral nutritional support, which is considered before home enteral tube feeding (HETF) and home parenteral nutrition (HPN).

Oral Nutritional Support

Dietary Counseling and Fortification

Dietary counseling, usually provided by a dietitian, is an integral part of oral nutritional support. It includes advice on dietary fortification, which is often the first-line treatment of malnutrition in the home and other care settings. Counseling may involve advice on eating patterns (e.g., eating certain types of snacks at particular times of day) or addition of energy- and protein-rich food ingredients (e.g., cream, milk, oil, butter, sugar, and skimmed milk powder) to meals. Commercial energy- and protein-containing supplements can also be used to improve intake without substantially altering the intake from normal food and drink. The use of nutritionally fortified food snacks as part of the diet may improve both the intake and the status of micronutrients. However, the success of these dietary strategies is limited in patients with severe anorexia, those living in poverty and due to other social factors, and in those with inadequate motivation. Thus, patients may find it difficult to purchase, manipulate, or prepare their meals.

Financial or other forms of social support, such as help with shopping, cooking (or provision of 'meals on wheels'), and help with eating, may do much to improve intake in some individuals. Although dietary counseling, with or without dietary fortification, is widely used in clinical practice, there is little research supporting its clinical efficacy in patients at risk of malnutrition in developed countries.

Oral Nutritional Supplements

Mixed macro- and micronutrient liquid sip feeds and other oral nutritional supplements (bars, powders, and puddings) are widely used in the treatment of malnutrition in the community setting. A systematic review of 78 randomized controlled trials (RCTs) (including 44 RCTs from the community setting) suggests that oral nutritional supplements can improve energy and nutrient intakes, improve body weight (or attenuate weight loss), and improve a number of functional and clinical outcomes in various patient groups (Table 2). Meta-analyses of RCTs involving hospital and community settings suggest significantly lower mortality and complications in favor of oral nutritional supplements (typically 1.05–2.5 MJ day^{-1} (250–600 kcal) daily). The evidence base in care homes is more limited, but a recent systematic review suggests that oral nutritional supplements can produce clinical or functional benefits, such as improved healing of pressure ulcers and fewer infections.

For some patients, nutrition via the oral route is either unable to meet the nutritional requirements (e.g., patients with a poor appetite) or contraindicated (e.g., a cerebrovascular accident patient with aspiration and intestinal failure). For such patients, HETF and HPN may be required, although the treatment is usually initiated in hospital.

Artificial Nutritional Support: Home Parenteral Nutrition and Home Enteral Tube Feeding

Patients suffering from chronic conditions often prefer to be treated in the familiar surroundings of their home rather than in hospital. When the treatment involves sophisticated techniques, it is essential that either the patient or the caregiver is adequately trained to distinguish between problems that can be easily remedied at home and those that need expert advice and treatment in hospital. With the increasing pressure for hospital beds and the increasing cost of hospital care, many forms of treatment that were previously restricted to the hospital environment have extended to the community, including renal dialysis, cytotoxic drug therapy, HETF, and HPN. HETF has grown rapidly so that its prevalence in several developed countries is now several times greater than in hospital. In contrast, PN is still practiced less commonly outside hospital than in hospital and is likely to remain so in the foreseeable future. Both forms of treatment have led to the development of professional teams specializing in nutritional support in both the hospital and the community. These teams deal with problems ranging from simple day-to-day management issues to difficult ethical problems, such as concerning withholding or withdrawing nutritional support.

Origins and Development

The first report of HPN appeared in 1970 in North America and in the late 1970s in Europe. The number of people receiving HPN has increased considerably since then but remains substantially lower than that for HETF (Figure 2).

Table 2 Summary of significant functional and clinical outcome improvements following oral nutritional supplementation in community patients from randomized controlled trials

Patient group	*Functional/clinical outcome*
Chronic obstructive pulmonary disease	Respiratory muscle function Hand grip strength Walking distances
Elderly	Reduced number of falls Increased activities of daily living Muscle power
HIV/AIDS	Cognitive function
Liver disease	Lower incidence of severe infections Lower frequency of hospitalization
Malignancy	Immunological benefits
Osteoarthritis	Increased activities of daily living[a] Improved osteoarthritis index[a]

Source: Reproduced from Stratton RJ, Green CJ, and Elia M (2003) *Disease-Related Malnutrition: An Evidence-Based Approach to Treatment*. Oxford: CABI Publishing.
[a]Nutritional supplement also containing immunoglobulin G (90 mg).

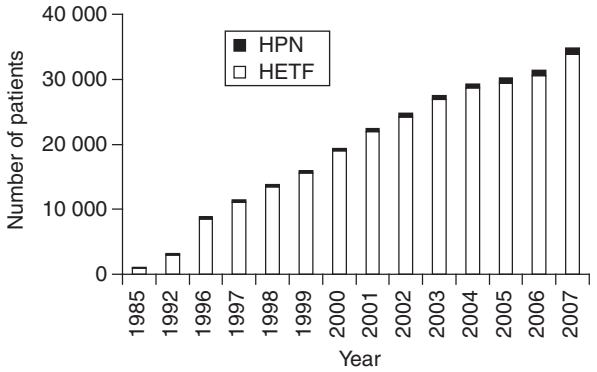


Figure 2 Estimated growth in point prevalence (amount of feeding taking place at a given point in time) in home enteral tube feeding (HETF) and home parenteral nutrition (HPN) in the United Kingdom.

HETF is a much older technique than HPN, with the first reports appearing centuries ago. Accurate information on the number of people receiving HETF is difficult to obtain, because HETF tends to be initiated from many centers, and centralized reporting and record keeping in most countries are not fully established. There has been rapid growth in HETF attributable to developments in tube technology (flexible fine-bore tubes) and endoscopic procedures for placement of gastrostomy tubes (facilitating easier initiation and management of long-term feeding), as well as the development of home care services provided by commercial enteral feeding companies. In many developed countries, there is considerably more ETF taking place in the community than in the hospital. In the United Kingdom, there has been a steady growth in the number of people receiving HETF, and in 2007, it was estimated that there were approximately 30 000 people receiving this treatment at a given point in time (point prevalence). As with HPN, HETF is less common in Europe than in North America and is practiced substantially less in Eastern Europe, India, and China than in high-income industrialized countries in the West.

In addition to differences in the prevalence of HETF and HPN between countries, there are also differences in practice of HETF and HPN within countries, which are unlikely to be due to chance. For example, the point prevalence of HPN in the United Kingdom varied from 3.7 to 22.5 patients per million of the population in 2006. This is likely to be due to availability of expertise and support staff, resources to fund treatment, reimbursement policies, and attitudes/policies toward the use of artificial nutrition.

The wide variations in the prevalence of home artificial feeding throughout the world are related to health-care economies. There is a relationship between expenditure on health care, as a percentage of gross domestic product (GDP), and the incidence of HPN and HETF. In India, Pakistan, and Africa, where spending on health is low, home artificial nutrition is less common. In Western Europe, where health care accounts for a greater proportion of GDP, home artificial nutritional support is more common. In the United States, with an even greater expenditure on health care, the prevalence of HPN and HETF is higher than anywhere else in the world.

Indications

Home Enteral Tube Feeding

The indications for HETF are different for adults and children. In adults, the most common indications are neurological disorders of swallowing resulting from cerebrovascular accidents, Parkinson disease, and obstructive lesions of the upper gastrointestinal tract. These mainly affect older individuals so that in various countries approximately half of HETF is administered to individuals aged 65 years or older. In children, HETF is usually used in conditions that lead to failure to thrive, such as cerebral palsy, cystic fibrosis, congenital malformation, and metabolic disorders. As with HPN, one of the main differences in the indications for HETF between countries concerns malignant disease. In North America, ~40% of people receiving HETF have been reported to have malignant disease and more than 50% in Italy. In the United Kingdom, as in many other countries, the proportion of patients receiving HETF with cancer has steadily grown over time. The main malignancies in the United Kingdom are head and neck tumors and malignancies of the upper gastrointestinal tract (mainly obstructive oropharyngeal and esophageal cancers). The age distribution of people receiving HETF is influenced by the indications. Because disorders of swallowing (due to strokes, motor neurone disease, and other neurological conditions) and cancer of the upper gastrointestinal tract tend to occur in older age groups, adults receiving HETF tend to be elderly (with more than 65% of those in the United Kingdom being older than 60 years, 43% older than 70 years, and 19% older than 80 years in 2007). In recent years, there has been a trend to provide HETF to an older and more disabled population. Recent surveys in the United Kingdom suggest that approximately one-third of patients are house or bed bound and nearly half require total help to manage their tube feeding. Because the majority of these patients with high levels of disability are at home (spending <1% of their time in hospital), there are resource implications associated with the provision of health care by the unrecognized and underappreciated voluntary caregivers. Approximately 20% of those receiving HETF are children, and many children who started HETF because of cerebral palsy or congenital handicap continue tube feeding into adulthood.

Home Parenteral Nutrition

The main indications for HPN are Crohn disease, mesenteric vascular ischemia, motility disorders, surgical complications (e.g., enterocutaneous fistulae), and malignant disease. Patients receiving HPN are usually younger than those who receive HETF, although there is an overlap. There are also differences between the practices of HPN in different countries. One of the main differences concerns malignant disease. In the United States, 40–50% of patients receiving HPN have been reported to have cancer, and similar, if not higher, percentages have been reported in some European countries, such as Italy. Early reports from the United Kingdom and Denmark suggested that only a small proportion of HPN (~5%) involved patients with cancer, although this has increased with time. For example, in the United Kingdom, it has steadily increased so that by 2007 more than 15% of patients starting HPN had cancer. There has also been an increase in age (due at least partly to the increasing use of HPN in patients with cancer and mesenteric vascular ischemia) so that by 2007 nearly one-third of all patients receiving HPN in the United Kingdom were older than 60 years.

Organization

The organization and management of HETF and HPN have evolved over time. For example, delivery of feeds and equipment to the first patients who received HPN or HETF was undertaken by the hospitals that initiated the treatment. As the number of patients receiving such treatment increased, commercial organizations have established an organizational infrastructure for delivering feed and ancillary equipment through a national and international network. Some companies employ doctors, nurses, and other staff so that they can provide most of the care, although this practice varies from country to country. In many countries, there is joint care between commercial companies and the national health-care systems.

HETF is initiated by many centers or hospitals, and some patients are followed up as outpatients. However, it is impractical to follow up many severely disabled patients in hospital, because they are house bound. Patients receiving HPN are often managed by centers with expertise in nutritional problems (e.g., in France, Denmark, and the United Kingdom). It has been suggested that all patients on HPN should be managed at such centers, but traveling to distant centers may require considerable time, effort, and expense. It is possible for patients to be managed more locally, especially if they are uncomplicated. It remains to be demonstrated if locally managed patients have better satisfaction and similar outcomes as those managed by larger centers. Of course, it is possible to have a system that combines local care and more distant specialist care when required.

Funding arrangements also vary. In several countries, home nutritional support is either totally or partially funded by the National Health Service, but payment may also be provided by private insurance and individual patients. Sometimes, confusion exists about the funding arrangements even in the same country, and this may limit and delay the use of HETF or HPN.

Patient organizations have developed in some countries, such as Patients on Intravenous and Nasogastric Nutrition Therapy (PINNT) in the United Kingdom. This organization provides support and information to people on home feeding, and it contributes to all levels of the operation of the British Association for Parenteral and Enteral Nutrition (BAPEN), through which it influences policy and decision making. Furthermore, as the feeding equipment for use at home was found to be impractical because it was originally designed for hospital use, PINNT has redesigned the equipment specifically for home use.

Standards of Care

Several surveys have identified inadequacies in training, support, and follow-up of patients receiving HETF and HPN. Specific problems include lack of written instructions about how to manage simple problems that may arise during feeding, lack of telephone contacts for use in emergency, lack of confidence, and inadequacy of equipment for home use. Such surveys have also highlighted the importance of a multidisciplinary approach and the need to undertake home visits to assess the status of severely disabled patients who cannot easily attend a hospital. Pressure on hospital beds has meant that some patients are discharged home before they have been adequately trained, and the care of such patients is sometimes passed on to other health-care workers who have little experience of home nutritional support. Because HPN is relatively uncommon in the population, general practitioners may have never encountered patients on this form of therapy and are therefore poorly equipped to manage them. The needs of patients may change during the course of their treatment; therefore, there is a need to establish an organizational infrastructure for continuity of care for HETF and HPN over time and from one health-care setting to another. Many hospitals do not have a nutrition team or policies that embrace the needs of people receiving artificial nutrition at home.

A series of guidelines for the management of artificial nutrition in the community have been developed by BAPEN (**Tables 3 and 4**). The guidelines cover aspects of training before discharge from hospital (although training can take place at home) and the support required from trained specialist staff once the patient is at home. A national and local organizational structure for delivering the support would aid the process.

Monitoring

The basic elements of monitoring are similar for both HETF and HPN. They include an assessment of the activity of the underlying disease, the nutritional and metabolic state of the patient, and complications associated with nutritional support (**Table 5**). The clinical history alerts the attending health professional to the general well-being, as well as the likelihood of specific problems,

Table 3 Standards of practice for home enteral tube feeding (HETF)

Structure	*Process*	*Outcome*
There will be a training program for the health-care professionals involved in the care of patients receiving HETF.	Discharge planning will be performed only by professionals who have the necessary experience or who have undertaken a course of training in the topic.	The patient has confidence in the hospital team planning his or her discharge.
There will be a model of care for patients needing HETF.	The members of the multidisciplinary team will be involved in writing the 'mission statement' on which the model is based.	The patient will know the benefits, aims, and objectives of the HETF team.
There will be a relaxed, quiet area suitable for private discussion.	There will be a caring and compassionate atmosphere with adequate time for discussion.	The patient will feel able to express his or her fears and expectations.
The discharge planning documentation will include sections on domestic, family, and social circumstances.	The nutrition team will evaluate, with the patient and family, how HETF will alter his or her way of life.	The patient will believe that the feeding system can be integrated into an acceptable way of life.
There will be written patient/carer learning goals of HETF.	A designated nurse or dietitian will be responsible for teaching the patient according to his or her individual capacity for learning.	The patient will be able to demonstrate the necessary skills and achieve all the learning goals.
There will be an instruction manual for HETF.	Information and procedures will be regularly updated in order to reflect developments and innovations in tube feeding, access, nutrients, and delivery systems.	The patient will perform therapy based on current practice.
A relative, friend, or appropriately trained health-care professional will be available to deliver therapy if the patient is unable to do so.	The nurse/dietitian will help the patient identify the most appropriate carer. A community nurse will be given the opportunity to visit the patient in hospital and observe therapy before the patient is discharged.	The patient has confidence that safe care will be available at home.
Access to the gastrointestinal tract will be achieved by a tube suitable for long-term use.	The patient, nurse, and doctor will choose the most appropriate tube and access site.	The patient will use a feeding tube that is acceptable and accessible.
There will be a policy for sharing care with the patient's general practitioner (GP).	The GP will be contacted and a shared care protocol agreed.	The patient will know the responsibility of each health-care professional.
Written information describing HETF will be available for the GP.	The hospital team will provide the GP with the information before the patient is discharged, together with the discharge date and on-call telephone numbers.	The patient will have confidence in his or her GP's knowledge of HETF.
There will be written procedures for the management of feeding tubes.	The nurse/dietitian will adapt the procedures according to the patient's physical skills and domestic circumstances.	The patient's daily life will not be restricted by prolonged inappropriate procedures.
There will be a written prescription for the enteral feed (and other prescribable items).	The patients's GP will be contacted and advised on how to prescribe the feed.	The patient will have the enteral feed available at home on the day of discharge.
There will be an on-call system for providing expert advice to the patient by telephone day and night.	The nurse/dietitian/doctor will explain the system to the patient and identify the professions involved.	The patient will know the names and telephone numbers of health-care professionals to contact in case of emergency by day or night.
Information will be available describing how the nutrient solutions and supplies will be provided following discharge.	The nurse/dietitian will explain the ordering system and discuss storage, depending on the patient's home circumstances.	The patient will know how to obtain supplies and store them, and dispose of unwanted material.
There will be a postdischarge monitoring protocol, established by the nutrition team.	Monitoring will be performed by a designated health professional as defined by the protocol.	The patient will know what the follow-up arrangements are.

such as dehydration, electrolyte imbalance (e.g., diarrhea), local infection (e.g., local redness and swelling near the catheter exit site or peristomal area), blocked tubes and catheters, and so on. Catheter-related sepsis is an important complication of PN, and aspiration pneumonia is an important complication of ETF. The patient/caregiver should have written instructions about basic procedures, which aim to reduce complication rates, and how to deal with simple problems and to recognize those that they cannot readily deal with. Specialist advice should be available 24 h a day. The frequency of complications depends at least partly on the support provided by health professionals.

Dietary intake should be monitored, especially in patients whose clinical status is changing. Appropriate dietary advice may facilitate return to normal oral feeding in some patients. In those with a swallowing difficulty, it may be necessary to assess whether swallowing has improved, with input from speech and language therapists, so that unnecessary HETF is not continued when full oral feeding becomes possible. Studies in the United Kingdom suggest that 15% of patients receiving HETF can revert to full oral feeding after 1 year. Blood tests should be carried out at intervals to check for metabolic stability and specific nutrient deficiencies (e.g., vitamins, minerals, and trace elements) and toxicities. The frequency with which tests are carried out depends on

Table 4 Standards of practice for home parenteral nutrition (HPN)

Structure	*Process*	*Outcome*
There will be a training program for health-care professionals involved in the care of patients receiving HPN.	Discharge planning will be performed only by professionals who have the necessary experience or who have undertaken a course of training in the topic.	The patient has confidence in the hospital team planning his or her discharge.
There will be a mode of care for patients needing home intravenous nutrition.	All members of the multidisciplinary team will be involved in writing the 'mission statement' on which the model is based.	The patient will know the beliefs, aims, and objectives of the HPN Care Team.
There will be relaxed, quiet area suitable for private discussion.	There will be a caring and compassionate atmosphere with adequate time for discussion.	The patient will feel able to express his or her fears and expectations.
The discharge planning documentation will include sections on domestic, family, and social circumstances.	The nutrition team will evaluate with the patient and family how the HPN will alter his or her way of life.	The patient will believe that the feeding system can be integrated into an acceptable way of life.
There will be written patient/carer learning goals for HPN.	A designated nurse will be responsible for teaching the patient according to his or her capacity for learning.	The patient/carer will be able to demonstrate the necessary skills and achieve all the individual learning goals.
There will be an instruction manual for HPN.	Information and procedures will be regularly updated to reflect developments and innovations in venous access, nutrient solutions, and delivery systems.	The patient will perform therapy based on current practice.
A relative, friend, or appropriate health-care professional will be available to deliver therapy if the patient is unable to do so (e.g., parent or guardian of a child).	The health-care professional will help the patient to identify the most appropriate carer. The district nurse will be given the opportunity to visit the patient in hospital and observe therapy before the patient is discharged.	The patient has confidence that safe care will be available at home.
Venous access will be achieved by a central venous catheter suitable for long-term use.	The patient, nurse, and doctor will choose the most appropriate catheter and access site.	The patient will use a central venous catheter that is acceptable and accessible.
There will be written procedures for the management of central venous catheters.	The nurse will adapt the procedures according to the patient's physical skills and domestic circumstances.	The patients's daily life will not be restricted by prolonged inappropriate procedures.
There will be a policy for sharing care with the patient's general practitioner (GP).	The GP will be contacted and a shared care protocol agreed.	The patient will know the responsibility of each health-care professional.
Written information describing HPN will be available for the GP.	The hospital teams will provide the GP with the information before the patient is discharged, together with the discharge date, and on-call telephone numbers.	The patient will have confidence in his or her GP's knowledge of HPN.
There will be a written prescription for the nutrition solutions (and other prescribable items).	The patient's GP will be contacted and advised on how to prescribe the feed.	The patient will have the feeding solution available at home on the day of discharge.
There will be a list of the required equipment, e.g., refrigerator, infusion pump, syringes, sterile gloves, and telephone.	Before discharge, the patient's home health authority will be provided with the list and asked to arrange supply by making local arrangements or establishing a contract with a commercial supplier.	The patient will have all the necessary supplies at home on the day of discharge.
There will be an on-call system for providing expert advice to the patient by telephone day and night.	The nurse will explain the system to the patient and identify the professions involved.	The patient/carer will know the names and telephone numbers to contact in case of emergency by day or night.
Information will be available describing how the nutrient solutions and supplies will be provided following discharge.	The nurse will explain the chosen supply system and discuss storage depending on the patient's home circumstances.	The patient will know how to obtain supplies, store them, and dispose of unwanted material.
There will be a postdischarge monitoring protocol, established by the nutrition team.	Monitoring will be supervised by the nutrition team.	The patient will know the date of the first outpatient visit and what monitoring will be performed.

the patient (e.g., whether the patient is receiving HETF or HPN), the duration of feeding, the extent of oral intake, and disease activity.

Outcome

The most important predictor of the outcome in patients receiving home artificial nutritional support (enteral or parenteral) is the underlying disease. Therefore, mortality statistics strongly depend on the initial indications. Nevertheless, a few conclusions can be

Table 5 Some complications associated with parenteral nutrition and enteral tube feeding

	Parenteral nutrition	*Enteral tube feeding*
Mechanical	Catheter malposition	Tube malposition (e.g., into lung)
	Insertion trauma (e.g., pneumothorax, brachial plexus injury, cardiac arrhythmia)	Insertion trauma; perforation of esophagus, stomach, and small bowel; peritonitis and peristomal leakage and inflammation
	Catheter blockage, kinking, or occlusion	Tube blockage, e.g., kinking or occlusion
	Catheter embolus	
	Air embolus	
	Clot embolus (from catheter tip)	
	Lack of access site	
Feed/flow	Nutrient overload (e.g., hyperglycemia, infusional hyperlipidemia)	Diarrhea or constipation
Infections	Catheter-related sepsis	Bloated abdomen/cramps
	Infected feed/administration set	Regurgitation/aspiration of feed
Metabolic	Fluid and electrolyte disturbances	Infected feed administration set
	Hyperglycemia	Infection around gastrostomy
	Deficiency syndromes, e.g., trace elements and vitamins	Fluid and electrolyte disturbances
	Nutrient overload (see above) and toxicity (e.g., some trace elements)	Deficiency syndromes (rate with standard feeds given to typical patients)
		Hyper/hypoglycemia
Organ tissue dysfunction	Abnormal liver function, intestinal atrophy, metabolic bone disease	Mainly disease-related, abnormal liver function
		Aspiration pneumonia
Psychological	Anxiety, depression, disturbance in self-image, social isolation	Anxiety, depression, disturbance in self-image, social isolation
Financial	Economic issues vary from center to center and country to country	Economic issues vary from center to center and country to country

made. First, the complications associated with artificial nutritional support vary but are reported to be responsible for less than 3–5% of deaths. Second, the outcome is dependent not only on the type of disease but also on the stage of the disease (e.g., patients with advanced HIV who start HPN are only expected to survive a few months, whereas patients with less-advanced disease are expected to survive longer). Third, the outcome of patients receiving HPN and HETF for a variety of conditions is available from the British Artificial Nutrition Survey (Table 6). For patients on HPN, overall mortality at 1 year is 13%, with 19% returning to oral feeding and the majority continuing with HPN. Patients with Crohn disease often have a good prognosis (with 2.5% mortality and 38% returning to oral feeding within 1 year). For patients on HETF, typically an older patient group, mortality is higher overall (36% at 1 year) and the outcome varies according to age and condition. The outcome data for two common conditions in adults and children receiving HETF are shown in Table 6.

Table 6 Twelve-month outcomes for patients receiving home enteral tube feeding (HETF) and home parenteral nutrition (HPN)

Continuing		*Discontinuing*			
Continues (%)		*In hospital (%)*	*Transferred to oral (%)*	*Withdrawn/refused (%)*	*Died (%)*
HETF					
All adults (*n*=33 955)	47.1	0.6	15.3	0.8	36.3
Cerebrovascular accident (*n*=11 621)	51.1	0.4	9.6	0.5	38.3
Esophageal cancer (*n*=2406)	23.8	1.0	24.9	1.3	49
All children (*n*=6946)	70.6	0.5	20.8	0.7	7.3
Cerebral palsy (*n*=977)	87.9	0.3	5.6	0.2	5.9
Congenital handicap (*n*=607)	85.5	1	7.1	0	6.4
HPN					
All adults (*n*=538)	63.6	1.1	19.3	3.2	12.8
All children (*n*=79)	64.6	1.3	19.0	0	15.2

Assessments of quality of life, using EuroQol, suggest that the majority of patients receiving HETF and HPN have some problems (moderate or extreme) with mobility, self-care, usual activities, pain/discomfort, and anxiety/depression (five EuroQol dimensions). Mean quality-of-life scores (0, 'worst imaginable health state'; 100, 'best imaginable health state') in adults receiving HPN (53±18) are higher than those for adults receiving HETF (42±27), but both are considerably lower than the scores obtained from the general population, even when adjusting for age. For HETF patients, quality-of-life scores have been found to be similar for those living at home and those in nursing care.

Intestinal Transplantation

In some patients with irreversible intestinal failure, intestinal transplantation can be considered as an alternative to long-term PN. The first intestinal transplantation in humans was undertaken in the early 1960s. Limitations in technical expertise and immunosuppressive therapy meant that none of the original patients survived beyond 76 days. From 1985 to 1990, a series of 20 patients were given cyclosporine, but only two patients were able to resume normal nutrition and most of the grafts failed. The development of new immunosuppressive agents, particularly tacrolimus, resulted in renewed interest in intestinal transplantation. Furthermore, since 1990, there has been greater standardization of patient selection, operative procedures, and postoperative care mainly in centers specializing in intestinal transplantation. The total international experience is still limited with less than 2000 patients reported in 2009 (1031 children and 733 adults). Some of the transplants were isolated intestinal grafts, others were intestinal–liver transplants, and the remaining few were multivisceral transplants that included the small intestine. Better graft and patient survival rates have been reported more recently, especially in the more experienced centers. For example, in North America, the 1-year graft survival rate for intestinal and multiorgan transplants increased from 52% in 1997 to 75% in 2005. The associated patient survival increased from 57% to 87%, but in more recent years centers with greater experience have reported a 1-year survival rate more than 90%. Early referral appears to be associated with better outcome. One study that assessed the quality of life reported an improvement after transplantation and less dependency on narcotics.

It appears that intestinal transplantation has become a realistic lifesaving option for some people who cannot be maintained on HPN. However, it is not yet the treatment of choice in patients who can be successfully maintained on HPN without noteworthy complications. Nor is it the treatment of choice in patients who are likely to deteriorate rapidly from other causes, such as aggressive multisystem disease, or likely to improve so that they can resume oral nutrition (e.g., patients with healing intestinal fistula or those with short bowel syndrome, in which benefits from intestinal adaptation may continue for up to 1–3 years). A better understanding of the immune response to the transplanted intestine and better immunosuppressive therapy, surgical techniques, and postoperative management are required. Appropriate selection and referral of patients to specialist centers are also important criteria that affect clinical outcomes.

Ethical Issues

The provision of nutritional support to people who are chronically sick, who have rapidly progressive disabling diseases, or who are terminally ill raises many ethical questions. Opinions about withholding or withdrawing artificial nutritional support vary from country to country because of different clinical, religious, and social beliefs and differences in national economies, some of which cannot support large-scale expensive long-term treatments. Thus, there is little home artificial nutrition in countries with poor economies. In more developed economies, the types of patients being fed may also vary considerably. For example, parenteral and enteral nutrition in patients with cancer are used more frequently in Italy than in the United Kingdom, suggesting that clinical attitudes to this type of nutritional support vary. The sanctity of human life is a belief that is strongly held by many religions, but when these conflict with medical judgment, public policies normally override personal religious beliefs. A common ethical controversy concerns the need to provide food and fluid to prolong life in severely disabled patients, such as those with severe neurological problems (e.g., cerebrovascular accident) or those approaching the ends of their lives. Although health professionals have a duty to prolong life, it seems inappropriate to prolong suffering. There has been controversy as to whether the provision of food and fluid by a feeding tube placed in the stomach or small intestine should be regarded as an essential part of care or medical treatment. The highest legal authorities in countries such as the United States and England have ruled that this is medical treatment. From an ethical perspective, there is no difference between withholding and withdrawing treatment, but in practice it is often more difficult to withdraw treatment once it has begun than to not initiate it. Joint discussions at the outset between mentally capable patients, family members, and health-care workers can do much to prevent future ethical dilemmas.

Conclusions

Home nutritional support, including both oral and artificial (enteral and parenteral) methods of feeding, is an important modality of treatment that is being used for an increasing number of people with disease and disability who are managed in the community. The identification of individuals who are at increased risk of malnutrition and who may benefit from additional nutritional support is a vital first step, which can be undertaken using a validated screening tool (such as MUST; **Figure 1**). Oral nutritional support, including liquid multinutrient supplements, is of value in improving the nutritional intake and functional well-being of patients with malnutrition in the community. Without ETF, many patients with persistent swallowing difficulties would die; similarly,

without PN, many patients with persistent intestinal failure would not survive. Although these forms of home therapy can be life-saving, they may restrict normal lifestyle and lead to life-threatening complications. These complications can be prevented or treated by establishing an adequate organizational infrastructure. This should include education and training of both health workers and patients/caregivers as well as a management structure that allows all patients to be followed up and, when necessary, admitting patients to the hospital for more intensive investigations and therapy. Ethical difficulties about withholding or withdrawing artificial nutritional support are likely to continue and to vary with time and from country to country. Intestinal transplantation is becoming a potentially realistic option for a few patients with irreversible intestinal failure who cannot be adequately maintained on long-term PN, but it has not yet become part of routine clinical care in the same way as renal transplantation has become routine in patients with renal failure, who would otherwise receive a lifelong treatment with dialysis.

Further Reading

Elia M (2003) The 'MUST' report. Nutritional screening for adults: A multidisciplinary responsibility. Development and use of the 'Malnutrition Universal Screening Tool' ('MUST') for adults. *A Report by the Malnutrition Advisory Group of the British Association for Parenteral and Enteral Nutrition*. Redditch, UK: BAPEN. Report No.: ISBN 1 899476 70X.

Elia M and Russell CA (2009) Combating malnutrition: Recommendations for action. *A Report from the Advisory Group on Malnutrition, Led by BAPEN*. Redditch, UK: BAPEN. Report No.: 978 1 899467 36 5.

Fishbein, TM., 2009. Intestinal transplantation. The New England Journal of Medicine 361 (10), 998–1008.

General Medical Council (2010) Treatment and care towards the end of life: Good practice in decision making. London: General Medical Council (code: GMC/EOL/0510).

Jones B, Holden C, Micklewright A, and British Artificial Nutrition Survey (BANS) Committee (2007) Trends in artificial nutrition support in the UK between 2000 and 2006. Redditch, UK: BAPEN. ISBN 1 899 467 16 5.

Stratton RJ and Elia M (2010) Encouraging appropriate, evidence-based use of oral nutritional supplements *Proceedings of the Nutrition Society* 1–11 (E-pub).

Stratton, RJ, Green, CJ, Elia, M., 2003. Disease-Related Malnutrition: An Evidence-Based Approach to Treatment. CABI Publishing (CAB International), Oxford.

Nutritional support: Infants and children, parenteral

S. Collier, Harvard School of Public Health, Boston, MA, USA
C. Lo, Harvard Medical School, Boston, MA, USA

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of S. Collier, C. Lo, Nutritional Support: Infants and Children, Parenteral, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 264–268, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00202-6.

History and Introduction **574**
Indications for PN 575
Dextrose 575
Protein 575
Lipid Emulsions 576
Micronutrients 576
Metabolic Complications **577**
Liver Disease 577
Bone Disease 577
Micronutrient Deficiency and Excess 577
Catheter Complications 578
Infections 578
Summary **579**
Further Reading **579**

Glossary

Central venous lines or catheters (CVL) Indwelling intravenous catheters, either silastic or polyurethane, placed with the tip in the superior vena cava to allow long-term provision of hypertonic dextrose solutions.

Enteral nutrition (EN) Nutrition through the gastrointestinal tract, either orally or through tube feedings.

Essential fatty acids Fatty acids which cannot be synthesized by humans and which are required in the diet, including linoleic, arachadonic, linolenic acids, docosahexanoic acid (DHA), and eicosapentanoic acid (EPA).

Macronutrients Carbohydrates, fats, or proteins required in large quantities in the diet, mainly for energy.

Micronutrients Vitamins or minerals essential in small quantities for normal physiologic functions but unable to be synthesized by humans and required in the diet.

Multivitamin infusion (MVI) A solution containing recommended requirements of essential vitamins and minerals.

Parenteral nutrition (PN) Nutrition delivered intravenously, not enterally.

History and Introduction

From its beginnings in the late 1960s, PN has expanded rapidly to become available for many patients in the USA and in Europe, but its expense and complications preclude wide availability in many countries. Before the availability of parenteral nutrition, as many as 30–50% of hospitalized patients had unrecognized malnutrition from chronic diseases and would remain for weeks without adequate nutrition to maintain weight or lean body mass, making them susceptible to infections, poor wound healing, and other complications. Surprisingly, it has been difficult to demonstrate substantial reductions in morbidity or mortality with parenteral nutrition except in moderately or severely malnourished patients or those with long-term intestinal failure. It is difficult to estimate the exact impact, but an American registry, the Oley Foundation, enumerated 10 035 Medicare beneficiaries on home parenteral nutrition in 1992, giving a rough estimate of 40 000 patients on home parenteral nutrition in the USA. Approximately 15–20% of these patients were children.

One of the first attempts at parenteral nutrition was carried out by Sir Christopher Wren in 1656. He infused ale, opium, and beer intravenously into animals. Parenteral nutrition that we are most familiar with for patient support has been available for approximately 40 years. The research carried out by Dr Stanley Dudrick and others allowed the support of the first pediatric patient on intravenous nutrition. The provision of intravenous nutrition was challenged by the development of several factors, before its completed use in patient support, including catheter access, sterility of solutions, and the optimal form of each macro- and micronutrient.

Indications for PN

There are many circumstances where PN is necessary and life sustaining. The indications for use have not changed dramatically over the years since the development of PN. Congenital malformation of the intestine, specifically small bowel atresia, was the diagnosis in the first cases PN was used in this age group. Congenital malformations of the gastrointestinal tract continue to be one of the leading reasons for its use as well as acquired diseases including necrotizing enterocolitis. Other indications include severe malabsorption, intestinal dysmotility, other congenital defects, and those with hematology–oncology diseases (Table 1).

Dextrose

The primary source of energy during intravenous therapy is usually provided by dextrose (D-glucose). This is especially true in infants and children when higher energy requirements often necessitate glucose infusion rates of up to 15 mg kg^{-1} min^{-1} or more. Not until 1945 did Zimmerman report the first attempt at infusing intravenous solutions through a catheter placed in the superior vena cava. Experiments performed by Dudrick in Beagle puppies advanced the glucose infusion solutions closer to what is utilized currently with hypertonic dextrose solutions. In current practice hypertonic solutions are infused through a catheter with its tip centrally located in the superior or inferior vena cava. It continues to be the major energy component of intravenous support.

Initial doses of glucose should be approximately 5–7 mg carbohydrate kg^{-1} min^{-1} with incremental increases of 2–5 mg kg^{-1} min^{-1}. Recommended maximum glucose infusion rate is 12–14 mg carbohydrate kg^{-1} min^{-1} for infants. Frequent monitoring of blood glucose and urine for glucosuria are important parameters to follow to assess tolerance to increasing glucose infusion rates. It is important to avoid excessive carbohydrate intake to minimize complications from potential hyperglycemia with subsequent osmotic diuresis. In addition, hepatic steatosis can occur with overfeeding. Hyperglycemia may ensue even without excess carbohydrate infusion in certain clinical situations, such as sepsis, renal failure, and certain medication including corticosteroids. Glucose infusion rates should be decreased if hyperglycemia ensues; however, it may still be necessary to add insulin to control blood glucose to provide adequate support.

Protein

Another vital macronutrient that needs to be provided is protein. Initial experiments from the 1930s were done with plasma as the protein source, where investigators achieved positive nitrogen balance. In the early 1900s research began into the development of protein hydrolysates and crystalline amino acids.

Vitrum, a company in Sweden, produced the first commercially available casein hydrolysate solution. It was developed by Arvid Wretlind who hydrolyzed casein enzymatically then dialyzed the mixture to remove large polypeptides Wretlind went on to modify it further and eventually replaced the hydrolysates in the 1970s.

The development of amino acid solutions specifically for infants took place in the early 1980s. These solutions provided conditionally required amino acids for the immature organ systems of premature infants and newborns. They were formulated based on the postprandial plasma amino acid levels of breast fed infants. Special amino acid solutions for renal or liver failure are also available, which have increased amounts of branched chain amino acids, Studies with the solutions for liver failure have shown they may be beneficial in adult patients with encephalopathy. Glutamine is a much-researched amino acid that could not initially be added to PN solutions due to shelf instability in liquid form. If it is added as a dipeptide it has been found to be more stable. Not all studies have shown clear benefit to its addition in patients for gut adaptation or prevention of bacterial translocation.

Recommendations for initiation and advancement of protein is 1–2 g kg^{-1} day^{-1} and advance by 1 g kg^{-1} day^{-1} to goal (see Table 1 for protein requirements). Initiation at the higher dose is more common practice for premature infants based on more recent studies that have demonstrated good tolerance and optimal metabolic profiles with earlier and higher protein within the first

Table 1 Conditions commonly requiring parenteral nutrition

Conditions	*Examples/Comments*
Surgical gastrointestinal disorders	Gastroschisis, omphalocele, tracheoesophageal fistula, intestinal atresias, meconium ileus, peritonitis, malrotation and volvulus, diaphragmatic hernia, prolonged postoperative ileus, Hirschsprung's disease, and intestinal dysmotility
Short bowel syndrome	
Prematurity	
Congenital heart disease	
Pancreatitis	
Gastrointestinal fistulas	
Bone marrow transplantation	
Acute intestinal disease	Antibiotic colitis, necrotizing enterocolitis, inflammatory bowel disease, chronic or secretory diarrhea
Hypermetabolic states	Burns, multiple trauma
Chronic idiopathic intestinal pseudoobstruction	

Source: Adapted with permission from Hendricks KM, Duggan C, and Walker WA (eds.) (2000) *Manual of Pediatric Nutrition*, 3rd edn. London: BC Decker.

24–48 h of life. Blood–urea nitrogen is monitored for tolerance to amino acid infusion. Prealbumin levels are helpful to monitor adequacy of protein intake.

Lipid Emulsions

Glucose was the only nonprotein source of energy until intravenous lipids were developed between 1920 and the 1960s. The first emulsion available for clinical use was Lipomul, a cottonseed oil based formulation. Because there were many adverse effects from its use it was withdrawn from clinical use in the mid 1960s. Wretlind, after extensive testing, developed Intralipid, a soybean-based emulsion, in 1961. It was well tolerated and is the most familiar intravenous fat emulsion currently available. It is available in 10%, 20%, and 30% solutions. The advantage of 20% and 30% over the 10% is the lower ratio of phospholipids to triglyceride, which minimizes the increase in plasma lipoprotein X levels. Lipid emulsions provide essential fatty acids, in addition to providing a concentrated energy source, particularly advantageous for patients requiring fluid restriction. Trials are underway for the use of emulsions that contain a blend of long chain fats with medium chain fats and those with fish oil blends. Also structured fat emulsions are being studied for clinical use. These specialized emulsions may have advantages in patients with liver disease and those with sepsis.

Lipid emulsions are usually initiated at 1 g kg^{-1} day^{-1} and advanced to 2–3 g kg^{-1} day^{-1} or 30–50% of total energy. More recent practice has been to limit to 1 g fat kg^{-1} day^{-1} for patients who are anticipated to need PN for greater than two weeks as a hepatoprotective measure. In addition, recent investigational use of a fish oil based emulsion has been found to be efficacious in reducing serum bilirubin levels in patients who have required long-term PN support. The mechanism is unclear, however a couple of proposed theories is less productions of inflammatory thromboxanes, prostaglandins, and leukotrienes and the absence of phytosterols. Serum TG levels are monitored for tolerance. Hypertriglyceridemia may occur in situations of stress, sepsis, and renal and liver insufficiency/failure. In addition, a number of medications can cause hypertriglyceridemia. In these situations a reduction in fat infusion is warranted, usually by infusing over 18–20 h instead of 24 h.

A minimum of 3–5% of total energy requirements is necessary to meet essential fatty acid requirements.

Micronutrients

To provide complete nutritional support, micronutrients, electrolytes, and minerals also need to be included in the parenteral solution. Addition of adequate amounts of calcium and phosphate together in one solution may be particularly problematic without precipitation occurring. Solubility guidelines are available accounting for the brand and percentage of amino acids and other salts, which impact the pH of the solution. Compounding guidelines for the order of addition of calcium and phosphorus, amounts of other additives and the temperature of the solution are other factors to optimize the solubility. Filters in the delivery system also help in minimizing the risk of occlusion of the catheter if a solution should precipitate, especially with so-called '3-in-1' solutions, where lipids are mixed with the glucose/amino acid solution. Further studies have evaluated the stability of the variety of nutrient components in these solutions.

Before the availability of vitamins and minerals, plasma levels of micronutrients fell rapidly while infusing only macronutrients. The only commercial preparation initially available was a trace element solution that had iron and iodide. The first commercial preparation of multivitamins for intravenous use was in the 1960s and lacked folic acid, vitamin B_{12} and K, and biotin. It also had very high concentrations of vitamins A and D and thiamin. Because of the variability in practice, there was increased risk of toxicity to vitamins A and D and deficiencies of other vitamins. Recommendations were made for intravenous pediatric and adult intravenous preparations in 1975. By 1978 there was a commercial multivitamin preparation that met these recommendations. Current preparations available contain all vitamins for which there are Dietary Reference Intake values, with the exception of choline. A Food and Drug Administration mandate now requires the addition of vitamin K to all preparations. Differences between the pediatric and adult forms of multivitamins (MVI) include amounts of B vitamins and vitamin D (see **Table 4**). There is currently no multivitamin preparation designed specifically for the premature infant. Dosing recommendations of Pediatric MVI for this group are based on weight (1/3 vial for <500 g; 2/3 vial for 500–1000 g and full vial for over 1000 g).

Trace element deficiencies have been noted in patients receiving long-term parenteral-nutrition support. The first case of chromium depletion was reported in 1977; selenium deficiency in 1979 and molybdenum in 1981. There are now many trace element solutions available with a variety of combinations of minerals and range from solutions appropriate to meet the needs of premature infants through the adult population. They are also available as single elements to tailor a solution as necessary. Contamination of trace elements can occur in parenteral solutions. Aluminum is one element that has been under recent scrutiny with the Food and Drug Administration mandate to minimize the amount patients receive. It was initially found to be in high concentrations in the casein hydrolysates and continues to be found in high concentrations in a variety of intravenous preparations. Over time aluminum can deposit in the bone, interfering with bone calcium uptake and deposition in the brain may impair neurological development.

Parenteral iron supplementation is controversial due to its potential risk of anaphylaxis and possible effect of providing a nutrient source for bacteria during sepsis episodes. However, if the enteral route is contraindicated for prolonged course or malabsorption and blood transfusions are not being given, consideration for iron dextran supplementation is warranted (**Table 2**).

Table 2 Pediatric parenteral nutritional requirements

	<2000 g	*0–4 years*	*5–18 years*
Energy (kcal kg^{-1} d^{-1})	100	80–90	40–70
Protein (g kg^{-1} d^{-1})	3–4	2.0–3.0	1–1.5
Fat (g kg^{-1} d^{-1})	<3	<3	<2
Sodium (mEq kg^{-1} d^{-1})	2–3	2–4	2–4
Potassium (mEq kg^{-1} d^{-1})	2–3	2–4	2–4
Chloride (mEq kg^{-1} d^{-1})	2–3	2–4	2–4
Calcium (mEq kg^{-1} d^{-1})	3–4.5	2–3	.5–2.5
(mg kg^{-1} d^{-1})	60–90	40–60	10–50
Magnesium (mEq kg^{-1} d^{-1})	.35–.6	.25–.5	.25–.5
Phosphate (mM kg^{-1} d^{-1})	1.5–2.5	1–2	1–2
Zinc (mcg kg^{-1} d^{-1})	400	300	100
Selenium (mcg kg^{-1} d^{-1})	1–3	1–3	1–2
Trace elements (ml l^{-1})	2	2	2
Multivitamins (ml d^{-1})	5	5	5–10

Metabolic Complications

Liver Disease

Although PN may be life-sustaining, long-term use may be detrimental to the liver. The severity of injury ranges from reversible transaminase elevations to severe cholestasis and cirrhosis, especially in infants with short bowel syndrome. It is still not clear whether this is due mainly to a nutrient deficiency, toxicity, or some physiological process missing because of the lack of enteral feeding. Prevention and treatment strategies continue to include minimizing or preventing episodes of sepsis, providing enteral feedings, moderating energy intake to provide adequate for growth, but not to overfeed, cycling parenteral nutrition infusion, reduction of copper and manganese, use of an amino acid solution developed for infants, treatment/prophylaxis for bacterial overgrowth and the use of ursodeoxycholic acid. Our common practice has been to reduce lipid exposure to 1 g kg^{-1} day^{-1} among infants likely to require long-term PN. Intravenous lipid emulsions are a rich source of linoleic acid, an omega- 6 polyunsaturated fatty acid, and may enhance production of the proinflammatory cytokines. Increased leukotriene B4 synthesis by the hepatic macrophages will draw additional polymorphonuclear leukocytes that intensify the inflammatory response to endotoxin by release of reactive oxygen species.

Bone Disease

The development of osteopenia is another complication that is common with long-term parenteral-nutrition support. The reasons are multifactorial and include relative immobility, inability to provide adequate calcium and phosphorus with solubility limitations, and hypercalciuria. It has also been suggested that the dose of vitamin D in the multivitamin preparation may contribute to bone disease. Excessive vitamin D may suppress parathyroid hormone secretion and directly cause bone resorption. Although aluminum is still present in some intravenous solutions, including calcium gluconate, vitamins, trace elements, the amounts are much less than those seen with the casein hydrolysates and are not believed to be a significant contributor toward the development of metabolic bone disease. Prevention and treatment strategies include maximizing calcium and phosphorus in parenteral-nutrition solutions, especially in growing children, providing enteral supplementation of these minerals as feasible and provide weight bearing physical therapy as possible.

Micronutrient Deficiency and Excess

If a patient is entirely parenteral-nutrition dependent, it is known that certain micronutrients need to be provided. Some parenteral-nutrition solutions require the addition of carnitine and selenium (if not provided in multi trace element solutions) and iron dextran (if the patient is not receiving transfusions or tolerating enteral iron). All serum levels should be monitored on a monthly basis or every 6–12 months if in the long-term phase of support. There may be other micronutrients not yet identified that may be deficient in the purified PN solution, hence another reason to begin enteral feedings as soon as feasible. Monitoring for excess losses is also important. For example, with increased stool/ostomy losses, the patient may require increased zinc in the parenteral-nutrition solution (Table 3).

Excess micronutrients can be caused by contamination, such as the case with aluminum, discussed earlier, or because of clearance. Copper and manganese can accumulate and become directly hepatotoxic because both elements depend on the biliary pathway for excretion. Therefore, in the presence of cholestasis, there will be increased intrahepatic accumulation. Manganese has also been reported to deposit in brain tissue, so copper and manganese levels should be monitored routinely.

Table 3 Suggested monitoring schedule for inpatients receiving parenteral nutrition

Parameter	*Daily*	*Weekly*[a]	*Periodically*[a]
Weight	x		
Fluid balance	x		
Vital signs	x		
Urine sugar	x		
Catheter site/function	x		
Laboratory (serum)			
Sodium		x	
Potassium		x	
Chloride		x	
Bicarbonate		x	
Glucose		x	
Urea Nitrogen		x	
Creatinine		x	
Triglycerides		x	
Calcium		x	
Magnesium		x	
Phosphorus		x	
Albumin or prealbumin		x	
Transaminases		x	
Bilirubin		x	
Selenium			x
Copper			x
Zinc			x
Iron			x

[a]Weekly or more often as necessitated by clinical course.

Catheter Complications

Complications with central venous catheters most frequently include obstructions, infections, and occasional leakage and perforation. Although parenteral nutrition can be temporarily provided though peripheral intravenous catheters, the high osmolarity of intravenous glucose-electrolyte solutions often cause phlebitis and loss of access. Therefore, long-term access has required placement of a central venous catheter, placed via the internal or external jugular vein or a subclavian vein. There is also increased placement of peripherally inserted cathethers by a team of specially trained staff and by an interventional radiologist. Tip position in the superior vena cava or SVC-right atrial junction should be verified radiographically to reduce complications from venous thrombosis or rare perforations. Central placement allows rapid dilution of hypertonic solutions in a large-diameter vein to minimize obstruction or thrombosis. Catheters for central venous access have been made of polyvinyl chloride, polyurethane, and silastic, often with a Teflon cuff to anchor the catheter subcutaneously. However, formation of a fibrin sheath is common, often with a biofilm, which may harbor infectious organisms and prevent penetration of antibiotics. Central catheter obstructions can often be visualized by ultrasound or inserting radioopaque dye in the catheter. A thrombus can often be lysed with installation of a small bolus of tissue plasminogen activator. Long-term anticoagulation with coumadin, low-dose coumadin, or low molecular weight heparin has been advocated by some to avoid repeated catheter obstruction, venous thrombosis, superior vena cava obstruction, and potential pulmonary emboli.

Obstructions caused by precipitation of calcium–phosphate salts or medications may be susceptible to installation of a small amount of dilute acid, and those due to fatty material might be dissolved with dilute ethanol. For long-term home parenteral use, some patients have preferred use of implantable ports, which can be accessed through the skin daily with a special needle. Recently, peripherally inserted central catheters (PICC) have been used for longer periods up to a month or more without requiring a surgical procedure.

Infections

Patients who require parenteral nutrition are often predisposed to infectious complications. The catheter hub is often the entry site with skin flora such as *Staphylococcus epidermidis*, *Staphylococcus aureus*, or *Candida* being the most common organisms, along with Gram negative enteric bacteria possibly from bacterial translocation. Antibiotic treatment through the central line is often successful without replacement of the catheter, using antibiotic combinations such as vancomycin and gentamicin, or with periodic indwelling infusions of a 70% ethanol lock (Table 4).

Table 4 Comparison of parenteral multivitamin preparations for pediatric and adult

Vitamin	*MVI Pediatric (Hospira)*	*Infuvite (Baxter) and MVI adult with vitamin K (Hospira)*
A (IU)	2300	3300
D (IU)	400	200
E (IU)	7	10
K (mCg)	200	150
Ascorbic acid (mg)	80	200
Thiamine (mg)	1.2	6
Riboflavin (mg)	1.4	3.6
Niacin (mg)	17	40
Pantothenate (mg)	5	15
Pyridoxine (mg)	1	6
B_{12} (mCg)	1	5
Biotin (mCg)	20	60
Folate (mCg)	140	600

Source: Adapted with permission from Hendricks KM, Duggan C, and Walker WA (eds.) (2000) *Manual of Pediatric Nutrition*, 3rd edn. London: BC Decker.

Summary

Advancements in the technology, production and manufacturing of intravenous solutions have progressed over the past 40 years. Along with improvement in the solutions available for dextrose, amino acids and lipid emulsions, there has also been progress with the delivery systems, catheters, and improved sterile techniques for line and skin care to reduce overall complications.

Ongoing research and product development in areas associated with long-term PN support is vital for future patient management to be able to continue to provide optimal support with minimal risk for those patients for whom PN is life sustaining.

Further Reading

American Gastroenterological Association Clinical Practice Committee, 2001. AGA technical review on parenteral nutrition. Gastroenterology 121, 970–1001.

American Society of Parenteral and Enteral Nutrition (ASPEN), 2002. Guidelines for the use of parenteral and enteral nutrition in adult and pediatric patients. Journal of Parenteral and Enteral Nutrition 26S, 1–152.

Buchman, A., 2002. Total parenteral nutrition associated liver disease. Journal of Parenteral and Enteral Nutrition 26, S43–S48.

Dudrick, SJ., 2003. Early developments and clinical applications of total parenteral nutrition. Journal of Parenteral and Enteral Nutrition 27, 291–299.

Dudrick, S.J., Wilmore, D.W., Vars, H.M., Rhoads, JE., 1968. Long term parenteral nutrition and growth, development and positive nitrogen balance. Surgery 64, 134–142.

Duggan, C., Watkins, J., Walker, WA., 2009. Nutrition in Pediatrics, 3rd edn. People's Medical Publishing, Shelton, CT.

Forchielli, M.L., Gura, K., Anessi-Pessina, E., Richardson, D., Cai, W., Lo, C., 2000. Success rates and cost-effectiveness of antibiotic combinations for initial treatment of central venous line infections during total parenteral nutrition. Journal of Parenteral and Enteral Nutrition 24, 119–125.

Kaufman, SS., 2002. Prevention of parenteral associated liver disease in children. Pediatric Transplantation 6, 37–42.

Kinney, JM., 2000. Clinical Nutrition Parenteral Nutrition, 3rd edn. WB Saunders, Philadelphia. pp. 1–20.

Kleinman, R.D., Barness, L.A., Finberg, L., 2003. History of parenteral nutrition and fluid therapy. Pediatric Research 54, 762–772.

Oley Foundation (1994) North American home parenteral and enteral nutrition patient registry. Annual Report With Outcomes Profiles 1985–1992. Albany, NY: Oley Foundation.

Seidner, DL., 2002. Parenteral nutrition associated metabolic bone disease. Journal of Parenteral and Enteral Nutrition 26, S37–S42.

Shils, ME., 2000. Recalling a 63 year nutrition odyssey. Nutrition 16, 582–628.

Veterans Affairs Total Parenteral Nutrition Cooperative Study Group, 1991. Perioperative total parenteral nutrition in surgical patients. The New England Journal of Medicine 325, 525–532.

Vinnars, E., Wilmore, D., 2003. History of parenteral nutrition. Journal of Parenteral and Enteral Nutrition 27, 225–232.

Obesity: Childhood obesity

G. Bueno-Lozano[a,c,d,e] **and L.A. Moreno Aznar**[b,c,d,e], [a] Department of Pediatrics and Pediatric Endocrinology, University of Zaragoza, Hospital Clínico Universitario Lozano Blesa, Zaragoza, Spain; [b] Faculty of Health Sciences, Instituto Agroalimentario de Aragón (IA2), University of Zaragoza, Zaragoza, Spain; [c] Instituto de Investigación Sanitaria de Aragón (IIS Aragón), Zaragoza, Spain; [d] GENUD (Growth, Exercise, NUtrition and Development) Research Group, Zaragoza, Spain; and [e] CIBER Fisiopatología Obesidad y Nutrición, Instituto de Salud Carlos III, Madrid, Spain

© 2023 Elsevier Ltd. All rights reserved.

This is an update of E.M.E. Poskitt, Obesity: childhood obesity, Editor(s): Benjamin Caballero, *Encyclopedia of Human Nutrition* (Third Edition), Academic Press, 2013, Pages 336–342, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00209-9.

Introduction **581**
Epidemiology **581**
Assessment of overweight and obesity in childhood **581**
Genetics factors **582**
Risk factors for childhood obesity **582**
Critical periods in the development of obesity 582
Early feeding 582
Familial obesity 583
Socioeconomic and environmental deprivation 583
Diet and dietary change 583
Physical activity 583
Sleep 583
Clinical approaches to childhood obesity **583**
Common or polygenic obesity 584
Monogenic obesity 584
Secondary obesity 585
Complications of childhood obesity **585**
Cardiometabolic and cardiovascular system 585
Endocrine system 585
Respiratory system 585
Gastrointestinal system 585
Musculoskeletal involvement 586
Psychosocial aspects 586
Dermatologic alterations 586
Neurological aspects 586
Long-term risks 586
Clinical evaluation of children with obesity **586**
Treatment of childhood obesity **587**
Dietary management 587
Physical activity 587
Behavioral treatment strategies 588
Pharmacologic treatment 588
Surgical treatment 588
New treatments 589
Prevention of obesity during childhood **589**
Conclusion **589**
References **589**

Key points

- Childhood obesity has increased worldwide in epidemic proportions
- Children with obesity are at high risk for multiple comorbidities previously considered to be "adult" diseases.
- Lifestyle modifications including dietary changes aimed at decreasing total energy intake, increasing physical activity and decreasing sedentary time are crucial for weight management.
- Pharmacotherapy may have a role in the treatment of pediatric obesity, but evidence is limited.

- Bariatric surgery is effective in achieving weight loss and improving comorbidities in adolescents with severe obesity.
- Primary prevention and family-based lifestyle interventions for weight loss are recommended.

List of abbreviations

BDNF brain-derived neurotrophic factor
FTO fat mass and obesity–associated protein
IGF-I insulin-like growth factor I
IGFBP-3 insulin-like growth factor binding protein 3
IR insulin resistance
LEPR leptin receptor
MC4R melanocortin receptor 4
MS metabolic syndrome
NTRK neurotrophic tyrosine receptor kinase
PCSK1 proprotein convertase type subtilisin/kexin 1
POMC proopiomelanocortin
PPARγ peroxisome proliferator-activated receptor γ
SIM1 single-minded homolog 1
T2DM type 2 diabetes mellitus

Introduction

Childhood obesity has emerged as one of the most important medical and public health problems. The prevalence is unacceptably high, and the rate of increase in severe obesity continues to climb. The etiology of the disease is multifactorial and complex, assuming an interaction among genetic, biological and environmental factors. Childhood obesity is associated with several comorbidities that affect physical and mental health. The serious comorbidities associated with childhood obesity needs effective treatment options. A stratified approach to treatment is recommended. Epidemiology, clinical features, complications, management and prevention of childhood obesity are reviewed.

Epidemiology

The prevalence of obesity has progressively increased during the past 30 years in all developed countries including the United States and most European countries, with these changes being more apparent in Mediterranean countries (with a prevalence between 20% and 40%) than in northern countries. Childhood obesity is also increasing in prevalence among the less developed countries.

In 2013–14, the worldwide number of children and adolescents (2–19 years of age) with obesity was estimated to be 110 million; this number has doubled since 1980. 22.8% of preschool children (age, 2–5 years), 34.2% of school-aged children (age, 6–11 years), and 34.5% of adolescents (age, 12–19 years) had overweight or obesity and 8.4% of preschool children (age, 2–5 years), 17.7% of school-aged children (age, 6–11 years), and 20.5% of adolescents (age, 12–19 years) had obesity (Ogden et al., 2014). Analyses of trends in overweight and obesity for both boys and girls (11–15 years of age) in North America and countries in Europe from 2002 to 2010 evidenced stabilization in overweight prevalence, but overall rates of overweight in many countries are high. In Spain, recent cross-sectional studies reports a joint prevalence of overweight and obesity of 44% in the population aged between 3 and 24 years (Aranceta-Bartrina et al., 2020).

Unfortunately, a high percentage of children with obesity carry their adiposity into adulthood. The tracking of obesity into adulthood is affected by age of the child, severity of obesity, and presence of parental obesity. Older age is associated with greater persistence of obesity into adulthood, and therefore most adolescents with obesity will continue to have obesity during the adult life (Kumar and Kelly, 2017).

Assessment of overweight and obesity in childhood

The term obesity refers to body fat excess. Because of the unavailability and high cost of techniques that directly measure body fat, body mass index (BMI), derived from the body weight and height, has emerged as the accepted clinical standard measure of overweight and obesity for children 2 years and older. Body mass index is calculated by dividing the body weight in kilograms by the

height in meters squared. In general, BMI provides a reasonable estimate of adiposity in the healthy pediatric population. However, BMI may overestimate fatness in children who are short or who have relatively high muscle mass and may underestimate adiposity in a substantial proportion of children, such as those with reduced muscle mass due to low levels of physical activity. BMI should be viewed as a surrogate measure of adiposity and its strengths and limitations should be considered when used in clinical and research settings.

For children younger than 2 years, weight for length is the accepted measure of overweight and obesity. Because of normal growth and development during childhood, the absolute level of BMI in children vary with age and sex. In children and adolescents, standardized values of BMI as a function of age and sex for their reference population should be used. This has generated intense controversy regarding the establishment of "cutoff points" and the reference populations that should be used (Martos-Moreno et al., 2014). In terms of use for epidemiological studies and international comparisons, the International Obesity Task Force reference standards should be used. They have defined childhood overweight and obesity considering those cut-off points equivalent to BMI, 25 and 30 kg m^2, in adults, for every sex and age value until 18 years (Martos-Moreno et al., 2014).

In clinical settings and according to local standards, the following BMI-based definitions are used for overweight and obesity for children and adolescents between 2 and 20 years of age:

- Overweight: BMI at or greater than 85th to less than 95th percentile for age and sex
- Obesity: BMI at or greater than 95th percentile for age and sex
- Severe obesity: BMI at or greater than 120% of the 95th percentile, or BMI at or above 35 kg/m^2 (whichever is lower). This corresponds to approximately the 99th percentile, or BMI z score at or above 2.3 above the mean (Kumar and Kelly, 2017).

There is also no consensus regarding the definition of morbid obesity in children and adolescents. Some authors suggest a BMI >3 SDS or 200% of the ideal body weight for height as possible cutoff points or BMI at or above 35 kg/m^2.

The measurement of total body fat and its distribution is important because different adipose depots have distinct roles in the development of metabolic complications. However, access to techniques for directly measuring body fat content is limited in routine clinical practice. Indirect estimations of visceral adipose content can be made by performing waist measurements and by using international references that are classified according to ethnic group (Martos-Moreno et al., 2014).

Although consensus has not been reached regarding cutoff points for normal waist circumference measurements in childhood, high waist circumference seems to predispose to developing obesity-related comorbidities.

Genetics factors

Heritable factors appear to be responsible for 30%–50% of the variation in adiposity. Large-scale genome-wide association studies (GWAS) that test the association of millions of common genetic variants without a prior hypothesis about their presumed role have identified >300 genetic loci for obesity traits. For example, variants in the primer intron of the gene coding for the fat mass and obesity–associated protein (FTO) have been associated with an elevation in BMI equivalent to approximately +0.4 kg/m^2 for each risk allele This can explain, at least in part, the large heritable influences observed in obesity (González-Muniesa et al., 2017).

Risk factors for childhood obesity

Childhood obesity is the consequence of an interaction among a complex set of factors that are related to, genetics, environment and ecological such as the family, school and community.

Critical periods in the development of obesity

The first thousand days, from conception until the end of the second year of life, mark the first critical period in the development of obesity. Body mass index (BMI) usually increases until 7 months of age, when it reaches a temporary maximum (the so-called infant BMI peak). Between 5 and 7 years of age, the BMI reaches a minimum in children with adequate growth and development, after which it starts to rise again (that is, the adiposity rebound). During adolescence, BMI changes are substantially associated with puberty. Body weight at these critical periods is associated with later body composition High maternal BMI, high gestational weight gain, rapid postnatal growth, early adiposity rebound, and early pubertal development are risk factors for childhood obesity (Poskitt, 2013; Rolland-Cachera et al., 2006).

Early feeding

Although there is no consistent evidence that breast feeding protects children from later obesity, the process of breast feeding may have positive influences on mothers' attitudes to child nurturing.

Familial obesity

Hereditary factors have a strong effect on the development of obesity in children. There is evidence that obesity in one parent increases the risk of obesity in the child by 2- to 3-fold, and up to 15-fold if both parents have obesity (Poskitt, 2013).

Socioeconomic and environmental deprivation

Deprived environments and families in which child care and nurture are deficient show the greatest predisposition to childhood obesity.

Diet and dietary change

Eating behaviors in children and parental feeding styles have been shown to be associated with a high risk of childhood obesity. Several factors in the current "obesogenic" environment have resulted in increased energy consumption, such as increasing use of sugar sweetened beverages, sweet snacks, fast foods containing excess fat, large portion sizes, and high glycemic foods. Consumption of sugar-sweetened beverages (including fruit juices) has been postulated as an important contributor to the development of obesity in children. Consumption of fast food has also been purported to contribute to the increasing prevalence of obesity (Moreno et al., 2010).

Physical activity

The changes contributing to increased energy intake have been accompanied by factors predisposing to decreased energy expenditure such as reduced levels of physical activity and increasing time spent in sedentary activities such as use of television, computers, phones, and tablets. The amount of time spent watching television and the presence of a television in a child's bedroom have been shown to be directly related to the prevalence of obesity in children and adolescents. This association can be explained by several potential mechanisms including displacement of physical activity and adverse effects on the quality and quantity of foods consumed. The use of electronic devices has also been associated with obesity during childhood (Seral-Cortés et al., 2021).

Sleep

There is increasing evidence for an association between shortened sleep duration and/or poor sleep quality and obesity (Kumar and Kelly, 2017). Sleep may also have an association with decreased insulin sensitivity, independent of the association with adiposity.

Clinical approaches to childhood obesity

Obesity in children has diverse causes, which can result in radically different pathologic entities. Although polygenic obesity is by far the most commonly observed, several single gene defects and syndromes associated with obesity have been identified. However, these account for less than 1% of childhood obesity (**Table 1**) (Martos-Moreno et al., 2014).

Table 1 Clinical description of the main monogenic forms of obesity.

MC4R mutations	Severe early-onset obesity with hyperphagia Binge eating Tall stature and advanced bone age during childhood
Leptin deficiency	Severe early-onset obesity with low circulating leptin concentrations Hypogonadotrophic hypogonadism reversible after recombinant leptin administration
Leptin receptor mutations	Severe early-onset obesity Hypogonadotrophic hypogonadism, low TRH and GnRH concentrations
POMC deficiency	Severe early-onset obesity Low plasma cortisol concentrations Common impairment of skin and hair pigmentation
PCSK1 deficiency	Severe early-onset obesity Increased POMC and proinsulin and low insulin and cortisol concentrations Hypogonadism possible

GnRH, gonadotrophin-releasing hormone; MC4R, melanocortin receptor number 4; PCSK1, proprotein convertase type subtilisin/kexin 1; POMC, proopiomelanocortin; TRH, thyrotropin-releasing hormone.

Common or polygenic obesity

Exogenous or "common" obesity is the most frequent form of obesity in children and adolescents. Hypercaloric nutrition and reduced physical activity coexist resulting in the accumulation of excess energy in the form of adipose tissue. Familial obesity is frequent. Unfortunately, this is also influenced by the individual's socioeconomic situation. Prepubertal children with obesity are relatively tall for age. Advanced growth may be associated with advanced maturity of bones (advanced bone age) and early onset of puberty.

Monogenic obesity

Monogenic obesity results from an alteration in a single gene and represents a minority of the total population of children with obesity. A common feature of these patients is severe early-onset obesity. The forms of monogenic obesity known to date can be classified into the following 3 categories:

1. *Genes involved in the adipocyte-hypothalamic system (leptin/melanocortin axis)*. The most important genes included in this category are those coding for leptin (LEP), leptin receptor (LEPR), proopiomelanocortin (POMC), proconvertase 1 (PCSK1), proconvertase 2 (PCSK2), melanocortin receptor 4 (MC4R), and peroxisome proliferator-activated receptor γ (PPARγ). The most common single gene defect currently identified in children with obesity is mutations in the melanocortin 4 receptor. The clinical features of patients affected by these types of monogenic obesities are summarized in **Table 1** (Kumar and Kelly, 2017; Martos-Moreno et al., 2014).

Table 2 Secondary causes of childhood obesity.

Monogenic disorders	*Endocrine*
• Melanocortin 4 receptor haploinsufficiency • Leptin deficiency • Leptin receptor deficiency • Proopiomelanocortin deficiency • Proprotein convertase 1	• Hypothyroidism • Neonatal hyperinsulinism • Glucocorticoid excess (cushing syndrome) • Growth hormone deficiency • Pseudohypoparathyroidism
Syndromes	*Psychological*
• Alstrom-Wolfram (OMIM: 203800) • Bardet-Biedl (OMIM: 209900) • Beckwith-Wiedemann (OMIM: 130650) • Börjesson-Forssman-Lehmann (OMIM: 301900) • Carpenter (OMIM: 201000) Cohen (OMIM: 216550) • Down (OMIM: 190685) • MEHMO (OMIM: 300148) • MOMO (OMIM:157980) • Prader-Willi (OMIM:176270) • Smith-Magenis (OMIM: 182290) • Wilson-Turner (OMIM: 309585) • WAGRO (OMIM: 612469)	• Depression • Eating disorders (binge eating disorder and night eating disorder) Drug induced • Tricyclic antidepressants • Glucocorticoids • Antipsychotic drugs • Antiepileptic drugs • Sulfonylureas
Neurologic	*Hypothalamic causes*
• Brain injury • Brain tumor • After cranial irradiation • Hypothalamic obesity	• Tumor • After brain surgery/radiation (craniopharyngioma) • ROHHAD/ROHHADNET syndrome

ROHHAD, rapid-onset obesity with hypothalamic dysfunction, hypoventilation, and autonomic dysregulation; ROHHADNET, rapid-onset obesity with hypothalamic dysfunction, hypoventilation, and autonomic dysregulation with neural crest tumors.

2. *Genes associated with hypothalamic development.* Anomalies in at least 3 genes associated with hypothalamic development, including single-minded homolog 1 (SIM1), brain derived neurotrophic factor (BDNF), and neurotrophic tyrosine receptor kinase 2 (NTRK2), have been shown to correspond to obesity. To date, the reported number of cases involving these genes is small (Martos-Moreno et al., 2014).
3. *Obesity associated with syndromes.* There are a number of syndromes in which obesity is among the characteristic feature (**Table 2**). Children with genetic syndromes associated with obesity typically have early-onset obesity and characteristic features on physical examination, such as short stature, dysmorphic features, developmental delay, or intellectual disability (mental retardation), retinal changes, or deafness (Martos-Moreno et al., 2014).

Prader-Willi syndrome is the most common syndrome associated with obesity, and children have hypotonia and feeding difficulties during infancy (often with failure to thrive), followed by hyperphagia and subsequent development of obesity.

Secondary obesity

Obesity can also be the consequence of a disease or its treatment. Endocrine diseases (Cushing syndrome) and pathologic processes or therapeutic protocols that affect the pituitary hypothalamic axis can promote weight gain. Likewise, some pharmacologic treatments, especially those with psychoactive ingredients, can result in obesity. The most frequent causes of secondary obesity are shown in **Table 2**.

Complications of childhood obesity

Childhood obesity is associated with relevant comorbidities affecting almost every system in the body including, the cardiovascular, endocrine, pulmonary, gastrointestinal, and musculoskeletal systems. Many of the comorbidities encountered in youth with obesity, including type 2 diabetes mellitus (T2DM), dyslipidemia, obstructive sleep apnea (OSA), and steatohepatitis, used to be previously considered "adult" diseases.

Cardiometabolic and cardiovascular system

Children with obesity are at an increased risk of hyperinsulinemia, insulin resistance, prediabetes, and subsequently T2DM. The prevalence of prediabetes and T2DM varies with severity of obesity, race, ethnicity, and age of the child. Those who present with T2DM during adolescence appear to have more rapid deterioration of glycemic control and progression of diabetes-related complications such as microalbuminuria, dyslipidemia, and hypertension as compared with those who present it later in life. Children with obesity also have a high prevalence of other cardiometabolic risk factors including elevated blood pressure, low levels of high-density lipoprotein cholesterol, and elevated levels of triglycerides. Echocardiographic findings include left ventricular hypertrophy, increased left ventricular and left atrial diameter, and systolic and diastolic dysfunction (Kumar and Kelly, 2017; Styne et al., 2017).

Endocrine system

Obesity may be associated with early onset of sexual maturation in girls and with accelerated linear growth and advanced skeletal maturation. Adolescent girls are also at high risk of developing hyperandrogenism and polycystic ovary syndrome. Manifestations of polycystic ovary syndrome includes menstrual irregularities, acne, and hirsutism (Kumar and Kelly, 2017; Styne et al., 2017).

Respiratory system

Children with obesity have a considerably higher prevalence of OSA than do normal-weight children. The prevalence and severity of OSA increase with increasing BMI. Children with severe obesity may also have alveolar hypoventilation associated with severe oxygen desaturation. Childhood obesity has also been shown to be associated with asthma (Kumar and Kelly, 2017; Poskitt, 2013; Styne et al., 2017).

Gastrointestinal system

Nonalcoholic fatty liver disease (NAFLD) in children is strongly associated with obesity. The spectrum of NAFLD can range from simple steatosis to progressive steatohepatitis and cirrhosis. Nonalcoholic fatty liver disease is now the most common cause of liver disease in children. Although most patients with NAFLD are asymptomatic, laboratory abnormalities include elevations in levels of liver transaminases (alanine aminotransferase and aspartate aminotransferase), alkaline phosphatase, and gamma-glutamyl transpeptidase (Styne et al., 2017).

Musculoskeletal involvement

Childhood obesity increases the risk of various musculoskeletal problems including impairment in mobility, increased prevalence of fractures, lower extremity joint pain, and lower extremity malalignment. Obesity is also a risk factor for unilateral or bilateral slipped capital femoral epiphysis and for tibia vara (Kumar and Kelly, 2017; Styne et al., 2017).

Psychosocial aspects

Psychosocial consequences of childhood obesity are common and include poor self-esteem, anxiety, depression, and decreased health-related quality of life. Children with obesity are more likely to become victims of bullying and discrimination (Kumar and Kelly, 2017).

Dermatologic alterations

Acanthosis nigricans, a marker of insulin resistance, is a common finding in children with obesity. Other skin abnormalities include intertrigo, hidradenitis suppurativa, furunculosis, and stretch marks (Kumar and Kelly, 2017).

Neurological aspects

Childhood obesity is associated with a higher risk of idiopathic intracranial hypertension (pseudotumor cerebri). Clinical symptoms include headache, vomiting, retro-ocular eye pain, and visual loss (Styne et al., 2017).

Long-term risks

Children whose obesity persists into adulthood have a significantly increased risk of T2DM, hypertension, dyslipidemia, and carotid-artery atherosclerosis than do adults who never had obesity. Higher BMI during childhood has also been associated with an increased risk of fatal and nonfatal cardiovascular events during adulthood in both men and women, though this may be partially mediated by the association between childhood obesity and adult obesity (Kumar and Kelly, 2017; Poskitt, 2013; Styne et al., 2017).

Clinical evaluation of children with obesity

The clinical evaluation of children with obesity is directed at identifying the cause of obesity and its obesity-related comorbidities. The evaluation includes a complete history and physical examination.

Dietary history should consist of details of eating habits including frequency, content, and location of meals and snacks as well as intake of energy dense foods such as fruit juices and others. Physical activity assessment should include details of time spent in unstructured play, organized sports, school recess, and physical education as well as screen time (television, video games, mobile phones, and tablets). Medical history should include details about medications that may cause weight gain such as glucocorticoids, antipsychotic drugs, and antiepileptic drugs. A developmental history is important as developmental delay may point toward a chromosomal or genetic cause for obesity. A complete review of systems is helpful in determining an underlying etiology for the weight gain, such as Cushing syndrome or hypothalamic tumor.

The review of symptoms is also helpful in screening for obesity-related comorbidities such as OSA. Family history of obesity and obesity-related comorbidities is a predictor of persistence of obesity into adulthood. Performing a comprehensive psychosocial screening including collecting details related to depression, peer relationships, and disordered eating habits is crucial.

Physical examination should include measurement of height and assessment for dysmorphic features suggestive of a chromosomal or monogenic cause and for Cushingoid features.

Most children with exogenous obesity are tall, whereas children with genetic and endocrine causes of obesity tend to have short stature. Acceleration of bone age in respect to chronologic age, but adequate for the child's height, is frequent in obesity.

Blood pressure should be measured with an appropriate sized cuff (Styne et al., 2017).

There is lack of standardization and consensus on when to screen and the types of laboratory screening tests to perform in children with obesity. Most experts recommend that children with overweight, that is, BMI between the 85th and 95th percentiles, who are free from risk factors should have measurement of a fasting lipid profile (Styne et al., 2017). These children should also undergo measurement of fasting blood glucose or hemoglobin A1C and aspartate aminotransferase and alanine aminotransferase levels if they are 10 years and older and have 1 or more of the following risk factors: elevated blood pressure, elevated lipid levels, currently using tobacco, or having a family history of obesity-related diseases (Martos-Moreno et al., 2014; Styne et al., 2017).

Because the appearance of insulin resistance (IR) and subsequent alterations in carbohydrate and lipid metabolism is gradual, it is possible that hyperinsulinemia may exist in the absence of alterations in fasting glycemia. Thus, it could be necessary to analyze fasting insulinemia in children and adolescents with obesity and to calculate the HOMA index [fasting glucose (mmol/L) x insulin (mU/mL)/22.5] as an indicator of IR. When the patient belongs to an ethnic group with an elevated risk (Hispanic, African

American) or there are alterations in fasting glycemia (>100 mg/dL) or insulinemia (>15 mU/mL), dyslipidemia, hypertension, family history of T2DM, or conditions associated with IR such as acanthosis nigricans or symptoms of polycystic ovary syndrome, an oral glucose-tolerance test should be performed (Martos-Moreno et al., 2014).

A fasting lipid profile is recommended for all children with BMI >95th percentile even in the absence of risk factors. If the results of the fasting lipid profile are normal, repeated screening is recommended every 2 years (Kumar and Kelly, 2017; Styne et al., 2017).

Sodium and potassium concentrations can be altered in hypercortisolism, calcium and phosphorus in pseudohypoparathyroidism, and urea and creatinine with the presence of microalbuminuria in nephropathies. It is also recommended that a Vitamin D, hemogram and iron and ferritin concentrations be studied because of possible associated deficiencies (Styne et al., 2017).

Hormonal studies should be performed to rule out the possibility of hypothyroidism (free thyroxin (T4) and thyroid stimulating hormone] or hypercortisolism (24-h urinary cortisol concentrations). If growth hormone deficiency or insensitivity is suspected, serum insulin-like growth factor I (IGF-I) and insulin-like growth factor binding protein-3 (IGFBP-3) concentrations should be determined (Martos-Moreno et al., 2014).

Abdominal ultrasound is the technique of choice if non-alcoholic steatohepatitis or polycystic ovary syndrome is suspected (Styne et al., 2017). Imaging techniques may confirm the presence of fatty liver, indicated by increased echogenicity on ultrasonography, but liver biopsy is the only way to reliably distinguish between simple steatosis, steatohepatitis, and fibrosis and can also be helpful in excluding other causes of elevated levels of serum aminotransferases.

Children with signs and symptoms suggestive of a genetic or endocrine cause for the weight gain may need specific testing. In addition, children with signs and symptoms suggestive of comorbidity such as OSA may need specific testing such as an overnight polysomnogram (Styne et al., 2017).

Treatment of childhood obesity

The treatment of obesity in children is based on the reorganization of dietary habits, increase physical activity, and behavioral modification. In this age group, the indications for pharmacologic or surgical treatment remain limited, and it is impossible to establish specific recommendations for their use (Kumar and Kelly, 2017; Styne et al., 2017). Treatment of the children with obesity should be supervised by an experienced professional with frequent visits, especially during the initial treatment period. It is recommended that treatment be individualized on the basis of the characteristics of the child and his or her family.

Weight loss goals should be determined by the child's age and severity of obesity and related comorbidities. Weight maintenance might be an appropriate goal for children who have mild obesity because BMI will decrease as children gain height. In contrast, weight loss is recommended in children with severe obesity and those with comorbidities. It has been suggested that a 500–1000 g weight loss per month is safe in children and adolescents with severe obesity and comorbidities (Kumar and Kelly, 2017; Styne et al., 2017).

Dietary management

Nutritional intervention includes combined strategies aimed at achieving changes in lifestyle. It is impossible to establish specific recommendations regarding nutritional intervention in childhood obesity, because controlled studies with long-term follow-up are still required, including evaluation of diets with different degrees of energy restriction and changes in macronutrient composition (limitation of fats and carbohydrates or increases in fiber or proteins) or micronutrients (calcium). Diets with modified carbohydrate intake such as low glycemic index diets and low carbohydrate diets have been shown to be as effective as standard portion-controlled diets for weight management in children with obesity. However, adherence to the modified carbohydrate diets may be low and children may be unable to follow these types of regimens, particularly in the long term. Very low-energy diets have been used quite successfully for short-term weight reduction. However, such diets carry some risks for nutrition and growth, and are unacceptable to many children with obesity (Kumar and Kelly, 2017; Poskitt, 2013; Styne et al., 2017). There is no consistent evidence that reduction of any particular energy source is more effective than any other in promoting fat loss. Current recommendations are focused on programming intake and avoidance of energy excess, suggesting that children and adolescents with obesity consume a mixed and varied diet (balanced diet) with a moderate energy restriction, and decreasing portion size. It is recommended to encourage intake of five servings of fruits and vegetables daily, decrease intake of energy-dense foods such as saturated fats, salty snacks, and high glycemic foods such as candy, minimize intake of sugar-containing beverages, minimize eating outside home and fast food in particular, eat breakfast daily and avoid skipping meals (Moreno et al., 2010).

To support children understand these recommendations, programs such as the "traffic light" are useful. The "traffic light" format labels food as red, yellow, and green on the basis of the energy density of the foods (red foods being most calorie dense and green foods being least calorie dense). Children are encouraged to eat green foods more often and red foods rarely. Nutritional intervention strategies are strongly suggested to be combined (Poskitt, 2013).

Physical activity

Children and adolescents with obesity perform less physical activity than their normal-weight peers. Reducing sedentary activities and increasing physical activity have positive effects on body composition and obesity-associated metabolic comorbidities (Poskitt, 2013; Styne et al., 2017) and are highly recommended in available clinical practice guides. Programming of collective family physical activities is also recommended (Styne et al., 2017). To achieve the first objective, the removal of the television and/or computer

from the child's room is recommended (Seral-Cortés et al., 2021; Styne et al., 2017). The time and type of sedentary activity should be limited and previously programmed with the parents. Walking to all destinations and avoiding the use of mechanical means are recommended when possible. The type of physical activity should be adapted to the child's age and be attractive and fun for him or her. In accordance with the acquisition of skills and improved physical condition, the duration and intensity of the activity can be gradually increased, including team activities.

It is recommended that children 6 years or older participate in 60 min or more of physical activity per day with a minimum of 20 min of moderate to vigorous physical activity daily (Kumar and Kelly, 2017; Styne et al., 2017).

Unstructured physical activity, including outdoor play, should be encouraged in younger children, whereas older children should be encouraged to participate in structured physical activity such as after-school sports. It is also recommended that "screen time" (other than homework) be limited to less than 2 h/d for children older than 2 years, whereas those younger than 2 years should avoid "screen time" altogether. Because of the increasing evidence for an association between shortened sleep duration and obesity, good sleep hygiene and adequate amount of sleep (10–13 h a night for preschoolers and 8–10 h a night for teenagers) should be recommended (Styne et al., 2017).

Behavioral treatment strategies

Behavioral modification in lifestyle programs is an important part of weight loss programs. Behavioral strategies targeted at decreasing overall energy intake, decreasing sedentary time, and increasing physical activity are the objective of pediatric weight management. Parental participation has been shown to be effective in the treatment of childhood obesity and noted to give better results as compared with participation of the child alone.

Behavioral modification interventions include self-monitoring of food and physical activity as well as control of stimuli that contribute to or elicit unhealthy behaviors.

Unfortunately, behavioral interventions for treating overweight and obesity in children and youth are associated with a low to moderate treatment effect. The reported weight loss and BMI reductions have been modest, ranging from 1 to 3 kg/m^2 of BMI (Kumar and Kelly, 2017).

Pharmacologic treatment

The role of pharmacological therapy in the treatment of obesity in children and adolescents is limited.

Orlistat has been the only medication currently approved for the treatment of obesity in adolescents (age >12 years). Orlistat is a lipase inhibitor that blocks absorption of

about one-third of the fat ingested in a meal. The recommended dose of orlistat is 120 mg, 3 times a day with meals. Orlistat is also available as an over-the-counter medication at a lower dose of 60 mg, 3 times a day. The efficacy of orlistat is modest: 1-year placebo-subtracted changes in BMI less than 1 kg/m^2 (4% of weight loss) (Kumar and Kelly, 2017; Styne et al., 2017). Adverse effects limiting the use of orlistat include diarrhea, abdominal pain, flatulence, and greasy stools. Orlistat blocks absorption of fat-soluble vitamins and therefore administration of a multivitamin is recommended.

Metformin, a drug approved for the treatment of T2DM in children 10 years and older, has been used off-label for weight loss in several trials but results in modest reductions in BMI only (placebo-subtracted BMI decrease of 1.1–1.4 kg/m^2) (Kumar and Kelly, 2017; Styne et al., 2017).

Liraglutide was recently approved for the treatment of obesity in adolescents >12 years. Liraglutide is a glucagon-like receptor 1 agonist, which reduces food intake. In adults, treatment is associated with 6% of weight loss, 13% reduction in major adverse cardiovascular events and a 15% reduction in all-cause mortality (González-Muniesa et al., 2017).

Other medications that have been used off-label for the treatment of obesity in children include topiramate.

More trials are needed to better understand the efficacy and safety of these medications for the treatment of pediatric obesity.

Surgical treatment

There are few available data regarding bariatric surgery in children and adolescents and they are limited to the publication of case studies and expert consensus statements, with no controlled studies or data regarding the long-term results. The retrospective series analyzed by the Collaborative Group for the Study of Pediatric Bariatric Surgery is based on gastric bypass surgery, the most used technique in the United States, whereas in Europe and Australia adjustable gastric banding has been more widely used. Thus, specific recommendations regarding its use in children and adolescents cannot be made, and it should be limited to extreme cases and performed only in highly specialized centers (Kumar and Kelly, 2017; Styne et al., 2017).

For an adolescent to be candidate for bariatric surgery, expert committees require that, in addition to the anthropometric requirements (BMI > 40 kg/m^2) and the presence of severe associated comorbidities, the patient's maturity, both physical (estimated by bone maturation) and cognitive, as well as his/her capacity for decision making and the family structure should be evaluated to reduce possible adverse effects both during surgery and follow-up. An additional requirement includes previous failure of intense weight-loss programs during a minimum of 6 months (Styne et al., 2017). The percentage of children and adolescents who are candidates for bariatric surgery is very low.

New treatments

The fact that the administration of recombinant leptin to patients with leptin deficiency reverses their clinical picture opens the door to the development of new treatments for specific conditions, such as the alteration of MC4R (Martos-Moreno et al., 2014). To date, clinical trials data are limited, with moderate secondary effects (flushing, nausea, vomiting, headache, insomnia, and alterations in taste) being reported. There are new drugs that have been extensively reviewed in recent publications (González-Muniesa et al., 2017; Martos-Moreno et al., 2014).

Prevention of obesity during childhood

Although systematic reviews on the effectiveness of childhood obesity prevention programs have shown diverse and often conflicting results, some strategies were identified as useful; implementation of health education on nutrition, increased number of physical activity hours at school and the provision of healthy food at school. Limited evidence for the effectiveness of policies focusing on the child's environment and health information technologies focusing on children and parents is available. Performing interventions in primary care settings could improve parenting practices and promote healthy eating habits, but available studies showed only limited effect (Poskitt, 2013). The complexity of obesity requires multilevel and multicomponent interventions that take a systems approach. Multilevel approaches focus on changing health behaviors by acting on multiple frameworks: in practice, going from the individual, children or adolescents, schools or occupational environments and the communities. Multicomponent interventions incorporate more than one strategy to modify behaviors, within the same level. For the development of the interventions, health promotion models should consider information on the determinants, such as the target population and opportunities in the corresponding societies. Factors to be taken into account include maternal–child health, socioeconomic status, nutritional education and health literacy, parents' perceptions of healthy infant growth and development, peer pressure, family eating and physical activity behavior, and the role of food and built environments. Despite the many identified limitations of obesity prevention programs, policy efforts to prevent childhood obesity should continue.

Conclusion

Childhood obesity remains an ongoing serious international health concern. Childhood obesity has its basis in genetic susceptibilities influenced by an obesogenic environment starting in utero and extending through childhood and adolescence. Endocrine etiologies for obesity are rare and usually are accompanied by attenuated growth patterns. Pediatric comorbidities are common and long-term health complications often result. The prevention of pediatric obesity by promoting healthy diets, physical activity, and school and residential environments should be a primary goal, in order to achieve effective results. Although some behavioral and pharmacotherapy studies reported modest success, additional research into accessible and effective methods for preventing and treating pediatric obesity is much needed. Stratified treatment is recommended, with initial management in primary care with focus on healthy eating habits and active lifestyle. Those with poor response to interventions in primary care and with significant health risks should be managed by a multidisciplinary team with expertise in childhood obesity. Pharmacotherapy and/or bariatric surgery should be considered in adolescence if there has been no response to structured weight management with a multidisciplinary team. Additional research is needed to evaluate the efficacy and safety of these treatments. The role of genetics, epigenetics, nutrigenomics or personalized nutrition remains to be better understood to improve obesity management.

References

Aranceta-Bartrina, J., Gianzo- Citores, M., Pérez-Rodrigo, C., 2020. Prevalence of overweight, obesity and abdominal obesity in the Spanish population aged 3 to 24 years. The ENPE study. Revista Española Cardiología 73, 290–299. https://doi.org/10.1016/j.recesp.2019.07.011.

González-Muniesa, P., Mártinez-González, M.-A., Hu, F.B., Després, J.-P., Matsuzawa, Y., Loos, Ruth J.F., Moreno, L.A., Bray, G.A., Alfredo Martinez, J., 2017. Obesity. Nat. Rev. Dis. Prim. 3, 1–18.

Kumar, S., Kelly, A.S., 2017. Review of childhood obesity: from epidemiology, etiology, and comorbidities to clinical assessment and treatment. Mayo Clin. Proc. 92, 251–265. https://doi.org/10.1016/j.mayocp.2016.09.017.

Martos-Moreno, G.A., Barrios, V., Muñoz-Calvo, M.T., Pozo, J., Chowen, J.A., Argente, J., 2014. Principles and pitfalls in the differential diagnosis and management of childhood obesities. Adv. Nutr. 5, 299S–305S. https://doi.org/10.3945/an.113.004853.

Moreno, L.A., Rodriguez, G., Fleta, J., Bueno-Lozano, M., Lazaro Bueno, A.G., 2010. Trends of dietary habits in adolescents. Crit. Rev. Food Sci. Nutr. 50, 106–112. https://doi.org/10.1080/10408390903467480.

Ogden, C.L., Carroll, M.D., Kit, B.K., Flegal, K.M., 2014. Prevalence of childhood and adult obesity in the United States, 2011–2012. JAMA 311, 806–814.

Poskitt, E., 2013. Childhood obesity. In: Allen, L.H., Prentice, A., Caballero, B. (Eds.), Encyclopedia of Human Nutrition, third ed. Elsevier Ltd, London, pp. 336–342.

Rolland-Cachera, M.F., Deheeger, M., Maillot, M., Bellisle, F., 2006. Early adiposity rebound: causes and consequences for obesity in children and adults. Int. J. Obes. 30, S11–S17.

Seral-Cortés, M., Sabroso-Lasa, S., Bailo-Aysa, A., et al., 2021. Mediterranean diet, screen-time-based sedentary behavior and their interaction effect on adiposity in European adolescents: the HELENA study. Nutrients 13, 474. https://doi.org/10.3390/nu13020474.

Styne, D.M., Arslanian, S.A., Connor, E.L., et al., 2017. Pediatric obesity-assessment, treatment, and prevention: an endocrine society clinical practice guideline. J. Clin. Endocrinol. Metab. 102, 709–757. https://doi.org/10.1210/jc.2016-2573.

Obesity: Complications

R.L. Atkinson, Virginia Commonwealth University, VA, USA

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of R.L. Atkinson, Obesity: Complications, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 343–349, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00210-5.

Role of Distribution of Body Fat in the Complications of Obesity **591**
Metabolic and Organ System Complications of Obesity **591**
Metabolic Syndrome 591
Type 2 Diabetes 591
Insulin Resistance and Hyperinsulinemia 592
Dyslipidemia 593
Gout 593
Abnormalities of Hormones and Other Circulating Factors 593
Growth Hormone 593
The Hypothalamic–Pituitary–Adrenal Axis 593
Adipokines (Cytokines) 593
Renin–Angiotensin System 594
Leptin 594
Ghrelin and Obestatin 594
Diseases of Organ Systems **594**
Atherosclerotic and Arteriosclerotic Vascular Diseases 594
Coronary Heart Disease 594
Hypertension 594
Congestive Heart Failure 595
Cerebrovascular Disease 595
Thromboembolic Disease 595
Respiratory System 595
Digestive System 595
Gallbladder Disease 595
Hepatic Disease 595
Reproductive System 595
Hormonal Complications: Males 595
Hormonal Complications: Females 596
Obstetric Complications 596
Nervous System 596
Pseudotumor Cerebri 596
Adiposis Dolorosa 596
Alzheimer Disease 596
Immune System 596
Skin 596
Eye Disease 597
Cancer **597**
Mechanical Complications of Obesity **597**
Arthritis 597
Intraabdominal Pressure 597
Surgical Complications **598**
Psychosocial Complications **598**
Psychological Complications 598
Social Complications 598
Economic Impact 598
Further Reading **598**

Glossary

Adipokines Cytokines present in adipose tissue.
Adiposis dolorosa Painful fat syndrome.
Dyslipidemia Abnormal blood fats.
Hyperinsulinemia High blood insulin.
Visceral fat Intraabdominal adipose tissue.

Obesity is a serious chronic disease associated with complications and comorbidities that involve most systems of the body (Table 1). The common factor in all obese people is the presence of excess adipose tissue stores and an increased percent body fat. Even in the absence of complications and comorbidities, obesity increases the risk of early mortality. The estimates of obesity-related excess deaths in the USA each year range from 125 000 to 400 000. In addition to the medical complications, obesity is associated with psychological and social problems that may overshadow the medical problems in the quality of life for many obese people. This article will discuss the role of distribution of body fat in the complications of obesity, then will summarize the major categories of complications and comorbidities.

Role of Distribution of Body Fat in the Complications of Obesity

The distribution of excess adipose tissue contributes to many of the complications of obesity. Obese individuals may be divided into those whose excess fat predominantly is deposited in the upper body versus those with increased lower body obesity, or both upper and lower body obesity. Upper body obesity may be localized to the subcutaneous space versus the intraabdominal space (visceral fat). Waist circumference and the ratio of waist-to-hip circumferences correlate with the morbidity and mortality of obesity. Individuals with increased visceral fat, as measured by the cross-sectional area on computerized tomography (CT) or magnetic resonance imaging (MRI) scanning, are at greater risk for systemic complications of obesity compared to people with fat localized to abdominal subcutaneous depots or to the lower body. The mechanisms of these differences are not clear, but research has shown that visceral fat has a higher triglyceride turnover rate and releases greater amounts of fatty acids into the circulation than do other adipose tissue depots. Also, visceral fat has a higher concentration of adipokines (cytokines), substances that affect function of the immune system. Because blood vessels from the visceral fat drain into the portal vein, some investigators postulate that exposure of the liver to high levels of free fatty acids or these adipokines produces insulin resistance, which is known to be correlated with many of the complications of obesity described below. There are significant racial differences in deposition of visceral fat. Asians and Hispanics tend to selectively deposit fat in the abdominal cavity with excess energy intake whereas Blacks have less visceral fat than other groups.

Metabolic and Organ System Complications of Obesity

Obesity is a syndrome that resembles premature aging. Multiple metabolic, hormonal, and organ system dysfunctions occur in aging. Similar changes occur in obesity, but at an earlier age. Below we will review generalized metabolic changes that occur with obesity, and then examine individual organ systems.

Metabolic Syndrome

The term 'metabolic syndrome' has been given to a cluster of abnormalities that classically includes insulin resistance, glucose intolerance, hypertension, and dyslipidemia. Several other abnormalities such as sleep apnea, gout, and pseudotumor cerebri have been associated with insulin resistance and the 'metabolic syndrome.' The classic abnormalities will be reviewed below and the other abnormalities later in the article.

Type 2 Diabetes

Type 2 diabetes mellitus (DM): a strong association of obesity with the prevalence of Type 2 DM is well documented. The risk of developing Type 2 DM increases with the degree and duration of obesity; as much as 50-fold with severe obesity. The US National Diabetes Commission reported that the risk of diabetes doubles for every 20% of excess body weight. The risk of Type 2 DM is greater with visceral obesity. Type 2 DM is frequently associated with other complications such as hypertension and dyslipidemia, resulting in additive risks for atherosclerosis and cardiovascular disease. Poor glycemic control in Type 2 DM may lead to severe microvascular complications, including nephropathy, retinopathy, and neuropathy. Weight loss is a very effective treatment for Type 2 DM and can prevent the onset of Type 2 DM in susceptible individuals. Type 2 DM, once extremely rare in children, has increased greatly in prevalence with the obesity epidemic.

Table 1 Complications of obesity

1. Metabolic complications
   a. Metabolic syndrome
   b. Noninsulin-dependent diabetes
   c. Insulin resistance, hyperinsulinemia
   d. Dyslipidemia
   e. Gout
   f. Abnormalities of hormones and other circulating factors:
      1) Growth hormone (GH)
      2) The hypothalamic-pituitary-adrenal (HPA) axis
      3) Cytokines
      4) Renin–angiotensin system
      5) Leptin
      6) Ghrelin
2. Diseases of organ systems
   a. Cardiac and vascular diseases
      1) Coronary heart disease (CHD)
      2) Hypertension
      3) Congestive heart failure
      4) Cerebrovascular disease
      5) Thromboembolic disease
   b. Respiratory system abnormalities
      1) Obesity–hypoventilation syndrome
      2) Sleep apnea
   c. Digestive system abnormalities
      1) Gallbladder disease
      2) Hepatic disease
   d. Reproductive system abnormalities
      1) Hormonal complications: males
      2) Hormonal complications: females
      3) Obstetric complications
   e. Nervous system
      1) Pseudotumor cerebri
      2) Adiposis dolorosa
      3) Alzheimer disease
   f. Immune system dysfunction
   g. Skin diseases
   h. Eye disease
3. Cancer
   a. Breast
   b. Uterus
   c. Gallbladder
   d. Colon
   e. Prostate
   f. Others
4. Mechanical complications of obesity
   a. Arthritis
   b. Increased intraabdominal pressure
5. Surgical complications:
   a. Perioperative risks: anesthesia, wound complications, infections
   b. Incisional hernias
6. Psychosocial complications
   a. Psychological complications
   b. Social complications
   c. Economic Impact

Insulin Resistance and Hyperinsulinemia

Insulin resistance refers to the phenomenon of insensitivity of the body cells to insulin's actions. Insulin resistance is usually associated with hyperinsulinemia. Different tissues may have differential insulin sensitivities. For example, adipose tissue may be more sensitive to insulin than muscle tissue, thus favoring the deposition of fatty acids in adipose tissue and diminished fatty acid

oxidation in muscle. There is a reduced efficiency of insulin to inhibit hepatic glucose production and stimulate glucose use in skeletal muscle and adipose tissue that leads to hyperglycemia.

Hyperinsulinemia is an independent marker that predicts the development of atherosclerosis. A causal relationship between hypertension and hyperinsulinemia has not yet been well established. Hypertension associated with hyperinsulinemia could be due to increased renal sodium retention, increased intracellular free calcium, increased sympathetic nervous system activity, or increased intraabdominal pressure because of increased visceral fat deposition.

The mechanisms of insulin resistance with increasing obesity are not clear, but increased production of cytokines such as tumor necrosis factor-α (TNF-α) and interleukin-6 (IL-6) are thought to play a role. Basal insulin levels increase with the degree of overweight, perhaps due to increased insulin secretion or reduced clearance by the liver. A reduced receptor number or post insulin-receptor defects may play a role in insulin resistance. Both basal hyperinsulinemia and insulin resistance decrease with weight reduction.

Dyslipidemia

Obesity, particularly visceral obesity, is associated with increased serum levels of cholesterol, triglycerides, low-density lipoproteins (LDLs), very-low-density lipoproteins (VLDLs), apolipoprotein B, and reduced levels of high-density lipoprotein-cholesterol (HDL-c). There is a change in the composition of LDLs to increased small, dense LDLs, which are more atherogenic. The lower amounts of HDL-c compromise antiatherogenic function. Every 10% increase in relative bodyweight is associated with a 12 mg dl^{-1} increase in serum cholesterol concentration. The correlation of serum cholesterol with body mass index (BMI=kg m^{-2}) is greater for men than women. Increased serum triglycerides with weight gain may be due to increased intake of fats, hyperinsulinemia, and impaired removal of triglycerides into tissues because of low levels of lipoprotein lipase activity. Insulin resistance promotes lipolysis and increased circulating free fatty acids, which enhance the formation of VLDLs in the liver. Dyslipidemia contributes to increased atherosclerosis in obesity. Weight reduction usually reduces triglycerides, changes the composition of serum cholesterol to less small, dense LDLs, and increases HDL-c. These changes, if maintained over time, may reduce atherosclerosis.

Gout

Serum uric acid and the prevalence of gout correlate positively with BMI. High serum uric acid levels correlate with insulin resistance and an increased risk of atherosclerotic cardiovascular disease in obesity. Serum uric acid levels may temporarily increase with acute weight loss, but usually decrease with large amounts of weight loss. The lower uric acid levels are maintained with continued weight loss.

Abnormalities of Hormones and Other Circulating Factors

Multiple different hormones are altered in obesity. Below are a few that are felt to be important in altering the physiology and contributing to the complications of obesity.

Growth Hormone

Obesity is typically accompanied by a decrease in growth hormone (GH) levels and an increase in GH-binding protein (GHBP) levels. GH is released by the anterior pituitary and affects lipid, carbohydrate, and protein metabolism. An inverse relation exists between GH levels and percent fat mass. GH is lipolytic in adipose tissue. Animal studies show enhanced catecholamine-induced lipolysis and increased beta adrenoreceptors in adipocytes of GH-treated animals. The rises of GH after meals, with sleep, and in response to secretogogues such as arginine or levodopa are blunted in obese people. GH stimulates secretion of insulin-like growth factor-1 (IGF-1). However, IGF-1 is increased in obesity, suggesting a difference in sensitivity to GH. The defects in GH and IGF-1 are reversed by weight reduction.

The Hypothalamic–Pituitary–Adrenal Axis

The hypothalamic–pituitary–adrenal (HPA) axis may be abnormal in obesity, with similarities to patients with Cushing's syndrome such as insulin resistance, impaired glucose homeostasis, hypertension, and lipid abnormalities. High levels of emotional or physical stress are thought to increase cortisol secretion or turnover, and thereby increase visceral obesity. Another potential mechanism involves the peripheral metabolism of cortisol. The enzyme 11 beta-hydroxysteroid dehydrogenase-1 (11 beta-HSD-1), which converts steroid precursors to cortisol, is expressed in adipose tissue. With increasing obesity, more cortisol is derived from cortisone in adipose tissue due to the increased activity of this hormone. Urine studies in obesity also show an increase in the ratio of tetrahydrocortisol to tetrahydrocortisone, indicating a relative increase in the pathways leading to cortisol formation.

Adipokines (Cytokines)

Adipose tissue secretes a number of cytokines such as TNF-α, ILs, plasminogen activator inhibitor-1 (PAI-1), and retinol-binding protein 4 (RBP-4) that may play a role in fat metabolism and insulin resistance. Adipokine secretion is higher with obesity. TNF-α has been shown to alter basal and glucose-stimulated insulin secretion and produce insulin resistance in isolated cell lines. Adipocytes also produce IL-6, 10, and 11, which stimulate C-reactive protein, a systemic marker of inflammation. All of these ILs are

increased in obesity. IL-6 and PAI-1 have been postulated to play an etiologic role in the increased risk of thromboembolism observed in obese patients. Plasma IL-8 is increased in normoglycemic obese subjects, and is related to fat mass and TNF-α levels. Circulating IL-8 is also acutely up-regulated by hyperinsulinemia. An increase in circulating IL-8 may be one of the factors linking obesity with greater cardiovascular risk. Adiponectin is the only adipokine whose circulating concentration decreases with increased fat mass and, in line with this, is inversely correlated with the metabolic syndrome.

Renin–Angiotensin System

Several components of the renin–angiotensin system are expressed by the adipose tissue. Angiotensinogen levels are increased and have been linked to hypertension and increased cardiovascular risk in obesity.

Leptin

Leptin is a cytokine made predominantly in the adipose tissue. Leptin binds to receptors in the hypothalamus to signal the state of adipose tissue stores and act by inhibiting neuropeptide Y. Serum leptin levels correlate positively with body fat stores, and are higher in obese people, denoting leptin resistance in obesity. Females have higher serum leptin levels than males, but this association does not appear to be due to estrogen levels. Leptin is found in greater concentrations in abdominal subcutaneous fat compared to visceral fat, and some studies link leptin to regulation of energy expenditure.

Ghrelin and Obestatin

Ghrelin and obestatin come from the same prohormone, preproghrelin. Ghrelin is a potent GH secretagogue that is produced mainly by the stomach and is present in several forms in serum. Administration of ghrelin increases food intake, and ghrelin levels increase with dieting and weight loss. However, serum ghrelin has a negative correlation with percent body fat, so levels in obese people are lower than in leans. Obestastin is thought by some to have an anorexic effect but its function and the interplay with ghrelin is not well understood.

Diseases of Organ Systems

Atherosclerotic and Arteriosclerotic Vascular Diseases

Diseases of the vascular system make the greatest contribution to the increased mortality associated with obesity. In both sexes, the excess mortality due to vascular disease increases linearly with BMIs greater than 25 kg m^{-2}. The vascular complications of obesity can be categorized into four major groups:

1. Coronary heart disease (CHD)
2. Hypertension
3. Congestive heart failure
4. Cerebrovascular disease

Coronary Heart Disease

Longitudinal studies show a positive correlation of BMI with CHD, and obesity is an independent predictor of CHD. However, in the presence of other risk factors such as hypertension, lipid abnormalities, and insulin resistance, all of which are increased by obesity, the risk of atherosclerotic CHD increases dramatically. Weight loss reduces all of these risk factors associated with cardiovascular disease, but there are only few long-term studies of changes in cardiovascular mortality due to weight loss. A very low-fat diet (10% of total calories as fat) has been shown to reduce the size of atherosclerotic plaques in coronary arteries.

Hypertension

The prevalence of hypertension among overweight adults in the USA is 2.9 times higher than that of nonoverweight individuals. Every 10 kg increase in bodyweight is associated with increases of 3 and 2 mm Hg in systolic and diastolic blood pressures, respectively. Persistent hypertension can contribute to the development of left ventricular hypertrophy, coronary ischemia, and stroke.

The etiology of the association between hypertension and obesity is unclear. Some of the mechanisms offered to explain the association of obesity and hypertension are:

1. Hyperinsulinemia due to insulin resistance leading to increased renal reabsorption of sodium.
2. Sodium and fluid retention due to a decreased renal filtration rate, increased intraabdominal pressure, or increased plasma renin activity.
3. Increased sympathetic nervous system activity.
4. Increased cytokines and inflammation.

Except in long-standing cases, weight reduction is usually accompanied with a decrease in blood pressure. The reductions in blood pressure with weight loss are not dependent on decreases in salt intake. Many studies have shown that even modest weight losses, in the range of 5–10% of initial bodyweight, may produce reductions or even normalization of blood pressure in obese individuals.

Congestive Heart Failure

Total blood volume increases with excess body weight. Higher oxygen consumption in obesity and increased blood flow to the splanchnic bed and adipose tissue increase cardiac output. Also, the transverse diameter of heart, thickness of the posterior wall, and thickness of the interventricular septum increase with body weight. Left ventricular mass is a stronger predictor of morbidity and mortality than blood pressure. A combination of these factors may result in the congestive heart failure seen in severely obese people. The heart rate, stroke volume, blood volume, cardiac output, and left ventricular work return toward normal with weight reduction. One study that compared weight loss by dieting to treatment with antihypertensive drugs demonstrated a greater improvement in cardiac hypertrophy with weight loss, despite similar reductions in blood pressure.

Cerebrovascular Disease

Obesity-related atherosclerosis, arteriosclerosis, and hypertension increase the risk of cerebrovascular disease and strokes. Obesity is an independent risk factor for strokes, even in the absence of other comorbidities.

Thromboembolic Disease

The risks of venous stasis, deep vein thrombosis, and pulmonary embolism are increased in obesity, particularly in persons with abdominal obesity. Lower extremity venous disease may result from increased intraabdominal pressure, impaired fibrinolysis, and the increase in inflammatory mediators described above.

Respiratory System

Obesity is associated with reduced lung volume, altered respiratory patterns, and an overall reduction in the compliance of the respiratory system, including a diminished vital capacity and total lung capacity. More severe obesity is associated with the obesity–hypoventilation syndrome, which is characterized by excessive daytime sleepiness and hypoventilation. The increased work required to move the chest wall, a decrease in arterial oxygenation in the lungs, and a diminished sensitivity of the respiratory center to the stimulatory effect of carbon dioxide are postulated to contribute to the obesity–hypoventilation syndrome.

The obesity–hypoventilation syndrome may be associated with, or exacerbated by, obstructive sleep apnea, a syndrome characterized by repeated collapse of the upper airway and cessation of breathing with sleep. Obstructive sleep apnea occurs when the tongue obstructs the glottis and prevents entry of air into the trachea. Up to 50% of massively obese people have sleep apnea. The risk of arrhythmias and sudden death increases during apneic episodes. Weight reduction usually reduces the severity of sleep apnea, and massive weight reduction, such as that after gastric bypass surgery, eliminates the disease in most patients.

Digestive System

Gallbladder Disease

The risk of gallbladder disease, particularly gallstone formation, is increased in obesity, and occurs with greater frequency in women. The prevalence of gallbladder disease in obese individuals increases with age, bodyweight, and parity. The etiology of increased gallstones is unclear, but genetic factors play a role. Increased cholesterol production, which leads to increased excretion of cholesterol in bile, is known to occur in obesity and correlates with increases in body weight. Many obese people skip meals and the reduced number of meals may result in less frequent emptying of the gallbladder. The resulting bile stasis may contribute to gallstone formation. Although long-term weight loss and maintenance may reduce the occurrence of gallbladder disease, the risk of gallstone formation actually increases during the active weight loss phase. The etiology of this increase is thought to be the mobilization of cholesterol from adipose tissue during rapid weight loss. This increased load of cholesterol in the circulation produces supersaturation of the bile, leading to gallbladder sludge in approximately 25% of patients and to symptomatic disease in approximately 1–3%. Treatment with ursodeoxycholic acid reduces or eliminates the risk of gallstone formation during weight loss.

Hepatic Disease

Abnormalities in hepatic function are commonly reported in obese people. The frequency of fatty liver has been reported to be as high as 94% in very obese subjects, many of whom have elevated liver function tests. A small number of very obese subjects will develop micronodular cirrhosis. Weight loss results in disappearance of the excess fat and normalization of liver function tests.

Reproductive System

Hormonal Complications: Males

Obese men have elevated levels of plasma estrone and estradiol that correlate with the degree of obesity. Plasma total testosterone and free testosterone (the biologically active moiety) are reduced in obese men, and the reductions correlate negatively with the degree of obesity. The reduced levels of free and total testosterone are not generally accompanied by hypogonadism or a decrease in libido, potency, or sperm count in obese men. Free and total plasma testosterone levels normalize on significant weight reduction. Also, estrogen levels are normalized if the individuals attain normal weight, but not if the weight loss is modest and significant obesity persists.

Hormonal Complications: Females

Obese women have normal levels of total plasma estradiol, but reduced levels of sex hormone-binding globulins (SHBGs). Thus, free estradiol (the biological active moiety) is significantly elevated. The high levels of free estradiol are postulated to increase the risks of endometrial and breast cancer and to reduce fertility. Estrone, derived in adipose tissue from androgen precursors, is also increased in obesity. Obesity in women is associated with the polycystic ovary syndrome (PCOS), characterized by hyperestrogenism, hyperandrogenism, polycystic ovaries, oligomenorrhea or amenorrhea, hirsutism, and infertility. Women with PCOS also have insulin resistance and are at high risk for developing impaired glucose tolerance and DM. Weight loss usually normalizes SHBG and estradiol levels for individuals with simple obesity, but weight loss may not restore fertility to patients with severe PCOS.

Obstetric Complications

Obesity increases the risk of complications during pregnancy and childbirth. Increased bodyweight, hypertension, and fluid retention during pregnancy can lead to toxemia of pregnancy. Heavier women have a longer duration of labor and a greater frequency of abnormal labor and caesarian sections.

Nervous System

Pseudotumor Cerebri

This syndrome is characterized by increased intracranial pressure, headaches, blurred vision or loss of vision, and papilledema. It is most common in massively obese individuals and may be seen in association with sleep apnea or with the obesity-hypoventilation syndrome. It may be associated with retinal hemorrhage or loss of vision from severe papilledema. Some investigators believe that increased intraabdominal pressure with massive obesity is an etiologic factor for pseudotumor cerebri. Major weight loss, particularly after obesity surgery, results in dramatic improvement.

Adiposis Dolorosa

This is a syndrome of unknown etiology characterized by pain in subcutaneous adipose tissue. Adiposis dolorosa occurs predominantly in postmenopausal women (female to male ratio of approximately 30:1), and has been described over all areas of the body. The painful areas of fat may occur as subcutaneous lumps on physical examination, but more commonly there are no differences from normal adipose tissue. The disease usually begins gradually with mild pain and tenderness of the area involved, but may progress to severe pain, particularly with movement or exercise. Intravenous infusions of lidocaine are reported to relieve pain, short term or even permanently. The mechanism involved in the relief of pain from lidocaine is unknown.

Alzheimer Disease

Obesity has been linked to an increased prevalence of Alzheimer disease. The etiology of this increase is unknown.

Immune System

Animal studies have shown an increased rate of infection and mortality in obese dogs compared to lean animals experimentally infected with canine distemper virus. Cell-mediated immune response is impaired in obese individuals. Maturation of monocytes into macrophages after *in vitro* incubation is significantly less for obese compared to lean subjects. Impaired cell-mediated immune response in children was demonstrated to be due to subclinical deficiencies of zinc and copper. The impairment in the immune response was reversed after 4 weeks of zinc and copper supplements. As described above, there are changes in numerous cytokines with obesity. The role of these changes in immune function is not clear.

Skin

Obese people may have several disorders of the skin. The most common is stasis changes of the skin of the lower legs in massively obese people. The etiology of this finding is venous stasis, edema, and breakdown of the skin. Fragilitas cutis inguinalis is a condition of fragile skin in the inguinal area of obese people. This condition is diagnosed by stretching the skin of the inguinal area. A linear tear appears at right angles to an applied force that is insufficient to tear the skin of a normal person. This condition is unrelated to the sex and age of the person.

Acanthosis nigricans, seen occasionally in obesity, is characterized by darkening of the skin in the creases of the neck, axillary regions, and over the knuckles. An association between acanthosis nigricans and insulin resistance is reported in persons who have circulating antibodies to the insulin receptors. Because acanthosis nigricans also may be associated with highly malignant cancers such as intraabdominal adenocarcinomas, physicians should be alert to this possibility and not attribute the condition simply to the presence of obesity.

Eye Disease

Obesity is associated with an increased prevalence of cataracts. Persons with abdominal obesity are at a greater risk than those with lower body obesity. Insulin resistance may be involved in the pathogenesis of cataract formation, and diabetes is a well-known risk factor.

Cancer

Obesity increases the risk of many cancers, including breast, colon, prostate, endometrium, cervix, ovary, kidney, gallbladder, liver, pancreas, rectum, brain, esophagus, and non-Hodgkins lymphoma. Table 2 shows the increased risk of selective cancers.

Although there are many theories about how obesity increases cancer risk, the exact mechanisms are not known. The mechanisms may be different for different types of cancer. Also, because obesity develops through a complex interaction of heredity and lifestyle factors, researchers may not be able to tell whether the obesity or other factors led to the development of cancer.

Mechanical Complications of Obesity

Arthritis

Obesity is frequently complicated by degenerative arthritis (DJD). Increased bodyweight leads to trauma of the weight-bearing joints and speeds the development of osteoarthritis in obesity. Knee and hip joints are particularly affected. However, obese patients have increased DJD of the hands, perhaps due to cytokines produced by adipose tissue, which may damage the cartilage in joints. Flattening of the arc of the planter surface of the feet (flat feet) occurs more frequently in obese people, presumably due to the stress of carrying excess body weight. Flat feet may lead to unsteady gait and aches and pains after walking. Increased fat deposition, particularly in the abdominal region, can change the natural curvature of the spine, causing lordosis and resulting in backache in obese people.

Intraabdominal Pressure

In severely obese people, the excess visceral fat is thought to increase intraabdominal pressure. Animal research shows that experimentally induced acute increases in intraabdominal pressure to the levels seen in the abdomens of very obese people cause increases in pleural pressure, intracranial pressure, and central venous pressure. The investigators postulated that in humans, increased intraabdominal pressure may contribute to hypertension, insulin resistance and Type 2 DM, obesity–hypoventilation syndrome, pseudotumor cerebri, incisional hernia, and urinary incontinence. Massive weight loss following obesity surgery normalizes the increased intraabdominal pressure and reduces or eliminates all the symptoms listed above.

Table 2 Deaths from cancer in women by BMI

	BMI (kg m^{-2})					
Type of cancer	*18.5–24.9*	*25.0–29.9*	*30.0–34.9*	*35.0–39.9*	*≥40.0*	*P for trend*
All cancers						
Deaths per 10 000 women[a]	329.3	339.75	382.62	419.59	522.51	
RR (95% CI)[b]	1	1.08	1.23	1.32	1.62	<0.001
Breast cancer						
Deaths per 10 000 women[a]	39.1	51.13	60.65	67.56	84.86	
RR (95% CI)[b]	1	1.34	1.63	1.7	2.12	<0.001
Uterine cancer						
Deaths per 10 000 women[a]	10.68	15.68	26.05	30.16	60.83	
RR (95% CI)[b]	1	1.5	2.53	2.77	6.25	<0.001
Ovarian cancer						
Deaths per 10 000 women[a]	27.88	31.44	31.85	44.49	–	
RR (95% CI)[b]	1	1.15	1.16	1.51	–	<0.001
Colorectal cancer						
Deaths per 10 000 women[a]	38.67	43.28	53.81	56.14	63.11	
RR (95% CI)[b]	1	1.1	1.33	1.36	1.46	<0.001

Source: Adapted with permission from Calle EE, Rodriguez C, Walker-Thurmond K, and Thun MJ (2003) Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *New England Journal of Medicine* 348(17): 1625–1638.

[a]Deaths per 100 000 women.

[b]RR, Relative risk of death compared to the reference group (BMI=18.5–24.9 kg m^{-2}).

Surgical Complications

Obese patients are at an increased risk of surgical and perisurgical complications. These risks include an increased risk of complications and death from anesthesia, longer operating times, delayed wound healing, increased postoperative wound infections and pneumonia, and a higher frequency of incisional hernias after surgeries involving the abdominal wall. Many surgeons recommend weight reduction before elective surgery, but there are few data to document that acute weight reduction improves the outcome of surgery.

Psychosocial Complications

Psychological Complications

Obesity is associated with negative emotions, low self-esteem, decreased marital satisfaction, and body image disparagement. All of these conditions and beliefs show improvement with weight reduction.

Dieting efforts correlate positively with the prevalence of eating disorders, particularly binge eating. A correlation of eating disorders with abuse of drugs and alcohol has been shown. In strictly dieting female college freshmen who were not alcohol abusers at baseline, the frequency of alcohol abuse was reported to increase after a year compared to nondieters.

Social Complications

Obesity carries a social stigma that dramatically affects the quality of life for obese individuals, particularly for women. Factors contributing to the social bias against obese people are beliefs that obesity is merely due to overeating and therefore obese people lack will power. Many members of the general public, and even health professionals, ignore the evidence for the medical and genetic contribution to obesity, believe that obese people are responsible for their own plight, and believe that they do not deserve sympathy for their disability. Despite similar intelligence (as judged by IQ values and the Scholastic Aptitude Test scores), a significantly lower number of obese females were admitted to certain colleges compared to nonobese females. The choice of mates is adversely affected by obesity. Obese individuals tend to marry mates with less education and from a lower socioeconomic class. It is more difficult for an obese person to find a job or to be promoted once hired, so lower earnings and a lower socioeconomic status are correlated with obesity. Obese employees are viewed as less competent, less productive, inactive, disorganized, and less successful by employers, regardless of actual productivity.

The bias against obesity has been shown to begin in early childhood. Obese children are considered lazy, stupid, slow, and self-indulgent by both children and adults. Because of these societal attitudes, many obese children and adolescents have lower levels of self-esteem than do their nonobese counterparts

Economic Impact

In the USA, the direct cost of obesity has been estimated at almost 150 billion dollars per year. The indirect costs of early retirement and increased risk for disability requiring financial support are also considerable. Because obese people have more health problems, health-care costs for the obese are higher than for lean individuals.

Further Reading

Anandacoomarasamy, A, Caterson, I, Sambrook, P, Fransen, M, March, L, 2008. The impact of obesity on the musculoskeletal system. International Journal of Obesity 32 (2), 211–222.

Atkinson, RL, 1982. Intravenous lidocaine for the treatment of intractable pain of adiposis dolorosa. International Journal of Obesity 6, 351–357.

Berrington de Gonzalez, A, Hartge, P, Cerhan, JR, et al., 2010. Body-mass index and mortality among 1.46 million white adults. The New England Journal of Medicine 363 (23), 2211–2219.

Calle, EE, Rodriguez, C, Walker-Thurmond, K, Thun, MJ, 2003. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. New England Journal of Medicine 348 (17), 1625–1638.

Catalano, PM, 2010. Obesity, insulin resistance, and pregnancy outcome. Reproduction 140 (3), 365–371.

Hevener, AL, Febbraio, MA, Stock Conference Working Group, 2010. The 2009 stock conference report: Inflammation, obesity and metabolic disease. Obesity Reviews 11 (9), 635–644.

Hunley, TE, Ma, LJ, Kon, V, 2010. Scope and mechanisms of obesity-related renal disease. Current Opinion in Nephrology and Hypertension 19 (3), 227–234.

Jornayvaz, FR, Samuel, VT, Shulman, GI, 2010. The role of muscle insulin resistance in the pathogenesis of atherogenic dyslipidemia and nonalcoholic fatty liver disease associated with the metabolic syndrome. Annual Review of Nutrition 30, 273–290.

Mathieu, P, Lemieux, I, Després, JP, 2010. Obesity, inflammation, and cardiovascular risk. Clinical Pharmacology and Therapeutics 87 (4), 407–416.

Moore, JB, 2010. Non-alcoholic fatty liver disease: The hepatic consequence of obesity and the metabolic syndrome. Proceedings of the Nutrition Society 69 (2), 211–220.

Murugan, AT, Sharma, G, 2008. Obesity and respiratory diseases. Chronic Respiratory Disease 5 (4), 233–242.

Pi-Sunyer, X, 2009. The medical risks of obesity. Postgraduate Medicine 121 (6), 21–33.

Shelton, RC, Miller, AH, 2010. Eating ourselves to death (and despair): The contribution of adiposity and inflammation to depression. Progress in Neurobiology 91 (4), 275–299.

Sugerman, HJ, Felton 3rd, WL, Salvant Jr, JB, Sismanis, A, Kellum, JM, 1995. Effects of surgically induced weight loss on idiopathic intracranial hypertension in morbid obesity. Neurology 45, 1655–1659.

Withrow, D, Alter, DA, 2011. The economic burden of obesity worldwide: A systematic review of the direct costs of obesity. Obesity Reviews 12 (2), 131–141.

Obesity: Definition, etiology, and assessment

S. Hawkesworth, London School of Hygiene and Tropical Medicine, London, UK

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of S. Hawkesworth, Obesity: Definition, Etiology, and Assessment, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 350–353, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00211-7.

Definition 599
Health Impacts 601
Etiology 601
Assessment 601
Measures of Body Fatness 602
Cadaver Analysis 602
Imaging Techniques 602
Densitometry 602
Bioimpedance Analysis 602
Anthropometry 602
Further Reading 602

Glossary

Body composition Relative proportion of tissue types within an individual. At its simplest level the body can be divided into the relative proportion of fat and fat-free mass.

Body mass index (BMI) Index of weight independent of height used as a proxy measure of body fatness; calculated as weight(kg)/height(m)^2.

Metabolic syndrome A cluster of risk factors (including abdominal obesity) associated with increased risk of cardiovascular disease and diabetes.

Nutrition transition Population-level shifts in diet and physical activity patterns often associated with modern lifestyles; in the latter stages characterized by high-fat/high-sugar/low-fibre diets and lower levels of activity.

Obesity Excess accumulation of body fat, which can be associated with a range of co-morbidities.

Definition

Obesity is the excess accumulation of body fat or adiposity. It is commonly and most easily assessed by individuals' body weight independent of their height; the body mass index (BMI), calculated as weight (kg)/height (m)^2. The World Health Organization (WHO) classifies individuals with a BMI of 25–29.9 kg m^{-2} as overweight and over 30 kg m^{-2} as obese (Table 1). These classifications are themselves based on increased mortality and morbidity associated with excess body weight for height. Although the WHO recommends the use of these International Classification cut-offs for all countries there is evidence that the relationship between BMI and health may vary for different populations partly as a consequence of different associations with body fat and fat distribution. Lower BMI cut-offs for Asian populations have been suggested although the relationships are also not uniform across ethnic groups. Additional cut-off points are therefore suggested as a useful public health reference in certain settings (Table 1).

Women are predisposed to higher body weight for the purposes of reproduction and in all populations mean BMI is higher for females compared to males. Data from 2010 collated by the International Association for the Study of Obesity (IASO) indicate that globally more than 1 billion individuals were overweight and 475 million obese. Rapid increases in BMI and the prevalence of obesity have occurred across all world regions; data from 199 countries complied by the Global Burden of Metabolic Risk Factors of Chronic Diseases Collaborating Group indicate that between 1980 and 2008, mean BMI increased by 0.4 kg m^{-2} per decade for men and 0.5 kg m^{-2} per decade for women worldwide. These figures mask large variations in rate of change; certain regions, such as Oceania, experienced rates of increase of over 2 kg m^{-2} per decade during this time frame. The prevalence of overweight and obesity is now so high that the condition is increasingly recognized as a pandemic, with projections of further increases particularly within lower- and middle-income countries in the next decades. In children, overweight and obesity are assessed by age and sex specific cut-offs that predict adult BMI (Table 2). Globally, over 200 million school-aged children are classified as overweight and obese.

Table 1 International classification of adult overweight and obesity according to body mass index (BMI)

Classification	*BMI (kg m^{-2})*	
	Principal cut-off points	*Additional cut-off points*
Normal range	18.50–24.99	18.50–22.99 23.00–24.99
Overweight	25.00–29.99	25.00–27.49 27.50–29.99
Obese	≥30.00	
Obese class I	30.00–34.99	30.00–32.49 32.50–34.99
Obese class II	35.00–39.99	35.00–37.49 37.50–39.99
Obese class III	≥40.00	

Source: Modified from WHO Expert Consultation (2004) Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* 363: 157–163.

Table 2 International cut-off points for body mass index (BMI) for overweight and obesity by sex between 2 and 18 years

Age (years)	*BMI 25 kg m^{-2}*		*BMI 30 kg m^{-2}*	
	Males	*Females*	*Males*	*Females*
2	18.41	18.02	20.09	19.81
2.5	18.13	17.76	19.80	19.55
3	17.89	17.56	19.57	19.36
3.5	17.69	17.40	19.39	19.23
4	17.55	17.28	19.29	19.15
4.5	17.47	17.19	19.26	19.12
5	17.42	17.15	19.30	19.17
5.5	17.45	17.20	19.47	19.34
6	17.55	17.34	19.79	19.65
6.5	17.71	17.53	20.23	20.08
7	17.92	17.75	20.63	20.51
7.5	18.16	18.03	21.09	21.01
8	18.44	18.35	21.60	21.57
8.5	18.76	18.69	22.17	22.18
9	19.10	19.07	22.77	22.81
9.5	19.46	19.45	23.39	23.46
10	19.84	19.86	24.00	24.11
10.5	20.20	20.29	24.57	24.77
11	20.55	20.74	25.10	25.42
11.5	20.89	21.20	25.58	26.05
12	21.22	21.68	26.02	26.67
12.5	21.56	22.14	26.43	27.24
13	21.91	22.58	26.84	27.76
13.5	22.27	22.98	27.25	28.20
14	22.62	23.34	27.63	28.57
14.5	22.96	23.66	27.98	27.87
15	23.29	23.94	28.30	29.11
15.5	23.60	24.17	28.60	29.29
16	23.90	24.37	28.88	29.43
16.5	24.19	24.54	29.14	29.56
17	24.46	24.70	29.14	29.69
17.5	24.73	24.85	29.70	29.84
18	25	25	30	30

Source: Modified with permission from Cole TJ, Bellizzi MC, Flegal KM, and Dietz WH (2000) Establishing a standard definition for child overweight and obesity worldwide: International survey. *British Medical Journal* 320: 1240–1243.

Health Impacts

The WHO classifies overweight or obesity as the fifth leading risk factor for death, responsible for 7% of deaths globally. It is estimated that at 40 years of age, an obese person will live up to 7 years less than someone of normal weight. Obesity is associated with a range of comorbidities, many of which fall under the heading of the 'metabolic syndrome,' which is characterized by insulin resistance and dyslipidemia as well as by excess body weight and by increased body fat in the abdominal region. The risks of diabetes, hypertension, and dyslipidemia all increase with increasing body weight from a BMI as low as 21 kg m^{-2}. One of the closest associations between obesity and disease is with type II diabetes, where up to 90% of individuals with the condition are classified as overweight. Hypertension is up to five times higher in individuals who are overweight or obese. Excess body fatness is convincingly associated with a range of cancers and reducing obesity was one of the key recommendations of the 2007 World Cancer Research Fund report on cancer prevention. Other important comorbidities associated with body fatness include gall-bladder disease, nonalcoholic fatty liver disease, sleep apnea, and osteoarthritis. Increasingly, obesity is associated with psychological effects and stigma, which have important implications for quality of life. Owing to the wide range of associated morbidities, the direct and indirect costs of obesity are substantial and the health-care costs of treatment alone confer a massive burden on societies across the world. Middle- and lower-income countries are projected to experience the greatest rise in obesity rates but may lack the resources to cope with associated health costs.

Etiology

Obesity is primarily caused by an imbalance between energy intake and expenditure, leading to weight gain. The components of the diet and the amount of energy expended are thus important determinants. From a substantial review of the evidence in 2007, the World Cancer Research Fund concluded that there was convincing or probable evidence that a sedentary lifestyle (including excess television viewing) and the consumption of energy-dense foods, sugary drinks, and fast foods were associated with increased risk of weight gain, overweight, and obesity.

There is increasing recognition of the role played by societal and environmental forces in influencing both sides of the energy balance equation and obesity is no longer viewed solely as preventable by the individual. The rapid rise in obesity rates across the globe has been attributed to correspondingly rapid changes in population demographics and dietary behavior. The so-called 'nutrition transition' from traditional forms of diet and activity to more 'Western' diets and lifestyles, characterized by increased energy density and high fat is likely to be a particularly important driver of obesity rates. Increasing urbanization is occurring across the globe, often with corresponding increases in obesity-inducing behavior such as a reduction in physical activity and access to highly processed urban food supplies.

Patterns of overweight and obesity are not uniform within populations and are strongly related to socioeconomic status. As countries become wealthier this pattern changes so that in the countries with the highest income, obesity rates are the highest amongst the poorest members of society whereas in lower-income countries the prevalence of obesity is highest amongst the highest socioeconomic class. This pattern reflects the changing distribution of obesity risk factors that occurs as countries progress through the nutrition transition.

The relative contributions of genetics and the environment to the etiology of obesity have been evaluated in several studies with inconsistent results and estimates of the inheritance of obesity range from 64% to 84%. Only very few individuals have genetically determined obesity relating to single-gene effects but complex gene–environment interactions are likely to predispose certain individuals to obesity within the context of particular diet and activity patterns. Investigation of the polymorphisms that underlie these interactions may help to understand the etiology of obesity and inform potential treatment and prevention strategies. However it is important to recognize that changes to the environment, rather than to genes, underlie the current worldwide obesity epidemic and addressing these environmental causes will be the most important component of effective prevention strategies.

Assessment

One of the goals of assessment of overweight/obesity is to decide whom to treat. Three main issues must be evaluated: whether treatment is indicated, whether treatment is safe for the patient, and whether the patient is ready and motivated to lose weight. In addition, routine assessment of eating and activity patterns in adults as well as in children must be considered. Recognition of excessive weight gain relative to linear growth is essential throughout childhood. Motivation for weight loss in obese individuals is a key component of effective treatment but is often lacking partly because it depends on the acceptance and recognition that obesity is a medical disorder. Motivation for change may also be reduced if the obese individuals are yet to develop comorbidities and do not yet conceive that their health is compromised.

Proper identification and classification of obesity through body composition assessment are important steps to initiate before beginning weight-loss treatment. Dietary management, physical activity, surgery, pharmacotherapy, and psychosocial and familial support must be considered together as part of obesity assessment. Before beginning a weight-loss program, patients should be evaluated for the number and severity of cardiovascular risk factors, conditions that may require treatment in addition to weight-loss strategies.

Measures of Body Fatness

Although BMI is a useful indicator of risk that is widely used at a population level, it is only a proxy measurement of body fatness and does not accurately reflect adiposity at an individual level. In addition, individuals with the same BMI may have very different body shapes depending on the distribution of fat; visceral fat in the abdominal region is associated with greater metabolic abnormalities. Many different techniques can be used to assess body fatness, each with its own advantages and disadvantages.

Cadaver Analysis

The only direct measure of body fat but clearly not widely applicable. The main use of cadaver studies is to validate methods that can be used to study patients *in vivo*. All remaining techniques are indirect assessments that require certain assumptions to calculate body fatness.

Imaging Techniques

Total adipose tissue and its distribution can be quantified using imaging techniques such as computed tomography (CT) and magnetic resonance imaging (MRI). Both methods produce high-resolution cross-sectional images from signals resulting from exposure of the subject to an X-ray source (CT) or electromagnetic field (MRI). Total body fat volume, total fat mass, and percentage fat mass can be estimated and visceral fat depots can be measured with good accuracy. These techniques are very expensive and may be problematic for people with claustrophobia. Dual-energy X-ray absorptiometry (DXA) is a widely used imaging technique that utilizes the principle that transmitted X-rays at two energy levels are differentially attenuated by bone mass and soft tissue mass. It is possible to obtain abdominal fat estimates with DXA, although these cannot be separated into subcutaneous and visceral components.

Densitometry

Total body fat can be assessed by measuring body density based on the Archimedes principles of water displacement, assuming two body compartments (fat and fat-free tissue) of distinct densities. Body density was traditionally measured using the technique of under-water weighing, which is both expensive and not widely acceptable to patients. Air displacement techniques are becoming more popular although these techniques also require a participant burden and may be problematic for individuals with claustrophobia.

Bioimpedance Analysis

This predictive technique of assessing body fatness is becoming widely used for individuals and research studies. The measurement is based on the relative impedance of tissues to a small electric current that is passed through the body. The impedance is converted into an estimate of total body water, which is used to calculate fat-free mass and then fat mass by difference from body weight. Because the estimate of total body water is crude, the estimation of fat mass by this technique is relatively weak.

Anthropometry

In addition to height and weight measurements for the calculation of BMI, other anthropometric assessments can provide useful predictions of body fatness and metabolic risk. Body fat can be estimated by measuring skinfold thickness directly using a caliper at different sites on the body. These often include the upper arm (biceps and triceps), under the scapular (subscapular), and above the iliac crest (suprailliac). The raw data provide a direct estimate of body fat, with an increasing number of measurement sites correcting for differences in fat distribution. Data can also be used to predict body density and thence fatness through prediction equations. These equations are population-specific however and may not be widely generalizable, suggesting that the raw skinfold thicknesses are often more useful.

Waist circumference, or the ratio of waist to hip circumference, has recently been recognized as a simple measure that has important risk prediction properties. Waist circumference is only minimally related to height and is a good predictor of visceral and total fat mass as well as disease risk. It is relatively easily measured within the primary care and research settings, which makes it a useful addition to the assessment of obesity by BMI. WHO guidelines specify that the circumference should be measured at the midpoint between the lower margin of the last rib and the top of the hip bone. Waist circumference cut-points associated with increased risk have not been fully defined as yet.

Further Reading

Cole, T.J., Bellizzi, M.C., Flegal, K.M., Dietz, W.H., 2000. Establishing a standard definition for child overweight and obesity worldwide: International survey. British Medical Journal 320, 1240–1243.

Finucane, M.A., Stevens, G.A., Cowan, M.J., et al., 2011. National, regional, and global trends in body-mass index since 1980: Systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. Lancet 377, 557–567.

Han, T.S., Sttar, N., Lean, M., 2006. Assessment of obesity and its clinical implications. British Medical Journal 333, 695–698.

Haslam, D.W., James, W.P.T., 2005. Obesity. Lancet 366, 1197–1209.

International Association for the Study of Obesity: (http://www.iaso.org)

Jebb SA, Kopelman P, and Butland B (2007) Foresight 'Tackling Obesities: Future Choices' Project. *Obesity Reviews* 8: vi–ix.

Kipping, R.R., Jago, R., Lawlor, D.A., 2008. Obesity in children. Part 1: Epidemiology, measurement, risk factors and screening. British Medical Journal 337, a1824.

Popkin, B.M., Gordon-Larsen, P., 2004. The nutrition transition: Worldwide obesity dynamics and their determinants. International Journal of Obesity Related Metabolic Disorders 28, S2–S9.

WHO Expert Consultation, 2004. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 363, 157–163.

World Cancer Research Fund/American Institute for Cancer Research (2007) Determinants of weight gain, overweight and obesity. In: *Food, Nutrition, Physical Activity, and the Prevention of Cancer: A Global Perspective*, ch. 8, pp. 322–341. Washington, DC: American Institute for Cancer Research.

Obesity: Prevention

T.P. Gill, University of Sydney, NSW, Australia

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of T.P. Gill, Obesity: Prevention, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 367–373, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00213-0.

Principles of Obesity Prevention **604**
Rationale for Obesity Prevention 604
Objectives of Obesity Prevention 605
Importance of Weight Gain Prevention in Adults 605
Who Should Obesity Prevention Strategies Target? **605**
Whole Community 606
Family Focus 606
High-Risk Groups 606
Those with an Existing Weight Problem 607
Key Elements of a Weight Gain Prevention Plan 607
Diet and Physical Activity Behaviors 608
Reducing Energy Intake 608
Reducing the Intake of High Energy-Dense Foods (i.e., foods high in fat/sugar) 608
Increasing the Intake of High-Fiber, Energy-Dilute Foods (especially vegetables and fruits) 608
Reducing the Consumption of Sugar-Sweetened Soft Drinks and Juices 609
Reducing the Level of Food Prepared Outside Home 609
Reducing Portion Sizes 609
Increasing Energy Expenditure 609
Regular Physical Activity 609
Reduced Time Spent in Sedentary Behaviors (especially TV watching) 609
Creating Supportive Environments 609
Identifying Effective Obesity Prevention Interventions 610
Lessons from Other Prevention Efforts 610
Further Reading **611**

The health impact of obesity is considerable, and obesity impacts on both quality and length of life. Overweight and obesity are associated with a wide range of chronic conditions, such as diabetes, hypertension, cardiovascular disease (CVD), and certain cancers, as well as non-lifethreatening but painful conditions, such as arthritis, back pain, and breathlessness. Obesity also places enormous financial burdens on governments and individuals, and accounts for a significant proportion of total health care expenditure in developed countries. Analyses suggest that obesity is fast approaching cigarette smoking as the major preventable cause of mortality.

In recent years, our understanding of the epidemiology and causation of obesity has improved dramatically and there is an acceptance that urgent action is required to address the problem. There is also a limited, but growing, body of successful, large-scale obesity prevention initiatives from across the world. These experiences, together with understandings gained from smaller-scale obesity prevention initiatives and experiences from the management of other epidemics of noncommunicable diseases, provide useful guidance on effective planning and implementation of obesity prevention programs.

Principles of Obesity Prevention

Rationale for Obesity Prevention

There are a number of reasons why prevention is likely to be the only effective way of tackling the problem of overweight and obesity. First, obesity develops over time, and once it has done so, it is very difficult to treat. A number of analyses have identified the limited success of current obesity treatments (with the possible exception of surgical interventions) to achieve long-term weight loss. Second, the health consequences associated with obesity result from the cumulative metabolic and physical stress of excess weight over a long period of time and may not be fully reversible by weight loss. Third, the proportion of the population that is either overweight or obese in many countries is now so large that there are no longer sufficient health care resources to offer treatment to all. It can be argued, therefore, that the prevention of weight gain (or the reversal of small gains) and the maintenance of a healthy weight would be easier, less expensive, and potentially more effective than to treat obesity after it has fully developed.

Objectives of Obesity Prevention

There remains a great deal of confusion regarding the appropriate objectives of an obesity prevention program. It is often assumed that to be effective, any intervention to address the problem of excess weight in the community should result in a reduction in the prevalence of overweight and obesity. However, such an objective is unrealistic and may be counterproductive. Most communities are experiencing significant increases in the average weight of the population as a result of a sizeable energy surplus resulting from reduced energy expenditure combined with an increased energy intake. This is leading to rapidly escalating rates of overweight and obesity. To reverse this trend will require not only the removal of this energy surplus but also the creation of a negative energy balance that will need to be maintained by the whole population for a significant period of time. Few (if any) interventions are capable of reducing energy intake, or increasing energy expenditure sufficiently, or are sustained long enough and with sufficient reach to achieve this effect. More appropriate objectives would relate to a reduction in the level of weight gain or the maintenance of weight stability in adults and the achievement of appropriate growth and development in children. The achievement of these objectives would result in a slowing in the rate of increase, followed by stabilization, and then an eventual decline in the level of overweight and obesity in the community.

However, even the goal of weight stability within a population may be difficult to achieve in the short term because it would require the maintenance or reestablishment of energy balance in a time of significant energy surplus. Therefore, it may be necessary to identify more sensitive short- and medium-term outcomes to evaluate obesity prevention programs. Such process outcomes may relate to the achievement of appropriate changes in energy intakes or outputs, food or physical activity behaviors, or changes to the environment that are significant enough to positively impact upon the achievement of energy balance.

Importance of Weight Gain Prevention in Adults

There are a number of important reasons why it is preferable to focus on weight gain prevention as the key individual and population objective of obesity prevention initiatives in adults (**Box 1**). The association between elevated body mass index (BMI) and increased risk of ill health is clear and consistent. However, research has demonstrated that weight gain *per se* is also associated with increased health risk, and that this risk is independent of absolute BMI (provided a person is not underweight). A number of studies have shown strong relationships between weight gain and increasing levels of diabetes, hypertension, gall bladder disease, and coronary heart disease. Therefore, a large weight gain in a lean individual may carry equivalent risk in maintaining a stable but slightly elevated BMI in an overweight individual. The combination of an elevated BMI and ongoing weight gain, however, leads to greatly magnified levels of risk.

Who Should Obesity Prevention Strategies Target?

Deciding where to invest limited time and resources in obesity prevention is a difficult task but finite health resources make this a necessity. WHO has identified three distinct but equally valid and complementary levels of obesity prevention (**Figure 1**). The specific 'targeted' approach directed at very high-risk individuals with existing weight problems is represented by the core of the figure, the 'selective' approach directed at individuals and groups with above average risk is represented by the middle layer, and the broader universal or populationwide prevention approach is represented by the outer layer. This replaces the more traditional classification of disease prevention (primary, secondary, and tertiary), which can be confusing when applied to a complex multifactorial condition such as obesity.

Box 1 Why focus on weight gain prevention?

- Weight gain in adulthood carries an independent risk of ill health.
- Risk for chronic disease begins to increase from low BMI levels and significant weight gain can occur within normal limits.
- Extended periods of weight gain are difficult to reverse.
- Weight gain in adulthood is mostly fat gain.
- The relationship between absolute BMI and health risk varies with age and ethnicity, but no such variations occur in the relationship between weight gain and ill health.
- A focus on weight gain prevention avoids exacerbation of inappropriate dieting behaviors.
- Weight maintenance can serve as a first stage goal for weight treatment programs.
- The message is equally relevant to all sections of the adult population.
- It avoids further stigmatization of people with an existing weight problem.
- It avoids reference to poorly understood terms such as 'healthy weight.'



Figure 1 Levels of obesity prevention intervention. Adapted from Gill TP (1997) Key issues in the prevention of obesity. *British Medical Bulletin* 53(2): 359–388, with permission from BMJ.

Universal prevention is the domain of public health, whereas selective and targeted prevention are predominantly dealt with in community and health care service settings. Community settings include schools, colleges, worksites, community centers, and shopping outlets.

Whole Community

Overweight and obesity are public health problems of relevance to the whole community and require strategies that focus on populationwide change rather than attempting to address individuals or small groups in isolation of the community in which they live. An effective population strategy needs to both improve population knowledge about obesity and its management and reduce the exposure of the community to obesity-promoting factors in the environment. Action at a population level requires coordination at a central level and the investment of resources to be maintained over a long period of time to achieve population change.

Family Focus

There are numerous reasons why children should be a major focus of any obesity prevention strategy. There is strong evidence that a high proportion of overweight or obese children will become obese adults. Childhood obesity also has immediate effects on health, and weight-related conditions are becoming more prevalent and their effect more pronounced as the rates of childhood obesity increase. However, children grow rapidly and increase the level of lean body mass as they age, and so reducing or keeping fat mass constant allows the normalization of weight over time. Thus, childhood (particularly younger children) is a period during which prevention efforts have a higher chance of success.

However, children also have little direct control over the environment in which they live. Parents and other caregivers mostly control decisions regarding the food available and the opportunities for activity. In addition, the behaviors of parents and other siblings have a profound effect on the diet and physical activity behaviors of children. For this reason, it is preferable to focus childhood obesity prevention efforts on the family environment rather than directly on children.

High-Risk Groups

There are a number of groups that appear to be at higher risk of developing overweight and obesity (**Table 1**). These groups warrant special attention and include the following:

- Those with a family history of weight problems.
- Socially disadvantaged and isolated communities.
- Certain ethnic groups.
- Smokers who have recently quit smoking.
- Those who have recently lost weight.

Table 1 Identifying at-risk groups for obesity

Critical ages and life stages	*Reason for increased risk*
Prenatal	There is strong evidence that *in utero* development has permanent effects on later growth and energy regulation.
Adiposity rebound (5–7 years)	BMI begins to increase rapidly after a period of reduced adiposity during preschool years. Food and activity patterns change as a result of exposure to other children and school. Early and rapid weight rebound often precedes the development of obesity.
Adolescence	Period of increased autonomy that is often associated with irregular meals, changed food habits, and periods of inactivity during leisure combined with physiological changes that promote increased fat deposition, particularly in females.
Early adulthood	Early adulthood usually correlates with a period of marked reduction in physical activity, increased alcohol consumption, and poor diets.
Pregnancy	Excessive weight gain during pregnancy often results in retention of weight after delivery, particularly with early cessation of breast feeding. Inappropriately large as well as inadequate weight gain may both contribute to weight problems for the developing child.
Menopause	In Western societies weight generally increases with age, but it is not certain why menopausal women are particularly prone to rapid weight gain. The loss of the menstrual cycle does affect food intake and reduce metabolic rate slightly.
High-risk groups	
Family history of weight problems	There is no longer any doubt that given the same environment some individuals are more prone to depositing fat. The basis of these differences in individual susceptibility to obesity is yet to be fully elucidated, but is believed to involve a number of physiological processes associated with fat deposition, oxidation, and involuntary energy expenditure.
Recent migrants and refugees	Recent migrants and refugees often find it difficult to maintain traditional diet and physical activity patterns and may replace them with inappropriate foods.
Socially or economically disadvantaged	In many developed countries, there is an inverse association between income and education level and obesity, which is most pronounced among women. It is argued that cheaper foodstuffs are usually high in fat and energy dense and those with less financial resources spend more time in sedentary activities such as watching TV.
Shift workers	Disrupted sleep patterns and altered daily routine have been associated with poorer quality diets, reduced opportunities to exercise and weight gain in those working night shifts.
Recent successful weight reducers	Successful weight reduction is usually followed by the regain of one-half to one-third of the weight loss over the following year. It is believed that biological and behavioral processes act to drive body weight back to baseline levels.
Recent past smokers	Smokers are usually thinner than nonsmokers because smoking tends to depress appetite, increase the basal metabolic rate, and, after each cigarette, induce a surge in heart and metabolic rate. The effect on metabolism of smoking 24 cigarettes per day has been estimated at approximately 200 kcal per day.

Source: Adapted with permission from Gill TP (1997) Key issues in the prevention of obesity. *British Medical Bulletin* 53(2): 359–388.

In addition, there are certain times in a person's life when the person is more prone to weight gain (Table 1). These age groups could be considered for selective prevention interventions. These times include the following:

- Prenatal
- Adiposity rebound (5–7 years)
- Adolescence
- Early adulthood
- Pregnancy
- Menopause

Those with an Existing Weight Problem

In developing weight gain prevention strategies, it is important not to neglect those with an existing weight problem who could benefit from more intensive efforts to help prevent further weight gain.

Key Elements of a Weight Gain Prevention Plan

Weight gain and obesity develop when the energy intake from food and drink exceeds energy expenditure from physical activity and other metabolic processes. It is often assumed that the prevention of weight gain should focus solely on attempting to alter these behaviors within individuals and communities. However, research has consistently shown that numerous and diverse factors,

Table 2 Summary of the strengths of evidence of factors that may promote or protect against weight gain and obesity

Evidence	*Decreases risk*	*Increases risk*
Convincing	Regular physical activity	High intake of energy-dense foods[a]
	High dietary fiber intake	Sedentary lifestyle
Probable	Home and school environments that support healthly food choices for children	Heavy marketing of energy-dense foods and fast food outlets
	Promoting linear growth	Adverse social and economic conditions in developed countries (especially for women)
		Sugar-sweetened soft drinks and juices
Possible	Low glycemic index foods	Large portion sizes
	Breast feeding	High proportion of food prepared outside of home
		Rigid restraint/periodic disinhibition eating patterns
Insufficient	Increased eating frequency	Alcohol

Source: Adapted with permission from World Health Organization (2003) *Joint WHO/FAO Expert Report on Diet, Nutrition and the Prevention of Chronic Disease*, WHO Technical Report Series 916. Geneva: WHO.

[a]Energy-dense foods are high in fat/sugar and energy-dilute foods are high in fiber and water, such as vegetables, fruits, legumes, and whole grain cereals.

including environmental and social factors, influence the behaviors that lead to excessive weight gain. Addressing aspects of the obesogenic (obesity-promoting) environment, as well as individuals' eating and physical activity patterns, is considered to be critical to the success of any obesity prevention program.

The 2003 WHO report on diet, nutrition, and the prevention of chronic disease undertook a detailed review of the literature and identified a range of key factors that either increase or decrease the risk of weight gain and the development of obesity (**Table 2**). These factors were rated on the quality of evidence available to support their contributory role. This analysis serves as a very useful guide as to the focus of weight gain prevention initiatives.

Diet and Physical Activity Behaviors

The WHO analysis and subsequent reviews have identified a number of key dietary and physical activity behaviors, amenable to change, that could conceivably influence energy balance sufficiently to contribute to the prevention of weight gain and obesity. Behaviors that reduced the risk of obesity included regular physical activity, high dietary fiber intake, and possibly breast feeding, and low glycemic index diets. Behaviors that increased the risk of obesity included a high intake of energy-dense foods, a high intake of sugar-sweetened drinks and juices, time spent in sedentary behaviors, and possibly large portion sizes, a high intake of fast foods, and a restrained eating pattern.

The area of dietary and physical activity antecedents to weight gain and obesity is still incompletely understood and new research findings, which help clarify our understanding, are being presented on a regular basis. In addition, different behaviors are more prevalent or pronounced in different regions of the world. It is therefore difficult to give definitive recommendations on the most important and useful behaviors to target in obesity prevention strategies. However, strong evidence exists to support the inclusion of some key behaviors.

Reducing Energy Intake

Reducing the Intake of High Energy-Dense Foods (i.e., foods high in fat/sugar)

There is a high level of agreement that the overconsumption of energy-dense foods is a major contributor to excess energy intake and weight gain and that the restriction of energy-dense food items is a useful strategy for the prevention of weight gain. However, discussion continues as to whether fat or refined carbohydrate is the major contributor to energy density in the modern diet and thus should be the target of programs to control weight. The debate is being fueled by dietary data from many developed countries showing that dietary fat intakes have leveled out or declined slightly and intakes of carbohydrates have increased dramatically. However, research has shown that dietary fat (along with water and fiber) is a major contributor to the energy density of foods and that *ad libitum* low-fat diet plans are an effective dietary approach to weight gain prevention or moderate weight loss. There is also strong evidence that excess refined carbohydrate, particularly high glycemic index carbohydrate, contributes to weight gain and its restriction aids weight loss and improves cardiovascular risk factor profiles.

Increasing the Intake of High-Fiber, Energy-Dilute Foods (especially vegetables and fruits)

There is less evidence on the effectiveness of increasing the intake of energy-dilute foods such as vegetables and some fruits in the diet. Such a strategy would assist weight gain prevention only if the inclusion of such foods leads to a reduction in the intake of more energy-dense alternatives and thus creates a reduction in energy intake. Few studies have addressed this issue in a comprehensive manner, but the additional health benefits of these foods make such a strategy low risk in nutritional terms.

Reducing the Consumption of Sugar-Sweetened Soft Drinks and Juices

Evidence is accumulating from a variety of studies that energy consumed as sweetened drinks is less well compensated for than energy consumed as solid food. Many, but not all, longitudinal studies have indicated that sweetened drinks (soft drinks or sodas) are associated with weight gain in both children and adults. Recent work has also demonstrated that the simple strategy of reducing the intake of sweetened drinks can be effective in preventing or limiting inappropriate weight gain.

Reducing the Level of Food Prepared Outside Home

The proportion of food purchased and consumed at food outlets outside home has increased dramatically in recent decades in both developed and developing nations. In the United States, approximately 40% of the household food budget is spent on food eaten away from home, and much of this is spent at fast food outlets. A number of analyses has linked increased consumption of fast food with increased risk of obesity. Although only a limited number of studies has evaluated the effect of reducing the consumption of fast food, it would seem to be a valuable strategy with few nutritional negatives.

Reducing Portion Sizes

The portion size of packaged foods and snacks, as well as restaurant serving sizes, has increased rapidly in recent times and has been identified as an important factor in the consumption of excess energy. Evidence suggests that people will consume the portion of food they are provided rather than respond to satiety signals to stop eating and leave food. Also, as the serving size increases, the ability of consumers to estimate accurately how much they have consumed decreases. Reducing portion sizes is a simple but immediately effective mechanism for reducing energy intake.

Increasing Energy Expenditure

Regular Physical Activity

Although it is difficult to obtain accurate assessments of physical activity, there is little doubt that energy expenditure from activity has decreased in the past 50 years in most countries throughout the world. In contrast to popular belief, participation rates in organized leisure-time physical activity have not declined in recent times and may have increased in many countries. This supports the contention that the greatest contributor to this reduction in energy expenditure is associated with substantial changes in occupational and incidental physical activity. Changes in employment patterns and work practices together with a reliance on motorized transport and the removal of almost all manual labor from our daily lives have led to a dramatic reduction in daily physical exertion.

Studies that have examined the association between physical activity and weight gain and the impact of increasing physical activity on weight gain prevention have been limited by the ability to accurately measure physical activity across the whole day and to engage people in sufficient levels of physical activity to prevent weight gain. However, there is sufficient evidence to support an important role for increasing physical activity in any weight gain prevention strategy, although questions remain about how much exercise is necessary and what type of exercise is appropriate to promote. The issue of the amount of extra time that people should spend in moderate physical activity to prevent weight gain remains hotly debated, but it is clearly substantially more than the 30 min on 5 days or more each week recommended by experts to reduce CVD risk. The type of exercise that should be the focus of weight gain prevention strategies is also under review. It has been suggested that the most effective ways to include regular physical activity in daily living are through increased incidental activity, increased participation in active recreation, and increased use of active transport.

Reduced Time Spent in Sedentary Behaviors (especially TV watching)

Changes in societal structures and improvements in technology have allowed a reduction in time spent at work or on domestic chores, leaving a greater proportion of the day for leisure. At the same time, most of the entertainment options developed to fill this time, such as watching television, playing video games, and using computers, are sedentary activities that require very little energy expenditure. These forms of entertainment, which initially complemented other forms of leisure activity, are occupying more hours of the day and are displacing more active pursuits and games. As a consequence, a number of studies have identified clear links between time spent in this sedentary behavior and weight gain. However, it is important to make a distinction between a lack of physical activity and sedentary behavior because their mechanisms for impacting on body weight may be different and a person with a high level of physical activity can also have a high level of sedentary behavior. Although the precise pathway by which sedentariness influences weight gain is not known, it is believed to involve both a reduction in physical activity and physiological reductions in energy expenditure together with an increase in dietary energy intake through inappropriate food intake that is often stimulated by and accompanies sedentary activities.

Some studies in children have shown that programs that seek to reduce time spent in sedentary behaviors are more effective in controlling weight than programs that aim to increase physical activity alone. In some cases, a simple program to reduce the amount of time spent watching television was sufficient to significantly limit inappropriate weight gain in children.

Creating Supportive Environments

The external physical, social, political, and economic environments in which people exist have a profound effect on their attitudes and behaviors. Each day, people interact with a wide range of services, systems, and pressures in settings such as schools, the

workplace, home, restaurants, and fast food outlets. In addition, laws, policies, economic imperatives, and the views of governments, industry, and society as a whole influence these settings. Each of the features of this complex system, which shapes the environment in which we live, has the capacity to inhibit or encourage appropriate dietary and physical activity patterns. The availability of open space, access to public transport, the design of suburbs, access to buildings, the perceived level of safety, provision of lighting, and many other factors influence our capacity and desire to be more physically active in our daily lives. Similarly, advertising pressures, access to appropriate food choices, school food policies, and nutrition information, and labeling all potentially influence food selection. Today, there is also a large commercial drive to promote obesogenic behaviors (cars and food are the two most advertised products on television). The Foresight Program of the UK Government Office for Science recently produced a complex conceptual model with 108 variables known as the obesity systems map to illustrate the complexity and interaction of the broad range of variables that drive the development of obesity.

Trying to motivate people to make healthy choices when the external environment works against such behaviors is a recipe for failure. **Figure 2** illustrates the role that the social environment plays in assisting or inhibiting personal behavior choices made by individuals, which ultimately has an impact on their health. Great success is likely to be achieved by creating a supportive environment and then promoting the healthy dietary and physical activity choices within such an environment.

Identifying Effective Obesity Prevention Interventions

Identifying the most efficacious programs to prevent obesity from the published literature has proved problematic. A number of systematic reviews have assessed the current scientific literature of reported programs addressing the prevention of obesity in both children and adults and have identified only a limited number of evaluated programs to assess. Reviews of childhood obesity prevention initiatives indicated that certain approaches appear to be associated with greater success. An intensive intervention in small groups was a successful management strategy in children, as was involving all the family. Reducing levels of inactivity was successful at both treating and preventing weight gain. Some interventions which increased time spent in formal physical activity were successful in controlling weight gain but generally multicomponent programs which addressed a range of strategies were deemed to hold the most potential. The Ensemble Prévenons l'obésité des Enfants program is a multicomponent, community-based childhood obesity program which has been implemented in a number of local government regions in France and throughout Europe and has produced promising reduction in measured obesity rates.

However, most reviews conclude that there was simply too small a body of research conducted in a limited number of settings to provide firm guidance on consistently effective interventions. To address the limitation, some researchers have proposed a system that allows the integration of the available information from the literature together with other forms of evidence including experience from past public health and health promotion action to identify the target groups and settings most likely to have produced effective action on obesity. This is achieved by producing a classification system which is based on potential for change rather than demonstrated effectiveness. To achieve this, interventions are selected and assessed in terms of how 'promising' they may be in addressing population weight gain, using a health gain/risk framework. This allows the selection of interventions to be based on the best available evidence, while not excluding untried but promising strategies (see **Table 3**). The return or health gain can be defined in terms of demonstrated or modeled efficacy (from previous studies), potential population reach, and likely uptake (estimated). Uncertainty or risk can be defined in terms of the level of information or evidence to support the effectiveness of the intervention.

Lessons from Other Prevention Efforts

Although the number of successful large-scale obesity prevention programs is limited, there is a wealth of information from past public health programs that can be used to address other chronic diseases and risk factors. The International Obesity Task Force identified 10 key principles on which efforts to prevent obesity at a population level should be based. These are presented in

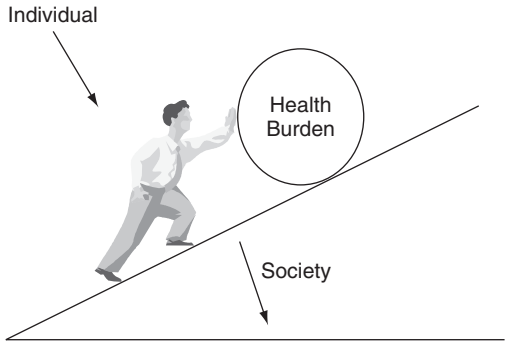


Figure 2 Influence of societal and environmental factors on development of obesity. Reproduced from House of Commons Health Committee (2004) *Obesity: Third Report of Sessions 2003–04. Volume 1. Report Together with Formal Minutes.* London: The Stationery Office Ltd.

Table 3 Matrix for determining the 'promise' of an intervention

	Potential population impact^ (Return)		
Certainty of effectiveness (Risk)*	*Low*	*Moderate*	*High*
Quite low	Least promising	Less promising	Promising
Medium	Less promising	Promising	Very promising
Quite high	Promising	Very promising	Most promising

Source: Adapted from Gill TP, King L, Webb K, and NSW Centre for Public Health Nutrition (2004) *Best Options for Promoting Healthy Weight and Preventing Weight Gain in NSW.* Sydney: NSW Health.

Box 2 IOTF principles for the development of population obesity prevention initiatives

1. Education alone is not sufficient to change weight-related behaviors. Environmental and societal intervention is also required to promote and support behavior change.
2. Action must be taken to integrate physical activity into daily life, not just to increase leisure time exercise.
3. Sustainability of programs is crucial to enable positive change in diet, activity, and obesity levels over time.
4. Political support, intersectoral collaboration, and community participation are essential for success.
5. Acting locally, even in national initiatives, allows programs to be tailored to meet real needs, expectations, and opportunities.
6. All parts of the community must be reached, not just the motivated healthy.
7. Programs must be adequately resourced.
8. Where appropriate, programs should be integrated into existing initiatives.
9. Programs should build on existing theory and evidence.
10. Programs should be properly monitored, evaluated, and documented. This is important for dissemination and transfer of experiences.

Box 2 and are drawn from experiences addressing CVD, smoking, alcohol and drug problems, dental disease, road accidents, and other public health issues.

Although much has yet to be elucidated about the development of obesity and its effective management and prevention, there is a consensus that action to address the problem must not be delayed. Efforts to prevent weight gain need to be well designed, comprehensive, and appropriately resourced and evaluated so that the knowledge base improves with each new program.

Further Reading

European Public Health Alliance, 2008. EPODE – Together Let's Prevent Childhood Obesity www.epha.org/a/3149.

ForeSight Committee, 2008. Tackling Obesities: Future Choices. Government Office for Science, London.

Gill, T.P., 1997. Key issues in the prevention of obesity. British Medical Bulletin 53 (2), 359–388.

Gill, T.P., King, L., Webb, K., NSW Centre for Public Health Nutrition, 2004. Best Options for Promoting Healthy Weight and Preventing Weight Gain in NSW. NSW Health, Sydney.

Heath, G.W., Brownson, R.C., Kruger, J., et al., 2006. The effectiveness of urban design and land use and transport policies and practices to increase physical activity: A systematic review. Journal of Physical and Activity and Health 3 (Suppl 1), S55–S76.

Hesketh, K.D., Campbell, K.J., 2010. Interventions to prevent obesity in 0–5 year olds: An updated systematic review of the literature. Obesity 18 (Suppl 1), S27–S35.

House of Commons Health Committee, 2004. Obesity: Third Report of Sessions 2003–04. Volume 1. Report Together with Formal Minutes. The Stationery Office Ltd., London.

Kumanyika, S.K., Jeffery, R.J., Morabia, A., Ritenbaugh, C., Antipatis, V.J., 2002. A report of the Public Health Approaches to the Prevention of Obesity (PHAPO) Working Group of the International Obesity Task Force (IOTF). International Journal of Obesity 26, 425–436.

Lemmens, V.E., Oenema, A., Klepp, K.I., Henriksen, H.B., Brug, J., 2008. A systematic review of the evidence regarding efficacy of obesity prevention interventions among adults. Obesity Review 9 (5), 446–455.

Sanigorski, A.M., Bell, A.C., Kremer, P.J., Cuttler, R., Swinburn, B.A., 2008. Reducing unhealthy weight gain in children through community capacity-building: Results of a quasi-experimental intervention program, Be Active Eat Well. International Journal of Obesity 32 (7), 1060–1067.

Summerbell, C.D., Hillier, F., 2010. Community interventions and initiatives to prevent obesity. In: Crawford, D., Jeffery, R., Ball, K., Brug, H. (Eds.), Obesity Epidemiology: From Aetiology to Public Health. Oxford University Press, Oxford, pp. 395–408.

Swinburn, B., Gill, T.P., Kumanyika, S., 2005. Obesity prevention: A proposed framework for translating evidence into action. Obesity Reviews 6, 23–33.

Waters, E., de Silva-Sanigorski, A., Hall, B.J., et al., 2011. Interventions for preventing obesity in children. Cochrane Database of Systematic Reviews Issue 12. https://doi.org/10.1002/14651858.CD001871.pub3.

World Health Organization (2000) *Obesity: Preventing and Managing the Global Epidemic. Report of a WHO Consultation*, WHO Technical Report Series 894. Geneva: WHO.

World Health Organization (2003) *Joint WHO/FAO Expert Report on Diet, Nutrition and the Prevention of Chronic Disease*, WHO Technical Report Series 916. Geneva: WHO.

Osteoporosis nutritional factors

Alexa Barad and Kimberly O. O'Brien, Cornell University, Division of Nutritional Sciences, Ithaca, NY, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of K.O. O'Brien, Osteoporosis: Nutritional Factors, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 418–424, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00221-X.

Introduction 612
Dietary intake and body mass 613
Calcium 613
Vitamin D 615
Magnesium 616
Zinc and copper 616
Vitamin K 616
Phosphorus 616
Sodium and potassium 617
Protein 617
Phytoestrogens 617
Homocysteine 618
Other lifestyle factors 618
Targeting groups at risk 618
Nutrient/gene interactions 618
Interactions of the bone and other physiological systems 619
References 620
Further reading 621
Relevant websites 621

Key points

- Review the role of key nutrients, foods, and dietary patterns on risk of osteoporosis and fractures.
- Discuss the impact of modifiable and non-modifiable factors on bone health.
- Highlight novel interactions between nutritional health of other physiological systems and bone health.

Glossary

Dual energy x-ray absorptiometry (DXA) This non-invasive test uses the degree to which x-rays are attenuated as they pass through the human body to evaluate bone mineral content and density of the entire skeleton or at selected sites (typically the lumbar spine, hip and cervical vertebrae)

Osteocyte Osteocytes comprise 95% of all bone cells and are derived from osteoblasts that have become imbedded in mineralized bone. These cells respond to mechanical forces and play an important role in the regulation of bone turnover

Osteopenia Osteopenia is a term used to denote low bone mass. It is diagnosed when the bone mineral density T Score (from a DXA test) is between −1.0 and −2.5

Osteoporosis A term used to denote a significant loss of bone. It is typically defined when the bone mineral density T Score (from a DXA test) is less than −2.5

Parathyroid Hormone (PTH) Parathyroid hormone is secreted by the parathyroid glands in response to low serum ionized calcium. It functions to maintain circulating calcium concentrations within a narrow range

Phosphatonin Phosphatonins are compounds that regulate phosphorus metabolism using PTH-independent mechanisms

Introduction

Optimal dietary intake is essential for bone health. During childhood and the pubertal growth spurt, nutrients are needed to fully consolidate skeletal mass and to insure the attainment of a peak bone mass consistent with one's genetic potential. After peak bone

mass is obtained, nutrition continues to play an essential role in skeletal health. If key nutrients are not consumed at required levels, mineral may be lost from bone or essential bone proteins may not be fully functional.

Osteoporosis and osteopenia are currently significant public health problems. Once a substantial amount of bone mineral and matrix is lost, bone becomes more susceptible to fracture. Osteopenia is defined when adult bone mineral density values are 1–2.5 SD below the mean peak value observed in a young adult. Osteoporosis is defined when adult bone mineral density falls 2.5 SD or more below that observed in a young adult. Recent estimates available from 2017 to 2018 suggest that more than 10 million adults over the age of 50 y in the United States have osteoporosis, and another 43 million have osteopenia and are at risk for osteoporosis and fracture (Sarafrazi et al., 2021). The increasing impact of low bone mass on public health can be demonstrated by the large increase in the age-specific rate of fractures (Rosengren et al., 2015). Recognition of this chronic disease is an increasing concern globally as the population ages and mean life span increases. Because bone loss is not fully reversible, the most effective strategies for treating osteoporosis should focus on prevention with nutrition playing a key role.

Dietary intake and body mass

A balanced diet is important to promote health and to maintain an appropriate body weight. An individual's body weight is one of the strongest determinants of bone mass because of the skeleton's responsiveness to the load that is placed upon it. Individuals with small body frames or those that are excessively thin have an increased risk of osteoporosis due to a lower overall skeletal reserve to draw on for calcium needed to offset the annual loss of bone that occurs later in life. At the extreme end of this spectrum, individuals with anorexia nervosa are at risk of osteoporosis because of alterations in hormonal status and amenorrhea in addition to insufficient dietary intake of nutrients required for bone health.

Excess body weight is typically associated with a greater skeletal mass, but more recent evidence suggests that obesity may have negative effects on bone health. Excessive fat accumulation may adversely affect bone health by a variety of mechanisms including overall and adipocyte-specific hormonal dysregulation, and increased oxidative stress and inflammation, eventually compromising bone quality (Shapses et al., 2017). Insulin resistance has been proposed as a key mediator of the relationship between adiposity and bone mass (Kindler et al., 2019). Additionally, obesity has been associated with lower 25-hydroxyvitamin D (25OHD) and higher parathyroid hormone concentrations (Pereira-Santos et al., 2015). These changes are thought to occur in part due to sequestration of vitamin D in adipose tissue or volumetric dilution. Bariatric surgery as a treatment for severe obesity is becoming more common and leads to both a loss of body weight and bone mass. Available evidence from epidemiological studies suggests that the type of procedure (restrictive, malabsorptive, or mixed) plays a role in determining the risk of fractures, with mixed procedures such as Roux-en-Y gastric bypass being associated with the highest risk of fractures (Gagnon and Schafer, 2018). Overall, the observed negative effects of bariatric surgery on bone health are multifactorial, involving nutritional deficiencies, mechanical unloading, hormonal shifts, and changes to total body composition. At present, the long-term impact of this surgery on skeletal health has not been fully elucidated as fracture risk only begins to manifest 2–5 years post-surgery (Gagnon and Schafer, 2018).

While overall caloric intake impacts body weight, many individual nutrients and dietary components have been studied in relation to their impact on bone health (**Table 1**). Several of these key nutrients and components of the diet have key roles in bone health and skeletal homeostasis.

Calcium

Calcium is the most abundant mineral found in bone and comprises approximately 33% of bone mineral. Optimal calcium intakes are essential across the lifecycle to meet the daily intrinsic requirements of calcium needed for skeletal growth and to offset urinary,

Table 1 Nutritional, lifestyle, and physiological parameters that may influence bone health.

Minerals	*Vitamins/hormones*	*Lifestyle and environmental factors*	*Dietary components*	*Interactions of the bone and other physiological systems*
Calcium	Vitamin D	Body mass index	Protein	Muscle mass
Phosphorus	Vitamin K	Exercise	Soy/phytoestrogens	Gut microbiome
Magnesium	Vitamin A	Cigarette smoking	Fatty acids	Adipose tissue
Sodium	Vitamin C	Alcohol intake	Homocysteine	Immune function
Zinc	Vitamin B_{12}	Sleep		
Copper	Vitamin B_6			
Iron	Folate			
Boron				
Manganese				
Fluorine				
Potassium				
Silicon				

dermal and endogenous fecal calcium losses. When dietary intakes of calcium are not sufficient to maintain circulating calcium concentrations and/or when the losses of calcium from the body are excessive, bone calcium will be resorbed to maintain calcium homeostasis. Because calcium is essential for the structural integrity of bone, deficiencies or inadequate intakes of this mineral will have a detrimental impact on bone mass and quality (Weaver and Peacock, 2019).

Skeletal mass peaks at approximately age 20–30 years with the majority of this gain (nearly 50%) being accrued during the pubertal growth spurt (Baxter-Jones et al., 2011). Thus, this period of skeletal accretion can be viewed as a window of opportunity to maximize skeletal mass. Calcium supplementation studies in children found increased bone mass following supplementation, an effect that is most pronounced when implemented during the prepubertal period (Weaver et al., 2016b). However, supplementation appears to primarily impact the tempo at which peak bone mass is achieved because gains in bone mass are not typically maintained after supplementation ends and this does not appear to result in a net difference in peak bone mass. The data on the efficacy of calcium supplementation with or without vitamin D in reducing risk of fractures in adults are controversial. In 2013, the US Preventive Services Task Force (USPSTF) concluded that there was insufficient evidence to recommend daily vitamin D and calcium supplementation for the prevention of fractures in healthy adults (Moyer, 2013). In 2016, the National Osteoporosis Foundation (NOF) conducted a meta-analysis of randomized-controlled trials that supported the use of calcium and vitamin D supplementation as an intervention for fracture risk reduction in middle-aged to older adults (Weaver et al., 2016a). Subsequently, in 2018 the USPSTF published an updated review, which concluded that vitamin D and calcium supplementation or calcium supplementation alone is not associated with lower incidence of fractures in healthy adults, however, the quality of evidence was deemed to be low (Kahwati et al., 2018). In high-risk adults, calcium supplementation has been found to have mild beneficial effects on bone mass and reduced fracture risk (Harvey et al., 2017) and may have the greatest impact in individuals whose habitual dietary calcium intakes are below 400 mg (10 mmol) d^{-1}.

In 2010 the National Academy of Medicine (NAM, previously known as the Institute of Medicine) updated the dietary recommendations for calcium and vitamin D (**Table 2**) (Ross et al., 2011). Using newly available data, the committee was able to set an RDA (recommended dietary allowance) for calcium and vitamin D for all age groups except infants. In infants during the 1st year of life available data were only sufficient to establish an adequate intake (AI). As in the previous recommendations, due to the importance of calcium in bone mineralization, the recommended daily allowance of calcium is highest (1300 mg (32.5 mmol) d^{-1}) between the ages of 9–18 y. Because of the known decreased efficiency of intestinal absorption coupled with increased losses of calcium that occur in older women as they progress through menopause, calcium intakes were increased to 1200 mg (30 mmol) d^{-1} in women ages 50 y and older but intakes for men remain at 1000 mg until the age of 70 y at which time they also increase to 1200 mg (30 mmol) d^{-1}. Recent dietary intake data suggest most adults 19 y and older do not consume adequate amounts of calcium or vitamin D. The Dietary Guidelines for Americans (DGA) 2020–2025 highlighted both calcium and vitamin D as nutrients of public health concern for the general US population, including children and adults, as low intakes of these nutrients are associated with health concerns (USDA and USHHS, 2020). The DGA recommends all age groups over 2 y increasing consumption of nutrient-dense, calcium- and vitamin D-rich foods, such as low-fat dairy products or soy alternatives. This

Table 2 2010 Dietary reference intakes for Calcium and vitamin D.

Group	*Ca RDA (mg)*	*Vitamin D RDA (IU)*
Infants		
Birth–6 months	200 (AI)	400 (AI)
7 months–1 year	260 (AI)	400 (AI)
Children		
1–3 years	700	600
4–8 years	1000	600
Male adolescents/adults		
9–18 years	1300	600
19–70 years	1000	600
>70 years	1200	800
Female adolescents/adults		
9–18 years	1300	600
19–50 years	1000	600
51–70 years	1200	600
>70 years	1200	800
Pregnancy		
14–18 years	1300	600
≥19 years	1000	600
Lactation		
14–18 years	1300	600
≥19 years	1000	600

Only adequate intake (AI) recommendations are available for those under the age of 1 year.

recommendation is particularly important for adolescents during peak bone mass accrual, for adults during the time when peak bone mass accrual is still active, and in post-menopausal women and elderly adults.

Because the majority of dietary calcium is obtained from dairy products, those with low dairy intakes or with other factors such as lactose intolerance, dieting or altered appetite and food consumption patterns may need to rely more on fortified food products or calcium supplements. An increasing variety of calcium fortified food products are now also available. Individuals with lactose intolerance may improve intake of calcium by use of lactose-free dairy products or lactase pills. Whenever possible, increased calcium intake should be obtained from dietary vs. supplemental sources in order to obtain additional nutrients needed for bone health including protein, magnesium, zinc, phosphorus and vitamin D (Geiker et al., 2020).

Use of calcium supplements is currently common in the United States, particularly among older women. Several forms of calcium supplements are commercially available with calcium carbonate and citrate being the forms most commonly consumed. Existing supplemental forms differ slightly with respect to their relative calcium content per tablet and their absorbability; however the magnitude of these differences is minor and may not be biologically significant. Several calcium supplements now also contain additional nutrients required for bone health including vitamins D and K. Because the fraction of calcium absorbed decreases as calcium intake increases, little additional benefit per dose is achieved when taking supplemental calcium sources containing more than 500 mg (12.5 mmol) per dose (Heaney et al., 1975).

The increased availability of calcium-fortified food products may increase the likelihood of excessive intakes of calcium. The tolerable upper intake level (UL) for calcium ranges between 2000 and 3000 mg d^{-1} depending on the age group in question. These limits for adults were based on data that found increased risk of kidney stones (mainly among postmenopausal women ingesting calcium supplements) (Ross et al., 2011).

Vitamin D

Vitamin D is integral to bone health primarily due to the role it plays in calcium and phosphorus homeostasis. The active form of vitamin D, calcitriol, stimulates calcium and phosphorus absorption, and vitamin D along with parathyroid hormone (PTH) plays a regulatory role in renal calcium reabsorption and in calcium release from bone.

Vitamin D can be obtained either from the diet (as D_2 or D_3) or is produced in the skin (as D_3) following cutaneous exposure to sunlight. These two forms of vitamin D may have similar effects on target tissues at low doses, but some data suggest that high doses of D_2 may be less effective than D_3 for endocrine effects on target tissues while D_2 may be more available for autocrine/intracrine effects because of its lower affinity for vitamin D binding protein (Chun et al., 2019; Tripkovic et al., 2012). Circulating 25OHD is the best indicator of vitamin D status because this form is not tightly regulated, and its production is indicative of available D_2 and D_3.

Lack of sufficient endogenous production of vitamin D in the skin is influenced by geographical location (northern latitudes have a shorter season over which the wavelength needed for vitamin D synthesis is available), increased use of sunscreen, and cosmetics and skin care products containing sunscreen (sunscreens when applied properly may limit the dermal production of vitamin D) and by lifestyle factors that decrease exposure to sunlight.

Calcium absorption is believed to be maximal when 25OHD concentrations are between 12 and 20 ng mL^{-1} (30 and 50 nmol L^{-1}) (Rosen et al., 2012). A serum 25OHD level of 16 ng mL^{-1} (40 nmol L^{-1}) was identified by the 2010 NAM committee as that desired for the population median and a concentration of 20 ng mL^{-1} (50 nmol L^{-1}) was identified as sufficient to meet the needs of "coverage" for the population (Ross et al., 2011). Recognizing the importance of vitamin D on skeletal health, the 2010 RDA for vitamin D was tripled over the earlier AI of 200 IU to the new RDA value of 600 IU for those between the ages of 1–70 y. In those over the age of 70 y this was raised to 800 IU to account for the increased concerns of bone health in this age group and for the variability in the physiology of aging that may impact renal function, cutaneous synthesis of vitamin D, altered body composition and increased PTH concentrations.

The UL for vitamin D was assessed using the indicators of hypercalcemia and related toxicity in adults and on retarded longitudinal growth in infants. Using these endpoints, the UL was set between 1000 and 3000 IU for infants and children through 8 y of age after which time it is set at 4000 IU over the remaining life course.

Achieving this level of vitamin D may be challenging from diet alone unless fortified food products are consumed. Vitamin D supplements may be useful in meeting requirements in those with low intake of dairy products. Several supplements are on the market containing various amounts of either D_2 or D_3.

While both calcium and vitamin D have been found to have a positive impact on bone mass, their independent effect on bone is challenging to isolate because supplementation studies typically administer both calcium and vitamin D. It is important to note that the recent calcium and vitamin D dietary reference intakes (DRI), as are all DRI recommendations, are based on the needs of healthy populations and are not meant to displace medical recommendations targeted to those with diseases that impact bone health.

Practice guidelines from the Endocrine Society were developed in 2011 for individuals at high risk of vitamin D deficiency, including those with chronic diseases that affect vitamin D status as well as those taking medications that affect vitamin D homeostasis (Holick et al., 2011). The Endocrine Practice Guidelines Committee recommends routine screening for individuals at high risk of vitamin D deficiency, a higher daily intake for those at high risk of deficiency than that recommended by the NAM for the general healthy population, and a more aggressive supplementation regimen for individuals with a vitamin D deficiency, with all guidelines to be applied under medical supervision.

Because many nutrients are important in bone health and there is still much to be learned about individual nutrients and bone physiology, it is essential that a well-balanced diet, containing grains, fruits and vegetables, protein and calcium-rich products be consumed. Many nutrients, in addition to calcium and vitamin D, are particularly important for maintaining bone homeostasis as detailed below.

Magnesium

More than half of the magnesium found in the body is located in bone. In addition to its presence in bone, magnesium is important in calcium metabolism and bone health because it is required for parathyroid hormone (PTH) secretion. Parathyroid hormone is integral to bone health because it increases the production of the active form of vitamin D (1,25-dihydroxyvitamin D) and plays a role in the tubular reabsorption of calcium and phosphorus.

Although magnesium deficiency is associated with abnormalities in vitamin D metabolism, hypocalcemia and impaired PTH secretion, epidemiological studies linking magnesium intakes to measures of skeletal health, such as incidence of fractures, have produced conflicting results. Observational studies have shown a positive association between magnesium intake and bone mineral density (Farsinejad-Marj et al., 2016). However, relationships between magnesium status and bone mass may be more challenging to elucidate due to the lack of a highly sensitive indicator of magnesium status (Razzaque, 2018).

Studies have indicated that typical magnesium intakes in healthy adolescents may not be sufficient to maintain magnesium balance (Abrams et al., 1997). Other groups, including older adults, those with gastrointestinal conditions, diabetes mellitus, alcoholism, or those being treated with diuretics are also at increased risk of magnesium deficiency (Razzaque, 2018). Because dietary intakes fall below recommended levels in several age groups and due to the known relationships between magnesium and hormones integral to bone health, increased attention should be focused on optimal magnesium intakes in relation to bone homeostasis particularly during the period of maximal bone acquisition.

Zinc and copper

Zinc and copper play important roles in bone metabolism and bone health in part due to the roles they play as cofactors for various enzymes required for the synthesis or modification of bone matrix constituents. Zinc is a co-factor for many enzymes in the body, including alkaline phosphatase. Alkaline phosphatase is synthesized by osteoblasts and is essential for bone mineralization. Zinc also plays a role in the osteoblast via its involvement in aminoacyl-tRNA synthetase. Copper is a necessary co-factor for lysyl oxidase, an enzyme that is involved in collagen cross-linking. Both copper and zinc are found as components of superoxide dismutase and may protect bone from oxidative damage. Genetic defects that cause zinc deficiency (acrodermatitis enteropathica) or copper deficiency (Menkes disease) result in growth retardation, stunting and impaired bone growth.

Vitamin K

Many proteins integral to bone health are dependent on vitamin K for the carboxylation of g-carboxyglutamyl (Gla) residues. Osteocalcin, one such vitamin K dependent protein, is the most abundant non-collagenous protein in bone. Osteocalcin contains three Gla residues that require vitamin K for carboxylation. The ability of osteocalcin to bind to the hydroxyapatite fraction of bone is dependent on its degree of carboxylation. Deficiency of vitamin K increases the fraction of undercarboxylated osteocalcin in the circulation. In addition to osteocalcin, other vitamin K dependent proteins (including matrix Gla protein and protein S) are found in bone and cartilage. Continued research is needed to elucidate the impact of vitamin K deficiency on risk of osteoporosis and fracture. Although vitamin K status correlates with bone mineral density, a systematic review and meta-analysis of randomized controlled trials concluded that vitamin K supplementation appears to have little to no effect in the risk reduction of osteoporosis and fracture (Mott et al., 2019). However, because of the known relationship between vitamin K and several crucial bone proteins, optimal status of this vitamin should be achieved to promote skeletal health.

Phosphorus

Phosphorus, like calcium, is an integral component of hydroxyapatite in bone. Bone contains 85% of the phosphorus found in the body, and together calcium and phosphorus comprise the major fraction of bone mineral. While sufficient phosphorus intake is necessary to support bone mineralization, phosphorus homeostasis can be maintained across a range of intakes and ratios of calcium to phosphorus in the diet. Diets high in phosphorus and low in calcium may be detrimental to bone health, as they have been associated with higher PTH levels and lower serum calcium concentrations in healthy individuals (Vorland et al., 2017).

Much attention has been focused recently on regulation of phosphorus homeostasis by PTH-independent mechanisms that are mediated via phosphatonins. To date, at least four phosphatonins have been identified. Of the phosphatonins identified to date, fibroblast growth factor 23 (FGF23) is the phosphatonin that contributes to phosphate homeostasis. Synthesis of FGF23 increases

in response to elevations in plasma phosphorus and calcitriol and FGF23 suppresses 1-α hydroxylase activity. These effects work in combination to reduce the body of excess phosphorus that is released during bone resorption. The role of FGF23 in bone physiology has opened up new possibilities for treatment of phosphorus related diseases that adversely impact bone mass (Pool and Wolf, 2017).

The phosphoric acid and phosphorus content of soda is often discussed in relation to bone health. The impact of these products on bone health is thought to be caused by their displacement of other more nutritive beverages (such as dairy products) from the diet. Phosphoric acid and other phosphorus containing-additives primarily present in sodas and processed foods are more readily absorbed than organic phosphorus naturally present in foods. Increased soda consumption may displace more nutritive beverages from the diet and may result in an unbalanced ratio of calcium to phosphorus in the diet, thus excessive soda intakes should be avoided (Vorland et al., 2017).

Sodium and potassium

Some dietary components influence the retention of nutrients required for optimal bone health. Sodium is one of the strongest determinants of urinary calcium excretion. Increased dietary sodium concentrations elevate urinary calcium losses. Every 2300 mg (100 mmol) increase in dietary sodium increases the urinary excretion of calcium by roughly 40 mg (1 mmol) (Nordin et al., 1993). Thus, excessive intakes of sodium (such as those that may occur in individuals who consume large amounts of processed food, salt food heavily or consume foods high in sodium) increase the obligatory losses of calcium from the body. During the growth phase, this could potentially limit the amount of calcium that can be utilized for bone mineralization. The long-term impact of variation in sodium intake on bone mass and fracture risk has been difficult to quantify because of a lack of sufficient information on how dietary effects on urinary calcium loss are counterbalanced and because other dietary components may modify this response.

Dietary potassium may help buffer calcium losses in urine. Increased fruit and vegetable intake will assist in increasing dietary potassium intake while providing additional nutrients and antioxidants that have been linked to overall skeletal health.

Protein

Protein is essential for the formation of the organic matrix of bone and optimal intakes are required for normal skeletal development, growth, and maintenance. Protein intakes are also needed to maintain muscle mass and help limit the involuntary loss of muscle (sarcopenia) that can occur as aging progresses. The importance of protein in bone health is well known, however, conflicting reports exist on the relative impact of extremes of protein intake on bone health. Many proteins are rich in sulfur amino acids. The resulting protein-induced acid load must be buffered before excretion from the body. Calcium is a positive cation and can be utilized to buffer increased dietary acid loads from high protein intakes. On average, for every 1 g increase in dietary protein intake, urinary calcium excretion increases by approximately 1 mg (Heaney and Layman, 2008).

Differences in habitual protein intakes have been related to bone mass and risk of fracture. Several recent systematic reviews and meta-analyses examining the relation between dietary protein intake and indices of bone health, such as bone mineral density and risk of fractures, have been published (Darling et al., 2009, 2019; Santesso et al., 2012; Shams-White et al., 2017, 2018; Wallace and Frankenfeld, 2017; Wu et al., 2015). Weak positive associations or no associations between total protein intake and bone health have been reported, and studies have found no overall adverse effects of higher protein intakes. Urinary calcium excretion increases in response to acute increases in protein intake, but intestinal calcium absorption also increases by an amount nearly comparable to that lost in urine. Insufficient intakes of protein can adversely impact muscle mass and function. In addition, low dietary protein intake has been associated with reductions in serum insulin-like growth factor 1 (IGF-1) concentrations. IFG-1 plays an essential role in skeletal health via its impact of osteoblast formation and bone growth. Evidence suggests that protein and calcium intake have a positive and synergistic effect on bone health, with higher intakes of protein coupled with higher intakes of calcium being associated with lower incidence of fractures and a slower rate of bone loss in adults (Dawson-Hughes and Harris, 2002; Sahni et al., 2010). Recent systematic reviews and meta-analyses reported that limited evidence showed no differences in the effect of plant vs. animal proteins on bone health (Darling et al., 2019; Shams-White et al., 2018; Wallace and Frankenfeld, 2017). More research is required to address the relative impact of the quantity and type of protein (animal vs. plant) on skeletal health.

Phytoestrogens

Phytoestrogens are dietary components that have a chemical structure similar to that of endogenous estrogens. The primary phytoestrogens in the diet are obtained from soybean isoflavones (including genistein and daidzein). These compounds appear to be able to weakly mediate some of the genomic and non-genomic effects of estrogen and may function as agonists or antagonists depending on the tissue and type of estrogen receptor involved. To date, supplemental sources of these compounds have not been found to decrease fracture risk (Abdi et al., 2016; Lagari and Levis, 2013). Additional clinical data will assist in definition of the long-term impact of phytoestrogens on bone health and fracture risk.

Homocysteine

For some time, it has been known that individuals with a genetic defect in homocysteine metabolism (homocystinuria) have an increased risk of early onset osteoporosis but less was known about the potential impact of circulating homocysteine concentrations on bone health among the general population. Much interest in this topic was generated by studies reporting significant relationships between serum homocysteine and lower bone mineral density in adults (Bailey et al., 2015; Enneman et al., 2014; Gerdhem et al., 2007; Zhang et al., 2014). The strength of the relationship observed is substantial and is comparable to the relationship found between serum homocysteine concentrations and cardiovascular disease. At present the mechanisms responsible for the impact of homocysteine concentrations on fracture risk are not known. Increased homocysteine concentrations could possibly interfere with normal collagen production and other effects of elevated homocysteine concentrations on bone health may be indirect. Further research will assist in identifying the mechanisms and relationships between homocysteine and bone health and the degree to which this relationship is influenced by nutritional status (particularly folate, vitamin B_{12} and vitamin B_6), sex, age, and ethnicity.

Other lifestyle factors

Lifestyle choices such as smoking, alcohol abuse, physical activity and sleep also impact overall bone health. Excessive alcohol intake is a risk factor for low bone mass (Cheraghi et al., 2019; Maurel et al., 2012). This finding may be a consequence of poor dietary quality in chronic alcoholics and may also be related to adverse effects of excessive alcohol intake on osteoblast function. Cigarette smoking also adversely impacts bone health (Yoon et al., 2012). Smokers may be leaner, and female smokers may experience an earlier menopause and have lower post-menopausal estrogen levels. Smoking may also have adverse effects on bone cells themselves either directly or indirectly through an increase in oxidative stress.

Exercise is known to positively influence bone mass (Kemmler et al., 2020). During exercise the strain placed on bone stimulates the osteocyte to positively influence the balance in bone remodeling. Many studies have found positive associations between exercise and bone mass at several sites, especially the hip and the spine. The impact of exercise on bone mass is related to the intensity of the exercise and is associated with the degree to which it increases the habitual physical activity level of the individual. The impact of exercise on bone mass is also influenced by diet and may be most efficacious when calcium intake is optimal (Daly et al., 2014). Exercise not only impacts bone mass but also influences muscle strength, muscle mass, balance and coordination. These improvements in muscle strength may also lead to improvements in posture, balance, flexibility, coordination and gait stability that influence the risk of falls. Robust evidence exists on the importance of physical activity in children and adolescents, particularly during peak bone mass accrual. Currently, only one-half of youth meet the recommendations for daily 60 min of moderate to vigorous intensity physical activity. Efforts should be directed to promoting physical activity during this crucial time of bone development (Weaver et al., 2016b).

A growing body of literature suggests that sleep may also have an impact on bone health (Moradi et al., 2017; Wang et al., 2018). Observational studies have shown a U-shaped association between sleep duration and risk of osteoporosis, with short (<8 h) or long (>8–9 h) sleep durations being associated with a higher risk of osteoporosis in adults. Currently, the mechanisms by which sleep affects bone health are unknown. One potential mechanism is the alteration of normal circadian rhythmicity in bone cells, as both osteoclasts and osteocytes express clock-controlled genes (Swanson et al., 2018). More research is needed to understand the relationship between sleep and bone health as it may be mediated by factors such as physical inactivity and hormonal imbalances.

Targeting groups at risk

The World Health Organization has evaluated risk factors for fracture using global epidemiology data. Using these data, an electronic web-based program called FRAX® has been developed that allows individuals to calculate their 10-y risk of major osteoporotic fracture using their individualized risk factors and other attributes that have been linked to osteoporotic fracture. Many of these factors are identified in **Table 3**.

Nutrient/gene interactions

Optimal nutrition is needed to supply the necessary substrates for bone but an individual's ability to utilize a given nutrient is influenced by their genetic makeup as a substantial amount of bone mineral acquisition (up to 80%) is genetically determined.

Many candidate genes have been associated or linked with the risk of osteoporosis or fracture, including genes coding for hormones (parathyroid hormone), receptors (including the parathyroid hormone, vitamin D, estrogen, glucocorticoid and calcitonin receptors), cytokines and growth factors (including the insulin-like growth factor 1, transforming growth factor B, epidermal growth factor, interleukin 4 and interleukin 6) and bone matrix proteins (such as osteocalcin, collagen type I ($\alpha 1$ and $\alpha 2$) and collagen type II ($\alpha 1$)). While many of these genes have obvious roles in bone metabolism, other candidate genes (such as those coding for apolipoprotein E and methylenetetrahydrofolate reductase) have less obvious relationships to bone mass (Yuan et al., 2019).

Table 3 Factors that have been associated with risk of osteoporotic fracture.

Nonmodifiable risk factors
Height
Shorter stature is associated with a lower net bone mass
Age
Risk of osteoporotic fracture in women doubles every 7–8 years after the age of 50
Sex
Women are at greater risk from lower peak bone mass and loss that occurs during menopause
Low femoral neck BMD
A 1 SD decrease in BMD increases fracture risk by approximately 1.5–2.6 fold
Prior fragility fracture
Is associated with a nearly two-fold increased risk of subsequent fracture
Parental history of fracture
Up to 80% of the variability in bone mass is due to genetic factors
Modifiable risk factors
Weight
Low weight ($<$127 lb or BMI $<$21 kg m^{-2}) is a risk factor for low bone mass
Current tobacco smoking
Smoking is associated with leanness, earlier menopause and may be toxic to bone cells
Alcohol
Alcohol intake greater than two units per day
Physical activity
Sedentary behavior and lack of weight bearing activity
Dietary intake
Low intake of Ca and vitamin D and a poorly balanced diet low in fruits/vegetables and protein and high in sodium
Disease-related risk factors
Medication use
Oral glucocorticoids, GnRH agonists, depot medroxyprogesterone acetate (MPS), aromatase inhibitors, heparin, anticonvulsants (phenytoin)
Diseases
Rheumatoid arthritis, lupus, hyperthyroidism, type 1 diabetes, ankylosing spondylitis, Cushing's, renal failure, total gastrectomy, gastrointestinal diseases (IBD and celiac), diseases that reduce mobility (stroke, Parkinson's, multiple sclerosis), organ transplant
Other clinical factors
History of fainting, falls, or dizziness
Muscle weakness
Neuropathy of the lower extremities
Impaired vision

Diet may influence interactions between genotype and environmental factors. For instance, the impact of exercise on bone can be influenced by the habitual dietary calcium intake and the individuals' genotype (such as the vitamin D receptor genotype). Further research into the genetic control of bone mineral acquisition and loss will be invaluable in targeting groups at risk for low bone mass and may eventually be useful is setting genotype-specific intakes of bone-related nutrients to maximize skeletal health throughout the lifecycle. Newer work is utilizing genome wide association studies (GWAS) to identify genes associated with bone mass in various pathways such as those involved in vitamin D metabolism. Currently, ~30 GWAS studies that have been conducted have identified over 100 independent loci that are associated with bone mineral density (Estrada et al., 2012; Sabik and Farber, 2017). This trait has a complex genetic architecture, and more research is needed to identify which loci harbor variants that possess a causative role on the risk of osteoporosis and associated fractures. These studies will shed new light on genetic determinants of bone with the long-term goal of identifying those at risk so interventions and nutritional recommendations can be initiated prior to the development of osteoporosis.

Interactions of the bone and other physiological systems

Bone health is integrally related to other physiological systems, and perturbations in these systems that are caused by under- or over-nutrition may also impact bone health. Excess adiposity, insufficient muscle mass, a dysbiotic gut microbiome, altered gut barrier function, or impaired immune system, can all have a negative impact on bone mass and/or quality. Excess adiposity adversely affects bone health by a variety of mechanisms involving adipose-derived endocrine factors, such as the adipokines leptin and adiponectin, and obesity-induced inflammation and oxidative stress (Shapses et al., 2017). Substantial crosstalk between bone and muscle occurs with various factors, such as genes, nutrition, mechanical pressure, inflammation, and endocrine factors affecting both bone and skeletal muscle simultaneously. Additionally, the muscle secretes a panel of myokines that act on the bone in an endocrine fashion. Myokines, such as myostatin, transforming-growth factor-β (TGF-β), and IL-6 negatively affect bone formation,

while IGF-1, FGF-2 and irisin promote bone formation (Kaji, 2016). Key nutrients, such as vitamin D, may play a key role in mediating the relationship between muscle and bone. Vitamin D inhibits myostatin production in muscle tissue and decreased myostatin results in greater bone mass (Gunton et al., 2015). Coordinated regulation of bone and muscle by neuronal pathways has also been explored (Houweling et al., 2015). Identifying unifying factors mediating these relationships may be an important avenue toward understanding and treating both osteoporosis and sarcopenia.

Increasing evidence obtained from studies using germ-free mice has demonstrated a connection between gut microbiota and bone homeostasis (Lu et al., 2021). Although the mechanisms underlying this relationship are not fully elucidated, the gut-microbiome/bone relationship may be mediated by the gut-microbiome's effect on the immune system. More research is needed to fully understand the overall impact of nutritional status and individual nutrients on the health of other physiological systems that are also integral to skeletal health.

References

Abdi, F., Alimoradi, Z., Haqi, P., Mahdizad, F., 2016. Effects of phytoestrogens on bone mineral density during the menopause transition: a systematic review of randomized, controlled trials. Climacteric 19 (6), 535–545.

Abrams, S.A., Grusak, M.A., Stuff, J., O'Brien, K.O., 1997. Calcium and magnesium balance in 9–14-y-old children. Am. J. Clin. Nutr. 66 (5), 1172–1177.

Bailey, R.L., Looker, A.C., Lu, Z., et al., 2015. B-vitamin status and bone mineral density and risk of lumbar osteoporosis in older females in the United States. Am. J. Clin. Nutr. 102 (3), 687–694.

Baxter-Jones, A.D., Faulkner, R.A., Forwood, M.R., Mirwald, R.L., Bailey, D.A., 2011. Bone mineral accrual from 8 to 30 years of age: an estimation of peak bone mass. J. Bone Miner. Res. 26 (8), 1729–1739.

Cheraghi, Z., Doosti-Irani, A., Almasi-Hashiani, A., et al., 2019. The effect of alcohol on osteoporosis: a systematic review and meta-analysis. Drug Alcohol Depend. 197, 197–202.

Chun, R.F., Shieh, A., Gottlieb, C., et al., 2019. Vitamin D binding protein and the biological activity of vitamin D. Front. Endocrinol. 10, 718.

Daly, R.M., Duckham, R.L., Gianoudis, J., 2014. Evidence for an interaction between exercise and nutrition for improving bone and muscle health. Curr. Osteoporos. Rep. 12 (2), 219–226.

Darling, A.L., Millward, D.J., Torgerson, D.J., Hewitt, C.E., Lanham-New, S.A., 2009. Dietary protein and bone health: a systematic review and meta-analysis. Am. J. Clin. Nutr. 90 (6), 1674–1692.

Darling, A.L., Manders, R.J.F., Sahni, S., Zhu, K., Hewitt, C.E., Prince, R.L., Millward, D.J., Lanham-New, S.A., 2019. Dietary protein and bone health across the life-course: an updated systematic review and meta-analysis over 40 years. Osteoporos. Int. 30 (4), 741–761.

Dawson-Hughes, B., Harris, S.S., 2002. Calcium intake influences the association of protein intake with rates of bone loss in elderly men and women. Am. J. Clin. Nutr. 75 (4), 773–779.

Enneman, A.W., Swart, K.M., Zillikens, M.C., et al., 2014. The association between plasma homocysteine levels and bone quality and bone mineral density parameters in older persons. Bone 63, 141–146.

Estrada, K., Styrkarsdottir, U., Evangelou, E., et al., 2012. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nat. Genet. 44 (5), 491–501.

Farsinejad-Marj, M., Saneei, P., Esmaillzadeh, A., 2016. Dietary magnesium intake, bone mineral density and risk of fracture: a systematic review and meta-analysis. Osteoporos. Int. 27 (4), 1389–1399.

Gagnon, C., Schafer, A.L., 2018. Bone health after bariatric surgery. JBMR Plus 2 (3), 121–133.

Geiker, N.R.W., Mølgaard, C., Iuliano, S., et al., 2020. Impact of whole dairy matrix on musculoskeletal health and aging-current knowledge and research gaps. Osteoporos. Int. 31 (4), 601–615.

Gerdhem, P., Ivaska, K.K., Isaksson, A., et al., 2007. Associations between homocysteine, bone turnover, BMD, mortality, and fracture risk in elderly women. J. Bone Miner. Res. 22 (1), 127–134.

Gunton, J.E., Girgis, C.M., Baldock, P.A., Lips, P., 2015. Bone muscle interactions and vitamin D. Bone 80, 89–94.

Harvey, N.C., Biver, E., Kaufman, J.M., et al., 2017. The role of calcium supplementation in healthy musculoskeletal ageing : an expert consensus meeting of the European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases (ESCEO) and the International Foundation for Osteoporosis (IOF). Osteoporos. Int. 28 (2), 447–462.

Heaney, R.P., Layman, D.K., 2008. Amount and type of protein influences bone health. Am. J. Clin. Nutr. 87 (5), 1567s–1570s.

Heaney, R.P., Saville, P.D., Recker, R.R., 1975. Calcium absorption as a function of calcium intake. J. Lab. Clin. Med. 85 (6), 881–890.

Holick, M.F., Binkley, N.C., Bischoff-Ferrari, H.A., et al., 2011. Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline. J. Clin. Endocrinol. Metab. 96 (7), 1911–1930.

Houweling, P., Kulkarni, R.N., Baldock, P.A., 2015. Neuronal control of bone and muscle. Bone 80, 95–100.

Kahwati, L.C., Weber, R.P., Pan, H., et al., 2018. Vitamin D, calcium, or combined supplementation for the primary prevention of fractures in community-dwelling adults: evidence report and systematic review for the US Preventive Services Task force. JAMA 319 (15), 1600–1612.

Kaji, H., 2016. Effects of myokines on bone. Bonekey Rep. 5, 826.

Kemmler, W., Shojaa, M., Kohl, M., von Stengel, S., 2020. Effects of different types of exercise on bone mineral density in postmenopausal women: a systematic review and meta-analysis. Calcif. Tissue Int. 107 (5), 409–439.

Kindler, J.M., Lobene, A.J., Vogel, K.A., et al., 2019. Adiposity, insulin resistance, and bone mass in children and adolescents. J. Clin. Endocrinol. Metab. 104 (3), 892–899.

Lagari, V.S., Levis, S., 2013. Phytoestrogens in the prevention of postmenopausal bone loss. J. Clin. Densitom. 16 (4), 445–449.

Lu, L., Chen, X., Liu, Y., Yu, X., 2021. Gut microbiota and bone metabolism. FASEB J. 35 (7), e21740.

Maurel, D.B., Boisseau, N., Benhamou, C.L., Jaffre, C., 2012. Alcohol and bone: review of dose effects and mechanisms. Osteoporos. Int. 23 (1), 1–16.

Moradi, S., Shab-Bidar, S., Alizadeh, S., Djafarian, K., 2017. Association between sleep duration and osteoporosis risk in middle-aged and elderly women: a systematic review and meta-analysis of observational studies. Metabolism 69, 199–206.

Mott, A., Bradley, T., Wright, K., et al., 2019. Effect of vitamin K on bone mineral density and fractures in adults: an updated systematic review and meta-analysis of randomised controlled trials. Osteoporos. Int. 30 (8), 1543–1559.

Moyer, V.A., 2013. Vitamin D and calcium supplementation to prevent fractures in adults: U.S. Preventive Services Task Force recommendation statement. Ann. Intern. Med. 158 (9), 691–696.

Nordin, B.E., Need, A.G., Morris, H.A., Horowitz, M., 1993. The nature and significance of the relationship between urinary sodium and urinary calcium in women. J. Nutr. 123 (9), 1615–1622.

Pereira-Santos, M., Costa, P.R., Assis, A.M., Santos, C.A., Santos, D.B., 2015. Obesity and vitamin D deficiency: a systematic review and meta-analysis. Obes. Rev. 16 (4), 341–349.

Pool, L.R., Wolf, M., 2017. FGF23 and nutritional metabolism. Annu. Rev. Nutr. 37, 247–268.
Razzaque, M.S., 2018. Magnesium: are we consuming enough? Nutrients 10 (12).
Rosen, C.J., Abrams, S.A., Aloia, J.F., et al., 2012. IOM committee members respond to Endocrine Society vitamin D guideline. J. Clin. Endocrinol. Metab. 97 (4), 1146–1152.
Rosengren, B.E., Karlsson, M., Petersson, I., Englund, M., 2015. The 21st-century landscape of adult fractures: cohort study of a complete adult regional population. J. Bone Miner. Res. 30 (3), 535–542.
Ross, C.A., Taylor, C.L., Yaktine, A.L., Del Valle, H.B. (Eds.), 2011. Dietary Reference Intakes for Calcium and Vitamin D. National Academy of Medicine. National Academies Press (US), Washington (DC).
Sabik, O.L., Farber, C.R., 2017. Using GWAS to identify novel therapeutic targets for osteoporosis. Transl. Res. 181, 15–26.
Sahni, S., Cupples, L.A., McLean, R.R., et al., 2010. Protective effect of high protein and calcium intake on the risk of hip fracture in the Framingham offspring cohort. J. Bone Miner. Res. 25 (12), 2770–2776.
Santesso, N., Akl, E.A., et al., 2012. Effects of higher- versus lower-protein diets on health outcomes: a systematic review and meta-analysis. Eur. J. Clin. Nutr. 66 (7), 780–788.
Sarafrazi, N., Wambogo, A.E., Shepherd, J.A., 2021. Osteoporosis or Low Bone Mass in Older Adults: United States, 2017–2018. National Center for Health Statistics, Hyattsville, MD.
Shams-White, M.M., Chung, M., Du, M., et al., 2017. Dietary protein and bone health: a systematic review and meta-analysis from the National Osteoporosis Foundation. Am. J. Clin. Nutr. 105 (6), 1528–1543.
Shams-White, M.M., Chung, M., Fu, Z., et al., 2018. Animal versus plant protein and adult bone health: a systematic review and meta-analysis from the National Osteoporosis Foundation. PLoS One 13 (2), e0192459.
Shapses, S.A., Pop, L.C., Wang, Y., 2017. Obesity is a concern for bone health with aging. Nutr. Res. 39, 1–13.
Swanson, C.M., Kohrt, W.M., Buxton, O.M., et al., 2018. The importance of the circadian system & sleep for bone health. Metabolism 84, 28–43.
Tripkovic, L., Lambert, H., Hart, K., et al., 2012. Comparison of vitamin D2 and vitamin D3 supplementation in raising serum 25-hydroxyvitamin D status: a systematic review and meta-analysis. Am. J. Clin. Nutr. 95 (6), 1357–1364.
USDA & USHHS, 2020. Dietary Guidelines for Americans, 2020–2025, ninth ed. Washington (DC).
Vorland, C.J., Stremke, E.R., Moorthi, R.N., Hill Gallant, K.M., 2017. Effects of excessive dietary phosphorus intake on bone health. Curr. Osteoporos. Rep. 15 (5), 473–482.
Wallace, T.C., Frankenfeld, C.L., 2017. Dietary protein intake above the current RDA and bone health: a systematic review and meta-analysis. J. Am. Coll. Nutr. 36 (6), 481–496.
Wang, D., Ruan, W., Peng, Y., Li, W., 2018. Sleep duration and the risk of osteoporosis among middle-aged and elderly adults: a dose-response meta-analysis. Osteoporos. Int. 29 (8), 1689–1695.
Weaver, C.M., Peacock, M., 2019. Calcium. Adv. Nutr. 10 (3), 546–548.
Weaver, C.M., Alexander, D.D., Boushey, C.J., et al., 2016a. Calcium plus vitamin D supplementation and risk of fractures: an updated meta-analysis from the National Osteoporosis Foundation. Osteoporos. Int. 27 (1), 367–376.
Weaver, C.M., Gordon, C.M., Janz, K.F., et al., 2016b. The National Osteoporosis Foundation's position statement on peak bone mass development and lifestyle factors: a systematic review and implementation recommendations. Osteoporos. Int. 27 (4), 1281–1386.
Wu, A.M., Sun, X.L., Lv, Q.B., et al., 2015. The relationship between dietary protein consumption and risk of fracture: a subgroup and dose-response meta-analysis of prospective cohort studies. Sci. Rep. 5, 9151.
Yoon, V., Maalouf, N.M., Sakhaee, K., 2012. The effects of smoking on bone metabolism. Osteoporos. Int. 23 (8), 2081–2092.
Yuan, J., Tickner, J., Mullin, B.H., et al., 2019. Advanced genetic approaches in discovery and characterization of genes involved with osteoporosis in mouse and human. Front. Genet. 10, 288.
Zhang, H., Tao, X., Wu, J., 2014. Association of homocysteine, vitamin B12, and folate with bone mineral density in postmenopausal women: a meta-analysis. Arch. Gynecol. Obstet. 289 (5), 1003–1009.

Further reading

Gallagher, J.C., Sai, A.J., 2010. Molecular biology of bone remodeling: implications for new therapeutic targets for osteoporosis. Maturitas 65, 301–307.
Holick, M.F., Dawson-Hughes, B. (Eds.), 2015. Nutrition and Bone Health, second ed. Humana Press, Totowa, NJ.
Rizzoli, R., Biver, E., Brennan-Speranza, T.C., 2021. Nutritional intake and bone health. Lancet Diabetes Endocrinol. 21, 00119-4.
Ross, C., Manson, J.E., Abrams, S.A., et al., 2010. The 2011 report on dietary reference intakes for calcium and vitamin D from the Institute of Medicine: what clinicians need to know. J. Clin. Endocrinol. Metab. 96 (1), 53–58.
The Board of Trustees of The North American Menopause Society, 2010. Management of osteoporosis in postmenopausal women: position statement of the north American Menopause Society. Menopause 17 (1), 25–54.
Tucker, K.L., 2009. Osteoporosis prevention and nutrition. Curr. Osteoporos. Rep. 7 (4), 114–117.
U.S. Department of Health and Human Services, 2004. The 2004 Surgeon General's Report on Bone Health and Osteoporosis. U.S. Department of Health and Human Services, Office of the Surgeon General, Washington, DC.
World Health Organization, 1994. Assessment of Fracture Risk and Its Application to Screening for Postmenopausal Osteoporosis. Technical Report Series 843. World Health Organization, Geneva.

Relevant websites

American Society of Bone and Mineral Research—Bone Curriculum. http://www.asbmr.org.
FRAX—World Health Organization Fracture Risk Assessment Tool. http://www.shef.ac.uk/FRAX/.
National Osteoporosis Foundation. www.nof.org.
Surgeon General's Report on Bone Health. http://www.surgeongeneral.gov.

Pediatric feeding disorders: Feeding children who can't or won't eat

Sarah E. Fleet[a,b] **and Ryan D. Davidson**[b,c], [a] Growth and Nutrition Program, Division of Gastroenterology, Hepatology & Nutrition, Boston Children's Hospital, Boston, MA, United States; [b] Harvard Medical School, Boston, MA, United States; and [c] Boston Children's Hospital, Boston, MA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of R.M. Katz, J.K. Hyche, E.K. Wingert, Pediatric Feeding Disorders: Feeding Children Who Can't or Won't Eat, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 21–27, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00157-4.

Introduction 622
Classification of feeding disorders in children 623
Feeding and swallowing physiology 623
Evaluation and treatment of pediatric feeding disorders 623
Evaluation 623
Physicians 624
Feeding specialists 627
Dietitians 627
Behavioral psychologists 628
Social workers 628
Cultural considerations within evaluations 629
Treatment of feeding disorders 629
Multidisciplinary approach 629
Medical intervention 629
Nutritional support 629
Feeding skill development 629
Behavioral interventions 630
Caregiver involvement 630
Treatment settings 631
Conclusion 632
Further reading 632

Key points

- Define pediatric feeding disorder using current guidelines
- Be familiar with normal feeding and digestion physiology
- Understand the multidisciplinary approach to evaluation and management of pediatric feeding disorders
- Identify medical comorbidities common with pediatric feeding disorders
- Appreciate psychosocial contributors to successful evaluation and treatment

Introduction

Feeding is the process by which children accept and digest food in amounts adequate to meet their nutritional needs for proper growth. Though seemingly a simple instinctive act, feeding children is actually a complex process requiring a combination of successful caregiver interaction, the child's oral–motor skills, and intact gastrointestinal motility and absorption. Feeding disorders have traditionally escaped recognition by primary providers due to a "normal" growth curve or an assumption that feeding disorders are a normal, transient occurrence in childhood which impedes the opportunity for early intervention.

The term "feeding disorder" is used when young children are unable to eat enough to maintain their nutritional needs. Feeding disorders are surprisingly common in children, and it has been reported that 25–35% of typically developing children will have mild feeding disorders, and up to 75% of children with disabilities, including Autism Spectrum Disorder, will have more severe feeding problems. Clinical manifestations include food refusal/selectivity, gagging, vomiting, swallowing difficulty, poor weight gain, or failure to thrive. Conventionally, these disorders have been grouped into medical, oral–motor, and behavioral categories with most children exhibiting overlapping problems. Therefore, it is more accurate to think of these disorders as a continuum from strictly physiologic defects to primarily behavioral, with the vast majority of these disorders as a combination of many etiologies.

Classification of feeding disorders in children

Given the complexity of feeding disorders and various etiologies, numerous attempts at classifying feeding disorders have been made based on apparent etiology, physical condition, or associated behaviors. Currently, there are two primary definitions using slightly different frameworks. The first, *Pediatric Feeding Disorder* (*PFD*) is based on the framework of the International Classification of Functioning, Disability, and Health, and defined as "impaired oral intake that is not age-appropriate, and is associated with medical, nutritional, feeding skill, and/or psychological dysfunction." The second, *Avoidant/Restrictive Food Intake Disorder* (*ARFID*) based on Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition (DSM-V) is defined as "an eating or feeding disturbance (e.g., apparent lack of interest in eating or food; avoidance based on sensory characteristics of food; concern about aversive consequences of eating) as manifested by persistent failure to meet appropriate nutrition and/or energy needs associated with one (or more) of the following: (1) significant weight loss (or failure to achieve expected weight gain or faltering growth in children); (2) significant nutritional deficiency; (3) dependence on enteral feeding or oral nutritional supplements; (4) marked interference with psychological functioning." In addition, the disturbance cannot be attributed solely to a medical condition or other mental disorder, lack of available food, and there is no evidence of concern with how the individual experiences body weight or shape. Though there are two frameworks with which to view feeding disorders under, they both have an emphasis on functioning and consider medical, nutritional, oral-motor skill, and psychological factors.

Feeding and swallowing physiology

Understanding the mechanisms involved in feeding is useful in understanding why a child refuses to eat. The swallowing process is conventionally divided into three phases: oral, pharyngeal, and esophageal. The oral phase includes the oral preparatory phase and the oral phase of the swallow. The oral preparatory phase occurs when food is manipulated in the mouth and masticated if necessary, reducing it to a consistency ready for swallowing. This phase of swallowing includes the transit of the bolus over the tongue posteriorly until the pharyngeal swallow is triggered. In the newborn and young infant, all phases are driven reflexively by typical rooting and sucking behavior. The pharyngeal phase begins when swallowing is triggered and the bolus is moved through the pharynx into the esophagus. The esophageal phase begins with the opening of the upper esophageal sphincter, and then peristalsis then carries the bolus into the stomach. After passing the lower esophageal sphincter, food enters the stomach, beginning the gastrointestinal and absorptive phase of feeding. Food is emptied from the stomach into the small intestine based on the volume, nutrient composition, and caloric density of the meal. Most nutrient absorption then occurs in the small intestine, while the large intestine generally resorbs fluids that are ingested and secreted during normal digestion.

Poor coordination, timing, or motor dysfunction during any phase of swallowing may lead to aspiration or other feeding difficulties, and medical conditions affecting any portion of the GI tract may lead to discomfort either during or after feeding which may also contribute to feeding difficulties. To understand feeding and swallowing disorders, one must recognize that there are dynamically changing developmental skills and social abilities in the growing child. Progression through the normal stages of feeding (see **Table 1**) requires the attainment of physical abilities such as postural stability, oral–motor coordination, and sensory awareness. In addition, factors such as emerging cognitive skills play an important role in an effective caregiver–child feeding interaction.

It is also important to understand that as very young children age, the oral phase of sucking, chewing, and managing food comes under more voluntary control, requiring cortical integration of sensory/motor input to coordinate the complex patterns of jaw, tongue, and oral movements, which is why some feeding difficulties may not present until after the newborn period.

Evaluation and treatment of pediatric feeding disorders

Due to the diverse and complex factors that contribute to feeding disorders, literature supports the use of a multidisciplinary team for evaluation and treatment of pediatric feeding disorders whenever possible. This team ideally would include a variety of pediatric specialists, including physicians (e.g., general pediatricians, developmental pediatricians, pediatric gastroenterologists, allergists, and otolaryngologists), nurse practitioners, dietitians, feeding specialists (e.g., speech and language pathologists, occupational therapists etc.) psychologists, and social workers.

Evaluation

An appropriate evaluation is the critical first step in initiating treatment for a feeding disorder. The assembled team must begin its approach to diagnosis and therapy with a comprehensive assessment completed by all disciplines. These include a careful prenatal, birth, and neonatal history. Assessment of nutrition and medical status of the child should be accompanied by an appropriate psychological and/or developmental evaluation.

Table 1 Developmental progression of feeding summary.

Age	*Developmental skill*	*Progression of textures*	*Behavioral expectations*
Birth to 4 months	Needs to be fully supported Suck-swallow-breath Rooting and sucking	Bottle/breastfeeding/formula—liquids only	Eats every few hours Volume increases from ~1/2 oz to 4–5 oz
4–6 months	Supported sitting Hand to mouth Holds own bottle Sucking Gag reflux starts to decrease	Liquids only appropriate Closer to 6 months—smooth state 1 or 2 baby pureed foods (by spoon)	Volume increases from 4 oz to 8 oz
6–8 months (after holding head up)	Can sit upright with support Unilateral hand to mouth Vertical munching of easily dissolvable solids	Homemade purees (smooth) Hard teething foods (with supervision) Finger foods may be introduced May start baby-led weaning, but should be cautious as skills may not be present	8 oz every 4–5 h Touching and exploring food is important Placed in high chair 1–2/day
8–9 months (after hand to mouth)	Needs sitting support Introducing cup and straw drinking Grasps small items Rotary jaw movements emerging	Dissolvable solids *Breastmilk/formula still primary source of nutrition	Palming foods Meals are MESSY Distractible Might not want to sit in high chair
10–12 months	Minimal assistance for sitting Mostly bottle, but cup drinks Teething started Self-feeding and assisted feeding	Beginning table foods/soft mashable foods	Accidental dropping foods, though sometimes intentional Bottle refusal common
12–18 months	Sits upright without support Self-feeding with spoon emerging Chewing	Intermediate soft table foods (some mixed textures)	Sits in high chair ~10–15 min Intentional throwing behaviors emerge
18–24 months	Mostly self-feeding Grasp and pinch present Cup drinking Efficient chewing Most teeth in	Broad range of table foods Difficult, tough, and chewy foods	Expresses preferences (can be strong expressions)
2–4 years	2: intense preferences Circulatory jaw rotations Holds cup with 1 hand Self-feeds with fork 3–4: individuation, toilet training	High risk choking foods still monitored	15–20 min meals (max 30 min) 5–6 "mini meals" a day Tantrums (during meals and other times) Lots of "no" statements
School age	Reasoning, negotiating Still concrete thinking		20–30 min meals Starts to understand behavior/consequences
High school	Individuation Personal identity Social autonomy		Exploring foods outside of family/culture Increased eating outside of home setting Skipping meals, increasing snacking Dietary restrictions

Physicians

Medical issues (i.e., gastroesophageal reflux, milk protein allergy, cleft palate, etc.) that occur very early in infancy may be the initial cause for food refusal. Consequently, for the majority of children with a feeding disorder, an early food avoidance pattern is common. The early food avoidance pattern may significantly impact the caregiver–child interaction surrounding feeding and meals. For example, because of severe gastroesophageal reflux, the child may learn to associate eating with pain. Consequently, when the caregivers attempt to feed the child, they will encounter food refusal behavior, which leads most caregivers to terminate the meal prematurely, or can lead to coercive feeding practices, both of which are detrimental. Thus, even when the medical condition is medically managed, the child will still have the learned history of pain associated with eating, but now also has the new history of having refusal behaviors to escape the meal, or, further negative association with eating due to coercive feeding practices. Due to this connection, certain groups may be at a higher risk for feeding difficulties. For example, children with food allergies, gastroesophageal reflux, eosinophilic gastroenteropathy, and motility disorders may demonstrate increased food refusal. A variety of medical conditions such as cardiopulmonary, genetic, and metabolic disorders can also lead to poor appetite and slow weight gain.

The most important role for the physician is to assess for any of these conditions that could contribute to feeding difficulties as an initial cause, as well as identify any disorder that would require treatment before the implementation of a therapeutic treatment program for feeding refusal (see Table 2 for comorbid conditions). A complete history, including past medical history, family

Table 2 Medical conditions associated with pediatric feeding disorders.

Disorders of the oral and pharyngeal phases of swallowing anatomic lesions
Cleft lip or palate
Pierre-Robin sequence
Choanal atresia
Laryngeal clefts
Macroglossia
CHARGE association
Acquired structural abnormalities
Dental caries
Tonsillar hypertrophy
Viral/inflammatory stomatitis
Retropharyngeal mass
Candida stomatitis
Cardiopulmonary disorders
Chronic lung disease
Complex congenital heart disease
Heart failure
Reactive airway disease
Tachypnea
Neuromuscular disorders
Familial dysautonomia
Cerebral palsy
Pseudo-bulbar palsy Bulbar atresia or palsy
Cranial nerve anomalies
Muscular dystrophic disorders
Arnold–Chiari malformation
Myelomeningocele
Intracranial mass lesions
Disorders of the esophageal phase of swallowing anatomic lesions
Esophageal atresia
Cricopharyngeal achalasia
Tracheoesophageal fistula
Esophageal mass
Esophageal stricture
Esophageal web
Esophageal rings
Vascular rings/aberrant vessels
Foreign bodies
Gastroenterological disorders
Peptic esophagitis
Candida esophagitis
Viral esophagitis
Pill esophagitis
Inflammatory bowel disease
Behcet syndrome
Motility disorders
Achalasia
Diffuse esophageal spasm
Chronic pseudo-obstruction
Gastroesophageal reflux
Constipation
Gas-bloat syndrome
Dumping syndrome
Genetic disorders
Prader–Willi syndrome
Trisomy 21
Cornelia de Lange syndrome
Velo-cardio-facial syndrome
Rett syndrome
Metabolic disorders
Urea cycle abnormalities
Hereditary fructose intolerance

(Continued)

Table 2 Medical conditions associated with pediatric feeding disorders.—cont'd

Hypothyroidism
Miscellaneous
Food allergies
Sensory loss (visual/auditory impairment)
Systemic lupus erythematosis
Polymyositis

history, social history, and medical review of systems may provide clues to any of these conditions. Physical examination of the child includes a general survey examination for the determination of any underlying medical disorders that may preclude safe feeding. It should include an assessment of nutrition status, evaluation of tongue and jaw movement, dentition, airway sounds, speech, and oral cavity assessment as well as a complete cardiac, pulmonary, and abdominal exam. For children with severely limited intake, an assessment of hydration status should also be made.

A full laboratory evaluation should be considered in children with feeding disorders to conduct a nutrition assessment as well as identify possible comorbid conditions. A blood count, iron panel and 25-hydroxy vitamin D can help identify dietary deficiencies in these nutrients. Based on a diet recall, other micronutrient levels may be needed, and the physician should work with the dietitian to identify these. Inflammatory markers, electrolytes, liver function tests, celiac serologies, thyroid function tests, and urinalysis may each be helpful to identify a possible co-morbid disorder. Infants may require a stool guaiac to assess cow's milk protein intolerance, and further stool analysis may be required if malabsorption is suspected.

Based on the history, physical exam and direct clinical feeding observation, diagnostic evaluations may be warranted to better assess swallowing and anatomy (see **Table 3**). The decision to obtain further study of swallowing function is often the result of a collaborative discussion between a physician and feeding specialist. The modified barium swallow (MBS) study is the most widely used procedure to assess oral, pharyngeal, and upper esophageal phases of swallowing. Positioning, food texture, bolus size, rate, and the amount of food presented can be manipulated during the performance of the MBS study to determine the safest and most efficient method of feeding. Clinical evaluation before the MBS is essential so that the food textures, liquid consistencies, and treatment techniques can be included at the time of the study. Changes in head and neck position, such as chin tuck, should be tried before the actual study to better correlate clinical and radiologic findings. The child's level of cooperation should also be assessed before the MBS.

Additionally, a standard upper gastrointestinal contrast series utilizing barium may be required for assessment of anatomy of the gastrointestinal tract (**Fig. 1**). Children with food allergies, repetitive vomiting, or abdominal pain may also need endoscopic evaluation, and some children will also need colonoscopy to rule out the possibility of underlying inflammatory disease such as eosinophilic gastroenteropathy. Gastroenterology consultation can help determine the utility of these studies. Depending on symptomatology and physical exam, some children may need cranial imaging to look for evidence of intracranial mass lesions, hydrocephaly, or posterior fossa anomalies, such as the Chiari malformation. If allergic disease is suspected, consultation with a pediatric allergist is recommended.

Table 3 Possible diagnostic evaluations for patients with feeding disorders.

Detailed history and physical examination
Upper gastrointestinal contrast radiography
Esophogram
Small bowel follow-through
Modified barium swallow study
Gastric emptying study
pH monitoring
Esophagogastroduodenoscopy with biopsies
Antroduodenal manometry
Fiberoptic endoscopic evaluation of swallowing (FEES)
CBC
Comprehensive metabolic panel
Thyroid function
Celiac screening markers
RAST analysis for food allergies
Skin test for food allergies
Plasma amino acids
Urine organic acids
Karyotype

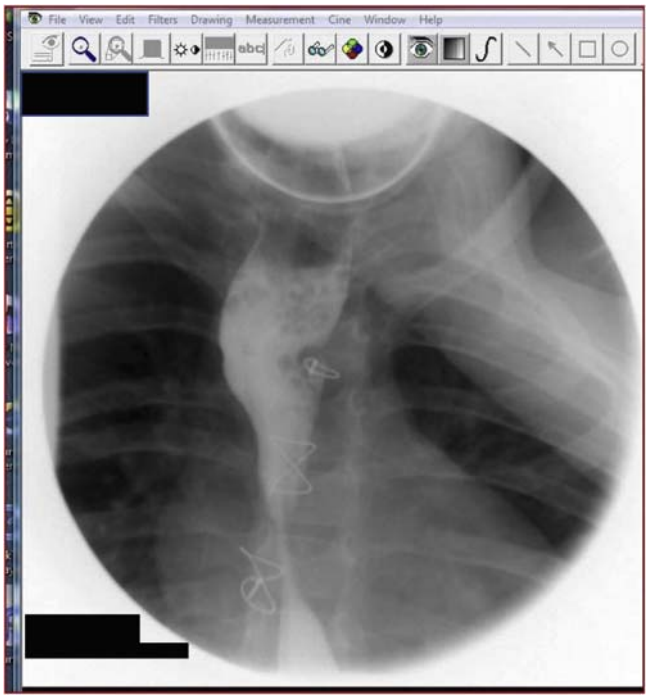


Fig. 1 General behavioral strategies to address mealtime hygiene. Content derived from Satter, E., 1995. Feeding dynamics: helping children to eat well. J. Pediatr. Health Care, 9, 178–184.

Feeding specialists

Oral–motor and swallowing problems are commonly seen in children with congenital and acquired neurologic conditions such as cerebral palsy, structural abnormalities, or traumatic brain injury. Premature and medically fragile infants may miss sensitive periods of oral–motor development resulting in delayed acquisition of feeding skills. This early interruption of feeding skill development and lack of experience can lead to oral aversion or more serious feeding disorders such as food refusal.

The feeding specialist refers to the team member whose responsibilities include assessing and treating oral–motor and swallowing dysfunction and performing instrumental evaluations of swallowing when warranted. The exact credentials of a feeding specialist may vary by location and institution, but often include speech and language pathologists, occupational therapists, and developmental specialists. Occupational therapists may also evaluate fine motor, sensorimotor, and visual motor function as well as positioning and the need for adaptive equipment. Speech/language pathologists may also evaluate and make recommendations for communication skills when necessary.

As part of the initial evaluation, an observation of a feeding session between the child and the primary caregiver will often offer insight into the feeding problem, especially from an oral–motor/sensory and behavioral perspective. Clinical signs of oral–motor dysfunction, length of meals, and nature of the caregiver–child interaction are noted. Observation of the muscle tone, posture and positioning, as well as special seating systems and feeding devices are routine, as these can give insight into the child's overall neurologic functioning.

After a clinical feeding observation, a feeding specialist may decide that further evaluation of swallowing anatomy and function is warranted. A modified barium swallow (MBS) may be recommended and can obtained with collaboration between the patient's feeding therapist and physician. A feeding specialist may also pursue fiberoptic endoscopic evaluation of swallowing (FEES) which allows for direct visualization of the nasal, pharyngeal, and laryngeal structures. This procedure enables evaluation of events occurring immediately before and after the pharyngeal phase of swallow. FEES allows for visualization of any residue in the epiglottic valleculae and pyriform sinuses, and at times, it may be possible to see aspirated materials below the level of the true vocal folds. This procedure does not provide information on the oral phase of swallowing. FEES may also be combined with sensory testing (FEESST). Pulses of air are directed at mechano-receptors within the larynx and reactions are observed that show potential sites of impairment. In older children, this could guide different swallowing techniques to prevent coughing, choking, or aspiration.

Dietitians

Dietitians dedicated to pediatric care are also essential members of the diagnostic team. The dietitian's assessment of the current nutritional status, anticipated growth, and recommended energy intake (that is both age and diagnosis appropriate) is an essential

part of the diagnostic and therapeutic process. Children with feeding disorders are less likely to take in the recommended amounts of vitamins and minerals, so careful diet recalls including precise foods and estimated amounts are important to identify risks for nutrient deficiency, and adequacy of caloric intake. Obtaining a list of preferred foods and beverages can be helpful when making recommendations for more calorie or nutrient dense meal and snack options. Assessing accurate volumes of fluid intake are equally important to identify risk for dehydration. A dietitian's assessment of nutrition status is crucial to determining the acuity of a patient's feeding disorder and identification of protein-calorie malnutrition is important as these patients may be at risk for refeeding syndrome if energy intake is increased significantly and suddenly.

It is also important to consider potential comorbid medical conditions that may increase energy needs. For example, children with cystic fibrosis or other pulmonary disorders may require an increased energy intake which can further exacerbate feeding difficulties if families are already struggling with the child accepting even their standard energy needs. The dietitian is instrumental in predicting energy and macronutrient requirements in these patients, and monitoring growth trends.

Behavioral psychologists

Behavioral difficulties such as food refusal or selectivity are not isolated problems and are often multifactorial in nature. Oftentimes a medical illness may adversely affect feeding patterns, making mealtimes unenjoyable for children and impacting child-caregiver interactions as families struggle to improve feeding behaviors. When there are concerns for failure to thrive or malnutrition, immediate solutions (e.g., highly caloric supplemental beverages), or supplemental enteral feeds (via nasogastric or gastronomy tube), may exacerbate feeding difficulties. These strategies, though necessary medically, can adversely impact hunger, feeding experience, endurance during mealtimes, and inadvertently reinforce escape behaviors around feeding.

Within the setting of pediatric feeding disorders, behavioral psychologists, assess for emotional and behavioral patterns to feeding. They may complete detailed behavioral observations, extensive interviews focusing on antecedents and consequences of behaviors, developmental screenings and assessments, and psychiatric review of systems to assess for comorbid conditions (behavioral and psychiatric) as a part of the assessment process. As a part of the functional behavioral assessment, psychologists assess for which behaviors are potentially reinforced by escape from negative physical experiences associated with medical or skill-based concerns; associated behaviors may have been inadvertently reinforced by social responses. In many cases challenging feeding behaviors are reinforced by both medical/skill-based and social factors.

Given the context of pediatric feeding difficulties, assessments by behavioral psychologists are often completed from the perspective of family systems approaches to help identify a variety of individual and family factors which may contribute to feeding challenges. As the child demonstrates increased distress around food and mealtime, caregivers may engage in a variety of accommodation strategies (modifying routines, offering preferred foods, limiting going elsewhere for meals). Though these strategies are understandable and in response to the child's behavior and the strategies may reduce distress in the short-term, the child does not learn new strategies to cope with the feeding difficulties and the symptoms continue. One of the behavioral psychologists' goals in the initial evaluation is to begin to understand this accommodation cycle.

Many children with pediatric feeding disorders also have comorbid developmental, behavioral, or psychiatric disorders which can exacerbate feeding difficulties. As a part of the developmental screening and/or assessment as well as the psychiatric review of systems, the behavioral psychologists assess for whether additional neurodevelopmental or neuropsychological assessment may provide diagnostic clarification. In addition, they can assess for whether additional in-home or outpatient therapy is required to address comorbid conditions which can impact feeding, such as broader anxiety disorders, Attention Deficit Hyperactivity Disorder or if there are concerns regarding body image disturbances.

Social workers

Since medical issues, behavioral needs, and family psychodynamics all play a central role in the development of the abnormal feeding patterns, a clinical social work evaluation is necessary for assessment and intervention of underlying familial interactions and support systems. Caring for children with feeding disorders negatively impacts caregivers quality of life and pediatric feeding disorders are associated with increased caregiver-related distress, which can also impact adherence to medical recommendations as seen in other pediatric populations. Social workers can assess for caregiver distress and help guide the medical team to develop family-centered treatment plans. They can then help provide additional support and resources to the family as needed to enact these plans This approach can help to ensure continued success once the child has returned to the home environment.

Additionally, assessment of food and housing security is needed to ensure that the reason for selective eating is not due to lack of availability of food, or inability to store food. Globally, the number of undernourished individuals was decreasing up until 2014, but has been increasing since. In the United States, prior to the COVID-19 pandemic, food insecurity continued to decrease but over 11% of the population still experienced food insecurity. Since the COVID-19 pandemic, food insecurity concerns have risen drastically around the globe, particularly for women and those with lower socioeconomic status. Social workers can meet with families to help assess whether there are concerns for food insecurity. In addition, social workers' familiarity with local resources in the region allows them to help families obtain support from community, state, and federal resources.

Specific resources and interventions to address food insecurity may vary based on location. In high income countries, insecurity is often a sign of unequal distribution of resources. For these families, community and federal resources such as WIC, SNAP, and food banks are often recommended by social workers, in addition to school-based feeding programs. Often in lower income countries, food insecurity is a result of lack of access to food resources and addressing strategies for families and communities to be prepared for natural disasters is important for prevention. When there are already concerns for access or in response to natural disasters, income assistance, provision of tools, cash transfers, or emergency aid are important interventions.

The role of social workers extends beyond working with individual families. Social workers can also help the medical team collaborate with community members. In places such as the United States this may include collaborating with Early Intervention, Head Start, private daycares, and schools and as children get older. Social workers are paramount to ensuring consistency in plan across all environments. In other countries, social workers may work on developing community education, peer support groups, conducting home visits, and engaging in advocacy work.

Cultural considerations within evaluations

All specialties should be aware of cultural considerations throughout the evaluation. Meals are inherently social activities and food has significant cultural implications within family structures. Not only are there differences in acceptable foods across cultures, but there are differences in timing, location, and mealtime expectations based on cultural preferences. Within pediatric feeding disorders, perception of weight concerns and concern with picky feeding are another area in which culture plays an important role. Food and feeding are such important aspects of culture that there are also differences in the role of food across cultures. It is important for providers to ask questions to help understand each family's values, culture around feeding, and to understand caregiver perceptions and concerns about weight and food intake.

Treatment of feeding disorders

Multidisciplinary approach

Throughout the treatment process, a multidisciplinary approach is paramount. The initial and perhaps most important part of any therapeutic approach to introducing or increasing oral food intake is to establish the safety of eating as well as the types and textures of food the child can consume most efficiently. The physician, dietician, feeding specialist, behavioral psychologist, and social worker work together to identify potential safety concerns and discuss barriers to treatment, including feeding skill deficits, behavioral, developmental, or psychological comorbidities, and family risk and protective factors.

Medical intervention

The goal of all therapy is directed toward allowing caregivers to safely feed their children in a developmentally appropriate manner. The physician in the treatment team must ensure that all appropriate diagnostic studies have been performed to determine whether an underlying medical condition has predisposed a child to developing an unusual feeding pattern. This includes appropriate utilization of consultants and diagnostic modalities (Table 3). Once these studies have been performed, the physician must coordinate all the resources and direct care so that feeding therapy may proceed with minimal risk to the patient, keeping the child safe from aspiration and other complications.

Nutritional support

Throughout the intervention process, dietitians monitor concerns for malnutrition, micronutrient deficiencies, and dehydration. They may make recommendations regarding necessary volume and types of food intake required to maintain appropriate nutritional status. In addition, they can help families identify strategies to increase caloric intake, whether through supplementing current diet with foods found in the grocery story or through use of high calorie supplemental beverages. Dietitians can help caregivers identify acceptable sources of nutrients using store-bought foods or over-the-counter supplements for patients with nutrient deficiencies.

In the case where there are urgent concerns regarding malnutrition, dangerous micronutrient deficiencies, or significant safety concerns with the act of feeding, providers may decide that enteral feeding support through the use of a nasogastric or gastrostomy tube is needed, for which hospital admission may be necessary to expedite this process. The feeding tube allows for safe delivery of nutrients while the patient and family work with the care team on implementation of the treatment plan, with the eventual goal of tube removal and full oral feeding.

Feeding skill development

If there are safety concerns with chewing skills or the child has not been cleared for PO intake, feeding specialists may use more nonnutritive approaches. These may include oral stimulation, which is performed to reduce hypersensitivity, facilitate management

of secretions, establish or retrain the swallowing mechanism, and maintain coordination of breathing and swallowing. As the child's skills improve or they are cleared for oral intake by medical providers, the feeding specialist may gradually transition to nutritive approaches.

Oral–motor techniques and positioning to improve muscle strength and postural control as a foundation for feeding and swallowing are largely based on a neurodevelopmental framework. The use of adaptive seating systems is a key component to feeding a child with physical disabilities that require external devices to provide head, neck, and trunk support. Attention must be paid as to how positioning affects the feeding process, as a change in head and neck posture and oral–motor structures may affect oral–motor control.

Once airway safety, sensitivity, and positioning have been assessed and a management plan has been implemented, a variety of treatment approaches have been suggested for children with feeding disorders. Goals often include advancing accepted food textures and flavors (including increasing volume of nonpreferred food), and improvement in self-feeding, including utensil use.

Behavioral interventions

Feeding specialists and the behavioral psychologist work together to develop and implement a behavioral plan. Behavioral interventions for pediatric feeding disorders are the most common modality of therapy and are often included within a multidisciplinary team approach that also addresses physiology, oral–motor functioning, caregiver–child interactions, and community or social support.

In general, behavioral strategies are presented in a sequence beginning with the use of the most positive and least intensive strategies that have a reasonable expectation of being effective. These less intensive strategies have the advantage of minimizing deviations from the norm in mealtimes for most families. More intensive behavioral strategies are utilized as needed. Although these more intensive strategies do require significant deviations from the norm of typical family meals and can be a burden on caregivers, these procedures can usually be faded over time.

Many of the less intensive strategies involve mealtime "hygiene" strategies (see Fig. 2). These include the establishment of mealtime structure (same time and place for meals, minimizing snacking or grazing in between meals) and minimizing distracted mealtimes (transitioning screens/electronics to after the meal). Other strategies include the types of foods offered (modifying textures of food offered), including at least 1 preferred food at each meal with 1 nonpreferred or a challenge food offered regularly. If children engage in escape behaviors, such as throwing food, identifying a more acceptable behavior to communicate their preferences can be helpful. To expand variety of accepted foods, interventions often focus on opportunities for the child to engage with new or nonpreferred foods in positive or neutral settings.

Creating a neutral or positive mealtime environment can be challenging as both child and caregivers have experienced potentially thousands of negative mealtimes. It is important to work with caregiver to stay calm during mealtimes. Increased pressure, emotion (positive or negative), and stress can make mealtimes less enjoyable for both child and caregiver, thus worsening feeding difficulties. Many families understandably engage in behaviors which may increase accepted bites in the short-term, but inadvertently reinforce feeding difficulties in the long-term, such as hiding nonpreferred foods in preferred foods, extending mealtimes, or continuing to ask for additional bites during a meal. However, these behaviors can lead to increased mistrust by the child and can make behavioral interventions more challenging later on.

As feeding behaviors become more challenging, or if children need additional behavioral support, intensive behavioral interventions often include a combination of exposure and reinforcement principles. Exposure based interventions may be conducted by behavioral psychologist, feeding specialist, or a combination and include graduated exposures to more challenging foods. Exposures are often paired with reinforcement-based interventions including rewarding appropriate eating behavior (i.e., positive reinforcement), ignoring food refusal behavior (i.e., escape extinction), or removing positive reinforcement in response to food refusal behavior (i.e., response cost). Thus, if a child accepted a bite, he or she would be rewarded with attention or a concrete reinforcer, such as a preferred toy, music, or video. If the child engaged in food refusal behavior, such as batting at the spoon or turning their head away from the food, potential attentional reinforcers would be limited through withdrawal of attention or limiting access to a toy. As the behavioral psychologist and feeding specialist work with the child and caregivers through these trials and interventions, they continue engage in data collection, observation, and assessment to update their conceptualization and then modify exposures and reinforcers to increase chance of success for the child.

Caregiver involvement

It is critical that caregivers are involved throughout all of the treatment process, but especially with behavioral approaches to feeding interventions. The behavioral psychologist and social worker can work together to identify what caregiver's abilities are to implement behavioral interventions based on caregiver factors, family structure, family work structure, and which caregivers are involved in mealtimes. Solely providing didactic information to caregivers regarding which strategies to use, is not as effective as skill-based caregiver training which involves step-by-step criteria-based training. When possible, caregiver training, including instruction, discussion, handouts, role-playing, feedback, and the practice of techniques and videotape with a trained clinician can result in increased caregiver treatment integrity.

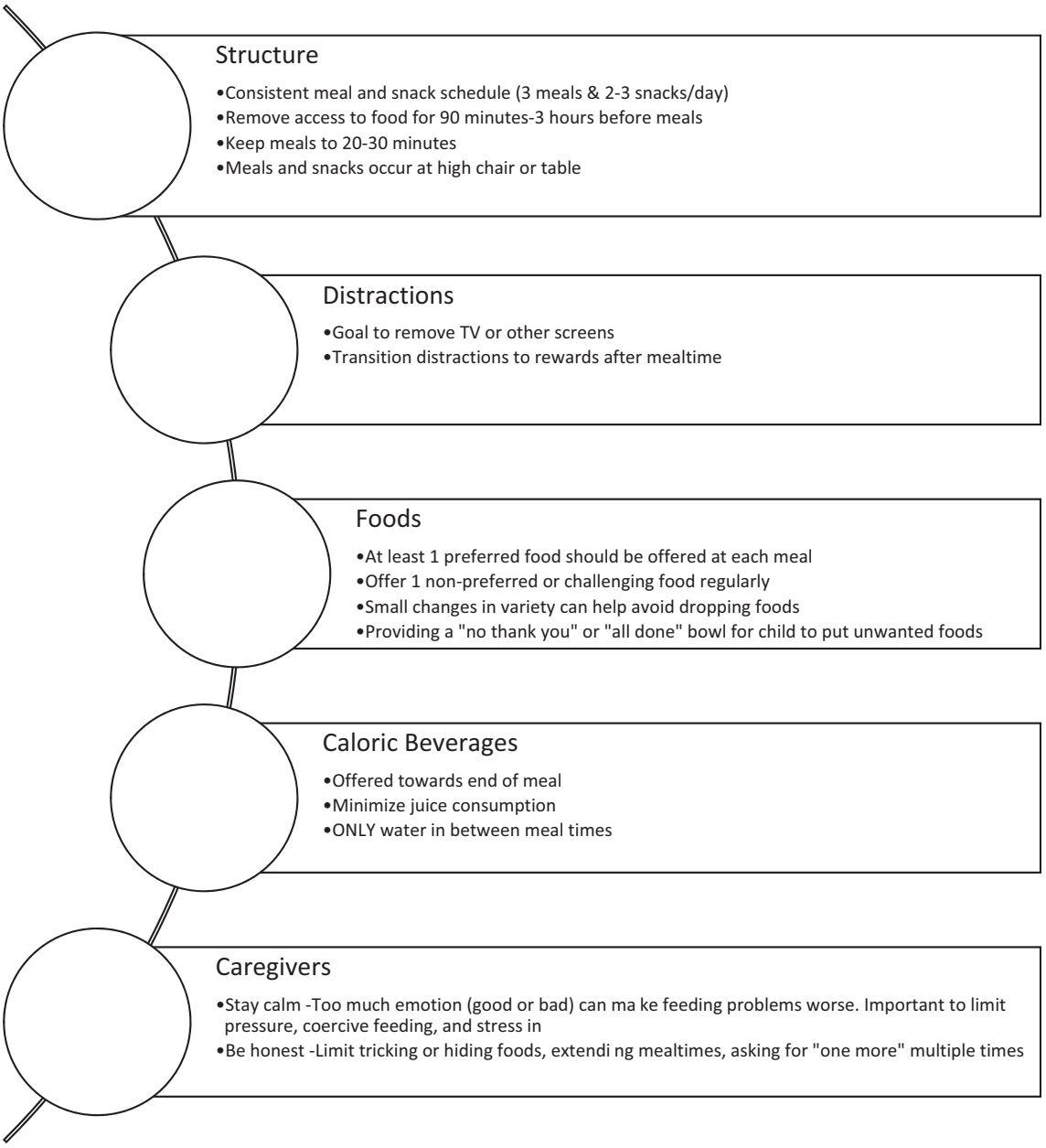


Fig. 2 Abnormal esophagus secondary to esophageal stricture following repair of esophageal atresia.

It is also important to work with caregivers from a stance of cultural awareness and sensitivity. Whether associated with cultural or individual family preferences, reinforcement systems can be a topic met with hesitancy by caregivers. Working with the family to understand their beliefs and caregiver values and providing psychoeducation and/or modifying interventions to meet the family's value systems is important for maintaining caregiver engagement.

Treatment settings

The above-mentioned strategies may be implemented in a variety of settings, depending on illness severity, geographic location, comorbid medical conditions, and psychosocial needs of the family. Across the United States, multidisciplinary programs generally fall into the following categories: inpatient, partial-inpatient, outpatient intensive, or outpatient consultative, with the inpatient program being the most intensive. While many patients would benefit from an inpatient program, the availability of these programs

is limited, and waiting lists are often long. Outpatient programs may not offer the initial intensity of an inpatient program, but do offer long-term continuity of care. Ultimately, the choice in care setting should be approached using the family-centered model of a multidisciplinary team.

Conclusion

Children who can't or won't eat require a systematic diagnostic and therapeutic approach by a team of dedicated professionals from many specialties. Proper evaluation for underlying medical conditions, deficits in feeding skills, and behavioral contributions are necessary for successful treatment. There is an important psychosocial domain that must be addressed and treatment plans must be family-centered. The goal of safe oral feeding is attainable in most children when those involved in the care of children understand the complexity of eating and the associated medical and psychological conditions that comprise a feeding disorder. Helping these children to eat will allow independence from artificial sources of nutrition such as gastrostomy feeds and parenteral nutrition, and ultimately reduce family mealtime stress.

Further reading

American Psychiatric Association, 2013. DSM-5 Diagnostic Classification. In: Diagnostic and Statistical Manual of Mental Disorders, vol. 10.
Arvedson, J.C., 2006. Swallowing and feeding in infants and young children. GI Motil. https://doi.org/10.1038/gim017.
Arvedson, J.C., 2008. Assessment of pediatric dysphagia and feeding disorders: clinical and instrumental approaches. Dev. Disabil. 14 (2), 118–127.
Austin, Texas: PRO-ED.
Babbitt, R.L., Hoch, T.A., Coe, D.A., et al., 1994. Behavioral assessment and treatment of pediatric feeding disorders. J. Dev. Behav. Pediatr. 15, 278–291.
Burklow, K., Phelps, A., Schultz, J., et al., 1998. Classifying complex pediatric feeding problems. J. Pediatr. Gastroenterol. Nutr. 27, 143–147.
Butler, A.M., Titus, C., 2015. Systematic review of engagement in culturally adapted parent training for disruptive behavior. J. Early Interv. 37 (4), 300–318.
Cafiero, C., Viviani, S., Nord, M., 2018. Food security measurement in a global context: the food insecurity experience scale. Measurement 116, 146–152.
Coleman-Jensen, A., Rabbitt, M.P., Gregory, C.A., Singh, A., 2018. Statistical Supplement to Household Food Security in the United States in 2017 (No. 2238-2019-2890).
Dellert, S.F., Hyams, J.S., Treem, W.R., Geertsma, M.A., 1993. Feeding resistance and gastroesophageal reflux in infancy. J. Pediatr. Gastroenterol. Nutr. 17, 66–71.
Evans, A., Seth, J.G., Smith, S., et al., 2011. Parental feeding practices and concerns related to child underweight, picky eating, and using food to calm differ according to ethnicity/race, acculturation, and income. Matern. Child Health J. 15 (7), 899–909.
Feinberg, E., Kavanagh, P.L., Young, R.L., Prudent, N., 2008. Food insecurity and compensatory feeding practices among urban black families. Pediatrics 122 (4), e854–e860.
Gholampour, T., Noroozi, M., Zavoshy, R., et al., 2020. Relationship between household food insecurity and growth disorders in children aged 3 to 6 in Qazvin City, Iran. Pediatr. Gastroenterol. Hepatol. Nutr. 23 (5), 447–456. https://doi.org/10.5223/pghn.2020.23.5.447.
Goday, P.S., Huh, S.Y., Silverman, A., et al., 2019. Pediatric feeding disorder: consensus definition and conceptual framework. J. Pediatr. Gastroenterol. Nutr. 68 (1), 124.
Gosa, M.M., Dodrill, P., Lefton-Greif, M.A., Silverman, A., 2020. A multidisciplinary approach to pediatric feeding disorders: roles of the speech-language pathologist and behavioral psychologist. Am. J. Speech Lang. Pathol. 29 (2S), 956–966.
Gundersen, C., Hake, M., Dewey, A., Engelhard, E., 2021. Food insecurity during COVID-19. Appl. Econ. Perspect. Pol. 43 (1), 153–161.
Harshman, S.G., Wons, O., Rogers, M.S., et al., 2019. A diet high in processed foods, total carbohydrates and added sugars, and low in vegetables and protein is characteristic of youth with avoidant/restrictive food intake disorder. Nutrients 11 (9), 2013.
Kerwin, M.E., 1999. Empirically supported treatments in pediatric psychology: severe feeding problems. J. Pediatr. Psychol. 24 (3), 193–214.
Lau, A.S., 2006. Making the case for selective and directed cultural adaptations of evidence-based treatments: examples from parent training. Clin. Psychol. Sci. Pract. 13 (4), 295.
Lefton-Greif, M.A., Okelo, S.O., Wright, J.M., et al., 2014. Impact of children's feeding/swallowing problems: validation of a new caregiver instrument. Dysphagia 29 (6), 671–677.
Lefton-Greif, M.A., 2008. Pediatric dysphagia. Phys. Med. Rehabil. Clin. 19 (4), 837–851 ix.
Logemann, J., 1983. Evaluation and Treatment of Swallowing Disorders. College Hill Press, San Diego, CA.
Logemann, J., 1998. Evaluation and Treatment of Swallowing Disorders, second ed.
Manikam, R., Perman, J., 2000. Pediatric feeding disorders. J. Clin. Gastroenterol. 30 (1), 34–45.
Morris, S.E., Klein, M.D., 2000. Pre Feeding Skills. A Comprehensive Resource for Mealtime Development, third ed. Therapy Skill Builders.
Munk, D.D., Repp, A.C., 1994. Behavioral assessment of feeding problems of individuals with severe disabilities. J. Appl. Behav. Anal. 27, 241–250.
Pak, L., Allen, P.J., 2012. The impact of maternal depression on children with asthma. Pediatr. Nurs. 38 (1), 11.
Piazza, C.C., Fisher, W.W., Brown, K.A., et al., 2003. Functional analysis of inappropriate mealtime behaviors. J. Appl. Behav. Anal. 37, 187–204.
Picchioni, F., Goulao, L.F., Roberfroid, D., 2021. The impact of COVID-19 on diet quality, food security and nutrition in Low and Middle Income Countries: a systematic review of the evidence. Clin. Nutr. https://doi.org/10.1016/j.clnu.2021.08.015.
Rommel, N., DeMeyer, A.M., Feenstra, L., Veereman-Wauters, G., 2003. The complexity of feeding problems in 700 infants and young children presenting to a tertiary care institution. J. Pediatr. Gastroenterol. Nutr. 7, 75–84.
Rudolph, C., Link, D., 2002. Feeding disorders in infants and children. Pediatr. Clin. 49 (1), 97–112.
Satter, E., 1995. Feeding dynamics: helping children to eat well. J. Pediatr. Health Care 9 (4), 178–184.
Sevin, B.M., Gulotta, C.S., Sierp, B.J., et al., 2002. Analysis of a response class hierarchy of food refusal behavior. J. Appl. Behav. Anal. 35 (1), 65–68.
Shore, B.A., Piazza, C.C., 1997. Pediatric feeding disorders. In: Konarski, E.A., Favell, J.E. (Eds.), Manual for the Assessment and Treatment of the Behavior Disorders of People With Mental Retardation. Western Carolina Center Foundation, Morganton, NC.
Silverman, A.H., Erato, G., Goday, P., 2021. The relationship between chronic paediatric feeding disorders and caregiver stress. J. Child Health Care 25 (1), 69–80.
Tuchman, D., 1994. Physiology of the swallowing apparatus. In: Tuchman, D. (Ed.), Disorders of Feeding and Swallowing in Infants and Children. Singular Publishing, San Diego, CA.
Wolf, L.,S., Glass, R., 1992. Feeding and Swallowing Disorders in Infancy—Assessment and Management. Therapy Skill Builders, Tucson.
World Health Organization, 2020. The State of Food Security and Nutrition in the World 2020: Transforming Food Systems for Affordable Healthy Diets, vol. 2020. Food & Agriculture Org.
Yagc-Küpeli, B., Akyüz, C., Küpeli, S., Büyükpamukçu, M., 2012. Health-related quality of life in pediatric cancer survivors: a multifactorial assessment including parental factors. J. Pediatr. Hematol. Oncol. 34 (3), 194–199.
Yu, K., Wu, F., Eisenberg Colman, M.H., 2021. Low-income caregivers' attitudes and behaviors on children's diets: emergent themes on cultural influences and perceived value of nutrition information from healthcare providers. J. Prim. Care Community Health 12, 21501327211003686.

Sarcopenia: Molecular mechanism and current nutritional approach

Kunihiro Sakuma[a], **Akihiko Yamaguchi**[b], **and Muneshige Shimizu**[c], [a] Institute for Liberal Arts, Environment and Society, Tokyo Institute of Technology, Meguro-ku, Tokyo, Japan; [b] Department of Physical Therapy, Health Sciences University of Hokkaido, Kanazawa, Ishikari-Tobetsu, Hokkaido, Japan; and [c] Tokai University, Faculty of Marine Science, Department of Fisheries, Shimizu-ku, Shizuoka, Tokyo, Japan

© 2023 Elsevier Ltd. All rights reserved.

Introduction 633
Molecular mechanism regulating sarcopenia 634
No significant change of ubiquitin-proteasome system (UPS) in sarcopenic muscle 634
Adaptation of autophagy-linked signaling in muscle with age 634
Supplementation with protein and amino acids 635
Amino acids 635
HMB 636
Polyphenol 637
Catechin 637
Soy isoflavone 637
Resveratrol 638
Creatine supplementation 638
Vitamin D 639
Ursolic acid 639
omega-3 PUFAs 639
Caloric restriction (CR) 640
Conclusion 640
Acknowledgments 641
References 641

Key points

- The most possible mechanism for sarcopenia is the defect of protein degradation through autophagy-lysosome signaling.
- Milder CR (15–25%) would also be effective for age-related muscle atrophy in humans.
- Many supplemental candidates for possible combating sarcopenia is needed systematic and fundamental research using rodent models of sarcopenia.

Introduction

The contraction of skeletal muscle is the driving force behind the body's movement and plays an important role in maintaining homeostasis. The skeletal muscle makes up almost half of the human body. Therefore, a decline in the metabolic and contractile properties of skeletal muscle can have very serious consequences for human health. Sarcopenia occurs as a result of age-related loss of muscle strength, quality, and quantity. It often refers to cellular processes (denervation, inflammation, mitochondrial dysfunction) and the resulting loss of muscle strength, function, and mobility and increased risk of falls. Sarcopenia is considered "primary" (or age-related) when there is no apparent cause other than aging (Cruz-Jentoft et al., 2010), especially when it is associated with decreased physical activity or work inactivity. Secondary sarcopenia usually occurs in the presence of several causes. This condition would be plausible for acute or chronic diseases that are common in the elderly, such as stroke, hip fracture, diabetes mellitus, chronic heart failure, and chronic obstructive pulmonary disease (von Haehling et al., 2010). Muscle mass loss is more pronounced in the lower extremity muscle groups, and at the myofiber level, it is characterized by atrophy and fibronecrosis of certain type II muscle fibers. Sarcopenia is characterized by atrophy and fibrous necrosis of specific type II muscle fibers at the muscle fiber level.

The mechanism of age-related muscle atrophy is thought to be influenced by multiple factors simultaneously, including inadequate nutritional intake, hormonal changes, poor muscle regeneration, and increased oxidative stress. Although the specific contribution of these factors is still not elucidated, it has been reported that the regulatory factors of muscle hypertrophy decrease reactively and quantitatively with aging, and that autophagy dysfunction becomes apparent (Sakuma and Yamaguchi, 2010; Wohlgemuth et al., 2010). On the other hand, no apparent enhancement of negative regulators [calpain, NF-κB (nuclear factor-κB), atrophy gene-1 (atrogin-1)] is observed in aging mammalian muscles (Sakuma et al., 2017). It is widely accepted that strategies such as nutritional supplementation and physical training (both aerobic and resistance exercise) are important interventions for

maintaining skeletal muscle mass. In particular, resistance training combining with amino acid supplementation is known to be the most effective in reducing the progression of sarcopenia (Sakuma and Yamaguchi, 2018). On the other hand, sarcopenia has been found to be most attenuated by mild caloric restriction (CR) in all mammals. Recent studies have also pointed to the potential application of new supplements (soy isoflavones, ursolic acid, etc.) in the prevention of muscle atrophy. In this article, we outline the molecular mechanisms of sarcopenia and detail recent nutritional approaches to inhibit this.

Molecular mechanism regulating sarcopenia

No significant change of ubiquitin-proteasome system (UPS) in sarcopenic muscle

UPS is essential for the control of proteolysis, as (1) substrates are covalently bound by multiple ubiquitin molecules and (2) proteins tagged by the 26S proteasome complex are degraded and released. Three enzymes regulate the ubiquitination of protein: ubiquitin activating enzyme (E1), ubiquitin binding enzyme (E2), and ubiquitin ligase (E3). Most famous E3 ubiquitin ligases (MuRF1 and atrogin-1) are increased in various skeletal muscle atrophy models (in vivo) such as cancer, diabetes, renal failure, denervation, and glucocorticoid administration (Bodine et al., 2001). These ubiquitin ligases are particularly important in muscle atrophy because loss of atrogin-1 or MuRF1 in mice inhibits fasting, denervation, and dexamethasone-induced muscle atrophy (Baehr et al., 2011; Cong et al., 2011).

Numerous studies in a variety of mammalian muscles have not yielded a concordant view of atrogin-1 or MuRF1 mRNA and protein changes associated with aging (Sakuma et al., 2017; Edström et al., 2006). A marked upregulation of phosphorylated Akt and forkhead box O (FOXO) 4 appears to occur in the gastrocnemius muscle of aging female rats. In addition, Léger et al. (2008) showed a decrease in nuclear FOXO1 and FOXO3a in 70-year-old humans, despite the absence of obvious age-related changes in these atrogen's mRNA. Sandri et al. (2013) demonstrated that mice with null mutation of atrogin-1 have a greater loss of muscle mass with aging (Sandri et al., 2013). In addition, although muscle mass is at least partially maintained in MuRF1-deficient mice, the age-related decline in muscle strength is more pronounced than in control mice (Hwee et al., 2014). As Sandri et al. (2013) showed, chronic inhibition of these atrogens exacerbates the decline rather than preventing it. As Sandri et al. (2013) have shown, chronic attenuation of these atrogens can't be considered an effective therapeutic target to combat sarcopenia, as it exacerbates decline rather than preventing muscle weakness.

Adaptation of autophagy-linked signaling in muscle with age

Using the lysosomal-dependent pathway, macroautophagy (autophagy) is a catabolic process that extensively degrade cytoplasmic components. This process is characterized by the incorporation of a portion of the cytoplasm into autophagosomes. The autophagosome then fuses with the lysosome to form an autophagolysosome, where the material in the cytoplasm is degraded and the degradation products are recycled.

Autophagy was thought to be a crude and non-selective degradation system. However, autophagy is an extremely precise way to recover altered organelles and abnormal protein aggregates. Autophagy is selective for specific cargo receptors such as Nix (Bnip3L), Nbr1, p62/SQSTM1, and optineurin (Neel et al., 2013). Three molecular complexes regulate the autophagosome formation at least. These are the microtubule-binding protein 1 light chain 3 (LC3)-coupled system and the regulatory complexes governed by beclin-1 and unc51-like kinase-1. The LC3-binding system is composed of various proteins encoded by autophagy-related genes (Atg).

In invertebrates and higher organisms, autophagy declines with normal aging; In Drosophila skeletal muscle, the proteolytic function of the autophagy/lysosome system has been shown to decline with age (Demontis and Perrimon, 2010), suggesting a gradual accumulation of polyubiquitin protein aggregates in aging Drosophila muscle. Interestingly, overexpression of FOXO increases the expression of many autophagy molecules and inhibits the accumulation of polyubiquitin protein aggregates in sarcopenic Drosophila muscle (Demontis and Perrimon, 2010). There are multiple reports on age-related autophagic changes in mammalian skeletal muscle (Wohlgemuth et al., 2010; Gaugler et al., 2011; Wenz et al., 2009). The amount of Beclin-1 was significantly increased in plantar muscles than young Fischer 344 rats (Wohlgemuth et al., 2010). Using immunofluorescence microscopy and Western blotting of fractionated homogenates, we found that p62/SQSTM1 and Beclin-1 selectively accumulated in the cytoplasm of myofibers in sarcopenic mice (Sakuma et al., 2016) (**Fig. 1**).

On the other hand, the amount of Atg7 and Atg9 proteins was not modulated in aged rat plantaris muscles (Wohlgemuth et al., 2010). Furthermore, Wohlgemuth et al. (2010) suggested a significant increase in the amount of LC3 in muscle with aging, but failed to detect a significant increase with age in the ratio of LC3-II to LC3-I, a more suitable marker for assessing autophagy progression. On the other hand, Wenz et al. (2009) pointed out that the ratio of LC3-II to LC3-I increased significantly with age in the biceps femoris of wild-type mice. The transcriptional levels of autophagy-linked molecules did not show a significant increase with age (Wohlgemuth et al., 2010; Gaugler et al., 2011). Not all of the substances involved in autophagy signaling are similarly altered in aging skeletal muscle at both the mRNA and protein levels. Interestingly, human studies (Carnio et al., 2014) using muscle samples clearly demonstrated age-dependent autophagy defects in older human volunteers, including reduced amounts of Atg7 protein and LC3-II/LC3-I protein ratios. Therefore, sarcopenia would be involved in a marked impairment of autophagy signaling, but more detailed investigation is needed in this area. **Fig. 2** is a hypothetical illustration of the autophagic dysfunctional situation in sarcopenic muscle fibers.

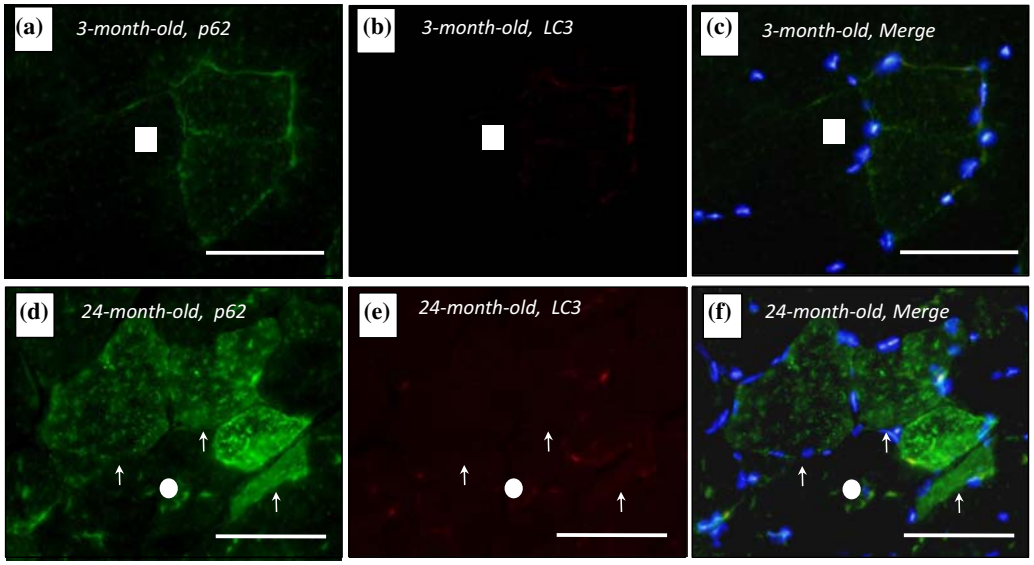


Fig. 1 Serial cryosections of the quadriceps muscle of 3- and 24-month-old mice. p62/SQSTM1 and LC3 immunoreactivity. In young quadriceps muscle, immunofluorescence labeling showed that p62/SQSTM1 was present in the membrane and at a low level in the cytosol of several muscle fibers (A). Marked increases of p62/SQSTM1 immunoreactivity were observed in the membrane and the cytosol of aged muscle fibers (D). No apparent difference in LC3 immunoreactivity was observed in the muscle between 3- and 24-month-old mice (B and E). White arrows denote the muscle fibers possessing p62/SQSTM1. Bar = 50 μm. Data from Sakuma et al. (2016).

Supplementation with protein and amino acids

Amino acids

Dietary protein provides amino acids necessary for muscle protein synthesis and, as an anabolic stimulus, acts directly on protein synthesis. Studies of young adults fed dietary protein suggest that more than half of the amino acids derived from protein are available for circulation within 5 h after a meal, and that ~11% of these amino acids are incorporated into muscle proteins (Groen et al., 2015). The slowing of the anabolic response to protein intake in the elderly is a problem, suggesting that more protein may be needed to prevent loss of muscle strength and muscle mass (Wolfe et al., 2008). Epidemiological studies suggest that there is a positive correlation between maintaining muscle mass and protein intake. A number of reviews suggest that nutritional interventions such as high protein intake and intake of leucine, a branched-chain amino acid, combined with resistance training would be effective in reducing muscle fiber atrophy in sarcopenia (Dickinson et al., 2017; Timmerman and Volpi, 2008). In particular, leucine is considered to be an amino acid with unique characteristics and has been the most studied in recent years. In particular, leucine has been the most studied in recent years and plays a variety of roles, including nitrogen donation for alanine and glutamine synthesis in muscle, glucose homeostasis, and translational regulation of protein synthesis.

Amino acid supplementation after acute resistance exercise has a synergistic effect on contraction-induced escalation of muscle protein synthesis (Walker et al., 2011). In human studies, amino acids have been shown to phosphorylate eukaryotic initiation factors 4F and p70S6K via mTOR. Resistance training and amino acid supplementation have been shown to result in additive hypertrophy (Esmarck et al., 2005). In fact, a study of frail elderly men and women showed that lean body mass increased in the protein supplementation group (in addition to resistance-type exercise training), but did not change in the placebo (exercise only) group (Tieland et al., 2012). Furthermore, the administration of multiple essential amino acids also increases muscle mass and protein synthesis, even under normal conditions without resistance training. (Nicastro et al., 2011). On the other hand, some reports have failed to prove the synergistic effects of long-term supplementation of several amino acids and carbohydrates with resistance training (Godard et al., 2002). Although the combination of both resistance training and amino acid supplementation has been recommended for the prevention of sarcopenia (Sakuma and Yamaguchi, 2018; Wakabayashi and Sakuma, 2014), recent systematic reviews (Beaudart et al., 2017; Thomas et al., 2016) often deny the positive effects of amino acid supplementation. Beaudart et al. (2017) found that the additional effect of nutritional interventions on Beaudart et al. (2017) noted that the additional effect of nutritional interventions on muscle mass was only found in 8 (23.5%) high quality RCTs. Exercise interventions in the elderly increased muscle strength in almost all studies (29/25 RCTs), while nutritional supplements had no beneficial effect in the majority of studies (8/35 RCTs, 22.8%). Therefore, the effects of amino acid supplementation on muscle mass, strength, and body performance may only be seen in elderly people with low nutrition and sarcopenia.

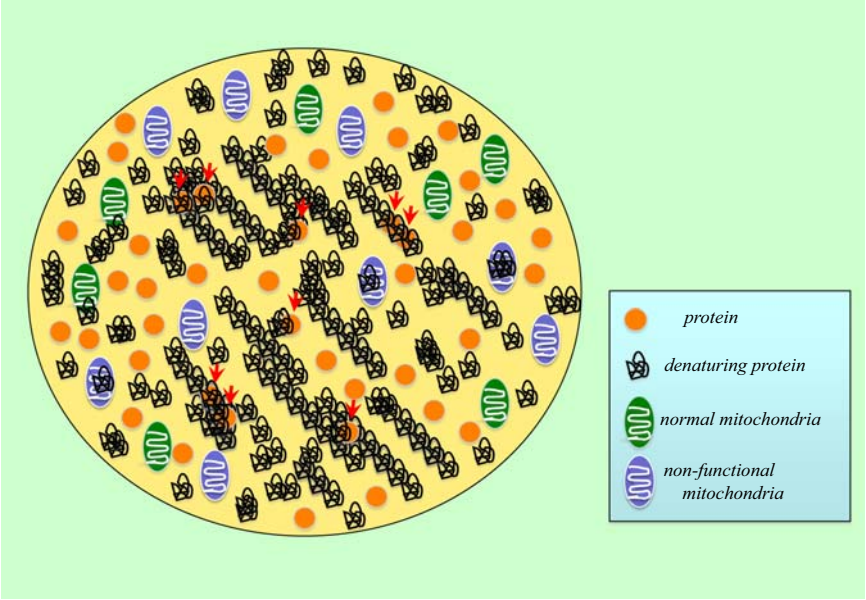


Fig. 2 The hypothesis of the cytosol condition of sarcopenic muscle fibers (autophagic defect). In the cytoplasm of sarcopenic muscle fibers, the normal autophagy-dependent system does not work and large amounts of degenerated proteins and nonfunctional mitochondria accumulate. The denatured proteins can penetrate inside normal mitochondria and become entangled with normal proteins. Some dysfunctional mitochondria left in cells would release substances that are toxic to muscle cells.

HMB

HMB is a metabolite of the branched-chain amino acid (BCAA) leucine. The initial reaction of HMB synthesized in the body is the transamination of leucine to α-ketoisocaproic acid (KID) by BCAA aminotransferases, which occurs primarily in skeletal muscle. HMB has been shown to stimulate protein synthesis via mTOR, a central regulator of mRNA translation efficiency (Wilkinson et al., 2013), reduce proteasome expression and enzyme activity, and decrease apoptosis in myonuclei (Gerlinger-Romero et al., 2011; Hao et al., 2011). He et al. (2016) have demonstrated that HMB supplementation increases mitochondrial biosynthesis and fat oxidation. Indeed, mitochondrial dysfunction and oxidative stress-induced anabolic resistance are observed in the muscles of the elderly (Holecek, 2017).

There are also a certain number of papers on the effects of HMB supplementation in humans, Deutz et al. (2013) used 3 g of HMB during 10 days of bed rest and obtained interesting results. Their results showed that compared to a control group, HMB reduced muscle weakness during bed rest and maintained muscle strength during the rehabilitation period. This was a well-designed study, as the amount of HMB supplementation was comparable to that frequently used in academic research, and the diet of the subjects was controlled. On the other hand, Hsieh et al. (2010) demonstrated that HMB supplementation had a positive effect during a 2–4 week intervention period. The authors found a decrease in blood urea and urinary urea excretion, suggesting a decrease in proteolysis, although the difference in anthropometric parameters was not significant. However, this study has some limitations, including a short supplementation period. On the other hand, a meta-analysis of several randomized controlled trials found that HMB supplementation can prevent loss of lean body mass in older adults without changing fat mass (Wu et al., 2015). HMB supplementation may be maximized when combined with other substances such as arginine and lysine. For example, Flakoll et al. (2004) demonstrated that consumption of an HMB/Arg/Lys mixture (2/5/1.5 g per day) for 12 weeks increased limb circumference, leg and hand grip strength, and whole body protein synthesis. Since neither group showed any improvement in muscle strength, they ruled out an effect of HMB supplementation on muscle strength. One systematic review of high quality supported that there was no difference between HMB and placebo on muscle strength (Wu et al., 2015). However, this meta-analysis presented many evidence to support HMB supplementation on muscle mass (Wu et al., 2015). Taken together, these data suggest that HMB supplementation in the elderly has a positive effect on muscle mass, but no clear effect on physical performance and muscle strength. Two reviews with somewhat insufficient evidence also found that it is not possible to determine whether HMB alone is effective in improving muscle strength, physical performance, and/or mass (Wu et al., 2015; Malafarina et al., 2013), and more research on the effects of HMB supplementation on reducing sarcopenia is needed.

Although it is not clearly elucidated whether HMB alone attenuate sarcopenic symptom, HMB plus vibration improve significantly this. Whole body vibration and local vibration have been implemented as treatment for muscle weakness in the elderly. One RCT using sarcopenic men (mean age 88.6 years) indicated that 8-week vibration (12–16 Hz, 3–5 mm) significantly improved muscle strength and physical performance (gait speed, five-times-sit-to-stand test, and timed-up-and-go test) (Zhu et al., 2019). Although systematic reviews and meta-analyses have shown that vibration therapy may be a positive strategy for improving physical

performance and muscle strength in sarcopenia, caution is needed in interpreting the data due to the limited number of relevant studies (Wu et al., 2020). Very intriguingly, Wang et al. (2020) recently showed that combined low-magnitude high-frequency vibration (LMHFV) and HMB intervention markedly enhanced muscle strength (twitch and tetanic force) and decreased intramuscular fat infiltration (lower oil red O area) in aging SAMP8 mice. Although this study uses an aging mouse model, I feel that the combination of LMHFV and HMB for sarcopenia patients is a very interesting approach.

Polyphenol

Among the natural compounds, polyphenols appear to be of interest to modulate mitochondrial function and protect against metabolic disease by activating mitochondrial biogenesis through SIRT1 and PGC-1α.

Catechin

Tea (Camellia sinensis), especially green tea whose main active ingredient is catechins, contains many phenolic hydroxyl groups (-OH). There are four monomers of catechins: epigallocatechin gallate (EGCG), epicatechin (EC), epigallocatechin (EGC), and epicatechin gallate (ECG). Recently, it has been shown that green tea catechins have the potential to modulate skeletal muscle cell homeostasis and inhibit muscle mass loss. In fact, ECs have enhanced mitochondrial capacity in animal and human studies on cardiac, skeletal muscle, and neuronal cells (Daussin et al., 2021). 8 weeks of EGCG (5 mg/kg, 4x/week) appears to reduce fibrosis and necrotic myofibers in muscular dystrophy mice (Nakae et al., 2008). EGCG supplementation also decreases protein carbonyl content, a marker of oxidative stress, in aged (34 months) male Wistar rats (Senthil Kumaran et al., 2008). Furthermore, in vitro and in rodent model studies, EGCG supplementation leads to increases in myonuclear number, myotube formation, cross-sectional area, and muscle mass (Luk et al., 2020). On the other hand, some results show that the effects of catechin supplementation on muscle adaptation are disparate; Si et al. (2019) conducted an experiment in which 20-month-old male C57/BL mice were fed EC or EGCG (0.25% w/v in drinking water) and a standard diet. The results showed that the survival rate significantly increased from 39% to 69% after 37 weeks of EC intake, whereas no significant effect was observed with EGCG. In addition, EC improved physical activity, slowed skeletal muscle degeneration, reversed age-related mRNA and protein expression of the extracellular matrix and growth factor-activated receptor pathways in skeletal muscle, and reversed age-related decreases in the nicotinic acid and nicotinamide pathways in both serum and skeletal muscle Thus, the effects of EC on skeletal muscle have been demonstrated. Thus, the effect of EC on skeletal muscle was more pronounced than that of EGCG.

That Oligonol, a lychee extract rich in flavanols consisting of phenolic compounds such as catechins and procyanidins, prevents obesity caused by a high-fat diet and kidney damage caused by diabetes (Liu et al., 2016). Feeding 32-week-old senescence-accelerated mouse prone 8 mice (SAMP8) with a chow diet containing 200 mg Kg-1 oligonol for 8 weeks showed that oligonol prevented high-fat diet-induced obesity and diabetes-induced renal damage. Addition of oligonol resulted in decreased expression of mitochondrial biosynthetic genes (PGC-1α etc.) and mitochondrial fusion genes (Mfn2 etc.), increased expression of Atg13, LC3-II and p62, and abundant accumulation of autophagosomes and lysosomes (Chang et al., 2019). Although much has been demonstrated about the effects of supplementation with green tea extracts (EC, EGCG, oligonols, etc.) on mitochondria and myofibers, almost all studies have been conducted using rodent models. In order to apply the results of animal studies to humans, it is necessary to examine the safety, bioavailability, and efficacy of green tea in sarcopenia patients. Most clinical studies on the effects of green tea extract on sarcopenia patients are not randomized controlled clinical trials, and the current limited human studies do not provide sufficient evidence to confirm a beneficial effect on sarcopenia patients.

Soy isoflavone

Isoflavones, a type of flavonoid found in soybeans, are one of the most naturally occurring organic compounds. Structurally similar to estrogen, they bind to estrogen receptors and exert physiological functions (such as antioxidant effects). Estrogen receptors (ERs) include ERα and ERβ, and daidzein, a type of soy isoflavone, acts on ERβ rather than ERα. For example, when isoflavones were added to a high-fat diet for 12 weeks, muscle mass increased in mice after ovariectomy (Beekmann et al., 2015). On the other hand, when isoflavones were fed for as long as 120 days, fat accumulation in skeletal muscle of male mice was suppressed (Kurrat et al., 2015). Thus, it was suggested that isoflavones affect muscle mass and function not only in females but also in male mice. Unfortunately, many researchers have investigated the effects of isoflavone supplementation by assessing only muscle mass, not muscle fiber size (Aoyama et al., 2016; Ogawa et al., 2017). This method of assessment is likely incorrect when large amounts of fat and connective tissue accumulate in atrophied tissues (Wall et al., 2015). In other words, the muscle mass maintenance effect of isoflavones, as assessed by these muscle mass alone, may not accurately reflect changes in muscle fiber size. Abe et al. (2013) examined the effect of isoflavone administration on muscle atrophy by assessing muscle fiber size. The results showed that isoflavone (20% of the diet) administration significantly suppressed fibroatrophy in the tibialis anterior muscle 4 days after denervation. At this time, the expression of IRS-1 and p-Akt1 proteins was significantly increased in the muscles of denervated mice after isoflavone supplementation. Hirasaka et al. (2014) also found that isoflavone inhibited the damage to acetylcholine receptors caused by denervation, suggesting that isoflavone intake may have a protective effect on the neuromuscular junction. Furthermore, an in vitro

study using cells reported that isoflavones suppressed the transcriptional activity of MuRF-1 and inhibited TNF-α-induced myotube atrophy (Hirasaka et al., 2013).

Many experiments on isoflavone supplementation in vivo have utilized amounts exceeding 1% of the dietary intake (Beekmann et al., 2015; Kurrat et al., 2015). Because it is very difficult for humans to consume such amounts of isoflavones at every meal, the supplemental data of Abe et al. (2013) are not realistic for determining the effect of isoflavones on muscle atrophy. Our group (Tabata et al., 2018) demonstrated that administration of 0.6% isoflavone (AglyMax) modulates the degree of denervation-induced muscle fiber atrophy in mice. AglyMax is an isoflavone aglycone that is a mixture of daidzein, genistein, and glycitein in a 7:1:2 ratio, and in humans It has been noted to have a higher absorption rate and quantity than glucosides. This effect is thought to be due to reduced signaling in an apoptosis-dependent system, rather than an atrophy gene-1 (atrophy gene-1)-dependent system.

Daidzein is thought to be effective in older women. 8-Prenylnaringenin, which has estrogenic effects similar to daidzein, is thought to inhibit denervation-induced atrophy of skeletal muscle by promoting phosphorylation of Akt (Mukai et al., 2012). In addition, quercetin, a type of flavonoid found in fruits and vegetables, scavenges ROS. This effect of quercetin is noteworthy because ROS activates NFκB pathway and Foxo pathway and induces the expression of E3 ubiquitin ligase. In addition, in a hindlimb unloading model, quercetin administration to the gastrocnemius muscle decreased the expression of atrogin-1 and MuRF-1 and prevented the loss of skeletal muscle mass (Mukai et al., 2010). Whether daidzein, quercetin, and AglyMax inhibit age-related muscle atrophy needs to be tested, but it seems likely.

Resveratrol

Resveratrol (3,5,4′-trihydroxystilbene), which is found in grape skin, seeds, and peanuts, is a polyphenolic phytoalexin produced by plant species. When young female mice were immobilized on one limb for 7 days, resveratrol prevented loss of muscle mass, muscle fiber CSA, and muscle strength, while simultaneously increasing satellite cell content during the 7-day recovery period (Bennett et al., 2013). In different atrophy model, we (Asami et al., 2018) demonstrated that 0.5% resveratrol diet for young mice significantly prevented decreases in muscle weight and fiber atrophy as well as the increase in atrogin-1 and p62 immunoreactivity after 2 weeks of denervation. Administration of resveratrol causes activation of SIRT1-AMPKα-PGC-1α signaling, which induces a variety of biological responses important for healthy aging (Gülcin, 2010). In fact, four weeks of resveratrol supplementation induced the increase of mitochondrial mass and function, suppression of lipid peroxide, and increases in activity of catalase and superoxide dismutase, causing increased physical endurance in mice (Muhammad and Allam, 2018); however, the effects of resveratrol on sarcopenia have not yielded stable results. Low to moderate daily intake of resveratrol for 6–7 weeks has no effect on muscle strength and function in older mice (Baumann et al., 2014; Zhou et al., 2019). Similarly, long-term consumption (10 months) of low to moderate resveratrol has no beneficial effects on muscle function and mass in aging mice. Overall, resveratrol appears to have a protective effect against sarcopenia. However, more detailed studies are needed, especially in humans, since these effects are contradicted by the age of the experimental species and the variation of dose and duration with the supplementation.

Creatine supplementation

Creatine is a guanidine compound produced endogenously in the liver, kidney and pancreas from arginine, glycine and methionine. Creatine monohydrate (CrM) is one of the most widely used and researched supplements. At the cellular level, CrM has been suggested to reduce oxidative stress, prevent motor neuron shedding, and enhance mitochondrial function. Since phosphocreatine plays an important role in supporting metabolism during high-intensity exercise, this metabolic disturbance is thought to inhibit muscle performance and reduce muscle mass. It is still unclear whether creatine levels change with age, and several studies have failed to show a beneficial effect of CrM supplementation during resistance training in older adults (Bermon et al., 1998; Rawson et al., 1999). On the other hand, CrM supplementation during resistance training in the elderly has been reported to increase muscle mass and strength, endurance, and power (Rawson et al., 1999; Brose et al., 2003). Administration of CrM in a 2-month resistance exercise program (67–80 years old) did not affect the increase in total volume or muscle strength from training (Bermon et al., 1998). On the other hand, longer term (>4 months) studies have demonstrated the beneficial effects of CrM supplementation to further increase muscle strength and muscle mass after resistance training (Rawson et al., 1999; Brose et al., 2003). Some previous studies have shown little benefit of CrM supplementation without resistance training in older adults (Rawson et al., 1999; Parise et al., 2001). More recently, two other systematic reviews of low to moderate quality found some evidence supporting no difference between CrM supplementation combined with progressive resistance training and exercise alone with respect to physical performance, but with respect to muscle mass and strength, some evidence supporting the combined. However, for muscle mass and strength, there was some evidence to support the combined intervention (Wu et al., 2015; Denison et al., 2015).

Vitamin D

Vitamin D has traditionally been considered an important regulator of bone metabolism and phosphorus and calcium homeostasis. Muscle mass loss and vitamin D deficiency often occur together; the prevalence of low vitamin D levels in persons 65 years of age and older is estimated to be about 50%. (Wicherts et al., 2007). The mechanisms by which vitamin D affects muscle function and strength is not fully understood, but is probably vitamin D receptor (VDR)-mediated. Interestingly, the expression of vitamin D receptors in skeletal muscle is also known to decrease with age (Snijder et al., 2006). Long-term vitamin D deficiency causes severe muscle weakness, which improves with vitamin D supplementation.

There is considerable clinical and epidemiological evidence linking vitamin D status to differences in muscle strength and function in the elderly; Annweiler et al. (2009) found that vitamin D supplementation improved muscle strength and performance in a double-blind study of elderly people living in areas with low vitamin D levels and nursing home residents They found that vitamin D supplementation improved muscle strength and performance. In a double-blind study of older adults and nursing home residents living in areas with low vitamin D levels (Annweiler et al., 2009), found that vitamin D supplementation improved muscle strength and performance and reduced the risk of falls. In a double-blind study, they found that vitamin D supplementation improved muscle strength and performance and reduced the risk of falls. Consistent with this observation, a number of epidemiological studies have shown reduced lower extremity function, including longer walking times and sit-to-stand times, in older adults with hypovitaminosis D status (Wicherts et al., 2007; Bischoff-Ferrari et al., 2004). However, the relationship between vitamin D and muscle mass and composition is less clear in comparison to the large body of evidence on the benefits of high vitamin D status on skeletal muscle function. A more comprehensive understanding of the effects of vitamin D on skeletal muscle function, muscle strength, and muscle mass may be a very effective means of preventing muscle atrophy (sarcopenia) in the elderly. However, many studies have shown negative results of treatment with vitamin D in elderly patients without vitamin D deficiency. In the absence of vitamin D deficiency in elderly patients, vitamin D is unlikely to be effective in controlling sarcopenia. In fact, a very recent meta-analysis (Prokopidis et al., 2022) reported that vitamin D intake does not attenuate sarcopenia in community-dwelling older adults.

Ursolic acid

Ursolic acid (UA), found in plants and fruits such as apples, is a natural pentacyclic triterpenoid carboxylic acid; UA has anticarcinogenic, antioxidant, anti-inflammatory, and anti-obesity properties (Prokopidis et al., 2022), and besides, UA has confirmed beneficial effects in animal models of hyperlipidemia and diabetes (Katashima et al., 2017). Supplementation with UA was reported to increase skeletal muscle and muscle strength in a mouse model (Wang et al., 2009). A systematic review, including animal studies, concluded that UA supplementation increases physical fitness (ex. muscle strength) and muscle mass by increasing sirtuin 1 expression and muscle satellite cells and activating the mTOR pathway in muscle. Indeed, Kunkel et al. (2012) found that ursolic acid reduced two different types of muscle atrophy, fasting and muscle atrophy-induced stress, by decreasing the mRNA levels of the muscle atrophy genes atrogin-1 and MuRF1. In addition, UA supplementation increased Akt phosphorylation in skeletal muscle in vivo. Intriguingly, ursolic acid may not have a direct effect on skeletal muscle, because UA alone was not sufficient to promote activation of IGF-I and insulin receptors. More recently, Yu et al. (Kunkel et al., 2011) experimented with long-term (21 days) administration of UA (orally 100 mg/kg) in a chronic kidney disease (CKD) mouse model. They found that administration of ursolic acid significantly inhibited CKD-induced muscle atrophy (tibialis anterior) by decreasing the expression of inflammatory cytokines (IL-6, TNF-α, etc.) and ubiquitin E3 ligases (MuRF-1, atrogin-1, MUSA1, etc.). They found that the decrease in the phosphorylation levels of NF-κB (p65), STAT3, and p38 were significantly inhibited by UA in CKD-induced atrophic muscle.

The hypertrophic effects of UA in humans are less clear. For example, in healthy adults, UA supplementation (50 mg/day provided by ingesting 500 mg of loquat leaf extract) did not improve muscle mass and strength compared to placebo ingestion (Yu et al., 2017). Furthermore, UA (3 g) for resistance-trained men did not modify activation of anabolic signaling pathways after one RT (Cho et al., 2016). Eight weeks of UA supplementation (450 mg/day) combined with RT increased IGF-I levels and muscle strength compared to RT alone, however the authors did not observe a substantial increase in LBM (Church et al., 2016). In summary, UA supplementation may have anabolic effects, but these effects have not been consistently confirmed in humans (Bang et al., 2014). Therefore, further studies are needed to determine the effects of UA supplementation on sarcopenia in humans.

omega-3 PUFAs

The major omega-3 PUFAs are alpha-linolenic acid (ALA), eicosapentaenoic acid (EPA), docosapentaenoic acid (DPA), and docosahexaenoic acid (DHA). In order to ensure structural cell function, the concentration of EPA and DHA in cell membranes is important. Major dietary sources of EPA and DHA are fatty fish (e.g., mackerel) and seafood. ALA is found in plant foods such as chia seeds, vegetable oils, and nuts. Although about 500 mg of EPA and DHA per day is recommended, most adults in the United States do not achieve the recommended intake of DHA and EPA (Lobo et al., 2021).

Supplementation with omega-3 PUFAs provides an interesting alternative opportunity because of low risk of adverse events and anti-inflammatory properties. Supplementation with omega-3 PUFAs may be beneficial for a variety of aging processes, including

bone and eye health, and cognition. Omega-3 PUFAs have also been suggested to have multiple beneficial effects in muscle such as increased muscle fiber size and improved mitochondria function (Vannice and Rasmussen, 2014). In addition, previous randomized controlled trials (RCTs) demonstrated that muscle protein synthesis in the elderly is promoted by omega-3 PUFA supplementation (Gray and Mittendorfer, 2018). When resistance training was combined with omega-3 PUFA supplementation, the improvement in muscle quality and function obtained from resistance training was also significantly enhanced (Smith et al., 2015). Cross-sectional studies have also been conducted on the association between dietary intake of omega-3 fatty acids and physical performance. The association between grip strength and dietary intake was investigated in 2893 men and women aged (59–73 years) included in the Hertfordshire Cohort study. This study indicated that each additional serving of fatty fish was associated with an increase in grip strength of men (0.43 kg) and women (0.48 kg) (Da Boit et al., 2017).

It is generally accepted for the anti-inflammatory effects of omega-3 PUFAs. Supplementation with omega-3 PUFAs have decreased in CRP and IL-6 in middle-aged and elderly people (Robinson et al., 2008). A recent RCT also investigated the effects of DHA and EPA treatment (8 weeks) in the elderly and found significant reductions in levels of IL-6, IL-1β, and TNFα (Custodero et al., 2018). Since chronic low-level inflammation may be involved in the development of sarcopenia, this suppression of inflammation by omega-3 PUFA may be one mechanism to combat sarcopenia. On the other hand, it has also been reported that omega-3 PUFA supplementation does not affect inflammatory markers (Tan et al., 2018), and further studies in sarcopenic older adults are needed.

The mTOR pathway plays an important role in skeletal muscle generation and muscle protein synthesis. Sakuma found increased activation of the mTOR-p70s6k signaling pathway in response to increased amino acid and insulin supply after 8 weeks of omega-3 PUFAs supplementation compared to the placebo group (Tan et al., 2018). It has also been suggested that supplementation with omega-3 PUFAs decreases the expression of inhibitory pathways to mTOR and causes important changes in gene expression that favor skeletal muscle anabolism (Smith et al., 2011). Therefore, omega-3 PUFAs have the potential to overcome age-related anabolic resistance through activation of mTOR signaling.

Caloric restriction (CR)

CR has been recognized as the most powerful intervention to extend lifespan by slowing both primary aging (natural decline associated with aging) and secondary aging (accelerated aging due to disease or negative lifestyle factors) in many species. It has been recognized as the most powerful intervention to extend life by slowing both primary aging (natural decline associated with aging) and secondary aging (accelerated aging due to disease and negative lifestyle factors). It has been recognized as the most powerful intervention to extend lifespan by slowing both primary aging (natural decline associated with aging) and secondary aging (accelerated aging due to disease or negative lifestyle factors). The protective effects of CR, which is recognized as the most potent intervention to extend lifespan in many species, may be attributed to its ability to reduce the occurrence of mitochondrial abnormalities and attenuate oxidative stress CR reduces the total amount of muscle protein synthesis and the fractional synthesis rate of individual proteins in rodents, but not the age-related CR was shown to have no effect on the decrease in mitochondrial density (Smith et al., 2015). These findings suggest that CR does not increase mitochondrial biosynthesis, but rather maintains mitochondrial function by protecting existing cellular components (Smith et al., 2015). Furthermore, CRs are resistant to increased apoptotic signaling in skeletal muscle with aging (Lanza et al., 2012). Notably, CR regulates most of the apoptotic pathways that occur in aging muscle, including mitochondria, cytokines/receptors, and Ca2+/ER stress-mediated signals (Lanza et al., 2012).

How does CR regulate sarcopenia, independent of mitochondrial function and apoptosis? CR increases PGC-1α in various organs such as brain, heart, liver, and visceral adipose tissue. PGC-1α plays a variety of roles, including enhancement of mitochondrial biosynthesis, fatty acid oxidation, autophagy activation, neuromuscular junction gene induction, and myokine secretion. Indeed, Valdez et al. (Dirks and Leeuwenburgh, 2004) showed that lifelong CR in mice significantly suppressed abnormalities at the synapse due to age-related loss protection of motoneurons by PGC-1α. It is known that autophagy in skeletal muscle decreases with age (Wohlgemuth et al., 2010), and normal function of autophagy by CR may inhibit age-related atrophy of muscle fibers.

A large study of CR in primates was conducted and showed favorable results for both all-cause and age-related mortality (Valdez et al., 2010). On the other hand, another study failed to find any significant difference by CR in either all-cause mortality or age-related mortality (Colman et al., 2014). The inhibitory effect of CR on sarcopenia in mice, rats and rhesus monkeys is well known (Mattison et al., 2012; McKiernan et al., 2011). We would like to clarify whether CR suppresses sarcopenia in humans and to what extent dietary restriction is possible.

Fig. 3 represents an overview of nutritional intervention for sarcopenia.

Conclusion

The recent advances in our understanding of muscle biology have led to possible candidate for molecular mechanism and nutritional treatment of sarcopenia. The most possible mechanism for sarcopenia is the defect of protein degradation through autophagy-lysosome signaling. Although resistance training combined with amino acid-containing supplementation is usually recommended to prevent age-related muscle wasting and weakness (Sakuma and Yamaguchi, 2018), supplementation with proteins (amino acids) only did not influence sarcopenic symptoms. Milder CR (15–25%) would also be effective for age-related muscle

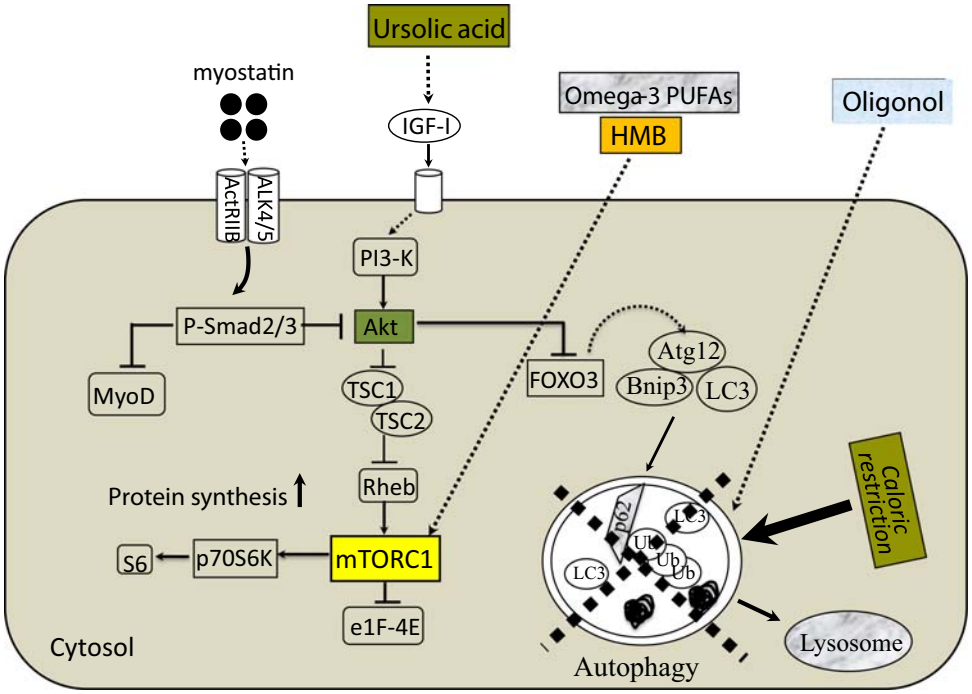


Fig. 3 Nutritional interventions affect different mediators in sarcopenic muscle. Recent findings suggest that the myostatin-Smad pathway inhibits protein synthesis probably due to blocking the functional role of Akt. Treatment with an ursolic acid upregulates the amount of IGF-I and then stimulates protein synthesis by activating the Akt/mTORC1/p70S6K pathway. Administration of omega-3 PUFA or HMB may work to prevent sarcopenia by activating mTORC1 Sarcopenic muscle exhibits a marked defect of autophagy-dependent signaling, which is effectively ameliorated by caloric restriction and oligonol supplementation. ALK activin receptor-like kinase, ActRIIB activin receptor IIB, IGF-I insulin-like growth factor I, *TSC* tuberous sclerosis complex, TORC1 component of TOR signaling complex 1, Rheb Ras homolog enriched in brain, mTORC1 mammalian target of rapamycin complex 1, eIF4E eukaryotic initiation factor 4E, FOXO forkhead box O, *LC3* microtubule-associated protein light chain 3, *BNIP* BCL2/adenovirus E1B 19 kd-interacting protein, *Atg* autophagy-related genes.

atrophy in humans, although some evidence-based modification is necessary. Many candidates such as catechins, soy isoflavones, and ursolic acid may combat sarcopenia, although systematic and fundamental research on this treatment has not been conducted, even in rodent models of sarcopenia (McKiernan et al., 2012).

Acknowledgments

This work was supported by a research Grant-in-Aid for Scientific Research C (No. 20K11359) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Conflict of interest.

Kunihiro Sakuma, Akihiko Yamaguchi, and Muneshige Shimizu declare that they have no conflict of interest.

References

Abe, T., Kohno, S., Yama, T., Ochi, A., Suto, T., Hirasaka, K., Ohno, A., Teshima-Kondo, S., Okumura, Y., Oarada, M., Choi, I., Mukai, R., Terao, J., Nikawa, T., 2013. Soy glycinin contains a functional inhibitory sequence against muscle-atrophy-associated ubiquitin ligase Cbl-b. Int. J. Endocrinol. 2013, 907565.

Annweiler, C., Schott, A.M., Berrut, G., Fantino, B., Beauchet, O., 2009. Vitamin D-related changes in physical performance: a systemic review. J. Nutr. Health Aging 13, 893–898.

Aoyama, S., Jia, H., Nakazawa, K., Yamamura, J., Saito, K., Kato, H., 2016. Dietary genistein prevents denervation-induced muscle atrophy in male rodents via effects on estrogen receptor-α. J. Nutr. 146, 1147–1154.

Asami, Y., Aizawa, M., Kinoshita, M., Ishikawa, J., Sakuma, K., 2018. Resveratrol attenuates denervation-induced muscle atrophy due to the blockade of atrogin-1 and p62 accumulation. Int. J. Med. Sci. 15, 628–637.

Baehr, L.M., Furlow, J.D., Bodine, S.C., 2011. Muscle sparing in muscle RING finger 1 null mice: response to synthetic glucocorticoids. J. Physiol. 589, 4759–4776.

Bang, H.S., Seo, D.Y., Chung, Y.M., Oh, K.M., Park, J.J., Arturo, F., Jeong, S.H., Kim, N., Han, J., 2014. Ursolic acid-induced elevation of serum irisin augments muscle strength during resistance training in men. Korean J. Physiol. Pharmacol. 18, 441–446.

Baumann, C.W., Rogers, R.G., Lees, S.J., Otis, J.S., 2014. Muscular strength is unaffected by shortterm resveratrol supplementation in aged mouse muscle. Int. J. Clin. Exp. Pathol. 1, 253–257.

Beaudart, C., Dawson, A., Shaw, S.C., Harvey, N.C., Kanis, J.A., Binkley, N., Reginster, J.Y., Chapurlat, R., Chan, D.C., Bruyére, O., Rizzoli, R., Cooper, C., Dennison, E.M., IOF-EXCEO Sarcopenia Working Group, 2017. Nutrition and physical activity in the prevention and treatment of sarcopenia: systematic review. Osteoporos. Int. 28, 1817–1833.

Beekmann, K., de Haan, L.H., Actis-Goretta, L., Houtman, R., van Bladeren, P.J., Rietjens, I.M., 2015. The effect of glucuronidation on isoflavone induced estrogen receptor (ER) α and ERβ mediated coregulator interactions. J. Steroid Biochem. Mol. Biol. 154, 245–253.
Bennett, B.T., Mohamed, J.S., Always, S.E., 2013. Effects of resveratrol on the recovery of muscle mass following disuse in the plantaris muscle of aged rats. PLoS One 8, e83518.
Bermon, S., Venembre, P., Sachet, C., Valour, S., Dolisi, C., 1998. Effects of creatine monohydrate ingestion in sedentary and weight-trained older adults. Acta Physiol. Scand. 164, 147–155.
Bischoff-Ferrari, H.A., Dietrich, T., Orav, E.J., Hu, F.B., Zhang, Y., Karlson, E.W., Dawson-Hughes, B., 2004. Higher 25-hydroxyvitamin D concentrations are associated with better lower-extremity function in both active and inactive persons aged > or =60y. Am. J. Clin. Nutr. 80, 752–758.
Bodine, S.C., Latres, E., Baumhueter, S., Lai, V.K., Nunez, L., Clarke, B.A., Poueymirou, W.T., Panaro, F.J., Na, E., Dharmarajan, K., Pan, Z.Q., Valenzuela, D.M., DeChiara, T.M., Stitt, T.N., Yancopoulos, G.D., Glass, D.J., 2001. Identification of ubiquitin ligases required for skeletal muscle atrophy. Science 294, 1704–1708.
Brose, A., Parise, G., Tarnopolsky, M.A., 2003. Creatine supplementation enhances isometric strength and body composition improvements following strength exercise training in older adults. J. Gerontol. A Biol. Sci. Med. Sci. 58, 11–19.
Carnio, S., LoVerso, F., Baraibar, M.A., Longa, E., Khan, M.M., Maffei, M., Reischl, M., Canepari, M., Loefler, S., Kern, H., Blaauw, B., Friguet, B., Bottinelli, R., Rudolf, R., Sandri, M., 2014. Autophagy impairment in muscle induces neuromuscular junction degeneration and precocious aging. Cell Rep. 8, 1509–1521.
Chang, Y.C., Chen, Y.T., Liu, H.W., Chan, Y.C., Liu, M.Y., Hu, S.H., Tseng, W.T., Wu, H.L., Wang, M.F., Chang, S.J., 2019. Oligonol alleviates sarcopenia by regulation of signaling pathways involved in protein turnover and mitochondrial quality. Mol. Nutr. Food Res. 63, 1801102.
Cho, Y.H., Lee, S.Y., Kim, C.M., Kim, N.D., Choe, S., Lee, C.H., Shin, J.H., 2016. Effect of loquat leaf extract on muscle strength, muscle mass, and muscle function in healthy adults: a randomized, double-blinded, and placebo-controlled trial. Evid. Based Complement. Alternat. Med. 2016, 4301621.
Church, D.D., Schwarz, N.A., Spillane, M.B., McKinley-Barnard, S.K., Andre, T.L., Ramirez, A.J., Willoughby, D.S., 2016. l-Leucine increases skeletal muscle IGF-1 but does not differentially increase Akt/mTORC1 signaling and serum IGF-1 compared to ursolic acid in response to resistance exercise in resistance-trained men. J. Am. Coll. Nutr. 35, 627–638.
Colman, R.J., Beasley, T.M., Kemnitz, J.W., Johnson, S.C., Weindruch, R., Anderson, R.M., 2014. Caloric restriction reduces age-related and all-cause mortality in rhesus monkeys. Nat. Commun. 5, 3557.
Cong, H., Sun, L., Liu, C., Tien, P., 2011. Inhibition of atrogin-1/MAFbx expression by adenovirus-delivered small hairpin RNAs attenuates muscle atrophy in fasting mice. Hum. Gene Ther. 22, 313–324.
Cruz-Jentoft, A.J., Baeyens, J.P., Bauer, J.M., Boirie, Y., Cederholm, T., Landi, F., Martin, F.C., Michel, J.P., Rolland, Y., Schneider, S.M., Topinková, E., Vandewoude, M., Zamboni, M., 2010. Sarcopenia: European consensus on definition and diagnosis: report of the European working group on sarcopenia in older people. Age Ageing 39, 412–423.
Custodero, C., Mankowski, R.T., Lee, S.A., Chen, Z., Wu, S., Manini, T.M., Hincapie Echeverri, J., Sabbà, C., Beavers, D.P., Cauley, J.A., Espeland, M.A., Fielding, R.A., Kritchevsky, S.B., Liu, C.K., McDermott, M.M., Miller, M.E., Tracy, R.P., Newman, A.B., Ambrosius, W.T., Pahor, M., Anton, S.D., 2018. Evidence-based nutritional and pharmacological interventions targeting chronic low-grade inflammation in middle-age and older adults: a systematic review and meta-analysis. Ageing Res. Rev. 46, 42–59.
Da Boit, M., Sibson, R., Sivasubramaniam, S., Meakin, J.R., Greig, C.A., Aspden, R.M., Thies, F., Jeromson, S., Hamilton, D.L., Speakman, J.R., Hambly, C., Mangoni, A.A., Preston, T., Gray, S.R., 2017. Sex differences in the effect of fish-oil supplementation on the adaptive response to resistance exercise training in older people: a randomized controlled trial. Am. J. Clin. Nutr. 105, 151–158.
Daussin, F.N., Heyman, E., Burelle, Y., 2021. Effects of (−)-epicatechin on mitochondria. Nutr. Rev. 79, 25–41.
Demontis, F., Perrimon, N., 2010. FOXO/4E-BP signaling in Drosophila muscles regulates organism-wide proteostasis during aging. Cell 143, 813–825.
Denison, H.J., Cooper, C., Sayer, A.A., Robinson, S.M., 2015. Prevention and optimal management of sarcopenia: a review of combined exercise and nutrition interventions to improve muscle outcomes in older people. Clin. Interv. Aging 10, 859–869.
Deutz, N.E.P., Pereira, S.L., Hays, N.P., Oliver, J.S., Edens, N.K., Evans, C.M., Wolfe, R.R., 2013. Effect of β-hydroxy-β-methylbutyrate (HMB) on lean body mass during 10 days of bed rest in older adults. Clin. Nutr. 32, 704–712.
Dickinson, J.M., Reidy, P.T., Gundermann, D.M., Borack, M.S., Walker, D.K., D'Lugos, A.C., Volpi, E., Rasmussen, B.B., 2017. The impact of postexercise essential amino acid ingestion on the ubiquitin proteasome and autophagosomal-lysosomal systems in skeletal muscle of older men. J. Appl. Physiol. 122, 620–630.
Dirks, A.J., Leeuwenburgh, C., 2004. Aging and lifelong calorie restriction result in adaptations of skeletal muscle apoptosis repressor, apoptosis-inducing factor, X-linked inhibitor of apoptosis, caspase-3, and caspase-12. Free Radic. Biol. Med. 36, 27–39.
Edström, E., Altun, M., Hägglund, M., Ulfhake, B., 2006. Atrogin-1/MAFbx and MuRF1 are downregulated in ageing-related loss of skeletal muscle. J. Gerontol. A Biol. Sci. Med. Sci. 61, 663–674.
Esmarck, B., Andersen, J.L., Olsen, S., Richter, E.A., Mizuno, M., Kjaer, M., 2005. Timing of postexercise protein intake is important for muscle hypertrophy with resistance training in elderly humans. J. Physiol. 535, 301–331.
Flakoll, P., Sharp, R., Baier, S., Levenhagen, D., Carr, C., Nissen, S., 2004. Effect of beta-hydroxyl-beta-methylbutyrate, arginine, and lysine supplementation on strength, functionality, body composition, and protein metabolism in elderly women. Nutrition 20, 445–451.
Gaugler, M., Brown, A., Merrell, E., DiSanto-Rose, M., Rathmacher, J.A., Reynolds IV, T.H., 2011. PKB signaling and atrogene expression in skeletal muscle of aged mice. J. Appl. Physiol. 111, 192–199.
Gerlinger-Romero, F., Guimaraes-Ferreira, L., Giannocco, G., Nunes, M.T., 2011. Chronic supplementation of beta-hydroxyl-beta-methylbutylate (HMB) increases the activity of the GH/IGF-I axis and induces hyperinsulinemia in rats. Growth Horm. IGF Res. 21, 57–62.
Godard, M.P., Williamson, D.L., Trappe, S.W., 2002. Oral amino-acid provision does not affect muscle strength or size gains in older men. Med. Sci. Sports Exerc. 34, 1126–1131.
Gray, S.R., Mittendorfer, B., 2018. Fish oil-derived n-3 polyunsaturated fatty acids for the prevention and treatment of sarcopenia. Curr. Opin. Clin. Nutr. Metab. Care 21, 104–109.
Groen, B.B.L., Horstman, A.M., Hamer, H.M., de Haan, M., van Kranenburg, J., Bierau, J., Poeze, M., Wodzig, W.K., Rasmussen, B.B., van Loon, L.J., 2015. Post-prandial protein handling: you are what you just ate. PLoS One 10, e0141582.
Gülcin, I., 2010. Antioxidant properties of resveratrol: a structure-activity insight. Innovat. Food Sci. Emerg. Technol. 11, 210–218.
Hao, Y., Jackson, J.R., Wang, Y., Edens, N., Pereira, S.L., Always, S.E., 2011. β-hydroxyl-β-methylbutyrate reduces myonuclear apoptosis during recovery from hind limb suspension-induced muscle fiber atrophy in aged rats. Am. J. Physiol. Regul. Integr. Comp. Physiol. 301, R701–R715.
He, X., Duan, Y., Yao, K., Li, F., Hou, Y., Wu, G., 2016. β-hydroxyl-β-methylbutyrate, mitochondrial biogenesis, and skeletal muscle health. Amino Acids 48, 653–664.
Hirasaka, K., Maeda, T., Ikeda, C., Haruna, M., Kohno, S., Abe, T., Ochi, A., Mukai, R., Oarada, M., Eshima-Kondo, S., Ohno, A., Okumura, Y., Terao, J., Nikawa, T., 2013. Isoflavones derived from soy beans prevent MuRF1-mediated muscle atrophy in C2C12 myotubes through SIRT1 activation. J. Nutr. Sci. Vitaminol. 59, 317–324.
Hirasaka, K., Maeda, T., Haruna, M., Abe, T., Ochi, A., Ohno-Maita, A., Teshima-Kondo, S., Taniyama, S., Tachibana, K., Nikawa, T., 2014. Effects of isoflavones derived from soy beans on muscle atrophy. Soy Protein Res. 17, 150–155.
Holecek, M., 2017. Beta-hydroxy-beta-methylbutyrate supplementation and skeletal muscle in healthy and muscle-wasting conditions. J. Cachexia Sarcopniea Muscle 8, 529–541.
Hsieh, L.C., Chow, C.J., Chang, W.C., Liu, T.H., Chang, C.K., 2010. Effect of beta-hydroxy-beta-methylbutyrate on protein metabolism in bed-ridden elderly receiving tube feeding. Asia Pac. J. Clin. Nutr. 19, 200–208.
Hwee, D.T., Baehr, L.M., Philp, A., Baar, K., Bodine, S.C., 2014. Maintenance of muscle mass and load-induced growth in Muscle RING Finger 1 null mice with age. Aging Cell 13, 92–101.
Katashima, C.K., Silva, V.R., Gomes, T.L., Pichard, C., Pimentel, G.D., 2017. Ursolic acid and mechanisms of actions on adipose and muscle tissue: a systematic review. Obes. Rev. 18, 700–711.

Kunkel, S.D., Suneja, M., Ebert, S.M., Bongers, K.S., Fox, D.K., Malmberg, S.E., Alipour, F., Shields, R.K., Adams, C.M., 2011. mRNA expression signatures of human skeletal muscle atrophy identify a natural compound that increases muscle mass. Cell Metab. 13, 627–638.
Kunkel, S.D., Elmore, C.J., Bongers, K.S., Ebert, S.M., Fox, D.K., Dyle, M.C., Bullard, S.A., Adams, C.M., 2012. Ursolic acid increases skeletal muscle and brown fat and decreases diet-induced obesity, glucose intolerance and fatty liver disease. PLoS One 7, e39332.
Kurrat, A., Blei, T., Kluxen, F.M., Mueller, D.R., Piechotta, M., Soukup, S.T., Kulling, S.E., Diel, P., 2015. Lifelong exposure to dietary isoflavones reduces risk of obesity in ovariectomized Wistar rats. Mol. Nutr. Food Res. 59, 2407–2418.
Lanza, I.R., Zabielski, P., Klaus, K.A., Morse, D.M., Heppelmann, C.J., Bergen III, H.R., Dasari, S., Walrand, S., Short, K.R., Johnson, M.L., Robinson, M.M., Schimke, J.M., Jakaitis, D.R., Asmann, Y.W., Sun, Z., Nair, K.S., 2012. Chronic caloric restriction preserves mitochondrial function in senescence without increasing mitochondrial biogenesis. Cell Metab. 16, 777–788.
Léger, B., Derave, W., De Bock, K., Hespel, P., Russell, A.P., 2008. Human sarcopenia reveals an increase in SOCS-3 and myostatin and a reduced efficiency of Akt phosphorylation. Rejuvenat. Res. 11, 163–175B.
Liu, H.W., Wei, C.C., Chen, Y.J., Chen, Y.A., Chang, S.J., 2016. Flavanol-rich lychee fruit extract alleviates diet-induced insulin resistance via suppressing mTOR/SREBP-1 mediated lipogenesis in liver and restoring insulin signaling in skeletal muscle. Mol. Nutr. Food Res. 60, 2288–2296.
Lobo, P.C.B., Vieira, I.P., Pichard, C., Marques, B.S., Gentil, P., da Silva, E.L., Pimentel, G.D., 2021. Ursolic acid has no additional effect on muscle strength and mass in active men undergoing a high-protein diet and resistance training: a double-blind and placebo-controlled trial. Clin. Nutr. 40, 581–589.
Luk, H.Y., Appell, C., Chyu, M.C., Chen, C.H., Wang, C.Y., Yang, R.S., Shen, C.L., 2020. Impacts of green tea on joint and skeletal muscle health: prospects of translational nutrition. Antioxidants 9, 1050.
Malafarina, V., Uriz-Otano, F., Iniesta, R., Gil-Guerrero, L., 2013. Effectiveness of nutritional supplementation on muscle mass in treatment of sarcopenia in old age: a systematic review. J. Am. Med. Dir. Assoc. 14, 10–17.
Mattison, J.A., Roth, G.S., Beasley, T.M., Tilmont, E.M., Handy, A.M., Herbert, R.L., Longo, D.L., Allison, D.B., Young, J.E., Bryant, M., Barnard, D., Ward, W.F., Qi, W., Ingram, D.K., de Cabo, R., 2012. Impact of caloric restriction on health and survival in rhesus monkeys from the NIA study. Nature 489, 318–322.
McKiernan, S.H., Colman, R.J., Lopez, M., Beasley, T.M., Aiken, J.M., Anderson, R.M., Weindruch, R., 2011. Caloric restriction delays aging-induced cellular phenotypes in rhesus monkey skeletal muscle. Exp. Gerontol. 46, 23–29.
McKiernan, S.H., Colman, R.J., Aiken, E., Evans, T.D., Beasley, T.M., Aiken, J.M., Weindruch, R., Anderson, R.M., 2012. Cellular adaptation contributes to calorie restriction-induced preservation of skeletal muscle in aged rhesus monkeys. Exp. Gerontol. 47, 229–236.
Muhammad, M.F., Allam, M.M., 2018. Resveratrol and/or exercise training counteract ageing-associated decline of physical endurance in aged mice; targeting mitochondrial biogenesis and function. J. Physiol. Sci. 68, 681–688.
Mukai, R., Nakao, R., Yamamoto, H., Nikawa, T., Takeda, E., Terao, J., 2010. Quercetin prevents unloading-derived disused muscle atrophy by attenuating the induction of ubiquitin ligases in tail-suspension mice. J. Nat. Prod. 73, 1708–1710.
Mukai, R., Horikawa, H., Fujikura, Y., Kawamura, T., Nemoto, H., Nikawa, T., Terao, J., 2012. Prevention of disuse muscle atrophy by dietary ingestion of 8-prenylnaringenin in denervated mice. PLoS One 7, e45048.
Nakae, Y., Hirasaka, K., Goto, J., Nikawa, T., Shono, M., Yoshida, M., Stoward, P.J., 2008. Subcutaneous injection, from birth, of epigallocatechin-3-gallate, a component of green tea, limits the onset of muscular dystrophy in mdx mice: a quantitative histological, immunohistochemical and electrophysiological study. Histochem. Cell Biol. 129, 489–501.
Neel, B.A., Lin, Y., Pessin, J.E., 2013. Skeletal muscle autophagy: a new metabolic regulator. Trends Endocrinol. Metab. 24, 635–643.
Nicastro, H., Artioli, G.G., Dos Santos Costa, A., Sollis, M.Y., Da Luz, C.R., Blachier, F., Lancha Jr., A.H., 2011. An overview of the therapeutic effects of leucine supplementation on skeletal muscle under atrophic conditions. Amino Acids 40, 287–300.
Ogawa, M., Kitano, T., Kawata, N., Sugihira, T., Kitakaze, T., Harada, N., Yamaji, R., 2017. Daidzein down-regulates ubiquitin-specific protease 19 expression through estrogen receptor β and increases skeletal muscle mass in young female mice. J. Nutr. Biochem. 49, 63–70.
Parise, G., Mihic, S., MacLennan, D., Yarasheski, K.E., Tarnopolsky, M.A., 2001. Effects of acute creatine monohydrate supplementation on leucine kinetics and mixed-muscle protein synthesis. J. Appl. Physiol. 91, 1041–1047.
Prokopidis, K., Giannos, P., Triantafyllidis, K.K., Kechagias, K.S., Mesinovic, J., Witard, O.C., Scott, D., 2022. Effect of vitamin D monotherapy on indices of sarcopenia in community-dwelling older adults: a systematic review and meta-analysis. J. Cachexia Sarcopenia Muscle. https://doi.org/10.1002/jcsm.12976 (in press).
Rawson, E.S., Wehnert, M.L., Clarkson, P.M., 1999. Effects of 30 days of creatine ingestion in older men. Eur. J. Appl. Physiol. Occup. Physiol. 80, 139–144.
Robinson, S.M., Jameson, K.A., Batelaan, S.F., Martin, H.J., Syddall, H.E., Dennison, E.M., Cooper, C., Sayer, A.A., Hertfordshire Cohort Study Group, 2008. Diet and its relationship with grip strength in community-dwelling older men and women: the Hertfordshire cohort study. J. Am. Geriatr. Soc. 56, 84–90.
Sakuma, K., Yamaguchi, A., 2010. Molecular mechanisms in aging and current strategies to counteract sarcopenia. Curr. Aging Sci. 3, 90–101.
Sakuma, K., Yamaguchi, A., 2018. Recent advances in pharmacological, hormonal, nutritional intervention for sarcopenia. Pflügers Arch. 470, 449–460.
Sakuma, K., Kinoshita, M., Ito, Y., Aizawa, M., Aoi, W., Yamaguchi, A., 2016. p62/SQSTM1 but not LC3 is accumulated in sarcopenic muscle of mice. J. Cachexia Sarcopenia Muscle 7, 204–212.
Sakuma, K., Aoi, W., Yamaguchi, A., 2017. Molecular mechanism of sarcopenia and cachexia: recent research advances. Pflügers Arch. 469, 573–591.
Sandri, M., Barberi, L., Bijlsma, A.Y., Blaauw, B., Dyar, K.A., Milan, G., Mammucari, C., Meskers, C.G., Pallafacchina, G., Paoli, A., Pion, D., Roceri, M., Romanello, V., Serrano, A.L., Toniolo, L., Larsson, L., Maier, A.B., Muñoz-Cánoves, P., Musarò, A., Pende, M., Reggiani, C., Rizzuto, R., Schiaffino, S., 2013. Signaling pathways regulating muscle mass in ageing skeletal muscle. The role of IGF-1-Akt-mTOR-FoxO pathway. Biogerontology 14, 303–323.
Senthil Kumaran, V., Arulmathi, K., Srividhya, R., Kalaiselvi, P., 2008. Repletion of antioxidant status by EGCG and retardation of oxidative damage induced macromolecular anomalies in aged rats. Exp. Gerontol. 43, 176–183.
Si, H., Wang, X., Zhang, L., Parnell, L.D., Admed, B., LeRoith, T., Ansah, T.A., Zhang, L., Li, J., Ordovás, J.M., Si, H., Liu, D., Lai, C.Q., 2019. Dietary epicatechin improves survival and delays skeletal muscle degeneration in aged mice. FASEB J. 33, 965–977.
Smith, G.I., Atherton, P., Reeds, D.N., Mohammed, B.S., Rankin, D., Rennie, M.J., Mittendorfer, B., 2011. Dietary omega-3 fatty acid supplementation increases the rate of muscle protein synthesis in older adults: a randomized controlled trial. Am. J. Clin. Nutr. 93, 402–412.
Smith, G.I., Julliand, S., Reeds, D.N., et al., 2015. Fish oil-derived n-3 PUFA therapy increases muscle mass and function in healthy older adults. Am. J. Clin. Nutr. 102, 115–122.
Snijder, M.B., Van Schoor, N.M., Pluijm, S.M., van Dam, R.M., Visser, M., Lips, P., 2006. Vitamin D status in relation to one-year risk of recurrent falling in older men and women. J. Clin. Endocrinol. Metab. 91, 2980–2985.
Tabata, S., Aizawa, M., Kinoshita, M., Ito, Y., Kawamura, Y., Takebe, M., Pan, W., Sakuma, K., 2018. The influence of isoflavone for denervation-induced muscle atrophy. Eur. J. Nutr. 58, 291–300.
Tan, A., Sullenbarger, B., Prakash, R., McDaniel, J.C., 2018. Supplementation with eicosapentaenoic acid and docosahexaenoic acid reduces high levels of circulating proinflammatory cytokines in aging adults: a randomized, controlled study. Prostaglandins Leukot. Essent. Fatty Acids 132, 23–29.
Thomas, D.K., Quinn, M.A., Saunders, D.H., Greig, C.A., 2016. Protein supplementation does not significantly augment the effects of resistance exercise training in older adults: a systematic review. J. Am. Med. Dir. Assoc. 17, 959.e1-959959.e9.
Tieland, M., van de Rest, O., Dirks, M.L., van der Zwaluw, N., Mensink, M., van Loon, L.J., Groot, L.C., 2012. Protein supplementation improves physical performance in frail elderly people: a randomized double-blind, placebo-controlled trial. J. Am. Med. Dir. Assoc. 13, 720–726.
Timmerman, K.L., Volpi, E., 2008. Amino acid metabolism and regulatory effets in aging. Curr. Opin. Clin. Nutr. Metab. Care 11, 45–49.
Valdez, G., Tapia, J.C., Kang, H., Clemenson Jr., G.D., Gage, F.H., Lichtman, J.W., Sanes, J.R., 2010. Attenuation of age-related changes in mouse neuromuscular synapses by caloric restriction and exercise. Proc. Natl. Acad. Sci. U. S. A. 107, 14863–14868.

Vannice, G., Rasmussen, H., 2014. Position of the academy of nutrition and dietetics: dietary fatty acids for healthy adults. J. Acad. Nutr. Diet. 114, 136–153.
von Haehling, S., Morley, J.E., Anker, S.D., 2010. An overview of sarcopenia: facts and numbers on prevalence and clinical impact. J. Cachexia Sarcopenia Muscle 1, 129–133.
Wakabayashi, H., Sakuma, K., 2014. Comprehensive approach to sarcopenia treatment. Curr. Clin. Pharmacol. 9, 171–180.
Walker, D.K., Dickinson, J.M., Timmerman, K.L., Drummond, M.J., Reidy, P.T., Fry, C.S., Gundermann, D.M., Rasmussen, B.B., 2011. Exercise, amino acids, and aging in the control of human muscle protein synthesis. Med. Sci. Sports Exerc. 43, 2249–2258.
Wall, B.T., Dirks, M.L., Snijders, T., Stephens, F.B., Senden, J.M., Verscheijden, M.L., van Loon, L.J., 2015. Short-term muscle disuse atrophy is not associated with increased intramuscular lipid deposition or a decline in the maximal activity of key mitochondrial enzymes in young and older males. Exp. Gerontol. 61, 76–83.
Wang, Z.H., Hsu, C.C., Huang, C.N., Yin, M.C., 2009. Anti-glycative effects of oleanolic acid and ursolic acid in kidney of diabetic mice. Eur. J. Pharmacol. 628, 255–260.
Wang, J., Cui, C., Chim, Y.N., Yao, H., Shi, L., Xu, J., Wang, J., Wong, R.M.Y., Leung, K.S., Chow, S.K., Cheung, W.H., 2020. Vibration and beta-hydroxy-beta-methylbutyrate treatment suppresses intramuscular fat infiltration and adipogenic differentiation in sarcopenic mice. J. Cachexia Sarcopenia Muscle 11, 564–577.
Wenz, T., Rossi, S.G., Rotundo, R.L., Spiegelman, B.M., Moraes, C.T., 2009. Increased muscle PGC-1alpha expression protects from sarcopenia and metabolic disease during aging. Proc. Natl. Acad. Sci. U. S. A. 106, 20405–20410.
Wicherts, I.S., van Schoor, N.M., Boeke, A.J., Visser, M., Deeg, D.J., Smit, J., Knol, D.L., Lips, P., 2007. Vitamin D status predicts physical performance and its decline in older persons. J. Clin. Endocrinol. Metab. 92, 2058–2065.
Wilkinson, D.J., Hossain, T., Hill, D.S., Phillips, B.E., Crossland, H., Williams, J., Loughna, P., Churchward-Venne, T.A., Breen, L., Phillips, S.M., Etheridge, T., Rathmacher, J.A., Smith, K., Szewczyk, N.J., Atherton, P.J., 2013. Effects of leucine and its metabolite beta-hydroxyl-beta-methylbutylate on human skeletal muscle protein metabolism. J. Physiol. 591, 2911–2923.
Wohlgemuth, S.E., Seo, A.Y., Marzetti, E., Lees, H.A., Leeuwenburgh, C., 2010. Skeletal muscle autophagy and apoptosis during aging: effects of calorie restriction and life-long exercise. Exp. Gerontol. 45, 138–148.
Wolfe, R.R., Miller, S.L., Miller, K.B., 2008. Optimal protein intake in the elderly. Clin. Nutr. 7, 675–684.
Wu, H., Xia, Y., Jiang, J., Du, H., Guo, X., Liu, X., Li, C., Huang, G., Niu, K., 2015. Effect of beta-hydroxyl-beta-methylbutyrate supplementation on muscle loss in older adults: a systematic review and meta-analysis. Arch. Gerontol. Geriatr. 61, 168–175.
Wu, S., Ning, H.T., Xiao, S.M., Hu, M.Y., Wu, X.Y., Deng, H.W., Feng, H., 2020. Effects of vibration therapy on muscle mass, muscle strength and physical function in older adults with sarcopenia: a systematic review and meta-analysis. Eur. J. Aging Phys. Act. 17, 14.
Yu, R., Chen, J.A., Xu, J., Cao, J., Wang, Y., Thomas, S.S., Hu, Z., 2017. Suppression of muscle wasting by the plant-derived compound ursolic acid in a model of chronic kidney disease. J. Cachexia Sarcopenia Muscle 8, 327–341.
Zhou, J., Liao, Z., Jia, J., Chen, J.L., Xiao, Q., 2019. The effects of resveratrol feeding and exercise training on the skeletal muscle function and transcriptome of aged rats. PeerJ 7, e7199.
Zhu, Y.Q., Peng, N., Zhou, M., Liu, P.P., Qi, X.L., Wang, N., Wang, G., Wu, Z.P., 2019. Tai Chi and whole-body vibrating therapy in sarcopenic men in advanced old age: a clinical randomized controlled trial. Eur. J. Ageing 16, 273–282.

Starvation and fasting: Biochemical aspects

Kathleen H. McGrath[a,b], **Wolfram Haller**[c], **and Julie E. Bines**[a,b,d], [a] Department of Gastroenterology and Clinical Nutrition, The Royal Children's Hospital, Melbourne, VIC, Australia; [b] Department of Pediatrics, The University of Melbourne, Melbourne, VIC, Australia; [c] Department of Gastroenterology, Birmingham Women's and Children's Hospital, Birmingham, United Kingdom; and [d] Murdoch Children's Research Institute, Melbourne, VIC, Australia

© 2023 Elsevier Ltd. All rights reserved.

This is an update of W. Haller, J.E. Bines, Starvation and fasting: Biochemical Aspects, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 209–218, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00254-3.

Introduction **646**
The feeding/fasting cycle **646**
Energy requirements and metabolism 646
Energy production 646
Carbohydrate metabolism 647
Protein metabolism 649
Fat metabolism 650
Metabolic consequences of fasting and starvation **650**
Overview 650
Metabolic adaptation in the post absorptive state **651**
Metabolic adaptation in prolonged fasting 651
Metabolic adaptation in starvation 652
Regulation of adaptation in starvation 653
Neural and endocrine regulation 653
Nutrient sensing mechanisms 654
Conclusion **655**
References **656**
Further reading **656**

Key points

- Metabolic homeostasis relies on a complex and integrated system to adapt to changes in nutrient provision and utilization.
- Most cells use glucose for ATP synthesis and energy supply however other fuels (e.g., fatty acids, amino acids and ketone bodies) have an important energy supply role in metabolic adaptation in response to prolonged fasting or starvation.
- Thyroid hormones are known as potent hypothalamic regulators of whole-body energy homeostasis.
- AMP-activated protein kinase (AMPK) and the evolutionary-conserved NAD^+-dependent histone deacetylase SIRT1 axis integrates multiple hormonal and nutritional signals (e.g., glucagon, leptin, adiponectin, glycogen, and free fatty acids) to regulate cellular and whole body energy homeostasis during starvation

Glossary

Adapted starvation Relates to the state of steady fat store depletion and minimal proteolysis beyond three weeks of fast during which ketone bodies and fatty acids are the main energy substrates

AMPK The AMP-activated protein kinase is an important sensor of nutrient supply to cells and intracellular energy status. Under conditions of low intracellular energy charge, AMP and ATP are bound in a competitive manner and activation of AMPK leads to downregulation of anabolic and upregulation of catabolic enzymatic pathways with the aim to increase intracellular ATP availability

Cori cycle The metabolic pathway in which lactate—generated through anaerobic glycolysis in the muscle—is used for hepatic glucose synthesis, which then is transported back to the muscle as an energy substrate

FOXO Forkhead box O is a family of transcription factors. FOXOs are elemental regulators of metabolism, upregulate key enzymes of hepatic gluconeogenesis during early starvation, interfere with adipocyte differentiation and increase protein catabolism

Intermediate phase This relates to the necessary transcriptional transition to ketogenesis and ketolysis as well as fatty acid oxidation during two to three weeks after initiation of fasting

Kwashiorkor Kwashiorkor is one of the major clinical syndromes of severe childhood malnutrition (SCM, see also *Marasmus*). While wasting is characteristic for all syndromes of SCM, Kwashiorkor has additional clinical features of protein energy malnutrition such as hypoalbuminemia, generalized edema, skin changes (dermatitis, hypopigmentation). The underlying etiology in comparison to marasmus is still unclear, additional stressors such as infections and environmental toxins might be involved. The overall mortality is high

Marasmus Marasmus is one of the major clinical syndromes of severe childhood malnutrition (SCM, see also *kwashiorkor*) characterized by non-edematous slow weight loss and wasting of muscle protein compartment through months and years. The serum albumin is normal and associated overall mortality is low following reintroductions of adequate nutrition

mTOR Mammalian target of rapamycin is a serine/threonine protein kinase with pleiotropic intracellular anabolic effects and functions. mTOR is involved in regulation of protein synthesis, cell proliferation, cell survival and inflammatory reactions. Upstream activators are insulin and insulin-like growth factors (IGF-1), inhibition is mediated through the *AMPK-FOXO* pathway

PGC1α PPARγ coactivator-1α is a transcriptional coactivator involved in mitochondrial metabolism. It is accepted as the central node in starvation-induced transcriptional co-activation and regulates large clusters of genes involved in oxidative phosphorylation and fatty acid metabolism

PPAR Peroxisome proliferator-activated receptor is a family of ligand-activated transcription factors involving three main subclasses, PPARα, β and γ. PPARα acts as an intracellular fatty acid sensor and is one of the master regulators of cellular fatty acid oxidation and ketone body production

SIRT1 An evolutionary-conserved NAD^+-dependent histone deacetylase which is an important element of metabolic homeostasis and energy conservation during fed-fasted-refed transitions

Introduction

Human metabolism and body function rely on the essential provision and effective utilization of energy at a cellular level. However, on a daily basis, there are numerous fluctuations in nutrient intake and changes in metabolic demand such as variations in physical activity and environmental factors. A highly integrated system allows the human body to adapt to these changes in daily life and maintain a state of homeostasis.

Critical illness, chronic disease and other environmental factors may lead to periods of prolonged fasting (total absence of nutrient intake) and starvation (prolonged period of inadequate food intake) that challenge homeostatic mechanisms and result in depletion of preexisting fuel stores. In response, a complex interplay of regulatory factors and systems is required to stimulate a sequence of further changes, called metabolic adaptation and provide ongoing energy supply to the body's organs.

This article provides a review of normal metabolic processes and responses in the feeding-fasting cycle and the biochemical adaptations that occur in response to prolonged fasting and starvation in humans.

The feeding/fasting cycle

Energy requirements and metabolism

Energy is essential for many important body functions, including the maintenance of cellular integrity and function, new tissue synthesis and growth, thermoregulation and adaption to physical activity and other metabolic stressors. The energy requirements of an individual vary with age, sex, body composition, physical activity, and stress. In the normal adult at basal state, approximately 75% of energy requirements reflect the energy needs of major organs (brain ~20%, skeletal muscle 18–22%, abdominal organs ~25%, and heart ~11%). In children, up to 50% of resting metabolic energy is expended by the brain. During normal daily activity, the total energy requirement and the proportion of energy needed by different tissues may vary considerably.

Energy production

Adequate availability of metabolic fuels for energy production has to be maintained in both the fed and the fasting states. The body derives energy from combustion of carbohydrate, fat, and protein provided exogenously in the fed state with the surplus being stored as glycogen in liver and muscle and as triacylglycerol in adipose tissue. These stores help to endogenously buffer periods of low-energy intake such as the post absorptive and fasting state. A mixture of metabolic fuels including glucose, triacylglycerols, ketone bodies, non-esterified fatty acids, and amino acids is present in the circulation. The proportion of these energy substrates in the blood at any one time depends on the fed or fasting state of the individual, the extent of fuel stores, and recent or current metabolic demand and stressors. Many of these are interconvertible and thus allow easier adaptation to situations of changing exogenous energy supply. In a normal, non-obese 70 kg adult, there are approximately 500 MJ (120,000 kcal) contained in adipose tissue, 100 MJ (24,000 kcal) stored in muscle and visceral proteins, and 8 MJ (2000 kcal) stored as liver and muscle glycogen. During

a normal day, half of the total energy requirement is met by carbohydrate metabolism. At this rate, glycogen stores would be exhausted after 1–2 days of fasting. However, glycogen stores are maintained for a longer period owing to the production of glucose from gluconeogenesis (Fig. 1). Gluconeogenic substrates such as lactate, amino acids (alanine, glutamine), glycerol, and β-hydroxybutyrate are shuttled from muscle and adipose tissue toward liver and kidney where they act as core substrates for gluconeogenesis. During prolonged fasting, the liver will account for approximately 80% of total body glucose production, whereas the remaining 20% is made by the kidney.

Carbohydrate metabolism

Glucose plays a key role in body metabolism. It is the preferred source of energy for the majority of tissues and is particularly important for the retina and the brain where it is the main energy substrate. Owing to a lack of mitochondria, red blood cells and the renal medulla are completely reliant on glucose as the sole metabolic fuel at all times. Therefore, the metabolic changes that occur during fasting and starvation target the necessity to spare blood glucose and liver glycogen for use by brain and other glucose-dependent tissues by providing alternative fuels to other tissues (Table 1).

Glucose enters the bloodstream through absorption from the intestine, the breakdown of liver glycogen, and hepatic gluconeogenesis from glucogenic precursors (e.g., lactate, alanine). To produce energy from glucose, three metabolic pathways are involved (Fig. 2). Glucose is first oxidized to form pyruvate via the glycolytic pathway. Pyruvate then enters the Krebs cycle and is completely oxidized to form NADH + H, FADH2, and carbon dioxide. The NADH + H transports hydrogen to the respiratory chain where it is used to reduce oxygen to water through oxidative phosphorylation. The net yield of energy from the metabolism of 1 molecule of glucose is 38 molecules of ATP. As the combustion of glucose within the human body is coupled to the energy-consuming synthesis of ATP, the overall efficiency of energy extraction from glucose within the cell is only approximately 50%.

Of all energy substrates, blood glucose concentration is the most constant. Because of the strict glucose requirements of the brain, the circulating blood glucose pool is tightly controlled at approximately 16 g or 5 mmol L^{-1}. Three important mechanisms are responsible for this regulation.

1. Insulin, as an anabolic hormone, enhances glucose uptake into muscle and fat and stimulates glycogen synthesis. It also inhibits lipolysis, glycogenolysis, and gluconeogenesis. High insulin levels will decrease blood glucose levels. Conversely, low insulin levels will cause a rise in blood glucose by decreased inhibition of glycogenolysis and reduced peripheral uptake of glucose.
2. Glucagon, a 29-amino acid peptide released from pancreatic α-cells, counteracts the insulin effect and increases liver glycogen breakdown, gluconeogenesis, and ketogenesis from fatty acids. It also stimulates lipolysis from adipocytes in extrahepatic tissue.

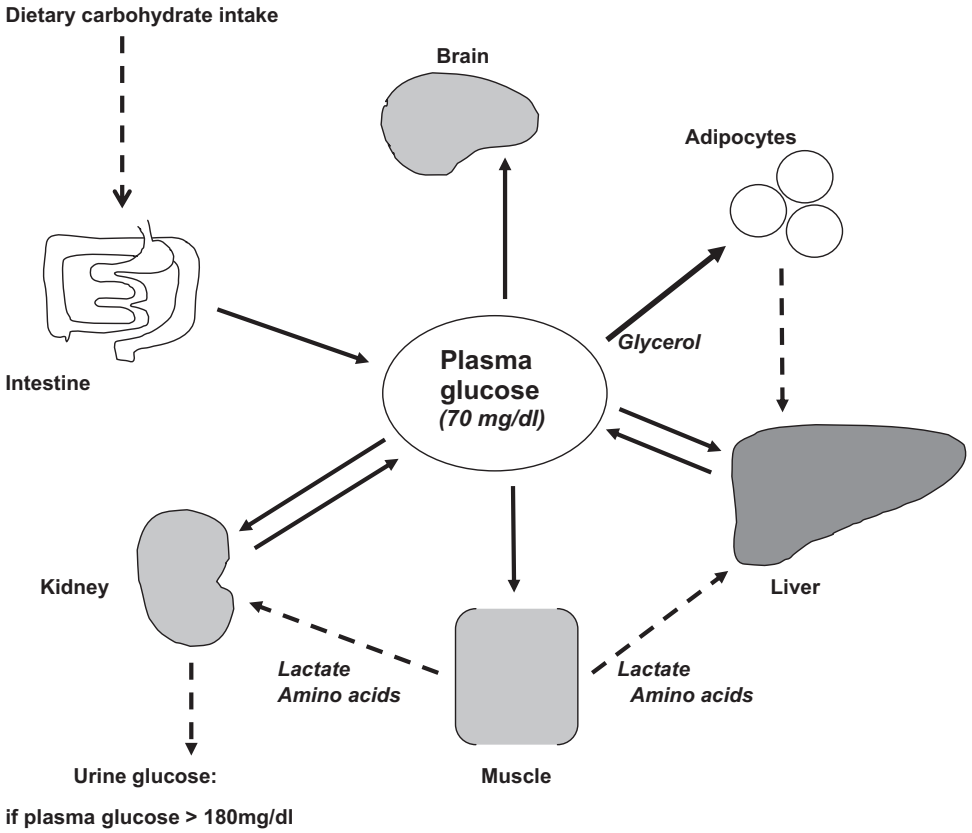


Fig. 1 The metabolism of energy substrates to maintain glucose homeostasis.

Table 1 Energy sources of different parts of the human body.

Organ	*Preferred energy sources (well-fed state)*	*Alternative energy sources (starvation state)*
Liver	Glucose, fatty acids	Fatty acids
Skeletal muscle	Glucose, fatty acids (resting muscle)	Fatty acids, ketone bodies
Cardiac muscle	Fatty acids	Fatty acids, amino acids and ketone bodies
Adipose tissue	Glucose	Fatty acids
Kidney	Glucose, fatty acids	Fatty acids
Brain[a]	Glucose	Glucose, later ketone bodies
Red blood cells[a]	Glucose	Glucose
Outer retina and lens[a]	Glucose	Glucose

[a]Glucose-dependent tissues.

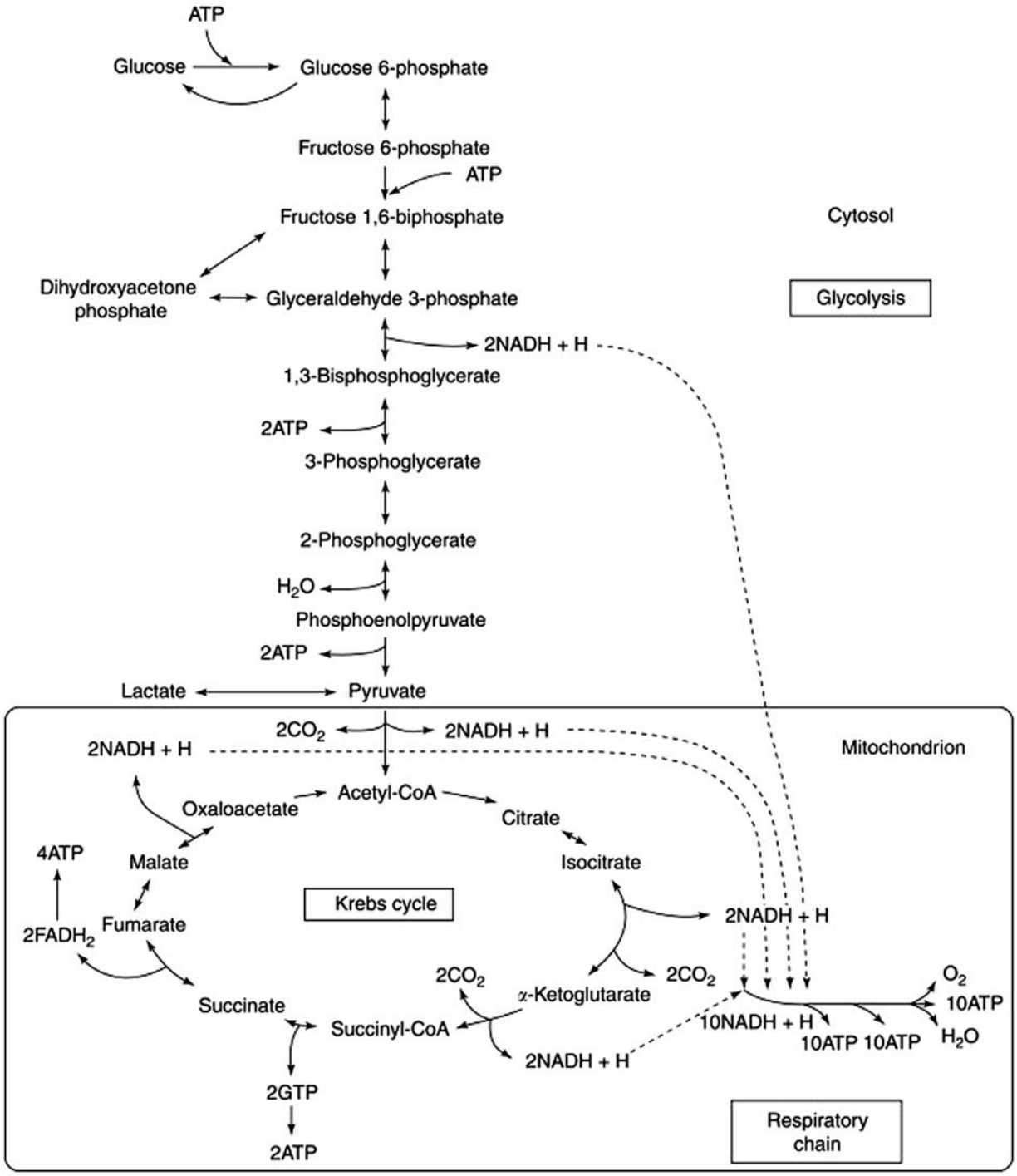


Fig. 2 The production of energy from glucose via the glycolytic pathway, the Krebs cycle, and the respiratory chain.

The net result of glucagon activity is an increase in blood glucose concentration, which helps to maintain blood glucose levels despite the effect of insulin.

3. Neuroendocrine counter regulatory responses to glucose deprivation in the brain with sympathoadrenal upregulation and release of cortisol, catecholamines, and growth hormone act to rapidly normalize blood glucose by increasing hepatic gluconeogenesis and glycogenolysis and curtailing peripheral tissue glucose uptake and utilization.

The fed state is characterized by increased blood concentrations of glucose, amino acids, and fat. Insulin secretion is stimulated while glucagon levels remain unchanged or are decreased resulting in increased glucose uptake into tissues and enhanced glycogen, protein, and triacylglycerol syntheses. Glucagon balances this effect by stimulating glycogen breakdown to maintain blood glucose levels. By this mechanism, blood glucose levels are regulated during periods of surplus carbohydrate ingestion and excess glucose is stored as glycogen or fat.

Glycogen is a complex, hydrated gel-like polymer of glucose arranged in a highly branched spherical form. It allows glucose to be stored in large amounts and a relatively small volume without causing osmotic shifts and acts as a rapidly available short-term energy buffer in the post absorptive and fasting state. The terminal glucose molecules within this branching structure are accessible to the enzymes mediating glycogen breakdown to allow the rapid release of glucose in times of stress. The glycogen molecule expands in size after a carbohydrate-rich meal to approximately 40 nm in diameter and shrinks to 10 nm in diameter or less between meals. An adult man receiving a normal carbohydrate-containing diet has approximately 70 g of liver glycogen and 200 g of muscle glycogen. The final enzymatic step required to complete glycogen breakdown to glucose, glucose-6-phosphatase, is present only in the liver. Muscle glycogen is metabolized by anaerobic glycolysis to form pyruvate and lactate. Lactate is then transported to the liver where it acts as a precursor for gluconeogenesis ("Cori cycle") (Fig. 3). The Cori cycle contributes to approximately 40% of the normal plasma glucose turnover. It has the advantage of providing energy (net 3 molecules of ATP) without the loss of glucose molecules. The energy required for the resynthesis of glucose in the liver is derived from fatty acid oxidation. The total body glycogen stores can meet the energy needs of the brain for approximately 3 days, after which, alternative sources of metabolic fuel must be found.

Protein metabolism

Body nitrogen resides in two main compartments. Approximately half of the body's nitrogen is contained in extracellular tissues such as collagen. The nitrogen present within these tissues is relatively fixed and does not change significantly with starvation. The nitrogen turnover within this compartment can be assessed by the measurement of hydroxyproline excretion. The remaining nitrogen is present in the lean muscle mass, comprising skeletal and visceral muscles. The proteins within these tissues are constantly being broken down and resynthesized at a rate of between 3 and 3.5 g kg^{-1} day^{-1} in a young adult. Measurement of urinary 3-methylhistidine and creatinine excretion can be used to estimate the fractional catabolic rate of skeletal muscle. Similar to fatty acids and glucose, amino acids can be completely oxidized. However, the body's protein compartments—contrary to liver glycogen or adipose tissue—are not primarily energy stores. Although amino acids serve as important energy substrates in gluconeogenesis and Krebs cycle oxidation during times of inadequate energy intake, the loss of functional body protein is the life-limiting factor in prolonged fasting, with associated clinical sequelae.

In the fed state, amino acids digested and absorbed in excess of the body's immediate requirements for incorporation into proteins or other molecules are either oxidized for energy or metabolized to glycogen or fat. Protein provides approximately 17 kJ g^{-1} (4 kcal g^{-1}) of energy when metabolized as an energy source.

Prolonged fasting results in depletion of liver and muscle glycogen stores. In this clinical setting, the conversion of amino acids to glucose in liver and kidney contributes to the glucose requirements of the brain. The transition to metabolism of amino acids as

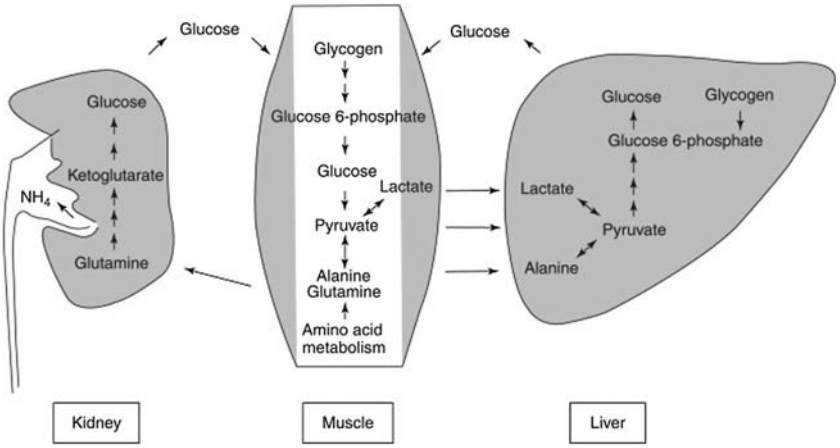


Fig. 3 The metabolism of muscle glycogen and protein to form glucose, involving the Cori cycle (lactic acid cycle) and the Cahill cycle (glucose–alanine cycle).

an energy source is mediated by an alteration in the balance of insulin and glucagon. The breakdown of tissue protein to provide glucose results in a sustained loss of body nitrogen of approximately 12 g day^{-1}. Experimentally, this loss of body nitrogen can be prevented by the administration of glucose. As a result of muscle protein breakdown, amino acids—predominantly alanine and glutamine—are released into the circulation. However, the amount of alanine released exceeds the alanine content of the muscle protein. This is because approximately one-third of the alanine released from muscle originates directly from the muscle protein, whereas the remaining two-thirds is derived from pyruvate. Pyruvate is formed by the metabolism of muscle glycogen or by the transamination of other amino acids contained within the muscle protein. Alanine is then transported to the liver where it is rapidly taken up and converted to glucose: this is known as the Cahill or glucose–alanine cycle (Fig. 3). Despite the increased release of alanine from muscle, plasma alanine levels fall in early fasting. This results from the rapid uptake and conversion of alanine by the liver to glucose. Similar to the glucose–alanine cycle between muscle and liver, there is a steady exchange of glutamine and glucose between muscle and kidney: glutamine released from muscle tissue acts as one of the main renal gluconeogenic precursors besides lactate in prolonged fasting.

Two major metabolic pathways accomplish regulated protein catabolism in prolonged fasting: the ubiquitin–proteasome system (UPS) and the autophagy–lysosomal system. The calcium-dependent calpain–calpastatin system and the caspase–proteolytic system is less well described but likely also play a role. The UPS acts through a cascade of enzymes that conjugate polyubiquitin chains with the target protein that are thereby marked for further degradation by the so-called "proteasome." The ubiquitinated proteins are bound by two 19S-regulatory proteins, unfolded and transported to the proteolytic centers of the 20S proteosomal core. The UPS serves as the primary route for degradation of short-lived proteins with high selectivity, allowing fine-tuning of the steady state of many regulatory and rate-limiting enzymes and degrading defective proteins. UPS-mediated catabolism is an important adaptation mechanism during acute starvation to maintain intracellular amino acid pools as a substrate source for gluconeogenesis. The enzyme complex is activated by a number of metabolic changes seen in fasting and starvation such as the decrease in insulin and other anabolic hormones, an increased concentration of intracellular reactive oxygen species, a decrease in intracellular amino acid concentration, and a reduction in AMP/ATP ratio.

Autophagy describes a catabolic process in which cell constituents (organelles, proteins) are engulfed by a phagophore and delivered to the lysosomal compartment for further degradation. It is primarily responsible for the degradation of long-lived proteins and those too large for other surveillance pathways and proceeds at a basal rate in nearly all eukaryotic cells in a "housekeeping role". Autophagy also plays a significant role in response to conditions of chronic nutrient starvation by inactivation of mTOR in response to metabolic changes (e.g., increased energetic or oxidative intracellular stress or a decrease in intracytoplasmic amino acids) leading to generation of amino acids able to be recycled for essential cellular functions. The UPS and autophagy systems are closely interrelated, share activating trigger factors and intracellular signaling pathways, as well as protein substrates, and can compensate for the compromise of the other system.

Fat metabolism

Fat is an efficient store of energy providing approximately 38 kJ g^{-1} (9 kcal g^{-1}). Fat is predominantly stored as triacylglycerols within adipocytes. Contrary to carbohydrate metabolism with blood glucose concentrations being tightly regulated, the concentrations of triacylglycerols and nonesterified fatty acids are much more variable between individuals related to variation in fat stores and gender. In the fed state, insulin stimulates triacylglycerol synthesis by allowing acyl-CoA from excess carbohydrates, protein, or fat being stored as adipose tissue. During fasting, triacylglycerol is converted to fatty acids and glycerol (Fig. 4). Within days, glycerol and palmitate release increases by two to three times fed levels, regulated by hormone-sensitive lipase. Owing to the absence of glycerol kinase in white adipose tissue, glycerol cannot be completely metabolized within the adipocytes and is transported to the liver where it is converted into glucose by gluconeogenesis. Fatty acids released from adipocytes are either delivered to the liver to be oxidized or re-esterified with glycerol-3-phosphate to re-enter the cycle to form triacylglycerol (Fig. 4). The energy cost of re-esterification of fatty acids in starvation may account for 2–3% of the resting energy expenditure.

Fatty acid oxidation is stimulated by activation of carnitine/acyl carnitine translocase acyltransferase, which influences the transport of long-chain fatty acids into the mitochondria. Most of the acetyl-CoA produced from fatty acid oxidation is metabolized to acetoacetate, which, in turn, may be converted to β-hydroxybutyrate and acetone. These products are known as "ketone bodies" and are water-soluble intermediates. In the fed state, ketone bodies are only produced in small quantities, are generally metabolized by the liver, and are not released into the circulation. During fasting, the rate of production of acetoacetate and β-hydroxybutyrate significantly increases to allow them to be used by the brain and other tissues as an alternative energy source.

Metabolic consequences of fasting and starvation

Overview

The transition from the post absorptive state and the early phases of adaptation to the steady state of adapted starvation can take up to 3 weeks (Table 2).

The length of survival (Fig. 5) will finally be determined by the total endogenous caloric reserve (calories per kilogram of bodyweight) that is the magnitude of preexisting stores of fat and protein, as well as the rate at which the endogenous substrates are utilized in the fasting state defined by the total energy expenditure (calories per kilogram of bodyweight per day). Owing to a different body composition with a much higher percentage of total body water and smaller subcutaneous fat stores, the neonatal

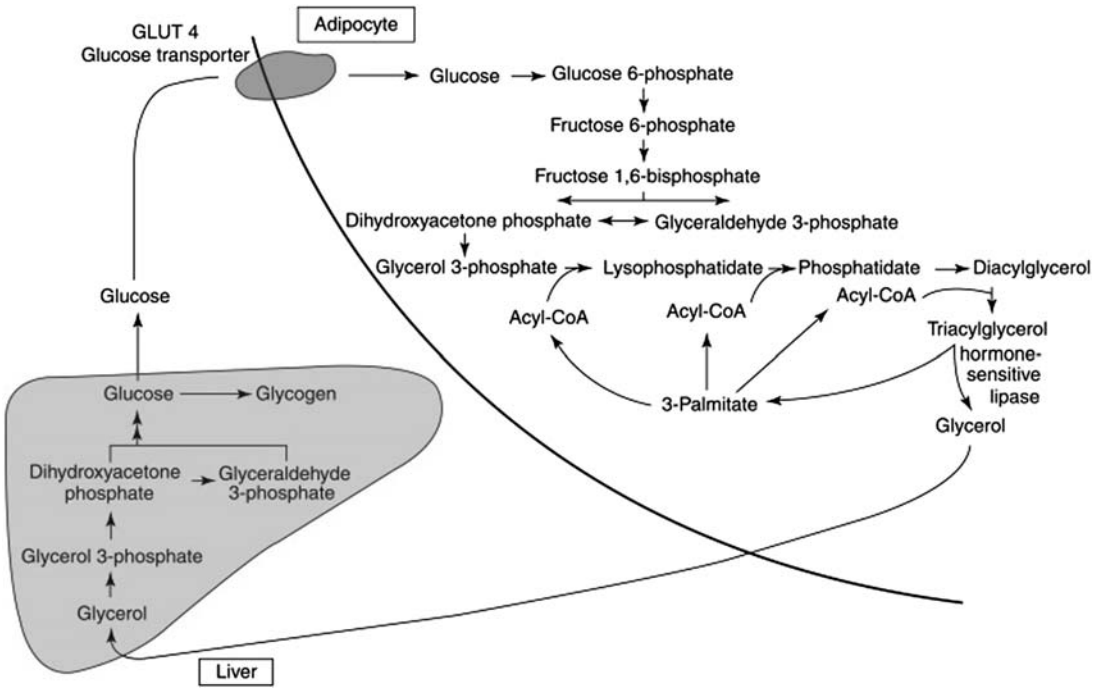


Fig. 4 The triacylglycerol/fatty acid cycle.

Table 2 Time course of the adaptive metabolic response to starvation.

Metabolic adaptation	*Duration*	*Pathway*	*Effect*
Post absorptive state	12 h	Glycogenolysis	Carbohydrate mobilization
Early starvation	2–3 days	Gluconeogenesis	Glycogenic AA mobilization
Intermediate phase	2–3 weeks	Ketogenesis/fatty acid oxidation	Transcriptional transition to fatty acid metabolism
Adapted starvation	>3 weeks	Ketogenesis/fatty acid oxidation	Steady fat store depletion economized proteolysis

overall caloric reserve is significantly lower compared with that of the adult. The neonate is further disadvantaged in its ability to tolerate starvation by its higher energy expenditure that is negatively related to the gestational age. Thermal stress, as a consequence of a 2.5–3 times higher surface area to bodyweight ratio in comparison to an adult, and the limited insulating capacity from subcutaneous fat are important determinants of increased neonatal energy expenditure and thus susceptibility to starvation.

Metabolic adaptation in the post absorptive state

The post absorptive state commences when the last nutrient is absorbed from the previous meal and continues until the next meal or for approximately 12 h during a normal overnight fast. Decrease in the insulin/glycogen ratio prompts metabolic transition from exogenous energy consumption to reliance on endogenous energy sources/breakdown of liver glycogen. It is, therefore, called the glycogenolytic phase of starvation. The hepatic release of approximately 200–250 g of glucose per day or 8–10 g h^{-1} balances the rate of glucose utilization of the brain and other tissues. A minor part of glucose formation derives from non-carbohydrate sources including glycerol (from triacylglycerols) and pyruvate and lactate (from muscle).

Metabolic adaptation in prolonged fasting

Fasting beyond 12 h will lead to the gluconeogenic phase of starvation represented by the transition from glycogen to metabolism of glucogenic amino acids as the main source of energy. This is mediated by a further decrease in the insulin/glucagon ratio. As a result, blood levels of the branched-chain amino acids, alanine, and glutamine double after 3–5 days of fasting. The glucose–alanine cycle provides glucose to the muscle in exchange for alanine provided to the liver as a precursor for gluconeogenesis (see Fig. 3). The intestine preferentially takes up glutamine released from the muscle during fasting where it is used as an energy source, and by the kidney where it is also used for renal ammonia production. Although the metabolism of amino acids to glucose is a very important step of metabolic adaptation to fasting, it only provides approximately 45 g glucose per day. This amount alone

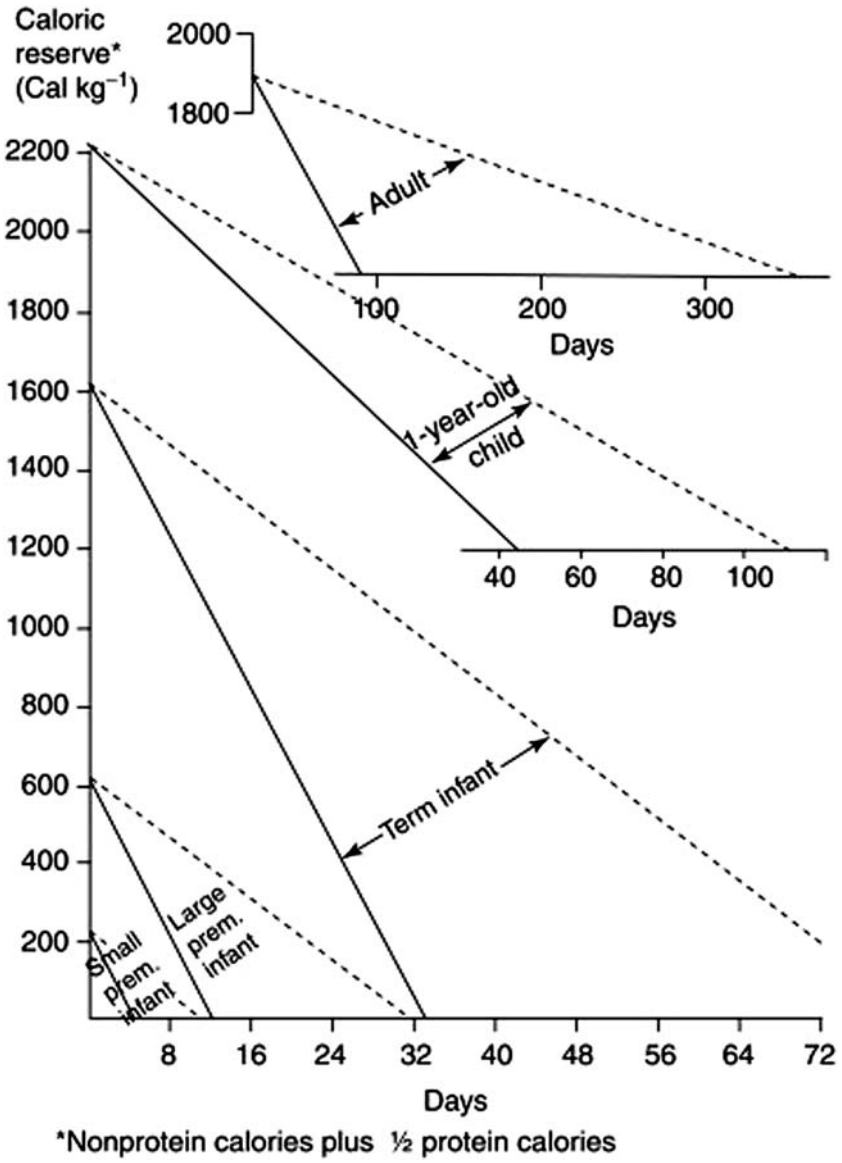


Fig. 5 Duration of survival in different age groups expected in starvation (solid line) and semi starvation (dashed line). Reproduced from Heird et al. (1972).

is insufficient to meet the glucose requirements of the brain and must be supplemented by energy produced from fat metabolism. Gluconeogenesis occurs at the expense of the functional protein compartment and provides energy substrates until the lipolytic and ketogenic machinery has fully adapted. Increased efficiency of the adaptive metabolic switch to fat and ketone body utilization is reflected by a fall in plasma amino acids if fasting is prolonged.

Metabolic adaptation in starvation

During periods of prolonged inadequate nutrient intake, the human body needs to mobilize preexisting stores of nutrients for continued functioning. This is done through down-regulation of anabolic energy consuming processes (fatty acid/triglyceride synthesis, glycogenesis) and activation of catabolic reactions (glycolysis, fatty acid oxidation, and proteolysis). The process of metabolic adaptation to starvation involves three main principles:

1. Replacement of glucose by fat as the main source of energy.
2. Promotion of ketone bodies as a fuel alternative in glycolytic tissues.
3. Maintenance of minimum of glucose supply by gluconeogenesis from glucogenic amino acids and glycerol, and by the shutdown of irreversible glucose oxidation in the citric acid cycle.

The mobilization of triacylglycerol stores to provide energy during starvation is regulated by a number of factors as outlined in the section below. Non-esterified fatty acids are released into the circulation and free fatty acid concentrations increase from 0.5–0.8 to 1.2–1.6 mmol L^{-1} within the first few days of fasting. Fatty acids circulate bound to albumin and can be oxidized in the liver or other tissues to produce energy. The switch to using ketone bodies as an energy source by the brain appears to be primarily controlled by the blood concentration of ketone bodies rather than a hormonal effect. Ketone body production by the liver peaks after 3–4 days of fasting. However, blood ketone body levels continue to rise rapidly for the first 7–10 days before stabilizing at approximately 6–8 mM at 2–3 weeks. The continued rise in blood ketone body levels despite achieving maximal liver production early in fasting is due to decreased renal excretion of ketone bodies and increasing muscle fatty acid oxidation (Fig. 6).

As fatty acid oxidation and ketone body formation increase, there is a reduction in glucose production and oxidation mediated by downregulation of the pyruvate dehydrogenase complex activity. After a 3-week fast, a marked reduction in glucose metabolism throughout the brain is observed using positron emission tomography. Glucose uptake of the brain is more than halved after a fast of 5 weeks.

After a period of fasting longer than 3 weeks, the process of metabolic adaptation to starvation is complete. Gluconeogenesis and glycolysis have been minimized paralleled by increase in hepatic ketone body production. The kidney becomes the major gluconeogenic organ and produces half of the body's glucose requirements. Glutamine is the predominant substrate for kidney gluconeogenesis, and the nitrogen product of this process provides the ammonia needed to buffer ketoacids in the urine. This saves energy compared to the energy-intensive ammonia disposal through the hepatic urea cycle. As a result, urinary nitrogen losses decrease to 4–6 g day^{-1}. Two-thirds of brain fuel consumption consists of ketone bodies, thereby markedly diminishing the need for muscle proteolysis to provide gluconeogenic precursors. With prolonged fasting, muscles change from ketone body production to fatty acid oxidation.

At completion of adaption, there is slow and ongoing depletion of the protein compartment and breakdown of the adipose tissue. Death will occur when there is a failure to replenish fuel stores through refeeding and insufficient available energy to maintain essential bodily functions. As fat is the predominant source of energy, the time until death in uncomplicated fasting will depend on the size of the prefasting fat stores. In a normal adult, fat stores will be sufficient to sustain life for approximately 60–70 days. The extent of protein loss is also linked to survival, and a loss of more than half of the lean body mass compartment (approximately half of total body protein) is predictive of death.

Regulation of adaptation in starvation

Metabolic regulation in starvation involves a complex interplay of nutrient sensing mechanisms and neural/endocrine adaptive processes on a tissue and cellular level, leading to structural, biochemical, and functional modifications.

Neural and endocrine regulation

The hypothalamus, in particular the arcuate nucleus (ARC), is acknowledged as a major center for integrating nutritionally relevant information from peripheral organs into an energy code setting the level of whole-body energy expenditure. Reduction in resting energy expenditure during fasting cannot be solely explained by decrease in lean body mass, and coordination of hormonal changes on the hypothalamic level plays an additional important role. Catecholamine secretion and turnover are decreased in uncomplicated starvation (Table 3). This is clinically recognized as a reduction of core temperature, heart rate, and blood pressure during starvation.

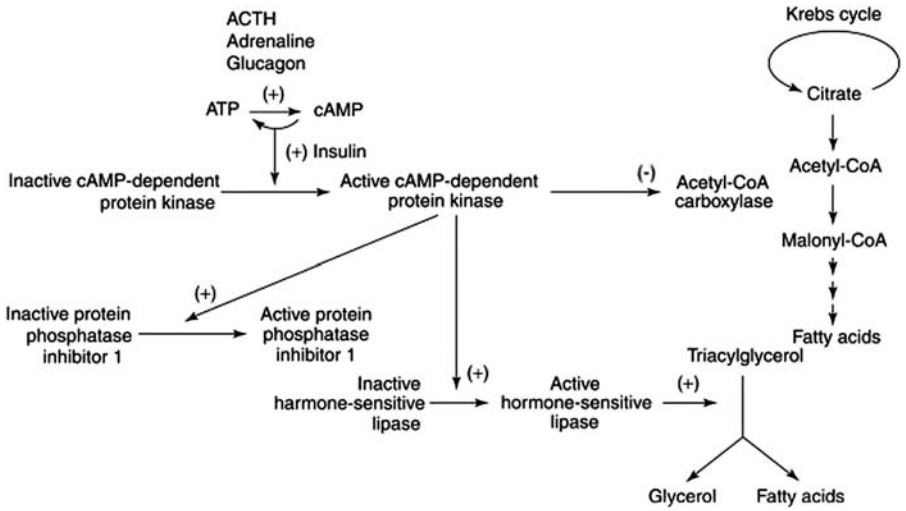


Fig. 6 Lipolysis is stimulated by the action of glucagon, ACTH, and adrenaline. This effect is mediated by cyclic AMP-dependent protein kinase.

Table 3 Effects of the major hormones in the steady (fed) state, during fasting and starvation.

Hormonal change and metabolic consequences		*Steady (fed) state* *Exogenous energy consumption*	*Fasting state* *Transition to endogenous energy sources*	*Starvation* *Metabolic adaptation: use of alternative endogenous energy sources; reduction in energy utilization and expenditure*
Hormone	Insulin	↑	↓	↓
	Glucagon	↓ Or unchanged	↑	↑
	Cortisol	Normal diurnal variation	↑	↑
	Growth hormone	Normal diurnal variation	↑	↑[†]
	Epinephrine	Normal variation in response to stressors	↑	↑/↓
	Thyroid hormones	Normal range	↓	↓
Metabolic consequences		Increased glucose uptake Glycogenesis* Lipogenesis Protein synthesis	Reduced peripheral glucose uptake/usage Glycogenolysis* gluconeogenesis Ketogenesis Lipolysis Proteolysis	Reduced peripheral glucose uptake/usage Gluconeogenesis*# Ketogenesis*^ Lipolysis Proteolysis Glycogenolysis

[†]Loss of IGF-1 mediated anabolic effect.
*Predominant metabolic process.
#During early starvation.
^During late starvation.

Thyroid hormones are known as potent hypothalamic regulators of whole-body energy homeostasis. Fasting results in marked suppression of diencephalic thyroid-stimulating hormone (TSH) secretion as a consequence of a fall in leptin (an adipocyte-derived hormone, which induces biosynthesis and release of hypothalamic TRH under non-fasting conditions through an increased hypothalamic type 2 de-iodinase-dependent triiodothyronine (T_3) production). Decreased activity of 5′-monodeiodinase in the liver and peripheral tissues resulting in a reduction in the conversion of thyroxine (T_4) to the metabolically active form T_3, has been observed within hours to days in patients on a starvation diet. However, the mechanism linking low-circulating T_3 levels to decreased resting energy expenditure in starvation is not well understood yet.

Glucagon plays a crucial role in the early phase of adaptation to starvation (Table 2) and stimulates hepatic glucose output by increasing hepatic glycogenolysis and gluconeogenesis through its canonical cAMP/PKA pathway, facilitating phosphorylation and allosteric change of key metabolic enzymes and nuclear factors (e.g., CREB, cAMP response element-binding protein).

In prolonged starvation, cortisol increases hormone-sensitive lipase synthesis. Insulin levels fall by 35% within 24 h of fasting. This is associated with a 50–80% increase in the rate of lipolysis. Low-circulating insulin levels cause a reduction in the uptake of glucose into adipocytes by altering the function of the GLUT4 glucose transporter (Fig. 4). Adequate amounts of glycerol-3-phosphate are therefore unavailable for the re-esterification of fatty acids produced from triacylglycerol breakdown.

Nutrient sensing mechanisms

The coupling of the body's metabolic environment to the energy status of the cell is crucial for its adaptive response to starvation. Cells have sensors detecting limitation of nutrient availability during starvation, initiating homeostatic mechanisms in order to tailor their metabolic needs to nutrient fluctuations and aiming to restrict overall energy expenditure. The AMP-activated protein kinase (AMPK) and the evolutionary-conserved NAD^+-dependent histone deacetylase SIRT1 axis integrates multiple hormonal and nutritional signals (e.g., glucagon, leptin, adiponectin, glycogen, and free fatty acids) to regulate cellular and whole body energy homeostasis during starvation (Table 4).

AMPK is a heterotrimeric complex composed of a catalytic α-subunit and two regulatory (β and γ) subunits with multiple isomers. Different subunit compositions (AMPK complexes) may have specific roles and respond to different types of stress stimuli. AMPK is activated under conditions of low-energy charge by the binding of AMP (and less so, ADP) to the γ-subunit leading to: (1) possible stimulation of Thr172 phosphorylation (on the α-subunit) by an upstream kinase (LKB1); (2) inhibition of dephosphorylation of Thr172 by protecting it from phosphorylases; and (3) allosteric activation of AMPK already phosphorylated on Thr172. Thr172 may also be directly phosphorylated by increased intracellular calcium through the calcium-sensitive kinase CAMKK2. Recent studies identified AMPK is localized in distinct compartmentalized pools with different targets and in lysosomes, is activated by an AMP-independent mechanism in response to low glucose levels, when aldolase binding fructose-1,6-biphosphate regulates AMPK-axin binding, resulting in co-regulation of AMPK and mTOR. Activated AMPK has essential roles in mitochondrial regulation, directing metabolism toward increased catabolism (increased energy production) and decreased anabolism (reduced energy usage) (Table 4).

Table 4 Effects of AMPK-SIRT1 axis on metabolism during starvation.

Inhibition of anabolism	*Stimulation of catabolism*
↓ Lipid and sterol synthesis - Inhibitory phosphorylation of acetyl-CoA carboxylases ACC1 and ACC2, HMG-CoA reductase ↓ Glycogen storage - Inhibitory phosphorylation of glycogen synthases GYS1 and GYS2 - Inhibitory hexosamine synthesis - Inhibitory O-G1cNAc action by phosphorylation of GFAT1 ↓ Protein synthesis - Reduced mTOR activity - Phosphorylates Eef2k Modified transcriptional regulation of biosynthetic pathways - ↓ Gluconeogenesis [cAMP response element-binding protein, FOXO] - Phosphorylates key regulators of lipid and glucose metabolism [SREBP1, ChREBP, HNF4α, phosphor late histone acetyltransferase p300] Regulation of other growth related pathways - Hedgehog signaling - Hippo pathway - JAK-STAT pathway - p53 tumor suppressor pathway	↑ Mobilization lipid stores - Stimulates lipases to release fatty acids for β-oxidation in mitochondria - Modulates acyl transferase (CPT1) activity to import fatty acids into mitochondria ↑ Glucose utilization - Increases plasma cell localization of GLUT1 and GLUT4 - Phosphorylation of phospholipase D1, PFKFB3 Modified transcriptional regulation of biosynthetic pathways - Upregulates genes involved in mitochondrial biogenesis, autophagy and lysosomal degradation, mitophagy

Source: Herzig and Shaw (2018).

mTOR, mammalian target of rapamycin; FOXO, forkhead box protein O; SREBP1, sterol regulatory element-binding protein 1; ChREBP, carbohydrate-responsive element-binding protein; HNP4α, hepatocyte nuclear factor 4α; STAT, signal transducer and activator of transcription; PFKFB3, 6-phosphofructo-2-kinase/fructose-2, 6-biphosphatase 3.

SIRT1 is induced and activated by intracellular NAD^+ levels (nicotinamide adenine dinucleotide) that increase similar to AMP with fasting and nutrient restriction leading to restitution of intracellular energy balance by switching off energy-intensive biosynthetic pathways (such as protein synthesis, glycogen synthesis, fatty acid, and sterol synthesis) and favoring ATP production (through, e.g., lipolysis, fatty acid oxidation, and mitochondrial biogenesis). Further, deacetylation of lysine residues helps SIRT1 to increase the degree of chromatin compaction, leading to direct repression of transcriptional activity. On a structural level, this translates into the observed muscular plasticity during starvation, with a switch from fast-twitching glycolytic type II fibers to slow-twitching oxidative type I fibers.

The Forkhead box O (FOXO) transcription factor family has received considerable attention as elemental regulators of metabolism. FOXOs play a crucial role in transcriptional upregulation of key enzymes of hepatic gluconeogenesis in early starvation (G6Pase, PEPCK), blockage of adipocyte differentiation, decrease in protein synthesis, and increase in protein degradation (through UPS and autophagy; see above). Its activity and cellular localization are controlled and increased by its acetylation (through SIRT1) and phosphorylation (through AMPK) states.

The peroxisome proliferator-activated receptors (PPARs) are another family of important ligand-activated transcription factors. PPARα, for example, acts as an intracellular fatty acid sensor which is one of the master regulators of cellular fatty acid oxidation and regulates ketogenesis. The estrogen receptor-related receptor (ERR) family (α, β, γ) represents a further important downstream target of AMPK. Activation is associated with increased expression of fatty acid oxidation genes and enzymes of the oxidative phosphorylation cascade (OXPHOS). It also enhances pyruvate dehydrogenase kinase 4 (PDK4), the crucial enzyme involved in the starvation-induced shutdown of the pyruvate dehydrogenase complex. Although these nuclear transcription factors confer a first level of specificity to transcriptional adaptive processes during times of energy deficit, coregulators are required for the transcriptional machinery to become fully activated, significantly increasing the diversity of interacting partners and thus the complexity of this regulatory process. The PPARγ coactivator-1alpha (PGC-1α) is generally accepted as the central node in starvation-induced transcriptional coactivation and is involved in the regulation of large clusters of genes controlling oxidative phosphorylation and fatty acid metabolism. It drives, for example, the switch from glycolytic to oxidative muscle fibers. It exerts its effect in conjunction with the transcription factors FOXO, PPAR, and ERR. PGC-1α is controlled by the AMPK-SIRT1 axis and activated—similar to FOXO—by deacetylation and phosphorylation.

Fibroblast growth factor 21 (FGF21) is a fasting induced hormone produced primarily in the liver as well as other tissues including skeletal muscle. It is thought to contribute to the late stages of adaptive starvation by regulating fuel derived from tissue breakdown although data is conflicting and mostly generated from mice, with its exact role in humans to be defined.

Conclusion

Metabolic adaptation is a complex and coordinated process which stimulates ongoing provision of energy in situations of adverse energy supply such as prolonged fasting and starvation. It includes a highly evolved and integrated network of energy sensing

mechanisms, neural and endocrine regulatory factors and modifications to human energy utilization pathways within the body to preserve the function of critical organs.

References

Heird, W.C., Driscoll Jr., J.M., Schullinger, J.N., Grebin, B., Winters, R.W., 1972. Intravenous alimentation in pediatric patients. J. Pediatr. 80 (3), 353.
Herzig, S., Shaw, R.J., 2018. AMPK: guardian of metabolism and mitochondrial homeostasis. Nat. Rev. Mol. Cell Biol. 19 (2), 121–135.

Further reading

Ahima, R.S., Prabakaran, D., Mantzoros, C., et al., 1996. Role of leptin in the neuroendocrine response to fasting. Nature 382, 250–252.
Cahill, G.F., 1970. Starvation in man. N. Engl. J. Med. 282, 668–675.
Cano, N., 2002. Bench to bedside review: glucose production from the kidney. Crit. Care 6, 317–321.
Canto, C., Auwerx, J., 2010. AMP-activated protein kinase and its downstream transcriptional pathways. Cell. Mol. Life Sci. 67, 3407–3423.
Feige, J.N., Auwerx, J., 2007. Transcriptional coregulators in the control of energy homeostasis. Trends Cell Biol. 17 (6), 292–301.
Feldman, M., Friedman, L.S., Brandt, L.J., 2010. Gastrointestinal and Liver Disease: Pathophysiology, Diagnosis, Management, ninth ed., vol. 1. Saunders Elsevier, Philadelphia.
Frayn, K.N., 2009. Metabolic Regulation—A Human Perspective, third ed. Wiley Blackwell, Chichester.
Gross, D.N., Wan, M., Birnbaum, M.J., 2009. The role of FOXO in the regulation of metabolism. Curr. Diabetes Rep. 9, 208–214.
Hardie, G.D., 2007. AMP-activated/SNF1 protein kinases: conserved guardians of cellular energy. Nat. Rev. Mol. Cell Biol. 8, 774–785.
Kadowaki, M., Kanazawa, T., 2003. Amino acids as regulators of proteolysis. J. Nutr. 133 (Suppl. 1), 2052S–2056S.
Klein, S., Sakurai, Y., Romijn, J.A., Carroll, R.M., 1993. Progressive alterations in lipid and glucose metabolism during short-term fasting in young adult men. Am. J. Physiol. 265, E801–E806.
MacDonald, I.A., Webber, J., 1995. Feeding, fasting and starvation: factors affecting fuel utilization. Proc. Nutr. Soc. 54, 267–274.
Mehrpour, M., Esclatine, A., Beau, I., Codogno, P., 2010. Autophagy in health and disease. 1. Regulation and significance of autophagy: an overview. Am. J. Physiol. Cell Physiol. 298, C776–C785.
Murayama, A., Ohmori, K., Fujimura, A., et al., 2009. Epigenetic control of rDNA loci in response to intracellular energy status. Cell 133, 627–639.
Naimi, M., Arous, C., Van Obberghen, E., 2010. Energetic cell sensors: a key to metabolic homeostasis. Trends Endocrinol. Metab. 21 (2), 75–82.
Nedelsky, N., Todd, P.K., Taylor, J.P., 2008. Autophagy and the ubiquitin-proteasome system: collaborators in neuroprotection. Biochim. Biophys. Acta 1782 (12), 691–699.
Randle, P.J., 1995. Metabolic fuel selection: general integration at the whole body level. Proc. Nutr. Soc. 54, 317–327.
Robinson, A.M., Williamson, D.H., 1980. Physiological roles of ketones as substrates and signals in mammalian tissues. Physiol. Rev. 60, 143–187.
Sandri, M., 2008. Signaling in muscle atrophy and hypertrophy. Physiology 23, 160–170.
Steinberg, G.R., Kemp, B.E., 2009. AMPK in health and disease. Physiol. Rev. 89, 1025–1078.
Stipanuk, M.H., 2009. Macroautophagy and its role in nutrient homeostasis. Nutr. Rev. 67 (12), 677–689.
Yu, J., Auwerx, J., 2009. The role sirtuins in the control of metabolic homeostasis. Ann. N. Y. Acad. Sci. 1173, E10–E19.

Additional sources to be added into further reading list

Berg, J.M., Tymoczko, J.L., Stryer, L., 2002. Each organ has a unique metabolic profile. In: Biochemistry, fifth ed. WH Freeman and Company, New York. Section 30.2.
Chattopadhyay, T., Maniyadath, B., Bagul, H.P., et al., 2020. Spatiotemporal gating of SIRT1 functions by O-GlcNAcylation is essential for liver metabolic switching and prevents hyperglycemia. Proc. Natl. Acad. Sci. U. S. A. 117 (12), 6890–6900.
El Bacha, T., Da Poian, A.T., 2010. Dynamic adaptation of nutrient utilization in humans. Nat. Educ. 3 (9), 8.
Fazeli, P.K., Lun, M., Kim, S.M., et al., 2015. FGF21 and the late adaptive response to starvation in humans. J. Clin. Invest. 125 (12), 4601–4611.
Inagaki, T., Dutchak, P., Zhao, G., et al., 2007. Endocrine regulation of the fasting response by PPARalpha- mediated induction of fibroblast growth factor 21. Cell Metab. 5, 415–425.
Soeters, M.R., Soeters, P.B., Schooneman, M.G., et al., 2012. Adaptive reciprocity of lipid and glucose metabolism in humans in short-term starvation. Am. J. Physiol. Endocrinol. Metab. 303, E1397–E1407.
Yorimitsu, T., Klionsky, D.J., 2005. Autophagy: molecular machinery for self-eating. Cell Death Differ. 12, 1542–1552.
Zong, Y., Zhang, C.-S., Li, M., et al., 2019. Hierarchical activation of compartmentalized pools of AMPK depends on severity of nutrient or energy stress. Cell Res. 29, 460–473.

Stroke nutritional management

L. Perry, University of Technology Sydney, NSW, Australia

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of L. Perry, Stroke Nutritional Management, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 219–230, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00256-7.

Introduction **658**
Risks of Protein-Energy Malnutrition 658
Prestroke Nutritional Status 658
Poststroke Eating Problems 658
Organizational Factors 659
Management of Stroke-related Psychosocial and Physical Problems Impairing Food Consumption **659**
Evidence-based Guideline Recommendations 659
Psychosocial Problems 660
Communication Problems 660
Impairments of Arm Movement and Posture 660
Visual Field Loss and Visual Neglect 661
Attention Span, Short-Term Memory 661
Swallowing Difficulties 661
Screening and Assessment 661
Nutritional Management and Treatment 663
Nutrient Requirements 665
Artificial Nutritional Support 667
Evaluation of Nutritional Support 668
Further Reading **668**

Glossary

Aphasia Acquired language disorder; may include difficulty in producing or comprehending spoken or written language. Originally aphasia referred to total inability to communicate, dysphasia to partial impairment; over time, aphasia has come to mean either both partial and total language impairment in common use.

Apraxia Loss of the ability to execute or carry out learned purposeful movements, despite the desire and physical ability to do so. It is a disorder of motor planning, not caused by inco-ordination, sensory loss, or failure to comprehend. The activity may be performed as a conditioned response (e.g., if handed a cup, the patient may take it and drink, although may not be able to do this to command).

Dysphagia Difficulty in swallowing, can be neurological, structural, or functional in origin. This article refers only to neurological dysphagia. Although classified under 'signs and symptoms' of the ICD-10 classification, it is often discussed as a disease-state *per se*.

Naso-Gastric Tube A means to deliver liquid feed directly into the stomach for people who are unable or unsafe to take food orally. A fine-bore, usually silicone or polyurethane feeding tube with a guidewire to aid placement through the nose, the esophagus, and into the stomach. After checking its position, the guidewire is removed. The tube may be retained in place using adhesive tape applied to the nose, a clip attached to the tube may be secured to the face, or a nasal bridle may be used: a device that secures the tube in place by tapes that hold the tube in the nasopharynx and exit through both nostrils. Most often used for short-term feeding, but may be used longer-term.

Percutaneous Endoscopic Gastrostomy (PEG) A means to deliver liquid feed directly into the stomach for people who are unable or unsafe to take food orally. A feeding tube placed into the stomach through the abdominal wall, usually using a combination of local anesthetic infiltration into the abdominal wall and intravenous sedation. During the procedure an endoscope is passed through the mouth to help locate the trans-abdominal puncture and placement of the feeding tube. Once the tract is established, most PEG tubes need only their retaining devices (flanges or balloons that are part of the tubing) to maintain position. May be retained in place for long periods.

Introduction

Stroke is a common and devastating event, the incidence rising with age. Approximately 125 000 and 500 000 new or recurrent strokes affect individuals each year in the UK and the US, respectively, creating a significant burden of long-term disability in survivors. Stroke is a syndrome that is sudden in onset, featuring signs of cerebral dysfunction that are vascular in origin. In first strokes, thrombo-embolic infarction is the underlying pathophysiological event in approximately 80% of cases, with 10% attributed to primary intracerebral hemorrhage and 10% to other/uncertain types. Resulting neurological and functional impairments vary in range, combination, and severity. These can include altered consciousness, motor and somato-sensory, auditory, visual, smell and taste disturbance, and loss. Impairments of memory, speech, language, continence, and higher-order cognitive functioning such as planning and decision-making also occur. Overall, the impairments and disabilities that result from stroke can exert a variable, negative impact on eating, mobility, mood, ability to selfmanage and selfcare, social role-function, and quality of life.

The challenge of nutritional management is apparent in the scope and complexity of issues presented. These include the physical, psychological, and social impact of stroke on appetite and eating, as well as prestroke nutritional status. Potential effects of metabolic injury responses, immobilization, and infective complications on energy, nitrogen, and micronutrient requirements must be considered. In individuals in whom oral feeding cannot be established, parenteral hydration and artificial nutritional support may be necessary, by the enteral route if possible, utilizing nasogastric tubes (NGT) or percutaneous endoscopic gastrostomy (PEG).

Effective nutritional management of stroke requires the integrated skills and knowledge of a multidisciplinary team. Complex situations entailing ethical and legal issues relating to hydration and nutritional support can arise, and involvement of patient and carers in all aspects of planning is vital.

Risks of Protein-Energy Malnutrition

The reported frequency of malnutrition after acute stroke has ranged from 6% to 62%, measured via multiple assessment methods, some of which have been previously validated (e.g., Subjective Global Assessment and Mini Nutritional Assessment). This is important because the presence of under-nutrition shortly after admission has been independently associated with significantly greater mortality and likelihood of developing pneumonia, infections, pressure sores and gastrointestinal bleeds in hospital, with increased risk of death or dependency poststroke.

A number of factors may interact to impair the nutritional status of stroke patients. These include factors that led to deterioration in prestroke nutritional status, direct physical and psychosocial effects of stroke on the consumption of food and fluids following hospital admission, and organizational factors that can hinder efficient, effective meal delivery and consumption in institutional and residential settings.

Prestroke Nutritional Status

Risk factors for malnutrition identified at the time of hospital admission have included increasing age, living alone, dementia, and inadequate dental status. Findings are consistent with those from surveys of elderly populations, where a wide range of personal characteristics and social and environmental features have been linked with poorer appetite and nutritional status (see related articles) The presence of malnutrition in older adults may be associated with impaired immune responses, increasing vulnerability to pneumonia and sepsis. The significance of a low serum albumin concentration in predicting clinical outcomes has been noted in stroke and older populations, where it may be more a reflection of disease severity than an indicator of nutritional status.

Poststroke Eating Problems

Neurological and functional impairments can result in eating problems following stroke, which can lead to an increased risk of protein-energy malnutrition. During the acute phase of recovery, eating disability has been associated with inadequate consumption of food and fluids and deterioration in body mass index, triceps skin fold thickness, mid-arm muscle circumference, and serum protein concentrations. Four months poststroke weight loss >3 kg (mean 6.6 kg) was found in 24% patients, and in 26% (mean loss 8.3 kg) at 1 year. Eating difficulties and low prealbumin values were associated with weight loss at both time points; at six months poststroke 66% of survivors were found with some degree of enduring disablement that affected eating. Specific eating problems include impaired lip closure leading to oral leakage of food and fluids, dysphagia, inability to manipulate utensils linked to loss of motor skills in eating, and difficulties in maintaining an upright posture at mealtimes. Visual field or perceptual deficits can result in inability to see or perceive the contents of a meal tray, whereas aphasia and dysarthria can hinder or prevent expression of dietary needs and preferences. Loss of concentration, short-term memory and cognitive impairments such as apraxias are common sequelae to stroke and can make it difficult for individuals to sustain the sequence of activities necessary to complete a meal, or even to remember to eat and drink (**Table 1**). Taste and smell dysfunctions have been reported but impact on eating has not been investigated. A number of assessment instruments have been developed to enable health professionals to identify the extent of eating disability and the nature of support needed.

Table 1 Eating disabilities of a cohort of acute stroke admissions to a South London hospital, March 1998–December 1999

Eating disabilities[a] at hospital	*3–5 days after admission to*	*Number (%)*
Posture control	No functional impairment	325 (56)
	Mild–moderate impairment	219 (37)
	Severe impairment	43 (7)
Arm movement	No functional impairment	159 (27)
	Mild–moderate impairment	295 (50)
	Severe	133 (23)
Lip closure	No functional impairment	449 (76)
	Partial impairment	109 (19)
	Severe impairment	29 (5)
Chewing	No functional impairment	369 (63)
	Partial impairment	172 (29)
	Severe impairment	46 (8)
Swallowing	No functional impairment	341 (58)
	Partial: Cannot tolerate 1 of 3 textures	66 (11)
	Severe: Cannot tolerate 2 of 3 textures	64 (11)
	Aspiration/high risk; nil orally	116 (20)
Communication	No functional impairment	276 (47)
	Partial impairment	149 (25)
	Severe impairment	162 (28)
Attention and praxis	No functional impairment	417 (71)
	Partial impairment	130 (22)
	Severe impairment	39 (7)
Visual field/ perceptual loss/ neglect	No functional impairment	428 (73)
	Partial impairment	132 (23)
	Severe impairment	25 (4)

[a]Total 670 stroke patients, of whom 587 (586 for attention and praxis; 585 for visual fields) were able to be assessed.

Organizational Factors

Organizational factors can hinder dietary provision and consumption in institutional settings. These include an inadequate mealtime environment, marked by poor lighting, noise, distractions (e.g., television, consultant ward rounds), unpleasant smells, temperature extremes, and lack of facilities or facilities that deter social eating. Lack of adequate assistance for dietary selection, meal delivery, and supervision may result in a meal is that inappropriate in relation to texture, portion size, patient preference and swallowing ability, or, with around half needing help to eat in the early stages poststroke, which the patient is unable to eat without help, which may not be available.

Management of Stroke-related Psychosocial and Physical Problems Impairing Food Consumption

Evidence-based Guideline Recommendations

The provision of effective nutritional care requires a concerted approach by health professionals in developing locally tailored evidence-based standards and guidelines for nutritional screening, assessment, and dietary support. These should be linked to national standards and guidelines (where they exist), clinical audit and processes for practice development. Recognition of the need for such guidance has led to the development and dissemination of evidence-based guidelines by interprofessional expert groups, designed to inform professional judgment and bench-mark care within wider stroke management. Implementation of guidelines can best be achieved by multifaceted strategies, for example, combining education of health professionals with leadership and sensitive change management approaches.

Following acute stroke, guideline recommendations emphasize the need for screening and assessment for nutritional risk to be undertaken on admission to hospital and repeated at regular intervals (e.g., weekly for hospital inpatients) using a valid, reliable instrument by appropriately trained personnel. Individuals who are malnourished at the point of admission, or likely to become so, can then be referred to dietitians for further assessment and prompt initiation of nutritional support. Early identification of

individuals whose swallow function may be unsafe with oral food and fluids is vital, as well as referral for specialist swallow assessment, usually by a speech pathologist. Screening of swallowing function should be undertaken using a validated and reliable method by appropriately trained personnel, initially as soon as the patient is able to be assessed poststroke or on admission to hospital, before patients are given food or drink. Following detailed assessment, modification of dietary textures may be advised to ensure safe eating in those with some degree of swallowing impairment. In others, artificial nutritional support using enteral routes may be necessary, owing to the severity of dysphagia or cognitive impairments. In individuals who are capable of taking food orally, the provision of support for physical, functional, and psychosocial problems is an essential aspect of nutritional management. To achieve this, appropriate specialist assessment should be requested, e.g., occupational therapy, physiotherapy, and psychology.

Psychosocial Problems

In the acute phase following stroke, approximately 30% of patients develop clinical signs of depression, 30% are anxious, and a similar proportion report loss of confidence as a major psychological problem. However, mood dysfunction may occur at any time, from immediately after stroke onset to many months later. Depression may result from an interaction of factors, including a direct result of the stroke lesion. Despite extensive investigation, there is no robust evidence that depressed mood after stroke is caused by a lesion in any particular area or side of the brain. Depression is more common amongst people with chronic disease than matched otherwise healthy groups, and associated with reactions to physical loss, altered lives and sense of identity. Comparatively little is known of interactions between depression, anorexia, and nutritional status in the early stages of recovery following stroke in individuals with and without physical eating problems. However, at 6 months patterns of behavioral disturbance characterized by depressed mood, anorexia, and insomnia have been identified and associated with weight loss. Anxiety-evoking experiences relating to being fed, or choking in the presence of dysphagia, may also result in avoidance or withdrawal from eating. General approaches to the treatment of poststroke depression are not different to the nonstroke population, and include antidepressant drugs, and behavioral and psychotherapeutic techniques. Use of therapeutic skills in communication, eating assistance, and provision of emotional support are vital in alleviating mealtime anxiety and increasing interest in food.

The enjoyment of eating as a social activity can be affected adversely by impairments of speech, lip closure, chewing and swallowing, and manual dexterity. Severely disabled individuals who are relearning eating and swallowing skills initially require privacy and a quiet environment. As rehabilitation progresses, social integration at mealtimes can be achieved.

Communication Problems

Whilst dysphasia refers to difficulty, and aphasia to inability to communicate verbally, aphasia is often used as a relative term. It affects 20–38% of acute stroke patients with possibly as many as 40% affected by dysarthria (speech difficulties due to oromotor dysfunction), resulting in variable difficulty or inability to express thoughts in language (expressive aphasia) or to comprehend language (receptive aphasia). Expressive aphasia, also known as Broca's aphasia, results from strokes affecting the prefrontal gyrus, whereas Wernicke's receptive aphasia results from lesions of the central sulcus. Dysarthria results from neurological damage affecting neuromuscular systems that control speech production; because these systems are also concerned with swallowing, dysarthria often coexists with difficulty swallowing (dysphagia).

Communication problems can result in inabilities/difficulties in expressing hunger, thirst, meal preferences (aphasia), reading a menu, or writing preferences (aphasia can occur alongside dyslexia and dysgraphia). Receptive aphasia can impair comprehension of instructions at mealtimes and thus affects response to information and rehabilitative advice. If paralysis and visual field and perceptual deficits are combined with expressive communication deficits, nonverbal communication can also be limited, affecting ability to signal assent or dissent by nodding the head or to use gestures or point to food items/utensils. Early involvement of speech pathologists is vital to enable individuals to regain lost functions in speech and language. In selected patients, visual material, e.g., pictures and symbols, can be helpful. Use of short sentences with single topics, simple terms, no jargon, normal volume speech, a clear light with a good view of the speaker's face and appropriate gestures, and patience in allowing individuals time to respond are helpful in general communication.

Impairments of Arm Movement and Posture

Stroke can affect any of the neural mechanisms controlling voluntary movement and posture. These include the motor and sensory cortex and associated pathways, cerebellum, basal ganglia, and brain stem. The impact on eating skills can be considerable, because weakness or paralysis affecting the arm occurs in 80% of strokes. Loss of coordination, spatial awareness, abnormal muscle tone, and sensory loss may also occur. Common problems resulting from this are difficulties manipulating cutlery, lifting/loading food onto utensils, cutting food, inserting food in the mouth, drinking from a cup, or discerning the spatial relationships between objects.

If one arm is unaffected, then some degree of compensation is possible, particularly if this is dominant (unless this has been intentionally immobilized: Constraint Induced Therapy). Use of the unaffected hand is important in detecting temperatures of food and liquids where sensation is impaired. An occupational therapy assessment is necessary to identify specific deficits,

rehabilitation techniques, and appropriate aids to eating. Lightweight plastic cups and cutlery with molded or built-up handles, plate-guards and nonslip mats can be provided. Where upper limb impairment is severe, individuals may require assistance or to be fed.

Postural impairment following stroke can result in inability to maintain an upright sitting position, required for safe and effective food preparation, eating and swallowing. A physiotherapy assessment can identify the most effective techniques for rehabilitation of muscle weakness and to counteract abnormal muscle tone (spasticity). Appropriate aids to seated balance can include molded seating and supports.

Visual Field Loss and Visual Neglect

Between 30% and 60% of individuals who sustain an acute stroke suffer from visual field loss due to partial or complete hemianopia. Neurological damage affecting the parietal or temporal lobes and involving the sensory pathway between the optic chiasma and visual cortex underlies this problem. The impact of loss in up to half the visual field is that food items on a meal tray may not be seen and therefore may remain uneaten. Compensatory interventions include instruction in scanning the visual field, or placing items within it for those who are unable to do this. Consistent placement of items on a meal tray and verbal identification of contents using a clock system is also helpful.

Neurological damage affecting the visual cortex of the occipital lobe, often following right hemisphere strokes, can result in neglect of half the visual space. A classic feature is failure to eat food on the left side of a plate. Affected individuals need reminding to focus on food items in the neglected space; placing a colored marker on one side of the plate can be helpful. This problem may occur in conjunction with visual field loss.

Attention Span, Short-Term Memory

Impairment of attention span and short-term memory of a few minutes duration are common following acute stroke. Attention deficits result in an inability either to focus on immediate events or to establish a new focus unless a current stimulus is removed. As a consequence, an activity that requires a sequence of steps, such as eating a meal with two or three courses, may not be completed unaided. Lack of concentration also deters relearning eating patterns. Removing or minimizing distractions at mealtimes, simplifying the complexity of information necessary to regain eating skills, and providing verbal, written, or auditory alarms as reminders to eat are important in overcoming this problem.

Swallowing Difficulties

Screening and Assessment

Dysphagia affects approximately 45–60% of people with acute stroke. Variable in severity, it is characterized by sensory or motor loss affecting one or more of the stages of swallowing, i.e., oral preparation, oral transport, pharyngeal transport, and reflex swallowing (Table 2). The effects of stroke on esophageal peristalsis have been little examined. It has been estimated that approximately 50% of dysphagic patients either die or recover their swallow spontaneously within the first 2 weeks of stroke onset. For most of the others, at least some degree of functional swallowing can be restored with time. Approximately 10–20% will still be affected by 6 months; a small number never recover swallow function, although for some, recovery may take years. Swallowing impairment that prohibits eating normal texture food can exert a negative impact on functional recovery and quality of life.

Typical clinical features of dysphagia include delayed oral and pharyngeal transit times, retention of food in the cheek cavities, uncontrolled leakage of food/fluids out through the lips or onwards into the pharynx, causing choking and regurgitation of food/fluids through the nose and mouth. Tongue coordination may be poor, triggering of swallow and laryngeal cartilage elevation may be delayed or absent, gag may be abnormal. After swallow there may be a wet, 'gurgly' voice or coughing. Alternatively, the patient may aspirate silently, with no overt signs. Complications resulting from dysphagia can be life threatening, i.e., aspiration of food/fluid into the respiratory tract resulting in pulmonary infection, and dehydration. Longer hospital stay, strong inverse correlations with functional capacity, and an increased mortality have been associated with dysphagia. Early identification of the problem is vital, encompassing screening as soon as possible after admission, clinical bedside assessment (CBA), and, if necessary, instrumental assessment using, for example, videofluoroscopic swallowing studies (VSS) or fibreoptic endoscopic examination of swallowing (FEES).

Screening is a procedure intended to identify patients with potential swallowing problems, who can then be referred for more detailed assessment of phases of swallowing, together with judgment of extent of dysfunction and risk of aspiration. Systematic reviews have identified a number of screening methods of varying validity and reliability, which combine identification of clinical features of dysphagia with or without swallowing water. Prescreen assessment of conscious level, oromotor and laryngeal function, signs of existing respiratory aspiration and the extent to which the patient can safely cooperate with the examination are necessary. Aspiration of food or fluid into the respiratory tract may be accompanied by choking, indicated by voice changes (wet, hoarse, gurgling) or breathlessness, or it may be silent. Loss of swallowing and protective gag reflexes or the presence of features of dysphagia or aspiration when attempting to swallow water at an initial screen are indications that nil should be given by mouth. Detailed investigation is then necessary. During the acute phase of stroke the patient's condition may not be stable, and even if there

Table 2 Stages of swallowing: effects of stroke

Stage	*Effects of stroke*
(1) Oral preparation	Inadequate lip seal causes leakage of food/fluid chewing slower, food impacts in oral sulci
Duration variable	Hyposecretion of saliva
Lip closure forms anterior seal	Paralysis of tongue impairs bolus formation
Chewing of food by mandibular and maxillary teeth	Sensory loss leads to impaired bolus lateralization
Salivation evoked by parasympathetic nervous system	
Bolus formation controlled by tongue	
Sensory feedback from oral mucosa on volume & consistency determine timing of bolus ejection	
(2) Oral transport	Slowed transport
Duration 1 s	Bolus localization, separation and formation impaired; can lead to food retention in oral cavity
Bolus of 5–15 cm^{-1} separated, moved to tongue midline	Lack of fine motor coordination may lead to loss of liquid bolus control; risk of aspiration
Oral cavity sealed, mandible raised, pressure exerted by tongue against palate propels bolus to posterior oral cavity	Abnormal positioning of bolus; diminished tongue elevation; inadequate bolus propulsion
(3) Pharyngeal transport/reflex swallowing	Events may occur in abnormal sequence/timing
Duration 0.5–0.6 s	Impaired sensation, delay/absent swallowing reflex
Bolus impacts on sensory receptors in tissues of soft palate, pharynx, tongue, fauces	Velopharyngeal closure impaired; food regurgitated through nose/mouth
Swallowing reflex stimulated; elevation/closure of velopharyngeal mechanism, elevation of larynx, closure of vocal cords, pharyngeal peristalsis, relaxation of esophageal sphincter	Incomplete laryngeal elevation/vocal cord adduction
	Swallowing reflex delay/absence leading to coughing, aspiration
Respiration transiently ceases as bolus enters esophagus; breathing resumed; soft palate returned to resting position	
(4) Esophageal transport	Effects of stroke little investigated
Duration 8–20 s	Aging results in slight impairment of peristaltic amplitude
Peristalsis moves bolus to stomach	

is no initial indication of problems, a high index of suspicion should be maintained and screening repeated if the patient's condition deteriorates.

There is no standardized format for clinical bedside assessment (CBA) but speech pathologists and other specially trained health professionals generally employ a combination of methods with demonstrated validity and reliability. These encompass the medical history relating to onset of swallowing problems; oral sensory and motor testing; laryngeal and pharyngeal assessment; presence/absence of swallowing, cough, and gag reflexes; cognitive and language function; alertness, attention span, and ability to follow instructions. Swallow function is usually assessed by observing the patient and feeling for swallow/laryngeal elevation with a graded sequence of foods and fluid consistencies. Adjunct assessments may be used, such as concurrent pulse oximetry to monitor oxygen saturation during swallowing, cervical auscultation to assess swallow function from swallow sounds, or assessment of muscle activity by cervical surface electromyography.

CBA provides limited detail of swallowing and functions poorly in relation to subtle signs of dysphagia and silent aspiration of food, fluids, or saliva into the respiratory tract: more invasive assessment may be required. VSS, using a modified barium swallow procedure, provides a detailed radiological assessment of the oral, pharyngeal, and upper esophageal phases of swallowing; can detect functional impairments resulting in aspiration, evaluate optimal head/neck positioning during swallowing, and determine the impact of food textures on the process. Limitations include its labor-intensive nature, exposure to radiation, problems transporting disabled stroke patients to radiology departments, variable protocol standardization with respect to volumes, consistencies, or textures of food and fluids and screening positions adopted. Further, views represent a snapshot of swallowing under ideal, rather than normal, circumstances, and variability in the reliability of reporting has been identified between and within raters of VSS outcomes. FEES entails passage of a flexible endoscope nasally, over the velum and into the pharynx. Its advantages lie in observation of bolus transit through the hypopharynx and identification of laryngeal penetration and aspiration; being portable it can be performed in a range of settings, overcoming the transport problems of VSS. However, it cannot investigate the oral stage of swallowing. The clinical utility of other methods are being explored, such as assessment of breathing patterns, to detect respiratory phases in relationship to swallows, and abnormal breathing patterns, such as swallow apnea followed by inspiration rather than expiration; impedance pharyngography, based on changes in the electrical impedance of the neck during swallowing.

Table 3 Standardized classifications of texture modified foods and fluids

Modified foods	*Au texture A – Soft/UK Texture E*	*Au texture B – Minced and Moist/UK Texture D*	*Au texture C – Smooth Pureed/UK Texture C*
Description	May be naturally soft (e.g., ripe banana), or may be cooked or cut to alter its texture	Soft and moist, may have some variation in texture, should easily form into a ball	Smooth and lump-free; may have a grainy quality, but no lumps
Characteristics	Soft foods can be chewed but should not need to be bitten. Minimal cutting required – easily broken up with a fork. Food should be moist or served with a sauce or gravy to increase moisture content	Easily mashed with a fork. May be presented as a thick puree with obvious lumps, but lumps are soft and rounded, not hard or sharp. Small lumps can be broken by the tongue	Moist and cohesive enough to hold its shape on a spoon (i.e., when placed side by side on a plate these consistencies maintain their position without 'bleeding' into one another). Can be molded, layered, piped. Au Runny puree/UK Textures A/B: Smooth, uniform consistency, will 'bleed' into one another; may be poured (A only); cannot be molded, layered, piped
Modified Fluids	*Au Mildly thick/UK Stage 1*	*Au Moderately thick/UK Stage 2*	*Au Extremely thick/UK Stage 3*
Description	Thicker than naturally thick fluids such as fruit nectars and commercial sip feeds.	Similar to room temperature pouring honey or a thickshake	Similar to pudding or mousse
Flow rate	Steady, fast flow; runs fast through the prongs of a fork, but leaves a mild coating on the prongs.	Slow flow; slowly drips in dollops through the prongs of a fork.	No flow; sits on and does not flow through the prongs of a fork.
Characteristics	Pours quickly from a cup but slower than regular fluids. May leave a coating in the cup/on the back of a spoon after being poured. Can be drunk from a cup, or a standard-bore straw with effort.	Cohesive, pours slowly. Can be drunk from a cup although flows very slowly/difficult or impossible to drink using a straw, even a wide bore straw	Cohesive and holds its shape on a spoon. Cannot pour from a cup; needs to be taken by spoon. Too thick if the spoon is able to stand upright in it.

Nutritional Management and Treatment

Use of modified food textures and fluid consistencies aims to reduce the level of challenge posed in terms of manipulation of the food bolus (i.e., by reducing the numbers of different textures presented and increasing the viscosity) and to alter the rate at which it passes through the pharynx, to maximize patient control of swallowing and reduce the risk of aspiration. National professional groups world-wide have developed classification systems for modification in the texture and consistency of foods and fluids. These usually cite approximately three grades, e.g., soft, minced, and pureed foods; mildly, moderately, and extremely thickened fluids (Tables 3–7).

It is commonly the responsibility of the speech pathologist to recommend the most appropriate food textures and maneuvers to promote safe eating for a dysphagic stroke patient (Tables 4 and 8). Dietitians ensure that texture-modified meals are adequate to meet nutrient requirements, offer choice, and are palatable. Skilled nursing assistance and interventions at mealtimes can also aid eating and mealtime processes. Impaired oral preparation may be compensated for by positioning food on, and tilting the head

Table 4 Australian standardised terminology and definitions for texture modified fluids 2007

	Mildly thick	*Moderately thick*	*Extremely thick*
Description	Thicker than naturally thick fluids such as fruit nectars, but for example, not as thick as a thickshake	Similar to the thickness of room temperature honey or a thickshake	Similar to the thickness of pudding or mousse
Flow rate	Steady, fast flow	Slow flow	No flow
Characteristics	Pours quickly from a cup but slower than regular, unmodified fluids. May leave a coating film of residue in the cup after being poured. Drink this fluid thickness from a cup. Effort required to take this thickness via a standard bore straw	Cohesive and pours slowly. Possible to drink directly from a cup although fluid flows very slowly. Difficult to drink using a straw, even if using a wide-bore straw	Cohesive and holds its shape on a spoon. It is not possible to pour this type of fluid from a cup into the mouth. Spoon is the optimal method for taking this type of fluid. This fluid is too thick if the spoon is able to stand upright in it
Testing Information	Subjectively, fluids at this thickness run fast through the prongs of a fork, but leave a mild coating on the prongs[a]	Subjectively, fluids at this thickness slowly drip in dollops through the prongs of a fork	Subjectively, fluids at this thickness sit on and do not flow through the prongs of a fork

[a]Testing scales for viscosity exist but are not formalised or standardised, and therefore are not included.

Table 5 Texture A—Soft: Australian standardised terminology and definitions for texture modified foods and fluids 2007

Food type	*Recommend*	*Avoid*
Bread, cereals, rice, pasta, noodles	• Soft sandwiches(a) with very moist fillings, for example egg and mayonnaise, hummus (remove crusts and avoid breads with seeds and grains) • Breakfast cereals well moistened with milk(b) • Soft pasta(a) and noodles • Rice (well cooked) • Soft pastry, for example quiche with a pastry base • Other, soft, cooked grains	• Dry or crusty breads, breads with hard seeds or grains, hard pasty, pizza • Sandwiches that are not thoroughly moist • Course or hard breakfast cereals that do not moisten easily, for example toasted muesli, bran cereals • Cereals with nuts, seeds and dried fruit
Vegetables, legumes	• Well cooked vegetables (a) served in small pieces or soft enough to be mashed or broken up with a fork • Soft canned vegetables, for example peas • Well cooked legumes (the outer skin must be soft), for example baked beans	• All raw vegetables (including chopped and shredded) • Hard, fibrous or stringy vegetables and legumes, for example sweet corn, broccoli stalks
Fruit	• Fresh fruit pieces that are naturally soft, for example banana, well-ripened pawpaw • Stewed and canned fruits in small pieces • Pureed fruit	• Large/round fruit pieces that pose a choking risk, for example whole grapes, cherries • Dried fruit, seeds and fruit peel • Fibrous fruits, for example pineapple
Milk, yoghurt, cheese	• Milk, milkshakes, smoothies (b) • Yoghurt (may contain soft fruit)(b) • Soft cheeses, (a) for example Camembert, ricotta	• Yoghurt with seeds, nuts, muesli or hard pieces of fruit • Hard cheeses, for example cheddar and hardened/ crispy cooked cheese
Meat, fish, poultry, eggs, nuts, legumes	• Casseroles with small pieces of tender meat(a) • Moist fish (easily broken up with the edge of a fork) • Eggs (a) (all types except fried) • Well cooked legumes (the outer skin must be soft), for example baked beans • Soft tofu, for example small pieces, crumbled	• Dry, tough, chewy, or crispy meats • Meat with gristle • Fried eggs • Hard or fibrous legumes • Pizza
Desserts	• Puddings, dairy desserts, (b) custards, (b) yoghurt (b) and ice-cream (b) (may have pieces of soft fruit) • Moist cakes (extra moisture, e.g., custard may be required) • Soft fruit-based desserts without hard bases, crumbly or flaky pastry or coconut, for example apple crumble • Creamed rice, moist bread and butter pudding	• Dry cakes, pastry, nuts, seeds, coconut, dried fruit, pineapple
Miscellaneous	• Soup (b) – (may contain small soft lumps, e.g., pasta) • Soft fruit jellies or nonchewy lollies(a) • Soft, smooth, chocolate • Jams and condiments without seeds or dried fruit	• Soups with large pieces of meats or vegetables, corn, or rice • Sticky or chewy foods, for example toffee • Popcorn, chips, biscuits, crackers, nuts, edible seeds

(a) These foods require case-by-case consideration.

(b) These foods may need modification for individuals requiring thickened fluids.

Source: Reproduced with permission from Australian Standardised Terminology and Definitions for Texture modified foods and fluids (2007) *Nutrition & Dietetics* 64 (Suppl. 2): S53–S76.

towards, the unaffected side; posterior positioning of food on the tongue may promote oral transport. If pharyngeal transport/reflex swallowing is impaired, it is important to ensure upright posture, head stable in the midline with a slight forward flexion to protect the airway; prompting of synchronization of the sequence of inspiration, breath-holding, swallowing, possibly repeatedly, expiration, and coughing on expiration to clear food debris may be helpful. Maintaining an upright posture for at least 30 min after meals minimizes risk of regurgitation/aspiration.

Many nondietary, therapeutic approaches to manage dysphagia have been identified including: oral electrical, thermal and chemical pharyngeal stimulation; high-intensity swallowing therapy; exercises to improve laryngeal closure, labial/mandibular closure, tongue elevation, and lateralization; use of palatal training devices and prostheses to assist triggering of the swallowing reflex or lower the palatal vault to improve bolus formation; drug therapy (nifedipine); use of biofeedback involving mirrors and VSS. Although benefits have been described for many of these interventions, lack of randomized, controlled clinical trials with adequate power limits conclusions on effectiveness.

Table 6 Texture B—Minced and Moist: Australian standardised terminology and definitions for texture modified foods and fluids 2007

Food type	*Recommend*	*Avoid (in addition to the Foods to Avoid listed for Texture A – Soft)*
Bread, cereals, rice, pasta, noodles	• Breakfast cereal with small moist lumps, for example porridge or wheat flake biscuits soaked in milk • Gelled bread • Small, moist pieces of soft pasta, for example moist macaroni cheese (some pasta dishes may require blending or mashing)	• All breads, sandwiches, pastries, crackers, and dry biscuits • Gelled breads that are not soaked through the entire food portion • Rice that does not hold together, for example parboiled, long-grain, basmati • Crispy or dry pasta, for example edges of a pasta bake or lasagne
Vegetables, legumes	• Tender cooked vegetables that are easily mashed with a fork • Well cooked legumes (partially mashed or blended)	• Vegetable pieces larger than 0.5 cm or too hard to be mashed with a fork • Fibrous vegetables that require chewing, for example peas
Fruit	• Mashed soft fresh fruits, for example banana, mango • Finely diced soft pieces of canned or stewed fruit • Pureed fruit • Fruit juice(a)	• Fruit pieces larger than 0.5 cm • Fruit that is too hard to be mashed with a fork
Milk, yoghurt, cheese	• Milk, milkshakes, smoothies(a) • Yoghurt(a) (may have small soft fruit pieces) • Very soft cheeses with small lumps, for example cottage cheese	• Soft cheese that is sticky or chewy, for example Camembert
Meat, fish, poultry, eggs, nuts, legumes	• Coarsely minced, tender, meats with a sauce. • Casseroles dishes may be blended to reduce the particle size • Coarsely blended or mashed fish with a sauce • Very soft and moist egg dishes, for example scrambled eggs, soft quiches • Well cooked legumes (partially mashed or blended) • Soft tofu, for example small soft pieces or crumbled	• Casserole or mince dishes with hard or fibrous particles, for example peas, onion • Dry, tough, chewy, or crispy egg dishes or those that cannot be easily mashed
Desserts	• Smooth puddings, dairy desserts,(a) custards,(a) yoghurt(a) and ice-cream(a) (may have small pieces of soft fruit) • Soft moist sponge cake desserts with lots of custard, cream or ice-cream, for example trifle, tiramisu • Soft fruit-based desserts without hard bases, crumbly or flaky pastry or coconut, for example apple crumble with custard • Creamed rice	• Desserts with large, hard or fibrous fruit particles (e.g., sultanas), seeds or coconut • Pastry and hard crumble • Bread-based puddings
Miscellaneous	• Soup(a) – (may contain small soft lumps, e.g., pasta) • Plain biscuits dunked in hot tea or coffee and completely saturated • Salsa's, sauces and dips with small soft lumps • Very soft, smooth, chocolate • Jams and condiments without seeds or dried fruit	• Soups with large pieces of meats or vegetables, corn, or rice • Sweets and lollies including fruit jellies and marshmallow

(a) These foods may require modification for individuals requiring thickened fluids.

Source: Reproduced with permission from Australian Standardised Terminology and Definitions for Texture modified foods and fluids (2007) *Nutrition and Dietetics* 64 (Suppl. 2): S53–S76.

Nutrient Requirements

Diabetes mellitus, hypertension, and renal failure are common in the stroke population; their disease-related dietary management requirements need to be borne in mind for affected stroke patients. Many different approaches can be used to estimate energy requirements. Resting energy expenditure can be estimated on the basis of body weight, height, age, and sex, with modifications to accommodate activity and injury factors, and supplemental values to replenish malnourished individuals. Estimation of resting energy expenditure by indirect calorimetry using a portable metabolic monitor provides more accurate estimates derived from the respiratory quotient. Values do not consider periods of activity, pyrexia, pain, or energy increments necessary for nutritional repletion, so further corrections are necessary.

Following stroke a number of factors may affect energy requirements. Inactivity caused by paralysis reduces energy expenditure, but muscular paresis increases the relative energy cost of activity. Infections, which commonly complicate stroke, will increase

Table 7 Texture C—Smooth Pureed: Australian standardised terminology and definitions for texture modified foods and fluids 2007

Food type	*Recommend*	*Avoid (in addition to the Foods to Avoid listed for Texture B – Minced and moist)*
Bread, cereals, rice, pasta, noodles	• Smooth lump-free breakfast cereals, for example semolina, pureed porridge • Gelled bread • Pureed pasta or noodles • Pureed rice	• Cereals with course lumps or fibrous particles, for example all dry cereals, porridge • Gelled breads that are not soaked through the entire food portion
Vegetables, legumes	• Pureed vegetables • Mashed potato • Pureed legumes, for example baked beans (ensuring no husks in final puree) • Vegetable soups that have been blended or strained to remove lumps(a)	• Coarsely mashed vegetables • Particles of vegetable fiber or hard skin
Fruit	• Pureed fruits, for example commercial pureed fruits, vitamised fresh fruits • Well mashed banana • Fruit Juice(a) without pulp	• Pureed fruit with visible lumps
Milk, yoghurt, cheese	• Milk, milkshakes, smoothies(a) • Yoghurt(a) (lump-free), for example plain or vanilla • Smooth cheese pastes, for example smooth ricotta • Cheese and milk-based sauces(a)	• All solid and semisolid cheese including cottage cheese
Meat, fish, poultry, eggs, nuts, legumes	• Pureed meat/fish (pureed with sauce/gravy to achieve a thick moist texture) • Soufflés and mousses, for example salmon mousse • Pureed legumes, hummus • Soft silken tofu • Pureed scrambled eggs	• Minced or partially pureed meats • Scrambled eggs that have not been pureed • Sticky or very cohesive foods, for example peanut butter
Desserts	• Smooth puddings, dairy desserts,(a) custards,(a) yoghurt(a) and ice-cream(a) • Gelled cakes or cake slurry, for example fine sponge cake saturated with jelly • Soft meringue • Cream(a), sirup dessert toppings(a)	• Desserts with fruit pieces, seeds, nuts, crumble, pastry or nonpureed garnishes • Gelled cakes or cake slurries that are not soaked through the entire food portion
Miscellaneous	• Soup(a) – vitamised or strained to remove lumps • Smooth jams, condiments and sauces	• Soup with lumps • Jams and condiments with seeds, pulps or lumps

(a) These foods may require modification for individuals requiring thickened fluids.

Source: Reproduced with permission from Australian Standardised Terminology and Definitions for Texture modified foods and fluids (2007) *Nutrition & Dietetics* 64 (Suppl. 2): S53–S76.

energy expenditure, with each 1 °C rise in core temperature raising it by 13%. The impact of the cerebral injury on poststroke metabolism, i.e., resting energy expenditure at different levels of stroke severity, has not been fully investigated. Evidence for metabolic injury responses based on hormonal profiles and changes in blood glucose concentration in the acute phase is limited. Hyperglycemia is common following stroke and has been associated with increased morbidity and mortality. Hyperglycemia can be attributed to overt or latent diabetes mellitus, stress responses, and effects of glucose intolerance in elderly subjects. Elevated plasma cortisol and catecholamine concentrations representing a transient stress response have been reported in the first 72 h following stroke. Nitrogen requirements following stroke can be estimated using reference ranges based on body weight, or using nitrogen balance studies.

Studies of nutritional supplementation of stroke patients have not conclusively demonstrated improved outcomes. One major international trial clearly showed no benefit from isocaloric supplements for unselected stroke populations, but early indications are that targeted protein-rich supplementation may be useful.

Fluid balance requirements require careful attention, because dehydration is a serious risk in individuals with dysphagia and physical disabilities that impair drinking, with potential to compromise already impaired cerebral circulation. Oral intake of fluid is contraindicated where the swallowing and gag reflex are lost or swallowing and level of consciousness are impaired. Parenteral (intravenous or other routes) fluid replacement therapy is then necessary, usually short term. Fluid requirements can be calculated on the basis of 35 ml per kg body weight daily in adults, but sepsis and fever can increase needs. Fluid intake and output in conjunction with insensible losses should be monitored on a daily basis together with the symptoms of dehydration, i.e., urine specific gravity, thirst, dry mucous membranes, and loss of skin turgor.

Table 8 Compensatory strategies and restorative therapies for dysphagia

Stage of swallow	*Swallow disorder*	*Compensatory strategy*	*Restorative/rehabilitative exercises/therapy*
Oral preparatory	Poor lip seal	Supported lip and jaw closure	Lip exercises
Oral preparatory	Poor cheek tone	Intra-oral prosthesis, cheek hold technique (apply pressure to weak side), tilt head towards unaffected side	Cheek tone exercises
Oral preparatory	Poor sensation in oral cavity	Increase bolus taste, volume, density, temperature, carbonated drinks	Sensory awareness program
Oral preparatory	Poor tongue movement	Modify consistency of bolus, pace rate of bolus presentation, avoid mixed consistencies, remove residue from oral cavity post swallow	Tongue lateralization exercises
Oral preparatory	Poor chewing/ jaw closure	Jaw support, diet modification	Chewing exercises
Pharyngeal	Delayed swallow	Adapted cutlery and crockery to assist in self feeding, chin tuck posture, increase bolus taste, volume, density, temperature, fizzy drinks	Thermal stimulation, PNF to the fauceal arches
Pharyngeal	Reduced base of tongue movement	Chin tuck, clearing swallows, effortful swallow, decrease bolus size, increase bolus consistency	Tongue hold technique, gargle & yawn exercises, supersupraglottic swallow
Pharyngeal	Unilateral pharyngeal paresis	Head rotation to damaged side, head tilt to unaffected side, back or side lying, clearing swallows, liquid wash down	
Pharyngeal	Unilateral tongue and pharyngeal paresis	Head tilt to unaffected side, clearing swallows	
Pharyngeal	Reduced laryngeal closure	Chin tuck, head rotation to damaged side, supraglottic swallow, supersupraglottic swallow, alter bolus consistency	Supraglottic swallow, supersupraglottic swallow, breath hold maneuver, push–pull voicing
Pharyngeal	Reduced laryngeal elevation	Chin tuck & lie on side/back, supersupraglottic swallow, Mendelssohn maneuver, clearing swallows	Falsetto voicing, Mendelssohn maneuver, Shaker technique, surface electromyography
Pharyngeal	Cricopharyngeal dysfunction/reduced anterior movement of hyolaryngeal structure	Head rotation, avoid mixed consistencies	Shaker technique
Fatigue		Nutritional supplements, decrease meal size, increase frequency of meals	

PNF: proprioceptive neuromuscular facilitation.

Source: Reproduced with permission from Perry L and Boaden E (2010) Nutritional aspects of stroke care. In: Williams J, Perry L, and Watkins (Eds) *Acute Stroke Nursing*. London: Wiley–Blackwell.

Artificial Nutritional Support

The presence of severe dysphagia and cognitive and complex physical impairments may render oral feeding unsafe or insufficient for nutritional requirements. If the gastrointestinal tract is functional, the options for delivering enteral nutritional support are either a fine-bore NGT or a PEG or radiologically inserted gastrostomy.

Decisions concerning choice of route are influenced by the anticipated duration of dysphagia and benefits versus risks. Impact on nutritional status, rehabilitation, quality of life, safety, tolerance, flexibility, ease of use, costs of insertion, removal, and maintenance are all important considerations. In the majority of cases dysphagia resolves within the acute phase of stroke, i.e., approximately 2–3 weeks. For relatively short time periods, enteral feeding by fine-bore NGT is recommended, and a major international trial of early insertion of PEG has not shown benefit. Advantages of NGT are the technical simplicity of intubation, maintenance and removal, low cost and ease of use. Accidental displacement and repeated removal by confused or cognitively impaired patients are not uncommon; nasal bridles to maintain secure positioning are showing promise. Placement positioning difficulties and malposition, discomfort, and risks of aspiration are potential complications, with aspiration occurring in up to 10% of patients.

For feeding over longer periods, a PEG inserted under local anesthetic and sedation offers greater comfort, toleration, ease of use, and reported improvements in nutritional status. However, costs are greater and this more invasive procedure carries a 1–2% technique-related fatality, often more for delayed insertion in malnourished patients. Minor complications include stomal sepsis, leakage, and blockage. Peritonitis, perforation, gastrointestinal bleeding, and intestinal obstruction can rarely occur. Most enterostomy catheters are made from nonacid-hardening polyurethane or silicone and can be left *in situ* for 6 months and longer.

No consensus exists concerning the time period within which gastrostomy feeding should be initiated following stroke, but it should be considered where dysphagia persists beyond 14 days, or the patient cannot tolerate a NGT. Rarely, enteral nutrition may be contraindicated following stroke owing to gastrointestinal bleeding resulting from severe stress ulceration; nonstroke-related contraindications may also be present, e.g., ascites, bleeding disorders. Parenteral nutrition should then be considered.

Issues relating to the optimum timing of commencement of artificial feeding can be challenging for a patient group that may experience sudden and major cognitive and physical impairments with uncertain recovery potential. Nutritional deterioration can be rapid in acutely ill patients unable to take food orally. Decision-making can be difficult and stressful, and require extensive and sensitive communication between healthcare team members, patient (where possible) and families (where available). Failure to establish timely agreed nutritional care planning for patients who have eaten little or nothing for more than 5 days and are likely to eat little or nothing for the next 5 days or longer, who are unable to take in nutrients properly, or who have increased nutritional needs is unacceptable.

Evaluation of Nutritional Support

It is vital that nutritional status is monitored in the acute phase of recovery and that dietary intakes are readjusted accordingly. Appropriate dietary, anthropometric, and clinical assessments, which can be performed on a weekly basis, are discussed. Other important components of monitoring include recovery of physical functions related to independence in eating, including swallowing capacity, and observing for complications of enteral support techniques. Effective nutritional management following stroke requires coordination of the professional skills of doctor, nurse, speech pathologist, dietitian, occupational therapist, and physiotherapist, ideally within the context of a nutrition support team. Dynamic leadership, referral policies, clear accountability and lines of communication are essential for the team to deliver effective support. Follow-up services in the community are also necessary to prevent deterioration in nutritional status in the later stages of rehabilitation.

Further Reading

Finestone, H.M., Greene-Finestone, L.S., 2003. Rehabilitation medicine: 2. Diagnosis of dysphagia and its nutritional management for stroke patients. Canadian Medical Association Journal 169 (10), 1041–1050.

Green, C.J., 1999. Existence, causes and consequences of hospital and community malnutrition; clinical and financial benefits of nutritional intervention. Clinical Nutrition 18 (supplement 2), 3–28.

Jönsson, A.C., Lindgren, I., Norrving, B, Lindgren, A., 2008. Weight loss after stroke: A population-based study from the Lund Stroke Register. Stroke 39, 918–923.

Miller A (2004) *The Neuroscientific Principles of Swallowing and Dysphagia*. Singular Publications.

National Collaborating Centre for Acute Care (2006) *Nutrition Support in Adults Oral Nutrition Support, Enteral Tube Feeding and Parenteral Nutrition*. National Collaborating Centre for Acute Care, London. At http://www.rcseng.ac.uk/new_rcseng/content/publications/docs/nutrition_support_guidelines.html

National Stroke Foundation (2010) *Clinical Guidelines for Stroke Management 2010*. Melbourne Australia. At http://www.strokefoundation.com.au/clinical-guidelines

Perry, L., McLaren, S., 2003. Eating difficulties after stroke. Journal of Advanced Nursing 43 (4), 360–369.

Perry, L., McLaren, S., 2004. An exploration of nutrition and eating disabilities in relation to quality of life at 6 months post-stroke. Health and Social Care in the Community 12 (4), 288–297.

Rabadi, M.H., Coar, P.L., Lukin, M., Lesser, M., Blass, J.P., 2008. Intensive nutritional supplements can improve outcomes in stroke rehabilitation. Neurology 71, 1856–1861.

Royal College of Physicians and British Society of Gastroenterology (2010) *Oral Feeding Difficulties and Dilemmas: A Guide to Practical Care, Particularly Towards the End of Life*. London: Royal College of Physicians. At https://bookshop.rcplondon.ac.uk/contents/68e931b6-e3bf-41bd-99a4-3900569d4973.pdf

The role of nutrition in inflammatory bowel disease: Disease associations, management of active disease and maintenance of remission

Gabrielle Wark[a,b], **Dorit Samocha-Bonet**[a,c], **Simon Ghaly**[a,b], **and Mark Danta**[a,b], [a] St Vincent's Clinical School, UNSW, Sydney, NSW, Australia; [b] Department of Gastroenterology and Hepatology, St Vincent's Hospital, Sydney, NSW, Australia; and [c] Garvan Institute of Medical Research, Sydney, NSW, Australia

© 2023 Elsevier Ltd. All rights reserved.

Introduction **671**
Inflammatory bowel disease (IBD) background **672**
Pathogenic factors in IBD **672**
Genetic factors in IBD 672
Microbiome in IBD 673
Immunological abnormalities 673
Environmental influences 673
Nutrition, diet and IBD pathogenesis **673**
Breastfeeding 675
Dietary fiber 675
Sugar 678
FODMAPs, wheat and gluten 678
Fat 680
Red meat 682
Food additives 683
Emulsifiers 683
Nanoparticles 684
Malnutrition assessment and management considerations in IBD **684**
Macronutrient deficiencies 684
Micronutrient deficiencies 684
Iron 686
Vitamin B12 686
Folate 686
Calcium 686
Vitamin D 688
Zinc 688
Diet as therapy for active disease in IBD **688**
Exclusive Enteral Nutrition (EEN) 688
Pediatric cohort 690
Adult cohort 690
Complicated CD: fistulizing and stricturing disease 690
Use of EEN in UC 690
Differences between EN formulas 690
EEN key points 691
Partial Enteral Nutrition (PEN) 691
PEN with restricted diet- Crohn's disease exclusion diet (CDED) 691
PEN key points 692
Targeted CDED 692
Other elimination diets 692
Elimination diets key points 692
Future directions and conclusions **692**
References **693**

Key points

- Inflammatory Bowel Disease (IBD): Background
- Pathogenic factors in IBD
- Nutrition, diet and IBD pathogenesis
- Malnutrition assessment and management considerations in IBD

- Diet as prescribed therapy for active disease in IBD
- Future directions and conclusions

Glossary

5- ASA Aminosalicylic Acid
AI Autoimmune
AID Anti-Inflammatory Diet
AIP Autoimmune Protocol Diet
Anti-TNF Anti-Tumor Necrosis Factor
APD Autoimmune Protocol Diet
ASUC Acute Severe Ulcerative Colitis
ATG 16L1 Autophagy Related-16-Like 1, a gene that codes for proteins involved in autophagy
BMD Bone Mineral Density
BMI Body Mass Index
CC-UK Crohn's & Colitis-UK
CD Crohn's Disease
CDAI Crohn's Disease Activity Index
CDED Crohn's Disease Exclusion Diet
CD-TREAT Crohn's Disease Treatment-with EATing diet; a solid food-based diet designed to replicate the nutrients and food ingredients composition of Exclusive Enteral Nutrition
CI Confidence Interval
CRP C-Reactive Protein
CS Corticosteroids
DDS Dextran Sulfate Sodium, a murine animal model of colitis
DHA Docosahexaenoic Acid
DNA Deoxyribonucleic Acid
ECCO European Crohn's and Colitis Organization
EEN Exclusive Enteral Nutrition
EN Enteral Nutrition
EIM Extra Intestinal Manifestation
EPIC European Prospective Investigation into Cancer
EPIC-IBD European Prospective Investigation into Cancer-Inflammatory Bowel Disease
ESPEN European Society for Clinical Nutrition and Metabolism
FA Fatty Acid
FC Fecal Calprotectin
FFQ Food Frequency Questionnaire
FODMAPs Fermentable Oligosaccharides, Disaccharides Monosaccharides and Polyols
GI Gastrointestinal
GWAS Genome Wide Association Study
HR Hazard Ratio
IBD Inflammatory Bowel Disease
IBD-AID Inflammatory Bowel Disease-Anti-Inflammatory Diet
IBS Irritable Bowel Syndrome
IECs Intestinal Epithelial Cells
IFN-gamma (IFA-γ) Interferon-gamma
Ig Immunoglobulin
IgA Immunoglobulin-A
IL-6 Interleukin- 6
IL-23R Interleukin-23 Receptor Gene
IL-1β Interleukin-1-Beta
IOF International Osteoporosis Foundation
IPAA Ileal Pouch Anal Anastomosis

IRR Incident Rate Ratio
JAK Janus Kinase, JAK inhibitors are a form of immune suppression used to treat IBD
LPS Lipopolysaccharide
MalX gene Amino Acid Sequence on *Escherichia coli* Facilitating Maltase Production
M-SHIME Mucosal Stimulated Human Intestinal Microbiota Ecosystem
MD Mediterranean Diet
MIRT Malnutrition Inflammation Risk Tool
MUFA Monounsaturated Fatty Acid
MUST Malnutrition Universal Screening Tool
n Number
n-3 Describes the position of the double bond on carbon 3, counting the methyl carbon as carbon number 1
n-6 Describes the position of the double bond on carbon 6, counting the methyl carbon as carbon 1
NAT Nutrition Assessment Tool
NCGS Non-Coeliac Gluten Sensitivity
NGT Naso-Gastric Tube
NHMRC National Health and Medical Research Council
$NLRP_3$ Nucleotide Leucine Rich Pyrin- Receptor-3
NOD2 Nucleotide Binding Oligomerization Domain Containing Protein -2
NRI Nutrition Risk Index
NST Nutrition Screening Tool
OR Odds Ratio
Oxford criteria in patients admitted to hospital with ASUC, at day 3 of IV corticosteroids those with stool frequency >8/day or stool frequency 3–8/day plus CRP >45 mg/L had a 85% chance of inpatient colectomy and would therefore benefit from rescue therapy
p p value, probability that an observed difference could have occurred just by random chance
p-80 Polysorbate 80
PCR Polymerase chain reaction
PEN Partial Enteral Nutrition
PN Parental Nutrition
PPGR Post-Prandial Glycemic Response
PUFA Polyunsaturated Fatty Acid
RCT Randomized Controlled Trial
RR Relative Risk
SBO Small Bowel Obstruction
SCD Specific Carbohydrate Diet
SCFAs Short Chain Fatty Acids
SGA Subjective Global Assessment
SK-IBD Saskatchewan Nutrition Assessment Tool
SOC Standard of Care
SVD Semi-Vegetarian Diet
TG Triglycerides
Th17 T-Cell Helper-17, a subset of lymphocytes implicated in IBD
TNF Tumor Necrosis Factor
UC Ulcerative Colitis
Ultra-processed foods Formulations of ingredients made by a series of industrial steps that require sophisticated equipment and technology. They are aimed at making the final product hyperpalatable and with a long shelf life.

Introduction

Inflammatory bowel disease (IBD) may cause significant symptoms and morbidity if left untreated, including malnutrition in the form of macro and micronutrient deficiencies. Current standard of care includes therapies designed to suppress the immune response, with inherent risk of infection and undesired side effects. In addition, IBD is often diagnosed in childhood and early adulthood, and the disease, as well as therapies used to manage it need to balanced during this important period of growth and fertility. A

treatment approach that focuses on dietary management is appealing to patients due to the favorable risk and side effect profile. In this article we will explore the associations between of diet and nutrition and the risk of developing IBD, the influence diet has on disease activity, common nutritional deficiencies and how these can be identified and managed. Lastly, proposed methods to manipulate the diet to manage active disease are discussed.

Inflammatory bowel disease (IBD) background

The Inflammatory Bowel Diseases (IBD), Ulcerative Colitis (UC) and Crohn's disease (CD), result in chronic relapsing and remitting inflammation of the gastrointestinal tract (Lamb et al., 2019). These diseases are thought to be caused by the interplay between genetic susceptibility, environmental factors, and intestinal microflora, resulting in an abnormal mucosal immune response and a compromised gut epithelial barrier function (Guan, 2019).

CD can affect all segments of the bowel, the most common being the terminal ileum and colon (Torres et al., 2017). The inflammation is typically segmental, asymmetrical, and transmural. UC, in contrast, results in inflammation limited to the mucosal layer, starting at the rectum and extending in a continuous fashion to proximal segments of the colon (Ungaro et al., 2017).

The clinical presentation of IBD relates to the areas of the gastrointestinal (GI) tract that are affected; but commonly includes abdominal pain, chronic diarrhea, and weight loss. Extraintestinal manifestations (EIM) including arthritis, skin rashes, eye disease such as episcleritis and anterior uveitis and metabolic bone disease; are not uncommon, with 50% of patients with IBD experiencing at least one EIM throughout the course of their illness. The EIM appears before the diagnosis of IBD 25% of the time (Harbord et al., 2015; Vavrika et al., 2015). The onset of IBD can occur at any age, but is most common between the ages of 10–40 years (Torres et al., 2017; Ungaro et al., 2017).

The diagnosis of IBD requires endoscopy and biopsy with characteristic histological features in a patient with suggestive clinical symptoms and where alternate diagnosis such as infection have been excluded. Due to the risk of significant complications and absence of a cure, appropriate long-term management to achieve and then maintain remission is necessary (Ungaro et al., 2017; Torres et al., 2017).

Current standard of medical care in IBD is determined by subtype, severity, and location in the gastrointestinal tract. In mild to moderate UC, Aminosalicylates are the first line standard treatment option, with immunosuppressants and biological agents reserved for those who do not respond, those who present with severe disease or those with a diagnosis of CD (Ungaro et al., 2017; Torres et al., 2017; Lamb et al., 2019; Raine et al., 2021). Immunosuppressive agents used as part of the standard treatment algorithm include anti-metabolites (Thiopurines and Methotrexate), anti-tumor necrosis factor alpha (anti TNF-α e.g., Infliximab, Adalimumab), $\alpha 4\beta 7$ ligand blocker Vedolizumab, the Interleukin (IL)-12 and IL-23 blocker Ustekinumab and Janus Kinase (JAK) inhibitors (Tofacitinib) (Raine et al., 2021; Torres et al., 2019). Aside from Aminosalicylates, the unifying mechanism of action of these medications is immune suppression, with inherent increased risk of infection as well as many other potential risks and side effects that are medication specific.

Surgery may be required in the setting of disease-related complications such as fistulae, perforation, or medically refractory disease. In CD most patients present with an inflammatory phenotype at diagnosis but, over time complications such as strictures, fistulae and abscesses may develop and up to half of patients require surgery at some point in the course of their disease. Surgery is less commonly required in UC, with a 10-year risk of 15.6%, but when required almost universally necessitates a total colectomy (Frolkis et al., 2013).

The pathogenesis of IBD is complex and associated with genetic susceptibility of the host, intestinal microbial dysbiosis and immunological abnormalities as well as other environmental triggering factors such as the diet (Guan, 2019).

This article will give an overview of the factors contributing to the pathogenesis of IBD before outlining the important aspects of nutrition in more detail. Specific to diet and IBD we will outline:

1. The proposed mechanisms of diet-microbial-immune system interactions that occur with important dietary components in the pathogenesis of IBD.
2. Established associations between elements of the diet and the development of IBD
3. Proposed assessment tools for malnutrition as well as micro and macro nutrient deficiencies
4. Dietary recommendations in the management of IBD for maintenance of remission and treatment of active disease.

An in-depth review of food supplements, probiotics and nutraceuticals is beyond the scope of this article.

Pathogenic factors in IBD

Genetic factors in IBD

At a genetic level, genome wide association studies (GWAS) have identified over 240 genetic risk loci for IBD (Park, 2019). Many of these genes play important roles in maintaining intestinal mucosal homeostasis, particularly involved in bacterial handling. Others code for epithelial barrier function, innate mucosal defense, immune regulation, cell migration, autophagy, and adaptive immunity. Examples of genes implicated in the development of IBD are NOD2, which codes for a protein involved in the innate immune response to bacterial peptidoglycan, ATG16L1, a gene that codes for a protein responsible for autophagy and bacterial clearance

and IL-23 Receptor (IL-23R) gene which is involved in the activation of adaptive immune responses. It should be noted that these are at-risk loci and only a small proportion of those who carry these genes develop IBD (Guan, 2019). Furthermore, twin studies demonstrate modest concordance rates in IBD, even among monozygotic twins, with rates in CD of 2.0–55.0% and 6.3–17.0% in UC (Gordon et al., 2015). Thus, while genetics play a role in the predisposition of a person to IBD, other environmental factors are needed for the development of clinical disease.

Microbiome in IBD

The microbiome, defined as the collection of microorganisms and their genes found in the host has become integral to our understanding of health and disease (Valdes et al., 2018). Dysbiosis, defined as an imbalance in the composition of the gut microbiome, is a hallmark of IBD. Evidence to support the role of the gut microbiota in IBD include the predisposition of inflammation for anatomical regions with relative fecal stasis (terminal ileum and rectum) and the effectiveness of fecal diversion as treatment for CD (Ni et al., 2017). The dysbiosis observed in IBD is characterized by loss of microbial diversity, particularly a loss of favorable species involved in the production of short chain fatty acids (SCFA), such as *Faecalibacterium prausnitzii* and Bacteroides, as well as an increase in invasive and adherent pathogenic species such as *Escherichia coli* (Zuo and Ng, 2018). Further studies have also demonstrated that there are differences in function of the gut microbes, as measured by characterizing metabolomic products of the microbiome; and these functional differences may be even more pronounced and influential than the changes in the microbial community structure (Franzosa et al., 2019; Santoru et al., 2017).

Immunological abnormalities

The immunological dysregulation that occurs in IBD encompasses abnormalities in both the innate and adaptive immune responses (Guan, 2019, Porter et al., 2020). It is characterized by epithelial damage, including abnormal mucus production and defective repair, expansion of inflammation driven by intestinal flora, and large numbers of inflammatory cells infiltrating the lamina propria of the gut mucosa. There is failure of immune regulation to control the inflammatory response and the activated lamina propria cells produce high levels of proinflammatory cytokines including TNF, IL-1β, IFN-γ and cytokines in the IL-23/Th17 pathway. This increase in proinflammatory cytokines is coupled with reduced tolerogenic responses, which results in the initiation and perpetuation of the inflammatory response.

It is postulated that the combination of dysbiosis, altered mucosal function and impaired mucosal barrier function found in IBD, results in mucosal inflammation.

Environmental influences

Epidemiological studies suggest a significant role of environmental factors in the pathogenesis of IBD. IBD is more prevalent in developed nations; however incidence rates are accelerating in more recently westernized countries such as Brazil and Taiwan (Ng et al., 2017). There is also a higher risk in first-generation immigrants from developing countries who have relocated to developed nations, suggesting that the typical western lifestyle, including diet and other environmental factors, may contribute to the development of IBD (M'Koma, 2013).

Environmental factors that have been identified as risk factors for the development of IBD include smoking, antibiotic use (particularly in childhood), vitamin D deficiency and oral contraceptive use (Piovani et al., 2019; Ungaro et al., 2014; Mahid et al., 2006; Cornish et al., 2008).

Physical activity, breastfeeding, high plasma vitamin D were found to be associated with a decreased risk in the development of IBD in a large umbrella meta-analysis (Piovani et al., 2019). Diet is another modifiable environmental factor that has an established association with the risk of developing IBD. It has gained particular interest due to its modifiable nature.

Nutrition, diet and IBD pathogenesis

Due to the risks associated with the current medical and surgical treatment options for IBD as well as the need for long term medical therapy there is substantial interest in the role of diet as a disease management strategy, but there is a variable quality of evidence to support clear recommendations (Sasson et al., 2021). Patient survey data demonstrates that implementation of a restrictive diet and avoidance of specific foods in the belief that it could prevent relapse was utilized by 68% of individuals with IBD; whist in another study 89% of IBD patients felt that dietary guidance in disease management was important. Only 16% of participants, however, felt they received adequate information on the recommended diet (Limdi et al., 2016; Wong et al., 2012).

Large epidemiological studies have established multiple dietary associations with the development of IBD (Ananthakrishnan et al., 2013, 2015; Shoda et al., 1996; John et al., 2010; Jantchou et al., 2010; Dong et al., 2020). Mechanistically, diet can impact IBD in several ways, including changing the structure of the microbiome; modifying gut microbial defenses; and directly modulating inflammatory responses. We further outline these interactions on an individual nutrient level (See **Table 1**) and a summary of the proposed positive and negative impacts are outlined in **Figs. 1** and **2**.

Table 1 Associations between dietary components and risk of developing IBD and proposed mechanisms.

		Risk of IBD development	
Dietary component	*Mechanism*	*CD*	*UC*
Breastfeeding	- Modulates microbiome (Azad et al., 2013) - Reduces infant infection and antibiotic use (IBD risk factor) (Parigi et al., 2015; Piovani et al., 2019) - Contains secretory IgA: strengthens mucosal barrier function (Rogier et al., 2014)	Reduced risk of developing IBD, dose response, with greatest reduction >12 months of breast feeding (Xu et al., 2017)	Reduced risk of developing IBD, dose response, with greatest reduction >12 months of breastfeeding (Xu et al., 2017)
Fiber	Fermented to SCFA by colonic bacteria, downstream effects include: - Primary energy source for colonocytes - Maintain integrity of protective mucus barrier - Inhibiting recruitment and activation of pro inflammatory cells - Increasing the number and activity of regulatory T cells (Prada Venegas et al., 2019; Goncalves et al., 2018)	Fiber, especially fruit derived decreased the risk of CD (Ananthakrishnan, et al., 2013, 2015; Andersen et al., 2018; Li, F. et al., 2015; Amre et al., 2007)	Fruit and vegetable consumption associated with reduced UC risk (Li et al., 2015a,b) Other studies have not shown a relationship with fiber intake and development of UC (Wang et al., 2017; Ananthakrishnan, et al., 2013, 2015)
Sugar	- Decreased thickness of colonic mucus layer - Increased gut permeability - Decreased microbial diversity - Decreased fecal SCFA (Laffin et al., 2019; Montrose et al., 2021)	Soft drink and refined sugar intake increase CD risk (Narula et al., 2021) Sugar intake had no impact on CD risk (Racine et al., 2015)	Soft drink and refined sugar intake increase UC risk (Narula et al., 2021) Sugar and soft drink intake as well as sucrose intake are associated with increased risk (Racine et al., 2015; Wang et al., 2017)
Fat	- Diet high in saturated fat increases sulphite-reducing pathogens (Devkota et al., 2012) - Diet high in animal fat (as well as sugar) resulted in dysbiosis (increased pro inflammatory organisms) and decreased favorable SCFA (Agus et al., 2015) - High animal fat diet decreased secretory IgA (Muhomah et al., 2019)	- Associations between total fat, animal fat and n-6 PUFA and incidence of CD (Shoda et al., 1996) - High n-3 PUFA and n-3:n-6 PUFA ratio associated with lower risk of CD (Amre et al., 2007)	- Consumption of the n-3 PUFA, DHA associated with reduced risk of UC (John et al., 2010) - Linoleic acid (a n-6 PUFA) associated with an increased risk of UC (Tjonneland et al., 2009) Recent meta-analysis showed no association between fat total or subtype intake and disease risk (Wang et al., 2016)
Red meat	- Altered microbiome - Increased intestinal permeability (Li et al., 2021)	Higher animal protein and red & processed meat intake associated with increased risk of IBD (Narula et al., 2018; Jantchou et al., 2010; Ge et al., 2015)	Higher animal protein and red & processed meat intake associated with increased risk of IBD (Jantchou et al., 2010; Dong et al., 2020; Ge et al., 2015; Narula et al., 2021)
Emulsifiers and nanoparticles	- Reduce microbial diversity - Reduce mucus thickness - Increase bacterial encroachment on the epithelium - Increasing intestinal permeability (Bancil et al., 2021; Ruiz et al., 2017)	Association between emulsifiers and ultra-processed food intake and development of CD (Narula et al., 2021; Roberts et al., 2013)	Association between ultra-processed food intake and development of UC (Narula et al., 2021)

CD: Crohn's Disease, DHA: Docosahexaenoic Acid, IBD: Inflammatory Bowel Disease, IgA: Immunoglobulin-A, n-3: Describes the position of the double bond on carbon 3, counting the methyl carbon as carbon number 1, n-6: Describes the position of the double bond on carbon 6, counting the methyl carbon as carbon 1, SCFAs: Short Chain Fatty Acids, UC: Ulcerative Colitis.

Agus, A., Denizot, J., Thevenot, J., et al., 2015. Western diet induces a shift in microbiota composition enhancing susceptibility to adherent-invasive *E. Coli* infection and intestinal inflammation. Sci. Rep. 6; Amre, D., D'Souza, S., Morgan, K., 2007. Imbalances in dietary consumption of fatty acids, vegetables and fruits are associated with risk for Crohn's Disease in children. Am. J. Gastroenterol. 102, 2016–2025; Ananthakrishnan, A., Khalili, H., Konijeti, G., et al., 2013. A prospective study of long-term intake of dietary fiber and risk of Crohn's disease and ulcerative colitis. Gastroenterology 145, 970–977; Ananthakrishnan, A., Khalili, H., Song, M., et al., 2015. High school diet and the risk of Crohn's disease and ulcerative colitis. Inflamm. Bowel Dis. 21, 2311–2319; Andersen, V., Chan, S., Luben, R., et al., 2018. Fiber intake and the development of inflammatory bowel disease: a European prospective multicentre cohort study (EPIC-IBD). J. Crohn's Colitis 24, 129–136; Azad, M., Konya, T., Maughan, H., et al., 2013. Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. Can. Med. Assoc. J. 185, 385–394; Bancil, A., Sandall, A., Rossi, M., et al., 2021. Food additive emulsifiers and their impact on gut microbiome, permeability and inflammation: mechanistic insights in inflammatory bowel disease. J. Crohn's Colitis 15, 1068–1079; Devkota, S., Wang, Y., Musch, M., et al., 2012. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice. Nat. Microbiol. 487; Dong, C., Mahamat-Saleh, Y., Racine, A., et al., 2020. OP17: protein intakes and risk of inflammatory bowel disease in the European Prospective Invesigation into Cancer and Nutrition cohort (EPIC-IBD). 15th Congress of ECCO. vol. 14; Ge, J., Han, T., Liu, J., et al., 2015. Meat intake and risk of inflammatory bowel disease: a meta-analysis. Turkish J. Gastroenterol. 26, 492–7; Goncalves, P., Ricardo Araujo, J., Di Santo, J.,

et al., 2018. A cross-talk between microbiota derived short chain fatty acids and the host mucosal immune system regulates intestinal homeostasis and inflammatory bowel disease. Inflamm. Bowel Dis. 24, 558–572; Jantchou, P., Morois, S., Clavel-Chapelon, F., et al., 2010. Animal protein intake and risk of inflammatory bowel disease: the E3N prospective study. Am. J. Gastroenterol. 105, 2195–2201; John, S., Luben, R., Shrestha, S., et al., 2010. Dietary n-3 polyunsaturated fatty acids and the etiology of ulcerative colitis: a UK prospective cohort study. Eur. J. Gastroenterol. Hepatol. 22, 602–6; Laffin, M., Fedorak, R., Zalaski, A., et al., 2019. A high-sugar diet rapidly enhances susceptibility to colitis via depletion of luminal short-chain fatty acids in mice. Sci. Rep. 9; Li, D., Cui, M., Tan, F., et al., 2021. High red meat intake exacerbates dextran sulfate induced colitis by altering gut microbiota in mice. Front. Nutr. 8; Li, F., Liu, X., Wang, W., et al., 2015. Consumption of vegetables and fruit and the risk of inflammatory bowel disease: a meta-analysis. Eur. J. Gastroenterol. Hepatol. 27, 623–630; Montrose, D., Nishiguchi, R., Basu, S., et al., 2021. Dietary fructose alters the composition, localization and metabolism of gut microbiota in association with worsening colitis. Cell. Mol. Gastroenterol. Hepatol. 11, 525–550; Muhomah, T., Nishino, N., Katsumata, E., et al., 2019. High-fat diet reduces the level of secretory immunoglobilin A coating of commensal gut microbiota. Biosci. of Microb. Food Health 38, 55–64; Narula, N., Dhillon, A., Zhang, D., et al., 2018. Enteral nutritional therapy for induction of remission in crohn's disease. Cochrane Database Syst. Rev. 1; Narula, N., Wong, E., Dehghan, M., et al., 2021. Association of ultra-processed food intake with risk of inflammatory bowel disease: prospective cohort study. Br. Med. J. 374; Parigi, S., Eldh, M., Larssen, P., et al., 2015. Breast milk and solid food shaping intestinal immunity. Front. Immunol. 6; Piovani, D., Danese, S., Peyrin-Biroulet, L., et al., 2019. Environmental factors for inflammatory bowel diseases: an umbrella review of meta-analysis. Gastroenterology 157, 647–659; Prada Venegas, D., De la Fuente, M., Landskron, G., et al., 2019. Short Chain Fatty Acids (SCFAs)- mediated gut epithelial and immune regulation and its relevance for inflamamtion. Front. Immunol. 10; Racine, A., Carbonnel, F., Chan, S., et al., 2015. Dietary patterns and risk of inflammatory bowel disease in Europe: results from the EPIC study. Inflamm. Bowel Dis. 22, 345–354; Roberts, C., Rushworth, S., Richman, E., et al., 2013. Hypothesis: Increased consumption of emulsifiers as an explanation for the rising incidence of Crohn's disease. J. Crohn's Colitis 7, 338–341; Rogier, E., Frantz, A., Bruno, M., et al., 2014. Lessons from mother: long-term impact of antibodies in breast milk on the gut microbiota and intestinal immune system of breastfed offspring. Gut Microbes 5, 663–668; Ruiz, P., Moron, B., Becker, H., et al., 2017. Titanium dioxide nanoparticles exacerbate DSS-induced colitis: role of the NLRP3 inflammasome. Gut 66, 1216–1224; Shoda, R., Matseuda, K., Yamato, S., et al., 1996. Epidemiologic analysis of Crohn disease in Japan: increased dietary intake of n-6 polyunsaturated fatty acids and animal protein relates to the increased incidence of Crohn disease in Japan. Am. J. Clin. Nutr. 63, 741–5; Tjonneland, A., Overvad, K., Bergmann, M., et al., 2009. Linoleic acid, a dietary n-6 polyunsaturated fatty acid, and the etiology of ulcerative colitis: a nexted case-control study within a European prospective cohort study. Gut 58, 1606–11; Wang, F., Feng, J., Gao, Q., et al., 2017. Carbohydrate and protein intake and the risk of ulcerative colitis: Systematic review and dose-response meta-analysis of epidemiological studies. Clin. Nutr. 36, 1259–1265; Wang, F., Lin, X., Zhao, Q., 2016. Fat intake and risk of ulcerative colitis: systematic review and dose-response meta-analysis of epidemiological studies. J. Gastroenterol. Hepatol. 32, 19–27; Xu, L., Lochhead, P., Ko, Y., et al., 2017. Systematic review with meta-analysis: breastfeeding and the risk of Crohn's disease and ulcerative colitis. Aliment. Pharmacol. Therapeut. 46, 780–789.

Breastfeeding

Environmental exposures during the first few years of life may influence the risk of disease development, particularly that of allergy and asthma (Cole and Ownby, 2017). This is hypothesized to be due to alteration in the gut microbiota of children, which is important for the development of immune regulatory responses, so called "microbial-immune" cross talk. Stable microbial structure is reached by age 3–4 years, and breastfeeding is the most significant factor associated with microbiome structure (Stewart et al., 2018). Breastfeeding exerts a strong influence on intestinal microbial composition of the infant, with different abundances of taxa, particularly lower in pathogens such as *Clostridium difficile* and Escherichia-Shigella identified in breastfed infants when compared to those who were formula fed (Azad et al., 2013).

Breastmilk contains multiple immune modulating compounds, including secretory IgA which strengthens the intestinal barrier and regulates the commensal gut microbiota, promoting a healthy balance between pro and anti-inflammatory gene expression later in life (Rogier et al., 2014; M'Rabet et al., 2008). This protects the infant from the first external pathogenic insults, which is important to reduce neonatal infection and subsequent antibiotic use, which is an established risk factor for IBD. In addition, breastmilk promotes immune tolerance to commensal bacteria, which can avoid undesired inflammatory responses against these innocuous antigens. This regulatory response is missing in IBD and manifests in dysregulated immune responses (Parigi et al., 2015; Piovani et al., 2019).

A large meta-analysis has established the association between breastfeeding and lower risk of development of both UC and CD across many ethnic groups. Breastfeeding duration has a dose-dependent association, with the greatest reduction in risk observed in infants who were breastfed for at least 12-months (Xu et al., 2017). See **Table 1** for summary of proposed mechanisms of influence of dietary components on gastrointestinal inflammation and epidemiological studies to support associations with IBD risk.

Dietary fiber

Dietary fiber provides a substrate for the bacteria that inhabit the distal gut, and these bacteria are integral to its handling. While human intestinal epithelial cells produce approximately 17 enzymes to digest fiber, bacteria produce thousands of complementary enzymes to depolymerize and ferment dietary polysaccharides to form host absorbable SCFAs such as butyrate, propionate and acetate. These SCFAs have an array of functions that are significant to the maintenance of gastrointestinal health, including provision of the primary energy source for colonocytes; inhibiting recruitment and pro-inflammatory activity of neutrophils, macrophages, dendritic cells and effector T-cells; and increasing the number and activity of regulatory T-cells (Prada Venegas et al., 2019; Goncalves et al., 2018). In IBD, bacteria that possess this capability from the Firmicutes phylum, such as Faecalibacterium prausnitzii, are less abundant, leading to lower stool SCFA concentration compared with healthy controls, which may contribute to increased intestinal inflammation (Sokol et al., 2009; Takaishi et al., 2008).

Studies in animal models have demonstrated the protective effect of fiber when given before the development of colitis (Silveira et al., 2017). In the absence of fiber, the intestinal microbiota in mice resorted to consuming the mucus layer glycoproteins as a nutrient source, leading to erosion of this barrier of defense against enteric pathogens (Desai et al., 2016). It is therefore plausible

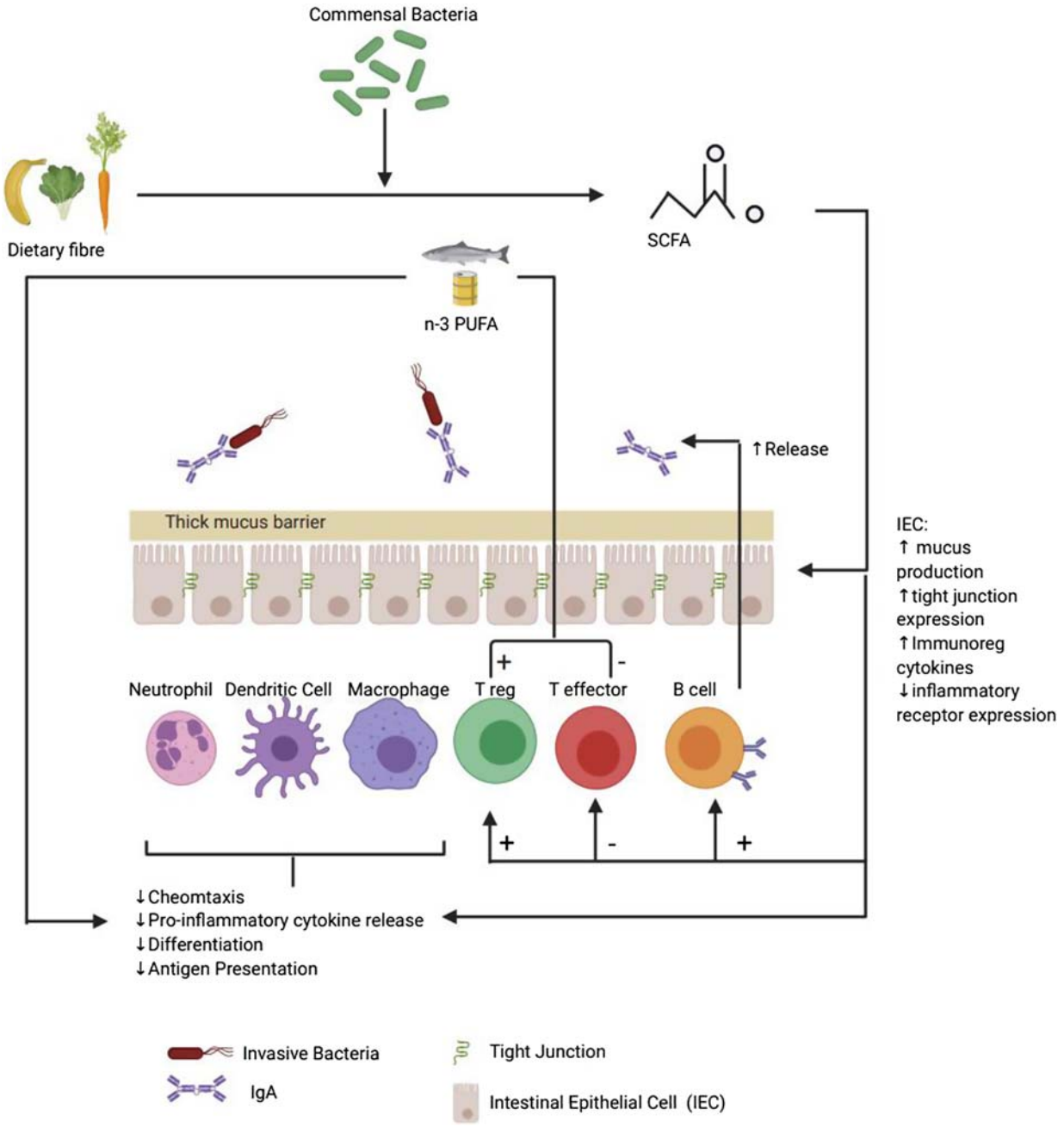


Fig. 1 Schematic representation of proposed mechanism of favorable diet-gut-immune interactions that may protect from the development of IBD and improve disease activity. (Created with BioRender). PUFA: Polyunsaturated Fatty Acid, SCFA: short-chain fatty acid, IBD: Inflammatory Bowel Disease, Ig: Immunoglobulin. Reproduced from original article: Wark et al. (2021).

that fiber might help to prevent IBD, or once inflammation is established it may help to restore, and then maintain, intestinal integrity (Gill et al., 2021).

Large cohort studies and several systematic reviews have explored the relationship between the consumption of fiber, including the source of fiber, and the risk of developing IBD. The Nurses' Health Study prospectively followed up 170,776 women for 26 years and demonstrated that women with an intake of fruit-derived fiber in the highest quartile (median 24.3 g/day) was associated with a 40% reduction in risk of CD (multivariate HR 0.59, CI 0.39–0.90) compared with women in the lowest quartile of fiber consumption (median 11.6 g/day). This apparent risk reduction appeared greatest for fiber derived from fruits, whereas fiber from cereals, whole grains or legumes did not modify the risk. In contrast, neither total intake of dietary fiber nor intake of fiber from specific sources modified the risk of UC (Ananthakrishnan et al., 2013).

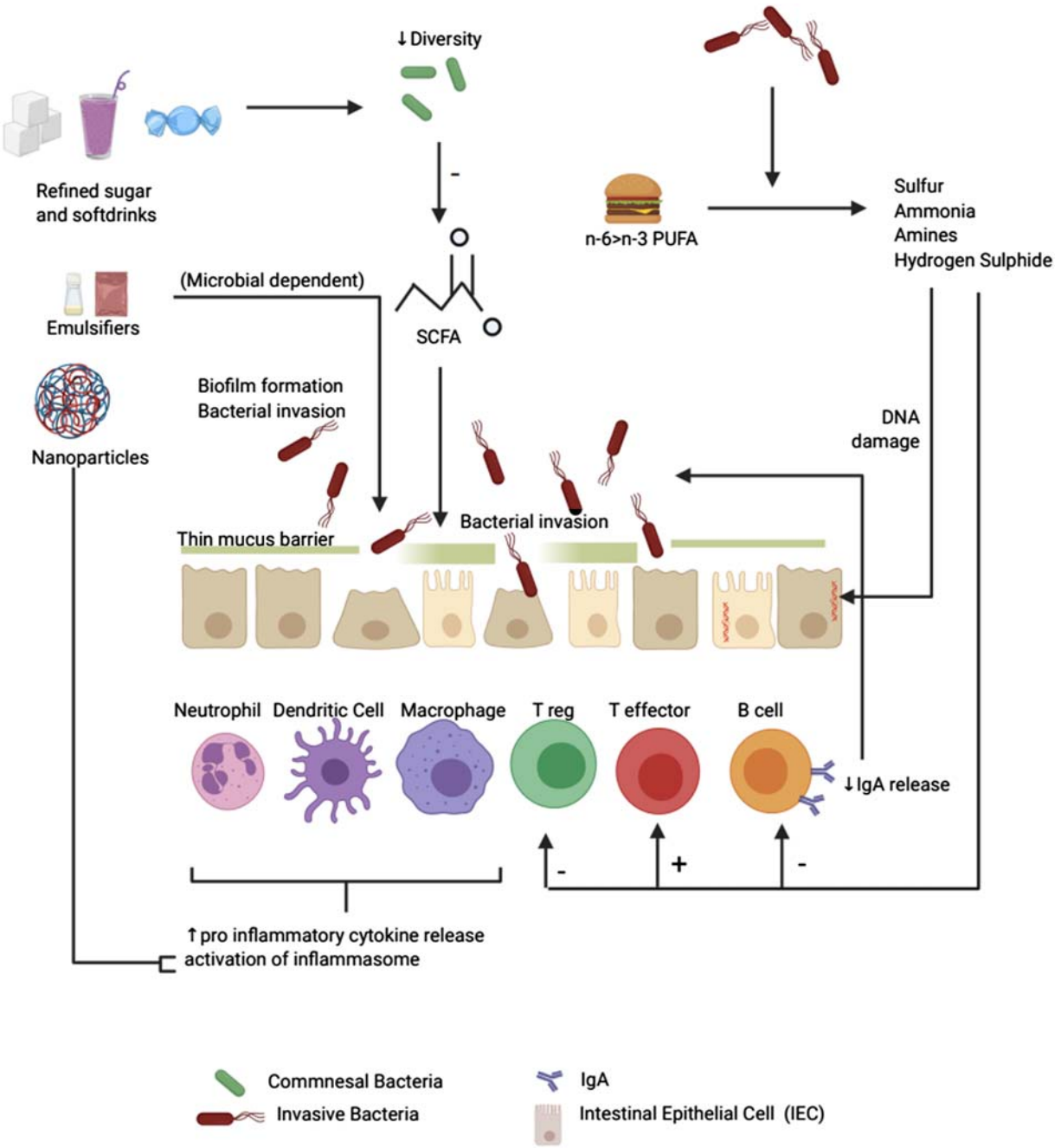


Fig. 2 Schematic representation of proposed mechanisms of pathological diet-immune interactions that may contribute to IBD risk and disease activity (Created with BioRender). PUFA: Polyunsaturated Fatty Acid, SCFA: short-chain fatty acid, IBD: Inflammatory Bowel Disease, Ig: Immunoglobulin. Reproduced from original article: Wark et al. (2021).

A large subgroup (n = 39,511) of women from this cohort who also completed a retrospective validated food frequency questionnaire (FFQ) recalling intake from their high school years was examined. A trend for women in the highest quartile of fiber intake to be less likely to develop CD than women in the lowest quartile (26.2 vs. 16.8 g fiber per day, p = 0.06) was identified. When combined with a diet high in vegetables and fish in addition to fiber (quantified by a "prudent" diet score), these women had a significantly lower risk of incident CD (HR 0.47, 95% CI 0.23–0.98, p = 0.04) (Ananthakrishnan et al., 2015). Again, this association was not found for UC, nor was it found in a more recent systematic review of nine studies examining the risk of UC in 240,327 people (Wang et al., 2017).

A large European prospective cohort study (n = 401,326) showed an inverse relationship between cereal fiber intake and the development of CD in non-smokers but no associations between other sources of fiber and the development of CD or UC

(Andersen et al., 2018). It should be noted that the median age of this cohort was middle age (median age at recruitment 49.6 years), which is older than the median incident age for IBD, suggesting that earlier exposures may be the most critical in modulating disease risk (Torres et al., 2017; Ungaro et al., 2017).

A meta-analysis of 14 case-control studies focusing on fruit and vegetable intake alone, demonstrated that a higher consumption of fruit was associated with a lower risk of development of both UC (OR = 0.69, 95% CI 0.49–0.96) and CD (OR = 0.57, 95% CI 0.44–0.74), whereas a higher consumption of vegetables was associated with a reduced risk of UC (OR 0.71, 95% CI 0.58–0.88) but not CD (OR 0.66, 95% CI 0.4–1.09) (Li, F. et al., 2015). Similar associations between higher fruit and vegetable intake and lower risk of CD were demonstrated in a smaller case control study performed in a pediatric population (Amre et al., 2007).

Several small clinical trials have reported a favorable effect of the use of fiber on disease outcomes in IBD patients but overall, the benefit of fiber demonstrated in clinical trials has been modest. One study (n = 102) reported that dietary fiber in the form of *Plantago ovata* seed was equivalent to mesalazine at maintaining remission at 12-months (40% vs. 35%, respectively) (Fernandez-Banares et al., 1999), and another small (n = 31) open label trial demonstrated that the addition of fiber to mesalazine resulted in reduced rate of relapse at 12-months when compared to mesalazine alone (Copaci et al., 2000).

A systematic review examined six randomized controlled trials in patients with CD prescribed high fiber diets or fiber supplements. They found there was no benefit in the use of dietary fiber in active disease nor in the maintenance of remission (Wedlake et al., 2014).

In the setting of pouchitis following an ileal pouch-anal anastomosis (IPAA), a small well-designed cross over double blind placebo controlled study of 20 patients, reported a positive outcome, with improvement in both clinical and endoscopic parameters following a 3-week fiber supplemented diet (Welters et al., 2002). **Table 2** summarizes the dietary factors proposed to influence IBD disease activity. **Fig. 3** summarizes the dietary management approaches used across the phases of IBD disease activity.

In summary, current data supports the association between fiber intake and the development of IBD, however the data to support its use as a specific IBD therapy is much weaker. Despite this, clinicians should recommend adequate dietary fiber intake due to the many other proven benefits, such as the positive association between fiber intake and a reduced risk of cardiovascular disease, stroke, rectal cancer and diabetes (Gill et al., 2021). The exception to this is in the setting of a patient with known intestinal strictures, where excessive fiber may lead to intestinal obstruction.

Sugar

Animal data support an association between sugar intake and IBD. Mice fed diets high in sucrose and fructose have demonstrated proinflammatory changes consistent with IBD. They were shown to have a decrease in the thickness and quality of colonic mucus, an increase in serum lipopolysaccharide (LPS), which is a bacterial endotoxin and a driver of inflammation, as well as decreased microbial diversity and fecal SCFA content (Laffin et al., 2019; Montrose et al., 2021).

Findings in human studies are less conclusive. In a large nested matched case-control study among 366,351 participants where incident IBD cases were matched to controls, the authors reviewed the association between consumption of a diet high in sugar and soft drinks with risk of a subsequent diagnosis of IBD. The study found a small but significant increased UC risk in the group of patients with a high intake of sugar and soft drinks (Incidence Rate Ratio (IRR) of sugar intake in the fifth vs. the first quintile 1.68, 95% CI 1.00–2.82, P = −0.02) (Racine et al., 2015). This risk was apparent only in those with a concurrent vegetable intake below the population median. No significant associations were found in the development of CD.

A systematic review examined the contribution of carbohydrate and sugar consumption to the risk of a person developing UC. The only sugar associated with a small but statistically significant increased risk of UC was sucrose, with a Relative Risk (RR) per 10 g increment/day of 1.098 (95% CI 1.024–1.177). The same relationship was not found with total sugar intake or other sugar types including fructose, other disaccharides and starch (Wang et al., 2017). A more recent prospective cohort study which included 116,087 patients across seven geographic regions identified associations between the consumption of soft drinks (HR 1.94, p = 0.001, CI 1.42–2.66) and refined, sweetened foods (HR 2.58, p = 0.003, CI 1.44–4.62) and the risk of IBD (Narula et al., 2021).

Studies specifically addressing the impact of sugar intake on active disease and maintenance of remission are lacking, though restricting sugar intake is a component of the Autoimmune protocol diet (AIP), IBD-Anti Inflammatory Diet (IBD-AID) and Specific Carbohydrate Diet (SCD). These diets will be discussed in further detail in following sections.

FODMAPs, wheat and gluten

Consumption of fermentable oligo-, di-, and monosaccharides, and polyols, collectively known as FODMAPs leads to increased GI water secretion, increased fermentation in the colon, and increased gas production. This can lead to luminal distension and can trigger meal-related symptoms in patients with Irritable Bowel Syndrome (IBS) (Lacy et al., 2021). Gluten is a protein found in wheat, and many wheat-based foods are also high in FODMAPs (Rhys-Jones, 2021).

Mouse models have demonstrated that chow and enteral nutrition fortified with gluten compared with gluten-free diets resulted in an increased incidence and severity of terminal ileitis. Polymerase chain reaction (PCR) analysis of distal ileal tissues they found increased expression of the genes coding for the proinflammatory cytokines TNF and Interferon gamma (IFN-gamma) in mice consuming gluten containing chow (Wagner et al., 2013).

Diets low in FODMAPs have been used with success in improving symptoms in patients in IBS and as such are part of international guidelines as an evidence-based management strategy (Altobelli et al., 2017; Lacy et al., 2021; Vasant et al., 2021). There is

Table 2 Dietary components in the management of IBD.

Dietary component	*Active disease*		*Maintenance of remission*	
	CD	*UC*	*CD*	*UC*
Fiber	No benefit of addition to the diet: avoid in stricturing disease (Bischoff et al., 2020)	No evidence of benefit, but no evidence of harm in active UC Pouchitis: Improved clinical and endoscopic parameters with 3 weeks of inulin (Welters et al., 2002)	No impact on risk of relapse (Wedlake et al., 2014)	Equivalent to mesalazine in maintenance of remission at 12-months. Added to mesalazine improved rates of remission at 12 months vs. mesalazine alone (Copaci et al., 2000; Wedlake et al., 2014, Fernandez-Banares et al., 1999)
Sugar	Small trials suggest improvement with restriction/avoidance as part of broader restrictive diets such as SCD/AIP/IBD-AID: see below		No data	
FODMAPs and gluten	Restriction associated with improved symptoms and lead to reduction in fecal calprotectin in patients with mildly active disease (Bodini et al., 2019)		In patients with quiescent disease and ongoing functional symptoms low FODMAPs diet improved symptom scores (Cox et al., 2020; Prince et al., 2016; Herfarth et al., 2014)	
Fat	Not reported	Not reported	A diet with a ratio of n-3:n-6 PUFA of >0.65 associated with reduced flare risk (Uchiyama et al., 2010)	Myristic acid associated with increased risk of disease flare (Barnes et al., 2017) A diet with a ratio of n-3:n-6 PUFA of >0.65 associated with reduced flare risk (Uchiyama et al., 2010)
Red meat			Restricting red meat to fortnightly decreased rate of relapse at 2 years 94% SVD vs. 33% omnivorous (Chiba et al., 2010) Larger study did not show a difference in relapse between meat ≥2/week vs. ≤1/month (Albenberg et al., 2019)	Meat, in particular red and processed meat increased risk of relapse (Jowett et al., 2004)
Emulsifiers and nanoparticles	Positive results in a 4 month pilot study (n = 20) excluding microparticles in addition to steroid treatment, not replicated in larger study (n = 88) (Lomer et al., 2001, 2005) EEN which is efficacious in active CD contains multiple food additives including a variety of emulsifiers (such as polysorbate 80 used in animal trials) and is associated with positive clinical outcomes (Logan et al., 2020)	Increased titanium dioxide nanoparticles in serum of patients with active UC (Ruiz et al., 2017)	Not reported	Not reported

AIP: Autoimmune Protocol Diet, CD: Crohn's Disease, EEN: Exclusive Enteral Diet, FODMAPs: Fermentable. Oligosaccharides, Disaccharides, Monosaccharides and Polyols, IBD: Inflammatory Bowel Disease, IBD-AID: Inflammatory Bowel Disease- Autoimmune Diet: n: Number, n-3: Describes the position of the double bond on carbon 3, counting the methyl carbon as carbon number 1, n-6: Describes the position of the double bond on carbon 6, counting the methyl carbon as carbon 1. PUFA: Polyunsaturated Fatty Acid, SCD: Specific Carbohydrate Diet, SVD: Semi-Vegetarian Diet, UC: Ulcerative Colitis.

Albenberg, L., Brensinger, C., Wu, Q., et al., 2019. A diet low in red and processed meat does not reduce rate of Crohn's disease flares. Gastroenterology 157, 128–136; Barnes, E., Nestor, M., Onyewadume, L., et al., 2017. High dietary intake of specific fatty acids increases risk of flares in patients with ulcerative colitis in remission during treatment with aminosalicylates. Clin. Gastroenterol. Hepatol. 15, 1390–1396; Bischoff, S., Escher, J., Hebuterne, X., et al., 2020. ESPEN practical guideline: clinical nutrition in inflammatory bowel disease. Clin. Nutr. 39, 632–653; Bodini, G., Zanella, C., Crespi, M., et al., 2019. A randomized 6-wk trial of low FODMAP diet in patients with inflammatory bowel disease. Nutrition 67–68; Chiba, M., Abe, T., Tsuda, H., et al., 2010. Lifestyle-related disease in Crohn's disease: relapse prevention by a semi vegetarian diet. World J. Gastroenterol. 16, 2484–2495; Copaci, I., Chira, C., Rovinaru, I., et al., 2000. Maintenance of remission of ulcerative colitis (UC): mesalamine, dietary fiber, S. boulardi. Dig. Liver Dis. 32(suppl); Cox, S., O'Lindsay, J., Fromentin, S., 2020. Effects of low FODMAP diet on symptoms, fecal microbiome, and markers of inflammation in patients with quiescent inflammatory bowel disease in a randomized trial. Gastroenterology 158, 176–188; Fernandez-Banares, F., Hinojosa, J., Sanchez-Lombrana, J., et al., 1999. Randomized clinical trial of *Plantago ovata* seeds (dietary fiber) as compared with mesalamine in maintaining remission in ulcerative colitis. Spanish group for the study of Crohn's Disease and Ulcerative Colitis (GETECCU). Am. J. Gastroenterol. 94,

427–33; Herfarth, H., Martin, C., Sandler, R., et al., 2014. Prevalence of a gluten-free diet and improvement of clinical symptoms in patients with inflammatory bowel diseases. Inflamm. Bowel Dis. 20, 1194–7; Jowett, S., Seal, C., Pearce, M., et al., 2004. Influence of dietary factors on the clinical course of ulcerative colitis: a prospective cohort study. Gut 53, 1479–1484; Logan, M., Gkikas, K., Svolos, V., et al., 2020. Analysis of 61 exclusive entral nutrition formulas used in the management of active Crohn's disease-new insights into dietary disease triggers. Aliment. Pharmacol. Therapeut. 51, 935–947; Lomer, M., Grainger, S., Ede, R., et al., 2005. Lack of efficacy of a reduced microparticle diet in multicentred trial of patients with active disease. Eur. J. Gastroenterol. Hepatol. 17, 377–384; Lomer, M., Harvey, R., Evans, S., et al., 2001. Efficacy and tolerability of a low microparticule diet in a double blind randomized pilot study in Crohn's Disease. Eur. J. Gastroenterol. Hepatol. 13, 101–106; Prince, A., Myers, C., Joyce, T., et al., 2016. Fermentable Carbohydrate Restriction (Low FODMAP Diet) in clinical practice improves functional gastrointestinal symptoms in patients with inflammatory bowel disease. Inflamm. Bowel Dis. 22, 1129–36; Ruiz, P., Moron, B., Becker, H., et al., 2017. Titanium dioxide nanoparticles exacerbate DSS-induced colitis: role of the NLRP3 inflammasome. Gut 66, 1216–1224; Uchiyama, K., Nakamura, M., Odahara, S., et al., 2010. N-3 polyunsaturated fatty acid diet therapy for patients with inflammatory bowel disease. Inflamm. Bowel Dis. 16, 1696–1707; Wedlake, L., Slack, N., Andreyev, H.J.N., Whelan, K., 2014. Fiber in the treatment and maintenance of inflammatory bowel disease: a systematic review of randomized controlled trials. Inflamm. Bowel Dis. 20, 576–86; Welters, C., Heineman, E., Thunnissen, F., et al., 2002. Effect of dietary inulin supplementation on inflammation of pouch mucosa in pateints with ileal pouch-anal anastamosis. Dis. Colon Rectum 45, 621–627.

evidence that this approach has a place for patients with quiescent IBD and concurrent functional symptoms consistent with IBS. The prevalence of self-reported non-coeliac gluten sensitivity (NCGS) is 23.6–27.8% in IBD, with NCGS being more common in patients with stricturing disease and those with recent self-reported flares (Aziz et al., 2015) (Limketkai et al., 2018).

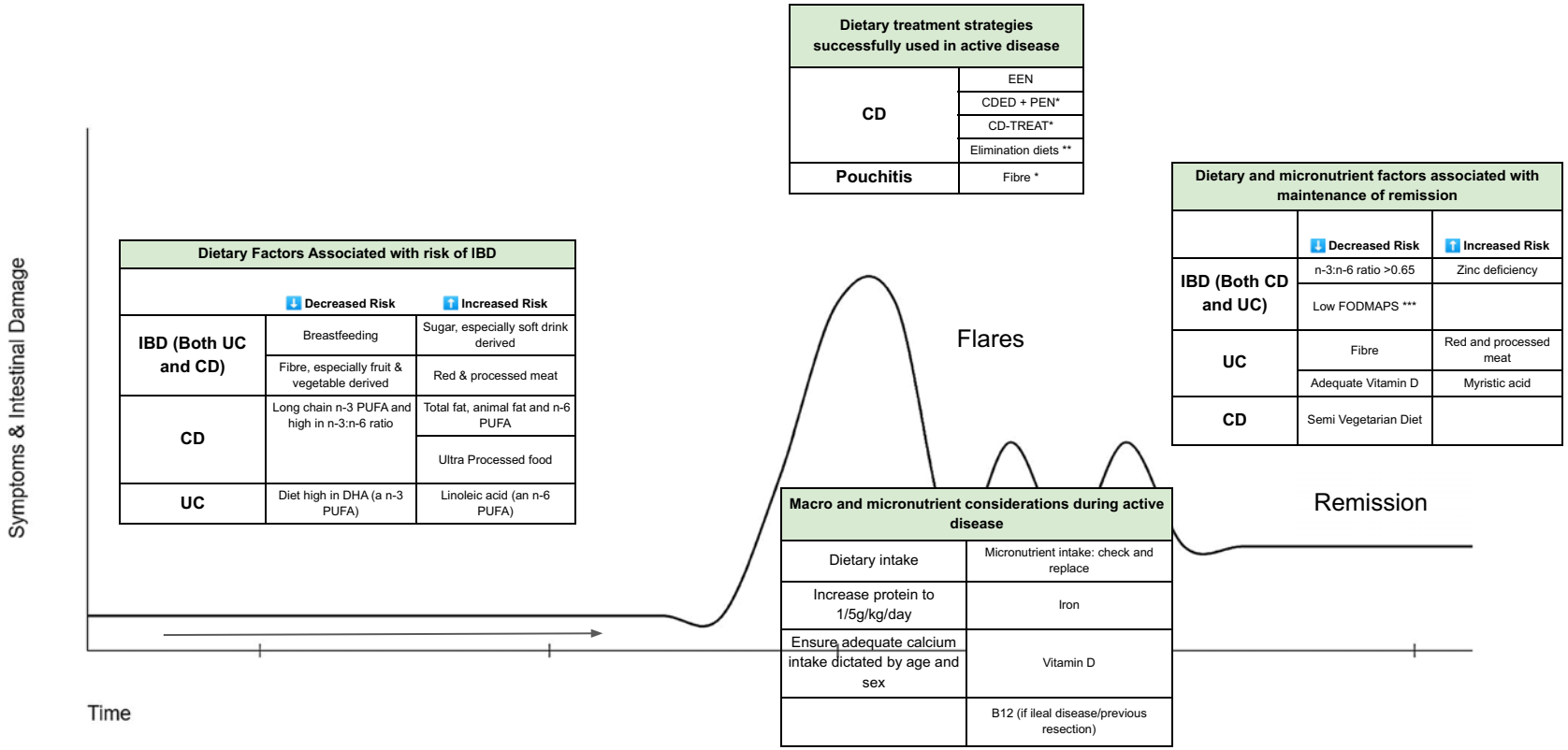
A randomized, double blind, placebo-controlled study demonstrated higher symptoms scores for pain, bloating and urgency in 32 patients with quiescent IBD and concurrent IBS challenged with FODMAPs (Cox et al., 2017). Several trials of patients with quiescent IBD following a low FODMAP approach have demonstrated an improvement in symptom severity scores, reduced fatigue, less self-reported flares and a higher quality of life without objective change in markers of inflammation (Cox et al., 2020; Prince et al., 2016; Herfarth et al., 2014). Interestingly, one study performed in 55 patients with IBD which included patients in remission or with mild disease activity on a stable medication regimen were commenced on a low FODMAP diet. In some patients there was a slight reduction in the fecal calprotectin, though the clinical significance of this is uncertain and these findings have not yet been replicated (Bodini et al., 2019).

Fat

Triglycerides (TG)s make up the largest proportion of dietary fat in the human diet. A TG is composed of a glycerol backbone and three fatty acids (FA)s esterified to the glycerol. The FAs in the human diet vary in the carbon chain length and in the number and position of the double bonds along the hydrocarbon chain. Most dietary oils contain TGs with FA containing 4–26 carbons. The type of dietary fat is defined by the saturation, the position of the first double bond relative to the methyl end (marked with "n" or "omega") in the carbon chain and the isomerization of the double bond (cis vs. trans). These characteristics determine the physical properties of the fats, for example, whether they are solid or liquid at room temperature (Ross, 2014).

1. Monounsaturated FAs (MUFAs): At least 12 carbon atoms long, typically with a double bond at the n-9 or n-7 position. Most common in the diet is oleic acid (C18:1, n-9 cis), found in olive oil
2. Polyunsaturated FAs (PUFA): Contains 2–6 double bonds between carbons within the fatty acid chain (the number of double bonds is indicated by a number to the right of the number of carbon atoms, e.g., C22:6 is a FA with 22 carbon atoms and 6 double bonds in the molecule). The position of the first double bond would be at the n-3, n-6 or n-9 if longer than 18 carbons in length and located at the n-7 position for a 16-carbon atom FA. FAs with double bonds at the n-3 and n-6 position are essential in the diet since the human biosynthetic enzymes cannot form double bonds at any position closer to the methyl end other than n-9. Example of a common n-6 PUFA in the diet is linoleic acid (C18:2, n-6,9 all cis) found in seed oils and n-3 includes alfa-linoleic acid (C18:3, n-3,6,9 all cis) found in vegetable oils, nuts and docosahexaenoic acid (C22:6, n-3,6,9,12,15,18 all cis) found in fish.
3. Saturated FAs have no double bonds and are found in animal food sources such as meat, dairy, palm oil and pastries (due to the use of butter or margarine).
4. Trans FAs are found in small quantities naturally in animal source fat, but can also be produced by partial hydrogenation of polyunsaturated oils by the food industry.

Healthy eating guidelines recommend limiting foods high in saturated and trans-fat due to the link with metabolic disease, cancer and even poorer mental health outcomes (NHMRC, 2013, Forouhi et al., 2018). Animal models have supported the role saturated fat may have in promoting bowel inflammation. Mouse models have demonstrated that a diet high in saturated fat, but not PUFA, increased the proinflammatory T-cell responses, as well as the incidence of colitis. This was thought to be due to the fat promoted taurine conjugation of bile acids facilitating an increased availability of sulfur to sulphite-reducing microorganisms such as Bilophilia wadsworthia, which is associated with intestinal inflammation (Devkota et al., 2012). Furthermore, mice fed a high fat-high sugar diet developed gastrointestinal dysbiosis characterized by increased proinflammatory microorganisms and decreased favorable SCFA that made them more susceptible to chemically-induced colitis (Agus et al., 2015). In another study, mice fed high fat diets had a reduced level of secretory immunoglobulin A coating the gut microbiota, which is an important component of gut barrier function (Muhomah et al., 2019).



Dietary Factors Associated with risk of IBD		
	⬇ Decreased Risk	⬆ Increased Risk
IBD (Both UC and CD)	Breastfeeding	Sugar, especially soft drink derived
	Fibre, especially fruit & vegetable derived	Red & processed meat
CD	Long chain n-3 PUFA and high in n-3:n-6 ratio	Total fat, animal fat and n-6 PUFA
		Ultra Processed food
UC	Diet high in DHA (a n-3 PUFA)	Linoleic acid (an n-6 PUFA)

Dietary treatment strategies successfully used in active disease	
CD	EEN
	CDED + PEN*
	CD-TREAT*
	Elimination diets **
Pouchitis	Fibre *

Dietary and micronutrient factors associated with maintenance of remission		
	⬇ Decreased Risk	⬆ Increased Risk
IBD (Both CD and UC)	n-3:n-6 ratio >0.65	Zinc deficiency
	Low FODMAPS ***	
UC	Fibre	Red and processed meat
	Adequate Vitamin D	Myristic acid
CD	Semi Vegetarian Diet	

Macro and micronutrient considerations during active disease	
Dietary intake	Micronutrient intake: check and replace
Increase protein to 1/5g/kg/day	Iron
Ensure adequate calcium intake dictated by age and sex	Vitamin D
	B12 (if ileal disease/previous resection)

*Trials with small numbers, ** Autoimmune Protocol Diet, Specific Carbohydrate Diet and Mediterranean Diet show symptomatic benefit, but no objective improvement in inflammatory activity, *** Indicated in the management of IBD patients with concurrent functional symptoms during remission

Fig. 3 Dietary influences on the risks of a person developing IBD, dietary measures used in its treatment and for the maintenance of remission. CD: Crohn's Disease, CDED: Crohn's Disease Exclusion Diet, CD-TREAT: Crohn's Disease Treatment-with EATing diet; a solid food-based diet designed to replicate the nutrients and food ingredients composition of Exclusive Enteral Nutrition, DHA: Docosahexaenoic Acid, EEN: Exclusive Enteral Nutrition, FODMAPs: Fermentable Oligosaccharides, Disaccharides Monosaccharides and Polyols, IBD: Inflammatory Bowel Disease, n-3: Describes the position of the double bond on carbon 3, counting the methyl carbon as carbon number 1, n-6: Describes the position of the double bond on carbon 6, counting the methyl carbon as carbon 1, PEN: Partial Enteral Nutrition, PN: Parental Nutrition, PUFA: Polyunsaturated Fatty Acid, UC: Ulcerative Colitis, Ultra processed foods: Formulations of ingredients made by a series of industrial steps that require sophisticated equipment and technology. They are aimed at making the final product hyperpalatable and with a long shelf life.

N-3 PUFAs typically from fish have anti-inflammatory functions. They are precursors to anti-inflammatory cytokines and suppress inflammatory T-cell responses (Chapkin et al., 2007). In contrast, it has been suggested that n-6 PUFA found in edible seed oils such as sunflower oil, serve as precursors to pro-inflammatory signaling molecules such as cytokines, prostaglandins and leukotrienes (Patterson et al., 2012) (Gonzalez-Becerra et al., 2019).

The data on the effect of fat intake on incidence of IBD are variable. Dietary intake of n-3 PUFA, n-6 PUFA and the ratio between them appears to be important in both IBD risk and disease control. Shoda et al. correlated the incidence of CD with population daily average nutrient intake in a Japanese population with minimal migration, and therefore stable genetics (Shoda et al., 1996). They found that the incidence of CD strongly correlated with the average daily intake of total fat, animal fat and n-6 PUFA (Shoda et al., 1996).

A case-control study in children found that a higher consumption of long chain n-3 FA and a higher ratio of dietary n-3:n-6 FAs was associated with a lower risk of CD (Amre et al., 2007). A nested case-control study within the EPIC cohort of 203,193 adults demonstrated an association between the consumption of dietary linoleic acid (C18:2, n-6) and an increased risk of UC, with an Odds Ratio (OR) of consumption in the highest vs. lowest quartile of 2.49 (95% CI 1.23–5.07, $p = 0.01$) (Tjonneland et al., 2009).

With regards to UC, a prospective cohort study of 25,639 participants in the UK demonstrated that there was a protective effect of higher consumption of the n-3 PUFA docosahexaenoic acid (C22:6) and incidence of UC (John et al., 2010). A more recent meta-analysis of nine studies including 172,555 participants however, failed to demonstrate an association between fat intake or fat categories and UC risk (Wang et al., 2016).

Fat intake may modulate the risk of disease flare in those with established IBD. In a prospective study of 412 patients in remission from UC on Aminosalicylate, higher intake of the saturated FA myristic acid (C14:0, commonly found in coconut oil, palm oil and dairy products) was associated with an increased risk of a UC flare (OR 3.01, 95% CI 1.17–7.74) (Barnes et al., 2017). In a dietary intervention, 230 IBD patients who had been induced into remission via conventional pharmacotherapy were commenced on a dietary regimen that targeted a dietary fat intake with a n-3/n-6 ratio of 1. This was achieved through restricting the consumption of dietary sources of n-6 PUFA (in particular food rich in vegetable oil, e.g., mayonnaize and fried foods) and promoting the consumption of foods high in n-3 PUFA (such as salmon and sardines) (Uchiyama et al., 2010). Objective assessment of the n-3/n-6 ratio was achieved by analyzing the ratio between n-3 and n-6 FA in the erythrocyte membrane (an objective measure reflecting the ratio between these FAs in the diet). They found that those who maintained remission for at least 12 months had a significantly higher n-3/n-6 FA ratio than those who relapsed.

Studies examining dietary cholesterol consumption have not been conducted, though due to its presence in similar food groups, levels of consumption tend to mirror saturated FA intake, and as a result the stand-alone effect of cholesterol intake difficult to assess (Xu et al., 2018). No studies have specifically examined dietary intake of fat in the setting of active disease.

Red meat

Red, but more significantly, processed meat consumption has been suggested to have proinflammatory effects in the gastrointestinal tract. Proposed mechanisms are the generation of DNA damaging amines in the cooking process, as well as the fermentation of meat derived protein by the gut microbiota to produce DNA damaging toxic substances such as ammonia, amines and nitrous compounds (Le Leu and Young, 2007, Sugimura et al., 2004; Lewin et al., 2006). Another contributing factor is the concurrent presence of saturated fat, emulsifiers and food additives in processed meats; with their established deleterious effects (Ge et al., 2015; Bancil et al., 2021; Narula et al., 2021).

Unprocessed red meat also showed evidence of harm in animal models. Li and colleagues reported that mice fed the diet highest in unprocessed red meat developed an altered microbiome in addition to increased intestinal permeability; they also suffered more severe colitis than mice fed a diet containing less red meat (Li et al., 2021).

A recent large prospective cohort study of 116,087 participants found that a processed meat intake of $\geq$1 serving (55 g) per day was associated with a higher risk of development of CD than an intake of $\leq$1 serve per week (HR 2.07, 95% CI 1.14–3.76) (Narula et al., 2021). No significant associations were found with the intake of unprocessed red or white meat intake and IBD risk. Another study that examined dietary factors in a cohort of 67,581 French women reported an association between a high animal protein intake, in the form of meat or fish rather than eggs or dairy, and an increased risk of IBD (third vs. first tertile of intake HR 3.31, 95% CI 1.45–6.34) (Jantchou et al., 2010).

The EPIC-IBD Cohort found that protein intake from an animal source, and meat consumption were associated with a higher risk of IBD, whereas protein intake from a vegetable source was not (Ge et al., 2015). This finding could be confounded by the protective effect of dietary fiber that accompanies the intake of vegetable based protein (Dong et al., 2020). Finally, a meta-analysis of nine studies investigating the associations between meat consumption and IBD risk found that red meat consumption was associated with an increased risk of IBD (RR 2.37, 95% CI 1.4–3.99), while white meat consumption was not (Ge et al., 2015).

Jowett and colleagues performed a prospective cohort study of 191 patients to determine dietary risk factors for relapse in UC (Jowett et al., 2004). They identified that the consumption of meat (OR 3.2, 95% CI 1.3–7.8), particularly red and processed meat (OR 5.19, 95% 2.1–12.9), was associated with an increased risk of relapse at 12 -months. A recent interventional study which randomized CD patients in remission ($n = 214$) to consume a high meat ($\geq$2 serves per week) or a low meat ($\leq$1 serve/month) diet for 49 weeks revealed that relapse occurred in 62% of the participants in the high meat intake group and 42% in the low meat intake group, with the difference in relapse not reaching statistical significance between the groups ($p = 0.61$) (Albenberg et al., 2019).

From a broader health perspective, the consumption of processed, and to a lesser extent, unprocessed red meat has been associated with detrimental health outcomes such as multiple forms of cancer, cardiovascular disease and dementia (Vernooij et al., 2019; Zhang et al., 2021; Huang et al., 2021). Conversely, the consumption of a vegetarian diet results in positive alterations to the gut microbiome compared with an omnivorous diet and has been associated with health benefits including reduced risk of ischemic heart disease and multiple types of cancer (Dinu et al., 2017). Semi vegetarian diets (SVD), which include infrequent consumption of meat, have been associated with similar health benefits, which may improve long term adherence (Derbyshire, 2016; Aykan, 2015).

There is promising data on the use of low meat, semi-vegetarian diets in maintenance therapy after medically or surgically induced remission in CD. Adherence to these diets are postulated to increase beneficial bacteria through the provision of fiber and prebiotic food (Chiba et al., 2015). The semi-vegetarian diet proposed includes red meat every 2 weeks, fish once a week and encourages daily consumption of brown rice, green tea, fermented foods, eggs, dairy products and vegetables. A pilot study of 22 patients reported an impressive 2-year clinical remission rate of 92% and diet adherence rate of 72% (Chiba et al., 2010). Studies in larger cohorts are necessary to verify this finding.

Food additives

Emulsifiers

Emulsifiers are over 60 different food additives that are used to stabilize the consistency of food products and prevent unappetizing separation of oil and water in the product (Bancil et al., 2021). Emulsifiers lengthen the shelf life and optimize the appearance and texture of food and have become an integral part of the "Western" diet which is rich in heavily processed food. Examples include lecithin found in chocolate, xanthan gum found in mayonnaize, carrageenan found in flavored milks and polysorbates found in ice cream (McClements, 2016, Cox et al., 2021).

An eloquent review by (Bancil et al., 2021) and colleagues summarized the many studies in animal models that have demonstrated the impact of emulsifiers on intestinal inflammation, which include: decreasing microbial diversity, reducing mucus thickness, increasing bacterial encroachment on the epithelium, increasing intestinal permeability and inducing the production of pro inflammatory mediators in the gut. The addition of emulsifiers such as carrageenan and methylcellulose to the diet have also been found to increase risk and severity of colitis in animal models. It should be acknowledged that the experimental doses of emulsifiers used in these animal models can be much higher than average human consumption, although this limitation may be mitigated by the fact that studies in animal models test the acute effect of emulsifiers, whereas the intake in humans is over many years and therefore the cumulative dose in the human diet is likely to exceed the doses tested in animals, although not directly comparable.

An in vitro model which stimulated the lumen-associated and mucus associated human intestinal microbial ecosystems, termed the "mucosal simulator of the human intestinal microbial ecosystem" (M-SHIME) was generated. Its design utilized a series of pH and temperature-controlled vessels that were inoculated with human fecal suspension and maintained to mimic conditions of the human GI tract. When polysorbate 80 (P80) was added to this model there was a 50% decrease in microbial species diversity and an increase in microbial genes coded for flagellin expression which is a marker of bacterial adhesion and invasion that can directly activate an inflammatory response. Transfer of the emulsifier treated M-SHINE microbiota into a germ-free mouse host increased IL-6 (a pro inflammatory cytokine) expression suggesting that host-microbial interactions may be responsible for driving intestinal inflammation (Chassaing et al., 2017).

In a study of patients with CD vs. healthy controls it was found that patients with CD demonstrated an increased expression of malX gene (essential for maltodextrin metabolism), which has been associated with E Coli adhesion and postulated to be a driver of inflammation (Nickerson and McDonald, 2012). This suggests a mechanism by which Western diets and the consumption of food additives may promote dysbiosis and contribute to the initiation and perpetuation of disease in susceptible populations.

A large prospective cohort study of 116,087 participants, spanning seven geographical regions examined the association between the consumption of a diet rich in ultra-processed food and incident CD. Ultra-processed food includes all types of packaged and formulated foods that contain emulsifiers, flavorings, colors and thickeners used in processed meat, sauces, mass produced biscuits and pastries, see glossary for formal definition (Monteiro et al., 2019). They found that the consumption of $\geq$5 serves of ultra-processed food per day was associated with an increased risk of CD with a HR of 1.83 (95% CI 1.22–2.72) (Narula et al., 2021). Roberts and colleagues also demonstrated a positive correlation between emulsifier consumption and incidence of CD (Roberts et al., 2013).

A double blind randomized controlled study was conducted in 12 patients in remission from their UC (Bhattacharyya et al., 2017). For 12 months all participants were instructed to follow a "no-carrageenan diet," while 5 were randomized to receive 200 mg/day carrageenan capsules, the other 7 were randomized to a placebo dextrose tablet. Three of the 5 patients receiving the carrageenan capsules relapsed vs. none of the 7 randomized to receive the placebo capsules ($p = 0.046$). The carrageenan group also had a significant increase in circulating IL-6 levels and fecal calprotectin between the beginning and the end of the study, while the placebo group did not. This is a very small study that supports the notion that carrageenan may impact the risk of disease relapse.

On the other hand, while food additive consumption in the context of whole foods common in Western diet may increase the incidence of IBD, formulas used in Exclusive Enteral Nutrition (EEN), an effective treatment for CD outlined in further detail below, all contain food additives such as modified food starch, inorganic phosphates, maltodextrin and carrageenan, yet have positive clinical efficacy in active disease (Logan et al., 2020). This challenges the perception that these ingredients are universally harmful.

Nanoparticles

Titanium dioxide nanoparticles are widely used as food additives and in pharmaceutical products. Ruiz and colleagues demonstrated in a mouse model as well as ex vivo that administered titanium dioxide nanoparticles accumulated in intestinal epithelial cells (IECs) and resulted in increased intestinal permeability and worsened acute colitis through a mechanism involving the NLRP3 inflammasome (Ruiz et al., 2017). An increased titanium level was found in the blood of patients with active UC. A pilot study of 20 patients with active ileal or ileocolonic CD were treated with standard corticosteroids at the clinicians' discretion. Patients were randomized to two groups, the first group was given advice to exclude food and pharmaceuticals that contained nanoparticles in addition to recommendation to avoid fibrous fruit and vegetables, which was part of standard dietary advice for CD at the time. The control group were given standard dietary advice alone. At 4 months there was a significant improvement in the Crohn's Disease Activity Index (CDAI) in the treatment group. (Lomer et al., 2001). A larger study including 88 IBD patients did not replicate these findings (Lomer et al., 2005).

Malnutrition assessment and management considerations in IBD

Malnutrition is more common in IBD patients than in the general population, with an adjusted OR of 5.57 (Nguyen et al., 2008). When compared to the Australian Nutrient reference values, the most significantly deficient macro and micro nutrients consumed in people with IBD include total energy, fiber, folate, calcium; in particular consumption from the food groups of breads, cereals, legumes, fruit, vegetables and dairy (Lambert et al., 2021). As outlined in the European Society for Clinical Nutrition and Metabolism (ESPEN) guidelines on clinical nutrition in IBD (Bischoff et al., 2020), malnutrition presents as a more common issue in CD than UC given its ability to affect any part of the gastrointestinal tract, whereas UC is limited to the colon and has fewer direct malabsorptive effects. Malnutrition is most prevalent in active disease and is associated with a worse prognosis when present; with increased rates of hospitalization, surgery, and mortality (Nguyen et al., 2008; Gajendran et al., 2016; Wallaert et al., 2012). Assessment of energy, macro- and micro-nutrient intake requires screening, assessment and management to prevent inadequate intake, which can lead to sequelae such as loss of muscle mass, osteoporosis and, in the setting of the pediatric population, sub optimal growth and development.

Macronutrient deficiencies

For malnutrition to be managed, it first must be recognized. There are two malnutrition screening tools (NST) that have been designed for use in the IBD population. Several NSTs designed for use in broader population groups have been validated in IBD populations, where associations with clinical outcomes such as flares, hospitalization and need for surgery have been identified, see **Table 3** for NSTs and their components (Russell et al., 2022; Li et al., 2019; Elia, 2003, Haskey et al., 2018; Kondrup et al., 2003; Jansen et al., 2016; Sumi et al., 2016). The Malnutrition Universal Screening Tool (MUST) is one such screening score, which considers Body Mass Index (BMI), weight loss, patient overall wellbeing and food intake and has been shown to be an accurate patient administered screening tool and can be carried out prior to their clinic review, which increases its ease of utilization (Sandhu et al., 2016). Once a patient is identified as being at risk, a more detailed and time consuming nutrition assessment tools (NAT), such as the Subjective Global Assessment (SGA) can be utilized for more detailed objective measurement (Detsky et al., 1987; Garry and Vellas, 1999). Limited data has shown that the scores obtained from the briefer nutrition screening tools associated well with a diagnosis of malnutrition compared with the more resource heavy and time consuming assessment tools (Li et al., 2019).

When protein calorie malnutrition is detected, oral nutritional support in the form of supplements or EEN are the preferred options in the setting of macronutrient deficiency and active disease. During active disease there is a proteolytic and catabolic response, as such there is a recommendation for an increase in daily protein intake from 1 g/kg/day to 1.2–1.5 g/kg/day (Bischoff et al., 2020). Outside of this setting, broad energy requirements and dietary recommendations are similar to those of the healthy population; i.e., there is no such thing as an "oral IBD diet" that can generally be recommended to promote remission, however, some of the approaches trialed and supporting data are outlined in following sections (Bischoff et al., 2020).

Under nutrition can negatively impact surgical outcomes. It is therefore recommended that patients with severe nutritional risk, defined as a weight loss >10–15 kg in the last 5 months, BMI <18.5 kg/m2 and/or albumin <30 g/L, should receive nutritional therapy prior to surgery to optimize outcomes (Bischoff et al., 2020).

A large study conducted in the USA found that even after adjusting for confounding factors such as malnutrition, surgery, age and comorbidity; the use of Parenteral Nutrition (PN) had a 2.5- fold higher odds of death and a twofold increase in length of hospital stay compared to enteral nutrition (Nguyen et al., 2007). Consequently, PN is reserved for those who have the most complicated disease, such as those who have decreased gut absorptive capacity (short gut), specific surgical cases of bowel obstruction, those who have attempted, but are intolerant to or cannot meet their nutritional requirements with Enteral Nutrition (EN) alone (Bischoff et al., 2020).

Micronutrient deficiencies

Micronutrient deficiencies can present with a vast array of signs and symptoms which can be screened for via taking a patient history, however, the most accurate way to assess for most deficiencies is via a blood test to quantify the serum level of the micronutrients

Table 3 Nutrition screening tools and nutritional assessment tools used in assessment of patients with IBD.

Name	*History*	*Examination*	*Investigations*	*Comment*
Malnutrition Universal Screening Tool (MUST) (Elia, 2003; Sandhu et al., 2016)	• Unintentional weight loss last 6 months as (% of initial body weight) • Acute illness • Adequacy of food intake over last 5 days	Height Weight BMI		• Designed for IBD cohort • Recommended for the ambulatory care setting • Has been validated as a patient collected pre clinic screening tool to improve efficiency (Sandhu et al., 2016)
Saskatchewan IBD-Nutrition Risk Tool (SK-IBD) (Haskey et al., 2018)	Questionnaire: • Experienced GI symptoms for >2 weeks • Unintentional weight loss in last month • Eating poorly due to decreased appetite • Restriction of food groups			• Designed for IBD cohort • Omits BMI: Felt to be a poor marker of malnutrition risk
Nutritional Risk Screening 2002 (NRS-2002) (Kondrup et al., 2003)	• Food intake as a % of normal in last week • Severity of disease: (mild, moderate, severe)	Height Weight BMI		Correlates with clinical outcomes in IBD (Li et al., 2019)
Malnutrition Inflammation Risk Tool (MIRT) (Jansen et al., 2016)	Unintentional weight loss last 3 months (% body weight)	Height Weight BMI	CRP	Correlates with clinical outcomes in IBD (Li et al., 2019)
Nutrition Risk Index (NRI) (Sumi et al., 2016)	Baseline weight	Current weight	Albumin	Correlated with clinical outcomes in IBD (Li et al., 2019)
Subjective Global Assessment (SGA) (Detsky et al., 1987)	• Weight change last 6 months • Weight change last 2 weeks • Adequacy of dietary intake • Gastrointestinal symptoms • Functional capacity	• Weight • Subcutaneous fat: under eyes, triceps, biceps • Muscle wasting: temple, clavicle,shoulder, scapular, ribs, quadriceps, calf, knee, hand • edema • Ascites		An example of a Nutrition Assessment Tool (NAT) prompted by high-risk nutrition screening tool: more detailed and time consuming
Mini Nutritional Assessment score (MNA) (Garry and Vellas, 1999)	• Decline in food intake • Full meals • Quantify consumption of: dairy, protein sources fruit and vegetables, fluid • Weight loss • Mobility and living status (independent/care) • Recent psychological stress or mental health diagnosis • Use of > 3 prescription drugs/day • Skin sores • Self-view of nutritional status	• Height • Weight • BMI • Mid arm circumference • Calf circumference		More tailored to geriatric population

BMI: Body Mass Index, CRP: C-Reactive Protein, GI: Gastrointestinal Tract, IBD: Inflammatory Bowel Disease.

Detsky, A., McLaughlin, J., Baker, J., et al., 1987. What is subjective global assessment of nutritional status? J. Parenter. Enter. Nutr. 11, 8–13; Elia, M., 2003. The "MUST" Report. Nutritional Screening of Adults: A Multidiciplinary Responsibility Redditch, UK BAPEN: Multidiciplinary Advisory Group; Garry, P.J., Vellas, B., 1999. Practical and validated use of the mini nutritional assessment in geriatric evaluation. Nutr. Clin. Care 2, 146–154; Haskey, N., Pena-Sanchez, J., Jones, J., et al., 2018. Development of a screening tool to detect nutrition risk in patients with inflammatory bowel disease. Asia Pac. J. Clin. Nutr. 27, 756–762; Jansen, I., Prager, M., Valentini, L., et al., 2016. Inflammation-driven malnutrition: a new screening tool predicts outcome in Crohn's disease. Br. J. Nutr. 116, 1061–1067; Kondrup, J., Rasmussen, H., Hamberg, O., et al., 2003. Nutritional risk screening (NRS 2002): a new method based on an analysis of controlled clinical trials. Clin. Nutr. 22, 321–336; Li, S., Ney, M., Eslamparast, T., et al., 2019. Systematic review of nutrition screening and assessment in inflammatory bowel disease. World J. Gastroenterol. 25, 3823–3837; Sandhu, A., Mosli, M., Yan, B., et al., 2016. Self-screening for malnutrition risk in outpatient inflammatory bowel disease patients using the malnutrition universal screening tool (MUST). J. Parenter. Enter. Nutr. 40, 507–10; Sumi, R., Nakajima, K., Iijima, H., et al., 2016. Influence of nutritonal status on the therapeutic effect of infliximab in patients with Crohn's disease. Surg. Today 46, 922–929.

that are present in the vascular compartment. When interpreting blood results of micronutrients and trace elements it is important to consider that serum values may be increased or decreased by a change in inflammatory activity and this can impact the interpretation of results in this context (Dingnass et al., 2015; Bischoff et al., 2020). For summary of important micronutrients and suggested indications for testing see **Table 4**.

Iron

Iron deficiency anemia is a common form of anemia in IBD patients with active disease and can be caused by blood loss from mucosal inflammation, malnutrition with reduced iron intake or impaired uptake through the duodeno-jejunal mucosa (Dingnass et al., 2015). It should be screened for at least every 3-months with a ferritin level during periods of active disease (Bischoff et al., 2020). In the presence of inflammation, a serum ferritin of <100 μg/L may be considered indicative of iron deficiency and supplementation is warranted. ECCO guidelines recommend IV replacement due to its superior speed of replacement and tolerability compared to oral iron (Dingnass et al., 2015). Oral iron can be considered in those with clinically inactive disease, mild deficiency and those who have not been intolerant to oral iron (See **Table 3**).

Vitamin B12

A systematic review of the literature has demonstrated that the primary risk factor for Vitamin B12 deficiency in IBD is ileal resection or areas of small intestinal disease >20 cm (Duerksen et al., 2006). UC does not predispose to deficiency with the possible exception of deficiency following restorative proctocolectomy (Battat et al., 2014). Patients with risk factors for B12 deficiency should be screened annually.

Administration of vitamin B12 via the intramuscular route to replace and maintain vitamin B12 levels is the current standard of care, however delivery via the oral and sublingual route is being increasingly explored and practice may change in the coming years (Bischoff et al., 2020; Gomollon et al., 2017). If vitamin B12 levels are low in patients without these IBD specific risk factors for deficiency then an alternative explanation such as autoimmune gastritis should be sought (Battat et al., 2014).

Folate

Folate is present in animal and plant foods and fortification in food is now common practice in bread and breakfast cereal (NHMRC, 2006; Zealand, 2016). Folate is absorbed in the proximal small intestine which is a less common site of CD related inflammation, but deficiency can be contributed to by therapeutic agents such sulfasalazine and methotrexate. A study of 257 patients with IBD vs. healthy controls demonstrated that patients with CD, in particular with disease duration of >5 years had lower folate levels than those patients with UC and healthy controls (Huang et al., 2017). A meta-analysis also demonstrated that serum folate concentrations were lower in IBD patients, interestingly, this was more marked in patients with UC in this study (Pan et al., 2017).

No guidelines surrounding routine testing are available, however, it would be logical to test in patients with risk factors for deficiency or patients who would face significant complications of deficiency such as those on sulfasalazine or Methotrexate, patients with malnutrition and pregnant patients with IBD.

Calcium

It is clear from large population based studies that IBD is associated with an increased risk of osteoporosis and osteopenia, resulting in an elevated incidence of low trauma fractures (Etzel et al., 2011; van Staa et al., 2003). Fracture propensity is increased by chronic inflammation, corticosteroid use, vitamin D deficiency and low dietary calcium intake (Vernia et al., 2013).

Serum calcium concentrations are kept at a tight set point via hormonal pathways, including parathyroid hormone, vitamin D and fibroblast growth factor 23. These mechanisms result in mobilization of skeletal calcium and resorption of calcium from the distal nephron of the kidney to maintain consistent serum calcium levels. As such, serum calcium measurements are not helpful in assessing the adequacy of a patient's calcium status (Houillier et al., 2005).

The minimum recommended daily calcium intake is 1000 mg for men and pre-menopausal women and 1300 mg for post-menopausal women and men >60 years (NHMRC, 2006). There are helpful online calculators such as the tool available from the International Osteoporosis Foundation that can assist in assessing sufficiency of dietary calcium intake (International Osteoporosis Foundation, 2022). A large study assessing calcium intake in patients with IBD compared with healthy controls found the intake of calcium was significantly lower in the IBD group and below minimum dietary recommendations in one-third of patients (Vernia et al., 2013). Self-reported lactose intolerance leading to dietary restriction was a major contributing factor. Low calcium intake represents a reversible risk factor for osteoporosis. Corticosteroids can also reduce calcium absorption from the gut and increase urinary losses leading to secondary hyperparathyroidism with a resultant increased bone resorption (Lamb et al., 2019).

Oral calcium supplementation should be taken with caution following a long-term cohort study that demonstrated that a calcium intake of >1400 mg daily was associated with an increase in all-cause mortality, in particular cardiovascular disease and ischemic heart disease (Michaelsson et al., 2013). This was most marked in people with a high dietary calcium intake plus oral supplementation rather than a high dietary calcium intake alone. As such, long term calcium supplementation is not recommended unless dietary calcium intake is <800 mg/day (Lamb et al., 2019). The suggested dose of daily oral calcium supplement in IBD patients with insufficient dietary intake is 500–1000 mg/day (Harbord et al., 2015).

Table 4 Key points in assessment and management of micronutrient deficiencies.

Vitamin/ minerals	*Blood test frequency in active disease*	*Blood test frequency in remission*	*Route of replacement*	*Comment*
Iron	3 monthly 3-monthly during pregnancy, even in IBD remission (Bischoff et al., 2020)	6–12 monthly	Oral if: - Clinically inactive disease - Mild deficiency - Tolerate oral IV if: - Active IBD - Intolerant to oral - Hb <100 g/L (Dingnass et al., 2015)	Ensure bloods are interpreted in the context of disease activity: In active disease: Ferritin <100 μg/L considered deficient In remission: Ferritin <30 μg/L considered deficient
B12	If clinical symptoms suggestive of deficiency or anemia	Annually if risk factors present	If deficient: 1000 μg intramuscularly second daily for 2 weeks then monthly indefinitely If >20 cm small bowel resection: 1000 μg intramuscularly monthly indefinitely (Bischoff et al., 2020) High dose oral therapy and sub lingual routes of delivery being explored (National Institutes of Health, 2022)	Patients at increased risk of deficiency: - Ileal resection or active disease >20 cm sbnd UC with restorative proctocolectomy (Duerksen et al., 2006; Battat et al., 2014) in patients with low levels and no risk factors, alternative causes such as AI gastritis should be considered
Folate	No guidelines but consider in those with risk factors: - Malnutrition - Sulfasalazine or Methotrexate therapy - Disease duration >5 years - Pregnancy (Bischoff et al., 2020)	No guidelines, annual test alongside B12 or if a patient develops anemia or symptoms suggestive of deficiency	Oral	Monitor and supplement in women contemplating pregnancy and breastfeeding. In particular, supplement those taking Sulfasalazine
Vitamin D	3 monthly	Annually 3 monthly if deficient and replacing	Oral replacement, dose guided by degree of deficiency	
Calcium	Guided by dietary assessment rather than bloods (Houillier et al., 2005) Helpful assessment tool: (International Osteoporosis Foundation, 2022)	Serum calcium level does not reflect stores and does not fluctuate with changes in calcium intake	Diet: minimum recommended daily calcium intake is 1000 mg for men and pre-menopausal women and 1200–1500 mg for post-menopausal women and men >60 (NHMRC, 2006) Oral supplementation if dietary calcium intake is <800 mg/day, max dose 1000 mg/day. (Harbord et al., 2015)	
Zinc	3 monthly (Siva et al., 2017)	Annually	Dietary replacement Consider oral supplementation if deficient and unable to meet dietary requirements	Limited evidence, based on clinical experience

AI: Autoimmune Gastritis, IBD: Inflammatory Bowel Disease, IOF: International Osteoporosis Ulcerative Colitis, IV: Intravenous.

Battat, R., Kopylov, U., Szilagyi, A., et al., 2014. Vitamin B12 deficiency in IBD: prevelence, risk factors, evaluation and management. Inflamm. Bowel Dis. 20, 1120–1128; Bischoff, S., Escher, J., Hebuterne, X., et al., 2020. ESPEN practical guideline: clinical Nutrition in inflammatory bowel disease. Clin. Nutr. 39, 632–653; Dingnass, A., Gasche, C., Bettenworth, D., et al., 2015. European consensus on the diagnosis and management of iron deficiency and anemia in inflammatory bowel diseases. J. Crohn's Colitis 9, 211–222; Duerksen, D., Fallows, G., Bernstein, C., 2006. Vitamin B12 malabsorption in patients with limited ileal resection. Nutrition 22, 1210–2; Harbord, M., Annese, V., Vavrika, S., et al., 2015. The first European evidence based consensus on extra-intestinal manifestations in inflammatory bowel disease. J. Crohn's Colitis 10, 239–254; Houillier, P., Froissart, M., Maruani, G., et al., 2005. What serum calcium can tell us and what it can't. Nephrol. Dial. Transplant. 21, 29–32; International Osteoporosis Foundation, 2022. Calcium Calculator [Online]. Available: osteoporosis.foundation/educational-hub/topic/calcium-calculator [Accessed 30 March, 2022]; NHMRC 2006. Nutrient Reference Values for Australia and New Zealand. In: NHMRC (Ed.). Canberra, Australia; Services, U. D. o. H. a. H. 2022. Vitamin B12: Fact Sheet for Health Professionals. In: ONLINE, N. I. O. H. O. O. D. S. (Ed.). Online; Siva, S., Rubin, D., Gulotta, G., et al., 2017. Zinc deficiency is associated with poor clinical outcomes in patients with inflammatory bowel disease. Inflamm. Bowel Dis. 23, 152–157.

Vitamin D

The National Osteoporosis Society (UK) and ESPEN guidelines recommend that vitamin D be checked and replaced according to standard clinical cut offs in those with malabsorptive conditions such as IBD, in order to optimize bone health. This is particularly pertinent in those with additional risk factors for low Bone Mineral Density (BMD) and fracture risk such as female gender, Asian ethnicity and significant cumulative corticosteroid use (Francis et al., 2015; Bischoff et al., 2020).

Vitamin D is also thought to have immunomodulatory properties including antibacterial and anti-inflammatory actions on cells from the innate and adaptive immune system that modulate gastrointestinal inflammation as well as having a role in the maintenance of gastrointestinal mucosal barrier function (Fletcher et al., 2019). It is postulated that deficiency may contribute to increased IBD disease activity (Ham et al., 2014). This proposition was supported in a study of 70 UC patients in remission which demonstrated that those who had vitamin D levels lower than a cut off of 35 nmol/L were at a higher risk of flaring within the following 12-months than in patients with a level above this threshold (Gubatan et al., 2017). In practice, monitoring of Vitamin D on an at least annual basis for those in remission and 3-monthly in individuals who are at risk of deficiency is logical, with replacement therapy initiated as appropriate (Ananthakrishnan, 2016, Pirotta et al., 2019).

Zinc

Zinc is an essential trace element which is absorbed predominately in the proximal small intestine and is proposed to have immunoregulatory properties as well as decreasing the severity of colitis in animal models (Maares and Haase, 2020). There is evidence from a prospective registry of 996 patients with IBD that there is an association between zinc deficiency and disease related complications such as hospitalization and surgery and those who normalized their zinc over serial measurements had improved outcomes (Siva et al., 2017). Patients with CD, even in remission consume diets lower in zinc, with a prevalence of deficiency ranging from 15 to 40% (Taylor et al., 2018; Siva et al., 2017).

The question of causation vs. association remains, however, to replace deficiencies, if found, seems appropriate. There are no guidelines around testing for zinc deficiency in active disease nor remission.

Malnutrition in IBD key points:

- More common in CD than UC and relates to extent of disease
- Most prevalent in active disease
- A result of anorexia, reduced small bowel absorption
- Confers a poorer prognosis
- Screen at diagnosis and during periods of increased disease activity
- Broad macronutrient nutritional recommendations are the same for IBD as the general population however protein requirements are increased to 1.2–1.5 g/kg/d in active disease
- Oral nutrition, be it by food or EN formulas is preferred, PN should be reserved as last resort
- Consider micronutrient deficiencies: see **Table 3** for further detail.

Diet as therapy for active disease in IBD

There are several prescriptive dietary approaches that have been studied for use in clinical practice to manage IBD flares (See **Table 5** for summary of approaches, key inclusions and exclusions and relevant trial data). Due to the restrictive nature of these dietary regimens, they are largely reserved for the most unwell patients or those wanting to avoid the use of steroid-based medications, particularly children in whom steroids can impact growth trajectory (Ajebab et al., 2017).

Exclusive Enteral Nutrition (EEN)

EEN is a treatment strategy which has been shown to be effective in patients with CD, with sparse data in UC. EEN involves the use of a complete liquid formula that provides all the patient's nutritional requirements as the sole source of food. EEN can be given orally or via a nasogastric tube (NGT) (Mitrev et al., 2021; van Rheenen et al., 2020; Fardet et al., 2012). The proposed mechanism by which EEN improves outcomes in CD includes: improved nutritional parameters; modification of the gut microbiome with reduced antigenic stimulus driving inflammation; and increased expression of tight junction proteins between epithelial cells reversing the increased permeability to intestinal bacteria, which is seen in patients with CD (Mitrev et al., 2021).

The use of EEN in place of corticosteroids avoids the deleterious impacts of steroids, such as increased infection risk, decreased bone mineral density, hypothalamic-pituitary-adrenal axis suppression, mood and cognitive changes, altered glycemic control and acne (Patschan et al., 2001; Prantera and Marconi, 2013, Narula et al., 2018). EEN is usually given for 6–8 weeks during an acute flare of CD, with the aim of inducing remission. EEN may also be considered in the preoperative setting or in the context of fistulizing or stricturing disease as an alternative to corticosteroids which have been shown to increase the risk of perioperative infection and anastomotic break down with a higher doses conferring a greater risk of these complications (Subramanian et al., 2008).

The role of EEN in the treatment algorithm for IBD is influenced by the age of the patient, with significantly more data to support its use available in the pediatric population.

Table 5 Elimination diets proposed for use in the treatment of active CD.

Diet name	*Components excluded*	*Components allowed*	*Key study findings*
Autoimmune protocol (AIP)	Refined sugars, gluten, grains, legumes, nightshades (e.g., potato, tomato, capsicum, eggplant), dairy, eggs, coffee, alcohol, nuts, seeds, oils, food additives (Konijeti et al., 2017) 6-week elimination 5-week maintenance Staged re introduction	Fresh foods, bone broth, fermented food (Konijeti et al., 2017)	Prospective active disease, CD and UC N = 15 Concurrent medication use No control arm Outcomes: clinical parameters: improved biochemical and endoscopic assessments: trends to improvement did not reach statistical significance Adverse events: increase symptoms and SBO in patients with known small bowel strictures (Konijeti et al., 2017)
Anti-inflammatory Diet (IBD-AID)	Refined sugar, gluten containing grains, lactose, saturated fat (Olendzki et al., 2014)	Pre biotic and pro biotic foods including soluble fiber (e.g., encouraged: Leeks, onions, fermented foods, banana, oats, flaxseed meal, legume flour lean meat, poultry, fish) Textures altered according to symptoms and advanced as tolerated (Olendzki et al., 2014)	Retrospective case series N = 27 No control arm UC and CD active disease 4 weeks of intervention 24 out of 27 had a clinical response All patients able to discontinue at least one IBD medication and clinical parameters were improved No biochemical or endoscopic measures assessed (Olendzki et al., 2014)
Specific Carbohydrate Diet (SCD)	Complex carbohydrates, lactose Sucrose (e.g., table sugar), Maltose (e.g., wheat and grains) Grains Potato Soy Preservatives and processed foods (Cohen et al., 2014)	Monosaccharides e.g., honey Homemade fermented (lactose free) yogurt Low lactose dairy Most fresh fruits Most non starchy vegetables Eggs Oil (Cohen et al., 2014)	Patient surveys, retrospective case series and case reports in patients with UC and CD demonstrate improvement in clinical parameters of up to 42% at 12 months (Kakodkar et al., 2015; Suskind et al., 2016) Prospective pediatric cohort with active CD N = 9 Lead to improvement in clinical scores as well as capsule endoscopy activity scores at 12 and 52 weeks (Cohen et al., 2014) Adverse impacts: below recommended calcium intake (Kakodkar and Mutlu, 2017, Sasson et al., 2021)
Mediterranean Diet (MD)	Moderate consumption of: diary Alcohol: wine, with meals Low consumption of: Meat Saturated fat (Lewis et al., 2021; Hoevenaar-Blom et al., 2012)	Fresh fruit Vegetables Legumes Nuts Fish Wholegrains Olive oil predominant fat source (Lewis et al., 2021; Hoevenaar-Blom et al., 2012)	Equivalent to the more restrictive SCD in symptomatic improvement measured by CDAI (Lewis et al., 2021)

CD: Crohn's Disease, CDAI: Chron's Disease Activity Index, IBD: Inflammatory Bowel Disease, n: number, SBO: Small Bowel Obstruction, UC: Ulcerative Colitis.

Cohen, S., Gold, B., Oliva, S., et al., 2014. Clincal and mucosal improvement with specific carbohydrate diet in pediatric Crohn's disease J. Pediatr. Gastroenteronterol. Nutr. 59, 516–521; Hoevenaar-Blom, M., Nooyens, A., Kromhout, D., et al., 2012. Mediterranean style diet and 12-year incidence of cardiovascular diseases: the EPIC-NL cohort study. PLOS One 7, e45458; Kakodkar, S., Faroqui, A., Mikolatis, S., et al., 2015. The specific carbohydrate diet for inflammatory bowel disease: a case series. J. Acad. Nutr. Diet. 115, 1226–1232; Kakodkar, S., Mutlu, E.A., 2017. Diet as a therapeutic option for adult inflammatory bowel disease. Gastroenterol. Clin. N. Am. 46, 745–767; Konijeti, G., Kim, N., Lewis, J., et al., 2017. Efficacy of the autoimmune protocol diet for inflammatory bowel disease. Inflamm. Bowel Dis. 23, 2054–2060; Lewis, J., Sandler, R., Brotherton, C., et al., 2021. A randomized trial comparing the specific carbohydrate diet to a mediterranean diet in adults with Crohn's disease. Gastroenterology 161, 837–852; Olendzki, B., Silverstein, D., Persuitte, G., et al., 2014. An anti-inflammatory diet as treatment for inflammatory bowel disease: a case series report. Nutr. J. 13; Sasson, A.N., Ananthakrishnan, A.N., Raman, M., 2021. Diet in the treatment of inflammatory bowel diseases. Clin. Gastroenterol. Hepatol. 19, 425–435; Suskind, D., Wahbeh, G, Cohen, S. et al., 2016. Patients perceive clinical benefit with the specific carbohydrate diet for inflamamtory bowel disease. Dig. Dis. Sci. 61, 3255–3260.

Pediatric cohort

In pediatric patients EEN is first-line standard of care for induction of remission of CD and the effectiveness of EEN has been examined in several large meta-analyses (Swaminath et al., 2017; Narula et al., 2018; van Rheenen et al., 2020; Yu et al., 2019). EEN was found to be at least as equivalent to corticosteroids in achieving clinical remission and normalizing of biomarkers C-reactive protein (CRP) and Fecal Calprotectin but more effective at achieving mucosal healing and histological remission than corticosteroids. The use of EEN also avoids the additional pediatric specific corticosteroid side effects of slowed growth trajectory and development (Lo et al., 2020; Ajebab et al., 2017).

It should be noted that despite its clinical benefits, patients on EEN were more likely to withdraw from allocated treatment than those on corticosteroids (23% vs. 6%). The reasons cited include: unpalatable formulations, poor acceptance of NGT and side effects such as diarrhea and vomiting (Narula et al., 2018). There is some evidence that EEN is more effective in treating newly diagnosed CD rather than patients with established disease (Frivolt et al., 2014). A case review of 27 Australian children with CD managed with EEN reported 12/15 (80%) of newly diagnosed CD patients entering clinical remission after 8 weeks of EEN, compared with 7/12 (58%) of children with longstanding disease (Day et al., 2006).

Adult cohort

The data supporting the use of EEN in the adult population is more contentious (Sasson et al., 2021; Wall et al., 2013). A systematic review found that corticosteroids were more effective than EEN in induction of clinical remission (73% vs. 45%), though the evidence was low quality (Narula et al., 2018). Another older but relatively large ($n = 107$) multi-center study of adult patients with active Crohn's showed lower rates of clinical remission in the EEN group compared with corticosteroid plus sulfasalazine, (53% vs. 79%, $p < 0.001$) and the median time to remission was substantially longer with dietary therapy than steroids (30.7 days with EEN vs. 8.2 days with medical therapy) (Lochs et al., 1991). Further research with larger cohorts is required, particularly given the nature of the intervention, where blinding is not possible.

Complicated CD: fistulizing and stricturing disease

Several small, largely observational studies have demonstrated the benefit of EEN in fistulizing and stricturing disease. In patients with both luminal and enterocutaneous fistulae the delivery of EEN has been shown to induce both symptomatic and radiologic remission and even result in cutaneous fistula closure in up to 62% of patients (Yan et al., 2014; Hu et al., 2014; Yang et al., 2017). As a preoperative intervention EEN has been shown to decrease surgical time, length of hospital stay and risk of post-operative complications as well as reduce the risk of endoscopically detectable post-operative recurrence at 6 months (Li, G. et al., 2014, Li, Y. et al., 2015; Ge et al., 2019). A study examining the effectiveness of EEN prescribed for preoperative optimization in patients with stricturing or penetrating disease found it was so effective that up to 25% of patients no longer required surgical intervention and were able to remain controlled with medical management (Heerasing et al., 2017).

Use of EEN in UC

Early publications discounted the role of EEN in UC and as a result the literature has largely focused on its use in CD (Shaoul et al., 2018). Studies in pediatric populations have indicated that the clinical EEN response rates were no worse in patients with a colonic distribution of CD, giving promise to its use in UC, which has a solely colonic distribution (Buchanan et al., 2009; Shaoul et al., 2018). With this in mind, a recent single center open label randomized controlled trial ($n = 62$) looking at the use of EEN in addition to standard of care (SOC) in patients with Acute Severe Ulcerative Colitis (ASUC) was designed (Sahu et al., 2021). SOC was consistent with the ECCO guidelines and included intravenous corticosteroids and initiation of rescue therapy on day 5 (Infliximab or Ciclosporin) if Oxford criteria were met (Harbord et al., 2017). Decision and timing of colectomy followed a joint patient-medical-surgical team assessment. It was found that compared to SOC, the EEN group had a higher response rate to corticosteroid therapy that reached statistical significance on per protocol analysis (per protocol analysis corticosteroid failure in EEN group 19% vs. SOC 43%, $p = 0.04$). Response rates to rescue therapy rate of colectomy index admission were equivalent. Rates of subsequent hospitalization or colectomy at 6 months were lower in the EEN group (EEN 16%, SOC 39%, $p = 0.045$). This gives promise to the use of EEN in UC patients with ASUC, but more research in this group of patients is warranted.

Differences between EN formulas

Enteral nutrition has been used in various forms, all preparations contain protein, fat and carbohydrate, but are classified based on the protein constituent of the formula; primarily elemental, semi-elemental and polymeric (Limketkai et al., 2019). In addition to their carbohydrate and fat components, elemental formulas are a mixture of single amino acids that are broken down to their simplest form, theoretically making them easier to digest and entirely antigen free. Semi-elemental formulas are made with protein hydrolyzates and have a mean peptide chain length of 4–5 amino acids, which are partially predigested and too short for antigen recognition. Polymeric formulas contain whole protein from sources such as milk, egg and soy that do require digestion and have antigenic potential (Narula et al., 2018). Polymeric formulas are considered more palatable and cost effective (Mitrev et al., 2021). A meta-analysis of 11 trials including 378 patients (both adult and pediatric) demonstrated no significant differences between elemental and non-elemental (semi-elemental and polymeric) diets for active Crohn's disease in remission rates (64% vs. 62%) or adverse event rates (17% in both groups) (Narula et al., 2018). A study which examined the compositional analysis and evidence of efficacy of 61 formulas demonstrated that despite their significant variability in composition, they produced similar rates of

remission (Logan et al., 2020). This suggests that no specific formula type is superior to another, and formula selection should be based on local availability, cost and palatability.

EEN key points

- There is good evidence to show that EEN is superior to corticosteroids in treatment of active CD in children, as supported by the current guidelines.
- Minimal data for use in UC but further trials warranted
- It may be more effective in the earlier stages of the disease
- In adults, ENN does not appear to be superior to corticosteroids but its side effect profile is favorable and maybe considered in those who have contraindications to corticosteroids.
- EEN also appears to be beneficial in stricturing and fistulizing CD in the preoperative setting
- There does not appear to be a difference in clinical benefit between the type of EEN and the focus should be on the palatability to improve adherence to the regimen.

Partial Enteral Nutrition (PEN)

The limited tolerability of EEN led to the trial of approaches to improve adherence to therapy. Partial Enteral Nutrition (PEN) is a diet that combines EEN with an unrestricted oral diet. Unfortunately, PEN does not seem to be as effective in treating active disease as EEN. In an observational study of 90 children with active CD who were allocated to PEN with unrestricted diet, EEN or anti-TNF at the clinician's discretion, fecal calprotectin reductions to <250 μg/g and rates of clinical response were measured by Pediatric disease indices. A fecal calprotectin level of 250 μg/g was selected as an appropriate cutoff as this has previously been shown to correlate well with endoscopic remission in CD (D'Haens et al., 2012). The improvements in these parameters were less in the PEN group compared to those receiving EEN and anti-TNF therapy (FC to <250 PEN 14%, EEN 45%, anti-TNF 62%, $p < 0.001$, clinical response PEN 64%, EEN 88%, anti TNF 84%, $p = 0.08$) (Lee et al., 2015). In a randomized trial of 50 children with active CD, EEN was compared to PEN. The remission rates, were significantly lower in the PEN group (15% vs. 42% $p = 0.035$) and only the EEN group demonstrated improvements in biochemical indices of hemoglobin, ESR and albumin (Johnson et al., 2006).

Despite the lack of efficacy of PEN in active CD, it does show promise as an adjunct to medical therapy in both induction and maintenance phases of disease management. In a retrospective study of 47 children who were in remission after EEN, those who continued nocturnal supplementary PEN feeds in addition to an unrestricted daytime diet were compared to those who ceased the PEN feeds. The relapse rate at 12-months was significantly higher in the group that were on normal diet alone compared to the children who continued supplemental overnight feeds (79% vs. 42%, $p < 0.02$) (Wilschanski et al., 1996). It should be noted that this was in an era prior to the widespread availability of maintenance therapy. A randomized controlled study of 51 Japanese adults with CD in remission on maintenance therapy with mesalazine with or without azathioprine was conducted. One group were instructed to consume half their daily estimated calories via elemental diet, with the remainder via an unrestricted diet; while the control group continued unrestricted diet alone. They observed a lower relapse rate in the patients on PEN vs. standard diet (34.6% vs. 64.0% HR 0.4, 95% CI 0.16–0.98) (Takagi et al., 2006). In another prospective study of 39 adult patients with CD in remission and on PEN of 35–50% of their daily estimated energy intake plus standard medical therapy (5-ASA ± Azathioprine and Prednisone) for CD demonstrated a higher clinical remission rate at 12 months compared with patients on standard therapy in the intention to treat analysis 48% vs. 22% ($p = 0.00{,}003$) (Verma et al., 2000).

A meta-analysis of four studies in Japanese patients with active CD ($n = 342$) compared treatment using PEN (at least 600 kcal/day of formula) in addition to Infliximab induction therapy vs. Infliximab monotherapy without dietary restrictions. It found that the addition of PEN to Infliximab monotherapy resulted in a higher rate of clinical remission (69.4% vs. 45.4%, $p < 0.01$) which was sustained at the 1 year follow up (75.4% vs. 49.2%, $p < 0.01$) (Nguyen et al., 2015). These studies need to be confirmed in a broader patient cohort to determine the generalizability of this strategy.

PEN with restricted diet- Crohn's disease exclusion diet (CDED)

It was postulated that the unrestricted nature of the whole food diet component of the PEN approach may have contributed to its failure as a treatment strategy to induce remission, suggesting that the mechanism of EEN depends on the exclusion of specific components of a free diet as opposed to the sole beneficial effect of the enteral nutrition alone. To address this, the Crohn's disease exclusion diet (CDED) was designed. It aimed to exclude dietary components that are hypothesized to degrade the mucus layer and increase intestinal permeability. Excluded foods include foods containing animal fats, gluten, maltodextrin, emulsifiers, sulfites and certain monosaccharides. A CDED also mandates the consumption of certain fruits (e.g., bananas and peeled apple) and animal source protein as well as food containing resistant starch (e.g., peeled fresh potatoes) (Sigall-Boneh et al., 2014).

A CDED has shown promise in a difficult to treat patient cohort, many of whom had undergone previous resections and despite co therapy and dose escalation had failed biologic therapy. In a retrospective cohort of 21 adult and pediatric patients treated with combined enteral nutrition and CDED, at the end of 6-week treatment regimen 61.9% of patients were in clinical remission and biochemical parameters had also improved (Sigall Boneh et al., 2017). A follow up study in 74 pediatric patients with CD that looked at two groups over two phases was performed. The first group had CDED with 50% EN for 6-weeks followed by CDED

with 25% EN for 6 weeks. The second group had EEN for 6 weeks followed by an unrestricted diet and 25% EN for 6 weeks. The CDED group had improved treatment tolerability (97.5% vs. 73.6%, $p = 0.002$). At week 6 the response rates in both groups were 75%. At week 12 the CDED group had a remission rate of 75% and a sustained reduction in CRP and fecal calprotectin, whereas the group where free diet had been re-introduced had a remission rate of 45.1% (Levine et al., 2019). These studies are small but do give promise to the concept that PEN plus a more select diet may be a plausible treatment strategy in patients with CD.

PEN key points

- Unrestricted Diet with PEN is not as effective as EEN in induction of remission
- Unrestricted Diet with PEN is an effective adjuvant to standard therapy in improving response rates to biological therapy and in maintenance of remission
- The use of CDED in conjunction with PEN may improve the efficacy of this approach in active disease

Targeted CDED

Although EEN is successful at achieving mucosal healing, its patient acceptability is limited which can reduce its utility in the real-world setting. Using PEN with an unrestricted diet to induce remission does not seem as effective as EEN and the need to continue supplement drinks for an extended period may impact upon adherence and therefore efficacy of therapy.

To bridge this gap, Salvos and colleagues designed a prescriptive and personalized diet that recreates EEN by the exclusion of certain dietary components, such as gluten, lactose and alcohol, and matches others as close as possible to ordinary food, termed CD-TREAT (Svolos et al., 2019). They demonstrated in a group of healthy controls that when CD-TREAT was compared to EEN it was more palatable and elicited similar changes in the gut microbiome and metabolome to that of EEN. Its use also reduced ileitis in an animal model and reduced clinical and biochemical disease activity in a pilot study of 4 out of 5 children with active CD. The CD-TREAT diet approach shows promise but needs to be reproduced in a larger study before being used in routine clinical practice.

Other elimination diets

Other diets proposed to treat active disease include exclusion diets with varied degrees of restriction such as the Autoimmune Protocol (APD), Anti-Inflammatory Diet (IBD-AID), Specific Carbohydrate Diet (SCD) and Mediterranean Diet (MD), see **Table 5** for a summary of commonly used dietary approaches (Konijeti et al., 2017; Olendzki et al., 2014; Kakodkar and Mutlu, 2017, Cohen et al., 2014; Chiba et al., 2010; Bodini et al., 2019; Sasson et al., 2021; Lewis et al., 2021).

On the whole, they exclude refined sugars and grains which are postulated to provide substrates to pathogenic bacteria in the gut lumen and enable overgrowth (Sasson et al., 2021). In largely uncontrolled trials these diets have shown improved rates of clinical remission, however, these studies were small and many lacked a comparator group and there is a paucity of data demonstrating an improvement in objective markers of disease activity. It should also be noted that up to a third of patients decline to start these interventions due to the restrictive nature of the diet and they can lead to significant obsessive food focus, whereby adhering to the prescribed dietary inclusions and exclusions can impact on social life and psychological wellbeing (Olendzki et al., 2014; Kakodkar and Mutlu, 2017).

A randomized controlled trial comparing SCD with the Mediterranean Diet (MD) in a cohort of patients with mild to moderately active CD demonstrated equivalent improvements in disease activity indices as well as quality of life and fatigue measures (Lewis et al., 2021). The MD has the appeal of being less restrictive as well as offering broader health benefits in the prevention of cardiovascular disease, as such, if dietary therapy is being pursued, a MD may be a good wholistic approach (Hoevenaar-Blom et al., 2012).

In summary, there currently is not enough evidence to recommend elimination diets above standard therapies for active IBD. Elimination diets may be reasonable to consider as an adjunct to standard therapies in patients who are looking to add dietary strategies or in patients who decline medical therapy. As deficiencies in key vitamins have been detected in patients on these diets, patients should start them under professional nutritional supervision (Kakodkar and Mutlu, 2017).

Elimination diets key points

- Elimination diets have shown improvement in clinical parameters but largely in small, uncontrolled, non-randomized trials
- Evidence of improvement in objective markers of disease activity is lacking
- These approaches may be considered as adjunct or in patients declining SOC or those seeking a dietary adjunct to SOC under professional supervision to ensure nutritional adequacy.

Future directions and conclusions

While diet has a role in modulating the risk of developing IBD and is a factor that can be modified to manage disease, it is apparent that a one size fits all approach is not effective, nor appropriate. With improved knowledge and utilization of technology we may be

able to tailor an individual's treatment based on specific individual factors (termed "precision medicine"). In IBD, factors such as age, diet, genetics, microbiome, clinical parameters, circulating biomarkers in addition to patient preferences will need to be considered in order to design the most effective treatment strategy. The principles of this approach were demonstrated in a landmark work which explored the relationship between diet and postprandial glycemic response (PPGR) in a generally healthy population (Zeevi et al., 2015). The PPGRs of the study participants were accurately predicted using an algorithm that considered individual participant features including blood tests, diet, physical activity and the gut microbiota, among others. The researchers were then able to generate personalized dietary interventions that lead to improved PPGRs and observed shifts in the microbiome in a way that had previously been associated with lowered metabolic risk. Their findings have now been replicated in a different cohort (Mendes-Soares et al., 2019). This suggests that the application of machine learning principles in tailoring individual dietary treatment strategies in order to improve disease outcomes may be within reach (Laing et al., 2019).

In summary, components of the diet may exert proinflammatory and anti-inflammatory effects in the context of IBD which are mediated via influences on mucosal barrier function, direct immune system interactions and the gastrointestinal microbiome. It is clear that the use of diet in the prevention and treatment of IBD is a desirable treatment strategy.

Epidemiological evidence in humans from large population-based cohort studies suggest that breast feeding, fruit and vegetable derived fiber and n-3 PUFAs have a protective effect, while a diet rich in sugar, red and processed meat and ultra-processed food may increase a person's risk of developing IBD. Current evidence-based SOC in IBD is directed at suppressing the immune response, a strategy which carries risks of significant side effects. Dietary factors that may be protective against flares are high dietary n-3/n-6 FA ratio as well as adopting a semi vegetarian diet. Treatment strategies that have been successful in small cohorts of patients with active disease are EEN, PEN plus CDED and CD-TREAT. Macro- and micro-nutrient deficiencies should be screened for and adequately treated, particularly zinc and vitamin D which if maintained at adequate levels are associated with decreased risk of a disease flare.

In conclusion, dietary factors should gain more prominence as a preventative strategy as well as a therapeutic intervention as part of a personalized treatment approach that takes into account individual treatment factors and minimizes unwanted risks and side effects in order to achieve optimal disease control.

References

Agus, A., Denizot, J., Thevenot, J., et al., 2015. Western diet induces a shift in microbiota composition enhancing susceptibility to adherent-invasive *E. Coli* infection and intestinal inflammation. Sci. Rep. 6.

Ajebab, F., Choonara, I., Conroy, S., 2017. Systematic review of the toxicity of long-course oral corticosteroids in children. PLoS One 12.

Albenberg, L., Brensinger, C., Wu, Q., et al., 2019. A diet low in red and processed meat does not reduce rate of Crohn's disease flares. Gastroenterology 157, 128–136.

Altobelli, E., Del Negro, V., MatteoAngeletti, P., et al., 2017. Low-FODMAP diet improves irritable bowel syndrome symptoms: a meta-analysis. Nutrients 9.

Amre, D., D'Souza, S., Morgan, K., 2007. Imbalances in dietary consumption of fatty acids, vegetables and fruits are associated with risk for Crohn's disease in children. Am. J. Gastroenterol. 102, 2016–2025.

Ananthakrishnan, A., Khalili, H., Konijeti, G., et al., 2013. A prospective study of long-term intake of dietary fiber and risk of Crohn's disease and ulcerative colitis. Gastroenterology 145, 970–977.

Ananthakrishnan, A., Khalili, H., Song, M., et al., 2015. High school diet and the risk of Crohn's disease and ulcerative colitis. Inflamm. Bowel Dis. 21, 2311–2319.

Ananthakrishnan, A., 2016. Vitamin D and inflammatory bowel disease. Gastroenterol. Hepatol. 12, 513–515.

Andersen, V., Chan, S., Luben, R., et al., 2018. Fibre intake and the development of inflammatory bowel disease: a European prospective multicentre cohort study (EPIC-IBD). J. Crohns Colitis 24, 129–136.

Aykan, N., 2015. Red meat and colorectal cancer. Oncol. Rev. 10, 288.

Azad, M., Konya, T., Maughan, H., et al., 2013. Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. Can. Med. Assoc. J. 185, 385–394.

Aziz, I., Branchi, F., Pearson, K., et al., 2015. A study evaluating the bidirectional relationship between inflammatory bowel disease and self-reported non-celiac gluten sensitivity. Inflamm. Bowel Dis. 21, 847–853.

Bancil, A., Sandall, A., Rossi, M., et al., 2021. Food additive emulsifiers and their impact on gut microbiome, permeability and inflammation: mechanistic insights in inflammatory bowel disease. J. Crohns Colitis 15, 1068–1079.

Barnes, E., Nestor, M., Onyewadume, L., et al., 2017. High dietary intake of specific fatty acids increases risk of flares in patients with ulcerative colitis in remission during treatment with aminosalicylates. Clin. Gastroenterol. Hepatol. 15, 1390–1396.

Battat, R., Kopylov, U., Szilagyi, A., et al., 2014. Vitamin B12 deficiency in IBD: prevelence, risk factors, evaluation and management. Inflamm. Bowel Dis. 20, 1120–1128.

Bhattacharyya, S., Shumard, T., Xie, H., et al., 2017. A randomised trial of the effects of the no-carrageenan diet on ulcerative colitis disease activity. Nutr. Healthy Aging 4, 181–192.

Bischoff, S., Escher, J., Hebuterne, X., et al., 2020. ESPEN practical guideline: clinical nutrition in inflammatory bowel disease. Clin. Nutr. 39, 632–653.

Bodini, G., Zanella, C., Crespi, M., et al., 2019. A randomised 6-wk trial of low FODMAP diet in patients with inflammatory bowel disease. Nutrition 67–68.

Buchanan, E., Gaunt, W., Cardigan, T., et al., 2009. The use of exclusive enteral nutrition for induction of remission in children with Crohn's disease demonstrates that disease phenotype does not influence clinical remission. Aliment. Pharmacol. Therapeut. 30, 501–507.

Chapkin, R., Davidson, L., Ly, L., et al., 2007. Immunomodulatory effects of (n-3) fatty acids: putative link to inflammation and colon cancer. J. Nutr. 137, 200s–204s.

Chassaing, B., Van de Wiele, T., De Bodt, J., et al., 2017. Dietary emulsifiers directly alter human microbiota composition and gene expression ex vivo potentiating intestinal inflammation. Gut 66, 1414–1427.

Chiba, M., Abe, T., Tsuda, H., et al., 2010. Lifestyle-related disease in Crohn's disease: relapse prevention by a semi vegetarian diet. World J. Gastroenterol. 16, 2484–2495.

Chiba, M., Tsuji, T., Nakane, K., et al., 2015. High amount of dietary fiber not harmful but favorable for Crohn disease. Perm. J. 19, 58–61.

Cohen, S., Gold, B., Oliva, S., et al., 2014. Clincal and mucosal improvement with specific carbohydrate diet in paediatric Crohn's disease. J. Pediatr. Gastroenterol. Nutr. 59, 516–521.

Cole, C., Ownby, D., 2017. The infant gut bacterial microbiota and risk of pediatric asthma and allergic diseases. Transl. Res. 179, 60–70.

Copaci, I., Chira, C., Rovinaru, I., et al., 2000. Maintenance of remission of ulcerative colitis (UC): mesalamine, dietary fiber, S. boulardi. Dig. Liver Dis. 32 (Suppl. I).

Cornish, J., Tan, E., Simillis, C., et al., 2008. The risk of oral contraceptives in the etiology of inflammatory bowel disease: a meta-analysis. Am. J. Gastroenterol. 103, 2394–2400.
Cox, S., Prince, A., Myers, C., et al., 2017. Fermentable carbohydrates [FODMAPs] exacerbate functional gastrointestinal symptoms in patients with inflammatory bowel disease: a randomised, double blind, placebo controlled cross-over, re-challenge trial. J. Crohns Colitis 11, 1420–1429.
Cox, S., O'Lindsay, J., Fromentin, S., 2020. Effects of low FODMAP diet on symptoms, fecal microbiome, and markers of inflammation in patients with quiescent inflammatory bowel disease in a randomized trial. Gastroenterology 158, 176–188.
Cox, S., Sandall, A., Smith, L., et al., 2021. Food additive emulsifiers: a review of their role in foods, legislation and classifications, presence in food supply, dietary exposure and safety assessment. Nutr. Rev. 79, 726–741.
Day, A.S., Whitten, K.E., Lemberg, D.A., et al., 2006. Exclusive enteral feeding as primary therapy for Crohn's disease in Australian children and adolescents: a feasible and effective approach. J. Gastroenterol. Hepatol. 21, 1609–1614.
Derbyshire, E., 2016. Flexitarian diets and health: a review of the evidence based literature. Front. Nutr. 3, 55.
Desai, M., Seekatz, A., Koropatkin, N., et al., 2016. Dietary-fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. Cell 167, 1339–1353.
Detsky, A., McLaughlin, J., Baker, J., et al., 1987. What is subjective global assessment of nutritional status? J. Parenter. Enteral Nutr. 11, 8–13.
Devkota, S., Wang, Y., Musch, M., et al., 2012. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice. Nat. Microbiol. 487.
D'Haens, F.,M., Vermeire, S., et al., 2012. Fecal calprotectin is a surrogate marker for endoscopic lesions in inflammatory bowel disease. Inflamm. Bowel Dis. 18, 2218–2224.
Dingnass, A., Gasche, C., Bettenworth, D., et al., 2015. European consensus on the diagnosis and management of iron deficiency and anaemia in inflammatory bowel diseases. J. Crohns Colitis 9, 211–222.
Dinu, M., Abbate, R., Gensini, G., et al., 2017. Vegetarian, vegan diets and multiple health outcomes: a systematic review with meta-analysis of observational studies. Crit. Rev. Food Sci. Nutr. 57, 3640–3649.
Dong, C., Mahamat-Saleh, Y., Racine, A., et al., 2020. OP17: protein intakes and risk of inflammatory bowel disease in the European Prospective Invesigation into Cancer and Nutrition cohort (EPIC-IBD). In: 15th Congress of ECCO, vol. 14.
Duerksen, D., Fallows, G., Bernstein, C., 2006. Vitamin B12 malabsorption in patients with limited ileal resection. Nutrition 22, 1210–1212.
Elia, M., 2003. The "MUST" Report. Nutritional Screening of Adults: A Multidiciplinary Responsibility Redditch. UK BAPEN: Multidiciplinary Advisory Group.
Etzel, J., Larson, M., Anawait, B., et al., 2011. Assessment and management of low bone mineral density in inflammatory bowel disease and performance of professional society guidelines. Inflamm. Bowel Dis. 17, 2122–2129.
Fardet, L., Petersen, I., Nazareth, I., 2012. Suicidal behaviour and severe neuropsychiatric disorders following glucocorticoid therapy in primary care. Am. J. Psychiatr. 169.
Fernandez-Banares, F., Hinojosa, J., Sanchez-Lombrana, J., et al., 1999. Randomized clinical trial of *Plantago ovata* seeds (dietary fiber) as compared with mesalamine in maintaining remission in ulcerative colitis. Spanish group for the study of Crohn's Disease and Ulcerative Colitis (GETECCU). Am. J. Gastroenterol. 94, 427–433.
Fletcher, J., Cooper, S., Ghosh, S., et al., 2019. The role of vitamin D in inflammatory bowel disease: mechanism to management. Nutrients 11.
Food Standards Australia New Zealand, 2016. Folic acid fortification. In: N. Z. (Ed.), New Zealand Food Safety. Government Publisher, Wellington, New Zealand.
Forouhi, N., Krauss, R., Taubes, G., et al., 2018. Dietary fat and cardiometabolic health: evidence, controversies and consensus for guidance. Br. Med. J. 361.
Francis, R., Aspray, T., Bowring, C., et al., 2015. National Osteoporosis Society practical clinical guideline on Vitamin D and bone health. Maturitas 80, 119–121.
Franzosa, E., Sirota-Madi, A., Avila-Pacheco, J., et al., 2019. Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat. Microbiol. 4, 293–305.
Frivolt, K., Schwerd, T., Werkstetter, K.J., et al., 2014. Repeated exclusive enteral nutrition on the treatment of paediatric Crohn's disease: predictors of efficacy and outcome. Aliment. Pharmacol. Therapeut. 39, 1398–1407.
Frolkis, A., Dykeman, J., Negron, M., et al., 2013. Risk of surgery for inflammatory bowel diseases has decreased over time: a systematic review and meta-analysis of population-based studies. Gastroenterology 145, 996–1006.
Gajendran, M., Umapathy, C., Loganathan, P., et al., 2016. Analysis of hospital based emergency department visits for inflammatory bowel disease in the USA. Dig. Dis. Sci. 61, 389–399.
Garry, P.J., Vellas, B., 1999. Practical and validated use of the mini nutritional assessment in geriatric evaluation. Nutr. Clin. Care 2, 146–154.
Ge, J., Han, T., Liu, J., et al., 2015. Meat intake and risk of inflammatory bowel disease: a meta-analysis. Turk. J. Gastroenterol. 26, 492–497.
Ge, X., Tang, S., Yang, X., et al., 2019. The role of exclusive enteral nutrition in the preoperative optimization of laparoscopic surgery for patients with Crohn's disease: a cohort study. Int. J. Surg. 65, 39–44.
Gill, S., Rossi, M., Bajka, B., et al., 2021. Dietary fibre in gastrointestinal health and disease. Nat. Rev. Gastroenterol. Hepatol. 18, 101–116.
Gomollon, F., Gargallo, C., Munoz, J., et al., 2017. Oral Cyanocobalamin is effective treatment of vitamin B12 deficiency in Crohn's disease. Nutrients 9.
Goncalves, P., Ricardo Araujo, J., Di Santo, J., et al., 2018. A cross-talk between microbiota derived short chain fatty acids and the host mucosal immune system regulates intestinal homeostasis and inflammatory bowel disease. Inflamm. Bowel Dis. 24, 558–572.
Gonzalez-Becerra, K., Ramos-Lopez, O., Barron-Cabrera, E., et al., 2019. Fatty acids, epigenetic mechanisms and chronic diseases: a systematic review. Lipids Health Dis. 18.
Gordon, H., Trier Moller, F., Andersen, V., et al., 2015. Heritability in inflammatory bowel disease. Inflamm. Bowel Dis. 21, 1428–1434.
Guan, Q., 2019. A comprehensive review and update on the pathogenesis of inflammatory bowel disease. J. Immunol. Res. 7247238.
Gubatan, J., Mitsuhashi, S., Zenlea, T., et al., 2017. Low serum vitamin D during remission increases risk of clinical relapse in patients with ulcerative colitis. Clin. Gastroenterol. Hepatol. 15, 240–246.
Ham, M., Longhi, M., Lahiff, C., et al., 2014. Low serum vitamin D levels in adults with Crohn's disease are responsive to disease activity and treatment. Inflamm. Bowel Dis. 20, 856–860.
Harbord, M., Annese, V., Vavrika, S., et al., 2015. The first European evidence based consensus on extra-intestinal manifestations in inflammatory bowel disease. J. Crohns Colitis 10, 239–254.
Harbord, M., Elikim, R., Bettenworth, D., et al., 2017. Third European evidence based consensus on diagnosis and management of ulcerative colitis. Part 2: current management. J. Crohns Colitis 11, 769–784.
Haskey, N., Pena-Sanchez, J., Jones, J., et al., 2018. Development of a screening tool to detect nutrition risk in patients with inflammatory bowel disease. Asia Pac. J. Clin. Nutr. 27, 756–762.
Heerasing, N., Thompson, B., Hendy, G., et al., 2017. Exclusive enteral nutrition provides an effective bridge to safer interval elective surgery for adults with crohn's disease. Aliment. Pharmacol. Therapeut. 45, 660–669.
Herfarth, H., Martin, C., Sandler, R., et al., 2014. Prevalence of a gluten-free diet and improvement of clinical symptoms in patients with inflammatory bowel diseases. Inflamm. Bowel Dis. 20, 1194–1197.
Hoevenaar-Blom, M., Nooyens, A., Kromhout, D., et al., 2012. Mediterranean style diet and 12-year incidence of cardiovascular diseases: the EPIC-NL cohort study. PLoS One 7, e45458.
Houillier, P., Froissart, M., Maruani, G., et al., 2005. What serum calcium can tell us and what it can't. Nephrol. Dial. Transplant. 21, 29–32.
Hu, D., Ren, J., Wang, G., et al., 2014. Exclusive enteral nutritional therapy can relieve inflammatory bowel stricture in crohn's disease. J. Clin. Gastroenterol. 48, 790–795.
Huang, S., Ma, J., Zhu, M., et al., 2017. Status of serum vitamin B12 and folate in patients with inflammatory bowel disease in China. Intest. Res. 15, 103–108.
Huang, Y., Cao, H., CHen, Z., et al., 2021. Red and processed meat consumption and cancer outcomes: umbrella review. Food Chem. 356.
International Osteoporosis Foundation, 2022. Calcium Calculator. Available. (Accessed 30 March 2022).
Jansen, I., Prager, M., Valentini, L., et al., 2016. Inflammation-driven malnutrition: a new screening tool predicts outcome in Crohn's disease. Br. J. Nutr. 116, 1061–1067.

Jantchou, P., Morois, S., Clavel-Chapelon, F., et al., 2010. Animal protein intake and risk of inflammatory bowel disease: the E3N prospective study. Am. J. Gastroenterol. 105, 2195–2201.
John, S., Luben, R., Shrestha, S., et al., 2010. Dietary n-3 polyunsaturated fatty acids and the aetiology of ulcerative colitis: a UK prospective cohort study. Eur. J. Gastroenterol. Hepatol. 22, 602–606.
Johnson, T., Macdonald, S., Hill, S.M., Thomas, A., Murphy, M.S., 2006. Treatment of active Crohn's disease in children using partial enteral nutrition with liquid formula: a randomised controlled trial. Inflamm. Bowel Dis. 55, 336–361.
Jowett, S., Seal, C., Pearce, M., et al., 2004. Influence of dietary factors on the clinical course of ulcerative colitis: a prospective cohort study. Gut 53, 1479–1484.
Kakodkar, S., Farooqui, A., Mikolatis, S., Mutlu, E., 2015. The specific carbohydrate diet for inflammatory bowel disease: a case series. J. Acad. Nutr. Diet. 115 (8), 1226–1232.
Kakodkar, S., Mutlu, E.A., 2017. Diet as a therapeutic option for adult inflammatory bowel disease. Gastroenterol. Clin. N. Am. 46, 745–767.
Kondrup, J., Rasmussen, H., Hamberg, O., et al., 2003. Nutritional risk screening (NRS 2002): a new method based on an analysis of controlled clinical trials. Clin. Nutr. 22, 321–336.
Konijeti, G., Kim, N., Lewis, J., et al., 2017. Efficacy of the autoimmune protocol diet for inflammatory bowel disease. Inflamm. Bowel Dis. 23, 2054–2060.
Lacy, B., Pimentel, M., Breener, D., et al., 2021. ACG clinical guideline: management of irritable bowel syndrome. Am. J. Gastroenterol. 116, 17–44.
Laffin, M., Fedorak, R., Zalaski, A., et al., 2019. A high-sugar diet rapidly enhances susceptibility to colitis via depletion of luminal short-chain fatty acids in mice. Sci. Rep. 9.
Laing, B., Lim, G., Ferguson, L., 2019. A personalised dietary approach- a way forward to manage nutrient deficiency, effects of the western det, and food intolerances in inflammatory bowel disease. Nutrients 11.
Lamb, C., Kennedy, N., Raine, T., et al., 2019. British Society of Gastroenterology consensus guidelines on the management of inflamamtory bowel disease in adults. Gut 68, s1–s106.
Lambert, K., Pappas, D., Miglioretto, C., et al., 2021. Systematic review with meta-analysis: dietary intake in adults with inflammatory bowel disease. Aliment. Pharmacol. Therapeut. 56, 742–754.
Le Leu, R.K., Young, G.P., 2007. Fermentation of starch and protein in the colon: implications for genomic instability. Cancer Biol. Ther. 2, 259–260.
Lee, D., Baldassano, R.N., Otley, A.R., et al., 2015. Comparative effectiveness of nutritional and biological therapy in North American children with active Crohn's disease. Inflamm. Bowel Dis. 21, 1786–1794.
Levine, A., Wine, E., Assa, A., et al., 2019. Crohn's disease exclusion diet plus partial enteral nutrition induces sustained remission in a randomised controlled trial. Gastroenterology 157, 440–450.
Lewin, M., Bailey, N., Bandaletova, T., et al., 2006. Red meat enhances the colonic formation of the DNA adduct 06-carboxymethyl guanine: implications for colorectal cancer risk. Cancer Res. 66, 1859–1865.
Lewis, J., Sandler, R., Brotherton, C., et al., 2021. A randomised trial comparing the specific carbohydrate diet to a mediterranean diet in adults with Crohn's disease. Gastroenterology 161, 837–852.
Li, G., Ren, J., Wang, G., et al., 2014. Preoperative exclusive enteral nutrition reduces the postoperative septic complications of fistulizing Crohn's disease. Eur. J. Clin. Nutr. 68, 441–446.
Li, F., Liu, X., Wang, W., et al., 2015a. Consumption of vegetables and fruit and the risk of inflammatory bowel disease: a meta-analysis. Eur. J. Gastroenterol. Hepatol. 27, 623–630.
Li, Y., Zuo, L., Weiming, Z., et al., 2015b. Role of exclusive enteral nutrition in the preoperative optimization of patients with Crohn's disease following immunosuppressive therapy. Medicine 94, e487.
Li, S., Ney, M., Eslamparast, T., et al., 2019. Systematic review of nutrition screening and assessment in inflammatory bowel disease. World J. Gastroenterol. 25, 3823–3837.
Li, D., Cui, M., Tan, F., et al., 2021. High red meat intake exacerbates dextran sulfate induced colitis by altering gut microbiota in mice. Front. Nutr. 8.
Limdi, K., Aggarwal, D., McLaughlin, J., et al., 2016. Dietary practices and beliefs in patients with inflammatory bowel disease. Inflamm. Bowel Dis. 22, 164–170.
Limketkai, B., Sepulveda, R., Hing, T., et al., 2018. Prevalence and factors associated with gluten sensitivity in inflammatory bowel disease. Scand. J. Gastroenterol. 53, 147–151.
Limketkai, B., Shah, N., Sheikh, G., et al., 2019. Classifying enteral nutrition: tailored for clinical practice. Nutr. Obes. 21.
Lo, B., Holm, J.P., Vester-Andersen, M.K., et al., 2020. Incidence, risk factors and evaluation of osteoporosis in patients with inflammatory bowel disease: a Danish population-based inception cohort with 10 years of follow-up. J. Crohns Colitis 14, 904–914.
Lochs, H., Steinhardt, H.J., Klaus-Wentz, B., et al., 1991. Comparison of enteral nutrition and drug treatment in active crohn's disease. Results of the European cooperative Crohn's disease study. IV. Gastroenterology 101, 881–888.
Logan, M., Gkikas, K., Svolos, V., et al., 2020. Analysis of 61 exclusive entral nutrition formulas used in the management of active Crohn's disease- new insights into dietary disease triggers. Aliment. Pharmacol. Therapeut. 51, 935–947.
Lomer, M., Harvey, R., Evans, S., et al., 2001. Efficacy and tolerability of a low microparticule diet in a double blind randomized pilot study in Crohn's disease. Eur. J. Gastroenterol. Hepatol. 13, 101–106.
Lomer, M., Grainger, S., Ede, R., et al., 2005. Lack of efficacy of a reduced microparticle diet in multicentred trial of patients with active disease. Eur. J. Gastroenterol. Hepatol. 17, 377–384.
Maares, M., Haase, H., 2020. A guide to human zinc absorption: general overview and recent advances of in vitro intestinal models. Nutrients 12.
Mahid, S., Minor, K., Soto, R., et al., 2006. Smoking and inflammatory bowel disease: a meta-analysis. Mayo Clin. Proc. 81, 1462–1471.
McClements, D.J., 2016. Food Emulsions: Principles, Practices and Techniques. NW CRC Press, Broken Sound Parkway.
Mendes-Soares, H., Raveh-Sadka, T., Azulay, S., et al., 2019. Model of personalised post prandial glycaemic response to food for an Israeli cohort predicts responses in Midwestern American individuals. Am. J. Clin. Nutr. 110, 63–75.
Michaelsson, K., Melhus, H., Warensjo Lemming, E., et al., 2013. Long term calcium intake and rates of all cause and cardiovascular mortality: community based prospective longitudinal cohort study. Br. Med. J. 346, f228.
Mitrev, N., Huang, H., Hannah, B., Kariyawasam, V.C., 2021. Review of exclusive entral therapy in adult Crohn's disease. Br. Med. J. Open Gastroenterol. 8, e000745.
M'Koma, A., 2013. Inflammatory bowel disease: an expanding global health problem. Clin. Med. Insights Gastroenterol. 33–47.
Monteiro, C., Cannon, G., Lawrence, M., et al., 2019. Ultra-processed Foods, Diet Quality, and Health Using the NOVA Classification System. FAO, Rome.
Montrose, D., Nishiguchi, R., Basu, S., et al., 2021. Dietary fructose alters the composition, localisation and metabolism of gut microbiota in association with worsening colitis. Cell. Mol. Gastroenterol. Hepatol. 11, 525–550.
M'Rabet, L., Vos, A., Boehm, G., et al., 2008. Breast-Feeding and its role in early development of the immune system in infants: consequences for health later in life. J. Nutr. 138, 1782S–1790S.
Muhomah, T., Nishino, N., Katsumata, E., et al., 2019. High-fat diet reduces the level of secretory immunoglobilin A coating of commensal gut microbiota. Biosci. Microb. Food Health 38, 55–64.
Narula, N., Dhillon, A., Zhang, D., et al., 2018. Enteral nutritional therapy for induction of remission in crohn's disease. Cochrane Database Syst. Rev. 1.
Narula, N., Wong, E., Dehghan, M., et al., 2021. Association of ultra-processed food intake with risk of inflammatory bowel disease: prospective cohort study. Br. Med. J. 374.
National Institutes of Health, 2022. Dietary Supplement Fact Sheet: Vitamin B 12. http://ods.od.nih.gov/factsheets/VitaminB12-HealthProfessional/. (Accessed 3 June 2022).
Ng, S., Yun Shi, H., Hamidi, N., et al., 2017. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet 390, 2769–2778.
Nguyen, G., Laveist, T., Brant, S., 2007. The utilization of parenteral nutrition during the in-patient management of inflammatory bowel disease in the United States: a national survey. Aliment. Pharmacol. Therapeut. 26, 1499–1507.

Nguyen, G., Munsell, M., Harris, M., 2008. Nationwide prevalence and prognostic significance of clinically diagnosable protein-calorie malnutrition in hospitalized inflamamtory bowel disease patients. Inflamm. Bowel Dis. 14, 1105–1111.
Nguyen, D., Palmer, L., Nguyen, E., et al., 2015. Specialised enteral nutrition therapy in crohn's disease patients on maintinance infliximab therapy. Therapeut. Adv. Gastroenterol. 8, 168–175.
NHMRC and NZMOH working party, 2006. Nutrient Reference Values for Australia and New Zealand: Including Recommended Dietary Intakes. National Health and Medical Research Council of Australia, Canberra, Australia.
NHMRC working committee, 2013. Australian Dietary Guidelines: Eat for Health. National Health and Medical Research Council, Canberra, Australia.
Ni, J., Wu, G., Albenberg, L., et al., 2017. Gut microbiota and IBD: causation or correlation. Nat. Rev. Gastroenterol. Hepatol. 14, 573–584.
Nickerson, K., McDonald, C., 2012. Crohn's disease-associated adherent-invasive *Escherichia coli* adhesion is enhanced by exposure to the ubiquitous dietary polysaccharide maltodextrin. PLoS One 7.
Olendzki, B., Silverstein, D., Persuitte, G., et al., 2014. An anti-inflammatory diet as treatment for inflammatory bowel disease: a case series report. Nutr. J. 13.
Pan, Y., Liu, Y., Guo, H., et al., 2017. Associations between folate and vitamin B12 levels and inflammatory bowel disease: a meta-analysis. Nutrients 9.
Parigi, S., Eldh, M., Larssen, P., et al., 2015. Breast milk and solid food shaping intestinal immunity. Front. Immunol. 6.
Park, S.J.,Y., 2019. Genetic studies of inflammatory bowel disease- focusing on Asian patients. Cells 8.
Patschan, D., Loddenkemper, K., Buttgereit, F., 2001. Molecular mechanisms of glucocorticoid-induced osteoporosis. Bone 29, 498–500.
Patterson, E., Wall, R., Fitzgerald, G., et al., 2012. Health implications of high dietary omega-6 polyunsaturated fatty acids. J. Nutr. Metab. 539426.
Piovani, D., Danese, S., Peyrin-Biroulet, L., et al., 2019. Environmental factors for inflammatory bowel diseases: an umbrella review of meta-analysis. Gastroenterology 157, 647–659.
Pirotta, M., Doogue, M., Dooley, S., et al., 2019. Therapeutic Guidelines: Bone and Metabolism: Vitamin D Deficiency Online.
Porter, R., Kalla, R., Ho, G., et al., 2020. Ulcerative colitis: recent advances in the understanding of disease pathogenesis. F1000Research 9.
Prada Venegas, D., De la Fuente, M., Landskron, G., et al., 2019. Short Chain Fatty Acids (SCFAs)- mediated gut epithelial and immune regulation and its relevance for inflamamtion. Front. Immunol. 10.
Prantera, C., Marconi, S., 2013. Glucocorticoids in the treatment of inflammatory bowel disease and approaches to minimizing systemic activity. Therapeut. Adv. Gastroenterol. 6, 137–156.
Prince, A., Myers, C., Joyce, T., et al., 2016. Fermentable carbohydrate restriction (Low FODMAP Diet) in clinical practice improves functional gastrointestinal symptoms in patients with inflammatory bowel disease. Inflamm. Bowel Dis. 22, 1129–1136.
Racine, A., Carbonnel, F., Chan, S., et al., 2015. Dietary patterns and risk of inflammatory bowel disease in Europe: results from the EPIC study. Inflamm. Bowel Dis. 22, 345–354.
Raine, T., Bonvas, S., Burisch, J., 2021. ECCO guidelines on therapeutics in ulcerative colitis: medical treatment. J. Crohns Colitis 1–16.
Rhys-Jones, D., 2021. Gluten and IBS. Monash University. Available. monashfodmap.com. (Accessed 30 March 2022).
Roberts, C., Rushworth, S., Richman, E., et al., 2013. Hypothesis: increased consumption of emulsifiers as an explanation for the rising incidence of Crohn's disease. J. Crohns Colitis 7, 338–341.
Rogier, E., Frantz, A., Bruno, M., et al., 2014. Lessons from mother: long-term impact of antibodies in breast milk on the gut microbiota and intestinal immune system of breastfed offspring. Gut Microb. 5, 663–668.
Ross, C., 2014. Modern Nutrition in Health and Disease. Wolters Kluwer Heath/Lippincott Willaims & Wilkins, Philadelphia.
Ruiz, P., Moron, B., Becker, H., et al., 2017. Titanium dioxide nanoparticles exacerbate DSS-induced colitis: role of the NLRP3 inflammasome. Gut 66, 1216–1224.
Russell, L., Balart, M., Serrano, P., et al., 2022. The complexities of approaching nutrition in inflammatory bowel disease: current recommendations and future directions. Nutr. Rev. 10, 215–229.
Sahu, P., Kedia, S., Vuyyuru, S., et al., 2021. Randomised clinical trial: exclusive enteral nutrition versus standard of care for acute severe colitis. Aliment. Pharmacol. Therapeut. 53, 567–576.
Sandhu, A., Mosli, M., Yan, B., et al., 2016. Self-screening for malnutrition risk in outpatient inflammatory bowel disease patients using the malnutrition universal screening tool (MUST). J. Parenter. Enteral Nutr. 40, 507–510.
Santoru, M., Piras, C., Murgia, A., et al., 2017. Cross sectional evaluation of the gut-microbiome metabolome axis in an Italian cohort of IBD patients. Sci. Rep. 7.
Sasson, A.N., Ananthakrishnan, A.N., Raman, M., 2021. Diet in the treatment of inflammatory bowel diseases. Clin. Gastroenterol. Hepatol. 19, 425–435.
Shaoul, R., Brown, S., Day, A., et al., 2018. Dietary treatment for ulcerative colitis. J. Pediatr. Gastroenterol. Nutr. 66.
Shoda, R., Matseuda, K., Yamato, S., et al., 1996. Epidemiologic analysis of Crohn disease in Japan: increased dietary intake of n-6 polyunsaturated fatty acids and animal protein relates to the increased incidence of Crohn disease in Japan. Am. J. Clin. Nutr. 63, 741–745.
Sigall Boneh, R., Shabat, C.S., Yanai, H., et al., 2017. Dietary therapy with Crohn's disease exclusion diet is a successful strategy for induction of remission in children and adults failing biological therapy. J. Crohns Colitis 11, 1205–1212.
Sigall-Boneh, R., Pfeffer-Gik, T., Segal, I., et al., 2014. Partial enteral nutrition with a Crohn's disease exclusion diet is effective for induction and remission in children and young adults with Crohn's disease. Inflamm. Bowel Dis. 20, 1353–1360.
Silveira, A., Ferreira, A., de Oliveria, M., et al., 2017. Preventative rather than therapeutic treatment with high fiber diet attenuates clinical and inflammatory markers of acute and chronic DSS-induced colitis in mice. Eur. J. Nutr. 56, 179–191.
Siva, S., Rubin, D., Gulotta, G., et al., 2017. Zinc deficiency is associated with poor clinical outcomes in patients with inflammatory bowel disease. Inflamm. Bowel Dis. 23, 152–157.
Sokol, H., Seksik, P., Furet, J., et al., 2009. Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. Inflamm. Bowel Dis. 15, 1183–1189.
Stewart, C., Ajami, N., O'Brien, J., et al., 2018. Temporal development of the gut microbiome in early childhood from the TEDDY study. Nat. Microbiol. 562, 583–588.
Subramanian, V., Saxena, S., Kang, J., et al., 2008. Peroperative steroid use and risk of postoperative complications in patients with inflammatory bowel disease undergoing abdominal surgery. Am. J. Gastroenterol. 103, 2373–2381.
Sugimura, T., Wakabayashi, K., Nakagama, H., et al., 2004. Heterocyclic amines: mutagens/carcinogens produced during cooking of meat and fish. Cancer Sci. 95, 290–299.
Sumi, R., Nakajima, K., Iijima, H., et al., 2016. Influence of nutritonal status on the therapeutic effect of infliximab in patients with Crohn's disease. Surg. Today 46, 922–929.
Suskind, D., Wahbeh, G., Cohen, S., et al., 2016. Patients perceive clinical benefit with the specific carbohydrate diet for inflammatory bowel disease. Dig. Dis. Sci. 61 (11), 3255–3260.
Svolos, V., Hansen, R., Nichols, B., et al., 2019. Treatment of active Crohn's disease with an ordinary food-based diet that replicates exclusive enteral nutrition. Gastroenterology 156, 1354–1367.
Swaminath, A., Feathers, A., Anathakrishnan, A.N., et al., 2017. Systematic review with meta analysis: enteral nutrition therapy for the induction of remission in paediatric crohn's disease. Aliment. Pharmacol. Therapeut. 46, 645–656.
Takagi, S., Utsunomiya, K., Kuriyama, S., et al., 2006. Effectiveness of an "half elemental diet" as maintenance therapy for Crohn's disease: a randomized-controlled trial. Aliment. Pharmacol. Therapeut. 24, 1333–1340.
Takaishi, H., Matsuki, T., Nakazawa, A., et al., 2008. Imbalance in intestinal microflora constitution could be involved in the pathogenesis of inflammatory bowel disease. Int. J. Med. Microbiol. 298, 463–472.
Taylor, L., Almutairdi, A., Shommu, N., et al., 2018. Cross-sectional analysis of overall dietary intake and mediterranean dietary pattern in patients with Crohn's disease. Nutrients 10.

Tjonneland, A., Overvad, K., Bergmann, M., et al., 2009. Linoleic acid, a dietary n-6 polyunsaturated fatty acid, and the aetiology of ulcerative colitis: a nexted case-control study within a European prospective cohort study. Gut 58, 1606–1611.
Torres, J., Mehandru, S., Colombel, J., et al., 2017. Crohn's disease. Lancet 389, 1741–1755.
Torres, J., Bonovas, S., Doherty, G., et al., 2019. ECCO guidelines on therapeutics in Crohn's disease: medical treatment. J. Crohns Colitis 14, 4–22.
Uchiyama, K., Nakamura, M., Odahara, S., et al., 2010. N-3 polyunsaturated fatty acid diet therapy for patients with inflammatory bowel disease. Inflamm. Bowel Dis. 16, 1696–1707.
Ungaro, R., Bernstein, C., Gearry, R., et al., 2014. Antibiotics associated with increased risk of new-onset Crohn's disease but not ulcerative colitis: a meta-analysis. Am. J. Gastroenterol. 109, 1728–1738.
Ungaro, R., Mehandru, S., Allen, P., et al., 2017. Ulcerative colitis. Lancet 389, 1756–1770.
Valdes, A., Walter, J., Segal, E., et al., 2018. Role of the gut microbiota in nutrition and health. Br. Med. J. 361 (Suppl. 1).
van Rheenen, P.F., Aloi, M., Assa, A., et al., 2020. The medical management of paediatric crohn's disease: an ECCO-ESPGHAN guideline update. J. Crohns Colitis 15, 171–194.
van Staa, T., Cooper, C., Brusse, L., et al., 2003. Inflammatory bowel disease and the risk of fracture. Gastroenterology 125, 1591–1597.
Vasant, D., Paine, P., Black, C., et al., 2021. British Society of Gastroenterology guidelines on the management of irritable bowel syndrome. Gut 70, 1214–1240.
Vavrika, S., Rogler, G., Gantenbein, C., et al., 2015. Chronological order of appearance of extra intestinal manifestations relative to the time of IBD diagnosis in the swiss inflammatory disease cohort. Inflamm. Bowel Dis. 21, 1794–1800.
Verma, S., Kirkwood, B., Brown, S., Giaffer, M.H., 2000. Oral nutritional supplementation is effective in the maintinance of remission in Crohn's disease. Dig. Liver Dis. 32, 769–774.
Vernia, P., Loizos, P., Di Giuseppantonio, I., et al., 2013. Dietary calcium intake in patients with inflammatory bowel disease. J. Crohns Colitis 8, 312–317.
Vernooij, R., Zeraatkar, D., Han, M., et al., 2019. Patterns of red and processed meat consumption and risk of cardiometabolic and cancer outcomes: a systematic review and meta analysis of cohort studies. Ann. Int. Med. 171, 732–741.
Wagner, S., Schmidt, A., Effenberger, M., et al., 2013. Semisynthetic diet ameliorates crohn's disease like ileitis in TNF mice through antigen-independent mechanisms of gluten. Inflamm. Bowel Dis. 19, 1285–1294.
Wall, C.L., Day, A.S., Gearry, R.B., 2013. Use of exclusive enteral nutrition in adults with crohn's disease: a review. World J. Gastroenterol. 19, 7652–7660.
Wallaert, J., De Martino, R., Marsicovetere, P., et al., 2012. Venous thromboembolism after surgery for inflammatory bowel disease: are there modifiable risk factors? Data from ACS NSQIP. Dis. Colon Rectum 55, 1138–1144.
Wang, F., Lin, X., Zhao, Q., 2016. Fat intake and risk of ulcerative colitis: systematic review and dose-response meta-analysis of epidemiological studies. J. Gastroenterol. Hepatol. 32, 19–27.
Wang, F., Feng, J., Gao, Q., et al., 2017. Carbohydrate and protein intake and the risk of ulcerative colitis: systematic review and dose-response meta-analysis of epidemiological studies. Clin. Nutr. 36, 1259–1265.
Wark, G., Samocha-Bonet, D., Ghaly, S., Danta, M., 2021. The role of diet in the pathogenesis and management of inflammatory bowel disease: a review. Nutrients 13.
Wedlake, L., Slack, N., Andreyev, H.J.N., Whelan, K., 2014. Fiber in the treatment and maintenance of inflammatory bowel disease: a systematic review of randomised controlled trials. Inflamm. Bowel Dis. 20, 576–586.
Welters, C., Heineman, E., Thunnissen, F., et al., 2002. Effect of dietary inulin supplementation on inflammation of pouch mucosa in pateints with ileal pouch-anal anastamosis. Dis. Colon Rectum 45, 621–627.
Wilschanski, M., Sherman, P., Percharz, P., et al., 1996. Supplementary enteral nutrition maintains remission in paediatric Crohn's disease. Gut 38, 543–548.
Wong, S., Walker, J., Carr, R., et al., 2012. The information needs and preferences of persons with longstanding inflamamtory bowel disease. Can. J. Gastroenterol. Hepatol. 26, 525–531.
Xu, L., Lochhead, P., Ko, Y., et al., 2017. Systematic review with meta-analysis: breastfeeding and the risk of Crohn's disease and ulcerative colitis. Aliment. Pharmacol. Therapeut. 46, 780–789.
Xu, Z., McClure, S., Appel, L., 2018. Dietary cholesterol intake and sources among U.S adults: results from national health and nutrition examination surveys (NHANES), 2001–2014. Nutrients 10.
Yan, D., Ren, J., Wang, S., et al., 2014. Predictors of response to enteral nutrition in abdominal enterocutaneous fistula patients with Crohn's disease. Eur. J. Clin. Nutr. 68, 959–963.
Yang, Y., Gao, X., Chen, H., et al., 2017. Efficacy of exclusive enteral nutrition in complicated Crohn's disease. Scand. J. Gastroenterol. 52, 995–1001.
Yu, Y., Kang-Chen, C., Chen, J., 2019. Exclusive enteral nutrition versus corticosteroids for treatment of paediatric Crohn's disease: a meta analysis. World J. Pediatr. 15, 26–36.
Zeevi, D., Korem, T., Zmora, N., et al., 2015. Personalized nutrition by prediction of glycaemic responses. Cell 163, 1079–1094.
Zhang, H., Greenwood, D., Risch, H., et al., 2021. Meat consumption and risk of incident dementia: cohort study of 493, 888 UK Biobank participants. Am. J. Clin. Nutr. 114, 175–184.
Zuo, T., Ng, S., 2018. The gut microbiota in the pathogenesis and therapeutics of inflammatory bowel disease. Front. Microbiol. 9.

Tuberculosis: Nutritional management

Pranay Sinha[a] **and J. Peter Cegielski**[b], [a] Boston Medical Center, Boston, MA, United States; and [b] Emory University Rollins School of Public Health, Atlanta, GA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of J.P. Cegielski, D.N. McMurray, Tuberculosis: Nutritional Management, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 293–298, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00269-5.

Introduction 698
- Nutritional status of tuberculosis patients 699
- Effect of nutritional factors on the course of TB 700
- Controlled intervention studies of nutritional supplements in the management of TB 700
- Nutritional management of TB 701
- Nutrient-drug interactions 703

Conclusions 703
Disclaimer 703
References 704
Further reading 705

Key points

- Undernutrition is associated with increased severity of TB disease and poorer treatment outcomes
- Nutritional support, once the mainstay of TB treatment, is being increasingly recognized as an important adjunct to pharmacological therapy
- Nutritional support may improve adherence to TB therapy and promote treatment completion
- Micronutrient supplementation may accelerate clinical improvement and, in HIV-coinfected individuals, has improved treatment outcomes

Glossary

Tuberculosis (TB) An infectious disease caused by *Mycobacterium tuberculosis* which usually affects the lungs, but can be found elsewhere (extrapulmonary disease)

Antimicrobial drug resistance The inability of a drug to kill or slow the growth of a microbe due to genetic mutations in the microbe

Protein-energy malnutrition (PEM) A potentially fatal body-depletion disorder of which there is inadequate protein intake

Multi-drug resistant (MDR) TB Drug resistant TB that is resistant to at least rifampin and isoniazid

Bacillary load The amount of disease-causing bacteria found within a medium (i.e., sputum) usually associated with severity of disease

Introduction

Globally, an estimated 1.5 million tuberculosis (TB) deaths and 10 million new TB cases occur each year (WHO, 2021). One-quarter of the world's population may be latently infected with *Mycobacterium tuberculosis* and at risk for reactivation TB (Cohen et al., 2019). Before the advent of specific *anti*-TB drugs, nutritional support was a mainstay of the treatment of TB. Highly effective *anti*-TB drugs were first developed in the 1940s and 1950s (McMillen, 2015). In the 1970s, the combination of isoniazid, rifampin, and pyrazinamide enabled the duration of treatment to be shortened to 6–9 months. However, TB staged a dramatic comeback in the late 20th century in both affluent and developing countries, especially in countries of the former Soviet Bloc and in countries with a high prevalence of HIV infection in sub-Saharan Africa and parts of Southeast Asia. Today, TB is the leading cause of death among persons with HIV infection, one of the top five leading infectious cause of death, and one of the leading causes of maternal mortality worldwide.

TB patients may be difficult to cure for many reasons, including antimicrobial drug resistance, drug toxicity and intolerance, advanced TB disease, and comorbidities such as HIV infection. Therefore, there is renewed interest in nutritional support in the management of TB.

Nutritional status of tuberculosis patients

Undernutrition is an important risk factor for developing TB, and TB causes anorexia, weight loss and cachexia. Weight loss and wasting in TB may result from the production of tumor necrosis factor-alpha (TNF-α) and other pro-inflammatory cytokines that play a critical in protection against TB but also precipitate anorexia. Using the lactose-mannitol excretion test, a study in Brazil found evidence of altered intestinal mucosal integrity in persons with TB as compared to healthy controls which suggests that malabsorption may also drive weight loss among persons with TB (Pinheiro et al., 2006). Protein utilization may be altered by the cytokine milieu as evidenced by negative nitrogen balance. Studies in the United Kingdom (UK) documented that TB patients had significantly lower serum albumin levels than controls (mean, 37 g/L vs. 46 g/L), suggesting protein undernutrition and/or a systemic inflammatory response. Anabolic pathways may be functionally blocked due to preferential oxidation of ingested amino acids for energy rather than for protein synthesis, contributing to wasting despite nutritional support (MacAllan et al., 1998).

Weight loss among persons with TB can be severe. Compared to healthy individuals, TB patients have significantly lower body mass index (BMI), skin fold thickness, limb circumference, and overall proportion of body fat. In the United States, weight loss was present at diagnosis in 45% of patients. In Tanzania, among 200 consecutive adults with sputum smear-positive pulmonary TB, 77% of males and 58% of females had a BMI < 18.5 kg/m^2, while 20% had a BMI < 16 kg/m^2 (Kennedy et al., 1996). In India, the median BMI among men with TB was 15.6 kg/m^2; the median BMI among women with TB was 16.4 kg/m^2 (Padmapriyadarsini et al., 2016). In Malawi, TB patients were substantially weaker than controls as measured by hand-grip dynamometry, suggesting loss of skeletal muscle protein (Harries et al., 1988). In addition, these patients had 35% lower fat mass and 19% lower lean body mass than the control group. More extensive TB disease and longer duration of symptoms were associated with lower BMI. BMI among Asian TB patients living in the UK (19.3 kg/m^2) was lower than controls (22.2 kg/m^2), skin fold thickness was 13% lower, and arm muscle circumference was 20% lower. Patients with concurrent HIV infection tend to be even more undernourished.

Protein-energy malnutrition (PEM) rarely occurs without micronutrient deficiencies as well. TB patients have been found to be deficient in vitamins A, B6, and D as well as zinc, copper, iron, and selenium, although serum levels of fat-soluble vitamins A, carotene, and D, and certain micronutrients like zinc, iron, and selenium also fall while serum copper levels rise with systemic inflammation. In Ecuador, Koyanagi et al. observed that TB patients had significantly lower serum concentrations of zinc, retinol and selenium (Koyanagi et al., 2004). More than 800 TB patients in Malawi demonstrated deficiencies in circulating selenium, carotenoids and vitamin A (van Lettow et al., 2004). These deficiencies were exacerbated in the most severely wasted group (BMI < 16). There were no significant differences between the HIV-infected and HIV-uninfected TB patients. Low plasma selenium levels were associated with anemia.

In a study of 155 Ethiopian TB patients, HIV co-infection was associated with lower serum zinc and selenium concentrations and an elevated copper/zinc ratio compared with HIV-negative TB patients (Kassu et al., 2006). After the intensive phase of antibiotic therapy, serum levels of both selenium and zinc improved in both patient groups. Another study in Ethiopia reported serum concentrations of vitamins C, E and A were significantly lower in TB patients than in healthy controls (Madebo et al., 2003). High malondialdehyde concentrations, an indicator of overall oxidant stress, were associated with increased clinical severity of TB, and these parameters were exacerbated in HIV co-infected individuals (Madebo et al., 2003).

Wiid and coworkers observed significantly lower total antioxidant status (TAS) in TB patients compared with community controls, and TAS values increased during anti-mycobacterial chemotherapy (Wiid et al., 2004). Similar results were seen with vitamin A and zinc levels, but not with vitamin E. The vitamin A status of 100 TB patients was studied in Tanzania before and after the intensive phase of *anti*-TB therapy (Jeremiah et al., 2014). Vitamin A levels were low in TB patients and improved with therapy in HIV-negative, but not in HIV infected patients. HIV infection was also associated with low vitamin A status in otherwise asymptomatic controls.

Ramachandran et al. observed low serum vitamin A levels in 47 newly-diagnosed TB patients compared with household contacts and healthy controls (Ramachandran et al., 2004). Their vitamin A status improved significantly following *anti*-TB therapy without the need for vitamin A supplementation. Pediatric TB patients in India had markedly reduced levels of plasma zinc, irrespective of their general nutritional status, and there was significant improvement after 6 months of *anti*-TB therapy (Ray et al., 1998). Turkish investigators also observed a significant improvement in serum zinc (which increased) and copper/zinc ratios (which decreased) after 2 months of *anti*-TB therapy in adult patients (Ciftci et al., 2003).

Vitamin D is critical for induction of innate macrophage functions via Toll-like receptor ligation of mycobacterial cell surface molecules, mycobacteria-specific activation of T lymphocytes by infected macrophages, and fusion of phagosomes containing mycobacteria with lysosomes within infected macrophages. Thus, having adequate vitamin D at the time of initial infection may be important in developing the immune response that prevents the progression to active TB disease. Vitamin D receptor genetic polymorphisms are associated with vitamin D deficiency and increased incidence of TB. TB patients of Asian and African origin in the UK were significantly vitamin D deficient. Studies in India and Africa also found vitamin D deficiency associated with receptor polymorphisms in TB patients (Wilkinson et al., 2000).

Effect of nutritional factors on the course of TB

Several observational studies have found an association between undernutrition and increased disease severity. A study of 173 persons with drug-susceptible TB in India found that after controlling for diabetes and cavitary disease, patients who were severely undernourished had 11% more of their lungs affected on chest x-ray (95% confidence interval [CI]: 4.0–13.3) as compared to those with normal BMI (Hoyt et al., 2019). Severely undernourished individuals also had a four-fold higher risk of cavitary lung disease (odds ratio [OR]: 4.6; 95% CI: 1.5–14.1). This association has also been noted in persons with MDR-TB. A Latvian study comprising 995 persons with TB found that individuals with BMI < 18.5 kg/m^2 were more likely to have smear positive TB (adjusted odds ratio [aOR]: 1.7; 95% 1–2.9) and more likely to have bilateral cavitation (aOR: 2.1; 95% CI: 1.3–3.5) (Podewils et al., 2011).

Studies have also found an association between undernutrition at the time of diagnosis and poor treatment outcomes. A large study of 1695 persons with drug-susceptible TB from central India found a dose-dependent decrease in mortality as the BMI increased (aOR: 0.78; 95% CI: 0.68–0.90) (Bhargava et al., 2013). Chemotherapy is less effective in MDR-TB, and the impact of nutritional deficits may be more pronounced. In the aforementioned MDR-TB study from Latvia, patients with BMI < 18.5 kg/m^2 had a higher risk of mortality (adjusted hazard ratio (aHR): 1.9; 95% CI: 1.1–3.5) (Podewils et al., 2011). Although low BMI may be a sign of worse disease, these studies controlled for disease severity by multivariable regression. Low serum albumin, anemia, weight loss, and lack of weight gain were associated with the severity and clinical course of TB. Among 373 patients hospitalized for TB in Ecuador, hypoalbuminemia increased the odds of in-hospital death >3-fold after controlling for HIV infection and other co-morbidities (Matos and Moreira Lemos, 2006). A prospective cohort study in the Philippines of 439 adults with pulmonary MDR TB found that failure to gain weight during the first 6 months of treatment strongly predicted poor treatment outcomes (Gler et al., 2013).

The timing of mortality may also be predicted based on baseline BMI. In one study of nearly 1200 TB patients followed prospectively, 10.9% of patients with moderate to severe undernutrition died in the first 4 weeks compared with 6.5% of the patients with normal nutritional status or mild undernutrition. Another study found that TB patients with a body mass index <17.0 kg/m^2 were at increased risk of early death. In children, weight for age is an important indicator of prognosis (Pranay et al., 2021).

The severity of undernutrition is an important indicator of the progress of the disease, and normalization of body weight in response to treatment is a positive sign. In Brazil, a study of 547 persons with TB found that the risk of unsuccessful outcomes decreased by 12% (95% CI: 5%–19%) per kilogram increase in weight at the end of the first two months (Peetluk et al., 2020). Conversely, a study of 650 persons with TB in Peru found that those who did not gain 5% or more of baseline weight at the end of treatment had a higher likelihood of treatment failure (risk ratio [RR] 2.05; 95% CI 1.10–3.80) (Krapp et al., 2008). In the US, Khan and coworkers demonstrated that TB patients who were underweight at diagnosis had a 4-fold increased risk of relapse within 2 years after completing treatment, and patients who did not gain weight had a 2-fold increase in risk of relapse (Khan et al., 2006).

Controlled intervention studies of nutritional supplements in the management of TB

Historically, the use of cod-liver oil for the treatment of TB was the taproot from which grew the broader field of nutritional management of TB. In 18th and 19th century Europe, TB was responsible for 25% of adult deaths. Survival after diagnosis was approximately 2 years (Tiemersma et al., 2011). Treatment for TB was revolutionized by the use of cod-liver oil which had been used for its medicinal properties (Grad, 2004). Treating TB patients with cod-liver oil in the 18th century resulted in weight gain and increased survival rates from 2 to 8 years. In the early 20th Century, one US study reported that TB patients treated with cod-liver oil gained weight and only 10% died in contrast with weight loss and 70% mortality in the comparison group during approximately 1 year of observation. Cod-liver oil contains vitamins A and D which are important in host defense against TB.

With the advent of highly effective *anti*-TB drugs in the 1940s and 1950s, interest in cod-liver oil and other nutritional interventions waned. An influential clinical trial was carried out in Madras, India, in the 1950s to compare sanatorium treatment with outpatient treatment with regard to nutritional influences on treatment outcome (Tuberculosis Chemotherapy, 1959). Patients treated in the sanatorium had substantially better diets and gained more weight than home-treated patients. Improvement was slightly faster in the sanatorium treated group, but the outcomes were nearly the same in the two groups after 12 months after controlling for baseline differences in disease severity. Chemotherapy was so effective that the effects of nutritional support apparently were overshadowed.

Chemotherapy is less effective in MDR TB and HIV-associated TB, renewing interest in nutritional interventions. In the past decade, trials of nutritional intervention during chemotherapy of TB patients have shown modest to no benefits. Eighty Indonesian TB patients with low BMI, low plasma retinol, and low plasma zinc were treated with a retinol and zinc supplement versus placebo in addition to standard *anti*-TB drugs (Karyadi et al., 2002). Sputum conversion, radiographic improvement, and increased plasma retinol levels were observed in the treated group after 6 months of therapy. Two weeks post-therapy, the percentage of patients with negative sputum smears was significantly higher ($p < 0.01$) in the micronutrient-treated group (23%) compared with the placebo group (13%). Lesion area was significantly reduced in the treated group after two months of therapy ($p < 0.01$). Plasma retinol concentrations were correlated inversely with reduction in mean lesion size at 6 months ($r = -0.367$; $p = 0.02$).

A much larger randomized control trial (RCT) in Tanzania examined the effect on treatment of supplemental zinc alone, multiple micronutrients (MMN: vitamins A, B, C, D, E and minerals Se, Cu), MMN + zinc, or a placebo (PrayGod et al., 2011). Approximately 43% of each group was HIV-infected, and all received standard *anti*-TB chemotherapy. After 8 weeks of therapy,

neither supplement had a significant effect on sputum culture positivity; however, patients receiving the MMN experienced a significant improvement in body weight. HIV status had no influence on the outcome. On the other hand, after 8 months of therapy, the MMN + zinc group had significantly reduced mortality (RR = 0.29; 95% confidence interval 0.10–0.80), but only in the TB-HIV co-infected patients.

Several RCTs have examined the impact of supplemental zinc, iron or vitamin D on the outcome of chemotherapy in TB patients. Sixty-six HIV-infected TB patients in Singapore who were receiving antiretroviral and *anti*-TB therapies were assigned to 28 days of oral zinc sulfate supplements or placebo. Zinc supplements had no effects on mycobacterial antigen-stimulated IFNγ production; however, nearly all (94%) of the subjects had normal plasma zinc levels at baseline (Green et al., 2005). In the second study, 117 anemic, adult male TB patients in India were enrolled in an RCT of iron supplementation during the first 2 months of conventional *anti*-TB therapy. During 6 months of follow up, hematological status improved as the TB disease improved, but supplemental iron had no additional benefit as determined by the extent of radiographic abnormalities (Devi et al., 2003). In Egypt, 24 newly diagnosed pediatric TB patients were enrolled in an RCT to examine the effect of vitamin D supplementation (1000 IU/day for 8 weeks) on the outcome of *anti*-TB therapy (Wejse et al., 2009). Most of the children were vitamin D-deficient at baseline and serum levels of 1,25 (OH)2D3 improved in both groups during chemotherapy, but, supplementation did not affect this parameter. The supplemented group showed significant radiological and clinical improvement at follow-up with significantly increased body weights (3.3 kg) compared with the placebo group (2.2 kg) ($p < 0.05$). Two placebo-controlled trails of vitamin D supplementation in India and the Republic of Georgia found that although serum 25(OH)D levels increased to normal among those with baseline vitamin D deficiency, there were no significant difference in the mean to sputum culture conversion between the placebo and vitamin D groups in both studies (Tukvadze et al., 2015).

Hanekom et al. conducted a RCT of vitamin A supplementation in 85 South African children with TB who were not co-infected with HIV (Hanekom et al., 1997). Children were given either 200,000 IU of retinyl palmitate or placebo on day 0 and day 1 and then followed up during 3 months of conventional *anti*-TB therapy. Nearly two-thirds of the patients were vitamin A-deficient at the beginning of the study, and the deficiency was more pronounced in children with extra-pulmonary disease. Vitamin A status improved in both groups, but supplementation had no significant effect on treatment outcome. Vitamin A supplementation was associated with significant decreases in a plasma protein biomarker of a Th2-type cytokine response which may indicate that vitamin A supplementation promoted a protective Th1-type 1 cytokine profile.

Inducible nitric oxide (NO) is a critical proximal mediator of anti-mycobacterial resistance in rodent models of TB, while its role in human TB remains controversial. Arginine is the substrate for inducible nitric oxide synthase, the enzyme that produces NO in macrophages. An RCT of oral arginine supplementation (1 g/day) was conducted in 120 HIV-infected and HIV–uninfected Ethiopian TB patients for 4 weeks along with standard *anti*-TB therapy (Schön et al., 2003). At 8 weeks, arginine supplementation resulted in significant improvement in serum arginine levels, weight gain, sputum conversion, and reduction of symptoms compared with the placebo group, but these benefits were observed only in HIV-negative patients. No treatment effect was observed in HIV co-infected patients.

The cholesterol content of macrophage vesicle membranes (i.e., phagosomes, lysosomes) has been shown to affect the ability of the phagocytes to suppress intracellular growth of mycobacteria. In a small RCT in Mexico City, investigators compared the clinical responses of 10 TB patients who received a cholesterol-rich diet (800 mg/day) with 11 TB patients who consumed a control diet (250 mg/day) during the first 8 weeks of standard *anti*-TB chemotherapy (Pérez-Guzmán et al., 2005). Respiratory symptoms improved in both groups: however, culture-negative sputum at 2 weeks was more frequent in patients consuming a high-cholesterol diet (91%) compared to the placebo group (20%) ($p < 0.002$). The bacillary load in sputum was much lower in the cholesterol-supplemented patients (0.05 colony forming units [cfu]) than in the placebo patients (3.4 cfu) ($p < 0.0002$).

One mechanism by which nutritional supplementation may improve treatment outcomes is by improving adherence to TB therapy. For instance, a retrospective cohort study of 573 persons with smear-positive TB who were living below the poverty line in India found that those receiving nutritional support had a 50% reduced risk of unsuccessful treatment outcome than those who did not receive nutritional support (95% CI: 30–86%) (Samuel et al., 2016). In the nutritional support group, only 1 individual (1%) was lost to follow up whereas 41 individuals (10%) were lost to follow up. Similarly, in Senegal, a small study of 26 HIV-TB individuals found that nutritional supplementation was associated with 95% adherence to both TB and HIV treatments. All subjects in this study completed therapy successfully.

Nutritional management of TB

The goal of nutritional interventions is to (1) compensate for the elevated resting energy expenditure and catabolic state associated with TB, (2) support the extensive cellular proliferation and protein production associated with anti-mycobacterial immune responses, (3) allow repair of diseased tissues, and (4) replenish somatic reserves. Supplementation with specific nutrients (e.g., vitamin A, vitamin D, zinc) may be required to correct specific deficiencies. The research reviewed above provides limited support for the use of dietary supplementation with specific macro- and micronutrients for their beneficial impact on both nutritional status and treatment outcomes in TB.

While this article is not intended to offer recommendations for medical practice, some general guidelines are discussed. The expertise of clinical nutritionists and dieticians should be sought in managing persons with TB with complex nutritional

requirements. In the absence of such expertise, reference works focusing on the nutritional management of patients with infectious diseases should be consulted. In 2013, the World Health Organization published a guidance document to guide nutritional assessment and care for persons with TB which can also serve as a roadmap.

Careful assessment of nutritional status is the starting point, including the medical history and physical examination, anthropometric data, dietary information, as well as laboratory tests. Nutritional interventions to correct specific nutrient deficiencies can be based on this assessment. In endemic regions, it may be reasonable to ask about a history of intestinal parasitic infections and diarrheal diseases or severe anemia as important contributing factors to nutritional status.

The medical history and physical examination should include questions to identify unintentional weight loss which is associated with poorer treatment outcomes, higher risk of relapse, and increased mortality. Anorexia, abdominal discomfort, nausea, vomiting, and diarrhea will affect nutritional status and nutritional support. Co-morbidities that have nutritional implications such as diabetes mellitus, hepatitis, alcohol abuse, and HIV infection should be identified. In endemic regions, it may be reasonable to screen for intestinal parasitic infections among persons with TB who are manifesting symptoms such as diarrhea or anemia.

Fever affects resting energy expenditure and caloric requirements. The expert clinician's subjective global assessment is one of the critical aspects of evaluating nutritional status. The clinical manifestations of specific nutrient deficiencies have been described in many other reference works to which the interested reader is referred. Peripheral neuropathy deserves special mention because it is a common side effect of isoniazid, one of the primary *anti*-TB drugs (see below), and vitamin B6 supplements are routinely prescribed with *anti*-TB treatment to prevent peripheral neuropathy.

Anthropometric data should include weight and height in relation to age, sex, and reference standards. BMI can be calculated to determine the overall macronutrient deficit. Skin fold thicknesses, linear and circumferential measurements of specific body parts, and bioelectric impedance may help categorize the nutritional deficit as involving energy, protein, or essential fats.

A dietary survey should include questions regarding recent patterns and quantities of food consumption pre- and post-illness (e.g., food availability and intake, dietary restrictions, preferences). Standardized tools include 24 h diet recall, 72 h food diaries, and food frequency questionnaires used in clinical and epidemiological research, but they also may be useful in patient care. The information is translated into nutrient intakes based on the known composition of foods and estimates of the quantities consumed. Dietary information should include specific requirements and restrictions based on age, co-existing medical conditions, cultural and religious practices, personal preferences, food allergies, and intolerance of certain foods. Dietary assessment will set specific boundaries around possible dietary recommendations and, the availability and cost of foods and nutritional supplements will affect these boundaries.

Basic hematology and biochemistry laboratory tests can be supplemented as indicated by the history, clinical examination, and abnormalities identified in these basic tests. Anemia is common in inflammatory conditions such as TB. However, anemia may also result from deficiencies of iron, folate, or vitamin B12. Apart from anemia, serum albumin level may be the most important predictor of nutritional risk for poor outcomes associated with TB. Since albumin is synthesized in the liver with a half-life of 21 days, hypoalbuminemia may reflect inadequate protein intake over a period of weeks or longer, although hypoalbuminemia is also a marker of systemic inflammation which is minimally affected by nutritional support until the inflammatory response remits.

Nutritional support of the TB patients should include a varied diet containing ~50% of calories from carbohydrates, 20%–30% from proteins (with an emphasis on high quality proteins), and 20%–30% from fats. Energy requirements start at a basal level of approximately 30 kcal/kg body mass, increasing to 40 kcal/kg or more for persons with significant deficits or with increased resting energy expenditure (e.g., fever). Respiratory function may be modestly or severely compromised, including both oxygenation and ventilation, in TB. As an energy source, dietary carbohydrates generate 20% more carbon dioxide than proteins and 30% more than fats. Thus, for patients who are short of breath, hypercapnic, or lack adequate ventilation, fats and proteins may be preferred dietary sources of energy. This is generally only a problem when the carbohydrate content of the diet exceeds the energy expenditure where the respiratory quotient can exceed 1.0 and substantially increase CO2 production.

Inflammation drives the utilization of both endogenous and exogenous proteins for energy due to hepatic gluconeogenesis and an anabolic block in protein synthesis created by the immune cell cytokines. Protein requirements may be as high as 1.5 gm/kg body mass/day. TB patients who are acutely ill, have advanced disease, or substantially compromised nutritional status, and whose dietary intake has been inadequate for an extended period of time may benefit from a gradual increase in dietary macronutrients up to the recommended amounts. Overfeeding is not recommended.

Omega-3 polyunsaturated fatty acids in the diet have anti-inflammatory and immunomodulatory effects, including increased susceptibility to TB in guinea pigs and mice, while dietary omega-6 polyunsaturated fatty acids have pro-inflammatory qualities, either of which may prove to be undesirable in the management of TB. The effects of different fatty acids on the nutritional management of TB remain to be established.

A multivitamin and mineral supplement is recommended to ensure the availability of micronutrients. Higher doses of individual components to treat specific micronutrient deficiencies should follow guidelines established for the treatment of those conditions. Otherwise, doses of vitamins and minerals above those demonstrated to be safe and beneficial are not recommended.

Iron deficiency and iron replacement also merits special mention. Increased severity of TB has been observed in individuals with hemochromatosis, an iron overload syndrome, and in indigenous societies with high levels of iron intake from drinking a type of traditional beer fermented in iron vessels. Moreover, epidemics of malaria have been reported during refeeding programs in refugee and famine situations, related to the increased availability of iron for the parasite from supplements provided for the treatment of

iron deficiency. *M. tuberculosis* acquires iron by means of highly developed scavenging mechanisms that must out-compete the host's iron-binding proteins (e.g., lactoferrin, transferrin). Iron deficiency is not necessarily more deleterious to the host than to the pathogen, and iron replacement therapy may benefit the microbe as well as the patient.

Nutrient-drug interactions

The standard treatment for all newly diagnosed patients with active, drug-susceptible TB includes four drugs, isoniazid, rifampin, pyrazinamide, and ethambutol, taken for two to three months followed by two of these drugs, isoniazid and rifampicin, taken for an additional four to 6 months. Isoniazid (INH) interferes with the metabolism of vitamin B6 which includes pyridoxine, pyridoxal, and pyridoxamine. Isoniazid combines with pyridoxal or pyridoxal phosphate to form potent inhibitors of pyridoxal kinase, thus, blocking formation of the coenzyme form of the vitamin. In the absence of vitamin B6 supplements, TB patients treated with INH may experience peripheral neuritis, manifested as tingling, numbness, or a painful "prickly" sensation in a stocking and glove distribution. Approximately 20% of patients treated with high doses of INH or patients who are otherwise predisposed to peripheral neuropathy (e.g., alcoholics, diabetes mellitus) will develop peripheral neuritis. Administration of 25–50 mg daily of vitamin B6 prevents peripheral neuritis in nearly all TB patients treated with INH. Importantly, excess vitamin B6 may also precipitate symptoms of peripheral neuropathy.

Patients treated with cycloserine, an important second-line drug used in the treatment of MDR TB, should also receive vitamin B6 supplements in doses of 200 mg–300 mg/day because of central nervous system side-effects (psychosis, depression) also related to pyridoxine metabolism.

Other *anti*-TB drugs that have adverse consequences on nutrition include para-aminosalicylic acid (PAS) and ethionamide, because these drugs commonly cause nausea, abdominal pain, anorexia, or vomiting. Such side effects will have a significant negative impact on the patient's nutritional status and well-being.

Treatment of TB may induce other problems that affect nutritional status and nutrient intake. Three of the first line drugs, isoniazid, rifampicin, and pyrazinamide all carry a small risk of chemical hepatitis characterized by anorexia, nausea, vomiting and decreased nutrient intake. In its more severe forms, drug-induced hepatitis results in disturbances in carbohydrate, protein and lipid homeostasis that clearly impact the patient's metabolic and nutritional status. Although nutritional factors do not contribute to the cause, TB drug-induced hepatitis has many important consequences affecting nutritional status and nutrient intake.

Conversely, poor nutritional status has important implications for the pharmacokinetics of antimycobacterial drugs. After conducting the lactulose-mannitol excretion assay among 41 persons with TB and 28 healthy controls, a study in Brazil found that, compared to healthy controls, persons with TB had evidence of altered intestinal mucosal integrity (Pinheiro et al., 2006). In a study of 84 Indian children with HIV/TB, stunted and underweight children had significantly lower serum peak concentrations of rifampin and isoniazid (Ramachandran et al., 2013). Similarly, in a Tanzanian study of 51 HIV-negative children, undernutrition was significantly associated with lower estimated peak concentrations of rifampin and isoniazid in multivariate analyses (Justine et al., 2020). However, as undernutrition decreases fat-free mass, it can also result in supratherapeutic doses of drugs, such as ethambutol and aminoglycosides, and increase the risk of drug toxicity (Ter Beek et al., 2019). There may be a role for ensuring safe and effective drug concentrations among undernourished persons through therapeutic drug monitoring.

Conclusions

Nutritional support is an important component of comprehensive care for persons with TB which must be adapted to each geographic region and socioeconomic context. Specific nutritional recommendations should be adapted to each patient depending on their clinical condition, nutritional status, and the practical possibilities of supplementation. Simplified, more generic approaches may be most suitable in some circumstances. These considerations notwithstanding, all TB patients should be offered nutritional/dietary advice based upon their nutritional status and accompanying diseases. While a specific, often costly diet that targets the individual nutritional needs of a TB patient may improve treatment outcome, it may be difficult to achieve this level of care in the absence of sufficient economic resources. This may be an entry point for policy interventions in countries with large burdens of both food insecurity and TB. Some national TB programs such as India's do provide monthly cash transfers to improve the nutritional status of persons with TB. More remains to be done. However, we believe a deeper understanding of the essential role of nutrition in TB pathogenesis and treatment will help improve current TB treatment practices and improve outcomes of TB patients.

Disclaimer

The views and opinions expressed in this article are those of the authors and do not necessarily represent an official position of the Centers for Disease Control and Prevention, the US government.

See Also: Nutrition and susceptibility to tuberculosis; Vitamin A: Deficiency and interventions; Vitamin A: Physiology, dietary sources and requirements; Vitamin B_6; Vitamin D: Role in chronic and acute diseases; Selenium

References

Bhargava, A., Chatterjee, M., Jain, Y., et al., 2013. Nutritional status of adult patients with pulmonary tuberculosis in rural central India and its association with mortality. PLoS One 8 (10), e77979.

Ciftci, T.U., Ciftci, B., Yis, Ö., Guney, Y., Bilgihan, A., Ogretensoy, M., 2003. Changes in serum selenium, copper, zinc levels and Cu/Zn ratio in patients with pulmonary tuberculosis during therapy. Biol. Trace Elem. Res. 95 (1), 65–71.

Cohen, A., Mathiasen, V.D., Schön, T., Wejse, C., 2019. The global prevalence of latent tuberculosis: a systematic review and meta-analysis. Eur. Respir. J. 54 (3).

Devi, U., Rao, C.M., Srivastava, V.K., Rath, P.K., Das, B.S., 2003. Effect of iron supplementation on mild to moderate anaemia in pulmonary tuberculosis. Br. J. Nutr. 90 (3), 541–550.

Gler, M.T., Guilatco, R., Caoili, J.C., Ershova, J., Cegielski, P., Johnson, J.L., 2013. Weight gain and response to treatment for multidrug-resistant tuberculosis. Am. J. Trop. Med. Hyg. 89 (5), 943–949.

Grad, R., 2004. Cod and the consumptive: a brief history of cod-liver oil in the treatment of pulmonary tuberculosis. Pharm. Hist. 46 (3), 106–120.

Green, J., Lewin, S., Wightman, F., Lee, M., Ravindran, S., Paton, N., 2005. A randomised controlled trial of oral zinc on the immune response to tuberculosis in HIV-infected patients. Int. J. Tubercul. Lung Dis. 9 (12), 1378–1384.

Hanekom, W.A., Potgieter, S., Hughes, E.J., Malan, H., Kessow, G., Hussey, G.D., 1997. Vitamin A status and therapy in childhood pulmonary tuberculosis. J. Pediatr. 131 (6), 925–927.

Harries, A., Nkhoma, W., Thompson, P., Nyangulu, D., Wirima, J., 1988. Nutritional status in Malawian patients with pulmonary tuberculosis and response to chemotherapy. Eur. J. Clin. Nutr. 42 (5), 445–450.

Hoyt, K.J., Sarkar, S., White, L., et al., 2019. Effect of malnutrition on radiographic findings and mycobacterial burden in pulmonary tuberculosis. PLoS One 14 (3), e0214011.

Jeremiah, K., Denti, P., Chigutsa, E., et al., 2014. Nutritional supplementation increases rifampin exposure among tuberculosis patients coinfected with HIV. Antimicrob. Agents Chemother. 58 (6), 3468–3474.

Justine, M., Yeconia, A., Nicodemu, I., et al., 2020. Pharmacokinetics of first-line drugs among children with tuberculosis in rural Tanzania. J. Pediatr. Infect. Dis. Soc. 9 (1), 14–20. https://doi.org/10.1093/jpids/piy106.

Karyadi, E., West, C.E., Schultink, W., et al., 2002. A double-blind, placebo-controlled study of vitamin A and zinc supplementation in persons with tuberculosis in Indonesia: effects on clinical response and nutritional status. Am. J. Clin. Nutr. 75 (4), 720–727.

Kassu, A., Yabutani, T., Mahmud, Z.H., et al., 2006. Alterations in serum levels of trace elements in tuberculosis and HIV infections. Eur. J. Clin. Nutr. 60 (5), 580–586.

Kennedy, N., Ramsay, A., Uiso, L., Gutmann, J., Ngowi, F., Gillespie, S., 1996. Nutritional status and weight gain in patients with pulmonary tuberculosis in Tanzania. Trans. R. Soc. Trop. Med. Hyg. 90 (2), 162–166.

Khan, A., Sterling, T.R., Reves, R., Vernon, A., Horsburgh, C.R., Consortium, T.T., 2006. Lack of weight gain and relapse risk in a large tuberculosis treatment trial. Am. J. Respir. Crit. Care Med. 174 (3), 344–348.

Koyanagi, A., Kuffó, D., Gresely, L., Shenkin, A., Cuevas, L.E., 2004. Relationships between serum concentrations of C-reactive protein and micronutrients, in patients with tuberculosis. Ann. Trop. Med. Parasitol. 98 (4), 391–399.

Krapp, F., Veliz, J., Cornejo, E., Gotuzzo, E., Seas, C., 2008. Bodyweight gain to predict treatment outcome in patients with pulmonary tuberculosis in Peru. Int. J. Tubercul. Lung Dis. 12 (10), 1153–1159.

MacAllan, D.C., McNurlan, M.A., Kurpad, A.V., et al., 1998. Whole body protein metabolism in human pulmonary tuberculosis and undernutrition: evidence for anabolic block in tuberculosis. Clin. Sci. 94 (3), 321–331.

Madebo, T., Lindtjørn, B., Aukrust, P., Berge, R.K., 2003. Circulating antioxidants and lipid peroxidation products in untreated tuberculosis patients in Ethiopia. Am. J. Clin. Nutr. 78 (1), 117–122.

Matos, E.D., Moreira Lemos, A.C., 2006. Association between serum albumin levels and in-hospital deaths due to tuberculosis. Int. J. Tubercul. Lung Dis. 10 (12), 1360–1366.

McMillen, C.W., 2015. Discovering Tuberculosis: A Global History, 1900 to the Present. Yale University Press.

Padmapriyadarsini, C., Shobana, M., Lakshmi, M., Beena, T., Swaminathan, S., 2016. Undernutrition & tuberculosis in India: situation analysis & the way forward. Indian J. Med. Res. 144 (1), 11.

Peetluk, L.S., Rebeiro, P.F., Cordeiro-Santos, M., et al., 2020. Lack of weight gain during the first 2 months of treatment and human immunodeficiency virus independently predict unsuccessful treatment outcomes in tuberculosis. J. Infect. Dis. 221 (9), 1416–1424.

Pérez-Guzmán, C., Vargas, M.H., Quiñonez, F., Bazavilvazo, N., Aguilar, A., 2005. A cholesterol-rich diet accelerates bacteriologic sterilization in pulmonary tuberculosis. Chest 127 (2), 643–651.

Pinheiro, V.G., Ramos, L., Monteiro, H.S., et al., 2006. Intestinal permeability and malabsorption of rifampin and isoniazid in active pulmonary tuberculosis. Braz. J. Infect. Dis. 10 (6), 374–379.

Podewils, L., Holtz, T., Riekstina, V., et al., 2011. Impact of malnutrition on clinical presentation, clinical course, and mortality in MDR-TB patients. Epidemiol. Infect. 139 (1), 113–120.

PrayGod, G., Range, N., Faurholt-Jepsen, D., et al., 2011. Daily multi-micronutrient supplementation during tuberculosis treatment increases weight and grip strength among HIV-uninfected but not HIV-infected patients in Mwanza, Tanzania. J. Nutr. 141 (4), 685–691.

Ramachandran, G., Santha, T., Garg, R., et al., 2004. Vitamin A levels in sputum-positive pulmonary tuberculosis patients in comparison with household contacts and healthy "normals". Int. J. Tubercul. Lung Dis. 8 (9), 1130–1133.

Ramachandran, G., Hemanth Kumar, A., Bhavani, P., et al., 2013. Age, nutritional status and INH acetylator status affect pharmacokinetics of anti-tuberculosis drugs in children. Int. J. Tubercul. Lung Dis. 17 (6), 800–806.

Ray, M., Kumar, L., Prasad, R., 1998. Plasma zinc status in Indian childhood tuberculosis: impact of antituberculosis therapy. Int. J. Tubercul. Lung Dis. 2 (9), 719–725.

Samuel, B., Volkmann, T., Cornelius, S., et al., 2016. Relationship between nutritional support and tuberculosis treatment outcomes in West Bengal, India. J. Tubercul. Res. 4 (4), 213–219.

Schön, T., Elias, D., Moges, F., et al., 2003. Arginine as an adjuvant to chemotherapy improves clinical outcome in active tuberculosis. Eur. Respir. J. 21 (3), 483–488.

Sinha, P., Lönnroth, K., Bhargava, A., Heysell, S.K., Sarkar, S., Salgame, P., Rudgard, W., Boccia, D., Van Aartsen, D., Hochberg, N.S., 2021. Food for thought: addressing undernutrition to end TB. Lancet Infect. Dis. 21 (10), e318–e325. https://doi.org/10.1016/S1473-3099(20)30792-1.

Ter Beek, L., Alffenaar, J.-W.C., Bolhuis, M.S., van der Werf, T.S., Akkerman, O.W., 2019. Tuberculosis-related malnutrition: public Health implications. J. Infect. Dis. 220 (2), 340–341.

Tiemersma, E.W., van der Werf, M.J., Borgdorff, M.W., Williams, B.G., Nagelkerke, N.J., 2011. Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. PLoS One 6 (4), e17601.

Tuberculosis Chemotherapy Centre, 1959. A concurrent comparison of home and sanatorium treatment of pulmonary tuberculosis in South India. Bull. World Health Organ. 21 (1), 51–144.
Tukvadze, N., Sanikidze, E., Kipiani, M., et al., 2015. High-dose vitamin D3 in adults with pulmonary tuberculosis: a double-blind randomized controlled trial. Am. J. Clin. Nutr. 102 (5), 1059–1069.
van Lettow, M., Harries, A.D., Kumwenda, J.J., et al., 2004. Micronutrient malnutrition and wasting in adults with pulmonary tuberculosis with and without HIV co-infection in Malawi. BMC Infect. Dis. 4 (1), 61.
Wejse, C., Gomes, V.F., Rabna, P., et al., 2009. Vitamin D as supplementary treatment for tuberculosis: a double-blind, randomized, placebo-controlled trial. Am. J. Respir. Crit. Care Med. 179 (9), 843–850.
WHO, 2021. Global Tuberculosis Report 2021. WHO, Geneva.
Wiid, I., Seaman, T., Hoal, E., Benade, A., Van Helden, P.D., 2004. Total antioxidant levels are low during active TB and rise with anti-tuberculosis therapy. IUBMB Life 56 (2), 101–106.
Wilkinson, R.J., Llewelyn, M., Toossi, Z., et al., 2000. Influence of vitamin D deficiency and vitamin D receptor polymorphisms on tuberculosis among Gujarati Asians in west London: a case-control study. Lancet 355 (9204), 618–621.

Further reading

Bruns, H., Stenger, S., 2014. New insights into the interaction of *Mycobacterium tuberculosis* and human macrophages. Future Microbiol. 9, 327–341.
Chandra, G., Selvaraj, P., Jawahar, M.S., et al., 2004. Effect of vitamin D3 on phagocytic potential of macrophages with live *Mycobacterium tuberculosis* and lymphoproliferative response in pulmonary tuberculosis. J. Clin. Immunol. 24, 249–257.
Coussens, A.K., Wilkinson, R.J., Hanifa, Y., et al., 2012. Vitamin D accelerates resolution of inflammatory responses during tuberculosis treatment. Proc. Natl. Acad. Sci. U. S. A. 109, 15449–15554.
Daley, P., et al., 2015. Adjunctive vitamin D for treatment of active tuberculosis in India: a randomised, double-blind, placebo-controlled trial. Lancet Infect. Dis. 15 (5), 528–534.
Devi, U., Rao, C.M., Srivastava, V.K., et al., 2003. Effect of iron supplementation on mild to moderate anaemia in pulmonary tuberculosis. Br. J. Nutr. 90, 541–550.
Green, J.A., Lewin, S.R., Wightman, F., et al., 2005. A randomised controlled trial of oral zinc on the immune response to tuberculosis in HIV-infected patients. Int. J. Tubercul. Lung Dis. 9, 1378–1384.
Hmama, Z., Sendide, K., Talal, A., et al., 2003. Quantitative analysis of phagolysosomal fusion in intact cells: inhibition by mycobacterial lipoarabinomannan and rescue by an 1α, 25-dihydroxyvitamin D3-phosphoinositide 3-kinase pathway. J. Cell Sci. 117, 2131–2139.
Hanekom, W.A., Potgieter, S., Hughes, E.J., et al., 1997. Vitamin A status and therapy in childhood pulmonary tuberculosis. J. Pediatr. 131, 925–927.
Harries, A.D., Nkhoma, W.A., Thompson, P.J., Nyangulu, D.S., Wirima, J.J., 1988. Nutritional status in Malawian patients with pulmonary tuberculosis and response to chemotherapy. Eur. J. Clin. Nutr. 42, 445–450.
Karyadi, E., West, C.E., Schultnik, W., et al., 2002. A double-blind, placebo-controlled study of vitamin A and zinc supplementation in persons with tuberculosis in Indonesia: effects on clinical response and nutritional status. Am. J. Clin. Nutr. 75, 720–727.
Kennedy, N., Ramsay, A., Uiso, L., et al., 1996. Nutritional status and weight gain in patients with pulmonary tuberculosis in Tanzania. Trans. R. Soc. Trop. Med. Hyg. 90, 162–166.
Koyanagi, A., Kuffo, D., Gresely, L., et al., 2004. Relationships between serum concentrations of C-reactive protein and micronutrients in patients with tuberculosis. Ann. Trop. Med. Parasitol. 98, 391–399.
Lodha, R., Mukherjee, A., Singh, V., et al., 2014. Delhi Pediatric TB Study Group. Effect of micronutrient supplementation on treatment outcomes in children with intrathoracic tuberculosis: a randomized controlled trial. Am. J. Clin. Nutr. 100, 1287–1297.
Madebo, T., Lindtjorn, B., Aukrust, P., et al., 2003. Circulating antioxidants and lipid peroxidation products in untreated tuberculosis patients in Ethiopia. Am. J. Clin. Nutr. 78, 117–122.
Mugusi, F.M., Rusizoka, O., Habib, N., et al., 2003. Vitamin A status of patients presenting with pulmonary tuberculosis and asymptomatic HIV-infected individuals, Dar es Salaam, Tanzania. Int. J. Tubercul. Lung Dis. 7, 804–807.
Nathan, C., 2002. Inducible nitric oxide synthase in the tuberculous human lung. Am. J. Respir. Crit. Care Med. 166, 130–131.
Onwubalili, J.K., 1988. Malnutrition among tuberculosis patients in Harrow, England. Eur. J. Clin. Nutr. 363–366.
Perez-Guzman, C., Vargas, M.H., Quinonez, F., et al., 2005. A cholesterol-rich diet accelerates bacteriologic sterilisation in pulmonary tuberculosis. Chest 127, 643–651.
Ramakrishnan, C.V., Rajendran, K., Jacob, P.G., Fox, W., Radhakrishna, S., 1961. The role of diet in the treatment of pulmonary tuberculosis. An evaluation in a controlled chemotherapy study in home and sanatorium patients in South India. Bull. World Health Organ. 25, 339–359.
Range, N., Changalucha, J., Krarup, H., et al., 2006. The effect of multi-vitamin/mineral supplementation on mortality during treatment of pulmonary tuberculosis: a randomized two-by-two factorial trial in Mwanza, Tanzania. Br. J. Nutr. 95, 762–770.
Ralph, A.P., Waramori, G., Pontororing, G.J., et al., 2013. L-arginine and vitamin D adjunctive therapies in pulmonary tuberculosis: a randomised, double-blind, placebo-controlled trial. PLoS One 8, e70032.
Salahuddin, N., Ali, F., Hasan, Z., Rao, N., Aqeel, M., Mahmood, F., 2013. Vitamin D accelerates clinical recovery from tuberculosis: results of the SUCCINCT Study [Supplementary Cholecalciferol in recovery from tuberculosis]. A randomized, placebo-controlled, clinical trial of vitamin D supplementation in patients with pulmonary tuberculosis. BMC Infect. Dis. 13, 22.
Schon, T., Elias, D., Moges, F., et al., 2003. Arginine as an adjuvant to chemotherapy improves clinical outcome in active tuberculosis. Eur. Respir. J. 21, 483–488.
Selvaraj, P., Narayanan, P.R., Reetha, A.M., 2000. Association of vitamin D receptor genotypes with susceptibility to pulmonary tuberculosis in female patients and resistance in female contacts. Indian J. Med. Res. 111, 172–179.
Sinclair, D., Abba, K., Grobler, L., Sudarsanam, T.D., 2011. Nutritional supplements for people being treated for active tuberculosis. Cochrane Database Syst. Rev. (11) Art. No.: CD006086.
Tukvadze, N., et al., 2015. High-dose vitamin D3 in adults with pulmonary tuberculosis: a double-blind randomized controlled trial. Am. J. Clin. Nutr. 102 (5), 1059–1069.
Ulukavak, T., Ciftci, B., Yis, O., et al., 2003. Changes in serum selenium, copper, zinc levels and Cu/Zn ratio in patients with pulmonary tuberculosis during therapy. Biol. Trace Elem. Res. 95, 65–71.
Ustianowski, A., Shaffer, R., Collin, S., et al., 2005. Prevalence and associations of vitamin D deficiency in foreign-born persons with tuberculosis in London. J. Infect. 50, 432–437.
Wiid, I., Seaman, T., Hoal, E.G., et al., 2004. Total antioxidant levels are low during active TB and rise with anti-tuberculosis therapy. IUBMB Life 56, 101–106.
Wilkinson, R.J., Martin, L., Toossi, Z., et al., 2000. Influence of vitamin D deficiency and vitamin D receptor polymorphisms on tuberculosis among Gujarati Asians in west London: a case control study. Lancet 355, 618–621.

Water security and nutrition

Joshua D. Miller[a] **and Sera L. Young**[b,c], [a] Department of Nutrition, University of North Carolina, Chapel Hill, NC, United States; [b] Department of Anthropology, Northwestern University, Evanston, IL, United States; and [c] Institute for Policy Research, Northwestern University, Evanston, IL, United States

© 2023 Elsevier Ltd. All rights reserved.

Introduction 706
Defining and measuring water insecurity 707
Domains of water insecurity 707
Tools for quantifying water insecurity 708
Linkages between water insecurity and nutrition 708
Water: an essential nutrient 708
The nutritional importance of safe drinking water 709
Water security for food production 710
Water security for a healthy diet 710
Cooking and meal preparation 710
Human milk production and feeding 711
Complementary feeding 711
Pathogens and pollutants: threats to water security and health 711
Pathogens, heavy metals, and emerging pollutants 711
Diarrhea and environmental enteropathy 712
Altered microbiome and inflammation 712
Conclusion 712
Acknowledgments 713
References 713
Relevant website 716

Key points

- Water insecurity, like food insecurity, is a multidimensional concept that captures availability, accessibility, use, and stability across time
- The nutrition community has primarily focused on water in the context of waterborne diseases and hydration, but water security is critical for many other aspects of diet and physical well-being
- Water security shapes food production and preparation; infant and young child feeding; development and maintenance of the gut microbiome; and sanitation and hygiene practices that limit exposure to harmful substances that undermine nutrient uptake and function
- Validated measures of household and individual water insecurity should be implemented concurrently with other nutritional assessments to advance our understanding of the linkages between water insecurity and nutrition

Glossary

Water insecurity
The inability to reliably access water in appropriate quantities and acceptable quality for all household uses (Jepson et al., 2017).

Introduction

Water is life. It is an essential nutrient that serves a critical role in absorbing, storing, and metabolizing other nutrients (Jéquier and Constant, 2010). Reliable access to water in sufficient quantities and quality for all household uses, or "water security" (Jepson et al., 2017), is also integral for food production and preparation, personal hygiene, and psychological health (Young et al., 2021a). In other words, water security creates an enabling environment for good nutrition (Miller et al., 2021).

Multifarious and growing threats to water security, including prolonged droughts, excess flooding, and water contamination (Kummu et al., 2016; Mekonnen and Hoekstra, 2016), present substantial challenges to achieving global nutrition targets. Yet few studies to date have explored the interconnections between water security and nutrition, including how water needs and challenges change throughout the life course (Rosinger and Young, 2020b). We therefore describe how water security is conceptualized, operationalized, and quantified for empirical analysis. We then synthesize available evidence about the linkages between water (security) and nutrition and conclude by identifying future research opportunities.

Defining and measuring water insecurity

Water has traditionally been quantified in terms of physical availability at the national, state, watershed, or service-provider level by physical scientists and engineers. In recent decades, social scientists have increasingly examined the ways by which diverse water problems manifest at the individual and household level. This shift in unit of analysis was catalyzed, in part, by a seminal 1998 report that introduced "water insecurity" as a multidimensional concept that considers the dual role of water as both a nutrient for maintaining good health and a critical resource with sociocultural, political, and spiritual significance (Webb and Iskandarani, 1998). Although researchers continue to debate the degree to which definitions of water insecurity should emphasize environmental welfare, infrastructure development and management, or human capabilities (Bigas, 2013; Cook and Bakker, 2012; Slaymaker et al., 2020), there is broad consensus that a more holistic understanding of water beyond availability is required to address the global water crisis.

Experience-based constructs and metrics of food insecurity have gained prominence in global public health as a useful way to explore how resource insecurities shape behavior and well-being (Pérez-Escamilla, 2012; Pérez-Escamilla and Segall-Corrêa, 2008), and there is growing recognition of their utility for understanding water insecurity (Jepson et al., 2017; Miller et al., 2021; Young et al., 2021a). Such household- and individual-level measures are often more proximal to outcomes of importance to nutrition researchers, dieticians, practitioners, and policymakers than those measured at the community or national level (Young,2021). By analogy, the development and widespread implementation of experienced-based food insecurity scales has exposed nutrition issues and inequities that are masked by less granular data on food availability or use, such as food balance sheets or calories per capita (Barrett, 2010; Jones et al., 2013; Leroy et al., 2015). Our focus herein is therefore on experiential water insecurity.

Domains of water insecurity

Definitions of water insecurity consider one or more of the following: water availability, access, use, and stability (**Fig. 1**) (Rosinger and Young, 2020b; Slaymaker et al., 2020). These domains align with those used in the conceptualization of food security (Jones et al., 2013).

Availability refers to the physical existence of water in the environment. Water security can also be conceived of in terms of access, or whether water can be acquired (e.g., affordable) through socially appropriate means. Availability is necessary, but not sufficient, for access. Access, in turn, is necessary but not sufficient for use, which considers if there is enough acceptable and safe water for all personal or households needs. Finally, stability (sometimes referred to as reliability) considers continuity in availability, access, and use across time.

Water availability is a physical phenomenon that can be measured directly, but access is a more complex construct that is shaped by structural, social, and individual factors. For example, physical access may be hampered by infrastructural problems (e.g., groundwater is available but inaccessible due to power cuts) and age, pregnancy status, or disease that limit one's ability to travel to water sources or haul heavy containers. Economic access, or water affordability, can be hampered by poverty or vagaries in water pricing. Additionally, cultural proscriptions and sociopolitical exclusion, biases, or discrimination may be barriers to equitable water access. For example, in North America, inequity in water access is predominantly experienced by indigenous people and people of color due to environmental racism and water policies that disproportionately benefit majority white communities (Meehan et al., 2020a,b).

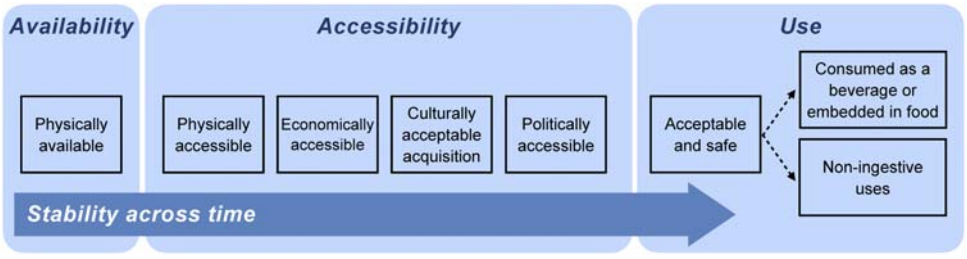


Fig. 1 Conceptual framework detailing the primary domains of experiential water security, including water availability, accessibility, use, and stability across time. Adapted with permission from Rosinger and Young (2020b), Slaymaker et al. (2020).

Use is also a multi-faceted domain, in part because water serves many purposes and the quality, quantity, or provenance of water considered acceptable for each use varies. We therefore distinguish between acceptability for consumptive and non-ingestive uses (**Fig. 1**). Consumption-related water use includes that which is drunk or eaten, including water that is embedded in foods ("virtual water") (Virtual Water Trade: Proceedings of the International Expert Meeting on Virtual Water Trade, 2003). Non-ingestive uses of water are numerous and include sanitation and hygiene practices (e.g., washing hands, food, utensils, or clothing), recreational activities, and income-generating activities that are integral to the economic productivity of many households. For instance, water is necessary for raising crops and livestock as well as producing goods that can be sold for profit.

Water acceptability is based on myriad factors including objective or subjective safety, organoleptic properties (e.g., taste, smell), spiritual and cultural beliefs, trust in water providers and public institutions, and the intended use for the water (Collins et al., 2019; Doria, 2010; Pierce and Gonzalez, 2017). Many households globally use multiple water sources since water for one activity may deemed to be of inappropriate quality for another (e.g., different water sources for drinking and washing clothes) (Elliott et al., 2019).

The fourth domain of water security is stability (or reliability). Social, economic, environmental, and political issues can affect the constancy of water security across time, such as change in water availability due to seasonal variations or climatic shocks. Further, water access can be shaped by fluctuations in health, income, infrastructural damage, and relationships with those who control water delivery (i.e., perturbations in social and economic capital). Diurnal stability can be shaped by queues for water (Collins et al., 2019) or power shortages (Biggs et al., 2015).

It is important to note that myriad resource-based conceptualizations of water are used interchangeably with experiential water insecurity, although they are distinct phenomena (Octavianti and Staddon, 2021). Sometimes the concepts are defined, such as in the case of water stress, water scarcity, and water risk (Schulte and Morrison, 2019). Other times such terms are used without explicit operationalization. This can create ambiguity about whether issues with water quantity, access, or quality are being assessed (Young, 2021).

Tools for quantifying water insecurity

Household or individual water insecurity is commonly measured using tools that ask respondents to report how frequently they experienced limitations to or deficiencies in water availability, accessibility, and use over a defined period (Jepson et al., 2017). Most items in experience-based scales probe about access and use (e.g., being unable to bathe or prepare foods because of problems with water) because these domains are contingent upon water availability (**Fig. 1**). Specific emotions related to one's water situation, such as worry or anger about insufficiency, are also sometimes asked about. Since issues with acceptability or safety manifest in suboptimal use or negative emotions, experiential scales implicitly capture these subdomains. Stability across time can be captured by asking respondents to recall the frequency of experiences across multiple timepoints (e.g., water insecurity in the prior month and year) in cross-sectional studies, or by using repeated measures in longitudinal assessments.

Experiential measures of food insecurity have been assessed by governments and international agencies in large-scale surveys for decades (Jones et al., 2013; Young et al., 2019a). In contrast, tools for measuring experiential water insecurity have only recently been developed, and their implementation has not been nearly as broad (Young et al., 2021a). The first effort at quantifying experiential water insecurity was done in a 2004–2005 study among 72 household heads in an urban setting in Bolivia (Wutich and Ragsdale, 2008). Since then, numerous scales have been developed to measure either household or individual experiences with water insecurity in specific settings (Octavianti and Staddon, 2021), such as western Kenya (Boateng et al., 2018), Ethiopia (Stevenson et al., 2012), and Nepal (Aihara et al., 2015). Typically, these scales ask about the frequency of a small subset of water-related activities in a specified timeframe, usually the prior month. The first scale validated from cross-cultural comparison is the 12-item Household Water Insecurity Experiences (HWISE) Scale (Young et al., 2019a,b), which was finalized in 2019. It is now being used in a range of venues, including nationally representative surveys to estimate water insecurity globally (Young et al., 2021b); in clinical studies to understand links to health and well-being (Miller et al., 2021); by NGOs like Oxfam to understand the impacts of interventions (Miller et al., 2020; Vonk, 2019); and by multi-lateral agencies such as IFPRI and UNICEF to assess the determinants and impacts of COVID-19 (Barooah et al., 2020).

Linkages between water insecurity and nutrition

Given that tools for measuring experiential water insecurity have only recently been developed, few studies have concurrently examined water availability, accessibility, use, and stability at the household or individual level in relation to nutrition. To illustrate the potential linkages between water insecurity and nutrition (**Fig. 2**), we therefore present findings from studies that either measure water insecurity directly or consider at least one of its four sub domains.

Water: an essential nutrient

The most obvious use of water is for drinking. Water is needed to maintain hydration and support normal physiological function. Chemically, water has a unique structure and properties that make it an essential nutrient (Jéquier and Constant, 2010). In the gut, water aids in digestion and serves as a universal solvent for maintaining concentration gradients to facilitate the uptake, transport,

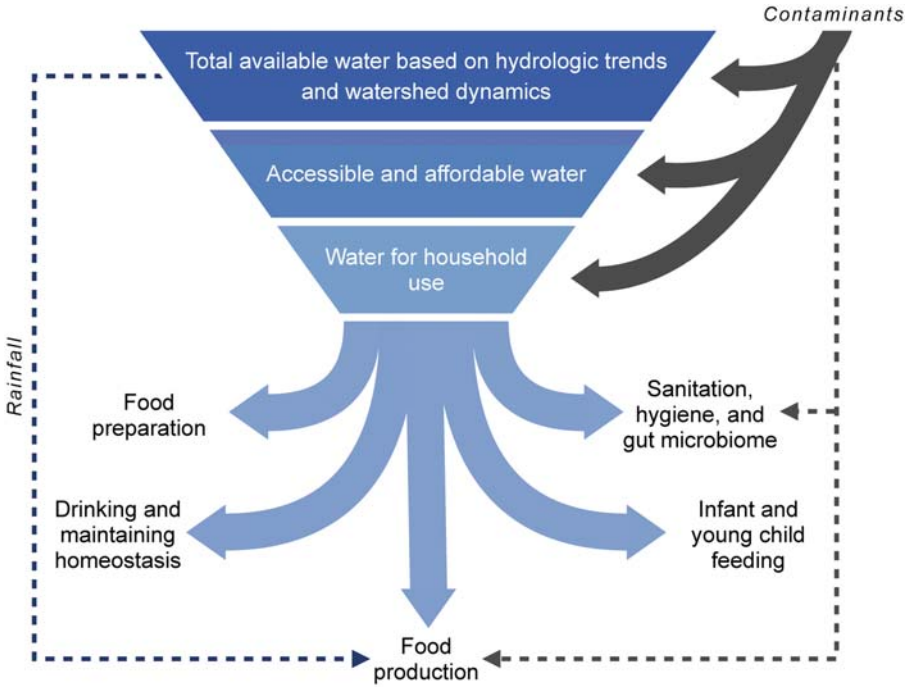


Fig. 2 Conceptual framework demonstrating the relationship between water security and nutrition. Water security is shaped at multiple ecological levels, from environmental factors (dark blue) to intra-household dynamics (light blue). Problems at any level can have negative impacts on downstream health and well-being, including key nutrition-related activities highlighted here. Adapted with permission from Miller et al. (2021).

and storage of other nutrients. Water is also used in many metabolic processes and hydrolytic reactions throughout the body. Further, it serves structural functions, including providing protection as a physical shock absorber at the cellular and tissue levels. Thermoregulation and waste excretion additionally depend on sufficient water stores (Jéquier and Constant, 2010; Popkin et al., 2010). Simply put, life cannot occur without water.

There are numerous and sometimes redundant pathways to ensure optimal total body water. Fluid balance is primarily maintained by matching outputs (e.g., urine excretion, insensible water loss through respiration) with inputs (e.g., direct fluid intake, consumption of foods that contain water) (Jéquier and Constant, 2010). Water needs vary based on individual characteristics, level of physical activity (Kenefick and Cheuvront, 2012), and across critical developmental periods (Rosinger, 2020). Net water loss results in dehydration and increases the risk of numerous morbidities, from headaches to kidney stones (Manz and Wentz, 2005; Popkin et al., 2010). Prolonged water restriction results in altered osmoregulation, hypotension, organ damage, and eventually death (Rosinger, 2020).

The nutritional importance of safe drinking water

Plain drinking water sometimes contains dissolved minerals in concentrations that are important for micronutrient metabolism (World Health Organization, 2005). For instance, millions of people drink "hard" groundwater that has biologically relevant concentrations of calcium and magnesium. Greater magnesium in drinking water is associated with lower risk of coronary heart disease and stroke mortality (Jiang et al., 2016; Rosanoff, 2013); higher calcium concentrations are associated with greater bone mineral density and lower risk of hip fracture (Dahl et al., 2015; Sengupta, 2013). Recommended sodium intakes can typically be achieved through food alone, such that drinking water with a high salt concentration is often associated with excess sodium consumption and concomitant hypertension (Choi et al., 2015; Talukder et al., 2017). Sodium is typically found in low concentrations in drinking water, but the salinity of many drinking water sources is increasing globally due to greater saltwater intrusion from excessive groundwater withdrawal and sea level rise (Vineis et al., 2011).

Although daily consumption of water is paramount for good health, few studies or national reporting agencies systematically measure hydration status or collect water intake data because of measurement challenges, including biases associated with dietary recall and limitations with biomarkers of hydration status (Cheuvront and Kenefick, 2014; Committee, 2020). Current recommended daily intakes can therefore only be considered crude estimates because they do not account for variations by climate, physical activity, body composition, medication use, or disease states (Howard et al., 2020). Indeed, the current MyPlate initiative to help inform healthy eating patterns in the United States does not include explicit recommendations about how much water to drink per day (although it encourages individuals to replace sugar-sweetened beverages [SSBs] with water) (Zizza et al., 2015). Taken together,

this has resulted in substantial confusion within the public about how much plain drinking water is required for good health (Stookey and Kavouras, 2020).

Non-water beverages have varied impacts on hydration status. Drinks with alcohol and caffeine can induce diuresis, with varied impacts on total body water (Maughan and Griffin, 2003). Juices and sugar-sweetened beverages (SSBs) can help restore hydration but can also contribute to excess caloric intake, thereby increasing the risk of overweight and obesity (Hu, 2013). Greater water intake may therefore complement weight maintenance strategies by decreasing overall energy intake and increasing fat oxidation (Stookey, 2016).

Non-water beverages may be preferentially consumed for their taste or sociocultural importance, but also because of distrust of drinking water (Onufrak et al., 2014). Non-water beverages may also be consumed in the face of water insecurity because of their convenience. For instance, one study found that individuals in rural Alaska who lacked in-home piped water consumed SSBs more frequently than their counterparts with greater water accessibility (Mosites et al., 2020). Infrastructural and economic disparities (Meehan et al., 2020a) may thereby contribute to disparate trends in SSB and plain water intake between racial groups and socio-economic strata (Rosinger and Young, 2020a; Vieux et al., 2020).

Water acquisition can also have nutritional consequences. Traveling to and fetching water from off-premises water sources is a physically demanding activity that necessitates considerable energy expenditure (Geere et al., 2018). One study among individuals living in a rural village in Laos estimated that, on average, 8.7% of daily calories consumed were spent on water fetching during the wet season and 12.8% during the dry season (Frenierre, 2017), meaning that having an off-premises water source may increase one's risk of undernutrition. Hauling heavy water is also associated with greater risk of injury (e.g., cervical spinal cord compression due to head-loading water, falling) (Geere et al., 2018; Venkataramanan et al., 2020b) that can indirectly impact nutrition by taking away time from income-generating activities or limiting an individual's ability to access or prepare preferred foods. This is substantial given that millions of households must travel more than 30 min roundtrip to collect water (WHO/UNICEF, 2017). More research is needed to understand how experiential water insecurity informs decisions about drinking water intake and associated health impacts.

Water security for food production

Just as water is necessary for human well-being, water is also fundamental for the growth and well-being of all plants and animals. As such, water is a critical resource for agricultural and aquacultural food production. More than two-thirds of freshwater withdrawals are currently used for agricultural activities like irrigating crops and watering livestock (Jury and Vaux, 2007). To support a growing global population in the context of climate change and intensifying water scarcity, innovative strategies will be required to grow more food with less water (Ringler and Paulo, 2020).

Smallholder farmers produce most domestic food supplies in sub-Saharan Africa and South Asia (IFAD and UNEP, 2013). The majority of these farmers rely on rainfed agriculture, which is less productive than irrigated operations (which more efficiently distribute water to crops and livestock) and more susceptible to climatic shocks (e.g., droughts) and changing weather patterns (Rockström et al., 2010). Expanding access to the financial and infrastructural resources required for irrigation could reduce market volatility and increase agricultural yields (Rosa et al., 2020). Greater food availability could in turn reduce rates of hunger and undernutrition in these regions, which are the highest globally (FAO et al., 2020).

Irrigation technologies can also indirectly improve nutritional well-being. Greater agricultural yields can increase a household's income, allowing individuals to purchase more (diverse) food (Ricciardi et al., 2020). Further, irrigation systems that produce potable water can also be used for drinking and other household activities, such as the cleaning and preparation of foods (Ringler et al., 2018).

Water security for a healthy diet

Cooking and meal preparation

Water is needed for the safe preparation and serving of foods. Washing fruits and vegetables with water can remove potentially harmful pesticides and residual soil matter, which can contain parasitic helminths that cause intestinal bleeding or other pathogens, and reduce one's ability to absorb nutrients (Stephenson et al., 2000). Water is also needed to wash hands and clean utensils before eating. Limited access to safe water can therefore increase the risk of foodborne illness and diarrhea (Gil et al., 2014).

Water is also needed to improve the palatability and digestibility of starchy staples. For instance, households in Kenya reporting having porridge but were unable to boil it because they lacked water (Collins et al., 2019; Zolnikov and Blodgett-Salafia, 2016). Other foods require water to remove toxins. Consumption of cassava that has not been boiled or soaked in water exposes individuals to glucosides that the body breaks down to neurotoxic cyanide, resulting in a disease known as konzo (Boivin et al., 2013). Individuals may therefore cope with water insecurity by consuming less food or substituting preferred foods with alternatives that require little or no water to prepare, including meals prepared outside the home (Venkataramanan et al., 2020a). For these reasons, water and food insecurity are strongly associated (Brewis et al., 2020; Miller et al., 2021; Young et al., 2021a), and water insecurity may precipitate future food insecurity (Boateng et al., 2020).

Human milk production and feeding

Water security is important for infant and young child feeding. Approximately 88% of human milk is comprised of water (Martin et al., 2016). Lactating individuals have greater water needs (Montgomery, 2002), particularly if they live in hot-humid climates where insensible water loss is greater (Rosinger, 2015). It is plausible that fluid intake and milk production are directly linked, such that even mild dehydration could result in lower output, but no human studies to date have found an association. This could be because most studies on this topic were conducted among small samples without substantial variation in fluid intake (Bentley, 1998). Alternatively, it is possible that humans have evolved to prioritize limited water for milk production, even during periods of dehydration, to ensure offspring survival (Kuzawa, 2020). More research is clearly needed to understand the dynamics between fluid restriction and human milk synthesis (Rosinger, 2020).

Water quality can also impact initiation and duration of human milk feeding. When drinking water is contaminated with heavy metals (e.g., mercury, lead), toxic compounds can bioaccumulate in human milk (Samiee et al., 2019). This is significant because repeated heavy metal exposure during sensitive periods of brain development, particularly early infancy, can result in lifelong neurocognitive deficits (Sanders et al., 2015). Exclusive human milk feeding, however, remains the preferred feeding method, even when drinking water quality is suboptimal, because infant formula and other foods prepared with contaminated water can expose infants to greater heavy metal concentrations (i.e., lactating individuals partially metabolize these substances before incorporation into human milk) (Rebelo and Caldas, 2016). Additionally, contaminated drinking water can contain waterborne pathogens or chemicals that cannot pass from caregiver to child through human milk (Baisley et al., 2011). As such, healthcare providers should address persistent caregiver misconceptions that water supplementation is needed to prevent human milk fed infants from becoming dehydrated (Nsiah-Asamoah et al., 2020; Yotebieng et al., 2013).

Other experiences with water insecurity may also negatively impact human milk feeding (Schuster et al., 2020). Greater water insecurity is associated with worse mental health and greater stress (Wutich et al., 2020; Young et al., 2019a), which can increase maternal sympathetic nervous system activity and thereby impair lactogenesis (Stuebe et al., 2013). Water insecurity may also exacerbate feelings of milk insufficiency or inadequacy, leading caregivers to introduce non-human milk foods before six months (Khatun et al., 2018). The physical burdens and opportunity costs associated with water insecurity also present barriers to caretakers' abilities and time to feed their infants (Schuster et al., 2020).

Complementary feeding

Water insecurity shapes complementary feeding as well. To date, most studies looking at water and complementary feeding have focused on pathogen transmission (Makasi et al., 2020). But as described above, problems with water can also limit the diversity and quantities of foods a household is able to purchase, produce, or prepare (Schuster et al., 2020). An analysis of DHS data from India, for instance, found that optimal water access, compared to intermediate or basic access, was associated with approximately a 2% higher probability of an infant meeting minimum dietary diversity (Choudhary et al., 2020). The small magnitude of effect may be due to the operationalization of water access, which did not capture other salient components of water insecurity, such as whether the water was of sufficient quantity or quality for preparing foods. Indeed, qualitative evidence from 19 sites globally suggests that caregivers experiencing issues with water may substitute preferred dishes with less nutrient-dense foods (Schuster et al., 2020). Caregivers who are experiencing water insecurity may also have more limited time and resources to notice or respond to feeding cues, which can influence an infant's lifelong dietary patterns and eating practices (Black and Aboud, 2011). More data are clearly needed to better characterize the relationships between water insecurity and infant and young child feeding.

Pathogens and pollutants: threats to water security and health

Pathogens, heavy metals, and emerging pollutants

Despite increasing global access to centralized piped water systems, unsafe water continues to substantially contribute to the global burden of disease, even in high-income countries (Murray et al., 2020). Drinking contaminated water can cause gut dysbiosis, limit nutrient absorption and increase gut leakage, and contribute to a pro-inflammatory state that is detrimental to overall well-being (Millward, 2017). Waterborne pathogens can cause illness after brief exposures, whereas chemical contaminants are typically harmful over longer periods of exposure (Damania et al., 2020).

There are hundreds of known waterborne pathogens, which include viruses, bacteria, parasitic protozoa and helminths, and fungi (Damania et al., 2020). Among these, viruses and bacteria are the most common cause of gastrointestinal distress globally (Gibson, 2014). Inhalation of mist containing bacteria can also cause respiratory disease. Outbreaks of Legionnaires' disease, for instance, are often attributed to poor water treatment and infrastructure maintenance (e.g., infrequent cleaning of HVAC systems), including in communities with centralized piped water networks (Phin et al., 2014). Antibiotic-resistant bacteria are also increasingly found in water sources, due in part to inappropriate disposal of antibiotics, and can transfer resistance to human pathogens (Sanganyado and Gwenzi, 2019).

Water can also be problematic because of heavy metal (e.g., cadmium, lead, mercury, and arsenic) contamination. Broadly, heavy metals can be detrimental by acting as competitive inhibitors and interfere with, for example, erythropoiesis and bone formation (Fernández-Luqueño et al., 2013; Zhang et al., 2020). Heavy metals can also alter the composition of the gut microbiota and induce dysbiosis (Duan et al., 2020).

"Contaminants of emerging concern" are those that are not commonly monitored or regulated but have known or suspected human health risks (Daughton, 2004). There are hundreds of such contaminants of emerging concern; they are found in household

goods, personal care products, and industrial additives and solvents (Geissen et al., 2015). Our understanding about the impacts of these emerging contaminants on nutritional health is extremely limited due to their novelty and the limited methods for their assessment. It is plausible that the relationship is bidirectional: contaminants may affect nutrient absorption, and nutritional status may modulate the effect of these pollutants (Hennig et al., 2012).

It is difficult to track water contamination due to numerous testing constraints (Thavarajah et al., 2020). Many national reporting agencies only assess the primary water source used by households, although use of multiple water sources—each of which may have its own suite of contaminants—is common globally (Daly et al., 2021). Additionally, few studies have investigated the use of inter-household water sharing networks, despite their ubiquity (Rosinger et al., 2020).

Diarrhea and environmental enteropathy

Disability-adjusted life-years attributable to diarrhea remain high despite decades-long efforts to limit exposure to pathogen-contaminated drinking water. The issue is particularly severe for children under five: unsafe water and sanitation caused nearly three-quarters of the almost 450,000 diarrheal deaths in 2016 (Troeger et al., 2018) and 16% of stunting among children under five in low- and middle-income countries (Prüss-Ustün et al., 2019).

Whereas diarrhea is most often an acute health hazard, environmental enteric dysfunction (EED) is a chronic, subclinical condition characterized by increased intestinal inflammation, gut permeability, and flattened villi that may severely undermine child well-being (Humphrey, 2009). EED has gained prominence within the water, sanitation, and hygiene (WASH) sector as a potentially under-appreciated mediator between environmental pathogens and poor child development. Most likely, EED is the result of repeated exposure to one or more (waterborne) pathogens that ultimately alter the structure and function of the gut (Owino et al., 2016). Numerous observational studies have found an association between indicators of EED and suboptimal nutrient absorption, stunted linear growth, restricted early childhood development, and lower oral vaccine effectiveness (Kosek et al., 2017).

Given its prevalence and adverse health effects, ameliorating the burden of EED is considered a public health priority. To that end, three large-scale randomized trials (WASH Benefits Bangladesh; WASH Benefits Kenya; Sanitation, Hygiene, Infant Nutrition Efficacy in Zimbabwe) in the 2010s received considerable attention for their efforts to reduce stunting by limiting environmental exposures to pathogens through improvements in both drinking water quality and sanitation and hygiene practices (Cumming et al., 2019). Many within the WASH sector were disappointed, however, when the interventions were found to only have mixed impacts on diarrhea and no effect on child linear growth, although improvements in water quality likely had beneficial impacts on other unmeasured aspects of child well-being (Leroy and Frongillo, 2019). The implementers of these studies concluded that strategies that address additional routes of exposure and other dimensions of water insecurity beyond quality (i.e., "transformative WASH") may be needed to meaningfully reduce the risk and impacts of EED (Cumming et al., 2019). It will be important to assess whether improvements in experiential water insecurity, as broadly conceptualized herein, decrease the risk of EED and related outcomes.

Altered microbiome and inflammation

The gut microbiome serves numerous critical functions, including providing energy to enterocytes, producing essential micronutrients, and preventing the colonization of pathogenic organisms (D'Argenio and Salvatore, 2015). It is broadly understood that the composition and stability of the gut microbiome is responsive to dietary and environmental factors, but the role of water (security) is only starting to be investigated.

Each drinking water source contains a signature microbiome that can directly alter the composition of an individual's gut microbiota or indirectly by changing the gut ecology (Jha et al., 2018; Pinto et al., 2012). Similarly, the chemical properties of drinking water may also influence the gut microbiota. For instance, work in the United Kingdom has found regional differences in the number and richness of species within fecal samples (i.e., alpha diversity) by the sodium, sulfate, and chloride content of local tap water sources (Bowyer et al., 2020). Enteric pathogen contamination may also shape gut colonization. In Nicaragua, investigators found that infants and young children living in households with higher levels of total coliforms in their drinking water had lower alpha diversity and a greater relative abundance of potentially predatory or pathogenic gut bacteria compared to those in households using water sources with low coliform concentrations (Piperata et al., 2020). This suggests that exposure to unsafe water may render the gut more susceptible to harmful bacteria.

Diet is a strong predictor of gut microbiota development and colonization (Scott et al., 2013). Given that water insecurity may cause food insecurity (Boateng et al., 2020) and lead to altered diets (Schuster et al., 2020), water problems can indirectly impact the gut microbiota. For instance, greater consumption of foods away from home and SSBs in response to limited water availability may favor the development of obesogenic gut microbiota (Ley, 2010). Gut microbiota can also be adversely affected by chronic stress (Carabotti et al., 2015; Zheng et al., 2013); the psychological distress that accompanies water insecurity (Workman et al., 2021; Wutich et al., 2020; Young et al., 2019a) may therefore shape the gut microbiome. Without better data on exposure to water insecurity and the microbiome, these relationships cannot be accurately ascertained.

Conclusion

Water security is essential for good nutrition. Recently developed tools for measuring experiential water insecurity are helping to advance our understanding of the many ways by which water availability, access, use, and stability influence diet and health.

Primary linkages between water security and nutrition include food production and preparation, sanitation and hygiene, psychosocial well-being, and the gut microbiome. Going forward, strategies to improve nutrition must consider the diverse ways by which water insecurity can be compromised. Sustained investment in interdisciplinary research, including systematic collection of high-resolution data on water insecurity in nutrition studies (Miller et al., 2021; Young, 2021), will advance our understanding of the global water and food crises and identify where resources should be targeted for greatest impact.

Acknowledgments

SLY was supported by the Andrew Carnegie Fellows Program.

Conflict of interest disclosure: No conflicts of interest to disclose.

References

Aihara, Y., Shrestha, S., Kazama, F., Nishida, K., 2015. Validation of household water insecurity scale in urban Nepal. Water Pol. 17, 1019–1032. https://doi.org/10.2166/wp.2015.116.

Baisley, K., Sarenje, K., Simuyandi, M., Clasen, T., Filteau, S., Peletz, R., Kelly, P., 2011. Drinking water quality, feeding practices, and diarrhea among children under 2 Years of HIV-positive mothers in peri-urban Zambia. Am. J. Trop. Med. Hyg. 85, 318–326. https://doi.org/10.4269/ajtmh.2011.11-0140.

Barooah, P., Wouterse, F., Saini, S., Alvi, M., Ringler, C., 2020. COVID-19 Challenges to Equity: Insights From Rural Nepal and Senegal. https://www.agrilinks.org/post/covid-19-challenges-equity-insights-rural-nepal-and-senegal.

Barrett, C., 2010. Measuring food insecurity. Science 327, 825–828. https://doi.org/10.1126/science.1182768.

Bentley, G.R., 1998. Hydration as a limiting factor in lactation. Am. J. Hum. Biol. 10, 151–161. https://doi.org/10.1002/(sici)1520-6300(1998)10:2<151::aid-ajhb2>3.0.co;2-o.

Bigas, H., 2013. Water Security and the Global Water Agenda: A UN-Water Analytical Brief. United Nations University, Hamilton, Ontario.

Biggs, E.M., Bruce, E., Boruff, B., Duncan, J.M.A., Horsley, J., Pauli, N., McNeill, K., Neef, A., Ogtrop, F.V., Curnow, J., Haworth, B., Duce, S., Imanari, Y., 2015. Sustainable development and the water–energy–food nexus: a perspective on livelihoods. Environ. Sci. Pol. 54, 389–397. https://doi.org/10.1016/j.envsci.2015.08.002.

Black, M.M., Aboud, F.E., 2011. Responsive feeding is embedded in a theoretical framework of responsive parenting. J. Nutr. 141, 490–494. https://doi.org/10.3945/jn.110.129973.

Boateng, G.O., Collins, S.M., Mbullo, P., Wekesa, P., Onono, M., Neilands, T.B., Young, S.L., Gnambs, T., 2018. A novel household water insecurity scale: procedures and psychometric analysis among postpartum women in western Kenya. PLoS One 13, e0198591. https://doi.org/10.1371/journal.pone.0198591.

Boateng, G., Workman, C., Miller, J., Onono, M., Neilands, T., Young, S., 2020. The syndemic effects of food insecurity, water insecurity, and HIV on depressive symptomatology among Kenyan women. Soc. Sci. Med.

Boivin, M.J., Okitundu, D., Bumoko, G.M.-M., Sombo, M.-T., Mumba, D., Tylleskar, T., Page, C.F., Muyembe, J.-J.T., Tshala-Katumbay, D., 2013. Neuropsychological effects of konzo: a neuromotor disease associated with poorly processed cassava. Pediatrics 131, e1231–e1239. https://doi.org/10.1542/peds.2012-3011.

Bowyer, R.C.E., Schillereff, D.N., Jackson, M.A., Roy, C.L., Wells, P.M., Spector, T.D., Steves, C.J., 2020. Associations between UK tap water and gut microbiota composition suggest the gut microbiome as a potential mediator of health differences linked to water quality. Sci. Total Environ. 739, 139697. https://doi.org/10.1016/j.scitotenv.2020.139697.

Brewis, A., Workman, C., Wutich, A., Jepson, W., Young, S., HWISE-RCN, Adams, E., Ahmed, J.F., Alexander, M., Balogun, M., Boivin, M., Carrillo, G., Chapman, K., Cole, S., Collins, S., Figueroa, L., Freeman, M., Gershim, A., Ghattas, H., Hagaman, A., Jamaluddine, Z., Jepson, W., Tshala-Katumbay, D., Krishnakumar, D., Maes, K., Mathad, J., Maupin, J., Mbullo, P., Miller, J., Muslin, I.M., Niesluchowski, M., Omidvar, N., Pearson, A., Melgar-Quiñonez, H., Sanchez-Rodríguez, C., Rosinger, A., Santoso, M.V., Schuster, R., Srivastava, S., Staddon, C., Stoler, J., Sullivan, A., Tesfaye, Y., Triviño, N., Trowell, A., Tutu, R., Escobar-Vargar, J., Zinab, H., 2020. Household water insecurity is strongly associated with food insecurity: evidence from 27 sites in low- and middle-income countries. Am. J. Hum. Biol. 32. https://doi.org/10.1002/ajhb.23309.

Carabotti, M., Scirocco, A., Maselli, M.A., Severi, C., 2015. The gut-brain axis: interactions between enteric microbiota, central and enteric nervoussystems. Ann. Gastroenterol. 28, 203–209.

Cheuvront, S.N., Kenefick, R.W., 2014. Dehydration: physiology, assessment, and performance effects. Comp. Physiol. 4, 29.

Choi, H.Y., Park, H.C., Ha, S.K., 2015. Salt sensitivity and hypertension: a paradigm shift from kidney malfunction to vascular endothelial dysfunction. Electrolyte Blood Press 13, 7. https://doi.org/10.5049/ebp.2015.13.1.7.

Choudhary, N., Schuster, R., Brewis, A., Wutich, A., 2020. Water insecurity potentially undermines dietary diversity of children aged 6–23 months: evidence from India. Matern. Child Nutr. 16. https://doi.org/10.1111/mcn.12929.

Collins, S.M., Owuor, P.M., Miller, J.D., Boateng, G.O., Wekesa, P., Onono, M., Young, S.L., 2019. "I know how stressful it is to lack water!" Exploring the lived experiences of household water insecurity among pregnant and postpartum women in western Kenya. Global Publ. Health 14, 649–662. https://doi.org/10.1080/17441692.2018.1521861.

Committee, D.G.A., 2020. Scientific Report of the 2020 Dietary Guidelines Advisory Committee: Advisory Report to the Secretary of Agriculture and the Secretary of Health and Human Services. Department of Agriculture, Agricultural Research Service, Washington, DC.

Cook, C., Bakker, K., 2012. Water security: debating an emerging paradigm. Glob. Environ. Change 22, 94–102. https://doi.org/10.1016/j.gloenvcha.2011.10.011.

Cumming, O., Arnold, B.F., Ban, R., Clasen, T., Mills, J.E., Freeman, M.C., Gordon, B., Guiteras, R., Howard, G., Hunter, P.R., Johnston, R.B., Pickering, A.J., Prendergast, A.J., Prüss-Ustün, A., Rosenboom, J.W., Spears, D., Sundberg, S., Wolf, J., Null, C., Luby, S.P., Humphrey, J.H., Colford, J.M., 2019. The implications of three major new trials for the effect of water, sanitation and hygiene on childhood diarrhea and stunting: a consensus statement. BMC Med. 17, 173. https://doi.org/10.1186/s12916-019-1410-x.

Dahl, C., Søgaard, A.J., Tell, G.S., Forsén, L., Flaten, T.P., Hongve, D., Omsland, T.K., Holvik, K., Meyer, H.E., Aamodt, G., 2015. Population data on calcium in drinking water and hip fracture: an association may depend on other minerals in water. A NOREPOS 1 1 Norwegian Epidemiologic Osteoporosis Studies. Bone 81, 292–299. https://doi.org/10.1016/j.bone.2015.07.020.

Daly, S., Lowe, J., Hornsby, G., Harris, A., 2021. Multiple water source use in low- and middle-income countries: a systematic review. J. Water Health. https://doi.org/10.2166/wh.2021.205.

Damania, R., Desbureaux, S., Rodella, A.-S., Russ, J., Zaveri, E., 2020. Quality Unknown: The Invisible Water Crisis. World Bank Group, Washington.

D'Argenio, V., Salvatore, F., 2015. The role of the gut microbiome in the healthy adult status. Clin. Chim. Acta 451, 97–102. https://doi.org/10.1016/j.cca.2015.01.003.

Daughton, C.G., 2004. Non-regulated water contaminants: emerging research. Environ. Impact Assess. 24, 711–732. https://doi.org/10.1016/j.eiar.2004.06.003.

Doria, M. de F., 2010. Factors influencing public perception of drinking water quality. Water Pol. 12, 1–19. https://doi.org/10.2166/wp.2009.051.

Duan, H., Yu, L., Tian, F., Zhai, Q., Fan, L., Chen, W., 2020. Gut microbiota: a target for heavy metal toxicity and a probiotic protective strategy. Sci. Total Environ. 742, 140429. https://doi.org/10.1016/j.scitotenv.2020.140429.

Elliott, M., Foster, T., MacDonald, M.C., Harris, A.R., Schwab, K.J., Hadwen, W.L., 2019. Addressing how multiple household water sources and uses build water resilience and support sustainable development. NPJ Clean Water 2, 6. https://doi.org/10.1038/s41545-019-0031-4.

FAO, IFAD, UNICEF, WFP, WHO, 2020. The State of Food Security and Nutrition in the World 2020.

Fernández-Luqueño, F., López-Valdez, F., Gamero-Melo, P., Luna-Suárez, S., Aguilera-González, E., Martínez, A., García-Guillermo, M., Hernández-Martínez, G., Herrera-Mendoza, R., Álvarez-Garza, M., Pérez-Velázquez, I., 2013. Heavy metal pollution in drinking water—a global risk for human health: a review. Afr. J. Environ. Sci. Technol. 7. https://doi.org/10.5897/ajest12.197.

Frenierre, J.L., 2017. Caloric expenditure as an indicator of access to water. wH_2O 5, 32–50.

Geere, J.-A.L., Cortobius, M., Geere, J.H., Hammer, C.C., Hunter, P.R., 2018. Is water carriage associated with the water carrier's health? A systematic review of quantitative and qualitative evidence. BMJ Glob. Health 3, e000764. https://doi.org/10.1136/bmjgh-2018-000764.

Geissen, V., Mol, H., Klumpp, E., Umlauf, G., Nadal, M., van der Ploeg, M., Van de Zee, S.E.A.T.M., Ritsema, C.J., 2015. Emerging pollutants in the environment: a challenge for water resource management. Int. Soil Water Conserv. Res. 3, 57–65. https://doi.org/10.1016/j.iswcr.2015.03.002.

Gibson, K.E., 2014. Viral pathogens in water: occurrence, public health impact, and available control strategies. Curr. Opin. Virol. 4, 50–57. https://doi.org/10.1016/j.coviro.2013.12.005.

Gil, A.I., Lanata, C.F., Hartinger, S.M., Mäusezahl, D., Padilla, B., Ochoa, T.J., Lozada, M., Pineda, I., Verastegui, H., 2014. Fecal contamination of food, water, hands, and kitchen utensils at the household level in rural areas of Peru. J. Environ. Health 76, 102–106.

Hennig, B., Ormsbee, L., McClain, C.J., Watkins, B.A., Blumberg, B., Bachas, L.G., Sanderson, W., Thompson, C., Suk, W.A., 2012. Nutrition can modulate the toxicity of environmental pollutants: implications in risk assessment and human health. Environ. Health Perspect. 120, 771–774. https://doi.org/10.1289/ehp.1104712.

Howard, G., Bartram, J., Williams, A., Overbo, A., Fuente, D., Geere, J.-A., 2020. Domestic Water Quantity, Service Level and Health, second ed. World Health Organization, Geneva.

Hu, F.B., 2013. Resolved: there is sufficient scientific evidence that decreasing sugar-sweetened beverage consumption will reduce the prevalence of obesity and obesity-related diseases: sugar-sweetened beverages and risk of obesity. Obes. Rev. 14, 606–619. https://doi.org/10.1111/obr.12040.

Humphrey, J.H., 2009. Child undernutrition, tropical enteropathy, toilets, and handwashing. Lancet 374, 1032–1035. https://doi.org/10.1016/s0140-6736(09)60950-8.

IFAD, UNEP, 2013. Smallholders, Food Security, and the Environment. IFAD, UNEP, Rome, Italy.

Jepson, W.E., Wutich, A., Colllins, S.M., Boateng, G.O., Young, S.L., 2017. Progress in household water insecurity metrics: a cross-disciplinary approach. Wiley Interdiscip. Rev. Water 4, e1214. https://doi.org/10.1002/wat2.1214.

Jéquier, E., Constant, F., 2010. Water as an essential nutrient: the physiological basis of hydration. Eur. J. Clin. Nutr. 64, 115–123. https://doi.org/10.1038/ejcn.2009.111.

Jha, A.R., Davenport, E.R., Gautam, Y., Bhandari, D., Tandukar, S., Ng, K.M., Fragiadakis, G.K., Holmes, S., Gautam, G.P., Leach, J., Sherchand, J.B., Bustamante, C.D., Sonnenburg, J.L., Metcalf, J., 2018. Gut microbiome transition across a lifestyle gradient in Himalaya. PLoS Biol. 16, e2005396. https://doi.org/10.1371/journal.pbio.2005396.

Jiang, L., He, P., Chen, J., Liu, Y., Liu, D., Qin, G., Tan, N., 2016. Magnesium levels in drinking water and coronary heart disease mortality risk: a meta-analysis. Nutrients 8, 5. https://doi.org/10.3390/nu8010005.

Jones, A.D., Ngure, F.M., Pelto, G., Young, S.L., 2013. What are we assessing when we measure food security? A compendium and review of current metrics. Adv. Nutr. Int. Rev. J. 4, 481–505. https://doi.org/10.3945/an.113.004119.

Jury, W.A., Vaux, H.J., 2007. The emerging global water crisis: managing scarcity and conflict between water users. Adv. Agron. 1–76.

Kenefick, R.W., Cheuvront, S.N., 2012. Hydration for recreational sport and physical activity. Nutr. Rev. 70, S137–S142. https://doi.org/10.1111/j.1753-4887.2012.00523.x.

Khatun, H., Comins, C.A., Shah, R., Islam, M.M., Choudhury, N., Ahmed, T., 2018. Uncovering the barriers to exclusive breastfeeding for mothers living in Dhaka's slums: a mixed method study. Int. Breastfeed. J. 13, 44. https://doi.org/10.1186/s13006-018-0186-5.

Kosek, M.N., Ahmed, T., Bhutta, Z., et al., 2017. Causal pathways from enteropathogens to environmental enteropathy: findings from the MAL-ED Birth Cohort Study. Ebiomedicine 18, 109–117. https://doi.org/10.1016/j.ebiom.2017.02.024.

Kummu, M., Guillaume, J.H.A., Moel, H. de, Eisner, S., Flörke, M., Porkka, M., Siebert, S., Veldkamp, T.I.E., Ward, P.J., 2016. The world's road to water scarcity: shortage and stress in the 20th century and pathways towards sustainability. Sci. Rep. 6, 38495. https://doi.org/10.1038/srep38495.

Kuzawa, C.W., 2020. Pregnancy as an intergenerational conduit of adversity: how nutritional and psychosocial stressors reflect different historical timescales of maternal experience. Curr. Opin. Behav. Sci. 36, 42–47. https://doi.org/10.1016/j.cobeha.2020.07.001.

Leroy, J.L., Frongillo, E.A., 2019. Perspective: what does stunting really mean? A critical review of the evidence. Adv. Nutr. 10, 196–204. https://doi.org/10.1093/advances/nmy101.

Leroy, J.L., Ruel, M., Frongillo, E.A., Harris, J., Ballard, T.J., 2015. Measuring the food access dimension of food security. Food Nutr. Bull. 36, 167–195. https://doi.org/10.1177/0379572115587274.

Ley, R.E., 2010. Obesity and the human microbiome. Curr. Opin. Gastroenterol. 26, 5–11. https://doi.org/10.1097/mog.0b013e328333d751.

Makasi, R.R., Humphrey, J.H., Michaelsen, K.F., Neufeld, L.M., Prentice, A.M., 2020. Summarizing the child growth and diarrhea findings of the water, sanitation, and hygiene benefits and sanitation hygiene infant nutrition efficacy trials. In: Nestlé Nutrition Institute Workshop Series. S. Karger AG, pp. 153–166.

Manz, F., Wentz, A., 2005. The importance of good hydration for the prevention of chronic diseases. Nutr. Rev. 63, S2–S5. https://doi.org/10.1111/j.1753-4887.2005.tb00150.x.

Martin, C., Ling, P.-R., Blackburn, G., 2016. Review of infant feeding: key features of breast milk and infant formula. Nutrients 8, 279.

Maughan, R.J., Griffin, J., 2003. Caffeine ingestion and fluid balance: a review. J. Hum. Nutr. Diet. 16, 411–420. https://doi.org/10.1046/j.1365-277x.2003.00477.x.

Meehan, K., Jepson, W., Harris, L.M., Wutich, A., Beresford, M., Fencl, A., London, J., Pierce, G., Radonic, L., Wells, C., Wilson, N.J., Adams, E.A., Arsenault, R., Brewis, A., Harrington, V., Lambrinidou, Y., McGregor, D., Patrick, R., Pauli, B., Pearson, A.L., Shah, S., Splichalova, D., Workman, C., Young, S., 2020a. Exposing the myths of household water insecurity in the global north: a critical review. Wiley Interdiscip. Rev. Water 7. https://doi.org/10.1002/wat2.1486.

Meehan, K., Jurjevich, J.R., Chun, N.M.J.W., Sherrill, J., 2020b. Geographies of insecure water access and the housing–water nexus in US cities. Proc. Natl. Acad. Sci. U. S. A. 117, 28700–28707. https://doi.org/10.1073/pnas.2007361117.

Mekonnen, M.M., Hoekstra, A.Y., 2016. Four billion people facing severe water scarcity. Sci. Adv. 2, e1500323. https://doi.org/10.1126/sciadv.1500323.

Miller, J., Vonk, J., Staddon, C., Young, S., 2020. Is household water insecurity a link between water governance and well-being? A multi-site analysis. J. Water Sanit. Hyg. Dev.

Miller, J.D., Frongillo, E.A., Weke, E., Burger, R., Wekesa, P., Sheira, L.A., Mocello, A.R., Bukusi, E.A., Otieno, P., Cohen, C.R., Weiser, S.D., Young, S.L., 2021. Household water and food insecurity are positively associated with poor mental and physical health among adults living with HIV in western Kenya. J. Nutr. https://doi.org/10.1093/jn/nxab030.

Miller, J.D., Workman, C.L., Panchang, S.V., Sneegas, G., Adams, E.A., Young, S.L., Thompson, A.L., 2021. Water Security and Nutrition: Current Knowledge and Research Opportunities. Adv. Nutr. https://doi.org/10.1093/advances/nmab075.

Millward, D.J., 2017. Nutrition, infection and stunting: the roles of deficiencies of individual nutrients and foods, and of inflammation, as determinants of reduced linear growth of children. Nutr. Res. Rev. 30, 50–72. https://doi.org/10.1017/s0954422416000238.

Montgomery, K.S., 2002. Nutrition column an update on water needs during pregnancy and beyond. J. Perinat. Educ. 11, 40–42. https://doi.org/10.1624/105812402x88830.

Mosites, E., Seeman, S., Fenaughty, A., Fink, K., Eichelberger, L., Holck, P., Thomas, T.K., Bruce, M.G., Hennessy, T.W., 2020. Lack of in-home piped water and reported consumption of sugar-sweetened beverages among adults in rural Alaska. Publ. Health Nutr. 23, 861–868. https://doi.org/10.1017/s1368980019002477.

Murray, C.J.L., Aravkin, A.Y., Zheng, P., et al., 2020. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 396, 1223–1249. https://doi.org/10.1016/s0140-6736(20)30752-2.

Nsiah-Asamoah, C., Doku, D.T., Agblorti, S., Enuameh, Y.A.K., 2020. Mothers' and Grandmothers' misconceptions and socio-cultural factors as barriers to exclusive breastfeeding: a qualitative study involving Health Workers in two rural districts of Ghana. PLoS One 15, e0239278. https://doi.org/10.1371/journal.pone.0239278.

Octavianti, T., Staddon, C., 2021. A review of 80 assessment tools measuring water security. Wiley Interdiscip. Rev. Water 8. https://doi.org/10.1002/wat2.1516.

Onufrak, S., Park, S., Sharkey, J., Sherry, B., 2014. The relationship of perceptions of tap water safety with intake of sugar-sweetened beverages and plain water among US adults. Publ. Health Nutr. 17, 179–185. https://doi.org/10.1017/s1368980012004600.

Owino, V., Ahmed, T., Freemark, M., Kelly, P., Loy, A., Manary, M., Loechl, C., 2016. Environmental enteric dysfunction and growth failure/stunting in global child health. Pediatrics 138. https://doi.org/10.1542/peds.2016-0641.
Pérez-Escamilla, R., Segall-Corrêa, A.M., 2008. Food insecurity measurement and indicators. Rev. Nutr. 21, 15s–26s. https://doi.org/10.1590/s1415-52732008000700003.
Pérez-Escamilla, R., 2012. Can experience-based household food security scales help improve food security governance? Glob. Food Sec. 1, 120–125. https://doi.org/10.1016/j.gfs.2012.10.006.
Phin, N., Parry-Ford, F., Harrison, T., Stagg, H.R., Zhang, N., Kumar, K., Lortholary, O., Zumla, A., Abubakar, I., 2014. Epidemiology and clinical management of Legionnaires' disease. Lancet Infect. Dis. 14, 1011–1021. https://doi.org/10.1016/s1473-3099(14)70713-3.
Pierce, G., Gonzalez, S., 2017. Mistrust at the tap? Factors contributing to public drinking water (mis)perception across US households. Water Pol. 19, 1–12. https://doi.org/10.2166/wp.2016.143.
Pinto, A.J., Xi, C., Raskin, L., 2012. Bacterial community structure in the drinking water microbiome is governed by filtration processes. Environ. Sci. Technol. 46, 8851–8859. https://doi.org/10.1021/es302042t.
Piperata, B.A., Lee, S., Apaza, A.C.M., Cary, A., Vilchez, S., Oruganti, P., Garabed, R., Wilson, W., Lee, J., 2020. Characterization of the gut microbiota of Nicaraguan children in a water insecure context. Am. J. Hum. Biol. 32. https://doi.org/10.1002/ajhb.23371.
Popkin, B.M., D'Anci, K.E., Rosenberg, I.H., 2010. Water, hydration, and health. Nutr. Rev. 68, 439–458. https://doi.org/10.1111/j.1753-4887.2010.00304.x.
Prüss-Ustün, A., Wolf, J., Bartram, J., Clasen, T., Cumming, O., Freeman, M.C., Gordon, B., Hunter, P.R., Medlicott, K., Johnston, R., 2019. Burden of disease from inadequate water, sanitation and hygiene for selected adverse health outcomes: an updated analysis with a focus on low- and middle-income countries. Int. J. Hyg. Environ. Health 222, 765–777. https://doi.org/10.1016/j.ijheh.2019.05.004.
Rebelo, F.M., Caldas, E.D., 2016. Arsenic, lead, mercury and cadmium: toxicity, levels in breast milk and the risks for breastfed infants. Environ. Res. 151, 671–688. https://doi.org/10.1016/j.envres.2016.08.027.
Ricciardi, V., Wane, A., Sidhu, B.S., Godde, C., Solomon, D., McCullough, E., Diekmann, F., Porciello, J., Jain, M., Randall, N., Mehrabi, Z., 2020. A scoping review of research funding for small-scale farmers in water scarce regions. Nat. Sustain. 3, 836–844. https://doi.org/10.1038/s41893-020-00623-0.
Ringler, C., Paulo, D., 2020. Water and Nutrition: Harmonizing Actions for the United Nations Decade of Action on Nutrition and the United Nations Water Action Decade. UNSCN.
Ringler, C., Choufani, J., Chase, C., McCartney, M., Mateo-Sagasta, J., Mekonnen, D., Dickens, C., 2018. Meeting the nutrition and water targets of the Sustainable Development Goals: achieving progress through linked interventions. In: International Water Management Institute (IWMI). CGIAR Research Program on Water, Land and Ecosystems (WLE). The World Bank.
Rockström, J., Karlberg, L., Wani, S.P., Barron, J., Hatibu, N., Oweis, T., Bruggeman, A., Farahani, J., Qiang, Z., 2010. Managing water in rainfed agriculture—the need for a paradigm shift. Agric. Water Manag. 97, 543–550. https://doi.org/10.1016/j.agwat.2009.09.009.
Rosa, L., Chiarelli, D.D., Rulli, M.C., Dell'Angelo, J., D'Odorico, P., 2020. Global agricultural economic water scarcity. Sci. Adv. 6, eaaz6031. https://doi.org/10.1126/sciadv.aaz6031.
Rosanoff, A., 2013. The high heart health value of drinking-water magnesium. Med. Hypotheses 81, 1063–1065. https://doi.org/10.1016/j.mehy.2013.10.003.
Rosinger, A., Young, S., 2020a. In-home tap water consumption trends changed among U.S. Children, but not adults, between 2007 and 2016. Water Resour. Res. 56. https://doi.org/10.1029/2020wr027657.
Rosinger, A.Y., Young, S.L., 2020b. The toll of household water insecurity on health and human biology: current understandings and future directions. Wiley Interdiscip. Rev. Water 7. https://doi.org/10.1002/wat2.1468.
Rosinger, A.Y., Brewis, A., Wutich, A., Jepson, W., Staddon, C., Stoler, J., Young, S.L., 2020. Water borrowing is consistently practiced globally and is associated with water-related system failures across diverse environments. Glob. Environ. Change 64, 102148. https://doi.org/10.1016/j.gloenvcha.2020.102148.
Rosinger, A., 2015. Dehydration among lactating mothers in the Amazon: a neglected problem: dehydration among lactating mothers in the Amazon. Am. J. Hum. Biol. 27, 576–578. https://doi.org/10.1002/ajhb.22672.
Rosinger, A.Y., 2020. Biobehavioral variation in human water needs: how adaptations, early life environments, and the life course affect body water homeostasis. Am. J. Hum. Biol. 32. https://doi.org/10.1002/ajhb.23338.
Rzepa, A., Young, S.L., Makariagakis, A., 2020. Monitoring the World's Water Access and Use. https://news.gallup.com/opinion/gallup/320825/monitoring-world-water-access.aspx.
Samiee, F., Vahidinia, A., Javad, M.T., Leili, M., 2019. Exposure to heavy metals released to the environment through breastfeeding: a probabilistic risk estimation. Sci. Total Environ. 650, 3075–3083. https://doi.org/10.1016/j.scitotenv.2018.10.059.
Sanders, A.P., Henn, B.C., Wright, R.O., 2015. Perinatal and childhood exposure to cadmium, manganese, and metal mixtures and effects on cognition and behavior: a review of recent literature. Curr. Environ. Heal Rep. 2, 284–294. https://doi.org/10.1007/s40572-015-0058-8.
Sanganyado, E., Gwenzi, W., 2019. Antibiotic resistance in drinking water systems: occurrence, removal, and human health risks. Sci. Total Environ. 669, 785–797. https://doi.org/10.1016/j.scitotenv.2019.03.162.
Schulte, P., Morrison, J., 2019. Driving Harmonization of Water-Related Terminology. Pacific Institute, Oakland, CA.
Schuster, R.C., Butler, M.S., Wutich, A., et al., 2020. "If there is no water, we cannot feed our children": the far-reaching consequences of water insecurity on infant feeding practices and infant health across 16 low- and middle-income countries. Am. J. Hum. Biol. 32. https://doi.org/10.1002/ajhb.23357.
Scott, K.P., Gratz, S.W., Sheridan, P.O., Flint, H.J., Duncan, S.H., 2013. The influence of diet on the gut microbiota. Pharmacol. Res. 69, 52–60. https://doi.org/10.1016/j.phrs.2012.10.020.
Sengupta, P., 2013. Potential health impacts of hard water. Int. J. Prev. Med. 4, 866–875.
Slaymaker, T., Johnston, R., Young, S., Miller, J., Staddon, C., 2020. Detailed Review of a Recent Publication: An Innovative Measure of the Experience of Water Insecurity Can Help Identify Vulnerable Households and Evaluate Interventions. WaSH Policy Research Digest.
Stephenson, L.S., Latham, M.C., Ottesen, E.A., 2000. Malnutrition and parasitic helminth infections. Parasitology 121, S23–S38. https://doi.org/10.1017/s0031182000006491.
Stevenson, E.G.J., Greene, L.E., Maes, K.C., Ambelu, A., Tesfaye, Y.A., Rheingans, R., Hadley, C., 2012. Water insecurity in 3 dimensions: an anthropological perspective on water and women's psychosocial distress in Ethiopia. Soc. Sci. Med. 75, 392–400. https://doi.org/10.1016/j.socscimed.2012.03.022.
Stookey, J.D., Kavouras, S.A., 2020. Water researchers do not have a strategic plan for gathering evidence to inform water intake recommendations to prevent chronic disease. Nutrients 12, 3359. https://doi.org/10.3390/nu12113359.
Stookey, J.J.D., 2016. Negative, null and beneficial effects of drinking water on energy intake, energy expenditure, fat oxidation and weight change in randomized trials: a qualitative review. Nutrients 8, 19. https://doi.org/10.3390/nu8010019.
Stuebe, A.M., Grewen, K., Meltzer-Brody, S., 2013. Association between maternal mood and oxytocin response to breastfeeding. J. Womens Health 22, 352–361. https://doi.org/10.1089/jwh.2012.3768.
Talukder, M.R.R., Rutherford, S., Huang, C., Phung, D., Islam, M.Z., Chu, C., 2017. Drinking water salinity and risk of hypertension: a systematic review and meta-analysis. Arch. Environ. Occup. Health 72, 126–138. https://doi.org/10.1080/19338244.2016.1175413.
Thavarajah, W., Verosloff, M.S., Jung, J.K., Alam, K.K., Miller, J.D., Jewett, M.C., Young, S.L., Lucks, J.B., 2020. A primer on emerging field-deployable synthetic biology tools for global water quality monitoring. NPJ Clean Water 3, 18. https://doi.org/10.1038/s41545-020-0064-8.
Troeger, C., Blacker, B.F., Khalil, I.A., et al., 2018. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of diarrhoea in 195 countries: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Infect. Dis. 18, 1211–1228. https://doi.org/10.1016/s1473-3099(18)30362-1.
Venkataramanan, V., Collins, S.M., Clark, K.A., Yeam, J., Nowakowski, V.G., Young, S.L., 2020a. Coping strategies for individual and household-level water insecurity: a systematic review. Wiley Interdiscip. Rev. Water 7. https://doi.org/10.1002/wat2.1477.

Venkataramanan, V., Geere, J.-A.L., Thomae, B., Stoler, J., Hunter, P.R., Young, S.L., 2020b. In pursuit of "safe" water: the burden of personal injury from water fetching in 21 low-income and middle-income countries. BMJ Glob. Health 5, e003328. https://doi.org/10.1136/bmjgh-2020-003328.

Vieux, F., Maillot, M., Rehm, C.D., Barrios, P., Drewnowski, A., 2020. Opposing consumption trends for sugar-sweetened beverages and plain drinking water: analyses of NHANES 2011–16 data. Front. Nutr. 7, 587123. https://doi.org/10.3389/fnut.2020.587123.

Vineis, P., Chan, Q., Khan, A., 2011. Climate change impacts on water salinity and health. J. Epidemiol. Glob. Heal 1, 5. https://doi.org/10.1016/j.jegh.2011.09.001.

Virtual Water Trade: Proceedings of the International Expert Meeting on Virtual Water Trade, 2003. Value of Water Research Report Series No. 12. The Netherlands.

Vonk, J., 2019. Sustainable Water and Sanitation in Zambia: Impact Evaluation of the "Urban WASH" Project, Effectiveness Review Series. Oxfam GB.

Webb, P., Iskandarani, M., 1998. Water Insecurity and the Poor: Issues and Research Needs, ZEF—Discussion Papers on Development Policy No. 2. Universität Bonn, Bonn, Germany.

WHO/UNICEF, 2017. Progress on Drinking Water, Sanitation and Hygiene: 2017 Update and SDG Baselines. World Health Organization (WHO) and the United Nations Children's Fund (UNICEF), Geneva.

Workman, C.L., Brewis, A., Wutich, A., Young, S., Stoler, J., Kearns, J., 2021. Understanding biopsychosocial health outcomes of syndemic water and food insecurity: applications for global health. Am. J. Trop. Med. Hyg. 104, 8–11. https://doi.org/10.4269/ajtmh.20-0513.

World Health Organization, 2005. Nutrients in Drinking Water. Water, Sanitation, and Health Protection and the Human Environment. World Health Organization, Geneva, Switzerland.

Wutich, A., Ragsdale, K., 2008. Water insecurity and emotional distress: coping with supply, access, and seasonal variability of water in a Bolivian squatter settlement. Soc. Sci. Med. 67, 2116–2125. https://doi.org/10.1016/j.socscimed.2008.09.042.

Wutich, A., Brewis, A., Tsai, A., 2020. Water and mental health. Wiley Interdiscip. Rev. Water 7. https://doi.org/10.1002/wat2.1461.

Yotebieng, M., Chalachala, J.L., Labbok, M., Behets, F., 2013. Infant feeding practices and determinants of poor breastfeeding behavior in Kinshasa, Democratic Republic of Congo: a descriptive study. Int. Breastfeed. J. 8, 11. https://doi.org/10.1186/1746-4358-8-11.

Young, S., Boateng, G.O., Jamaluddine, Z., Miller, J.D., Frongillo, E.A., Neilands, T.B., Collins, S.M., Wutich, A., Jepson, W.E., Stoler, J., 2019a. The Household Water InSecurity Experiences (HWISE) Scale: development and validation of a household water insecurity measure for low-income and middle-income countries. BMJ Glob. Health 4, e001750. https://doi.org/10.1136/bmjgh-2019-001750.

Young, S., Collins, S., Boateng, G., Neilands, T., Jamaluddine, Z., Miller, J., Brewis, A., Frongillo, E., Jepson, W., Melgar-Quiñonez, H., Schuster, R., Stoler, J., Wutich, A., Network, H.R.C., 2019b. Development and validation protocol for an instrument to measure household water insecurity across cultures and ecologies: the Household Water InSecurity Experiences (HWISE) Scale. BMJ Open 9. https://doi.org/10.1136/bmjopen-2018-023558.

Young, S.L., Frongillo, E.A., Jamaluddine, Z., Melgar-Quiñonez, H., Pérez-Escamilla, R., Ringler, C., Rosinger, A.Y., 2021a. Perspective: the importance of water security for ensuring food security, good nutrition, and well-being. Adv. Nutr. https://doi.org/10.1093/advances/nmab003.

Young, S.L., Bethancourt, H.J., Ritter, Z., Frongillo, E.A., 2021b. The individual water insecurity experiences (IWISE) scale: reliability, equivalence, and validity of an individual-level measure of water security. BMJ Glob. Health 6, e006460. https://doi.org/10.1136/bmjgh-2021-006460.

Young, S.L., 2021. The measurement of water access and use is key for more effective food and nutrition policy. Food Pol. https://doi.org/10.1016/j.foodpol.2021.102138.

Zhang, S., Sun, L., Zhang, J., Liu, S., Han, J., Liu, Y., 2020. Adverse impact of heavy metals on bone cells and bone metabolism dependently and independently through anemia. Adv. Sci. 7, 2000383. https://doi.org/10.1002/advs.202000383.

Zheng, G., Wu, S.-P., Hu, Y., Smith, D.E., Wiley, J.W., Hong, S., 2013. Corticosterone mediates stress-related increased intestinal permeability in a region-specific manner: corticosterone and colon tight junction protein. Neurogastroenterol. Motil. 25, e127–e139.

Zizza, C.A., Sebastian, R.S., Enns, C.W., Isik, Z., Goldman, J.D., Moshfegh, A.J., 2015. The contribution of beverages to intakes of energy and MyPlate components by current, former, and never smokers in the United States. J. Acad. Nutr. Diet. 115, 1939–1949. https://doi.org/10.1016/j.jand.2015.07.015.

Zolnikov, T.R., Blodgett-Salafia, E., 2016. Access to water provides economic relief through enhanced relationships in Kenya. J. Public Health fdw001. https://doi.org/10.1093/pubmed/fdw001.

Relevant website

hwise.org.

Weight management: Weight cycling

L. Lissner, University of Gothenburg, Göteborg, Sweden
B.L. Heitmann, Institute of Preventive Medicine, Frederiksberg, Denmark

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of L. Lissner, B.L. Heitmann, Weight Management: Weight Cycling, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 410–415, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00283-X.

Weight Cycling – A Health Risk? 717
The Metabolic Hypothesis 717
Weight Cycling and Mortality 719
Weight Cycling and Cardiovascular Disease 720
Other Health Outcomes: Cancer and Diabetes 721
Psychological Consequences 721
Methodological Issues 721
Conclusions 722
Further Reading 722

Is it better to have lost and re-gained than never to have lost at all? (Weight cycling refrain, loosely adapted from Alfred, Lord Tennyson)

Weight Cycling – A Health Risk?

The term weight cycling is used in the fields of nutrition and obesity research to refer to losses and subsequent regains of body weight typically, but not exclusively, occurring in association with dieting. Interest in this phenomenon was initially based on the observation that conventional weight loss programs are often unsuccessful in the long term. Dieting recidivism thus sets the stage for weight cycling, popularly referred to as yo-yo dieting, whereby dieters undergo multiple cycles of weight loss and regain in pursuit of their ideal body weights.

There seems to be little disagreement that weight cycling is one of the most difficult therapeutic aspects in the management of obesity. Most of the obese subjects seeking treatment have previously experienced cycles of weight loss and regain. Examples of weight cycling in two case studies are shown in **Figure 1**. These figures illustrate several phenomena that are important when considering weight cycling in humans: Weight losses and subsequent regains occur in conjunction with intentional weight loss dieting; however, weight changes vary in magnitude, and important fluctuations may be missed if weight measurements are taken at infrequent time intervals. Moreover, these two cases illustrate the common observation that intentional weight losses are frequently followed by regains in excess of the original body weight, and that true weight stability may be difficult to achieve.

Dieting to control body weight is not confined to overweight individuals, but has been widely reported even among men and women who have never been overweight. As described by Jeffery in 1984, 72.5% and 43.7% of surveyed women and men, respectively, had dieted to lose weight; even among women who had never been overweight, the majority reported having been on weight-loss diets. Although it is traditionally assumed that adherence to weight-reduction diets is beneficial to health, the high rates of dieting and weight regain, among the nonobese as well as the obese, have naturally created concern regarding potentially negative health consequences. However, weight loss can be intentional or unintentional, and the most recent generation of research on weight cycling has specifically focused on intentional weight loss as the risk factor of relevance to the weight-cycling debate, as reviewed by Simonsen.

In this article, some of the main points of this debate will be highlighted. The first epidemiological studies suggesting health implications of weight cycling were reviewed by Lissner and Brownell, when the majority of available observational evidence indicated adverse consequences. Subsequently, this topic became a source of considerable controversy, and a number of investigators continued to examine this issue, focusing on possible effects of weight cycling on metabolism, chronic disease, and mental health. This article will offer an overview of knowledge in this area, together with some methodological controversies surrounding existing research on weight cycling.

The Metabolic Hypothesis

Given the fact that most people who lose weight are unable to sustain their losses, a 'metabolic' hypothesis was formulated. It was proposed that if weight loss dieting caused permanent decreases in metabolic rate, the weight would be easily regained and every subsequent weight loss attempt would be more difficult. In 1994, the National Task Force on the Prevention and Treatment of Obesity in the US reviewed the evidence and reported an overall lack of support for the hypothesis that weight cycling promoted obesity, increased body fat, or had permanent effects on metabolism. This report also concluded that the majority of available data in animals did not independently link weight cycling to any parameter of energy balance (food intake, body composition, or energy

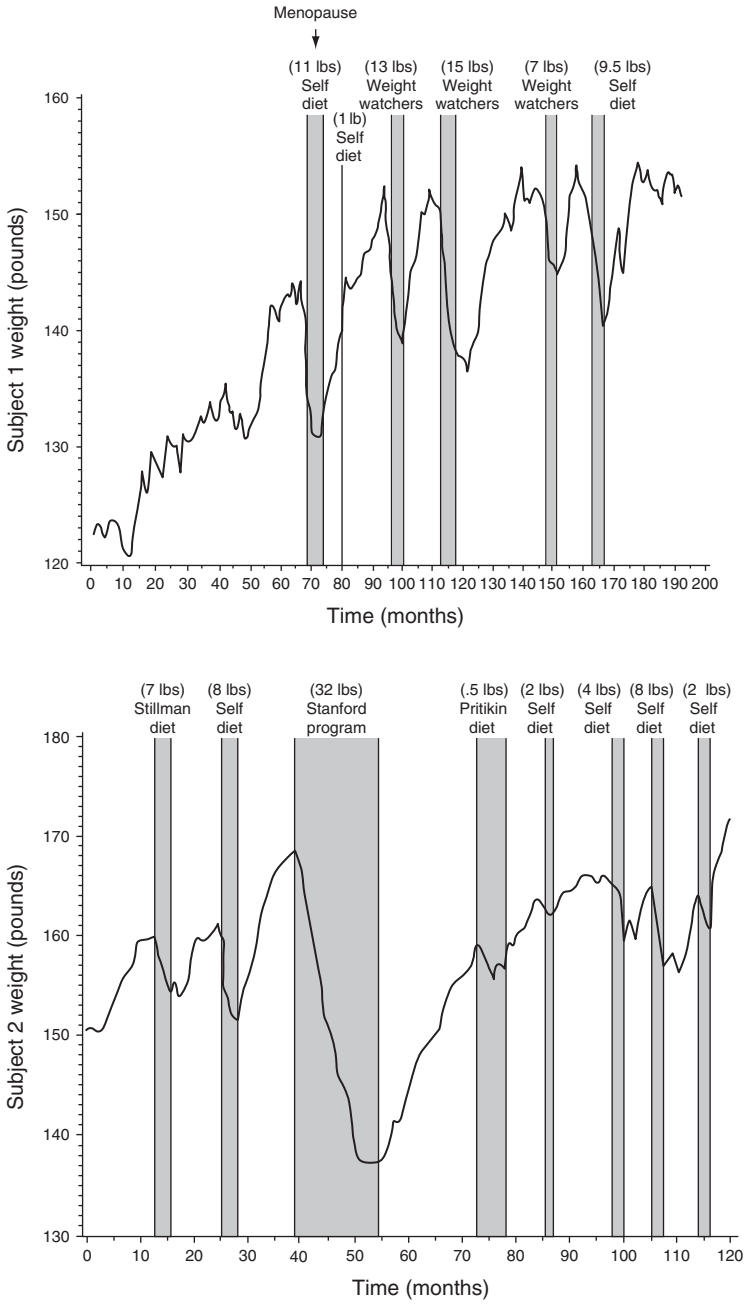


Figure 1 Monthly body weights (average of daily weights) of two female subjects, self-monitored over time. Reproduced from Black DR, Pack DJ, and Hovell MF (1991) A time-series analysis of longitudinal weight changes in two adult women. *International Journal of Obesity* 15: 623–633, with permission from Nature.

expenditure). This conclusion was supported by studies in humans, using a variety of designs, which failed to document irreversible effects of weight loss on metabolic rate, body composition, or adipose tissue distribution after regain. Also of interest in this context is the observation by Field in 2001 that weight-cycling women, in spite of their regains of lost weight, tended to gain less weight over time than their nonweight-cycling peers. In contrast, Van Wye reported that a history of weight cycling in healthy subjects did not seem to modify the risk of long-term weight gain in men and was associated with marginally more weight gain in women.

Weight Cycling and Mortality

Although most studies did not bear out the original idea that weight cycling alters metabolic rate, the possibility that weight fluctuation predicts chronic disease and death has been more difficult to discount. A number of prospective epidemiological studies have shown that an individual's own variations in body weight over time, a proxy for weight cycling, can statistically predict subsequent risk of mortality and disease. Positive associations have been reported between body weight fluctuation and all-cause mortality in several but not all such studies, as reviewed by the National Task Force on the Prevention and Treatment of Obesity in 1994. These findings are often expressed in terms of relative risk estimates, which represent the mortality rates in a weight-fluctuating group compared with the rates in a weight-stable group. The relative risk estimates for all-cause mortality have been reported to be as high as 2, indicating a double excess mortality risk in the weight-fluctuating individuals. Some investigators have reported that significant associations are restricted to certain types of individual, that is, nonobese or nonsmoking subgroups. For instance, data from the Multiple Risk Factors Intervention Trial (1993) concluded that any adverse effects of weight fluctuation were occurring in relatively normal-weight subjects.

By way of example, results are shown from a reanalysis from a longitudinal population study of Swedish women, started in 1968 when subjects were 38–60 years old. Women were weighed on three occasions over a 12-year observation period, based on which four subcategories of weight change could be created: stable, weight gain, weight loss, and weight cycle. Specifically, these categories were defined as: (1) women whose weights remained stable within plus or minus 3 kg; those who; (2) gained or; (3) lost at least 3 kg between the first and last observation; and (iv) those who had lost and then regained at least 3 kg, or gained then lost 3 kg, without an overall change of more than 3 kg from the first to final observation. When these groups were followed up for an additional 20 years, it was found that the weight-loss group and weight-cycling group both had approximately double risk of mortality, compared with the weight-stable women (**Figure 2**). It may be argued that the weight losses and subsequent gains may not be voluntary but rather reflect preexisting diseases. However, after exclusion of women with prevalent or incident cancer, diabetes, or cardiovascular disease during the entire period of weight observations, the excess mortality in the fluctuating group was not attenuated, suggesting that morbidity from these conditions was not the underlying cause of the fluctuations or reason for the association. These findings and similar observations in a number of other populations have not been adequately explained by biologically plausible mechanisms.

A systematic literature review focusing on intentional weight loss in healthy men and women, rather than weight cycling per se, identified nine studies that could address this issue. The review yielded no consistent evidence that intentional weight loss either increased or decreased mortality. Simonsen concluded that it is still not possible to make secure recommendations that intentional weight loss will increase longevity. This is in contrast to research on severely obese bariatric surgery patients showing improved

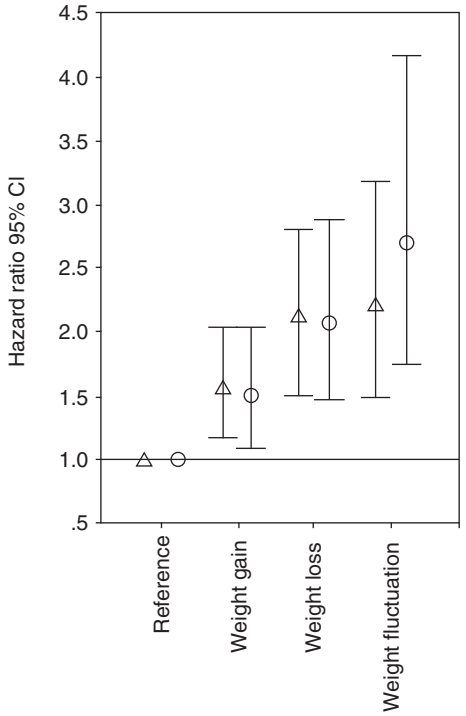


Figure 2 Relative risks of 20-year all-cause mortality, in relation to previous 12-year weight changes in the Prospective Population Study of Women in Göteborg. Triangular symbols refer to risks of total mortality, adjusted for age and final body mass index in 800 women. Circular symbols display results after excluding 99 prevalent cases of coronary heart disease, diabetes, or cancer during 12 years of weight observation.

longevity following elective weight-loss surgery, as reviewed in a recent meta-analysis by Pontiroli. Other factors may influence the associations between intentional weight loss and survival. The age at which weight fluctuations/cycles occurs may be of some relevance as weight loss among the younger adults may be less hazardous than in old age, when even apparently successful weight loss may in fact result from underlying disease. Similarly, weight gain early in adulthood may be more deleterious to overall health and mortality than weight gain later on. However, more research is needed to examine the influence of timing on the relation between weight variations and later disease and death.

Weight Cycling and Cardiovascular Disease

Most investigators have considered it more informative to focus the association between weight fluctuation and specific diseases, and causes of death, and have frequently observed positive associations between weight fluctuation and cardiovascular disease endpoints. However, as reviewed by Lissner and Brownell, the results were not always in agreement; data from the Framingham study showed excess cardiovascular disease among male and female weight-fluctuating individuals, whereas in men from the Baltimore Longitudinal Study on Aging there was no association between weight fluctuation and coronary heart disease. Two additional studies in male populations are illustrated in **Figure 3**. Both show a pattern of elevated risk of mortality from cardiovascular or coronary heart disease among weight-cycling men, and consistently lowest risk associated with stable body weight, based on which Jeffery in 1996 concluded that "stable weight over time is associated with best health. All patterns of weight change other than stable weight – gains, losses, and both combined – appear to be associated with increased mortality risk."

It has been hypothesized that some of the observed associations between weight fluctuation and cardiovascular disease might be explained by changes in cardiovascular risk factors occurring during weight gain that are not fully reversible with weight loss. This possibility has been explored using longitudinal data on body weight and risk factors that are concurrently measured on multiple occasions. A systematic review of these studies by the National Task force on the Prevention and Treatment of Obesity (1994) revealed no consistent associations between weight fluctuation and concomitant increases in traditional cardiovascular disease risk factors such as blood pressure and serum cholesterol.

In particular, hypertension has been examined as an endpoint in a number of studies with somewhat mixed results. Using data from the Nurses' Health Study II, Field and coworkers in 1999 reported that intentional weight cycling was not associated with significant excess risk of development of hypertension. In contrast, a retrospective study by Guagnano indicated that a positive history of weight cycling among obese women as well as the sum of weight regained increased the likelihood of being hypertensive. Interestingly, Stevens reported a possible beneficial effect of one weight cycle: In this blood pressure reduction trial, weight returned to baseline levels after 3 years, although blood pressure remained well below control levels. Control subjects in this study experienced a net weight gain, and one interpretation of this study is that a weight cycle did not predispose further hypertension, but rather, seemed to deter further weight gains at control levels, as also observed in the Nurses' Health Study II (2001). However, intentional weight loss may not have persisting beneficial effects on blood pressure, as shown by the Swedish Obesity Study (2000) where a large weight loss induced by gastric surgery and maintained over 8 years had no effect on the 8-year incidence of hypertension. In contrast, the beneficial effect on blood pressure was of a transitory nature, with dramatic initial reductions in both systolic and diastolic blood pressure during the initial 12 months of weight loss, and although an 18- to 30-kg weight loss was maintained over the subsequent 7 years, both the systolic and diastolic blood pressure of the surgically treated group increased. After 8 years

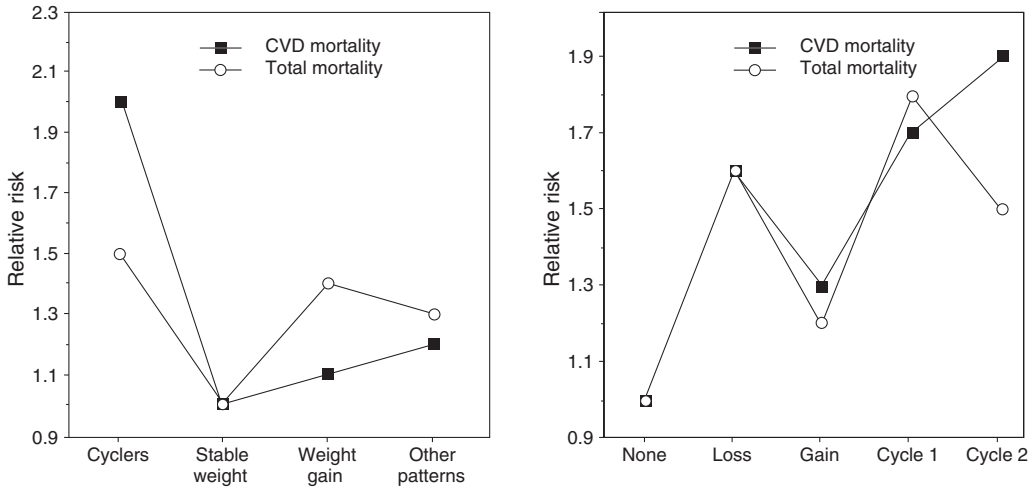


Figure 3 Total and cardiovascular disease mortality in male subjects with different weight change patterns. On the left, data from the Chicago Western Electric Study (Hamm, 1989) and on the right data from MRFIT (Blair, 1991).

there was no difference in systolic blood pressure between the surgically treated and the control groups, and diastolic blood pressure ended up higher in the surgical group than in the controls.

Other Health Outcomes: Cancer and Diabetes

Although several of the original weight-cycling studies also tested associations between weight cycling and cancer, cancer endpoints have typically not followed the same patterns as cardiovascular disease. It has been observed by Trentham-Dietz in 2000 that temporary weight cycling (weight loss followed by weight gain) was not associated with increased risk of postmenopausal breast cancer.

A number of studies have examined associations between weight cycling and diabetes, and yielded little evidence of a relation. In 1997, a study by Podar monitored glucose tolerance and weight fluctuations in obese patients, and reported no deterioration directly associated with weight cycling. In the Nurses' Health Study in 2004, no association was found between weight fluctuation and diabetes incidence. Interestingly, the Diabetes Prevention Program Research Group in 2002 found that the diabetes reduction achieved for more than 4 years with a lifestyle intervention was not diminished with the gradual regaining of more than half of the weight lost. This observation is an indication that a period of weight reduction may exert a net benefit for diabetes, even if weight is subsequently regained.

Psychological Consequences

It has often been assumed that the experience of dieting followed by involuntary regain of the lost weight must take a psychological toll, independent of any medical consequences of weight fluctuation. The possible psychological effects of weight cycling among obese persons were the topic of a literature review by Foster, who reported that weight cycling was not associated with depression, other psychopathology, or depressogenic cognitive styles. However, it was observed that weight cycling was associated with decreased perceptions of health and well-being, decreased eating self-efficacy, and weak increases in binge eating severity. Subsequently, Friedman concluded that an individual's perception of being a weight cycler may be more related to psychological problems than the actual number of pounds lost and regained over time. In 2000, the National Task Force on the Prevention and Treatment of Obesity concluded that concerns that dieting induces eating disorders or other psychological dysfunction in overweight and obese adults are generally not supported by empirical studies. A more recent prospective study of weight changes in relation to women's self-rated health found that that weight gain was associated with significant worsening of self-rated health but that weight loss did not result in an improvement in the same index. The observation that initially poor self-rated health actually predicted more weight gain suggests that causality is not simple, but that the association is more likely to be bidirectional.

Methodological Issues

Although a number of studies have shown that increased weight fluctuations are associated with subsequent occurrence of adverse health outcomes, a number of methodological problems make interpretation of these findings difficult. For example, weight gain, weight loss, and weight cycling are often considered separately, but it is almost impossible to determine their degree of overlap in observational studies. An individual who is observed to be systematically gaining weight at two points in time may experience a number of unmeasured fluctuations in the interim. However, as both loss and gain have been shown to increase risk of dying prematurely, it is possible that the increased mortality risk associated with weight cycling depends on the cumulated risk contributions from repeated weight loss and gain. The statistical complexities in defining weight cycling were reviewed in 1994 by the National Task Force on the Prevention and Treatment of Obesity.

A common observation surrounding this type of research is that different studies may be measuring quite different kinds of weight change – voluntary and involuntary. Involuntary changes may reflect serious underlying illness, depression, and other non-dieting phenomena. However, intentional dieting may also occur by a variety of dietary methods, some of which are more detrimental to health than others, as discussed by French in 1993. The issue of volition may shed light on the problem, but the specific impact of previous and current illness on epidemiological associations between weight cycling and longevity is still not fully understood. Other covarying factors besides intentionality of weight change and underlying illness may be producing artifactual associations in observational studies. These include aging, smoking and other lifestyle choices, degree and regional pattern of adiposity, and psychological factors. In epidemiological analyses, various types of statistical corrections can be made for potential confounding factors of this type, although adjustment may be incomplete. Finally, biological plausibility is always an issue to consider when reviewing any epidemiological evidence and has been a particular concern when considering the observations of excess risk in association with weight fluctuations.

Table 1 Summary of hypothesized weight cycling effects and selected sources of evidence

Hypothesized adverse health consequences of weight cycling	*Comments (with suggested reading)*
Psychological consequences	Not supported by evidence (reviewed by the US National Task Force, 2000)
Metabolic rate and body composition	Not supported by evidence (reviewed by the US National Task Force, 1994)
All-cause mortality	Supported by most prospective studies, for example, Hamm, 1989; Lissner, 1991; Blair, 1993 (reviewed in Obesity 1992, and by the US National Task Force, 1994)
Cardiovascular disease	Supported by most prospective studies, for example, Hamm, 1989; Lissner, 1991; Blair, 1993 (reviewed in Obesity 1992 and the US National Task Force, 1994)
Diabetes	Not supported by epidemiological studies (Field, 2004). Mixed evidence from clinical setting (Guagnano, 2000)
Cancer	Not supported by epidemiological studies (Lissner, 1991; Trentham-Dietz, 2000)

Conclusions

As summarized in Table 1, many – but not all – epidemiological studies indicate that men and women undergoing body weight fluctuations are at higher risk of mortality and/or cardiovascular disease than individuals experiencing less fluctuation, but the lack of biologically plausible mechanisms to explain these associations has limited the conclusions that can be drawn. Moreover, the evidence for effects on psychological and other health endpoints is inconclusive. Uncontrolled confounding from disease states resulting in a loss–gain or gain–loss pattern must be considered a plausible explanation for some of these findings, underscoring the difficulty of using observational data to study weight cycling. Some studies are suggesting that even limited periods of weight reduction may be beneficial in the long run, whereas others are not. Experimental data and intervention studies are required for confirmation of the weight-cycling hypothesis. The published observational studies of subjects whose weight changes are known to be caused by dieting have been important contributions to the critical discussion of the weight-cycling phenomenon, but additional studies are needed in which weight fluctuations are assessed in a more controlled manner.

Regarding the hypothesis that dieting exacerbates the problem of obesity and weight gain, most studies have failed to demonstrate that weight fluctuation per se depresses metabolic rate. Available research on the effects of weight cycling on both metabolism and disease thus provides little basis to discourage overweight patients from losing weight. Nevertheless, one of the conclusions of the 1994 report from the National Task Force on the Prevention and Treatment of Obesity was that individuals who are not obese and who have no risk factors for obesity-related illness should not undertake weight loss efforts. The only uncontroversial message of the weight-cycling research is that overweight individuals need to be counseled in skills to maintain weight loss, and that relapse prevention should be a more central focus of weight loss programs.

In conclusion, although this research has drawn attention to the necessity of developing improved behavioral and nutritional strategies for sustaining weight reductions and thus preventing weight cycling, the evidence relating weight cycling to adverse health outcomes must be considered equivocal. Regarding the hypothesis that weight cycling exacerbates weight problems, a 1995 opinion survey of obesity researchers conducted by Bray concluded that weight cycling was not considered a very important cause of obesity, and little convincing evidence has emerged in subsequent years to change that conclusion. With regard to the question in the 'weight cycling refrain' regarding possible consequences of weight cycling, the current knowledge would suggest that it is probably not worse to have lost and regained than never to have lost at all. However, there remain some curious and persistent results in the experimental as well as epidemiological literature suggesting that we do not completely understand the phenomenon.

Further Reading

Black, D.R., Pack, D.J., Hovell, M.F., 1991. A time-series analysis of longitudinal weight changes in two adult women. International Journal of Obesity 15, 623–633.
Blair, S., Shaten, J., Brownell, K., Collins, G., Lissner, L., 1993. Body weight change, all cause, and cause-specific mortality in the Multiple Risk Factor Intervention Trial. Annals of Internal Medicine 119, 749–757.
Bray, G.A., DeLany, J., 1995. Opinions of obesity experts on the causes and treatment of obesity –a new survey. Obesity Research 3, 419S–423S.
Diabetes Prevention Program Research Group, 2002. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. The New England Journal of Medicine 346, 393–403.
Field, A.E., Byers, T., Hunter, D.J., et al., 1999. Weight cycling, weight gain, and risk of hypertension in women. American Journal of Epidemiology 150, 573–579.
Field, A.E., Manson, J.E., Laird, N., et al., 2004. Weight cycling and risk of developing type 2 diabetes among adult women in the United States. Obesity Research 12, 267–274.
Field, A.E., Wing, R.R., Manson, J.E., Spiegelman, D.L., Willett, W.C., 2001. Relationship of a large weight loss to long-term weight change among young and middle-aged US women. International Journal of Obesity 25, 1113–1121.
Foster, G.D., Sarwer, D.B., Wadden, T.A., 1997. Psychological effects of weight cycling in obese persons: A review and research agenda. Obesity Research 5, 474–488.

French, S.A., Jeffery, R.W., Folsom, A.R., Williamson, D.F., Byers, T., 1995. History of intentional and unintentional weight loss in a population-based sample of women aged 55–69 years. Obesity Research 3, 163–170.

Friedman, M.A., Schwartz, M.B., Brownell, K., 1998. Differential relation of psychological functioning with the history and experience of weight cycling. Journal of Consulting and Clinical Psychology 66, 646–650.

Guagnano, M.T., Ballone, E., Pace-Palitti, V., et al., 2000. Risk factors for hypertension in obese women. The role of weight cycling. European Journal of Clinical Nutrition 54, 356–360.

Hamm, P.B., Shekelle, R.B., Stamler, J., 1989. Large fluctuations in body weight during young adulthood and 25-year risk of coronary death in men. American Journal of Epidemiology 129, 312–318.

Jeffery, R.W., 1996. Does weight cycling present a health risk? The American Journal of Clinical Nutrition 63 (supplement), 452S–455S.

Jeffery, R.W., Folsom, A.R., Luepker, R.V., et al., 1984. Prevalence of overweight and weight loss behavior in a metropolitan adult population: The Minnesota Heart Survey experience. The American Journal of Public Health 74, 349–352.

Lissner, L., Brownell, K., 1992. Weight cycling, mortality, and cardiovascular disease: A review of epidemiologic findings. In: Brodoff, B., Björntorp, P. (Eds.), Obesity. JB Lippincott, Philadelphia ch. 56, pp. 653–661.

Lissner, L., Odell, P., D'Agostino, R., et al., 1991. Variability of body weight and health outcomes in the Framingham population. The New England Journal of Medicine 324, 1839–1844.

National Task Force on the Prevention and Treatment of obesity, 1994. Weight Cycling. Journal of the American Medical Association 272, 1196–1202.

National Task Force on the Prevention and Treatment of Obesity, 2000. Dieting and the developing of eating disorders in overweight and obese adults. Archives of Internal Medicine 160, 2581–2589.

Podar, T., Solntsev, A., Vali, M., Vinogradova, T., Podar, I., 1997. No deterioration of glucose tolerance in weight cycling obese. Indian Journal of Ophthalmology 20, 921–924.

Pontiroli, A.E., Morabito, A., 2011. Long-term prevention of mortality in morbid obesity through bariatric survery. A systematic review and meta-analysis of trials performed with gastric banding and gastric bypass. Annals of Surgery 3, 484–487.

Simonsen, M.K., Hundrup, Y.A., Groenbaek, M., Heitmann, B.L., 2008. A prospective study of the association between weight changes and self-rated health. BMC Women's Health 8, 13–20.

Simonsen, M.K., Hundrup, Y.A., Obesl, E.B., Gronbaek, M., Heitmann, B., 2008. Intentional weight loss and mortality among initially healthy men and women. Nutrition Reviews 66, 375–386.

Sjöström, C.D., Peltonen, M., Wedel, H., Sjöström, L., 2000. Differentiated long-term effects of intentional weight loss on diabetes and hypertension. Hypertension 36, 20–25.

Stevens, V.J., Obarzanek, E., Cook, N.R., et al., 2001. Long-term weight loss and changes in blood pressure: Results of the trials of hypertension prevention, phase II. Annals of Internal Medicine 134, 1–11.

Trentham-Dietz, A., Newcomb, P.A., Egan, K.M., et al., 2000. Weight change and risk of postmenopausal breast cancer (United States). Cancer Causes Control 11, 533–542.

Van Wye, G., Dubin, J.A., Blair, S.N., DiPietro, L., 2007. Weight cycling and 6-year weight change in healthy adults: The Aerobics Center Longitudinal Study. Obesity 3, 731–739.

Weight management: Weight maintenance

Lauren A. Griffiths, Anna M. Jackson, Elizabeth T. Anderson Steeves, and Hollie A. Raynor, Department of Nutrition, University of Tennessee Knoxville, Knoxville, TN, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of H.A. Raynor, E.A. Steeves, Weight Management: Weight Maintenance, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 416–421, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00284-1.

Introduction 724
Definition of successful weight loss maintenance 725
Data on prevalence of long-term weight loss maintenance 725
Why is weight loss maintenance difficult? 725
Observational studies of weight loss maintenance 725
The National Weight Control Registry 726
Psychological factors associated with weight loss maintenance 726
Experimental studies examining weight loss maintenance 727
Energy balance 727
Diet 727
Structured low-calorie diets 727
Very low-calorie diets 730
Diet composition 730
Intermittent fasting 730
Conclusions on dietary strategies and weight loss maintenance 731
Physical activity 731
Strategies for improving maintenance of physical activity 731
Behavioral component 731
Obesity as a chronic disease 731
Extending professional contact 732
Social support 732
Self-monitoring 732
Technology 733
Incentives 733
Systems-level programs 733
Conclusion 734
References 734

Key points

- Successful weight loss maintenance is generally defined as an intentional loss of 5–10% of initial body weight maintained for at least 1 year.
- The complex interaction of physiological, environmental, and psychological factors associated with weight status make long-term maintenance of weight loss challenging.
- In general, successful weight loss maintainers consume a low-calorie diet, eat breakfast regularly, self-monitor weight frequently, and maintain consistent eating patterns across time. Successful maintainers also engage in substantial levels of physical activity (approximately 1 h per day).
- No optimal diet has been identified for weight loss maintenance. The best dietary strategy is one that provides fewer calories than the individual consumed prior to weight loss and that the individual can adhere to long-term.
- Obesity and weight management should be treated using a chronic care approach that emphasizes ongoing behavioral support over time, such as extended professional contact and social support.

Introduction

Nearly 3 out of 4 adults in the United States have overweight or obesity (Fryar et al., 2020). Obesity is a risk factor for numerous serious diseases, including type 2 diabetes, hypertension, and some cancers, and the condition constitutes a significant economic burden in developed countries (Bianchini et al., 2002; Tremmel et al., 2017; Paeratakul et al., 2002). Thus, understanding how

individuals can achieve and maintain weight loss is an important public health issue. The goal of this article is to review current research related to successful weight loss maintenance. A definition of weight loss maintenance is provided and is discussed within the context of the field's evolving perspective of obesity as a chronic disease. Next, current explanations for why weight loss maintenance is difficult are described. Finally, factors important for successful weight loss maintenance are identified from observational and experimental research examining the behavioral, psychological, and environmental determinants of maintenance. The article closes with recommendations for achieving weight loss maintenance, as well as directions for future research.

Definition of successful weight loss maintenance

There is currently no standard definition for weight loss maintenance. In general, an intentional weight loss of 5–10% of initial body weight that is sustained for at least 1 year is considered successful weight loss maintenance (Rössner, 1997; Wing and Hill, 2001). This magnitude of weight loss is supported by the 2013 AHA/ACC/TOS Guideline for the Management of Overweight and Obesity in Adults, which recommends 5–10% weight loss for clinically meaningful improvements in cardiometabolic health (Jensen et al., 2014). However, losses as low as 3% may also result in modest improvements in cardiovascular risk factors, with greater weight loss resulting in greater benefits (Jensen et al., 2014).

Data on prevalence of long-term weight loss maintenance

Obesity treatment studies with long-term follow up provide important insight into the maintenance of weight loss over time. A systematic review of different weight loss interventions (e.g., lifestyle, meal replacements, pharmaceutical) with at least 1-year follow up found that initial weight losses ranged from 4.8% to 8.0%, with sustained weight losses of 3.0%–4.3% maintained up to 4 years later (Franz et al., 2007). In general, weight loss tended to plateau around 6 months, at which time partial weight regain occurred. In behavioral lifestyle interventions (which include diet, physical activity, and behavior modification components), 35%–60% of participants maintain a loss of at least 5% of initial body weight for 2 or more years (Jensen et al., 2014).

Weight control registries in the United States and Europe, which enroll adults who are successful weight loss maintainers, also provide information on the prevalence of weight loss maintenance. While each registry has unique enrollment criteria (an initial loss of 5 kg, 13.6 kg, or 10% of body weight maintained for at least 1 year or 2 years), successful maintenance (defined as maintaining at least 3%–10% weight loss, depending on the registry) was observed in 60.5%–88.0% of participants at follow-up 1–10 years later. It is unclear whether or in which type of intervention registry members participated for either initial weight loss or the weight loss maintenance period (Paixão et al., 2020).

Why is weight loss maintenance difficult?

A complex interaction of physiological, environmental, and psychological determinants contributes to the challenges associated with weight loss maintenance (**Fig. 1**) (Huang et al., 2009; Varkevisser et al., 2019). Researchers speculate that weight loss, in response to dieting, produces physiological (e.g., decrease in energy expenditure) and psychological changes that may result in consequent weight regain (MacLean et al., 2015). The interplay of environmental cues, such as exposure to unhealthy foods or lack of sidewalks, may affect energy balance through encouraging increased energy intake and decreased energy expenditure (Li et al., 2009; McCormack and Shiell, 2011). The psychological processes (e.g., motivation) that are needed to supersede unhealthy behavior cues in an obesogenic environment may also be difficult to maintain over a long period of time (Metzgar et al., 2015). Additionally, greater weight loss earlier in treatment predicts better outcomes at the end of treatment, and the weight loss attributed to diet and physical activity changes may provide positive reinforcement that supports treatment adherence (Handjieva-Darlenska et al., 2012). During the maintenance period, however, there is no continued weight loss (and often at least some weight regain), which may reduce reinforcement and decrease the behaviors associated with healthy eating and regular physical activity. Finally, accountability and social support may also wane over time, making weight loss maintenance more difficult (Metzgar et al., 2015).

Observational studies of weight loss maintenance

Observational research uses non-experimental study designs to identify factors associated with weight loss maintenance. In observational research, the variable of interest is observed as it naturally occurs and is not manipulated by the researchers. These study designs provide important information about the relationships between behavioral, psychological, and environmental factors and maintenance of weight loss over time. However, since researchers do not have control over the variable of interest, the ability to determine causality is limited.

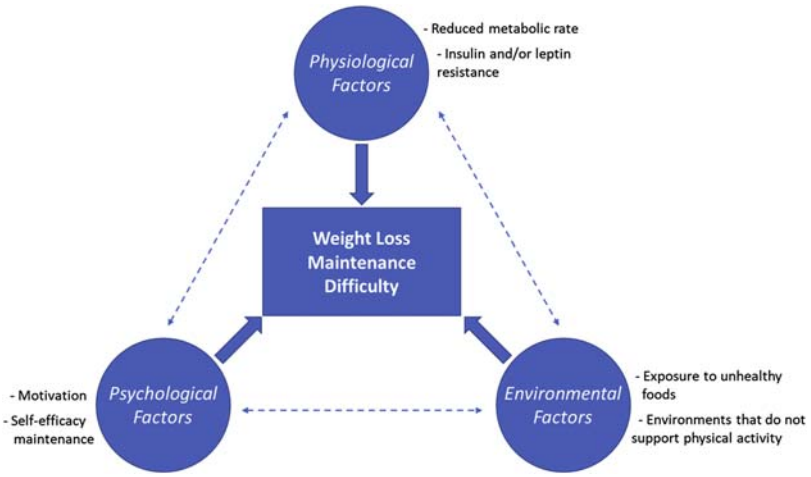


Fig. 1 Conceptual model of the complex interaction of physiological, psychological, and environmental factors that influence weight loss maintenance.

The National Weight Control Registry

The National Weight Control Registry (NWCR) is the largest prospective observational study investigating weight loss maintenance to date (The National Weight Control Registry). NWCR was established in 1994 and now tracks more than 10,000 individuals identified as successful weight loss maintainers. To be eligible for enrollment in NWCR, adults aged 18 years or older must have a previous weight loss of at least 13.6 kg (30 pounds) that has been maintained for at least 1 year. The participants in NWCR are predominantly women (77.5%) and white (95.7%).

A 10-year follow up study was conducted among NWCR members (Thomas et al., 2014). With an average weight loss of 31.3 kg at enrollment, participants maintained an average weight loss of 23.8 kg at 5 years and 23.1 kg at 10 years. Additionally, participants with greater initial weight loss at enrollment maintained substantially greater weight loss at 10 years, suggesting that the magnitude of initial weight loss is important in for successful weight loss maintenance over the long term.

Despite heterogenous approaches to weight loss maintenance reported by NWCR members, certain behaviors have been consistently identified among successful maintainers. These include consuming a low-calorie diet, eating breakfast regularly, self-monitoring weight frequently, and engaging in consistent eating patterns across time (Wing and Phelan, 2005). In a 2014 study, registry members reported consuming around 1410 kilocalories per day, with 29% of daily calories coming from fat, 19% from protein, and 50% from carbohydrates (Catenacci et al., 2014). This marks a shift in macronutrient composition from two decades earlier, when registry members reported percent of daily calories from fat and carbohydrates to be around 24% and 56%, respectively (protein intake stayed relatively constant) (Klem et al., 1997). This shift toward fat and away from carbohydrates is consistent with national trends in the United States and may reflect dieting trends that emphasize lower-carbohydrate recommendations (e.g., keto diet) (Klem et al., 1997; Shan et al., 2019). Additionally, more than three-quarters of registry participants report eating breakfast daily, and self-reported dining at restaurants and fast-food establishments is limited (Wing and Phelan, 2005). More than half (59%) of NWCR members report maintaining a consistent eating pattern across weekdays and weekends, as well as during holidays and vacations (Wing and Phelan, 2005). Additionally, self-monitoring of weight is an important determinant of successful weight loss maintenance. Nearly half (44%) of NWCR members report weighing themselves every day, and nearly a third (31%) more weigh themselves at least once per week (Wing and Phelan, 2005). This frequent self-monitoring of weight may contribute to increased awareness of weight fluctuations, allowing individuals to self-evaluate how diet and activity behaviors are affecting their weight and to adjust as needed.

Physical activity is also an important part of long-term weight loss maintenance (Catenacci et al., 2008). The vast majority of NWCR members (85%) report using physical activity as a strategy for maintaining their weight loss. NWCR members report an average energy expenditure of about 2600 kilocalories per week from physical activity, which equates to approximately 1 h of moderate-intensity physical activity per day. Walking is the most reported leisure time physical activity, with cycling and cardiovascular exercise machines also being popular activities among both women and men (Catenacci et al., 2008).

Psychological factors associated with weight loss maintenance

Observational studies have also examined psychological factors related to successful weight loss maintenance. NWCR members with higher levels of depression, binge eating, and disinhibition (i.e., loss of control while eating) experienced greater weight

regain at 1-year follow up (McGuire et al., 1999). Research in other populations suggests that individuals who feel confident and in control of their weight-related behaviors are more likely to be successful maintainers. Having higher self-efficacy (i.e., a person's confidence in his or her ability to perform behaviors related to weight management) and an internal locus of control (i.e., believing that success or failure is related to personal attributes rather than external factors outside of one's own control) have been associated with better weight loss maintenance, as have higher levels of autonomous motivation (Teixeira et al., 2015; Anastasiou et al., 2015). While behavioral lifestyle interventions focus on building skills related to self-regulation and self-efficacy, more research is needed investigating how weight interventions can address other psychological determinants of weight loss maintenance.

Experimental studies examining weight loss maintenance

In experimental research, variables thought to be related to weight loss maintenance are manipulated by the researchers, and weight is measured to examine change over time. Randomized controlled trials (RCTs), a type of experimental design, use rigorous methodology to investigate treatment factors associated with weight loss maintenance (Miller et al., 2020; Fisher, 1937). The process of randomizing study participants to treatment groups reduces the influence of both measured and unmeasured confounding variables, resulting in improved ability to determine causality (Miller et al., 2020; Fisher, 1937). Experimental studies in the weight loss maintenance literature typically investigate long-term weight loss (e.g., weight change from baseline to study completion) rather than the maintenance of weight loss from the end of the initial treatment (usually 6 months) to study completion (Look AHEAD Research Group, 2014). While recent long-term weight loss trials have increased in sample size, these interventions are still often limited in study duration (usually not more than 1–2 years in length). **Table 1** outlines three major experimental trials examining behavioral lifestyle interventions for successful maintenance of weight loss.

Energy balance

Interventions for the treatment of obesity encompass a broad spectrum of dietary and physical activity strategies. For any diet to cause weight loss, it must produce an energy deficit in which energy consumed is less than energy expended through lifestyle and activity. To successfully maintain weight loss, an individual must sustain energy balance (energy intake equals energy expenditure) at a lower energy intake than prior to initial weight loss. Early dietary research in this area focused primarily on caloric restriction and dietary structure (Cordain et al., 2005). However, the scope of recent research has expanded to include regulation of metabolic pathways through adaptations of macronutrient composition (Freire, 2020; Cordain et al., 2005). Research has also shown that behavioral lifestyle interventions, which combine diet, physical activity, and behavioral strategies, are the most effective non-surgical treatment for long-term maintenance of weight loss.

Diet

Behavioral lifestyle interventions have traditionally recommended a balanced low-fat, low-calorie diet for weight loss (Baker, 2006; Expert Panel on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults, 1998). This diet includes goals to consume 1000–1500 kilocalories per day, depending on initial body weight and to reduce dietary fat intake to 20–30% of total calories. The MyPlate method, developed by the United States Department of Agriculture, offers dietary recommendations that prioritize portion control (Chang and Koegel, 2017). MyPlate encourages consumption of all macronutrients and has no strict guidelines related to food group restriction or inclusion. In general, the method emphasizes consumption of whole grains, fruits and vegetables, low-fat dairy, and lean proteins and recommends limiting added sugars, saturated fat, and sodium (Chang and Koegel, 2017).

Structured low-calorie diets

Dietary structure refers to the amount of control imposed on the types and/or amount of food consumed. A structured dietary regimen can be achieved through a variety of strategies, including limited food choice variety, structured meal plans, meal replacements, and the use of comprehensive grocery lists (Ditschuneit and Flechtner-Mors, 2001). In general, weight loss outcomes tend to improve with increasing dietary structure (Raynor and Champagne, 2016). One RCT investigated the effect of providing portion-controlled lunch and dinner entrées compared to self-selection of foods during participation in a behavioral lifestyle intervention. The provision of portion-controlled entrées promoted greater weight loss and fat mass loss at 12 weeks compared to control [8.6 kg (3.9%), 7.8 kg (5.1%), and 6.0 kg (4.4%), respectively] (Rock et al., 2016).

Meal replacements (MRs) can be characterized as portion-controlled food products or beverages that replace complementary foods. The strategy can be a cost-effective approach for weight loss and weight loss maintenance, including in populations with economic limitations (Huerta et al., 2004). MRs are commonly used to reduce total caloric intake to simultaneously promote

Table 1 Major experimental studies examining weight loss maintenance.

Trial name (author, year)	*Sample characteristics (sample size, sex, race/ethnicity, mean BMI, mean age)*	*Study design, duration, assessment timepoints*	*Intervention components*					
			Study phases & conditions	*Diet*	*Physical activity*	*Behavior*	*Frequency of contact*	*Key findings*
Look AHEAD (action for health in diabetes) trial (Look AHEAD Research Group, 2014)	N = 5145 69% female 16% African American 13% hispanic 63% non-hispanic white 7% other 36.0 kg m^{-2} 58.7 yrs	RCT 8 yrs Assessments: 0, 1, 4, 8 yrs	*Phase 1 (Yr 1):* Wt loss *Phase 2 (Yrs 2+):* Wt maintenance **Experimental:** Intensive Lifestyle Intervention (ILI) **Active comparator:** Diabetes Support and Education (DSE)	**ILI:** *Phase 1:* *Mos 1–6:* 1200–1800 kcal d^{-1} (depending on body wt); ≤30% kcal from fat (≤10% from saturated fat), ≥15% kcal from protein; structured meal plans; MRs (replace 2 meals and 1 snack per d) *Mos 7–12:* 500 kcal d^{-1} energy deficit; MRs (replace 1 meal or snack per d); eat a low energy-dense diet[a] *Phase 2:* MRs (replace 1 meal or snack per d) **DSE:** basic nutrition education for type 2 diabetes	**ILI:** *Phase 1:* *Mos 1–6:* Incrementally increasing goal to ≥175 min d^{-1} of moderate-intensity PA by wk 26 *Mos 6–12:* ≥ 175–200 min d^{-1} of moderate-intensity PA; 10,000 steps d^{-1} *Phase 2:* ≥175 min d^{-1} of moderate-intensity PA; 10,000 steps d^{-1} **DSE:** basic PA education for type 2 diabetes	**ILI:** Self-monitoring, problem solving, motivational interviewing, goal setting, coping skills, relapse prevention, cognitive restructuring **DSE:** none	**ILI:** *Phase 1:* *Mos 1–6:* 60–75 min group mtgs for first 3 wks of mo plus 20–30 min individual mtg every fourth wk *Mos 7–12:* 2 group mtgs per mo plus 1 individual mtg per mo *Phase 2:* *Yrs 2-4: 1,* 20–30 min individual mtg per mo plus telephone or email f/u 2 wks later *Yrs 5+:* 1 individual mtg per mo **DSE:** *Yrs 1-4:* 3, 1 h group mtgs per yr *Yrs 5-8:* 1 group mtg per yr	ILI lost a greater % of initial wt than DSE at all f/u timepoints: *Yr 1:* 8.5% vs −0.6% *Yr 4:* 4.4% vs −0.7% *Yr 8:* 4.7% vs −2.1% ILI reported greater kcal expenditure wk^{-1} from PA at all f/u timepoints, including yr 8 (1040 kcal wk^{-1} vs. 853 kcal wk^{-1}). ILI was more likely than DSE to weigh themselves daily or weekly, and participants who weighed themselves more frequently were more likely to maintain ≥10% wt loss at yr 8.

Weight loss maintenance trial (Svetkey et al., 2008)	N = 1032 63% female 38% African American 62% non-African American Ethnicity not reported 34.1 kg m^{-2} 55.6 yrs	RCT 30 mos Assessments: 0, 6, 12, 18, 24, 30 mos	*Phase 1 (Mos 1–6):* Wt loss *Phase 2 (Mos 7–36):* Wt maintenance; participants losing ≥4 kg randomized to one of the following groups: **Experimental:** Personal Contact (PC) Interactive Technology (IT) **Active comparator:** Self-Directed (SD)	*Phase 1:* Reduce kcal intake; adopt DASH dietary pattern[b] *Phase 2:* Individualized calorie goals; DASH dietary pattern	*Phase 1:* ≥180 min wk^{-1} of moderate-intensity PA *Phase 2:* ≥225 min wk^{-1} of moderate-intensity PA	*Phase 1:* Social support, goal setting, decision making, problem solving, self-monitoring, contingency planning *Phase 2:* **PC & IT:** self-monitoring, feedback, problem-solving, relapse prevention, social support, motivational interviewing **SD:** none	*Phase 1:* Weekly 1.5 h grp mtgs *Phase 2:* **PC:** 1 mtg per mo (45 min face-to-face mtg every 4 mos, telephone contacts in between) **IT:** 45 min orientation mtg at randomization plus 1 "re-orientation" mtg at 12 mos; unlimited access to intervention website; automated telephone calls and emails **SD:** minimal contact (only at f/u visits)	All groups maintained significant wt loss from study entry to 30 mos: *SD:* 2.9 kg *IT:* 3.3 kg *PC:* 4.2 kg PC regained less wt than SD and IT at 24 mos (group diff of −2.0 and −1.1 kg) and 30 mos (group diff of −1.5 and −1.2 kg). Wt loss was greater in IT than SD for all f/u assessments (6, 12, 18, 24 mos) except 30 mos.
Study to prevent regain (STOP regain) trial (Wing et al., 2006)	N = 314 81.2% female Race/ethnicity not reported 28.6 kg m^{-2} 51.3 yrs	RCT 18 mos Assessments: 0, 6, 12, 18 mos	No phases. To be eligible, individuals were required to have lost ≥10% of body wt in prior 2 yrs. **Experimental:** Face-to-Face (F2F) Internet-Based (IB) **Comparator:** no-intervention control	**F2F & IB:** for participants with wt regain of ≤2.2 kg, no specific strategies recommended; for participants with wt regain ≥2.3 kg, either initial wt loss approach or standard low-kcal/low-fat diet with MRs recommended	**F2F & IB:** ≥60 min d^{-1} of PA **Control:** none	**F2F & IB:** self-monitoring, positive reinforcement **Control:** none	**F2F & IB:** weekly group mtgs for 1 mo, then monthly mtgs for remaining 18 mos (in-person for F2F and online via chat room for IB); additional individual counseling for participants with ≥2.3 kg wt regain (telephone for F2F and email for IB) **Control:** none	F2F, but not IB, had less wt regain at 18 mos compared to control (2.5 kg in F2F, 4.7 kg in IB, and 4.9 kg in control). Fewer participants in F2F and IB regained ≥2.3 kg compared to control at 18 mos (45.7% in F2F, 54.8% in IB, and 72.4% in control).

Abbreviations: BMI = body mass index; d = day; DASH = dietary approaches to stop hypertension; diff = differences; f/u = follow up; hr = hour; kcal = kilocalorie; kg = kilogram; m = meter; min = minute; mo = month; MR = meal replacement; mtg = meeting; PA = physical activity; RCT = randomized controlled trial; wk = week; wt = weight; yr = year.

[a]Participants were taught an approach to eating a low energy-dense diet following the book *Volumetrics* (Rolls and Barnett, 2000).

[b]The DASH dietary pattern encourages increased consumption of fruits, vegetables, low-fat dairy and reduced consumption of saturated and total fats.

weight loss over time. MRs used in weight loss interventions have been considered a helpful strategy as delivery can occur within a community or via an over-the-counter purchase without external support from healthcare systems. A recent systematic review and meta-analysis of MRs found that MR use resulted in greater weight loss over one year compared to alternative interventions using conventional foods only (Astbury et al., 2019).

Very low-calorie diets

Very low-calorie diets (VLCDs) were designed to produce rapid weight loss while minimizing loss of lean muscle mass (Expert Panel on the Identification, 1998; Saris, 2001). In VLCDs, calorie intake is restricted to 450–800 kcal per day, and a protein-enriched diet (often in the form of liquid formulation or specially formulated bars) is provided to maintain lean body mass. The diet is usually prescribed for no more than 12 weeks, and patients receive intensive medical supervision during this period to monitor for health complications or adherence issues (Expert Panel on the Identification, 1998; Saris, 2001). VLCDs can induce large initial weight losses of 14.2–21.0 kg over 11–14 weeks (Jensen et al., 2014). Despite this large initial weight loss, a period of rapid weight regain is typically observed after stopping the VLCD, and VLCDs do not necessarily result in better long-term weight loss outcomes compared to LCDs with less severe calorie restriction (Franz et al., 2007; Jensen et al., 2014).

Diet composition

A large body of research has investigated how dietary composition affects weight loss and weight loss maintenance. Different macronutrients may influence perceptions of satiety, hunger, and fullness differently and may, therefore, also affect energy intake (Hill and Blundell, 1986). Despite these differing qualities, a large behavioral lifestyle intervention trial examining four diets with distinct macronutrient compositions (i.e., differing percent of energy from fats, carbohydrates, and proteins) found no differences among the diets in either initial magnitude of weight loss or weight loss maintenance two years later (Sacks et al., 2009). However, a body of research suggests that consuming a diet high in protein may have a significant beneficial effect on the prevention of weight regain (van Baak and Mariman, 2019).

Another potentially useful dietary strategy for weight loss maintenance is consumption of a diet high in low-energy dense foods. Energy density is a measure of kilocalories per gram of food. Fewer calories are consumed when eating the same volume of a lower energy density food as compared a higher energy density food (Rolls, 2009). Reducing the energy density of the diet has been shown to produce weight loss, even in the absence of explicit goals for calorie restriction (Ello-Martin et al., 2007). One year-long intervention trial conducted in women with obesity found that the group with goals to reduce fat and consume water-rich foods ate a diet with lower energy density, were less hungry, and had greater weight loss (7.9 vs. 6.4 kg lost, respectively) than the group with goals to reduce fat intake only (Ello-Martin et al., 2007). Furthermore, a secondary analysis of two separate intervention trials found that successful weight loss maintainers reported consuming a diet with lower energy density than participants who were both overweight and healthy weight (Raynor et al., 2011).

At this time, little experimental research has looked at the effectiveness of food group-related dietary patterns (e.g., Mediterranean diet) for long-term weight loss maintenance. One weight loss maintenance trial implemented the Dietary Approaches to Stop Hypertension (DASH) diet during the initial weight loss phase (Svetkey et al., 2008). Results showed that all treatment conditions maintained significant weight loss at 30 months; however, groups receiving extended contact experienced less weight regain. The Mediterranean diet, which emphasizes consumption of plant-based foods and healthy fats and seafood, has been shown to be effective at producing weight loss at $\geq$12 months and, in one systematic review, performed better than low-fat diets (Mancini et al., 2016).

Intermittent fasting

In recent years, there has been increased interest in nutrient timing as a mechanism for weight loss and improved metabolic health (Templeman et al., 2020; Patterson and Sears, 2017). Intermittent fasting (IF) encompasses a broad spectrum of dietary timing approaches that partially or completely restrict energy intake within a consistent, recurring temporal period (e.g., alternate day fasting, fasting up to 4 days per week) (Patterson and Sears, 2017; Harris et al., 2018; Templeman et al., 2020). One such strategy is time-restricted feeding (TRF). TRF eating patterns restrict caloric consumption to a specific period of 8–10 h or less for every day of the week, and common variations include early- and mid-day TRF (Rynders et al., 2019). An evaluation of recent evidence related to intermittent energy restriction strategies as a treatment for overweight and obesity found equivalent results when intermittent energy restriction (IF or TRF) was compared to continuous energy restriction (Rynders et al., 2019). Nine out of eleven studies included in the review reported no significant between-group differences in weight change or total fat loss.

Few long-term studies of IF diets have been conducted. A 12-month RCT compared continuous energy restriction (CER) to two different IF strategies: week-on-week-off energy restriction (alternating each week between the same energy restriction as CER or habitual diet) and a 5:2 diet (habitual diet on 5 days per week and a modified fast of approximately 500 and 600 kcal on 2 days per week for women and men, respectively) (Headland et al., 2019). Mean weight loss was not significantly different between groups at 12 months (−5.0 kg for 5:2, −5.1 kg for week-on-week-off, and −6.6 kg for CER), and discontinuation rates were also similar, indicating that adherence to IF does not appear to be more difficult than traditional

weight loss diets. A separate 12-month RCT in adults with type 2 diabetes also compared CER (1200–1600 kcal/day) to a 5:2 IF diet (two non-consecutive fasting days with energy intake limited to 500 and 600 kcal for women and men, respectively) (Carter et al., 2018). At the end of the 12-month treatment period, the 5:2 condition lost significantly more weight than the CER condition (−6.8 kg vs. −5.0 kg, respectively); however, there were no differences in change in weight from baseline at 12 months post-treatment (Carter et al., 2019).

Conclusions on dietary strategies and weight loss maintenance

A larger initial weight loss is associated with better weight loss maintenance over time. However, any diet that produces a calorie deficit can induce weight loss, and current research does not clearly support the recommendation of one dietary strategy over another for the weight loss phase. Similarly, no optimal dietary strategy has been identified for long-term successful weight loss maintenance. Instead, identifying a dietary strategy that an individual is able to follow closely over time may be most important, as adherence, regardless of dietary strategy, is associated with improved weight outcomes (Alhassan et al., 2008).

Physical activity

Observational research suggests that physical activity (PA) is one of the strongest predictors of long-term weight maintenance. In addition to increasing energy expenditure, PA has other benefits, such as elevated mood and reduced hunger (Kanning and Schlicht, 2010; Vatansever-Ozen et al., 2011). Multiple robust RCTs have associated volume of PA with successful maintenance of weight loss (Jeffery et al., 2003).

The 2018 Physical Activity Guidelines for Americans recommend that adults participate in at least 150 min–300 min of moderate-intensity exercise per week or 75 min–150 min of vigorous-intensity erobic activity per week for good health (Piercy et al., 2018). They also recommend that adults engage in 2 or more days of muscle-strength training per week. These guidelines emphasize that adults who even modestly increase their moderate-to-vigorous activity will experience beneficial effects for weight loss. However, even greater levels of PA may be needed to maintain weight loss long-term. The American College of Sports Medicine (ACSM) recommends even higher levels of PA—200–300 min per week—to prevent weight regain after weight loss, and successful maintainers in NWCR report engaging in approximately 1 h of moderate-intensity activity per day, which exceeds the recommendation in the Physical Activity Guidelines for Americans (Donnelly et al., 2009).

Strategies for improving maintenance of physical activity

While the best diet for long-term obesity treatment is still to be identified, research consistently indicates that moderate-to-vigorous PA is integral to long-term weight loss maintenance success. Research in this area primarily focuses on maintaining increases in moderate-to-vigorous intensity PA.

For individuals with obesity, maintenance of increased PA levels is complicated by physiological, psychological, and environmental factors. The main perceived barriers to PA adherence include low fitness, pain, lack of stimuli, body image dissatisfaction, time, and weather constraints (Patel et al., 2013). Behavioral and cognitive strategies may mitigate these challenges to PA adherence (Patel et al., 2013; Dalle Grave et al., 2011). One beneficial technique may be intermittent bouts of exercise, such as 4 short (∼10 min) bouts of exercise for at least 5 days a week, vs. one long bout of continuous exercise (Jakicic et al., 1995). Though one RCT showed promise using this strategy, larger studies were unable to produce comparable outcomes in long-term adherence to intermittent exercise (Jakicic et al., 1995). A more recent study investigated how and when weight loss maintainers engage in PA throughout a single day and across 7 consecutive days (Creasy et al., 2021). Weight loss maintainers were found to engage in high amounts of moderate-to-vigorous PA ($\geq$60 min/day) on more days of the week and accumulated more moderate-to-vigorous PA in uninterrupted bouts, and during morning hours compared to the control group. Further research on temporal patterns of PA across days and weeks among weight loss maintainers may help inform behavioral interventions targeted at long-term PA adherence.

A wide array of other types of support to achieve PA adherence has also been tested, such as personal trainers, home exercise equipment, and financial incentives (Mitchell et al., 2013; Burns et al., 2012). Research suggests that as financial incentives increase, attendance at group exercise sessions increases, at least in the short term or during periods when the provision of incentives is active (Mitchell et al., 2013; Wall et al., 2006; Paul-Ebhohimhen and Avenell, 2008). However, programs that provide financial incentive may not be sustainable over time do prohibitive implementation costs.

Behavioral component

Obesity as a chronic disease

Healthcare's perspective on obesity has changed substantially in recent decades. In 1998, the National Institutes of Health published treatment guidelines that described obesity as a "complex multifactorial chronic disease" (Expert Panel on the Identification, 1998). However, viewing obesity as a disease was not widely accepted at that time (Kyle et al., 2016). Over the next 15 years, research began

to identify genetic and environmental factors that contribute to obesity, and consequently professional organizations and governmental agencies began to recognize the condition as a chronic disease rather than simply a cosmetic issue due to poor personal choices (Kyle et al., 2016; American Society for Metabolic and Bariatric Surgery). Concurrent with this perspective shift, the field has also moved toward a chronic care approach to obesity treatment. The Chronic Care Model (CCM) emphasizes several areas related to care of patients with chronic illnesses: (1) quality improvement is ongoing and incentivized within the care delivery system, (2) behavioral self-management support is provided to patients that increases their self-efficacy to manage their illness, (3) care teams and systems function in a way to meet the needs of patients with chronic illnesses, (4) evidence-based guidelines are implemented for care, and (5) information systems are improved and disease registries are developed to provide feedback on treatment performance (Wagner, 1998). Thus, the CCM highlights the need for ongoing, high-quality care that focuses on improving health care delivery systems to support successful disease management in patients with chronic illness. Similarly, the 2013 AHA/ACC/TOS Guideline for the Management of Overweight and Obesity in Adults provides a treatment algorithm for physicians treating patients with obesity that uses a "Chronic Disease Management Model" (Jensen et al., 2014). The algorithm emphasizes frequent monitoring of patient weight and related risk factors, as well as referral to intensive lifestyle intervention when warranted.

Extending professional contact

Research suggests that obesity treatment should extend beyond the initial weight loss phase and should include ongoing professional contact, behavioral skill development, social support, and physical activity adherence (Perri and Corsica, 2002). The healthy behaviors acquired through behavioral lifestyle interventions require continuous effort and practice to achieve long-term maintenance goals and sustained health benefits (Perri and Corsica, 2002; Latner et al., 2000).

An advisable recommendation is the extension of program delivery (Latner et al., 2000). Duration of treatment is correlated with weight loss following intervention, suggesting that the poor long-term maintenance of weight loss may be attributed to loss of support after termination of intervention. An RCT was implemented to compare two weight loss maintenance interventions with an individual-driven control (Svetkey et al., 2008). After the weight loss phase, participants were randomly assigned to either a monthly personal contact, a technology-based intervention, or a control group for 30 months (Svetkey et al., 2008). Following the intervention, participants in the personal-contact group regained less weight (4.0 kg) compared to the control group (5.5 kg). At 18 months, the interactive-technology group had lower weight regain compared to the self-directed group (mean difference, −1.1 kg); however, no difference was observed at 30 months.

The use of an extended care model that includes clinic-based follow-up is warranted. A recent meta-analysis assessed the evidence related to extended care and long-term weight maintenance (Ross Middleton et al., 2012). The effect of the extended care was additional weight loss maintenance of 3.2 kg over 17.6 months compared to both no-contact and education control groups. The domains deemed most effective at supporting weight loss maintenance included regular contact with an interventionist, group sessions coupled with reinforcement of behavioral strategies, and use of problem-solving skills to overcome perceived barriers.

Social support

Both dimensions of social support, structural (e.g., support-benefactor) and functional (e.g., perceived support), are thought to produce beneficial health outcomes such as weight maintenance (Karfopoulou et al., 2016). The obesity intervention literature consistently emphasizes the role of social support; however, the degree of effectiveness in weight loss maintenance remains unclear (Karfopoulou et al., 2016). A cross-sectional analysis examined if participants who experienced weight loss and weight maintenance differ by the amount and type of support they received. One cross-sectional study found that functional support may be a contributor to weight loss maintenance, such that maintainers perceived a greater degree of social support regarding healthy eating and physical activity. The authors concluded that social support should be focused on positive and not instructive feedback from peers and family members (Karfopoulou et al., 2016).

Self-monitoring

Self-monitoring is an important strategy for successful weight loss maintenance. Regularly weighing oneself has consistently been linked to more successful weight loss maintenance over time (Butryn et al., 2007; O'Neil and Brown, 2005). Almost half (44%) of NWCR members weigh themselves daily, and nearly a third (31%) more weigh themselves at least once per week (Wing and Phelan, 2005). Self-monitoring of diet and physical activity is also associated with greater success in maintaining weight loss (Laitner et al., 2016; Akers et al., 2012; Burke et al., 2011). Frequent self-monitoring can increase awareness of energy balance-related behaviors (i.e., diet and physical activity) and provide information on how these behaviors affect weight, allowing individuals to make timely adjustments to their diet and activity routines when they experience a weight increase.

Technology

Recent advances in technology offer unique opportunities for the field of weight management, such as streamlined intervention components (e.g., digital food and beverage tracking) and decreased barriers to treatment (e.g., bypassing issues due to lack of vehicle access). The efficacy of web-based interventions for weight loss has been studied for decades, and web-based interventions have consistently been shown to be effective in producing greater weight loss than control or minimal intervention conditions (Sorgente et al., 2017). However, research is mixed regarding whether web-based interventions are equally as effective as face-to-face delivery (Sorgente et al., 2017). The Weight Loss Maintenance Trial, which enrolled more than 1000 adults who had successfully lost 4 kg of body weight during an initial 6-month phase, found that both the personal-contact and interactive technology-based conditions regained less weight than the self-directed condition at 18 and 24 months (Svetkey et al., 2008). However, only the personal-contact condition regained less weight than the self-directed group at 30 months, with no difference between the interactive technology-based and self-directed conditions at this timepoint.

Furthermore, intervention technologies have begun to move beyond simply providing traditional treatment in an online format. One RCT tested the effects of using virtual reality (VR) in combination with a cognitive behavioral therapy approach on weight outcomes (Manzoni et al., 2016). Multiple VR modules were implemented, such as simulated environments important for maintenance and relapse prevention (e.g., supermarkets, pubs) that allowed participants to practice problem solving and decision-making skills. Participants in the VR-enhanced cognitive behavioral therapy condition had increased odds of weight maintenance and improved weight at 1-year post-intervention than both the standard behavioral inpatient program and the standard behavioral inpatient program plus standard cognitive behavioral therapy. Another increasingly common technology-based strategy is the use of avatars, or graphical representations of the user within the virtual environment, to elicit healthy behaviors. While evidence for the utility of this strategy is promising, additional research is needed to better understand the effect of avatars on health behaviors and outcomes, particularly compared to interventions that do not use them (Rheu et al., 2020).

Advances in smartphone technology have also provided novel approaches to support weight loss maintenance. Dozens of apps are now available in the commercial app market that allow users to track their food and beverage consumption and macro- and micronutrient intake more quickly and easily than conventional pen-and-paper methods (Wharton et al., 2014). Algorithms that provide real-time feedback on dietary choices within dietary self-monitoring apps are in early stages of testing (Burke et al., 2017). Additionally, text messages provide a convenient, low-cost method for engaging individuals in extended contact that can improve long-term weight maintenance (Spark et al., 2015). More sophisticated applications of phone-based technology to weight research include ecological momentary assessment (EMA). EMA allows researchers to gather information about individuals' behaviors and psychological state while in real-world settings, at or very near the moment that a behavior of interest takes place. EMA has been used for a variety of research-related purposes, from improving the accuracy of self-reported dietary data (users submit pictures of the foods and drinks they are about to eat) to understanding the contextual factors that lead to dietary lapses (users respond to prompts about lapses sent to their phones) (Forman et al., 2017; Loth et al., 2021).

Incentives

A systematic review found that randomized control trials investigating the efficacy of material incentives on weight loss and obesity-related behavior change showed incentives can increase targeted outcomes (e.g., physical activity, weight change, and program attendance) (Burns et al., 2012). However, researchers were less likely to see effects on outcomes that were not directly targeted by the interventionist (Burns et al., 2012). Employers have also increasingly used health insurance premium adjustments to incentivize healthy behavior change, reducing the financial contribution required from employees who meet the targets (Mattke et al., 2013). However, empirical evidence to support this strategy is limited. An RCT examining the effectiveness of offering a $550 premium reduction found no differences in weight loss from control group (Patel et al., 2016). Premium adjustments, which are by nature delayed, may not be valued in the same way as cash or other tangible incentives that can be offered more immediately.

Systems-level programs

As the prevalence of overweight and obesity increases, medical practice and public health interventions have shifted to consider societal and environmental factors. The literature suggests that multi-level interventions, which address factors at one or more levels of influence (e.g., individual, relationship, community, and society/policy levels), be implemented to address the physiological, environmental, and psychosocial determinants that influence weight loss maintenance (MacLean et al., 2015). Optimal environments can strengthen an individuals' capacity to make sustainable healthy choices, ultimately influencing their health and behavioral outcomes. Integrative primary care models deliver comprehensive healthcare that comprises conventional and complementary domains (e.g., population health, equity, governance, culture, advocacy) to reflect a holistic standard of care approach. Moreover, intervening across multiple levels of the socioecological model to encompass communities, organizations (e.g., school systems and worksites), and legislation would support individuals in the maintenance of healthy eating and physical activity behaviors.

Table 2 Strategies for supporting successful weight loss maintenance identified from the literature.

Area	*Strategy*
Diet	• Consume a low-calorie (~1400 kcal per day) diet • Choose a dietary strategy that can be adhered to over the long-term
Physical activity	• Be active at a moderate-intensity level for at least 200–300 min/week
Self-monitoring	• Weigh frequently (at least daily or weekly) • Track dietary intake • Track physical activity
Professional contact	• Extend professional contact beyond initial treatment—contact should be ongoing in keeping with the chronic care model of disease management
Technology	• Internet- and smartphone-based technologies may be helpful for extending professional contact, as well as increasing the ease of self-monitoring

Conclusion

Obesity is now recognized as a chronic disease that results from a milieu of physiological, psychological, environmental factors. It is well-known that moderate weight loss confers meaningful improvements in metabolic health; however, long-term weight loss maintenance can be challenging. Information from weight loss registries and results from experimental studies have helped to identify several strategies that support successful weight loss maintenance in the long term. Critically, individuals wishing to maintain their weight loss must engage in behaviors that support a favorable energy balance for weight maintenance, such as eating fewer calories and being more active than before their weight loss. If these behaviors are not sustained, weight regain will occur.

Experimental studies show that a variety of dietary approaches can be effective during the weight loss phase, if an energy deficit is achieved. For weight loss maintenance, energy intake needs to be lower than what occurred prior to weight loss, as well as, balanced with energy expenditure, such that energy balanced is achieved. No one specific dietary strategy has been found to be more effective for weight loss maintenance, and as with dietary intake during weight loss, an individual's ability to adhere to the dietary approach appears to be key. Engaging in 200 or more minutes of physical activity per week is also recommended to support successful weight loss maintenance. Additionally, extending contact beyond the initial intensive phase of intervention is vitally important for the maintenance of weight loss over time, which aligns with a chronic care model of obesity treatment. Technological advancements also continue to offer new approaches for improving weight loss maintenance, including web-based interventions that decrease access barriers, text messaging that can extend contact at low cost, and apps that streamline self-monitoring of diet and physical activity behaviors (Table 2).

Current research suggests that long-term treatment of obesity requires multiple intervention strategies and extended contact. Future research should view obesity through a chronic care lens and examine the frequency of contact and intensity of support needed after the initial weight loss phase to support optimal long term weight loss maintenance. Identifying intervention components that assist in maintaining energy-balance behaviors long-term should be emphasized. Additionally, novel approaches to using technology to improve both intervention outcomes and how research data are collected should continue to be investigated.

References

Akers, J.D., Cornett, R.A., Savla, J.S., Davy, K.P., Davy, B.M., 2012. Daily self-monitoring of body weight, step count, fruit/vegetable intake, and water consumption: a feasible and effective long-term weight loss maintenance approach. J. Acad. Nutr. Diet. 112, 685–692.e2.

Alhassan, S., Kim, S., Bersamin, A., King, A., Gardner, C., 2008. Dietary adherence and weight loss success among overweight women: results from the A to Z weight loss study. Int. J. Obes. 32, 985–991.

American Society for Metabolic and Bariatric Surgery. Available https://asmbs.org/, [Accessed October 6 2021].

Anastasiou, C.A., Fappa, E., Karfopoulou, E., Gkza, A., Yannakoulia, M., 2015. Weight loss maintenance in relation to locus of control: the MedWeight study. Behav. Res. Ther. 71, 40–44.

Astbury, N.M., Piernas, C., Hartmann-Boyce, J., Lapworth, S., Aveyard, P., Jebb, S.A., 2019. A systematic review and meta-analysis of the effectiveness of meal replacements for weight loss. Obes. Rev. 20, 569–587.

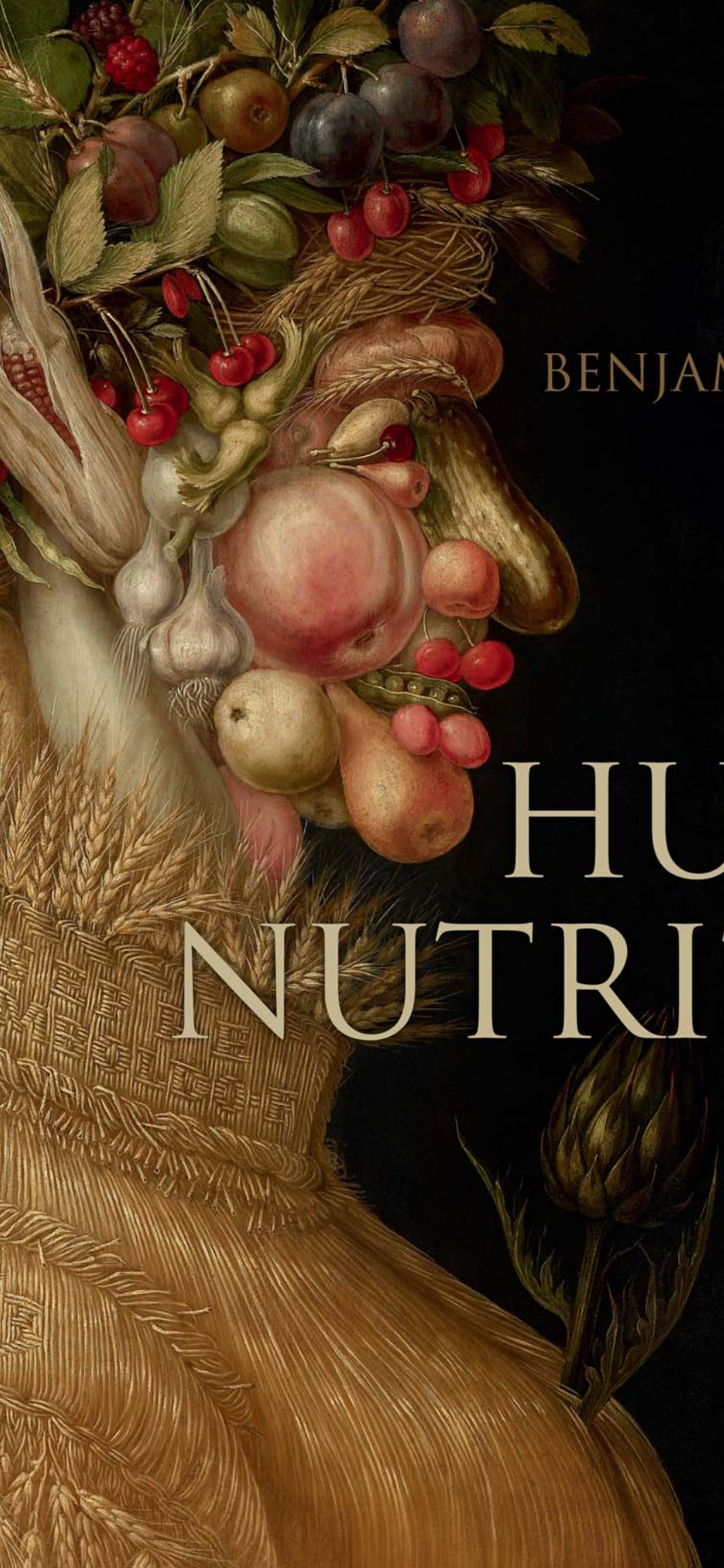
Baker, B., 2006. Weight loss and diet plans: several types of diet plans produce at least short-term weight loss; portion size may matter more than what we eat. Am. J. Nurs. 106, 52–59.

Bianchini, F., Kaaks, R., Vainio, H., 2002. Overweight, obesity, and cancer risk. Lancet Oncol. 3, 565–574.

Burke, L.E., Wang, J., Sevick, M.A., 2011. Self-monitoring in weight loss: a systematic review of the literature. J. Am. Diet Assoc. 111, 92–102.

Burke, L.E., Zheng, Y., Ma, Q., Mancino, J., Music, E., Styn, M., Ewing, L., French, B., Sieworek, D., Smailagic, A., 2017. The SMARTER pilot study: testing feasibility of real-time feedback for dietary self-monitoring. Prev. Med. Rep. 6, 278–285.
Burns, R.J., Donovan, A.S., Ackermann, R.T., Finch, E.A., Rothman, A.J., Jeffery, R.W., 2012. A theoretically grounded systematic review of material incentives for weight loss: implications for interventions. Ann. Behav. Med. 44, 375–388.
Butryn, M.L., Phelan, S., Hill, J.O., Wing, R.R., 2007. Consistent self-monitoring of weight: a key component of successful weight loss maintenance. Obesity 15, 3091–3096.
Carter, S., Clifton, P.M., Keogh, J.B., 2018. Effect of intermittent compared with continuous energy restricted diet on glycemic control in patients with type 2 diabetes: a randomized noninferiority trial. JAMA Netw. Open 1, e180756.
Carter, S., Clifton, P., Keogh, J., 2019. The effect of intermittent compared with continuous energy restriction on glycaemic control in patients with type 2 diabetes: 24-month follow-up of a randomized noninferiority trial. Diabetes Res. Clin. Pract. 151, 11–19.
Catenacci, V.A., Ogden, L.G., Stuht, J., Phelan, S., Wing, R.R., Hill, J.O., Wyatt, H.R., 2008. Physical activity patterns in the national weight control registry. Obesity 16, 153–161.
Catenacci, V.A., Odgen, L., Phelan, S., Thomas, J.G., Hill, J., Wing, R.R., Wyatt, H., 2014. Dietary habits and weight maintenance success in high versus low exercisers in the National Weight Control Registry. J. Phys. Activ. Health 11, 1540–1548.
Chang, S., Koegel, K., 2017. Back to basics: all about MyPlate food groups. J. Acad. Nutr. Diet. 117, 1351–1353.
Cordain, L., Eaton, S.B., Sebastian, A., Mann, N., Lindeberg, S., Watkins, B.A., O'keefe, J.H., Brand-Miller, J., 2005. Origins and evolution of the Western diet: health implications for the 21st century. Am. J. Clin. Nutr. 81, 341–354.
Creasy, S.A., Hibbing, P.R., Cotton, E., Lyden, K., Ostendorf, D.M., Willis, E.A., Pan, Z., Melanson, E.L., Catenacci, V.A., 2021. Temporal patterns of physical activity in successful weight loss maintainers. Int. J. Obes. 1–9.
Dalle Grave, R., Calugi, S., Centis, E., El Ghoch, M., Marchesini, G., 2011. Cognitive-behavioral strategies to increase the adherence to exercise in the management of obesity. J. Obes. 2011.
Ditschuneit, H.H., Flechtner-Mors, M., 2001. Value of structured meals for weight management: risk factors and long-term weight maintenance. Obes. Res. 9, 284S–289S.
Donnelly, J.E., Blair, S.N., Jakicic, J.M., Manore, M.M., Rankin, J.W., Smith, B.K., 2009. Appropriate physical activity intervention strategies for weight loss and prevention of weight regain for adults. Med. Sci. Sports Exerc. 41, 459–471.
Ello-Martin, J.A., Roe, L.S., Ledikwe, J.H., Beach, A.M., Rolls, B.J., 2007. Dietary energy density in the treatment of obesity: a year-long trial comparing 2 weight-loss diets. Am. J. Clin. Nutr. 85, 1465–1477.
Expert Panel on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults (U.S.), 1998. Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults: The Evidence Report. National Institutes of Health, National Heart, Lung, and Blood Institute: in cooperation with the National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD.
Fisher, R.A., 1937. The Design of Experiments.
Forman, E.M., Schumacher, L.M., Crosby, R., Manasse, S.M., Goldstein, S.P., Butryn, M.L., Wyckoff, E.P., Graham Thomas, J., 2017. Ecological momentary assessment of dietary lapses across behavioral weight loss treatment: characteristics, predictors, and relationships with weight change. Ann. Behav. Med. 51, 741–753.
Franz, M.J., Vanwormer, J.J., Crain, A.L., Boucher, J.L., Histon, T., Caplan, W., Bowman, J.D., Pronk, N.P., 2007. Weight-loss outcomes: a systematic review and meta-analysis of weight-loss clinical trials with a minimum 1-year follow-up. J. Am. Diet Assoc. 107, 1755–1767.
Freire, R., 2020. Scientific evidence of diets for weight loss: different macronutrient composition, intermittent fasting, and popular diets. Nutrition 69, 110549.
Fryar, C.D., Carroll, M.D., Afful, J., 2020. Prevalence of Overweight, Obesity, and Severe Obesity Among Adults Aged 20 and Over: United States, 1960–1962 Through 2017–2018. NCHS Health E-Stats.
Handjieva-Darlenska, T., Holst, C., Grau, K., Blaak, E., Martinez, J.A., Oppert, J.-M., Taylor, M.A., Sørensen, T.I., Astrup, A., 2012. Clinical correlates of weight loss and attrition during a 10-week dietary intervention study: results from the NUGENOB project. Obes. Facts 5, 928–936.
Harris, L., Hamilton, S., Azevedo, L.B., Olajide, J., De Brún, C., Waller, G., Whittaker, V., Sharp, T., Lean, M., Hankey, C., 2018. Intermittent fasting interventions for treatment of overweight and obesity in adults: a systematic review and meta-analysis. JBI Evid. Synthesis 16, 507–547.
Headland, M.L., Clifton, P.M., Keogh, J.B., 2019. Effect of intermittent compared to continuous energy restriction on weight loss and weight maintenance after 12 months in healthy overweight or obese adults. Int. J. Obes. 43, 2028–2036.
Hill, A.J., Blundell, J.E., 1986. Macronutrients and Satiety: The Effects of a High-Protein or High-Carbohydrate Meal on Subjective Motivation to Eat and Food Preferences. Nutrition and Behavior (USA).
Huang, T.T., Drewnowski, A., Kumanyika, S.K., Glass, T.A., 2009. A systems-oriented multilevel framework for addressing obesity in the 21st century. Prev. Chronic Dis. 6.
Huerta, S., Li, Z., Li, H., Hu, M., Yu, C., Heber, D., 2004. Feasibility of a partial meal replacement plan for weight loss in low-income patients. Int. J. Obes. 28, 1575–1579.
Jakicic, J.M., Wing, R., Butler, B., Robertson, R., 1995. Prescribing exercise in multiple short bouts versus one continuous bout: effects on adherence, cardiorespiratory fitness, and weight loss in overweight women. Int. J. Obes. Relat. Metab. Disord. 19, 893–901.
Jeffery, R.W., Wing, R.R., Sherwood, N.E., Tate, D.F., 2003. Physical activity and weight loss: does prescribing higher physical activity goals improve outcome? Am. J. Clin. Nutr. 78, 684–689.
Jensen, M.D., Ryan, D.H., Apovian, C.M., Ard, J.D., Comuzzie, A.G., Donato, K.A., Hu, F.B., Hubbard, V.S., Jakicic, J.M., Kushner, R.F., Loria, C.M., Millen, B.E., Nonas, C.A., Pi-Sunyer, F.X., Stevens, J., Stevens, V.J., Wadden, T.A., Wolfe, B.M., Yanovski, S.Z., 2014. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Obesity Society. J. Am. Coll. Cardiol. 63, 2985–3023.
Kanning, M., Schlicht, W., 2010. Be active and become happy: an ecological momentary assessment of physical activity and mood. J. Sport Exerc. Psychol. 32, 253–261.
Karfopoulou, E., Anastasiou, C.A., Avgeraki, E., Kosmidis, M.H., Yannakoulia, M., 2016. The role of social support in weight loss maintenance: results from the MedWeight study. J. Behav. Med. 39, 511–518.
Klem, M.L., Wing, R.R., Mcguire, M.T., Seagle, H.M., Hill, J.O., 1997. A descriptive study of individuals successful at long-term maintenance of substantial weight loss. Am. J. Clin. Nutr. 66, 239–246.
Kyle, T.K., Dhurandhar, E.J., Allison, D.B., 2016. Regarding obesity as a disease: evolving policies and their implications. Endocrinol. Metab. Clin. N. Am. 45, 511.
Laitner, M.H., Minski, S.A., Perri, M.G., 2016. The role of self-monitoring in the maintenance of weight loss success. Eat. Behav. 21, 193–197.
Latner, J., Stunkard, A., Wilson, G., Jackson, M., Zelitch, D., Labouvie, E., 2000. Effective long-term treatment of obesity: a continuing care model. Int. J. Obes. 24, 893–898.
Li, F., Harmer, P., Cardinal, B.J., Bosworth, M., Johnson-Shelton, D., 2009. Obesity and the built environment: does the density of neighborhood fast-food outlets matter? Am. J. Health Promot. 23, 203–209.
Look Ahead Research Group, 2014. Eight-year weight losses with an intensive lifestyle intervention: the Look AHEAD study. Obesity 22, 5–13.
Loth, K.A., Fertig, A., Trofholz, A., Harnack, L., Crosby, R., Neumark-Sztainer, D., Berge, J.M., 2021. Concordance of children's intake of selected food groups as reported by parents via 24-h dietary recall and ecological momentary assessment. Publ. Health Nutr. 24, 22–33.
Maclean, P.S., Wing, R.R., Davidson, T., Epstein, L., Goodpaster, B., Hall, K.D., Levin, B.E., Perri, M.G., Rolls, B.J., Rosenbaum, M., 2015. NIH working group report: innovative research to improve maintenance of weight loss. Obesity 23, 7–15.
Mancini, J.G., Filion, K.B., Atallah, R., Eisenberg, M.J., 2016. Systematic review of the Mediterranean diet for long-term weight loss. Am. J. Med. 129, 407–415.e4.
Manzoni, G.M., Cesa, G.L., Bacchetta, M., Castelnuovo, G., Conti, S., Gaggioli, A., Mantovani, F., Molinari, E., Cárdenas-López, G., Riva, G., 2016. Virtual reality-enhanced cognitive-behavioral therapy for morbid obesity: a randomized controlled study with 1 year follow-up. Cyberpsychol. Behav. Soc. Netw. 19, 134–140.
Mattke, S., Liu, H., Caloyeras, J., Huang, C.Y., Van Busum, K.R., Khodyakov, D., Shier, V., 2013. Workplace wellness programs study. Rand Health Q. 3.

Mccormack, G.R., Shiell, A., 2011. In search of causality: a systematic review of the relationship between the built environment and physical activity among adults. Int. J. Behav. Nutr. Phys. Activ. 8, 1–11.
Mcguire, M.T., Wing, R.R., Klem, M.L., Lang, W., Hill, J.O., 1999. What predicts weight regain in a group of successful weight losers? J. Consult. Clin. Psychol. 67, 177.
Metzgar, C., Preston, A., Miller, D., Nickols-Richardson, S., 2015. Facilitators and barriers to weight loss and weight loss maintenance: a qualitative exploration. J. Hum. Nutr. Diet. 28, 593–603.
Miller, C.J., Smith, S.N., Pugatch, M., 2020. Experimental and quasi-experimental designs in implementation research. Psychiatr. Res. 283, 112452.
Mitchell, M.S., Goodman, J.M., Alter, D.A., John, L.K., Oh, P.I., Pakosh, M.T., Faulkner, G.E., 2013. Financial incentives for exercise adherence in adults: systematic review and meta-analysis. Am. J. Prev. Med. 45, 658–667.
O'Neil, P.M., Brown, J.D., 2005. Weighing the evidence: benefits of regular weight monitoring for weight control. J. Nutr. Educ. Behav. 37, 319–322.
Paeratakul, S., Lovejoy, J.C., Ryan, D.H., Bray, G.A., 2002. The relation of gender, race and socioeconomic status to obesity and obesity comorbidities in a sample of US adults. Int. J. Obes. 26, 1205–1210.
Paixão, C., Dias, C.M., Jorge, R., Carraça, E.V., Yannakoulia, M., De Zwaan, M., Soini, S., Hill, J.O., Teixeira, P.J., Santos, I., 2020. Successful weight loss maintenance: a systematic review of weight control registries. Obes. Rev. 21, e13003.
Patel, A., Schofield, G.M., Kolt, G.S., Keogh, J.W., 2013. Perceived barriers, benefits, and motives for physical activity: two primary-care physical activity prescription programs. J. Aging Phys. Activ. 21, 85–99.
Patel, M.S., Asch, D.A., Troxel, A.B., Fletcher, M., Osman-Koss, R., Brady, J., Wesby, L., Hilbert, V., Zhu, J., Wang, W., 2016. Premium-based financial incentives did not promote workplace weight loss in a 2013–15 study. Health Aff. 35, 71–79.
Patterson, R.E., Sears, D.D., 2017. Metabolic effects of intermittent fasting. Annu. Rev. Nutr. 37.
Paul-Ebhohimhen, V., Avenell, A., 2008. Systematic review of the use of financial incentives in treatments for obesity and overweight. Obes. Rev. 9, 355–367.
Perri, M.G., Corsica, J.A., 2002. Improving the maintenance of weight lost in behavioral treatment of obesity. In: Wadden, T.A., Stunkard, A.J. (Eds.), Handbook of Obesity Treatment. Guilford Press.
Piercy, K.L., Troiano, R.P., Ballard, R.M., Carlson, S.A., Fulton, J.E., Galuska, D.A., George, S.M., Olson, R.D., 2018. The physical activity guidelines for Americans. JAMA 320, 2020–2028.
Raynor, H.A., Champagne, C.M., 2016. Position of the Academy of Nutrition and Dietetics: interventions for the treatment of overweight and obesity in adults. J. Acad. Nutr. Diet. 116, 129–147.
Raynor, H.A., Van Walleghen, E.L., Bachman, J.L., Looney, S.M., Phelan, S., Wing, R.R., 2011. Dietary energy density and successful weight loss maintenance. Eat. Behav. 12, 119–125.
Rheu, M., Jang, Y., Peng, W., 2020. Enhancing healthy behaviors through virtual self: a systematic review of health interventions using avatars. Game. Health J. 9, 85–94.
Rock, C.L., Flatt, S.W., Pakiz, B., Barkai, H.S., Heath, D.D., Krumhar, K.C., 2016. Randomized clinical trial of portion-controlled prepackaged foods to promote weight loss. Obesity 24, 1230–1237.
Rolls, B.J., Barnett, R.A., 2000. The Volumetrics Weight-Control Plan: Feel Full on Fewer Calories. Quill, New York.
Rolls, B.J., 2009. The relationship between dietary energy density and energy intake. Physiol. Behav. 97, 609–615.
Ross Middleton, K., Patidar, S., Perri, M., 2012. The impact of extended care on the long-term maintenance of weight loss: a systematic review and meta-analysis. Obes. Rev. 13, 509–517.
Rössner, S., 1997. Defining success in obesity management. Int. J. Obes. Relat. Metab. Disord. 21, S2–S4.
Rynders, C.A., Thomas, E.A., Zaman, A., Pan, Z., Catenacci, V.A., Melanson, E.L., 2019. Effectiveness of intermittent fasting and time-restricted feeding compared to continuous energy restriction for weight loss. Nutrients 11, 2442.
Sacks, F.M., Bray, G.A., Carey, V.J., Smith, S.R., Ryan, D.H., Anton, S.D., Mcmanus, K., Champagne, C.M., Bishop, L.M., Laranjo, N., 2009. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. N. Engl. J. Med. 360, 859–873.
Saris, W.H., 2001. Very-low-calorie diets and sustained weight loss. Obes. Res. 9, 295S–301S.
Shan, Z., Rehm, C.D., Rogers, G., Ruan, M., Wang, D.D., Hu, F.B., Mozaffarian, D., Zhang, F.F., Bhupathiraju, S.N., 2019. Trends in dietary carbohydrate, protein, and fat intake and diet quality among US adults, 1999–2016. JAMA 322, 1178–1187.
Sorgente, A., Pietrabissa, G., Manzoni, G.M., Re, F., Simpson, S., Perona, S., Rossi, A., Cattivelli, R., Innamorati, M., Jackson, J.B., 2017. Web-based interventions for weight loss or weight loss maintenance in overweight and obese people: a systematic review of systematic reviews. J. Med. Internet Res. 19, e229.
Spark, L.C., Fjeldsoe, B.S., Eakin, E.G., Reeves, M.M., 2015. Efficacy of a text message-delivered extended contact intervention on maintenance of weight loss, physical activity, and dietary behavior change. JMIR mHealth uHealth 3, e4114.
Svetkey, L.P., Stevens, V.J., Brantley, P.J., Appel, L.J., Hollis, J.F., Loria, C.M., Vollmer, W.M., Gullion, C.M., Funk, K., Smith, P., 2008. Comparison of strategies for sustaining weight loss: the weight loss maintenance randomized controlled trial. JAMA 299, 1139–1148.
Teixeira, P.J., Carraça, E.V., Marques, M.M., Rutter, H., Oppert, J.-M., De Bourdeaudhuij, I., Lakerveld, J., Brug, J., 2015. Successful behavior change in obesity interventions in adults: a systematic review of self-regulation mediators. BMC Med. 13, 1–16.
Templeman, I., Gonzalez, J.T., Thompson, D., Betts, J.A., 2020. The role of intermittent fasting and meal timing in weight management and metabolic health. Proc. Nutr. Soc. 79, 76–87.
The National Weight Control Registry. Available http://www.nwcr.ws/default.htm, [Accessed October 21 2021].
Thomas, J.G., Bond, D.S., Phelan, S., Hill, J.O., Wing, R.R., 2014. Weight-loss maintenance for 10 years in the National Weight Control Registry. Am. J. Prev. Med. 46, 17–23.
Tremmel, M., Gerdtham, U.-G., Nilsson, P.M., Saha, S., 2017. Economic burden of obesity: a systematic literature review. Int. J. Environ. Res. Publ. Health 14, 435.
Van Baak, M.A., Mariman, E., 2019. Dietary strategies for weight loss maintenance. Nutrients 11, 1916.
Varkevisser, R., Van Stralen, M., Kroeze, W., Ket, J., Steenhuis, I., 2019. Determinants of weight loss maintenance: a systematic review. Obes. Rev. 20, 171–211.
Vatansever-Ozen, S., Tiryaki-Sonmez, G., Bugdayci, G., Ozen, G., 2011. The effects of exercise on food intake and hunger: relationship with acylated ghrelin and leptin. J. Sports Sci. Med. 10, 283.
Wagner, E.H., 1998. Chronic disease management: what will it take to improve care for chronic illness? Effect Clin. Pract. 1.
Wall, J., Mhurchu, C.N., Blakely, T., Rodgers, A., Wilton, J., 2006. Effectiveness of monetary incentives in modifying dietary behavior: a review of randomized, controlled trials. Nutr. Rev. 64, 518–531.
Wharton, C.M., Johnston, C.S., Cunningham, B.K., Sterner, D., 2014. Dietary self-monitoring, but not dietary quality, improves with use of smartphone app technology in an 8-week weight loss trial. J. Nutr. Educ. Behav. 46, 440–444.
Wing, R.R., Hill, J.O., 2001. Successful weight loss maintenance. Annu. Rev. Nutr. 21, 323–341.
Wing, R.R., Phelan, S., 2005. Long-term weight loss maintenance. Am. J. Clin. Nutr. 82, 222S–225S.
Wing, R.R., Tate, D.F., Gorin, A.A., Raynor, H.A., Fava, J.L., 2006. A self-regulation program for maintenance of weight loss. N. Engl. J. Med. 355, 1563–1571.



EDITOR IN CHIEF

BENJAMIN CABALLERO

VOLUME FOUR

Encyclopedia of HUMAN NUTRITION

FOURTH EDITION

ENCYCLOPEDIA OF HUMAN NUTRITION

FOURTH EDITION

This page intentionally left blank

ENCYCLOPEDIA OF HUMAN NUTRITION

FOURTH EDITION

EDITOR IN CHIEF

Professor Benjamin Caballero,

Emeritus Professor, Bloomberg School of Public Health and School of Medicine, Johns Hopkins University
Baltimore, MD, United States

VOLUME 4

Section 6: Nutrition in the Life Cycle

Edited by Lawrence J. Cheskin

Department of Nutrition and Food Studies, George Mason University, Fairfax, VA, United States and Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, United States

Section 7: Nutrition Epidemiology

Edited by Manuel Franco

Associate Professor at the Department of Surgery, Medical and Social Sciences, University of Alcalá, Madrid, Spain, and adjunct faculty at the Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Section 8: Global aspects of Human Nutrition

Edited by Daniel J. Hoffman

Department of Nutritional Sciences and director of the Program in International Nutrition at Rutgers University, New Brunswick, NJ, United States

ELSEVIER

AMSTERDAM ● BOSTON ● HEIDELBERG ● LONDON ● NEW YORK ● OXFORD
PARIS ● SAN DIEGO ● SAN FRANCISCO ● SINGAPORE ● SYDNEY ● TOKYO
Academic Press is an imprint of Elsevier

ACADEMIC PRESS

Academic Press is an imprint of Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge MA 02139, United States

Copyright © 2023 Elsevier Ltd. All rights reserved

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers may always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN 978-0-12-821848-8

For information on all publications visit our website at http://store.elsevier.com



Publisher: Oliver Walter
Acquisitions Editors: Emma Hayes and Michael Nicholls
Content Project Managers: Richard Berryman and Sindhuja Vadlamani
Associate Content Project Manager: N. Kiruthigadevi
Designer: Matthew Limbert

CONTENTS OF VOLUME 4

Contributors to Volume 4 *ix*

Editor Biographies *xiii*

Preface *xvii*

NUTRITION IN THE LIFE CYCLE

Adolescents: Nutrient requirements 1
CHS Ruxton and E Derbyshire

Aging 12
P Hyland, Y Barnett, and LH Allen

Breastfeeding 20
Chessa K Lutter

Children: Nutritional requirements 30
I Thorsdottir and B Thorisdottir

Complementary feeding 43
Nirupama Shivakumar

Early origins of disease 50
DB Hardy and DJ Hoffman

Early origins of disease: Non-fetal 58
Daniel J Hoffman and Alessandra D Sommer

Growth and development: Physiological aspects 66
William W Hay, Jr.

Growth monitoring and promotion 83
Leila M Larson, Edward A Frongillo, and Jef L Leroy

Lactation: Dietary requirements 91
H Lochen and H Tymann

Lactation: Physiology 98
JL McManaman and MC Neville

Low birth weight and preterm infants: Nutritional management 107
H Tymann and H Lochen

Low birth weight and preterm infants: Causes, prevalence, and prevention 116
M Merialdi and JH Requejo

Nutrition during the preschool years 121
Samantha L Huey, Julia L Finkelstein, and Saurabh Mehta

Nutrition for sport and exercise 127
RJ Maughan

Nutrition guidance for infants: Nutrient-based reference intakes and feeding recommendations 133
Stephanie A Atkinson

Nutritional problems of adolescents 143
S Gordon and CHS Ruxton

Older people: Nutritional requirements 150
N Solomons

Older people: Physiological changes 157
N Solomons

Pregnancy: Energy requirements and metabolic adaptations 163
GR Goldberg

Pregnancy: Nutrient requirements 168
Melissa F Young and Usha Ramakrishnan

Pregnancy: Placental regulation of nutrient delivery to the fetus 177
P Haggarty

Pregnancy: Pre-eclampsia and diet 185
Doris González-Fernández, Kristine G Koski, and Zulfiqar Bhutta

Pregnancy: Prevention of neural tube defects 198
Helene McNulty, Aoife Caffrey, and Kristina Pentieva

Pregnancy: Safe diets 212
S Stanner and H Gibson-Moore

Pregnancy: Weight gain 222
Melissa F Young and Usha Ramakrishnan

Stunting: Prevalence and prevention 230
Nandita Perumal, Diego G Bassani, and Daniel E Roth

NUTRITION EPIDEMIOLOGY

Coronary heart disease: Dietary patterns 241
Danielle E Haslam, Shatabdi Goon, Saloni Gautam, and Shilpa N Bhupathiraju

Dietary issues in coronary heart disease prevention 261
Rajeev Gupta and Shalini Singhal

Food culture 270
Johanna T Dwyer

Food environment 282
Manuel Franco and Julia Díez

Food fortification programs 288
Helena Pachón

Food labeling 301
Barbara Schneeman

Food security 306
Saskia de Pee and Rafael Pérez-Escamilla

Harmonizing the approach to deriving nutrient requirements 316
Ann L Yaktine

Health inequities 327
R Perez-Escamilla

Religious customs, influence on diet 336
K Albala

School food systems 341
Manuel Franco and Marta Fajó-Pascual

Seasonality of nutrition 350
Ilana R Cliffer, Anastasia Marshak, Kate R Schneider, Aishwarya Venkat, and Elena N Naumova

GLOBAL ASPECTS OF HUMAN NUTRITION

Climate change: Effects on health and nutrition 369
Shweta Khandelwal, Sarah Parwez, and Mahima Mehra

Dietary guidelines, international perspectives 379
B Schneeman

Dietary surveys: Surveys of food intake in groups and individuals 385
Katherine L Tucker

Displaced populations 392
Daniel J Hoffman and Hannah Posluszny

Double burden of malnutrition 401
Daniel J Hoffman, Angelica A Pozzoli, and Hannah R Posluszny

Famine: Causes, consequences and responses 411
Shannon Doocy, Shelley Walton, and Keith P West, Jr.

Food security and nutrition surveillance in low- and middle-income countries 420
Kathrin M Demmler, Valerie M Friesen, and Lynnette M Neufeld

Nutrition transition, diet change, and its implications 435
Daniel J Hoffman and Hannah Posluszny

Nutritional surveillance: Developed countries 444
EW Harris

Supplementation: Developing countries 454
R Shrimpton

Index *459*

This page intentionally left blank

CONTRIBUTORS TO VOLUME 4

K Albala
University of the Pacific, Stockton, CA, USA

LH Allen
ARS Western Human Nutrition Research Center, Davis, CA, USA

Stephanie A Atkinson
Department of Pediatrics, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

Y Barnett
Nottingham Trent University, Nottingham, UK

Diego G Bassani
Center for Global Child Health, Hospital for Sick Children, Toronto, ON, Canada; Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; and Department of Pediatrics, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

Shilpa N Bhupathiraju
Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States; and Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, United States

Zulfiqar Bhutta
Centre for Global Child Health, The Hospital for Sick Children (SickKids), Toronto, ON, Canada

Aoife Caffrey
Nutrition Innovation Centre for Food and Health (NICHE), School of Biomedical Sciences, Ulster University, Coleraine, United Kingdom

Ilana R Cliffer
Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, United States

Kathrin M Demmler
Global Alliance for Improved Nutrition, London, United Kingdom

Saskia de Pee
Systems Analysis for Nutrition, the United Nations World Food Program, Rome, Italy

E Derbyshire
Manchester Metropolitan University, Manchester, UK

Shannon Doocy
Department of International Health, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Julia Díez
Public Health and Epidemiology Research Group, School of Medicine and Health Sciences, Universidad de Alcalá, Madrid, Spain

Johanna T Dwyer
Department of Medicine and Community Health, Tufts University School of Medicine, Friedman School of Nutrition Science and Policy, Boston, MA, United States

Marta Fajó-Pascual
Faculty of Health and Sports Sciences, Universidad de Zaragoza, Instituto Agroalimentario de Aragón (IA2), Huesca, Spain

Julia L Finkelstein
Division of Nutritional Sciences, Cornell University, Ithaca, NY, United States; and Institute for Nutritional Sciences, Global Health, and Technology, Cornell University, Ithaca, NY, United States

Manuel Franco
Public Health and Epidemiology Research Group, School of Medicine and Health Sciences, Universidad de Alcalá, Madrid, Spain; and Universidad de Alcalá, Facultad de Medicina y Ciencias de la Salud, Grupo de Investigación en Epidemiología y Salud Pública, Alcalá de Henares, Madrid, España

Valerie M Friesen
Global Alliance for Improved Nutrition, Geneva, Switzerland

Edward A Frongillo
Department of Health Promotion, Education, and Behavior, University of South Carolina, Columbia, SC, United States

Saloni Gautam
Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, United States

H Gibson-Moore
British Nutrition Foundation, London, UK

GR Goldberg
MRC Human Nutrition Research, Cambridge, UK

Doris González-Fernández
Centre for Global Child Health, The Hospital for Sick Children (SickKids), Toronto, ON, Canada

Shatabdi Goon
Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

S Gordon
Ulster University, Coleraine, Northern Ireland

Rajeev Gupta
Academic Research Development Unit, Rajasthan University of Health Sciences, Jaipur, Rajasthan, India; and Department of Preventive Cardiology & Medicine, Eternal Heart Care Centre & Research Institute, Jaipur, Rajasthan, India

P Haggarty
University of Aberdeen, Aberdeen, UK

DB Hardy
The Children's Health Research Institute and the Lawson Health Research Institute, The University of Western Ontario, London, ON, Canada

EW Harris
Beltsville Human Nutrition Research Center, ARS, USDA, Beltsville, MD, USA

Danielle E Haslam
Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States; and Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, United States

William W Hay, Jr.
University of Colorado, Denver, CO, United States

Daniel J Hoffman
Department of Nutritional Sciences, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

Samantha L Huey
Division of Nutritional Sciences, Cornell University, Ithaca, NY, United States

P Hyland
Nottingham Trent University, Nottingham, UK

Shweta Khandelwal
Public Health Foundation of India, Gurugram, Haryana, India

Kristine G Koski
School of Human Nutrition, McGill University (Macdonald Campus), Ste-Anne-de-Bellevue, QC, Canada

Leila M Larson
Department of Health Promotion, Education, and Behavior, University of South Carolina, Columbia, SC, United States

Jef L Leroy
International Food Policy Research Institute, Washington, DC, United States

H Lochen
Johns Hopkins Children's Hospital, Baltimore, MD, United States

Chessa K Lutter
RTI International, Washington, DC, United States

Anastasia Marshak
Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, United States

RJ Maughan
Loughborough University, Loughborough, UK

JL McManaman
University of Colorado Anschutz Medical Campus, Aurora, CO, USA

Helene McNulty
Nutrition Innovation Centre for Food and Health (NICHE), School of Biomedical Sciences, Ulster University, Coleraine, United Kingdom

Mahima Mehra
Institute of Development Studies (IDS), University of Sussex, Brighton, United Kingdom

Saurabh Mehta
Division of Nutritional Sciences, Cornell University, Ithaca, NY, United States; and Institute for Nutritional Sciences, Global Health, and Technology, Cornell University, Ithaca, NY, United States

M Merialdi
WHO, Avenue Appia, Geneva, Switzerland

Elena N Naumova
Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, United States

Lynnette M Neufeld
Food and Agriculture Organization of the United Nations (FAO), Rome, Italy

MC Neville
University of Colorado Anschutz Medical Campus, Aurora, CO, USA

Helena Pachón
Food Fortification Initiative, Emory University, Atlanta, GA, United States

Sarah Parwez
Jamia Millia Islamia, Department of Hospital Management and Hospice Studies, New Delhi, India

Kristina Pentieva
Nutrition Innovation Centre for Food and Health (NICHE), School of Biomedical Sciences, Ulster University, Coleraine, United Kingdom

Rafael Pérez-Escamilla
Department of Social and Behavioral Sciences, Yale School of Public Health, New Haven, CT, United States

Nandita Perumal
Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA, United States

Hannah R Posluszny
Rutgers, the State University of New Jersey, New Brunswick, NJ, United States

Angelica A Pozzoli
Rutgers University, New Brunswick, NJ, United States

Usha Ramakrishnan
Hubert Department of Global Health, Emory University, Atlanta, GA, United States

JH Requejo
Johns Hopkins Bloomberg School of Public Health, MD, USA

Daniel E Roth
Center for Global Child Health, Hospital for Sick Children, Toronto, ON, Canada; Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; Department of Pediatrics, Faculty of Medicine, University of Toronto, Toronto, ON, Canada; and Department of Nutrition, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

CHS Ruxton
Nutrition Communications, Cupar, United Kingdom

Barbara Schneeman
University of California, Davis, CA, United States

Kate R Schneider
Paul H. Nitze School of Advanced International Studies, Johns Hopkins University, Washington, DC, United States

Nirupama Shivakumar
Division of Nutrition, St. John's Research Institute, Koramangala, Bengaluru, India

R Shrimpton
Independent Consultant, Olhão, Portugal

Shalini Singhal
Dr. Shalini's Diet & Wellness Clinic, Dwarka, New Delhi, India

N Solomons
Aging and Metabolism (CeSSIAM), Guatemala City, Guatemala

Alessandra D Sommer
Rutgers University, New Brunswick, NJ, United States

S Stanner
British Nutrition Foundation, London, UK

B Thorisdottir
Unit for Nutrition Research, Health Science Institute, School of Health Sciences, University of Iceland, Reykjavík, Iceland

I Thorsdottir
Unit for Nutrition Research, Health Science Institute, School of Health Sciences, University of Iceland, Reykjavík, Iceland

Katherine L Tucker
Department of Biomedical and Nutritional Sciences, Center for Population Health, University of Massachusetts Lowell, Lowell, MA, United States

H Tymann
Johns Hopkins Children's Hospital, Baltimore, MD, United States

Aishwarya Venkat
Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, United States

Shelley Walton
Department of International Health, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Keith P West, Jr.
Department of International Health, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

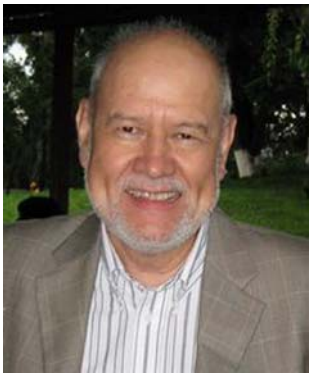
Ann L Yaktine
Food and Nutrition Board, Health and Medicine Division, The National Academies of Sciences, Engineering, and Medicine, Washington, DC, United States

Melissa F Young
Hubert Department of Global Health, Emory University, Atlanta, GA, United States

EDITOR BIOGRAPHIES

Editor in Chief

Professor Benjamin Caballero
Emeritus Professor, Bloomberg School of Public Health and School of Medicine, Johns Hopkins University, Baltimore, MD, United States.



Dr. Caballero is Professor Emeritus at the Department of International Health, Bloomberg School of Public Health, with joint appointment at the Department of Pediatrics, School of Medicine, Johns Hopkins University. He obtained his MD from the University of Buenos Aires, Argentina, and his PhD (in neuroendocrine regulation) from MIT, in Cambridge, Massachusetts. He started his academic career at Boston Children's Hospital, Harvard Medical School, and subsequently became the Founding Director of the Center for Human Nutrition at Johns Hopkins University.

Dr. Caballero has focused his research on child nutrition and health in developing countries. In particular, he has explored the combination of undernutrition and overweight that has become increasingly prevalent in low- and middle-income countries.

He is currently a member of the Council of the International Union of Nutritional Sciences. He has served on the Food and Nutrition Board of the US National Academy of Medicine and on a number of expert panels, including the Dietary Reference Intakes Committee, the Expert Panel on Macronutrient Requirements, and the Childhood Obesity Task Force. He was also a member of the U.S. Dietary Guidelines for Americans Advisory Committee, of the Scientific Advisory Board of the Food and Drug Administration, and of advisory committees of the National Institutes of Health and the Department of Agriculture. He is a Fellow of the American Society for Nutrition and of the Royal Society of Medicine (UK), and a member of the Spanish Academy of Nutritional Sciences.

He is the Editor-in-Chief of the *Encyclopedia of Food Sciences and Nutrition*, a 10-volume work on food production, consumption, and biological effects. He is also Editor-in-Chief of the *Encyclopedia of Human Nutrition*, which received the Book of the Year Award from the British Medical Association. His *Guide to Dietary Supplements* summarizes the current scientific basis for the use of mineral and vitamin supplements. He also co-edited a widely used textbook on human nutrition, *Modern Nutrition in Health and Disease*.

Section Editors

Section 1: *The Foundations of Human Nutrition*

Professor Angel Gil

Honorary President of the Iberoamerican Nutrition Foundation (FINUT), Emeritus Professor, Department of Biochemistry and Molecular Biology II, Institute of Nutrition and Food Technology, Centre of Biomedical Research, University of Granada, Campus de la Salud, Avda. del Conocimiento, 18100 Armilla, Granada, Spain.



Angel Gil is an Emeritus Professor at the University of Granada and a Full Professor in the Department of Biochemistry and Molecular Biology. He is Doctor Honoris Causa at the Autonomous University of Nuevo León, Monterrey, México, and a visitant Professor of the University of Chile, Santiago de Chile, and University of San Miguel de Tucumán, Argentina. He got a major degree in Biology (1973) and PhD in Biochemistry and Molecular Biology (1978).

Prof. Gil is an internationally recognized authority in Food and Nutrition: His expertise extends from the study of human milk composition to the molecular effects of food bioactive compounds and probiotics and the design and development of novel products for infant and clinical nutrition. He conducted pioneering and innovative research, leading 7 international, 27 national, and more than 50 projects and 120 contracts; he has taught since 1981, supervising more than 50 PhD students.

Prof. Gil has several areas of interest that include evaluating the role of dietary nucleotides in early life and the development of infant nutrition products. Besides, the isolation, identification, and description of the mechanism of action of probiotics and the metabolic, molecular, and genetic factors involved in obesity and the early onset of metabolic syndrome (MS) in childhood; and the design, development, and evaluation of enteral clinical nutrition products. What describes Prof. Gil best is the variety of fields and problems he has faced during his professional carrier and his significant ability to combine his knowledge and expertise in Food Science and Human Biochemistry. This has allowed him to design, develop, innovate, and evaluate exclusive products for Human Nutrition, which are demonstrated in his published articles and his patents' impact.

The multi- and interdisciplinary nature of his work is reflected in the variety of international journals in which he has published 546 articles. Also, he has published 28 books and about 180 book chapters. His five volumes *Treatise of Nutrition*, 3rd Edition, Ed. Medica Panamericana, 2017, with more than 3500 pages, is the "bedside" book for the study of Nutritional Sciences in Spain and all Latin American countries.

He has also been the Chairman of the International Union of Nutritional Sciences (IUNS) 21st International Congress of Nutrition (2013) and the Executive Director of the 23rd International Congress of Nutrition (2017) and has been engaged in the organization of other renowned international congresses. He is a member of prestigious international and national nutrition societies and Honorary President of the Iberoamerican Nutrition Foundation (FINUT), a nonprofit organization promoted by the IUNS, in which the main goal is to contribute to the formation of young scientists in Food and Nutrition in the setting of Iberoamerica. He has received 42 National and International Awards for his contribution to Nutrition and Food Science, among them, the Class Fellow 2022 of the American Society of Nutrition; the Sir David Cuthbertson Lecture Award of the European Society of Clinical Nutrition and Metabolism for scientific achievement in clinical nutrition on 2021; the Award "Granada, City of Science and Innovation" 2021 to the Scientific Career; the Gregorio Marañón Award 2018 to the best Spanish Scientist in the field of Food Science and Nutrition; the Institute Danone Spain Award 2017; the Award of the Spanish Federation of Dairy Industries, 2015; the Nutra Excellence Award 2014, Nutra India Summitt; the UIB Honorary Award of 2013; and the NAOS Strategy Prize 2012, Special recognition for his extensive professional experience in the field of nutrition and obesity, Spanish Ministry of Health, Social Services and Equality (AESAN).

Section 2: *Molecular Mechanisms for the Interaction of Nutrients and Health*

Dr. Noel W. Solomons

Center for the Studies of Sensory Impairment, Aging, and Metabolism (CeSSIAM), Guatemala City, Guatemala, Guatemala



Noel W. Solomons was Assistant and Associate Professor from 1977 to 1984 in the former Department of Nutrition and Food Science of MIT. A resident of Guatemala since 1975, he cofounded the Center for Studies of Sensory Impairment, Aging and Metabolism (CeSSIAM) in 1985, where he has been Scientific Director ever since. He is an Adjunct Professor at Tufts University in both the Friedman School of Nutrition Science and Policy and the Department of Public Health and Community Medicine.

Dr. Solomons received his MD degree from Harvard Medical School and subsequently undertook clinical and research training in infectious diseases and in gastroenterology and clinical nutrition at the Universities of Pennsylvania and University of Chicago, respectively. He was Editor-in-Chief of the *Food and Nutrition Bulletin* from 2016–2020. He has 364 publications indexed on PUBMED. In addition, he has edited 2 books and contributed over 100 articles, reviews, editorials, and commentaries in nonindexed venues and over 50 book chapters. These are dedicated to the scientific and academic interests of his career including: clinical nutrition; human growth and body composition; lactose maldigestion; dietary intake, nutritional status, intestinal absorption, and food fortification related to various micronutrients (vitamins, trace elements, and essential fatty acids); complementary feeding; nutrition in aging and chronic disease; and the interaction of malnutrition and infection. Also, he has supervised doctoral dissertations for 12 PhD candidates from the USA, Canada, Germany, Spain, the UK and the Netherlands through CeSSIAM.

Section 3: *Diet Composition*
Professor Anura Kurpad
Department of Physiology, St John's Medical College, Bengaluru, India



Anura Kurpad works at St. John's Medical College, Bengaluru. He received his MBBS and MD from St. John's, and his PhD from Bangalore University in 1992. He was a Postdoctoral Fellow at the Rowett Research Institute and at Cambridge University, UK, and a visiting Scientist at MIT, Cambridge, United States. He is an elected Fellow of the Royal College of Physicians (London), Fellow of the Indian National Academy of Medical Sciences, Indian Academy of Sciences, and International Union of Nutritional Sciences. His interests are in human, clinical, and public health aspects of nutrition, applied throughout the life cycle. His research focuses on the physiology and clinical aspects of human energy and protein requirements and metabolism, with more recent interests in micronutrient status and metabolism. He is an Associate Editor of *The American Journal of Clinical Nutrition*. He was the President of the Nutrition Society of India from 2012 to 2016 and is the current President of the Asia Pacific Clinical Nutrition Society. He has been, and is, on many national and international advisory bodies, including being the Chair of the Indian Nutrient Requirements Committee.

Section 4: *Disorders Directly Related to Inadequate Nutrient Intake*
Dr. Katherine L. Tucker
Department of Biomedical and Nutritional Sciences, Center for Population Health, University of Massachusetts Lowell, Lowell, MA, United States



Katherine L. Tucker, PhD, is University Distinguished Professor of Nutritional Epidemiology in the Department of Biomedical and Nutritional Sciences, and Director of the Center for Population Health, at the University of Massachusetts Lowell. She holds an adjunct appointment at the University of Massachusetts Medical School. She received her PhD from Cornell University and her undergraduate degree from the University of Connecticut, both in nutritional sciences. Between these degrees, she spent 2 years as a Peace Corps volunteer in the Philippines. Before joining UMass Lowell, she was at the USDA Human Nutrition Research Center on Aging at Tufts University, and McGill University. Dr. Tucker has contributed to more than 450 articles in scientific journals. Her research focuses on dietary intake and risk of chronic disease, including osteoporosis, cognitive decline, obesity, metabolic syndrome, and heart disease, with an emphasis on health disparities. She is the PI of the Boston Puerto Rican Health Study, an ongoing cohort study, to examine the roles of diet, health behaviors, stress, and genetic predisposition in relation to chronic conditions, including heart disease, cognitive decline, and bone health; and is actively involved as a scientific advisor for the NHLBI Jackson Heart Study. She served two terms on the Food and Nutrition Board of the National Academies of Science and Engineering. She is a Fellow of the American Society for Nutrition (ASN), the Gerontological Society of America, and the American Society for Bone Mineral Research and is currently the Editor-in-Chief of *Advances in Nutrition*, the international review journal of the ASN, and Senior Editor of the forthcoming 12th edition of the textbook, *Modern Nutrition in Health and Disease*.

Section 5: *Nutrition in Disease States*
Dr. Paolo M. Suter
Department of Endocrinology, Diabetology and Clinical Nutrition, University Hospital, Zurich, Switzerland.



Paolo M. Suter, MD, MS, is a Professor of Medicine presently affiliated with the Department of Endocrinology, Diabetology and Clinical Nutrition at the University Hospital Zurich (Switzerland). He received his MD at the University of Zurich and an MS in Nutrition at Tufts University (Boston, USA). He specialized in Internal Medicine and was a Faculty Member at the Medical Policlinic of the University Hospital Zurich, where he was directing the well-known Hypertension and Obesity Outpatient Consultation. His research activities focused on vitamin nutriture in the elderly, alcohol metabolism and obesity, blood pressure and hypertension, as well as nutrition and lifestyle in non-communicable disease prevention. During his clinical work he focused on the ideal combination of pharmacological therapy with nonpharmacological strategies and on the therapy and prevention of diseases especially hypertension and obesity. Besides many research publications and review articles he authored a widely used textbook entitled *Checkliste Ernährung* (*Checklist Nutrition*). He served on different national and international boards in the area of his expertise.

Section 6: *Nutrition in the Life Cycle*
Professor Lawrence J. Cheskin
Department of Nutrition and Food Studies, George Mason University, Fairfax, VA, United States
And
Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, United States



Lawrence J. Cheskin, MD, FACP, FTOS, is Professor and Chair of the Department of Nutrition and Food Studies at George Mason University in Fairfax, VA, and an Adjunct Faculty Member of the Department of Medicine at the Johns Hopkins School of Medicine in Baltimore, MD, United States. He received his medical degree from Dartmouth and was a Postdoctoral Fellow in gastroenterology and liver diseases at Yale University-New Haven Hospital. Dr. Cheskin's work is at the intersection of public health nutrition and clinical medicine and has dedicated his career to research, education, and program building in the service of combating obesity, through both treatment and prevention. He founded the Johns Hopkins Weight Management Center, a multidisciplinary clinical research and treatment program. He was also Director of Clinical Research of the Global Obesity Prevention Center at Johns Hopkins and directed its Pilot Studies Core, which evaluated systems-focused proposals worldwide to study such areas as school policies and the built environment in preventing childhood obesity. He is co-PI of the Mason Cohort Study of entering college freshmen and has authored over 240 peer-reviewed journal articles and 8 books. He is committed to mentoring the next generation of public health and clinical scholars dedicated to nutrition and obesity.

Section 7: *Nutrition Epidemiology*
Dr. Manuel Franco
Associate Professor at the Department of Surgery, Medical and Social Sciences, University of Alcalá, Madrid, Spain, and adjunct faculty at the Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States.



Manuel Franco, MD, PhD, Epidemiologist, is an Associate Faculty at the University of Alcalá and an Adjunct Faculty at the Johns Hopkins Bloomberg School of Public Health.

He is the Principal Investigator of the European Research Council Starting Grant Heart Healthy Hoods (http://www.hhhproject.es) studying urban characteristics in relation to eating patterns, physical activity levels, smoking, and alcohol consumption. He is also the PI of the Participatory Project on food in the cities: Photovoice Villaverde (https://youtu.be/VIiFggKzVas).

Manuel Franco trained in Medicine both in Madrid and Berlin. As a Fulbright Scholar he joined the Johns Hopkins Bloomberg School of Public Health Department of Epidemiology to obtain his PhD and Postdoctoral Fellowship in the fields of social epidemiology and urban health.

His work focuses on the prevention of chronic diseases and their major risk factors as nutrition. He has published 94 international peer-reviewed articles and has led 15 studies as PI raising over 2.9 million € in competitive international research bids. His methodological interests include the measurement of urban characteristics related to chronic diseases and nutrition, the use of mixed methods, and the conduction of participatory action research methods.

Section 8: *Global aspects of Human Nutrition*
Professor Daniel J. Hoffman
Department of Nutritional Sciences and director of the Program in International Nutrition at Rutgers University, New Brunswick, NJ, United States



Daniel J. Hoffman, PhD, is a Professor in the Department of Nutritional Sciences and Director of the Program in International Nutrition at Rutgers University, New Brunswick, NJ, United States. He graduated from St. John's University, Collegeville, MN, United States, and received his PhD in Human Nutrition from the Tufts University Friedman School of Nutrition Science and Policy, Boston, MA, United States. For his doctoral training, Dr. Hoffman studied the metabolism of children living in the shantytowns of São Paulo, Brazil. He then completed a Postdoctoral Fellowship at Columbia University, New York, NY, United States, before joining the faculty at Rutgers University.

Dr. Hoffman's research focuses on the long-term impact of undernutrition early in life on body composition and metabolism later in life. This work is complemented by research on sustainable methods to improve diet and growth in lower- and middle-income countries, including Brazil, Mexico, Kenya, South Sudan, and Vietnam. He also directs the Center for Childhood Nutrition Research at the New Jersey Institute for Food, Nutrition, and Health where members conduct interdisciplinary research on childhood growth and development. Dr. Hoffman is an Associate Editor of the *Food and Nutrition Bulletin* and a Section Editor for the *Annals of Human Biology*. His other responsibilities include serving on the editorial board of the *Journal of Nutrition*, serving in leadership positions for the American Society for Nutrition, and advising international and national agencies on matters related to childhood nutrition. Aside from his academic life, Dr. Hoffman is a published documentary photographer whose portfolio includes work on the life of Roma in Europe, urban landscape of São Paulo, Brazil, gentrification of Times Square in New York City, and punk music as the "American Mosaic."

PREFACE

By the middle of last century, the science of nutrition had identified most of the essential nutrients and had provided evidence to propose specific dietary intake recommendations for many diet constituents, with the practical aim of preventing nutrient deficiencies. As chronic, noncommunicable diseases such as cardiovascular disease, cancer, etc., began to emerge as an important causal factor for disability and early death, scientists turned their attention to the potential effects of nutrients and diet patterns on chronic disease risk. Pioneering studies by Burkitt, Keys, Breslow, and others were followed by a large number of studies on the role of dietary patterns and constituents on certain chronic diseases. Many important studies were completed over the second part of the century, providing the evidence to support specific dietary recommendations to reduce disease risk.

The 21st century ushered the next transition in nutrition science, this time centered on the interrelationships between nutrients, dietary patterns, and the human genome. Over the past few decades, advances in our understanding of the human genome and on the molecular tools to explore it have permitted to probe those interactions in increasing detail. In turn, findings from nutrient–gene interaction studies have informed population-wide and clinical and metabolic studies, further advancing our understanding of the effects of diets on human health at the molecular level. This understanding of the links between genotype, phenotype, and nutrient/dietary intake became a key contributor to the emerging area of personalized nutrition/precision medicine.

All those phases of research emphasis, to different degree, continue to exist today and result in a vast, multidisciplinary, ever-expanding amount of information reaching the peer-reviewed literature. This massive amount of information needs to be organized and summarized in a way that makes it accessible to experts, teachers, and, as much as possible, the general public. This has been and continues to be the goal of the *Encyclopedia of Human Nutrition* since its first edition, over 20 years ago.

Such an ambitious task can only be achieved by the collective work of many people. We all have experienced the challenge of writing an article that combines focus and relevance with conciseness, so we are very appreciative of the work of our contributors. Their effort was backed up by an excellent editorial board, which reviewed and provided feedback on every manuscript. Finally, we must acknowledge the outstanding support of the Major Reference Works division at Elsevier. A publication like this *Encyclopedia* has a lot of moving parts, and it is a great privilege to be able to concentrate on the content, knowing that the other parts of the process are in the hands of excellent professionals.

We hope that this book will help satisfy the need for accurate and concise information to the many students and professionals who are committed to use nutrition science as a tool to improve people's quality of life.

Benjamin Caballero, MD, PhD
Editor In Chief

This page intentionally left blank

Adolescents: Nutrient requirements

C.H.S Ruxton, Freelance Dietitian, Cupar, UK
E. Derbyshire, Manchester Metropolitan University, Manchester, UK

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of C.H.S. Ruxton, E. Derbyshire, Adolescents: Requirements for Growth and Optimal Health, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 23–32, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00003-9.

Introduction 2
Physical Changes During Adolescence 2
Growth 2
Adipose Stores 3
Sexual Development 3
Dietary Recommendations 3
Dietary Intakes 4
Energy, Protein, and Salt 4
Fats 4
Carbohydrates and Fiber 8
Micronutrients 8
Impact of Lifestyle on Nutrition 9
Breakfast 9
Food Choices at School 9
Fruit and Vegetable Consumption 9
Snacking and Soft Drink Consumption 9
Smoking 9
Alcohol Consumption 10
Socioeconomic Status 10
Physical Activity 10
Sleep 11
Conclusions 11
Acknowledgments 11
Further Reading 11

Glossary

Cross-sectional study A study that provides a one-off observation of a population.
Growth percentile A way of comparing individual children to a reference group using percentages.
Longitudinal study A study that provides a series of observations of a population, or follows a specific cohort of individuals for a defined period of time.
Osteoporosis A progressive disease which causes bone loss, leading to fractures.
Reference nutrient intake A nutrient recommendation which would be expected to meet the needs of 97.3% of the population.
Underreporting Self-reporting of food intake that deliberately or unconsciously underestimates food, energy or nutrient intakes.

Abbreviations

BMI Body mass index
EAR Estimated average requirement
NMES Nonmilk extrinsic sugars
NSP Nonstarch polysaccharide
RNI Reference nutrient intake

Introduction

Adolescence is the period of transition between childhood and adulthood. This reflects not only the physical and emotional changes experienced by the adolescent, but also the development of dietary behaviors. Although younger children tend to resist new foods (i.e., neophobia), adolescents may use food to assert their independence, but not always in a beneficial way. This section covers development during adolescence and highlight nutrients that are important during this time. Information on adolescent energy and nutrient intakes from a broad range of countries is presented. The findings are put in context with dietary recommendations.

Physical Changes During Adolescence

Adolescence is generally assumed to be the period of human development from 10 to 18 years, a time during which rapid growth and physical maturity take place.

Growth

During prepubescent childhood, the growth of boys and girls follows a similar trajectory, although boys may be slightly taller and heavier than girls. The pubertal growth spurt begins around the ninth year in girls, lasting up to 3½ years, with boys starting their growth spurt approximately 2 years later. This means that girls can reach their full height 2 years before boys do. UK standards for height and weight during adolescence are presented in **Table 1**. International reference growth standards are developed by the World Health Organization (WHO). Maximum height velocity is generally seen in the year preceding menarche for girls and at around 14 years of age for boys. On average, weight velocity peaks at 12.9 years of age for girls and 14.3 years of age for boys. Annual growth rates during adolescence can be as much as 9 cm/8.8 kg in girls and 10.3 cm/9.8 kg in boys. It is not fully known when growth ceases. Certainly, height gains of up to 2 cm can still occur between 17 and 28 years of age.

Energy and protein intakes per kilogram body weight have been observed to peak during maximal growth, suggesting increased requirements during adolescence. Undernutrition in this crucial window of development can result in a slow height increment, lower peak bone mass, and delayed puberty. Overnutrition is not without its risks. It is believed that obesity in young girls can bring about an early menarche, which then increases the risk of breast cancer in adulthood. Menarche is deemed precocious if it occurs before the age of 8 years. Rising childhood obesity levels in Western countries have resulted in more girls experiencing precocious menarche.

Important nutrients for growth include protein, iron, calcium, vitamin C, vitamin D, and zinc. Calcium, in particular, has a key role in bone development, and huge increments in bone density are seen during adolescence under the influence of sex hormones. Bone density peaks in the early twenties and a low bone density at this time is related to an increased risk of osteoporosis in later life, especially for women. Studies have suggested that body mass index (BMI) in adolescence is the best predictor of adult bone density, explaining why children who experience anorexia nervosa are likely to have a higher risk of osteoporosis.

Table 1 Percentiles for height, weight and body mass index

(a) Boys

	Height (cm)			*Weight (kg)*			*Body mass index*		
Age (years)	*2nd*	*50th*	*98th*	*2nd*	*50th*	*98th*	*2nd*	*50th*	*98th*
11	130.1	143.4	156.7	25.8	34.6	52.6	13.99	16.89	22.58
16	158.2	173.4	188.6	44.5	60.6	85.7	16.34	19.94	26.73
18	163.1	177.1	191.1	51.7	68.8	90.3	17.24	21.05	28.02

(b) Girls

	Height (cm)			*Weight (kg)*			*Body mass index*		
Age (years)	*2nd*	*50th*	*98th*	*2nd*	*50th*	*98th*	*2nd*	*50th*	*98th*
11	130.3	144.1	157.9	25.6	36	55.4	14.02	17.48	23.88
16	151	163.2	175.4	42.3	55.5	75.8	16.37	20.44	27.76
18	151.5	163.6	175.6	44.2	57.5	78.4	16.99	21.19	28.62

For more information, contact the Child Growth Foundation: childgrowthfoundation.org.

Adipose Stores

There are few differences in body fat between boys and girls in the prepubertal stage. However, during puberty, girls develop adipose tissue at a greater rate than boys, laying down stores in the breast and hip regions. The pattern for boys is rather different and tends toward a more central deposition. Methods for estimating fatness in adolescents include weight for height, BMI (weight in kilograms/height in square metres), skinfold thickness measures, bioelectrical impedance analysis, densitometry, magnetic resonance imaging, dual–energy X-ray absorptiometry, and computed tomography. Waist circumference is gaining popularity as a useful proxy of fatness in the field. Many researchers argue that it is a better predictor than BMI of the central adipose stores that place the individual most at risk from later obesity, diabetes, and coronary heart disease.

UK standards for BMI and waist circumference are shown in **Table 1**. As adolescents are still growing, it is important to use age- and sex-appropriate appropriate standards to assess over- and underweight. The 85th percentile of BMI is often used as the lower cutoff point for classification of overweight, whereas the 5th is taken as an upper cutoff for underweight. Some surveys have suggested that adult obesity risk can be tracked from childhood, citing BMI at adolescence as a strong predictor of adult obesity. However, this information should be used with caution in practice because adolescents have not yet reached their full height and may still revert to a normal BMI without dietary intervention.

Sexual Development

In girls, the onset of menarche at around 13 years is triggered by the attainment of a specific level of body fat, with taller, heavier girls more likely to experience an early menarche. Vigorous exercise, for example, gymnastics and endurance running, can delay the menarche, both due to the physiological effects of regular training and the depletion of body fat. Iron becomes more important for girls as menstrual periods become regular and heavier, and there is good evidence that the iron status of many teenage girls is inadequate. Low iron status is due to a combination of higher requirements (i.e., menstrual periods and growth) and poor nutritional practices, such as dieting, missing breakfast; and avoiding red meat.

Dietary Recommendations

In general, each country has its own nutritional recommendations for adolescents, which are developed by expert bodies using a combination of deficiency studies and extrapolations from adult studies. In the UK, US, and Canada, guidelines have evolved from a simple Recommended Dietary Intake to the more complex bell-shaped distribution with a mean representing the intake likely to satisfy the needs of 50% of the population. The upper extreme, at the 97.5th centile, represents the intake likely to meet the needs of the majority of the population, whereas the lower extreme, at the 2.5th centile, represents the lowest acceptable

Table 2 UK Dietary guidelines for adolescents

(a) Dietary Reference Values for macronutrients

Age group	*Sex*	*Energy (MJ)*	*Protein (g)*	*NSP (g)*	*Total fat% energy*	*Saturated fat% energy*	*Starch/intrinsic sugars% energy*	*NMES% energy*
11–14 yr	M	9.27	42.1	18	35	11	39	11
	F	7.92	41.2	18	35	11	39	11
15–18 yr	M	11.51	55.2	18	35	11	39	11
	F	8.83	45.0	18	35	11	39	11

(b) Reference Nutrient Intakes for vitamins and minerals

Age group	*Sex*	*Vitamin B_2 (mg)*	*Vitamin B_2 (mg)*	*Niacin (mg)*	*Vitamin B_6 (mg)*	*Vitamin B_{12} (mcg)*	*Folate (mcg)*	*Vitamin C (mg)*	*Vitamin A (mcg)*	*Calcium (mg)*	*Iron (mg)*	*Zinc (mg)*
11–14 yr	M	0.9	1.2	15	1.2	1.2	200	35	600	1000	11.3	9.0
	F	0.7	1.1	12	1.0	1.2	200	35	600	800	14.8	9.0
15–18 yr	M	1.1	1.3	18	1.5	1.5	200	40	700	1000	11.3	9.5
	F	0.8	1.1	14	1.2	1.5	200	40	600	800	14.8	7.0

Key: y, years; MJ, megajoules; mg, milligrams; mcg, micrograms; %energy, percentage of food energy; NMES, nonmilk extrinsic sugars (similar to added sugars).

intake. Current UK Reference Nutrient Intakes (RNI), presented in Table 2, cover a range of nutrients from fat and sugars to the main micronutrients. Dietary guidelines are an important reference point for nutrition scientists and dietitians, but it must also be borne in mind that they relate to the average needs of populations, rather than to the needs of individuals.

As well as numerical recommendations, many nations have adopted more descriptive or visual methods of promoting the ideal diet. This makes sense as recommended nutrient intakes are poorly understood by the public and need to be put into context by health professionals. Communication tools such as the plate model, pyramid system, food groups, and traffic light systems can help to get healthy eating messages across to adolescents.

Dietary Intakes

It is often assumed that most adolescents in Western countries have a nutritionally inadequate diet yet, despite reported low intakes of some micronutrients in surveys, there is little evidence of widespread clinical deficiencies, or indications that adolescents are failing to achieve appropriate heights and weights. Iron is the exception, where mean intakes are low and clinical markers suggest deficiency across several age groups, particularly in girls. There is also justifiable concern about the general healthiness of diets eaten by 'at risk' subgroups of adolescents such as dieters, smokers, vegans, and those who regularly consume alcohol.

Mean daily intakes of energy and selected micronutrients from a selection of major international surveys of adolescents are presented in Table 3. Caution should be taken when interpreting data from dietary surveys because underreporting can be a feature of these. Selective underreporting, often focused on energy-dense or high-fat foods, can partially explain low reported intakes of energy and certain micronutrients. It is also difficult to make comparisons between the data from different countries given the range of dietary assessment methods used. There is normally a trade-off between the sample size and methodology that sees the larger surveys favoring less precise methods such as 24 h recalls or food frequency questionnaires in order to make data collection more economical.

Energy, Protein, and Salt

Despite height and weight data that are consistent with expected results, mean energy intakes appear to be low when compared with dietary recommendations in many studies of adolescents, particularly in lower income groups. Although low energy intakes can be of concern in individuals, on a population level, this phenomenon may be due to a number of reasons. These include dieting, low physical activity levels, and underreporting, where subjects subconsciously or deliberately misreport dietary intakes. Popular sources of energy in the adolescent diet included cereal products (providing around one-third of energy), savory snacks, potatoes, meat/meat products, white bread, milk/dairy products, biscuits/cakes, spreading fats, and confectionery. Beverages (i.e., soft drinks, juices, and alcohol) provide a significant amount of energy due to their popularity with adolescents and this has led to concern about their impact on obesity risk. In UK dietary surveys, beverages not including milk provided 9% of total daily calories in 11–18 years old.

In Western countries, average protein intakes are considerably in excess of requirements for all ages and both sexes. The main sources in adolescent diets are meat and meat products (contributing around one-third of protein), cereals, bread, and dairy products. It is believed that protein requirements in adolescents are between 0.8 and 1.0 g kg^{-1} body mass, although this does not take into account any additional needs that may relate to regular participation in sport and exercise. As a proportion of energy, protein intakes tend to be higher in Southern European countries, Australia, and New Zealand compared with intakes in the United States and Northern European countries.

High intakes of salt are a risk factor for abnormal blood pressure which, in turn, increases the risk of heart disease. Health experts believe that reductions in salt from childhood can help to maintain normal vascular health. UK dietary surveys of 11–18 years old report daily sodium intakes of approximately 2280 mg in girls and 2970 mg in boys. This equates to daily salt intakes of 5 and 6.5 g, respectively, with the target being less than 6 g day^{-1}. Meat products, such as burgers and sausages, provide around one-third of salt intakes with bread and processed foods also contributing significantly.

Fats

Average fat intakes as a proportion of energy vary considerably across Western countries. In the UK, intakes are close to the recommendation of 35% energy, suggesting increased public awareness as previous surveys reported intakes of 38–40% energy from fat. However, intakes of saturated fat, at 14% energy, still exceed the target of 11% food energy. It is worth noting that population averages hide subgroups with more extreme intakes. In the case of saturated fat, a considerable number of adolescents have intakes approximately 17% energy that may increase their risk of heart disease. Main sources of saturated fat in the adolescent diet include meat and meat products (approximately 20%), savory snacks and fried foods. Fat intakes in other European countries are 36–38% energy, with the highest fat intakes reported in Finland, Greece, Belgium, Germany, Switzerland, and Spain. In the United States, where the dietary target for fat is 30% of energy, intakes are approximately 32% energy from fat.

Not all dietary fats need to be reduced. There is increasing evidence that long-chain ω-3 fats from marine sources, specifically docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA), are beneficial for health. Rich sources of these are oily fish, such as salmon, trout, mackerel, and sardines, whereas shellfish also make an important contribution. Studies show that high intakes of oily fish, DHA, and EPA can help maintain heart health and cognitive function in adults. Emerging data suggest a role for

Table 3 Key international surveys of adolescent dietary intakes

Country Date carried out	Sex (age, years)	Energy (MJ)	Energy (kcal)	Protein % en	CHO % en	Sugars (g)	Fat % en	Iron (mg)	Calcium (mg)	A (mcg)	B_1 (mg)	B_2 (mg)	B_6 (mcg)	B_{12} (mcg)	Niacin (mg)	Folate (mcg)	Carbon (mg)
Australia 24 h DR 1995	M 12–15	11.59	2777	15.1	50.9	33.5	24.7	16.1	1093	1296	2.4	3.0	–	–	46.0	271	121
	M 16–18	13.53	3233	15.4	49.6	32.9	24.5	17.9	1280	1186	2.3	3.0	–	–	53.5	313	154
	F 12–15	8.53	2038	8.5	51.1	33.1	25.6	11.0	784	1130	1.5	2.0	–	–	33.4	206	124
	F 16–18	8.69	2076	8.7	50.1	32.1	24.0	11.1	801	877	1.5	1.8	–	–	35.3	217	126
Austria 7DUR, 24 h DR 1991, 2002	M 11–14	9.49	2268	13.2	48.2	–	35.2	13.0	903	–	1.4	1.6	1.5	5.7	–	229	113
	M 15–18	11.65	2784	12.9	50.0	–	37.2	15.4	1002	–	1.4	1.7	1.5	–	–	247	140
	F 10–14	9.49	2268	12.6	49.4	–	35.8	10.2	834	–	1.1	1.4	1.3	5.0	–	217	132
	F 15–18	8.49	2029	12.7	49.5	–	33.5	13.4	784	–	1.0	1.2	1.2	4.0	–	201	99
Belgium 3DUR, FFQ 1991, 1995	M 11–12	11.49	2746	11.6	–	–	–	–	–	–	–	–	–	–	–	–	–
	M 12–18	13.06	3122	13.0	48.6	149	37.2	13.4	913	–	1.5	1.7	1.6	–	–	–	83
	F 11–12	11.72	2802	11.6	–	–	–	–	–	–	1.0	–	–	–	–	–	–
	F 1–18	9.44	2256	14.9	48.8	112	36.7	8.2	805	–	1.2	1.3	1.2	–	–	–	78
Canada, 24 h DR 2004	M 9–13	10.28	2467	14.6	54.5	–	31.0	16.5	1219	–	2.1	2.4	1.8	4.6	–	–	157
	M 14–18	12.08	2901	15.2	52.7	–	31.5	19.1	1300	–	2.4	2.7	2.2	5.5	–	–	163
	F 9–13	8.49	2037	14.0	55.5	–	30.5	13.5	993	–	1.7	2.0	1.5	3.5	–	–	146
	F 14–18	8.53	2048	14.4	54.3	–	30.9	13.1	917	–	1.6	1.9	1.5	3.3	–	–	147
Denmark 7DUR 1995	M 11–14	10.90	2605	–	51.0	–	35.0	–	1286	–	1.5	2.2	1.7	6.7	27.0	304	79
	M 15–18	12.15	2903	14.0	–	–	35.0	–	1362	–	1.5	2.3	–	7.1	30.0	295	80
	F 11–14	8.70	2079	–	51.0	–	34.0	–	1061	–	1.1	1.7	1.4	5.1	23.0	238	72
	F 15–18	9.70	2318	14.0	–	–	34.0	–	1121	–	1.2	1.8	1.5	5.5	23.0	266	79
Finland 48 h DR, 2007	M 13–14	8.3	1978	16.7	53.2	–	30.0	10	1273	524	1.3	2.2	2.1	5.3	28	203	87
	F 13–14	6.7	1602	16.0	54.1	–	29.8	8.9	1032	537	1.1	1.7	1.6	3.9	22	190	93
France DH, 1DWR 1988, 1993–4	M 10–13	–	–	–	47.8	142.5	–	–	1250	–	1.2	2.1	1.7	11.0	–	–	88
	M 11–14	10.83	2587	15.4	–	–	36.5	12.6	835	–	1.4	1.8	1.8	5.6	–	253	91
	M 11–18	–	–	15.7	–	–	–	–	–	–	1.0	2.2	–	–	–	–	–
	M 13–18	12.10	2892	14.9	48.8	126.8	36.0	12.5	1300	–	1.4	–	2.0	7.0	–	–	127
	F 10–13	–	–	–	47.7	113.3	–	–	1100	–	1.0	1.8	1.5	7.5	–	–	99
	F 11–14	8.84	2112	15.9	–	–	–	11.4	835	–	–	1.8	1.8	5.6	17.0	253	91
	F 11–18	–	–	16.1	–	–	–	–	–	–	1.3	–	–	–	–	–	–
	F 13–18	9.16	2188	16.1	45.7	98.2	–	10.4	1100	–	–	1.7	1.4	7.0	–	–	112

(Continued)

Table 3 Key international surveys of adolescent dietary intakes—cont'd

Country Date carried out	*Sex (age, years)*	*Energy (MJ)*	*Energy (kcal)*	*Protein % en*	*CHO % en*	*Sugars (g)*	*Fat % en*	*Iron (mg)*	*Calcium (mg)*	*A (mcg)*	*B_1 (mg)*	*B_2 (mg)*	*B_6 (mcg)*	*B_{12} (mcg)*	*Niacin (mg)*	*Folate (mcg)*	*Carbon (mg)*
Germany DH, 3–d/7–d Recall, 1DWR 1985–95, 1998	M 10–12	9.08	2170	12.9	46.0	–	38.0	12.4	795	–	1.1	1.2	1.3	4.4	24.0	221	87
	M 13–14	10.41	2487	13.2	45.6	–	37.5	14.3	893	–	1.3	1.5	1.4	5.5	28.6	245	98
	M 15–18	11.14	2661	13.2	46.9	–	38.3	14.8	902	–	1.4	1.7	1.6	5.7	30.4	263	97
	F 10–12	7.78	1860	12.9	47.6	–	36.4	11.1	681	–	1.0	1.2	1.1	4.0	20.4	203	87
	F 13–14	8.49	2028	12.6	45.1	–	39.2	12.2	754	–	1.1	1.1	1.2	4.2	24.1	210	98
	F 15–18	8.59	2052	13.1	46.5	–	37.0	12.3	728	–	1.0	1.3	1.5	4.4	24.6	216	92
Greece 1DWR, 24 h DR 1993–4, 1999	M 10–11	–	–	15.7	44.0	–		11.0	963	–	–	–	–	–	–	–	119
	M 12–14	8.90	2126	15.6	45.0	97	40.0	13.5	1011	–	1.7	2.1	1.8	4.7	18.0	226	112
	M 14–16	9.00	2151	15.0	47.2	–	–	13.8	871	–	2.5	2.0	1.9	4.3	19.2	251	123
	F 10–11	–	–	15.6	44.0	–	41.8	10.0	851	–	–	–	–	–	–	–	108
	F 12–14	9.70	2318	14.7	48.0	88	–	10.1	748	–	1.4	1.6	–	4.1	–	212	108
	F 14–16	7.08	1692	14.5	46.0	–		9.4	771	–	2.4	1.5	1.3	3.2	13.2	217	118
Ireland DH 1988	M 11–14	11.3	2700	–	50.3	–	36.3	14.7	1208	–	1.8	2.5	2.2	4.9	40.2	246	76
	M 15–17	14.0	3346	14.2	49.3	–	36.0	19.3	1549	–	2.2	3.1	2.6	7.2	51.7	306	95
	F 11–14	9.10	2174	–	50.2	–	36.0	–	962	–	–	1.9	1.7	3.9	32.0	198	76
	F 12–15	–	–	13.9	48.9	–	37.1	12.4	–	–	1.4	–	–	–	–	–	–
	F 15–17	8.90	2127					11.6	950	–	1.3	1.8	1.6	4.0	32.0	182	79
Japan UR n/a	M 15–19	10.6	2545	14.4	51.9	–	28.3	8.6	633	978	1.2	1.4	1.3	8.1	16.6	303	89
	F 15–19	8.0	1918	15.1	50.4	–	29.7	7.4	516	875	0.9	1.2	1.1	6.4	13.2	268	91
Netherlands 2DUR 1997–8	M 13–16	10.9	2605	13.1	51.2	188	35.5	10.9	1045	778	1.2	1.6	1.6	3.9	–	–	79
	M 16–19	11.6	2772	13.3	49.5	184	35.4	11.5	1095	972	1.3	1.6	1.8	4.4	–	–	71
	F 13–16	8.7	2079	13.7	50.3	146	35.9	9.0	904	724	1.0	1.4	1.3	3.4	–	–	81
	F 16–19	9.1	2175	13.4	50.3	152	35.5	9.9	908	754	1.2	1.4	1.4	3.4	–	–	81
New Zealand 24 h DR, FFQ 1997	M 15–18	12.4	2963	15.0	49.0	82	35.0	15.2	957	505	1.8	2.1	1.8	4.9	43.0	280	155
	F 15–18	8.86	2117	14.0	51.0	69	34.0	10.4	783	342	1.3	1.5	1.1	3.2	28.0	203	120
Norway 1DWR, FFQ n/a	M 11–14	–	–	–	–	–	31.1	–	–	–	–	–	–	–	–	–	–
	M 13–14	15.0	3585	–	–	–	–	–	1625	–	2.1	2.8	–	–	–	–	110
	M 13–15	–	–	13.4	–	–	–	–	–	–	–	–	–	–	–	–	–
	F 11–14	10.9	2605	–	54.9	–	28.9	–	–	–	–	–	–	–	–	–	–
	F 13–14	–	–	–	–	–	–	–	1142	–	1.6	2.1	–	–	–	–	104
	F 13–15	–	–	13.7	–	–	–	–	–	–	–	–	–	–	–	–	–

Table 3 Key international surveys of adolescent dietary intakes—cont'd

Country Date carried out	Sex (age, years)	Energy (MJ)	Energy (kcal)	Protein % en	CHO % en	Sugars (g)	Fat % en	Iron (mg)	Calcium (mg)	A (mcg)	B_1 (mg)	B_2 (mg)	B_6 (mcg)	B_{12} (mcg)	Niacin (mg)	Folate (mcg)	Carbon (mg)
Portugal 24 h DR, 1995	M 12–18	8.86	2117	–	49.1	–	–	–	890	–	–	–	–	–	–	–	–
	M 13–17	9.41	2248	17.6	–	–	–	–	–	–	–	–	–	–	–	–	77
	F 12–18	9.40	2248	–	53.4	–	33.3	–	853	–	–	–	–	–	–	–	–
	F 13–17	8.14	1945	17.8	–	–	–	–	–	–	–	–	–	–	–	–	99
Spain 24 h DR 1998–2000	M 10–13	9.63	2302	–	–	–	–	15.1	1010	521	1.5	1.9	1.9	8.3	24.8	162	70.0
	M 14–17	10.7	2565	–	–	–	–	16.6	1031	551	1.6	1.9	2.0	8.9	26.2	178	76.3
	F 10–13	8.15	1949	–	–	–	–	12.7	862	453	1.3	1.6	1.5	6.7	20.1	140	69.7
	F 14–17	8.28	1979	–	–	–	–	12.5	823	415	1.3	1.5	1.5	7.2	21.1	149	73.7
Sweden 7DUR 1989–90, 1993–4	M 13–14	–	–	–	–	–	–	17.4	1279	–	–	–	–	–	–	–	–
	M 14–16	8.90	2127	–	52.6	–	32.1	18.2	1406	–	1.8	2.4	2.0	6.6	33.5	178	68
	M 17–18	10.50	2509	14.7	–	–	–	–	1472	–	1.8	2.8	2.2	8.7	36	138	77
	F 13–14	–	–	–	49.4	–	–	–	1061	–	–	–	–	–	–	–	–
	F 14–15	7.21	1722	–	–	–	–	13.4	1046	–	1.4	1.8	1.5	4.9	24.9	144	68
	F 17–18	7.88	1884	14.2	54.1	–	–	13.3	966	–	1.2	1.8	1.5	5.5	23.0	105	77
Switzerland 7DUR 1994–5	M 11–12	–	–	13.3	46.1	–	–	–	–	–	–	–	–	–	–	–	–
	M 13–14	11.98	2863	–	–	–	40.1	16.0	1311	–	1.5	2.2	–	–	–	–	185
	M 15–18	12.56	3001	–	–	–	35.0	–	1157	–	1.3	1.8	–	–	–	–	163
	F 11–12	–	–	–	49.4	–	–	–	–	–	–	–	–	–	–	–	–
	F 13–14	7.90	1887	–	–	–	37.4	9.3	819	–	–	1.3	–	–	–	–	110
	F 15–18	8.12	1939	–	–	–	35.8	–	832	–	1.5	1.3	–	–	–	–	146
Turkey 3DUR, 2004	M 12–14	8210	1961	16.8	49.3	–	34.3	11.2	715	428	0.8	1.4	1.2	–	–	121[a]	86.3
	M 15–17	8880	2122	17.6	48.8	–	33.3	12.2	732	488	0.8	1.5	1.2	–	–	130[a]	76.1
	F 12–14	6960	1664	15.2	50.4	–	34.4	9.6	632	430	0.7	1.1	1.2	–	–	218[a]	85.0
	F 15–17	6530	1560	15.4	49.7	–	34.8	8.8	600	409	0.6	1.1	0.9	–	–	206[a]	69.3
USA 24 h DR 2007–8	M 12–19	60.6	2424	15	52	152	33.0	16.6	1173	680	1.9	2.6	2.3	6.7	28.9	198[b]	86.6
	F 12–19	7.75	1861	14	54	116	33.0	13.8	878	528	1.5	1.8	1.6	4.1	20.8	154[b]	73.8

Key: CHO, carbohydrate; DR, Dietary recall; DH, Diet history; FFQ, Food frequency questionnaire; UR, Unweighed record; WR, Weighed record.

Vitamin A=micrograms retinol equivalent.

Dates refer to when the surveys were carried out. For some countries, data refer to a combination of more than one survey.

[a]Folic acid data only.

[b]Folate from food sources only.

DHA and EPA in brain development in infants, and maintenance of the normal cognitive function in school-age children. Adolescents often have poor intakes of oily fish and, unless offered alterative sources such as marine oil supplements, may miss out on the benefits offered by DHA and EPA.

Carbohydrates and Fiber

Average carbohydrate intakes in Western populations are close to the recommendation of 50% energy. The main sources in the adolescent diet are breakfast cereals, bread, savory snacks, vegetables and potatoes. Fiber intakes, expressed as nonstarch polysaccharide (NSP), are 10–13 g per day in the UK, which is approximately 70% of the adult guideline. Vegetables, potatoes, and savory snacks together contribute 40% of NSP. Interestingly studies have not always found consistent relationships between fiber intakes and bowel movements in young people. Sugar intakes peak in adolescence before declining to adult levels and are higher in boys compared with girls. In UK surveys, mean intakes of added sugars are 16% of energy and the largest contributor is nondiet soft drinks, providing 40% of sugars, followed by confectionery and preserves. Children from lower income households tended to have lower intakes of protein and NSP but higher intakes of sugars and fats compared with their more affluent peers.

Recommendations to reduce fat are often accompanied by those urging a decrease in added sugars due to concerns about obesity, dental health, and micronutrient adequacy. However, inverse relationships between fat and sugars are often seen in observational surveys suggesting that reductions in dietary fat may occur, in part, due to increases in foods containing added sugars. Most cross-sectional surveys also report inverse relationships between sugar intakes and BMI, suggesting that leaner people have higher sugar intakes, but lower fat intakes. This may seem counterintuitive given concerns about the impact of sugar on obesity risk; however, an explanation could be that heavier people either restrict their sugar intake or underreport sugar-containing foods. Nevertheless, apart from one intervention study relating to soft drinks, the available evidence does not support a relationship between higher sugar intakes and an increased risk of obesity. This was confirmed by the European Food Safety Authority in 2010 following a review of the evidence.

Several reviews have also considered the evidence linking sugar consumption with diet adequacy, responding to concerns that higher sugar intakes result in diets, which are low or inadequate in vitamins and minerals. The findings, based on surveys of children and adolescents in the United States and Europe, suggest that a broad range of sugar consumption is consistent with adequate micronutrient intakes. This may be due to fortification of popular sugar-containing foods, for example breakfast cereals and dairy foods. Lower levels of vitamins and minerals tend to be seen only at the upper and lower extremes of sugar consumption, suggesting that these diets lack variety.

Micronutrients

The main contributors to sources of vitamins and minerals are breakfast cereals, milk, bread, chips/potatoes, and eggs. Surveys that report comparisons between intakes and recommendations have found satisfactory intakes for most micronutrients when population averages are considered. Intakes of vitamins B_1, B_2, B_6, B_{12}, vitamin C, and niacin greatly exceeded recommendations in the UK, perhaps reflecting high-protein intakes and the fortification of breakfast cereals, bread, and beverages. Average intakes that fall below recommendations have been seen for iron, calcium, magnesium, and zinc. In girls, mean selenium and iodine intakes can also be lower than recommendations.

As mentioned earlier, averages hide subgroups of adolescents whose nutrient intakes fall below acceptable levels for health. UK data suggest that 3–47% of adolescents have intakes of vitamin A, iron, calcium, magnesium, potassium, zinc, selenium, and iodine that fall below the population minimum, suggesting that they may be at risk from deficiency. The nutrients of most concern in adolescence are iron and calcium, which are important for the normal growth and development. Intakes of folate are particularly important for girls, some of whom may fall pregnant in their teens. Certain practices, such as smoking and drinking alcohol, can increase requirements for micronutrients, suggesting that specific groups of adolescents may be more at risk from a poor nutritional status.

Mean iron intakes are particularly low in 11–18 year old girls, at 58–63% of the recommendation (see **Table 4**), reflecting avoidance of iron-rich foods, such as red meat and offal. It is important to address low iron intakes because these can lead to poor iron status and, in some case, anemia. The best source of iron is red meat which provides the easily absorbed form, called heme iron. However, most iron in the adolescent diet comes from foods, such as breakfast cereals, which supply the less-well-absorbed nonhaem iron. Iron absorption can be improved by consuming vitamin C rich foods or beverages at meal times.

An adequate calcium intake during childhood and adolescence is important for establishing an optimal peak bone mass that can be maintained throughout adulthood. Low intakes of calcium are a risk factor for osteoporosis in later life. Although average calcium intakes tend to be close to recommended levels, there are groups of adolescents with intakes below adequate levels. In UK 11–14 year olds, 12% of boys and 24% of girls had intakes below the lower reference nutrient intake, while in 15–18 year olds, the figures were 9% and 19%, respectively. This suggests a risk of deficiency. Good sources of calcium are milk, cheese, yogurt, soya products, tinned fish and, in many countries, fortified grain products.

Concern has been expressed that the rise in soft drink consumption has displaced milk from the diets of adolescents and this could be contributing to the low calcium intakes found in many surveys. Fluid milk consumption has fallen dramatically over the last decade in Western countries and this is due to a range of factors including preference for other beverages, concerns about

weight management, and the perception that milk is for younger children. Although improving diet is important, it should not be forgotten that physical activity is also important for optimal bone health.

Impact of Lifestyle on Nutrition

Young people consume particular foods and diets for a variety of reasons, mostly unrelated to their nutritional content. These include weight control (whether justified or not), peer group pressure, celebrity endorsement, convenience, personal ideologies (e.g., veganism), or enhancement of sporting prowess. As energy and nutrient intakes are influenced by eating patterns, it is important to consider lifestyle when interpreting dietary information or development health promotion messages.

Breakfast

Breakfast can be a nutrient-dense, low fat meal, yet is often omitted by adolescents. Around 10% of younger children miss breakfast, rising to 20% as adulthood is approached. Boys are more likely than girls to eat breakfast, and favor cereals rather than bread or cooked foods. Data on breakfast habits reveal higher intakes of sugars, fiber and micronutrients, such as folate, niacin, iron, calcium, and zinc, amongst regular consumers of breakfast cereals. Fat intakes, as a proportion of energy, are lower when breakfast cereals are consumed. Surveys of adolescents have found an inverse relationship between breakfast cereal consumption and body mass index, suggesting that eating breakfast is a useful strategy for weight control.

Food Choices at School

Although the popularity of school lunches has diminished over the last 10 years, they are still eaten regularly by many children, particularly those from lower socioeconomic groups. School lunches have been criticized in the past for containing a high proportion of fat and delivering low levels of vitamins, minerals, and ω-3 fatty acids. The introduction of school meal standards in a number of countries has significantly improved the nutritional content of school lunches. However, this has not always benefited adolescents who often prefer to assert their independence by buying food out with the school environment. Foods purchased from cafes and take-aways tend to be less healthy than the meals offered at school, and opportunities to choose fruits and vegetables are few.

Fruit and Vegetable Consumption

Data on intakes of fruit and vegetables show that adolescents, particularly girls, have lower intakes than adults and younger children. In the UK, where the recommendation is 400 g (expressed as five portions of fruit/vegetables per day), average intakes in 15–18 year old are only 200 g. Indeed, only 22% of boys and 7% of girls meet the five portions-a-day target.

Snacking and Soft Drink Consumption

There has been a general shift over the last few decades toward more meals eaten outside the home and a greater proportion of daily energy consumed as snacks and soft drinks. Concerns about the possible impact of snacks on diet quality and the risk of obesity are not always borne out by the evidence, although dietary assessment is hindered by a lack of consensus on what constitutes a 'snack'. Observational studies have found that frequent snackers have similar nutrient intakes to those who snack infrequently. With respect to body size, snacking is often associated with a lower, rather than a higher, BMI. The few intervention studies that have examined snacking behavior report full or partial compensation for the additional calories provided by snacks by a reduction in the energy from meals. This suggests that snacking itself is not harmful but that adolescents should choose their snacks wisely, focusing on those that make some contribution to micronutrient intakes. Intervention studies looking at the impact of sugar-sweetened soft drinks give a different picture and tend to find less compensation for the additional calories and a reduction in nutrient density of the diet. This suggests that high intakes of sweetened soft drinks, as are often found in adolescence, pose a risk for obesity and could be detrimental for diet adequacy.

Smoking

The proportion of adolescent smokers rises with age and is between 8 and 20% with an average exposure, in older children, of around 40 cigarettes per week. Since the 1980s, smoking has decreased in adolescent boys but not in girls. Smokers tend to have different dietary habits from nonsmokers and this is reflected in their nutrient intakes. Studies have found that smokers consume fewer dairy foods, wholemeal bread, fruit, breakfast cereals, and more coffee, alcohol, and chips. Smokers' diets tend to be lower in fiber, vitamin B_1 and vitamin C compared with nonsmokers. In a study of 18 year olds, male smokers had a higher percentage energy from fat and lower intakes of sugars and iron. Contrary to beliefs, there is no evidence that smoking helps to control body weight in young people.

Table 4 Average daily intakes of UK adolescents from UK National Diet and Nutrition Surveys

Sex (age) *Sample size*	*Energy* *MJ*	*Protein* *% en*	*CHO%* *en*	*NMES%* *en*	*Fat%* *en*	*NSP* *(g)*	*Iron* *(mg)*	*Calcium* *(mmg)*	*A (mcg)*	*B_1 (mg)*	*B_2 (mg)*	*B_6* *(mcg)*	*B_{12}* *(mcg)*	*Niacin* *(mg)*	*Folate* *(mcg)*	*Carbon* *(mg)*
National Diet and Nutrition Survey (Bates, *et al.*, 2010)[a]																
M 11–18 yr	9.07	14.8	50.7	16.3	34.5	13.1	11.1	919	776	1.69	1.72	2.6	4.8	39.4	256	94.4
N=114	–						98% RNI	92% RNI	120% RNI	170% RNI	137% RNI	192% RNI	359% RNI	238% RNI	128% RNI	252% RNI
F 11–18 yr	7.02	14.4	49.4	14.8	35.4	10.8	8.5	702	619	1.26	1.28	2.1	3.9	30.7	193	72.4
N=110	–						58% RNI	88% RNI	103% RNI	169% RNI	117% RNI	191% RNI	289% RNI	237% RNI	97% RNI	194% RNI
Low Income National Diet and Nutrition Survey (Nelson, *et al.*, 2007)[b]																
M 11–18 yr	9.36	13.1	50.5	17.2	36.4	12.6	11.4	913	625	1.82	1.70	2.3	4.6	35.7	232	74.5
N=200	93% EAR						101% RNI	91% RNI	98% RNI	187% RNI	137% RNI	176% RNI	346% RNI	220% RNI	116% RNI	201% RNI
F 11–18 yr	7.85	13.3	50.4	16.3	36.3	11.5	9.3	723	568	1.47	1.30	1.9	3.7	29.1	201	78.0
N=215	97% EAR						63% RNI	90% RNI	95% RNI	201% RNI	118% RNI	181% RNI	286% RNI	229% RNI	101% RNI	213% RNI

Key: CHO, carbohydrate; EAR, Estimated Average Requirement; RNI, Reference Nutrient Intake; NMES, Non–milk extrinsic sugars (similar to added sugars); NSP, Non starch polysaccharide.

[a]Bates B, Lennox A, and Swan G (2010) *National Diet and Nutrition Survey Headline results from year 1 of the rolling program (2008–09).* London: Food Standard Agency and the Department of Health.

[b]Nelson M, Erens B, Bates B, *et al.* (2007) *Low Income Diet and Nutrition Survey. Three Volume Survey, Executive Summary.* London: The Stationary Office.

Alcohol Consumption

In the UK, alcohol is consumed by 10% of 11–14 years old, and 37–46% of 15–18 year olds with older boys most likely to drink alcohol. Other European surveys have found higher proportions, 60–90% in 14–18 year old males, Although US surveys have found similar proportions to the UK. The average contribution of alcohol to energy intakes in the NDNS is just over 1%, with higher contributions reported by Danish and Irish studies (around 2–5% energy). Excess alcohol intake can increase micronutrient requirements but few younger adolescents fall into this category. However, regular consumption of alcohol contributes to obesity because the energy provided by alcoholic drinks rarely displaces energy from other food sources.

Socioeconomic Status

Differences in diet are sometimes seen between children from different social classes or income groups. In the UK, children from a lower socioeconomic background consume fewer low fat dairy foods, fruit juice, salad vegetables, high fiber cereals, fruit juices, and fruit than children from a higher socioeconomic background. This impacts on mean daily nutrient intakes with poorer children consuming lower amounts of protein, sugars, carbohydrate, vitamin C and fiber. Some surveys have also found higher fat intakes and a greater risk of obesity in children from lower socioeconomic backgrounds.

Physical Activity

Regular physical activity impacts on nutrition and health in a number of ways, for example, by maintaining normal energy balance that lowers the risk of obesity, by supporting the normal heart function, and by promoting bone density. However, adolescents, particularly girls, have become far less physically active in recent years. The European Youth Heart Study found that 82% of 15-

Table 5 Average daily intakes of adolescents from the European Nutrition & Health Report (2009)[a]

Sex (age)	*Protein % en*	*CHO % en*	*Sucrose% en*	*Fat % en*	*Fiber (g)*	*Iron (mg)*	*Calcium (mg)*	*Vit D (mcg)*	*Folate (mcg)*	*Iodine (mcg)*	*Sodium (g)*
M 1–15 yr	12–17	43–56	5–29	28–41	9–24	5.7–14.0	554–1104	1.4–5.3	116–304	48–299	1.6–6.3
F 1–15 yr	12–17	42–55	5–29	28–42	6–21	5.4–12.1	560–1049	1.2–6.3	109–278	51–299	1.5–5.4
M 14–24 yr	13–18	42–54	13–16	31–40	14–26	11.8–16.3	675–1362	1.3–5.4	175–312	93–133	2.4–4.1
F 14–24 yr	12–17	42–55	12–18	29–40	14–22	8.9–12.8	659–1212	1.5–3.4	161–266	78–106	2.2–3.2

% en, as percentage of total daily energy.

CHO, carbohydrate.

[a]Data presented as ranges.

year old boys, but only 62% of girls of that age, met physical activity targets. At present, WHO recommends that children and young people are physically activity for at least 60 min per day^{-1}. This includes sports and exercise, but also walking. Reducing sedentary behaviors, for example, TV viewing, video games, is also important for obesity prevention.

Sleep

Although sufficient sleep is vital for normal growth and cognitive development in childhood and adolescence, short sleep duration can also impact on nutritional intake and obesity risk. Studies in adolescents and young adults have found associations between a lack of nightime sleep and increases in BMI. Short sleep duration could influence body weight by disrupting appetite control, reducing physical activity levels (e.g., due to tiredness), or affecting thermoregulation.

Conclusions

Most adolescents in Western countries are consuming adequate energy and protein to support normal growth. However, intakes of micronutrients in subgroups of the population are lower than recommendations and may not be sufficient for optimal health. These include iron, calcium, zinc, folate, and vitamin A. For iron, there is good evidence of clinical deficiency in low iron consumers, particularly girls. As energy intake is the best predictor of micronutrient adequacy, care must be taken when advising adolescents about weight management because calorie restriction can reduce intakes of vitamins and minerals.

Longitudinal studies that attempt to link early diet with the incidence of later disease suggest that high intakes of fruit, vegetables, wholegrains, and oily fish are markers of good health in later life. However, this beneficial dietary pattern has not been adopted by most adolescents. Indeed, despite the public health efforts of government and health professionals, adolescents remain a hard to reach group. (Table 5).

Acknowledgments

The authors are very grateful to the dietitians and nutritionists around the world who provided data, both published and unpublished, for Table 3.

Further Reading

Bates, B., Lennox, A., Swan, G., 2010. National Diet and Nutrition Survey Headline Results from Year 1 of the Rolling Program (2008-09). Food Standards Agency and the Department of Health, London.

Elmadfa, I., Meyer, A., Nowak, V., et al., 2009. European nutrition and health report 2009. Annals of Nutrition and Metabolism 55 (Supplement 2), 1–40.

Lambert, J., Agostoni, C., Elmadfa, I., et al., 2004. Diet intake and nutritional status of children and adolescents in Europe. British Journal of Nutrition 92 (Supplement 2), S147–S211.

Nelson, M., Erens, B., Bates, B., et al., 2007. Low Income Diet and Nutrition Survey. Three Volume Survey, Executive Summary. The Stationary Office, London.

Prentice, A., Schoenmakers, I., Laskey, M.A., et al., 2006. Nutrition and bone growth and development. Proceedings of the Nutrition Society 65, 348–360.

Ramakrishnan, U., Imhoff-Kunsch, B., DiGirolamo, A.M., 2009. Role of docosahexaenoic acid in maternal and child mental health. American Journal of Clinical Nutrition 89, 958S–962S.

Ruxton, C.H.S., Gardner, E.J., McNulty, H.M., 2010. Is sugar consumption detrimental to health? A review of the evidence 1995–2006. Critical Reviews in Food Science & Nutrition 50, 1–19.

Shepherd, J., Harden, A., Rees, R., et al., 2006. Young people and healthy eating: A systematic review of research on barriers and facilitators. Health Education Research 21, 239–257.

USDA (United States Department of Agriculture) database (2010) *What we eat in America NHANES 2007-8.* Available at: http://www.ars.usda.gov accessed August 2010.

World Health Organization. Child Growth Standards. Available at: http://www.who.int/childgrowth/en/ (Accessed February 2011).

Aging

P. Hyland and Y. Barnett, Nottingham Trent University, Nottingham, UK
L.H. Allen, ARS Western Human Nutrition Research Center, Davis, CA, USA

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of P. Hyland, Y. Barnett, L.H. Allen, Aging, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 33–39, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00004-0.

Introduction 12
Theories of Aging 13
Programed and Genetic Theories 13
Longevity Genes 13
Accelerated Aging Syndromes 14
Neuroendocrine Theories 14
Immunologic Theory and Immunosenescence 14
Cellular Senescence 14
Disposable Soma Theory 14
Damage Accumulation (Stochastic) Theories 14
Somatic Mutation and DNA Repair 15
Error Catastrophe 15
Cross-Linking 15
Free Radicals 15
Mitochondrial DNA Damage 15
Age-Related Diseases 16
Modification of the Aging Process 17
Nutritional Modification 17
Vitamins and Micronutrients 18
Dietary Energy Restriction 19
Molecular Biological Interventions and the Aging Process 19
Further Reading 19

Glossary

Aging A progressive sequence of detrimental age-related changes that occur in every individual of a given species, at varying rates. These changes lead to a breakdown in normal homeostatic mechanisms so that the functional capacity of the body and its ability to respond to a wide variety of extrinsic and intrinsic agents is often reduced. This causes degradation of structural elements within the cells, tissues, and organs of the body, leading eventually to the onset of age-related disorders and ultimately death.

Apoptosis Programed cell death which is important for changes that occur during development, and for cell turnover. Approximately 50–70 billion cells per day die through apoptosis in the human adult. Apoptosis is a normal event and differs from necrosis, which is cell death due to injury. Insufficient atrophy results in excessive cell proliferation such as in cancer.

Hayflick phenomenon (or Hayflick limit) The number of times a cell population can divide before it dies.

Senescence The process of becoming old.

Somatic mutation Changes in DNA that can occur in any cell except sperm or eggs. They cannot be passed on to the next generation, but can cause cancer or other diseases.

Introduction

The aging processes, and interventions to ameliorate them, have fascinated humans since the dawn of civilization. Research into aging is now a vital area of human endeavor, as our species reaches the limits of its longevity and faces the prospect of an aging population.

An individual's life expectancy is contributed to by the interaction of intrinsic (genetic and epigenetic) factors with extrinsic (environmental and life style) factors (**Figure 1**). In the world's more developed countries (MDCs) life expectancy at birth in the 1900s was around 47 years. By the end of the twentieth century this rose to a mean of 78 and 76 years in Western Europe and North America, respectively, with many individuals living much longer. This dramatic increase in average life expectancy has been largely

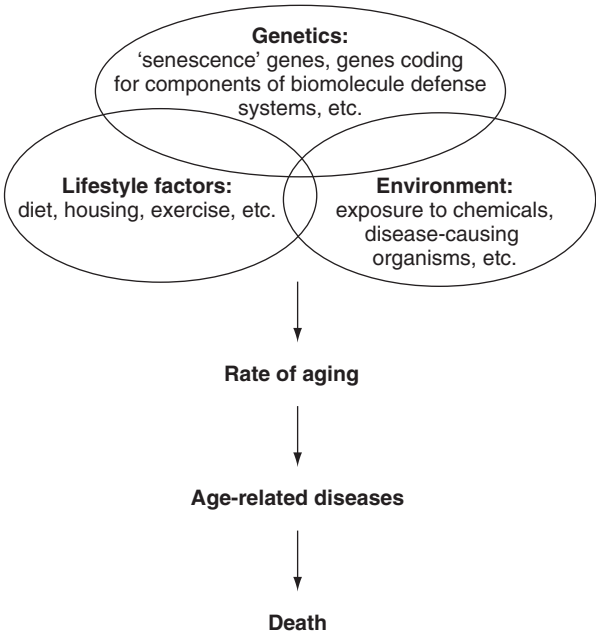


Figure 1 Interactive factors that contribute to the aging process. Reproduced from Barnett YA (1994) Nutrition and the ageing process. *British Journal of Biomedical Sciences* 51: 278–287.

due to improvements in environmental conditions such as nutrition, housing, sanitation, and medical and social services, leading to a large increase in the number of older people around the world. This change in the age structure of society is compounded by the decreasing fertility levels in the world's populations leading to large gains in worldwide median population ages. Our aging populations have a growing number and proportion of older people and, importantly, a growing number and proportion of very elderly people.

Based on current rates and trends in population growth by the year 2025 the elderly population (aged 65 and above) in the world's MDCs will increase by more than 50%, and will more than double worldwide. The very elderly (aged 80 and above) is the fastest growing section of the elderly population. This changing demographic picture will result in a large increased prevalence worldwide of long-term illness, disability, and the degenerative diseases associated with aging. These alterations in the proportions of the population of working age and those beyond working age will significantly impact the funding and costs of healthcare for all nations, making research into aging of critical international importance.

Theories of Aging

The biological manifestations that occur with aging affect the entire hierarchical structure of living systems. Age-related effects are seen in the accumulation of damaged cellular biomolecules (e.g., advanced glycosylation end products, lipid peroxidation products, genetic damage, and mutation), damaged organelles (mitochondria), and loss of cellular function, which contributes to dysfunction of the body's tissues, organs, and systems. These hierarchical changes have paved the way for more than 300 theories in an attempt to explain how and why aging occurs. They have previously been broadly categorized into: (1) programed or genetic theories; and (2) damage accumulation (stochastic) theories. However, these categories are not proven to be entirely comprehensive or mutually exclusive and it is likely that there is a shifting range throughout the life span that reflects a decreasing influence of genetic factors and an increasing influence of stochastic events.

Programed and Genetic Theories

Programed and genetic theories propose that the process of aging follows a biological timetable, perhaps a continuation of the one that regulates childhood growth and development. There are a number of lines of evidence supporting these theories.

Longevity Genes

It is clear that aging is controlled to some extent by genetic mechanisms. The distinct differences in life span among species are a direct indication of genetic control, at least at the species level. A number of genes have been identified in yeast, nematode worms (*Caenorhabditis elegans*), and fruit flies (*Drosophila melanogaster*) that significantly increase the organism's potential maximum life span. The products of these genes are involved in stress response and resistance, development, signal transduction, transcriptional

regulation, and metabolic activity. However, the genetics of longevity have not been as revealing in mammalian studies. In mouse systems genes involved with immune response have been implicated in longevity, as has the 'longevity gene' $p66^{shc}$, which is involved in signal transduction pathways that regulate the cellular response to oxidative stress. In humans, a number of mitochondrial DNA polymorphisms are associated with longevity. Linkage analysis in human systems has associated certain genes on chromosome 4 with exceptional longevity. Further support for human longevity genes may be provided by the observation that siblings and parents of centenarians live longer. The major histocompatibility complex (MHC), the master genetic control of the immune system, may also be a gene system controlling aging, because a number of genetic defects that cause immunodeficiency shorten the life span of humans. Certain MHC phenotypes have also been associated with malignancy, autoimmune disease, Alzheimer disease, and xeroderma pigmentosum in humans.

Accelerated Aging Syndromes

No distinct phenocopy exists for normal aging, but several genetic diseases/syndromes display some features of accelerated aging, including Hutchinson-Gilford syndrome (classic early onset Progeria), Werner syndrome, and Down syndrome. Patients with these syndromes suffer from many signs of premature aging including hair loss, early graying, and skin atrophy, and premature age-related diseases such as atherosclerosis, osteoporosis, and glucose intolerance. The defined genetics involved in these syndromes provides strong evidence for the genetic basis of aging.

Neuroendocrine Theories

These theories propose that functional decrements in neurons and their associated hormones are pivotal to the aging process. An important version of this theory suggests that the hypothalamic–pituitary–adrenal (HPA) axis is the key regulator of mammalian aging. The neuroendocrine system regulates early development, growth, puberty, the reproductive system, metabolism, and many normal physiological functions. Functional changes to this system could exert effects of aging throughout an organism. However, the cells of the neuroendocrine system are subject to the normal cellular aging processes found in all cells, and changes in the neuroendocrine system may be secondary expressions of the aging phenotype.

Immunologic Theory and Immunosenescence

Deterioration of the immune system with aging ('immunosenescence') may contribute to morbidity and mortality due to decreased resistance to infection and, possibly, certain cancers in the aged. T-cell function decreases and autoimmune phenomena increase in elderly. Although the immune system obviously plays a central role in health status and survival, its cells are subject to the normal cellular aging processes found in all cells. Changes to the immune system may be secondary expressions of the aging phenotype.

Cellular Senescence

At the cellular level, most, if not all, somatic cell types have a limited replicative capacity *in vitro* before they senesce and die. The number of cell population doublings *in vitro* is inversely correlated with donor age. This is called the 'Hayflick phenomenon' after the scientist credited with its discovery. This limit in the capacity of a cell type or tissue to divide and replenish itself would have major repercussions *in vivo*. There is evidence that replicative senescence is related to *in vivo* aging, but definitive evidence that senescent cells accumulate *in vivo* is lacking to date. Many alterations to normal cellular physiology are exhibited with the senescent phenotype, indicating that senescent cells exist in a growth state that is quite distinct from that of young cells and are subject to a complex alteration to their cellular physiology. Several explanations for limiting the number of cell population doublings have been proposed, including a tumor suppressive mechanism. One is that the shortening of telomeres, the sequences of noncoding DNA located at the end of chromosomes, is a measure of the number of cell divisions that a cell has experienced. These telomeres may act as specialized regions of the genome, a sacrificial 'sentinel' zone, for the detection of DNA damage being noncoding, more prone to damage, and less prone to repair than the genome as a whole. Damage to telomeres transposes to telomere shortening, and loss of telomere higher order structure may trigger senescence and/or apoptosis.

Studies involving fusion of normal cells (subject to senescence) with immortal cell lines *in vitro* have clearly demonstrated that the senescent phenotype is dominant, and that unlimited division potential results from changes in normal growth control mechanisms. These fusion studies have also revealed the existence of several dominant genes associated with the process of cellular senescence. These genes reside on a number of chromosomes, including 1, 4, and X.

Disposable Soma Theory

The disposable soma theory suggests that aging is due to stochastic background damage to the organism, i.e., damage that is not repaired efficiently because the energy resources of the somatic cells are limited. So, instead of wasting large amounts of energy in maintaining the whole body in good condition, it is far more economical to simply repair the heritable stem cell genetic material, in order to ensure the survival of the species. In this way the future of the species is secured at the expense of individual lives. When the somatic energy supply is exhausted, the body ages and dies, but the genetic material survives (in the next generation).

Damage Accumulation (Stochastic) Theories

The 'damage' or 'error' theories emphasize intrinsic and environmental insults to our cellular components that accumulate throughout life and gradually cause alterations in biological function and the physiological decline associated with aging.

Somatic Mutation and DNA Repair

Damage to DNA occurs throughout the lifetime of a cell. If this damage is not repaired or removed then mutations may result. Mutations may result in the synthesis of aberrant proteins with altered or absent biological function; alterations to the transcriptional and translational machinery of a cell; and deregulation of gene control. The accumulation of mutations on their own, or in combination with other age-related changes, may lead to alterations in cellular function and ultimately the onset of age-related disease.

Error Catastrophe

This theory suggests that damage to mechanisms that synthesize proteins results in faulty proteins, which accumulate to a level that causes catastrophic damage to cells, tissues, and organs. Altered protein structure has been clearly demonstrated to occur with age; however, most of these changes are posttranslational in nature, and hence do not support this theory of aging. Such changes to protein structure may result in progressive loss of 'self-recognition' by the cells of the immune system and thus increase the likelihood that the immune system would identify self-cells as foreign and launch an immune attack. Indeed, the incidence of autoimmune episodes is known to increase with age.

Cross-Linking

The cross-linking theory states that an accumulation of cross-linked biomolecules caused by covalent or hydrogen bonds damages cellular and tissue function through molecular aggregation and decreased mobility. The modified malfunctional biomolecules accumulate and become increasingly resistant to degradation processes and may represent a physical impairment to the functioning of organs. There is evidence *in vitro* for such cross-linking over time in collagen and in other proteins, and in DNA. Many agents exist within the body that have the potential to act as cross-linking agents, e.g., aldehydes, antibodies, free radicals, quinones, citric acid, and polyvalent metals, to name but a few.

Free Radicals

The most popular, widely tested and influential of the damage accumulation theories of aging is the 'free radical' theory, first proposed by Harman in 1956. Free radicals from intrinsic and extrinsic sources (**Table 1**) can lead to activation of cytoplasmic and/or nuclear signal transduction pathways, modulation of gene and protein expression, and also alterations to the structure and ultimately the function of biomolecules. Free radicals may thus induce alterations to normal cell, tissue, and organ functions, which may result in a breakdown of homeostatic mechanisms and lead to the onset of age-related disorders and ultimately death. It can be predicted from this theory that the life span of an organism may be increased by slowing down the rate of initiation of random free radical reactions or by decreasing their chain length. Studies have demonstrated that it is possible to increase the life span of cells *in vitro* by culturing them with various antioxidants or free radical scavengers. Antioxidant supplementation with a spin-trapping agent has been demonstrated to increase the life span of the senescence accelerated mouse, although as yet there is little evidence for increasing the life span of a normal mammalian species by such strategies.

Mitochondrial DNA Damage

This hypothesis combines elements of several theories, covering both the stochastic and genetic classes of aging theories. It is proposed that free radical reactive oxygen species generated in the mitochondria contribute significantly to the somatic accumulation of mitochondrial DNA mutations. This leads to a downward spiral wherein mitochondrial DNA damage results in defective mitochondrial respiration that further enhances oxygen free radical production, mitochondrial DNA damage, and mutation. This leads to the loss of vital bioenergetic capacity eventually resulting in aging and cell death.

Table 1 Extrinsic and intrinsic sources of free radicals

Extrinsic sources	*Intrinsic sources*
Radiation: ionizing, ultraviolet	Plasma membrane: lipoxygenase, cycloxygenase, NADPH oxidase
Drug oxidation: paracetamol, carbon tetrachloride, cocaine	Mitochondria: electron transport, ubiquinone, NADH dehydrogenase
Oxidizing gases: oxygen, ozone, nitrogen dioxide	Microsomes: electron transport, cytochrome p450, cytochrome b_5
Xenobiotic elements: arsenic (As), lead (Pb), mercury (Hg), cadmium (Cd)	Peroxisomes: oxidases, flavoproteins
Redox cycling substances: paraquat, diquat, alloxan, doxorubicin	Phagocytic cells: neutrophils, macrophytes, eosinophils, endothelial cells
Heat shock	Auto-oxidation reactions: Metal catalyzed reactions
Cigarette smoke and combustion products	Other: hemoglobin, flavins, xanthine oxidase, monoamine oxidase, galactose oxidase, indolamine dioxygenase, tryptophan dioxygenase
	Ischemia – reperfusion

The absence of evidence that exclusively supports any one theory leaves no doubt that aging is due to many processes, interactive and interdependent, that determine life span and death.

Age-Related Diseases

Regardless of the molecular mechanisms that underlie the aging process, a number of well-characterized changes to the structure and therefore the function of the major cellular biomolecules (lipids, proteins, carbohydrates, and nucleic acids) are known to occur with age (Table 2). The age-related alterations to the structure and therefore the function of cellular biomolecules have physiological consequences and may directly cause or lead to an increased susceptibility to the development of a number of diseases (Figure 2).

Cellular biomolecules are constantly exposed to a variety of extrinsic and intrinsic agents that have the potential to cause damage. A number of defense systems exist, e.g., antioxidant enzymes and DNA repair systems, which aim to reduce, remove, or repair damaged biomolecules. These defense systems are not perfect, however, and biomolecular damage may still occur. Such damage can result in the degradation of structural elements within the cells, tissues, and organs of the body, leading to a decline in biological function and eventually to disease and death.

The physiological alterations with age proceed at different rates in different individuals. Some of the common changes seen in humans are: the function of the immune system decreases by the age of 30 years, reducing defenses against infection or tumor establishment and increasing the likelihood of autoimmune disorders; metabolism starts to slow down at around 25 years of age; kidney and liver function decline; blood vessels lose their elasticity; bone mass peaks at the age of 30 years and drops approximately 1% per year thereafter; the senses fade; the epidermis becomes dry and the dermis thins; the quality of and need for sleep diminish; and the brain loses 20% of its weight, slowing recall and mental performance. A number of age-related diseases may develop due to the tissue, organ, and system deterioration (Table 3).

Table 2 Major age-related alterations in biomolecule structure and the resultant physiological consequences of such structural changes

Biomolecule	*Alteration*	*Physiological consequence*
Lipids	Lipid peroxidation	Oxidized membranes become rigid, lose selective permeability and integrity. Cell death may occur Peroxidation products can act as cross-linking agents and may play a role in protein aggregation, the generation of DNA damage and mutations, and the age-related pigment lipofuscin
Proteins	Racemization, deamination, oxidation, and carbamylation	Alterations to long-lived proteins may contribute to aging and/or pathologies. For example, modified crystallins may aggregate in the lens of the eye thus leading to the formation of cataracts Cross-linking and formation of advance glycosylation end-products (AGEs), which can severely affect protein structure and function Effects on the maintenance of cellular homeostasis
Carbohydrates	Fragmentation, depolymerization glucose auto-oxidation	Alters physical properties of connective tissue. Such alteration may be involved in the etiology and pathogenesis of osteoarthritis and other age-related joint disorders Glycosylation of proteins *in vivo* with subsequent alteration of biological function; for example, glycosylation of insulin in patients with diabetes may result in altered biological function of insulin and contribute to the pathogenesis of the disease
Nucleic acids	Strand breaks base adducts loss of 5-methyl cytosine from DNA	Damage might interfere with transcription, translation, and DNA replication, reducing a cell's capacity to synthesize vital polypeptides/proteins. In such circumstances cell death may occur. The accumulation of hits in critical cellular genes associated with the control of cell growth and division results in the process of carcinogenesis Dedifferentiation of cells (5-methylcytosine plays an important role in switching off genes as part of gene regulation) If viable, such dedifferentiated cells may have altered physiology and may contribute to altered tissue/organ function

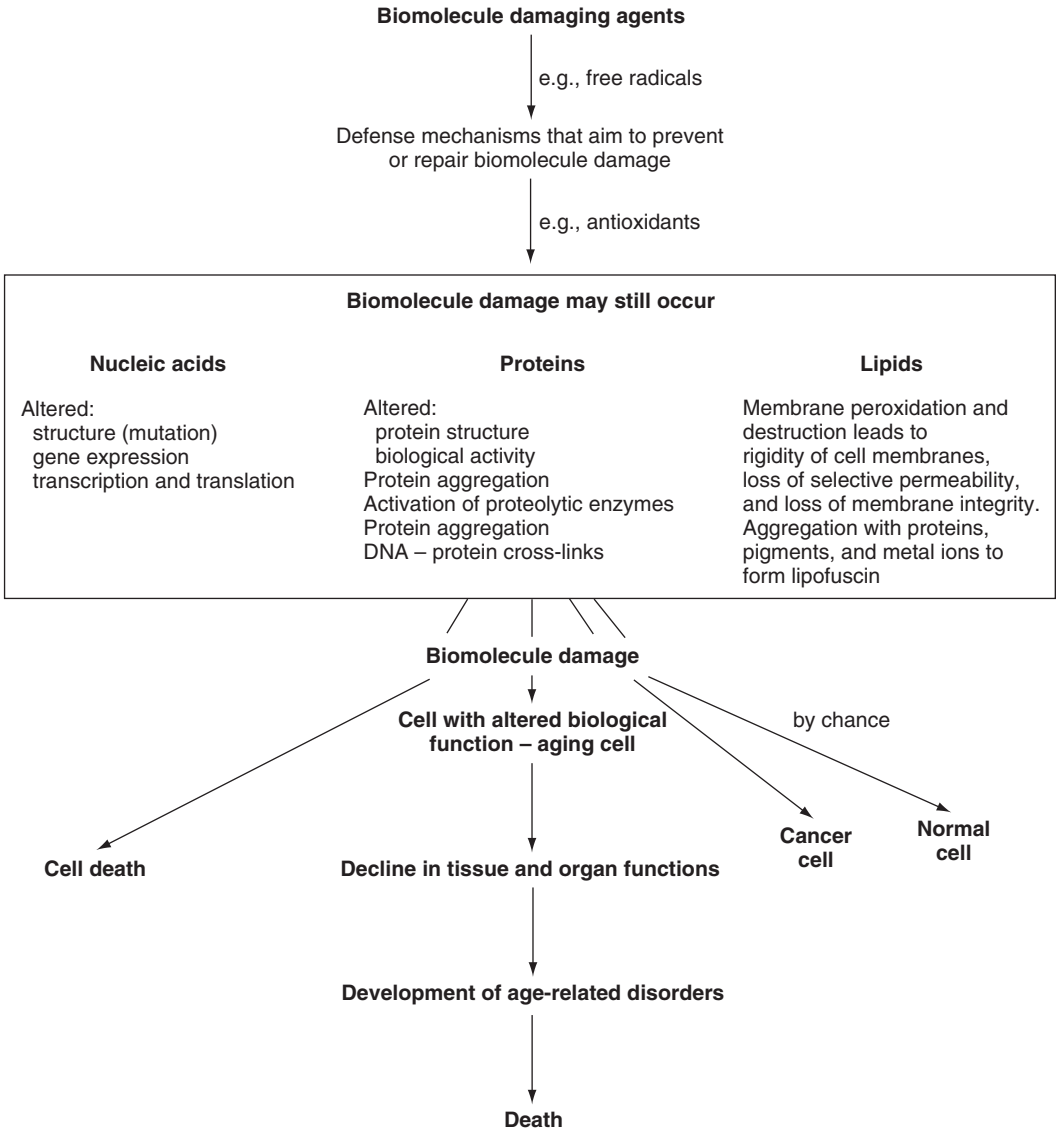


Figure 2 Biomolecule damage and the aging process. Reproduced from Barnett YA (1994) Nutrition and the aging process. *British Journal of Biomedical Sciences* 51: 278–287.

Modification of the Aging Process

Can the adverse consequences of aging be prevented? Through the ages many have pursued the elixir of life. Attempts to increase the average life expectancy and quality of life in the elderly can only succeed by slowing the aging process itself. In humans, the rate of functional decline associated with aging may be reduced through good nutrition, exercise, timely health care, and avoidance of risk factors for age-related disease.

Nutritional Modification

It is clear that diet contributes in substantial ways to the development of age-related diseases and that modification of the diet can contribute to their prevention and thus help to improve the quality of life in old age. Macronutrient intake levels can play a significant role in the progression of age-related diseases and affect the quality of life. For example, the total and proportional intakes of polyunsaturated fatty acids and saturated fatty acids in the Western diet may have an effect on the incidence of atherosclerosis and cardiovascular diseases.

Our dietary requirements also change as we age and if such changes are not properly addressed this could lead to suboptimal nutritional status. This challenge is compounded by a decrease in the body's ability to monitor food and nutrient intakes. Dietary

Table 3 Major age-related alterations *in vivo* and the resultant pathological conditions

Body system	*Pathological changes*
Cardiovascular	Atherosclerosis, coronary heart disease, hypertension
Central nervous system	Reduction of cognitive function, development of various dementias (e.g., Alzheimer's disease and Parkinson's disease)
Endocrine	Noninsulin-dependent diabetes, hypercortisolemia
Hematopoietic	Anemia, myelofibrosis
Immune	General decline in immune system function, particularly in T cells
Musculoskeletal	Osteoporosis, osteoarthritis, skeletal muscle atrophy
Renal	Glomerulosclerosis, interstitial fibrosis
Reproductive	Decreased spermatogenesis, hyalinization of semeniferous tubules
Respiratory	Interstitial fibrosis, decreased vital capacity, chronic obstructive pulmonary disease
Sense organs	Cataracts, senile macular degeneration, diabetic retinopathy
All systems	Cancer

intake and requirements are complex issues, intertwined with many health and life style issues. However, most research points toward the need for a varied diet as we age, with an increased emphasis on micronutrient intake levels.

An exemplary diet for healthy aging can be found in the traditional diet of Okinawa, Japan. Okinawans are the longest-living population in the world according to the World Health Organization, with low disability rates and the lowest frequencies of coronary heart disease, stroke, and cancer in the world. This has been attributed to healthy life style factors such as regular physical activity, minimal tobacco use, and developed social support networks as antistress mechanisms, all of which are underpinned by a varied diet low in salt and fat (with monosaturates as the principal fat) and high levels of micronutrient and antioxidant consumption.

Vitamins and Micronutrients

The mechanisms by which certain vitamins and micronutrients mediate their protective effect on age-related disorders is based on their abilities to prevent the formation of free radicals or scavenging them as they are formed, either directly (e.g., vitamins C, E, and β-carotene) or indirectly (e.g., copper/zinc superoxide dismutase, manganese-dependent superoxide dismutase, selenium-dependent glutathione peroxidase) (Table 4). Only by exploring more fully the underlying molecular mechanisms of aging and the major classes of antioxidants will it be possible to establish their role, and develop strategies for using various classes of antioxidants to reduce the effects of aging. Other dietary components may also have a beneficial effect in preventing or delaying the

Table 4 Effects of vitamins and micronutrients on age-related disorders

Vitamin or micronutrient	*Possible effect on age-related disorder*
Vitamins B_6, E copper, zinc, and selenium	Impairment of immune function in older humans if inadequate amounts
Vitamins C, E, and carotenoids	Increased amount in the diet is associated with delayed development of various forms of cataract
Carotenoids and zinc	Dietary supplementation associated with a decreased risk of age-related macular degeneration
Selenium	Absolute or relative deficiency associated with development of a number of cancers (not breast cancer)
Vitamin C, β-carotene, α-tocopherol, and zinc	Dietary supplementation may decrease the rate of development of atherosclerosis
Selenium, copper, zinc, lithium, vanadium, chromium, and magnesium	Dietary deficits are associated with an increased risk of cardiovascular disease
Vitamins B_{12}, B_6, and folate	Adequate levels throughout a lifetime may prevent some of the age-related decrease in cognitive function
Chromium	Deficiency is associated with an increased risk of Type 2 diabetes mellitus

onset of age-related disease. For example, as a deterrent against the onset of osteoporosis, adults should ensure adequate calcium and vitamin D intakes.

Dietary Energy Restriction

The effect of caloric restriction on life span has only been demonstrated convincingly in rodents. Feeding mice and rats diets deficient in energy (approximately 35% of that of animals fed ad libitum, after the initial period of growth) retards the aging of body tissues, inhibits the development of disease and tumors, and prolongs life span significantly. The exact mechanism of action of dietary energy restriction remains to be elucidated, but may involve modulation of free radical metabolism, or the reduced hormone excretion that occurs in dietary restricted animals may lower whole body metabolism resulting in less 'wear and tear' to body organs and tissues.

Current investigations into the effects of dietary energy restriction (by approximately 30%) on the life spans of primates, squirrels, and rhesus monkeys continue. Caloric restriction in rhesus monkeys leads to reductions in body temperature and energy expenditure consistent with the rodent studies. These investigations should have direct implications for a dietary energy restriction intervention aimed at slowing down the aging process in humans, should any humans wish to extend their life span at such a cost. Once the mechanisms of effects of caloric restriction on longevity are understood it may be possible to develop drugs that act through these mechanisms directly, mitigating the need for diets that interfere with the quality of life.

Molecular Biological Interventions and the Aging Process

Accelerated aging syndromes show degenerative characteristics similar to those appearing during normal aging. The mutations leading to these disorders are being identified and their roles in the aging process are being elucidated. Examining differences in the genetic material from normal elderly people and those with progeria should provide a better understanding of the genetic mechanisms of aging. Identification of a control gene or genes that inhibit the action of the genes producing the progeroid phenotype might make it possible to slow down aberrant protein production in normal people as well.

As an example, the genetic defect that predisposes individuals to the development of Werner syndrome has been elucidated. Individuals with this disease carry two copies of a mutant gene that codes for a helicase enzyme (helicases split apart or unwind the two strands of the DNA double helix). DNA helicases play a role in DNA replication and repair. In light of the biological function of these enzymes it has been proposed that premature aging in Werner syndrome is caused by defective helicase preventing DNA repair enzymes from removing background DNA damage, which thus becomes fixed as mutations, with consequent deleterious effects on cellular function. It remains to be determined whether increasing the fidelity or activity of helicases in cells will extend their life span.

Because loss of telomeric DNA sequences may lead to replicative senescence in dividing cells, in theory by preventing such telomere loss the life span of the cell could be extended. A naturally occurring enzyme, telomerase, exists to restore telomeric DNA sequences lost by replication. Telomerase is normally only functional in germ cells. Manipulating certain cell types (e.g., cells of the immune system) to regulate the expression of telomerase may extend their functional life span. Drugs that enhance telomerase activity in somatic cells are being developed. However, cellular senescence has been implicated as a tumor suppressor mechanism and it has been found that cancer cells express telomerase. An uncontrolled expression of this enzyme in somatic cells may lead to the onset of malignancy through uncontrolled cell proliferation. Thus, any intervention aiming to increase life span based on the cellular expression of telomerase must strike a balance between maintaining controlled cell division and uncontrolled proliferation.

A number of single gene mutations have been identified that affect metabolic function, hormonal signaling, and gene silencing pathways. In future it may be possible to develop drugs to mimic the antiaging effects that these genes exert.

Further Reading

Anon (1995) Somatic mutations and ageing: Cause or effect? *Mutation Research, DNAging* (special issue) 338: 1–234.
Barnett, Y.A., 1994. Nutrition and the ageing process. British Journal of Biomedical Sciences 51, 278–287.
Bellamy, D. (Ed.), 1995. Ageing: A Biomedical Perspective. Wiley, Chichester.
Esser, K., Martin, G.M., 1995. Molecular Aspects of Ageing. Wiley, Chichester.
Finch, C.E., 1991. Longevity, Senescence and the Genome. University of Chicago Press, Chicago.
Hayflick, L., 1993. Aspects of cellular ageing. Reviews in Clinical Gerontology 3, 207–222.
Kanungo, M.S., 1994. Genes and Ageing. Cambridge University Press, Cambridge.
Kirkland, J.L., 2002. The biology of senescence: Potential for the prevention of disease. Clinics in Geriatric Medicine 18, 383–405.
Kirkwood, T.B.L., 1992. Comparative lifespans of species: Why do species have the lifespans they do? American Journal of Clinical Nutrition 55, 1191S–1195S.
Mera, S.L., 1992. Senescence and pathology in ageing. Medical Laboratory Sciences 4, 271–282.
National Institute on Aging (2010) Can we prevent aging? http://www.nia.nih.gov/HealthInformation/Publications/preventaging.htm
Tominaga, K., Olgun, A., Smith, J.R., Periera-Smith, O.M., 2002. Genetics of cellular senescence. Mechanisms of Ageing and Development 123, 927–936.
Troen, B.R., 2003. The biology of ageing. The Mount Sinai Journal of Medicine 70 (1), 3–22.
US Bureau of the Census, 1999. Report WP/98, World Population Profile. US Government Printing Office, Washington, DC.
von Zglinicki, T., Bürkle, A., Kirkwood, T.B.L., 2001. Stress, DNA damage and ageing – An integrated approach. Experimental Gerontology 36, 1049–1062.
Weinert, B.T., Timiras, P.S., 2003. Invited review: Theories of aging. Journal of Applied Physiology 95, 1706–1716.

Breastfeeding

Chessa K. Lutter, RTI International, Washington, DC, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of C. Lutter, Breast Feeding, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 207–212, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00031-3.

Introduction 20
Breast milk composition 20
Breastfeeding and maternal and infant health 21
Global and community factors affecting breastfeeding 24
Health system 24
Community 25
Maternal 26
Breastfeeding protection, promotion, and support 26
Breastfeeding rates and trends 27
Conclusion 29
References 29
Further reading 29

Key points

- Breast milk is a unique bioactive substance that changes composition to suit the nutritional, developmental, and immunological needs of the growing infant and young child
- While breast milk and lactation can be viewed as a biological system, the act of breastfeeding is a behavior that involves complex social and cultural norms and maternal and infant attributes
- Successful breastfeeding is not the sole responsibility of a mother, but requires collective social responsibility and supportive public policies
- Breastfeeding not only benefits child health and development, it also holds benefits for maternal health
- Rates of exclusive breastfeeding are increasing; however, rates of continued breastfeeding are declining

Introduction

Incomparable to any other mammalian milk, breast milk is a unique bioactive substance that changes composition, within and between feedings and over time, to suit the nutritional, developmental, and immunological needs of the growing infant and young child. Its composition and biological functions, which have evolved over millions of years to meet the specific needs of a child, are more than the sum of their parts (Christian et al., 2021). As a perfectly adapted source of nutrition and immune protection, breast milk has been referred to as the ultimate personalized medicine (Victora et al., 2016). While breast milk and lactation can be viewed as a biological system, the act of breastfeeding is a behavior that involves complex social and cultural norms and maternal and infant attributes. As such, successful breastfeeding is not the sole responsibility of a mother, but requires collective social responsibility and supportive public policies (Rollins et al., 2016). Breastfeeding not only benefits child health and development, it also holds benefits for maternal health (Victora et al., 2016). This article provides a broad overview of the science of breast milk and the benefits and determinants of breastfeeding, including sections on breast milk composition, breastfeeding and maternal and infant health, global and community factors affecting breastfeeding, breastfeeding protection, promotion, and support, and breastfeeding rates and trends. Recommendations for breastfeeding by the World Health Organization (WHO) and the United Nations Children's Fund (UNICEF) are summarized in **Box 1**.

Breast milk composition

As a dynamic and bioactive fluid, the composition of breast milk changes during the course of lactation and is also affected by interactions between the mother-infant dyad (Ballard and Morrow, 2013; Christian et al., 2021). It is also influenced by maternal diet,

Box 1 WHO and UNICEF breastfeeding recommendations

- Early initiation of breastfeeding within 1 h of birth
- Exclusively breastfeed, defined as breast milk as the sole source of infant nutrition with no other liquids (including water) or food, except for oral rehydration solution, drops, syrups (e.g., vitamins, minerals, or medicines)
- Continue to breastfeed for 2 years of age or more
- Breastfed on demand day and night

health, and nutritional status. Breast milk is composed of fat, protein, carbohydrates, vitamins, minerals, and water. It also includes a large array of non-nutritive substances, which include immunoglobulins, hormones proteins, human milk oligosaccharides, white blood cells, anti-microbial peptides, cytokines, chemokines, micro RNAs, and commensal bacteria, all of which are highly effective in protecting infant health (Christian et al., 2021). Knowledge about the number and function of these non-nutritive substance continually increases with new analytic techniques.

Milk composition varies by the three stages of lactation: colostrum, transitional milk, and mature milk. Colostrum, produced in low quantities and secreted during the first 4–5 days after birth, is rich in immunoglobulins and other proteins, carotenoids and electrolytes (e.g., calcium and sodium), and developmental factors such as epidermal growth factor (Ballard and Morrow, 2013; Christian et al., 2021). Transitional milk, secreted between around the fifth to tenth day after birth, shares some characteristics of colostrum but increases in volume to support the rapidly increasing nutritional and developmental needs of the newborn. Mature milk, secreted when lactation is fully established at about day 10, is higher in fat and carbohydrates but still rich in immunological factors.

Breast milk contributes all the energy an infant needs for about 6 months and a substantial proportion of energy beyond 6 months of age (Taren and Lutter, 2017). In low- and middle-income countries, average breast milk intake between 12 and 23 months of age contributes between 35% and 45% of a young child's energy needs. Between 15 and 18 months, breast milk provides approximately 70% of a child's vitamin A requirements, 40% of his or her calcium requirements, and 37% of his or her riboflavin requirements. Breast milk is also an important source of choline, a nutrient critical for brain development and function.

Fats in breast milk contribute about 44% of total energy intake in young infants and provide essential vitamins, polyunsaturated fatty acids, and bioactive components. Fats are the most variable component of the three macronutrients in breast milk, with wide interindividual and intraindividual variation. They also change during a single breastfeeding episode, with markedly higher concentrations at the end of feeding compared to the beginning. Maternal fatty acid status affects breast milk concentrations of these lipids and there is some evidence to suggest that infants fed breast milk with a higher content of docosahexaenoic acid (DHA) have better vision and neurodevelopmental outcomes.

Proteins in breast milk contribute 45% of total energy intake in young infants, providing a source of amino acids and improving the bioavailability of vitamins, minerals, and trace elements. Proteins also provide immunologic benefits, stimulate intestinal growth and maturity, shape the microbiome, and contribute to learning and memory.

Carbohydrates in breast milk contribute about 8% of total energy intake in young infants, with the majority being lactose. Other carbohydrates include oligosaccharides, which positively affect neurodevelopmental outcomes, and fructose, the health and nutrition role of which is less well understood, though it may have an effect on infant growth. Maternal diet influences fructose concentrations in breast milk. Because fructose can negatively alter areas of the brain that support learning and memory, more research is needed to understand the role of fructose on infant health and nutrition.

Vitamins and minerals in breast milk are generally divided into two categories: those that are affected by maternal nutritional status and diet and those that are not affected by maternal nutrition (**Table 1**) However, there are some nutrients that are not affected by maternal nutritional status but are affected by maternal diet. In general, water-soluble vitamins are most affected by maternal nutrition, though there are some important exceptions such as thiamin and folate. Minerals are not generally affected by maternal nutrition, though iodine and selenium are important exceptions. Iron content of breast milk is low, though highly bioavailable. Exclusively breastfed infants, particularly those born preterm, are at risk of iron deficiency if they do not receive placental blood at delivery through delay of clamping of the umbilical cord and if their mothers had poor iron status during pregnancy. For full term infants, delayed cord clamping results in increased hematocrit and hemoglobin at 2–4 months of age and reduces the risk of iron deficiency. For the preterm infant, delayed cord clamping decreases the risk of intra-vernacular hemorrhage, necrotizing enterocolitis, and late-on-set sepsis among other benefits.

Breastfeeding and maternal and infant health

Breastfeeding has multiple health benefits for mothers and children in both low- and high-income settings (Victora et al., 2016). Systematic reviews of multiple health outcomes, briefly summarized below, are available in an open-access supplement in *Acta Paediatrica* (Grummer-Strawn and Rollins, 2015).

Table 1 Summary of effects of maternal nutritional status and diet on vitamins and minerals in breast milk.

Nutrient	*Affected by maternal nutritional status*	*Affected by maternal diet*
Thiamin	–	+
Riboflavin	−/+ (mixed evidence)	+
Vitamin B-6	+	+
Vitamin B-12	+	+
Folate	−	−
Choline	+	+
Vitamin C	−	−/+
Vitamin A	− (unless maternal reserves are depleted)	−/+ (+ if maternal reserves are inadequate)
Vitamin D	−/+ (conflicting data)	−/+ (diet may affect vitamin D_3 but not active 25 (OH)D)
Vitamin E	–	–
Vitamin K	–	–
Iron	–	–
Copper	–	–
Zinc	–	–
Calcium	–	−/+ (+ where habitual calcium intake is low)
Phosphorus	−/+ (+only in instances of genetic anomalies)	–
Magnesium	–	–
Iodine	–	+
Selenium	−/+ (weak correlation, if present)	+

Based on data presented in Table 1 in Dror, D., Allen, L.H., 2018. Overview of nutrients in human milk. Adv. Nutr. (Suppl. 1), 78A–294S.

In women, breastfeeding reduces a woman's risk of breast cancer, ovarian cancer, and type-2 diabetes (Victora et al., 2016) (Table 2). There is also some data suggesting that breastfeeding reduces a woman's risk of cardiovascular disease. A set of systematic reviews on maternal health outcomes associated with breastfeeding, summarized by Victora et al. showed that each 12-month period of breastfeeding in a woman's lifetime was associated with a reduction of 4.3% in breast cancer and did not vary substantially according to menopausal status (Victora et al., 2016). Current rates of breastfeeding are estimated to prevent nearly 20,000 excess deaths from breast cancer each year and research indicates that another 20,000 deaths could be prevented by improving breastfeeding rates. Highest versus lowest durations of breastfeeding were associated with a 30% reduction in ovarian duration cancer. Never versus ever breastfed, longer versus shorter duration of exclusive breastfeeding; or longer versus shorter duration of any breastfeeding was associated with 35% reduction of type-2 diabetes. As is well-known, breastfeeding is associated with lactational amenorrhea and, in the absence of use of modern contraceptives, contributes substantially to birth spacing (Victora et al., 2016). Breastfeeding had no discernible effect on a woman's risk of osteoporosis and evidence on post-partum weight is inconclusive.

Breastfeeding has short- and long-term effects on children's health and cognitive development (Victora et al., 2016). In low- and middle-income countries, exclusively breastfed infants had only a 12% risk of death compared to non-breastfed infants and, among children 6–23 months of age, any breastfeeding was associated with a 50% reduction in death (Victora et al., 2016). An estimated 820,000 lives a year could be saved though improved breastfeeding practices, 87% of them in infants less than 6 months of age. There is also some evidence that breastfeeding might also protect against death in high-

Table 2 The association of breastfeeding and maternal health outcomes.

Outcome	*Types of comparisons (# studies)*	*Pooled effect (95% CI)*
Breast cancer	Highest versus lowest duration of breastfeeding (76)	Odds ratio 0.81 (0.77–0.86)
Ovarian cancer	Highest versus lowest duration of breastfeeding (41)	Odds ratio 0.70 (0.64–0.75)
Osteoporosis (distal radius)	Highest versus lowest duration of breastfeeding (4)	Standard deviation score 0.132 (−0.260 to −0.003)
Osteoporosis (femoral neck)	Highest versus lowest duration of breastfeeding (5)	Standard deviation score −0.142 (−0.426 to 0.142)
Type 2 diabetes	Highest versus lowest duration of breastfeeding (6)	Risk ratio 0.68 (0.57–0.82)
Post-partum weight change	Qualitative review (45)	Not estimated because of different outcome measures at variable post-partum ages
Lactational amenorrhea	Highest versus lowest duration of breastfeeding (13)	Risk ratio 1.17 (1.04–1.32)

Based on Table in Victora, C., et al., 2016. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. Lancet 387, 1–16.

income countries, including from sudden infant death syndrome and necrotizing enterocolitis (Victora et al., 2016). A large number of research studies have examined the effect of different lengths of breastfeeding (more versus less breastfeeding and any breastfeeding versus no breastfeeding) on risk of diarrhea and respiratory infections. From these studies, researchers estimate that increased breastfeeding could prevent nearly half of all diarrhea episodes and one-third of all respiratory infections (Victora et al., 2016). Breastfeeding is also protective against malocclusion. Systematic reviews found no effect of breastfeeding on eczema or food allergies and evidence was inconclusive regarding asthma and wheezing. Breastfeeding appears to have a negative effect on dental caries. With respect to long-term effects, children who were breastfed were 35% less likely to contract type-2 diabetes and 26% less likely to suffer from overweight or obesity (**Table 3**). Breastfeeding has also been associated with a 19% reduction in childhood leukemia. It has no effect on systolic blood pressure, diastolic blood pressure, and total cholesterol. The effect of breastfeeding on intelligence has received particular attention from researchers, with the most recent systematic review showing consistently higher performance on intelligence tests in breastfed children and adolescents, with a pooled increase of 3.4 IQ points.

Globally, the estimated economic cost of lower cognitive ability associated with lower rates of breastfeeding amount to about US $300 billion, annually (Victora et al., 2016). Annual reductions in health care costs for mothers and children associated with breastfeeding have been estimated to be US $312 million in the United States, $48 million in the United Kingdom, US $6 million in Brazil, and US $30.3 million in urban China.

There are several maternal health conditions that affect breastfeeding and this article reviews only a few of them. Early in the human deficiency virus (HIV) epidemic researchers determined that the virus could be transmitted through breast milk. Risk of transmission is related to the health and viral load of the mother and breastfeeding conditions, such as whether the mother suffered from mastitis, breast abscesses, or nipple fissures. Current WHO guidance for infant feeding in the context of HIV recommends that mothers known to be HIV infected should be provided with lifelong antiretroviral therapy (ART) or antiretroviral prophylaxis interventions to reduce HIV transmission through breastfeeding and national or sub-national health authorities should decide whether health services will principally counsel mothers known to be HIV infected to either breastfeed and take antiretrovirals or to avoid all breastfeeding. In settings where national health authorities are recommending breastfeeding for HIV-infected mothers, guidance recommends that: (1) Mothers known to be HIV infected (and whose infants are HIV uninfected or of unknown HIV status) should exclusively breastfeed their infants for the first 6 months of life, introducing appropriate complementary foods thereafter, and continue breastfeeding; (2) Mothers living with HIV should breastfeed for at least 12 months and may continue breastfeeding for up to 24 months or longer while being fully supported to optimize ART adherence as outlined in the *WHO Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection*; (3) In settings where health services provide and support lifelong ART, including adherence counseling, and promotion and support of breastfeeding among women living with HIV, the duration of breastfeeding should not be restricted; and (4) Breastfeeding should only stop once a nutritionally adequate and safe diet without breast milk can be provided. The guidelines also stated that national and local health authorities should actively coordinate and implement services in health facilities and activities in workplaces, communities, and homes to protect, promote, and support breastfeeding among women living with HIV.

WHO and UNICEF guidance on acceptable medical reasons for use of breast-milk substitutes identifies other situations and circumstances when it is best to avoid breastfeeding. The American Academy of Pediatrics provides information on the transfer of drugs and therapeutics into breast milk that have implications for the safety of breastfeeding when a mother is taking certain medications.

Table 3 The association of breastfeeding and long-term child health outcomes.

Outcome	*Types of comparisons (# studies)*	*Pooled effect (95% CI)*
Systolic blood pressure	Never versus ever breastfed; or longer versus shorter breastfeeding duration (43)	−80 mm Hg (−1.17 to −0.43)
Diastolic blood pressure	Never versus ever breastfed; or longer versus shorter breastfeeding duration (38)	−0.24 mm Hg (−0.50 to 0.02)
Overweight or obesity	Never versus ever breastfed; or longer versus shorter breastfeeding duration; or longer versus shorter duration of breastfeeding (113)	Odds ratio 0.74 (0.70–0.78)
Total cholesterol	Never versus ever breastfed; or longer versus shorter breastfeeding duration (46)	−0.1 mmol/L (−0.05 to 0.02)
Type 2 diabetes	Never versus ever breastfed; or longer versus shorter breastfeeding duration of exclusive breastfeeding; or longer versus shorter duration of breastfeeding (11)	Odds ratio 0.65 (0.49–0.86)
Intelligence	Highest versus lowest duration of breastfeeding (16)	IQ points 3.44 (2.30–4.58)

Based on Table in Victora, V., et al., 2016. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. Lancet 387, 1–16.

The COVID-19 pandemic has had an effect on breastfeeding caused by reduced face-to-face professional support and the virtual elimination of face-to-face peer support. Some women also received inaccurate advice regarding the safety of breastfeeding with COVID even though WHO recommends that mothers with both suspected and confirmed cases should be encouraged to initiate and continue breastfeeding (WHO, 2019). In the United Kingdom, an online survey conducted among 1219 breastfeeding mothers with a child under 12 months of age indicated two different experiences: 42% of respondents felt that breastfeeding was supported during the lockdown whereas 27% reported a struggle to get support, while some stopped breastfeeding before they had planned to do so. Mothers with lower levels of education, those living in more difficult circumstances, those who are Black, and those with minority ethnic backgrounds were most negatively impacted.

Maternal obesity increasing in low-, middle-, and high-come countries, leads to adverse breastfeeding outcomes resulting from both biological and social factors. A recent study reported that mothers with a body mass index (BMI) greater than 25 were less likely to exclusively breastfeed at 6 weeks and at 6 months compared with mothers with a "healthy" BMI between 18.5 and 25 (Marshall et al., 2019). A dose response effect was also found; cessation of breastfeeding was higher with increasing maternal BMI.

Global and community factors affecting breastfeeding

The sociocultural model is useful to frame the multiple levels of factors that affect a woman's choice of how to feed her infant and, if she chooses to breastfeed, how long she is likely to continue (**Fig. 1**). It also illustrates that a comprehensive approach to support breastfeeding is needed. **Fig. 1** highlights four levels of support: public policies and declarations; the health system; the community and maternal. Relevant public policies and declarations are discussed in detail in the last section.

Health system

At the level of the health system, important interventions that have been proven to improve breastfeeding behaviors include the Baby Friendly Hospital Initiative (discussed in detail in the last section), immediate skin-to-skin contact at birth, rooming in throughout time in the health facility, kangaroo care when indicated for a preterm infant, health worker education and training, counseling support, and lactation management support when a mother needs more skilled care.

Following birth, immediate skin-to-skin contact between mothers and their healthy newborn infants, continuing until the first breastfeed, evokes neuro-behaviors that fulfill a basic biological need of the mother-infant dyad (Moore et al., 2016). Researchers hypothesize that this period may represent a sensitive period for programming future physiology and behavior. After mode of delivery, how a newborn is fed is the next most basic determinant of the infant microbiome, which fosters important immune response in the infant and possibly brain development and cognitive functioning (Victora et al., 2016). Breast milk, provided early in life when gene expression is being fine-tuned for life, has been called the most specific personalized medicine an infant is likely to receive (Victora et al., 2016). Immediate skin-to-skin contact also affects breastfeeding outcomes. A recent systematic review showed that women who experienced skin-to-skin contact breastfed their infants longer and were more likely to exclusively breastfeed between discharge and 1 month post birth and between discharge and 6 weeks post birth (Moore et al., 2016). Furthermore, such women had higher scores for breastfeeding effectiveness and their newborns were more likely to breastfeed successfully during their first feed. The authors concluded that the evidence reviewed supported the recommendation that early skin-to-skin contact be a normal practice for healthy newborns, including those born by cesarean section and as early as 35 weeks gestation. A systematic review of rooming-in throughout a woman's and baby's stay in a health facility found little evidence to support or refute the practice of rooming-in compared to mother-infant separation for different breastfeeding outcomes. Nonetheless, both WHO and UNICEF continue to recommend rooming-in as the standard of care for healthy newborns.

The kangaroo mother care method was developed in the mid-1980s to support the growth and development of preterm infants, particularly in settings where advanced neonatal care was not available. A systematic review found that among low birthweight newborns, kangaroo mother care was associated with a 36% lower mortality rate, a 47% decrease in neonatal sepsis, a 78% decrease in hypothermia, a 88% decrease in hypoglycemia, and a 58% decrease in hospital readmission after discharge when compared to conventional care. Furthermore, kangaroo mother care increased exclusive breastfeeding by 50%.

A systematic review found that counseling and support increased the duration and exclusivity of breastfeeding (McFadden et al., 2017). Characteristics of effective support included that it was offered as standard of care during antenatal and postnatal care by trained personnel, that it included ongoing scheduled visits so that women knew when support would be available, and that it was tailored to the setting and needs of the population group. Face-to-face support was more likely to succeed with women exclusively breastfeeding. Lastly, support could be provided by either professional or lay/peer counselors or both. Lactation management provided by highly skilled personnel is important for the management of breastfeeding complications, such as mastitis, when they arise. A systematic review to assess if lactation education and support programs using lactation consultation specialists would improve rates of initiation and duration of any breastfeeding and exclusive breastfeeding compared with the standard of practice found that breastfeeding interventions using lactation consultants and counselors increased the number of women initiating breastfeeding by 35%. The interventions improved any breastfeeding rates by 51%. The odds of exclusive breastfeeding rates up to 1 month compared to non-exclusive breastfeeding (i.e., breastfeeding with some sort of supplementation) increased by 71%.

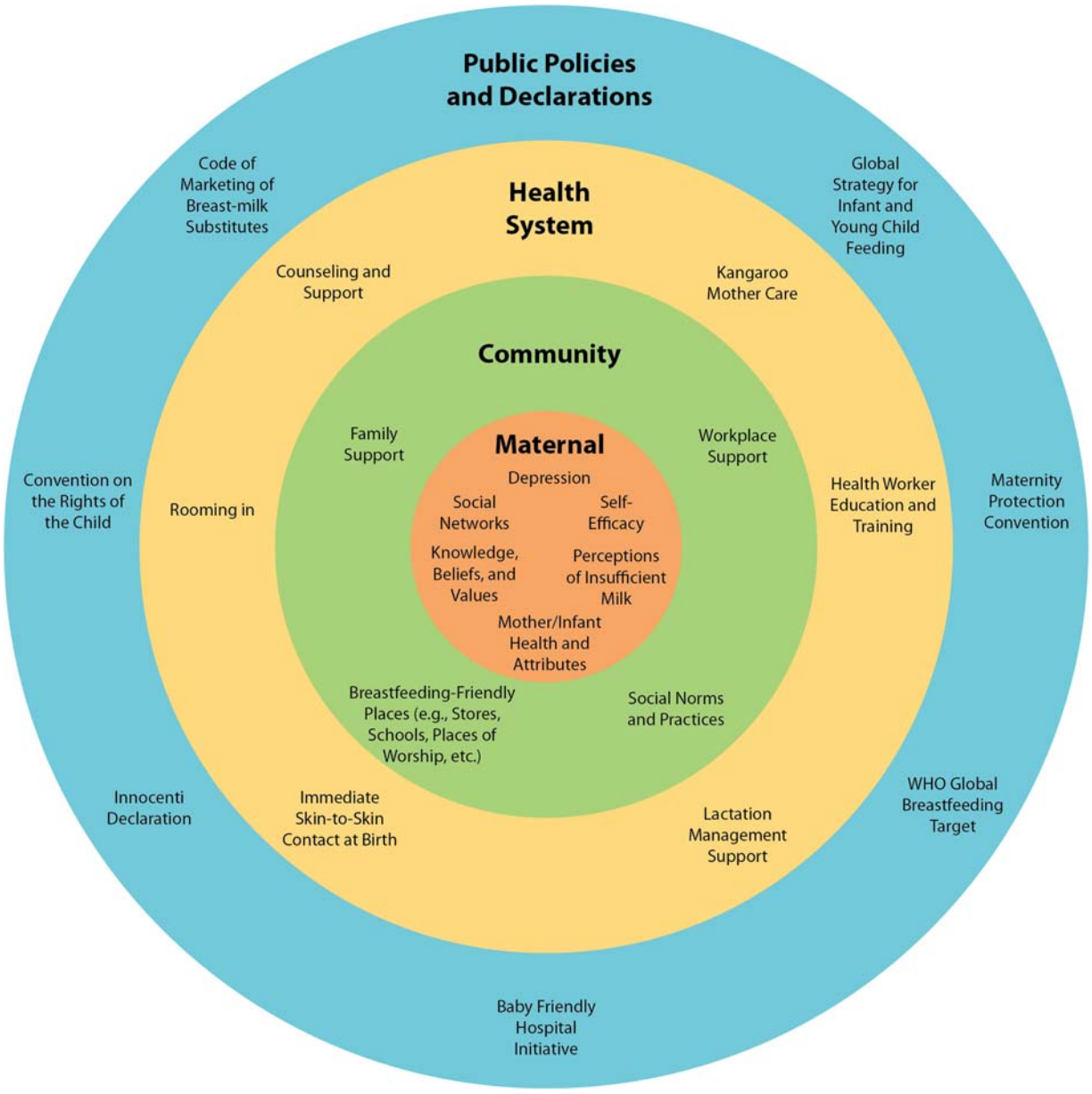


Fig. 1 Social-ecological model of breastfeeding.

Community

At the level of the community, important factors influencing breastfeeding include family support, workplace support, social norms and practices, and breastfeeding-friendly places that mothers frequent (e.g., stores, religious institutions, schools, etc.). Family support, especially from the child's father, has been shown to be important for positive breastfeeding outcomes. The results of a systematic review showed that father support was associated with a two-fold increase of exclusive breastfeeding at 6 months whereas lack of father support was associated with a risk of both full formula feeding and lactation-related problems. Returning to work outside the home is associated with early breastfeeding termination. A recent systematic review examined the effect of workplace-based breastfeeding/lactation support programs, policies, or interventions to promote breastfeeding among female employees. It found that mothers were more likely to maintain breastfeeding when employers provided the support that women need to do so, such as adequate space, tools, and time needed. It also found that it was possible for employers to implement a breastfeeding support program that fit their company's budget and resources.

Breastfeeding practices are influenced by descriptive social norms, or beliefs about the prevalence of breastfeeding by other women, and injunctive social norms, or beliefs about the degree of approval of the behavior. As part of a multiyear program to

improve infant and young child feeding practices in Bangladesh, an intensive social and behavior change communication intervention was implemented between 2009 and 2014. To evaluate the effectiveness, the program was implemented in both areas of intensive intervention and areas with less intensive intervention. Results of the evaluation showed that compared to mothers in the non-intensive areas, mothers in the intensive areas had larger social networks that increased over time. Compared to the non-intensive group, sharing of information by mothers in the intensive group increased by 17–23% points. Breastfeeding practices were also found to be associated with social networks, diffusion of information about infant and young child feeding, and social norms around breastfeeding. Although data are lacking, it seems intuitive that the more places that welcome breastfeeding mothers, the more likely she is to feel comfortable breastfeeding. Such places include stores, markets, schools, parks, religious institutions, etc. Pope Francis's support of breastfeeding in the Sistine Chapel in 2014 highlighted the importance of making mothers comfortable to breastfeed wherever they are.

Maternal

At the most basic level, maternal factors influence breastfeeding, as does the mother-infant relationship and attributes. A systematic review examined the associations between breastfeeding and mother-infant relationships, concluding that the association is highly complex but that how breastfeeding was carried out appeared to be a decisive factor. This was in turn influenced by a number of additional variables. Positive maternal knowledge, beliefs, and values about breastfeeding have long been known to be associated with good breastfeeding outcomes in multiple countries and settings. For example, in Ghana mothers with greater knowledge of exclusive breastfeeding were nearly 6 times more likely to practice exclusive breastfeeding. Higher scores of breastfeeding knowledge and attitude were also associated with exclusive breastfeeding. Maternal beliefs about what are normative infant and young child feeding practices, including breastfeeding, have been shown to be influenced by individuals within a mother's social network. Such networks include other mothers, husbands, mothers-in-law, grandparents, other family members or neighbors within the community, and healthcare providers. Social networks, diffusion of information, and social norms are also highly interrelated and can affect maternal infant feeding behaviors. Social networks developed through social media, which are becoming increasingly important, have been shown to lead to a longer duration of breastfeeding. Self-efficacy, a critically important attribute for breastfeeding, has been defined as an individual's belief in his or her capacity to execute behaviors necessary to produce specific performance attainments. A systematic review of observational studies found that maternal intention and breastfeeding-efficacy were important predictors of the duration of breastfeeding. Interventions to improve maternal breastfeeding self-efficacy have also been found to improve breastfeeding rates in mothers of full-term infants.

Researchers hypothesize that the role of breastfeeding in maternal depression has several effects. Breastfeeding can lead to hormonal changes that protect against postpartum depression, including by helping to regulate sleep and wake patterns of the mother and child and improve maternal self-efficacy and her emotional interaction with the child. At the same time, maternal depression can also lead to poorer breastfeeding outcomes, including decreased early initiation and duration of exclusive and any breastfeeding, increased difficulties, and decreased levels of self-efficacy. Interventions to prevent post-partum depression have been shown to be effective.

Insufficient breast milk is frequently cited as a reason for stopping exclusive and any breastfeeding. Although there is a biological basis for some instances of insufficient breast milk, the most common cause is perceived insufficiency by the mother. A recent systematic quantitative and qualitative review identified 11 studies in which mothers stopped practicing exclusive breastfeeding because of perceptions of insufficient breast milk to satisfy their infant's needs.

Breastfeeding protection, promotion, and support

Global policies for breastfeeding have been centered around three key themes: protection, promotion, and support, which are summarized in Box 2. The International Code of Marketing of Breast-milk Substitutes, commonly referred to as the Code, was approved by the World Health Assembly (WHA) in 1981 and is the key policy measure protecting breastfeeding. The Code resulted from compelling accounts of inappropriate and unethical marketing of breast-milk substitutes in low-income countries, which resulted in large numbers of young children becoming malnourished or dying as a result of contaminated or diluted infant formulas. It recognized that health workers, pregnant women, mothers of young children, and their families are susceptible to direct and indirect marketing strategies. The Code includes 11 articles and, since 1981, the WHA has passed a number of resolutions to clarify and strengthen aspects of the Code. Together these outline the responsibilities of governments, the health care system, and companies that manufacturer or market breast-milk substitutes. Although commonly thought to refer only to infant formula, the Code defines a breast-milk substitute as any food being marketed or otherwise represented as a partial or total replacement, whether or not suitable for that purpose. The Code represents the collective will of the member states of the WHO; however, countries must pass a regulation or law and an enforcement mechanism for it to be effective. Enforcement of country laws and regulations has generally been weak and violations frequently occur in many countries (WHO et al., 2020). Marketing by the large and growing breast-milk substitute industry continues to undermine breastfeeding as well as other aspects of young child feeding.

Breastfeeding is also protected through the Maternity Protection Convention, 2000 (No. 183) of the International Labor Organization (ILO), which recommends 14 weeks of maternity benefit for whom it applies. In addition, it stipulates women on

Box 2 Key policies and programs to protect, promote, and support breastfeeding

- WHO International Code of Marketing of Breast-milk Substitutes, 1981 and subsequent relevant World Health Assembly Resolutions
- Innocenti Declaration, 1989
- Baby Friendly Hospital Initiative, 1991 (updated in 2017)
- Convention on the Rights of the Child, 1997
- Maternity Protection Convention 183, 2000
- WHO/UNICEF Global Strategy for Infant and Young Child Feeding, 2003
- World Health Assembly Target for Exclusive Breastfeeding, 2014

maternity leave shall be entitled to a cash benefit, no less than two-thirds of her previous earnings, to ensure that she and her child can be maintained in proper living conditions. The convention also prohibits employers from terminating the employment of a woman during pregnancy and on maternity leave or during a period following her return to work, except on grounds not related to pregnancy, childbirth and its consequences, or breastfeeding. It further stipulates women must return to the same or equivalent position at the same pay. Lastly, it provides a woman with the right to one or more daily breaks or a daily reduction of hours of work to breastfeed. A subsequent ILO recommendation (No. 191) promotes the optimal maternity leave to be at least 18 weeks. Similar to the Code, countries must adopt the convention through legal or regulatory measures for it to take effect. Short maternity leaves, 6 weeks or less, increase the odds of not breastfeeding or stopping early by 400% and only 23% of countries meet or exceed the ILO's recommended 18 weeks of maternity leave (Victora et al., 2016).

The Innocenti Declaration on the Protection, Promotion, and Support of Breastfeeding, signed by multiple countries in 1990, set an international agenda with ambitious targets for action. The Declaration identified four operational targets: (1) appoint a national breastfeeding coordinator of appropriate authority and establish a multisectoral breastfeeding committee; (2) ensure that every facility providing maternity services practice all 10 steps to successful breastfeeding; (3) take action to give effect to the principles and aim of the Code and subsequent relevant WHA resolutions in their entirety, and (4) enact imaginative legislation protecting the breastfeeding rights of working women and established means for its enforcement. It also stated that breastfeeding of all infants should be a global goal for optimal maternal and child health and nutrition and promoted the practice of exclusive breastfeeding.

The Baby Friendly Hospital Initiative (BFHI), launched by WHO and UNICEF in 1991, promotes and supports breastfeeding in the health system, particularly around the time of birth. It outlines 10 steps facilities that provide maternity care should follow to become certified as "baby friendly." The recommendations to support the initiative were recently updated in which the 10 steps were separated into critical management procedures and key clinical practices. Detailed implementation guidance of the recommendations is also available. Health worker training has been a key component of the initiative and a course has been developed for this purpose. A recent systematic review on the impact of the BFHI on breastfeeding and child health outcomes concluded that the initiative led to a positive impact on short-term, medium-term, and long-term breastfeeding outcomes (Pérez-Escamilla et al., 2016). It also found a dose-response relationship between the number of steps a woman was exposed to and the likelihood of early initiation of breastfeeding, exclusive breastfeeding at hospital discharge, any breastfeeding, and the duration of exclusive breastfeeding. Community support, the 10th step, was found to be essential for sustaining the breastfeeding impacts in the long term.

Article 24 of the Convention on the Rights of the Child spells out the obligations to reduce infant and young child mortality and combat disease and malnutrition by taking measures to ensure that all sectors of society, including parents, know about the advantages of breastfeeding.

The aim of the Global Strategy for Infant and Young Child Feeding, endorsed by the WHA in 2002, is to improve young child nutrition through optimal feeding and included for the first time the recommendation to exclusively breastfeed for the first 6 months. The strategy is intended as a guide for action and identifies a set of evidence-based interventions and emphasizes the importance of providing mothers and families with the support they need to provide optimal nutrition for their young children. Similar to the Code, it also defines the obligations and responsibilities of governments, international organizations, and other important stakeholders.

In 2014, the WHA endorsed 6 global nutrition targets to be achieved by 2025. The importance of breastfeeding is reflected by the fact that exclusive breastfeeding is included as 1 of the 6 targets; specifically, to increase the rate of exclusive breastfeeding in the first 6 months up to at least 70%. A process indicator to measure interventions needed to achieve this goal is the number of births that occur in baby-friendly facilities.

Breastfeeding rates and trends

Endorsement by the WHA of the Code initiated a concerted global and national-level efforts to improve breastfeeding rates and provide conditions possible for women to breastfeed over the past 40 years. Efforts have especially focused on early

initiation of breastfeeding and exclusive breastfeeding. To measure progress, WHO and UNICEF have defined a set of indicators to measure breastfeeding rates, which include ever breastfed, early initiation of breastfeeding, exclusively breastfed for the first 2 days after birth, exclusive breastfeeding under 6 months, mixed-milk feeding under 6 months, and continued breastfeeding 12–15 months.

Improvements have been seen in some indicators, including early initiation of breastfeeding and exclusive breastfeeding, while declines have been seen in others such as breastfeeding 12–15 months (Rollins et al., 2016). Globally, exclusive breastfeeding among infants less than 6 months has increased from 35% in 2005 to 42% in 2018 though progress is uneven when measured by world regions (Fig. 2). As of about 2018, less than half of all newborns benefited from early initiation of breastfeeding (44%) and about the same percentage (43%) of newborns received only breast milk during the first 3 days (Fig. 3). Progress has also not been even among socioeconomic groups. Although the WHO global target for exclusive breastfeeding is that 70% of infants less than 6 months of age are exclusively breastfed, currently the rate is only 42%. Although all children should continue to be breastfed for 2 years or more, breastfeeding at 1 year (between 12 and 15 months) was 70% and breastfeeding at 2 years (between 20 and 23 months) only 45% (Fig. 3). Detailed information on early initiation of breastfeeding, exclusive breastfeeding, and continued breastfeeding, by country, are available from UNICEF, which tracks infant and young child feeding through its global databases.

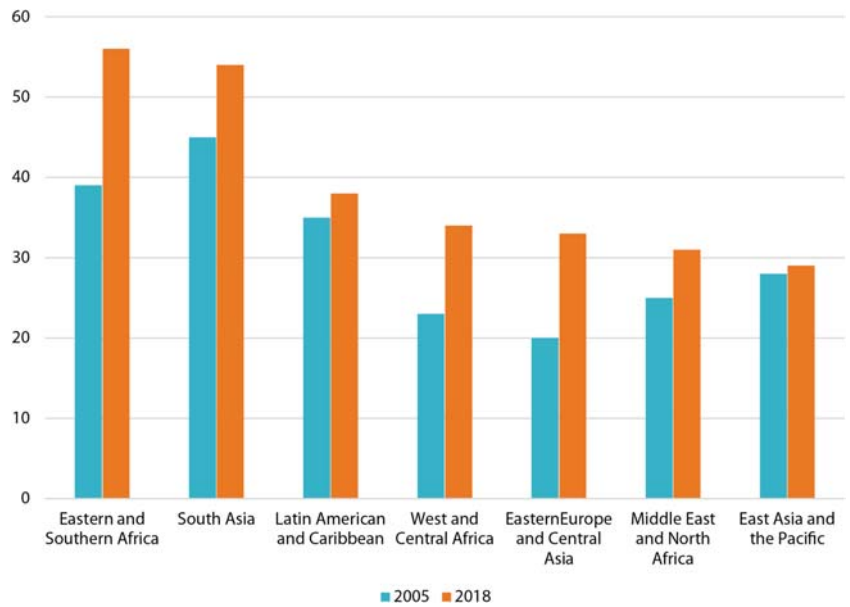


Fig. 2 Regional trends in exclusive breastfeeding, 2005–2018. Based on data presented in Figure 2.2 in UNICEF. 2019. The State of the World's Children 2019. Children, Food and Nutrition. Growing Well in a Changing World. UNICEF, New York.

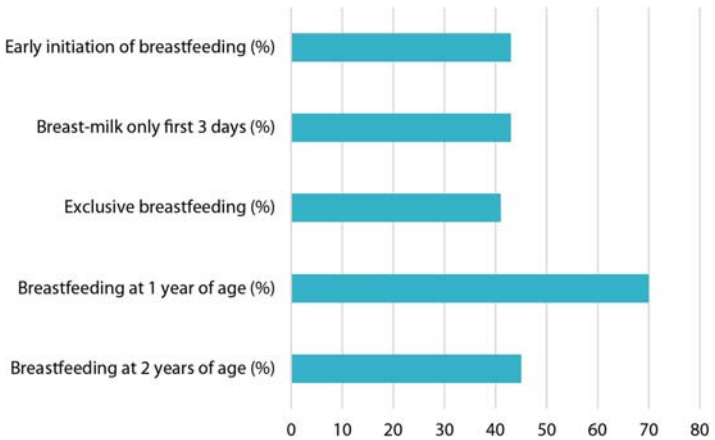


Fig. 3 Global rates of selected breastfeeding indicators, 2018. Based on data presented in UNICEF. 2019. The State of the World's Children 2019. Children, Food and Nutrition. Growing Well in a Changing World. UNICEF, New York, pp. 78–79.

The Global Breastfeeding Collective, coordinated by WHO and UNICEF and composed of numerous non-governmental organizations, has published a global breastfeeding scorecard every year since 2017. It also has defined 8 indicators to measure levels of investment, supportive policies and programs, and monitoring systems to gauge progress toward established goals to be measured annually at the country level. Progress by country, however, is uneven. For example, only 6% of donors contribute at least US $5 per newborn to support breastfeeding, only 18% fully implement the Code, only 14% of births are in Baby-friendly facilities, 56% of primary healthcare facilities provide counseling on infant and young child feeding, 47% of districts have community infant and young child feeding programs, 41% have had breastfeeding programs assessed in the last 5 years, and 36% have had breastfeeding data collected in the past 5 years.

Conclusion

Evidence continues to grow around the positive effects of breastfeeding on maternal health and child health and development in all countries, regardless of their level of wealth. However, for women to be made aware of breastfeeding's benefits and to meet their breastfeeding goals, an enabling environment is essential. Creating this environment is the responsibility of governments, employers, the community, and family members. Creating an enabling environment is also the responsibility of manufacturers and distributors of infant formula and other breast-milk substitutes; they must comply with the spirit and aim of The International Code of Marketing of Breastmilk Substitutes. The past 20 years have seen increases in early initiation of breastfeeding and exclusive breastfeeding; however, breastfeeding duration has decreased throughout much of the world. At present, countries are not on track to reach the global target for exclusive breastfeeding set by the WHA. Further investments in breastfeeding protection, promotion, and support are needed so that all mothers and children can benefit from the multiple health and developmental advantages provided by breast milk and through the act of breastfeeding.

References

Ballard, O., Morrow, A.L., 2013. Human milk composition: nutrients and bioactive factors. Pediatr. Clin. N. Am. 60, 49–74.

Christian, P., Smith, E.R., Lee, S.E., Vargas, A.J., Bremer, A.A., Raiten, D.J., 2021. The need to study human milk as a biological system. Am. J. Clin. Nutr. 113, 1063–1072.

Grummer-Strawn, L.M., Rollins, N., 2015. Impact of breastfeeding on maternal and child health. Acta Paediatr. 104, 1–134.

Marshall, N.E., Lau, B., Purnell, J.Q., Thornburg, K.L., 2019. Impact of maternal obesity and breastfeeding intention on lactation intensity and duration. Matern. Child Nutr. e12732.

McFadden, A., Gavine, A., Renfrew, M.J., Wade, A., Buchanan, P., Taylor, J.L., Veitch, E., Rennie, A.M., Crowther, S.A., Neiman, A., MacGillivray, S., Cochrane Pregnancy and Childbirth Group, 2017. Support for healthy breastfeeding mothers with healthy term babies. Cochrane Database Syst. Rev. 2, CD001141.

Moore, E.R., Bergman, N., Anderson, G.C., Medley, N., 2016. Early skin-to-skin contact for mothers and their healthy newborn infants. Cochrane Database Syst. Rev. 11.

Perez-Escamilla, R., Martinez, J.L., Segura-Perez, A., 2016. Impact of the Baby-friendly Hospital Initiative on breastfeeding and child health outcomes: a systematic review. Matern. Child Nutr. 12, 402–417.

Rollins, N.C., Bhandari, N., Hajeebhoy, N., Horton, S., Lutter, C.K., Martines, J.C., Piwoz, E.G., Richter, L.M., Victora, C.G., On Behalf of the Lancet Series Breastfeeding Group, 2016. Why invest, and what it will take to improve breastfeeding practices? Lancet 386.

Taren, D., Lutter, C.K., 2017. The role of breastfeeding protection, promotion and support in a developing world. In: De Pee, S., et al. (Eds.), Nutrition and Health in a Developing World. Springer Science+Business Media, New York.

Victora, C.G., Bahl, R., Barros, A.J.D., Franca, G.V.A., Horton, S., Krasevec, J., Murch, S., Sankar, M.J., Walker, N., Rollins, N.C., On Behalf of the Lancet Series Breastfeeding Group, 2016. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. Lancet 387, 1–16.

WHO, 2019. Breastfeeding and COVID-19. World Health Organization, Geneva.

WHO, UNICEF, IBFAN, 2020. National Implementation of the International Code, Status Report 2020. World Health Organization, Geneva.

Further reading

Baker, P., Melo, T., Neves, P.A.R., Machado, P., Smith, J., Piwoz, E., Barros, J.J.D., Victora, C.G., McCoy, D., 2020. First-food systems transformations and the ultra-processing of infant and young child diets: the determinants, dynamics and consequences of the global rise in commercial milk formula consumption. Matern. Child Nutr. e13097.

Bergera, P.K., Plowsa, J.F., Demerathb, E.W., Fields, D.A., 2020. Carbohydrate composition in breast milk and its effect on infant health. Curr. Opin. Clin. Nutr. Metab. Care 23, 1–5.

Black, R., McLaughlin, M., Giles, M., 2020. Women's experience of social media breastfeeding support and its impact on extended breastfeeding success: a social cognitive perspective. Br. J. Health Psychol. 25, 754–771.

Dinour, L.M., Szaro, J.M., 2017. Employer-based programs to support breastfeeding among working mothers: a systematic review. Breastfeed. Med. 12, 131–141.

Dror, D., Allen, L.H., 2018. Overview of nutrients in human milk. Adv. Nutr. 9, 278S–294S.

Koletzo, B., 2016. Human milk lipids. Ann. Nutr. Metab. 69, 28–40.

WHO, 2016. Updates on HIV and Infant Feeding Guideline. World Health Organization, Geneva.

WHO, UNICEF, 2018. Protecting, Promoting, and Supporting Breastfeeding in Facilities Providing Maternity and Newborn Services: Implementing the Revised Baby Friendly Hospital Initiative 2018. World Health Organization and the United Nations Children's Fund (UNICEF), Geneva.

Children: Nutritional requirements

I. Thorsdottir and B. Thorisdottir, Unit for Nutrition Research, Health Science Institute, School of Health Sciences, University of Iceland, Reykjavík, Iceland

© 2023 Elsevier Ltd. All rights reserved.

This is an update of I. Thorsdottir, Children: Nutritional Requirements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 326–334, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00050-7.

Introduction 31
Methodology for definition of nutrient requirements and recommended intakes 34
Energy 34
Carbohydrate 34
Fat 36
Proteins 37
Physical activity 38
Water 38
Sodium 38
Potassium 38
Bone minerals: phosphorous, calcium, and magnesium 38
Phosphorous 38
Calcium 38
Magnesium 38
Essential trace elements 39
Iron 39
Zinc 39
Iodine 39
Selenium 39
Copper 39
Water-soluble vitamins 40
Thiamine (vitamin B_1) 40
Riboflavin (vitamin B_2) 40
Niacin (vitamin B_3) 40
Vitamin B_6 40
Folate 40
Vitamin B_{12} 40
Vitamin C 40
Fat-soluble vitamins 41
Vitamin A 41
Vitamin D 41
Vitamin E 41
Conclusion 41
References 41

Key points

- Nutrient requirements and recommended intakes for children older than 1 year of age are discussed.
- Methodology used to construct the values is explained.
- Physiological role of nutrients and possible symptoms of deficiency and excessive intake of nutrients in children are mentioned.
- Nutritional terminology used in different countries and organizations is presented.
- A few reasons for similarities and dissimilarities in recommended nutrient intakes between these organizations are discussed.

List of abbreviations

µg microgram(s)
E% percent of total energy coming from the energy-giving nutrient

g gram(s)
kcal kilocalorie, 1 kcal = 4.184 kJ
kg kilogram
MJ megajoule, 1 MJ = 1000 kJ = 239 kcal
PUFA(s) polyunsaturated fatty acid(s)

Introduction

Proper nutrition in early life is very important. Nutritional status is known to affect the rate of growth, maturation, learning ability, and neurological development. It also can affect health in adulthood, such as risk of cardiovascular disease, obesity, and type 2 diabetes. Poor nutrient intake may affect the immune system, making one more susceptible to infections and unhealth.

This article will describe nutrient requirements for children between 1 and 18 years of age. It includes the methodology for determining nutrient requirements and recommended intakes and reviews the physiological importance of select nutrients, including the risks of inadequate and excessive intakes. **Table 1** shows the nomenclature used by authorities in select countries and institutes in setting nutrient reference values, a set of recommendations for daily nutritional intake for groups of individuals. Note that several different names exist for similar definitions. Additionally, most authorities set reference values for energy and recommended intake ranges for macronutrients.

Nutrient needs vary throughout the lifespan and correspond with growth velocity, stages of maturation, and other factors. The requirement for a particular nutrient is defined by the physiological role that it plays in the body and its bioavailability from the diet. If the requirement is not met, symptoms of nutrient deficiency may appear, and severe deficiency can lead to disease. The recommended dietary intake set for a nutrient meets the needs of 97.5% of individuals. **Tables 2–5** shows the daily recommended dietary intakes for minerals, trace elements, water-soluble vitamins, and fat-soluble vitamins set by authorities in select countries and institutes: Australia and New Zealand (National Health and Medical Research Council, 2006), the DACH countries Germany, Austria and Switzerland (Die Referenzwerte für die Nährstoffzufuhr, 2021), the European Food Safety Authority (EFSA) of the

Table 1 Terminology for nutrition recommendations.

Authority	*Intake threshold below which is risk of adverse effects*	*Intake meeting needs of 50% of individuals (average)*	*Intake meeting needs of 97–98% of individuals (average +2SD)*	*Less evidence-based data available*	*Intake threshold above which is risk of adverse effects*
Australia/ New Zealand		Estimated average requirement (EAR)	Recommended dietary intake (RDI)	Adequate intake (AI)	Upper level of intake (UL)
DACH[a]			Recommended nutrient intake (RNI)	Adequate intake (AI)	Tolerable upper intake levels (UL)
EFSA[b]	Lower threshold intake (LTI)	Average requirement (AR)	Population reference intake (PRI)	Adequate intake (AI)	Tolerable upper intake levels (UL)
NNR[c]	Lower intake level (LI)	Average requirement (AR)	Recommended intake (RI)		Upper intake level (UL)
UK	Lower reference nutrient intake (LRNI)	Estimated average requirement (EAR)	Reference nutrient intake (RNI)	Safe intake	Upper intake
USA/ Canada		Estimated average requirement (EAR)	Recommended dietary allowance (RDA)	Adequate intake (AI)	Tolerable upper intake level (UL)
WHO/FAO[d]		Estimated average requirement (EAR)	Recommended nutrient intake (RNI)		Upper tolerable nutrient intake level (UL)

[a]Recommendations for three central European countries: Germany (D), Austria (A), Switzerland (CH).
[b]The European Food Safety Authority (EFSA) of the European Union.
[c]Nordic Nutrition Recommendations (NNR) for Denmark, Finland, Iceland, Norway, and Sweden.
[d]World Health Organization (WHO) and Food and Agriculture Organization (FAO) of the United Nations.

Table 2 Recommendations for minerals for children, for all or for female/male (F/M).

Nutrient	*Age (y)*	*Australia/NZ*	*DACH*	*EFSA*	*NNR*	*UK*	*USA/Canada*	*WHO/FAO*
Sodium (mg day^{-1})[a,b]	1–3	200–400	400	1100	1400	500	800	
	4–6	300–600	500	1300	1400	700	1000	
	7–8	300–600	750	1700	1400	1200	1000	
	9	400–800	750	1700	1400	1200	1200	
	10	400–800	1100	1700	2400	1200	1200	
	11–12	400–800	1100	2000	2400	1600	1200	
	13	400–800	1400	2000	2400	1600	1200	
	14	460–920	1400	2000	2400	1600	1500	
	15–18	460–920	1500	2000	2400	1600	1500	
Potassium (mg day^{-1})[a,b]	1	2000	1100	800	1400	800	2000	
	2–3	2000	1100	800	1800	800	2000	
	4–5	2300	1300	1100	1800	1100	2300	
	6	2300	1300	1100	2000	1100	2300	
	7–8	2300	2000	1800	2000	2000	2300	
	9	2500/3000	2000	1800	2000	2000	2300/2500	
	10	2500/3000	2900	1800	2900/3300	2000	2300/2500	
	11–12	2500/3000	2900	2700	2900/3300	3100	2300/2500	
	13	2500/3000	3600	2700	2900/3300	3100	2300/2500	
	14	2600/3600	3600	2700	3100/3500	3100	2300/3000	
	15–18	2600/3600	4000	3500	3100/3500	3500	2300/3000	
Phosphorous (mg day^{-1})[b]	1–3	460	500	250	470	270	460	
	4–5	500	600	440	470	350	500	
	6	500	600	440	540	350	500	
	7–8	500	800	440	540	450	500	
	9	1250	800	440	540	450	1250	
	10	1250	1250	440	700	450	1250	
	11–17	1250	1250	640	700	625/775	1250	
	18	1250	1250	550	700	625/775	1250	
Calcium (mg day^{-1})	1–3	500	600	450	600	350	700	500
	4–5	700	750	800	600	450	1000	600
	6	700	750	800	700	450	1000	600
	7–8	700	900	800	700	550	1000	700
	9	1000–1300	900	800	700	550	1300	700
	10	1000–1300	1100	800	900	550	1300	1300
	11–12	1000–1300	1100	1150	900	800/1000	1300	1300
	13	1000–1300	1200	1150	900	800/1000	1300	1300
	14–17	1300	1200	1150	900	800/1000	1300	1300
	18	1300	1200	1000	900	800/1000	1300	1300
Magnesium (mg day^{-1})[b]	1	80	170	170	85	85	80	60
	2	80	170	170	120	85	80	60
	3	80	170	230	120	85	80	60
	4–5	130	190	230	120	120	130	76
	6	130	190	230	200	120	130	76
	7–8	130	240	230	200	200	130	100
	9	240	240	230	200	200	240	100
	10	240	230/260	250/300	280	200	240	220
	11–12	240	230/260	250/300	280	280	240	220
	13	240	240/280	250/300	280	280	240	220
	14	360/410	240/280	250/300	280/350	280	360/410	220
	15–17	360/410	260/330	250/300	280/350	300	360/410	220
	18	360/410	260/330	300/350	280/350	300	360/410	220

[a]Adequate intake (AI) for Australia/New Zealand and USA/Canada.

[b]Adequate intake (AI) for EFSA.

Data sources: see list of references.

Table 3 Recommendations for trace minerals for children, for all or for female/male (F/M).

Nutrient	*Age (y)*	*Australia/NZ*	*DACH*	*EFSA*	*NNR*	*UK*	*USA/Canada*	*WHO/FAO*
Iron (mg day^{-1})[a]	1–3	9	8	7	8	6.9	7	5.8
	4–5	10	8	7	8	6.1	10	6.3c
	6	10	8	7	9	6.1	10	6.3c
	7–8	10	10	11	9	8.7	10	8.9c
	9	8	10	11	9	8.7	8	8.9c
	10	8	15/12	11	11	8.7	8	14.0/14.6
	11	8	15/12	11	11	14.8/11.3	8	14.0/14.6
	12–13	8	15/12	13/11	11	14.8/11.3	8	32.7/14.6
	14	15/11	15/12	13/11	15/11	14.8/11.3	15/11	32.7/14.6
	15–17	15/11	15/12	13/11	15/11	14.8/11.3	15/11	31.0/18.8
	18	15/11	15/12	16/11	15/9	14.8/11.3	15/11	14.0/18.8
Zinc (mg day^{-1})	1	3	3	4.3	5	5	3	4.1[d]
	2–3	3	3	4.3	6	5	3	4.1[d]
	4–5	4	4	5.5	6	6.5	5	4.8[d]
	6	4	4	5.5	7	6.5	5	4.8[d]
	7–8	4	6	7.4	7	7	5	5.6[d]
	9	6	6	7.4	7	7	8	5.6[d]
	10	6	8/9	7.4	8/11	7	8	7.2/8.6[d]
	11–12	6	8/9	10.7	8/11	9	8	7.2/8.6[d]
	13	6	10/12	10.7	8/11	9	8	7.2/8.6[d]
	14	7/13	10/12	10.7	9/12	9	9/11	7.2/8.6[d]
	15–17	7/13	11/14	11.9/14.2	9/12	7/9.5	9/11	7.2/8.6[d]
	18	7/13	11/14	7.5–12.7/9.4–16.3[e]	7/9	7/9.5	9/11	7.2/8.6[d]
Iodine (μg day^{-1})[b]	1	90	100	90	70	70	90	90
	2–3	90	100	90	90	70	90	90
	4–5	90	120	90	90	100	90	90
	6	90	120	90	120	100	90	120
	7–8	90	140	90	120	110	90	120
	9	120	140	90	120	110	120	120
	10	120	180	90	150	110	120	120
	11–12	120	180	120	150	130	120	120
	13	120	200	120	150	130	120	150
	14	150	200	120	150	130	150	150
	15–17	150	200	130	150	140	150	150
	18	150	200	150	150	140	150	150
Selenium (μg day^{-1})[b]	1	25	15	15	20	15	20	17
	2–3	25	15	15	25	15	20	17
	4–5	30	20	20	25	20	30	22
	6	30	20	20	30	20	30	22
	7–8	30	30	35	30	30	30	21
	9	50	30	35	30	30	40	21
	10	50	45	35	40	30	40	26/32
	11–12	50	45	55	40	45	40	26/32
	13	50	60	55	40	45	40	26/32
	14	60/70	60	55	50/60	45	55	26/32
	15–18	60/70	60/70	70	50/60	60/70	55	26/32
Copper (mg day^{-1})[b,c]	1	0.7	0.5–1	0.7	0.3	0.4	0.34	0.56
	2	0.7	0.5–1	0.7	0.4	0.4	0.34	0.56
	3	0.7	0.5–1	1	0.4	0.4	0.34	0.56
	4–5	1	0.5–1	1	0.4	0.6	0.44	0.57
	6	1	0.5–1	1	0.5	0.6	0.44	0.57
	7–8	1	1–1.5	1	0.5	0.7	0.44	0.75
	9	1.1/1.3	1–1.5	1	0.5	0.7	0.7	0.75
	10	1.1/1.3	1–1.5	1.1/1.3	0.7	0.7	0.7	0.75
	11–13	1.1/1.3	1–1.5	1.1/1.3	0.7	0.8	0.7	1.0
	14	1.1/1.5	1–1.5	1.1/1.3	0.9	0.8	0.89	1.0
	15–17	1.1/1.5	1–1.5	1.1/1.3	0.9	1	0.89	1.15/1.33
	18	1.1/1.5	1–1.5	1.3/1.6	0.9	1	0.89	1.15/1.33

[a]Higher recommendation for girls after start of menarche.

[b]Adequate intake (AI) for EFSA.

European Union (Dietary reference values for nutrients: Summary report, 2017), the Nordic countries Denmark, Finland, Iceland, Norway, and Sweden (Nordic Nutrition Recommendations, 2012), United Kingdom (Government recommendations, 2016; Salt and Health), United States and Canada (Oria et al., 2019; Ross et al., 2011; Nutrition and Physical Activity), and the World Health Organization (WHO) and Food and Agriculture Organization (FAO) of the United Nations (Vitamin and mineral requirements, 2005). The age-groups in the tables may differ between nutrients since they are chosen to represent the exact ages set by different authorities. The origin of the figures in the tables can be found in the article List of references.

Methodology for definition of nutrient requirements and recommended intakes

Studies on specific energy levels and nutrient requirements for children are scarce. Variations between countries may be due to distinct dietary habits and nutrient bioavailability, differences in rate of childhood growth and average body size, different climates, or how the nutrient requirement is categorized. For example, countries categorize the needs for children by using different age ranges.

The estimated amounts of required nutrients for children are determined by several methods. Extrapolation from infant and adult data is the usual approach to estimate the required amount of nutrients in childhood. The factorial approach defines the requirement in a two-step equation that includes maintenance and growth. The amount required for maintenance is calculated from clinical trials, which estimate unavoidable losses during a period of negligible intake. The amount required for growth is a calculated accretion of the nutrient in the body. Another method of determining requirements is through balance studies, where nutrient intake is manipulated to equal losses. The required amount of a nutrient also can be established through avoiding deficiency symptoms that represent a biological nutrient inadequacy, such as iron deficiency anemia.

The World Health Organization has suggested using the grading of recommendations, assessment, development, and evaluation (GRADE) system for judging the quality of evidence for nutritional recommendations. The highest grades are on high-quality randomized clinical trials. The World Cancer Research Fund has developed recommendations based on relationships between nutrition and health or disease (e.g., cancer), which are evaluated as convincing or probable. Such an approach is especially valuable for recommendations regarding the energy-providing nutrients and development of food-based dietary guidelines.

Energy

Reference values for children are based on energy required per kilogram of body weight. The stage of childhood growth dictates specific energy requirements. Other components in calculating energy needs are the same as for adults, i.e., basal metabolic rate, diet-induced thermogenesis, and physical activity. Basal metabolic rate comprises the majority of energy requirements and is largely determined by fat-free mass, more so by the vital organs than skeletal muscles. Young children have twice or more the energy requirement of adults per kg body weight, depending on body composition and growth stage. Higher fat mass lowers the energy requirement, an important factor to consider when determining reference values. After the first year of life, the fraction of energy intake used for growth decreases rapidly. By 1–3 years of age, approximately 3% of energy intake is used for growth, and thereafter less than 2%. Although still important, the energy needed for growth is a relatively small proportion of the total energy intake for children after 1 year of age. However, the energy cost of growth still must be taken into consideration.

Basal metabolic rate may be estimated using equations based on age, weight, height, and gender. Total energy expenditure is estimated by adding the energy required for physical activity, diet-induced thermogenesis (approximately 10%), and growth (approximately 3% at the age of 1–3 years, and less than 2% in older children).

As energy requirements depend on several different factors, variation in reference values exists among countries. Growth standards have been used to estimate the long-term adequacy of energy intake. Feeding practices in infancy and childhood, genetic variations, and other possible determinants of growth also will contribute to differences in growth standards.

Carbohydrate

Daily requirements of carbohydrate intake (g per day or g per kilogram body weight per day) have not been defined as they have for total energy and essential nutrients. Very low carbohydrate intake causes ketosis and can be avoided by ingesting only a small

[c]Adequate intake (AI) for Australia/New Zealand.

[d]In settings of moderate bioavailability, mixed diets containing animal or fish protein.

[e]Estimated ARs and Population Reference Intakes (PRIs) for zinc are provided for phytate intake levels of 300, 600, 900 and 1200 mg/day, which cover the range of mean/median phytate intakes observed in European populations. PRIs for adults were estimated as the zinc requirement of individuals with a body weight at the 97.5th percentile for reference body weights for men and women.

Data sources: see list of references.

Table 4 Recommendations for water-soluble vitamins for children, for all or for female/male (F/M).

Nutrient	*Age (y)*	*Australia/NZ*	*DACH*	*EFSA*	*NNR*	*UK*	*USA/Canada*	*WHO/FAO*
Thiamine (mg day^{-1})	1	0.5	0.6	0.1	0.5	0.3	0.5	0.5
	2–3	0.5	0.6	0.1	0.6	0.4	0.5	0.5
	4–5	0.6	0.7	0.1	0.6	0.6	0.6	0.6
	6	0.6	0.7	0.1	0.9	0.6	0.6	0.6
	7–8	0.6	0.8/0.9	0.1	0.9	0.7	0.6	0.9
	9	0.9	0.8/0.9	0.1	0.9	0.7	0.9	0.9
	10	0.9	0.9/1	0.1	1.0/1.1	0.7	0.9	1.1/1.2
	11–12	0.9	0.9/1	0.1	1.0/1.1	0.8/1	0.9	1.1/1.2
	13	0.9	1/1.2	0.1	1.0/1.1	0.8/1	0.9	1.1/1.2
	14	1.1/1.2	1/1.2	0.1	1.2/1.4	0.8/1	1/1.2	1.1/1.2
	15–17	1.1/1.2	1.1/1.4	0.1	1.2/1.4	0.8/1	1/1.2	1.1/1.2
	18	1.1/1.2	1.1/1.4	0.1	1.1/1.4	0.8/1	1/1.2	1.1/1.2
Riboflavin (mg day^{-1})	1	0.5	0.7	0.6	0.6	0.6	0.5	0.5
	2–3	0.5	0.7	0.6	0.7	0.6	0.5	0.5
	4–5	0.6	0.8	0.7	0.7	0.8	0.6	0.6
	6	0.6	0.8	0.7	1.1	0.8	0.6	0.6
	7–8	0.6	0.9/1	1.0	1.1	1.0	0.6	0.9
	9	0.9	0.9/1	1.0	1.1	1.0	0.9	0.9
	10	0.9	1/1.1	1.0	1.2/1.3	1.0	0.9	1/1.3
	11–12	0.9	1/1.1	1.4	1.2/1.3	1.1/1.2	0.9	1/1.3
	13	0.9	1.1/1.4	1.4	1.2/1.3	1.1/1.2	0.9	1/1.3
	14	1.1/1.3	1.1/1.4	1.4	1.4/1.7	1.1/1.2	1/1.3	1/1.3
	15–17	1.1/1.3	1.2/1.6	1.6	1.4/1.7	1.1/1.3	1/1.3	1/1.3
	18	1.1/1.3	1.2/1.6	1.6	1.3/1.6	1.1/1.3	1/1.3	1/1.3
Niacin (mg day^{-1})	1	6	8	1.6	7	4.7/5.0	6	6
	2–3	6	8	1.6	9	6.6/7.2	6	6
	4–5	8	9	1.6	9	9.1/9.8	8	8
	6	8	9	1.6	12	9.1/9.8	8	8
	7–8	8	10/11	1.6	12	11.2/12	8	12
	9	12	10/11	1.6	12	11.2/12	12	12
	10	12	11/13	1.6	14/15	11.2/12	12	16
	11–12	12	11/13	1.6	14/15	13.2/16.5	12	16
	13	12	13/15	1.6	14/15	13.2/16.5	12	16
	14	14/16	13/15	1.6	16/19	13.2/16.5	14/16	16
	15–17	14/16	13/17	1.6	16/19	13.2/16.5	14/16	16
	18	14/16	13/17	1.6	15/19	13.2/16.5	14/16	16
Vitamin B6 (mg day^{-1})	1	0.5	0.6	0.6	0.5	0.7	0.5	0.5
	2–3	0.5	0.6	0.6	0.7	0.7	0.5	0.5
	4–5	0.6	0.7	0.7	0.7	0.9	0.6	0.6
	6	0.6	0.7	0.7	1	0.9	0.6	0.6
	7–8	0.6	1	1.0	1	1.0	0.6	1
	9	1	1	1.0	1	1.0	1	1
	10	1	1.2	1.0	1.1/1.2	1.0	1	1.2/1.3
	11–12	1	1.2	1.4	1.1/1.2	1/1.2	1	1.2/1.3
	13	1	1.4/1.5	1.4	1.1/1.2	1/1.2	1	1.2/1.3
	14	1.2/1.3	1.4/1.5	1.4	1.3/1.6	1/1.2	1.2/1.3	1.2/1.3
	15–17	1.2/1.3	1.4/1.6	1.6/1.7	1.3/1.6	1.2/1.5	1.2/1.3	1.2/1.3
	18	1.2/1.3	1.4/1.6	1.6/1.7	1.2/1.5	1.2/1.5	1.2/1.3	1.2/1.3
Folate (μg day^{-1})	1	150	120	120	60	70	150	150
	2–3	150	120	120	80	70	150	150
	4–5	200	140	140	80	100	200	200
	6	200	140	140	130	100	200	200
	7–8	200	180	200	130	150	200	300
	9	300	180	200	130	150	300	300
	10	300	240	200	200	150	300	400
	11–12	300	240	270	200	200	300	400
	13	300	300	270	200	200	300	400
	14	400	300	270	300	200	400	400
	15–17	400	300	330	300	200	400	400
	18	400	300	330	400/300	200	400	400

(Continued)

Table 4 Recommendations for water-soluble vitamins for children, for all or for female/male (F/M).—cont'd

Nutrient	*Age (y)*	*Australia/NZ*	*DACH*	*EFSA*	*NNR*	*UK*	*USA/Canada*	*WHO/FAO*
Vitamin B12 (μg day^{-1})	1	0.9	1.5	1.5	0.6	0.5	0.9	0.9
	2–3	0.9	1.5	1.5	0.8	0.5	0.9	0.9
	4–5	1.2	2	1.5	0.8	0.8	1.2	1.2
	6	1.2	2	1.5	1.3	0.8	1.2	1.2
	7–8	1.2	2.5	2.5	1.3	1.0	1.2	1.8
	9	1.8	2.5	2.5	1.3	1.0	1.8	1.8
	10	1.8	3.5	2.5	2	1.0	1.8	2.4
	11–12	1.8	3.5	3.5	2	1.2	1.8	2.4
	13	1.8	4	3.5	2	1.2	1.8	2.4
	14	2.4	4	3.5	2	1.2	2.4	2.4
	15–17	2.4	4	4	2	1.5	2.4	2.4
	18	2.4	4	4	2	1.5	2.4	2.4
Vitamin C (mg day^{-1})	1	35	20	20	25	30	15	30
	2–3	35	20	20	30	30	15	30
	4–5	35	30	30	30	30	25	30
	6	35	30	30	40	30	25	30
	7–8	35	45	45	40	30	25	35
	9	40	45	45	40	30	45	35
	10	40	65	45	50	30	45	40
	11–12	40	65	70	50	35	45	40
	13	40	85	70	50	35	45	40
	14	40	85	70	75	35	65/75	40
	15–17	40	90/105	90/100	75	40	65/75	40
	18	40	90/105	95/110	75	40	65/75	40

Data sources: see list of references.

amount of carbohydrates. It is generally advised that carbohydrates should comprise 45–60% of total energy intake (45–60E%). This is partly to avoid very high fat and protein intake, but high-quality carbohydrates are of key importance to obtain the beneficial health effects of high-carbohydrate diet. Early introduction of high-quality carbohydrates is believed to have an impact on dietary habits throughout life. The quality is based on where the carbohydrates originate and the nutrient density of the food.

Intake of added sugars may be unfavorable for young children, as they are especially vulnerable to developing dental caries. In addition, a high-sugar diet generally includes foods with low nutrient value, and sugar-sweetened beverages have been associated with increased risk of overweight and obesity. It is widely recommended for children to avoid added sugar, and it should be less than 10% of total energy intake (10E%).

Some health authorities recommend fiber intake to equal the age of the child plus five (i.e., 7-year-old child = 12 g fiber day^{-1}). This concept has been criticized when determining fiber intake for older children. From the age of 2 years 2–3 g/MJ can be advised and school children can easily adopt the same recommendation as adults, 3 g fiber MJ^{-1} or 25–35 g day^{-1}. Fiber intake should be increased gradually in early childhood to reach the level of 3 g/MJ slowly.

Fat

Dietary requirements have been defined for the essential polyunsaturated fatty acids (PUFAs): linoleic acid (C18:2, *n*-6) and α-linolenic acid (C18:3, *n*-3). Fat gives energy in a concentrated form, and essential fatty acids are involved in important physiological functions in the body. They are required for regulation of renal function, blood coagulation, inflammatory and immunological reactions, and blood pressure control. Human physiology does not have the enzymes necessary to introduce double bonds in the *n*-3 and *n*-6 positions; therefore, these fatty acids must be obtained from the diet. Studies strongly suggest that the *n*-3 fatty acids are important for normal brain and vision development in children. Although the *n*-3 and *n*-6 fatty acids are essential, excess intake is possible.

Adequate intake of fat is necessary to meet the needs for essential fat-soluble vitamins and fatty acids. Dietary fat intake, providing less than 20% of total energy intake, can be inadequate. High intake of saturated fat is associated with cardiovascular disease. Decreases in intake of saturated and *trans*-fatty acids in the Nordic countries after 1970 and 1980 was followed by a reduction in prevalence and mortality rate of ischemic heart disease.

The recommended combination of dietary fat in most countries is similar for children as for adults, but for young children the recommended total fat intake is sometimes higher. The Nordic Nutrition Recommendations advise a total fat intake of 30–40E% for

Table 5 Recommendations for fat-soluble vitamins for children, for all or for female/male (F/M).

Nutrient	*Age (y)*	*Australia/NZ*	*DACH*	*EFSA*	*NNR*	*UK*	*USA/Canada*	*WHO/FAO*
Vitamin A ($\mu g\ day^{-1}$)[a]	1	300	300	250	300	400	300	400
	2–3	300	300	250	350	400	300	400
	4–5	400	350	300	350	400	400	450
	6	400	350	300	400	400	400	450
	7–8	400	450	400	400	500	400	500
	9	600	450	400	400	500	600	500
	10	600	600	400	600	500	600	600
	11–12	600	600	600	600	600	600	600
	13	600	700/800	600	600	600	600	600
	14	700/900	700/800	600	700/900	600	700/900	600
	15–17	700/900	800/950	650/750	700/900	600/700	700/900	600
	18	700/900	800/950	650/750	700/900	600/700	700/900	600
Vitamin D ($\mu g\ day^{-1}$)[b]	1	5	20	15b	10	10	15	5
	2–3	5	20	15b	10	10	15	5
	4–5	5	20	15b	10	10	15	5
	6	5	20	15b	10	10	15	5
	7–8	5	20	15b	10	10	15	5
	9	5	20	15b	10	10	15	5
	10	5	20	15b	10	10	15	5
	11–12	5	20	15b	10	10	15	5
	13	5	20	15b	10	10	15	5
	14	5	20	15b	10	10	15	5
	15–17	5	20	15b	10	10	15	5
	18	5	20	15b	10	10	15	5
Vitamin E ($mg\ day^{-1}$)[b,c,d]	1	5	5/6	6	4		6	5
	2–3	5	5/6	6–9	5		6	5
	4–5	6	8	9	5		7	5
	6	6	8	9	6		7	5
	7–8	6	9/10	9	6		7	7
	9	8/9	9/10	9	6		11	7
	10	8/9	11/13	11/13	7/8		11	7.5/10
	11–12	8/9	11/13	11/13	7/8		11	7.5/10
	13	8/9	12/14	11/13	7/8		11	7.5/10
	14	8/10	12/14	11/13	8/10		15	7.5/10
	15–17	8/10	12/15	11/13	8/10		15	7.5/10
	18	8/10	12/15	11/13	8/10		15	7.5/10

[a]Retinol equivalents (RE); 1 RE = 1 μg retinol = 12 μg β-carotene.
[b]1 mg α-tocopherol equivalence = α-TE = D-α-tocopherol = RRR-α-tocopherol.
[c]Adequate intake (AI) for Australia/New Zealand and EFSA.
[d]Best estimate of requirements for WHO/FAO, due to insufficient evidence.
Data sources: see list of references.

1–2-year-olds, but 25–40E% after the age of 2 years. In addition, energy from saturated fat should be lower than 10E% and *trans*-fatty acids as low as possible, energy from PUFAs 10–20E% with at least 1E% *n*-3 fatty acids, and energy from monounsaturated fatty acids 10–15E%.

Proteins

Dietary protein requirements include the essential amino acids and protein needs to maintain a positive nitrogen balance necessary for growth. Nitrogen is a part of amino acids, which are the building blocks of proteins. Proteins are important for the transport of various substances in the body, antibody action, enzyme functions, repairing processes, and building cellular structures. The quality of proteins depends on the combination of essential amino acids it contains. A deficiency of protein can result in protein energy malnutrition, a serious condition causing muscle weakness, changes in hair and skin, and edema. Malnourished or wasted children have higher protein requirements than a child of normal body composition and normal growth. Early protein malnutrition may lead to permanent impairment of cognitive functions. Conversely, a high protein intake early in life has been associated with the onset of overweight and obesity.

Infants, children, and even adolescents who have requirements for catch-up growth due to an earlier malnutrition or stunting need a larger portion of total energy intake from protein than the ordinary healthy child. Adequate protein intake from 1 year of age is generally equivalent to 1 g kg^{-1} body weight, and from 2 years of age approximately 0.9 g kg^{-1} body weight. In the Nordic Nutrition Recommendations, the recommended protein intake is 10–15E% for one to two years old children, and 10–20E% after 2 years of age which is the same as for adults.

Physical activity

Physical activity comprises part of the energy balance and contributes to the prevention of many noncommunicable diseases, such as cardiovascular disease, osteoporosis, and certain types of cancer and mental illness. For children and adolescents, a minimum of 1 h day^{-1} of physical activity has been advised.

Water

Water is an essential nutrient, but very few health authorities have defined the daily requirements and recommended intake of water. A general estimated requirement for children is 1 mL water $kcal^{-1}$ energy intake, or approximately 1.5–2 L of fluid day^{-1}. The requirement for water is quite variable, and needs change based on climate, physical activity, and diet.

Sodium

Sodium is part of salt (sodium chloride), which is a common food ingredient. Sodium ions are essential for many metabolic processes. There is no recommendation for sodium, and the required amount of sodium intake for children is not well known. Excess water loss increases the risk for hyponatremia and dehydration. Most Western countries have health problems related to high intake of sodium. Clinical trials show that even in childhood, salt intake is associated with high blood pressure. Based on these findings, it is beneficial to limit sodium intake in children to avoid the preference for a high-salt diet and to prevent later hypertension due to high sodium intake.

Potassium

Potassium is the major intracellular cation. Deficiency is very unlikely, as the average diet provides adequate amounts of potassium. Prolonged diarrhea, vomiting, or abnormal use of laxatives can cause excessive loss of potassium. There is evidence that potassium supplementation can decrease blood pressure, and a balance between potassium and sodium is important for blood pressure regulation.

Bone minerals: phosphorous, calcium, and magnesium

Phosphorous

Phosphorous is essential for bone health and an important nutrient during growth. It is widespread in the diet, and deficiency is very seldom observed. Phosphorous requirement is closely related to calcium, as it is a major part of the skeleton in the form of hydroxyapatite, which contains phosphorous and calcium in the ratio 1:2.

Calcium

Calcium is stored in the bone as hydroxyapatite. The bone continuously undergoes remodeling. Bone formation exceeds bone resorption in children, and their rate of remodeling is higher than in adults. A small but important role of calcium is in the bloodstream, extracellular fluids, and all cells in the body. The absorption of calcium is dependent on 1,25-dihydroxyvitamin D_3, the hormonal form of vitamin D. Calcium is absorbed more efficiently in the body during periods of increased physiological need, i.e., infancy, early childhood, puberty, and when dietary intake is low. The absorption can be diminished by certain factors, such as phytic acid, oxalic acid, or phosphates. Physical inactivity also increases bone resorption and loss of calcium. Risk of osteoporosis has been related to high sodium, high protein, and a lack of physical activity. Very high calcium intake can result in hypercalcemia, kidney stones, and kidney damage.

Magnesium

Magnesium is involved in many metabolic reactions and depends on vitamin D for its absorption. Deficiency is rare, but the symptoms are hypokalemia, hypercalcemia, neuromuscular hyperexcitability, and cardiac arrythmias. Epidemiological studies have suggested the importance of magnesium to protect against noncommunicable diseases, but the evidence is weak. Excess magnesium intake can cause diarrhea. However, if kidney function is normal, it is highly unlikely.

Essential trace elements

Recommended daily intake values exist for the essential trace minerals iron, zinc, iodine, selenium, and copper. Other essential minerals include molybdenum, manganese, chromium, and fluorine, but there is still limited information available about the requirements for these.

Iron

Iron has many essential functions in the body. An important function of iron is to alternate between two oxidation states, ferrous iron (Fe^{2+}) and ferric iron (Fe^{3+}). Iron forms the oxygen-binding part of hemoglobin, which transports oxygen from the lungs to the tissues, and myoglobin, which transports oxygen within the muscle. Absorption of nonheme iron is increased in diets with meat or fish, as well as in diets including vitamin C. Factors that inhibit iron absorption are phytic acids, tannins, and calcium. Iron-deficiency anemia reduces work capacity and impairs cell-mediated immunological defenses. Prolonged iron deficiency in children slows mental development and permanently affects cognitive function. Iron overload is also possible, particularly in a hereditary condition of nonselective high absorption.

Infants can utilize the iron stores that they are born with for several months. After 6 months of age, they require more iron than can be obtained solely from breast milk, and this is the reason for iron fortification of infant formula and food. It is estimated that more than 40% of infants and children below 5 years of age worldwide suffer from anemia, most often due to iron deficiency. The recommendation for iron in childhood is relatively high and is based on the need for iron and proportional absorption.

Zinc

The largest portion of zinc is located in the body cells. It is an essential part of many enzymes involved in metabolism, comprises part of the cell nucleus, and participates in gene expression. Mild deficiency symptoms are skin lesions and hair loss. Severe zinc deficiency leads to growth retardation and delayed sexual maturation. Excess zinc intake can occur with dietary supplements, and high intakes can reduce the absorption of copper, another essential mineral.

Iodine

Iodine is an essential component of the thyroid hormones, tetraiodothyronine and triiodothyronine, necessary for normal growth, development, and metabolism. Iodine deficiency is one of the most common nutrition disorders in the world and the leading cause of preventable brain damage. When physiological requirements for iodine are not met, a series of functional and developmental abnormalities occur as a result of thyroid dysfunction. The main deficiency symptom is goiter, which is characterized by enlargement of the thyroid gland.

When iodine deficiency is severe, cretinism can occur, resulting in impaired growth, mental disorders, and disturbances in speech. The recommended intake for children is based on the amount needed to prevent goiter, urinary iodine excretion, and data from adults. Iodine fortification of salt has decreased the incidence of deficiency worldwide. Iodine toxicity may occur when the intake of iodine-fortified foods is increased drastically in a short time period.

Selenium

Selenium is a trace mineral that has antioxidant properties. The main biological function of selenium is thought to be mediated through glutathione peroxidase and other selenoproteins. Low intake of selenium has been associated with cardiomyopathy, affecting children in particular. The recommended intake is based on the levels needed to maximize glutathione peroxidase activity or derived from studies on selenium-deficient children and adolescents. Toxicity is rare; however, excess intake has been associated with nausea, nail and hair deformities, and in very severe cases, peripheral nerve and liver damage.

Copper

Copper plays a role in the formation of connective tissue, defense against free radicals, and is a part of several enzymes involved in energy metabolism. Copper deficiency is uncommon but has been seen in premature and malnourished infants and children with chronic diarrhea. The symptoms of copper deficiency are leukopenia, anemia, and abnormal hair and skin pigmentation. Some organ dysfunctions have also been observed. Risk of copper deficiency may increase after breastfeeding is discontinued, especially when coupled with low dietary intake. Acute copper toxicity causes gastric pain, nausea, vomiting, and diarrhea. Copper requirements have been calculated from adult reference values.

Water-soluble vitamins

Most data on physiological requirements and recommendations for water-soluble vitamins in childhood are extrapolated from adult data. Deficiency of water-soluble vitamins is rare. The vitamins are widely found in food and easily absorbed, as well as excreted. Therefore, the risk of inadequate or excess intake is minimal. However, diseases associated with water-soluble vitamin deficiencies have been well described and will be discussed in this section.

Thiamine (vitamin B_1)

Thiamine is converted to its biologically active form, thiamine pyrophosphate, in the liver. It is involved in nerve and muscle function, and is essential for the utilization of carbohydrates and branched-chain amino acids. Beriberi is caused by thiamine deficiency. Symptoms include nervous system dysfunction, heart failure, muscle weakness, anorexia, and weight loss. These are generally more severe in children than in adults. Risk of thiamine toxicity is very low.

Riboflavin (vitamin B_2)

Riboflavin is important for the coenzymes flavine mononucleotide and flavine adenine dinucleotide, which are oxidizing agents. Recommended values are set by calculating the amount of riboflavin per energy or protein unit. Deficiency has been associated with poor iron status, and symptoms include skin changes, glossitis, anemia, and mental disturbances. Toxicity from excess riboflavin intake is rare.

Niacin (vitamin B_3)

Niacin can be formed in the body from tryptophan. It functions as the coenzymes nicotinamide adenine dinucleotide and nicotinamide adenine dinucleotide phosphate, which are involved in a number of redox reactions in the metabolism of all the energy-given nutrients. A deficiency of niacin causes pellagra, which has mainly been observed in populations with a corn (maize)-based diet. There is no known toxic effect of niacin from food. However, high doses of niacin in the form of nicotinic acid may induce liver damage.

Vitamin B_6

The main role of vitamin B_6 is in protein metabolism, where it acts as a coenzyme in the metabolism of amino acids. Dietary deficiency of vitamin B_6 alone is not common, and it is usually seen with other B-vitamin deficiencies. Symptoms of deficiency in infants and children include epileptic convulsions, weight loss, gastrointestinal problems, and anemia. The recommended amount is based on protein intake (0.015 mg g^{-1} protein) and is the same for adults. Toxicity is not common but has been seen with prolonged intake at 15 mg/day.

Folate

Folate is indirectly involved in the metabolism of amino acids and the synthesis of nucleic acids required for normal cell division. Signs of deficiency are observed first in rapidly replicating tissues, such as blood cells and bone marrow. Folate deficiency is the most common cause of megaloblastic anemia in childhood. The Nordic recommendation is based on body weight at 5 μg per kg per day. Folate toxicity cannot occur from dietary intake alone. Excess use of folate supplements may mask vitamin B_{12} deficiency.

Vitamin B_{12}

Cobalamin or Vitamin B_{12} is necessary for normal red blood cell formation and neurological functions. Vitamin B_{12} must be bound to a protein called intrinsic factor to be absorbed properly. Deficiency results in macrocytic megaloblastic anemia and in severe cases may cause neurological changes. There is no known risk of vitamin B_{12} toxicity. Children who are breastfed by vegan mothers may be at risk for vitamin B_{12} deficiency if the mother is not taking supplements. In general, B_{12} deficiency in childhood is rare. The recommended intake for children is approximately 0.05 μg kg^{-1} body weight.

Vitamin C

Vitamin C (ascorbic acid) is a potent antioxidant with multiple roles in the body. Deficiency of vitamin C causes scurvy. Signs of mild deficiency include reduced antioxidant capacity, fatigue, and irritability. The requirement for children is extrapolated from adult values and a growth factor. Toxicity is rare, but high intakes can cause diarrhea and gastrointestinal disturbances. Excess intake over a long period increases oxalate formation and risk of kidney stones. Vitamin C is important in childhood due to its ability to enhance the absorption of nonheme iron from porridge and cereal foods commonly given to young children. The iron level in cow's

milk is very low and infantile scurvy is regularly reported; therefore, formulas must be fortified with vitamin C. Some studies indicate that low vitamin C intake may be associated with asthma in childhood.

Fat-soluble vitamins

Most countries have set recommendations for vitamin A, D, and E, and some have developed recommendations for vitamin K, an important factor in blood clotting. This section discusses the requirements for vitamin A, D, and E.

Vitamin A

Vitamin A is important for vision, maintenance of epithelial surface, immune competence, growth, development, and reproduction. Vitamin A deficiency is one of the most common nutrient deficiencies in the world. Epidemiological and intervention studies on children found low intake of vitamin A and poor vitamin A status to be associated with an increased rate and severity of infections, and increased mortality from infectious diseases, e.g., measles. Retinol is the active form of vitamin A and is found in foods of animal origin. Inactive forms, e.g., β-carotene, are found in plant sources and can be activated in the body. Conditions that can affect the bioavailability and bioconversion of retinol and carotenoids are protein energy malnutrition, zinc deficiency, infections, and degree of food processing. There are no direct studies on vitamin A requirements in childhood; thus, the recommended intakes are extrapolated from studies on adults. High intake of vitamin A can cause hepatotoxicity, and during pregnancy, it increases the risk of infant malformations.

Vitamin D

Vitamin D is a prohormone converted to 1,25-dihydroxyvitamin D_3 in the body. Active vitamin D is a steroid-like substance that can be synthesized in the skin from ultraviolet B light exposure. In areas where exposure to sunlight is limited, vitamin D must be obtained from diet. A sufficient level of vitamin D in the body ensures that the concentrations of calcium and phosphate in plasma are regulated. Vitamin D enhances the absorption of calcium from the intestine and, together with parathyroid hormone, stimulates the release of calcium from bone, resulting in increased concentration of calcium in plasma. Vitamin D is essential through this mechanism for normal bone mineralization.

Vitamin D deficiency in childhood causes rickets. Insufficient intake also has been associated with cancer, autoimmune diseases, infections, and decreased muscle strength. Supplemental vitamin D early in life has been recommended. The recommended daily intake has been increasing slowly for some decades, and recently, the US recommended dietary allowance was highly elevated. Vitamin D can be toxic in large amounts; possible adverse effects are hypocalcaemia, nephrocalcinosis, and possible kidney failure.

Vitamin E

Vitamin E comprises two groups of components: tocopherols and tocotrienols. It is a fat-soluble antioxidant essential for neurological function. Vitamin E deficiency in children can occur with high intake of iron and PUFAs, or if they suffer from protein energy malnutrition. The recommendations for children are generally based on dietary intake of PUFAs. Vitamin E is less toxic than the other fat-soluble vitamins. However, high intakes may interfere with blood coagulation.

Conclusion

Proper nutrition in early life is very important. This article discusses the physiological role of nutrients and possible symptoms of deficiency and excessive intake of nutrients in children and shows the dietary recommended dietary intakes set by select authorities for minerals, trace elements, water-soluble vitamins, and fat-soluble vitamins for children aged 1–18 years.

References

Die Referenzwerte für die Nährstoffzufuhr, Deutsche Gesellschaft für Ernährung (DGE), Österreichische Gesellschaft für Ernährung (ÖGE), Schweizerische Gesellschaft für Ernährung (SGE), 2021. Bonn, second ed., 7th updated ed. Link (Germany). https://www.dge.de/wissenschaft/referenzwerte/. https://www.sge-ssn.ch/grundlagen/lebensmittel-und-naehrstoffe/naehrstoffempfehlungen/dachreferenzwerte/.

Dietary Reference Values for Nutrients: Summary Report, 2017. European Food Safety Authority (EFSA) Supporting Publication, p. e15121.92. https://doi.org/10.2903/sp.efsa.2017.e15121.

Government Recommendations for Energy and Nutrients for Males and Females Aged 1–18 Years and 19+ Years. Public Health England, Published August 2016 Government Dietary Recommendations Link https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/618167/government_dietary_recommendations.pdf.

National Health and Medical Research Council, 2006. Nutrient Reference Values for Australia and New Zealand. Including Recommended Dietary Intakes. National Health and Medical Research Council, Australian Government Department of Health and Ageing, New Zealand Ministry of Health, Canberra. VERSION 1.2. Updated September 2017.

Nordic Nutrition Recommendations, 2012. Integrating Nutrition and Physical Activity, fifth ed. Nordic Council of Ministers, Copenhagen, Denmark.

Scientific Advisory Committee on Nutrition, 2003. Salt and Health. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/338782/SACN_Salt_and_Health_report.pdf.

Nutrition and Physical Activity. Dietary Guidelines: Dietary Reference Intakes (DRIs). Last updated August 2021. Office of Disease Prevention and Health Promotion, Office of the Assistant Secretary for Health, Office of the Secretary, U.S. Department of Human Health and Human Services.

Oria, M., Harrison, M., Stallings, V.A. (Eds.), 2019. Reference Intakes for Sodium and Potassium. National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Food and Nutrition Board; Committee to Review the Dietary Reference Intakes for Sodium and Potassium. National Academies Press (US), Washington (DC).

Ross, A.C., Taylor, C.L., Yaktine, A.L., et al. (Eds.), 2011. Dietary Reference Intakes for Calcium and Vitamin D. Institute of Medicine (US) Committee to Review Dietary Reference Intakes for Vitamin D and Calcium. National Academies Press (US), Washington (DC).

Vitamin and Mineral Requirements in Human Nutrition, second ed., 2005. World Health Organization https://apps.who.int/iris/handle/10665/42716.

Complementary feeding

Nirupama Shivakumar, Division of Nutrition, St. John's Research Institute, Koramangala, Bengaluru, India

© 2023 Elsevier Ltd. All rights reserved.

Introduction 43
Age of introduction of complementary foods 43
Nutrient needs from complementary foods 45
Meal frequency, energy density, and consistency of complementary foods 47
Meeting nutrient needs during the period of complementary feeding 47
Preparation and feeding of complementary foods 48
Introduction of allergenic food: early as opposed to late as previously suggested 48
What NOT to feed as complementary foods? 48
Baby led weaning (BLW) or "auto-weaning" 48
Conclusion 49
References 49
Further reading 49

Key points

- Appropriate and timely introduction of complementary feeding is crucial for child development and growth
- Developmental readiness includes motor coordination, gastrointestinal and renal maturation
- Taste preferences up to childhood can be modulated by developmental programming through exposure to specific tastes in utero and during infant feeding
- Parents play an important role in establishing good dietary habits with responsive feeding
- Meal frequency, nutrient density and consistency of complementary foods need to be tailored to suit an infant's requirements
- Foods to avoid include artificial, packaged, commercial, ultra-processed, refined products and those with an excess of sugar, salt and trans-fatty acids
- Baby led weaning as an alternative approach is gaining popularity over conventional parent spoon-feeding
- Allergenic foods should be introduced early as opposed to late, as previously suggested

Introduction

Complementary feeding has been defined as "the process starting when breast milk alone is no longer sufficient to meet the nutritional requirements of infants, and therefore other foods and liquids are needed, along with breast milk." The World Health Organization (WHO) recommends exclusive breastfeeding for 6 months and continued breastfeeding thereafter (along with the provision of safe and appropriate complementary foods) until 2 years of age or beyond. Therefore, the period of complementary feeding usually refers to the age range of 6–24 months. This is a crucial time as there is rapid growth and development with a high demand for nutrients, and equally higher vulnerability to growth faltering, micronutrient deficiencies and childhood illnesses. This article will present information mainly from three documents on guidelines related to complementary feeding; first, a joint publication by WHO and UNICEF in 1998, entitled Complementary Feeding of Young Children in Developing Countries: A Review of Current Scientific Knowledge, second, an update to the first document in 2003, and third, a separate document published by Pan American Health Organization and WHO, titled Guiding Principles for Complementary Feeding of the Breastfed Child (summarized in Table 1). Most of the material presented here are for breastfed children, for information on non-breastfed children, refer to the WHO document published in 2005, entitled Guiding Principles for Feeding Non-Breastfed children 6–24 months of age. Along with these, recent updates and additional information are from other resources provided in references and further reading.

Age of introduction of complementary foods

In 2001, the WHO Expert Consultation on the Optimal Duration of Exclusive Breastfeeding reviewed the evidence regarding the age of introduction of complementary foods and concluded that exclusive breast-feeding for 6 months is beneficial for both the infant and the mother. Initiating complementary foods too early or too late have adverse consequences on child growth and development. Risks arising from the early introduction of complementary foods include displacement of breastmilk by foods that are usually less

Table 1 Guiding principles for complementary feeding of the breastfed child.

1. Duration of exclusive breastfeeding and age of introduction of complementary foods

Practice exclusive breastfeeding from birth to 6 months of age, and introduce complementary foods at 6 months of age (180 days) while continuing to breastfeed.

2. Maintenance of breastfeeding

Continue frequent on-demand breastfeeding until 2 years of age or beyond.

3. Responsive feeding

Practice responsive feeding, applying the principles of psychosocial care. Specifically: (a) feed infants directly and assist older children when they feed themselves, being sensitive to their hunger and satiety cues; (b) feed slowly and patiently, and encourage children to eat, but do not force them; (c) if children refuse many foods, experiment with different food combinations, tastes, textures, and methods of encouragement; (d) minimize distractions during meals if the child loses interest easily; (e) remember that feeding times are periods of learning and love—talk to children during feeding, with eye-to-eye contact.

4. Safe preparation and storage of complementary foods

Practice good hygiene and proper food handling by (a) washing caregivers' and children's hands before food preparation and eating, (b) storing foods safely and serving foods immediately after preparation, (c) using clean utensils to prepare and serve food, (d) using clean cups and bowls when feeding children, and (e) avoiding the use of feeding bottles, which are difficult to keep clean.

5. Amount of complementary food needed

Start at 6 months of age with small amounts of food and increase the quantity as the child gets older, while maintaining frequent breastfeeding. The energy needs from complementary foods for infants with "average" breast-milk intake in developing countries are approximately 200 kcal day at 6–8 months of age, 300 kcal day^{-1} at 9–11 months of age, and 550 kcal day^{-1} at 12–23 months of age. In industrialized countries these estimates differ somewhat (130, 310, and 580 kcal day^{-1} at 6–8, 9–11, and 12–23 months, respectively) because of differences in average breast-milk intake.

6. Food consistency

Gradually increase food consistency and variety as the infant gets older, adapting to the infant's requirements and abilities. Infants can eat pureed, mashed, and semi-solid foods from the age of 6 months. By 8 months most infants can also eat "finger foods" (snacks that can be eaten by children alone). By 12 months, most children can eat the types of foods consumed by the rest of the family (keeping in mind the need for nutrient-dense foods, as explained in #8 below). Avoid foods that may cause choking (i.e., items that have a shape and/or consistency that may cause them to become lodged in the trachea, such as nuts, grapes, and raw carrots).

7. Meal frequency and energy density

Increase the number of times that the child is fed complementary foods as he/she gets older. The appropriate number of feedings depends on the energy density of the local foods and the usual amounts consumed at each feeding. For the average healthy breastfed infant, meals of complementary foods should be provided two or three times per day at 6–8 months of age and three or four times per day at 9–11 and 12–24 months of age. Additional nutritious snacks (such as a piece of fruit or bread or chapatti with nut paste) may be offered once or twice per day, as desired. Snacks are defined as foods eaten between meals—usually self-fed, convenient, and easy to prepare. If energy density or amount of food per meal is low, or the child is no longer breastfed, more frequent meals may be required.

8. Nutrient content of complementary foods

Feed a variety of foods to ensure that nutrient needs are met. Meat, poultry, fish, or eggs should be eaten daily, or as often as possible. Vegetarian diets cannot meet nutrient needs at this age unless nutrient supplements or fortified products are used (see #9 below). Fruits and vegetables rich in vitamin A should be eaten daily. Provide diets with adequate fat content. Avoid giving drinks with low nutrient value, such as tea, coffee, and sugary drinks such as soda. Limit the amount of juice offered to avoid displacing more nutrient-rich foods.

9. Use of vitamin–mineral supplements or fortified products for infant and mother

Use fortified complementary foods or vitamin–mineral supplements for the infant, as needed. In some populations, breastfeeding mothers may also need vitamin–mineral supplements or fortified products, both for their own health and to ensure normal concentrations of certain nutrients (particularly vitamins) in their breast milk. (Such products may also be beneficial for pre-pregnant and pregnant women.)

10. Feeding during and after illness

Increase fluid intake during illness, including more frequent breastfeeding, and encourage the child to eat soft, varied, appetizing, favorite foods. After illness, give food more often than usual and encourage the child to eat more.

Reproduced from Pan American Health Organization/World Health Organization, 2003. Guiding Principles for Complementary Feeding of the Breastfed Child. Washington, DC: Pan American Health Organization.

nutrient dense than breastmilk and a higher incidence of gastrointestinal infections. Late introduction of complementary foods results in risk of inadequate nutrient intakes, due to the gap in requirements with reducing nutrient contribution from breastmilk. By 6 months of age, infants are also developmentally ready to transition to semi-solids and progress to solid foods with advancing age. Developmentally motor coordination improves with voluntary control over sucking and swallowing, initiation of biting movements, decrease in tongue thrust tendency, and the eruption of teeth. At around 9 months infants can clear a spoon using lips and maneuver the tongue to move food between teeth. Additionally, the pancreatic amylase and lipase activity, and bile production increase from birth to first year of life (McClean and Weaver, 1993). The renal and gastrointestinal function are also sufficiently mature to metabolize nutrients from CF by the age of 4 months, and the gastrointestinal maturation is largely driven by the foods ingested (Fewtrell et al., 2017). There is also an innate and evolutionary preference for sweet and/or salt and dislike of bitter tastes, which is a potential disadvantage in the current obesogenic environment. However, these predispositions can be modulated by early experience, through fetal (exposure to specific tastes in utero) and infant (via breast milk) programming of taste acceptance. Parents thus play an important role in establishing good dietary habits, where it is necessary to introduce and reinforce, that is to be persistent in introducing a new flavor at least 8–10 times, with effects that last up to 6 years (Fewtrell et al., 2017). Nevertheless,

breastfeeding should be continued till 2 years and beyond, as it provides an additional source of energy, fat, protein, several micronutrients and other bioactive components. For the mother, breastfeeding delays maternal fertility (ensuring adequate spacing between births) and impacts faster post-partum weight loss.

At a population level, introducing complementary foods after 6 months has no adverse effect on growth and the risk of micronutrient deficiencies is low among infants born at term with normal birth weight and well-nourished mothers. In conditions of low nutrient stores in infants, either due to poor maternal stores or low transfer of nutrients from breastmilk, micronutrient deficiencies can manifest before 6 months of life. For example, iron and zinc status may be marginal in low-birth-weight infants or infants born to iron deficient mothers, where the recommendation is to provide iron or zinc supplements at a dose adjusted for body weight from 2 to 3 months of age, rather than early introduction of complementary foods. The concentrations of certain vitamins (e.g., vitamin A and many of the B vitamins) and trace elements (e.g., iodine and selenium) in breast milk may be lower than desirable due to poor diversity in maternal diets. In such situations, dietary counseling or supplements to the mother are the preferred approaches.

Nutrient needs from complementary foods

The amounts of nutrients provided daily to the baby by breast milk can be estimated by multiplying the average breast-milk intake by the concentration of each nutrient in human milk (Dewey, 2001). By subtracting these values from the total recommended nutrient intakes (RNIs) one can derive estimates of the amounts of nutrients needed from complementary foods after 6 months of age. Using this approach, **Table 2** lists these estimates for three age ranges: 6–8, 9–11, and 12–23 months. In this table, the RNIs for energy and protein are taken from the update report on complementary feeding published in 2003, and the RNIs for micronutrients were taken from the FAO/WHO estimates or the US dietary reference intakes. The estimated amount of each nutrient provided by breast milk is based on the average milk intake during each of the age intervals, calculated separately for infants in developing countries and in industrialized countries, using data from the studies compiled in the 1998 WHO/UNICEF document. Because of differences between developing and industrialized countries in average milk intake and in the assumed breast-milk concentration of vitamin A, the estimated amount of each nutrient provided by breast milk may vary. Within each column of **Table 2**, the first value listed refers to developing countries, and the second to industrialized countries.

The first row of **Table 2** shows the total energy requirements and the estimated amounts of energy obtained from breast milk and required from complementary foods at each age. In developing countries, the average expected energy intake from complementary foods is approximately 200 kcal (837 kJ) at 6–8 months, 300 kcal (1256 kJ) at 9–11 months, and 550 kcal (2302 kJ) at 12–23 months. These values represent 33%, 45%, and 61% of total energy needs, respectively. In industrialized countries, the corresponding values are approximately 130 kcal (544 kJ) at 6–8 months, 310 kcal (1298 kJ) at 9–11 months, and 580 kcal (2428 kJ) at 12–23 months (21%, 45%, and 65% of total energy needs, respectively). These values will differ if the child is consuming more or less breast milk than the average.

The second row of **Table 2** shows the same estimates for protein. Assuming average breast-milk intake, the amount of protein needed from complementary foods increases from about 2 g day^{-1} at 6–8 months to 5–6 g day^{-1} at 12–23 months, with the percentage from complementary foods increasing from 21% to about 50%. The remaining rows show the estimates for the key vitamins and minerals. For vitamin B12 and selenium, the amounts needed from complementary foods prior to 12 months are zero because human milk contains generous amounts of these nutrients if the mother is adequately nourished. For the other micronutrients, the percentage of the RNI needed from complementary foods varies widely. At 6–8 months, for example, complementary foods need to provide less than 30% of the RNI for vitamin A, folate, vitamin C, copper, and iodine but more than 70% of the RNI for niacin, vitamin B6, vitamin D, iron, and zinc. The values for niacin needed from complementary foods are high in all age intervals (75–88% of the RNI), but, because niacin needs can also be met by the contribution of tryptophan in the diet, niacin is not likely to be a limiting nutrient among infants who receive adequate protein. Similarly, the percentage of vitamin D needed from other sources is very high (more than 92%) because there is relatively little vitamin D in human milk; however, it should be noted that adequate exposure to sunlight can meet the child's needs for vitamin D even if complementary foods are not rich in this nutrient.

Complementary foods need to provide relatively large amounts (at least 80% of the RNI in all age intervals) of iron, zinc, and vitamin B6. Because the amount of iron in human milk is very low (even though what is present is well absorbed), it is likely to be one of the first limiting nutrients in the diets of infants who rely predominantly on breast milk. Dietary lipids are important not only as a source of essential fatty acids but also because they influence dietary energy density and sensory qualities. Breast milk is generally rich in fat (approximately 30–50% of energy) relative to most complementary foods, so as breast-milk intake declines with age, total fat intake is also likely to decline. If one assumes that the percentage of energy from fat in the total diet should be at least 30% and that the concentration of fat in breast milk averages the amount of fat needed from complementary foods (assuming average breast-milk intake) is zero at 6–8 months, approximately 3 g day^{-1} at 9–11 months, and 9–13 g day^{-1} at 12–23 months, or 0%, 5–8%, and 15–20% of the energy from complementary foods, respectively. As infants decrease their intake of breastmilk, they also need other good sources of essential fatty acids, such as fish, egg, liver, nut pastes, and most vegetable oils.

Table 2 Recommended nutrient intakes, average amount provided by breast milk, and amount needed from complementary foods at 6–8 months, 9–11 months, and 12–23 months. (Dewey and Brown, 2003; Food and Agriculture Organization/World Health Organization, 2002; IOM, 1991; Krebs et al., 1995).

	6–8 months				9–11 months				12–23 months			
	RNI [a]	Amount from breastmilk [b]	Amount needed from CF [c]	% from CF [c]	RNI [a]	Amount from breastmilk [b]	Amount needed from CF [c]	% from CF [c]	RNI [a]	Amount from breastmilk [b]	Amount needed from CF [c]	% from CF [c]
Energy (kcal d^{-1})	615	413; 486	202; 129	33; 21	686	379; 375	307; 311	45	894	346; 313	548; 581	61; 65
Protein (g d^{-1})	9.1	7.2	1.9	21	9.6	6.5; 5.6	3.1; 4.0	32; 42	10.9	5.8; 4.7	5.1; 6.2	47; 57
Vitamin A (μgRE d^{-1})	400	337; 461	63; 0	16; 0	400	308; 354	92; 46	23; 12	400	275; 300	125; 100	31; 25
Folate (μg d^{-1})	80	58	22	28	80	52; 45	28; 35	35; 44	160	47; 38	113; 122	71; 76
Niacin (mg d^{-1})	4	1	3.0	75	4	0.9; 0.8	3.1; 3.2	78; 80	6	0.8; 0.7	5.2; 5.3	87; 88
Riboflavin (mg d^{-1})	0.40	0.24	0.16	40	0.40	0.22; 0.19	0.18; 0.21	45; 53	0.50	0.19; 0.16	0.31; 0.34	62; 68
Thiamine (mg $d^{-)1}$)	0.30	0.14	0.16	53	0.30	0.12	0.18	60	0.50	0.11	0.39	78
Vitamin B_6 (mg d^{-1})	0.30	0.06	0.24	80	0.30	0.06	0.24	80	0.50	0.05	0.45	90
Vitamin B_{12} (μg d^{-1})	0.50	0.66	0	0	0.50	0.60; 0.51	0	0	0.90	0.53; 0.47	0.37; 0.43	41; 48
Vitamin C (mg d^{-1})	30	28	2	7	30	25; 21	5; 9	17; 30	30	22; 18	8; 12	27; 40
Vitamin D (μg d^{-1})	5	0.4	4.6	92	5	0.3	4.7	94	5	0.3; 0.2	4.7; 4.8	94; 96
Calcium (mg d^{-1})	270	191	79	29	270	172; 148	172; 148	36; 45	500	154; 125	346; 375	69; 75
Copper (mg d^{-1})	0.20	0.17	0.03	15	0.20	0.14	0.06	30	0.30	0.14; 0.11	0.16; 0.19	53; 63
Iodine (μg d^{-1})	90	75	15	17	90	68; 58	22; 32	24; 36	90	60; 49	30; 41	33; 46
Iron [d] (mg d^{-1})	9.3	0.2	9.1	98	9.3	0.2	9.1	98	5.8	0.2; 0.1	5.6; 5.7	97; 98
Magnesium (mg d^{-1})	54	24	30	56	54	22; 19	32	59	60	19; 16	41; 44	68; 73
Phosphorus (mg d^{-1})	275	95	180	65	275	86; 74	189; 201	69; 73	460	77; 63	383; 397	83; 86
Selenium (μg d^{-1})	10	14	0	0	10	12	0	0	17	11	6	35
Zinc (mg d^{-1})	3	0.6	2.4	80	3	0.5; 0.4	2.5; 2.6	83; 87	3	0.4; 0.3	2.6; 2.7	87; 90

[a] Recommended nutrient intakes, from FAO/WHO (2002) except for energy and protein (Dewey, K.G. and Brown, K.H., 2003) and calcium, copper, phosphorus, and zinc (from the US-Canada Dietary Reference Intakes).

[b] Based on average milk volumes of 674, 616, and 549 ml d^{-1} in developing countries and 688, 529, and 448 ml d^{-1} in industrialized countries for 6–8, 9–11, and 12–23 months, respectively (WHO 1998) and milk nutrient concentrations from the Institute of Medicine (IOM 1991), except for vitamin A in milk of women from developing countries (WHO 1998) and zinc (Krebs, N.F., Reidinger, C.J., Hartley, S., et al, 1995). For each nutrient, the first value refers to developing countries and the second value (after the semicolon) refers to industrialized countries, whenever there is a difference between the two.

[c] CF, complementary foods. For each nutrient, the first value refers to developing countries and the second value (after the semi-colon) refers to industrialized countries, whenever there is a difference between the two.

[d] Assuming medium bioavailability of iron.

Meal frequency, energy density, and consistency of complementary foods

The frequency (i.e., the number of meals per day) with which complementary foods need to be fed depends on the total quantity of food required and the amount of food that a child can consume at a single meal. The total quantity of food required is a function of the energy requirement, which varies with age and breast-milk intake, and the energy density of foods (i.e., kcal g^{-1}). The amount of food per meal that the child can consume depends on the functional gastric capacity of infants and young children, which is assumed to be 30 g kg^{-1} reference body weight. To cover the needs of nearly all children, these calculations use as a starting point the average total energy requirement plus two standard deviations (Pan American Health Organization/World Health Organization, 2003). For children with average breast-milk intake consuming foods with an energy density of at least 0.8 kcal g^{-1}, the number of meals required is two at 6–8 months and three thereafter. Meal frequency increases with low or no breastmilk intake complemented with meals with a lower energy density. These calculations assume that children are fed to their gastric capacity at each meal, which may not be the case. For this reason, the guidelines recommend that additional nutritious snacks be offered once or twice per day, as desired (see guiding principle 7, **Table 1**). It is useful to adhere to the meal-frequency guidelines as a greater than necessary frequency may lead to excessive displacement of breast milk and may also require more time and effort by caregivers (Pan American Health Organization/World Health Organization, 2003).

The consistency of complementary foods is important for two reasons, first it should be appropriate for the child's stage of neuromuscular development, second it determines the nutrient density, and in turn the quantity of feed per meal that can be consumed by the child. To start with semi-solid or puréed foods are provided, as young infants do not have the ability to chew and swallow thick or solid foods. By the age of 8 months most infants can eat "finger foods," and by 12 months they can generally consume "family foods" of a solid consistency. Thus, a gradual advancement is necessary in the consistency of foods offered between 6 and 12 months (see guiding principle 6, **Table 1**). There is some evidence to suggest that there may be a critical window for introducing lumpy solid foods, and that failure to introduce such foods by approximately 9–10 months of age is associated with an increased risk of feeding difficulties and reduced consumption of important food groups such as fruits and vegetables later. A new feeding approach is being adopted in few countries now, which challenge the traditional approach, see the section on "baby led weaning" for more details.

Meeting nutrient needs during the period of complementary feeding

As described above and shown in **Table 2**, breastfed infants need considerable amounts of certain nutrients from complementary foods after 6 months of age. It is a challenge to meet nutrient needs at this age because the amount of food consumed is relatively small. Thus, it is important to provide nutrient-dense complementary foods. The desired nutrient density (the amount of nutrient per kcal of food) can be calculated by dividing the quantity of each nutrient by the total energy from complementary foods (as shown in **Table 2**). When the desired nutrient densities are compared with the actual nutrient densities of the typical complementary foods consumed in various populations, is generally seen to be adequate but several micronutrients are limiting in diets and are considered problem nutrients. In addition to low nutrient density of certain nutrients, nutrient bioavailability, which includes digestion, absorption and utilization for body functions, also needs to be considered. Particularly, the quality of protein is important to determine the quantity of protein required to meet requirements (Shivakumar et al., 2019). In developing countries, poor sanitary environments adversely impact gastrointestinal functions, further compromising nutrient availability (Kelly, 2021). Selection of foods for complementary feeding should focus on these factors, where animal source foods with relatively higher nutrient density than plant-based foods are recommended.

In most developing and some industrialized countries, fats, iron, zinc, and vitamin B6 are problem nutrients. The fat composition or addition of extra fat (oil), especially when feeding a vegetarian or vegan diet, may be minimal, resulting in a risk for inadequate fat intake. This can be more prevalent in poorer and marginalized communities. Equally, with plant foods and when there is poor consumption of animal source foods, the inclusion of fat as an energy source will displace grains in the mix, and therefore reduce protein intake. Therefore, attention must be given to the use of legumes to improve protein quality in these circumstances. The consumption of iron is likely to be marginal in all populations unless iron-rich foods (e.g., organ meat, fish, eggs, green leafy vegetables), or fortified products are consumed. Riboflavin, niacin, thiamine, folate, calcium, vitamin A, and vitamin C may also be problem nutrients, depending on the local mix of complementary foods. There is insufficient information to determine the extent to which some of the other micronutrients, such as vitamin E, iodine, and selenium, may be problem nutrients. Guiding principles 8 and 9 (**Table 1**) provide general recommendations to help ensure that the nutrient density of complementary foods will be adequate. It is difficult to develop more specific dietary "prescriptions" to be used globally because of the great variability in foods consumed across populations. Purely plant-based diets are often high in phytate, which reduces the bioavailability of iron and zinc. Therefore, it is recommended that meat, fish, poultry, or egg be offered daily, if possible.

When the amount of animal-source food available locally is limited, the amounts of iron and zinc absorbed from the diet can be enhanced by, first, reducing the phytate concentrations of the staple complementary food through germination, fermentation, and/or soaking, second, reducing the intake of polyphenols (e.g., from coffee and tea), which are known to inhibit iron absorption, and, third, increasing the intake of enhancers of iron and zinc absorption, such as vitamin C (for iron) and other organic acids (for iron and zinc). Protein digestibility is found to improve through de-hulling and extrusion of the source food grain (Kashyap et al., 2019; Devi et al., 2020). Adequate calcium can be obtained from cheese, yoghurt, and other dairy products, however, feeding fresh unheated cow's milk is not recommended before 12 months because it is associated with cow's milk protein allergy, which could

lead to fecal blood loss and lower iron status. Some vegetables can also provide modest amounts of calcium, but the bioavailability of calcium from foods with high amounts of oxalate (such as spinach) is very low. Fruits and vegetables rich in vitamin A are recommended daily because of the importance of preventing vitamin A deficiency, which has been linked with excess child mortality and other adverse outcomes. The bioavailability of pro-vitamin A carotenoids can be enhanced by finely chopping or puréeing the food and serving it with a source of fat to facilitate absorption.

Preparation and feeding of complementary foods

Complementary feeding guidelines also involve directions on how to feed infants and young children. Firstly, it deals with safe preparation and storage of complementary foods (see guiding principle 4, **Table 1**) with attention to food hygiene during preparation (clean utensils) and feeding (clean hands), which is essential for preventing gastrointestinal illness. Feeding bottles are a major source of contamination if kept unclean/unsterile, and thus should be avoided in resource poor settings. Additionally, food should be served warm and immediately after cooking, and leftovers stored safely.

Secondly, there is emphasis on caregiver and infant interaction during feeds, termed responsive feeding (see guiding principle 3, **Table 1**). This appropriate feeding behavior is being sensitive to the child's hunger and satiety cues than either a laissez-faire style of feeding (the caregiver rarely encourages the child to eat) or, at the opposite extreme, a controlling style of feeding (the caregiver determines when and how much the child will eat, even to the point of force feeding). Responsive feeding also involves feeding slowly and patiently, minimizing distractions, and actively engaging the child in dialog. There is some evidence that indicates promotion of responsive feeding will enhance dietary intake, reduce undernutrition and obesity, and pave way for positive behavioral development.

Thirdly, it is important to know what to feed during and after illness (see guiding principle 10, **Table 1**). The need for fluids is often greater during illness, and for this reason in addition to preference for breastfeeding by children, it is critical to increase breastfeeding frequency and offer other fluids on demand. Although poor appetite is often a concern, continued consumption of complementary foods is recommended to enhance recovery. After illness, with improvement in appetite, the child needs more food than usual to make up for nutrient losses during illness and to allow for catch-up growth.

Introduction of allergenic food: early as opposed to late as previously suggested

An increase in prevalence of food allergy has been observed in higher-income countries despite restriction and/or delay in introduction of potentially allergenic foods, including cows' milk, egg, fish, gluten, peanut, and seeds. On the contrary countries in which peanuts are commonly used in complementary feeding have a low incidence of peanut allergy. Animal models suggest that immune tolerance to foods might be regulated by an early and regular exposure to food proteins during a critical window. In support, studies in human infants, show that the introduction of cereal <5 and a half months, fish <9 months and egg <11 months, peanuts between 4 and 6 months, as compared to later introductions, reduces the risk of asthma, allergic rhinitis and atopic sensitization at 5 years of age. Gluten introduction anytime between 4 and 12 months does not seem to impact celiac disease development, albeit consumption in large quantities is better avoided in infants, particularly those with a known risk for the disease. Overall, the evidence is not in favor of delayed introduction of allergenic foods, rather, early introduction with repeated exposure may prove beneficial. Furthermore, continued breastfeeding while introduction of allergenic complementary foods seems to modulate the development of tolerance (Fewtrell et al., 2017).

What NOT to feed as complementary foods?

Foods to avoid include, artificial, packaged, commercial, ultra-processed, refined, those with excess of sugar, salt, and trans-fatty acids. Avoid adding sugar or salt (in excess) to home prepared foods and beverages, due to the possibility of an increased risk of obesity or hypertension in later life. These practices prevent obesity or non-communicable diseases and promote a diverse and nutrient-rich diet. Constant vigilance on inappropriate marketing and promotion of commercial foods is required as a mandate at all levels of childcare. Use of commercial foods may undermine exclusive and continued breastfeeding, create dependency on commercial products and can undermine the value of home-prepared foods. Excessive animal protein intake $>15\%$ of total calories has also correlates with risk of overweight or obesity in later life (Fewtrell et al., 2017).

Baby led weaning (BLW) or "auto-weaning"

The baby led approach that is popular in some countries, promotes infant self-feeding from 6 months of age (deciding on what, when and how long to feed), instead of the conventional parent spoon-feeding, along with sharing of family foods and mealtimes. The BLW as a complementary feeding practice has gained popularity and grown substantially over the past 10 years. It aims to divide the responsibility of feeding between the infant and caregivers. With BLW, a complex engagement of oral and hand-coordinated

motor skills is required, where the infant is required to manipulate movement of food in the mouth, and consequently exposed to the textures and flavors of a variety of foods. Although, BLW has potential to positively shape interpersonal food environments, it may also carry inherent risks. Based on the current evidence BLW may reduce infant food fussiness and increase satiety responsiveness. The literature also indicates that there is no difference in energy, iron or zinc intake between BLW and traditional feeding practices, but concerns do exist. Recently, a modified version called Baby Led Introduction to SolidS (BLISS) has been developed that highlight introducing iron and energy-rich foods and avoiding foods that are a choking hazard. Further research on the effect of this method on nutrient intake, appetite regulation and health outcome such as growth and obesity are proposed (Fewtrell et al., 2017).

Conclusion

Appropriate and timely introduction of complementary feeding is crucial for child development and growth. The consensus is to introduce complementary feeds not early or later than 6 months (180 days) of age, with continued breast feeding till 2 year and beyond. Infants are developmentally ready by ~4 months of age with improved motor coordination, gastrointestinal and renal maturation. Infants from developing and industrialized countries differ in their requirements based on variation in breastmilk volume and a relatively higher pathogenic burden through unsanitary environment. Depending on the status of nutritional stores of mother and infant, there is potential for problem nutrients, which could be overcome by consuming diverse foods, particularly of animal source, fruits and vegetables. Thus, attention to nutrient density, meal frequency along with responsive feeding and food hygiene is key to successful complementary feeding. It is prudent to avoid foods that are artificial, packaged, commercial, ultra-processed, refined, and those with excess of sugar, salt, or trans-fatty acids. In recent times baby led weaning has gained popularity instead of the conventional parent spoon-feeding, which has potential to positively shape interpersonal food environments. There is supportive evidence for early introduction of allergenic foods to shape immune tolerance as opposed to late introduction as earlier suggested. Overall, the experience of complementary feeding should forge healthy parent-child interactions along with supporting age- and sex-appropriate growth and development.

References

Devi, S., Varkey, A., Dharmar, M., Holt, R.R., Allen, L.H., Sheshshayee, M.S., Preston, T., Keen, C.L., Kurpad, A.V., 2020. Amino acid digestibility of extruded chickpea and yellow pea protein is high and comparable in moderately stunted South Indian children with use of a dual stable isotope tracer method. J. Nutr. 150 (5), 1178–1185.

Dewey, K.G., Brown, K.H., 2003. Update on technical issues concerning complementary feeding of young children in developing countries and implications for intervention programs. Food Nutr. Bull. 24 (1), 5–28.

Dewey, K.G., 2001. Nutrition, growth, and complementary feeding of the breastfed infant. Pediatr. Clin. 48 (1), 87–104.

Fewtrell, M., Bronsky, J., Campoy, C., Domellöf, M., Embleton, N., Mis, N.F., Hojsak, I., Hulst, J.M., Indrio, F., Lapillonne, A., Molgaard, C., 2017. Complementary feeding: a position paper by the European Society for Paediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN) Committee on Nutrition. J. Pediatr. Gastroenterol. Nutr. 64 (1), 119–132.

Food and Agriculture Organization/World Health Organization, 2002. Vitamin and Mineral Requirements in Human Nutrition. World Health Organization, Geneva.

IOM, 1991. Nutrition During Lactation. National Academy Press, Washington, DC.

Kashyap, S., Varkey, A., Shivakumar, N., Devi, S., Reddy, B.H.,R., Thomas, T., Preston, T., Sreeman, S., Kurpad, A.V., 2019. True ileal digestibility of legumes determined by dual-isotope tracer method in Indian adults. Am. J. Clin. Nutr. 110 (4), 873–882.

Kelly, P., 2021. The contribution of environmental enteropathy to the global problem of micronutrient deficiency. Proc. Nutr. Soc. 1–8.

Krebs, N.F., Reidinger, C.J., Hartley, S., Robertson, A.D., Hambidge, K.M., 1995. Zinc supplementation during lactation: effects on maternal status and milk zinc concentrations. Am. J. Clin. Nutr. 61 (5), 1030–1036.

McClean, P., Weaver, L.T., 1993. Ontogeny of human pancreatic exocrine function. Arch. Dis. Child. 68 (1 Spec No), 62.

Pan American Health Organization/World Health Organization, 2003. Guiding Principles for Complementary Feeding of the Breastfed Child. Pan American Health Organization, Washington, DC.

Shivakumar, N., Kashyap, S., Kishore, S., Thomas, T., Varkey, A., Devi, S., Preston, T., Jahoor, F., Sheshshayee, M.S., Kurpad, A.V., 2019. Protein-quality evaluation of complementary foods in Indian children. Am. J. Clin. Nutr. 109 (5), 1319–1327.

World Health Organization/UNICEF, 1998. Complementary Feeding of Young Children in Developing Countries: A Review of Current Scientific Knowledge. WHO/NUT/98.1. World Health Organization, Geneva.

Further reading

Caulfield, L.E., Huffman, S.L., Piwoz, E.G., 1999. Interventions to improve intake of complementary foods by infants 6 to 12 months of age in developing countries: impact on growth and on the prevalence of malnutrition and potential contribution to child survival. Food Nutr. Bull. 20 (2), 183–200.

Dewey, K.G., Black, R., Michaelsen, K., 2002. Success of intervention programs to promote complementary feeding. In: Nestle Nutrition Workshop Series, vol. 48. Lippincott-Raven, Philadelphia, pp. 199–216, 1999.

Gibson, R.S., Ferguson, E.L., Lehrfeld, J., 1998. Complementary foods for infant feeding in developing countries: their nutrient adequacy and improvement. Eur. J. Clin. Nutr. 52 (10), 764–770.

Lutter, C.K., Dewey, K.G., 2003. Proposed nutrient composition for fortified complementary foods. J. Nutr. 133 (9), 3011S–3020S.

UNICEF, 2020. Improving Young Children's Diets During the Complementary Feeding Period. UNICEF Programming Guidance. UNICEF, New York, p. 76.

World Health Organization, 2016. Ending the Inappropriate Promotion of Foods for Infants and Young Children: A Primer on WHO Guidance.

World Health Organization/London School of Hygiene and Tropical Medicine, 2000. Complementary Feeding: Family Foods for Breastfed Children. WHO/NHD/00.1; WHO/FCH/CAH/00.6. World Health Organization, Geneva.

Early origins of disease

D.B. Hardy[a] **and D.J. Hoffman**[b], [a] The Children's Health Research Institute and the Lawson Health Research Institute, The University of Western Ontario, London, ON, Canada; and [b] Department of Nutritional Sciences, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of M.S. Martin-Gronert, S.E. Ozanne, Early origins of disease: Fetal, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 99–105, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00081-7.

Introduction 51
Epidemiological data 51
Developmental origins of health and disease hypothesis 52
Evidence from human studies 52
Animal models 53
Maternal undernutrition models 53
Maternal calorie restriction 53
Maternal protein restriction 54
Maternal overnutrition models 54
Maternal obesity 54
Maternal high-fat diet 54
Surgical model of—uteroplacental insufficiency 54
Pharmacological models 55
Maternal glucocorticoid exposure 55
Gestational diabetes 55
Cellular and molecular mechanisms 55
Epigenetic mechanisms 55
Mitochondrial dysfunction and oxidative stress 56
Excessive intrauterine exposure to lipids 56
Conclusions 57
Further reading 57

Key points

- To understand the history underlying the Developmental Origins of Health and Disease (DOHaD) hypothesis.
- To identify the clinical evidence supporting DOHaD
- To appreciate the role of preclinical animal models in mimicking the adverse *in utero* events leading to fetal growth restriction and long-term dysmetabolism.
- To understand the underlying molecular mechanisms linking a poor *in utero* environment to metabolic disease in postnatal life.

Glossary

DNA methylation Chemical modification of DNA in which methyl group is added usually to cytosine

Epigenetics The study of inherited changes in phenotype (appearance) or gene expression that occur without any changes to the underlying DNA sequence

Histones Proteins that package and order DNA

Impaired glucose tolerance Prediabetic state of increased levels of glucose in the blood

Insulin resistance Condition in which insulin is less effective in lowering blood sugar levels

Metabolic syndrome A combination of medical disorders including high blood pressure, high triglyceride levels, high cholesterol, obesity, and insulin resistance, that increase the risk of developing heart disease, type 2 diabetes, and stroke

Mitochondrial dysfunction Inability of the mitochondria to convert the energy derived from food into the adenosine triphosphate (ATP), which powers most cellular functions

Oxidative stress Imbalance between the production of reactive oxygen species within cells and the ability of the antioxidant machinery to detoxify the reactive intermediates, or to repair the cellular damage

Introduction

Chronic diseases such as cardiovascular disease, type 2 diabetes, liver fibrosis, and cancer account for 60% of all deaths worldwide. In recent years it emerged that these diseases, thought of predominantly as diseases of adulthood, are increasingly prevalent among children. Although a nutritionally unbalanced diet and sedentary lifestyle play a contributing role, a growing body of evidence suggests that the ability of an individual to respond to metabolic challenges encountered throughout its lifetime may be determined during its fetal life. Although fetal growth and development follows the route encoded within an individual's genome, the *in-utero* environment in which a fetus grows can influence this process. The intrauterine environment provides a forecast of conditions for the fetus after birth. The fetus can respond and adapt to a variety of stimuli or insults short-term. However, the adaptations include irreversible changes in the structure and function of the body, that are detrimental to the long-term health of an individual, especially if postnatal conditions differ to the environment experienced *in utero*.

Epidemiological data

It has been over 75 years since it was suggested, that the de-cline in overall death rates in Sweden and the UK between 1751 and 1930 was attributed to improvements in childhood nutrition and living conditions. Focus on the very early environment was prompted when David Barker and colleagues reported strong correlations between the prevalence of is-chemic heart disease and rates of mortality among newborns suggesting that increased risk of disease was linked to environmental factors affecting fetal development. The association between birth weight (a proxy for a compromised fetal growth) and development of type 2 diabetes was shown for the first time in a study of men born in Hertfordshire, UK, who were 64 years old at the time of study. The prevalence of impaired glucose tolerance among these men steadily in-creased as the birth weight decreased and men with lower birth weights were six times more likely to have type 2 diabetes than those born heavier. Subsequently, it was shown that for every 1 kg increase in birth weight there is 25% decrease in type 2 diabetes risk. In the original cohorts this was linear across the entire birth weight spectrum. However, in contemporary populations as rates of maternal obesity increase and the prevalence of women with gestational diabetes rises, more infants are born large for gestational age (44,000 g). These newborns have increased adiposity and are at a higher risk of developing obesity and features of the metabolic syndrome including impaired glucose tolerance. Therefore, infants with birth weight at both ends of the body weight spectra are more susceptible to the development of chronic diseases in later life, giving rise to U-shaped relationships. Suboptimal *in utero* conditions have now been linked to a broad spectrum of diseases (Table 1).

Table 1 Diseases associated with suboptimal intrauterine environment in humans.

Metabolic disorders
Impaired glucose tolerance insulin resistance
Obesity dyslipidemia type 2 diabetes
Nonalcoholic fatty liver disease
Cardiovascular disorders
Hypertension
Coronary heart disease stroke
Atherosclerosis coagulation disorders pre-eclampsia
Reproductive system disorders
Polycystic ovary syndrome early adrenarche/menarche early menopause
Endocrine disorders
Hypercortisolism hypothyroidism
Respiratory disorders
Chronic obstructive lung disease asthma
Nervous system disorders
Neurological disorders schizophrenia dementia
Skeletal system disorder
Osteoporosis
Renal disorders
Chronic renal failure
Cancer
Breast cancer

Developmental origins of health and disease hypothesis

On the basis of the epidemiological data, in 1992 Nick Hales and David Barker proposed the thrifty phenotype hypothesis to explain the relationship between fetal growth and development of diseases in later life. They postulated that in response to undernutrition, a growing fetus adopts strategies to ensure its immediate postnatal survival in conditions of continued poor nutrition. These strategies include redistribution of blood flow to preserve brain growth at the expense of other tissues such as the liver, kidney, endocrine pancreas, and skeletal muscle. In addition, alterations in hormone production (e.g., decrease in fetal insulin and insulin-like growth factor 1 (IGF-1) concentrations) and tissue sensitivity to hormones, as well as programming of whole body metabolism to promote storage of nutrients when available also occur. If the individual is born into conditions of poor nutrition these adaptations are beneficial for his survival. However, if the individual is born into an environment of plenty or adequate nutrition, the adaptations become detrimental to long-term health. Since the proposal of the thrifty phenotype hypothesis, it has become apparent that critical windows of development extend into postnatal life and that the nutrient overload during fetal life is also detrimental. Therefore, the concept that experiences of early life influence an individual's long-term health is now referred to as the "developmental origins of health and disease hypothesis (DOHaD)" (Fig. 1).

Evidence from human studies

Studies of twins have provided a very strong evidence in support of the link between the intrauterine environment and the risk of developing chronic disease. A study of Danish twin men discordant for type 2 diabetes revealed that in both monozygotic (identical) and dizygotic (nonidentical) twin pairs, the twin born with lower birth weight developed diabetes. As monozygotic twins are genetically identical, the differences in birth weight must reflect differences in the fetal environment. Similar observations were also made in younger Italian twin men (mean age 32 years). The link between nongenetic intrauterine factors and blood pressure has been proven to exist in a huge study involving over 20,000 Swedish twin pairs. Decreased birth weight was associated in this study with increased risk of hypertension independent of risk factors for hypertension in adulthood, including body mass index (BMI).

Aside from cardiovascular defects, several human studies have also reaffirmed the links between a poor *in utero* environment and dysmetabolism. The 2014 EPOCH study demonstrated that IUGR children with catch-up growth in postnatal life exhibit greater

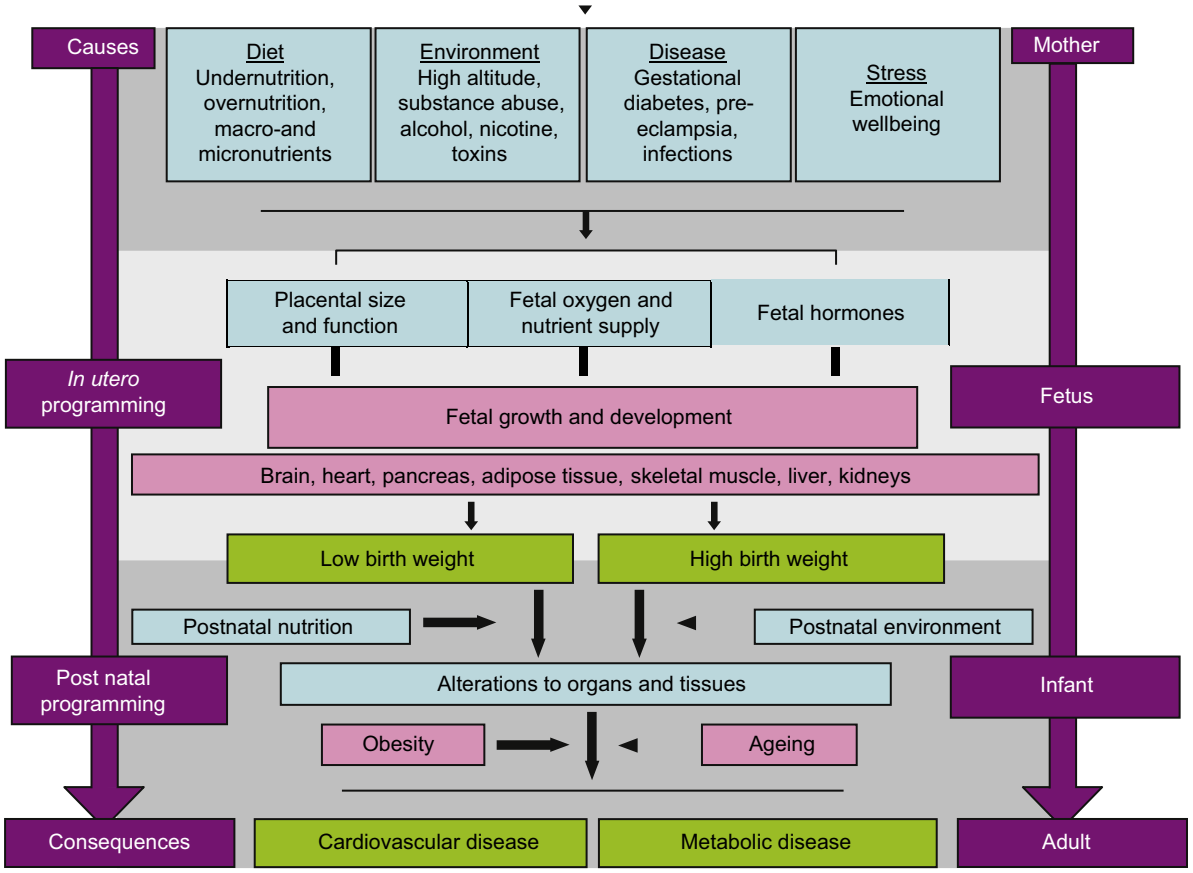


Fig. 1 Intrauterine programming of adult disease.

homeostatic model assessment of insulin resistance (HOMA-IR) and higher circulating insulin. Furthermore, neonatal catch-up growth in IUGR infants, likely attributed to growth-promoting formula diets, is also associated with higher LDL:HDL ratio, and an earlier incidence of dyslipidemia. With respect to triglycerides, fetal growth *in utero* is inversely associated with circulating triglycerides in postnatal life. Given dyslipidemia in the liver can develop into nonalcoholic fatty liver (NAFL)/nonalcoholic fatty liver disease (NAFLD), it is noteworthy that either low- or high-birth weight children exhibit increased risk of NAFLD later in life. This is of great concern considering liver fibrosis is thought to be responsible for up to 45% of deaths in the developed world.

In humans it is difficult to access the direct impact of maternal nutrition on the health of the offspring. Studies of individuals who underwent famine while *in utero* are there-fore invaluable in accessing the effect of undernutrition on long-term health. The Dutch famine occurred for 5 months at the end of World War II (from late November 1944 to early May 1945) and before the famine the population of the affected area was reasonably well-nourished. When comparison was carried out between the individuals born during the year before the famine and those *in utero* during the famine, the latter had impaired glucose tolerance as shown by raised plasma glucose concentrations following an oral glucose tolerance test. The timing of the exposure to famine has also proven to be very important. Obesity rates at 19 years of age were increased among individuals exposed to famine in the early gestations, whereas they were decreased among those exposed in late gestation. Exposure to famine in early gestation has also been associated with greater risk of developing coronary heart disease at the age of 50 years.

In addition to fetal undernutrition, fetal overnutrition can also influence an individual's long-term health. Some of the strongest evidence in support of the role of nutritional excess during pregnancy and subsequent development of diseases such as obesity and type 2 diabetes in adulthood have come from the study of the Pima Indians of Arizona, a population with very high incidence of type 2 diabetes, believed to be due to genetic inheritance. A six-fold increase in the prevalence of type 2 diabetes has been shown in this population among children born to diabetic mothers in comparison to children of nondiabetic women. This increase was still present after adjustments for paternal diabetes, age of onset of diabetes in parents and obesity in the offspring. Moreover, compared to siblings born before the mother developed diabetes, children born after her diagnosis were at higher risk of developing diabetes and obesity by early adulthood. The effect was observed in addition to genetic transmission and hence must be related to the diabetic milieu experienced *in utero*. Similar findings have also been obtained from studies in lower-risk populations and in the developing world. In addition to gestational diabetes, paternal and maternal body composition and diet can not only influence the first generation of offspring, but the second generation as well. US studies have demonstrated that a poor maternal diet, as assessed by a low score of the Healthy Eating Index (HEI) and/or a high fat diet led to increased neonatal adiposity. This is of great concern considering that large for gestational age newborns, as early as 2 days old, exhibit decreased insulin sensitivity. Moreover, higher pre-pregnancy body mass index (BMI) increases the risk of being overweight in postnatal life. Specifically, children born to mothers with pre-pregnancy BMI > 25 are 1.7 times more likely to be overweight in early life and twice as likely to be overweight in late childhood. A study of $\sim$250,000 offspring from the Nurses Health Study II have also found children (aged 9–14 years) born to obese mothers exhibit higher percentage body fat, systolic blood pressure, and fasting glucose compared to normal-weight women.

Animal models

To gain insight into the mechanisms underlying the association between the early environment and adult health, a number of animal models have been established. These include studies in large animals such as nonhuman primates, sheep, horses, or pigs. However, the majority of research has been conducted in smaller animals such as mice and rats to take advantage of their shorter gestation and lifespan.

Maternal undernutrition models

Maternal calorie restriction

In this model of prenatal undernutrition caloric intake of mothers is restricted and the outcome is dependent on the timing, length, and magnitude of the caloric restriction. Reduction in maternal nutritional intake to 50% of *ad libitum* during both pregnancy and lactation leads to the birth of low birth weight offspring, who remain smaller throughout adulthood. If such reduction is limited to gestation the off-spring are smaller than controls at birth but become heavier than control offspring by weaning. This is also known as postnatal catch-up growth. By the age of 9 months they have a greater percentage of body fat, increased plasma leptin and triglyceride levels, hyperglycemia, hyper-insulinaemia, and impaired glucose tolerance. This is due, in part, to increased circulating cortisol and higher expression of PEPCK, involved in gluconeogenesis, in the liver. A 50% decrease in maternal food intake during the first two weeks of pregnancy does not have adverse effect on insulin secretion and action in adult male offspring. However, if 50% maternal nutrition is given during the last week of pregnancy the off-spring are born with low birth weights and have decreased pancreatic b-cell mass and insulin content. Even if these animals are fed a standard laboratory diet in postnatal life, pancreatic deficiencies persist into adulthood. Much more severe maternal caloric restriction of 30% of *ad libitum* food intake results in the birth of growth-restricted offspring, who develop hyperphagia and adult-onset obesity with hyperinsulinaemia, hyperleptinaemia, and hypertension. In this model, the impact of decreased maternal and placental weight during pregnancy must also be taken in consideration.

Maternal protein restriction

This is one of the most extensively studied rodent animal models of developmental programming. The regime involves feeding pregnant dams a low (5–8%) protein (LP) diet during pregnancy and lactation in comparison to a 20% protein diet given to the control group. This maternal diet leads to a deficiency in fetal amino acids which are critical for growth. Moreover, this dietary manipulation does not lead to changes in maternal physiology (i.e., weight gain or food intake), conception rates or litter size. Offspring of LP rat dams have a 15% reduction in birth weight and if offspring are suckled by their mothers, permanent growth restriction occurs despite weaning the animals onto a control diet fed *ad libitum*. After the initial increase in insulin sensitivity in LP offspring in young adult life (6 weeks–3 months of age), these animals undergo an age-dependent loss of glucose tolerance in a sex-dependent manner. Male offspring of LP dams have impaired glucose tolerance by 15 months of age and frank diabetes by 17 months. Female LP offspring develop hyperglycemia and have impaired glucose tolerance at the older age of 21 months. The insulin resistance is associated with decreased protein expression of key insulin signaling proteins including insulin receptor substrate1 (IRS1), protein kinase zeta (PKC z), and glucose transporter GLUT4 in skeletal muscle; and the phosphatidylinositol 3-kinase (PI3K) catalytic subunit p110b in adipose tissue. The profile of insulin signaling molecule deficiencies identified in the LP model is strikingly similar to the pattern observed in tissues from low birth weight young men. As the changes observed in humans occurred prior to the development of insulin resistance or type 2 diabetes, they are unlikely to be a secondary consequence of hyperglycemia or hyperinsulinaemia and hence may help predict diabetes risk. Maternal protein restriction has also been linked to development of hypertension in the offspring due to the alterations in the activity of the renin–angiotensin system and to increased early mammary tumor risk among female LP offspring. After weaning, if maternal protein restricted offspring are placed on a normal 20% protein diet, they also exhibit visceral obesity, hypercholesterolemia (e.g., lower hepatic Cyp7a1 expression), alterations in hepatic drug metabolism (i.e., Cyp3a and Cyp2c11), and lower insulin growth factor 1 (Igf-1) despite no changes in postnatal food intake. These LP male offspring also exhibit decreases in circulating testosterone which might account for the sexual dimorphism often observed in this model.

Maternal overnutrition models

Maternal obesity

Maternal obesity has been associated with the development of insulin resistance and type 2 diabetes. Adult offspring of obese mouse dams are hyperphagic, have increased body weight, and raised fat-to-lean mass ratio. With age they develop beta cell failure. As with the maternal low protein model, they have alterations in key hepatic insulin signaling molecules. Diminished protein expression of IRS1 in the liver coupled with increased phosphorylation of IRS serine residues may there-fore contribute to the development of type 2 diabetes. Not surprisingly, these offspring develop insulin resistance by 3 months of age and impaired glucose tolerance was observed in males at 6 months of age. Insulin resistance also contributes to the pathogenesis of nonalcoholic fatty liver disease, signs of which have been observed in these animals. Additionally, hypertension and impaired endothelial cell function were reported in the offspring of obese dams implying that maternal obesity can predispose offspring to the development of cardiovascular disease.

Maternal high-fat diet

Offspring exposed to maternal high-fat diet while *in utero* became obese in adulthood and demonstrated abnormal cholesterol and lipid metabolism, hyperleptinemia, hyper-insulinemia, and insulin resistance. They have reduced muscle mass and persistent accumulation of lipid in the liver, which predisposes them to the development of nonalcoholic fatty liver disease (NAFLD). In addition, these animals have increased risk of developing hypertension and cardiovascular disease. The phenotypic characteristics of metabolic syndrome are present in most models of maternal overnutrition regardless of type of diet the offspring is exposed to in postnatal life.

Surgical model of—uteroplacental insufficiency

Uteroplacental insufficiency (UPI) is one of the most common causes (~8% of all pregnancies) of idiopathic intrauterine growth restriction (IUGR) in the western world. It has been attributed to maternal smoking, preeclampsia or abnormalities in the development of placenta. Unilateral and bilateral uterine artery ligation/ablation has been performed to model UPI by reducing oxygen and substrate availability to the fetus and to induce asymmetric IUGR. Following this intervention, fetuses shown to be were hypoxic, hypoglycaemic, and had reduced insulin and IGF1 levels, a profile also found in human growth restricted fetuses. Uterine artery ligation leads to the development of mild insulin resistance, defects in insulin synthesis, and secretion in early life. With time, as beta cells fail to compensate, diabetes ensues. The dysglycemia observed in these offspring is also attributed to increased glucocorticoid activity, impaired glucose transporter expression, augmented gluconeogenesis, and blunted-insulin suppression all within the liver. Moreover, this phenotype is propagated to the next generation when intrauterine artery-ligated females develop gestational diabetes during pregnancy. Of note. uterine ligated guinea pigs also exhibit hepatic perisinusoidal or periportal fibrosis (e.g., higher expression of SMAD4, TGFβ1, and MMP-2) underlying the development of NAFLD.

Pharmacological models

Maternal glucocorticoid exposure

Glucocorticoids are administered to pregnant women pre-dominantly to advance fetal maturation. In both humans and animals, fetal overexposure to glucocorticoids leads to the birth of IUGR offspring. In early gestation, exposure to glucocorticoids is minimal as the expression of placental 11b-hydroxysteroid dehydrogenase 2 (11b-HSD2)—an enzyme that catalyzes the conversion of active glucocorticoids into their inactive forms—is relatively high. It is noteworthy that children homozygous for mutations of *11β-HSD2* gene are lower in birth weight compared to heterozygous siblings. However, as the adrenal glands are activated in the late gestation, the synthesis of glucocorticoids increases to enable maturation of fetal tissues. Overexposure to glucocorticoids in late gestation leading to IUGR infants. This scenario can be mimicked by administration of synthetic compounds, for example, dexamethasone or inhibitors of 11β-HSD2. Treatment with dexamethasone in rats leads to decreased insulin content in fetal beta cells, due to the downregulation of the transcription factor PDX1—a critical factor involved in the development, differentiation, and function of pancreatic beta cells. In the nonhuman primates, the number of beta cells has also been shown to be decreased in the offspring following maternal dexamethasone treatment. Dexamethasone has proapoptotic effect on beta cells, which further explains its negative impact on endogenous pancreas. Maternal overexposure to glucocorticoids has been associated not only with the development of glucose intolerance and insulin resistance, but also hypertension in the offspring.

Gestational diabetes

Maternal administration of streptozotocin (STZ) is the most common pharmacological model for studying the effects of gestational diabetes on health of the offspring in rodents and sheep. STZ is a chemical that destroys pancreatic beta cells and acts in a dose-dependent manner. High doses cause severe maternal hyperglycemia and birth of IUGR offspring with hyperglycemia. The maternal hyperglycemia results in hyperstimulation of fetal pancreatic islets leading to their degranulation and consequently beta cell exhaustion. This leads to fetal hypoinsulinemia, which coupled with decreased insulin receptors on target cells, leads to diminished fetal glucose uptake. Low doses of STZ induce mild gestational diabetes and macrosomic offspring that have enhanced development of their endocrine pancreas with hyperplasia and hypertrophy of islets. These offspring also have an increased beta cell mass, increased pancreatic insulin content, and enhanced insulin secretion. However, in adulthood these animals develop a deficit in insulin secretion and impaired glucose tolerance.

Cellular and molecular mechanisms

The phenotypic outcomes of nutritional, surgical, and pharmacological insults designed to challenge the fetal environment are very similar, indicating the existence of common pathways and mechanisms linking early life experiences to long-term health. These are therefore a major focus of the research in the field to define the mechanisms involved.

Epigenetic mechanisms

Epigenetics refers to covalent modifications of DNA and his-tones that alter gene expression without affecting the DNA nucleotide sequence. Epigenetic alterations include DNA methylation and post-translational histone modifications such as methylation, acetylation, phosphorylation, ubiquitination, and sumoylation. In recent years, microRNAs (miRs) have also emerged as a potential epigenetic mechanism. These small, on average 22 nucleotides long noncoding RNAs have been traditionally associated with post-transcriptional gene regulation; however, recently they have been shown to also play a role in DNA methylation, thereby enabling them to further regulate transcription of their targets. The main role of epigenetic modifications is to heritably promote transcriptional silencing of specific gene regions so that varying gene expression levels can be achieved from identical DNA

A number of studies to date have shown that maternal nutrition during pregnancy can lead to permanent changes in the epigenome of offspring. One of the first studies showing the link between maternal nutrition and methylation status in the offspring was conducted in mice that carried the epigenetically sensitive Agouti viable yellow (*Avγ*) allele. Offspring of *Avγ* dams fed a diet supplemented with methyl donors (folate, vitamin B12, choline, or betaine) were leaner and had a different coat color (pseudo-Agouti) in comparison to obese and yellow offspring of normally fed dams. The difference in phenotype was caused by hypermethylation of a retro-transposon element downstream of the *Avγ* allele, which led to the silencing of the *Avγ* gene. Depending on the pattern of methylation, a wide variation in coat color ranging from yellow (unmethylated) to pseudo-Agouti (methylated) could be achieved. Moreover, there is greater hypermethylation of the *HNF4α* gene in $CD34^{+}$ hematopoietic stem and progenitor cells from cord blood from IUGR neonates. The decrease in HNF4α is linked to the development of type II diabetes. Maternal methyl donor supplementation has also been shown to increase in DNA methylation, of another epigenetically sensitive allele *Axin(Fu)* resulting in a 50% reduction in the incidence of tail kinking in the offspring. In addition, high levels of methyl vitamins in rodent pregnancy leads to hypermethylation of the *leptin* promoter contributing to obesity and insulin resistance. Hypomethylation has also been reported in sheep that were exposed to a diet deficient in methyl donors during the peri-conceptional period. These sheep developed insulin resistance and hypertension in adult life. Studies in the offspring of protein-restricted dams showed increase hepatic DNA methylation of the insulin-like growth factor 2 (*IGF2*) gene and parallel reduction in gene expression. The methylation status of gene promoters in the offspring can also be affected by material protein-restriction. Both glucocorticoid receptor (*GR*) and

peroxisome proliferator-activated receptor-a (*PPAR*a) gene promoters were found to be hypomethylated in the livers of protein-restricted offspring. Parallel changes in the expression of corresponding genes were also observed. Maternal diet has also been shown to influence epigenetic status in humans. Altered methylation of the *IGF2* gene was observed in the white blood cells of individuals who were exposed *in utero* to the Dutch Hunger Winter. Dietary induced changes in DNA methylation can also leads to transgenerational effects. For example, uterine-ligated IUGR offspring exhibit higher DNA methylation in the promoter of *Igf-1* which persists into the F2 generation even if F1 offspring are completely nourished.

Alterations in histone modifications are also emerging as an important mechanism of developmental programming. In general, euchromatin is associated with histones which are acetylated on specific residues (e.g., K9/K14 on Histone H3) whereas heterochromatin contains mainly hypoacetylated and/or methylated histones. Maternal nutrition has been shown to affect histone acetylation in the offspring. When the diet of *Avy* dams was supplemented with the phytoestrogen genistein (nonmethyl donor) the color of the offspring's coat was shifted toward pseudo-Agouti due to the increased DNA methylation of the *Avy* retrotransposon. It was proposed that genistein altered histone modifications consequently affecting the chromatic structure, DNA methylation, and gene expression. Recently it has been shown that hypomethylation in yellow *Avy* mice corresponds with enrichment of some activating histone acetylation modifications (H3 and H4 diacetylation), whereas hypermethylation corresponded with the repressive histone H4K20 tri-methylation modification. Therefore, it appears that histone modifications act in concert with DNA methylation to affect interindividual variation of epigenetically sensitive genes. Intrauterine artery ligation has been show to affect both DNA methylation and histone acetylation of the PDX 1 promoter. Additionally, losses in acetylation and an increase in dimethylation of histone, H3 at the GLUT4 locus persisting into adulthood, have been shown in the offspring of calorie-restricted dams. In low protein restricted offspring which exhibit hypercholesterolemia and lower hepatic expression of Cyp7a1 (involved in cholesterol catabolism), this was associated with higher trimethylated histone H3 [K9] and suppressed acetylation of histone 3 [K9,14] all surrounding the promoter of *Cyp7a1* from 3 weeks to 4 months. Interestingly, female low protein offspring do not exhibit these epigenetic changes in *Cyp7a1* or altered cholesterol levels. Recently, an association has been reported between periconceptional undernutrition in sheep and marked epigenetic changes in two hypothalamic genes: *GR* and an appetite-regulating neuropeptide, *proopiomelanocortin* (*POMC*). These epigenetic modifications could predispose the offspring to altered regulation of food intake, energy expenditure, and glucose homeostasis in later life.

As previously mentioned, the role of microRNAs (miRs) to epigenetic influence the developmental expression of metabolic genes has emerged in human and animal models of IUGR. In low birth weight humans, higher expression of miR-483-3p in the adipose is associated with higher insulin resistance, decreased lipid storage, and increased lipotoxicity. In low protein rat offspring, the entire miR-29 family (e.g., miR-29a/b/c) is increased in liver silencing the expression of insulin growth factor 1 (*Igf-1*). Moreover, in uterine-ligated IUGR guinea pig offspring, decreased miR-146a was linked to increased expression of its profibrotic target gene *smad4* implicated in NAFLD.

Mitochondrial dysfunction and oxidative stress

Mitochondria generate most of the cell's supply of adenosine triphosphate (ATP) and are the main source of highly de-structive reactive oxygen species (ROS). Evidence of oxidative stress in the fetus has been found following uteroplacental insufficiency. Pups born following this manipulation had in-creased pancreatic oxidative stress and impaired mitochondria function, which progressively worsens with age. This was associated with a decline in ATP production and accumulation of mitochondrial DNA damage. Pancreatic beta cells are especially vulnerable to ROS due to their high oxidative energy requirement and very low expression of antioxidant enzymes. In addition to the direct effects of the perinatal nutritional environment on oxidative stress/mitochondrial dysfunction, rapid postnatal catch-up growth may further contribute to oxidative stress and long-term dysmetabolism, For example in low protein restricted offspring, postnatal catch-up growth was *exclusively* associated with oxidative stress (i.e., elevated 4-HNE) altered aerobic metabolism (i.e., decreased citrate synthase and TFAM), along with endoplasmic reticulum (ER) stress in the liver. Oxidative stress affects not just mitochondrial DNA but also genomic DNA. ROS induced attrition of telomeres located at the ends of chromosomes has been associated with development of chronic disorders. Impaired antioxidant defense mechanisms and increased ROS production have also been observed when offspring were exposed to intrauterine nutritional excess. ROS production through xanthine oxidase (XO) activation was increased in cord plasma and placenta of neonates born to mothers with gestational diabetes. It has been proposed that exposure to oxidative stress can directly mediate DNA methylation and chromatic remodeling; however, further research is required to understand the underlying mechanisms.

Excessive intrauterine exposure to lipids

Obese and diabetic mothers often give birth to large for gestational age (44,000 g) babies. However, as only 25% of the differences in birth weight can be attributed to maternal hyperglycemia, the majority of large infants are born to nor-moglycemic mothers. This would suggest that other factors besides glucose may be involved. Indeed maternal pre-pregnancy BMI, maternal fasting triglyceride, and free fatty acid levels have all been implemented in mediating excessive fetal growth. In rodents, fetal exposure to excessive lipid levels leads to lipid accumulation in the adult offspring's liver, pre-disposing the offspring to nonalcoholic fatty liver disease. Deposition of lipids in liver and muscle can cause mitochondrial dysfunction. Increased fetal lipids may also promote formation of adipocytes over other cell types such as myocytes in early organogenesis. Circulating saturated fatty acids can activate kinases that cause an increase in IRS1 serine phosphorylation, an event that is associated with inhibition of insulin signaling and one of

the hallmarks of insulin resistance. Excessive fetal lipid concentrations may also affect hypothalamic regulators of appetite and satiety. For example, consumption of a high-fat diet during pregnancy in the non-human primate may compromise the development of the melanocortin system in the fetal hypothalamus. Finally, increased pancreatic beta cell mass and excess insulin secretion, which can lead to islet cell failure and contribute to the development of diabetes, can be found in the models of maternal obesity during pregnancy.

Conclusions

There is no doubt that environmental challenges experienced by a growing fetus can determine the risk of developing chronic diseases such as type 2 diabetes, obesity, hypertension, and other features of the metabolic syndrome in the individual. In light of the increasing prevalence of these disorders across the globe it is of critical importance to understand the mechanisms underlying the developmental origins of health and disease. Further research into the interactions between nutrition and chromatin dynamics and the role that particular nutrients play in fetal metabolic programming will be vital in allowing targeted early interventions to be developed. This may ultimately lead to the establishment of safe prevention strategies.

Further reading

Bouret, S.G., 2010. Neurodevelopmental actions of leptin. Brain Res. 1350, 2–9.
Fowden, A.L., Forhead, A.J., 2009. Endocrine regulation of feto-placental growth. Horm. Res. 72, 257–265.
Godfrey, K.M., Gluckman, P.D., Hanson, M.A., 2010. Developmental origins of metabolic disease: life course and intergenerational perspectives. Trends Endocrinol. Metabol. 21, 199–205.
Hales, C.N., Barker, D.J., 2001. The thrifty phenotype hypothesis. Br. Med. Bull. 60, 5–20.
Heerwagen, M.J., Miller, M.R., Barbour, L.A., Friedman, J.E., 2010. Maternal obesity and fetal metabolic programming: a fertile epigenetic soil. Am. J. Physiol. Regul. Integr. Comp. Physiol. 299, R711–R722.
Hoffman, D.J., Powell, T., Barrett, E., Hardy, D.B., 2021. Developmental origins of metabolic diseases. Physiol. Rev. https://doi.org/10.1152/physrev.00002.2020.
Jansson, T., Powell, T.L., 2007. Role of placenta in fetal programming: underlying mechanisms and potential interventional approaches. Clin. Sci. 113, 1–13.
McMillen, I.C., Robinson, J.S., 2005. Developmental origins of the metabolic syndrome: prediction, plasticity, and programming. Physiol. Rev. 85, 571–633.
Newnham, J.P., Ross, M.G. (Eds.), 2009. Early Life Origins of Human Health and Disease. S. Karger AG, Basel, Switzerland.
Pinney, S.E., Simmons, R.A., 2010. Epigenetic mechanisms in the development of type 2 diabetes. Trends Endocrinol. Metabol. 21, 223–229.
Plagemann, A., 2005. Perinatal programming and functional teratogenesis: impact on body weight regulation and obesity. Physiol. Behav. 86, 661–668.
Poston, L., 2010. Developmental programming and diabetes—the human experience and insight from animal models. Best Pract. Res. Clin. Endocrinol. Metabol. 24, 541–552.
Sohi, G., Marchand, K., Revesz, A., Arany, E., Hardy, D.B., 2011. Maternal protein restriction elevates cholesterol in adult rat offspring due to repressive changes in histone modifications at the cholesterol 7α-hydroxylase promoter. Mol. Endocrinol. 25 (5), 785–797.
Warner, M.J., Ozanne, S.E., 2010. Mechanisms involved in the developmental programming of adulthood disease. Biochem. J. 427, 333–347.

Early origins of disease: Non-fetal

Daniel J. Hoffman and Alessandra D. Sommer, Rutgers University, New Brunswick, NJ, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of L.S. Adair, Early Origins of Disease: Non-Fetal, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 106–112, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00082-9.

Introduction 58
Environmental exposures 59
Infant feeding and adult health 59
Selected outcomes related to infant feeding 59
Infant feeding and blood pressure 59
Infant feeding and body composition 60
Infant feeding and obesity 60
Autoimmune diseases 60
Postnatal growth and later risk of disease 60
Adiposity and obesity 62
Body composition and metabolism 62
Cancer 62
Blood pressure and coronary heart disease 63
Insulin resistance and diabetes 64
Summary 64
References 65
Further reading 65

Key points

- Nutritional insults early in life may have lifelong effects on health.
- Breastfeeding and proper complementary feeding protect against poor growth and chronic diseases in adulthood.
- Deficits in post-natal growth may promote chronic diseases later in life.
- Mechanisms to explain how early nutrition influences adult health are emerging with a focus on the role of epigenetics

Introduction

A substantial body of evidence supports the hypothesis that adult chronic diseases have origins in early life, formerly known as the Developmental Origins of Health and Disease or DOHaD (Hoffman et al., 2021; Bianco-Miotto et al., 2017; Bansal and Simmons, 2018). The basic premise of research in this field is that nutritional insufficiency during sensitive developmental periods results in structural changes or programming of metabolic functions. In the short term, such adaptations enhance survival and spare brain growth at the expense of other organs. In the long run, the cost of such adaptive responses may be an increased risk of chronic disease under specific environmental conditions, such as a diet high in ultra-processed foods.

The focus of DOHaD research has been primarily on the fetal origins of adult disease, but there remains substantial potential for nutritional programming of later disease risk during infancy and childhood. The young infant has high energy and nutrient needs to support rapid growth and development. Birth weight typically doubles in the first 4–6 months of life, and length increases by approximately 30% between birth and 6 months. Many organ systems continue to mature after birth, notably the immunologic, gastrointestinal, and renal systems. This combination of rapid growth and continued development make the infant highly susceptible to the effects of environmental exposures and suboptimal nutrition, which might affect the development of disease risk. Differentiating postnatal from fetal origins is challenging, however, owing to the inevitable link between pre- and postnatal growth.

Instances of purely postnatal effects relate primarily to infant feeding or exposure to pathogens or toxins. The potential effects of infant feeding relate to nutritional adequacy, and to exposure, or lack of exposure, to specific substances in human milk or human milk substitutes. Effects of feeding may occur independently of the infant's nutritional status at birth. This topic is discussed further in a separate section below. There is also a continuum of fetal and postnatal effects. Intrauterine growth-restricted infants may experience optimal or even excess postnatal nutrition, or they may continue to be exposed to nutritional insufficiency. The responses to postnatal challenges may be conditioned by the fetal nutritional history, such that there is an interaction or synergism of fetal and postnatal effects.

Prenatal nutritional insufficiency may be thought to result in "downsizing." It may produce smaller organs, for example, kidneys with a reduced nephron number, a pancreas with fewer islet cells, or a low skeletal muscle mass. Nutritional insufficiency may also alter metabolic or hormonal regulation, for example, hormone secretion or sensitivity of the hypothalamic–pituitary axis. In either case, the effects may be permanent, or subject to compensatory responses once nutritional or other insults are removed. For example, a permanently reduced nephron number is a hypothesized mechanism through which fetal growth restriction affects later blood pressure. Similarly, a reduced skeletal muscle mass may persist and affect insulin sensitivity in later life associated with a reduced number of insulin receptors. In such cases, the physiological capacity to respond to risk factors encountered later in life (e.g., diets high in sodium or excess calories relative to energy needs) may be compromised.

Alternatively, rapid growth following growth restriction (i.e., catch-up growth) or compensatory postnatal growth may occur. Many infants who were underweight for length at birth typically undergo a period of rapid postnatal compensatory growth in weight, whereas those who are relatively short at birth have larger length increments (see Fig. 1 for an example from a Philippines infant cohort). A central finding in many studies is that chronic disease risk is most likely to be elevated in individuals who were growth restricted in utero and thus small at birth, but relatively large at the time health outcomes were measured, leading to the conclusion that excess postnatal growth contributes to disease risk. Indeed, it has been shown that regardless of birth weight, children who grow more rapidly in length, independent of attained body mass, are more likely to have greater adipose tissue mass, a risk factor for many metabolic diseases. The extent to which rapid postnatal growth itself is a risk factor for the development of chronic disease has been the subject of extensive recent research. The relationship of early growth patterns to later disease risk is discussed in detail in a subsequent section.

Environmental exposures

Infant feeding and adult health

Much of the literature on the long-term effects of infant feeding is based on comparison of composition and health outcomes associated with human milk versus infant formula feeding however, effects found tend to be modest. Before discussing the results of these studies, it is important to raise several key methodological issues relevant to the interpretation of the literature. First, mothers who breast-feed may differ systematically or have varying health behaviors that contribute to their choice to breast feed. Unfortunately, most studies have insufficient data to adequately control statistically for these other behaviors, particularly because they are often unmeasured or poorly measured. Second, many studies use historical cohorts which may be subjected to recall bias and skewed information about breast-feeding duration and timing of introduction of other foods. Third, the composition of proprietary infant formulas has changed since their introduction in the 1920s. Therefore, results from older versus younger cohorts may differ either because true age-specific effects have emerged or because they were exposed to infant formula of different composition. The ideal study design for determining the long-term effects of infant feeding would require randomization to feeding regimens, and frequent follow-up of subjects up to the time when a disease risk factor or outcome is measured. Such designs are rarely ethical or feasible. Although many of the current studies have focused on neurodevelopment, some are now looking at other health outcomes.

Selected outcomes related to infant feeding

The following are examples of some chronic disease-related outcomes studied in relation to infant feeding. The selected outcomes are intended to be illustrative of a range of effects rather than a comprehensive treatment of all outcomes related to infant feeding.

Infant feeding and blood pressure

Differences in the sodium and fat content in breast milk versus formula are thought to be the relevant determinants of long-term effects of infant feeding on blood pressure. In a recent systematic review, data were compiled to compare exclusive breast-feeding to formula feeding. On average, subjects who were breast-fed had a modestly lower systolic blood pressure than those who had been

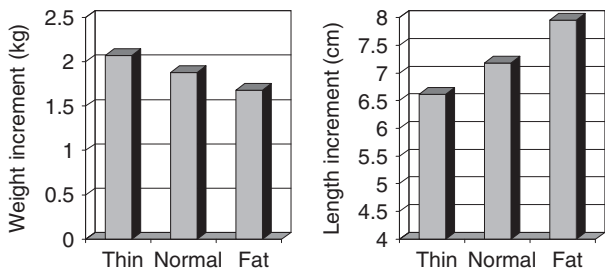


Fig. 1 Early growth of Filipino infants is associated with relative weight at birth. Mean growth increments from birth to 2 months of age in children who were relatively thin (BMI <10th sample percentile) or fat (BMI >90th sample percentile).

formula fed but studies showed no effects on diastolic blood pressure. One study on infants that included a 15-year follow-up, found that a diet of low or normal sodium intake resulted in slightly lowered blood pressure later in life. Another long-term follow-up study found blood pressure to be positively associated with dried formula milk supplement consumed in infancy compared to usual care, suggesting an effect on diet composition independent of growth. Mode of feeding may indirectly affect later disease risk through its effects on energy intake or aspects of metabolic regulation that affect growth and body composition. Numerous studies demonstrate different growth patterns in breast- and formula-fed infants that are hypothesized to reflect differences in nutrient intakes. In fact, evidence of systematic differences in breast- and formula-fed infants led the World Health Organization to undertake the production of growth charts for breast-fed infants.

Infant feeding and body composition

In one careful study of body composition, total energy intakes and weight velocity from 3 to 6 months of age were higher in formula-fed compared to breast-fed infants. Estimates of fat and fat-free mass also indicate higher adiposity in formula-fed infants, however, none of these differences persisted into the 2nd year of life. Similarly, in a study of just over 15,000 children who participated in the Growing Up Today Study, infants who were breastfed for 3 months or less and infants who were breastfed for at least 7 months were less likely to be overweight during later childhood and adolescence (Gilman et al., 2001).

Infant feeding and obesity

The available data on whether breast feeding provides protection against later overweight, or obesity is inconsistent. It has previously been shown that rapid weight gain during infancy is associated with increased risk of obesity later in life. On the other hand, a large prospective birth cohort study suggests that breast-fed and formula-fed infants have the same risk in childhood overweight. Some have proposed that the lipid and protein content in breastmilk is the primary driver of infant body composition. Thus, it is not clear, based on the available data, whether the effects of infant feeding are causal or whether breast-feeding serves as a marker for other health behaviors that may affect child and adolescent growth. Recent studies among siblings, which allow control for maternal characteristics, show no protective effects of breast-feeding on obesity in adolescents and young adults.

Autoimmune diseases

The infant's diet is the main source of exposure to antigens suspected to be related to the development of autoimmune diseases. Exposure to bovine proteins by feeding cow's milk, and to allergenic plant proteins, such as those found in wheat, is suspected to increase risk of developing diseases such as type 1 diabetes and celiac disease in genetically susceptible individuals. First, type 1 diabetes, one of the most prevalent chronic diseases with childhood onset and is characterized by autoimmunity to pancreatic islet cells. Early introduction of cow's milk has received a great deal of attention as a potential risk factor to type 1 diabetes. Some studies found no association while others suggest that eliminating cows' milk proteins in at-risk infants reduces risk of developing islet cell autoantibodies. In terms of mechanisms, it has been suggested that early enteroviral infections play a role in the etiology of type 1 diabetes in genetically susceptible individuals. In fact, some assert that the effect of cows' milk diabetes risk may depend on viral exposures. Moreover, recent studies suggest a role for other food antigens. For example, a study of at-risk German children found that feeding of gluten-containing foods before 3 months of age was associated with risk of having pancreatic islet cell autoantibodies (Ziegler et al., 2003). Another study in the US also found an increased risk of islet cell autoimmunity among at-risk children given cereal before 3 months or after 7 months of age. Furthermore, they found that risks associated with cereal introduction were reduced by breast-feeding. While more studies clearly need to be developed to better understand these relationships, it is clear that the development of autoimmune diseases may be preceded by nutritional insults that may affect the immune response of children.

Few studies have assessed the relationship of infant feeding to later development of T2D. Early feeding may affect patterns of insulin secretion in the newborn period, and thereby program subsequent development of metabolic control. Two studies in native American populations, one in Canada and one among Pima Indians, report a protective effect of breast-feeding on later development of T2D. In the Pima study, exclusive breast-feeding in the first 2 months of life was associated with a lower rate of T2D in children and adults (Pettitt et al., 1997). In the Canadian study, breast-feeding for more than 12 months was associated with decreased risk of T2D (Young et al., 2002). While the mechanisms for these observations are still poorly understood, the developing picture seems to suggest a broad impact of poor growth on a number of developing systems that are associated with normal substrate metabolism.

Based on the discussion above, it is clear that infant feeding, through nutritional adequacy, direct exposure to antigens, and protective substances provided in human milk, has the potential to alter response to subsequent exposures and to directly influence the beginning of disease processes. The real lesson for the future will be to ensure that all children have access to adequate feeding programs from birth through adolescence.

Postnatal growth and later risk of disease

Small body size in childhood may reflect nutritional insufficiency that may program adult disease like that observed in the fetal period. Independent of birth weight, low weight at 1 year of age has been associated with increased risk of cardiovascular disease in adult men. Similarly, poor childhood growth manifested as short stature has been linked with insulin resistance.

More attention has recently been paid to the effects of rapid childhood growth in height and weight. The observation in much of the fetal programming literature that effects of birth size emerge or are strengthened when current body size (typically represented as BMI) is taken into account suggests an important role for postnatal growth in the origins of adult disease. Individuals who are born small, but who end up relatively large (taller or heavier than their peers) have clearly experienced more rapid growth at some point between birth and when health outcomes and current size are assessed. Whether rapid growth is an independent risk factor or whether it confers increased risk only in individuals with a history of intrauterine growth restriction is a question requiring further research. Moreover, even when strong associations of growth rate and chronic disease risk are found, it is unclear whether the association is causal or whether growth serves as a marker for other underlying causal processes.

Postnatal growth is clearly related to prenatal growth. Some metabolic changes associated with prenatal nutritional sufficiency may affect postnatal physiology and behavior that, in turn, affect growth. In addition, there is intriguing evidence from animal studies that prenatal nutritional restriction alters appetite and induces hyperphagia, and also reduces physical activity in adult animals (see Fig. 2). If true in humans, this would be an important pathway by which disease risk is affected. Suggestive evidence comes from human infants whose cord blood leptin levels at birth were inversely related to weight gain in the first 4 months of life, independent of birth weight. Leptin may relate to subsequent growth by affecting appetite and energy intake.

Depending on the outcome under study, there are differences in whether linear growth or growth in weight, particularly weight relative to height, matters. Most often, more rapid weight gain is the risk factor, owing to the fact that excess adiposity is an important risk factor for many chronic diseases of adulthood. Another key issue concerns the timing of effects. There is controversy about whether early infancy compensatory growth following intrauterine growth restriction confers risk, or whether it is only later growth that matters. Indeed, one study of children followed from birth through early childhood found that the growth promoting hormones IGF-I and IGF-II were associated with postnatal catch-up growth (Fig. 3). This would suggest that IGF-1 levels in early childhood are positively influenced by rapid post-natal growth.

Where many potential adverse outcomes might be affected by postnatal growth, the following sections focus on adiposity, blood pressure, coronary heart disease, insulin resistance, diabetes, and cancer.

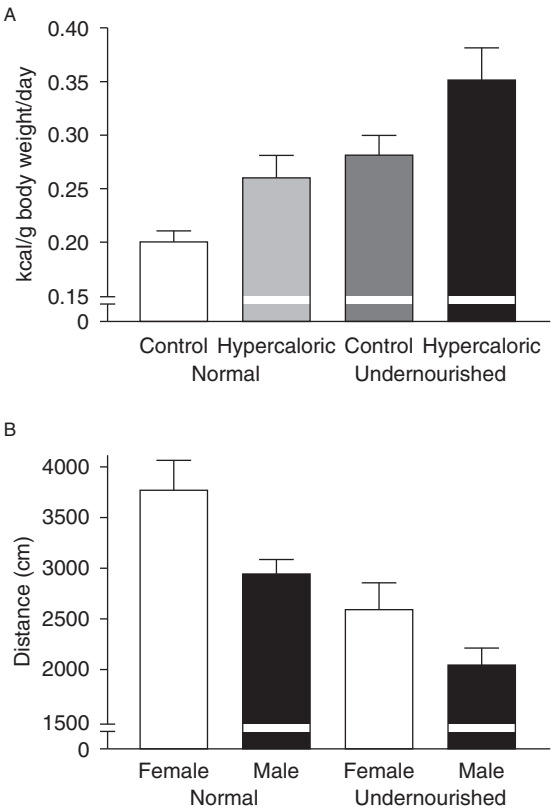


Fig. 2 Locomotor behavior and food intake in Wistar rats as a consequence of a normal or adverse fetal environment ($n = 6$–8/group). (A) Food intake (kcal per gram body weight per day over a 5-day period) in females at day 145; $P < 0.005$ for effect of fetal programming, $P < 0.05$ for postnatal hypercaloric diet. (B) Locomotor activity at 14 months in males and females; $P < 0.005$ for effect of fetal programming and gender. Data analyzed by factorial ANOVA, and data are shown as means ± SE. Reproduced from Vickers, M.H., Breier, B.H., McCarthy, D., Gluckman, P.D., 2003. Sedentary behavior during postnatal life is determined by the prenatal environment and exacerbated by postnatal hypercaloric nutrition. Am. J. Physiol. Regul. Integr. Comp. Physiol. 285(1), R271–R273, with permission from APS.

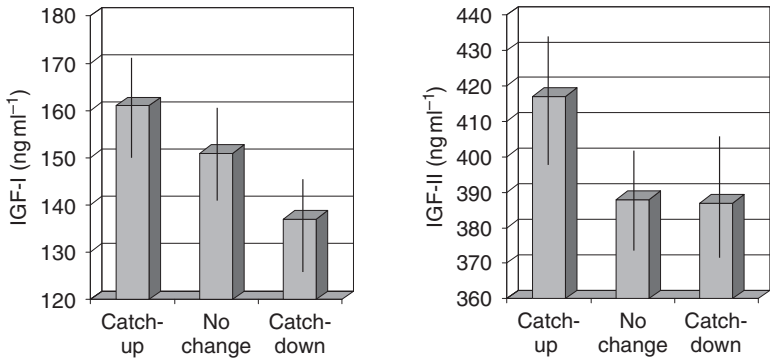


Fig. 3 Hormone levels at age 5 years by change in weight *Z*-score from birth to 2 years of children in the ALSPAC cohort: means and 95% confidence intervals of IGF-I and IGF-II, adjusted for fat mass and fat-free mass. Data drawn from Ong, K., Kratzsch, J., Kiess, W., Dunger, D., ALSPAC Study Team, 2002. Circulating IGF-I levels in childhood are related to both current body composition and early postnatal growth rate. J. Clin. Endocrinol. Metab. 87(3), 1041–1044.

Adiposity and obesity

Early undernutrition followed by later overnutrition as well as early overfeeding independent of prior growth restriction are thought to increase risk of later obesity. Rapid postnatal weight gain occurs in a significant proportion of infants who are born small for gestational age. Prospective studies from a number of cohorts from around the world show that rapid growth in early infancy increases later risk of overweight (Baird et al., 2005; Cameron et al., 2003; Monteiro and Victora, 2005; Okihiro et al., 2012). Longitudinal data from the US National Perinatal Collaborative study show that, independent of birth weight, one-third of obesity at age 20 is attributable to rapid weight gain in the first 4 months of life. In a Bristol, UK cohort, nearly one-third of children had an increased weight standard deviation (SD) score of more than 0.67 units from birth to age 2 years, and these children remained fatter, having more central fat distribution at age 5 years compared to children with lower early growth rates. Similarly, data from the South Africa Birth to Ten cohort showed that children with rapid weight gain in infancy were significantly lighter at birth and significantly taller, heavier, and fatter throughout childhood. Finally, a study of growth from birth to age 8 found that children who grew more rapidly than others were more likely to be taller and leaner, independent of current BMI and birth weight.

Body composition and metabolism

Clinical studies of children who experience growth deficits have revealed much information on the long-term consequences of poor growth. Briefly, it has been found that children born small are more likely to be obese compared to those born with average birth weight. However, most of these studies used BMI or BMI *Z*-score to assess obesity and did not specifically measure adiposity. In studies that used measures of body composition, children who were growth restricted had more truncal fat mass and more total body fat mass, independent of attained body size. While these differences between studies exist, it is important to recognize that comparing studies that used varying sample sizes and outcomes (body composition versus BMI), the more explicit studies support the conclusion that poor growth early in life is a risk factor for excess adiposity later in life. Several studies have explored the metabolic implications of poor growth. One set of studies from Guatemala and Brazil found no difference in resting metabolic rate (RMR), between growth restricted and normal height children, even after controlling for lean body mass (LBM). On the other hand, some studies have reported that RMR is lower in growth restricted children. Explanations for these apparently divergent results include inconsistent statistical methods as those studies which used RMR per unit body weight, did not properly adjust for body weight or LBM, the most important predictors of RMR.

In terms of specific aspects of energy expenditure, several studies have found that growth restriction is associated with metabolic adaptations that may protect a child during periods of food insecurity or illness but may predispose them to obesity under favorable environmental conditions. As presented in **Fig. 4**, studies conducted in different cohorts in Brazil, Russia, and South Korea consistently found that children and adults who experienced linear growth restriction had lower rates of fat oxidation compared to peers who had not suffered growth deficits (Hoffman, 2000; Lee et al., 2015; Leonard, 2009.) These studies support the general hypothesis that poor growth early in life is a risk factor for nutrition-related chronic diseases. While specific physiological mechanisms to explain these apparent metabolic differences following episodes of growth restriction, it is most reasonable to assume that a dysregulation of normal metabolism occurs that remains following recovery, perhaps due to poor lean tissue development or biochemical modifications that favor fat storage during periods of nutritional insults.

Cancer

A large body of literature relates adult height to cancer risk, with the largest volume of evidence on breast, prostate, and colorectal cancers. In each case, risk of disease is increased with taller stature. A role for accelerated childhood growth is inferred as taller

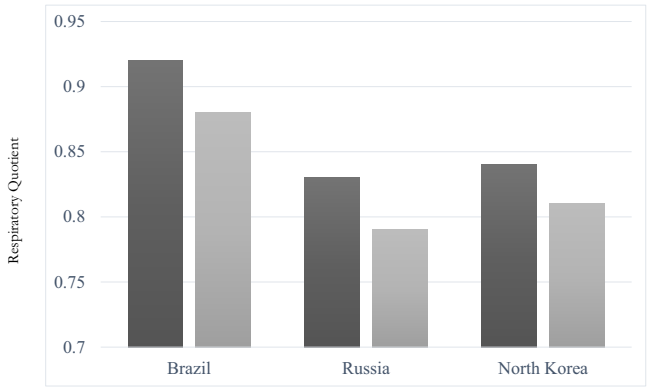


Fig. 4 Differences in respiratory quotient assessed by indirect calorimetry in growth restricted children compared to normal height peers in three cohorts. Growth Restricted ■ Control □. Data drawn from Hoffman, D.J., Sawaya, A.L., Verreschi, I., Tucker, K.L, Roberts, S.B., 2000. Why are nutritionally stunted children at increased risk of obesity? Studies of metabolic rate and fat oxidation in shantytown children from São Paulo, Brazil. Am. J. Clin. Nutr. 72(3), 702–707; Lee, S.K., Nam, S.Y., Hoffman, D.J., 2015. Growth retardation at early life and metabolic adaptation among North Korean children. J. Dev. Orig. Health Dis. 6(4), 291–298; Leonard, W.R., Sorensen, M.V., Mosher, M.J., Spitsyn, V., Comuzzie, A.G., 2009. Reduced fat oxidation and obesity risks among the Buryat of Southern Siberia. Am. J. Hum. Biol. 21(5), 664–670.

individuals have experienced more linear growth. Possible mechanisms fall into two categories: childhood growth as a marker for other exposures that influence risk (fetal exposures, infections, timing of puberty, and energy intake) or growth as a mediator of risk (effects of growth promoting hormones such as IGF-I and IGF-II).

Few studies have directly addressed the effects of childhood growth, owing to lack of longitudinal data. Based on data from the UK Boyd Orr cohort, a one SD difference in height was associated with a 42% higher risk of overall cancer mortality in later life among males, but no effects were found in females. In another UK birth cohort, risk for breast cancer was elevated among women who were large at birth and tall at age 7 years (Whitley et al., 2009). Based on data from the US Nurse's Health Study, rapid adolescent growth was associated with an increased risk of both pre- and postmenopausal breast cancer.

Blood pressure and coronary heart disease

Blood pressure is a well-studied outcome in the context of DOHaD, with consistent findings of a modest inverse relationship of birth weight to adult systolic blood pressure. Substantial evidence demonstrates a synergistic relationship of fetal growth restriction with rapid postnatal growth. Fig. 5 presents the classic picture for systolic blood pressure: the highest pressure is found among adolescent males who were relatively thin at birth, but relatively heavy as adolescents. Current BMI is typically the strongest anthropometric predictor of blood pressure, but at the same BMI, those with a history of fetal growth restriction have higher mean systolic blood pressure and increased risk of having high blood pressure. Using longitudinal growth data in Scandinavia, child growth trajectories can be traced for individuals with and without hypertension or other adverse outcomes such as coronary heart disease. As shown in Fig. 6, though initially smaller, adults with hypertension diverged in their BMI trajectory and were relatively heavier after age 7 compared to those without hypertension.

There remains controversy about the age at which faster growth rates pose risk of later disease. Some studies show elevated blood pressure in association with rapid weight gain in infancy, whereas other studies show no effect, or a protective effect (infants with larger weight increments have lower blood pressure as adults). The degree to which rapid infant growth represents risk may depend on whether it occurs in the context of recovery from fetal growth restriction and results in normalization of body weight versus

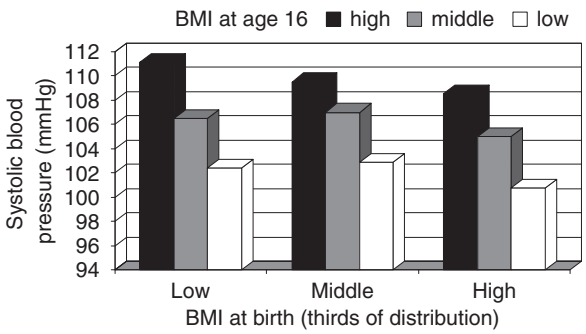


Fig. 5 Synergistic effect of BMI at birth and age 16 on systolic blood pressure of Cebu (Philippines) boys: ■ high; ■ middle; □ low BMI. Data from the Cebu Longitudinal Health and Nutrition Survey.

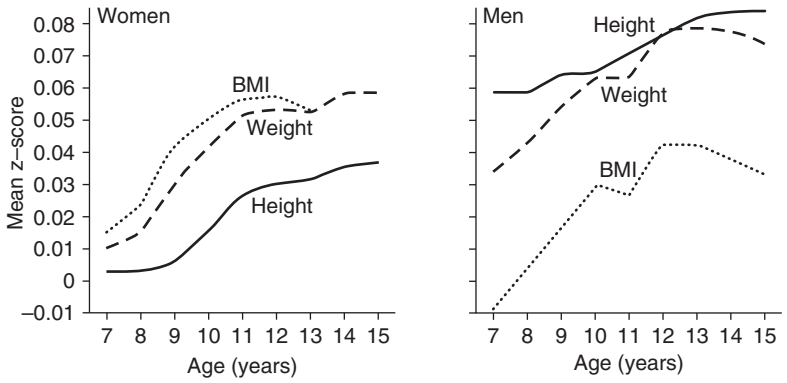


Fig. 6 *Z*-scores for height, weight, and BMI from 7 to 15 years in 975 boys and 983 girls who later developed hypertension. Mean values for all 7086 subjects in cohort are zero. Reproduced from Eriksson, J., Forsen, T., Tuomilehto, J., Osmond, C., Barker, D., 2000. Fetal and childhood growth and hypertension in adult life. Hypertension 36(5), 790–794, with permission from LWW.

excess growth leading to infant obesity. In a Philippine cohort, larger weight increments from age 8–15 years increased risk of high blood pressure in boys who were relatively thin at birth (Adair and Cole, 2003). However, higher childhood weight gain in the absence of fetal growth restriction was not a risk factor in this population.

The physiological changes associated with a higher risk of hypertension following fetal undernutrition may include a reduced number of nephrons. Such deficits may not increase disease risk in individuals who remain small, but excess growth may challenge the ability of the kidneys to effectively regulate blood pressure. Catch-up linear growth has not been consistently implicated as a risk factor for later elevated blood pressure. In fact, continued poor linear growth, particularly in association with more rapid weight gain, increases risk of later elevated blood pressure. Thus, it is important to consider both size at birth as well as post-natal growth patterns as possible mediating factors between poor growth and disease development.

Insulin resistance and diabetes

The relationship between poor growth and T2D suggest that both continued growth faltering in infancy and rapid growth rates are associated with an increased risk of insulin resistance and T2D. As well, postnatal faltering in length is also associated with impaired insulin metabolism. The highest risk of T2D is found when for persons with a small size at birth and rapid postnatal growth gain. In a well-studied cohort in Finland, men and women who developed T2D had lower birth weight, length, and ponderal index, and accelerated growth in weight and height from age 7–15 years. In terms of precursors of T2D, a follow-up study of British children who were born preterm, fasting split proinsulin and glucose concentration 30 min after a glucose load were highest in children with the greatest increase in weight centile between birth and time of measurement, regardless of early size. Thus, it is important to understand how the risk for T2D may vary between those exposed to nutritional insults in utero compared to post-natal periods.

To further illustrate this issue, several studies of the early origins of T2D have been conducted in India where the prevalence of T2D has increased dramatically in the past 20 years. Indian babies who are small at birth have a deficit in skeletal muscle, but not body fat compared to normal size infants. These infants tend to grow into adults that retain a lower skeletal muscle mass but have increased abdominal obesity, a key risk factor for T2D. Prospective studies of Indian children show an interaction between birth weight and subsequent growth. For example, children who were born small but were relatively large at age 4 had higher plasma glucose and insulin concentrations 30 min after an oral glucose load, and greater insulin resistance at age 8.

Higher growth rates in previously growth-restricted individuals may pose excessive demands on systems initially adapted to function in the face of limited resources, leading to increased risk of diseases, particularly those associated with the metabolic syndrome. Rapid growth in weight during infancy and childhood, and in particular, rapid growth following prenatal growth restriction, increases risk of developing obesity, especially abdominal obesity. Factors that contribute to early onset of obesity are therefore important to control, because obesity tracks from early life to adulthood, and is a well-recognized risk factor for diseases such as type 2 diabetes, hypertension, and coronary heart disease.

Summary

To summarize, the continued vulnerability and responsivity of the developing infant and child suggest the importance of a life course perspective on the development of diseases that are typically thought of as "adult onset." Based on the review of studies for this article, future research needs to focus on understanding the mechanisms of how infant feeding practices and growth in early childhood may predispose some children to develop chronic diseases later in life compared to those who were fed differently, were born normal size, or grew without delays.

See Also: Breastfeeding; Cancer: Epidemiology and associations between diet and cancer; Diabetes mellitus: Etiology and epidemiology; Hyperlipidemia; Hypertension: Dietary factors

References

Adair, L.S., Cole, T.J., 2003. Rapid child growth raises blood pressure in adolescent boys who were thin at birth. Hypertension 41 (3), 451–456.
Baird, J., Fisher, D., Lucas, P., Kleijnen, J., Roberts, H., Law, C., 2005. Being big or growing fast: systematic review of size and growth in infancy and later obesity. BMJ 331 (7522), 929.
Bansal, A., Simmons, R.A., 2018. Epigenetics and developmental origins of diabetes: correlation or causation? Am. J. Physiol. Endocrinol. Metab. 315 (1), E15–E28. Jul 1.
Bianco-Miotto, T., Craig, J.M., Gasser, Y.P., van Dijk, S.J., Ozanne, S.E., 2017. Epigenetics and DOHaD: from basics to birth and beyond. J. Dev. Orig. Health Dis. 8 (5), 513–519. https://doi.org/10.1017/S2040174417000733.
Cameron, N., Pettifor, J., De Wet, T., Norris, S., 2003. The relationship of rapid weight gain in infancy to obesity and skeletal maturity in childhood. Obes. Res. 11 (3), 457–460.
Gillman, M.W., Rifas-Shiman, S.L., Camargo Jr., C.A., Berkey, C.S., Frazier, A.L., Rockett, H.R., Field, A.E., Colditz, G.A., 2001. Risk of overweight among adolescents who were breastfed as infants. JAMA 285 (19), 2461–2467.
Hoffman, D.J., Sawaya, A.L., Verreschi, I., Tucker, K.L., Roberts, S.B., 2000. Why are nutritionally stunted children at increased risk of obesity? Studies of metabolic rate and fat oxidation in shantytown children from São Paulo, Brazil. Am. J. Clin. Nutr. 72 (3), 702–707.
Hoffman, D.J., Powell, T.L., Barrett, E.S., Hardy, D.B., 2021. Developmental origins of metabolic diseases. Physiol. Rev. 101 (3), 739–795. July 1.
Lee, S.K., Nam, S.Y., Hoffman, D.J., 2015. Growth retardation at early life and metabolic adaptation among North Korean children. J. Dev. Orig. Health Dis. 6 (4), 291–298.
Leonard, W.R., Sorensen, M.V., Mosher, M.J., Spitsyn, V., Comuzzie, A.G., 2009. Reduced fat oxidation and obesity risks among the Buryat of Southern Siberia. Am. J. Hum. Biol. 21 (5), 664–670.
Monteiro, P.O., Victora, C.G., 2005. Rapid growth in infancy and childhood and obesity in later life–a systematic review. Obes. Rev. 6 (2), 143–154.
Okihiro, M., Davis, J., White, L., Derauf, C., 2012. Rapid growth from 12 to 23 months of life predicts obesity in a population of Pacific Island children. Ethn. Dis. 22 (4), 439–444.
Pettitt, D.J., Forman, M.R., Hanson, R.L., Knowler, W.C., Bennett, P.H., 1997. Breastfeeding and incidence of non-insulin-dependent diabetes mellitus in Pima Indians. Lancet 350 (9072), 166–168.
Whitley, E., Martin, R.M., Smith, G.D., Holly, J.M., Gunnell, D., 2009. Childhood stature and adult cancer risk: the Boyd Orr cohort. Cancer Causes Control 20 (2), 243–251.
Young, T.K., Martens, P.J., Taback, S.P., Sellers, E.A.C., Dean, H.J., Math, M.C., Flett, B., 2002. Type 2 diabetes mellitus in children prenatal and early infancy risk factors among native Canadians. Arch. Pediatr. Adolesc. Med. 156 (7), 651–655.
Ziegler, A.G., Schmid, S., Huber, D., Hummel, M., Bonifacio, E., 2003. Early infant feeding and risk of developing type 1 diabetes-associated autoantibodies. JAMA 290 (13), 1721–1728. Oct 1.

Further reading

Dewey, K.G., 2003. Is breast feeding protective against child obesity? J. Hum. Lact. 19, 9–18.
Giannì, M.L., Roggero, P., Orsi, A., Piemontese, P., Garbarino, F., Bracco, B., Garavaglia, E., Agosti, M., Mosca, F., 2014. Body composition changes in the first 6 months of life according to method of feeding. J. Hum. Lact. 30 (2), 148–155.
Huxley, R.R., Shiell, A.W., Law, C.M., 2000. The role of size at birth and postnatal catch-up growth in determining systolic blood pressure: a systematic review of the literature. J. Hypertens. 18, 815–831.
Lucas, A., 1998. Programming by early nutrition: an experimental approach. J. Nutr. 128, 401s–406s.
Martorell, R., Stein, A.D., Schroeder, D.G., 2001. Early nutrition and later adiposity. J. Nutr. 131, 874s–880s.
Okasha, M., Gunnell, D., Holly, J., Davey Smith, G., 2002. Childhood growth and adult cancer. Best Pract. Res. Clin. Endocrinol. Metab. 16, 225–241.
Ong, K.K., Dunger, D.B., 2002. Perinatal growth failure: the road to obesity, insulin resistance and cardiovascular disease in adults. Best Pract. Res. Clin. Endocrinol. Metab. 16, 191–207.
Owen, C.G., Whincup, P.H., Odoki, K., Gilg, J.A., Cook, D.G., 2002. Infant feeding and blood cholesterol: a study in adolescents and a systematic review. Pediatrics 110, 597–608.
Owen, C.G., Whincup, P.H., Gilg, J.A., Cook, D.G., 2003. Effect of breast feeding in infancy on blood pressure in later life: a systematic review and meta-analysis. BMJ 327, 1189–1195.
Pekmez, C.T., Larsson, M.W., Lind, M.V., Vazquez Manjarrez, N., Yonemitsu, C., Larnkjaer, A., Bode, L., Mølgaard, C., Michaelsen, K.F., Dragsted, L.O., 2020. Breastmilk lipids and oligosaccharides influence branched short-chain fatty acid concentrations in infants with excessive weight gain. Mol. Nutr. Food Res. 64 (3), e1900977.
Swasmuth, H.E., Kolb, H., 2000. Cow's milk and immune-mediated diabetes. Proc. Nutr. Soc. 59, 573–579.
Yajnik, C.S., 2003. Nutrition, growth, and body size in relation to insulin resistance and type 2 diabetes. Curr. Diabetes Rep. 2, 108–114.

Growth and development: Physiological aspects

William W. Hay, Jr., University of Colorado, Denver, CO, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of W.W. Hay, Growth and Development: Physiological Aspects, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 399–407, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00139-2.

Introduction **67**
Growth of fetal size **67**
Developmental change of fetal body composition **67**
Water 69
Non-fat dry weight 69
Nitrogen balance, protein turnover, protein synthesis, and lean body mass 69
Glycogen 71
Fat 71
Caloric accretion in the fetus 72
Mineral accretion in the fetus 73
Regulation of fetal growth **73**
Genetic factors 73
Non-genetic maternal factors 74
Maternal nutrition 74
The placenta 74
Growth of the placenta and its transport capacity 75
Maternal endocrine influences on fetal growth 75
Fetal endocrine and autocrine/paracrine-acting growth factor influences on fetal growth **75**
Interpretation of growth curves **76**
Extremes of growth and development: intrauterine growth restriction and macrosomia **77**
Intrauterine growth restriction (IUGR) 78
Possible adult disorders resulting from intrauterine growth restriction 78
Macrosomia 80
Development of type II diabetes in later life in macrosomic offspring 80
Conclusion/summary/outlook **80**
References **80**
Further reading **81**

Key points

- The most rapid rate of growth and development in an organism's life occurs during the embryonic and fetal period of gestation.
- Measurements of fetal growth are difficult to make in utero; they are generally estimated from cross sectional anthropometric measurements of body weight, length, circumferences (head, chest, abdomen), and composition (e.g., fat mass and lean mass) of "normal" infants at birth at different gestational ages, complemented by fetal ultrasound estimates of length and circumferences.
- Fetal growth is regulated by placental size and the rate of placental transfer of oxygen and nutrients (amino acids for protein synthesis, glucose for energy production) to the fetus.
- Insulin and insulin-like growth factors contribute to fetal growth by promoting cell replication earlier in gestation and hypertrophy later in gestation.
- Principal disorders of fetal growth and development include intrauterine growth restriction (IUGR, slower than normal rate of fetal growth) and macrosomia (larger than normal body size, generally due to excess fat).
- Fetal IUGR and macrosomia may predispose the fetus to later life obesity, insulin resistance, and type 2 diabetes.

Glossary

GUR glucose utilization rate
GSIS glucose stimulated insulin secretion

IUGR intrauterine growth restriction (slower than normal rate of fetal growth)
AGA appropriate for gestational age
SGA small for gestational age
LGA large for gestational age
IGF insulin-like growth factor
Macrosomia larger than normal body size due to excess fat
Ponderal Index weight (grams)/[length (cm)]^3
Autocrine acting at the site of production, such as an autocrine growth factor
Paracrine acting adjacent to the site of production, such as a paracrine growth factor
Programming an insult or stimulus, when applied at a critical or sensitive stage in development, of sufficient magnitude and duration, that produces lasting, even lifelong, effects on the structure or function or both of the organism
Monotocous singleton fetus within the same uterus (e.g., in most human pregnancies)
Polytocous multiple fetuses within the same uterus (e.g., in rats, with 8–10 pups/litter)

Introduction

Growth and development refers to the growth of an organism in size as assessed by anthropometric measurements of body weight, length, and circumferences (head and body), as well as changes in body composition, primarily cell number and size, organ size, and the relative amounts of fat mass and lean mass (muscle, bone, organs). Most of the relevant concepts about growth and development apply to the fetal period of development, the focus of this review. This period encompasses the greatest changes in growth rate, body proportions, and body composition during the life of an organism. In the first third of gestation (first trimester), during the embryonic period, growth occurs primarily by increased cell number (hyperplasia). In the middle third of gestation (second trimester), cell size also increases (hypertrophy) while the rate of cell division stabilizes. In the last third of gestation (third trimester), the rate of cell division declines while cell size continues to increase.

Many terms are used to describe variations in growth. For example, human newborns are classified as having normal birth weight (greater than 2500 g), low birth weight (less than 2500 g), very low birth weight (less than 1500 g), or extremely low birth weight (less than 1000 g). Obviously, classification by weight alone says little about growth rate, as most infants with less than normal birth weights are the result of a shorter than normal gestation, i.e., they are born preterm (<37 weeks of gestational age). Furthermore, it is inappropriate to label newborns as abnormally grown when their birth weight is less than some arbitrarily determined "normal" birth weight, but their mother was quite small to begin with; such newborns are considered constitutionally small but not abnormal. Classifying newborns according to duration of gestation (e.g., preterm, term, or post term) on the basis of birth weight also is erroneous, because infants with intrauterine growth restriction (IUGR) are smaller and macrosomic infants of diabetic mothers are larger than normal at any gestational age.

Growth of fetal size

Under usual conditions, the fetus grows at its genetic potential. Small fetuses of small parents or large fetuses of large parents do not reflect fetal growth restriction or fetal overgrowth, respectively; in fact, their rates of growth are normal for their genome. The smaller (generally, shorter) the mother, the more she limits fetal growth by "maternal constraint," which represents a limitation of uterine size (Gluckman and Hanson, 2004). Uterine size is directly related to maternal height; thus, a shorter mother will have a smaller uterus with reduced endometrial surface area and the capacity for placental growth (Rochow et al., 2018). Placental size is the primary determinant of its capacity to supply oxygen and nutrients to the fetus. A clear example of maternal constraint is shown in **Fig. 1**, depicting the reduced rate of fetal growth of multiple fetuses in a monotocous species—human—that optimally supports only one fetus.

Fetal size tends to increase exponentially in the first third or trimester of gestation. It increases relatively linearly in the middle third or second trimester of gestation, and into the third part or third trimester of gestation, the period of largest increase in fetal mass. Fetal size then slows during the latter part of the third trimester, producing the typical S-shaped curve of fetal size vs. gestational age that is derived from cross-sectional measurements of birth weights at different gestational ages (**Fig. 2**). The length of gestation is more strongly related to the growth of neural tissue mass (range 0.015–0.033 $g^{1/3}$/day—a 2.2-fold range) than to the growth of the fetal body (range 0.033–0.25 $g^{1/3}$/day—a 7.6-fold range). The physiological significance of this relationship is not known, but intrauterine development of the relatively large brain in human fetuses is made possible by their relatively slower rate of somatic growth compared with other mammals.

Developmental change of fetal body composition

Fetal growth during the last third of gestation requires large increases in oxygen and nutrient supplies and appropriate utilization of oxygen and nutrients to produce net synthesis of complex molecules (e.g., protein) and cell replication and

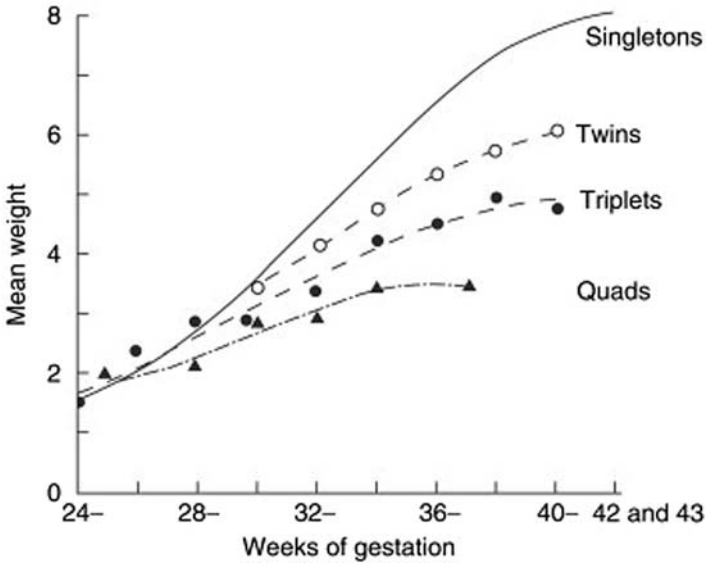


Fig. 1 Mean birth weight of single and multiple fetuses related to duration of gestation. Adapted from McKeown and Record (1952); Reproduced with permission from Ounsted and Ounsted (1973).

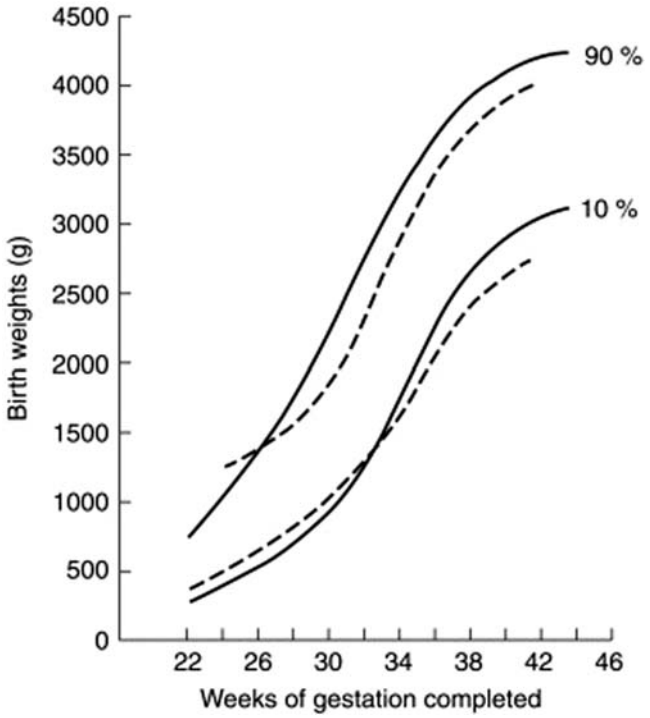


Fig. 2 Birth-weight percentiles for gestational age. Solid lines represent California total singleton live births, 1970–1976: dotted lines represent Colorado General Hospital (Denver, Colorado) live births, 1948–1960. Reproduced with permission from Creasy and Resnik (1999).

hypertrophy. Nutrient substrate supply is coupled with increased development of anabolic hormones and growth factors in fetal tissues and fetal plasma to produce increased nitrogen and carbon deposition in protein, carbohydrate deposition in glycogen, and fatty acid, glycerol, and triglyceride deposition as fat in adipose tissue. Growth of these tissues gradually replaces water in the fetal extracellular space.

Chemical composition studies of normal human infants are limited. Based on data from 15 studies that included 207 infants, nonfat dry weight and nitrogen content (predictors of protein content) show a linear relationship with fetal size and an exponential relationship with gestational age (**Fig. 3**) (Sparks, 1984). As gestation proceeds, larger fetuses grow faster than smaller fetuses, as do

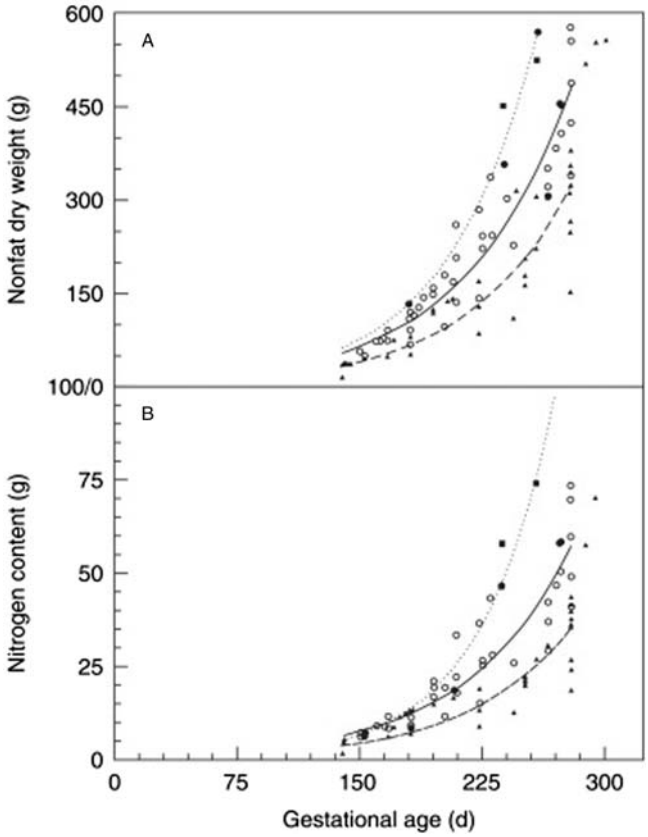


Fig. 3 Non-fat dry weight (A) and nitrogen content (B) are plotted against gestational age for LGA (■,· · · ·), AGA (◯, ———), and SGA (◆,——) infants. Reproduced with permission from Sparks (1984).

net protein accretion and tissue and organ mass. At each gestational age, however, fetal size (and thus birth weights) varies considerably, as do nonfat dry weight and nitrogen content; protein accretion and therefore protein nutritional requirements follow accordingly as gestational age advances and fetal size increases.

Water

Fetal water content increases directly with but not proportionally to body weight, as fetal body water, expressed as a fraction of body weight, decreases with advancing gestation. The relatively larger growth of adipose tissue in the human fetus compared with all other species of land mammals further dilutes the body concentration of water. Extracellular water as a fraction of fetal body weight also decreases more than intracellular water as gestation advances; this is mainly due to increasing cell number and increasing cell size rather than the intracellular concentrations of osmotic substances.

Non-fat dry weight

Comparative aspects of chemical and physical growth in fetuses of six different species are summarized in **Table 1**. Despite growth rate variations of up to 20-fold and weight-specific fat content variances at term of up to 16-fold among these species, nonfat dry weight and protein weight-specific contents as fractions of total weight at term are constant. Protein concentration is about 12% in all species at term and fetal protein content is linearly related to fetal weight; thus, protein accretion in the fetal rat occurs about 23 times faster than it does in the human. These species-related differences in growth rate are remarkable and require marked differences in the capacity of the placenta to transport nutrients to the fetus.

Nitrogen balance, protein turnover, protein synthesis, and lean body mass

According to animal data, only about 80% of the nitrogen content of the fetus is found in protein; the remainder is found in urea, ammonia, and free amino acids. Additional nitrogen requirements for urea excretion and for other possible nitrogen excretion products are not known for human fetuses. Protein accretion accounts for the largest fraction, up to 90%, of fetal growth, contributing to the replication (hyperplasia) and growth (hypertrophy) of all cells and all tissues and organs. Fetal protein

Table 1 Growth characteristics and chemical composition at term of selected mammals and a representative human fetus.

	Human	*Monkey*	*Sheep*	*Pig*	*Rabbit*	*Rat*
Gestation (days)	280	163	47	67	30	21.5
Number of fetuses	1	1	1	3–5	4–6	10–12
Growth rate (gm/day/kg)	15	44	60	70	300	350
Fetal weight (gm)	3500	500	4000	100	60	5
Dry weight (gm/% body wt)	1050/30	125/25	760/19	25/25	9/15	0.2/4
Nonfat dry weight						
(gm/% body wt)	490/14	–	640/16	14/14	–	–
Protein (gm/% body wt)	420/12	–	480/12	12/12	7.2/12	0.6/12

From McCance and Widdowson (1985).

synthesis, breakdown, and accretion rates have been measured in animal models, principally the fetal sheep, with isotopic tracers of selected amino acids, especially essential amino acids such as leucine and lysine. The tracers allow calculation of the fractional protein synthetic rate (K_S, or the fraction of body proteins that are synthesized/unit time), which can be compared with fractional growth rate (K_G). K_S and K_G have been calculated and compared in fetal sheep using two tracers, ^{14}C-leucine and ^{14}C-lysine, at different gestational ages (Kennaugh et al., 1987). As shown in **Fig. 4**, Ks and K_G decrease with gestational age, but the decline in K_S is greater. The 3–4 fold greater protein synthetic rate and 50% greater net protein accretion rate in the fetus at mid-gestation compared to near term are proportional to the greater amino acid nitrogen uptake rates, metabolic rates, and glucose uptake and utilization rates at this earlier stage of gestation. Estimates of amino acid requirements in mid-gestation fetal sheep range from 3.5 to 4.5 g/kg/day; when scaled to the ~55% slower human fetal growth rates at 20–28 weeks gestation, these amino acid requirements compare favorably with that of 4 g/kg/day in human fetuses as determined by the factorial method (Ziegler, 2007). The decreasing Ks and Kg with gestation also indicate that nutritional requirements, particularly for protein, decline at the same time. Thus, at 28–34 weeks, protein requirements decrease to 3.0–3.5 g/kg/day, and by term gestation to 1.5–2.0 g/kg day, a rate that is provided by mature maternal milk with full breastfeeding.

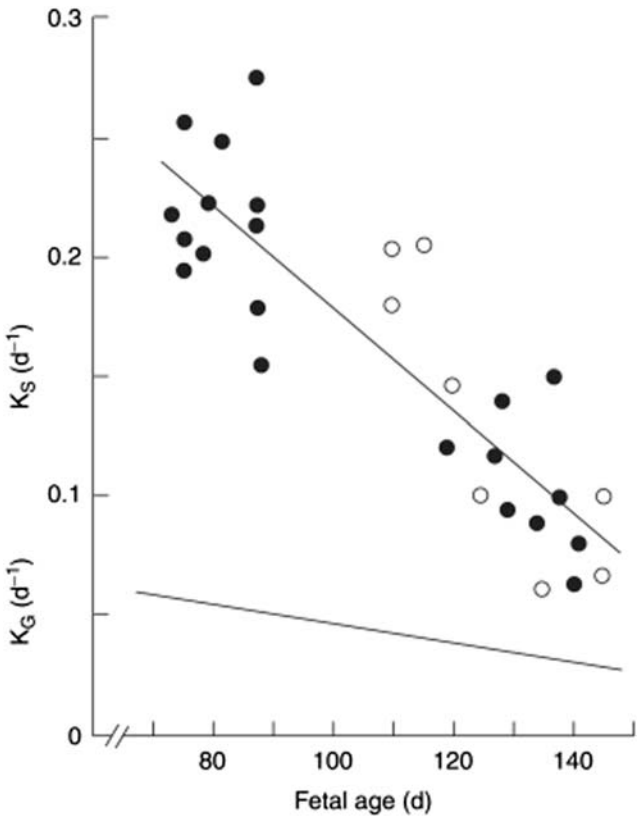


Fig. 4 Fractional rate of protein synthesis (K_s) over gestation in fetal sheep studied with leucine (●) and lysine (o) radioactive tracers compared with the fractional rate of growth (K_G) in the lower portion of the figure (——). Reproduced with permission from Hay and Regnault (2003).

Table 2 Fetal organ weight as percent of body weight[a].

	50% gestation	*67% gestation*	*90% gestation*
Liver	6.5	5.1	3.1
Kidneys	1.6	1.2	0.7
Heart	0.9	0.8	0.8
Brain	3.4	2.9	1.7
Hindquarters	14.5	15.1	22.0

[a]From Bell et al. (1987).

The decline in K_S and K_G with gestational age also indicate that the reduced rate of fetal growth in later gestation is an intrinsic quality of fetal development and not the result of insufficient nutrient supply. Even at term gestation, however, the fetus has higher protein synthetic rates than protein breakdown rates, reflecting an anabolic state with considerable net amino acid uptake and accretion (van den Akker et al., 2009). Mechanisms underlying the decrease in protein synthetic rate over gestation are not well understood, but a partial explanation for the continued anabolism can be derived from the changing proportion of body mass contributed by the major organs (**Table 2**). The mass of bone, skin, gastrointestinal tract, and skeletal muscle, for example, increase markedly during later gestation, but these organs have relatively lower fractional protein synthetic rates. Thus, whole body fractional protein synthetic rate should decrease despite the continued anabolism. It also is clear that a direct, causal relationship between continued anabolism but declining fractional protein synthetic rates cannot be made to anabolic endocrine-paracrine factors acting as principal regulators of fetal protein synthesis rate and thus fetal growth rate, (e.g., insulin, pituitary and placental growth hormone, placental lactogen, IGFs, epidermal growth factors), because the secretion and plasma concentrations of these growth hormones increase with advancing gestation.

Glycogen

Many tissues in the fetus, including brain, liver, lung, heart, and skeletal muscle, produce glycogen over the second half of gestation. Liver glycogen content, which increases with gestation, is the most important store of carbohydrate for systemic glucose needs immediately after birth, because only the liver contains sufficient glucose-6-phosphatase for release of glucose into the circulation. Skeletal muscle glycogen content increases during late gestation and forms a ready source of glucose for glycolysis within the myocytes. Lung glycogen content decreases in late gestation with loss of glycogen-containing alveolar epithelial cells, development of type II pneumocytes, and onset of surfactant production. Cardiac glycogen concentration decreases with gestation as cellular hypertrophy develops, but cardiac glycogen appears essential for postnatal cardiac energy metabolism and function. At term, fetal liver glycogen concentration in most species is about 80–120 mg/gm, at least twice the adult concentration. In the relatively slowly growing human fetus, glycogen synthesis rates are low (about 2 mg/day/gm), representing less than 2% of estimated whole body glucose utilization rates (Hay et al., 2016b).

Macrosomic fetuses of diabetic mothers have very high body and organ contents of glycogen. In IUGR fetuses, placental insufficiency and decreased placental glucose supply to the fetus would tend to decrease fetal organ glycogen content. This could be augmented by acute hypoxic stress, especially right before birth, that would increase catecholamine and glucagon secretion, both of which would increase glycogen breakdown. More recent studies have shown, however, that the tendency to reduced glycogen content in IUGR fetuses is balanced or even exceeded by increased glycogen formation in peripheral (non-hepatic) tissues and organs produced by increased insulin sensitivity and cellular glucose uptake capacity. These increases represent compensatory adaptations in IUGR fetuses to the chronically low glucose concentrations.

Fat

Fetal fat content as a fraction of fetal weight varies several fold among species (**Fig. 5**). The fat content of newborns at term of almost all land mammals is 1–3% and is considerably less than that of the human, 10–20%. Differences in body fat content among species are due primarily to the capacity of the placenta to transfer lipids, principally fatty acids, to the fetus and to the capacity of the fetus to synthesize triglycerides, produce fat, and store fat in adipose tissue cells that human fetuses develop in relatively large abundance by late gestation. Even in the species such as the human that take up fatty acids from the placenta and deposit fat in fetal tissues, the rate of fetal fatty acid oxidation is presumed to be low, because plasma concentrations of fatty acids are low and the carnitine palmitoyl transferase enzyme system is not sufficiently developed to transfer long chain fatty acids to the respiration pathway inside the mitochondria. Fetal fatty acid oxidation also is limited by the ready supply of glucose, lactate, and amino acids that are preferentially metabolized to oxidation in the tricarboxylic acid cycle. In the human fetus, calories produced by the complete oxidation of glucose, lactate, and amino acids can fully meet the amount of energy required for maintenance metabolism. In the human fetus between 26 and 30 weeks gestation, nonfat and fat components contribute equally to the carbon content of the fetal body. After that period, fat accumulation considerably exceeds that of

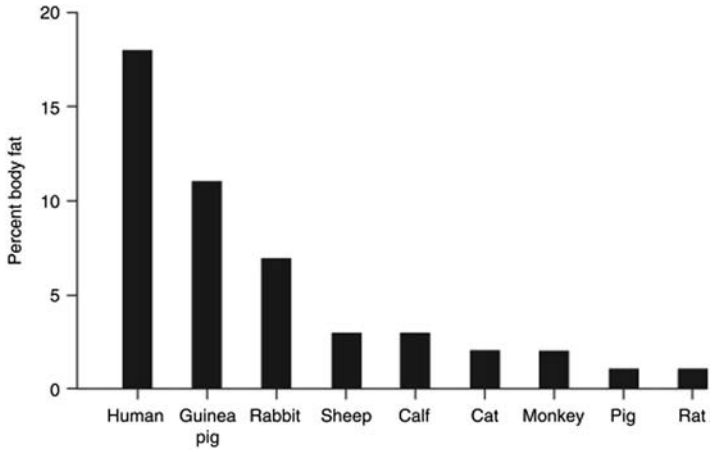


Fig. 5 Fetal fat content at term as a percent of fetal body weight among species. Reproduced with permission from Hay (2003).

the nonfat components. At 36 weeks gestation, 1.9 g of fat accumulates for each gram of nonfat daily weight gain, and by term, the deposition of fat accounts for over 90% of the carbon accumulated by the fetus. The rate of fat accretion is approximately linear between 36 and 40 weeks gestation, and by the end of gestation, fat accretion ranges from 1.6 to 3.4 g/day/kg. By term, the fat content of the human fetus is 15–20% of body weight, ranging from less than 10% in IUGR fetuses to 25% or more in macrosomic infants of diabetic mothers (Hay et al., 2016b).

Caloric accretion in the fetus

Fat has a high energy content, ~9.5 kcal/g, and a very high carbon content, approximately 78%. Thus, differences in fetal fat concentration among species lead to large differences in calculated caloric accretion rates and carbon requirements of the fetal tissues for growth. The caloric concentration of nonfat dry weight is fairly consistent across species and also within species at different developmental stages, indicating that the ratio of protein to non-protein substrates in the tissues is relatively constant. Thus, caloric accretion rate of any fetus can be estimated from the growth curve of that fetus and its change in whole body fat and water concentrations over time.

Data for caloric accretion and caloric distribution in the human fetus are shown in **Table 3** (Sparks et al., 1980). Because growth of fat and non-fat (protein plus other) tissues are metabolically linked through energy supply that is used for protein synthesis and the production of anabolic hormones that promote positive protein, fat, and carbohydrate growth, restriction of nutrient supply is likely to produce growth deficits of all tissues, not just fat, i.e., growth restriction involves limitation of neuronal and myocyte replication and brain and muscle growth as well as fat and glycogen deposition. Indeed, chronic experimental selective caloric (glucose) restriction in the fetal sheep leads to increased protein breakdown as well as to lower rates of fetal growth and lipid content. In contrast, as shown by the growth curves in **Fig. 6** from human infants born preterm at different times over the last third of gestation, there is a bias toward thinner infants with less fat content, indicating that in a species such as the human that does lay down considerable fetal fat during late gestation, differences in the caloric requirements for intrauterine growth later in gestation reflect fat deposition more than the growth of non-fat tissues. IUGR fetuses of all species, however, show reduced growth of skeletal muscle, and preterm infants tend to have less muscle mass with a slight excess in body fat mass, the latter likely the result of excess nutritional energy intakes. Importantly, the reduced skeletal muscle mass in normal preterm infants persists, but the excess fat mass does not (Hamatschek et al., 2020). An important long term adverse consequence of reduced muscle growth might involve diminished whole body insulin action and a tendency to hyperglycemia and type 2 diabetes, as skeletal muscle is the principal insulin-sensitive tissue in the body in older children and adults.

Table 3 Calculation of the caloric distribution in the term human infant.[a]

Wet weight	*Wet weight*	*Fat*	*Nonfat wet weight*	*Nonfat dry weight*
Weight (g)	3450	386	3064	511
Total calories (kcal)	5950	3650	2300	2300
Caloric concentration (kcal/g)	1.73	9.45	0.75	4.5

[a]The values in this table are from Ziegler et al. (1979).

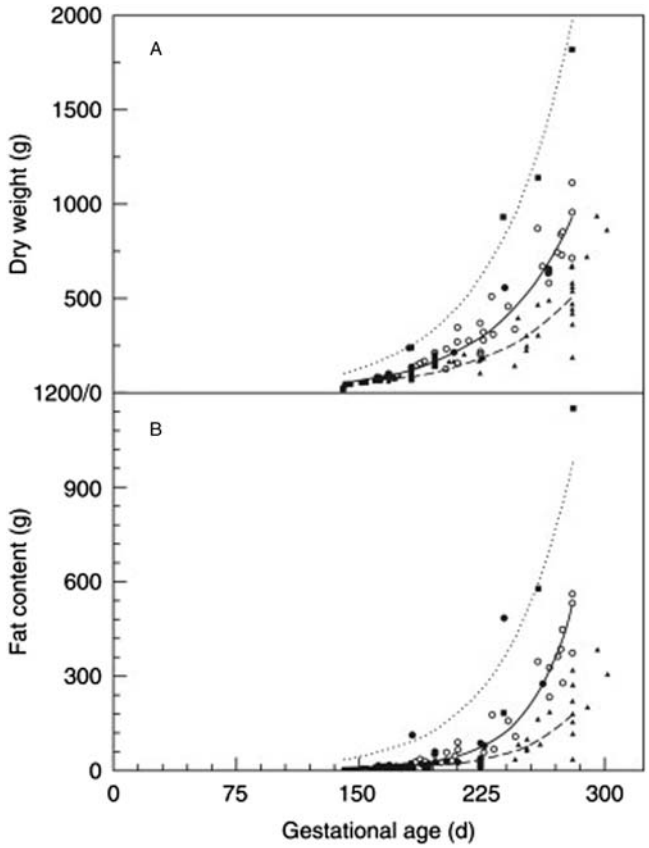


Fig. 6 Dry weight (A) and fat content (B) plotted against gestation age in the same newborn human infants shown in Figure 3 for LGA (■, ····), AGA (○, ———), and SGA (◆,——) infants. Reproduced with permission from Sparks (1984).

Mineral accretion in the fetus

Fetal calcium content is best correlated with fetal body length; this is true for both AGA and SGA infants. Using this index, fetal calcium content increases exponentially with a linear increase in length. Using this estimate, the human fetal rate of calcium accretion is about 85 mg/day/kg. Accretion of other minerals varies more directly with body weight, and according to the distribution of the minerals into extracellular (e.g., sodium) or intracellular (e.g., potassium) spaces.

Regulation of fetal growth

Regulation of fetal growth represents a mix of genetic mechanisms and environmental influences through which the genetic factors are expressed and modulated. The single most important environmental influence that affects fetal growth is the nutrition of the fetus, including oxygen as well as the principal nutrient substrates, amino acids, glucose, and lipids. Oxygen and nutrient substrate supply to the fetus and the resulting increases in fetal tissue and plasma concentrations of anabolic hormones and growth factors are regulated by maternal health, maternal nutrition, uterine growth (including uterine blood flow and endometrial surface area), and placental growth and function.

Genetic factors

Maternal genotype is more important than fetal genotype in the overall genetic regulation of fetal growth. Table 4 presents estimates of the quantitative contribution of fetal and parental factors to fetal growth and birth weight at term. The more modest regulation by the paternal genotype acts through its contribution to trophoblast development and thus placental-to-fetal nutrient transport capacity. More specific gene targeting studies have shown the importance of genomic imprinting on fetal growth. For example, in mice normal fetal and placental growth require that the IGF-II gene be paternal and the IGF-II receptor gene be maternal, and paternal disomy producing IGF-II gene over-expression results in fetal overgrowth while maternal disomy producing IGF-II under-expression results in fetal dwarfism. In humans, isopaternal inheritance of IGF-II alleles is associated with the Beckwith-Wiedemann syndrome that includes hyperinsulinism and fetal macrosomia.

Table 4 Factors determining variance in birth weight.

	Percent of total variance
Fetal	
Genotype	16
Sex	2
	18
Maternal	
Genotype	20
Maternal environment	24
Maternal age	1
Parity	7
	52
Unknown	30

Derived from Penrose (1954), Milner and Gluckman (1996).

Non-genetic maternal factors

There is a high correlation between birth weights of siblings that extends to cousins. The non-genetic, maternal nature of this effect is demonstrated by embryo transfer and cross-breeding experiments. For example, a small-breed embryo transplanted into a large-breed uterus will grow larger than a small-breed embryo remaining in a small-breed uterus (Ounsted and Ounsted, 1973). Furthermore, partial reduction in fetal number in a polytocous species such as the rat produces greater than normal birth weights in the remaining offspring. Conversely, transfer of a large-breed embryo into a small-breed uterus will result in a newborn that is smaller than in its natural large-breed environment. Such evidence demonstrates that fetal growth is normally constrained, and that this constraint comes from the maternal environment. This is a physiological process and includes the maternal-specific capacity of uterine size, placental implantation surface area of the uterus, and uterine circulation, which together support the growth of the placenta and its capacity to transport oxygen and nutrients to the fetus to support fetal growth.

Maternal nutrition

Normal variations in maternal nutrition have relatively little effect on fetal growth, because they do not markedly alter maternal plasma concentrations of nutrient substrates or the rate of uterine blood flow, the principal determinants of nutrient substrate delivery to and transport by the placenta. Human epidemiological data from conditions of prolonged starvation, as well as nutritional deprivation in experimental animals, indicate that even severe limitations in maternal nutrition limit fetal growth by 10%–20%, although muscle mass is preferentially reduced under such maternal dietary restrictions (Gauvin et al., 2020). Restriction of caloric and protein intakes to less than 50% of normal for a considerable portion of gestation are needed before marked reductions in fetal growth are observed; such severe conditions often result in fetal loss before the impact on late gestation fetal growth rate and fetal size at birth are manifested.

Fetal macrosomia is common in pregnancies complicated by maternal diets excessively rich in sugars and fats (Armitage et al., 2005). Fetal macrosomia also is common in pregnancies complicated by gestational diabetes mellitus in which maternal and fetal plasma hyperglycemia and hypertriglyceridemia plus fetal hyperinsulinemia combine to produce excessive fetal adiposity. More recent observations of increasing fetal birth weight and macrosomia in relation to increased maternal obesity, a potentially serious adverse impact of the world wide obesity epidemic, remain mechanistically unexplained, although there is some evidence that obese pregnant women also tend to be hypertriglyceridemic as well as slightly hyperglycemic, both of which would contribute to increased lipid and glucose supply to the fetus, increased fetal insulin secretion, and fetal fat production. The placentas in diabetic and obese pregnant mothers also have greater amounts of lipoprotein lipase, which hydrolyzes maternal plasma triglycerides into free fatty acids that are then transported to the fetus and contribute to fetal fat mass production (Heerwagen et al., 2018). Maternal diets lower in simple sugars, higher in complex carbohydrates, and limited in total lipid can contribute to less fetal fat mass gain and the potential to prevent later life obesity in these offspring (Hernandez et al., 2016).

The placenta

The placenta exerts strong control over fetal growth by providing nutrients directly or in metabolically altered form and amount. The size of the fetus is directly related to the size of the placenta (Molteni et al., 1978). Naturally and experimentally, placental growth precedes fetal growth, and failure of placental growth is directly associated with decreased fetal growth. There is considerable variation in this control, however. For example, experiments in sheep that limited placental growth did not result in proportionately reduced fetal weight, indicating that the capacity of the smaller placenta to transport nutrients to the fetus adaptively increased or that the fetus developed increased capacity to extract nutrients from the placenta and direct those nutrients to growth. More

characteristically, though, limitation in placental function to transfer nutrients to the fetus directly limits fetal growth. In fact, fetal growth restriction is seen as a natural and reproductively successful (though not perfect) compensatory adaptation to nutrient limitation. There is a direct relationship between fetal weight and placental weight in humans, although functional interactions between placenta and fetus also are important to fetal growth and development.

Growth of the placenta and its transport capacity

Placental nutrient transfer capacity increases over gestation by increased placental growth, primarily of membrane surface area and vascular development in the fetal villi, allowing for the increase in nutrient supply required for the growing fetus. Placental size, morphology, vascular surface area, and membrane transporter abundance are regulated by imprinted paternally-derived genes, such as the placental-specific Igf2-H19 gene complex (Reik and Walter, 2001). A larger paternal vs. maternal Igf2 gene allele supply leads to a larger placenta and the potential for a larger fetus. Activity of the imprinted genes also can be affected by epigenetic modification, which allows for considerable environmental influence over gene expression. Thus, DNA methylation can limit placental-specific Igf-2 gene activity, leading to intrauterine growth restriction (IUGR) of the placenta and, in turn, the fetus.

Maternal endocrine influences on fetal growth

Changes in maternal circulating growth hormone and growth hormone-like peptides such as placental lactogen, which increase during pregnancy, have combined effects that induce maternal insulin resistance and lead to higher circulating concentrations of glucose and lipids. These in turn are transported in increased amounts to the fetus where, combined with their stimulatory effects on fetal insulin and IGF-I and II, promote fetal adiposity (or macrosomia, as in the infant of the diabetic mother) and limit fetal protein breakdown, both of which promote fetal growth.

Fetal endocrine and autocrine/paracrine-acting growth factor influences on fetal growth

Pituitary Growth Hormone: Growth hormone, which classically acts as the major regulator of postnatal growth, has no demonstrable influence on fetal growth.

Insulin and IGFs: Both insulin and IGF-1 regulate protein synthesis through well-recognized intermediates in their signal transduction pathways, including the mammalian target of rapamycin (mTOR) and the eukaryotic initiation factors, which are active in all cells throughout their lifespan, even in the fetal period (Shen et al., 2002). mTOR is an evolutionarily-conserved protein that functions as a sensor for growth factors, nutrients, energy, and stress, and coordinates these signals to regulate cell growth and proliferation.

Insulin: Fetal insulin does regulate fetal growth, although the complete absence of insulin does not abolish fetal growth. In sheep, for example, fetal pancreatectomy in late gestation limits fetal growth rate only by 20%–30%, and pancreatic agenesis in humans produces IUGR fetuses that are 30%–50% less than normal weight near term. Insulin infusions into the fetus and excessive fetal insulin secretion enhance fetal glucose utilization and produce increased adiposity. Such hyperinsulinemic conditions also limit protein breakdown, which supports increased protein accretion, but overall there is little impact of excess insulin to promote excess growth of non-fat lean body mass in the fetus. The primary action of fetal insulin is to promote glucose utilization and, in turn, enhance protein accretion by providing more energy substrate to fuel protein synthesis and to substitute glucose carbon for amino acids to fuel oxidative metabolism. For example, reducing fetal glucose supply lowers fetal weight and oxygen consumption to the same extent, indicating that oxidative metabolism of glucose was responsible for the growth. Similarly, removal of insulin from the fetus increases fetal glucose concentration and the transfer of glucose from mother to fetus via the placenta, which reduces fetal glucose oxidative metabolism and results in reduced net fetal carbon accretion and reduced fetal growth.

Insulin also acts in the fetus to increase amino acid uptake by cells in insulin sensitive tissues, promoting direct amino acid synthesis into protein via both the mTOR pathway and the MAP-kinase pathway, and, when energy is deficient, into oxidative metabolism and energy production (Brown et al., 2021). In the fetus and the preterm newborn, the anabolic effects of insulin on protein kinetics depend on the adequacy of circulating amino acids and glucose. Insulin is anabolic and induces a net gain in protein balance in the fetus and preterm infant when euglycemia, maintained by glucose infusion, is complemented by a mixed amino acid infusion producing aminoacidemia (Thureen et al., 2000).

Amino acids, and particularly leucine, also function as direct-acting nutrient signals that activate protein synthesis, which is true in normal and growth restricted fetuses (Brown et al., 2012). Leucine in particular can stimulate mTOR, the key regulator of the protein synthetic pathway, independently of insulin or IGF-1. However, amino acids alone do not activate the phosphorylation of any of the proteins in the protein synthetic pathway. Thus, insulin appears to be a more effective regulator of the mTOR signaling pathway in the fetus, at least when administered acutely (Brown et al., 2009). These observations indicate that amino acids can independently up-regulate particular signal transduction proteins during late gestation fetal growth and emphasize, as does the data showing insulin activation of the MAP-kinase pathway, that nutrient-hormonal interaction is central for regulation of growth.

Insulin-like Growth Factors, IGF-I and IGF-II: Mice lacking the IGF-I gene have markedly reduced rates of fetal growth in late gestation (Woods et al., 1996). IGF-II knockouts also have delayed fetal growth that is more pronounced in early to mid-gestation. IGF-I receptor knockout mice are more growth restricted than either IGF-I or IGF-II knockouts alone. These IGF-I receptor knockouts are growth restricted to the same extent as mice in which both IGF-I and IGF-II genes are deleted, confirming that receptor activation is the principal growth-regulating step in IGF-I and IGF-II action. IGFs also regulate fetal growth by regulating placental growth. IGF-II gene knockout mice have smaller placentas and, in turn, lower IGF-I and IGF-II binding proteins. IGF binding proteins modulate the effects of IGF-I and IGF-II on fetal growth. (Abuzzahab et al., 2003).

IGF-I also has been shown to have anabolic effects on protein metabolism in the fetus. A series of studies in late gestation fetal sheep have shown that fetal IGF-1 infusion suppresses whole fetal protein breakdown and amino acid oxidation, and increases organ growth without promoting amino acid transfer from the placenta, together resulting in net fetal protein accretion (Stremming et al., 2021). The effect of IGF-1 on the suppression of proteolysis was dose-dependent, and when combined with an insulin infusion, both hormones resulted in increased fetal protein synthesis.

Interpretation of growth curves

Cross-sectional growth curves have been developed from anthropometric measurements in populations of infants born at different gestational ages. Such curves have been used to estimate whether growth of an individual fetus or preterm newborn is within or outside of the normal range of fetal growth, which has been most commonly but arbitrarily defined as between the 10th and 90th percentiles for the population of fetuses included in the growth curve analysis. Fetuses and newborn infants who are within the 10th and 90th percentiles for weight vs. gestational age are considered appropriate for gestational age (AGA), those who are less than the 10th percentile are considered small for gestational age (SGA), and those who are greater than the 90th percentile are considered large for gestational age (LGA). Importantly, fetuses and newborn infants who are <10th percentile or >90th percentile for anthropometric measurements can be perfectly normal, as their size is determined by parental stature and genetic background; they should not be assumed to have had fetal or intrauterine growth restriction or overgrowth simply based on their size.

Fetal and preterm neonatal growth curves represent the late second and the third trimester in humans. Each curve is based on defined, usually local populations with variable composition of maternal age, parity, socio-economic status, race, ethnic background, body size, degree of obesity or thinness, health, pregnancy-related problems, and nutrition, as well as the number of fetuses per mother, the number of infants included in the study, and how and how well measurements of body size and gestational age were made, all of which contribute to the wide range of anthropometric measurements at any gestation age. Variability also occurs because estimates of gestational age of newborn infants often are imprecize due to variations in maternal post-implantation bleeding and irregular menses, onset and appearance of physical features of maturation in the infant, and inter-observer assessments of an infant's developmental stage. The best of the many growth curve studies have excluded obviously abnormal infants, and have involved defined, relatively homogeneous populations. Even with such precautions, there has been no complete assurance that all of the infants included in the cross sectional measurements who were born preterm had grown normally in utero; therefore, it is not known for certain whether their anthropometric measurements at birth represent those of normally growing fetuses. It is likely, however, that inclusion of a relatively small number of slowly growing fetuses is balanced by an approximately equal number of more rapidly growing fetuses. Furthermore, anthropometric measurements have their own inherent range of accuracy, and cross-sectional, static size-for-gestational-age groupings of neonates may not accurately reflect the dynamics of fetal growth (Demerath and Fields, 2014). In addition, means of fetal growth rates from different populations do not necessarily reflect growth of a given fetus or the changes in body composition with growth that occur with advancing gestational age.

A large variety of fetal growth curves developed from cross sectional measurement of weight at birth are shown in **Fig. 7**, representing infants born in North America and Europe and from low and high socioeconomic backgrounds, multiple races and ethnic origins, "normally" short and tall mothers as well as the majority of normal sized mothers, and low (sea level) and moderately high (~1 mile, or ~1500 m) altitudes. The average fractional growth rate or growth velocity among this diverse population is 17 g/day/kg, ranging from 15 to 20 g/day/kg for average sized infants, slower for smaller infants and faster for larger infants. The Fenton growth curves are the most robust, with ~4 million infant measurements <37 weeks gestational age from several international populations screened for normal pregnancies and excluding abnormal infants (Fenton and Kim, 2013). The Olsen (~250,000 singleton "normal" preterm infants, 57.2% male, who survived to discharge), NICHD Fetal Growth Studies (~1700 infants, serial ultrasound and birth weight combinations, including data for race and ethnicity), and Mihatsch et al. (~1.8 million "normal" preterm infants in Germany) growth curves are similar, representing data from North America and Europe (Olsen et al., 2010; Buck Louis et al., 2015; Mihatsch et al., 2004). Given the commonality of growth rates from these diverse studies representing almost 8 million infants, it is unlikely that concerns regarding inclusion of abnormally growing infants are significant.

The Intergrowth 21st curves are projections into the fetal period from serially measured growth, primarily after birth through 2 years of age, of "healthy" infants of "healthy" mothers from pooled international populations that included smaller mothers and mothers from different racial and ethnic backgrounds (Kiserud et al., 2017). In marked contrast to the approximately 8 million infants in the fetal curves from Fenton, Olsen, NICHD, and Mihatsch, the actual fetal data used for the Intergrowth 21st curves

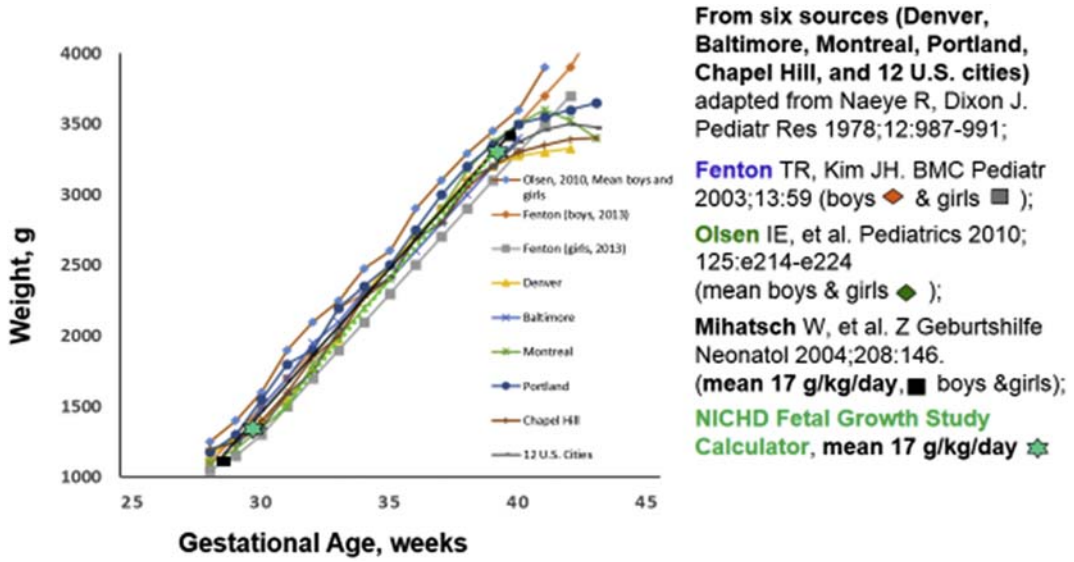


Fig. 7 Regardless of the growth curve, birth weight by gestational age is relatively constant from ~26 to 38 weeks gestation at an *average* growth rate of ~15–20 g/kg/day (mean of 17 g/kg/day). Adapted from Naeye and Dixon (1978), Fenton and Kim (2003), Olsen et al. (2010), Mihatsch et al. (2004), Buck Louis et al. (2015).

included only 201 infants <37 weeks, only 12 infants (9 males and 3 female) <32 weeks, and no female infants <30 weeks; the fetal data are not normally distributed, as the curves are projected from postnatal growth. Thus, the Intergrowth 21st fetal growth curves do not represent actual fetal growth as well as the very large and robust cross sectional growth curves noted in **Fig. 7**.

The two recent growth curves that have been derived in part from serial fetal ultrasound measurements during gestation, the NICHD Fetal Growth study and the Intergrowth-21st Project, provide continuous rather than cross-sectional growth patterns when combined with neonatal birth weight data at the completion of a pregnancy (Buck Louis et al., 2015; Kiserud et al., 2017). Curiously, there is an apparent conflict between such ultrasound-derived fetal growth curves that show a rate of fetal growth in late gestation that is continuous with postnatal growth rather than the actual slowing of fetal growth rate in late gestation (after 36 weeks in human gestation) that has been shown with actual measurements in humans and animal models (see **Fig. 2**).

There are, however, several advantages of serial ultrasound measurement of fetal growth. Serial ultrasounds theoretically permit examination of the change in fetal growth with time in normal pregnancies rather than establishing growth curves based only on birth weights of pregnancies that are by definition pathological because they ended prematurely. Ultrasound-derived fetal growth curves also may allow the early detection of fetuses who are not SGA, but who are experiencing growth faltering or IUGR, or those who are growing faster than normal in terms of circumferences, whether LGA or not, such as in infants of diabetic and/or obese mothers. Serial ultrasound measurements of fetal growth also can help determine how environmental factors can inhibit (for example, maternal under nutrition globally, or hypoglycemia specifically) or enhance (for example, maternal over nutrition globally or hyperglycemia and hypertriglyceridemia specifically) the rates of fetal growth.

There are several nutritional implications of fetal and preterm neonatal growth curves. Growth should proceed symmetrically following normal fetal growth for weight, length, head circumference, and body lean and fat mass components. Failure to provide sufficient protein and energy nutrition, from either maternal or neonatal under nutrition, leads to growth faltering that universally has been shown to produce later life shorter stature and abnormal neurodevelopment. Excess maternal protein intake does not increase fetal growth and actually may worsen growth restriction and lead to fetal death, especially when the fetus already is growth restricted (Rush et al., 1980). Excess energy, even when length and head circumference of the fetus or preterm infant are growing appropriately, leads to excess fat mass production. When this occurs in fetal life, it appears to predispose to later life obesity, whereas modest amounts of excess fat in the preterm infant do not appear to last. Mechanisms for this discrepancy are uncertain, but might relate to the development of excess adipocytes in the fetus, perhaps from mesenchymal stem cells that populate and then proliferate in peripheral adipose tissue, that does not occur after birth (Baker et al., 2017). When fetal IUGR is the result of placental insufficiency, the fetus also is deprived of oxygen and is relatively hypoxic, a condition that also might lead to later life excess adiposity. This would not likely occur in preterm infants after birth when oxygenation is normal.

Extremes of growth and development: intrauterine growth restriction and macrosomia

Extremes of fetal growth and body composition are important, as both IUGR and macrosomia may predispose the fetus to the same later life disorders of obesity, insulin resistance, and type 2 diabetes, despite their very different patterns of intrauterine growth (Godfrey et al., 2011).

Intrauterine growth restriction (IUGR)

In developed countries, three to seven percent of newborns are classified as IUGR. The percentages vary depending on arbitrary definitions, such as <10th percentile or >2 standard deviations below the mean for a population based on weight at birth at a given gestational age. Any definition of IUGR likely is an underestimate, however, as it is commonly based on infants who are SGA (e.g., <10th percentile weight for gestational age), whereas fetal growth restriction can occur within normal percentiles of weight, length, and circumferences. For example, an infant at the 50th percentile at 20 weeks gestation but falls to the 25th percentile by 30 weeks has experienced growth faltering or IUGR, but still is within normal weight for gestational age ranges. Attempts also have been made to improve the diagnosis of fetal growth faltering or IUGR by the use of standards that incorporate maternal and paternal size, which aim to distinguish fetuses that grow more slowly than normal and are smaller at any gestation because the parents are small from those with growth faltering resulting from a disease process (Robinson et al., 2000).

Most cases of fetal growth faltering occur when the fetus experiences deficits in oxygen and nutrient supplies from a less than optimally functioning placenta; these cases are identified as Placental Insufficiency IUGR (PI-IUGR). Most of the oxygen and nutrient deficits can be attributed initially to maldevelopment of the placenta, which not only is reduced in size but also has structural abnormalities in its vasculature and reductions in transport capacity. These changes in the placenta, with the additional factor of chronic hypoxia, can lead to circulatory changes in many fetal vascular beds and altered metabolism that produce a wide range in clinical severity of PI-IUGR. Animal based studies and human observational studies indicate, however, that the placenta is not always dysfunctional in IUGR pregnancies. Instead, adverse changes in placental transport observed in the fetal growth restriction placenta also represent adaptations to reduced fetal anabolic metabolism and growth in response to an inability of the mother to allocate resources to the fetus.

Recent studies in fetal sheep with placental insufficiency have clarified the mechanisms by which fetal hypoxemia leads to fetal growth faltering. Fetal hypoxemia increases blood flow to the fetal adrenal glands and increases adrenal norephinephrine output. The increased norephinephrine inhibits pancreatic insulin secretion, which in turn decreases fetal growth rate (Leos et al., 2010). Fetal oxygenation therefore plays a central role in the pathogenesis of all placentally mediated fetal IUGR. The reduced oxygen delivery in IUGR fetuses indicates impaired placental oxygen transport, whereas reduced fetal oxygen consumption presumably reflects metabolic adaptation to diminished substrate delivery that results in slower fetal growth.

Postnatal complications of the unique IUGR fetal metabolism are common and can be quantitatively significant. IUGR fetuses have up-regulated peripheral glucose and insulin sensitivity in response to low glucose and insulin concentrations, with maintained or even increased glucose transporter concentrations in insulin independent (brain—Glut 1) and insulin sensitive (skeletal muscle, heart—Glut 4) tissues. These changes maintain glucose utilization rate (GUR) and insulin sensitivity, i.e., GUR/kg is normal at less than normal glucose and insulin concentrations. Also, although glucose stimulated insulin secretion (GSIS) is reduced in IUGR fetuses, they also have increased in vitro fractional islet insulin secretion. In vivo, IUGR fetuses are relatively hypoxic and have increased plasma catecholamine concentrations. Fetal insulin concentrations increase with an adrenergic receptor blockade, proof that they suppress insulin secretion in IUGR fetuses. Increased peripheral glucose uptake capacity from increased or at least maintained glucose transporters persists well after birth. Increased fractional insulin secretion also persists, as does increased insulin secretion after reduction of catecholamine suppression of insulin secretion. When added to the greater brain to body (or liver) ratio and thus greater body weight-specific GUR in asymmetrically grown IUGR infants, persistent hypoglycemia clearly is a natural complication of normal adaptive mechanisms in the fetus that maintain glucose utilization at the expense of growth, thereby ensuring survival. After birth, persistence of this condition can lead to maladaptive hypoglycemia (Hay et al., 2016a).

Generally, chronically IUGR fetuses have less subcutaneous fat and skeletal muscle, with low weight/length ratios or low Ponderal Index values (Ponderal Index = weight (grams)/[length (cm)]^3 that are the result of reduced skeletal muscle mass as much as or more than reduced fat mass. IUGR fetuses, in fact, uniquely have reduced muscle mass from reduced insulin and amino acid activation of net protein synthesis and myocyte proliferation (Brown and Hay, 2016). IUGR imposes increased risks of specific types of fetal and neonatal morbidity and mortality (**Table 5**). The earlier fetal growth restriction develops the longer it lasts, and the more severe it becomes the more likely the infant will have persistent short stature after birth (Villar et al., 2021), even into adulthood. Muscle mass also is permanently limited, as myocyte proliferation stops with birth; this is true for both skeletal and cardiac myocytes, the former contributing to risk for diabetes and the latter to heart failure (Sayer et al., 2008; Louey et al., 2007). Brain growth also is diminished, including permanent deficits in neuronal number, dendritic formation and arborization, and synapse development, leading to permanent neurodevelopmental, cognitive, and behavioral deficits later in life (Smart, 1986).

Possible adult disorders resulting from intrauterine growth restriction

Interest in IUGR has been enhanced recently by retrospective epidemiological, clinical follow-up, and animal studies that indicate long-term consequences in adult life of chronically IUGR offspring, including higher incidences of obesity, insulin resistance, impaired glucose tolerance, enhanced hepatic glucose production, pancreatic insulin secretion deficiency, type 2 diabetes mellitus, hypertriglyceridemia, and cardiovascular disease, particularly hypertension (Bloomfield et al., 2006). These conditions, often called the Metabolic Syndrome, may represent an example of "programming," in which an insult, when applied at a critical or sensitive stage in development, produces lasting, even lifelong, effects on the structure or function

Table 5 Risks of specific types of fetal and neonatal morbidity and mortality in IUGR infants.

Problem	*Pathogenesis/pathophysiology*
Intrauterine death	Chronic hypoxia
	Placental insufficiency
	Growth failure
	Malformation
	Infection
	Infarction/abruption
	Preeclampsia
Asphyxia	Acute hypoxia/abruption
	Chronic hypoxia
	Placental insufficiency/preeclampsia
	Acidosis
	Glycogen depletion
Meconium aspiration	Hypoxia
Hypothermia	Cold stress
	Hypoxia
	Hypoglycemia
	Decreased fat stores
	Decreased subcutaneous insulation
	Increased surface area
	Catecholamine depletion
Persistent pulmonary hypertension	Chronic hypoxia
Hypoglycemia	Decreased hepatic/muscle glycogen
	Decreased alternative energy sources
	Heat loss
	Hypoxia
	Decreased gluconeogenesis
	Decreased counterregulatory hormones
	Increased insulin sensitivity
Hyperglycemia	Low insulin secretion rate
	Excessive glucose delivery
	Increased catecholamine and
	Glucagon effects
Polycythemia/hyperviscosity	Chronic hypoxia
	Maternal–fetal transfusion
	Increased erythropoiesis
Gastrointestinal perforation	Focal ischemia
	Hypoperistalsis
Acute renal failure	Hypoxia/ischemia
Immunodeficiency	Malnutrition
	Congenital infection

of the organism. Mechanisms responsible for these later-life morbidities are not fully established. There is relatively consistent evidence in human IUGR infants and those produced experimentally in animal models of diminished pancreatic growth and development, which might present later in life as pancreatic insufficiency when the adult starts and then continues eating a diet rich in simple carbohydrates and lipids. IUGR fetuses also tend to have increased peripheral insulin and glucose sensitivity, which would augment fat production and obesity with such a diet (Thorn et al., 2011). The fetal liver in response to lower plasma glucose concentrations and perhaps hypoxia develops net glucose production from hepatic insulin resistance, the opposite of the increased peripheral insulin sensitivity, with increased development and activation of the key regulatory gluconeogenic enzymes, particularly PEPCK (Wesolowski and Hay, 2016). In part, these changes in IUGR fetal metabolism represent a compensation to maintain glucose supply to the brain and heart as glucose concentrations decline. Over time, however, these changes in fetal metabolism are overcome by peripheral insulin resistance, presumed to be the result of developing obesity with reduction in the insulin signal transduction pathways in both adipocytes and myocytes that result from specific products of intracellular fat metabolism, leading to glucose intolerance. When coupled with pancreatic insufficiency, this would lead to type 2 diabetes. A common theme among these observations is that excessive weight gain of adipose tissue starting at any weight percentile and continuing or even starting after birth is the strongest predictor of the Metabolic Syndrome disorders. The greatest risk for obesity is excess fat mass gain after 6 months to one year after birth. Hypertension in adulthood in individuals who experienced IUGR, also part of the Metabolic Syndrome, may be the result of restricted renal and adrenal development.

Macrosomia

At the other end of the birth weight spectrum are macrosomic, large for gestational age (LGA) infants. These infants were exposed to excess nutrient supply in utero, principally of carbohydrates and lipids. Macrosomic newborns have increased specific morbidities primarily associated with metabolic complications of maternal diabetes mellitus during pregnancy and associated birth complications and birth injuries as a result of excessive fetal size.

Macrosomia is defined in a newborn as a birth weight more than two standard deviations above the mean percentile for gestational age, >90th percentile weight for gestational age, or a birth weight greater than 4000 g at term. Neonatal macrosomia has a strong ethnic predisposition affecting up to 50% of Latino and Native American pregnant women versus 19% of African–American pregnant women. Macrosomia is characteristic of infants of diabetic mothers (IDMs) who were hyperglycemic during pregnancy. The diabetes can be long standing, but the most common group producing macrosomic infants are women with gestational diabetes mellitus (GDM), most of whom also are obese, and obesity itself predisposes to fetal macrosomia (Heerwagen et al., 2010). The percentage of pregnant women who have some form of GDM has been increasing worldwide and now is well above the historical range of 3–5 % of all pregnancies, perhaps as high as 20–50%. The risk of macrosomia is not consistent across all classes of diabetes; it primarily reflects the degree and duration of maternal hyperglycemia and hypertriglyceridemia and particularly high spikes of these conditions following meals that are more common in gestational diabetes. Maternal hyperglycemia results in fetal hyperglycemia and hyperinsulinemia; maternal and fetal hypertriglyceridemia contribute to the effect of the excess glucose and insulin to produce excess fat deposition in the fetus.

Development of type II diabetes in later life in macrosomic offspring

IDMs, particularly those with macrosomia, have increased risk of developing Type II diabetes earlier in life. Mechanisms responsible for this sequence of events include insulin resistance and insufficient insulin secretion (β-cell dysfunction) in response to hyperglycemia. Typically, glucose intolerance from obesity and increased insulin resistance progress to fasting hyperglycemia and the inability of pancreatic β-cells to compensate by increasing their rate of insulin secretion. This form of β-cell failure appears to be reversible over short periods by improved glycemic control, but long-term exposure to hyperglycemia can lead to β-cell exhaustion and specific inhibition of insulin secretion. The insulin resistance also extends to the liver where glucose production increases. This triad of insulin resistance, reduced β-cell insulin secretion, and increased hepatic glucose production produces Type II diabetes.

Conclusion/summary/outlook

The fetal period of growth and development encompasses the greatest changes in growth rate and body proportions and composition during the life of an organism. Total body cell number and the cellular structure of all organs takes place during this period. Such growth and development require adequate supplies of oxygen, amino acids (both essential and non-essential), carbohydrates (principally glucose), lipids (principally essential fatty acids), water, and minerals. Such supplies are regulated by the size and function of the placenta. Growth of the placenta is dependent on paternal genes and the size of the uterus, which is directly related to maternal height. Restriction of fetal growth (IUGR, or a slower than normal rate of fetal growth) most commonly occurs due to a smaller, dysfunctional placenta, and in under developed areas, from maternal under nutrition. Maternal over nutrition, principally of sugar and fat, leads to fetal macrosomia (excess body fat), which is the most common outcome in pregnancies complicated by maternal diabetes and obesity. Both fetal growth restriction and fetal macrosomia are associated with later life metabolic disorders, including obesity, insulin resistance, and diabetes.

References

Abuzzahab, M.J., Schneider, A., Goddard, A., et al., 2003. IGF-I receptor mutations resulting in intrauterine and postnatal growth retardation. N. Engl. J. Med. 349, 2211–2222.
Armitage, J.A., Taylor, P.D., Poston, L., 2005. Experimental models of developmental programming: consequences of exposure to an energy rich diet during development. J. Physiol. 565, 3–8.
Baker II, P.R., Patinkin, Z., Shapiro, A.L., et al., 2017. Maternal obesity and increased neonatal adiposity correspond with altered infant mesenchymal stem cell metabolism. JCI Insight 2, e94200.
Bell, A.W., Battaglia, F.C., Meschia, G., 1987. Relation between metabolic rate and body size in the ovine fetus. J Nutr 1181–1186.
Bloomfield, F.H., Oliver, M.H., Harding, J.E., 2006. The late effects of fetal growth patterns. Arch. Dis. Child. Fetal Neonatal Ed. 91, F299–F304.
Brown, L.D., Hay Jr., W.W., 2016. Impact of placental insufficiency on fetal skeletal muscle growth. Mol. Cell. Endocrinol. 435, 69–77.
Brown, L.D., Rozance, P.J., Barry, J.S., Friedman, J.E., Hay Jr., W.W., 2009. Insulin is required for amino acid stimulation of dual pathways for translational control in skeletal muscle in the late-gestation ovine fetus. Am. J. Physiol. Endocrinol. Metab. 296, 56–63.
Brown, L.D., Rozance, P.J., Thorn, S.R., Friedman, J.E., Hay Jr., W.W., 2012. Acute supplementation of amino acids increases net protein accretion in IUGR fetal sheep. Am. J. Physiol. Endocrinol. Metab. 303, 352–364.
Brown, L.D., Lewis, R.M., Hay Jr., W.W., Regnault, T.R.H., 2021. Placental transfer and fetal requirements of amino acids. In: Polin, R.A., Abman, S.H., Rowitch, D.H., Benitz, W.E. (Eds.), Fetal and Neonatal Physiology, sixth ed. Elsevier, Philadelphia (in press).
Buck Louis, G.M., Grewal, J., Albert, P.S., et al., 2015. Racial/ethnic standards for fetal growth: the NICHD Fetal Growth Studies. Am. J. Obstet. Gynecol. 213, 449.e1–449.e41.
Creasy, R., Resnik, R., 1999. Intrauterine growth restriction. In: Creasy, R., Resnik, R. (Eds.), Maternal-Fetal Medicine, fourth ed. W. B. Saunders Co., Philadelphia, pp. 569–584.
Demerath, E.W., Fields, D.A., 2014. Body composition assessment in the infant. Am. J. Hum. Biol. 26, 291–304.

Fenton, T.R., Kim, J.H., 2013. A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. BMC Pediatr. 13, 59.
Gauvin, M.C., Pillai, S.M., Reed, S.A., et al., 2020. Poor maternal nutrition during gestation in sheep alters prenatal muscle growth and development in offspring. J. Anim. Sci. 98 (1), skz388.
Gluckman, P.D., Hanson, M.A., 2004. Maternal constraint of fetal growth and its consequences. Semin. Fetal Neonatal Med. 9, 419–425.
Godfrey, K.M., Inskip, H.M., Hanson, M.A., 2011. The long-term effects of prenatal development on growth and metabolism. Semin. Reprod. Med. 29, 257–265.
Hamatschek, C., Yousuf, E.I., Möllers, L.S., et al., 2020. Fat and fat-free mass of preterm and term infants from birth to six months: a review of current evidence. Nutrients 12, 288.
Hay Jr., W.W., Regnault, T.R.H., 2003. Fetal requirements and placental transfer of nitrogenous compounds. In: Polin, R.A., Fox, W.W., Abman, S.H. (Eds.), Fetal and Neonatal Physiology, third ed. W. B. Saunders Co., Philadelphia, pp. 509–527.
Hay Jr., W.W., Brown, L.D., Rozance, P.J., Wesolowski, S.R., Limesand, S.W., 2016a. Challenges in nourishing the IUGR fetus. Acta Paediatr. 105, 881–889.
Hay Jr., W.W., Brown, L.D., Thorn, S., Rozance, P.J., 2016b. Nutrition and development of the fetus: carbohydrate and lipid metabolism (Chapter 27). In: Duggan, C.P., Watkins, J.B., Koletzko, B., Walker, W.A. (Eds.), Nutrition in Pediatrics (Basic Science and Clinical Applications), fifth ed. People's Medical Publishing House-USA, Shelton, CT, USA, pp. 445–463.
Hay Jr., W.W., 2003. Nutrition and development of the fetus: carbohydrate and lipid metabolism. In: Walker, W.A., Watkins, J.B., Duggan, C. (Eds.), Nutrition in Pediatrics, third ed. B. C. Decker, Inc., Hamilton, pp. 449–470.
Heerwagen, M.J., Miller, M.R., Barbour, L.A., Friedman, J.E., 2010. Maternal obesity and fetal metabolic programming: a fertile epigenetic soil. Am. J. Physiol. Regul. Integr. Comp. Physiol. 299, R711–R722.
Heerwagen, M.J.R., Gumina, D.L., Hernandez, T.L., et al., 2018. Placental lipoprotein lipase activity is positively associated with newborn adiposity. Placenta 64, 53–60.
Hernandez, T.L., Van Pelt, R.E., Anderson, M.A., et al., 2016. Women with gestational diabetes mellitus randomized to a higher-complex carbohydrate/low-fat diet manifest lower adipose tissue insulin resistance, inflammation, glucose, and free fatty acids: a pilot study. Diabetes Care 39, 39–42.
Kennaugh, J.M., Bell, A.W., Teng, C., Meschia, G., Battaglia, F.C., 1987. Ontogenetic changes in the rates of protein synthesis and leucine oxidation during fetal life. Pediatr. Res. 22, 688–692.
Kiserud, T., Piaggio, G., Carroli, G., et al., 2017. The world health organization fetal growth charts: a multinational longitudinal study of ultrasound biometric measurements and estimated fetal weight. PLoS Med. 14 (1), e1002220. Erratum in: PLoS Med. 2017 Mar 24;14 (3):e1002284. PMID: 28118360.
Leos, R.A., Anderson, M.J., Chen, X., et al., 2010. Chronic exposure to elevated norepinephrine suppresses insulin secretion in fetal sheep with placental insufficiency and intrauterine growth restriction. Am. J. Physiol. Endocrinol. Metab. 298, E770–E778.
Louey, S., Jonker, S.S., Giraud, G.D., Thornburg, K.L., 2007. Placental insufficiency decreases cell cycle activity and terminal maturation in fetal sheep cardiomyocytes. J. Physiol. 580, 639–648.
McCance, Widdowson, 1985. In: Falkner, F., Tanner, J.M. (Eds.), Human Growth, 2nd ed., 1. Plenum Press, New York, p. 139.
Mckeown, T., Record, R.G., 1952. Observations on foetal growth in multiple pregancy in man. J. Endocrinol. 9, 386–401.
Mihatsch, W., Polandt, F., Koetting, K., Voigt, M., 2004. New and improved population-based German reference data for preterm infants growth. Z. Geburtshilfe Neonatol. 208, 146.
Milner, R.D.G., Gluckman, P.D., 1996. Regulation of intrauterine growth. In: Gluckman, P.D., Heymann, M.A. (Eds.), Pediatrics and Perinatology, 2nd ed. Arnold, London, p. 285.
Molteni, R.A., Stys, S.J., Battaglia, F.C., 1978. Relationship of fetal and placental weight in human beings: fetal/placental weight ratios at various gestational ages and birth weight distributions. J. Reprod. Med. 21, 327–334.
Naeye, R.L., Dixon, J.B., 1978. Distortions in fetal growth standards. Pediatr. Res. 12, 987–991.
Olsen, I.E., Groveman, S.A., Lawson, M.L., Clark, R.H., Zemel, B.S., 2010. New intrauterine growth curves based on United States data. Pediatrics 125, 214–224.
Ounsted, M., Ounsted, C., 1973. On Fetal Growth Rate. Clinics in Development Medicine No. 46. J. B. Lippincott Co, Philadelphia.
Penrose, L.S., 1954. Proc 9th Int Congr Genetics Part 1, 520.
Reik, W., Walter, J., 2001. Genomic imprinting: parental influence on the genome. Nat. Rev. Genet. 2, 21–32.
Robinson, J.S., Moore, V.M., Owens, J.A., McMillen, I.C., 2000. Origins of fetal growth restriction. Eur. J. Obstet. Gynecol. Reprod. Biol. 92, 13–19.
Rochow, N., Al Samnan, M., So, H.Y., et al., 2018. Maternal body height is a stronger predictor of birth weight than ethnicity: analysis of birth weight percentile charts. J. Perinat. Med. 47, 22–29.
Rush, D., Stein, Z., Susser, M., 1980. A randomized controlled trial of prenatal nutritional supplementation in New York City. Pediatrics 65, 683–697.
Sayer, A.A., Syddall, H., Martin, H., et al., 2008. The developmental origins of sarcopenia. J. Nutr. Health Aging 12, 427–432.
Shen, W., Mallon, D., Boyle, D.W., Liechty, E.A., 2002. IGF-I and insulin regulate eIF4F formation by different mechanisms in muscle and liver in the ovine fetus. Am. J. Physiol. Endocrinol. Metab. 283, 593–603.
Smart, J., 1986. Undernutrition, learning and memory: review of experimental studies. In: Taylor, T.G., Jenkins, N.K. (Eds.), Proceedings of XII International Congress of Nutrition. John Libbey, London.
Sparks, J.W., Girard, J.R., Battaglia, F.C., 1980. An estimate of the caloric requirements of the human fetus. Biol. Neonate 38, 113–119.
Sparks, J.W., 1984. Human intrauterine growth and nutrient accretion. Semin. Perinatol. 8, 74–93.
Stremming, J., Heard, S., White, A., et al., 2021. IGF-1 infusion to fetal sheep increases organ growth but not by stimulating nutrient transfer to the fetus. Am. J. Physiol. Endocrinol. Metab. https://doi.org/10.1152/ajpendo.00453.2020. Epub ahead of print. PMID: 33427051.
Thorn, S.R., Rozance, P.J., Brown, L.D., Hay Jr., W.W., 2011. The Intrauterine Growth Restriction (IUGR) phenotype: fetal adaptations and potential implications for later life insulin resistance and diabetes. Semin. Reprod. Med. 29, 225–236.
Thureen, P.J., Scheer, B., Anderson, S.M., et al., 2000. Effect of hyperinsulinemia on amino acid utilization in the ovine fetus. Am. J. Physiol. Endocrinol. Metab. 279, 1294–1304.
van den Akker, C.H., Schierbeek, H., Dorst, K.Y., et al., 2009. Human fetal amino acid metabolism at term gestation. Am. J. Clin. Nutr. 89, 153–160.
Villar, J., Restrepo-Méndez, M.C., McGready, R., et al., 2021. Association between preterm-birth phenotypes and differential morbidity, growth, and neurodevelopment at age 2 years: results from the INTERBIO-21st Newborn Study. JAMA Pediatr. https://doi.org/10.1001/jamapediatrics.2020.6087.
Wesolowski, S.R., Hay Jr., W.W., 2016. Role of placental insufficiency and intrauterine growth restriction on the activation of fetal hepatic glucose production. Mol. Cell. Endocrinol. 435, 61–68.
Woods, K.A., Camacho-Hubner, C., Savage, M.O., Clark, A.J., 1996. Intrauterine growth retardation and postnatal growth failure associated with deletion of the insulin-like growth factor I gene. N. Engl. J. Med. 335, 1363–1367.
Ziegler, E.E., O'Donnell, A.M., Nelson, S.E., Fomon, S.J., 1979. Body composition of the reference fetus. Growth 40, 329–341.
Ziegler, E.E., 2007. Protein requirements of very low birth weight infants. J. Pediatr. Gastroenterol. Nutr. 45, 170–174.

Further reading

Anderson, S., Hay Jr., W.W., 2005. Intrauterine growth restriction and the small-for-gestational-age infant. In: MacDonald, M.G., Seshia, M.M.K., Mullett, M.D. (Eds.), Neonatology: Pathophysiology and Management of the Newborn, sixth ed. Lippincott-Williams & Wilkins, Philadelphia.
Battaglia, F.C., Meschia, G., 1986. An Introduction to Fetal Physiology. Academic Press, Inc, Orlando.
Buckley, A.J., Jaquiery, A.L., Harding, J.E., 2005. Nutritional programming of adult disease. Cell Tissue Res. 322, 73–79.
Catalano, P.M., Kirwan, J.P., 2001. Maternal factors that determine neonatal size and body fat. Curr. Diabetes Reprod. 1, 71–77.

Claris, O., Beltrand, J., Levy-Marchal, C., 2010. Consequences of intrauterine growth and early neonatal catch-up growth. Semin. Perinatol. 34, 207–210.
Fowden, A.L., 1995. Endocrine regulation of fetal growth. Reprod. Fertil. Dev. 7, 351–363.
Fowden, A.L., 2003. The insulin-like growth factors and feto-placental growth. Placenta 24, 803–812.
Harding, R., Bocking, A. (Eds.), 2001. Fetal Growth and Development. Cambridge University Press, Cambridge.
Hay Jr., W.W., Catz, C.S., Grave, G.D., Yaffe, S.J., 1997. Workshop summary: fetal growth: its regulation and disorders. Pediatrics 99, 585–591.
Hay Jr., W.W., Philip, A.G.S., Stevenson, D.K., 2017. Intrauterine growth restriction. In: Stevenson, D.K., Benitz, W., Sunshine, P., Hintz, S., Druzin, M. (Eds.), Fetal and Neonatal Brain Injury, fifth ed. Cambridge University Press, Cambridge UK, pp. 89–115.
Hay Jr., W.W., 2006. Recent observations on the regulation of fetal metabolism by glucose. J. Physiol. 572, 17–24.
Hay Jr., W.W., 2008. Nutrition and development of the fetus: carbohydrate and lipid metabolism (Chapter 28). In: Walker, W.A., Watkins, J.B., Duggan, C.P. (Eds.), Nutrition in Pediatrics (Basic Science and Clinical Applications), fourth ed. BC Decker Inc. Publisher, Hamilton, Ontario, Canada.
Regnault, T.R.H., Limesand, S.W., Hay Jr., W.W., 2006. Aspects of fetoplacental nutrition in intrauterine growth restriction and macrosomia. In: Thureen, P.J., Hay Jr., W.W. (Eds.), Neonatal Nutrition and Metabolism, second ed. Cambridge University Press, Cambridge, pp. 32–46.
Regnault, T.R.H., de Vrijer, B., Hay Jr., W.W., Meschia, G., Battaglia, F.C., 2021. Circulatory and metabolic changes accompanying fetal growth restriction. In: Polin, R.A., Abman, S.H., Rowitch, D.H., Benitz, W.E. (Eds.), Fetal and Neonatal Physiology, sixth ed. Elsevier, Philadelphia (in press), 2021.
Sparks, J.W., Cetin, I., 2006. Determinants of intrauterine growth. In: Thureen, P.J., Hay Jr., W.W. (Eds.), Neonatal Nutrition and Metabolism, second ed. Cambridge University Press, Cambridge, pp. 23–31.
Yogev, Y., Catalano, P.M., 2009. Pregnancy and obesity. Obstet. Gynecol. Clin. N. Am. 36, 285–300.

Growth monitoring and promotion

Leila M. Larson[a], **Edward A. Frongillo**[a], **and Jef L. Leroy**[b], [a] Department of Health Promotion, Education, and Behavior, University of South Carolina, Columbia, SC, United States; and [b] International Food Policy Research Institute, Washington, DC, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of M. de Onis, Growth Monitoring, In Encyclopedia of Human Nutrition (Third Edition), edited by Benjamin Caballero, Academic Press, 2013, https://doi.org/10.1016/B978-0-12-375083-9.00140-9.

Introduction 84
Objectives and activities 84
Measurement of growth 85
Use and interpretation of growth measurements 85
Growth standards and references 85
Anthropometric indices 87
Indicators, cutoff points, and interpretation of anthropometric measurements 87
Challenges in conducting growth monitoring and promotion 89
Evidence of effectiveness 90
Conclusion 90
References 90
Further reading 90

Key points

- List the main objectives and activities linked to growth monitoring and promotion.
- Summarize the proper measurement of weight, length, and height.
- Explain the use and interpretation of growth measurements using a standard or reference population, an appropriate anthropometric index, and a correct cutoff point to define an anthropometric indicator.
- Summarize some challenges with growth monitoring and promotion.
- Describe the evidence for effectiveness of growth monitoring and promotion.

Glossary

Growth monitoring The serial weighing and measuring of the length/height (and head circumference if $\leq$2 years old) of a child and graphing these measurements on a growth chart

Growth promotion The process of assessing growth using repeated weight, length/height, or head circumference measurements and providing counseling and motivation for actions to improve patterns of growth, nutrition, and health

Growth surveillance Monitoring the growth status of a population. Usually, measurements of height and weight are taken periodically on a representative sample of children to monitor trends in growth status in the population over time

Growth standard Defines a recommended pattern of growth that has been associated empirically with specified health outcomes and minimization of long-term risks of disease. It represents "healthy" growth of a population and suggests a model or target pattern of growth for all children to achieve. A standard describes "what should be"

Growth reference Describes the expected growth pattern of a defined population, without making any claims about associated health outcomes. A reference describes "what is"

Growth velocity The average change in a specific anthropometric measure over a specific age interval. Growth velocity varies depending on the child's age and is a more sensitive measure of small changes in growth status than attained growth

Length/height The length or height of children is measured from the top of the child's head to the bottom of their heel. In children under 24 months of age, recumbent length is used; in children 24 months and above, standing height is used

Malnutrition Deficiencies, excesses, or imbalances in intake of energy, protein and/or other nutrients. The term malnutrition includes both undernutrition and overnutrition

Z-scores Z-scores are a statistical unit used to describe how far a measurement is from the standard or reference population average (i.e., mean or median). Percentiles are commonly used in the clinical setting. Use of z-scores is almost universal for population-based applications and research reporting. For comparison purposes, the 50th percentile is equal to a z-score of 0, the 15th and 85th percentiles approximate z-scores of -1 and $+1$ respectively, the 2.5th and 97.5th percentiles approximate z-scores of -2 and $+2$ respectively, and the 1st and 99th percentiles approximate z-scores of -3 and $+3$, respectively

Z-score	*Exact percentile*	*Rounded percentile*	*Z-score*	*Exact percentile*	*Rounded percentile*
0	50th	50th			
−1	15.9	15th	+1	84.1	85th
−2	2.3	3rd	+2	97.7	97th
−3	0.1	1st	+3	99.9	99th

Introduction

Growth monitoring and promotion is regarded as an essential component of primary health care for infants and children. Serial measurements of weight and height for all children, and head circumference for infants and toddlers, are compared with the growth of a large sample population of children depicted on an appropriate growth chart and are used to confirm a child's healthy growth and development. The objective of growth monitoring and promotion is to screen for potential nutritional or health problems which subsequently prompts action before a child's health is seriously compromised.

The consequences of undernutrition during the early years include compromised immunity, cognitive problems, and stunted growth. Excess energy intake may predispose to conditions such as obesity, diabetes, and metabolic syndrome later in life. When potential problems are identified early, health professionals, and parents can work together to initiate action before the child's nutritional status or health are significantly affected.

Despite the ubiquitous use of growth monitoring and promotion globally, how effective it is and the conditions that would make it effective are not known.

Objectives and activities

The main objectives of growth monitoring and promotion are to:

- Screen children at risk of poor nutrition and health.
- Initiate effective action in response to abnormal patterns of growth.
- Teach parents how nutrition, physical activity, genetics, and illness can affect growth and, in doing so, motivate and facilitate individual initiative and improved child-care practices.
- Provide regular contact with and facilitate use of primary health care services.

Activities linked to growth monitoring and promotion at the individual level include:

- Accurately measuring weight, length or height, and head circumference.
- Precisely plotting measurements on the appropriate growth chart.
- Correctly interpreting the child's pattern of growth.
- Discussing the child's growth pattern with the parent(s) or caregiver and agreeing on subsequent action when required.
- Assessing the child's response to the actions.

Adapting the objectives and activities of interventions to the environment in which they are used is important. In practice, growth monitoring is different in low-income and high-income settings due to differences in the burden of disease, resources, and training. In low- and middle-income countries, undernutrition and infectious diseases such as diarrhea and respiratory infections are believed to primarily affect growth outcomes in the first 5 years of life. In contrast, in high-income countries, such conditions are less common and milder, and the focus is more on growth disorders such as growth hormone deficiency or Turner's syndrome. This different focus affects the target age range when children are monitored. Whereas in low- and middle-income countries growth monitoring targets children under 5 years of age, in high-income countries it covers all of childhood up to and including puberty.

In addition to detecting disease and raising parental awareness at the individual level, growth monitoring theoretically can create community mobilization. With caregivers being more aware and concerned about their child's health, they may influence community leaders and collective action to improve the underlying social and economic determinants of poor health.

Growth surveillance (i.e., monitoring the growth of populations) provides information about average child growth and prevalence of growth faltering that is useful for comparison, policy, and planning. For example, monitoring the growth status of populations from different regions is useful for identifying areas where the prevalence of malnutrition is highest, which in turn allows resources and support staff to be effectively targeted.

Measurement of growth

During infancy, the measurement of weight is simple to do, requires reasonably inexpensive equipment, and provides a convenient global summary of the infant's size. Birth weight is a useful proxy for fetal growth. An advantage of weight is that it relates closely to the caregivers' own perception of her child's size. Although weight is the easiest anthropometric measure to obtain, weight lacks the biological specificity necessary to differentiate variation due to height from variation due to body mass at a given height.

The routine collection of length or height measurements (recumbent length before 24 months of age and standing height for older children) is important to enable the assessment of length or height itself and weight relative to length or height. Body mass index (BMI) (i.e., ratio of weight in kilograms to the square of height in meters) is valuable for monitoring both undernutrition and overweight and obesity in childhood. During infancy, length is difficult to measure for several reasons. The optimal equipment is a length board with a sliding footboard, which is expensive and needs regular calibration. Simpler equipment such as a tape measure increases the measurement error dramatically. Furthermore, proper length measurement requires two trained observers, one to hold the infant's head against the headboard and the other to position the footboard and take the measurement. Keeping an infant still long enough to take the measurement is challenging. For these reasons, infant length is often measured either poorly or not at all. Standing height is best measured with a stadiometer or a tape measure that has been carefully attached to a wall and a measuring block for the head.

Head circumference is another important measure in neonates and toddlers to detect abnormal patterns of growth due to conditions such as inadequate nutrition or hydrocephalus. Other available anthropometric measures that are used to describe growth status during childhood include mid-upper arm circumference, which is most commonly used in resource-poor settings to detect acute malnutrition. Other measurements such as triceps and subscapular skinfolds are useful proximate measures of body fat but are not widely used due, in part, to technical difficulties and the required skill of individuals to perform the measurements accurately and precisely.

The process of anthropometry requires attention to detail; suitable equipment that is regularly maintained and calibrated; observers who are trained in correct measurement technique; regular quality control sessions in which observers are checked against both themselves, each other, and an expert, for intra-rater and inter-rater reliability, measurement precision, and accuracy; correct recording and graphing of the results; and correct interpretation. Only in this way can valid conclusions be made.

Variability in infant and child measurements can result from several influences: the setting in which the measurements are taken (e.g., home or clinic), the degree of filling of the stomach or bladder (for weight), the behavior and cooperation of the child being measured, accuracy and precision of instruments, the technical ability (i.e., training, experience, and reliability) of the anthropometrist, and the methods for recording data. Adequate training on the use of appropriate measurement techniques, continual standardization, adherence to specified methods and procedures, and monitoring of data quality are essential for reducing measurement error and minimizing bias. A comprehensive description of the techniques used in the development of the WHO Child Growth Standards can be found under the Further Reading section.

Use and interpretation of growth measurements

The proper use of growth measurements requires the selection of (1) an appropriate standard or reference population with which to compare the child or community, (2) an appropriate anthropometric index, and (3) correct cutoff points to interpret anthropometric measurements and classify children according to the extent of malnutrition.

Growth standards and references

In the context of growth assessments, the distinction between a standard and a reference is important from theoretical and practical perspectives. A standard defines a recommended pattern of growth that has been associated empirically with specified good health outcomes and minimization of long-term risks of disease. A standard embodies the concept of a norm or target and thus leads to a value judgment about what should be. In contrast, a reference describes the expected growth pattern of a defined population without making any claims about associated health outcomes. A reference provides a means of grouping and analyzing data and making comparisons but does not involve a value judgment.

There is now broad international consensus about the utility of the World Health Organization (WHO) Child Growth Standards (www.who.int/childgrowth/en) for assessing the growth of preschool children. Based on an international sample of breastfed infants and young children, the WHO standards depict human growth of children up to 60 months of age under optimal environmental conditions and can be used to assess children everywhere, regardless of ethnicity, socioeconomic status, and type of feeding. A salient outcome of the WHO growth standards project is the striking similarity in length or height of the child populations from the six countries (Brazil, Ghana, India, Norway, Oman, and the USA) that participated in the study (**Fig. 1**). The WHO standards have been adopted by more than 120 countries worldwide. **Figs. 2**, **3** and **4** present examples of generic WHO growth charts for weight, height, and head circumference in the age group 0–60 months. Useful software to monitor the growth of individual children and populations can be found at www.who.int/tools/child-growth-standards/software.

After 5 years of age, the WHO reference for school-age children and adolescents is recommended by the WHO for both clinical and epidemiological use in the 5–19 years age group. The full set of tables and charts is available at www.who.int/toolkits/growth-reference-data-for-5to19-years, including application tools such as software for clinicians and public health specialists.

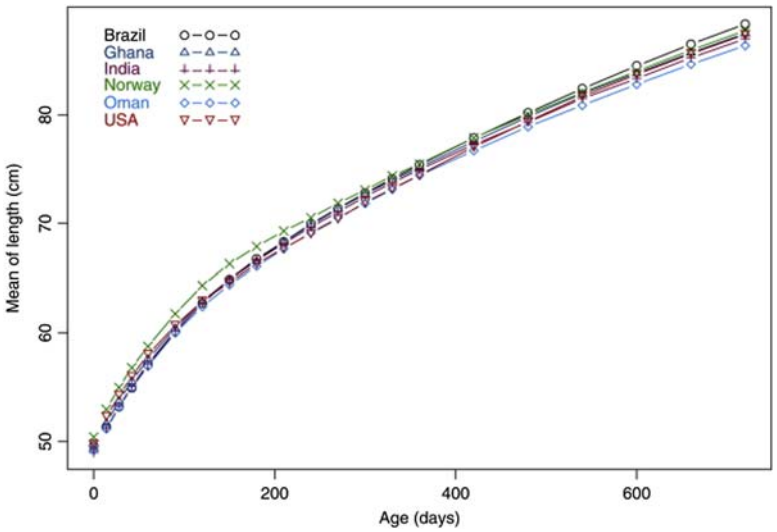


Fig. 1 Mean length (cm) from birth through 2 years for each of the six sites in the WHO Multicentre Growth Reference Study. Reproduced from WHO Multicentre Growth Reference Study Group. Acta Paediatr. 95 (2006): 56–65.

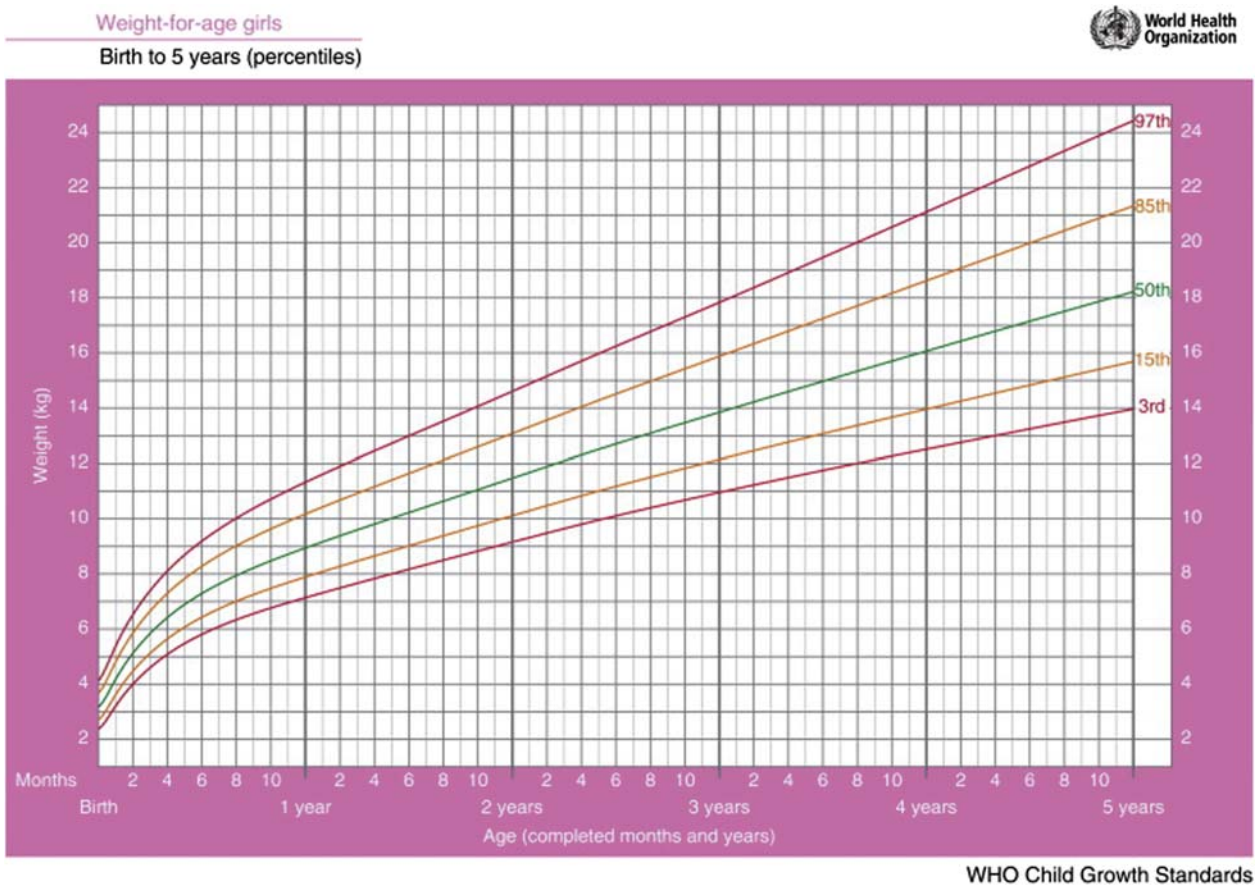


Fig. 2 Weight-for-age for girls from birth to 5 years (percentiles). Reproduced from WHO Child Growth Standards at https://www.who.int/tools/child-growth-standards/standards.

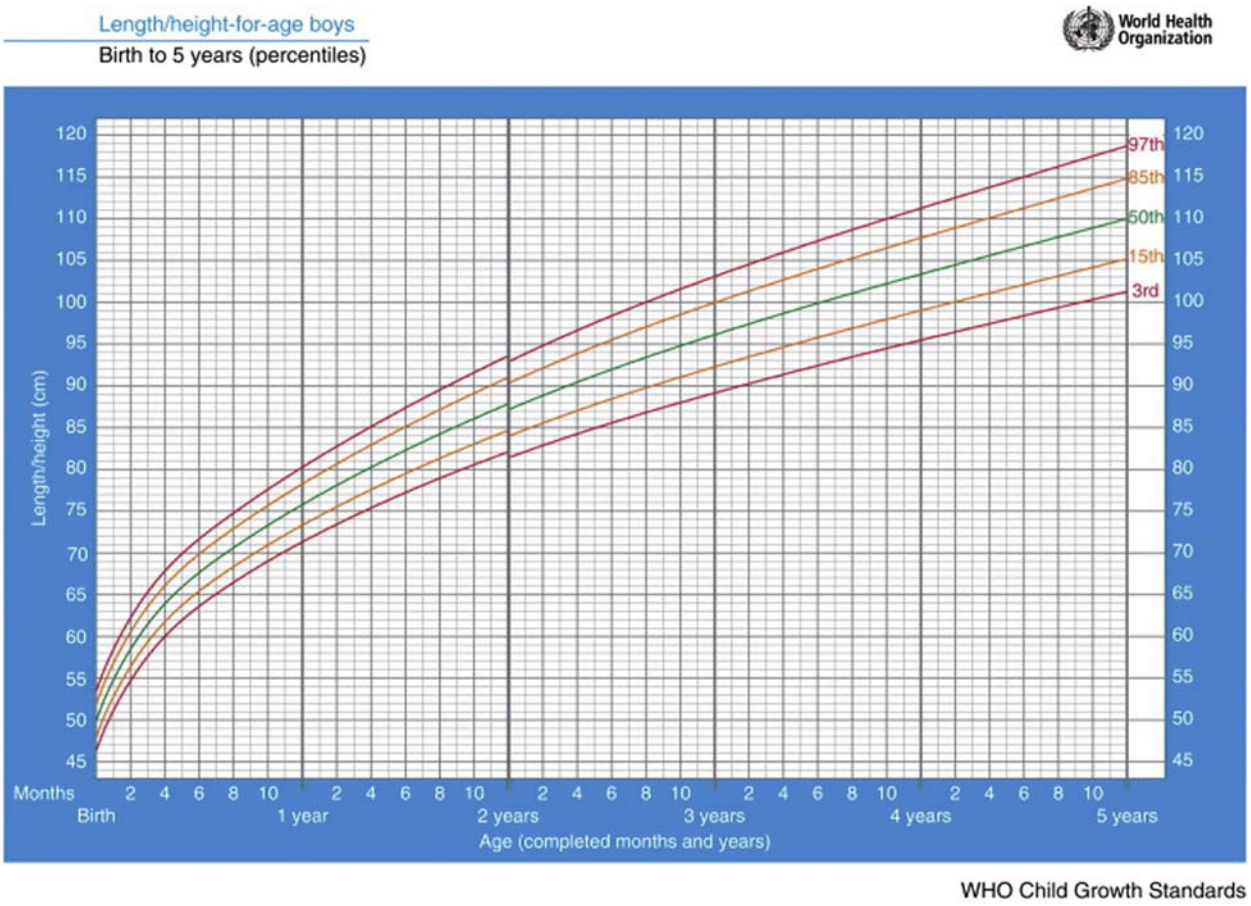


Fig. 3 Length/height-for-age for boys from birth to 5 years (percentiles). Reproduced from WHO Child Growth Standards at https://www.who.int/tools/child-growth-standards/standards.

Anthropometric indices

Anthropometric indices are combinations of measures and are essential for interpretation and comparison. In children the three most used indices to assess growth status are weight-for-age, height-for-age, and weight-for-height. Weight-for-age is the most used and, in many low-income countries, the sole anthropometric index used. These anthropometric indices can be expressed as Z-scores or percentiles relative to a standard or reference population. A Z-score (i.e., standard deviation score) is the difference between an individual measurement and the sex-specific median value of the standard or reference population, divided by the sex-specific standard deviation of the standard or reference population. A Z-score of zero corresponds to the measurement being at the median of the standard or reference population. Z-scores of $+1$ and -1 correspond to values one standard deviation above or below the median, respectively. Z-scores of $+2$ and -2 correspond to analogous values. Z-scores are comparable across ages, sexes, and other traits. For example, a Z-score of $+2$ always designates the same relative position within a normal distribution. This consistency has enormous advantages whenever comparisons are made.

Indicators, cutoff points, and interpretation of anthropometric measurements

Indicators are intended to provide guidance for assessment, referral, or intervention by selecting a cutoff point for a continuous anthropometric measure or index. Cutoffs can be based on the mathematical estimation of presumed risk depending strictly on the cutoff's distance from a given measurement's average (i.e., mean or median) or on a demonstrable link between the cutoff and a designated health outcome (e.g., illness or death).

Indicators are commonly expressed in terms of cutoffs in percentiles or multiples of a standard deviation (i.e., Z-score) of the targeted measure's distribution. The recommended cutoffs (percentiles and Z-scores) by the WHO for classification of undernutrition and overnutrition are presented in **Table 1**.

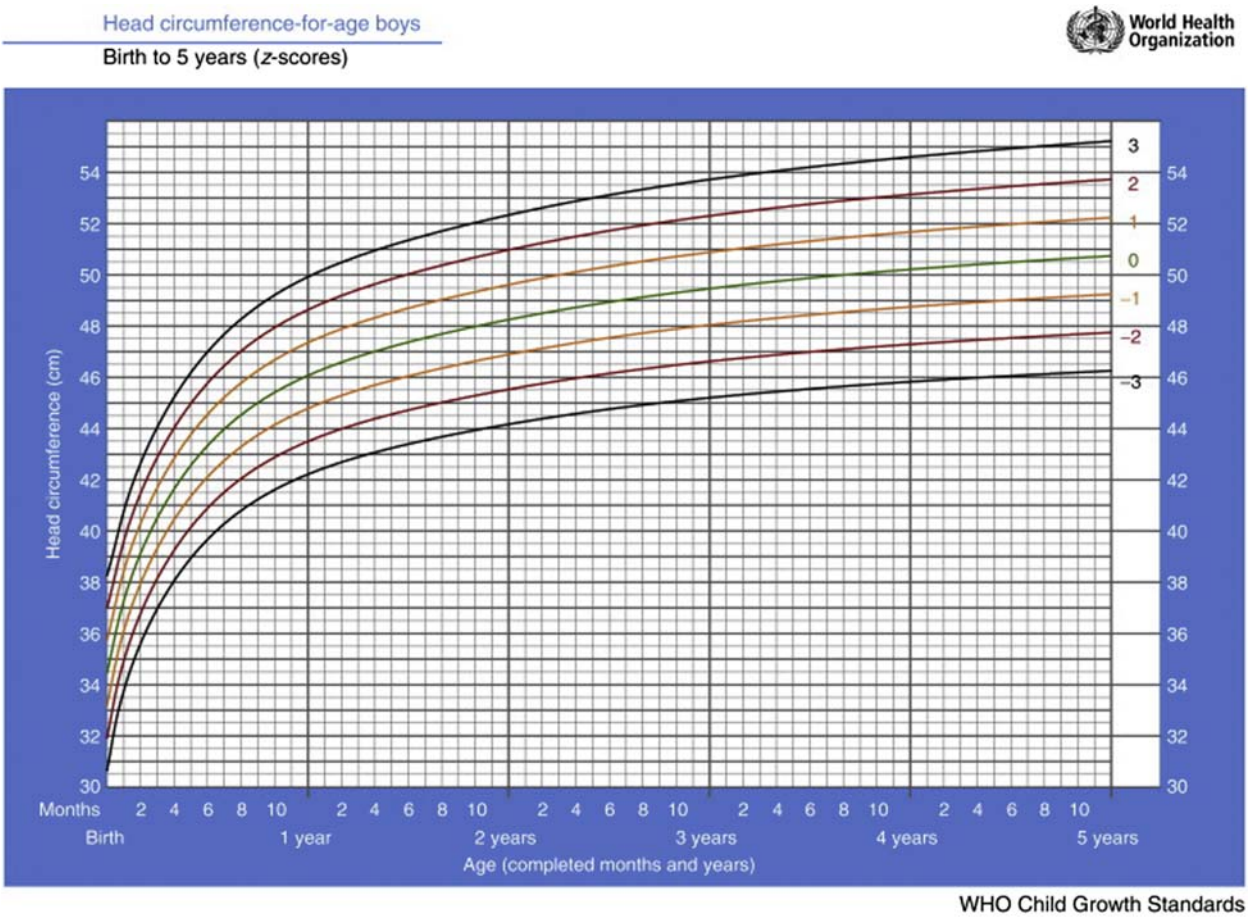


Fig. 4 Head circumference-for-age for boys from birth to 5 years (z-scores). Reproduced from WHO Child Growth Standards at https://www.who.int/tools/child-growth-standards/standards.

Table 1 Recommended cutoffs by the WHO for classification of undernutrition and overweight/obesity.

Indicator	*WHO child growth standards (birth to 5 years)*	*WHO reference 2007 (5–19 years)*
Underweight (i.e., low weight-for-age)	<−2 Z-scores or <3rd centile	<−2 Z-scores or <3rd centile
Stunted (i.e., low length-for-age or height-for-age)	<−2 Z-scores or <3rd centile	<−2 Z-scores or <3rd centile
Wasted (i.e., low weight-for-length/height or BMI-for-age)	<−2 Z-scores or <3rd centile	<−2 Z-scores or <3rd centile
Overweight (i.e., high weight-for-length/height or BMI-for-age)	>+2 Z-scores or >97th centile	>+1 Z-scores or >85th centile
Obese (i.e., very high weight-for-length/height or BMI-for-age)	>+3 Z-scores or >99.9th centile	>+2 Z-scores or >97th centile

An important concept in the use of anthropometric measurements for interpreting growth is the need for serial measurements. One-time measurements, taken and plotted accurately on a growth chart, reflect a child's size at one time and may be used to screen children for nutritional risk using the cutoffs indicated in **Table 1**. One-time measurements do not provide adequate information to assess a child's growth and are not sufficient to be considered growth monitoring (Mangasaryan et al., 2011). A series of weight and length or height measurements over time are more informative and reflect a child's growth pattern.

The interpretation of the commonly used anthropometric indicators (based on one-time measurements) is as follows:

Low weight-for-age: Weight-for-age reflects body mass relative to chronological age. It is influenced by both the child's height and his or her weight. Its composite nature makes interpretation uncertain. For example, weight-for-age fails to distinguish between short children of adequate body weight and tall, thin children. In the absence of significant acute malnutrition in a community, similar information is provided by weight-for-age and height-for-age in that both reflect an individual's or population's long-term health and nutritional experiences. Short-term changes, especially reductions in weight-for-age, reveal changes in weight relative to height. In general terms, the worldwide variation of low weight-for-age and its age distribution are like those of low height-for-age.

Low height-for-age: Stunted growth reflects a process of failure to reach linear growth potential because of suboptimal health or nutritional conditions. On a population basis, high levels of stunting are associated with poor socioeconomic conditions and increased risk of frequent and early exposure to adverse conditions such as illness or inappropriate feeding practices. Similarly, a decrease in the national stunting rate is usually indicative of improvements in overall socioeconomic conditions of a country. The worldwide variation of the prevalence of low height-for-age is considerable, ranging from 5% to 65% among low-income countries.

In populations without significant undernutrition, low height-for-age in individual children likely signals deprivation due to conditions not necessarily related to poverty. Thus, for example in high-income countries, it is necessary to also consider the possible metabolic or genetic mechanisms that may result in growth failure.

Low weight-for-height: Wasting or thinness results in most cases from a recent and severe process of weight loss, which is often associated with acute starvation or severe disease, but wasting also may be the result of chronic unfavorable conditions. Provided there is no severe food shortage, the prevalence of wasting is usually below 2.5%, even in poor countries.

High weight-for-height: Overweight is the preferred term for describing high weight-for-height. Even though there is a strong association between high weight-for-height and obesity as measured by adiposity, greater lean body mass can also contribute to high weight-for-height. On an individual basis, therefore, fatness or obesity should not be used to describe high weight-for-height.

High BMI-for-age: Another measure of body mass relative to height, BMI, provides a widely used indicator for classifying overweight and obesity for individual and population use. Although there is some reluctance to describe children as obese based on BMI alone, i.e., without considering more direct measures of body fat, recognition of the difficulties inherent in obtaining more proximate measures of body fat and lack of references to interpret them has resulted in BMI-for-age alone being used to define overweight and obesity. In its favor, high BMI-for-age in childhood and adolescence is associated with high percentages of body fat and known risk factors for cardiovascular disease.

Challenges in conducting growth monitoring and promotion

Conducting growth monitoring and promotion is challenging in practice. Some challenges are limited agreement about what growth monitoring and promotion actually means and what are its objectives (Mangasaryan et al., 2011), poor quality measurements and other implementation constraints (Mangasaryan et al., 2011; Roberfroid et al., 2005a; Scherdel et al., 2016), problems of low coverage (Mangasaryan et al., 2011; Roberfroid et al., 2005a), lack of guidance on serial measurements, and the absence of internationally agreed and validated algorithms to define abnormal growth (Scherdel et al., 2016; Roberfroid et al., 2005a).

The lack of agreement on a definition of growth monitoring and promotion and confusion around its objectives has resulted in discordant expectations and minimal evidence of its impact (Mangasaryan et al., 2011). For instance, there are no set agreed-upon activities that constitute growth promotion. Furthermore, other interventions are often built into growth monitoring and promotion activities, without distinguishing them as separate inputs with potentially separate objectives.

Quality of implementation of growth monitoring and promotion activities depends largely on the context. Experiences from large-scale programs in low- and middle-income settings have demonstrated important deficits in the technical abilities of staff to measure and chart growth but also in interpreting growth charts and initiating appropriate action. These deficits are particularly striking in situations where resources for the latter activities are lacking, making growth monitoring a futile exercise. Training of health workers on effective counseling and communication skills is essential to the quality of implementation. When conveying messages and recommendations to parents, health workers need to ensure this information leads to benefits rather than harm due to the potential for misinterpretation of information. For instance, growth monitoring may not be beneficial if parents who are assured that they do not have a stunted child then interpret this message as no longer needing to give as much attention to their child's nutrition and subsequently reduce the support provided to that child.

Coverage of growth monitoring and promotion on a large-scale, to reach all target children in a setting, has proved to be challenging. Attendance is low, particularly in older children and in families with the most at-risk children. Low attendance could be due to parents' lack of interest after completing the vaccination schedule, weakness of awareness campaigns to motivate caregivers, and barriers to using the information received during the monitoring sessions because of illiteracy and low food access (Roberfroid et al., 2005b).

Although growth monitoring entails periodic frequent anthropometric measurements, guidance is lacking on how to conduct and interpret the repeated measurements. For instance, no consensus exists on the recommended frequency for growth monitoring. UNICEF recommends monthly growth monitoring in the first year of life (with no recommendation thereafter), whereas WHO has no recommended schedule but indicates to follow country-specific guidance if available (WHO, 2008).

A review of 69 studies reporting algorithms for growth monitoring in children found seven algorithms used to define abnormal growth, but all suffered from lack of validation and poor sensitivity and specificity (Scherdel et al., 2016). The lack of validated algorithms is especially marked in high-income countries where the emphasis is on detecting primary growth disorders such as Turner syndrome, Celiac disease, cystic fibrosis, growth hormone deficiency, renal tubular acidosis, and small-for-gestational age with failure to catch up.

Evidence of effectiveness

Growth monitoring and promotion, along with other activities, is often seen as a catch-all for nutritional problems and an effective means of improving the nutritional status of children. No rigorous evidence exists, however, on the effectiveness of growth monitoring and promotion that justifies its use. A 2000 Cochrane Review attempted to answer the question "Does growth monitoring work?" The authors defined growth monitoring as "the regular recording of a child's weight, coupled with some specified remedial actions if the weight is abnormal in some way" (Panpanich and Garner, 2000). Only two randomized clinical trials testing the effectiveness of growth monitoring were identified. In India, there was no difference in mean weight or height gain between children whose growth was monitored for 30 months and control children (George et al., 1993). A study in Lesotho showed that mothers trained to use a growth chart were more knowledgeable about growth patterns after 4 months (Ruel et al., 1990). A recently conducted cluster-randomized trial in Zambia assessed the effectiveness of two growth monitoring models: home-based growth monitoring using a growth chart installed against a wall at home and community-based growth monitoring combined with a micronutrient-fortified soy and maize flour for children with stunted linear growth (Fink et al., 2017). No differences were found in child length or height growth between the two models. Compared to their peers in the control group who received no intervention, those who received community-based growth monitoring with nutrition supplementation performed worse on developmental tests.

Conclusion

Growth monitoring and promotion activities are meaningful if they empower parents to provide adequate nutrition, care, and support for their child. Despite the limited evidence on the effectiveness of growth monitoring and promotion, enormous resources are devoted to it each year in low-, middle-, and high-income countries. The near-universal implementation appears to be mostly motivated by the belief that growth monitoring is intrinsically beneficial for children and their parents.

References

Fink, G., Levenson, R., Tembo, S., Rockers, P.C., 2017. Home- and community-based growth monitoring to reduce early life growth faltering: an open-label, cluster-randomized controlled trial. Am. J. Clin. Nutr. 106, 1070–1077.

George, S.M., Latham, M.C., Abel, R., Ethirajan, N., Frongillo Jr., E.A., 1993. Evaluation of effectiveness of good growth monitoring in south Indian villages. Lancet 342, 348–352.

Mangasaryan, N., Arabi, M., Schultink, W., 2011. Revisiting the concept of growth monitoring and its possible role in community-based nutrition programs. Food Nutr. Bull. 32, 42–53.

Panpanich, R., Garner, P., 2000. Growth monitoring in children. Cochrane Database Syst. Rev. Cd001443.

Roberfroid, D., Kolsteren, P., Hoerée, T., Maire, B., 2005a. Do growth monitoring and promotion programs answer the performance criteria of a screening program? A critical analysis based on a systematic review. Trop. Med. Int. Health 10, 1121–1133.

Roberfroid, D., Lefèvre, P., Hoerée, T., Kolsteren, P., 2005b. Perceptions of growth monitoring and promotion among an international panel of district medical officers. J. Health Popul. Nutr. 23, 207–214.

Ruel, M.T., Pelletier, D.L., Habicht, J.P., Mason, J.B., Chobokoane, C.S., Maruping, A.P., 1990. Comparison of mothers' understanding of two child growth charts in Lesotho. Bull. World Health Organ. 68, 483–491.

Scherdel, P., Dunkel, L., Van Dommelen, P., Goulet, O., Salaün, J.F., Brauner, R., Heude, B., Chalumeau, M., 2016. Growth monitoring as an early detection tool: a systematic review. Lancet Diabetes Endocrinol. 4, 447–456.

UNICEF. Key practice: Monitoring growth and development of the child. https://www.unicef.org/uganda/key-practice-monitoring-growth-and-development-child.

WHO, 2008. Training Course on Child Growth Assessment: WHO Child Growth Standards. World Health Organization, Geneva, Switzerland. https://apps.who.int/iris/bitstream/handle/10665/43601/9789241595070_B_eng.pdf?sequence=2&isAllowed=y.

Further reading

American Academy of Pediatrics Policy Statement, 2003. Prevention of pediatric overweight and obesity. Pediatrics 112, 424–430.

de Onis, M., Onyango, A.W., Van den Broeck, J., Chumlea, C., Martorell, R., 2004. For the WHO Multicentre Growth Reference Study. Measurement and standardization protocols for anthropometry used in the construction of a new international growth reference. Food Nutr. Bull. 25 (Suppl. 1), S27–S36.

de Onis, M., Wijnhoven, T.M.A., Onyango, A.W., 2004. Worldwide practices in child growth monitoring. J. Pediatr. 144, 461–465.

de Onis, M., Blössner, M., Borghi, B., 2010. Global prevalence and trends of overweight and obesity among preschool children. Am. J. Clin. Nutr. 92, 1257–1264.

de Onis, M., 2013. Update on the implementation of the WHO child growth standards. World Rev. Nutr. Diet. 106, 75–82. https://doi.org/10.1159/000342550. Epub 2013 Feb 11. PMID: 23428684. https://www.karger.com/Article/Abstract/342550.

Jelliffe, D.B., Jelliffe, E.F.P., 1990. Growth Monitoring and Promotion in Young Children. Oxford University Press, Oxford.

Panpanich, R., Garner, P., 2003. Growth monitoring in children, update software. In: The Cochrane Library. Oxford [Cochrane Review], Issue 3.

Victora, C.G., de Onis, M., Hallal, P.C., Blössner, M., Shrimpton, R., 2010. Worldwide timing of growth faltering: revisiting implications for interventions using the World Health Organization growth standards. Pediatrics 125 (3), e473–e480.

WHO Child Growth Standards for Preschool Age Children at http://www.who.int/childgrowth.

WHO Global Database on Child Growth and Malnutrition at http://www.who.int/nutgrowthdb.

WHO Growth Reference for School Age Children and Adolescents at http://www.who.int/growthref.

WHO Training Course on Child Growth Assessment at http://www.who.int/childgrowth/training.

WHO, 1995. Physical Status: The Use and Interpretation of Anthropometry. Report of a WHO Expert Committee. Technical Report Series No. 854. World Health Organization, Geneva, Switzerland.

Lactation: Dietary requirements

H. Lochen and H. Tymann, Johns Hopkins Hospital, Baltimore, MD, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of L.H. Allen, Lactation: Dietary Requirements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 54–59, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00167-7.

Introduction 91
Rationale for recommended nutrient intakes 91
Requirements and dietary recommendations 92
Macronutrients 92
Energy 92
Fat 93
Protein 93
Minerals and vitamins 94
Calcium 94
Zinc 95
Vitamin A 96
Folate 96
Other B vitamins 96
Conclusions 96
References 97

Key points

- Nutrient needs during lactation are elevated secondary to the mobilization of tissue stores and transfer of nutrition through the milk
- The development of these recommendations and guidelines were created based on fundamental nutrient needs from the World Health Organization (WHO) and the Institution of Medicine
- Without adequate nutritional intake during lactation, nutrient depletion may occur due to the excessive mobilization of nutrient stores

Introduction

Human lactation causes a significant increase in the requirement for most nutrients. Lactation is primarily supported by mobilization of tissue stores, which affects weight and nutritional status. Without adequate nutritional intake during lactation, nutrient depletion may occur due to this excessive mobilization of nutrient stores (Allen, 2006). Nutritional status during lactation can affect milk quantity and composition. Recommended intakes of most nutrients for lactation are 10–90% higher (Institute of Medicine, 1991). The dietary recommendations for lactation are those of the Food and Agriculture Organization/World Health Organization Reports on fats (2008) and micronutrients (2001), and the dietary reference intakes (DRIs) of the Institute of Medicine (USA and Canada) for micronutrients (1997, 1998, 2000, 2001), and macronutrients (2002), with an update on calcium and vitamin D in 2011. The rationales for the recommended nutrient intakes, and requirements and dietary recommendations for energy, fats, protein, calcium, zinc, folate, and vitamin A are addressed specifically. Given calls from LGBTQ+ organizations and statement from the Academy of Breastfeeding Medicine, this chapter will use non-gendered language to support and affirm non-cisgender or gender diverse lactating parents (Bartick et al., 2021). While using the anatomical term "breast" or "breast milk" in this chapter, please note that the term may be triggering and therefore recommended to discuss preferred terms with each client/patient.

Rationale for recommended nutrient intakes

Recommendations on dietary nutrient intakes for lactation by different scientific authorities are typically based on the estimated total amount of each nutrient secreted daily into breast milk, the efficiency of milk synthesis, and the bioavailability of the nutrient. This estimate for each nutrient is then added to the recommended nutrient intake for nonpregnant, non-lactating assigned female at birth individuals (FAO/WHO, 2008). The onset of lactation after parturition is brought about by the major hormonal changes that occur in this period. During the first 2–7 days postpartum a thick yellow fluid (colostrum) is secreted. With the progress of lactation,

the volume of milk secreted increases and its nutrient composition changes. After approximately 21 days, the milk is considered mature milk (Darragh and Lönnerdal, 2011). The volume of breast milk secreted daily increases rapidly in the first postpartum days, from ~500 mL on day 5, ~650 mL at 1 month, and ~750 mL at 3 months, thereafter remaining relatively stable until decreasing during weaning. In industrialized countries, the average volume of breast milk produced is 750–800 mL day^{-1} in the first 4–5 months postpartum and decreases to 600 mL day^{-1} during 6–12 months after delivery (Neville et al., 1988). In this period, the volume of milk produced may be even lower and is quite variable, depending on the weaning practices for the infant. The FAO/WHO and DRI committees (2008) considered 750 and 780 mL, respectively, as the average milk volume produced during full lactation and the basis for recommendations. For most nutrients, average concentration in mature milk multiplied by the average milk volume was used to estimate the total amount of nutrient secreted daily into breast milk. A correction factor was then applied when the bioavailability of a nutrient in the diet is less than 100%, and where known, for the anabolic cost of milk synthesis. The final value was added to the recommended intake for nonpregnant, nonlactating assigned female at birth individuals. The stage of lactation was considered to be a factor for some nutrients and, where applicable, separate values were given according to the period of time postpartum (Medicine, 2002).

The volume of milk secreted during lactation is not influenced by nutritional status, unless undernutrition is severe. The composition of breast milk for most nutrients is also adequate to support infant growth and development regardless of nutritional status of the lactating individual. However, diet and nutritional status do have an influence on the concentration of some micronutrients such as vitamins A, D, thiamin, riboflavin, vitamins B_6 and B_{12}, choline, iodine, and selenium. Also, the fatty acid composition of breast milk can be affected by diet (Daniels et al., 2019).

When setting recommendations, the DRI committees take gestational parent age into account, so there are separate values for adolescent (≤18 years) and adult (19–50 years) lactating individuals. For some nutrients, adolescents may have greater requirements than adults because they are still growing and they need to cover their own nutrient demands. Recommended intakes of vitamins A and C, calcium, phosphorus, magnesium, iron, and zinc are also higher for adolescents than for adults during lactation (Institute of Medicine, 2011).

In general, there is considerable uncertainty about dietary nutrient recommendations for lactation due to high intra- and inter-individual variability in breast milk volume output, specific nutrient concentrations in milk, and the temporal changes in milk volume and nutrient concentrations during a day and across the period of lactation. The composition of human milk is affected by several factors depending on the nutrient, such as stage of lactation, changes during a feeding, diurnal rhythm, diet, gestational age of the infant at birth, and parity. Moreover, the total amount of nutrients secreted into breast milk depends on the extent and duration of breast feeding. In addition, physiological adaptations to the high nutrient demands for lactation may include increased nutrient absorption and conservation, and use of nutrient stores. These adaptations are quite specific for each nutrient and not easily quantified, which contributes to the degree of uncertainty. Gestational parental age and nutritional status during pregnancy and lactation may also influence the homeostatic adaptations during lactation including the efficiency of nutrient absorption and the degree of mobilization of maternal nutrient stores.

Requirements and dietary recommendations

Macronutrients

Energy

The dietary energy intake recommended for healthy adults is that required to maintain energy balance, considering gender, age, weight, height, and level of physical activity. The energy requirements during lactation include the additional energy that is necessary for milk production, which differ by the stage and extent of breastfeeding.

The energy density of human milk is mainly determined by its fat content, which represents 50–60% of the total energy in mature milk and is the most variable energy-yielding component. Protein and lactose contribute to approximately 5% and 38% of energy, respectively. The mean energy density of representative 24 h pooled mature milk samples from well-nourished individuals ranges from 0.64 to 0.74 k cal g^{-1} (2.7–3.1 kJ g^{-1}) (USDA, 2020).

The estimated energy requirements (EER) for lactation are set by the DRI committee and based mainly on studies done in the 1990s, using the doubly labeled water method. The main findings in individuals who were fully breastfeeding their infants up to 6 months of age were: total energy expenditure of 2109–2580 kcal day^{-1} (8860–10,840 kJ day^{-1}) or 35.8–41.0 kcal kg day^{-1} (150–172 kJ^{-1} kg^{-1} day^{-1}), milk energy output of 483–538 kcal day^{-1} (2030–2260 kJ day^{-1}), and energy mobilization from tissue stores of 72–287 kcal day^{-1} (300–1200 kJ day^{-1}) (DRI, 2005). It was concluded that the energy requirements of lactating, well-nourished individuals are met primarily from the diet and partially by mobilization of tissue stores, without evidence for adaptations in basal metabolism and physical activities. The EER for lactation during the first 6 months is calculated as the sum of the EER obtained from the equation for nonlactating adults (using current age, weight, and physical activity level), the energy secreted in milk energy (500 kcal day^{-1} or 2100 kJ day^{-1}), and subtracting the energy derived from tissue mobilization during the weight loss that normally occurs during lactation (170 kcal day^{-1} or 714 kJ day^{-1}). The committee assumed a milk production rate of 0.78 L day^{-1} from birth through 6 months of age, a milk energy density of 0.67 kcal g^{-1} (2.8 kJ day^{-1}), and an average gestational weight loss of 0.8 kg $month^{-1}$. For the second 6 months of lactation, the incremental EER is calculated assuming a milk energy output of 400 kcal day^{-1} or 1680 kJ day^{-1} (a milk production rate of 0.6 L day^{-1}) and no weight loss. The EER for lactating

adolescents (14–18 years) is calculated in the same manner as for lactating adults, but the increment for lactation is added to the appropriate equation for estimating the EER of nonlactating adolescents.

The acceptable macronutrient intake distribution ranges, expressed as percentage of total dietary energy, are the same as for the general adult population: 10–35% protein, 20–35% fat, and 45–65% carbohydrates (DRI, 2005). Natural simple sugars, such as those present in fruit, and complex carbohydrates (polysaccharides), such as in cereals (rice, wheat), cereal products (flour, pasta) and starchy roots, should be the preferred sources of carbohydrates in the diet. Added sugars, usually sucrose, should not be higher than 25% of dietary energy. Many of the energy-yielding carbohydrate food sources are also sources of dietary fiber which is beneficial for reducing the risk of coronary heart disease, ameliorating constipation, and other health outcomes. A total fiber intake of 29 g day^{-1} is recommended for lactating women. Whole grain cereals, nuts, legumes, and fruit are good fiber and energy sources, and are also nutrient-rich foods. Restriction of energy intake during lactation to values below 1800 kcal (7500 kJ) per day may lead to low intakes of other nutrients including vitamins and minerals.

Fat

Total fat content of human milk is affected by several factors, including stage of lactation, stage of feeding, and parity, while dietary intake of energy, fat, or fatty acids and overall nutritional status have little influence, except when there is long-term or severe maternal undernutrition (Jensen, 1989). Milk fat content is highly variable, averaging 35–40 g L^{-1} in mature milk from well-nourished individuals delivering at term gestation. The content of individual fatty acids in milk is also highly variable, especially for the long-chain polyunsaturated fatty acids (LCPUFA, especially docosahexenoic acid, DHA), and more dependent on diet than total fat. The concentration of DHA can vary from 0.2 to 1% of total fatty acids and is proportional to the usual dietary intake. Fatty acid intake and relative contribution of carbohydrate and fat to total energy intake, as well as the body stores and endogenous synthesis of the lactating individual, influence the fatty acid composition of human milk. In well-nourished persons, the polyunsaturated EFA linoleic acid (18: 2n-6) and α-linolenic acid (18: 3n-3) represent approximately 11 and 1% (wt/wt), respectively, of the total fatty acids in milk. LCPUFA of the n-6 and n-3 series account for 1.2 and 0.6% of the fatty acids, respectively (Delgado-Noguera, 2015).

The adequate transfer of polyunsaturated fatty acids from circulation of the lactating individual to the milk and the synthesis of LCPUFA, especially arachidonic acid (20: 4n-6), dihomo-γ-linolenic acid (20: 3n-6), eicosapentenoic acid (EPA, 20: 5n-3), and DHA (22: 6n-3), from their respective EFA precursors, are important for infant growth, neurodevelopment, and visual function. These polyunsaturated fatty acids are structural components of all cell membrane phospholipids. Arachidonic acid and DHA are the two quantitatively most important LCPUFA in the brain and retina, and the LCPUFA with 20 carbon atoms are precursors for the synthesis of eicosanoids, a group of signaling molecules. The major part of the polyunsaturated fatty acids in human milk (70–85% in an omnivorous diet) is derived from body stores which reflect dietary intake over the long term, and not from direct dietary transfer (Delgado-Noguera, 2015).

The metabolic fate of individual fatty acids depends on dietary energy intake and energy balance. Therefore, the intake and requirements for fat, EFA, and LCPUFA are usually expressed as a percentage of the total energy in the diet, rather than total intake. The fat intake recommended for lactating persons is in the range of 20–35% which is the same range as recommended for the adult population. Concerning the fatty acid intake, FAO/WHO (2008) recommends an additional intake of 1–2% as EFA (3–4 g day^{-1}) during the first 3 months of lactation, and up to 4% (about 5 g day^{-1}) thereafter due to depletion of fat stores. Based on the median linoleic and α-linolenic acid intakes of lactating individuals in the US, the DRI committee recommends an intake of 5–10% (average 13 g day^{-1}) of n-6 (as linoleic acid) and of 0.6–1.2% (average 1.3 g day^{-1}) of n-3 (as α-linolenic acid) polyunsaturated fatty acids throughout lactation, with a 10% contribution from LCPUFA in the n-6 and n-3 series to these ranges. The ratio of n-6: n-3 unsaturated fatty acids in the diet is important because these fatty acids are desaturated and elongated, and incorporated into membranes, using the same series of enzymes. Increased intake of linoleic acid reduces the conversion of α-linolenic acid to EPA and DHA, whereas the conversion of linoleic acid to arachidonic acid is inhibited by EPA and DHA, as well as by arachidonic acid, α-linolenic acid, and linoleic acid itself. The n-6:n-3 ratio recommended for adults by both DRI and FAO/WHO committees is 5:1 to 10:1. Vegetable oils are the main dietary source of n-6 fatty acids, and also of n-3 fatty acids although in lower amounts. Fish such as herring, mackerel, and salmon are good sources of n-3 fatty acids.

The intake of trans fatty acids (trans isomers of oleic and linoleic acid) present in hydrogenated food fats and oils, deep-fried foods, and meats are of special concern during lactation when intake is excessively high, or when EFA intake is low during pregnancy and lactation. An inverse correlation of arachidonic acid and DHA with trans fatty acids in plasma lipids has been reported in infants, suggesting impaired LCPUFA synthesis and metabolism.

Protein

The average protein content in colostrum is 15–20 g L^{-1} decreasing to approximately 8–10 g L^{-1} in mature human milk during the first 6 months of lactation. The protein concentration in human milk is not affected by diet, body composition, or undernutrition.

The recommended dietary allowance (RDA) of protein for adolescent and adults who are lactating set by the DRI committee is 1.1 g kg^{-1} of body weight per day. This corresponds to an increment of 25 g day^{-1} above the RDA for nonlactating persons assigned female at birth, and is the same as for pregnancy. Recent data have shown that protein intakes of 1 g/kg/day are able to maintain good milk production, and promote conservation of skeletal muscle apparently by down-regulating protein metabolism. The recommended range of percentage of energy from dietary protein is the same as for the general adult population (10–35%).

The factorial approach was used to estimate the protein RDA for lactation, assuming that the maintenance protein requirement of the lactating individual is not different from that of nonlactating individual assigned female at birth, and that the additional protein and/or amino acid requirements are proportional to milk production. The additional protein requirement for lactation is defined as the output of total protein and nonprotein nitrogen (the latter converted to protein by multiplying by 6.25) in milk. Nonprotein nitrogen represents 20–25% of total milk nitrogen, mainly as urea. It is taken into account because it is assumed that the nitrogen needed to cover the total nitrogen loss in milk should be derived from dietary protein. The total protein output, approximately 10 g day^{-1}, is divided by the incremental efficiency of nitrogen utilization (0.47), which is assumed to be the same in lactating adults and adolescents. The additional estimated average requirement due to milk production is therefore 21.2 g day^{-1}. After correction by the coefficient of variation and rounding off, the RDA for lactation is +25 g day^{-1}, which corresponds to +0.46 g protein kg day^{-1} (based on a reference of 57 kg individual) above the RDA for nonlactating individuals assigned female at birth.

Recommendations for individual indispensable amino acids for lactation by the DRI committee assume that the incremental needs correspond to the amino acids secreted in milk, because there are no specific data on the amino acid requirements of lactating individuals. Therefore, the RDA for amino acids in lactation is calculated by adding the average amounts of amino acids in human milk in the first 6 months of lactation (expressed as mg kg^{-1} day^{-1}) to the respective RDA for nonlactating adults assigned female at birth. Overall, recommendations for indispensable amino acids are 36% (histidine) to 80% (tryptophan) higher during lactation. High-quality protein from sources such as eggs, milk, meat, and fish provide the requirements for all indispensable amino acids. Individuals who restrict their diets to plant proteins (cereals, legumes, nuts, starchy roots, vegetables, and fruits) may be at risk of inadequate intakes of certain indispensable amino acids. However, adequate complementary mixtures of plant proteins, with increased digestibility through processing and preparation, can provide high-quality protein.

Minerals and vitamins

Daily requirements for several micronutrients (riboflavin, vitamins B_{12}, C, A, and E, copper, iodine, manganese, selenium, and zinc) are higher during lactation than during pregnancy, indicating that lactation is a very demanding process. The only micronutrients needed in lower amounts during lactation are iron, due to the small amount of iron secreted into human milk and to the usual amenorrhea, and folate. However, iron requirements may be high postpartum for those who need to replace major blood losses during delivery.

The recommended intakes for micronutrients during lactation established by FAO/WHO and DRI committees are summarized in Table 1. The percentages of change from the recommendations for nonpregnant nonlactating adults are also shown. To meet these intakes, those lactating should be guided to consume daily a large variety of foods rich in micronutrients, because food diversification contributes to improve the intake of limiting nutrients. Micronutrients most commonly at risk of inadequate intakes by lactating women are calcium, zinc, folate, and vitamin A.

Calcium

It is estimated that an average of 200 mg of calcium per day is secreted into mature human milk although this amount is variable, usually ranging from 150 to 300 mg day^{-1}. The diet does not affect human milk calcium concentration except when calcium intake is very low (<300 mg day^{-1}). The primary source of calcium for milk production is the increased mobilization of calcium from bone due to the increased bone resorption that occurs during lactation, favored by the low estrogen concentration. This results in a net loss of bone mass during lactation that is regained after weaning upon return of ovarian function. The decreased urinary calcium excretion during lactation also contributes to the calcium economy for milk secretion. The efficiency of intestinal calcium absorption is not increased during lactation and does not contribute to the extra calcium needed for milk production.

Calcium homeostasis is maintained during lactation and has been found to be independent of dietary calcium intake. It has been demonstrated that the loss of bone mass during lactation is not affected by calcium supplementation with habitual adequate dietary calcium intakes (Prentice, 2000; Polatti et al., 1999). Because the loss of bone calcium that occurs during lactation is not prevented by increased dietary calcium, and the calcium lost appears to be regained after weaning, the recommended intake of calcium during lactation is the same as that for nonpregnant nonlactating adults assigned female at birth of the same age, being 1000 and 1300 mg day^{-1} for adults and adolescents, respectively. Even if not increased during lactation, the recommended calcium intake may be difficult to obtain for those with a low habitual intake of dairy products. Therefore, those who are lactating should be guided to consume dairy products such as milk, yogurt, cheese, and other calcium-rich foods such as fish with edible bones, broccoli, and kale.

Lactating adolescents are a group of special concern regarding calcium intake due to the already high calcium requirements of non-pregnant nonlactating adolescents. They are still increasing their own bone density as well as needing the increased calcium requirement to support lactation. Studies are needed to investigate if they are able to regain bone after weaning to the same level as when they were nonpregnant and nonlactating, and if they would benefit from increased calcium intake.

Table 1 Daily recommended mineral and vitamin intakes for lactating adults.

	FAO/WHO[a]		IOM[b]	
Nutrient	*Recommended value*	*% change*[c]	*Recommended value*	*% change*[c]
Vitamin A (μg RAE)	–	–	1300	↑ 86
Vitamin A (μg RE)	850	↑ 70	–	–
Vitamin D (μg)	5	No change	15	No change
Vitamin E (mg α-TE)	7.5	No change	19	↑ 27
Vitamin K (μg)	55	No change	90	No change
Thiamin (mg)	1.5	↑ 36	1.4	↑ 27
Riboflavin (mg)	1.6	↑ 45	1.6	↑ 45
Niacin (mg NE)	17	↑ 21	17	↑ 21
Vitamin B_6 (mg)	2.0	↑ 54	2.0	↑ 54
Pantothenate (mg)	7.0	↑ 40	7.0	↑ 40
Biotin (μg)	35	↑ 17	35	↑ 17
Folate (μg DFE)	500	↑ 25	500	↑ 25
Vitamin B_{12} (μg)	2.8	↑ 17	2.8	↑ 17
Vitamin C (mg)	70	↑ 55	120	↑ 60
Calcium (mg)	1000	No change	1000	No change
Iodine (μg)	200	↑ 82	290	↑ 93
Iron (mg)	15[d]	↓ 49	9	↓ 50
Zinc (mg)	9.5[e]	↑ 94	12	↑ 50
	8.8[f]	↑ 80		
Magnesium (mg)	270	↑ 23	310	No change
Selenium (μg)	35	↑ 35	70	↑ 27
Chromium (μg)	–	–	45	↑ 80
Copper (μg)	–	–	1300	↑ 44
Fluoride (mg)	–	–	3	No change
Manganese (mg)	–	–	2.6	↑ 44
Molybdenum (μg)	–	–	50	↑ 11

RAE, retinol activity equivalent; α-TE, alpha-tocopherol equivalent; NE, niacin equivalent; DFE, dietary folate equivalent.

[a]FAO/WHO (2001) Human vitamin and mineral requirements. *Report of a Joint FAO/WHO Expert Consultation*. Rome: Food and Agriculture Organization.

[b]Institute of Medicine (IOM) (2001) *Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc*. Also (2011) *Dietary Reference Intakes for Calcium and Vitamin D*. Washington, DC: National Academies Press.

[c]Changes from recommendations for nonpregnant nonlactating adults: ↑, increase; ↓, decrease.

[d]Assuming 10% bioavailability.

[e]0–3 months postpartum, assuming moderate bioavailability.

[f]4–6 months postpartum, assuming moderate bioavailability.

Zinc

Zinc concentrations in human milk are highest in colostrum, decrease rapidly during the first 3 months postpartum, and continue to fall but more gradually at later stages of lactation. Typical milk zinc concentrations are 4 mg L^{-1} at 2 weeks, 3 mg L^{-1} at 4 weeks, 2 mg L^{-1} at 8 weeks and 1.2 mg L^{-1} at 24 weeks. These concentrations are not influenced by either dietary intake or zinc supplementation, at least in well-nourished individuals (Krebs, 1998). Less is known about the effect of low dietary zinc intakes on milk zinc concentrations, but some reports indicate that concentrations in developing countries may be slightly lower than those in developed countries at comparable times postpartum (Krebs, 1998).

Average daily losses of zinc in human milk range from 2.2 mg during the first month postpartum to 1 mg at 6 months (Dror and Allen, 2018). The average estimate of daily output of zinc in milk during the first 3 months of lactation is 1.6 mg, which would theoretically double the minimum endogenous zinc losses during lactation. However, homeostatic mechanisms such as enhanced zinc absorption and reduced urinary zinc excretion compensate for the secretion of zinc into human milk, independent of zinc intake. Intestinal conservation of endogenous fecal zinc appears to contribute to zinc homeostasis during lactation at low zinc intakes (<8 mg day^{-1}). Involution of the uterus, decreased blood volume, and increased resorption of trabecular bone in the postpartum period also contribute to mobilizable zinc pools to compensate for the increased needs. These sources appear to provide up to 0.5 mg day^{-1} of zinc during the first 3 months of lactation (Krebs, 1998). Taking all these adaptive mechanisms into account, the average estimate for the increased requirement for absorbed zinc during the first 6 months of lactation is 1.35 mg day^{-1}. Therefore, dietary zinc requirements during lactation are substantially increased, both for adults and adolescents.

Bioavailability is an important factor in setting dietary zinc recommendations because the efficiency of dietary zinc utilization may vary up to fivefold depending on the overall composition of the diet, and particularly on the negative effect of dietary phytate. The efficiency of absorption is inversely related to the level of dietary zinc, except when phytate intake is high. Dietary zinc recommendations during lactation are set at 12 mg day^{-1} for those consuming a mixed diet, but requirements will be higher for those

whose diet is high in phytate because it is based mainly on unrefined cereals and legumes. Red meat, milk, poultry, eggs, and seafood provide highly available zinc, and their consumption should be encouraged during lactation.

Vitamin A

Vitamin A is present in human milk as retinyl esters (95%) and free retinol. Vitamin A activity is also provided as carotenoid precursors, mainly as β-carotene, which accounts for up to 30% of total carotenoids in human milk. The concentration of vitamin A is high in early lactation (600–2000 μg L^{-1}) and declines thereafter to 200–1100 μg L^{-1} (Dror and Allen, 2018; Haskell and Brown, 1999). It is responsive to dietary intake, particularly in those with poor vitamin A status.

Dietary recommendations for vitamin A during lactation are based on replacing the amount of the vitamin secreted into breast milk during the first 6 months of lactation, while preserving vitamin A stores. Because the bioconversion of carotenoids in human milk is still uncertain, the contribution of carotenoids in breast milk to meeting the vitamin A lactation is not considered.

Based on the average milk vitamin A milk concentration of 485 μg L^{-1}, an additional intake of 400 μg retinol activity equivalents (RAE) per day is recommended during lactation, which represents an increase of over 70% compared to the recommendations for non-pregnant nonlactating adolescents and adults. A RAE is defined as 1 μg all-trans-retinol, 12 μg beta-carotene, or 24 μg α-carotene or β-cryptoxanthin. The amount of carotenoids equivalent to 1 RAE is double the equivalent to 1 RE (retinol equivalent).

The vitamin A intake recommended during lactation can be obtained as the preformed vitamin from foods of animal origin (primarily milk products, eggs, and liver) and as carotenoid precursors in green leafy vegetables and ripe, colored fruits. However, meeting the recommended intake by consumption of plant sources alone, as is the case in many developing countries, may be difficult unless the diet contains some carotenoid-rich foods such as sweet potatoes.

Folate

The concentration of folate in human milk increases during the lactation period, with lower values for colostrum (10–40 μg L^{-1}) than for mature milk (79–133 μg L^{-1}) (Dror and Allen, 2018). These concentrations are several-fold higher than in plasma, independent of folate status, indicating that the mammary gland actively transports and regulates the secretion of this vitamin into milk. Folate concentration in breast milk is maintained with the concomitant depletion of folate when dietary intake is low. Supplementation during lactation has little effect on milk folate but it benefits the folate status of the lactating individual (O'Connor, 1994). Dietary folate requirements during lactation are based on the average milk folate concentration of 85 μg L^{-1} and assume 50% absorption from a mixed diet. The average extra amount of dietary folate needed to cover lactation is thus estimated as 133 μg day^{-1}, an increase of approximately 40% above the nonpregnant nonlactating average folate requirements. Dietary folate recommendations during lactation are set at 500 μg dietary folate equivalents (DFEs) daily. A DFE is defined as 1 μg of food folate, or 0.6 μg of folic acid from fortified food or as a supplement taken with meals, or 0.5 μg of folic acid as a supplement taken on an empty stomach. Thus, in order to meet lactation requirements, less of this vitamin is needed when given as pure folic acid than as natural food folate.

Although folate is found in a variety of foods, such as fresh green vegetables, oranges, legumes and nuts, several servings per day of these foods are needed to meet recommended intake. Moreover, considerable losses of folate can occur during food harvesting, storage, and cooking. Fortification of wheat flour with folic acid has become mandatory or encouraged in many countries in order to reduce the risk of neural tube defects in those at risk of this condition.

Other B vitamins

Deficiency of thiamin, riboflavin, vitamin B_6, vitamin B_{12}, and choline can result in low concentrations in breast milk and subsequent infant deficiency. It has been estimated that intakes of infants exclusively breastfed by a deficient mother may be in the range of 16% (vitamin B_{12}) to 80% (vitamin B_6) of their recommended Adequate Intakes (Allen, 2012). Further information is needed about the status of these vitamins in lactation and their infants because intakes of most B vitamins are inadequate where animal source food consumption is low—a common situation in resource limited populations. Increasing dietary intake through supplementation increases the milk content of these vitamins and improves the status of the parent and infant.

Conclusions

Human lactation causes a significant increase in the requirement for most nutrients given the need to provide adequate nutrition to infant and to support the mass mobilization of tissue. Without adequate nutritional intake during this time, those lactating are at risk for nutrient depletion which can affect milk production and composition. The recommended dietary intakes of both macro- and micronutrients during lactation developed by the World Health Organization and Institute of Medicine account for the amount not only to maintain nutrient stores of the individual lactating but also to ensure optimal composition of human milk for the infant.

References

Allen, L.H., 2006. Pregnancy and lactation. In: Bowmen, A., Russell, R.M. (Eds.), Present Knowledge in Nutrition. International Life Science Institute Press, Washington, DC, pp. 529–543.

Allen, L.H., 2012. B vitamins in breast milk: relative importance of maternal status and intake, and effects on infant status and function. Adv. Nutr. 3 (3), 362–369.

Bartick, M., Stehel, E.K., Calhoun, S.L., Feldman-Winter, L., Zimmerman, D., Noble, L., Rosen-Carole, C., Kair, L.R., Academy of Breastfeeding Medicine, 2021. Academy of breastfeeding medicine position statement and guideline: infant feeding and lactation-related language and gender. Breastfeed. Med. 16 (8), 587–590.

Daniels, L., et al., 2019. Micronutrient intakes of lactating mothers and their association with breast milk concentrations and micronutrient adequacy of exclusively breastfed Indonesian infants. Am. J. Clin. Nutr. 110 (2), 391.

Darragh, A., Lönnerdal, B., 2011. Milk | human milk. In: Fuquay, J.W. (Ed.), Encyclopedia of Dairy Sciences, second ed. Academic Press, pp. 581–590.

Delgado-Noguera, M.F., Calvache, J.A., Bonfill Cosp, X., 2015. Supplementation With Long Chain Polyunsaturated Fatty Acids (LCPUFA) to Breastfeeding Mothers for Improving Child Growth and Development. Cochrane Database Syst Rev, Columbia.

Dietary Reference Intakes, 2005. Dietary Reference Intakes: For Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids (Macronutrients). National Academy Press, Washington, DC.

Dror, D.K., Allen, L.H., 2018. Overview of nutrients in human milk. Adv. Nutr. 9 (Suppl. l_1), 278S–294S.

FAO/WHO, 2008. Human Vitamin and Mineral Requirements. Food and Agriculture Organization, Rome.

Haskell, M.J., Brown, K.H., 1999. Maternal vitamin A nutriture and the vitamin A content of human milk. J. Mammary Gland Biol. Neoplasia 4 (3), 243–257.

Insitiute of Medicine, 2011. Dietary Reference Intakes for Calicum and Vitamin D. National Acadamies Press, Washington, DC.

Institute of Medicine (US), 1991. Subcommittee on lactation. Nutrition During Lactation, Committee on Nutritional Status During Pregnancy and Lactation, Food and Nutrition Board, Institute of Medicine, National Academy of Sciences. National Academy Press, Washington, DC.

Institute of Medicine, 2002. Dietary Reference Intakes for Energy, Carbohydrates, Fiber, Fat, Fatty Acids, Cholesterol, Protein and Amino Acids. Natyional Academies Press, Washington, DC.

Jensen, R., 1989. The Lipids of Human Milk. CRC Press, Boca Raton.

Krebs, N.F., 1998. Zinc supplementation during lactation. Am. J. Clin. Nutr. 68 (2), 509S–512S.

Nevillle, M., et al., 1988. Studies in human lactation: milk volumes in lactating women during the onset of lactation and full lactation. Am. J. Clin. Nutr. 48 (6), 1375–1386.

O'Connor, D.L., 1994. Folate status during pregnancy and lactation. In: Nutrient Regulation During Pregnancy, Lactation, and Infant Growth, pp. 157–172.

Polatti, F., Capuzzo, E., Viazzo, F., Colleoni, R., Klersy, C., 1999. Bone mineral changes during and after lactation. Obstet. Gynecol. 94 (1), 52–56.

Prentice, A., 2000. Calcium in pregnancy and lactation. Annu. Rev. Nutr. 20 (1), 249–272.

U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2020. Dietary Guidelines for Americans, ninth ed., pp. 2020–2025 [Online] Available at. https://www.dietaryguidelines.gov/. (Accessed 19 November 2021).

Lactation: Physiology

J.L. McManaman and M.C. Neville, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of J.L. McManaman. M.C. Neville, Lactation: Physiology, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 60–66, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00168-9.

Functional Anatomy of Lactation 99
Milk Composition 100
Synthesis and Secretion of Milk Components 100
Exocytotic Pathway (I) 101
Lipid Secretion Pathway (II) 101
Transcytosis Pathway (III) 102
Transmembrane Pathway (IV) 102
Paracellular Transport Pathway (V) 103
Regulation of Milk Synthesis, Secretion, and Ejection 103
Hormonal Control of Milk Synthesis and Secretion 103
Local Control of Synthesis and Secretion 103
Regulation of Milk Ejection 103
Initiation of Lactation 104
Changes in Milk Composition during Secretory Activation 104
Hormonal Control of Secretory Activation 105
Delays in Secretory Activation 106
Further Reading 106

Glossary

Arborizing ductal network The system of ducts composed of epithelial cells that connect lobuloalveolar units to the nipple.

Junctional complex Tight junctions branching networks of transmembrane protein strands that seal adjacent epithelial cells, preventing passage of ions and other substances from blood to milk.

Lobuloalveolar unit The portion of the mammary gland specialized for producing and secreting milk. The lobuloalveolar unit is composed of a single layer of secretory epithelial cells that synthesize and secrete milk. Secretory epithelial cells are surrounded by a network of myoepithelial cells that contract in response to oxytocin stimulation to stimulate milk ejection.

Paracellular transport The passage of substances between epithelial cells. Paracellular transport is regulated by tight junctions and is closed during lactation.

Secretory activation The hormone regulated process of maturation of gene expression, secretory mechanisms and transport pathways that occurs after parturition, and that is required for copious milk secretion. Secretory activation is also called lactogenesis II.

Transcytosis The endocytic uptake of substances at the basal membrane and their transport to the apical membrane for secretion.

Lactation is a uniquely mammalian physiological process in which the caloric and nutrient reserves of the mother are transformed into a complex fluid capable of supporting the nutritional demands of newborns for sustained periods. Milk, the product of lactation, is a mixture of solutes whose composition reflects the activities of distinct secretion and transport processes of the mammary gland and mirrors the differing nutritional requirements of mammalian neonates. In humans, this fluid is capable of providing the full-term infant with all the nutrients required for the first 4–6 months of life as well as offering significant protection against infectious disease. Although artificial formulas are widely utilized for human infant nutrition in developed countries, many components of human milk, including critical growth factors, long-chain polyunsaturated fatty acids (PUFA), antiinfectious oligosaccharides and glycoconjugates, and the protein lactoferrin, are not duplicated in formula. Although it is likely that such substances are beneficial even to healthy infants in well-protected environments, they are particularly important for infants living under conditions of inadequate sanitation, as well as for preterm infants and infants with feeding problems. Despite the obvious importance of milk to neonatal nutrition and the selective advantage of lactation in mammalian evolution, the physiological mechanisms underlying milk secretion and utilization are not well understood and the molecular mechanisms involved in the production of individual milk components are still poorly characterized. In this article, the functional anatomy of the mammary gland is described, followed by a brief description of human milk composition and a review of the transport mechanisms involved in the secretion of individual

milk components. The authors then summarize the functional differentiation of the mammary gland and the initiation of lactation – a process that involves a series of carefully programmed functional changes that transform a prepared, but nonsecretory, gland into a fully functioning organ during the first week postpartum in humans.

Functional Anatomy of Lactation

The lactating mammary gland consists of an arborizing ductal network that extends from the nipple and terminates in grape-like lobular clusters of alveoli forming the lobuloalveolar unit, which is the site of milk secretion. A stylized diagram of these structures is shown in **Figure 1**. Alveoli are composed of a single layer of polarized secretory epithelial cells that possess specialized features indicative of highly developed biosynthetic and secretory capacities, including numerous mitochondria, an extensive rough endoplasmic reticulum network, and a well-developed Golgi apparatus. Secretory components including lipid droplets and casein-containing secretory vesicles are found juxtaposed to the apical membrane of these cells. The epithelial cells are connected to each other through a junctional complex composed of adherens and tight junctional elements that function to inhibit the transfer of extracellular substances between the vascular system and milk compartments during lactation (**Figure 2**). The basal portion of alveolar epithelial cells is surrounded by a meshwork of myoepithelial cell processes that contract to bring about milk ejection and by a connective tissue stroma that supports and separates the lobules. The stromal component also contains lymphatics and becomes extensively vascularized during lactation to sustain the biosynthetic demands of alveolar epithelial cells. In nonpregnant, nonlactating animals, the stroma contains a large adipose component.

The nipple, which is the termination point of the mammary ductal network, is innervated by the fourth intercostal nerve. Afferent sensory stimuli from suckling are transmitted to the spinal cord and the brain, resulting in the release of prolactin and oxytocin from the pituitary. Prolactin, secreted from the anterior pituitary, acts directly on alveolar epithelial cells to foster the synthesis and secretion of milk components. Oxytocin, secreted from the posterior pituitary, stimulates contraction of the myoepithelial cells that surround the alveoli and ducts. This process, called the 'letdown reflex,' forces the milk from the alveoli through ductules into ducts draining several clusters of alveoli. In humans, the small ducts converge into 15–25 main ducts that drain sectors of the gland and open directly on the nipple. The secretory product is stored in the alveolar space until myoepithelial cell contractions force it through the ducts toward the nipple, where it is available to the suckling infant.

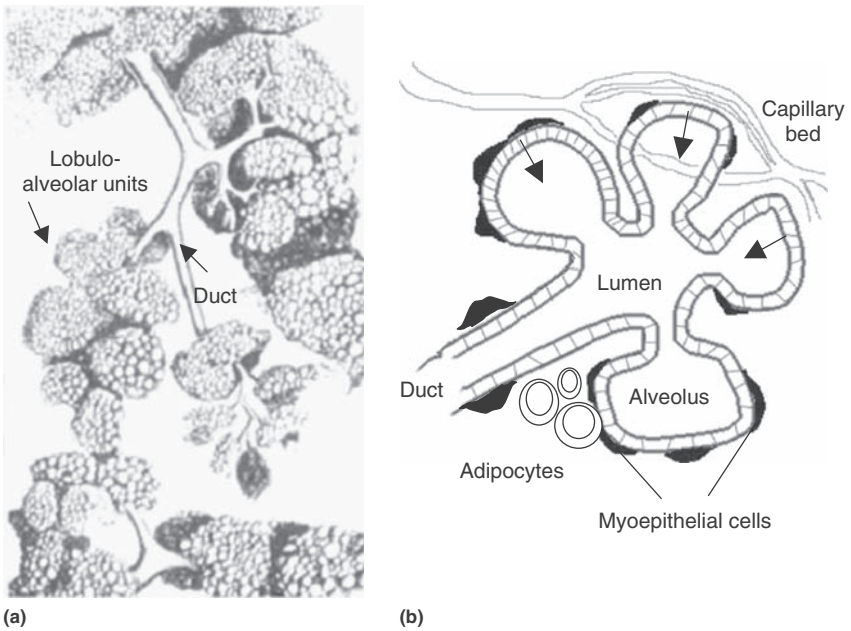


Figure 1 Camera lucida drawing of a section of the breast of a woman who died 2 days after last suckling her infant (a). The drawing clearly shows collecting ducts and the grape-like lobuloalveolar units, which are engorged with milk. (b) Cross-sectional diagram showing the relationships of the lobuloalveolar unit composed of milk-secreting alveoli and ducts with the other cellular compartments of the mammary gland. Arrows indicate milk secretion by the alveolar epithelial cells into the lumen. Camera lucida drawing is reproduced from Dabelow A (1941) Der Entfaltungsmechanismus derMamma. II. Die postnatale entwicklung der menschlichenmilchdruse und ihre korrelationen. *Morphology Journal* 85: 361–416, with permission from Wiley.

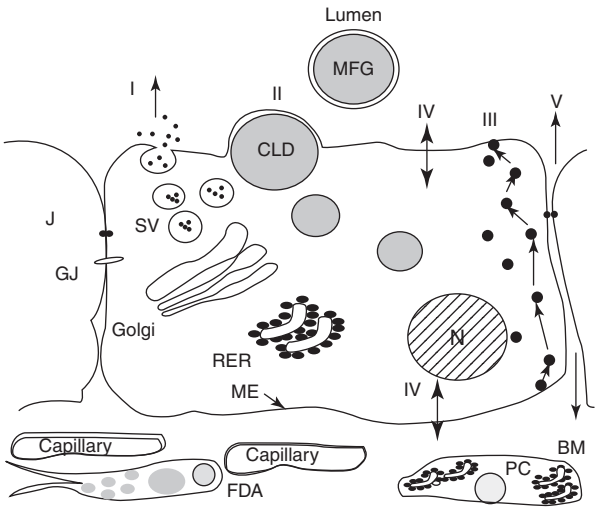


Figure 2 Diagram of a mammary epithelial cell showing pathways for milk secretion described in the text. Abbreviations: SV, secretory vesicle; RER, rough endoplasmic reticulum; BM, basement membrane; N, nucleus; PC, plasma cell; FDA, fat-depleted adipocyte; JC, junctional complex containing the tight and adherens junctions; GJ, gap junction; ME, myoepithelial cell; CLD, cytoplasmic lipid droplet; MFG, milk fat globule. Redrawn from Neville MC, Allen JC, and Watters C (1983) The mechanisms of milk secretion. In: Neville MC and Neifert MR (eds.) *Lactation: Physiology, Nutrition and Breast-Feeding*, p. 50. New York, NY: Plenum Press.

Milk Composition

The major macronutrients in milk are lactose (a disaccharide unique to milk); lipids; proteins, including casein, α-lactalbumin, lactoferrin, secretory immunoglobulin A (sIgA), and many others present at much lower concentrations; and minerals such as sodium, chloride, calcium, and magnesium. Other nutritionally important substances in milk are enzymes, vitamins, trace elements, and growth factors. The lipid content of milk varies considerably between species. In human and cow's milk, the fat accounts for approximately 4% of milk volume, whereas in whales and seals it can account for as much as 60% of milk volume. Milk fat is primarily composed of triglycerides, a major source of neonatal calories, and it also contains cholesterol and phospholipids, essential for early neonatal development. Casein micelles form a separate phase that can be pelleted by high-speed centrifugation or acidification. These micelles have a high calcium and phosphate content. The aqueous fraction of milk, often called whey, is a true solution that contains all the milk sugar as well as the major milk proteins lactoferrin, α-lactalbumin, and sIgA and nonprotein nitrogen compounds (mostly urea); the monovalent ions sodium, potassium, and chloride; citrate; calcium; free phosphate; and most of the water-soluble minor components of milk. The casein fraction from cow's milk, usually obtained by rennin precipitation, is used in cheese making, whereas the whey has a multiplicity of uses, most notably as the base for infant formula. Urea and other nonprotein nitrogen components of milk are a source of nitrogen for amino acid and protein synthesis. Isotope utilization studies indicate that on average 10–20% of urea nitrogen is converted into protein by breast-fed infants. Significantly higher utilization rates, however, have been measured in children recovering from infection, suggesting that alterations in urea nitrogen utilization may be a homeostatic response. Human and bovine milk differ primarily in their concentrations of lactose, mono- and divalent ions, and casein levels and the existence of antiinfectious agents in human milk (**Table 1**). These differences are related to the specific needs of these species. Human milk, for example, possesses higher concentrations of lactose and lower divalent ion concentrations than cow's milk. The high lactose concentration provides a large amount of 'free water,' *via* osmotic regulation, that serves as a reserve for temperature regulation *via* sweating in human infants. Human milk also contains a number of agents that protect against gastrointestinal and respiratory infections, including oligosaccharides that interact specifically with pathogen receptors, lactoferrin, and sIgA. Bovine milk, however, contains high concentrations of casein, which provides protein and associated calcium and phosphate needed to support the rapid growth of young calves.

Synthesis and Secretion of Milk Components

Solutes enter milk through five general pathways (**Figure 2**). Endogenously generated substances, including the major milk proteins, oligosaccharides, and nutrients such as lactose, citrate, phosphate, and calcium, are secreted through an exocytotic pathway (pathway I). Lipids and lipid-associated proteins are secreted by a process that is unique to mammary epithelial cells (pathway II). The transcytosis pathway (pathway III) transports a wide range of macromolecular substances derived from serum or stromal cells, including serum proteins such as immunoglobulins, albumin, and transferrin; endocrine hormones such as insulin, prolactin, and insulin-like growth factor-1; and stromal-derived agents such as immunoglobulin A (IgA) cytokines, and lipoprotein lipase. In addition, various membrane transport pathways (pathway IV) exist for the transfer of ions and small molecules, such as

Table 1 Comparison of the macronutrient contents of human and bovine milk

Component	*Human milk*	*Bovine milk*
Carbohydrates (g dl^{-1})[a]		
Lactose	7.3	4.0
Oligosaccharides	1.2	0.1
Proteins (g dl^{-1})[a]		
Caseins	0.2	2.6
α-Lactalbumin	0.2	0.2
Lactoferrin	0.2	Trace
Secretory IgA	0.2	Trace
β-Lactoglobulin	0	0.5
Nonprotein nitrogen (NPN) (g l^{-1})		
Total NPN	0.42[b]	0.29[c]
Urea	0.16[b]	0.14[c]
Milk lipids (%)[a]		
Triglycerides	4.0	4.0
Phospholipids	0.04	0.04
Minerals and other ionic constituents (mM)[a]		
Sodium	5.0	15
Potassium	15.0	43
Chloride	15.0	24
Calcium	7.5	30
Magnesium	1.4	5
Phosphate	1.8	11
Bicarbonate	6.0	5

[a]Reproduced from Neville MC (1998) Physiology of lactation. *Clinical Perinatology* 26: 251.
[b]Reproduced from Atkinson SA and Lonnerdal B (1995) In: Jensen RG (ed.) *Handbook of Milk Composition*. San Diego: Academic Press.
[c]Reproduced from Alston-Mills B (1995) In: Jensen RG (ed.) *Handbook of Milk Composition*. San Diego: Academic Press.

glucose, amino acids, and water across basal and apical plasma membranes. Finally, there is a paracellular pathway (pathway V) that provides a direct route for entry of serum and interstitial substances into milk. This pathway, however, closes during the first few days of lactation in humans. Transport through these pathways is affected by the functional state of the mammary gland and regulated by direct and indirect actions of hormones and growth factors. The general cellular and physiological properties of these pathways are summarized next.

Exocytotic Pathway (I)

Like exocytotic secretion mechanisms found in other cells, proteins, oligosaccharides, and nutrients such as lactose and citrate are packaged into secretory vesicles within the Golgi that are then transported to the apical region of the cell, where they fuse with the apical plasma membrane, discharging their contents into the extracellular space. A unique feature of this pathway in the mammary gland is the presence of high concentrations of lactose, phosphate, citrate, and calcium within the vesicles. Lactose is synthesized in the Golgi from UDP-galactose and glucose, which have entered from the cytoplasm using specific transporters, by the enzyme β-galactosidase, with α-lactalbumin acting as a cofactor. The high concentration of lactose present in the Golgi during lactation osmotically stimulates the influx of water that contributes to the fluidity of milk. Casein micelle formation begins in the terminal Golgi with condensation, and simultaneous phosphorylation, of casein molecules. The addition of calcium, possibly in the secretory vesicle, leads to maturation of casein micelles into particles sufficiently dense to be seen in the electron microscope. This complex thus delivers an efficient package of protein, calcium, and phosphate that provides the nutrients necessary for bone growth, among other things. Calcium enters the cytoplasm from the plasma by a poorly defined transport process. Cytoplasmic calcium is then transported into secretory vesicles by an ATP-dependent Ca^{2+} pump localized on Golgi and secretory membranes. The phosphate in secretory vesicles is derived from the hydrolysis of UDP-galactose during the synthesis of lactose. Citrate is generated endogenously within the cytoplasm of alveolar epithelial cells and transported into the Golgi lumen by an undefined process.

Lipid Secretion Pathway (II)

Estimates of the quantity of milk lipid secretion during lactation in humans and rodents indicate that in many species, the lactating mammary gland may be one of the most lipogenic organs in the body. In a fully lactating woman secreting 800 ml day^{-1} of milk

containing 4% fat, the mammary gland synthesizes approximately 32 g of triglyceride daily or approximately 6 kg, 10% of the weight of the woman, in a typical 6-month lactation. The fatty acids for triglyceride synthesis are synthesized from glucose or derived from the plasma lipids by the action of lipoprotein lipase. Once available in the mammary alveolar cells, fatty acids are either bound to a fatty acid-binding protein or activated by combination with coenzyme A (CoA) and then bound to an acyl-CoA-binding protein. Activated fatty acids are joined with glycerol-3-phosphate by transacylases located in the endoplasmic reticulum to form triglycerides, which enter the cytoplasm as protein-coated structures called cytoplasmic lipid droplets.

These structures are translocated to the apical membrane, where they are enveloped by a novel budding process that leads to their release as membrane-bound lipid droplets known as milk fat globules. The fatty acid composition of milk triglycerides reflects differences in maternal diet. Medium-chain (C8–14) fatty acids are synthesized only in the mammary gland using glucose (or acetate in ruminants) as a substrate, whereas long-chain fatty acids are derived from the plasma. Nigerian women who have high-carbohydrate, low-fat diets have significantly more medium-chain fatty acids in their milk than Western women who consume a high-fat diet (Table 2).

Transcytosis Pathway (III)

Transport of proteins and other macromolecules by transcytotic pathways involves endocytic uptake of substances at the basal membrane, formation and maturation of endosomes, and sorting to lysosomes for degradation or to the apical recycling compartment for exocytosis at the apical membrane. The best-studied molecule in this regard is IgA. IgA is synthesized by plasma cells in the interstitial spaces of the mammary gland or elsewhere in the body and binds to receptors on the basal surface of the mammary alveolar cell; the entire IgA–receptor complex is endocytosed and transferred to the apical membrane, where the extracellular portion of the receptor is cleaved and secreted together with the IgA. It is thought that many other proteins, hormones, and growth factors that find their way into milk from the plasma are secreted by a similar mechanism.

Transmembrane Pathway (IV)

Transport processes for sodium, potassium, and chloride exist on the basal and apical plasma membranes of alveolar epithelial cells. Uptake mechanisms for calcium, phosphate, and iodide, however, are thought to be limited to the basal membrane. The mammary

Table 2 Major fatty acids of human and bovine milk (wt%)

Fatty acid	*Human milk*		*Bovine milk*
	Western diet	*Nigerian diet*	
Saturated fatty acids			
Medium and intermediate chain (formed in the mammary gland)			
8:0, octanoic acid	0.46		1.3
10:0, decanoic acid	1.03	0.54	2.7
12:0, lauric acid	4.40	8.34	3.0
14:0, myristic acid	6.27	9.57	10.6
Long chain			
16:0, palmitic acid	22.0	23.35	28.2
18:0, stearic acid	8.06	10.15	12.6
Monounsaturated fatty acids			
16:1 n-7 (*cis*), palmitoleic acid	3.29	0.91	1.6
18:1 n-9 (*cis*), oleic acid	31.3	18.52	21.4
18:1 n-9 (*trans*), oleic acid	2.67	0.86	1.7
Polyunsaturated fatty acids (PUFA) (essential fatty acids)			
18:2 n-6, linoleic acid	10.76	11.06	2.9
18:3 n-3, linolenic acid	0.81	1.41	0.3
Long-chain PUFA (n-6)			
18:3 n-6, γ-linolenic acid	0.16	0.12	2.9
20:2 n-6,	0.34	0.26	0.03
20:3 n-6, dihomo-γ-linolenic acid	0.26	0.49	0.1
20:4 n-6, arachadonic acid	0.36	0.82	0.2
Long-chain PUFA (n-3)			
20:5 n-3, eicosapentenoic acid	0.04	0.48	0.08
22:5 n-3	0.17	0.39	
22:6 n-3, docashexenoic acid	0.22	0.93	0.09

epithelial cells possess a GLUT1 glucose transporter and a sodium-dependent glucose transporter. The GLUT1 transporter is thought to mediate glucose transport at the basal and Golgi membranes, but it does not contribute to glucose transport at the apical membrane. Both sodium-dependent and sodium-independent amino acid transport mechanisms analogous to those found in other organs are located in the basolateral component of the mammary epithelium. It is unclear whether apical membranes have similar transport mechanisms for amino acids, and it is unknown how amino acids enter milk.

Paracellular Transport Pathway (V)

Pathway V (Figure 2) involves the passage of substances between epithelial cells rather than through them, and for this reason it is designated the paracellular pathway. During full lactation the passage of even low-molecular-weight substances between alveolar cells is impeded by the gasket-like tight junction structures that join the epithelial cells tightly, one to another. During pregnancy, with mastitis and after involution, the tight junctions become leaky and allow components of the interstitial space, such as sodium and potassium, to pass unimpeded into the milk, which is sometimes useful in diagnosing breast-feeding problems.

Regulation of Milk Synthesis, Secretion, and Ejection

Milk volume production is a primary indicator of lactational function; the most precise methods for measuring the volume of milk produced involve weighing infants before and after each feed for 24 h or longer or using an isotope dilution technique with stable isotopes. Clinically, the amount of milk that can be expressed with a breast pump or the change in infant weight after a single feed can be used as a rough index. The volume of milk secreted by women exclusively breast-feeding a single infant at 6 months postpartum is remarkably constant at approximately 800 ml day^{-1} in populations throughout the world. Mothers of twins, and occasionally even triplets, are able to produce volumes of milk sufficient for complete nutrition of their multiple infants, and studies of wet nurses indicate that at least some women are capable of producing up to 3.5 l of milk per day. However, if infants are supplemented with foods other than breast milk, milk secretion is proportionately reduced. This point is illustrated in Figure 3, which shows that milk volumes gradually decline during weaning and increase as the feeding frequency increases. These observations illustrate the important principle that the volume of milk secretion in lactating women is regulated by infant demand. If milk cannot be removed from the breast, local mechanisms cause an inhibition of milk secretion and downregulation of milk synthetic machinery. With partial removal of milk on a consistent basis, these local factors adjust milk secretion to a new steady-state level. If milk removal ceases for extended periods, involution sets in and the gland loses its competency to secrete milk.

Hormonal Control of Milk Synthesis and Secretion

In most species, the presence of high levels of plasma prolactin appears to be essential for lactation. In rats, the ergot alkaloid bromocriptine (an inhibitor of prolactin release from the pituitary) inhibits lactation, and in women it inhibits the onset of lactation when given in appropriate doses. How prolactin influences lactation is not known in any detail. However, it appears to promote both the survival of mammary epithelial cells and the synthesis of macromolecular milk components. In addition, prolactin is an osmoregulator in some species of fish, birds, and amphibians and may function to maintain solute transport in the mammary gland. Maintaining high calcium levels in milk is also dependent on parathyroid hormone-related protein-dependent mobilization of calcium from maternal stores.

Local Control of Synthesis and Secretion

The neurotransmitter, serotonin, functions as an intrinsic homeostatic regulator of lactation, reducing milk production when production rates exceed removal rates, such as during weaning. Serotonin has been shown to be synthesized by mammary epithelial cells in mice and cattle and is found in milk. Elevated levels of serotonin have been demonstrated to directly reduce milk production and inhibit the expression of genes encoding milk protein components. In humans, elevated serotonin levels associated with the use of serotonin reuptake inhibitor-based antidepressants have been shown to delay secretory activation. Understanding this regulatory system may be very important in helping women to increase or maintain their milk supply, particularly in the postpartum period.

Regulation of Milk Ejection

When the infant is suckled, afferent impulses from sensory stimulation of nerve terminals in the areolus travel to the central nervous system, where they promote the release of oxytocin from the posterior pituitary. This neuroendocrine reflex can be conditioned, and in women, oxytocin release is often associated with stimuli such as the sight or the sound, or even the thought, of the infant. The oxytocin is carried through the bloodstream to the mammary gland, where it interacts with specific receptors on myoepithelial cells, initiating their contraction and expelling milk from the alveoli into the ducts and subareolar sinuses. The passage of milk through the ducts is facilitated by longitudinally arranged myoepithelial cell processes whose contraction shortens and widens the ducts, allowing free flow of milk to the nipple. Milk is removed from the nipple not so much by suction as by the stripping motion of the tongue against the hard palate. This motion carries milk through the teat into the baby's mouth. The letdown response is

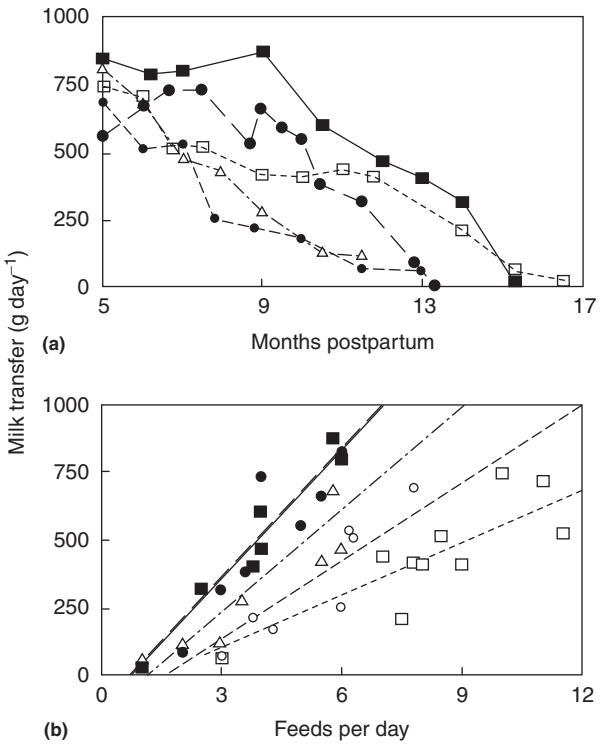


Figure 3 Changes in milk volume during weaning and in response to increased feeding frequency. (a) Milk volume transfer as a function of time postpartum. (b) Relation between feeding frequency (feeds per day) and the milk volume. Data from five breast-feeding dyads; each symbol represents an individual dyad. Reproduced from Neville MC, Allen JC, Archer PC, *et al.* (1991) Studies in human lactation: milk volume and nutrient composition during weaning and lactogenesis. *The American Journal of Clinical Nutrition* 54: 81–92.

decreased by psychological stress or pain, which interferes with oxytocin release. Oxytocin also appears to be involved in regulating maternal behavior in laboratory animals and may play a similar role in humans.

Initiation of Lactation

Pregnancy transforms the mammary gland from a simple ductal tree into a highly efficient exocrine organ with expansive lobuloalveolar structures. This transformation is hormonally regulated and involves changes in the cellular composition of the mammary gland and alterations in the structural, cellular, and biochemical properties of alveolar cells that are critical to the development of efficient solute transport and secretory functions. Alveolar epithelial cells begin to differentiate into secretory cells at midpregnancy in most species. The differentiation process occurs heterogeneously and has been divided into initiation and activation phases based on differences in the composition of mammary secretions, gene expression, and the structural and functional properties of alveolar cells. Alveolar cells become capable of limited secretion of some milk components during the initiation phase, which in humans is detected by measurement of increased concentrations of lactose and α-lactalbumin in the plasma. Copious milk secretion, however, is induced during the secretory activation phase (sometimes called lactogenesis II) that occurs in response to the decrease in serum progesterone levels. In rodents and ruminants, this decrease is closely associated with parturition; in humans, it occurs after parturition.

Changes in Milk Composition during Secretory Activation

Secretory activation is reflected in dramatic modifications of the solute composition of milk and increased secretory volume, which in turn reflect the maturation of secretory mechanisms and transport pathways during this period. In women, there are three temporally distinct changes in milk composition at the onset of lactation. The earliest is a decrease in sodium and chloride concentrations and an increase in the lactose concentration of milk (Figure 4). These modifications occur immediately after delivery and are largely complete by 72 h postpartum. They precede increases in milk volume by at least 24 h and can be explained by closure of the tight junctions that block the paracellular pathway. Blocking this pathway prevents lactose, generated by the epithelial cells, from passing from the lumen of the alveolus to the plasma, and it prevents sodium and chloride from directly entering the lumen from the interstitial space. These changes result in the reduction of sodium and chloride and an elevation of lactose concentrations in the

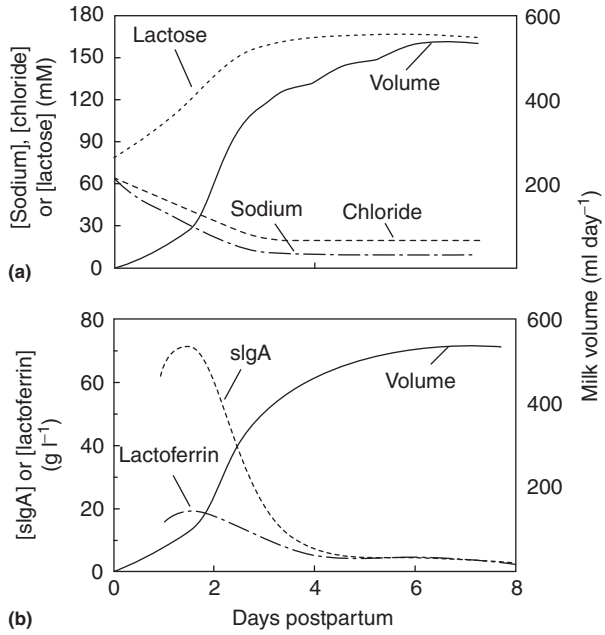


Figure 4 Changes in milk composition and volume in women during secretory activation and early lactation. Reproduced from McManaman JL and Neville MC (2003) Mammary physiology and milk secretion. *Advanced Drug Delivery Reviews* 55: 630–641.

mammary secretion. The increased lactose concentration is reflective of decreased water entering the lumina as monovalent ion secretion decreases rather than increasing the lactose secretion rate. Secondarily, the rates of secretion of sIgA and lactoferrin into milk of women are elevated soon after delivery. The concentrations of these two important protective proteins remain high, comprising as much as 10% of milk, for the first 48 h after birth. The concentration of each protein diminishes rapidly after day 2, both from dilution as milk volume secretion increases and from actual reduction in their rates of secretion, particularly of immunoglobulins.

Although both these proteins are found at high concentrations in colostrum, they are likely to be secreted by different mechanisms; lactoferrin, an endogenous protein of alveolar cells, is secreted by the exocytotic pathway (pathway I), whereas sIgA, a plasma-derived protein, is secreted by receptor-mediated transcytosis (pathway III). In addition, the peak secretion rate of lactoferrin occurs at the same time as that of lactose and the major milk proteins, whereas sIgA secretion peaks 1 day earlier, indicating the possibility that the exocytotic and transcytosis pathways are regulated differently during early lactation. The third phase occurs approximately 36 h postpartum and is associated with massive and concerted increases in milk volume and the rates of synthesis and secretion of almost all the components of mature milk, including, but not limited to, lactose, protein (mainly casein), lipid, calcium, sodium, magnesium, potassium, citrate, glucose, and free phosphate. Considering that the secretion of these substances involves the actions of several distinct transport pathways and biosynthetic processes, such tightly synchronized increases imply the presence of a common activation switch for coordinating their activities.

Hormonal Control of Secretory Activation

The decrease in progesterone around parturition is generally agreed to be required for the onset of milk secretion. In humans, it is known that removal of the placenta, the source of progesterone, is necessary for the initiation of milk secretion. In swine, timing of the increase in milk lactose correlates closely with timing of the decrease in plasma progesterone at parturition. Exogenous progesterone prevents lactose and lipid synthesis in mammary glands of pregnant rats and sheep after the removal of their ovaries, the source of progesterone in these species. Progesterone also suppresses β-casein expression in the rat mammary gland during pregnancy and the decrease in progesterone levels is linked to increased β-casein synthesis at parturition. Receptors for progesterone are not detected in lactating mammary tissues, which explains why progesterone does not inhibit established lactation. It is likely that the decline in progesterone is insufficient to activate secretion and that the actions of other hormones, including prolactin and glucocorticoids, are necessary to complete this process. In all *in vitro* mammary systems, insulin and corticoids, in addition to prolactin, are necessary to maintain the synthesis of milk components. Furthermore, cortisol replacement is required for maintenance of milk production in adrenalectomized animals. An early notion that a surge of glucocorticoids is the initiator of lactation is likely incorrect because the increase in cortisol seen in unanesthetized women associated with the stress of labor is complete by the time milk volume begins to increase to any extent. Because secretory activation proceeds at parturition in severely diabetic rats, a role for insulin in lactogenesis as opposed to metabolic adjustments during lactation seems improbable. In summary, the most

reasonable interpretation of the data from both animal and human studies is that the hormonal trigger for lactogenesis is a decrease in progesterone in the presence of maintained prolactin. Because postpartum prolactin levels are similar in both breast-feeding and non-breast-feeding women, the basic process appears to be initiated whether or not breast-feeding occurs. The caveat, of course, is that the mammary epithelium must be sufficiently prepared by the hormones of pregnancy to respond with milk synthesis.

Delays in Secretory Activation

A delay in the onset of milk secretion is a problem for the initiation of breast-feeding in a significant number of parturient women. A number of pathological conditions may delay secretory activation in women, including cesarean section, diabetes, obesity, and stress during parturition. The role of cesarean section is controversial, but if there is one, it is likely to have only a modest effect. However, poorly controlled diabetes, stress from delivery, or obesity are associated with significant decreases in early milk production. Because each of these conditions is related to higher blood glucose, hyperglycemia may be an underlying factor in the delay in lactation. However, once it is established, women with diabetes do not have a problem in maintaining lactation. Thus, compensatory factors may override initiation defects to ensure infant nutrition in these disorders.

Further Reading

Bissell, MJ, Polyak, K, Rosen, J (Eds.), 2011. The Mammary Gland as an Experimental Model. Cold Spring Harbor Press, New York, NY.
Jenson, RG, 1995. The Handbook of Milk Composition. Academic Press, London.
McManaman, JL, Reyland, ME, Thrower, EC, 2006. Secretion and fluid transport mechanisms in the mammary gland: Comparisons with the exocrine pancreas and the salivary gland. Journal of Mammary Gland Biology and Neoplasia 11, 249–268.
Neville, MC, Daniel, CW, 1987. The Mammary Gland: Development Regulation and Function. Plenum Press, New York, NY.
Neville, MC, Neifert, MR, 1983. Physiology, Nutrition and Breastfeeding. Plenum Press, New York, NY.
Peaker, M, 1977. Comparative Aspects of Lactation. Academic Press, London.
Richert, MM, Schwertfeger, KL, Ryder, JW, Anderson, SM, 2000. An atlas of mouse mammary gland development. Journal of Mammary Gland Biology and Neoplasia 5, 227–241.
Shennan, DB, Peaker, M, 2000. Transport of milk constituents by the mammary gland. Physiological Reviews 80, 925–951.

Low birth weight and preterm infants: Nutritional management

H. Tymann and H. Lochen, Johns Hopkins Children's Hospital, Baltimore, MD, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of J.L. Bosarge, Low Birth Weight and Preterm Infants: Nutritional Management, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 104–110, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00174-4.

Introduction 108
Growth 108
Energy needs 108
Parenteral nutrition 109
Components of PN 109
Glucose 109
Protein 110
Cysteine 110
Lipids 110
Carnitine 111
Electrolytes 111
Calcium, phosphorus, and magnesium 111
Trace minerals 111
Vitamins 111
Enteral nutrition 111
Trophic feedings 112
Feeding route 113
Feeding selection 113
Feeding delivery 113
Monitoring feeding tolerance 114
Vitamin, mineral, electrolyte supplementation 114
Monitoring nutritional status 114
Conclusions 115
References 115
Further reading 115

Key points

- Premature infants are high nutrition risk as they have minimal nutrient stores and are born during a time of accelerated growth
- Inadequate nutrition and growth failure during the neonatal period negatively affect cognitive and developmental outcomes
- Immature gastrointestinal tracts, medical instability, and inability to take food by mouth can make the preterm infant dependent on nutrition support
- Most preterm infants will require parenteral nutrition as a bridge until full enteral feeds are established

Glossary

Low birth weight infant Infant born weighing less than 2500 g; very low birth weight infants weigh less than 1500 g at birth; extremely low birth weight infants weigh less than 1000 g at birth

Necrotizing enterocolitis (NEC) A spectrum of disease characterized by ischemic necrosis to the mucosal barrier of a neonate's intestines, often associated with significant inflammation, pneumatosis, and possible rupture or perforation of the intestine.

Parenteral nutrition Intravenous infusion of nutrients and fluid

Premature infant Infant born before completing 37 weeks gestation

Trophic feedings Non-nutritive quantities (10–20 mL kg^{-1} day^{-1}) of enteral feeds given for a predetermined time (3–7 days) before beginning the advancement to full enteral feeds

Introduction

Infants born prematurely have numerous nutritional risk factors. Nutrient stores are accumulated during the third trimester of pregnancy; therefore, preterm infants have minimal energy, protein, fat, vitamin, and mineral reserves. The infant may also be a product of a pregnancy complicated by diminished uterine blood flow, thus further compromising the infant's nutrient stores at birth. The metabolic rate of the preterm infant is elevated due to the predominance of metabolically active tissue and minimal fat stores. Protein, fat, and glucose needs are very high to provide adequate energy for metabolism, fat deposition, and growth. The preterm infant has excessive evaporative losses and increased urinary losses, which increase fluid needs significantly compared to late preterm or term infants. The gastrointestinal tract of the preterm infant is very immature with delayed production of enzymes (Mobassaleh et al., 1985; Shulman et al., 1986), poor gastric emptying, and uncoordinated peristalsis. To further complicate the provision of nutrients, preterm infants have episodes of metabolic instability, including hypo- and hyperglycemia, poor lipid clearance, and electrolyte disturbances. The preterm infant also has high rates of stressful events, including respiratory distress, hypoxemia, hypercarbia, and sepsis. Infants born prematurely require nutrients to be delivered through parenteral and enteral nutrition support, due to the fact that they have an immature gastrointestinal tract and lack the ability to take food by mouth. Moreover, early nutrition support and growth in the early neonatal period have been associated with improved developmental and health outcomes.

Growth

Similar to full term infants, preterm and low birth weight infants lose 8–15% of their body weight in the first 4–6 days of life due to diuresis, and subsequently take approximately 2–3 weeks to regain birth weight. The smallest infants lose the greatest percentage of weight and subsequently can take the longest time to return to birth weight (Bell and Acarregui, 2008). The goal of nutrition support is to provide sufficient nutrients to achieve fetal growth rate. **Table 1** shows the goal growth velocity based on birth weight. Adequate and optimal growth remains a struggle for preterm infants due to their complex medical needs. Most preterm and low birth weight infants show significant delays in growth due to the inability to provide adequate nutrients, due to fluid status and medical instability, especially in the first few weeks following birth. Over the past several years, improvements in neonatal management and a more aggressive approach to nutrition have accelerated growth, but it still lags behind the fetal growth rate. Growth velocity in the infant is the greatest between 25- and 30-weeks gestation. If the infant is undernourished after birth, during this key period of accelerated growth, an increased supply of nutrients may be necessary to achieve catch-up growth to prevent the poor outcomes associated with postnatal growth failure. While protein and energy are key nutrients for growth, vitamins, minerals, and electrolytes must also be supplied in adequate amounts to contribute to optimal growth. For example, despite calorie and protein intake, insufficient sodium and zinc intake may lead to insufficient weight gain and linear growth (Alshaikh et al., 2021; Segar et al., 2018). Preterm infants require parenteral nutrition to supply nutrients with the gradual transition to a combination of parenteral and enteral nutrition, and finally to full enteral nutrition. The preterm infant is at highest risk for growth failure during the transition period. Extra attention needs to be made to ensure appropriate protein, energy and nutrients are supplied during this phase of nutrition.

Energy needs

Several factors impact the energy needs and metabolic rate of the preterm infant, these factors are summarized in **Table 2**. Resting metabolic rate accounts for the greatest percentage of energy needs. Resting metabolic rate is equivalent to basal metabolic rate plus some of the energy used for growth; estimates have ranged from 45 to 60 kcal kg^{-1} day^{-1} (188–251 kJ kg^{-1} day^{-1}) (AAP, 2009). The energy cost of activity ranges between 2% and 12% of the total energy expenditure. The smaller, more premature infants are at the lower end of the range whereas the older preterm infant has increased activity and therefore a higher expenditure. Although preterm infants are cared for in a thermoneutral environment within an isolette, energy is still lost to thermoregulation during

Table 1 Expected growth velocity.

Weight (g)	*Goal weight gain (g kg^{-1} day^{-1})*
500–700	21
700–900	20
900–1200	19
1200–1500	18
1500–1800	15
1800–2200	16

Adapted with permission from Table 1 in Ziegler et al. (2002).

Table 2 Energy needs of the growing preterm infant.

Energy factor	*Kcal kg^{-1} day^{-1} (EN vs. PN)*
Resting metabolic rate	50
Activity (0–30% above REE)	0–15
Thermoregulation	10
Thermic effect of food	10
Fecal losses	10
Energy storage (growth)	25–35
Total	85–130

AAP (2009).

nursing care and medical procedures. After infants are weaned from the isolette, there may also be energy lost to thermoregulation during bathing and feeding. The energy cost of growth includes tissue synthesis as well as the energy stored in tissues. The estimates for growth needs vary widely depending on the composition of weight gain in the infant. For the enterally fed infant, the thermic effect of food and fecal loss contribute to total energy need. The estimated calorie and protein intakes required to achieve fetal weight gain are listed in **Table 3**.

Parenteral nutrition

The gastrointestinal tract of the preterm infant is immature, requiring the gradual and/or delayed initiation of EN. During this period, enteral nutrition alone will not meet the nutritional needs of the preterm infant, thus requiring supplemental PN. Infants whose birth weights are less than 1500 g (g) are often at higher risk of delayed enteral nutrition; therefore, the preterm infant is at risk for long term PN dependence to prevent catabolism, maintain lean body mass, support metabolism, and achieve growth until adequate enteral feeds can be established.

In infants whose birth weight is between 1500 and 2000 g, may also require PN especially if the initiation or progression of enteral feeding is likely to be prolonged or delayed. Historically, PN was delayed for several days after birth due to metabolic instability of the infant and concern for tolerance of the components in the solution. More recently, the early use of PN has been recommended, ideally within 24 h after birth. This practice minimizes the interruption of nutrient delivery and the catabolism that occurs after preterm delivery.

PN can be administered by two different routes, a peripheral line or a central line. There are both risks and benefits associated with each route. The osmolarity is limited to less than 1000 osmoles per liter (Dugan et al., 2014). Peripheral line osmolarity limits can impede nutrient composition in PN. Peripheral lines also require vigilance on the part of staff to prevent infiltrates and line replacements. Central PN is recommended when it is anticipated that it will be used for greater than 5–7 days. Central access is intended for more long term management, however complications such as pneumothorax, pleural effusions, and increased risk of sepsis are associated with central lines. **Table 4** summarizes a comparison of peripheral PN and central PN.

Components of PN

PN solutions contain dextrose, amino acids, lipids, electrolytes, vitamins, and minerals.

Glucose

Glucose, provided as a dextrose solution, is the predominant energy source in PN. It is the main energy substrate for the fetus as well as the neonate after birth. Preterm infants often require more glucose than the term infant due to the higher brain to body weight ratio and the additional energy requirements for the central nervous system. Glycogen stores are very limited in the preterm and low

Table 3 Estimated calorie and protein intakes to achieve fetal weight gain.

Body weight (g)	*Energy (kcal kg^{-1} day^{-1}) (parenteral–enteral)*	*Protein (g kg^{-1} day^{-1}) (parenteral–enteral)*	*Protein/energy (g 100 kcal^{-1}) (parenteral–enteral)*
500–700	89–105	3.5–4	3.9–3.8
700–900	92–108	3.5–4	4.1–3.7
900–1200	101–199	3.5–4	3.5–3.4
1200–1500	108–127	3.4–3.9	3.1–3.1
1500–1800	109–128	3.2–3.6	2.9–2.8

Adapted with permission from Table 1 in Ziegler et al. (2002).

Table 4 Risks and benefits of parenteral nutrition routes.

Peripheral	*Central*
• Adequate for short-term use • Dextrose limited to 10–12.5% • Can provide 80–85 kcal kg^{-1} day^{-1} if adequate fluid available • Possible complications • Intravenous line can infiltrate and cause deep skin sloughing • Requires nursing vigilance to care for intravenous line • Can require multiple intravenous attempts	• Recommended with PN required for greater than 5–7 days • Requires placement of central line/PICC • Able to meet estimated needs if adequate fluid available • Possible complications • Sepsis • Line complications: pleural effusions, pneumothorax

birth weight infant; therefore, they require a continuous source of glucose to prevent hypoglycemia. This should be initiated at a rate of 4–6 mg kg^{-1} min^{-1} (0.033 mmol kg^{-1} min^{-1}) and can be advanced 1–2 mg kg^{-1} min^{-1} (0.0055–0.011 mmol kg^{-1} min^{-1}) each day to an optimum of 12–14 mg kg^{-1} min^{-1} (0.066–0.78 mmol kg^{-1} min^{-1}) as long as the infant does not become hyperglycemic. Above this rate, glucose is not used for energy, but rather fat deposition, an inefficient process that can result in increased energy expenditure and carbon dioxide production (Mesotten et al., 2018).

Difficulties with glucose stability are a common problem in preterm infants due to decreased energy stores, increased gluconeogenesis due to stress, decreased insulin secretion, and insulin resistance. When hyperglycemia occurs, the glucose infusion rate should be decreased; however, the rate should not be decreased below 4–6 mg kg^{-1} min^{-1} (0.022–0.33 mmol kg^{-1} min^{-1}) as this is the minimum supply rate necessary to provide adequate energy to the brain. Usually, the infusion of amino acids improves glucose tolerance by decreasing glucose production, stimulating insulin secretion, and enhancing insulin action. The use of continuous insulin infusions to treat hyperglycemia is controversial, however may be consider to treat hyperglycemia when GIR is already limited.

Protein

The early administration of protein to the preterm infant is one of the changes that have occurred over the last decade. Early studies of amino acid administration in preterm infants in the 1960s and 1970s raised the concern for protein toxicity because these infusions were associated with acidosis, azotemia, and hyperammonemia, thus causing a delay in the routine administration of protein. However, the above conditions were most likely the result of the preparations being casein or fibrin hydrolyzates. In the late 1980s, crystalline amino acid solutions specifically for use in infants were designed to mimic the plasma amino acid level comparable to that of a postprandial breast-fed infant. TrophAmine is currently the recommended amino acid solution for use with preterm and low birth weight infants as it contains taurine, a semi-essential amino acid in premature infants. It has a higher ratio of essential to nonessential amino acids, thus resulting in improved nitrogen balance and protein synthesis. Moreover, TrophAmine has been shown to reduce the incidence and degree of PN-associated cholestasis, and has a lower pH, thus enhancing calcium and phosphorus solubility (Lenz and Mikrut, 1988; Wright et al., 2003).

The early administration of amino acids is crucial as the preterm infant suffers protein losses of between 0.9 and 1 g kg^{-1} day^{-1} starting as soon as the placental supply is cut off (Embelton, 2007). The infusion of amino acids, along with glucose, decreases protein catabolism and prevents negative nitrogen balance with as little as 1–1.5 g kg^{-1} day^{-1} (Embelton, 2007). It has also been shown that the infusion of 3 g kg^{-1} day^{-1} within the first 2 days of life results in increased protein synthesis, suppressed protein breakdown, and produced plasma aminograms similar to those of the breast-fed infant (Denne and Poindexter, 2007).

Therefore, protein should be started on the first day of life, ideally, in the first few hours of life once intravenous access is obtained, at 3 g kg^{-1} day^{-1} and advanced to 3.5–4 g kg^{-1} day^{-1} to achieve *in utero* accretion rates.

Cysteine

The amino acid cysteine is a conditionally essential nutrient in preterm infants because they have low cystathionase activity. Cystathionase, an enzyme, is necessary to convert methionine to cysteine. This amino acid is unstable in liquid solutions, so commercially available crystalline amino acid solutions do not contain cysteine. Plasma levels of cysteine are low in infants receiving cysteine-free PN. Cysteine hydrochloride is soluble and stable in aqueous solutions for a short period of time and often added to PN solutions when prepared. An additional advantage is that the addition of cysteine decreases the pH of the PN solution, which allows for the addition of more calcium and phosphorus.

Lipids

Lipids are the most concentrated source of calories in the PN solution. Traditionally, lipid emulsions were comprised of soy bean and safflower oil. Recently, the utilization of alternative and blended lipids in preterm infants has gained popularity. These alternative lipids vary in composition and have been shown to reduce PN related cholestasis. The two widely used alternative lipids are SMOFlipid and Omegaven. SMOFlipid is 4-oil lipid emulsion and is a blend of soybean oil, medium chain triglycerides (MCT), olive oil, and fish oil. It has been shown to reduce the incidence of PN-related cholestasis in the infant population (Diamond

et al., 2017; Rayyan et al., 2012). Omegaven is 100% omega-3 fatty acids from fish oil and has been show to reverse the effects of PN related cholestasis (Diamond et al., 2009; Park et al., 2015).

When using standard lipid emulsions, 20% emulsions are recommended for use in preterm and low birth weight infants as they contain less phospholipid than the 10% emulsion. Phospholipid interferes with the rate of lipid hydrolysis, leading to elevated serum triglycerides. Lipids are critical for central nervous system development and when infused with the PN solution, they may prevent phlebitis. Lipids are dosed to prevent essential fatty acid deficiency and as an energy source. Maximum lipid clearance occurs when lipids are infused over 24 h. Starting recommendations vary, but it is generally accepted to start with 1 g kg^{-1} day^{-1} within the first 24 h of life and advance to an optimum of 3 g kg^{-1} day^{-1}; however, safety of initiation at 3 g kg^{-1} day^{-1} has been demonstrated (Ibrahim et al., 2004). Preterm infants have optimal protein retention when approximately 30–40% of calories are provided as lipids. Plasma triglycerides can be used to monitor lipid clearance. It is generally accepted that levels below 250 mg dL^{-1} indicate adequate clearance. Lipoprotein lipase and hepatic lipase are the primary enzymes for clearance of intravenous lipid. These activities are inducible by low-dose heparin, which is usually present in central PN solutions. In infants with hypertriglyceridemia, the provision of 0.5–1 g kg^{-1} day^{-1} of lipid is adequate to prevent essential fatty acid deficiency when a standard soybean lipid emulsion is used. When using alternative lipids or blends the minimum dosing as well as well as tolerable dose varies, therefore imperative to check the recommended dose from the specific manufacturer.

Carnitine

Carnitine is necessary for the transport of long-chain free fatty acids into the inner mitochondrial membrane, and for oxidation of fatty acids in the mitochondria. Carnitine is considered a conditionally essential nutrient because the preterm infant has decreased carnitine synthesis capability and has low plasma and tissue concentrations. Carnitine is found in human milk and formula; therefore, deficiencies can develop in 6–10 days after birth in less than 34 weeks' gestation infants without enteral feedings. There is no clear consensus as to the benefit to adding it to PN. Its use should be considered in all infants less than 34 weeks' gestation, those receiving long-term PN without enteral feedings, and those with hypertriglyceridemia.

Electrolytes

Electrolytes are added to the PN solution when the infant undergoes initial diuresis and is losing electrolytes in urine. Hyper- and hyponatremia within the first 48 h of life usually reflects fluid status and not excess or suboptimal provision of electrolytes. Very immature infants and those on diuretics may require additional amounts to maintain normal plasma concentrations. Chloride and acetate need to be dosed based on electrolyte levels. The very young preterm infant may need a higher proportion of acetate due to urinary bicarbonate losses. Later, when chronic diuretics are used, a greater proportion of chloride may be needed.

Calcium, phosphorus, and magnesium

Calcium and phosphorus are added to PN solutions to meet the minimum need for bone mineralization and cellular function. Calcium and phosphorus are relatively insoluble in solution together, making it difficult to provide adequate levels of these minerals to meet the needs of the preterm infant. Cysteine and TrophAmine are utilized in PN solutions to increase the solubility and improve the provision of calcium and phosphorus in PN. During the third trimester the average accretion rate of calcium in the fetus is normally about 100–120 mg kg^{-1} day^{-1} and 50–65 mg kg^{-1} day^{-1} for phosphorus. This level of accretion is difficult to match in the extrauterine environment, more-so in PN dependent infants given restrictions with volume and solubility. Infants on prolonged PN may develop osteopenia and fractures. For optimal absorption and utilization of both calcium and phosphorus, a ratio of 1.3–1.7 mg of calcium per every 1 mg of phosphors is recommended. The usual dose of magnesium is 0.3–0.5 mEq kg^{-1} day^{-1} (0.3–0.5 mmol kg^{-1} day^{-1}).

Trace minerals

Zinc and copper deficiencies occurred in some preterm infants before these trace elements were routinely added to PN solutions. There is very little research that defines the parenteral requirements of trace minerals in preterm infants. The current recommendations for trace minerals are summarized in **Table 5**.

Vitamins

Like trace minerals, the recommendations for intake of vitamins are not based on randomized trials, but are based on the best information available. Infants receiving these parenteral intakes in **Table 6** do not develop deficiencies or evidence of excessive intake.

The suggested initiation and advancement of PN in the preterm infant is summarized in **Table 7**.

Enteral nutrition

The early initiation of enteral nutrition has been shown to benefit premature infants. The rate of feeding advancement and type of feeds may result in feeding intolerance or necrotizing enterocolitis (NEC). NEC is a major cause of morbidity and mortality in preterm infants. The incidence of this disease is estimated to be between 8% and 10% of preterm infants. The cause of NEC is still fully understood, however likely multifactorial, including enteral feeds, hypoxia, ischemia, patent ductus arteriosus, and infection. It is known that delayed enteral feeding has a negative effect on gastrointestinal structure and function. Lack of enteral nutrition can

Table 5 Suggested parenteral intakes of trace minerals.

Trace mineral	*μg kg^{-1} day^{-1}*
Zinc	400
Copper	20
Selenium	1.5–2
Manganese	1
Chromium	0.05–0.2

Task Force for the Revision of Safe Practices for Parenteral Nutrition, Mirtallo, J., Canada, T., Johnson, D., Kumpf, V., Petersen, C., Sacks, G., Seres, D., Guenter, P., 2004. Safe practices for parenteral nutrition. J. Parenter. Enteral Nutr. 28(6), p. S39–S70.

Table 6 Suggested parenteral intake of vitamins.

Vitamin	*Amount (kg day^{-1})*
Vitamin A (μg)	280–500
Vitamin E (mg)	2.8
Vitamin K (μg)	100
Vitamin D (IU)	400
Ascorbic acid (mg)	25
Thiamin (μg)	350
Riboflavin (μg)	150
Pyridoxine (μg)	180
Niacin (mg)	6.8
Pantothenate (mg)	2
Biotin (μg)	6
Folate (μg)	56
Vitamin B_{12} (μg)	0.3

Total dose should not exceed the amounts provided by 5 mL of reconstituted MVI Pediatric (Armor Pharmaceutical Co., Chicago, IL, USA): 700 μg vitamin A, 7 μg vitamin E, 200 μg vitamin K, 10 μg vitamin D, 80 mg ascorbic acid, 1.2 mg thiamin, 1.4 mg riboflavin, 1.0 mg pyridoxine, 17 mg niacin, 5 mg pantothenic acid, 20 μg biotin, 140 μg folic acid, and 1 μg vitamin B_{12}.
Task Force for the Revision of Safe Practices for Parenteral Nutrition, Mirtallo, J., Canada, T., Johnson, D., Kumpf, V., Petersen, C., Sacks, G., Seres, D., Guenter, P., 2004. Safe practices for parenteral nutrition. J. Parenter. Enteral Nutr. 28(6), p. S39–S70.

Table 7 Suggested initiation and advancement of parenteral nutrition for the preterm infant.

Component	*Initial*	*Advancement per day*	*Goal*
Dextrose (mg kg^{-1} min^{-1})	6–8	1–2	12–14
Protein (g kg^{-1} day^{-1})	2–3	1	3.5–4
Lipids (g kg^{-1} day^{-1})	1	1	3

induce gastrointestinal atrophy, depresses gut hormone secretion, and delays the maturation of gastrointestinal motility. There benefits of early enteral feeding includes endocrine adaptation, the accelerated maturation of gut motility patterns, the provision of luminal nutrients, and possible benefits to the immune system. In fact, early enteral nutrition may enhance feeding tolerance and may actually decrease the incidence of NEC.

Trophic feedings

A strategy that has been extensively studied since the late 1980s is trophic feeding, also referred to as minimal enteral nutrition or gut priming. This method involves giving the infant small volumes of feedings, approximately 10–20 mL kg^{-1} day^{-1}, for a period of 3–7 days before beginning to advance to full enteral feedings. The benefits found are greater energy intake, earlier attainment of full enteral feedings, improved growth, decrease in PN-related complications, reduced risk of infection, and earlier hospital discharge. Furthermore, infants who receive trophic feedings have no increased incidence of NEC (Oddie et al., 2021). Many clinicians have adapted variations of this practice; some with a shortened period of trophic feeds, others reserving this practice for the smallest, most preterm and liable infants while employing advancement of feeds in larger, more stable infants. Once minimal enteral nutrition has been established and the infant is stable enough to advance feedings, it is generally considered a safe practice to increase

feedings by 20–30 mL kg^{-1} day^{-1} while using PN for the balance of intake until an adequate enteral intake has been established and tolerated.

Feeding route

Preterm infants lack the ability to orally feeds due to their inability to coordinate sucking, swallowing, and breathing until 32–34 weeks gestation therefore tube feedings must be used. Jejunal feeding was a popular method for feeding infants during the 1970s to early 1980s. It was felt that this method would minimize the risk of reflux and aspiration. This method is now generally reserved for infants in whom reflux and aspiration is complicating chronic lung disease or those who have poor gastric emptying. Now, most infants are fed using an orogastric or nasogastric tube. Orogastric tubes are often used in the lowest weight infants as the feeding tube may occlude one naris and impair nasal breathing.

Feeding selection

Human milk is the preferred type of feeding because of its well-established health benefits for the infant as well as those who lactate. Human milk is nutritionally superior to infant formula due to its increased bioavailability and protective effects against several comorbidities including NEC. Human milk contains many biologic components such as immunoglobulins, cytokines, growth factors, hormones, and oligosaccharides. Benefits to the infant also include improvements in gastrointestinal function, digestion and absorption, cognitive and visual development, host defense as well as enhanced parental-infant bonding.

However, despite human milk being the optimal feeding choice for infants, it is insufficient to meet the nutritional needs of the growing preterm infant alone. Unfortified human milk does not supply sufficient quantities of protein, energy, fatty acids, minerals, and other micronutrients that are needed for preterm infants to achieve optimal fetal growth and adequate bone mineralization. There are fortifiers available that can be added to human milk to improve nutrient intake. The use of these fortifiers has been associated with improved intake of protein and subsequent nitrogen retention, improved intake of minerals, subsequent bone mineralization, and improved growth.

If human milk is not available, the feeding of choice becomes either donated human milk or preterm infant formulas. The use of donated human milk has been associated with decreased incidence of NEC compared to preterm infant formulas however is costly and depending on state may require hospitals to acquire a tissue license. Preterm infant formulas have greater protein content and are cow's milk whey predominantly. The carbohydrate is a mixture of lactose and glucose polymers, and the fat a mixture of both long-chain and medium-chain triglycerides for improved nutrient absorption. The concentration of minerals, electrolytes, and vitamins is increased to meet the estimated nutrient needs of the preterm infant when fed in an amount to provide 120 kcal kg^{-1} day^{-1}. Infants fed preterm infant formulas have improved growth over those fed term formula.

Feeding delivery

Infants can be enterally fed via either continuous infusions or bolus. There are risks and advantages for both feeding modalities. It is commonly believed that continuous feedings are better tolerated. Continuously fed infants have been found to take longer to achieve full volume feeds and have decreased growth when compared with bolus fed infants (Wang et al., 2020). Bolus feedings have been associated with improved gastric emptying, and more mature intestinal motility patterns in addition to being the more physiologic option (Lucas et al., 1986). Because it causes infants to have cyclical surges of gastric hormones and insulin, promoting gastrointestinal tract development. Bolus or intermittent feeding may be more beneficial for low-birth weight infants, It is difficult to compare feeding tolerance between continuous and bolus feeds infants due to differences in the criteria used, more studies are needed to determine the best feeding delivery method for premature infants.

Table 8 summarizes the indications for bolus vs. continuous enteral feeding.

Table 8 Review of feeding delivery methods via NG or OG.

Bolus	*Continuous*
Advantages: • Appropriate for most premature infants • Simple delivery with minimal equipment • Physiologic feeds, ideal for those starting to take food by mouth • Can be done at home Possible contraindications: • Malabsorption • Dumping • High aspiration risk • Frequent emesis	Advantages: • Generally, well tolerated • Increased absorption and reduce dumping for those with intestinal disease • Can reduce aspiration risk for those who are high risk Possible contraindications: • Poor weight gain for infants fed human milk given the separation and adherence of fat to the tubing • More equipment and room for malfunction

Monitoring feeding tolerance

Feeding tolerance among preterm infants must be closely monitored given the high risk of gastrointestinal complications such as NEC. The presence of gastric residuals is one factor that is frequently used, but because preterm infants have poor gastric emptying, amounts less than 50% of a previous feed should not be considered significant. Other indicators include increase in abdominal girth, the absence of active bowel sounds, the presence of blood in the stool, a change in the number or quality of stools, and the presence of emesis. A careful exam is warranted if these symptoms are present.

Vitamin, mineral, electrolyte supplementation

Preterm and low birth weight infants who are fed fortified human milk or preterm infant formulas may require supplemental doses of vitamins, minerals, or electrolytes. Most commonly supplemented include iron, vitamin D, calcium, phosphorus, and sodium chloride. The specific dosing for each supplement should be based on recommended intake, laboratory results, and current provision. Table 9 provides enteral nutrient requirements of commonly supplemented vitamins and minerals for preterm or low birth weight infants with comparison of the RDA/AI for infants aged 0–6 months.

	Espghan 2010[a]	*RDA/AI (age 0–6 months)*[b]
Vitamins		
Vitamin A (mcg/kg)	400–1100	400
Vitamin D (mcg/d)	20–25	5
Vitamin E (mg/d)	2.2–11	4
Vitamin K (mcg/kg)	4.4–28	2
Minerals		
Calcium	120–140 mg/kg	210 mg/d
Phosphorus	60–90 mg/kg	100 mg/d
Magnesium	8–15 mg/kg	30 mg/d
Sodium	69–115 mg/kg	120 mg/d
Potassium	66–132 mg/kg	400 mg/d
Chloride	105–177 mg/kg	180 mg/d
Trace elements		
Zinc	1.1–2 mg/kg	2 mg/d
Iron	2–3 mg/kg	0.27 mg/d
Fluoride	1.5–60 mcg/kg	0.01 mg/d

[a] Source For ESPGHN: Agostoni, C., Buonocore, G., Carnielli, V.P., De Curtis, M., Darmaun, D., Decsi, T., Domellöf, M., Embleton, N.D., Fusch, C., Genzel-Boroviczeny, O. and Goulet, O., 2010. Enteral nutrient supply for preterm infants: commentary from the European Society of Pediatric Gastroenterology, Hepatology and Nutrition Committee on Nutrition. J. Pediatr. Gastroenterol. Nutr. 50(1), pp. 85–91.

[b] Source For RDA/AI: National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Food and Nutrition Board; Committee to Review the Dietary Reference Intakes for Sodium and Potassium; Oria, M., Harrison, M., Stallings, V.A., (Eds.) Washington (DC): National Academies Press (US); 2019 Mar 5.

Monitoring nutritional status

The nutritional status and growth of the preterm infant should be monitored throughout the hospitalization. The fluid and caloric intake should be monitored daily, body weight should be recorded daily, length and head circumference measured weekly, and all three measurements plotted on standardized growth charts. If growth is inadequate, the volume, caloric density, or protein content

Table 9 Periodic monitoring of nutritional status.

Indicator	*Frequency*
Weight	Daily
Length	Weekly
Head circumference	Weekly
Electrolytes	Daily until stable, then 1–2 times weekly
Albumin	Weekly
Bili/transaminases	Weekly (while on PN)
Calcium, phosphorus, magnesium, alkaline phosphatase	Weekly
Hemoglobin/hematocrit	Weekly
Reticulocyte	

of feeds should be increased. Supplementation of vitamins, minerals, or electrolytes may also be introduced/increased to support growth as deficiencies of iron, zinc, and sodium can lead to poor growth despite calorie intake. Biochemical measurements should also be assessed periodically. Table 9 summarizes the recommended schedule for monitoring anthropometrics and biochemical measurements.

Conclusions

Preterm infants have specialized nutritional needs. Nutrition support must be initiated as soon as possible after birth to achieve optimal outcomes. It needs to be continuously assessed to ensure that the best possible nutritional support is provided to promote optimal growth and subsequent developmental outcomes.

References

Alshaikh, B., Zeed, M.A., Yusuf, K., Guin, M., Fenton, T., 2021. Effect of enteral zinc supplementation on growth and neurodevelopment of preterm infants: a systematic review and meta-analysis. J. Perinatol. 1–10.

American Academy of the Pediatric (AAP) Committee on Nutrition, 2009. Nutritional needs of the preterm infant. In: Kleinman, R.D. (Ed.), Pediatric Nutrition Handbook, sixth ed. AAP, Elk Grove Village/IL, pp. 79–112.

Bell, E.F., Acarregui, M.J., 2008. Restricted vs. liberal water intake for preventing morbidity and mortality in preterm infants. Cochrane Database Syst. Rev. (1).

Denne, S.C., Poindexter, B.B., 2007. Evidence supporting early nutritional support with parenteral amino acid infusion. Semin. Perinatol. 31 (No. 2), 56–60.

Diamond, I.R., Sterescu, A., Pencharz, P.B., Kim, J.H., Wales, P.W., 2009. Changing the paradigm: omegaven for the treatment of liver failure in pediatric short bowel syndrome. J. Pediatr. Gastroenterol. Nutr. 48 (2), 209–215.

Diamond, I.R., Grant, R.C., Pencharz, P.B., de Silva, N., Feldman, B.M., Fitzgerald, P., Sigalet, D., Dicken, B., Turner, J., Marchand, V., Ling, S.C., 2017. Preventing the progression of intestinal failure–associated liver disease in infants using a composite lipid emulsion: a pilot randomized controlled trial of SMOFlipid. J. Parenter. Enteral Nutr. 41 (5), 866–877.

Dugan, S., Le, J., Jew, R.K., 2014. Maximum tolerated osmolarity for peripheral administration of parenteral nutrition in pediatric patients. J. Parenter. Enteral Nutr. 38 (7), 847–851.

Embleton, N.D., 2007. Optimal protein and energy intakes in preterm infants. Early Hum. Dev. 83 (12), 831–837.

Ibrahim, H.M., Jeroudi, M.A., Baier, R.J., Dhanireddy, R., Krouskop, R.W., 2004. Aggressive early total parental nutrition in low-birth-weight infants. J. Perinatol. 24 (8), 482–486.

Lenz, G.T., Mikrut, B.A., 1988. Calcium and phosphate solubility in neonatal parenteral nutrient solutions containing Aminosyn-PF or TrophAmine. Am. J. Hosp. Pharm. 45 (11), 2367–2371.

Lucas, A., Bloom, S.R., Aynsleygreen, A., 1986. Gut hormones and "minimal enteral feeding". Acta Paediatr. Scand. 75, 719–723.

Mesotten, D., Joosten, K., van Kempen, A., Verbruggen, S., Braegger, C., Bronsky, J., Cai, W., Campoy, C., Carnielli, V., Darmaun, D., Decsi, T., 2018. ESPGHAN/ESPEN/ESPR/CSPEN guidelines on pediatric parenteral nutrition: carbohydrates. Clin. Nutr. 37 (6), 2337–2343.

Mobassaleh, M., Montgomery, R.K., Biller, J.A., Grand, R.J., 1985. Development of carbohydrate absorption in the fetus and neonate. Pediatrics 75 (1), 160–166.

Oddie, S.J., Young, L., McGuire, W., 2021. Slow advancement of enteral feed volumes to prevent necrotising enterocolitis in very low birth weight infants. Cochrane Database Syst. Rev. (8).

Park, H.W., Lee, N.M., Kim, J.H., Kim, K.S., Kim, S.N., 2015. Parenteral fish oil–containing lipid emulsions may reverse parenteral nutrition–associated cholestasis in neonates: a systematic review and meta-analysis. J. Nutr. 145 (2), 277–283.

Rayyan, M., Devlieger, H., Jochum, F., Allegaert, K., 2012. Short-term use of parenteral nutrition with a lipid emulsion containing a mixture of soybean oil, olive oil, medium-chain triglycerides, and fish oil: a randomized double-blind study in preterm infants. J. Parenter. Enteral Nutr. 36, 81S–94S.

Segar, D.E., Segar, E.K., Harshman, L.A., Dagle, J.M., Carlson, S.J., Segar, J.L., 2018. Physiological approach to sodium supplementation in preterm infants. Am. J. Perinatol. 35 (10), 994–1000.

Shulman, R.J., Kerzner, B., Sloan, H.R., Boutton, T.W., Wong, W.W., Nichols, B.L., Klein, P.D., 1986. Absorption and oxidation of glucose polymers of different lengths in young infants. Pediatr. Res. 20 (8), 740–743.

Wang, Y., Zhu, W., Luo, B.R., 2020. Continuous feeding vs. intermittent bolus feeding for premature infants with low birth weight: a meta-analysis of randomized controlled trials. Eur. J. Clin. Nutr. 74, 775–783.

Wright, K., Ernst, K.D., Gaylord, M.S., Dawson, J.P., Burnette, T.M., 2003. Increased incidence of parenteral nutrition-associated cholestasis with Aminosyn PF compared to trophamine. J. Perinatol. 23 (6), 444–450.

Ziegler, E.E., Thureen, P.J., Carlson, S.J., 2002. Aggressive nutrition of the very low birthweight infant. Clin. Perinatol. 29 (2), 225–244.

Further reading

Groh-Wargo, S., Sapsford, A., 2009. Enteral nutrition support of the preterm infant in the neonatal intensive care unit. Nutr. Clin. Pract. 24, 363–376.

Groh-Wargo, S., Thompson, M., Hovasi Cox, J., Hartline, J.V., 2000. Nutritional Care for High-Risk Newborns. Precept Press, Inc, Chicago, IL.

Tsang, R.C., Uauy, R., Koletzko, B., Zlotkin, S.H., 2005. Nutrition of the Preterm Infant, Scientific Basis and Practical Guidelines, second ed. Digital Educational Publishing, Inc, Cincinnati, OH.

Valentine, C.J., Puthoff, T.D., 2007. Enhancing parenteral nutrition therapy for the neonate. Nutr. Clin. Pract. 22, 183–193.

Low birth weight and preterm infants: Causes, prevalence, and prevention

M. Merialdi, WHO, Avenue Appia, Geneva, Switzerland
J.H. Requejo, Johns Hopkins Bloomberg School of Public Health, MD, USA

© 2013 Elsevier Ltd. All rights reserved.
This is an reproduction of M. Merialdi, J.H. Requejo, Low Birth Weight and Preterm Infants: Causes, Prevalence, and Prevention, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 100–103, ISBN 780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00173-2.

Introduction 116
Assessment of Gestational Age and Fetal Growth. Methods and Limitations 116
Causes 117
Health Consequences 117
Epidemiology 118
Prevention 118
Nutritional Interventions to Prevent Preterm Delivery 119
Nutritional Interventions to Prevent IUGR 119
Conclusions 120

Introduction

It is widely accepted that weight at birth is a key indicator of fetal and neonatal health, both for individuals and populations. The strong association between low birth weight and perinatal mortality and morbidity is well recognized, and there is growing evidence about the different determinants and health consequences of conditions resulting in low birth weight. Knowledge of these epidemiological associations progressively increased over the course of the last 100 years. In the USA, the practice of weighing babies at birth was introduced at the end of the nineteenth century when low-birth-weight babies were categorized as *premature* and usually left unattended with minimal or no intervention attempted to prevent their deaths. This practice and the incorporation of information on birth weight and gestational age into the birth certificate in the mid-twentieth century resulted in irrefutable evidence that *prematurity* was the most significant cause of infant deaths at national level.

Awareness of the importance of low birth weight as a predictor of infant mortality led to the findings suggesting that low birth weight could be due to a restriction of the normal process of fetal growth, to delivery before the full term of gestation, or through a combination of both factors. Based on this evidence, the World Health Organization (WHO) made a distinction between the condition of low birth weight (birth weight less than 2500 g) and prematurity (delivery at less than 37 completed weeks, i.e., 259 days). A further development was the concept of small for gestational age (SGA) that better describes babies affected by intrauterine growth restriction (IUGR). SGA is defined as a birth weight below the 10th percentile for a given gestational age based on a sex-specific reference population. Although these distinctions and definitions are commonly applied in developed countries, their use is more problematic in developing countries where information on gestational age is often nonexistent or unreliable. This data limitation represents a major obstacle to the development of effective prevention and treatment efforts because IUGR and preterm delivery have different determinants and prognoses, as well as different epidemiological distributions which vary by country and socioecomomic status.

Before discussing the causes, prevalence, and prevention of low birth weight it is important to understand how its two components (gestational age and fetal growth) can be correctly identified and quantified for epidemiological and clinical purposes, and the major methodological limitations in capturing this information.

Assessment of Gestational Age and Fetal Growth. Methods and Limitations

Preterm birth is defined as delivery before 37 completed weeks (259 days). To accurately differentiate between preterm and term delivery it is crucial to have a reliable estimate of gestational age. Sonographic determination is presently the most accurate method to estimate gestational age. When ultrasonography is not available, gestational age can be determined by patient's recall of the time of last menstrual period, physical examination of the size of the uterus, and examination of the neonate. These alternate methods can be used alone or in combination, but are often inaccurate.

Early pregnancy sonographic estimation of gestational age is also crucial for estimation of fetal growth in utero, which is assessed by evaluating the size of several fetal anatomical parameters and comparing those measurements with the normal ranges at specific gestational ages obtained from reference populations with growth that can be considered unaffected by pathological conditions. Alternatively, fetal growth can be assessed by the anthropometrical evaluation of the neonate. Several classification systems have been proposed for newborn birth weight. The simplest categorizes newborns <2500 g as having a low birth weight, but this

classification does not enable differentiation between infants born SGA and infants who are small because they are born preterm. A second classification system based on reference charts of birth weight at different gestational ages groups infants into the categories of SGA, adequate for gestational age (AGA), and large for gestational age (LGA). Because these categories are based on percentile distributions of a reference population, a proportion of normal, constitutionally small, newborns in the lower tail of the normal fetal growth distribution will be miscategorized as growth restricted. The interpretation of the birth weight data using this system is also complicated by inaccuracies in the estimation of gestational age at delivery and by the pathological processes that could affect the size of infants born early in gestation.

Causes

Low birth weight results from either IUGR or preterm delivery, and, in some cases, from a combination of the two. These two conditions are likely caused by various and possibly independent etiopathological factors.

The definitive etiology of preterm delivery has not yet been determined, making it difficult to identify women at risk and to develop and implement effective preventive strategies. Available evidence shows that a complex range of factors such as pregnancy complications, health care practices, and socioeconomic conditions are implicated in preterm births. Preeclampsia, fetal distress, fetal growth restriction, abruptio placenta, fetal death, placenta previa, and multiple gestations, for example, are all associated with preterm delivery, either spontaneous or induced. Developments in obstetric and neonatal care and the consequent increase in obstetric interventions including infertility treatments have been linked with the increase in the rates of preterm delivery observed in recent years. Psychological stress and other socioeconomic factors such as poor nutrition, cigarette smoking, alcohol and drug abuse, young maternal age, poverty, and short stature have also been found to be possible causes. Genetic factors are likely to be involved in the etiopathogenesis of preterm delivery given that the condition tends to recur in families and that prevalence varies across races. The possible role of infection in triggering preterm delivery has been suggested by several studies, which show associations between preterm delivery and amniotic fluid and chorioamniotic infection, bacterial vaginosis, genitourinary clamydial infection, and periodontal disease. Despite the biological plausibility of these associations, their causal relationship has not been definitely proved by unequivocal scientific evidence.

Conditions associated with IUGR include but are not limited to fetal infections, congenital malformations, chromosomal abnormalities, chemical teratogenes, vascular disease such as preeclampsia, chronic renal disease, chronic hypoxia, placental and cord abnormalities, and multiple fetuses. Present knowledge of IUGR is limited by the challenges of differentiating between constitutional and environmental determinants of fetal growth. This limitation complicates the investigation of the role of maternal size and genetic factors in IUGR. Small women tend to have smaller babies. There is evidence that intergenerational effects on birth weight are transmitted through the maternal line, suggesting a genetic effect. However, poor maternal nutrition and social deprivation have also been proven to be related to small maternal size and impaired fetal growth. Similarly, the relationship between fetal size and race may be mediated by a combination of genetic and environmental factors. Carefully designed studies are needed to better determine the contribution of genetic and environmental determinants to the process of fetal growth.

Health Consequences

Low birth weight, either due to preterm delivery or IUGR is associated with poor neonatal health outcomes including higher rates of mortality. Neonatal mortality levels are indirectly associated with gestational age at delivery and birth weight.

Preterm birth is one of the major causes of neonatal mortality and morbidity. Of the estimated 8.795 million deaths in children younger than 5 years worldwide in 2008, 41% (3.575 million) occurred in neonates, and the most important single cause was preterm birth complications (12%, 1.033 million, UR 0.717 million–1.216 million). Mortality rates due to preterm birth are correlated with the overall level of neonatal mortality in a specific country. In low resource countries with high neonatal mortality rates (>45 neonatal deaths per 1000 live births), preterm birth is responsible for approximately 20% of all neonatal deaths with the other 80% attributable mainly to infections and birth asphyxia. In more developed countries, where neonatal mortality rates are below 15 deaths per 1000 live births, preterm birth is the cause of up to 40% of neonatal deaths since deaths due to infection and birth asphyxia are largely prevented. Although the proportion of neonatal deaths attributable to preterm birth is higher in developed countries, the majority of preterm birth related deaths occur in low resource settings because of lack of access to preventive and therapeutic interventions. This discrepancy between rich and poor countries is also reflected in the mortality differentials in late preterm births (between 32 and 37 weeks). Although late preterm infants in developed countries have a survival rate similar to full-term infants, the chance of survival of such infants in low resource settings is minimal.

The negative effects of preterm delivery and IUGR often persist throughout infancy and childhood, impacting the individual child, the entire family, the health care system, and society in general. Studies have also shown a relationship between low birth weight and an increased risk of cardiovascular disease, high blood pressure, obstructive lung disease, diabetes, high cholesterol concentrations and renal damage, indicating that the effects of low birth weight may extend into adulthood.

Epidemiology

Aggregate data on low birth weight rates show marked differences in underlying causes between geographical areas. Most low birth weight infants born in developed countries are due to preterm delivery, whereas a substantial proportion of low birth weight in developing countries is related to IUGR.

The first global and regional estimates of preterm birth rates were recently published. We now know that approximately 13 million newborns are born prematurely every year. This figure represents 9.6% of all births and is a conservative measure as it is based on estimates of the risk in normal pregnancies. **Figure 1** shows the estimated rates in the six major regions of the world. The highest rates are in North America and Africa. The greatest absolute numbers of preterm births (**Figure 2**) occur in Africa and Asia, the two world regions with the highest number of births and fertility rates.

These global and regional estimates mask important disparities between and within countries. In the US, for example, African and American women are more than twice as likely to deliver preterm than Caucasian women, a difference that accounts for a major proportion of the variation in infant mortality between the two racial groups.

Preterm birth rates are increasing in many developed countries. Although the causes of this trend are not fully elucidated, changes in clinical practices such as increased use of assisted reproductive techniques, and the rising rates of cesarean sections and induced deliveries have been associated with increases in preterm deliveries.

Prevention

Results from clinical trials provide the most powerful scientific evidence to guide policy and program development and implementation. Interventions aimed at preventing low birth weight targeted at preterm delivery and IUGR have not proven to be effective by randomized clinical trials. The multicausal nature of these conditions is likely responsible for single interventions not showing an effect of enough magnitude to be detected by medium sized clinical trials. Thus appropriate combinations of interventions should be a priority for evaluation in the context of large, methodologically sound trials. Available evidence shows that some interventions may be effective and their combined implementation may have a significant public health impact. Interventions likely to be beneficial to prevent IUGR are smoking cessation, antimalarial chemoprophylaxis in primigravide women, and balanced protein energy supplementation. Treatment of urinary tract infection, placement of circumferential stitches on a structurally weak uterine cervix (cerclage), and treatment of bacterial vaginosis in high-risk women have been shown to be effective in preventing preterm birth. These interventions are applicable only to a small number of high risk women and their overall effect on the general population is likely to be limited.

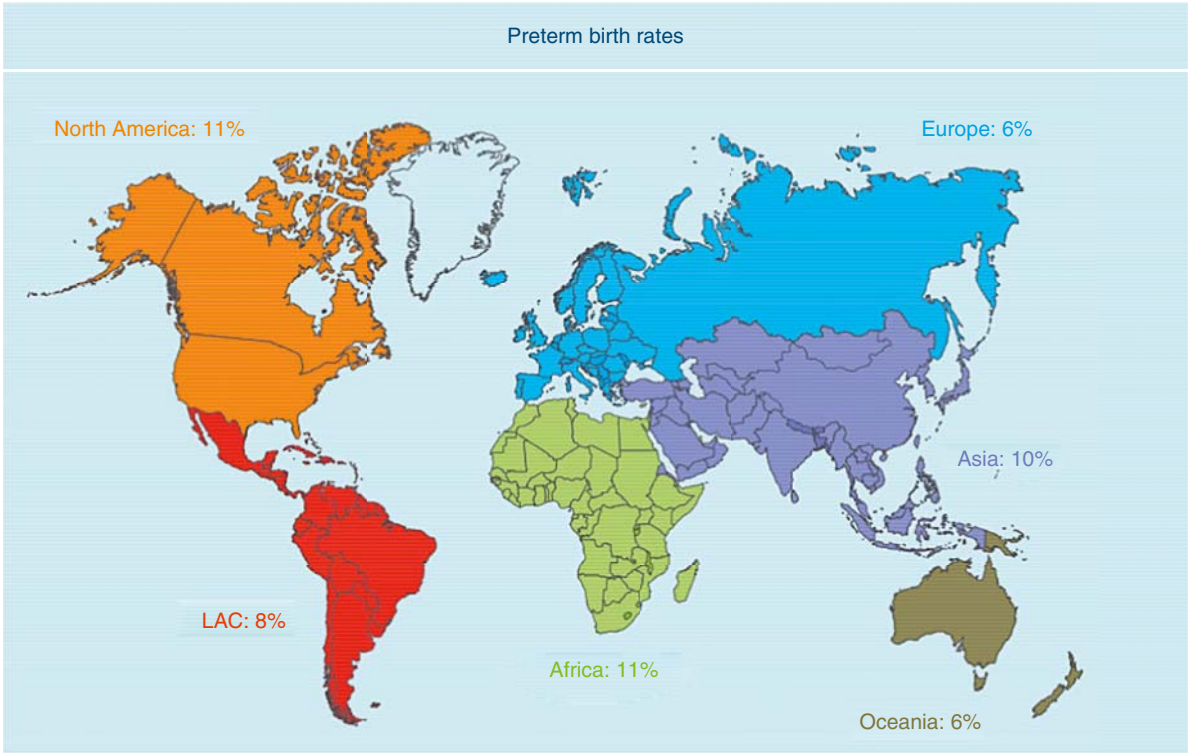


Figure 1 Preterm birthrates.

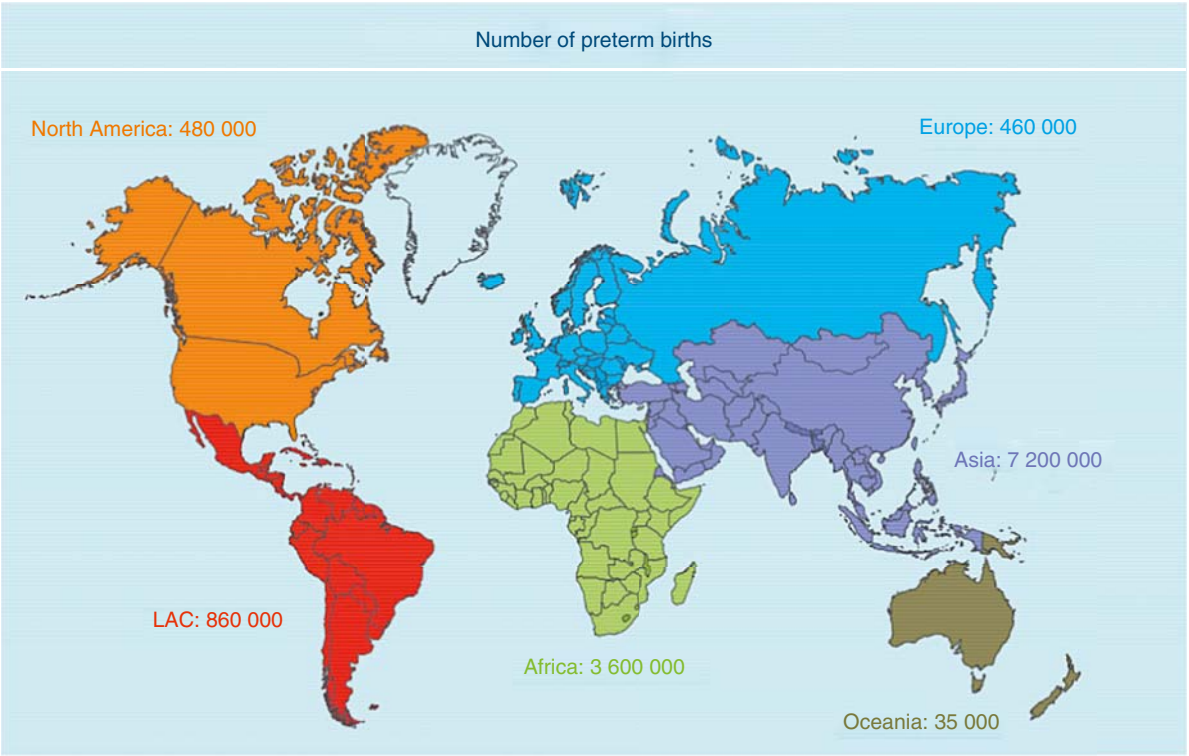


Figure 2 Number of preterm births.

In the next paragraphs, nutritional interventions to prevent preterm delivery and IUGR will be reviewed with the aim of identifying potentially effective interventions and suggesting possible mechanisms that may explain the link between maternal nutritional status and low birth weight. The focus will be on the review of randomized clinical trials which provide the most unbiased epidemiological evidence on the effectiveness of interventions. Clinical trials testing the same or similar interventions can be pooled together to estimate an overall effect by means of systematic reviews of published and unpublished studies and meta-analysis.

Nutritional Interventions to Prevent Preterm Delivery

Of the nutritional interventions during pregnancy that have been tested by clinical trials to prevent preterm delivery, only calcium and fish oil supplementation appear promising. Nutritional counseling and magnesium supplementation are likely to be effective, but methodological limitations in the analysis of the clinical trial results means definitive conclusions cannot yet be made. Most of the other interventions hypothesized to potentially prevent preterm delivery such as protein and energy supplementation, protein and energy restriction, salt restriction, iron or folate supplementation, zinc supplementation and vitamin A supplementation have not been proven effective.

Nutritional Interventions to Prevent IUGR

Of the interventions tested through randomized clinical trials to prevent IUGR, balanced energy protein supplementation has been shown to reduce the risk of SGA by approximately 30%. On the basis of these results it has been proposed that universal balanced energy supplementation should be provided to women in areas with high prevalence of maternal undernutrition to prevent impaired fetal growth. There is some evidence that magnesium and calcium supplementation may be effective. For calcium supplementation, the evidence is still not clear whether the observed effect on reducing the risk of low birth weight is due to a direct effect on fetal growth or mediated by a prolongation of gestational age at delivery. Other interventions such as nutritional counseling, energy protein restriction, salt restriction, iron or folate supplementation, fish oil supplementation, zinc supplementation, Vitamins E, C, and D supplementation have not shown any preventive effect. High protein supplementation in women of low socioeconomic status in the USA has been associated with an increase in the rate of SGA infants, suggesting that nutritional supplementation may, in some cases, have potentially harmful effects. This finding warrants further investigation.

Conclusions

Low birth weight, due to preterm delivery or IUGR, represents a major public health problem for developing and developed countries. Access to adequate obstetric and neonatal care has been shown to reduce the mortality and morbidity associated with these two conditions Public health efforts should aim at improving the quality and availability of such services, particularly in developing countries where the absolute numbers of low birth weight babies is highest and access to needed care is lowest. Among nutritional interventions to prevent low birth weight, only balanced energy protein supplementation has been shown to be effective in reducing the risk of SGA and should be provided to all women living in areas with high prevalence of maternal undernutrition.

Research efforts should focus on elucidating the etiological factors responsible for preterm delivery and IUGR. Despite the considerable burden of disease related to these conditions, very little progress has been achieved in identifying their causes. This information is essential for developing and implementing effective preventive and therapeutic interventions with universal application and that could benefit the most vulnerable populations of women and newborns.

Nutrition during the preschool years

Samantha L. Huey[a], **Julia L. Finkelstein**[a,b], **and Saurabh Mehta**[a,b], [a] Division of Nutritional Sciences, Cornell University, Ithaca, NY, United States; and [b] Institute for Nutritional Sciences, Global Health, and Technology, Cornell University, Ithaca, NY, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A.F. Williams, Nutritional Problems of pre-school children, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 244–249, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00049-0.

Introduction 121
Faltering growth 122
Developmental milestones in relation to diet 122
Dietary recommendations 122
Excess weight 123
Micronutrient deficiency 123
COVID-19 123
Iron deficiency and anemia 124
Rickets and vitamin D deficiency 124
Food allergy and intolerance 125
Conclusions 125
References 125
Further reading 126

Key points

- Children between one and five years of age are rapidly growing and developing, and attain several developmental milestones sensitive to dietary intakes during this period
- Stunting, underweight and wasting remain highly prevalent globally
- The proportion of children with excess weight continues to increase
- Iron deficiency and anemia are also persistent major issues in this age group
- Although COVID-19 affected only a small proportion of children directly, the ramifications of this pandemic include major indirect effects particularly in the context of nutrition

Introduction

The "preschool period" extends from one year to approximately 5 years of age. It is a time when children are changing from wholly dependent infants to mobile individuals who communicate their needs verbally and eat independently. Poor nutrition during this period has lifelong physical and psychological effects. Undernutrition was also directly or indirectly implicated in approximately 45% of 5.2 million deaths that occurred in children under five globally in 2019 (WHO, 2020a).

Young children face many challenges: their nutrient requirements are relatively high, yet intake is frequently threatened, particularly in low-income settings. Infections, particularly diarrheal disease, respiratory infections, and malaria may simultaneously induce loss of appetite and reduced intake while increasing energy expenditure and nutrient losses. Measles, case counts of which have been at an all-time high over the past few years, can severely impact young children; most of the 207,500 measles deaths in 2019 were among children under 5 years of age, particularly those living in low-resource settings and who are undernourished (WHO, 2020c). Additionally, the COVID-19 pandemic has had a profound negative impact on dietary quality and nutrition status in children and is expected to have ripple effects for years to come. The need to fuel catch-up growth after infection can further increase requirements.

Paradoxically, the rising prevalence of obesity in higher income countries and settings undergoing economic transition is also believed to have its origins in early life; accelerated weight gain in the preschool years is associated with an increased risk of cardiovascular and metabolic complications later in adult life. Therefore, the early years particularly those before the second birthday, i.e., the first 1000 days of life, represent a "window of opportunity" for investment in nutrition.

Faltering growth

Growth monitoring is a fundamental aspect of both child health surveillance and pediatric clinical care. It is based on the premise that the growth trajectory of a healthy child follows the percentile channels described by a reference or "standard" such as that described by the World Health Organization International Growth Reference of 2006 (WHO, 2009).

Overall healthy children become leaner and the amount of fat mass decreases while fat-free mass increases. There are also changes in the proportionate growth of organ systems. The brain continues to grow at a greater velocity than somatic tissues, particularly during the early preschool years. By age two it has attained approximately 80% of its adult size and by age five, and 90%. Linear growth velocity varies, approximately reaching 9 cm per year at the end of the second year, reaching 50% of adult height at two years of age; 9 cm per year at the end of the third year; and 5–8 cm per year thereafter until puberty. Average weight gain in healthy children is 7 g body weight per day by 24 months of age and remains at around 2–3 kg per year until adolescence.

Linear growth faltering in children under five continues to be a major global health issue despite reducing stunting from 33% to 22% over the past two decades (UNICEF/WHO/WB, 2021). Unfortunately, these achievements in the decrease of linear growth faltering, as well as the prevalence of underweight and wasting, are at risk of being undone due to the COVID-19 pandemic. Stunting reflects recurrent or long-term undernutrition, while underweight and wasting result from acute or recurring malnutrition or and/disease; it is thought among children who survive episodes of wasting during COVID-19 that linear growth is likely to be affected. Suboptimal growth in the preschool years also contributes to stunting in later childhood and adult short stature, which can perpetuate the intergenerational cycle of undernutrition in that shorter mothers give birth to smaller offspring. Further, some studies have shown that stunted children are likely to develop obesity and other chronic diseases in adulthood, in settings undergoing nutritional transition.

The causes of growth faltering are complex and partly determined by the environment in which the child lives. For example, in urban slums settings, challenges may include high mobility, recurrent acute infection, poor sanitation and food shortage. Infections contribute to poor nutrient intakes, increasing the risk of growth faltering. Low micronutrient status is common in this age group, particularly in lower-resource settings, and often manifests as deficiencies of iron, zinc, vitamins A and D. Additional risk factors for poor growth may include maternal education, child's gender in that most studies show more male children are malnourished than female children, household income, and family size.

Developmental milestones in relation to diet

Developmental progress through the preschool years is closely linked to nutritional status because it determines how the child eats, selects food, and expends energy. Equally, poor nutritional status is a common consequence when developmental progress is delayed or halted.

As they enter the preschool years children begin to walk. By age two children can climb stairs, by age three they can ride a tricycle, and by four they run with sufficient balance and skill to avoid obstacles. In keeping with these advances, physical activity level (PAL, the total energy expenditure divided by basal metabolic rate) rises from approximately 1.40 (sedentary) to 1.55 (light activity) between one to five years of age. Children also develop their eating skills in the preschool years. At one year, most children can drink from a cup without a lid, albeit with some spillage. Toward the end of the first year, children should be able to pick up food with their hands and transfer it to the mouth. At this age they can also begin to use a spoon. The ability to use spoon and fork together or chopsticks normally appears toward the end of the second year and by four they are used accurately but not until five years or so can the child be expected to use these utensils.

With the cessation of breastfeeding and/or formula and the introduction of solid foods, the child's gut microbiome—the bacteria present in the human gastrointestinal tract, as well as their genes—undergoes shifts in the number and types of bacteria present. These shifts may depend on type of weaning food consumed, baseline gut microbial composition, and other factors such as birth mode and infant feeding type. Research on the early childhood gut microbiome has increased over the last several years, with studies supporting the links between diet, the gut microbiome, and immune system development. For example, a child's immune cells may learn to distinguish pathogens from harmless antigens from dietary inputs which are processed by gut microbiota, in addition to environmental, and other exposures a child may encounter as they mature. Increasing dietary diversity may therefore benefit both the child and their gut microbiome.

Dietary recommendations

Infant and young child feeding guidelines recommend exclusive breastfeeding until age 6 months, at which point solid complementary foods may be introduced into the diet while partial breastfeeding continues until at least 12 months of age (WHO, 2021c). Between 15 and 24 months of age, approximately 60% of energy comes from solid table foods, with milk being the leading source (25%) of daily energy. Milk consumption decreases with older age, with increasing dietary diversity of solid foods. The American Academy of Pediatrics recommends that children 1–4 years of age be given three main meals and two to three snacks throughout the day. In an encouraging trend, country-specific guidelines are being formulated for this age group, including the recent 2020–2025 Dietary Guidelines for Americans recommendations.

Unfortunately, foods low in micronutrient density such as cookies, salty snacks, and candy often contribute significant amounts of energy as refined carbohydrate and fat along with salt in the diets of young children, as high as 28–48% of daily energy intake. Processed meats and bread may further increase saturated fat and salt intake. Indeed it can be difficult to keep salt intakes in diet of young children within the target population average (2 g per day, 1–3 years; 3 g per day, 4–6 years), particularly where there is reliance on commercially prepared foods. Sweetened beverages may additionally offer significant amounts of sugar; preferably water should be provided between meals and milk or fruit juice (diluted to 50%) with meals.

In terms of macronutrients, children one to three years of age are recommended to consume 1.05 g protein per kg per day, and at four years of age to decrease intake to 0.95 g/kg/day. Fat intake restriction should be avoided before two years of age due to developmental demands. After two years of age, fat should approximate 30% of total energy.

Adequate micronutrient intake is critical for proper growth and development. Approximately 21% of toddlers and 40% of preschool children are given some type of daily micronutrient supplement by their parents or guardians, however it is important to fully assess the child's diet before giving potentially unnecessary supplements. For children who cannot consume a well-balanced diet, vitamin and mineral supplements should be considered.

Excess weight

Overweight and obesity are defined as abnormal or excessive fat accumulation that may impair health (WHO, 2021d). In children under five years of age, obesity and overweight is defined as weight-for-height greater than three and two standard deviations above the WHO Child Growth Standards median, respectively. Despite becoming a concern several decades ago, the epidemic of childhood overweight and obesity remains an important issue. Over the last 45 years there has been a remarkable increase globally in the proportion of young children who are overweight or obese, including 39 million (5.7%) children under age five in 2020. Overweight and obesity is on the rise across both high-income countries as well as low- and middle-income countries, particularly in urban areas. In Africa, where more than one-quarter of overweight children live, the number of overweight children under five increased by almost 24%. Half of overweight or obese children under five lived in Asia between 2000 and 2019. With the rise of overweight and obesity comes an increase in hidden hunger—deficiencies in micronutrients with simultaneous overconsumption of energy in the form of empty calories, which may impact risk of and responses to infection; gut health; cognition and brain development, and body composition. Reduced physical activity, lower socioeconomic status and changing food systems (for example, food deserts and food swamps limiting access to healthy food) are part of an obesogenic environment, which promotes overweight and obesity.

Short term consequences of excess weight in childhood include a higher likelihood to suffer from psychological comorbidities, allergy and asthma, low-grade systemic inflammation, liver complications, and metabolic and cardiovascular risk factors. Overweight and obesity in childhood also brings long term consequences, including persistence of excess weight and increased risks of cardiovascular diseases, diabetes, cancers, and other musculoskeletal disorders.

The 2016 Report of the Commission on Ending Childhood Obesity has developed a comprehensive, integrated package of recommendations, including: (1) implement comprehensive programs that promote the intake of healthy foods and reduce the intake of unhealthy foods and sugar-sweetened beverages by children and adolescents; (2) implement comprehensive programs that promote physical activity and reduce sedentary behaviors in children and adolescents; (3) integrate and strengthen guidance for noncommunicable disease prevention with current guidance for preconception and antenatal care, to reduce the risk of childhood obesity; (4) provide guidance on, and support for, healthy diet, sleep and physical activity in early childhood to ensure children grow appropriately and develop healthy habits; (5) implement comprehensive programs that promote healthy school environments, health and nutrition literacy and physical activity among school-age children and adolescents; (6) provide family-based, multicomponent, lifestyle weight management services for children and young people who living with obesity. However, work is still ongoing to expand the evidence basis and develop additional policy approaches (WHO, 2021b). The World Health Organization Guideline Development Group is currently outlining recommendations for the treatment and management of children and adolescents with obesity.

Micronutrient deficiency

Globally, preschool children are particularly at risk of deficiency in many micronutrients including iron, zinc, iodine, vitamin A, and vitamin D, leading to substantial morbidity and mortality. Often, these deficiencies coexist. Impacts on dietary quality and food security due to the COVID-19 pandemic has resulted in additional micronutrient malnutrition, the full extent of which is currently unknown.

COVID-19

Though the coronavirus that causes COVID-19 affects only a small proportion of children directly, the ramifications of this pandemic include major indirect effects particularly in the context of nutrition. Social and economic inequalities, as well as food insecurity, already on the rise globally, have been further exacerbated by the COVID-19 pandemic. This has raised concerns

for both extremes of malnutrition: increasing pediatric obesity, particularly in middle- and high-income countries, as well as deepening undernutrition in poorer countries, undoing decades of global progress. It is estimated that undernutrition will increase by an additional 6.7 million children in 2020, as well as an additional 10,000 child deaths per month. Additionally, the lockdowns associated with the pandemic may result in a difficult-to-reverse sedentary lifestyle and poor dietary quality in almost all children, from those living in higher income settings to lower resource areas. Children who have migrated to Western countries may face initial undernutrition at arrival, and subsequently become overweight in these settings.

Among infants and children who have suspected or confirmed COVID-19, all recommended infant and young child feeding practices remain the same: exclusively breastfeed for the first 6 months, introduce complementary foods from 6 to 2 years of age, and continue breastfeeding up to or beyond 2 years of age (WHO, 2020b). Mothers with suspected or confirmed COVID-19 and isolated at home should continue with the aforementioned recommended feeding practices (UNICEF, 2020). Currently, there is no evidence of SARS-CoV-2 transmission through breastmilk, and preliminary studies have found that SARS-CoV-2 neutralizing IgA and IgG antibodies persist in breast milk for at least 6 weeks after vaccination against the virus.

Iron deficiency and anemia

Anemia (hemoglobin less than 11.0 g dL^{-1}) (WHO, 2011) is estimated to affect 42% of children under five globally (WHO, 2021a). In 2019 the prevalence of anemia in under-five children is particularly high (44 to nearly 80%) in Sub-Saharan Africa and South Asia (WHO, 2021e), while in the United States the figure is closer to 6%; in the UK, 15%; in South America, 12–35%, and in Australia and New Zealand, around 13–15% (WHO, 2021e). Iron deficiency is the most common cause of anemia, while deficiencies in folate, vitamin A, vitamin B12, as well as hemoglobinopathies, parasites, and infectious diseases contribute to anemia prevalence (WHO, 2021a).

By six months of age, an infant's iron stores are depleted and iron intake from external dietary sources is required. Early cessation of breastfeeding with dependence on cow's milk, and complementary foods low in iron content contribute to low intake, which is compounded by increased losses during gastrointestinal infection. Severe anemia may be a manifestation of malabsorption, for example, a consequence of celiac disease, particularly if associated with weight faltering. Iron deficiency occurs at twice the rate of iron-deficiency anemia; with or without anemia, depressed levels of body iron early in life has several short- and long-term consequences. On the cellular level, iron is crucial for respiration, energy production, DNA synthesis, and cell proliferation; functionally, iron plays a role in cognitive development, school achievement, immune function, and physical performance. Oral iron therapy raises iron status but must be used with caution in areas with a high prevalence of malaria. Some studies have shown potentially detrimental effects of oral iron supplements and fortificants on the gut microbiome of young children living in low-resource settings in Africa, such as increases in the abundance of potential pathogens, decreases in beneficial microbes such as Bifidobacteria, and gastrointestinal distress. Further research on this phenomenon is required.

Rickets and vitamin D deficiency

Given there are no national registrations of rickets cases, few population level data on the global prevalence of rickets exist. Further complicating this issue is the lack of internationally accepted diagnostic criteria, which can involve health history, physical exam, radiographs, and biochemical testing. In the US and Europe, the prevalence of rickets declined over the last century due to vitamin D fortification programs but over the last two decades a resurgence has been reported in the UK and other countries. Several factors are thought to have contributed to this increase in rickets prevalence. For example, less sun exposure (due to time indoors and sunscreen as well as air pollution); and global migration and rural to urban migration, particularly among those with higher skin melanin content living in temperate climates. Regional databases, spanning periods within 1967–2016 recently reviewed by WHO in 2019 (WHO, 2019), show a wide range of rickets prevalence, from 1% out of 20,000 children aged 1–15 years in 2008 in Bangladesh, to upwards of 50% in Kenya in 2014. Softening of the bone matrix in rickets predisposes to fractures and to bony deformities such as bowed legs and delayed closure of the cranial sutures. Deficiency of vitamin D may result in other clinical presentations in young children including hypocalcemia-related seizures and stridor. Low vitamin D status throughout pregnancy is associated with low infant vitamin D status at birth. Poor intake of vitamin D-rich foods or supplements through infancy may further increase risk.

Vitamin D deficiency is defined by the presence of a plasma 25-hydroxycholecalciferol concentration <50 nmol L^{-1}. Levels above this threshold are unlikely to be associated with bone disease. For other health outcomes such as immune function or atopy, concentrations below 75 nmol L^{-1} are considered insufficient; however, further research is required in this area. Prevention of vitamin D deficiency and rickets may be achieved by vitamin D supplements and adequate sun exposure. The National Academies of Science, Engineering and Medicine (NASEM) recommends that pregnant and breastfeeding women, as well as children over one year of age, consume a recommended dietary allowance (RDA) of 600 IU daily. In the US, the RDA is 10 mcg (400 IU) before age 1 and 15 μg (600 IU) thereafter. In India, the estimated average requirement for children beginning at age one year is 10 μg (400 IU) per day. Similarly, China recommends children from two weeks to two years old to receive 400 IU daily and be fed vitamin D-fortified food from age zero to three years.

Policy on the use of supplements in infants varies between countries. In the UK it is recommended that all breastfed infants should be given a supplement of 7 μg per day by seven months of age, starting earlier than this if the mother's vitamin D status

is uncertain, and continuing until at least four years of age. In high-risk groups, for example, those with dark skin or those with reduced sunlight exposure, supplementation should continue through childhood.

In some other European countries and the USA, supplementation from birth is recommended. In the US and Canada, vitamin D fortification of infant formula is mandatory (1–2.5 μg/100 kcal (40–100 IU) and 1–2 mcg/100 kcal (40–80 IU), respectively). Infants who are not breastfed do not require supplementation until consumption of infant formula (which is fortified with vitamin D) falls below 500 mL per day. Many foods including dairy products, orange juice, and plant milks are fortified with vitamin D (about 3 μg (120 IU) per cup), and breakfast cereals contain added vitamin D.

Food allergy and intolerance

"Food allergy" and "food intolerance" may both occur in young children and cause serious clinical manifestations. In both, the child exhibits a reproducible adverse response to a food or food component; in the case of allergy this is immunologically mediated, whereas a variety of mechanisms may explain intolerance. The latter include pharmacological responses to food components or inability to metabolize a food component (e.g., hereditary fructose intolerance or lactose intolerance).

Food allergy, like asthma, allergic rhinitis, and atopic dermatitis is usually mediated through release of immunoglobulin E (IgE) in response to antigen exposure, and thus manifests with immediate symptoms when the offending food is ingested: these may include swelling and erythema of the face, lips, and tongue, periorbital or generalized edema, and rarely anaphylactic shock. Acute gastrointestinal symptoms such as vomiting, pain, or diarrhea may also occur. Sometimes symptomatology is more chronic, for example, worsening of wheezing or eczema. Symptoms resolve when the food is omitted from the diet and recur when it is reintroduced. It is important, however, to stress that neither maneuver should be attempted without professional supervision.

Dietary challenge is not always necessary to establish the diagnosis, as skin prick testing or specific IgE measurement can be helpful if the clinical history is strongly suggestive. Most young children grow out of egg or milk allergies by school age, but nut, fish, soy, and wheat allergies are generally more persistent and may remain a problem into adult life. Serial measurement of specific IgE can be helpful in determining whether resolution is likely. Parents may need to be provided with an adrenaline injection device and instructed in its use. Desensitization through subcutaneous administration of specific antigens may prove an effective therapy where allergic symptoms persist.

In 2017, the National Institute of Allergy and Infectious Diseases, along with 25 professional organizations, federal agencies, and patient advocacy groups, published three sets of updated guidelines to address the prevention of peanut allergy across groups with varying risk factors for allergy. Infants who have severe eczema and/or egg allergy comprise the highest risk group, and are recommended to be introduced to peanut at 4–6 months of age, ideally after allergy testing for peanut-specific IgE or skin prick test (guideline #1). Infants with mild to moderate eczema should be introduced to peanut around 6 months of age (guideline #2). Finally, infants who do not have eczema or other food allergy, who are not at increased risk, are recommended to be exposed to peanut "freely" into the diet with other solid foods (guideline #3). These guidelines are a reversal from the 2000 American Academy of Pediatrics guidelines which advised high-risk infants to completely avoid peanut until age 3; however, there was no convincing evidence for delaying introduction to peanut beyond 4–6 months of age, resulting in updated guidance.

Conclusions

Undernutrition and micronutrient deficiency remain a major cause of morbidity and mortality in children between 1 and 5 years of age, while the number of children with excess weight continues to rise. These nutrition issues have been directly and indirectly exacerbated by the COVID-19 pandemic. Focusing efforts on improving child health, and supporting adequate dietary and nutrient intake, particularly in this age group, remains a priority.

References

UNICEF, 2020. Infant and Young Child Feeding in the Context of COVID-19.
UNICEF/WHO/WB, 2021. Joint child malnutrition estimates. WHO 24, 51–78.
WHO, 2009. WHO Child Growth Standards and the Identification of Severe Acute Malnutrition in Infants and Children. A Joint Statement by the WHO and UNICEF. WHO, UNICEF, Geneva, Switzerland.
WHO, 2011. Haemoglobin concentrations for the diagnosis of anemia and assessment of severity. In: Vitamin and Mineral Nutrition Information System.
WHO, 2019. Nutritional Rickets: A Review of Disease Burden, Causes, Diagnosis, Prevention and Treatment. World Health Organization, Geneva.
WHO, 2020a. Children: improving survival and well-being. Fact Sheets. Available from. https://www.who.int/news-room/fact-sheets/detail/children-reducing-mortality. (Accessed 8 September 2020).
WHO, 2020b. In: WHO (Ed.), Clinical Management of Severe Acute Respiratory Infection (SARI) When COVID-19 Disease is Suspected. Geneva.
WHO, 2020c. Worldwide Measles Deaths Climb 50% From 2016 to 2019 Claiming Over 207,500 Lives in 2019. World Health Organization, New York/Geneva/Atlanta. Available. https://www.who.int/news/item/12-11-2020-worldwide-measles-deaths-climb-50-from-2016-to-2019-claiming-over-207-500-lives-in-2019. (Accessed 18 June 2021).
WHO, 2021a. Anaemia. World Health Organization, Geneva. Available. https://www.who.int/health-topics/anemia#tab=tab_1. (Accessed 8 October 2021).
WHO, 2021b. In: WHO (Ed.), Draft Recommendations for the Prevention and Management of Obesity Over the Life Course, Including Potential Targets. World Health Organization, Geneva.

WHO, 2021c. Indicators for Assessing Infant and Young Child Feeding Practices: Definitions and Measurement Methods. Geneva, Switzerland.
WHO, 2021d. Obesity and Overweight. World Health Organization, Geneva. Available. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight. (Accessed 8 October 2021).
WHO, 2021e. Prevalence of Anemia in Children Aged 6–59 Months (%). World Health Organization, Geneva. Available. https://www.who.int/data/gho/data/indicators/indicator-details/GHO/prevalence-of-anemia-in-children-under-5-years-(-). (Accessed 8 October 2021).

Further reading

Abioye, A.I., Fawzi, W.W., 2020. Chapter 27—Nutritional Anemias. Elsevier Inc.
American Academy of Pediatrics, 2017. Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents. American Academy of Pediatrics, Elk Grove Village.
Bhutta, Z.A., Hurrell, R.F., Rosenberg, I.H., 2012. Safety of iron fortification and supplementation in malaria-endemic areas. Nestlé Nutr. Inst. Workshop Ser. 70, 117–217.
Camaschella, C., 2015. Iron-deficiency anemia. N. Engl. J. Med. 372, 1832–1843.
Centeno-Tablante, E., Medina-Rivera, M., Finkelstein, J.L., Rayco-Solon, P., Garcia-Casal, M.N., Rogers, L., Ghezzi-Kopel, K., Ridwan, P., Pena-Rosas, J.P., Mehta, S., 2021. Transmission of SARS-CoV-2 through breast milk and breastfeeding: a living systematic review. Ann. N. Y. Acad. Sci. 1484, 32–54.
Daniels, S.R., Greer, F.R., Committee on Nutrition, 2008. Lipid screening and cardiovascular health in childhood. Pediatrics 122, 198–208.
Dewey, K.G., Begum, K., 2011. Long-term consequences of stunting in early life. Matern. Child Nutr. 7 (Suppl. 3), 5–18.
Di Cesare, M., Soric, M., Bovet, P., Miranda, J.J., Bhutta, Z., Stevens, G.A., Laxmaiah, A., Kengne, A.P., Bentham, J., 2019. The epidemiological burden of obesity in childhood: a worldwide epidemic requiring urgent action. BMC Med. 17, 212.
Edwards, J.K., Thiongo, A., Van Den Bergh, R., Kizito, W., Kosgei, R.J., Sobry, A., Vandenbulcke, A., Zuniga, I., Reid, A.J., 2014. Preventable but neglected: rickets in an informal settlement, Nairobi, Kenya. Public Health Action 4, 122–127.
Goudet, S., Griffiths, P., Bogin, B., Madise, N., 2017. Interventions to tackle malnutrition and its risk factors in children living in slums: a scoping review. Ann. Hum. Biol. 44, 1–10.
Goudet, S.M., Bogin, B.A., Madise, N.J., Griffiths, P.L., 2019. Nutritional interventions for preventing stunting in children (birth to 59 months) living in urban slums in low- and middle-income countries (LMIC). Cochrane Database Syst. Rev. 6, CD011695.
Holick, M.F., Binkley, N.C., Bischoff-Ferrari, H.A., Gordon, C.M., Hanley, D.A., Heaney, R.P., Murad, M.H., Weaver, C.M., Endocrine Society, 2011. Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society Clinical Practice Guideline. J. Clin. Endocrinol. Metab. 96, 1911–1930.
Mehta, S., Huey, S.L., Mcdonald, D., Knight, R., Finkelstein, J.L., 2021. Nutritional interventions and the gut microbiome in children. Annu. Rev. Nutr. 41, 479–510.
Mwangome, M., Prentice, A.M., 2019. Tackling the triple threats of childhood malnutrition. BMC Med. 17, 210.
NASEM, 2020. Feeding Infants and Children From Birth to 24 Months. Washington, D.C.
National Institute of Nutrition, Indian Council of Medical Research, 2009. Nutrient Requirements and Recommended Dietary Allowances for Indians. National Institute of Nutrition, Indian Council of Medical Research, Hyderabad, India.
National Research Council (US) Subcommittee, 1989. Recommended Dietary Allowances, tenth ed. Washington, D.C.
Neuberger, A., Okebe, J., Yahav, D., Paul, M., 2016. Oral iron supplements for children in malaria-endemic areas. Cochrane Database Syst. Rev. 2, CD006589.
NHMRC, 2006. Nutrient Reference Values for Australia and New Zealand. Nutrient Reference Values for Australia and New Zealand Including Recommended Dietary Intakes. National Health and Medical Research Council, Australian Government, Commonwealth of Australia.
NIH Office of Dietary Supplements, 2021. Vitamin D—Health Professional. NIH Office of Dietary Supplements. Available. https://ods.od.nih.gov/factsheets/VitaminD-HealthProfessional/. (Accessed 8 October 2021).
Paganini, D., Zimmermann, M.B., 2017. The effects of iron fortification and supplementation on the gut microbiome and diarrhea in infants and children: a review. Am. J. Clin. Nutr. 106, 1688S–1693S.
Parks, E.P., Mascarenhas, M.R., Goh, V., 2020. Nutrient needs and requirements during growth. In: Present Knowledge in Nutrition, eleventh ed. Elsevier.
Perl, S.H., Uzan-Yulzari, A., Klainer, H., Asiskovich, L., Youngster, M., Rinott, E., Youngster, I., 2021. SARS-CoV-2-Specific antibodies in breast milk after COVID-19 vaccination of breastfeeding women. J. Am. Med. Assoc. 325, 2013–2014.
Ramakrishnan, U., Martorell, R., Schroeder, D.G., Flores, R., 1999. Role of intergenerational effects on linear growth. J. Nutr. 129, 544S–549S.
Ross, A.C., Taylor, C.L., Yaktine, A.L., Del Valle, H.B. (Eds.), 2011. Dietary Reference Intakes for Calcium and Vitamin D. Washington, D.C.
Roy, S., Rakib, R., Alam, N.H., Haque, S., Das, H.K., Ali, M., 2009. High Burden of Childhood Rickets in Bangladesh: The First National Prevalence Survey of Mineral Deficiency. International Centre for Diarrhoeal Disease Research, Bangladesh, Capgan Congress, Dhaka.
Services, USDoHAH, 2020. Dietary Guidelines for Americans, 2020–2025, ninth ed.
Shils, M.E., Olson, J.A., Shike, M., 1994. Modern Nutrition in Health and Disease. Lea & Febiger.
Stone, K.D., 2002. Atopic diseases of childhood. Curr. Opin. Pediatr. 14, 634–646.
Togias, A., Cooper, S.F., Acebal, M.L., Assa'ad, A., Baker Jr., J.R., Beck, L.A., Block, J., Byrd-Bredbenner, C., Chan, E.S., Eichenfield, L.F., Fleischer, D.M., Fuchs III, G.J., Furuta, G.T., Greenhawt, M.J., Gupta, R.S., Habich, M., Jones, S.M., Keaton, K., Muraro, A., Plaut, M., Rosenwasser, L.J., Rotrosen, D., Sampson, H.A., Schneider, L.C., Sicherer, S.H., Sidbury, R., Spergel, J., Stukus, D.R., Venter, C., Boyce, J.A., 2017. Addendum guidelines for the prevention of peanut allergy in the United States: report of the National Institute of Allergy and Infectious Diseases-sponsored expert panel. Ann. Allergy Asthma Immunol. 118, 166–173.e7.
Wang, L.L., Wang, H.Y., Wen, H.K., Tao, H.Q., Zhao, X.W., 2016. Vitamin D status among infants, children, and adolescents in southeastern China. J. Zhejiang Univ. Sci. B 17, 545–552.
WHO, 2002. The Optimal Duration of Exclusive Breastfeeding a Systematic Review. Geneva.
WHO, 2006. WHO Child Growth Standards Based on Length/Height, Weight and Age, vol. 450, pp. 76–85.
WHO, 2019. Nutritional Rickets: A Review of Disease Burden, Causes, Diagnosis, Prevention and Treatment. World Health Organization, Geneva.
WHO, 2021. Call for Experts: WHO Guideline Development Group on Treatment of Children and Adolescents With Obesity. World Health Organization, Geneva. Available. https://www.who.int/news-room/articles-detail/call-for-experts-who-guideline-development-group-treatment-of-children-and-adolescents-with-obesity.
Wong, S., Chan, K., Wong, V., Wong, W., 2002. Use of chopsticks in Chinese children. Child Care Health Dev. 28, 157–161.
Zemrani, B., Gehri, M., Masserey, E., Knob, C., Pellaton, R., 2021. A hidden side of the COVID-19 pandemic in children: the double burden of undernutrition and overnutrition. Int. J. Equity Health 20, 44.

Nutrition for sport and exercise

R.J. Maughan, Loughborough University, Loughborough, UK

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of R.J. Maughan, Sports and Exercise Nutrition, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 204–208, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00253-1.

Introduction 127
Nutrition for Training 128
Protein Requirements 128
Fat and CHO 129
Vitamins and Minerals 130
Water and Electrolyte Balance 130
Dietary Supplements 131
Nutrition for Competition 131
Further Reading 131
Relevant Websites 132

Glossary

Dietary supplements Dietary supplements may be single elements or nutrients, or may be complex mixtures, including herbals and botanicals of uncertain composition.

Fatigue Exercise, if sufficiently intense or prolonged, will inevitably result in fatigue, which has both subjective and objective components. The nature of fatigue depends on exercise type and on individual training status, but typically includes events localized to the active muscles as well as effects on the central nervous system.

Glycogen Glycogen, a long-chain, highly-branched polymer of glucose, is the storage form of carbohydrate in the human body. It is present primarily in skeletal muscle (up to 300–500 g) and in liver (typically 70–100 g in the fed state).

Hydration Euhydration refers to normal or desirable body water content and electrolyte balance. Positive (hyperhydration) and negative (hypohydration) excursions from euhydration are both associated with impaired performance and with health risks. Dehydration is the process of water loss from the body.

Training Training involves a systematic program of exercise with the aim of achieving an improvement in performance. The key characteristics of a training program are the intensity, duration and frequency of the individual training sessions. Performance outcomes are specific to the type of stimulus applied and proportional to the training load.

Introduction

In 2010, the International Olympic Committee issued a consensus statement on nutrition in sport, which began with the following words: "Diet significantly affects athletic performance". This is a bold and unambiguous statement, leaving little room for doubt, but there are other, more important factors that will determine the outcome of sporting contests. Talent, motivation, training, tactics, and many other considerations are much more important. It is clear, however, that when all else is equal, nutrition can make the difference between victory and defeat. The role of nutrition, and the eating strategies that should be adopted, will vary greatly between sports and will vary also according to the level of competition, the training load and the competition schedule, so athletes may need the help of qualified professionals to ensure good nutrition strategies. For those who exercise for health and enjoyment rather than in pursuit of fame and wealth, a sound nutrition strategy also has much to offer.

Many different issues arise in considering the interactions between diet and exercise. In considering the role of diet in the athlete's life, two main issues must be considered, each of which gives rise to many subordinate questions. The first question is how the demands of training affect the body's requirement for energy and nutrients: This then has implications for body composition (including the body content of fat, muscle, and bone), for the hormonal environment and the regulation of substrate metabolism, and for various disease states that are affected by body fatness, nutrient intake, and other related factors. The second question is how nutritional preparation can influence performance in competition.

Nutrition for Training

The aim of training is to improve performance, and the effectiveness of a training program will depend on the intensity, duration, and frequency of the training sessions that are completed. Training is designed to induce highly specific adaptations that address the limitations to exercise performance and move those boundaries so that performance improves. It was thought that the primary aim of nutrition support in training was to allow better recovery between training sessions so that the total training load could be increased. It is now recognized, however, that more training is not always better – it brings an increased risk of injury and chronic fatigue. Instead, nutrition strategies are aimed at allowing greater adaptations to the training stimulus, or to allow the same training adaptation with less training.

Sound nutrition strategies can also make exercise feel easier. This is especially important to those who exercise for health: If the exercise feels hard, it is unlikely to be repeated and the duration will be cut short. It is well established that some simple strategies, such as ensuring an adequate carbohydrate (CHO) status and maintaining good hydration status, will reduce the subjective perception of effort.

Protein Requirements

In most cases, training adaptations aim to change the structure and function of the muscles engaged in the exercise task, though all the body's tissues are affected. As the structure and function of muscles are dictated by their protein composition, the training response involves an increased net breakdown of proteins that are not required and an increased net synthesis of those proteins that contribute to exercise performance. The weightlifter wants bigger muscles with more actin and myosin to increase force-generating capacity. The marathon runner does not want bigger muscles: instead, a greater content of oxidative enzymes and more capillaries to increase oxygen and substrate delivery. Smaller muscle fibers – to reduce the diffusion distance from the capillary to the fiber – and a decrease in total mass are both desirable outcomes.

The idea that protein requirements are increased by physical activity is intuitively attractive, and high-protein diets are a common feature of the diets of many sportsmen and women. The available evidence does show an increased rate of oxidation of the carbon skeletons of amino acids during exercise, especially when CHO availability is low. Protein contributes only approximately 5% of total energy demand in endurance exercise, but the absolute rate of protein breakdown is higher than at rest (where protein contributes about the same fraction as the protein content of the diet, which is typically approximately 12–16% of the total energy intake) because of the higher energy turnover. It is often recommended that athletes engaged in endurance activities on a daily basis should aim to achieve a protein intake of approximately 1.2–1.4 g kg^{-1} day^{-1}, whereas athletes engaged in strength and power training may need as much as 1.6–1.7 g kg^{-1} day^{-1}. There is no evidence of benefits from a daily intake of more than 2 g kg^{-1}. This compares with an estimated average requirement of approximately 0.6 g kg^{-1} day^{-1} in sedentary people and a recommended intake of approximately 0.8–1.0 g kg^{-1} day^{-1} for those who take no exercise.

In strength and power sports such as weightlifting, sprinting, and bodybuilding, the use of high-protein diets and protein supplements is especially prevalent, and daily intakes in excess of 2–4 g kg^{-1} are not unusual. Scientific support for such high intakes is generally lacking, but those involved in these sports are adamant that such high levels of intake are necessary, not only to increase muscle mass, but also to maintain muscle mass. This apparent inconsistency may be explained by Millward's adaptive metabolic demand model, which proposes that the body adapts to either high or low levels of intake, and that this adjustment to changes in intake occurs only very slowly. This means that individuals, such as strength and power athletes, who consume a high-protein diet over many years, will find that any reduction in protein intake will result in a loss of muscle mass. This is because of an upregulation of the activity of the enzymes involved in protein oxidation to cope with the high intake: activity of these enzymes remains high when there is a sudden decrease in intake, leading to a net catabolic effect.

Protein synthesis and degradation are both enhanced for some hours after exercise, and the net effect on muscle mass will depend on the relative magnitude and duration of these effects. Several recent studies have shown that ingestion of small amounts of protein (typically approximately 20–40 g) or essential amino acids (approximately 10 g) either before or immediately after exercise will result in net protein synthesis in the hours after exercise, whereas net negative protein balance is observed if no source of amino acids is consumed in the immediate postexercise period. These observations have led to recommendations that approximately 20 g of mixed protein should be consumed immediately after exercise. It is important to recognize, though, that the control condition in most of these studies has involved a relatively prolonged (6–12 h) period of fasting before and after the exercise bout, and this does not reflect the normal behavior of athletes. Individuals who consume foods containing CHO and proteins in the hour or two before exercise may not further increase protein synthesis if additional amino acids or proteins are ingested immediately before, during, or after exercise.

Various high (30%) protein, high (30%) fat, low (40%) CHO diets have been promoted for weight loss, and some diets even suggest almost complete elimination of CHO from the diet. Some of these diets have been specifically targeted at athletes, accompanied by impressive claims and celebrity endorsements. Proposed mechanisms of action of these diets include reduced circulating insulin levels, increased fat catabolism and altered prostaglandin metabolism. The high-protein content of these diets may contribute to increased satiety, but it seems more likely that they may achieve weight loss simply by restricting dietary choice and therefore reducing energy intake. These diets can be effective in promoting short-term weight loss, primarily by restricting

energy intake (typically to 1000–2000 kcal day^{-1}). There is not any evidence to support improvements in exercise performance, and what evidence there is does not support the concept.

Fat and CHO

A more recent development has been the suggestion that training on a high-fat diet can enhance endurance performance. The theory is sound, but the experimental evidence does not support the theoretical advantage. CHO is an essential fuel for the brain, red blood cells, and a few other tissues and is also an important fuel for muscle during high-intensity exercise. At rest and during low-intensity exercise, most of the energy demands of skeletal muscle can be met by fat oxidation, but the contribution of CHO, and especially of the muscle glycogen, increases as the rate of energy demand increases. The muscle glycogen stores are small, however, and once the glycogen content of the exercising muscles reaches very low levels, the work rate must be reduced to a level that can be accommodated by fat oxidation. In high-intensity exercise, essentially all of the energy demands are met by CHO metabolism. Repeated short sprints therefore place high demands on the muscle CHO store, most of which can be converted to lactate within a few minutes.

CHO is stored in the body in the form of glycogen, primarily in the liver (approximately 70–100 g in the fed state) and in the skeletal muscles (approximately 300–500 g, depending on muscle mass and preceding diet). These stores are small relative to the body's requirements for CHO. CHO supplies approximately 45% of the energy in the typical Western diet. This amounts to approximately 200–300 g day^{-1} for the average sedentary individual, and is adequate for normal daily activities. In an hour of hard exercise, however, up to 200 g of CHO can be used, and sufficient CHO must be supplied by the diet to replace the amount used. Replacement of the glycogen stores is an essential part of the recovery process after exercise: if the muscle glycogen content is not replaced, the quality of training must be reduced, and the risks of illness and injury are increased. Low muscle glycogen levels are associated with an increase secretion of cortisol during exercise, with consequent negative implications for immune function.

Reducing the CHO availability forces the muscle to rely more on fat oxidation for energy supply and restricting the availability of CHO will result in an increased capacity of the muscle to oxidize fat. This can be achieved by feeding a low-CHO diet that meets the total energy demand through an increased fat intake or by feeding an energy-deficient diet. The training studies that have been completed suggest that training on a low-CHO diet impairs some of the adaptations that take place in muscle in response to training, even though the capacity for fat oxidation is increased. Even restoration of the muscle glycogen content by a short period of high-CHO intake does not allow the same performance capacity as when training was performed with adequate CHO availability. The recommendation that athletes consume a high-CHO diet during periods of intensive training therefore remains in place. When rapid recovery is a priority, especially when the interval between successive training sessions is no more than a few hours, replacement of CHO should begin as soon as possible after exercise with CHO foods that are convenient and appealing. Thereafter, the diet should supply sufficient CHO to replace the amount used in training and to meet ongoing demands of other tissues. Some recommendations for CHO intake after training or competition are shown in **Table 1**. It is important to remember that not all athletes need a high-CHO diet at all times: when training consists of mostly technical work, the total energy expenditure and the demand for CHO may be low. For the athlete with very high levels of energy expenditure in training, the exercise intensity will inevitably be reduced to a level where fatty acid oxidation will make a significant contribution to energy supply and fat will provide an important energy source in the diet. Once the requirements for protein and CHO are met, the balance of energy intake can be in the form of fat. Fat also serves other important functions in the diet. As well as providing essential fatty acids, it acts as a vehicle for the transport of fat-soluble nutrients. Some athletes try to minimize their fat intake, but this is not wise.

The high-CHO diet recommended for the physically active individual coincides with the recommendations of various expert committees that a healthy diet is one that is high in CHO (at least 55% of energy) and low in fat (less than 30% of energy). However, where energy intake is either very high or very low, it may be inappropriate to express the CHO requirement as a fraction of energy intake. With low-total energy intakes, the fraction of CHO in the diet must be high, but the endurance athlete with a very high energy intake may be able to tolerate a higher fat intake. Recommendations, as in **Table 1**, should be framed in absolute amounts relative to body mass, i.e., grams of CHO per kg body mass.

The type of CHO eaten is generally much less important than the amount. It is valuable to choose nutrient-rich CHO foods and to add other foods to recovery meals and snacks to provide a good source of protein and other nutrients. The presence of small amounts of protein in recovery meals may promote additional glycogen recovery when CHO intake is less than optimal or when frequent snacking is not possible. Protein at this time may also stimulate protein synthesis in muscles, as described above.

Table 1 Suggested carbohydrate intakes for athletes in training

Immediate post-exercise recovery (0–4 h): 1 g per kg body mass per hour, consisting of several small snacks
Daily recovery (moderate duration/low-intensity training): 5–7 g kg^{-1} day^{-1}
Daily recovery (moderate–heavy endurance training): 7–12 g kg^{-1} day^{-1}
Daily recovery (extreme training: 4–6 h or more per day): Up to 10–12 g kg^{-1} day^{-1}

CHO-rich foods with a moderate to high glycemic index (GI) provide a readily available source of CHO for glycogen synthesis, and should be the major fuel choices in recovery meals.

Vitamins and Minerals

Many micronutrients play key roles in energy metabolism, and high rates of energy turnover – up to 20–100 times the resting rate – may be required in the active muscles during hard exercise. Although an adequate vitamin and mineral status is essential for normal health, marginal deficiency states may be apparent only during periods of metabolic stress. Prolonged strenuous exercise performed on a regular basis may also result in increased losses of essential nutrients in sweat or urine, or may increase the rate of breakdown, resulting in the need for an increased dietary intake. An increased food intake to meet energy requirements will generally increase dietary micronutrient in proportion to energy intake, but not all athletes have high-energy intakes. Athletes who restrict food intake to control or reduce body fat levels may have low-energy intakes over prolonged periods. Some athletes may also eat monotonous diets, with a limited range of foods in the diet, thus increasing the risk of an inadequate micronutrient intake. Supplementation with micronutrients may be warranted in some instances, but normally only where specific deficiencies have been demonstrated by biochemical investigations and where dietary modification is not an option.

Individuals who are very active may need to pay particular attention to their intake of iron and calcium. Iron deficiency anemia affects some athletes engaged in intensive training and competition, but it seems that the prevalence is similar in athletic and sedentary populations, suggesting that exercise *per se* does not increase the risk. The implications of even mild anemia for exercise performance are, however, significant. A fall in the circulating hemoglobin concentration is associated with a reduction in oxygen carrying capacity and a decreased exercise performance. Low serum ferritin levels are not associated with impaired performance, however, and iron supplementation in the absence of frank anemia does not influence indices of fitness. Routine iron supplementation is not wise, as too much may be harmful.

Osteoporosis is now widely recognized as a problem for both men and, more especially, women, and an increased bone mineral content is one of the benefits of participation in an exercise program. Regular exercise results in increased mineralization of those bones subjected to stress and an increased peak bone mass may delay the onset of osteoporotic fractures; exercise may also delay the rate of bone loss. Estrogen plays an important role in the maintenance of bone mass in women, and prolonged strenuous activity may result in low estrogen levels, causing bone loss. Many very active women also have a low body fat content and may also have low-energy (and calcium) intakes in spite of their high-activity levels. All of these factors are a threat to bone health. The loss of bone in these women may result in an increased predisposition to stress fractures and other skeletal injury and must also raise concerns about bone health in later life. It should be emphasized, however, that this condition appears to affect only relatively few athletes, and that activity is generally beneficial for the skeleton.

In recent years it has increasingly been recognized that vitamin D may be needed in supplemental form when sun exposure is inadequate, and this may apply especially to those athletes who spend long periods training indoors and to those who live in high latitudes. Intakes of most other nutrients can be met from food sources, but athletes may need advice from qualified professions to identify their nutrition needs and to develop an eating strategy that will meet those needs.

Water and Electrolyte Balance

Prolonged strenuous exercise in a warm environment poses a major challenge to the body's homeostatic mechanisms. Only approximately 20–25% of the energy available from substrate catabolism is used to perform external work, with the remainder appearing as heat. At rest, the metabolic rate is low: Oxygen consumption is approximately 250 ml min^{-1}, corresponding to a rate of heat production of approximately 60 W. Heat production increases in proportion to metabolic demand, and reaches approximately 1 kW in strenuous activities such as marathon running (for a 70 kg runner at a speed that takes approximately 2½ hours to complete the race). To prevent a catastrophic rise in core temperature, heat loss must be increased correspondingly and this is achieved primarily by an increased rate of evaporation of sweat from the skin surface. In hard exercise in hot conditions, sweat rates can reach 3 l h^{-1}, and trained athletes can sustain sweat rates in excess of 2 l h^{-1} for many hours. This represents a much higher fractional turnover rate of water than that of most other body components. In the sedentary individual living in a temperate climate, approximately 5–10% of total body water may be lost and replaced on a daily basis. When prolonged exercise is performed in a hot environment, 20–40% of total body water can be turned over in a single day. In spite of this, the body water content is tightly regulated, and regulation by the kidneys is closely related to osmotic balance.

Along with water, a variety of minerals and organic components are lost in variable amounts in sweat. Sweat is invariably hypotonic relative to plasma and the main electrolytes lost are sodium and chloride, at concentrations of approximately 15–80 mmol l^{-1}. A range of other minerals, including potassium and magnesium, as well as trace elements, are lost in small amounts. Sweat rate and sweat composition both vary greatly between individuals. Some athletes may lose up to 10 g of salt (sodium chloride) in a single training session, and may train in these conditions twice per day. Others doing the same training will lose no more than 1 g of salt. High-salt losses may be related to development of muscle cramps in some, but not all, individuals, and additional salt intake may be helpful for susceptible individuals. Salt losses must be replaced from foods and drinks, though the use of salt supplements is seldom necessary.

Failure to maintain hydration status has serious consequences for the active individual. A body water deficit of as little as 1–2% of total body mass can result in a significant reduction in exercise capacity, especially in endurance exercise performed in warm weather. Endurance exercise is affected to a greater extent than high-intensity exercise, and muscle strength is not adversely affected until water losses reach 5% or more of body mass. Hypohydration greatly increases the risk of heat illness, and also abolishes the protection conferred by prior heat acclimation.

Many studies have shown that the ingestion of fluid during exercise can significantly improve performance. Adding an energy source in the form of CHO confers an additional benefit by providing an energy source for the working muscles. Addition of small amounts (perhaps approximately 2–8%) of CHO, in the form of glucose, sucrose, or maltodextrin, will promote water absorption in the small intestine as well as providing exogenous substrate that can spare stored CHO. Recent evidence suggests that addition of fructose in addition to glucose, sucrose, or maltodextrin will increase intestinal absorption of CHO and can enhance performance. The addition of too much CHO will slow gastric emptying and, if the solution is strongly hypertonic, may promote secretion of water into the intestinal lumen, thus delaying fluid availability. Voluntary fluid intake is seldom sufficient to match sweat losses, and palatability of fluids is therefore an important consideration. It is not necessary to consume enough fluid during exercise to match sweat losses, as a body mass deficit of 1–2% is unlikely to have adverse consequences. If exercise is prolonged and sweat losses high, the addition of sodium to drinks may be necessary to prevent the development of hyponatremia. Ingestion of large volumes of plain water is also likely to limit intake because of a fall in plasma osmolality leading to suppression of thirst.

Replacement of water and electrolyte losses incurred during exercise is an important part of the recovery process in the postexercise period. This requires ingestion of fluid in excess of the volume of sweat lost to allow for ongoing water losses from the body. Reestablishment of water balance requires replacement of solute, especially sodium, losses as well as volume replacement. If food containing electrolytes is not consumed at this time, electrolytes, especially sodium, must be added to drinks to prevent diuresis and loss of the ingested fluid.

Dietary Supplements

The use of nutritional supplements in athletes and in the health-conscious recreationally active population is widespread, as it is in the general population. Many different supplements are used by athletes with the aim of improving or maintaining general health and exercise performance. In particular, supplement use is often aimed at promoting tissue growth and repair, promoting fat loss, enhancing resistance to fatigue, and stimulating immune function. Most of the supplements that are sold to athletes have not been well researched, and both safety and efficacy remain open to question for many of these products. Anyone seeking to improve health or performance would be better advised to ensure that they consume a sound diet that meets energy needs and contains a variety of foods. A recent development of concern to athletes is the finding of various prohibited doping agents in what should be legitimate sports nutrition products. A wider concern are the recent reports of serious health issues related to the use of various supplements.

Supplements for which there is good evidence of beneficial effects on performance in some specific situations include caffeine, creatine, and buffering agents, but the risk of an inadvertent positive doping result must always be considered. Supplement use in young athletes is discouraged, and the focus should be on choosing a varied, nutrient-rich diet to provide all the nutrients essential for growth while maintaining a healthy body composition.

Nutrition for Competition

A detailed consideration of nutrition strategies for competition in all sports is beyond the scope of this brief review. For athletes preparing for competition, a reduction in the training load and the consumption of a high-CHO diet in the past few days are recommended: This will maximize the body's CHO stores, and should ensure optimum performance, not only in endurance activities, but also in events involving short-duration high-intensity exercise and in field games involving multiple sprints. Beginning competition in a well hydrated state is generally beneficial, and is essential in events lasting longer than approximately 30 min in warm environments. Regular intake of fluids should be based on the individual needs identified during training or previous competition.

Further Reading

American College of Sports Medicine, American Dietetic Association, and Dietitians of Canada, 2000. Joint Position Statement: Nutrition and athletic performance. Medicine and Science in Sports and Exercise 32, 2130–2145.

Burd, N.A., Tang, J.E., Moore, D.R., Phillips, S.M., 2009. Exercise training and protein metabolism: Influences of contraction, protein intake, and sex-based differences. Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology 106, 1692–1701.

Burke, L.M., Kiens, B., 2006. "Fat adaptation" for athletic performance – The nail in the coffin? Journal of Applied Physiology 100, 7–8.

Deakin, V., 2010. Measuring nutritional status of athletes: Clinical and research perspectives. In: Burke, L., Deakin, V. (Eds.), Clinical Sports Nutrition, 4th edn., Ch. 2. McGraw Hill, Sydney, pp. 18–43.

Hawley, J.A., Burke, L.M., 2010. Carbohydrate availability and training adaptation: Effects on cell metabolism and exercise capacity. Exercise and Sport Sciences Reviews 38, 152–160.

Maughan, R.J. (Ed.), 2011. Foods, Nutrition and Sports Performance 3. Routledge, London.

Maughan, R.J., Shirreffs, S.M., 2008. Development of individual hydration strategies for athletes. International Journal of Sport Nutrition and Exercise Metabolism 18, 457–472.
O'Connor, H., Olds, T., Maughan, R.J., 2007. Physique and performance for track and field events. Journal of Sports Sciences 25, S49–S60.

Relevant Websites

http://www.acsm.org. – American College of Sports Medicine.
http://www.eatright.org. – American Dietetic Association.
http://www.americanheart.org. – American Heart Association.
http://www.ais.org.au. – Australian Institute of Sport.
http://www.dietwatch.com. –DietWatch.com Inc.
http://www.sportsoracle.com. –Sportsoracle Limited.
http://www.dapa-toolkit.mrc.ac.uk. – UK Medical Research Council.
http://www.nal.usda.gov/fnic. – US Department of Agriculture.

Nutrition guidance for infants: Nutrient-based reference intakes and feeding recommendations

Stephanie A. Atkinson, Department of Pediatrics, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

© 2023 Elsevier Ltd. All rights reserved.

This is an update of S.A. Atkinson, Nutritional requirements of infants, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 250–257, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00158-6.

Introduction 134
Dietary reference intakes for infants 134
Summary of DRIs for infants: macronutrients—energy, carbohydrate, fat, protein, and amino acids 135
Energy 135
Carbohydrate 135
Fat 136
Linoleic acid (*n*-6) and α-linolenic acid (*n*-3) 136
Protein 136
Amino acids 136
Other macronutrients 136
Macrominerals: calcium, phosphorus, magnesium, and fluoride 137
Microminerals/trace elements 137
Fat-soluble vitamins: A, K, E, and D 138
Water-soluble B vitamins, folate, choline, and vitamin C 139
Water and electrolytes 139
Dietary guidelines/recommendations for infants 139
Research needs 141
Conclusion 141
References 142

Key points

- Available quantitative values for nutrient-based Dietary Reference Intakes (DRI) for infants age birth to one year have not been revised since 2005 except for calcium, vitamin D, sodium and potassium.
- In the United States and Canada, recent infant feeding recommendations have been based on evidence reviews with some harmonization across agencies and comparatively have general consistency in their recommendations within and between countries.
- Guidance on infant nutrient requirements and feeding practice recommendations is critical to optimize development during the early critical stages of growth. There is an urgent need to expand the evidence base upon which such guidance is developed.

Glossary

Adequate Intake (AI) The recommended average daily intake level based on observed or experimentally determined approximations or estimates of nutrient intake by a group (or groups) of apparently healthy people that are assumed to be adequate; used when a recommended dietary allowance cannot be determined

Dietary Reference Intakes (DRI) The nutrient reference values developed by the Institute of Medicine that comprise four nutrient-based reference values that can be used to assess or plan the diets of healthy people. The reference values include the four other Glossary terms defined here

Estimated Average Requirement (EAR) The average daily nutrient intake level that is sufficient to meet the nutrient requirements of half of the healthy individuals in a particular life stage and gender group.

Recommended Dietary Allowance (RDA) The average daily nutrient intake level that is sufficient to meet the nutrient requirements of nearly all (97–98%) healthy individuals in a particular life stage and gender group.

Tolerable Upper Level (UL) The highest average daily nutrient intake level that is likely to pose no risk of adverse health effects to almost all individuals in the general population. As intake increases above the UL, the potential risk of adverse effects may increase

Dietary Guidelines for Americans (DGA) The most recent DGA are for 2020–2025

Introduction

Optimal nutrition in the first 1000 days from conception to 24 months of age is a key factor in the attainment of normal trajectories of growth and development and to reduce risk of common non-communicable diseases (NCDs) (Christian et al., 2015; Garmendia et al., 2014). Recently, the importance of nutrition and lifestyle in the preconception period to optimize pregnancy and offspring health has also been emphasized (Stephenson et al., 2018). This article will focus on the state of knowledge of nutrition guidance after birth to two years for normal healthy newborn infants.

Nutrient-based reference values of dietary requirements often referred to as dietary standards are produced by many countries as well as key international agencies such as the Food and Agriculture Organization/World Health Organization (FAO/WHO) (United Nations University, 2004). For infants, the recommended reference values for nutrient intakes are usually intended for term-born, healthy, and normally growing infants who have a birth weight of more than 2500 g (and thus not small for gestational age). In this article, the nutrient requirements outlined reflect the Dietary Reference Intakes (DRIs) for the US and Canada as published by the Institute of Medicine (IOM). With the exception of calcium and vitamin D in 2011 (IOM, 2011) and sodium and potassium in 2019 (NASEM, 2019), the DRIs for all other nutrients have not been updated since 2005, although revision of the macronutrients is to begin in 2022.

The DRIs serve as the scientific basis for the development of many food-based dietary guidelines or recommendations. For infants, such feeding recommendations provide guidance on what, when and how to introduce specific foods to infants have been published by professional societies and some government agencies for many years. The National Academies of Sciences, Engineering and Medicine recently reviewed existing dietary recommendations in its report on Feeding Infants and Children from Birth to 24 Months: Summarizing Existing Guidance. A consensus study report (NASEM, 2020).

This article provides an overview of key concepts and examples of the nutrient-based DRIs specific for infants, the new approaches to establishing food-based feeding recommendations/guidelines for infants, and future needs for additional research to provide for a rigorous evidence-based review to underpin both nutrient requirements and feeding guidelines for infants.

Dietary reference intakes for infants

For infants, evaluation of evidence to establish the DRIs consistently revealed a paucity of appropriate studies on which to base an Estimated Average Requirement (EAR) or tolerable upper level (UL). A Recommended Dietary Allowance (RDA) could not be calculated if a value for the EAR was not established, in which case the recommended intake was based on an Adequate Intake (AI). The nutrient recommendations for infants from birth through 6 months of age for all nutrients except for energy were set as an AI, a value based on "an observed or experimentally determined estimate of nutrient intake by a group of infants who are apparently healthy and assumed to be maintaining normal growth". An AI value does not reflect an average requirement but rather an intake based on approximations or estimates of nutrient intakes that are assumed to be adequate.

For infants from birth to one year of age, human milk was employed as the reference model for setting AI values for breast-fed infants and no specific dietary recommendations were provided for formula-fed infants. The mean intake of a nutrient was calculated based on the average concentration of the nutrient in human milk from 2 to 6 months of lactation using consensus values from several reported studies, multiplied by an average volume (0.780 L day^{-1}) of human milk. The predicted daily volume of breast milk ingested by an infant was based on observational studies that used test weighing of full-term infants. For infants aged 7–12 months, the AI for many nutrients was based on mean observed nutrient intake from human milk in the second 6 months (0.6 L day^{-1}), in addition to the published values for intake of nutrients from complementary or weaning foods if such data were available. It is important to highlight that questions have arisen recently about the values used for both the nutrient composition of human milk and the volume of milk ingested at different ages in setting the original AIs for infants. In a recent review, Allen et al. (2018) compared milk micronutrient values and the assumptions taken to derive micronutrient intakes for infants and lactating women by the IOM, the WHO/FAO and United Kingdom and the European Food Safety Authority and concluded that currently available data on the micronutrient composition of human milk is inadequate to derive reliable reference values for the DRIs and for other countries. This was attributed in part to the fact the data used was based on small sample sizes leading to a wide range of concentrations within studies, limited longitudinal studies and variability in collection and analysis methods (Allen et al., 2018; Hampel et al., 2018). These authors concluded that reference values for human milk composition across lactation need to be developed based on a multi-center study in well-nourished women not reliant on nutrient supplements and followed for 4–6 months of lactation and called for the striking of a new multicenter study to establish reliable reference values for human milk nutrients (Allen et al., 2018).

The approach used to develop an EAR or AI for nutrients for which intake data were not available for those aged older than 6 months, was to extrapolate from intake estimates of older children or adults using the formula with adjustments for metabolic body size, growth, and variability:

$$\mathrm{EAR}_{\text{infant or child}} = \mathrm{EAR}_{\text{child or adult}} \times F,$$

where $F=(\text{weight}_{\text{infant or child}}/\text{weight}_{\text{child or adult}})^{0.75}(1+\text{growth factor})$ or occasionally by extrapolating up from intake of breast-fed infants with similar adjustments using the formula

$$\text{AI}_{6-11\ \text{months}} = \text{AI}_{0-5\ \text{months}} \times F,$$

where $F=(\text{weight}_{6-11\ \text{months}}/\text{weight}_{0-5\ \text{months}})^{0.75}$. The extrapolation model varied across nutrients (Atkinson and Koletzko, 2007). Finally, for a few nutrients, such as iron and zinc, sufficient metabolic data were available to derive an EAR using modeling or factorial methods.

When possible, a DRI called the tolerable UL was defined as "the highest level of daily nutrient intake that is likely to pose no risk of adverse health effects for almost all individuals in the general population" (IOM, 2007). Chronic consumption of nutrients above the UL increases the potential risk of adverse effects, the latter varying by nutrient. For infants, data were only available to reliably estimate ULs for vitamins A and D and the minerals fluoride, selenium, zinc, iron (IOM, 2007), and now revised for vitamin D and calcium in (IOM, 2011). Although adequate data were not available to define an UL for infants for other nutrients, it is important to note that intake nutrients for which an UL does not exist should only be consumed from food or formula and not from supplements. Also notable is that the UL for iron for infants is only relevant to intake from supplements and not foods.

Summary of DRIs for infants: macronutrients—energy, carbohydrate, fat, protein, and amino acids

Energy

The estimated energy requirement (EER) for infants was derived by summing the predicted total energy expenditure (TEE) and energy deposition for growth. Because the energy needs for growth decelerate with advancing age, an equation for EER was established for three age intervals during the first year of life (**Table 1**). The TEE is calculated using an equation based on energy expenditure measured by doubly labeled water and adjusted for weight of the child. The EER is then the sum of TEE for an individual child plus the predicted energy deposition for age (**Table 1**). No adjustment for physical activity was included in the EER for infants. Examples of the EER for males and females using the reference weights are shown in **Table 1** for infants at five age intervals during the first year of life. At most ages beyond the first 2 months of life, the values for EER exceed the average energy provided (500 kcal) by human milk assuming a volume of intake (0.780 L day^{-1}) from human milk.

The energy requirements of infants during the first year of life recommended in the FAO/WHO/UNU report (United Nations University, 2004) are remarkably similar to that of the DRI report (IOM, 2005) as the approaches included the same parameters. In both the reports, estimates of TEE were based on the analysis of experiments using doubly labeled water and calculated by monthly age intervals, to which an assessment of energy deposition for growth was added.

Carbohydrate

The AI for carbohydrate for infants through 1 year of age is based on the average carbohydrate intake from human milk and complementary foods for the 7–12-month age group (**Table 2**). Although the carbohydrate from human milk is almost exclusively lactose and that from infant's formula may be lactose, sucrose, or glucose polymers alone or in combination, there is no evidence that non-lactose-containing formulas vary from lactose present in human milk with regard to available energy.

Table 1 DRI estimated energy requirement (EER) for infants (IOM, 2005).

Equations	
0–3 months	(89 × weight of infant (kg)−100)+175 (kcal for energy deposition)
4–6 months	(89 × weight of infant (kg)−100)+56 (kcal for energy deposition)
7–12 months	(89 × weight of infant (kg)−100)+22 (kcal for energy deposition)

Calculated EER for age using reference weights for age		
Age (months)	*Males (kcal day^{-1})*	*Females (kcal day^{-1})*
1	472	438
3	572	521
6	645	593
9	746	678
12	844	768

DRI, Dietary Reference Intake; EER, Estimated Energy Requirement.

Table 2 DRI for macronutrients for infants—carbohydrate, protein, fat, and essential fatty acids (IOM, 2005).

	0–6 months	*7–12 months*
Nutrient[a]	*g day*$^{-1}$	
Carbohydrate, AI	60	95
Protein		
AI	9.1	–
RDA	–	13.5
Total fat, AI	31	30
Linoleic acid (*n*-6), AI	4.4	4.6
α-Linolenic acid (*n*-3), AI	0.5	0.5

DRI, Dietary Reference Intake; AI, Adequate Intake; RDA, Recommended Dietary Allowance.
[a]No upper levels of nutrients were set for any macronutrients.

Fat

As for other nutrients, the AI for fat intake is based on the average intake of fat from human milk alone or in addition to complementary foods after 7 months of age (Table 2). Although infant formulas are designed to contain a percentage of energy as fat similar to human milk (approximately 50%), the type of fat in formulas varies widely, including sources such as safflower, sunflower, soybean oil, and coconut and palm oils, usually in some combination.

Linoleic acid (n-6) and α-linolenic acid (n-3)

The *n*-6 (linoleic) and *n*-3 (linolenic) fatty acids are essential for maturation of the developing brain, retina and other organs both in utero and in early life of infants. Linoleic acid serves as a precursor of arachidonic acid (AA) and linolenic acid as a precursor of docosahexaenoic acid (DHA), which collectively are referred to as the long chain polyunsaturated fatty acids (LCPUFA). While humans have the capacity to metabolize the precursor fatty acids to AA and DHA, the capacity for such endogenous synthesis appears limited in the fetus and newborn infant during rapid development. Human milk is a natural source of both fatty acid families, including the long-chain polyenoic derivatives DHA and AA. The pattern of all fatty acids in human milk, including the polyenoic fatty acids, is dependent on the maternal diet. The AI established for infants for *n*-6 and *n*-3 fatty acids is based on the average content in human milk reported for North American women with the addition of that from complementary foods during months 7–12 (Table 2).

Since the DRI for *n*-6 and *n*-3 fatty acids DRI was developed in 2005 (IOM, 2005) numerous investigations have explored whether supplementation of infant formula with LCPUFA is of benefit to the development of infants. In a systematic review and meta-analysis of 15/31 randomized trials identified (N = 1889), for infants born full-term no beneficial effects or harms of the added LCPUFAs on neurodevelopment outcomes nor consistent benefits to visual acuity were identified (Jasani et al., 2017). While this suggests LCPUFAs are not necessary additions to term infant formula, many marketed formulas are available with added DHA and AA.

Protein

For infants aged 0–6 months, the AI for protein is based on the intake from human milk. For infants aged 7–12 months, sufficient information was available from nitrogen balance studies and protein deposition to derive an EAR based on the factorial method. For both males and females, this averaged to 1.1 g protein per kg body weight day^{-1}. The RDA was set as the EAR+2 standard deviations (based on coefficients of variation observed in adults), which yielded a value for protein intake of 1.5 g kg^{-1} day^{-1}. Because the absorption and digestibility of protein contained in the infant formula may be less efficient than that from the human milk, the quantity of protein contained in infant formulas may have to be adjusted depending on the protein source used (Table 2).

Amino acids

The DRI for the essential (indispensable) amino acids for infants was derived from the content of human milk for ages 0–6 months. For older infants, an EAR was derived for these amino acids using a factorial estimate that was based on the amino acid needs for growth or protein deposition, with adjustments for efficiency of protein deposition and maintenance requirement. The RDA was determined by adding the coefficient of variation derived for maintenance and protein deposition to the value for the EAR. No values were set for UL for any of the amino acids. A summary of the AI and RDA for the indispensable amino acids of infants is provided in Table 3.

Other macronutrients

For infants, no DRI was set for saturated fat, monounsaturated fat, *trans*-fatty acids, cholesterol or dietary fiber. Although some dietary fiber is present in the diet after solid foods are introduced, there are no data on fiber intakes in such young age groups and no theoretical basis exists, which establishes a need for fiber at less than 1 year of age.

Table 3 DRI for indispensable (essential) amino acids (IOM, 2005).

Amino acid[a]	*0–6 months AI (mg kg^{-1} day^{-1})*[b]	*7–12 months RDA (mg kg^{-1} day^{-1})*
Histidine	23	32
Isoleucine	88	43
Leucine	156	93
Lysine	107	89
Methionine + cysteine	59	43
Phenylalanine + tyrosine	135	84
Threonine	73	49
Tryptophan	28	13
Valine	87	58

DRI, Dietary Reference Intake; AI, Adequate Intake; RDA, Recommended Dietary Allowance.

[a]No upper levels were set for any of the indispensable amino acids.

[b]AI values shown as amino acid in mg kg^{-1} day^{-1} can be converted to milligram amino acid per day by multiplying with the reference weight of 6 kg for infants 0–6 months of age.

Macrominerals: calcium, phosphorus, magnesium, and fluoride

The AI for infants for the "bone" minerals are summarized in Table 4. The content of human milk was used as the basis to derive the AI for calcium, phosphorus, and magnesium for infants aged 0–6 months and with the addition of intake from complementary foods for those aged 7–12 months (IOM, 1998). For calcium, the values were updated in 2011 (IOM, 2011) but only changed to 200 mg day^{-1} compared to 220 mg day^{-1} likely due to a slightly different value used for the concentration of calcium in breast milk. The AI value of 200 mg day^{-1} for breast-fed infants from 0 to 6 months was substantiated by considering that average calcium absorption in infants is around 60% thus yielding retention of calcium of 120 mg day^{-1}, a value approximately 20% higher than the estimated accretion of calcium for an infant of about 100 mg day^{-1}. An UL for calcium for infants 0–6 months was set at 1000 mg day^{-1}, a conservative estimate given the paucity of data available (IOM, 2011). For infants aged 6–12 months, recent data on calcium intakes from solid foods for formula-fed infants were used to add to the intake from breast milk to yield an AI of 260 mg day^{-1} (IOM, 2011). A UL was set at 1500 mg calcium/day although based on only one study (IOM, 2011).

For fluoride, intake from human milk was the reference for the first 6 months only (IOM, 1998). After 6 months, the AI for fluoride was set at 0.5 mg day^{-1} based on the evidence of the benefit of fluoride intake for the prevention of dental caries (Table 4). The UL of 0.7 mg day^{-1} for 0–6 months, and 0.9 mg day^{-1} was based on risk of developing fluorosis of the anterior teeth.

Microminerals/trace elements

The DRIs for microminerals and trace elements as published in 2001 (IOM, 2001) are summarized in Table 5. The AI for iron for ages 0–6 months is based on the concentration of iron in human milk albeit low (approximately 0.35 mg L^{-1}) but assumes that the infant is born with maximal iron stores due to transplacental transfer of iron from an iron-replete mother. If the latter conditions do not apply, then an exogenous source of iron such as iron drops may be required. For infants aged 7–12 months, an EAR and RDA were developed based on a factorial modeling method that summed basal loss of iron with needs for growth, increasing hemoglobin mass, and iron stores. This value was then adjusted for iron bioavailability using a factor of 10% for infants due to a medium bioavailability of iron from infant cereals, which are generally the major dietary source of iron in weaning foods before meats are introduced. An UL was established (Table 5) for iron based on the risk of adverse gastrointestinal side effects from supplemental (not food) iron.

For zinc, an AI was based on the human milk model only for the 0–6 months age group (Table 5). The zinc content of human milk declines rapidly during the first 6 months (from 4 to 1.2 mg L^{-1}), so the AI was based on a milk zinc concentration of 2.5 mg L^{-1}. This value cannot be directly applied to infants being fed with cow milk- or soy-based infant formulas because zinc

Table 4 DRI for minerals for infants—calcium and phosphorus (IOM, 2011), magnesium and fluoride (IOM, 1997).

	0–6 months	*7–12 months*
Nutrient	*AI (mg day^{-1})*	
Calcium	200	260
Phosphorus	100	275
Magnesium	30	75
Fluoride	0.01	0.5

DRI, Dietary Reference Intake; AI, Adequate Intake; UL, Upper Level was not determinable due to lack of data of adverse effects in infants except for fluoride which was set at 0.7 mg day^{-1} for 0-6 months and 0.9 mg day^{-1} for 7-12 months.

Table 5 DRI for micronutrient/trace minerals for infants—chromium, copper, fluoride, iodine, iron, manganese, molybdenum, selenium, and zinc (IOM, 2001).

	0–6 months	*7–12 months*
Nutrient	*mg day^{-1}*	
Chromium		
AI	0.2	5.5
UL	ND	ND
Copper		
AI	200	220
UL	ND	ND
Iodine		
AI	110	130
UL	ND	ND
Iron		
AI	0.27	–
RDA	–	11
UL	40	40
Manganese		
AI	30	75
UL	ND	ND
Molybdenum		
AI	2	3
UL	ND	ND
Selenium		
AI	15	20
UL	45	60
Zinc		
AI	2	–
RDA	–	3
UL	4	5

DRI, Dietary Reference Intake; AI, Adequate Intake; ND, Not Determinable due to lack of data of adverse effects in infants; RDA, Recommended Dietary Allowance; UL, Upper Level.

absorption is significantly lower from these feeds compared with human milk. The EAR for the 7–12 months age group was set using a factorial method that summed obligatory losses with requirements for growth and adjusted for fractional absorption of dietary zinc from human milk and complementary foods. The RDA was derived by adding twice the coefficient of variation of 10% to the EAR (2.5 mg day^{-1} of zinc) for infants aged 7–12 months (Table 5). An UL was set for zinc based on the possibility of an adverse effect of high-zinc intakes on copper status.

For the trace elements, such as chromium, copper, iodine, manganese, molybdenum, and selenium, an AI was set for infants of age 0–6 months based on the human milk model (Table 5). For the age group 7–12 months, data on intake from complementary foods were only available to set an AI for chromium, copper, and selenium (Table 5). For iodine and molybdenum, the AI represents an extrapolation up from the AI values for the age group 0–6 months based on differences in metabolic body weight (kg$^{0.75}$). For manganese, the AI represents an extrapolation down from the AI for adults as described previously (Table 5). Owing to the lack of relevant information, no UL values for infants younger than 1 year were established for chromium, copper, iodine, manganese, or molybdenum, but intakes of these nutrients should be limited to foods and not supplements. An UL was established for selenium due to the known chronic toxicity of excessive selenium ingestion, which is manifested clinically as brittleness and loss of nails and hair. The UL was set for infants based on the highest known intake of selenium from human milk and adjusting for a reference infant weight (Table 5). The UL value pertains to intake from both foods and supplements.

The trace elements such as arsenic, boron, nickel, silicon, and vanadium are known to have a role in human metabolism, but due to lack of information, DRI values, including UL, could not be established for infants.

Fat-soluble vitamins: A, K, E, and D

For the DRIs for vitamins A, K, (IOM, 2001) and for vitamin E (IOM, 2000) (Table 6), the AI for infants aged 0–6 months was based on the human milk model as previously described. For vitamins A, K, and E, the AI for infants aged 7–12 months was extrapolated up from the values for infants aged 0–6 months using a reference weight for infants at this age. Notably, the AI was set assuming that the infants had received a prophylactic injection of vitamin K just after birth, a routine in North America since vitamin K is not readily transferred to the fetus while in utero, and human milk is relatively low in vitamin K.

Table 6 DRI for fat-soluble vitamins (IOM, 2000, 2001, 2011).

Nutrient	*0–6 months*	*7–12 months*
Vitamin A (mg day^{-1})		
AI	400	500
UL	600	600
Vitamin D (IU day^{-1})		
AI	400	400
UL	1000	1500
Vitamin E (mg day^{-1})		
AI	4	5
UL	ND	ND
Vitamin K (mg day^{-1})		
AI	2.0	2.5
UL	ND	ND

DRI, Dietary Reference Intake; AI, Adequate Intake; ND, Not Determinable due to lack of data of adverse effects in infants; UL, Upper Level.

A revision of the AI for vitamin D was made in 2011 (IOM, 2011). This AI was not based on the content of human milk as it contains only marginal amounts of vitamin D. The AI for vitamin D of 10 μg (400 IU) day^{-1} was set for infants from 0 to 12 months, based on this intake being associated with maintenance of serum 25-hydroxyvitamin D level higher than 30 nmol L^{-1} and likely closer to 50 nmol L^{-1}, which represents a vitamin D status that is above that usually associated with clinical rickets in infants. For breast-fed infants to meet the AI of 400 IU of vitamin D per day, they must be provided with a vitamin D supplement. For formula-fed infants, intake of nearly 1000 mL day^{-1} is required to achieve the AI for vitamin D because infant formulas in North America are regulated to contain 400 IU L^{-1} of liquid formula. The UL of vitamin D intake is 25 μg (1000 IU) day^{-1} for 0–6 months and 38 μg (1500 IU) day^{-1} for 7–12 months (**Table 6**) which included using an uncertainty factor of 0.5 (IOM, 2011).

Water-soluble B vitamins, folate, choline, and vitamin C

The AIs for infants aged 0–6 months for most water-soluble vitamins (**Table 7**) were based on the content of human milk (IOM, 1998). This approach may be problematic for water-soluble B vitamins, in which the milk content is dependent on maternal intake of vitamins (Allen et al., 2018; Hampel et al., 2018). An example of clinical relevance is a vegan mother who may have subclinical vitamin B_{12} deficiency and produce vitamin B_{12}-deficient milk. For vitamin C, the effect of maternal supplementation on milk content remains uncertain, but available reports do not indicate that excessive amounts of vitamin C are secreted in milk, even in mothers taking supplements of 1000 mg or more. For those aged 7–12 months, the AI for thiamin, riboflavin, niacin, folate, pantothenic acid, and choline was derived by extrapolating down from values for older children or adults due to a lack of information of dietary intake of these nutrients from solid foods. Tolerable ULs for infants were not established for any of the water-soluble vitamins.

Water and electrolytes

Optimal water intake in infants is more critical than at any other period of life. Not only do infants have a higher total body water content per body mass than children or adults but also they have a higher water turnover rate, a less well-developed sweating mechanism, and little ability to indicate when they are thirsty. The AI for water intake of infants aged 0–6 months is 0.7 L day^{-1} and is based on the water content of human milk. Assuming that the infants are breast fed on demand, infants will drink to meet their thirst needs; thus, even in hot and humid climates, supplemental water may not be required. The AI for water intake of 0.8 L day^{-1} set for infants aged 7–12 months is based on the sum of the water content of human milk, complementary foods, and beverages, the latter obtained from reported food intakes from surveys in the USA (IOM, 2005).

For sodium and potassium, the updated values for infants 0–6 months established in 2019 (NASEM, 2019) remain based on the human milk model with an AI of 110 mg day^{-1} for sodium and 400 mg day^{-1} for potassium. For 7–12 months, the AI for sodium is 370 mg day^{-1} and for potassium 860 mg day^{-1} based on the sum of observed intakes from human milk and complementary foods. No ULs were established for infants due to the lack of data on adverse effects of these nutrients on infant health. However, particularly because the renal excretory capacity of young infants may not be able to handle excessive amounts of ingested electrolytes, the DRI report notes that the intake of sodium, chloride, and potassium should be limited to human milk (or infant formula) and solid foods appropriate for age.

Dietary guidelines/recommendations for infants

Dietary guidelines or recommendations for infants provide food-based information on what, when and how to introduce complementary foods to infants and young children. The introduction of complementary foods, especially solids and eventually finger foods, is important for infants to develop normal oral and motor skills related to eating and to achieve the recommended nutrient requirements that may be low in breast milk (e.g., protein or iron). Such dietary guidelines are generally targeted to health care

Table 7 DRI for water-soluble vitamins (IOM, 1998).

	0–6 months	*7–12 months*
Nutrient	*mg day^{-1}*	
Vitamin C		
AI	40	50
UL	ND	ND
Thiamin		
AI	0.2	0.3
UL	ND	ND
Riboflavin		
AI	0.3	0.4
UL	ND	ND
Niacin		
AI	2	4
UL	ND	ND
Vitamin B_6		
AI	0.1	0.3
UL	ND	ND
Folate		
AI	65	80
UL	ND	ND
Vitamin B_{12}		
AI	0.4	0.5
UL	ND	ND
Pantothenic acid		
AI	1.7	1.8
UL	ND	ND
Biotin		
AI	5	6
UL	ND	ND
Choline		
AI	125	150
UL	ND	ND

DRI, Dietary Reference Intake; AI, Adequate Intake; Upper Level (UL) was not set for any water-soluble vitamin as a value was deemed not determinable (ND) due to lack of data of adverse effects in infants.

providers (e.g., physicians, nurse practitioners, dentists, dietitians) and or parents and guardians, early care and education providers, program administrators, or a combination of audiences.

The recent consensus study report from NASEM on Infant Feeding provided an insightful overview of existing food-based dietary guidelines that included review of 43 guideline documents on recommendations for what, when and how to feed infants and young children that were available on the internet from 26 agencies in 7 high income countries plus the WHO (NASEM, 2020). Overall, there was a high degree of consistency across countries in the recommendations on various aspects of infant as detailed in the report (NASEM, 2020) and summarized by Jimenez and colleagues (Jimenez et al., 2021). Canada (Health Canada et al., 2014, 2015) and the US (USDA and USDHHS, 2020) have some of the most recently published infant feeding recommendations. In general, such guidance on infant feeding has not been based on use of systematic review methodology and infrequently has involved a harmonized approach across agencies or professional societies either within a country or across countries. However, a trend is emerging for inclusion of both methodological processes. In the US, for the first time, the 2020–2025 Dietary Guidelines for Americans included nutrition guidance for pregnant women, infants and young toddlers from birth to 24 months which are based on systematic reviews that addressed specific questions and were published in peer reviewed nutrition journals (USDA and USDHHS, 2020). The 2014–2015 Canadian nutrition recommendations for infants from birth to 24 months were not developed using systematic reviews but were based on a duplicate search of the literature, followed by a consensus process that included extensive citations to substantiate the report's feeding recommendations which were targeted to health professionals, parents and caregivers. The Canadian recommendations are, however, an excellent example of a harmonized effort across agencies as they were developed by an Infant Feeding Joint Working Group, comprised of representatives from Health Canada, the Canadian Pediatric Society, the Dietitians of Canada, and the Breastfeeding Committee for Canada (Health Canada et al., 2014, 2015).

Despite the different underlying methods to develop the infant feeding recommendations, there was great consistency in key recommendations for most topic areas between the US and Canada as summarized in **Table 8**. Indeed, based on the summary of existing guidance on infant feeding across 26 agencies in seven countries conducted by NASEM (NASEM, 2020; Jimenez

Table 8 Comparative examples of current infant feeding recommendations between Canada (Health Canada et al., 2014, 2015) and the United States (USDA and USDHHS, 2020).

Recommendation topic	*Health Canada 2014 and 2015*	*Dietary Guidelines for Americans 2020–2025*
Age grouping	Specific recommendations for: - Birth–6 months - 6–24 months	Recommendations for birth to 23 months - Separate section on eating patterns for 12–23 months
Guidance on breastfeeding	Exclusive breastfeeding for 6 months; continued up to 2 years or longer with appropriate foods	Exclusive breastfeeding for about the first 6 months; continued to at least 1 year or longer if desired
Supplements	Vitamin D of 400 IU (10 mcg)/day with breastfeeding	Vitamin D of 400 IU (10 mcg)/day with breastfeeding
First complementary foods; timing and type	Begin at 6 months with a focus on iron-rich meat, meat alternatives, and iron-fortified cereal	Begin at about 6 months with nutrient dense foods rich in iron and zinc; to not introduce foods before 4 months or after 6 months Tips on responsive feeding to offer developmentally appropriate foods
Potentially allergenic foods	Foods containing common allergens (peanuts, fish, wheat, cow milk products, soy and eggs) may be fed from about 6 months	Introduction of allergenic foods (peanuts, egg, cow milk products, tree nuts, wheat, fish and soy) should be introduced
Cow milk as main beverage	Not before 9–12 months	Not until 12 months or later

et al., 2021), there were more consistencies than inconsistencies across recommendations. Any differences reflect slight variations in age ranges suggested.

Research needs

For nutrient-based dietary reference values for infants, there is an urgent need for revision since only the DRI for calcium, vitamin D, sodium and potassium have been revised since the first DRI report in 1997. The first step is to conduct evidence reviews to determine if sufficient new research is available to substantiate revising individual nutrient requirements. Each of the individual DRI reports highlighted research needs in a concluding article but whether such research has evolved remains to be determined. Because infants and children are not just "little adults," the DRI values must be carefully defined for the specific stages of growth and development and with consideration for nutritional programming that occurs in early life in response to dietary exposures as our knowledge of this area becomes more complete.

With regard to food-based infant feeding guidelines, evidence gaps related to lack of existing evidence or inconsistencies in existing research were identified in the NASEM report on infant feeding (NASEM, 2020) and summarized in a recent paper (Atkinson et al., 2021). Areas identified for future research include the following:

- the effects of breastfeeding beyond the first year
- the consequences of replacing infant formula with cow's milk at 9 months vs 12 months, among formula-fed infants and partially breastfed infants
- how to identify and appropriately treat infants at risk of iron deficiency
- does high dose vitamin D supplementation in lactating women meet infant needs
- how to achieve nutritional adequacy for infants weaned to vegetarian or vegan diets
- how to provide adequate iodine without excessive exposure to salt
- responsive feeding research to support guidance on elucidating reliable child development indicators that identify when infants are ready to be introduced to complementary foods.

Conclusion

Guidance on the specific nutrient needs and on how to meet those nutrient needs by types of foods and feeding practices for infants and young children is critical to ensure optimal growth and development, especially in the first two years of life (Darling et al., 2020; Stephenson et al., 2018). This requires both nutrient-based dietary reference values that provide quantitative amounts of nutrients such as the DRIs, and nutrition recommendations on what, when and how to feed infants. The DRIs are the scientific basis for the translation of the quantitative values into amounts of food (the what), for specific age groupings according to growth needs and developmental functioning (the when). Nutrition guidelines or recommendations must extend beyond these topics to provide guidance on the form and presentation manner of food for infants (the how). The paucity of past research in infant feeding has limited the scientific rigor with which both the DRIs and infant feeding recommendations were developed but there are signs the state of the science is improving. Global harmonization of derivation of nutrient-based requirements is encouraged (Yaktine et al., 2020). Harmonization of infant feeding recommendations/guidelines within countries (such as done in Canada) and across

countries would elevate translation and dissemination of information to a new level. The communication, dissemination and implementation of guidelines is critical and must consider ethnicity, socioeconomic status, as well as regional food availability and security (Jimenez et al., 2021).

References

Allen, L.H., Donohue, J.A., Dror, D.K., 2018. Limitations of evidence base used to set recommended nutrient intakes for infants and lactating women. Adv. Nutr. 9 (Suppl. l), 295S–312S.

Atkinson, S.A., Koletzko, B., 2007. Determining life-stage groups and extrapolating nutrient intake values (NIVs). Food Nutr. Bull. 28 (1 Suppl), S61–S76.

Atkinson, S.A., Jimenez, E.Y., Perez-Escamilla, R., 2021. Evidence gaps and research needs in current guidance on feeding children from birth to 24 months. Appl. Physiol. Nutr. Metabol. 46 (3), 294–297.

Christian, P., Mullany, L.C., Hurley, K.M., et al., 2015. Nutrition and maternal, neonatal, and child health. Semin. Perinatol. 39 (5), 361–372.

Darling, J.C., Bamidis, P.D., Burberry, J., Rudolf, M.C.J., 2020. The first thousand days: early, integrated and evidence-based approaches to improving child health: coming to a population near you? Arch. Dis. Child. 105 (9), 837–841.

Garmendia, M.L., Corvalan, C., Uauy, R., 2014. Assessing the public health impact of developmental origins of health and disease (DOHaD) nutrition interventions. Ann. Nutr. Metabol. 64 (3–4), 226–230.

Hampel, D., Dror, D.K., Allen, L.H., 2018. Micronutrients in human milk, analytical methods. Adv. Nutr. 9 (Suppl. 1), 313S–331S.

Health Canada, Canadian Paediatric Society, Dietitians of Canada, Breastfeeding Committee for Canada, 2014. Nutrition for Healthy Term Infants: Recommendations From Birth to Six Months. https://www.canada.ca/en/health-canada/services/canada-food-guide/resources/infant-feeding/nutrition-healthy-term-infants-recommendations-birth-six-months.html.

Health Canada, Canadian Paediatric Society, Dietitians of Canada, Breastfeeding Committee for Canada, 2015. Nutrition for Healthy Term Infants: Recommendations From Six to 24 Months, 2014. https://www.canada.ca/en/health-canada/services/canada-food-guide/resources/infant-feeding/nutrition-healthy-term-infants-recommendations-birth-six-months/6-24-months.html.

Institute of Medicine, 1997. Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D and Fluoride. National Academy Press, Washington, DC. Available at. http://www.nap.edu/catalog/5776.html.

Institute of Medicine, 1998. Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline. National Academy Press, Washington, DC.

Institute of Medicine, 2000. Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids. National Academy Press, Washington, DC.

Institute of Medicine, 2001. Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc. National Academy Press, Washington, DC.

Institute of Medicine, 2005. Dietary Reference Intakes for Water, Potassium, Sodium, Chloride, and Sulfate. The National Academies Press, Washington, DC.

Institute of Medicine, 2005. Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids. The National Academies Press, Washington, DC.

Institute of Medicine, 2007. Dietary Reference Intakes: The Essential Guide to Nutrient Requirements. National Academy Press, Washington, DC.

Institute of Medicine, 2011. Dietary Reference Intakes for Calcium and Vitamin D. The National Academies Press, Washington, DC.

Jasani, B., Simmer, K., Patole, S.K., Rao, S.C., 2017. Long chain polyunsaturated fatty acid supplementation in infants born at term (Review). Cochrane Database Syst. Rev. (3) Art. Mp.: CD000376.

Jimenez, E.Y., Atkinson, S.A., Perez-Escamilla, R., 2021. Existing guidance on feeding infants and children from birth to 24 Months: implications and next steps for registered dietitian nutritionists. J. Acad. Nutr. Diet. 121 (4), 647–654.

National Academies of Sciences, Engineering, and Medicine (NASEM), 2019. Dietary Reference Intakes for Sodium and Potassium. The National Academies Press, Washington, DC. https://www.nap.edu/catalog/25353/dietary-reference-intakes-for-sodium-and-potassium.

National Academies of Sciences, Engineering and Medicine (NASEM), 2020. Feeding Infants and Children From Birth to 24 Months: Summarizing Existing Guidance. The National Academies Press, Washington, DC. https://www.nap.edu/read/25747/chapter/1.

Stephenson, J., Heslehurst, N., Hall, J., Schoenaker, D.A.J.M., Hutchinson, J., Cade, J.E., Poston, L., Barrett, G., Crozier, S.R., Barker, M., Kumaran, K., Yajnik, C.S., Baird, J., Mishra, G.D., 2018. Before the beginning: nutrition and lifestyle in the preconception period and its importance for future health. Lancet 391, 1830–1841.

United Nations University, World Health Organization and Food and Agriculture Organization of the United Nations, 2004. Human Energy Requirements. https://www.fao.org/3/y5686e/y5686e00.htm.

U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2020. Dietary Guidelines for Americans 2020–2025. 9th Edition. December 2020. Available at: https://www.dietaryguidelines.gov/sites/default/files/2021-03/Dietary_Guidelines_for_Americans-2020-2025.pdf

Yaktine, A.L., King, J.C., Allen, L.H., 2020. Why the derivation of nutrient reference values should be harmonized and how it can be accomplished (Review). Adv. Nutr. 11, 1102–1107.

Nutritional problems of adolescents

S. Gordon[a] **and C.H.S. Ruxton**[b], [a] Ulster University, Coleraine, Northern Ireland; and [b] Nutrition Communications, Cupar, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of E.W. Evans, C. Lo, Nutritional Problems of Adolescents, In Encyclopedia of Human Nutrition (Third Edition), edited by Benjamin Caballero, Academic Press, 2013, https://doi.org/10.1016/B978-0-12-375083-9.00002-7.

Introduction 143
Physical growth 143
Obesity 144
Eating patterns 144
Less healthy options 145
Body image and eating disorders 145
Vegan and vegetarian diets 146
Nutrient inadequacy and potential consequences 146
Nutrients of concern: iron 146
Nutrients of concern: bone health 147
Nutrients of concern: folate 147
Nutrients of concern: iodine 147
Improving adolescents' diets 147
Conclusion 148
Appendix 1: examples of growth charts for adolescents 148
References 148
Further reading 149

Key points

- Adolescence marks a stage of physical, mental and intellectual development, with nutrient recommendations being higher than at any other stage throughout the life cycle.
- Increased nutritional requirements in adolescence, combined with a tendency toward less healthy eating patterns, make this group susceptible to vitamin and mineral inadequacies.
- Adolescence is an opportune time to instill healthy eating patterns that can persist into adulthood.
- Obesity in adolescence is not just restricted to high-income countries and has short and long-term implications for the individual.

Introduction

There are approximately 1.2 billion adolescents worldwide. Although they differ by nationality, culture, education and wealth, adolescents share similar developmental experiences. The normal transition from childhood to adolescence consists of hormonal changes, physical growth, sexual development and an expansion of emotional and cognitive skills. In addition, adolescents begin to seek independence and develop a social life that is separate from the family unit, both of which will influence eating habits. Skipping breakfast and eating away from the home are common among adolescents, as are the consumption of high energy snacks, drinks and fast foods, particularly in richer nations. These foods, however, are limited in important vitamins and minerals; and can be energy-dense which contributes to excess weight gain. As adolescence is a time of peak growth requiring additional energy, protein and key micronutrients, optimal nutrition is essential. There is also a need for nutrition education to guide adolescents toward balanced dietary habits that support future disease risk reduction, and can be maintained during adulthood.

Physical growth

Adolescence refers to the life stage from approximately 10 to 19 years during which there are changes in physical, cognitive, social and emotional development. Adolescence can be divided into two stages, early and late adolescence. During early adolescence, there is enhanced physical growth (height and weight), pubertal changes in males and females, and the onset of menstruation in girls. In late adolescence, physical growth continues for boys and girls, though at a much slower rate in girls.

Monitoring these changes can be achieved using growth charts. The World Health Organization (WHO) growth charts, originally designed for children up to 5 years of age, are universal as there are similar patterns of growth among subpopulations exposed to similar environments. However, genetic, nutritional and maternal factors can impact the growth of children and adolescents. Therefore, some countries have developed their own growth charts for adolescents, such as Canada, the United States and the United Kingdom. Examples of these are provided in **Appendix 1: examples of growth charts for adolescents** section.

Before puberty, girls and boys acquire bone mass at a similar rate. After puberty, boys tend to acquire a greater amount of bone mass than girls. Peak bone mass is the maximum amount of bone tissue a person achieves throughout their life, usually by age 18–25 years. Higher bone mass is thought to be protective against the risk of osteoporosis later in life—a condition that mainly affects females but also some males (National Institute of Health, 2018a). Peak bone mass is determined mainly by genetics, sex, nutritional intake, endocrine status, physical activity and health status during growth. With regards to nutrition, bone formation requires amino acids and an adequate supply of energy. The main bone forming minerals include calcium, phosphorus, magnesium and iodine. Also, vitamins C, D and K are required for the formation of mineral crystal and collagen—a type of protein found in the bone matrix. These vitamins are also involved in cartilage formation and bone metabolism. Other recommendations to optimize bone development include having a body weight in the healthy range and being physically active (Redmond and Schoenmakers, 2019). The WHO recommends children and adolescent to do at least 1 h per day of moderate to vigorous intensity, mostly aerobic physical activity. Exercises that strengthen muscle mass and bone should be incorporated into adolescent's lifestyles at least 3 times per week (WHO, 2020b).

Obesity

Overweight and obesity refer to excessive fat accumulation that may impair an individual's health. Body mass index (BMI) is a commonly used proxy of weight-for-height to classify overweight and obesity and is calculated as an individual's weight (kg) divided by height squared (m^2). BMI is typically used for assessing overweight and obesity in adults but is not useful on its own in children and adolescents due to the continuing growth that occurs throughout this period. Instead, BMI is only meaningful when plotted on sex- and age-specific growth charts. For adolescents, overweight is defined as BMI-for-age that is greater than one standard deviation above the WHO growth reference median. Obesity is defined as two or more standard deviations above the reference median. In country-specific growth charts, such as those for the UK, overweight and obesity are defined as being above the 91st and 98th centiles, respectively. In epidemiological studies, obesity and overweight may be defined at the 95th and 85th centiles, respectively (Stewart and Gillespie, 2019). Section **Appendix 1: examples of growth charts for adolescents** summarizes growth chart cut-offs for overweight and obesity for a selection of countries.

In 2016, over 340 million children and adolescents were overweight or obese and, according to the WHO report (2020a) on childhood obesity, the issue affects every country in the world. The most rapid growth is occurring in low- and middle-income countries, particularly in the Middle East, Southern Africa and the Pacific Islands. The number of young people aged 5–19 years living with obesity is estimated to rise to over 250 million by 2030, with China and India accounting for over a third of this increase. In the United States, approximately 20% of 6–19-year-olds are defined as obese.

Although the prevalence of obesity is starting to flatten in many countries, especially in Europe, most children and adolescents will not outgrow the condition. A longitudinal study that tracked BMI from childhood revealed that nearly 80% of children with severe obesity at one examination remained severely obese at the next (Freedman et al., 2018). Likewise, a meta-analysis of 37 studies investigated the use of BMI to predict the likelihood that obese children will becomes obese adults. The authors found that obesity persisted from childhood onwards, with most obese adolescents continuing to be obese in adulthood (Simmonds et al., 2015). The health consequences of overweight and obesity have been well documented. Obesity increases the risk of developing several non-communicable diseases, namely hypertension, coronary heart disease, type 2 diabetes and sleep apnea. Longitudinal research highlighted that adults have a better cardiovascular health profile if they were never overweight or obese, or at least only in childhood or adolescence (Quinte et al., 2019). In addition, adolescents can experience behavioral and emotional problems related to obesity, affecting their quality of life and societal opportunities.

A position paper by a European pediatric expert group (Verduci et al., 2021) concluded that certain dietary patterns enhanced the risk of obesity including consumption of SSB (but not 100% fruit juices at appropriate intakes), breakfast skipping, large portion sizes, snacking on energy-dense foods, sedentary behavior (including excess use of technology) and lack of regular physical activity. The position paper made recommendations to improve young people's diets as a means of preventing obesity. However, many adolescents around the world face obstacles accessing and affording healthy diets (UNICEF, 2019).

Eating patterns

Eating a varied and balanced diet and establishing healthy eating patterns can have a significant effect on adolescents' health by optimizing growth, supporting normal physical and intellectual performance, and helping to maintain a healthy weight. In contrast, there are a number of unhealthy eating patterns common in adolescence which can hinder health, such as skipping breakfast, excess consumption of fast foods and low consumption of fruit and vegetables. These eating patterns are influenced

by a range of factors including food preferences, cost, convenience and peer influences. Since eating habits and food preferences established in adolescence tend to continue into adulthood, it is a very opportune time to promote healthful eating patterns (WHO, 2016).

Less healthy options

In developed countries, adolescents' diets are characterized by less healthy food choices, especially sugar sweetened beverages (SSB), fast foods and high fat/sugar/salt snacks. Skipping breakfast is associated with an overall lower intake of folic acid, vitamin D, calcium and iron, as well as fiber. However, adolescents who regularly consume breakfast have been found to snack less and have a lower likelihood of being overweight or obese. Moreover, research (systematic reviews) has shown that breakfast consumption has beneficial effects on cognitive performance and academic outcomes. In spite of this, skipping breakfast is very common among adolescents worldwide with boys being more likely to eat breakfast than girls. At age 11, approximately 70% of boys and girls eat breakfast every weekday but by age 15, this reduces to 62% and 52% for boys and girls respectively (Dye, 2017; WHO, 2016).

Fast food is frequently consumed by adolescents and can be energy dense, high in sodium and sugar, and low in fiber and micronutrients. The prevalence of fast food consumption was examined in a large study with over 153,000 young adolescents from 54 low- and middle-income countries. The results showed that 55% ate fast food at least one day per week, while 10% ate fast food 4–7 days per week (Li et al., 2020). During 2015–2018, over 35% of children and adolescents in the US consumed fast food on a given day (Ostchega et al., 2020). In addition to the direct effects of consuming fast food, this dietary pattern is linked with higher consumption of SSB and low consumption of fruit and vegetables (Li et al., 2020). While the consumption of energy dense foods is certainly a risk factor for obesity, the Foresight Report shows that it is a complex, multifaceted system involving biological, social and environmental factors (Butland et al., 2007).

Numerous studies suggest that the consumption of SSB is directly associated with weight gain, overweight and obesity in adolescents (Keller and Bucher Della Torre, 2015). Worldwide, evidence shows that young people consume a large quantity of SSB; for example, almost 80% of adolescents surveyed consumed SSB daily in a nationally representative population of Americans. In China, two-thirds of over 53,000 children and adolescents reported that they consumed more than 7 servings of SSB per week. The European Academy of Pediatrics and the European Childhood Obesity Group recommend that the consumption of SSB should be limited in favor of water and non-sweetened drinks. Not only can SSB contribute to weight gain but also the high sugar content and low pH contributes to dental caries and enamel erosion (Dereń et al., 2019). The development of dental caries occurs due to the metabolization of sugar by oral bacteria, producing acid which demineralizes the hard tissues of the teeth. Limiting free sugar intake to less than 10% of total energy intake, ideally less than 5%, minimizes the risk of dental caries throughout life (WHO, 2017).

Body image and eating disorders

Adolescents typically have concerns about the appearance of their bodies, particularly in relation to their physical appearance and size. This concern is exacerbated by societal or peer pressure for thinness, dieting or musculature which can result in a minority of cases developing overly restrictive eating habits. Girls may experience pressure to become thin while boys feel they must be large and muscular, both of which have the potential to exacerbate disordered eating (Fleming et al., 2020), which often begins in adolescence (WHO, 2018a). Two sets of diagnostic criteria for eating disorders are in use: the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), published by the American Psychiatric Association, and the International Classification of Diseases (ICD-10), published by the WHO. Both manuals list the criteria for the major eating disorders: anorexia nervosa, bulimia nervosa and binge-eating disorder.

With anorexia nervosa, people severely restrict or avoid food, or eat very small quantities of particular foods. People with anorexia nervosa may see themselves as being overweight even when they are extremely underweight. Over time, this can cause thinning of the bones, anemia and muscle wasting. At the most extreme outcome, medical complications due to starvation may be the cause of mortality. People with bulimia nervosa have frequent episodes of consuming large amounts of food, followed by behaviors that compensate for the overeating such as fasting, forced vomiting, excessive exercise or use of laxatives or diuretics. People with bulimia nervosa may have a chronically inflamed throat, worn tooth enamel, gastrointestinal problems and electrolyte imbalance. Finally, binge-eating involves eating unusually large amounts of food but it is not followed by fasting, purging or excessive exercise. Consequently, those with binge-eating disorder are usually overweight or obese (National Institute of Health, 2018b).

The lifetime prevalence of eating disorders among US adolescents aged 13–18 years is 2.7% and is more prevalent among females (3.8%) than males (1.5%) (National Institute of Health, 2017). In Mexico, the lifetime prevalence of anorexia nervosa, bulimia nervosa and binge-eating disorder among adolescents is 0.5%, 1.0% and 1.4%, respectively (Kolar et al., 2016). In the UK, prevalence estimates for anorexia nervosa range from 0.3% to 0.6% (Petkova et al., 2019). In Portugal, the prevalence of anorexia nervosa among adolescents is 0.4% (Smink et al., 2012). Although, eating disorders are more common among girls, it is not clear if the disorder is under-detected in boys. Anxiety, depression and low self-esteem which are often seen in adolescence are associated with eating disorders. The long-term consequences of eating disorders can include detriment to the cardiovascular, gastrointestinal, endocrine, renal and neurological systems.

Vegan and vegetarian diets

Vegetarianism and veganism have increased in popularity and are growing social movement. Generally, vegetarians do not eat meat, poultry or fish but typically eat eggs and dairy foods. Vegans refrain from eating any animal-derived products and may extend this to other aspects of life e.g., avoiding use of animal-sourced clothing, medicines or household products. The prevalence of vegetarianism ranges from 5% in the US/UK and 8% in Canada, to 30% in India (Paslakis et al., 2020) which may be largely attributed to strict Hindus being lacto-vegetarians (vegetarian but includes dairy). In Europe, the prevalence of vegetarianism is on the rise. While specific data on adolescents are scarce, there has been an apparent rise in interest in young people according to social media and campaigns by vegan/vegetarian organizations. In a national health survey in Germany, 6% of girls and 2% of boys aged 14–17 years reported eating a vegetarian-type diet (Schürmann et al., 2017).

Both the British Dietetic Association and the US Academy of Nutrition and Dietetics agree that vegetarian and vegan diets are appropriate at all stages of the life cycle, although this is not a universal view and expert bodies from some European countries have raised concerns about vegan diets in younger age groups. The Spanish Pediatric Association suggest that is it safe to follow a vegetarian diet at any age but advise young children to follow an omnivorous diet or at the very least, an ovo-lacto-vegetarian diet (Redecillas-Ferreiro et al., 2020). With any restrictive diet, inadequate intakes can occur and vegetarianism or veganism are no different. There are a number of nutrients that may be a cause for concern if the diet is not adequately planned. Firstly, protein quality is a concern. While soy protein has a high digestibility, this is not the case for protein in cereals, pulses, nuts and seeds which need to be combined to ensure a broad range of essential amino acids. Hence vegan adolescents will require more protein or specific combinations of protein-rich foods than their vegetarian or omnivorous peers (Agnoli et al., 2017). Another issue in vegetarian and vegan diets relates to omega-3 fatty acids. Oily fish is an excellent source of very long chain omega-3 fatty acids, such as eicosapentaenoic acid and docosahexaenoic acid which are incorporated directly into body cells and which support the synthesis of anti-inflammatory immune compounds. Vegetarian and vegan sources of alpha-linolenic acid—a shorter chain omega-3 fatty acid—include nuts, seeds, dark green vegetables, soya products and rapeseed oil. Alpha-linolenic acid must be converted into eicosapentaenoic acid and docosahexaenoic acid in the body but the process is inefficient and inhibited by the amount of omega-6 fatty acids in the diet. Thus, plasma levels of very long chain omega-3 fatty acids are often lower in vegetarians and vegans compared with fish and meat eaters, although the health implications of this are unclear.

There are a number of minerals and vitamins that are essential for adolescents but may not be supplied in sufficient quantities by a vegetarian or vegan diet. Iodine can be found in plant foods but levels vary depending on soil quality. In general, foods tend to be higher in iodine if they are grown closer to the ocean. Vegans are advised not to use seaweed or kelp as a source of iodine but to consider a dietary supplement. For calcium, tofu and calcium-enriched milk alternatives are excellent sources as well as kale, wholegrain bread (fortified) and sesame seeds and almonds. Spinach is high in calcium but is poorly absorbed as the calcium is bound to oxalate. In contrast, the absorption of calcium from low-oxalate founds such as broccoli, kale and cauliflower can be up to 50%. Iron, which will be discussed in more detail later, is another mineral of particular concern for adolescents. Vegetarian diets mostly consist of non-heme sources of iron which is not as efficiently absorbed as heme sources of iron originating from animal origin foods. Vegetarians can access non-heme sources of iron such as cereals, nuts, seeds, legumes and fortified foods such as breakfast cereals, as well as foods rich in vitamin C which enhance the absorption of non-heme iron (Hood, 2019). Lastly, as vitamin B12 is present only in foods of animal or microbiological origin, vegans depend on fortified foods, fermented foods or supplements, whereas vegetarians can choose fortified foods as well as dairy products and eggs as a source of vitamin B12 (Agnoli et al., 2017).

Nutrient inadequacy and potential consequences

In countries where dietary surveys are available, the evidence suggests that adolescents have the least well-balanced diets of all the age groups. This is as a result of inappropriate food choices favoring high energy dense, low nutrient-dense foods and drinks. According to UNICEF (2019), only a third of school-aged adolescents in low- and middle-income countries consume fruit regularly while a fifth eat vegetables once a day or less. In contrast, over 40% consumed SSB at least once a day. This, combined with other eating behaviors during adolescence, may result in nutrient inadequacies. Likewise, in Europe and Canada, nearly 50% of adolescents reported that they do not eat fruits nor vegetables every day. Roughly 16% of adolescents consumed SSB every day. Not only can these dietary patterns drive excess body weight, but can put adolescents at greater risk of failing to achieve recommendations for micronutrients, particularly iron, calcium, iodine, vitamin D and folate.

Nutrients of concern: iron

Iron deficiency is an area of concern in adolescence. For normal physical development and rapid growth, adolescents require increased intakes of certain vitamins and minerals. Iron is particularly important, especially for post-menarche girls. Iron deficiency anemia (IDA), which affected roughly 619 million adolescents in 2013, can result in shortness of breath, fatigue or chest pain. Adolescents are at an even greater risk of anemia in areas where early marriage and childbearing is common. In addition to IDA, iron deficiency can be associated with frequent infections, behavioral problems and developmental delay.

The European Food Safety Authority (EFSA) has set a population reference intake (PRI) of iron for boys (11 mg/day) and girls (13 mg/day) aged 12–17 years old. The US National Institute of Health recommend 11 mg/day for boys and 15 mg/day for girls

aged 14–18 years old. The PRI set by EFSA are the same as reference nutrient intakes (RNIs) used in the UK and recommended dietary allowances (RDAs) used in the United States. The PRI refers to the amount of nutrient that is adequate for virtually all people within a population. Average intakes of iron among 11–17-year-old girls in Europe range from approximately 7.7 mg/day in Denmark to 14.9 mg/day in Germany. Including supplements, mean intakes ranged from 9 mg/day in the UK to 15.7 mg/day in Germany. For boys, mean daily dietary and supplemental intakes of iron ranged from 10.8 mg in the UK to 18.8 mg in Germany.

Nutrients of concern: bone health

As a consequence of accelerated skeletal development during adolescence, the demand for calcium during this period is higher than at any other stage of life. Approximately 99% of the body's calcium supply is stored in the bones and teeth as calcium hydroxyapatite. There is a constant remodeling process that occurs in bone through the removal of old bone (resorption) and formation of new bone (ossification). The balance between bone removal and bone deposition changes with age. As bone formation exceeds resorption during childhood and adolescence, it is critical that an adequate supply of calcium and other bone related nutrients is achieved to ensure an optimal peak bone mass. The PRI for calcium in Europe is 1150 mg/day for adolescents aged 11–17 years. There are differences between countries as some reflect calcium requirement differences for boys and girls, such as in the UK, whereas the RDA in Canada and the US is the same for males and females (1300 mg/day). Therefore, it is important that an individual follows their respective government's advice for all nutrition recommendations.

Vitamin D is well recognized due to its effect on bone health. The biologically active metabolite of vitamin D is 1,25 $(OH)_2D$ which is involved in the maintenance of calcium and phosphorus homeostasis. This is essential for bone health throughout the life cycle. Bone accretion and growth occurs in infancy and childhood, as well as rapid bone accrual during adolescence. Thus, vitamin D, along with calcium, is essential during adolescence and throughout adulthood to maintain healthy bones and prevent bone loss. Despite this, vitamin D deficiency has become a global public health problem, particularly among children and adolescents. Low vitamin D status may be due to limited endogenous synthesis of vitamin D (as a consequence of lack of appropriate sun exposure) and reduced intakes of vitamin D-rich foods, such as oily fish, eggs and fortified foods/drinks. Endogenous synthesis and dietary intake alone are normally insufficient for maintaining appropriate blood levels across the year. Hence, vitamin D supplementation is often required and some countries, such as the UK, overtly recommend year-round supplemental intakes. The recommended intake differs by country with dietary reference values for adolescents ranging from 5 μg to 20 μg (200–800 IU) per day.

Nutrients of concern: folate

Folate is a B vitamin which is essential for amino acid metabolism, methylation reactions and DNA/RNA biosynthesis. Rich food sources of naturally occurring folate include green leafy vegetables, citrus fruits, beans and legumes. Folic acid is the synthetic version of folate which is more stable and bioavailable. Thus, it is commonly found in supplements, and fortified foods including flour and breakfast cereals. Folate is particularly important for prepregnant females as it is a critical nutrient for the prevention of neural tube defects; the most common being spina bifida and anencephaly which often result in fetal death or severe disability. Therefore, girls of reproductive age who may become pregnant are advised to take a 400 μg supplement of folic acid preconceptually until the 12th week of pregnancy, on top of food sources of folate. However, there are a number of barriers to achieving this. Naturally occurring food folates have limited bioavailability and can be unstable during cooking, while many pregnancies in adolescence are unplanned resulting in girls failing to take folic acid supplements in the recommended window of development for the neural tube (McNulty et al., 2019). Lastly, skipping breakfast (source of folic acid if fortified cereals are chosen) combined with a low consumption of vegetables are common among adolescents.

Nutrients of concern: iodine

Iodine is a trace element that is an essential component of thyroid hormones, which regulate protein synthesis and are needed for skeletal and central nervous system development. Insufficient iodine intake causes the thyroid stimulating hormone to remain elevated, leading to goiter, an enlargement of the thyroid gland. This enlarged gland reflects the body's attempt to trap iodine and produce thyroid hormones (National Institute of Health, 2021). The consequences of iodine deficiency for adolescents may delay physical development and impair mental function (Zimmermann, 2009). The highest levels of iodine deficiency disorders occur in Sub-Saharan Africa, South Asia, the Middle East and North Africa regions (WHO, 2018b). Groups at risk of inadequate iodine status include vegans or those who eat negligible amounts of seafood, dairy products and eggs, since these are among the best sources. Other groups at risk include those who do not use iodized salt – a common public health strategy to control iodine deficiency in many countries (National Institute of Health, 2021).

Improving adolescents' diets

Investing the time and resources to promote healthy eating in adolescents not only improves health in this age group, for example by supporting normal growth, lowering the risk of obesity, and boosting nutrient adequacy, but helps to establish better eating habits in the long term which can help to prevent non-communicable diseases. Wholegrains, fruits, vegetables, oily fish, dairy

foods, lean meats, eggs, plant-based proteins and sugar-free fluids (water, tea) should be promoted over those foods currently preferred by adolescents in many countries, particularly SSB, fast foods, confectionery, savory snacks and sweetened baked goods. Dietary interventions should consider a whole dietary approach and focus on increasing whole grain, fruit and vegetable consumption (Doherty et al., 2021). However, diet is one component of a multifaceted problem. Interventions combining diet, physical activity and behavioral components are considered to be best practice in the prevention of obesity in adolescence. A systematic review assessed the effects of dietary, physical activity and behavioral interventions on overweight and obese adolescents, locating 44 randomized controlled trials which met the inclusion criteria. Body weight was lowered by approximately 3.7 kg in trials of holistic interventions with moderate quality evidence (Al-Khudairy et al., 2017).

A growing number of studies have started to use social media in nutrition interventions for adolescents and young adults. A systematic review of mostly randomized controlled trials found that 11 out of 16 interventions had short-term positive effects on nutrition-related, clinical or behavioral outcomes. Social media was used to facilitate communication and social support. Online discussions were also used in some studies, whereas others employed smaller group interactions using Facebook, Twitter and WhatsApp. Some of the positive results included weight loss, reduction in SSB intake and screen time, as well as improvements in fruit and vegetable intake (Chau et al., 2018).

Conclusion

In conclusion, the evidence on adolescence and associated nutritional problems reveals:

- Adolescents are prone to unhealthy eating patterns including an excess consumption of SSB and fast foods, as well as skipping breakfast and having low intakes of fruits, vegetables and nutrient-rich foods.
- Vegetarianism and veganism are increasing in popularity but may result in low intakes of essential nutrients if diets are not appropriately planned. Hence additional support may be needed for young people intending to follow these types of diets.
- Iron, calcium, iodine, vitamin D and folate are among the nutrients of particular concern for adolescents-in some cases supplementation may be warranted.
- Obesity in adolescence is a worldwide major public health issue which increases the risk of several non-communicable diseases.
- Holistic interventions combining advice on diet, physical activity and behavioral aspects seem to be most effective for preventing and treating overweight and obesity in adolescence.

Appendix 1: examples of growth charts for adolescents

Country	*Boys*	*Girls*	*Cut offs*
Canada	Boys (2–19 yrs) BMI	Girls (2–19 yrs) BMI	Overweight: >85th centile; obese: >97th centile
India	Boys (0–18 yrs) height and weight	Girls (0–18 yrs) height and weight	Overweight: 85th centile; obese: 97th centile
United Kingdom	Boys 2–18 yrs BMI 2–20 yrs	Girls 2–18 yrs BMI (2–20 yrs)	Overweight: >91st centile; obese: >98th centile; undernutrition/small build: <2nd centile
United States	Boys (2–20 yrs) stature for age weight for age BMI for age	Girls (2–20 yrs) stature for age weight for age BMI for age	BMI ranges: overweight: ≥85th centile; obese: ≥95th; normal/healthy: 5th–85th; underweight: <5th

References

Agnoli, C., Baroni, L., Bertini, I., et al., 2017. Position paper on vegetarian diets from the working group of the Italian Society of Human Nutrition. Nutr. Metabol. Cardiovasc. Dis. 27, 1037–1052.

Al-Khudairy, L., Loveman, E., Colquitt, J.L., et al., 2017. Diet, physical activity and behavioral interventions for the treatment of overweight or obese adolescents aged 12 to 17 years. Cochrane Database Syst. Rev. 6, 1–356.

Butland, B., Jebb, S., Kopelman, P., et al., 2007. Tackling Obesities: Future Choices-Project Report. Department of Innovation, Universities and Skills, London.

Chau, M.M., Burgermaster, M., Mamykina, L., 2018. The use of social media in nutrition interventions for adolescents and young adults—a systematic review. Int. J. Med. Inf. 120, 77–91.

Dereń, K., Weghuber, D., Caroli, M., et al., 2019. Consumption of sugar-sweetened beverages in paediatric age: a position paper of the European Academy of Paediatrics and the European Childhood Obesity Group. Ann. Nutr. Metabol. 74, 296–302.

Doherty, A.M., Lacko, A.M., Popkin, B.M., 2021. Sugar-sweetened beverage (SSB) consumption is associated with lower quality of the non-SSB diet in US adolescents and young adults. Am. J. Clin. Nutr. 113, 657–664.

Dye, L., 2017. The Importance of Breakfast in Europe: A Review of National Policies and Health Campaigns.

Fleming, C.A., De Oliveira, J.D., Hockey, K., et al., 2020. Food and Me: How Adolescents Experience Nutrition Across the World. A Companion Report to the State of the World's Children 2019. UNICEF.

Freedman, D.S., Lawman, H.G., Galuska, D.A., Goodman, A.B., Berenson, G.S., 2018. Tracking and variability in childhood levels of BMI: the Bogalusa Heart Study. Obesity 26, 1197–1202.

Hood, S., 2019. Vegetarianism and vegan diets. In: Gandy, J. (Ed.), Manual of Dietetic Practice, sixth ed. John Wiley & Sons, West Sussex, pp. 129–139.
Keller, A., Bucher Della Torre, S., 2015. Sugar-sweetened beverages and obesity among children and adolescents: a review of systematic literature reviews. Child. Obes. 11, 338–346.
Kolar, D.R., Rodriguez, D.L.M., Chams, M.M., Hoek, H.W., 2016. Epidemiology of eating disorders in Latin America: a systematic review and meta-analysis. Curr. Opin. Psychiatry 29, 363–371.
Li, L., Sun, N., Zhang, L., et al., 2020. Fast food consumption among young adolescents aged 12–15 years in 54 low-and middle-income countries. Glob. Health Action 13, 1795438.
McNulty, H., Ward, M., Hoey, L., Hughes, C.F., Pentieva, K., 2019. Addressing optimal folate and related B-vitamin status through the lifecycle: health impacts and challenges. Proc. Nutr. Soc. 78, 449–462.
National Institute of Health, 2017. Eating Disorders: Fact Sheet for Health Professionals. US Department of Health & Human Services.
National Institute of Health, 2018a. Osteoporosis: Peak Bone Mass in Women. US Department of Health & Human Services.
National Institute of Health, 2018b. Eating Disorders: About More than Food. US Department of Health & Human Services.
National Institute of Health, 2021. Iodine: Fact Sheet for Health Professionals. US Department of Health & Human Services.
Ostchega, Y., Fryar, C.D., Nwankwo, T., Nguyen, D.T., 2020. Fast food intake among children and adolescents in the United States, 2015–2018. NCHS Data Brief 375, 1–8.
Paslakis, G., Richardson, C., Nöhre, M., et al., 2020. Prevalence and psychopathology of vegetarians and vegans: results from a representative survey in Germany. Sci. Rep. 10, 1–10.
Petkova, H., Simic, M., Nicholls, D., et al., 2019. Incidence of anorexia nervosa in young people in the UK and Ireland: a National Surveillance Study. BMJ Open 9, 1–9.
Quinte, G.C., Barros, F., Gigante, D.P., et al., 2019. Overweight trajectory and cardio metabolic risk factors in young adults. BMC Pediatr. 19, 1–8.
Redecillas-Ferreiro, S., Moráis-López, A., Moreno-Villares, J.M., 2020. Position paper on vegetarian diets in infants and children. Committee on Nutrition and Breastfeeding of the Spanish Paediatric Association. Ann. Pediatr. 92, 1–306.
Redmond, J., Schoenmakers, I., 2019. Osteoporosis. In: Gandy, J. (Ed.), Manual of Dietetic Practice, sixth ed. John Wiley and Sons, West Sussex, pp. 612–622.
Schürmann, S., Kersting, M., Alexy, U., 2017. Vegetarian diets in children: a systematic review. Eur. J. Nutr. 56, 1797–1817.
Simmonds, M., Burch, J., Llewellyn, A., et al., 2015. The use of measures of obesity in childhood for predicting obesity and the development of obesity-related diseases in adulthood: a systematic review and meta-analysis. Health Technol. Assess. 19, 1–336.
Smink, F.R., Van Hoeken, D., Hoek, H.W., 2012. Epidemiology of eating disorders: incidence, prevalence and mortality rates. Curr. Psychiatr. Rep. 14, 406–414.
Stewart, L., Gillespie, J., 2019. Childhood obesity. In: Gandy, J. (Ed.), Manual of Dietetic Practice, sixth ed. John Wiley & Sons, West Sussex, pp. 755–760.
UNICEF, 2019. The State of the World's Children: Children, Food and Nutrition. https://www.unicef.org/media/60806/file/SOWC-2019.pdf.
Verduci, E., Bronsky, J., Embleton, N., , et al.ESPGHAN Committee on Nutrition, 2021. Role of dietary factors, food habits, and lifestyle in childhood obesity development: a position paper from the European Society for Paediatric Gastroenterology, Hepatology and Nutrition Committee on Nutrition. J. Pediatr. Gastroenterol. Nutr. 72 (5), 769–783.
WHO, 2016. Adolescents' Dietary Habits: Fact Sheet. World Health Organization. Available at. https://www.euro.who.int/__data/assets/pdf_file/0006/303477/HBSC-No.7_fact-sheet_Diet.pdf%3Fua%3D1.
WHO, 2017. Sugars and Dental Caries. WHO Technical Information Note. Available at. https://apps.who.int/iris/rest/bitstreams/1091242/retrieve.
WHO, 2018a. Taking Action on Childhood Obesity (No. WHO/NMH/PND/ECHO/18.1). World Health Organization. Available at. https://apps.who.int/iris/bitstream/handle/10665/274792/WHO-NMH-PND-ECHO-18.1-eng.pdf.
WHO, 2018b. Guideline: Implementing Effective Actions for Improving Adolescent Nutrition.
WHO, 2020a. Obesity and Overweight. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.
WHO, 2020b. WHO Guidelines on Physical Activity and Sedentary Behavior. Available at. https://apps.who.int/iris/rest/bitstreams/1315866/retrieve.
Zimmermann, M.B., 2009. Iodine deficiency. Endocr. Rev. 30, 376–408.

Further reading

Beal, T., Morris, S.S., Tumilowicz, A., 2019. Global patterns of adolescent fruit, vegetable, carbonated soft drink, and fast-food consumption: a meta-analysis of global school-based student health surveys. Food Nutr. Bull. 40, 444–459.
Darfour-Oduro, S.A., Buchner, D.M., Andrade, J.E., Grigsby-Toussaint, D.S., 2018. A comparative study of fruit and vegetable consumption and physical activity among adolescents in 49 Low-and-Middle-Income Countries. Sci. Rep. 8, 1–12.
Dror, D.K., 2014. Dairy consumption and pre-school, school-age and adolescent obesity in developed countries: a systematic review and meta-analysis. Obes. Rev. 15, 516–527.
Fidler Mis, N., Braegger, C., Bronsky, J., et al., 2017. Sugar in infants, children and adolescents: a position paper of the European Society for Paediatric Gastroenterology, Hepatology and Nutrition Committee on Nutrition. J. Pediatr. Gastroenterol. Nutr. 65 (6), 681–696.
Fleming, C.A., De Oliveira, J.D., Hockey, K., et al., 2020. Food and Me: How Adolescents Experience Nutrition Across the World. A Companion Report to the State of the World's Children 2019. UNICEF.
Hobbs, M., Pearson, N., Foster, P.J., Biddle, S.J., 2015. Sedentary behavior and diet across the lifespan: an updated systematic review. Br. J. Sports Med. 49, 1179–1188.
Keats, E.C., Rappaport, A.I., Shah, S., et al., 2018. The dietary intake and practices of adolescent girls in low-and middle-income countries: a systematic review. Nutrients 10, 1978.
Monzani, A., Ricotti, R., Caputo, M., et al., 2019. A systematic review of the association of skipping breakfast with weight and cardiometabolic risk factors in children and adolescents. What should we better investigate in the future? Nutrients 11, 387.
Moreno, L.A., Gottrand, F., Huybrechts, I., et al., 2014. Nutrition and lifestyle in European adolescents: the HELENA (Healthy Lifestyle in Europe by Nutrition in Adolescence) Study. Adv. Nutr. 5, 615–623.
OECD/European Union, 2020. Health at a Glance: Europe 2020: State of Health in the EU Cycle. OECD Publishing, Paris.
Rivera, J.Á., de Cossío, T.G., Pedraza, L.S., et al., 2014. Childhood and adolescent overweight and obesity in Latin America: a systematic review. Lancet Diabetes Endocrinol. 2, 321–332.
World Health Organization, 2017a. Adolescent Obesity and Related Behaviors: Trends and Inequalities in the WHO European Region, 2002–2014. WHO, Geneva.
World Health Organization, 2017b. Global Accelerated Action for the Health of Adolescents (AA-HA!): Guidance to Support Country Implementation. WHO, Geneva.
World Health Organization, 2018. Implementing Effective Actions for Improving Adolescent Nutrition. WHO, Geneva.

Older people: Nutritional requirements

N. Solomons, Aging and Metabolism (CeSSIAM), Guatemala City, Guatemala

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of N. Solomons, Older People: Nutritional Requirements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 393–399, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00217-8.

Human Aging and Nutrition 150
Successful Aging, Normative Aging, and Frailty 150
Overview of Specific Factors of Aging Influencing Nutritional Requirements 151
Nutrient Intake Recommendations in Later Life 151
Definitions Surrounding Recommended Intakes of Nutrients 151
Established Recommended Intakes for Older Persons 152
Macronutrients 152
Water 152
Micronutrients 153
Dietary Guidelines for Function, Health, and Disease Prevention 155
Barriers to Meeting Recommended Nutrient Intakes and Healthful Dietary Intake Patterns by Older Persons 155
Future Considerations 156
Further Reading 156

Human Aging and Nutrition

The World Health Organization (WHO) defines the 'elderly' as persons of 60 years of age and older. The elderly constitute a rapidly expanding segment of populations in both developed and developing countries. This is the combined result of ever-longer survival and dramatic reductions in fertility rates. Regardless of age, people must respond to their feelings of hunger and thirst by consuming foods and beverages. This eating and drinking behavior also serves to provide the nutrients to nourish the body. The amount of a nutrient that must be ingested and absorbed to maintain an adequate and appropriate body composition varies with age across the life span, depending on basic underlying physiological and metabolic processes specific to the chronological stage of life. Similarly, the degree to which we retain and conserve, or excrete or degrade, absorbed nutrients is influenced by chronological age and biological aging.

As a consequence of this new demographic reality, attention is being focused belatedly on gerontology and its nutritional biology; this, in turn, is reflected in very recent efforts to refine our knowledge of the amounts of various macro- and micronutrients that the aging body requires (nutrient requirements) and of the amounts that must be consumed in the diet to provide for sufficient uptake of these nutrients (nutrient recommendations).

Successful Aging, Normative Aging, and Frailty

From an epidemiologic and demographic, as well as an economic and humanitarian standpoint, the ideal contribution of life-long nutrition would be to a situation of 'compression of morbidity,' first enunciated by J. Fries. It strives to keep individuals functional, independent, and free of chronic illness, until the final moments of their lives, and thus reduces to a minimum the burden of disability and dependency suffered by older persons, their families, and the society that contributes to their maintenance.

A disclaimer has traditionally been appended to the official pronouncements of recommended nutrient intakes (RNIs); whether they are from national or international expert panels, the prescriptions are meant to apply to 'healthy' individuals. Nutrient needs in disease conditions are considered to be a clinical matter, and are related to the pathological processes in question.

When it comes to older persons, the exigency of being 'healthy' becomes immediately problematic. Advanced age is associated with increased susceptibility to chronic and degenerative illnesses. Most persons over 60 years of age have two or three diagnosed chronic illnesses, and are receiving multiple medications. Maintaining a rigid definition of 'healthy' for application of nutrient recommendations in later life would exclude almost everyone from coverage by general nutrient-intake standards.

In fact, the older the cohort of individuals examined, the more heterogeneous are individuals of the same chronological age in their physical and cognitive functioning. Over the last two decades, general domains of classifications have come into usage to embrace the heterogeneity of aging populations: successful aging; usual aging; and frail aging. Successful aging has been defined as multidimensional, 'encompassing the avoidance of disease and disability, the maintenance of high physical and cognitive function, and sustained engagement in social and productive activities.' It may involve aspects of resilience and wisdom, as well. Usual aging involves an accumulation of ailments and loss of function that is 'typical' or 'normative' for older persons surviving to later

life. Frailty is the far extreme of disability and dependency associated with major physical and cognitive decline in which disease and senescent processes become irreversibly established.

A prominent and optimistic school of thought suggests that exposures to behavioral and environmental factors that modify risk of disease and dysfunction determine one's position in these alternative outcomes in the aging process. In this view, more optimal nutrient intake, food selection, and life-style choices could reduce the heterogeneity, retaining more individuals in the successfully aged category for most of their life span. Others consider that genetic constitution may be as important in determining the course of aging as any positive or negative influences during our lifetime.

Overview of Specific Factors of Aging Influencing Nutritional Requirements

The discussion of nutrient requirements and recommended dietary intakes of nutrients in older persons has proceeded on both the theoretical and empirical level. Since the peak years for human reproduction occur before middle age, and well before older age begins, the forces of natural selection, governing fecundity in reproduction cannot exert themselves for the Darwinian selection of traits favoring longevity in the evolution for any traits related to longevity *per se* or physiological sustained function. Hence, there has been little or no evolutionary selection for nutrient requirements to achieve advanced age or for long-term survival. It is more for the preservation of comfort and function for those surviving to advanced age that the optimization of nutrition intakes for the elderly would apply, that is, for humanitarian and public health importance in the face of the physiological and anatomic changes of senescence.

As early as the 1970s, nutritional scientists advanced the proposition that requirements for different macro- and micronutrients changed with age. A large number of conjectures based on an emerging scientific understanding of senescent physiology have been advanced. It has been suggested that the decreased physical activity and physical conditioning associated with the body composition changes attendant to aging sets the stage for alterations in requirements in both amounts and relative proportions of protein and the energy-yielding macronutrients. Decreased gastric secretory capacity has a negative influence on the absorption of calcium, iron, and vitamin B_{12}. Changing intestinal motility and digestive function evoked considerations of distinct increases and decreases of nutrients to compensate for the senescence of the intestinal tract, with particular interest in dietary fiber. Attention to compensatory intake for all of the nutrients involved in skeletal mineralization has come to the fore in relation to the recognized tendency to bone mineral loss with advancing age.

The immune and host defense system has been a major focus of gerontological nutrition. Increased intakes of both vitamin E and zinc, well above the normally recommended level, have stimulated certain immune functions in studies involving older volunteers. More recently, evidence for enhanced immune function in older individuals from physiological doses of zinc has been reported. Zinc may act to fortify the responsiveness of lymphocytes to stress. Impaired interferon and interleukin-2 responses may be the mechanisms whereby low zinc status impairs immune function in elderly humans. Additionally, vitamin D seems to play a role in the regulation of the aging immune system in the area of antibody and cytokine responses. Cognitive function declines with advancing age, and it has even been suggested that the adjustment of nutrient intake can favorably affect the retention of memory and cognitive function in older persons. Results of association studies and intervention trials have been mixed and inconclusive. The adequate intake (AI) of B-complex vitamins, particularly those related to homocysteine metabolism (vitamin B_{12}, folic acid, vitamin B_6, riboflavin), are associated with mental function in older age. It has also been suggested that older individuals need more *n*-3 fatty acids for preserving cerebral cellular anatomy related to cognition. A prospective trial of B-vitamin and *n*-3 fatty acids in older French adults found positive effects in the subsegment with a history of cerebral strokes. Implications for a role of vitamin D in preserved cognition with advancing age have emerged in human studies. Animal research suggests that both vitamin D and K may be important in conserving central nervous system function with aging. For several nutrient effects in immunity, cognition, and other areas, an interaction with a biomarker of a genetic polymorphism seems to be important.

Nutrient Intake Recommendations in Later Life

Comprehensive recommendations for macro- and micronutrients with differential attention to older persons have arisen from a collaboration between the US and Canada, and from expert panels serving the United Nations (UN) System. Each panel has set out its methodology and definitions and then presented tables of quantitative estimates. The recommendations for persons considered elderly in the respective systems are outlined below.

Definitions Surrounding Recommended Intakes of Nutrients

An important advance in establishing nutrient intake recommendations relates to the semantics. There has been a refining of the operational definitions of terms related to nutrient intakes. RNIs are set by the agencies of the UN System and are considered to be the intakes of nutrients required to satisfy the requirements of nearly all healthy persons of a given age, sex, and physiological condition, and should be universal for all regions of the globe. In 2006, a way to calculate (retrofit) the population-relevant estimated average requirement (EAR) to the UN System was published.

The Food and Nutrition Board of the Institute of Medicine in the US took a new approach in 1997 in which they applied the new dietary reference intakes (DRI) to micro- and macronutrient intakes. This work was undertaken jointly with Canada. It began with an assessment, where possible, of the EAR, defined as "the average daily nutrient intake level estimated to meet the requirement of half the healthy individuals in a particular life stage and gender group." The EAR is critical for an assessment of the risk of a nutrient deficiency problem at the population level. The traditional criterion used for decades, the recommended dietary allowance (RDA), is preserved. It is defined in the DRI process as "the average daily nutrient intake level sufficient to meet the nutrient requirement of nearly all (97 to 98%) healthy individuals in a particular life stage and gender group." When an EAR cannot be established from which to derive a formal RDA, the DRI process has a 'fall-back' category known as AI; this is defined as "a recommended average daily nutrient intake level based on observed or experimentally determined approximations or estimates of nutrient intake by a group (or groups) of apparently healthy people that are assumed to be adequate." A new classification scheme involving a range of intakes was created specifically for energy, electrolytes, and liquids: the acceptable macronutrient distribution ranges (AMDRs).

For the first time, a specific and well-defined process to delimit levels of excess intake of nutrients and dietary substances was defined by the DRI process as the upper tolerable intake levels (UL). The UL is 'the highest average daily nutrient intake level likely to pose no risk of adverse health effects to almost all individuals in the general population.' It is considered that as intake increases above the UL, the potential risk of adverse effects increases. To date, the UN System's process has dealt much less explicitly with issues of excessive intake of nutrients and dietary substances.

Established Recommended Intakes for Older Persons

In earlier versions of the RDAs for the US population (up to the 10th edition in 1989), the nutrient recommendations for all healthy adults over 51 years of age were combined as a single value. For the UN System, the age threshold in the early editions was 50 years or older. Concerted efforts to refine our understanding of nutrient requirements for older adults have been made over the past two decades. This allowed the US–Canada DRI process to establish categories for men and women aged 70 years and older. For the WHO/Food and Agriculture Organization (FAO) process, a specific estimation for individuals over 65 years has been provided in the 2004 micronutrient recommendations.

Given the magnitude of the theoretical considerations regarding senescence and aging physiology that have been raised by various authors, what is really surprising is the paucity of specific instances in which the recommended intakes of nutrients for men or women in the 'elderly category' are considered to be different from persons in the next youngest age category. Composite tables for men (**Table 1**) and women (**Table 2**) are given for all of the nutrients and dietary substances expressed in the US–Canada DRIs and in the UN system for RNIs.

Macronutrients

In the DRI system, a universal, individual protein requirement was established as 0.80 g of good-quality protein per kilogram of body weight per day independent of age. No evidence for altered protein requirements with older age has been found. Moreover, it is recommended that the contribution of protein to total energy intake should not exceed 30%. The US Food and Nutrition Board also established an amino acid pattern in 2002. It specifies the density (mg g^{-1} protein) of seven indispensable (essential) amino acids (histidine, isoleucine, leucine, lysine, threonine, tryptophan, and valine) and for two amino acid combinations (methionine+-cysteine, phenylalanine+tyrosine). This pattern is universal from age 1 year to the extremes of older age without modification.

It has long been recognized that energy recommendations cannot be made on a group basis, as each individual has his or her own daily energy requirement dependent on the amount of energy one is forced to expend with metabolic reactions, food processing, and physical exertion. In the DRI process, this is recognized in an effort to individualize the estimation of energy intake. Estimated energy requirement (EER) is based on the amount of energy needed to maintain energy balance in relation to one's total energy expenditure. The DRI process for the US and Canada has published general EER equations (multidimensional nomograms) by which a reasonable estimate of an individual energy requirement can be calculated. There are general equations for adult men and women (over 19 years of age), based on consideration of physical activity level, weight, and height. In addition, there is an age term in the general EER, which is attached to a negative (minus sign) term in the equation. This signifies that energy requirements decline as a function of advancing years.

Although dietary fiber is not considered to be an 'essential' nutrient, the DRIs give a recommended level for intake. Curiously, in light of the active discussion of the role of fiber for the elderly in colonic function, the recommendations for intake by men decline from 38 to 30 g day^{-1} and in women from 25 to 21 g day^{-1} after 50 years. These are continued throughout the 70-year period, as well. This is a consequence of the fiber recommendations being pegged to total average energy intake.

Water

It is recommended in the DRI as an AI that males over the age of 70 require 2.6 l and females 2.1 l of water per day; this is a decline from the 51–70-year age group, where the daily water intake recommendations were 3.7 l and 2.6 l, respectively. It is further

Table 1 Nutrient intake recommendations for older males

	UL[a]	*EAR*[a]	*RDA/AI*[a] AMDR[a]	*EAR*[b]	*RNI*[c]
Macronutrients					
Water (l)	–	–	2.1[d]	–	–
Carbohydrate (g)	–	100	120	–	–
Protein (g)	–	46	56	–	–
Total fat (g)	–	–	20–35	–	–
n-6 PUFA (g)	–	–	14[d]	–	–
n-3 PUFA (g)	–	–	1.6[d]	–	–
Dietary fiber (g)	–	–	30	–	–
Vitamins					
Vitamin A (RAE)	3000	625	900	430	600 (μg RE)
Vitamin D (μg)	4000	–	20	–	15
Vitamin E (mg α-tocopherol)	1000	12	15	8	10 (mg α-TE)
Vitamin K (μg)	–	–	120[d]	–	65
Vitamin C (mg)	2000	75	90	38	45
Thiamin (mg)	–	1.0	1.2	1.0	1.2
Riboflavin (mg)	–	1.1	1.3	1.1	1.3
Niacin (mg)	35	12	16	12	16
Vitamin B_6 (mg)	100	1.4	1.7	8.0	1.7
Biotin (mg)	–	–	30[d]	–	–
Pantothenic acid (mg)	–	–	5[d]	–	5
Folic acid (μg)	1000	320	400	320	400
Vitamin B_{12} (μg)	–	2.0	2.4	2.0	2.4
Choline (mg)	3500	–	550[d]	–	–
Elements					
Sodium (g)	2.3	–	1.2[d]	–	–
Potassium (mg)	–	–	4.7[d]	–	–
Chloride (g)	3.6	–	1.8[d]	–	–
Calcium (mg)	2000	–	1200	1083	1300
Phosphorus (mg)	3000	580	700	–	–
Magnesium (mg)	(350)	350	420	–	230
Iron (mg)	45	6	8	10	14[e]
Zinc (mg)	40	9.4	11	5.8	7.0[f]
Iodine (μg)	1100	95	150	93	130
Copper (mg)	10	0.7	0.9	–	–
Fluoride (mg)	10	–	4[d]	–	–
Manganese (mg)	11	–	2.3[d]	–	–
Chromium (μg)	–	–	30[d]	–	–
Selenium (μg)	400	45	55	28	34
Molybdenum (μg)	2000	34	45	–	–

The figures in bold denote recommendations specifically modified for ageing (see text). UL, upper tolerable upper intake level; EAR, estimated average requirements; RDA, recommended dietary allowance; AI, adequate intake; AMDR, acceptable macronutrient distribution range; RNI, recommended nutrient intake; PUFA, polyunsaturated fatty acids; RAE, retinol activity equivalents; RE, retinol equivalents; α-TE, alpha-tocopherol.

[a]In DRIs 70 years plus is considered as 'older'.

[b]In UN System (WHO/FAO/IAEA) 65 years plus is considered as 'older'.

[c]The EARs estimated retrospectively from factors published by Allen L, de Benoist B, Dary O, and Hurrell R (2006) *Guidelines on Food Fortification with Micronutrients*. World Health Organization and Food and Agricultural Organization of the United Nations. Geneva: WHO.

[d]Recommendation in the form of adequate intake.

[e]Assumes a 10% bioavailability of iron from the diet.

[f]Based on the assumption of a moderate bioavailability of zinc.

suggested that males and females over 70 years of age derive 81% of their daily water allowance from beverages and 19% as the metabolic water from foods. This is consistent throughout adulthood from age 19 years. Hence, there is no consideration of a higher requirement for water intake with older age. With respect to the electrolytes, no differences in AIs exist across the ages in adulthood.

Micronutrients

A number of recommendations (RDAs or AIs) change with advancing age in the DRI system; this is indicated by the bold type in Tables 1 and 2. The change in recommendations occurs at either age 50 or 70 years. In women over 50 years, the RDA for dietary

Table 2 Nutrient intake recommendations for older females

	UL	*EAR*[a]	*RDA/AI*[a] AMDR[a]	*EAR*[b]	*RNI*[c]
Macronutrients					
Water (l)	–	–	2.6[d]	–	–
Carbohydrate (g)	–	100	120	–	–
Protein (g)	–	38	46	–	–
Total fat (g)	–	–	20–35	–	–
n-6 PUFA (g)	–	–	11[d]	–	–
n-3 PUFA (g)	–	–	1.3[d]	–	–
Dietary fiber (g)	–	–	21	–	–
Vitamins					
Vitamin A (RAE)	3000	500	700	430	600 (μg RE)
Vitamin D (μg)	4000	–	20	–	15
Vitamin E (mg α-tocopherol)	1000	12	15	6.2	7.5 (mg α-TE)
Vitamin K (μg)	–	–	90[d]	–	55
Vitamin C (mg)	2000	60	75	38	45
Thiamin (mg)	–	0.9	1.1	0.9	1.1
Riboflavin (mg)	–	0.9	1.1	0.9	1.1
Niacin (mg)	35	11	14	11	14
Vitamin B_6 (mg)	100	1.3	1.5	1.3	1.5
Biotin (mg)	–	–	30[d]	–	–
Pantothenic acid (mg)	–	–	5[d]	–	5
Folic acid (μg)	1000	320	400	320	400
Vitamin B_{12} (μg)	–	2.0	2.4	2.0	2.4
Choline (mg)	3500	–	425[d]	–	–
Elements					
Sodium (g)	2.3	–	1.2[d]	–	–
Potassium (mg)	–	–	4.7[d]	–	–
Chloride (g)	3.6	–	1.8[d]	–	–
Calcium (mg)	2000	–	1200	1083	1300
Phosphorus (mg)	3000	580	700	–	–
Magnesium (mg)	(350)	265	320	–	190
Iron (mg)	45	5	8	6.9	11[e]
Zinc (mg)	40	6.8	8	4.1	4.9[f]
Iodine (μg)	1100	95	150	79	110
Copper (mg)	10	0.7	0.9	–	–
Fluoride (mg)	10	–	3[d]	–	–
Manganese (mg)	11	–	1.8[d]	–	–
Chromium (μg)	–	–	20[d]	–	–
Selenium (μg)	400	45	55	22	26
Molybdenum (μg)	2000	34	45	–	–

The figures in bold denote recommendations specifically modified for ageing (see text). UL, upper tolerable upper intake level; EAR, estimated average requirements; RDA, recommended dietary allowance; AI, adequate intake; AMDR, acceptable macronutrient distribution range; RNI, recommended nutrient intake; PUFA, polyunsaturated fatty acids; RAE, retinol activity equivalents; RE, retinol equivalents; α-TE, alpha-tocopherol.

[a]In DRIs 70 years plus is considered as 'older'.

[b]In UN System (WHO/FAO/IAEA) 65 years plus is considered as 'older'.

[c]The EARs estimated retrospectively from factors published by Allen L, de Benoist B, Dary O, and Hurrell R (2006) *Guidelines on Food Fortification with Micronutrients*. World Health Organization and Food and Agricultural Organization of the United Nations. Geneva: WHO.

[d]Recommendation in the form of adequate intake.

[e]Assumes a 10% bioavailability of iron from the diet.

[f]Based on the assumption of a moderate bioavailability of zinc.

iron decreases from 18 to 8 mg day^{-1}; there is no change in requirement for the 70-year plus age group. This lower value is the recommendation for adult men of all ages. The fact that the menopause allows women to replete iron stores depleted by an adulthood of monthly menstrual blood loss accounts for this lower RDA in older women.

The senescence of the skeletal system and the reduction of bone mineral content with age are a major nutritional concern in gerontological nutrition. In recent revisions of the recommendations, evidence for the need for increases in both vitamin D and calcium for older persons has led to changes in the estimates of requirements for these nutrients in later life. Within the DRI system the RDA for vitamin D for males and females over 70 years has recently been raised from 15 to 20 μg. The RDA for adults over 51–70 years is now 15 mg, as it is for young adults as well. Similar increases in vitamin D intake with age are recommended by the FAO/WHO and represent a progression from 5 μg in young adulthood to 10 μg after mid-century to 15 μg after 65 years. With respect to

calcium, the recommended levels increase from 1000 mg for younger adults to 1200 mg at age 50 years and beyond in the DRI system, and from 1000 to 1300 mg in the FAO/WHO standards. These are justified based on the higher propensity for skeletal fractures after 70 years of age associated with epidemiological evidence of widespread vitamin D deficiency in this age group, and evidence showing a reduction in bone loss with daily calcium intakes exceeding 1000 mg after midlife.

With respect to chromium, it is interesting that the estimation of an AI declines with advancing age. The AI for persons over 70 years in the DRI is the same as that for individuals between 51 and 70 years, but it is 5 μg day^{-1} higher for the 19–50 years age range. This reduction is tied to the lower energy demands for individuals over 50 years of age.

The upper tolerable UL for phosphorus in the DRI system is 3000 mg day^{-1} for both men and women over 70 years as compared to 4000 mg day^{-1} for adults in the 19–70 years age group. This lower tolerance is explained by the greater prevalence of impaired renal function in advanced old age. Recently, moreover, the DRI has lowered the UL for calcium to 2000 mg beyond age 50 years, which is 500 mg below the level for ages 19–50 years.

Magnesium intake recommendations in the FAO/WHO guidelines decline for individuals over 65 years by 30 mg day^{-1} compared to those in the 51–65 years age group. An anomalous finding for the magnesium RDA in the DRI system, which applies to all adult age groups, is that the UL for magnesium has been set at 350 mg. This is only 30 mg higher than the 320 mg daily recommended for older women, and is 70 mg lower than the 420 mg daily intake recommended for older men.

Dietary Guidelines for Function, Health, and Disease Prevention

If indeed the motivation for our dietary intake is hunger and thirst, and the primary evolutionary purpose is fulfilling nutritional needs, the manner in which we eat has important consequences for function, health, and disease prevention. The discipline of nutritional epidemiology has emerged over the past 30 years to assess associations between the selection of foods in the diet and probability and risk of suffering from poor physical function or ill health. Various features of eating behavior, from the size of portions to the number of repasts consumed in a day, to the frequency of consumption of foods with protective or noxious characteristics, have been implicated in function and health. Since aging represents an independent risk factor for disease and dysfunction, the effects of the dietary pattern become ever more manifest as an individual ages. Recent epidemiological research has shown that compliance or behavior concordant with healthy eating guidelines are associated with lower later life incidences of certain cancers, cataracts, diabetes, hypertension, stroke, and cardiovascular diseases, as well as overall survival. Moreover, there is intense interest in whether and how dietary pattern in addition to nutrition influence the maintenance of memory and cognitive function with aging.

These suggestions and recommendations have been codified and promoted by various entities focused on specific pathological situations, such as guidance for a healthy cardiovascular system by the American Heart Association or cancer prevention by the World Cancer Research Fund. In its Technical Report 916, Diet, Nutrition and the Prevention of Chronic Diseases, an expert committee of the UN agencies provides generic guidelines for reducing the risk of six important noncommunicable diseases. National bodies have also established dietary guidelines for healthful eating. They are often projected to the public in the form of emblems such as plates, rainbows, or pyramids, with visual representation of the relative proportion of various food-groups to consume, with lesser intakes of foods with salt, saturated fats, and high-energy density foods, and greater intakes of grains, fruits, and vegetables. More elaborated justifications and instructions can be found in the published forms in paperback editions or online on the Internet. Such is the case for the Dietary Guidelines for Americans in its 2011 edition.

The notion is that one adheres throughout life, even from childhood, to the tenets of these guidelines. Robert Russell and colleagues constructed a food guidelines pyramid, which specifically focused on the health of the elderly with more generous allowances in some areas (e.g., vitamin E, dietary fiber, liquids) and more extreme restrictions (e.g., sodium, solid fats), but it was based more on a potpourri of published findings about nutrient and food associations than any systematic review with agency backing.

Barriers to Meeting Recommended Nutrient Intakes and Healthful Dietary Intake Patterns by Older Persons

The late Professor Doris Calloway, in the early 1970s, commented: "People eat food, not nutrients." This highlights the paradoxes in considering and enumerating the objectives of dietary intake at the level of the chemical composition, whereas most members of the general public are uninformed as to the nutrient composition of the foods and beverages in their diets.

Elderly persons face a number of challenges in meeting their RNIs. In the first instance, they are likely to be those with the least sophisticated or available knowledge of the nutrients required and the food sources to provide them. The social, economic, and physiological changes imposing on the lives of persons surviving to advanced age pose logistical problems for their selecting and purchasing a diet. Economic dependency and the limited incomes of older persons may restrict their access to high-quality foods. Social isolation, depression, and impaired mobility, as well as chewing difficulties, may limit the variety of items included in the diet with advancing age. In some circumstances, it may be that free-living and independent elders are relatively less able to optimize their nutrient intake and dietary pattern compared to the more dependent individuals served or fed in institutional settings.

The exigencies of consuming a healthful diet for the prevention of chronic diseases, emphasizing a plant-based diet rich in whole grains, fruits, and vegetables, limits the nutrient selection that would be obtained from an even wider variety of foods and food groups. Specific essential fatty acids, and certain minerals (calcium, zinc, selenium) and some vitamins are far less nutrient dense

in foods of vegetal origin, setting a dilemma between consuming for nutrient adequacy and prevention of degenerative disease. Two nutrients with accentuated requirement levels in later life – vitamin D and calcium – have so few rich dietary sources that the elderly may be able to afford or assimilate, that supplemental forms will most likely be required.

Future Considerations

The DRI recommendations are specifically derived for the populations of the US and Canada in North America. The RNIs of the UN System are meant to be universal across the entire world. With respect to meeting nutrient intakes, increased selection of fortified foods by older individuals may contribute to the closing of any intake gaps or deficits. Fortified food consumption, however, has a number of caveats. The need for iron is one nutrient requirement that decreases with advancing age, at least for postmenopausal women. It may actually be that both sexes would benefit from a lesser burden of exposure to the oxidant effects of iron with advancing age. Similarly, the folate requirement has been set with interest in reproductive matters (prevention of neural tube defects), which are of no biological relevance in later life. Some, as yet inconclusive, evidence that colonic and prostatic tissues may receive dangerous proliferative stimulation from folic acid, the pharmaceutical form of the vitamin, has been reported. On the safety side of the equation, however, the effective upper tolerable levels for certain nutrients in later life may prove to be lower than those for younger members of the adult population.

The slight majority of the living elderly are currently to be found in the low income, largely tropical regions of the world in which 80% of the global population reside; this shift is due to rise rapidly over the next two decades. With specific reference to low-income societies, a number of caveats apply to the estimation of nutrient intake recommendations for the elderly across the world. If the "applies only to healthy individuals" disclaimer were applied to the developing world, then virtually no older people would qualify as eligible for coverage by any nutrient recommendations system. However, rather than abandon the effort for nutrient intake guidance, an attempt should be made to take into account the influences of life-long climatic issues (heat, humidity) and ecological factors (parasites, recurrent infections) on nutrient needs in later life.

Nature *versus* nurture issues will also continue to be debated with regard to nutrient requirements, especially in later life. The revelation of the human genome (complete genetic code), has given rise to the issues of 'nutrigenomics' and 'nutrigenetics'; theoretically, it could soon be possible to understand individual variation in needs for and tolerances of essential and nonessential nutrients and dietary bioactive substances. The significance of this potential for the already aged person, however, is likely to be limited for two reasons. First, the accumulative effects of nutrient imbalance will already have been established. Second, the economic and intellectual wherewithal to access and execute such individualized prescriptions for nutrient intakes and dietary patterns will likely escape the majority of older persons with limited financial means. Hence, further refinements in recommended intakes for older persons are likely to remain at the public health level for this segment of the population, and will involve establishing evidence that increased intakes of specific nutrients and will have health-protective effects or function-enhancing properties.

Further Reading

Allen, L., de Benoist, B., Dary, O., Hurrell, R., 2006. Guidelines on Food Fortification with Micronutrients. World Health Organization and Food and Agricultural Organization of the United Nations. Geneva: WHO.

Brown, B., Huang, M.H., Karlamangla, A., Seeman, T., Kado, D., 2011. Do the effects of APOE-ε4 on cognitive function and decline depend upon vitamin status? MacArthur Studies of Successful Aging. Journal of Nutrition, Health, and Aging 15, 196–201.

Food and Nutrition Board, Institute of Medicine, 2010. Dietary Reference Intakes for Vitamin D and Calcium. National Academy Press, Washington DC.

Fries, J.F., Bruce, B., Chakravarty, E., 2011. Compression of morbidity 1980–2011: A focused review of paradigms and progress. Journal of Aging Research 2011, 261702.

Hausman, D.B., Fischer, J.G., Johnson, M.A., 2011. Nutrition in centenarians. Maturitas 68, 203–209.

James, W.P., 2010. 23rd Marabou Symposium: Nutrition and the Aging Brain. Nutrition Reviews 68 (supplement 2), S162–S167.

Rowe, J.M., Kahn, R.L., 2000. Successful aging and disease prevention. Advances in Renal Replacement Therapy 7, 70–77.

Russell, R.M., 2000. The aging process as a modifier of metabolism. American Journal of Clinical Nutrition 72 (supplement 2), 529S–532S.

Russell, R.M., Rasmussen, H., Lichtenstein, A.H., 1999. Modified food guide pyramid for people over seventy years of age. Journal of Nutrition 129, 751–753.

Solomons, N.W., 2002. Nutrition and the extremes of life: Dilemmas and enigmas of advanced old age. Asia Pacific Journal of Clinical Nutrition 11, 247–250.

Woo, J., 2011. Nutritional strategies for successful aging. Medical Clinics of North America 95, 477–493.

World Health Organization/Food and Agricultural Organization (2003) Diet, nutrition and the prevention of chronic diseases. *Report of the Joint WHO/FAO Expert Consultation*. WHO Technical Report Series, No. 916. Geneva: WHO.

Older people: Physiological changes

N. Solomons, Aging and Metabolism (CeSSIAM), Guatemala City, Guatemala

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of N. Solomons, Older People: Physiological Changes, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 400–404, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00219-1.

The Aging of the Population and its People **157**
The Nature of Senescence **157**
Cellular Senescence 158
Telomeres and Telomerase 158
Apoptosis 158
Mitochondrial Senescence and Oxidative Stress 158
Physiological Changes Occurring in Tissues and Organ Systems with Human Aging **158**
Integumentary Tissues 158
Pulmonary and Respiratory System 159
Cardiovascular and Circulatory System 159
Oral Cavity and Alimentary Tract 159
Musculoskeletal System 159
Renal and Urogenital System 160
Gonads and Reproductive System 160
Endocrine Systems and Metabolism 160
Hematopoietic and Immune System 160
Hematological Aging 160
Immunological Aging 160
Central and Peripheral Nervous System 161
Central Nervous System 161
Special Senses 161
Cognitive Function 161
Peripheral Nervous System 161
Drug Metabolism 162
Synthesis and Conclusion **162**
Further Reading **162**

The Aging of the Population and its People

The maximal human life span is approximately 120 years. Approaching this degree of longevity, however, was not a prominent feature in the evolutionary phases of our species, *Homo sapiens*. The imperative was to survive the various mortal hazards long enough to reproduce and provide initial care for the offspring. The twenty-first century has ushered in an unprecedented longevity. The life expectancy of infants born today in Western Europe or Japan is more than 75 years. The most rapidly increasing population segment in the world today is the centenarian. By the year 2020, there will be more than 1 billion people older than 60 years of age, constituting 13.3% of the global population, and three-quarters of them will be living in developing countries.

Many people are living a long time, but not all of them are healthy and functional throughout their lifespan. Chronic disability and the cost of health services and custodial care are a growing burden on the economies of developed and developing countries alike. To understand the pathological aspects of advancing age, the normative pattern of changes in physiological function in older persons is an essential benchmark.

The Nature of Senescence

Aging has been described as "a series of time-related processes that ultimately bring life to a close," that is, a process of physiological 'wearing out.' Physiology is the basis of human functionality as well as of our susceptibility to disease. The late gerontologist, Nathan Shock, established the principle of a progressive decline in physiological reserves as a consequence of 'normal' aging, recognizing that the rate of decline differed markedly among the body's organ systems. In fact, one cannot really separate the concept of the physiology of older persons from the physiology of the aging process itself. Similarly, the high prevalence of chronic diseases in older persons challenges our ability to discriminate 'normative' senescence from pathophysiological changes.

The origin of physiological changes in older persons begins within the domain of cellular senescence. The extension to tissue and organ levels originates in what we interpret to be the physiological changes of human aging. Major advances in our cellular and molecular understanding of basic aging processes have been made in recent years.

Cellular Senescence

In most tissues, with the notable exception of neural tissue, healthy cells are replicating cells, which are capable of mobilizing at least 20 enzymes and proteins that must be preassembled to initiate deoxyribonucleic acid (DNA) synthesis for cell division. An irreversible state of growth arrest known as replicative senescence is the fundamental basis of cellular aging. Such senescent cells remain viable and metabolically active, but their genomic function and protein expression are distinct from that of normal, proliferating cells. Iron accumulates in senescent cells, possibly contributing to the greater oxidative stress and cellular dysfunction seen in senescent cells. Senescent cells also express proinflammatory enzymes, an internal process that could possibly contribute to the aging process; intercellular adhesion molecules, which are part of the inflammatory response, are overexpressed in association with senescent cells and aging tissue.

Telomeres and Telomerase

Telomeres are small units composed of the tandem DNA repeats and associated proteins, which cap the end of linear chromosomes and are responsible for maintaining chromosome length. They provide stability to the chromosome and protect against DNA loss associated with cellular replication. The mechanism of replicative arrest of senescent cells has been related to changes in the function of telomerase, a nuclear enzyme that synthesizes and maintains the telomeres. Shortening and uncapping of these structures, related to the number of past cell divisions, renders the DNA strand incapable of replication.

Apoptosis

Another factor involved in aging at the cellular level is the orderly 'retirement' of cells. For every cell that divides in, another would somehow have to make space for the extra cell in order to maintain numerical stability in the organ. This is achieved by a process of programmed cell death, known as 'apoptosis.' Cell senescence disrupts these apoptotic processes. Necrosis, by contrast, is cell death due to injury or noxious stimuli. Diseases of aging may favor the necrotic process.

Mitochondrial Senescence and Oxidative Stress

The intracellular mitochondria, organelles involved in energy metabolism, are central to the process of cell senescence. They are also involved in regulating thermogenesis, calcium buffering, and integrating apoptosis. With aging, mitochondria become less efficient, partly due to mutations in the cell nucleus, derepressing the expression of proteins that compete with mitochondrial function. This disrupts energy metabolism for the cell and makes the mitochondria more porous, releasing reactive oxygen species into the rest of the cell. The mitochondrial production of reactive oxygen species is inversely proportional to longevity in animals. The oxidative activity also damages the mitochondria themselves. Mitochondria have their own DNA strands, and these accumulate mutations with age. In tissues dependent on progenitor (stem) cells, mitochondrial DNA mutations can disrupt replication.

Free radicals and reactive oxidative species can produce mutations in nuclear material and oxidize proteins and lipids throughout the cells. Aging involves an accumulation of oxidative damage at the cellular level, if not an increase in its intensity as well. The thiol-containing antioxidant mechanisms, typified by glutathione but represented by a number of sulfur-containing species, represent an important buffer against intracellular free radicals, but decline with age due to downregulation of their synthetic enzymes. Confirming the cellular trend to oxidative stress in aging cells, clinical biomarkers of oxidation, and antioxidant mechanisms reveal that systemic oxidative stress increases with aging characterized by lower concentrations of vitamins E and C and carotenes as well as lower activities of Cu–Zn–superoxide dismutase, catalase, and glutathione peroxidase.

Physiological Changes Occurring in Tissues and Organ Systems with Human Aging

Physiology has classically been organized around organ systems. According to this convention, the important features of the age-associated changes are enumerated and synthesized, with implications for human nutrition.

Integumentary Tissues

The integumentary tissues (skin, hair, and nails) cover and protect the body. Two of the more classical and reproducible manifestations of aging can be seen in this system. The depigmentation of hair to gray or white is an almost universal aging effect given its sufficient survival. Wrinkling of the skin, due to alteration in connective tissue composition, is another consequence of aging; it should be assessed by the changes in skin texture only in the nonsun-exposed regions of the body. Beyond the cosmetic consequences of the aging integumentary tissues, wound healing is a health-relevant consideration. Healing of wounds is slower with

increasing age, but the resulting scars have the same tensile strength. Reduced recruitment of vessels of the microvascular is a function of aging.

The skin is an endocrine organ. Vitamin D is produced from the conversion of 7-hydroxy-cholesterol to cholecalciferol in the dermis of the skin. The efficiency of vitamin D decreases with age, such that older persons need a longer exposure to solar radiation to produce a given quantity of the vitamin.

Pulmonary and Respiratory System

Compliance of the chest wall changes with age, which gets stiffer and less compliant. The muscular force of the diaphragm is reduced with advancing years. The combination of these two factors reduces the maximal amount of air that can be moved into and out of the lungs. This diminution in the so-called forced vital capacity of the lungs occurs as one gets older. There is less compliance, less recoil, and greater dead space. The original lung capacity, however, is sufficient to allow for sufficient gas exchange throughout life in the absence of underlying pulmonary disease. Nonetheless, the longitudinal Framingham Heart Study found an association between decrease in lung capacity and all-cause mortality.

The hygiene of the respiratory airways is somewhat compromised by a decreasing function of the microcilia of the bronchial epithelial cells. Because this mechanism is used to clear microbial pathogens, it has a direct influence on host defenses. Finally, because the basis of the respiratory system is an exchange of gases (oxygen, carbon dioxide, trace gases) with the bloodstream, any cardiovascular changes involving the side chambers of the heart will influence the overall gas exchange efficiency for the body.

Cardiovascular and Circulatory System

For this system, it is necessary to separate the aging effects on the cardiac muscle and its apparatus from the aging of the vessels of the circulatory system, which transports blood to and from the heart. A characteristic of aging is a diminished resting cardiac output, which can have the combined bases of lower force of the cardiac muscle and a lesser oxygen demand for metabolism with diminished active cell mass. Aging of the myocardium reduces its capacity for cellular repair and replacement. With aging, elevations of noradrenaline (norepinephrine) associated with downregulation of β-1 receptors mimic the process of the failing heart. The compliance of the arteries emanating from the heart decreases with age. Stiffening of these vessels produces a progressive rise in the systolic blood pressure.

It is the circulation through smaller blood vessels and the generation of new vessels (neovascularization) that is a major concern with advancing years. The process of angiogenesis, through which new blood vessels are formed, is impaired during aging. The integrity of endothelial cells lining the vessels, the cascade of coagulation factors, and growth factors and neurochemical mediators, and their respective receptors are all altered by aging in the neovascularization processes.

Oral Cavity and Alimentary Tract

The digestive tract is subject to functional changes with aging. Beginning in the oral cavity, loosening and loss of teeth is a frequent companion of aging. Saliva secretion decreases leading to relative degrees of xerostomia or dry mouth.

Reduced parietal cell function develops in older persons, but prior *Helicobacter pylori* infections are now thought to be a major cause of hypochlorhydria in later life. An important nutritional consequence of reduced gastric acid secretion is a lesser biological availability of iron. Because iron stores are generally replete in both men and women in later life, this has little practical nutritional impact. The reduced secretion of gastric intrinsic factor, however, contributes to vitamin B_{12} deficiency, which is an important nutritional problem of older persons.

The capacity of the liver for biliary secretion and the pancreas for digestive enzyme and bicarbonate secretion begins adult life with a >90% excess of the necessary minimum. Secretory function declines with increasing age, but rarely falls below the minimal reserve capacity. The metabolic and detoxifying capacity of the human liver also has a reserve capacity and is not usually compromised by normal aging.

Intestinal motility is reduced with aging as a result of functional changes in the visceral nerves. With decreased transit, the residence time of the chyme on the absorptive surfaces is longer, compensating for any senescence in the mucosal uptake itself. The reduction in motility produces the most noticeable and notorious of the manifestations of intestinal health in older persons, namely reduced frequency of defecations.

Musculoskeletal System

Bone mineral content declines with age; this aging process is known as 'osteopenia.' (It should be distinguished from the related pathological process in which bone architecture is altered, producing 'osteoporosis.') From the peak in the third and fourth decades, a 30% average decline in bone mineral density occurs through the ninth decade. In women, there is well-characterized acceleration of the rate of bone mineral loss immediately following the menopause. Decreasing levels of anabolic hormones may be associated with musculoskeletal atrophy and decrease in function that is observed in older women. This change in skeletal mineralization with aging is not associated with any apparent change in vitamin D nutriture as reflected in circulating levels of the vitamin.

The joints of the body undergo changes with the senescence of replacement of the cartilaginous substance, complicated by the pathological effects of cumulative use over the life span.

Recently, increasing attention has been given to the loss of muscle strength and substance with increasing age. Sarcopenia loss of lean body mass skeletal muscle mass replacement by fat mass decreased creatinine-to-height ratio in normative aging in healthy subjects diminished grip strength is a function of age. Reduction in muscle mass (sarcopenia obesity) is an important determinant of physical function and metabolic rate.

Renal and Urogenital System

Renal creatinine and inulin clearance decreases with aging have been demonstrated for decades. These functional changes in filtration are associated with changes in the glomerular structure in the kidney. Circulatory senescence decreases blood flow to the kidneys, which further reduces the efficiency of renal clearance. The reserve capacity of these organs is such, however, that age-associated glomerular decline per se does not compromise the net excretion of nitrogenous waste.

Urine flow at the outlet is another aging consideration. The male urogenital system undergoes a characteristic aging change in the hypertrophy of the prostate gland, associated with decreased secretion of prostatic fluid. The anatomical consequence is a constriction in the passage through which urine flows from the bladder.

Gonads and Reproductive System

It has been aptly stated by Harman that: "It is clear that aging results in alterations of endocrine physiology, which in turn appear to contribute to development of the senescent phenotype." Aging is associated with a decrease in pituitary hormone secretions. This decline explains, in part, the reduction in gonadal hormone production with aging. Primary aging of the testes and ovaries themselves accounts for the remainder of the changes. As the ovaries have a finite number of eggs, ovulation can only continue through the number of cycles that correspond to the original store of ova. Menopause ensues with the characteristic cessation of estrogenic hormone secretion. In both sexes, gonadal androgenic hormone production declines with consequent effects on libido.

Endocrine Systems and Metabolism

As stated above, the pituitary gland is the hub of endocrine regulation. Important among the decline stimulation within the axis is that growth hormone (GH) secretion declines with increasing age, a condition termed 'somatopause.' The changes in the GH/insulin-like growth factor axis with aging produce changes in function, metabolism, and body composition analogous to the pathological GH deficiency seen in younger adults. Another change with age is the efficiency with which physical activity stimulates the secretion of GH.

The availability of hormones is not the only variable in endocrine signaling. Cellular and intracellular receptor function is complementary. An attractive explanation for the disordering of hormonal axes is oxidative damage to cell membranes, compromising the function of receptors.

Basal and resting metabolism and diet-induced thermogenesis are all reduced with increasing age. Changes in body composition, and the replacement of lean tissue with fat and the increasing visceral distribution of fat, as well as decreasing physical activity, influence these metabolic changes of aging. Basal metabolic rate declines in aging more than can be attributed to body composition changes and intracellular mitochondrial senescence may explain part of this discrepancy. For practical purposes, the standard oxygen consumption value equivalent to one metabolic equivalent, that is, 3.5 ml min^{-1} kg^{-1}, is not appropriate for elderly people.

Hematopoietic and Immune System

The formation of new red and white blood cells and platelets is one of the most proliferation-dependent physiological processes of the body. The various classes of circulating white cells are the underpinning of the host defense system, together with tissue macrophages, hepatic proteins, and the alimentary tract's mucosa.

Hematological Aging

The blood-forming organ is the bone marrow. Aging is associated with fatty infiltration of the marrow spaces in the long bones, but enough marrow remains to support the turnover of erythrocytes and red blood cell lines. The circulating red blood cell mass neither changes normally with advancing age nor does the normative peripheral white cell count or platelet number. As noted, iron stores tend to be abundant in later life; nutritional problems influencing red blood cell production are based on alterations in gastric function (vitamin B_{12} malabsorption), which result in a macrocytic (megaloblastic) anemia.

Immunological Aging

Circulating phagocytic white blood cells counts do not reduce with aging but aging does influence the innate host defense system. Mucosal barrier functions are influenced by aging of the gut in its interaction with microflora. Although not reduced in number, aged macrophages and neutrophils have blunted intracellular signaling by specific receptors, decreased metabolic functions, and

impaired bacterial killing. Production of superoxide anion, chemotaxis, and orderly apoptosis of neutrophils is also disrupted by the disordered signaling. The tumor cell-destroying capacity of natural killer cells in the elderly is diminished.

More profound changes occur in the adaptive immune functions, which rely on the memory (T-cell) lymphocytic cell line. Life-long antigen exposure induces increases in the number of memory T-cells, but with enhanced reactivity against self-antigens, priming the individual for autoimmune disease. In healthy adults, immunoglobulin A concentration increases by $0.2\ g\ l^{-1}$ per decade throughout life. The T-lymphocytes, however, respond more poorly to ongoing antigen assault in later life. Thymic involution associated with neural and hormonal changes of aging is an impediment to T-cell maturation in older persons. The basis of intrinsic function deficits of memory cells, however, has been ascribed to defective signaling and includes hyporesponsiveness to mitogen-stimulated proliferation and decrease in genetic suppression, allowing increased stimulation of inflammatory cytokines; the balance between pro- and anti-inflammatory cytokines shifts with aging, favoring the inflammatory pole, especially with the greater expression of interleukin 6. This has a negative systemic effect on bone metabolism, as well as dysregulating overall immune function.

Aging of mitochondria in the immune cell lines produces increased intracellular reactive oxygen species burdens. Finally, there is diminished programmed death (apoptosis) of immune cells and dysregulation of apoptosis-dependent functions.

Central and Peripheral Nervous System

The integration of all senses and origins of all systemic coordination is a function of the brain and central nervous system. This is the one system in which proliferation of the primary cells (neurons) is not an issue after early childhood, although the supportive, nerve-tending (glial) cells continue to depend on replication and apoptosis for normal function.

Central Nervous System

The neurons of the brain continue to divide only through to the second year of life. Thereafter, the goal is to preserve the number and health of the cerebral nerve cell mass. Myelination of axons of nerve cells must be maintained throughout life. This is the function of the supporting cells (oligodendrocytes), which for more than 40 years continue to differentiate into myelin-producing cells. Free radicals pose a threat to these axon-tending cells, whose metabolic demands for producing the brain's cholesterol and maintaining its array of myelin sheaths render them particularly vulnerable to stress.

Positron emission tomography imaging of the aging brain has revealed and mapped the plethora of changes in blood flow and neurotransmitter metabolism that occurs with advancing years.

Special Senses

The special senses related directly to the cranial nerves (vision, hearing, taste, and smell) experience age-related change. With respect to vision, the most typical of all biological aging changes is presbyopia, or the loss of accommodation function for the ocular lens with loss of capacity of the associated musculature. The consequence is loss of near-vision, which leads to the need for reading glasses or bifocal spectacles. A more important aging change related to the lens is the opacification that leads to cataract formation. The eye is designed to translate light energy into visual images, but the energy of light, particularly the ultraviolet β-rays of solar energy, damages ocular tissue. Thus, there is as a strong environmental component to the disarranging of the laminar stacking of the fibrillar proteins of the lens, which imparts its clear, transparent basis; consumption of diets high in antioxidant vitamins has been associated with the delay in cataract formation.

Age-related hearing loss is a feature of biological aging. It affects the cochlear neural structures and leads to loss of acuity, especially for higher pitched tones. It is speculated that apoptosis of the most vital neural cells drives this hearing loss, based on mutations in the mitochondria due to life-long free-radical stress.

Taste and smell acuity decline with aging, both in sensitivity and in accuracy of recognition. Because these combined senses account for the recognition of flavors, their diminution with age could affect appetite and reduce the enjoyment of meals.

Cognitive Function

The intellectual, reasoning, and memory functions of the cerebral cortex decline with increasing age. This has been a universal observation in general elderly populations. The debate is whether this is a consequence of neurodegenerative diseases (pathological change) or a biological correlate of aging (senescence). Continued intellectual stimulation has been posited as an approach to retard cognitive decline, and a role for B-complex vitamins and antioxidants has been advanced.

Peripheral Nervous System

Vibratory perception in the peripheral extremities is the classical index of peripheral nervous decline with aging. Less well appreciated is the effect of aging on pain perception, in which there can be a numbing of sensation or, less commonly, an accentuation of perception. Pain perception from the visceral organs is often dulled, which can have adverse implications for the early detection of organic diseases. All of the peripheral nerve dysfunction can result from the compensatory sprouting of axonal limbs to compensate for the loss of motor neurons. This is well directed at first, but with further aging, the synaptic connections are poorly directed and motor function suffers as a consequence.

Drug Metabolism

The metabolism of drugs and pharmacological agents is not the purview of any single organ system. Older persons tend to be prescribed increasing numbers of medications with advancing age. Important changes in drug metabolism occur with aging. Metabolism and disposition of drugs change with age. This involves age-associated decrease in the function of some, but not all, cytochrome P-450 enzymes. Among the pharmacokinetic and pharmacodynamic changes that occur with advancing age are reductions in renal and hepatic clearance and an increased effective half-life of lipid-soluble drugs. The older population shows increased sensitivity to some psychotropic drugs and anticoagulants, with the frail elderly being more susceptible than healthy elders.

Synthesis and Conclusion

The number of older people is increasing in all regions and all societies of the world. Advancing age produces senescent changes in cellular function that are reflected in a declining capacity of all physiological systems. The increased prevalence of disease in older population aging is a major risk factor for disease but does not necessarily lead to age-related diseases.

All physiological systems are intrinsically interrelated in maintaining the health and function of the organism. Aging is associated with a loss of complexity in the dynamics of many physiological systems. It has been speculated that the basis for the syndrome of frailty in older persons may result from a reduced ability to adapt to internal and external stresses of daily life due to the loss of dynamic coordination among the interrelated physiological systems.

The alterations in physiological functions with aging have important implications for absorbing, retaining, and utilizing nutrients. The extent to which dietary patterns and nutrient intakes are accelerating or retarding the rates of functional decline is a matter of ongoing investigation in gerontological nutrition and physiology.

Further Reading

Ahluwalia, N., 2004. Aging, nutrition and immune function. Journal of Nutrition, Health and Aging 8, 2–6.
Balin, A.K. (Ed.), 1994. Practical Handbook of Human Biologic Age Determination. CRC Press, Boca Raton.
Harman, S.M., 2004. What do hormones have to do with aging? What does aging have to do with hormones? Annals of the New York Academy of Science 1019, 299–308.
Hayflick, L., 2003. Living forever and dying in the attempt. Experimental Gerontology 38, 1231–1241.
Hutchinson, M.L., Munro, H.N., 1986. Nutrition and Aging: Bristol-Meyer Nutrition Symposia, vol. 5. Academic Press.
Leveille, S.G., 2004. Musculoskeletal aging. Current Opinion in Rheumatology 16, 114–118.
Lipsitz, L.A., 2004. Physiological complexity, aging, and the path to frailty. Science of Aging Knowledge Environment 16, 16.
Mishra, S.K., Misra, V., 2003. Muscle sarcopenia: An overview. Acta Myol 22, 43–47.
Park, H.L., O'Connell, J.E., Thomson, R.G., 2003. A systematic review of cognitive decline in the general elderly population. Internation Journal of Geriatric Psychiatry 18, 1121–1134.
Timiras, P.S., 1994. Physiological Basis of Aging and Geriatrics, 2nd edn. CRC Press, Baton Raton.

Pregnancy: Energy requirements and metabolic adaptations

G.R. Goldberg, MRC Human Nutrition Research, Cambridge, UK

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of G.R. Goldberg, Pregnancy: Energy Requirements and Metabolic Adaptations, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 56–60, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00233-6.

Extra Energy Costs of Pregnancy **163**
Tissue Deposition 163
Maintenance Energy Costs of Pregnancy 163
Theoretical Total Metabolic Costs of Pregnancy 164
Longitudinal Studies of the Energy Costs of Pregnancy **164**
Fat Deposition 164
Basal Metabolic Rate 164
Diet-Induced Thermogenesis 165
Energy Cost of Activities 166
Behavioral Changes in Physical Activity 166
Between-Country Comparison of the Metabolic Costs of Pregnancy **166**
Individual Variability in the Total Energy Costs of Pregnancy **166**
Implications of Energy-Sparing Adaptations for Mother and Infant **166**
Further Reading **167**

The subject of energy metabolism in human pregnancy has received extensive consideration for more than 60 years, dating back to early work that assessed the contribution of fetal metabolism to the overall energy costs of pregnancy. Since then, the emphasis of much work has been on separating and quantifying the different components of gestational energy needs and on establishing appropriate recommendations for the energy requirements of pregnant women, with the intention of quantifying average amounts. Deviations from average values were mostly regarded as undesirable biological or measurement noise that needed to be overcome by studying large samples of women to get a more precise estimate of the mean values. These interindividual variations in the metabolic responses to pregnancy are increasingly recognized as biologically significant 'plasticity' that has true adaptive value in enabling women to carry a pregnancy to term under a wide range of nutritional conditions. The shorter- and longer-term consequences of such adaptations are being explored as part of fetal and infant origins of adult disease hypotheses.

Extra Energy Costs of Pregnancy

The question of how much extra dietary energy a pregnant woman needs is closely linked to the question of the amount of weight she should gain during pregnancy. This in turn is linked to her age and to her prepregnant body mass index as a proxy for energy status.

Hytten and Leitch's theoretical estimations of the overall energy costs of human pregnancy published more than 30 years ago have subsequently been experimentally validated as reasonable average values, and they have been adopted by many national and international bodies as a partial basis for developing recommended energy intakes in pregnancy. The costs can be divided into three main components: the energy deposited as new tissue in the conceptus, the energy deposited as fat, and the energy required to maintain this new tissue.

Tissue Deposition

Weight gain during pregnancy consists of the fetus, placenta, and amniotic fluid (the products of conception) and the extra growth of several maternal tissues. The deposition of fat in pregnancy is presumed to help meet the extra energy demands of lactation. The total energy deposited as new tissue, excluding maternal fat, averages approximately 49 MJ (11 700 kcal). If an average maternal fat gain of 2.6 kg is assumed, then the estimate of the total energy deposited as new tissue during an average pregnancy is approximately 174 MJ (41 600 kcal; **Table 1**).

Maintenance Energy Costs of Pregnancy

Because of the increase in tissue mass, the body's oxygen consumption also increases during pregnancy. Estimates suggest that the increase in oxygen consumption is equivalent to an extra 187 (45), 414 (100), 620 (148), and 951 (230) kJ day^{-1} (kcal) at 0–10,

10–20, 20–30, and 30–40 weeks of gestation, respectively. The total maintenance cost for an average human pregnancy is approximately 150 MJ (35 800 kcal); (Table 2).

Theoretical Total Metabolic Costs of Pregnancy

Compared to many other mammals, humans have a relatively small and usually single infant, which develops during a long gestation period. The energy stress to the mother is therefore low per unit time. The 49 MJ of energy deposited as the products of conception represents only 4 or 5 days of food intake for the mother. Humans also differ from most other mammals because their large fat stores can help meet some of these costs. The theoretical total metabolic costs (i.e., due to extra tissue and increased metabolism) of pregnancy are approximately 335 MJ (80 000 kcal), or 1.25 MJ day^{-1} (300 kcal). This value does not make any allowance for changes (increases or decreases) in energy expended on physical activity. It is assumed that the majority of the energy costs of human pregnancy are met by behavioral adjustments in energy metabolism rather than increased energy intake. This assumption has formed the basis for energy intake recommendations, some of which are summarized in Table 3. It should be noted that the 1985 estimates used by World Health Organization (WHO)/Food and Agriculture Organization (FAO)/United Nations University (UNU) are under revision. Future recommendations may separate the obligatory costs (e.g., by fixed increments for basal metabolic rate (BMR) and tissue deposition) and differences in physical activity (based on PAL values).

Longitudinal Studies of the Energy Costs of Pregnancy

Fat Deposition

The increase in maternal fat stores is by far the largest contributor to the energy cost of tissue deposition. It is also the most variable. Although the average increase for a well-nourished woman who has an uncomplicated pregnancy and healthy infant is approximately 3 kg, a large number of studies have reported ranges of −2 to 8 kg and standard deviations of 2–4 kg. There is also a wide range in fat deposition between different populations, particularly when those from developed and developing countries are compared. Fat is very energy dense and therefore changes in body fat stores have a large impact on the energy costs of pregnancy. A loss of 2 kg saves approximately 78 MJ (18 600 kcal), whereas a gain of 8 kg costs approximately 312 MJ (74 600 kcal). Women most likely to need an energy reserve to help meet the costs of lactation are often those who are least able to deposit spare energy as fat in pregnancy. Conversely, women who store large amounts of fat during pregnancy are least likely to need to use it during lactation. They are often able to increase food intake and/or decrease physical activity instead. Studies have shown that excess energy intake during pregnancy results in excess maternal weight (and fat) gain. Postpartum retention of excess fat has implications for the development of obesity and its comorbidities such as type 2 diabetes.

Basal Metabolic Rate

The cumulative increase in BMR can comprise a large part of the total energy costs of pregnancy. Although 150 MJ is a good estimate of the average energy cost of maintenance for a well-nourished woman, there is a very wide range. This has an important influence on the extra daily requirements for individual women. Studies in which BMR has been measured every 6 weeks from prepregnancy to 36 weeks of pregnancy have shown very marked differences. In some women, there is the expected response to pregnancy – an

Table 1 Protein and fat deposition during pregnancy for a reference woman

Site	*Protein*		*Fat*		*Water (kg)*	*Total*	
	kg	*MJ (kcals)*	*kg*	*MJ (kcals)*		*kg*	*MJ (kcals)*
Fetus	0.44	12.76 (3050)	0.44	20.24 (4840)	2.41	3.29	33.00 (7890)
Placenta	0.10	2.90 (690)	0.04	0.18 (43)	0.54	0.64	3.08 (740)
Amniotic fluid	0.003	0.09 (21)	0.00	0.00	0.79	0.79	0.09 (21)
Uterus	0.17	4.81 (1150)	0.04	0.18 (43)	0.80	0.97	5.00 (1200)
Breasts	0.08	2.35 (560)	0.12	0.55 (130)	0.30	0.40	2.90 (690)
Blood	0.14	3.92 (940)	0.02	0.92 (220)	1.29	1.44	4.84 (1157)
Water	0.00	0.00	0.00	0.00	1.50	1.50	0.00
Subtotal	0.93	26.83 (6400)	0.48	22.08 (5280)	7.63	9.04	48.9 (11 700)
Fat stores	0.07	1.94 (460)	2.68	123.10 (29 400)	0.60	3.35	125.04 (29 900)
Total	0.99	28.77 (6900)	3.16	145.18 (34 700)	8.24	12.38	173.94 (41 600)

Source: Adapted from Prentice AM, Spaaij CJK, Goldberg GR, *et al.* (1996) Energy requirements of pregnant and lactating women. *European Journal of Clinical Nutrition* 50(supplement 1): S82–S111, with permission from Nature.

Table 2 Increases in oxygen consumption during pregnancy

	ml min^{-1}			
	10 weeks	*20 weeks*	*30 weeks*	*40 weeks*
Cardiac output	4.5	6.8	6.8	6.8
Respiration	0.8	1.5	2.3	3.0
Kidneys	7.0	7.0	7.0	7.0
Breasts	0.1	0.6	1.2	1.4
Uterus	0.5	1.2	2.2	3.6
Placenta	0	0.5	2.2	3.7
Fetus	0	1.1	5.5	12.4

Source: Adapted from Hytten FE (1991) Nutrition; Weight gain in pregnancy. In: Hytten F and Chamberlain G (eds.) *Clinical Physiology in Obstetrics*, 2nd edn. Oxford: Blackwell Scientific, with permission from Wiley.

immediate and progressive increase in BMR. In other women, BMR actually decreases or increases only slightly in the early stages of pregnancy and does not increase substantially until late gestation. This offsets the later increase in BMR such that there is actually a slight net saving of energy over the entire gestation period in some of these 'energy-sparing' women. The total net cost of maintenance, estimated as the cumulative area under the curve represented by the rise in a mother's BMR above the prepregnancy baseline metabolic rate, is negative or only very small. Data indicate that this between-subject variability is found in women from both well-nourished and marginally nourished populations. However, 'energy-sparing' and 'energy-profligate' responses dominate in marginally and well-nourished women, respectively. There is a more than fivefold range between the most energy-profligate and the most energy-sparing women.

In addition to the wide variability in changes in BMR between individual women, there are also wide variations between different populations. Well-nourished affluent women from developed countries tend to show an energy-profligate increase in BMR. In marginally nourished thinner women from developing countries the increase in BMR is delayed and/or preceded by a decline in early pregnancy. The total maintenance costs of pregnancy in these studies range from +210 MJ (+50 000 kcal) to −45 MJ (−11 000 kcal).

Diet-Induced Thermogenesis

A reduction in diet-induced thermogenesis (DIT) may be a mechanism by which energy is saved during pregnancy. However, when expressed as a proportion of energy intake, DIT remains essentially unaltered during pregnancy and any changes are small and unlikely to be biologically significant.

Table 3 Examples of current recommendations for energy intakes during pregnancy

	Trimester(s)	*Increment, MJ day^{-1} (kcal day^{-1})*	*Total for pregnancy, MJ (kcal)*	*Qualifying comments*
FAO/WHO/UNU (1985)	All	1.20 (300)	336 (80 300)	
	All	0.84 (200)	235 (56 150)	For healthy women who reduce activity Energy and protein requirements are undergoing revision (interim report published 2004)
United Kingdom (1991)	Third	0.80 (190)	74 (17 000)	Underweight women and those not reducing activity may need more
United States and Canada (2002)	First	Adult EER+0		For women aged 19–50 years
	Second	Adult EER+160 kcal (8 kcal week^{-1}×20 weeks)+180 kcal		EERs for pregnant adolescents are based on EER for 14- to 18-year olds
	Third	Adult EER+272 kcal (8 kcal week^{-1}×34 weeks)+180 kcal		EER is based on total energy expenditure in the nonpregnant state; increments for pregnancy are 8 kcal week^{-1} for total energy expenditure and 180 kcal day^{-1} for tissue deposition

EER, estimated energy requirement.

Energy Cost of Activities

Results from a number of longitudinal studies have shown that the cost of non-weight-bearing activity changes little until very late pregnancy. From approximately 35 weeks, the gross costs (which include changes in BMR) increase by approximately 11% and net costs by approximately 6%. The gross and net costs of weight-bearing exercise (treadmill walking and standardized step testing) remain fairly constant during the first half of pregnancy and then increase progressively by approximately 15–20% at term.

Behavioral Changes in Physical Activity

It has frequently been assumed that a behavioral reduction in the energy expended on physical activity helps to counteract the increases in expenditure due to increased body weight, and in some women this leads to saving of energy that largely meets the costs of pregnancy. However, although relatively small changes in activity patterns can potentially result in significant energy savings, there is little evidence that this occurs to a large extent. A possible reason for this is that affluent women are habitually so sedentary that there is little scope for further reduction. In contrast, in developing countries habitual levels of physical activity are high and there is therefore more potential for behavioral reductions. However, many women are likely to be unable to reduce their physical activity because of the constraints imposed by a subsistence livelihood, where farm work is obligatory for survival.

This topic has been one of considerable debate in recent years, particularly because longitudinal studies that have measured total energy expenditure with doubly labeled water have shown that many women increase the energy expended on physical activity during pregnancy, and that any decreases are not sufficient to counterbalance the energy costs of pregnancy due to tissue (fat) deposition and maintenance energy metabolism. It has been recommended that the data used by the WHO should be revised to take account of changes in energy expended on physical activity and to separate these energy costs from those of maintenance and tissue deposition. The Dietary Reference Intakes for the United States and Canada have already incorporated these changes (**Table 3**).

Between-Country Comparison of the Metabolic Costs of Pregnancy

The average costs across different populations result in a wide range of energy needs from −30 MJ (−7000 kcal) to 523 MJ (125 000 kcal). Studies found that the average costs in the well-nourished groups were similar to the current international assumption of 336 MJ (80 000 kcal). These studies have also shown that the amount of prepregnancy body fat is strongly correlated with both the maintenance costs and the total metabolic costs of pregnancy. The combined costs of maintenance, fat deposition, and conceptus across studies from different countries drawn from emerging and affluent nations show that the energy cost of fat deposition also varies according to the state of affluence and is positively correlated with variations in maintenance requirements.

This flexibility in energy metabolism acts in a protective manner, with undernourished women showing significant energy-sparing adaptive strategies that tend to normalize energy balance. Body fat content is one of the measures of fitness for reproduction; fertility is suppressed in undernourished women. However, future unfavorable conditions cannot be anticipated and pre- or early pregnant fatness may be indicative of overall nutritional status and energy balance during pregnancy.

These relationships suggested the existence of a mechanism that can monitor the mother's prepregnancy energy status and adjust the homeorrhetic changes in maternal metabolism accordingly. The discovery of leptin provides a plausible mechanism by which peripheral energy status can be centrally monitored and may coordinate the metabolic responses to pregnancy. It is clear that in addition to its role in the regulation of adipose tissue, appetite, and metabolic rate; leptin plays a significant role in several components of the reproductive axis. Evidence suggests that it plays a key role in pregnancy, including the modulation of fetal growth.

Individual Variability in the Total Energy Costs of Pregnancy

Because of the marked differences between individuals in the different components of the energy costs of pregnancy (changes in BMR, body fat, and energy expended on physical activity), the total energy costs, and therefore energy requirements, are also variable. Studies of well-nourished women indicate that the total extra energy costs of pregnancy average 418 MJ (100 000 kcal), considerably higher than the estimates in **Table 3**, and there is a large range from 34 to 1200 MJ (8000–287 000 kcal). These values are probably representative of many women in developed countries. They show that it is impossible to prescribe energy intakes for individual women because it cannot be predicted how they will respond metabolically (BMR and fat) or behaviorally (physical activity and food intake) to pregnancy.

Implications of Energy-Sparing Adaptations for Mother and Infant

Human energy metabolism is particularly adaptable during pregnancy, with early/prepregnancy body 'fatness' being a major determinant. The adaptive strategies that maintain energy balance seem to be a coordinated biological system in which energy-sensitive modulations in metabolism help to sustain human pregnancies and protect fetal growth in highly marginal environmental circumstances. However, the existence of such mechanisms should not be misinterpreted as suggesting that maintenance of optimal

nutritional status in pregnant women is not a priority because the adaptive mechanisms of the women will cope. It cannot be assumed that pregnant women will have energy-sparing alterations in metabolism and/or that physical activity decreases. Any adaptations that do occur should not be overinterpreted as suggesting that this is the case. The possible long-term detrimental effects must also be considered. The biochemical and physiological processes that are downregulated in the mother causing the suppression in BMR are unknown and there may be long-term consequences to her health and that of her infant.

The associations between maintenance needs, pregnancy weight gain, and prepregnant fatness indicate that a target weight gain of 12.5 kg is associated with maintenance costs of approximately 160 MJ (38 000 kcal). Although individual women or populations may have lower maintenance requirements, these may be associated with inadequate weight gain and low-birth-weight infants. A major determinant of birth weight is maternal weight gain, and the single most important determinant of infant survival is birth weight. Although birth weight is relatively well preserved at different planes of nutrition, weight alone is an inadequate measure of an infant's overall condition at birth. Even subtle nutritional influences on the fetal environment may have long-term consequences.

As mentioned previously, pregnancy weight gain is a critical component of the overall energy costs of pregnancy. The issue of whether pregnancy weight gain drives, or is driven by, the metabolic changes is interesting, but it is clear that women who consume marginal diets have small weight gains and that women from poorer countries have much lower percentage weight gains despite having lower initial body weights. Extremes of weight gains during pregnancy may have several consequences, which may or may not be mediated directly through an effect on birth weight. Other effects of weight gain may be more subtle and may be mediated through qualitative effects on fetal growth and development at different stages of intrauterine growth. There is a considerable body of evidence that suggests that many chronic adult diseases have their origins in fetal and infant nutrition, which has refocused attention on early life as a critical period in human development.

Further Reading

Barash, I.A., Cheung, C.C., Weigle, D.S., et al., 1996. Leptin is a metabolic signal to the reproductive system. Endocrinology 137, 3144–3147.

Butte, N., Hopkinson, J., Nicholson, M., 1997. Leptin in human reproduction: Serum leptin levels in pregnant and lactating women. Journal of Clinical Endocrinology and Metabolism 82, 585–589.

Butte, N., Treuth, M., Mehta, N., et al., 2003. Energy requirements of women of reproductive age. American Journal of Clinical Nutrition 77, 630–638.

Department of Health (1991) Dietary reference values for Food energy and nutrients for the United Kingdom. *Report on Health and Social Subjects No. 41*. London: HMSO.

FAO/WHO/UNU (1985) Report of a joint expert consultation: Energy and protein requirements. *Technical Report Series 724*. Geneva: WHO.

Forsum, E., 2004. Energy requirements during pregnancy: Old questions and new findings. American Journal of Clinical Nutrition 79, 933–934.

Goldberg, G.R., 1997. Reproduction: A global nutritional challenge. Proceedings of the Nutrition Society 56, 319–333.

Goldberg, G.R., Prentice, A.M., Coward, W.A., et al., 1993. Longitudinal assessment of energy expenditure in pregnancy by the doubly-labelled water method. American Journal of Clinical Nutrition 57, 494–505.

Harigaya, A., Nagashima, K., Nako, Y., et al., 1997. Relationship between concentration of serum leptin and fetal growth. Journal of Clinical Endocrinology and Metabolism 82, 3281–3284.

Hytten, F.E., 1991. Nutrition; Weight gain in pregnancy. In: Hytten, F., Chamberlain, G. (Eds.), Clinical Physiology in Obstetrics, 2nd edn. Blackwell Scientific, Oxford.

Kopp-Hoolihan, L.E., Loan, M.V., Wong, W.W., et al., 1999. Longitudinal assessment of energy balance in well nourished, pregnant women. American Journal of Clinical Nutrition 69, 697–704.

National Academy of Sciences, 2002. Dietary Reference Intakes for Energy, Carbohydrates, Fiber, Fat, Protein and Amino Acids (Macronutrients). National Academy Press, Washington, DC.

Poppitt, S.D., Prentice, A.M., Goldberg, G.R., et al., 1994. Energy-sparing strategies to protect human fetal growth. American Journal of Obstetrics and Gynecology 171, 118–125.

Poppitt, S.D., Prentice, A.M., Jequier, E., et al., 1993. Evidence of energy-sparing in Gambian women during pregnancy: A longitudinal study using whole-body calorimetry. American Journal of Clinical Nutrition 57, 353–364.

Prentice, A.M., Goldberg, G.R., 2000. Energy adaptations in human pregnancy: Limits and long-term consequences. American Journal of Clinical Nutrition 71 (supplement), 1226S–1232S.

Prentice, A.M., Spaaij, C.J.K., Goldberg, G.R., et al., 1996. Energy requirements of pregnant and lactating women. European Journal of Clinical Nutrition 50 (supplement 1), S82–S111.

Scrimshaw, N.S., Waterlow, J.C., Schurch, B., 1996. Energy and protein requirements. Proceedings of an IDECG Workshop, European Journal of Clinical Nutrition 50 (supplement 1), S1–S197.

Pregnancy: Nutrient requirements

Melissa F. Young and Usha Ramakrishnan, Hubert Department of Global Health, Emory University, Atlanta, GA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of L.H. Allen, Pregnancy: Nutrient Requirements, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 61–67, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00229-4.

Introduction 168
Recommended nutrient intakes for pregnancy 168
Energy 169
Protein 170
Folate 170
Other B vitamins 171
Vitamin A 171
Vitamin D 171
Vitamin C (ascorbic acid) 172
Vitamin E 172
Calcium 172
Phosphorus 172
Magnesium 173
Iron 173
Zinc 173
Iodine 174
Trace elements: copper, selenium, chromium, fluoride, manganese, and molybdenum 174
Water and electrolytes 174
Summary 174
References 175
Further reading 175

Key points

- To review recommended macronutrient and micronutrient intakes during pregnancy
- To identify research gaps and opportunities.

Introduction

Maternal nutrition during pregnancy is a critical public health priority across the globe. The first 1000 days, starting at conception, marks a time period of enormous potential but also vulnerability. Inability to meet nutrient requirements during pregnancy can have serious short- and long-term effects on maternal and child health and development (Victora et al., 2021). Indeed, it is now understood that epigenetic effects of nutrient imbalances during pregnancy can affect the health of the offspring for the remainder of their life. Most of the research that provides information on nutrient requirements during pregnancy has been conducted in high-income countries, although trials in low income countries have been important in revealing the adverse effects of poor maternal nutrition and the benefits of nutrient interventions. There is an unacceptably high rate of pregnancy complications many of which may be prevented by improved maternal nutrition, including anemia, low birth weight, birth defects, and preeclampsia.

Recommended nutrient intakes for pregnancy

The most recent and best-described recommended intakes of nutrients during pregnancy are those of the Institute of Medicine, developed for the United States and Canada, which are the main set presented in this article (**Table 1**). Many other countries have their own sets of recommendations, as do organizations such as the Food and Agriculture Organization (FAO)/World Health Organization (WHO) and the European Economic Community. However, many of these, including FAO/WHO, do not provide EARs.

Table 1 Recommended Dietary Allowances (RDAs) for nonpregnant women and Estimated Average Requirements (EARs), RDAs, and upper limits of nutrients for pregnant women.

	AI/RDA,[a] *adult nonpregnant woman*	*EAR, pregnancy*	*AI/RDA,*[a] *pregnancy*	*Upper limit, pregnancy*
Energy (kcal)	2000–2200[b]	+340 (trimester 2) +452 (trimester 3)	–	–
Energy (MJ)	8.37–9.21[b]	+1.42 (trimester 2) +1.89 (trimester 3)	–	–
Protein (g kg^{-1})	0.8	0.88	+1.1	None
Vitamin A (μg RAE, retinol activity equivalents)	700	550	770	3000
Vitamin D	600 IU (15 μg)	400 IU (10 μg)	600 IU (15 μg)	4000 IU (100 μg)
Vitamin E (mg α-tocopherol)	15	12	15	1000
Vitamin K (μg)	90	–	90	None
Vitamin C (mg)	75	70	85	2000
Folate (μg dietary folate equivalents)	400	520	600	1000 from fortified food + supplements
Thiamin (mg)	1.1	1.2	1.4	None
Riboflavin (mg)	1.1	1.2	1.4	None
Vitamin B6 (mg)	1.3	1.6	1.9	100 as pyridoxine
Niacin (mg NE)	14	14	18	35
Vitamin B_{12} (μg)	2.4	2.2	2.6	None
Pantothenic acid (mg)	5	–	6	None
Biotin (μg)	30	–	30	None
Choline (mg)	425	–	450	3500
Calcium (mg)	1000	800	1000	2500
Phosphorus (mg)	700	580	700	3500
Magnesium (mg)	320	290	350	+350 from supplement
Iron (mg)	18	22	27	45
Zinc (mg)	8	9.5	11	40
Iodine (μg)	150	160	220	1100
Copper (μg)	900	800	1000	10,000
Selenium (μg)	55	49	60	400
Chromium (μg)	25	–	30	None
Fluoride (mg)	3	–	3	10
Manganese (mg)	1.8	–	2	11
Molybdenum (μg)	34	40	50	2000

[a]Values are Recommended Dietary Intakes (RDAs) except for pantothenic acid, biotin, and choline, where value is an Adequate Intake.

[b]Assuming moderately active woman. Actual requirements vary by weight and height. Requirements increase throughout pregnancy and the higher end of the range is recommended during the third trimester.

Adapted from Institute of Medicine, National Academies Press, for the United States and Canada (http://www.nap.edu).

The set of Dietary Reference Intake recommendations developed by the Institute of Medicine includes several values. The EAR is the intake required to meet the nutrient needs of 50% of a population group (e.g., pregnant women). It is an important value for two reasons. First, it is the value used to estimate the prevalence of inadequate intakes of a nutrient in a population group; the percentage of a group consuming less than the EAR of a nutrient is the percentage with an inadequate intake. For energy, the Estimated Energy Requirement (EER) is equivalent to the EAR because adding a margin of safety would lead to overweight. Second, the RDA is calculated by adding two standard deviations (assumed 20% of the EAR when unknown) to the EAR. The RDA should meet the requirements of 97.5% of a population group. The Tolerable Upper Level (UL) for a nutrient is the intake above which there is a risk of adverse effects. **Table 1** shows the RDAs for nonpregnant women for comparison, and the EARs, RDAs, and ULs for pregnant women.

Energy

Maternal energy requirements increase during pregnancy due to higher basal energy expenditure as well as energy deposition in maternal and fetal tissues. Basal metabolism of the mother is higher due to the increased work by the lungs and heart and because

of the metabolism of the fetus and uterus. A longitudinal study by Butte et al. found that basal metabolic rate increased by 10.7 ± 5.4 kcal per week of gestation, mostly in the second and third trimesters. On average, the fetus requires approximately 68 kcal day^{-1}. The substantial variability in basal energy expenditure among individual women is caused mainly by differences in fat-free mass (including maternal skeletal muscle mass and fetal tissue). The cumulative increase in basal energy expenditure during pregnancy is positively correlated with maternal fatness and weight gain. Energy requirements for the thermic effect of feeding are not different from those of nonpregnant women, nor is there much change in the total energy cost of activity. Although the increasing body weight of the mother means that the energy cost of each activity is higher, the net effect is canceled out by the fact that after approximately 25 weeks of gestation women tend to become less active. The longitudinal study by Butte et al. suggests that energy expenditure in physical activity decreases by approximately 100–200 kcal day^{-1} in women with a low or normal body mass index before pregnancy and by an average of more than 400 kcal day^{-1} in those with a high body mass index (>26 kg m^{-2}).

In deriving the recommendations for the United States and Canada, the EER during pregnancy is accepted to be the sum of the Total Energy Expenditure (TEE) of the nonpregnant woman, measured using a doubly labeled water technique, plus an estimated median change in TEE of 8 kcal per week, plus 180 kcal day^{-1} to cover energy deposited in maternal and fetal tissues. In the first trimester of pregnancy, TEE changes little and weight gain is small, so the energy requirement is increased only during the second (340 kcal/d) and third trimesters (452 kcal/d). There is no RDA or UL because energy intakes greater than the EER may lead to undesirable weight gain.

The EER for pregnancy is as follows:

Trimester 1: nonpregnant EER + 0 kcal
Trimester 2: nonpregnant EER + 160 kcal (based on 8 kcal per week × 20 weeks) + 180 kcal
Trimester 3: nonpregnant EER + 272 kcal (based on 8 kcal per week × 34 weeks) + 180 kcal

These formulas represent average requirements in trimesters 1 and 2. If a more precise estimate of requirements is needed at a specific stage of gestation, instead of the mean increment of 160 kcal in trimester 1 and 272 kcal in trimester 2, the actual weeks of gestation can be multiplied by 8 kcal per week.

Recently, there have been calls to provide additional guidance for women based on their prepregnancy nutritional status (National Academies of Sciences Engineering and Medicine, 2020; Most et al., 2019). Prepregnancy body size may influence energy storage requirements. Most et al. suggest women with obesity do not need to increase energy intake and may even need slight reduction in intake to meet weight gain recommendations. Further research is needed to guide clinical guidance (Most et al., 2019).

Protein

The turnover of body protein is higher after approximately 13 weeks of pregnancy, and the mother adjusts by losing less nitrogen as urea even during the first trimester. A woman who gains 12.5 kg of body weight has deposited 925 g of protein that comprises of the following components: the fetus 440 g, the uterus 166 g, expanded maternal blood volume 81 g, the placenta 100 g, and the increment in extracellular fluid, 135 g, and perhaps some additional protein in muscle. The EAR for all age groups is 0.88 g^{-1} kg^{-1} day^{-1} protein or 21 g of additional protein/day. The RDA is 1.1 g protein per kg per day or 25 g day^{-1}.

One-third of the total protein deposition during the 40 weeks of pregnancy occurs in the second trimester and two-thirds in the third trimester. By the end of the third trimester, the US–Canada recommendations assume that an additional consumption of 17 g protein per day is required to meet the needs for protein deposition, and as about half of this occurs during the second trimester this amounts to 8 g day^{-1}. It is also assumed that no additional protein is needed in trimester 1, but for the last two trimesters consumption of an additional 21 g day^{-1} (a total of 1.1 g kg^{-1} day^{-1}) is recommended.

New estimates of protein needs have recently been discussed using the new amino acid oxidation methodology (National Academies of Sciences Engineering and Medicine, 2020). Stephens et al. provides evidence to suggest increasing EAR for protein in early (1.22 g/kg/day) and late pregnancy (1.52 g/kg/day) (Stephens et al., 2015). Further research is required to confirm recommendation and to determine amino acid requirements during pregnancy (Elango and Ball, 2016). It should be noted that maternal balanced energy protein supplements are one of the recommended interventions for undernourished women during pregnancy. As highlighted in the 2021 Lancet Nutrition Series, balanced energy protein supplementation during pregnancy was associated with a 40% reduction in LBW, 61% reduction in stillbirths and a 29% reduction in SGA (Keats, 2021; Lassi, 2020) in these populations.

No UL has been set for protein, including for pregnancy, in the US–Canada recommendations due to lack of data on harmful effects. However, some earlier studies noted adverse pregnancy outcomes when high-protein supplements were given to relatively well-nourished pregnant women, so caution in this regard is certainly warranted.

Folate

Maternal folate requirements increase markedly during pregnancy due to utilization of the vitamin in cell division in the mother and fetus, single-carbon transfer reactions, and deposition in the fetus. The risk of women giving birth to an infant with a neural tube defect (NTD) is significantly reduced if women consume folic acid supplements before conception through approximately the first 4–6 weeks of pregnancy—during the time of neural tube closure. Some women are at greater risk of producing an infant with this birth defect, especially when their folate intake is rather low. Because such women are unaware of this risk unless they have had

a previous NTD delivery, the recommendation is that all women who are capable of becoming pregnant consume at least 400 μg of folic acid daily from supplements, fortified food, or both in addition to consuming food folate from a varied diet.

In pregnancy, the recommendation is for all women to consume an additional 200 μg of synthetic folic acid (equivalent to 400 μg of dietary folate due to the higher bioavailability of the synthetic form) in addition to the RDA for the nonpregnant woman of 400 μg day^{-1}. This amount was shown to prevent plasma homocysteine from becoming elevated during pregnancy and to maintain normal folate concentration in red blood cells. The UL of 1000 μg day^{-1}, the same as for nonpregnant women, is set to avoid potential exacerbation of vitamin B_{12} deficiency.

In addition to its importance for lowering risk of NTDs in the periconceptional period, there is evidence that adequate folate status, which is important for maintaining normal plasma homocysteine concentrations, lowers the risk of other delivery problems and birth defects, including preeclampsia, preterm delivery, very low birth weight, club foot, and placental abruption. Currently 63 countries across the globe that include the United States and Canada have mandatory wheat flour fortification with folic acid (Global Fortification Data Exchange, www.FortificationData.org). Despite strong evidence, it is estimated that only 23% of Folic Acid-Preventable Spina Bifida and Anencephaly is being prevented through folic acid programs (Kancherla et al., 2021). Further advocacy and prioritization is required for scaling up fortification efforts to vulnerable populations.

Other B vitamins

Several B vitamin deficiencies cause homocysteinemia, notably folic acid, vitamin B_{12}, riboflavin, and vitamin B_6. Importantly, homocysteinemia is associated with adverse pregnancy outcomes. In a large retrospective study in Norway, for example, women in the highest 25% of plasma homocysteine concentrations had significantly more placental abruption, stillbirths, very low birth weight and preterm infants, preeclampsia, club foot, and NTDs in their offspring compared to women with values in the lowest 25%. Supplementation with folic acid up to 500–600 μg day^{-1} lowers plasma homocysteine, but few studies have been done on the other B vitamins. Of these, it is most difficult for poor women to obtain their dietary vitamin B_{12} requirement because this vitamin in found only in animal source foods, such as meat and dairy products.

The recommended intakes of most B vitamins and choline are increased above nonpregnant values as shown in **Table 1**. The increases are based on evidence for higher maternal requirements (in the case of thiamin, riboflavin, niacin, and vitamin B_6) and for fetal and placental deposition of the vitamin (thiamin, riboflavin, niacin, vitamin B_6, vitamin B_{12}, and choline). There is growing research on the importance of choline during pregnancy for birth outcomes and child cognitive function (Caudill et al., 2018; National Academies of Sciences Engineering and Medicine, 2020). Further research on potentially raising the AI for choline may be merited.

UL values, the same as for nonpregnant women, have been set for niacin when consumed as nicotinic acid in supplements based on a "flushing" reaction and for choline based on cholinergic reactions and a fishy body odor.

Vitamin A

The increment in vitamin A requirements during pregnancy is based on the relatively small amount of the vitamin that is found in fetal liver at birth. The liver content is assumed to be 36 μg, mostly accumulated during the last 3 months of gestation. Using an estimated 70% absorption of the vitamin from the maternal diet, the EAR is 50 μg RAE above the 500 μg RAE requirement for the nonpregnant woman, and the 770 μg RAE RDA is 70 μg higher.

In wealthier regions of the world, vitamin A deficiency during pregnancy is rare. Rather, there is more concern about the potentially adverse effects of consuming excessive amounts of the vitamin. Based on the potential for retinol excess to cause birth defects (malformations), especially if high doses are consumed early in pregnancy, a UL of 3000 μg day^{-1} is set for all women who may become pregnant as well as those who are pregnant. This intake is unlikely to be achieved with natural food, although it would be possible if large amounts of liver, foods fortified with the vitamin, or supplements were consumed. The World Health Organization (WHO) does not recommend routine supplementation of vitamin A in the prenatal period. However, vitamin A supplementation during pregnancy is recommended in areas where vitamin A deficiency is a severe public health problem ($\geq$5% of women in a population have a history of night blindness, or if $\geq$20% of pregnant women have a serum retinol level $<$0.70 μmol/L) to prevent night blindness (WHO, 2016).

Vitamin D

Vitamin D is important for maintaining normal blood levels of calcium and phosphate for general cell functioning and bone health. Maternal transfer of vitamin D to fetus during pregnancy is essential for acquired infant stores at birth. Those women who obtain adequate exposure to ultraviolet light do not need higher amounts during pregnancy. However, if usual intake declines below $\approx$150 IU (3.8 μg) per day at high latitudes (where there is little ultraviolet radiation in the winter, such as in France), evidence of low maternal 25(OH)D and infant vitamin D depletion has been observed at delivery. Among women with more highly pigmented skin, such as African Americans, or those who are veiled, the prevalence of vitamin D deficiency near term (circulating 25(OH)D $<$ 25–50 nmol L^{-1}) is 30–70%, compared to 5–20% in lighter skinned populations.

The effect of maternal depletion in pregnancy on fetal bone mineralization remains controversial; the Institute of Medicine committee concluded that maternal vitamin D status does not have a major influence on calcium transfer to the fetus. Poor

maternal vitamin D status in pregnancy has been implicated as a risk factor in a number of diseases of childhood and adult life, in part because risk of the diseases varies with season. These include multiple sclerosis, schizophrenia, type I diabetes and some cancers. However, a definitive link with maternal vitamin D status has not been proven.

The IOM recommendations for vitamin D were last updated in 2011 (IOM, 2011). For both adolescent and adult women is to continue to consume the amount recommended for nonpregnant women, 15 μg (600 IU day^{-1}). The UL of 100 μg (4000 IU day^{-1}) is the same as before pregnancy, and is based on the fact that this level of intake will not cause serum 25(OH)vitamin D to exceed 125–150 nmol L^{-1}, which is at the high end of serum concentrations associated with the lowest risk of adverse events such as all-cause mortality.

The WHO does not recommend routine vitamin D supplementation during pregnancy (WHO, 2020, 2016; Palacios et al., 2019). Current guidelines recommend women meet requirement though a healthy balanced diet and advise that sunlight is the most important source of vitamin D. For pregnancy women with suspected vitamin D deficiency, including populations with limited direct sun exposure, WHO recommends providing 200 IU (5 μg) per day, which is the WHO and FAO recommended nutrient take (WHO and FAO of the UN, 2004). Note, this is lower than the current IOM recommendations. Vitamin D requirements and supplementation remains an active area of research with critical gaps in our current knowledge.

Vitamin C (ascorbic acid)

The EAR for nonpregnant women is based on the intake that attains the maximum neutrophil concentration of ascorbic acid. Maternal plasma vitamin C concentrations decline during pregnancy, probably as a result of normal hemodilution. Oxidized ascorbic acid is transferred from the maternal circulation to the fetus, where it is retained in the reduced form. Although vitamin C deficiency in pregnancy is rare in most situations, it has been associated with increased risk of premature rupture of the membranes and infections, preterm birth, and eclampsia. However, clinical trials have not shown a benefit of higher doses of vitamin C during pregnancy. Smokers have lower levels of ascorbic acid in their serum and amniotic fluid. Women who smoke more than 20 cigarettes per day and regular aspirin users may require twice as much, as may heavy users of alcohol and street drugs. The UL of 2000 mg day^{-1} is based on prevention of diarrhea and gastrointestinal disturbances that occur with high intakes.

Vitamin E

There is no increase in the recommended intake of vitamin E during pregnancy, so the RDA remains at 15 mg of α-tocopherol per day for all ages. There have been no reports of deficiency of vitamin E during pregnancy nor any evidence of benefit from maternal supplementation. The UL is 1000 μg day^{-1} of any form of the vitamin taken as a supplement, extrapolated from data showing that high levels cause hemorrhaging in rats.

Calcium

It is clear that changes in maternal calciotropic hormones and calcium metabolism (i.e., increased intestinal absorption and reduced urinary calcium excretion) enable the fetus to be supplied with adequate amounts of this mineral, and that little change in maternal intake is needed. Nor is there a correlation between the number of pregnancies a woman has and her risk of bone fracture. Thus, for the United States and Canada there is no increase in recommended calcium intakes for pregnancy and the RDA remains at 1300 mg day^{-1} for women aged 14–18 years and 1000 mg day^{-1} for the 19–50-year-old group (IOM, 2011).

The UL for calcium in pregnancy is the same as that for the nonpregnant woman, 2500 mg day^{-1}. This safe level is set based on the risk of kidney stones.

The WHO however recommends daily calcium supplementation (1.5–2.0 g) for pregnant women in settings with low dietary calcium intake based on evidence of benefit for improved birth outcomes (WHO, 2016). Recent systematic reviews of controlled trials confirm that calcium supplementation reduces risk of pre-eclampsia by 55% (RR 0.45, 0.19–1.06) (Keats, 2021). However, this intervention is not widespread and there remain gaps on the ideal formulations and implementation strategies to enhance program effectiveness.

Phosphorus

The efficiency of phosphorus absorption increases by 15% during pregnancy. The term infant contains approximately 17 g of phosphorus at birth, mostly in bone and water. The physiological adaptations of the mother that increase calcium retention also help to supply the fetus with more phosphorus. There is no evidence that the EAR needs to increase over that recommended for nonpregnant women, so the RDA for women aged 14–18 years is 1250 mg day^{-1} and for those aged 19–50 years it is 700 mg day^{-1}. Based on the need to avoid high serum phosphorus concentrations, and the fact that phosphorus absorption is more efficient in pregnancy, the UL is set at 3500 mg day^{-1}, slightly lower than the 4000 mg day^{-1} for nonpregnant women.

Magnesium

It is assumed that the gain in fat-free mass in pregnancy (7.5 kg) is associated with a greater deposition of magnesium. If this tissue contains 470 mg kg^{-1}, after adjustment for a bioavailability of 40%, the EAR is an increase of 35 mg day^{-1} for pregnant women of all ages, and the RDA is 10% higher than this; for women aged 14–18 years, the EAR and RDA, respectively, are 335 and 400 mg; for those aged 19–30 years, these values are 290 and 350 mg; and for those 31–50 years, they are 300 and 360 mg.

The UL for magnesium in pregnancy is set at 350 mg day^{-1} taken as a supplement, based on the potential for higher doses of magnesium salts to cause an osmotic diarrhea.

Iron

Incremental iron requirements for the mother and fetus are relatively well established, although how these requirements should be met is more controversial. It is generally accepted that the mother needs to absorb an additional 6 mg day^{-1} to supply the amount retained by the fetus (300 mg) and placenta (60 mg) and that used to synthesize additional maternal erythrocytes (450 mg) and replace blood loss during delivery (200 mg). Some iron is saved by the lack of menstruation in pregnancy. The fetus obtains iron from the placenta in a process that involves iron transfer from maternal transferrin to transferrin receptors on the placenta, endocytosis of holotransferrin, and release of iron into the fetal circulation. Maternal iron absorption and transfer to the fetus increases during the second and third trimesters. This process is upregulated if the mother is iron deficient; however, maternal iron deficiency does reduce the amount of fetal iron stored at birth and available to the fetus during the first months of life.

The EAR for pregnancy is set at 23 mg day^{-1} for adolescents and 22 mg day^{-1} for adult women, and the RDA is 27 mg day^{-1} for both groups. Although the requirement is mainly increased in the last trimester, it is important to build iron stores early and to avoid high doses later, so the higher intake recommendation is distributed throughout pregnancy. The UL is the same as that for the nonpregnant woman and is based on the need to avoid gastrointestinal distress.

Anemia is an important global health problem affecting 38% of all pregnant women and low maternal Hb (<110 g/L) has been associated with a range of poor birth outcomes (low birth weight, preterm birth, small-for-gestational-age (SGA), stillbirth, and perinatal and neonatal mortality) and adverse maternal outcomes (postpartum hemorrhage, preeclampsia, and blood transfusion) (Young et al., 2019; Stevens et al., 2013). In most countries, iron supplements are recommended routinely for all pregnant women as a strategy to reduce anemia during pregnancy, but only approximately 50% of anemia among pregnant women is responsive to iron supplementation (Stevens et al., 2013), and this can be highly variable across different settings. The amount of iron in prenatal supplements was recently reduced from 60 mg to 30 mg of elemental iron along with 400 g (0.4 mg) of folic acid based on evidence to prevent maternal anemia, puerperal sepsis, low birth weight, and preterm birth (WHO, 2016). The higher dose of 60 mg iron is however still recommended in settings with a high prevalence of anemia (>40%). The WHO also recommends intermittent oral iron and folic acid supplementation with 120 mg of elemental iron and 2800 g (2.8 mg) of folic acid once weekly for pregnant women if daily iron is not acceptable due to side-effects, and in populations where less than 20% of women are anemic. The WHO advises context-specific examination of the etiology of anemia be conducted in order tailor anemia reduction strategies to different settings. WHO hemoglobin cut-offs for defining anemia are currently being re-examined and remains an active area of research (Garcia-Casal et al., 2019; Young et al., 2019; Ohuma et al., 2020).

Zinc

The estimated additional zinc required for pregnancy is approximately 100 mg, equivalent to 5–7% of the mother's body zinc, part of which is obtained through more efficient intestinal zinc absorption. Approximately half of this is deposited in the fetus. The EAR for pregnant women is based on an additional requirement of 2.7 mg day^{-1} during the last 10 weeks of gestation. The UL is based on evidence of impaired copper status at high intakes, as for nonpregnant women.

Zinc plays critical roles in cell division, hormone metabolism, protein and carbohydrate metabolism, and immunocompetence. Because zinc deficiency in pregnant animals causes birth defects and fetal growth retardation, there has been considerable effort to determine the effects of human zinc status on pregnancy outcome, especially in developing countries, where zinc intakes are often inadequate.

In general, however, meeting recommended zinc intakes is more difficult but more critical for women whose diets are low in animal source foods and higher in fiber. High intakes (supplements) of iron and calcium may also impair zinc absorption and therefore increase requirements.

While the WHO does not recommend zinc supplementation as part of routine ANC, it is recommended in the context of rigorous research (WHO, 2021). Although a previous review of randomized controlled trials found a reduced risk of preterm birth (Ramakrishnan et al., 2014) current WHO recommendations were informed by a recent systematic review including over 18,000 women from 25 randomized controlled trials (Carducci et al., 2021) that concluded that there was inadequate evidence on zinc supplementation on maternal and newborn health outcomes. In the review, there was low-certainty evidence that zinc supplementation in pregnancy had little to no difference in reducing the risk of preterm births, stillbirths, or mortality. In addition, there remain critical gaps in our understanding of interactions of zinc supplementation with other iron and calcium supplements during pregnancy.

Iodine

In the many countries with endemic iodine deficiency, which include parts of the United States, Canada, and substantial areas of Europe and many other industrialized and developing countries, there is clear potential for the harmful effects of this deficiency to emerge during pregnancy. The most damaging effect of iodine deficiency is on the brain of the fetus because iodine is required for thyroid hormone, which in turn affects myelination and function of the developing central nervous system. The clinical expression of severe maternal iodine deficiency during pregnancy is cretinism, including severe mental retardation, deaf mutism, short stature, and spasticity. Injections of iodized oil before midpregnancy have markedly reduced cretinism and neonatal mortality in areas of severe iodine deficiency. Universal salt iodization is the recommended strategy for addressing iodine deficiency. In most countries, Universal Salt Iodization has reduced the prevalence of cretinism substantially. However, the effectiveness of salt iodization is dependent on country-level regulations (mandatory vs. voluntary), level and choice of fortificant, and coverage. WHO and UNICEF recommend iodine supplementation for pregnant women in countries where less than 20% of households have access to iodized salt (WHO and UNICEF, 2007). In settings where 20–90% of households have access to iodized salt WHO and UNICEF recommend efforts to accelerate salt iodization or providing an iodine supplement or fortified foods for vulnerable populations. In the United States, there is a worrisome trend of declining status and pregnant women have insufficient intake according to WHO criteria (Perrine et al., 2019).

The EAR for pregnancy is set at 160 μg day^{-1} and the RDA at 220 μg day^{-1} for the United States and Canada based on the amount needed to prevent increased thyroid size in previously deficient women. The UL is 1100 μg day^{-1}, the same as for nonpregnant, nonlactating women, and it is based on the need to avoid elevated thyroid-stimulating hormone concentrations. In 2008 the WHO increased its recommendation by 50 μg day^{-1}–250 μg day^{-1}, and established 500 μg day^{-1} as the intake above which no additional health benefit could be expected.

Trace elements: copper, selenium, chromium, fluoride, manganese, and molybdenum

Copper is required for the function of many enzymes, primarily oxidases. In pregnancy, an increased intake of this mineral is recommended to cover deposition of approximately 18 mg day^{-1}, most of which is in fetal liver. The UL (10,000 μg day^{-1}) is the same as for nonpregnant women, based on the need to prevent the liver damage that occurs with high intakes.

Recommended intakes of selenium for adults are based on the criterion of maximizing plasma glutathione peroxidase activity. Based on an estimated selenium content of the fetus of 1000 μg, across pregnancy this would require that an additional 4 μg day^{-1} be consumed. The EAR is therefore increased from 45 to 49 μg day^{-1} and the RDA from 55 to 60 μg day^{-1}. The UL is determined on the basis of hair loss and brittle nails, which occur at higher levels of intake, and is the same as that set for nonpregnant women.

Chromium is required for normal insulin metabolism. There are no data from which to derive a recommendation for pregnancy, so an increase of 5 μg day^{-1} is recommended (as an AI) based on the additional weight and tissue chromium gained in pregnancy. No UL was set due to lack of documented adverse effects in humans.

For fluoride, there is no evidence that increasing the AI in pregnancy above that for the nonpregnant woman will benefit fetal tooth or bone content or afford protection against later tooth decay in the child. The UL is set at 10 mg day^{-1} to avoid fluorosis (discoloration of tooth enamel, joint pain, and skeletal abnormalities).

Manganese is required for bone formation and the normal metabolism of amino acids, lipids, and carbohydrates. The AI for pregnancy, estimated from the manganese content of maternal weight gain, is 2 mg day^{-1}. The UL is based on avoidance of elevated blood manganese and neurotoxicity, and it is not increased for pregnancy.

Recommended molybdenum intakes, based on the mineral's role as a cofactor for several enzymes, increase by 16 mg day^{-1} in pregnancy to cover the increment in fetal and maternal weight. The UL is derived from adverse reproductive effects seen in animals.

Water and electrolytes

The US–Canada recommended intake of water for pregnant women is based on median intake from a large national survey in the United States. The AI of 3 L day^{-1} is anticipated to come from foods (0.7 L) and beverages (2.3 L). No UL was set because individuals stop drinking once their intake is adequate.

The AI for sodium in pregnancy is 1500 mg day^{-1} based on an intake level to cover daily losses, provide adequate intakes of other nutrients, and maintain normal function. The UL of 2300 mg day^{-1} is based on the adverse effects of higher intakes on blood pressure in susceptible members of the population.

The AI for potassium in pregnancy (4.7 g day^{-1}) is set at a level that will lower blood pressure, reduce the extent of salt sensitivity, and minimize the risk of kidney stones. There is no evidence that adverse effects of potassium are seen with high intakes from food and no UL was set, but potassium supplements can cause high blood potassium in some chronic diseases, such as renal disease and type 1 diabetes.

Summary

In the US–Canada recommendations, the recommended intakes are increased for most, but not all, nutrients during pregnancy. However, the recommendations are often based on less than ideal experimental data, in part due to the difficulty of conducting experiments on pregnant women.

For most nutrients, it is likely that some population groups may have higher requirements than those recommended in **Table 1**, notably women bearing more than one fetus or adolescents (see the Institute of Medicine volumes for specific recommendations for this age group). To meet the recommended nutrient increases, dietary quality often needs to be improved during pregnancy. It is often advised that pregnant women should also take iron supplements and/or a multiple vitamin–mineral supplement. The specific benefits of supplementation in pregnancy, optimal timing, and optimal doses are still somewhat controversial and the subject of ongoing research. Currently, some countries recommend routine supplementation for all pregnant women, whereas others recommend supplementation only when there is evidence of anemia, other nutritional deficiencies, a poor diet, or other problems such as drug or alcohol abuse.

See Also: Calcium; Choline and phosphatidylcholine; Chromium; Copper; Folate/folic acid; Iodine: Physiology, dietary sources, and requirements; Phosphorus; Potassium; Protein: Requirements and role in diet; Sodium: Physiology and dietary sources; Vitamin A: Physiology, dietary sources and requirements; Vitamin B_{12}: Physiology, dietary sources, and requirements; Vitamin E | metabolism and requirements

References

Carducci, B., Keats, E.C., Bhutta, Z.A., 2021. Zinc supplementation for improving pregnancy and infant outcome. Cochrane Database Syst. Rev. 3, CD000230.

Caudill, M.A., Strupp, B.J., Muscalu, L., Nevins, J.E.H., Canfield, R.L., 2018. Maternal choline supplementation during the third trimester of pregnancy improves infant information processing speed: a randomized, double-blind, controlled feeding study. Faseb. J. 32, 2172–2180.

Elango, R., Ball, R.O., 2016. Protein and amino acid requirements during pregnancy. Adv. Nutr. 7 (4), 839S–844S. https://doi.org/10.3945/an.115.011817.

Garcia-Casal, M.N., Pasricha, S.R., Sharma, A.J., Peña-Rosas, J.P., 2019. Use and interpretation of hemoglobin concentrations for assessing anemia status in individuals and populations: results from a WHO technical meeting. Ann. N. Y. Acad. Sci. 1450, 5–14.

Global Fortification Data Exchange, n.d., viewed, 4 October 2021, https://fortificationdata.org.

Institute of Medicine (IOM), 2011. Dietary reference intakes for calcium and vitamin D. The National Academies Press, Washington, DC.

Keats, E.C., Das, J.K., Salam, R.A., Lassi, Z.S., Imdad, A., Black, R.E., Bhutta, Z.A., 2021. Effective interventions to address maternal and child malnutrition: an update of the evidence. Lancet Child Adolesc. Health 5 (5), 367–384. https://doi.org/10.1016/S2352-4642(20)30274-1.

Kancherla, V., Wagh, K., Pachón, H., Oakley Jr., G.P., 2021. A 2019 global update on folic acid-preventable spina bifida and anencephaly. Birth Defects Res. 113, 77–89.

Lassi, Z.S., Padhani, Z.A., Rabbani, A., et al., 2020. Impact of dietary interventions during pregnancy on maternal, neonatal and child outcomes in lowand middle-income countries. Nutrients 12, 531.

Most, J., Dervis, S., Haman, F., Adamo, K.B., Redman, L.M., 2019. Energy intake requirements in pregnancy. Nutrients 11, 1812.

National Academies of Sciences, Engineering, and Medicine, 2020. Nutrition during pregnancy and lactation: exploring new evidence. In: Proceedings of a Workshop. The National Academies Press, Washington, DC. https://doi.org/10.17226/25841. Accessed September 2021.

Ohuma, E.O., Young, M.F., Martorell, R., et al., 2020. International values for haemoglobin distributions in healthy pregnant women. EClinicalMedicine 29, 100660.

Palacios, C., Kostiuk, L.K., Peña-Rosas, J.P., 2019. Vitamin D supplementation for women during pregnancy. Cochrane Database Syst. Rev. 7, CD008873.

Perrine, C.G., Herrick, K.A., Gupta, P.M., Caldwell, K.L., 2019. Iodine status of pregnant women and women of reproductive age in the United States. Thyroid 29, 153–154.

Ramakrishnan, U., Imhoff-Kunsch, B., Martorell, R., 2014. Maternal nutrition interventions to improve maternal, newborn, and child health outcomes. Nestle Nutr. Inst. Workshop Ser. 78, 71–80.

Stephens, T.V., Payne, M., Ball, R.O., Pencharz, P.B., Elango, R., 2015. Protein requirements of healthy pregnant women during early and late gestation are higher than current recommendations. J. Nutr. 145, 73–78.

Stevens, G.A., Finucane, M.M., De-Regil, L.M., et al., 2013. Global, regional, and national trends in haemoglobin concentration and prevalence of total and severe anaemia in children and pregnant and non-pregnant women for 1995–2011: a systematic analysis of population-representative data. Lancet Glob. Health 1, e16–25.

Victora, C.G., Christian, P., Vidaletti, L.P., et al., 2021. Revisiting maternal and child undernutrition in low-income and middle-income countries: variable progress towards an unfinished agenda. Lancet 397, 1388–1399.

WHO, UNICEF, 2007. Reaching Optimal Iodine Nutrition in Pregnant and Lactating Women and Young Children: A Joint Statement by WHO and UNICEF. https://www.who.int/docs/default-source/nutritionlibrary/reaching-optimal-iodine-nutrition-in-pregnant-and-lactating-women-and-young-children.pdf?sfvrsn=4f864061_2. (Accessed 11 October 2021).

WHO, 2016. WHO Recommendations on Antenatal Care for a Positive Pregnancy Experience. World Health Organization, Geneva. http://www.who.int/nutrition/publications/guidelines/antenatalcare-pregnancy-positive-experience/en/.

WHO, 2020. Nutritional Interventions Update: Vitamin D Supplements During Pregnancy. WHO Antenatal Care Recommendations for a Positive Pregnancy Experience. World Health Organization, Geneva.

WHO, 2021. Nutritional Interventions Update: Zinc Supplements During Pregnancy. WHO Antenatal Care Recommendations for a Positive Pregnancy Experience. World Health Organization, Geneva.

WHO, FAO of the UN, 2004. Vitamin and Mineral Requirements in Human Nutrition, second ed. World Health Organization, Geneva https://www.who.int/nutrition/publications/micronutrients/9241546123/en/. (Accessed 10 October 2021).

Young, M.F., Oaks, B.M., Tandon, S., et al., 2019. Maternal hemoglobin concentrations across pregnancy and maternal and child health: a systematic review and meta-analysis. Ann. N. Y. Acad. Sci. 1450, 47–68.

Further reading

Allen, L.H., 2000. Anemia and iron deficiency: effects on pregnancy outcome. Am. J. Clin. Nutr. 71, 1280S–1284S.

Allen, L.H., 2012. Maternal nutrient metabolism and requirements in pregnancy and lactation. In: Erdman, J. (Ed.), Present Knowledge in Nutrition, eleventh ed. Wiley-Blackwell, Oxford.

Berry, R.J., Li, Z., Erickson, J.D., et al., 1999. Prevention of neural-tube defects with folic acid in China. China–US collaborative project for neural tube defect prevention. N. Engl. J. Med. 341, 1485–1490.

Butte, N.F., Wong, W.W., Treuth, K.J., Ellis, E., O'Brian, S., 2004. Energy requirements during pregnancy based on total energy expenditure and energy deposition. Am. J. Clin. Nutr. 79, 1078–1087.

Dror, D.K., Allen, L.H., 2010. Vitamin D inadequacy in pregnancy: biology, outcomes, and interventions. Nutr. Rev. 68, 465–477.

Institute of Medicine, Six Volumes on Dietary Reference Intakes. http://nap.edu.

Kaiser, L., Allen, L.H., 2008. Position of the American Dietetic Association: nutrition and lifestyle for a healthy pregnancy outcome. J. Am. Diet Assoc. 108, 553–561.

Murphy, M.M., Scott, J.M., Arija, V., Molloy, A.M., Fernandez-Ballart, J.D., 2004. Maternal homocysteine before conception and throughout pregnancy predicts fetal homocysteine and birth weight. Clin. Chem. 50, 1406–1412.

Prentice, A., 2000. Maternal calcium metabolism and bone mineral status. Am. J. Clin. Nutr. 71, 1312S–1316S.

Scholl, T.O., Hediger, M.L., Bendich, A., et al., 1997. Use of multivitamin/mineral prenatal supplements: influence on the outcome of pregnancy. Am. J. Epidemiol. 146, 134–141.

West Jr., K.P., Christian, P., Labrique, A.B., et al., 2011. Effects of vitamin A or B-carotene supplementation on pregnancy-related mortality and infant-related mortality in rural Bangladesh: a cluster randomized trial. J. Am. Med. Assoc. 305, 1986–1995.

Pregnancy: Placental regulation of nutrient delivery to the fetus

P. Haggarty, University of Aberdeen, Aberdeen, UK

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of P. Haggarty, Pregnancy: Placental Regulation of Nutrient Delivery to the Fetus, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 68–74, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00236-1.

Fetal Nutrient Requirements 177
The Human Placenta 179
Methods Used to Study Placental Function 179
The Mechanisms of Placental Nutrient Transport 179
Epigenetics and the Placenta 181
Placental Selectivity 182
Placental Metabolic Activity 182
Placental Buffering of Maternal Dietary Intake 183
Placental Insufficiency and Fetal Growth 183
The Role of the Fetus 184
Further Reading 184

Glossary

Endothelium Thin layer of cells lining the blood vessels.
Imprinting Parent of origin specific marking or expression of genes.
In utero In the uterus.
Intrauterine growth restriction (IUGR) Pathological deviation from normal growth.
Trophoblast Placental cells formed from the outer part of the blastocyst.
Vesicle Spherical preparation of membrane bilayer used to study transport.
Villi Hair- or finger-like projection from a cell which increases the surface area.

Fetal Nutrient Requirements

Prenatal development can usefully be divided into two periods; the embryonic period, which covers the first eight weeks of life and the fetal period, which lasts from the 9th week of gestation until term. During the latter period the fetus is entirely dependent on the

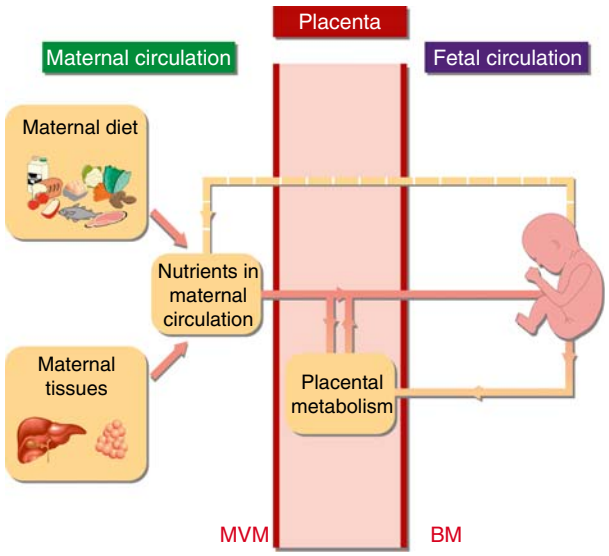


Figure 1 Nutrient exchanges between the maternal circulation, placenta, and fetus.

placenta for its supply of nutrients (Figure 1). The fetus has an absolute requirement for the same essential nutrients as the adult but the adequacy of supply is particularly critical during *in utero* life when all the structures of the body are being established. In addition, because of the particularly high demand for some strictly nonessential nutrients these may be considered as 'conditionally essential' if the rate of utilization exceeds the fetal capacity for *de novo* synthesis.

The placenta has to maintain the supply of all nutrients at a rate adequate to allow fetal growth to proceed along its optimum trajectory. It also has to provide an appropriate mix of nutrients to meet the needs of the fetus at the different stages of pregnancy. For example, in the first two-thirds of pregnancy the fetus deposits mainly protein whereas in late gestation fat takes over as the dominant form of deposition (Figure 2).

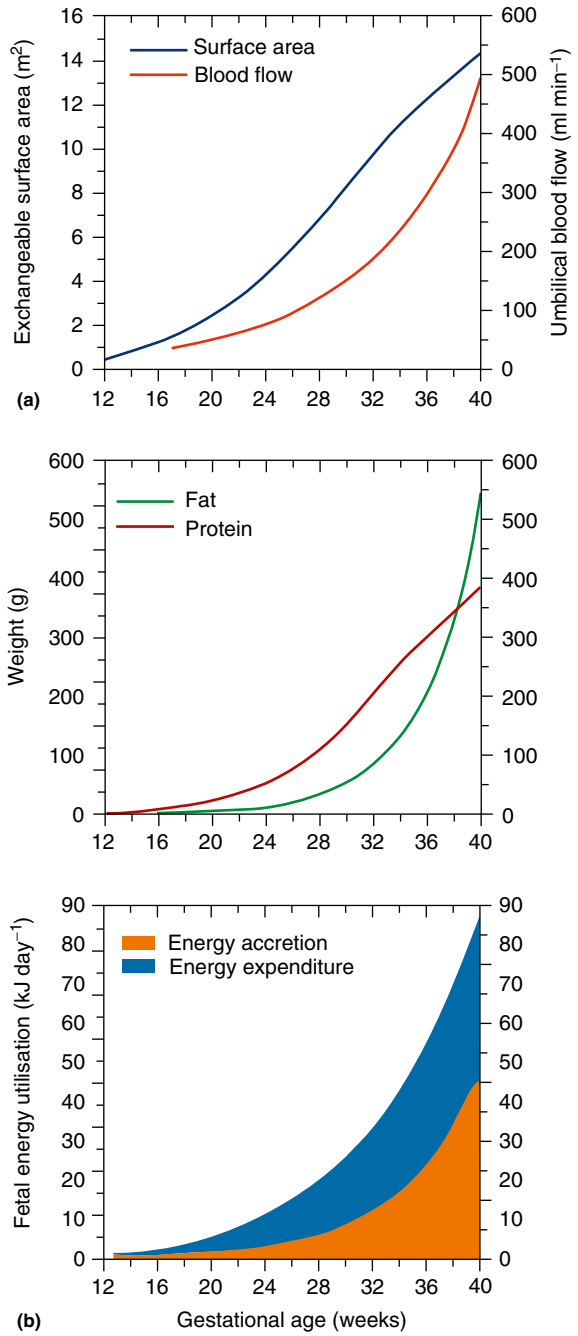


Figure 2 Changes with gestational age in placental exchangeable surface area and umbilical blood flow (a), accretion of fat and protein in the fetus (b), and the components of fetal energy requirements.

The availability of individual nutrients to the fetus depends not only on the maternal dietary intake but also on the function of the placenta and the many physiological and biochemical adaptations, which occur during pregnancy (Figure 2). An understanding of placental function and its interaction with diet is essential to the setting of appropriate dietary guidelines for pregnancy.

The Human Placenta

The human placenta is a hemochorial, villous type where the maternal blood enters the intervillous space *via* the spiral arteries and flows directly around the terminal villi of the fetal circulation without any intervening maternal vessel wall. The surface area available for exchange gradually increases throughout pregnancy, reaching approximately 10–15 m^2 in the last trimester (Figure 2). The nature of the exchangeable surface of the placenta also changes throughout gestation. Mature intermediate villi appear toward the end of the second trimester and the terminal villi – the main site of feto-maternal exchange – develop a few weeks later. The rate of fetal blood delivery to the placenta (umbilical flow) also increases with gestational age. It is approximately linearly related to fetal weight, and hence the fetal nutrient requirement, throughout gestation (Figure 2).

The human placenta typically weighs approximately half a kilogram at term. However, its physical bulk belies the delicate nature of the separation between the maternal and fetal circulations, which consists of only two cell layers; the syncytiotrophoblast and the capillary endothelium. The endothelium allows the passage of nutrients through pores within the interendothelial cleft and therefore is not a significant barrier to nutrient exchange between the maternal and fetal circulations. The effective barrier between the circulations is a thin trophoblastic sheet in the form of a syncytium (a tissue in which the cytoplasm of constituent cells is continuous), known as the syncytiotrophoblast. Between 10 weeks and term the thickness of the villous trophoblast falls from approximately 10 μm to 4 μm. Over the same period the overall materno–fetal diffusion distance drops from 40 μm to 5 μm. Any substance crossing between the maternal and fetal circulation has to pass though this barrier, which consists of two membranes; the microvillus membrane (MVM) facing the maternal blood and basal membrane (BM) facing the fetal blood. The surface area of the maternal facing MVM is approximately five- to six-times that of the fetal facing BM. There are other cell types and structures within the placenta, such as maternal myometrium and decidua, connective tissue, Hofbauer cells, and persisting cytotrophoblast cells, which contribute to the metabolic activity and nutrient requirements of the placenta but which are not thought to be significant barriers to transport.

Methods Used to Study Placental Function

Direct measurement of placental nutrient transport function in human pregnancy is practically and ethically extremely difficult to achieve and the available techniques require a tradeoff between physiological relevance and the quality of the information derived. Stable isotope labeled amino acids and fatty acids have been administered to the mother and their appearance measured in the cord blood. Such studies are clearly physiologically relevant but their interpretation is severely constrained by the number of sequential cord blood samples which can be taken. Placental function is often inferred by measurements of concentration differences in the maternal and fetal circulations. Measurements of arterio–venous differences across the umbilical cord can be made at cesarean section but, given the significant changes in placental function during development, their relevance to the younger placenta is not clear. Cord blood nutrient levels can be measured at earlier stages of development using the invasive method of cordocentesis but this procedure carries risks and may only be justified when carried out opportunistically as part of a clinical test, and this in turn may bias the population sample toward those with fetal or placental pathology. An important limitation to the interpretation of placental function from such 'snapshot' measurements is that the cord blood nutrient status is the net result of both placental delivery and fetal utilization. A number of *in vitro* approaches to the study of placental function are also available. The dually perfused placenta has the advantage that it retains the cellular structure and metabolic activity of the syncytiotrophoblast and the placental vascular structure. This preparation also allows the nutrient composition of the maternal and fetal circulation to be controlled and transfer rates to be measured dynamically using isotopic tracers. However, the placenta tends to be very mature, the efficiency of perfusion cannot be assumed to exactly mimic the *in vivo* situation. Also, care has to be taken to ensure that the composition of the maternal and fetal perfusates are made up in such a way that they are relevant to the form in which nutrients are actually transported in the maternal and fetal circulations. Vesicles formed from the syncytiotrophoblast are particularly well suited to the detailed study of nutrient transport mechanisms under highly controlled conditions. Finally, there is the identification and characterization of individual transport proteins.

The Mechanisms of Placental Nutrient Transport

The transport of individual nutrients across the placenta generally depends on the same principles and the presence of the same or similar transport systems to those in the tissues and organs of the adult although there are some additional factors specific to the placenta (Figure 3). In particular, unlike most tissues in the adult where either uptake or export dominate at any given time, the syncytiotrophoblast whose primary function is transport has to do both simultaneously.

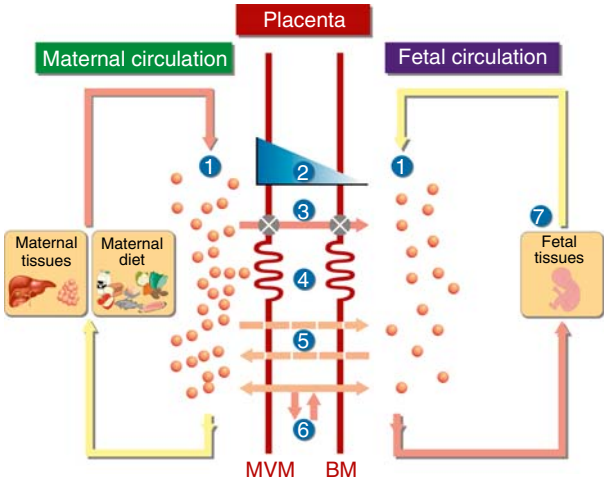


Figure 3 Factors affecting nutrient transfer across the syncytiotrophoblast. These include: (1) maternal and fetal blood flow, (2) the concentration gradient across the syncytiotrophoblast for nutrients and, where relevant, their transporters in the maternal and fetal circulations, (3) the concentration of transport proteins to facilitate or actively transport nutrients, (4) the exchangeable surface area, (5) the rate of diffusion of some nutrients across membranes without the intervention of transport proteins, (6) metabolism (utilization and *de novo* synthesis) within the placenta, and (7) the rate of nutrient utilization by the fetal tissues.

The placental transport systems for the macronutrients (carbohydrate, fat, and protein) have been extensively studied. Glucose transport within the placenta appears to be mediated exclusively by the GLUT1 transporter, which has been located on both the MVM and BM. GLUT3 and GLUT4 are also present in the placenta but not in the syncytiotrophoblast itself. They are located on the vascular endothelium and the intravillous stromal cells, respectively. The syncytiotrophoblast also contains a wide range of amino acid transporters; system A, ASC, Asc, B^0, $b^{0,+}$, L, N, Gly, y^+, y^+L, and X_{AG} and β. A number of fatty acid binding proteins are also found in the placenta. Of these proteins FAT/CD36 and FATP have been located on both the MVM and BM but there is also a placenta specific protein (p-FABPpm), which has been located exclusively on the MVM. This p-FABPpm is similar in size (~40 kDa) to the ubiquitous FABPpm found in most mammalian cells but it has a different amino acid composition.

The driving force, which results in the net transfer of nutrients to the fetus is different for different nutrients and this is reflected in their transplacental gradients (Figure 4). Where the nutrient concentrations are lower in the cord than maternal blood this has been cited as a reason to supplement the mother but in many cases it is precisely this gradient which drives placental nutrient transfer. Glucose is thought to flow down a concentration gradient from the mother to the fetus and this process of 'facilitated diffusion' is mediated by GLUT1. Unlike glucose the concentration of most amino acids in the fetal circulation is greater than that in the maternal circulation suggesting some form of active transport. For many amino acids the concentration is even higher within the placenta than the fetal circulation and the key gradient generating step for amino acids is the active transport across the MVM. The amino acids can then diffuse down a concentration gradient into the fetal circulation, and to some extent back to the mother. The concentration of water-soluble vitamins and lactate in the fetal circulation also exceeds that in the maternal circulation.

Like glucose, the fats and fat-soluble vitamins also flow down a concentration gradient from the mother to the fetus mediated by the various fatty acid transport proteins. However, unlike glucose or the amino acids, fat-soluble compounds can also cross the syncytiotrophoblast, and all other membranes for that matter, by simple diffusion and partition without the intervention of a carrier protein. The role of the fatty acid binding proteins appears to be to improve the efficiency of this process. The key factor in understanding the driving force for the placental transfer of fat-soluble nutrients is that these compounds are only sparingly soluble in water (13 μM for C18:0 at 37 °C) and have to be transported in the plasma in hydrophobic binding sites on carrier proteins. The partition of fats between the maternal and fetal circulations is largely determined by the relative abundance of available hydrophobic binding sites within those compartments. Because only NEFA is thought to cross membranes, it is the NEFA concentration gradient which is most relevant to the transplacental flow of fatty acids. The concentration of NEFA in the maternal plasma at term is approximately three-times that in the fetal circulation but the concentration of its primary carrier protein, albumin, is actually 10–20% higher in the fetal circulation. This results in a ratio of NEFA to albumin on the fetal side of the placenta of around a quarter of that on the maternal side at term. The fat-soluble vitamins (A, E, and D) are also present in the fetal circulation in lower concentrations than in the maternal circulation. These materno–fetal concentration differences for the macronutrients develop gradually throughout gestation.

It is less easy to generalize about the transplacental gradient for minerals as some are at a lower concentration in the fetal circulation (Se, Cu, Ba) some are higher (Ca, Zn, Be, Rb) and some are about the same (Co, Mg, Mo, Sn, Bi, Cd, Cs, La, Li, Pb). Iron is particularly important during pregnancy and its concentration in the fetal venous blood leaving the placenta is almost three-times that of the maternal serum. Iron is transported in the serum on the transport protein transferrin and, like the fats and fat-soluble vitamins, its rate of transfer may be influenced by the availability of free binding sites.

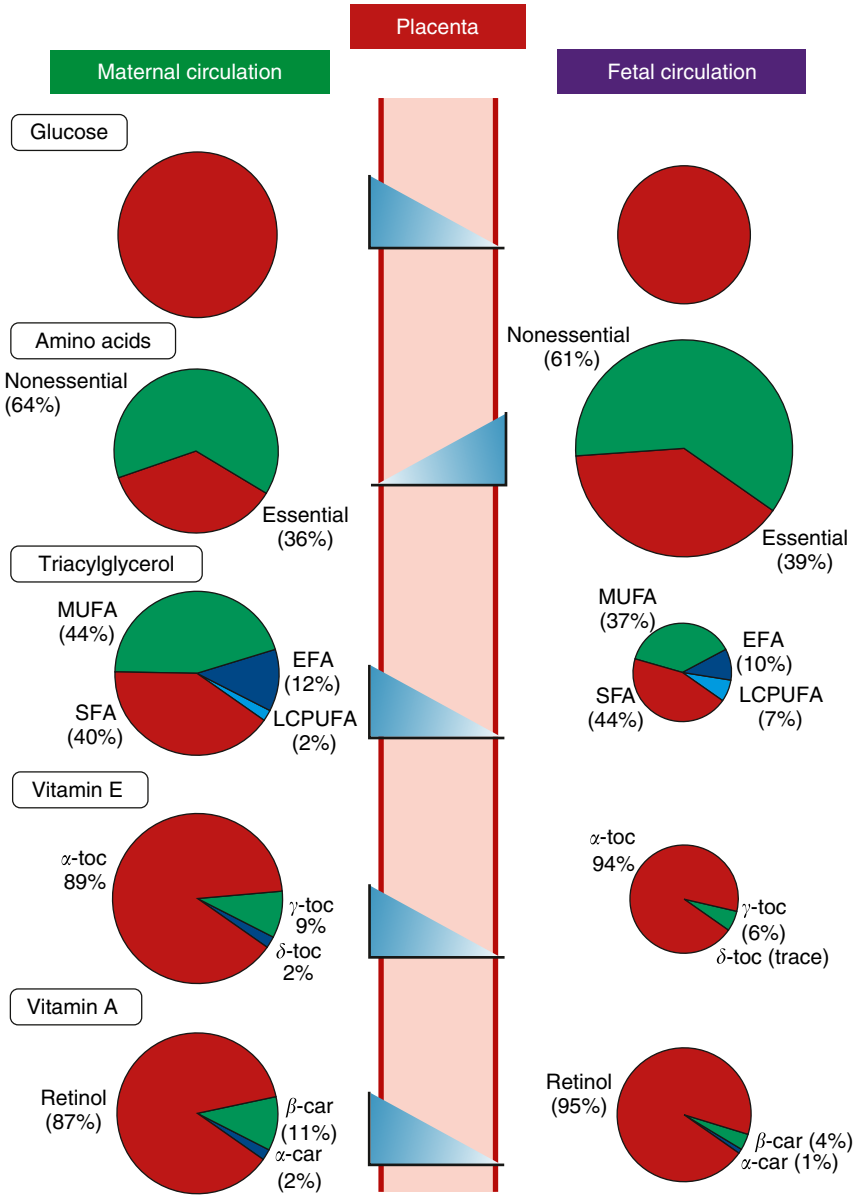


Figure 4 The relative concentration of nutrients in the maternal and fetal circulations. The concentration differences for each nutrient class are represented by the area of the circle in the fetal circulation relative to the maternal circulation. Apart from glucose the relative concentrations of individual nutrients within the nutrient groups are shown as segments of the circle. For triglyceride the fractions are saturated (SFA) monounsaturated (MUFA), essential (EFA) and long chain polyunsaturated fatty acids (LCPUFA). For vitamin E the fractions are α-tocopherol (α-toc), γ-tocopherol (γ-toc) and δ-tocopherol (δ-toc). For vitamin A the fractions are β-carotene, (β-car), and α-carotene (α-car).

Epigenetics and the Placenta

There is growing interest in the role of epigenetics, and imprinting in particular, in controlling placental function and nutrient delivery to the fetus. The vast majority of human autosomal genes are thought to be equally expressed from the two parental alleles. In the imprinted genes the expression pattern is different for the maternally and paternally derived alleles, with information on the parental origin of each allele being retained in the conceptus through epigenetic mechanisms. Epigenetics encompasses a collection of mechanisms that control gene function and chromatin structure without altering the nucleotide sequence of DNA and the most commonly studied is DNA methylation. Imprinted genes make up only approximately 1% of all human genes. Their main functions include control of placental function, fetal growth, and brain development but the way in which the imprint is acquired and propagated is not fully understood. Much of what is known about imprinting and developmental epigenetics comes from studies in mice, though data from human reproduction are increasingly available. Although there are important differences between species, particularly in the timing of epigenetic events, a number of themes appear to be universal. Following demethylation and

genome wide *de novo* methylation at the embryonic pregastrulation stage, striking differences in the methylation status of embryonic cell lineages within the early embryo are apparent. The trophoblast lineages, which give rise to the placenta, typically achieve a much lower level of methylation than the heavily methylated somatic cells, which give rise to the fetal tissues. Such differences, and the parent of origin specificity of the imprinted genes, have given rise to the main hypothesis that seeks to explain the phenomenon of imprinting in placental mammals. The 'conflict theory' of imprinting is based on the premise that the function of the imprinted genes is to control resource allocation to the fetus. The logic being that the mother is more likely to pass on her genes if she spreads her body resources over a number of pregnancies whereas the father's genes are more likely to be successful in evolutionary terms if they maximize resource allocation to his offspring, even at the mother's expense. Paternally expressed imprinted genes are generally thought to enhance fetal growth by promoting resource allocation to the conceptus whereas maternally expressed imprinted genes are thought to protect the maternal resources by suppressing fetal growth. Experimental data to support this model has been obtained in mice, but it has yet to be confirmed in detail for humans.

Placental Selectivity

One of the key functions of placental nutrient transport is to maintain the most appropriate balance of nutrients in the fetal circulation and the balance of nutrients transferred by the placenta may be as important as the overall transfer capacity in influencing the pattern of fetal growth. Nutrients such as the fatty acids and amino acids occur in many forms yet they are translocated across membranes by a relatively small number of transporter molecules (**Figure 3**). This nutritional 'bottleneck' results in competition for transfer and the possibility of placental selectivity. An example of the resulting change in nutrient quality can be seen in the increase in the relative proportion of the essential to nonessential amino acids in the fetal circulation compared to the maternal circulation (**Figure 4**). The same is true of the fat-soluble vitamins where the relative concentration of the most biologically active form is increased in the fetal circulation. In the case of the fatty acids it is the long chain polyunsaturated fatty acids (LCPUFAs) such as arachidonic acid [20:4 n-6; AA] and docosahexaenoic acid [22:6 n-3; DHA], which perform most of the essential functions in the fetus. Although the overall concentration of the lipid classes are greatly reduced in the fetal circulation the critical LCPUFA make up a greater proportion of total fatty acid in the fetal circulation. In the case of the fatty acids the placenta has multiple mechanisms including preferential binding of LCPUFA by p-FABPpm, selective uptake by the syncytiotrophoblast, intracellular metabolic channeling of individual fatty acids, and selective export to the fetal circulation, which allow it to preferentially deliver DHA and AA to the fetal circulation.

Placental Metabolic Activity

Although the barrier between the maternal and fetal circulation is effectively only one cell thick, the placenta is a substantial organ, made up of many cell types. It is extremely active metabolically and has its own requirement for nutrients and this is consistent with the observations that the surface area of the maternal facing membrane (MVM) is approximately five-times greater than that of the fetal facing membrane (BM), that the concentration of expression of GLUT1 is greater on the MVM, that the MVM contains additional fatty acid binding proteins, which are not present on the BM, and that the amino acid transporters act to produce the maximum amino acid gradient across the MVM. The metabolic transformations within the placenta are intimately linked to fetal metabolism and represent another way in which the placenta can regulate nutrient transport availability within the fetal circulation.

In late pregnancy the overall contribution of fat to whole body oxidation is reduced and this is thought to result from the preferential utilization of carbohydrate and amino acids such as glutamate as an energy source in the feto-placental unit and the sparing of fatty acids to maximize fetal accretion of the critical LCPUFA in particular. The inter-relationships between the placenta and fetus are particularly complex for the amino acids. The placenta is a net user of serine, glutamate, leucine, isoleucine, and valine and there is significant interconversion of alanine, pyruvate, and lactate between the placenta and fetal tissues. The concentration of lactate in the fetal circulation is considerably greater than that in the maternal circulation and a considerable proportion of the glucose taken up by the placenta is converted into lactate before export into the fetal circulation for use by the fetus. The placenta takes up serine from both the maternal and fetal circulation, converting this into glycine and exporting it into the fetal circulation for oxidation by the fetal liver and there is significant cycling of glutamate and glutamine between the placenta and fetal liver. This partition of the various segments of metabolic pathways between the placenta and fetal tissues is a general phenomenon and in many respects the feto-placental unit can be considered as a metabolic whole with the placenta acting as an extra fetal organ in addition to its role as a simple nutrient transporter. Metabolic activity in the feto-placental unit is also responsive to nutrient supply and fetal demand. For example, AA is an important precursor of the prostacyclins, prostaglandins, thromboxanes, and leukotrienes, which play key roles in pregnancy. When the maternal circulation of AA is low there is net uptake by the placenta of AA from the fetal circulation, presumably to maintain placental synthesis of these important compounds.

Placental Buffering of Maternal Dietary Intake

In cases where the increased demand for nutrients during pregnancy is not met by the diet alone the shortfall may be made up from the maternal stores and the placenta may play a role in orchestrating some of the maternal nutritional adaptations in pregnancy (Figure 3). For example, placentaly derived leptin is a potent stimulator of lipolysis and there is evidence that the rate of export into the maternal circulation is controlled to allow the placenta to modulate its own substrate supply in response to the fetal demand for fats. The various homeostatic mechanisms within the placenta and their interaction with maternal physiological adaptations during pregnancy act to ensure a constant supply of substrate to the fetus, free of large diurnal fluctuations corresponding to the timing of maternal meals, and to protect the fetus against a transiently poor intake during critical periods of fetal growth. These adaptations help the mother to meet the full fetal requirement for nutrients such as LCPUFA and iron while consuming apparently poor diets.

Placental Insufficiency and Fetal Growth

Low birth weight is a significant public health problem in developing countries where maternal nutrition may be marginal or poor. Low birth weight resulting from poor nutrition is also a concern in industrialized societies but maternal deficiency here is relatively rare and potentially a more important public health issue relating to nutrition in pregnancy is the apparent epidemiological association between birth weight and adult disease susceptibility (cardiovascular disease, diabetes, and hypertension). The highest risk is associated with the lowest birth weight but, because of the nature of the normal distribution, in terms of the numbers potentially affected in adult life it is the small variations in the normal birth weight range, which have the largest public health implications. A causal connection between birth weight and adult disease has been proposed in the 'fetal origins' hypothesis which is that fetal undernutrition in middle to late gestation, leads to disproportionate fetal growth and programs later disease susceptibility. The close association between birth weight and placental weight has led to speculation that the placenta may limit fetal growth within the normal weight range. However, the available evidence suggests that the capacity of the normal human placenta to transport macronutrients exceeds the fetal requirement and that a considerable proportion of transport function would have to be lost before it became limiting for fetal growth.

Intrauterine growth restriction (IUGR) resulting from utero–placental insufficiency is a serious pathology, which is associated with a greatly increased risk of adverse outcomes including perinatal mortality and morbidity, impaired mental, visual and aural

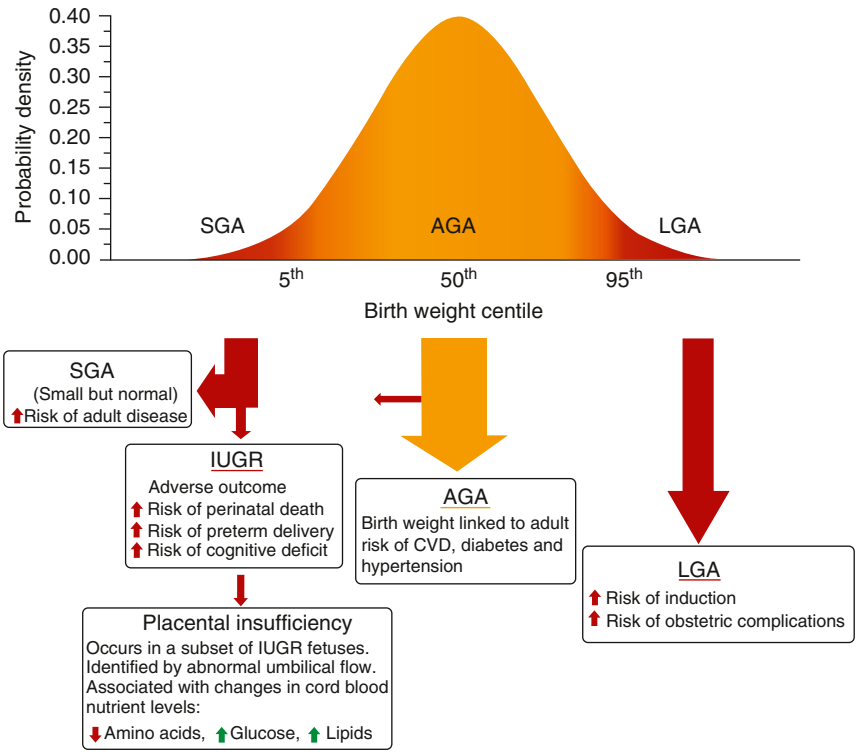


Figure 5 The normal distribution of birth weights and relative risks associated with babies who are small for gestational age (SGA), appropriate for gestational age (AGA), large for gestational age (LGA) and those subjected to intrauterine growth retardation (IUGR) and the relationship to placental insufficiency.

development, autism, and cerebral palsy and strongly associated with serious adverse maternal outcomes especially pre-eclampsia (Figure 5). IUGR is often detected indirectly by measuring abnormal umbilical artery flow velocity waveforms and abnormal fetal heart rate. The abnormal waveforms are thought to result from increased vascular resistance associated with abnormal arteriolar tree and villi branching and a reduction in the villous capillary tree. Pregnancies in which these abnormalities are observed are also associated with fetal hypoxia and reduced concentrations of glucose and amino acids in the fetal circulation and reduced activity of the system A amino acid transporter within the placenta. However, *in vitro* studies have shown that the hypoglycemia observed in some IUGR fetuses is not caused by a decreased glucose transport capacity within the placenta (expression and activity of GLUT1) and IUGR fetuses are actually hypertriglyceridemic compared to their appropriately grown counterparts. The fetal blood concentrations of the trace elements are also either normal or elevated in IUGR. Thus although it is possible that the placenta from IUGR fetuses may limit the supply of amino acids there is no evidence that placental delivery is the first limiting factor in the supply of glucose, lipids, or trace elements. IUGR is a complicated syndrome in which almost all aspects of placental and fetal metabolism are altered and many researchers have emphasized the primary importance of the fetal hypoxia and its effects on fetal metabolism rather than a simple limitation of placental nutrient transfer capacity.

There is considerable uncertainty about the magnitude of the problem of IUGR. The lowest 5–10% of weight for gestational age babies may be referred to as small for gestational age (SGA) but babies in this range need not be growth retarded. They may be naturally small and have no increased risk of adverse outcome. Conversely, a baby born within the apparently normal birth weight range could have suffered growth retardation *in utero* if its genetic potential was for a higher birth weight (Figure 5). The true incidence of IUGR resulting from utero–placental insufficiency is therefore unknown but if it is defined in relation to umbilical flow or fetal heart rate abnormalities then it is only a fraction of even those in the lowest 5% of weight for gestational age that are affected by utero–placental insufficiency. At the other end of the spectrum babies who are large for gestational age (LGA) are at higher risk of adverse obstetric outcomes and early developmental problems but there is no evidence that LGA or macrosomic babies are produced as a result of a primary alteration in the placenta.

The Role of the Fetus

The nutrient composition of the human diet varies enormously between populations yet the healthy human newborn is essentially the same the world over. The available evidence points to extensive homeostatic mechanisms at work within the placenta to ameliorate some of the variations in the quality of the maternal diet by regulating the mix of nutrients to the developing fetus. However, these mechanisms can only operate on the nutrients already available in the maternal circulation. The maternal diet and maternal circulating concentrations of many nutrients are major determinants of the concentrations in the fetal circulation and the fetus clearly has the ability to cope with relatively large variations in nutrient availability in the cord blood. The fetus also plays an active role in regulating placental nutrient transfer. The rate of placental nutrient transport is directly influenced by the transplacental concentration gradient, which is in turn largely determined by the rate of uptake by the fetal tissues. Another major determinant of placental nutrient transfer is the umbilical blood flow, which is approximately linearly related to the fetal weight, and hence the fetal nutrient requirement, throughout gestation. Finally, the most intimate connection between the fetus and the placenta is the way in which different parts of metabolic pathways and cycles are distributed between the placenta and fetal tissues, mainly the fetal liver. Thus although the placenta has to provide the correct mix of nutrients in sufficient quantities to support fetal growth and development throughout pregnancy it is the fetus itself, which ultimately regulates many key aspects of placental nutrient transfer function.

Further Reading

Cetin, I., 2003. Placental transport of amino acids in normal and growth-restricted pregnancies. European Journal of Obstetrics and Gynecology and Reproductive Biology 110 (supplement 1), S50–S54.

Duttaroy, A.K., 2009. Transport of fatty acids across the human placenta: A review. Progress in Lipid Research 48, 52–61.

Gagnon, R., 2003. Placental insufficiency and its consequences. European Journal of Obstetrics and Gynecology and Reproductive Biology 110, S99–S107.

Haggarty, P., 2010. Fatty Acid Supply to the Human Fetus. Annual Review of Nutrition 30, 237–255.

Hay Jr, W.W., 1995. Metabolic interrelationships of placenta and fetus. Placenta 16, 19–30.

Illsley, N.P., 2000. Glucose transporters in the human placenta. Placenta 21 (1), 14–22.

Jansson, T., Ylven, K., Wennergren, M., Powell, T.L., 2002. Glucose transport and system A activity in syncytiotrophoblast microvillous and basal plasma membranes in intrauterine growth restriction. Placenta 23 (5), 392–399.

Kaufmann, P., Scheffen, I., 1998. Placental development. In: Polin, R.A., Fox, W.W. (Eds.), Fetal and Neonatal Physiology. W.B. Saunders Company, Philadelphia, pp. 59–70.

Marconi, A.M., Paolini, C., Buscaglia, M., Zerbe, G., Battaglia, F.C., Pardi, G., 1996. The impact of gestational age and fetal growth on the maternal-fetal glucose concentration difference. Obstetrics and Gynecology 87, 937–942.

Pardi, G., Marconi, A.M., Cetin, I., 2002. Placental-fetal interrelationship in IUGR fetuses – A review. Placenta 23 (supplement A), S136–S141.

Regnault, T.R., de Vrijer, B., Battaglia, F.C., 2002. Transport and metabolism of amino acids in placenta. Endocrine 19 (1), 23–41.

Reik, W, Walter, J., 2001. Genomic imprinting: parental influence on the genome. Nature Reviews Genetics 2, 21–32.

Sparks, J.W., 1984. Human intrauterine growth and nutrient accretion. Seminars in Perinatology 8, 74–93.

Pregnancy: Pre-eclampsia and diet

Doris González-Fernández[a], **Kristine G. Koski**[b], **and Zulfiqar Bhutta**[a], [a] Centre for Global Child Health, The Hospital for Sick Children (SickKids), Toronto, ON, Canada; and [b] School of Human Nutrition, McGill University (Macdonald Campus), Ste-Anne-de-Bellevue, QC, Canada

© 2023 Elsevier Ltd. All rights reserved.

This is an update of E. Abalos, Pregnancy: Pre-eclampsia and Diet, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 75–80, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00235-X.

Introduction 185
Risk factors 185
Classification of HDPs 186
Pathophysiology of preeclampsia 187
Angiogenic and antiangiogenic factors 187
Inflammation 187
Oxidative stress 188
Abnormal lipid metabolism 189
Plasma volume dysregulation 190
Early detection of HDPs 190
Association between nutritional factors and preeclampsia 190
Overweight and obesity 190
Maternal undernutrition 191
Vitamins 191
Minerals and trace elements 192
Conclusion 193
References 193

Key points

- Overview of general and dietarian risk factors and classification of hypertensive disorders of pregnancy
- Synopsis of the pathophysiology of pre-eclampsia with emphasis in possible nutritional pathways involved
- Introduction to strategies for early detection of pre-eclampsia
- Current knowledge of associations between pre-eclampsia and nutritional factors

Introduction

Hypertensive disorders of pregnancy (HDPs) defined as systolic blood pressure ≥140 and/or diastolic blood pressure ≥90 after 20 weeks of gestation (Garovic et al., 2021) are one of the main causes of maternal and fetal/neonatal morbidity, increasing the odds of myocardial infarction, stroke, spontaneous coronary artery dissection, peripartum cardiomyopathy, small for gestational age, stillbirth, preterm delivery, placental abruption and postpartum hemorrhage, as well as the odds of maternal mortality (Garovic et al., 2021). The incidence of HDPs varies depending on the socio-demographic index, countries with low socio-demographic index having higher incidence (70.6×10^4 population) than those with high socio-demographic index (10.9×10^4 population) (Wang et al., 2021), and the prevalence of eclampsia, one of the most severe forms of HDPs, was reported as 50–151 per 10,000 deliveries in developing countries, but 1.6–10 per 10,000 deliveries in developed countries (Fishel Bartal and Sibai, 2020).

Risk factors

Age and race: HDPs affect most the extremes of reproductive age. Adolescent pregnancy (Leppalahti et al., 2013) is considered a risk factor for developing HDPs. On the other hand, women ≥35 years were found to have a higher risk of pre-eclampsia (PE) (Lamminpaa et al., 2012; Schimmel et al., 2015), and those ≥40 yrs, increased risk of late-onset post-partum PE (Bigelow et al., 2014; Koo et al., 2012; Khalil et al., 2013). African American (Shen et al., 2005; Liu et al., 2014), Hispanic (Bigelow et al., 2014) and Indigenous women (Johnson and Louis, 2020) are more likely to develop HDPs.

Obstetric and medical history: A review and meta-analysis of 92 studies described obstetric and medical conditions that increase the risk of PE, describing risk ratios [95% confidence interval (CI)] (Bartsch et al., 2016). Obstetric risk factors comprised history of

PE [8.4 (7.1–9.9)], multifetal pregnancy [2.9 (2.6–3.1)], prior stillbirth [2.4 (1.7–3.4)], first pregnancy [2.1 (1.9–2.4)], history of placental abruption [2.0 (1.4–2.7)] and assisted reproduction [1.8 (1.5–2.1)]. Maternal medical conditions such as chronic hypertension [5.1 (4–6.5)], pregestational diabetes mellitus [3.7 (3.1–4.3)], obesity [2.8 (2.6–3.1)], chronic kidney disease [1.8 (1.5–2.1)], and autoimmune disorders such as systemic lupus erythematosus [2.5 (1.0–6.3)] and antiphospholipid syndrome [2.8 (1.8–4.3)] have also been associated with increased the risk of PE (Bartsch et al., 2016).

Genetic and fetal factors: Family and racial predispositions to the disease have prompted an examination of polymorphisms at the maternal and fetal levels that might be increasing the risk of HDPs and/or PE (Broughton Pipkin, 1999). Some polymorphism variants in genes for angiotensin I and II (Li et al., 2015), nitric oxide synthase (Qi et al., 2013; Groten et al., 2014), methylenetetrahydrofolate reductase (Wang et al., 2013), coagulation factor V, leptin receptor (Fong et al., 2014), FMS-like tyrosine kinase 1, methylenetetrahydrofolate reductase, and vitamin-D receptor (Guan et al., 2022) among others have been associated with increased risk of PE. Fetal factors for the development of HDPs, such as carrying a male fetus and multiple pregnancies, have been reported; also, genetic disorders in the offspring such as Beckwith-Wiedemann syndrome and trisomy 13, have been related with higher risk of HDPs, whereas trisomy 21 might reduce the risk (Petry et al., 2014).

Diet: A Western dietary pattern The Dietary Approaches to Stop Hypertension (DASH) diet is known to improve blood pressure in non-pregnant populations (Wiertsema et al., 2021). It recommends a high daily intake of fruits, vegetables and grain products, the daily inclusion of low-fat dairy products and moderate intake of meats, poultry, and fish in the diet, as well as nuts, seeds, and legumes (Lin et al., 2003). A higher score in the intake of the components of the DASH diet was associated with lower diastolic blood pressure in mid-pregnancy and with better feto-placental vascular function in late pregnancy, but was not associated with HDPs in a low-risk European population (Wiertsema et al., 2021). However, a hospital-based study in China found an inverse relationship between the adherence to DASH diet and the odds of PE (Cao et al., 2020). A higher adherence to the New Nordic Diet, which promotes the high intake of organic legumes, vegetables, fruit, whole grains, seafood, potatoes, nuts, herbs and other Nordic-produced foods (Jensen et al., 2015), was associated with low relative risk of PE only in nulliparous women, but also was associated with preterm delivery in multiparous women (Hillesund et al., 2014).

The Mediterranean diet is typically known as being high in unprocessed plant foods (fruits, vegetables, legumes, nuts, whole grain cereals, and olive oil); low to moderate in fish/shellfish; infrequent intake of red meat, animal fats, vegetable oils, and processed foods and low intake of red wine (Villani et al., 2019). However, a recent meta-analysis of randomized-controlled trials involving 2277 subjects found no effect of the Mediterranean diet with PE (Zhang et al., 2021). In low-medium income countries, the evidence of a Western diet (high sugar, fried/processed goods) increasing the odds of HDPs is inconclusive; however, the low or no-intake of fruits respectively increased by 2.3 and 2.6 times the odds of PE compared with women who consumed fruits and vegetables during their pregnancies (Kinshella et al., 2021a). Moreover, evidence of a healthy diet pattern (high intake of fruits, vegetables and low-fat dairy) was associated with 51–82% reduced incidence of PE (Kinshella et al., 2021a).

Classification of HDPs

The classification of HDPs slightly differs among guidelines from North American, Latin American and Europe. In general, guidelines align with the International Society for the Study of Hypertension in Pregnancy (ISSHP) (Brown et al., 2018), that classifies HDPs in four categories:

1. Pre-existent chronic hypertension: Hypertension that is present prior to the pregnancy or before week 20 of gestation
2. Gestational hypertension: Hypertension after 20 weeks of gestation without manifestations of complications
3. Preeclampsia—*de novo* or superimposed on chronic hypertension: New onset of hypertension after week 20 of gestation, associated with proteinuria or one or more severe complications
4. White coat hypertension: Elevated blood pressure that is only present during the physical examination in a clinic or office; it does not require antihypertensive medication.

PE is the most severe form of HDPs (Cicero et al., 2015). PE is a multisystemic disease occurring exclusively during pregnancy after week 20 of gestation (Lambert et al., 2014), during the peri-partum time or up to 42 days after delivery (Vilchez et al., 2015). It is characterized by *de novo* appearance of hypertension, with blood pressure measured at least twice with an interval of at least 4 h (Lambert et al., 2014).

PE has also been subclassified using different criteria, such as the presence/absence of intra-uterine growth restriction, its origin (placental vs. maternal), its association with later maternal cardiovascular disease or immunological/genetic subtypes (Roberts et al., 2021). Most generally accepted, PE can be classified as severe when the systolic blood pressure is $\geq$160 or diastolic blood pressure $\geq$110 mmHg (Poon et al., 2019) or one of the following signs of systemic disease is present:

- Proteinuria, defined as $\geq$300 mg/24 h (gold standard) or $\geq$1+ on reagent strip, is a diagnostic criterion for PE in some country guidelines (Scott et al., 2020), but is not an absolute requirement for the diagnosis of PE (Magee et al., 2014; ACOG, 2013; Tranquilli et al., 2014), because pathophysiological features of PE can be present before the kidney involvement is severe enough to induce proteinuria (Chaiworapongsa et al., 2014), and because proteinuria might not be present in complications such as eclampsia, acute renal failure or pulmonary edema (Thornton et al., 2010).

- Other maternal organ involvement: Central nervous system (severe headache and/or hyperreflexia with clonus, altered mental status), liver (double or more liver transaminases), kidney [creatinine >90 μmol/L, or double of previous creatinine, or oliguria (<500 mL/d)], hematological complications [thrombocytopenia (platelet count <150,000/μL)] (Poon et al., 2019), or new onset of visual impairment with retinopathy, retinal detachment, or cortical blindness (Abu Samra, 2013).
- Utero-placental involvement: Fetal growth restriction, stillbirth or abnormal umbilical artery Doppler wave (Poon et al., 2019).

The detection and urgent treatment of those signs are important given that they predict the imminent development of two forms of severe PE that increase the risk of maternal and perinatal morbidity and mortality (Vigil-De Gracia et al., 2015):

- Eclampsia, the occurrence of generalized tonico-clonic seizures at the end of pregnancy, during labor or postpartum in women with PE (Lipstein et al., 2003), and
- HELLP syndrome, an acronym for Hemolysis, Elevated Liver enzymes and Low Platelets, results from a thrombotic micro-angiopathy, with consumptive thrombocytopenia and the formation of thrombi in small vessels (Balsak et al., 2015).

However, as PE can quickly worsen to produce complications, the term "severe PE" has been intentionally avoided from some country guidelines, to encourage a close survey of all women with PE (Scott et al., 2020).

Furthermore, PE can also be classified as early when starting <34 weeks, and late when starting ≥34 weeks of gestation, as they differ based on associated risk factors, biophysical and biochemical markers, uterine artery blood flow measured by artery Doppler velocimetry and in pathological lesions found in the placenta (Ogge et al., 2011). According to this classification, early onset PE would be characterized by deficient remodeling of uterine spiral arteries, impaired placental development and intra-uterine growth restriction, whereas late onset PE would be mainly characterized by maternal endothelial dysfunction, both early and late processes leading to systemic inflammatory response (Aneman et al., 2020). With advanced knowledge in the pathophysiology of PE, new sub-classifications involving different stages of the disease have been proposed and are currently under investigation (Staff, 2019).

Pathophysiology of preeclampsia

Despite extensive research, the exact etiology of PE has not yet been elucidated. Current knowledge of PE recognizes it as a multi-causal and heterogeneous syndrome (Redman, 2011) with two main possible etiologies depending on the early- or late-onset gestational age at diagnosis. Subsequently, two stages of PE have been classically recognized: (a) poor placentation, and (b) endothelial dysfunction and clinical syndrome (Tranquilli and Landi, 2010; Cheng and Wang, 2009). The first stage of misplacentation can start as early as two weeks gestation, where cytotrophoblast fails to properly invade and remodel maternal spiral arterioles, producing a shallow placentation with narrow maternal vessels and relative placental ischemia (Rana et al., 2019). Also, a premature unblocking of spiral arteries that normally occurs about 8 weeks could impair placental growth due to oxidative stress instead of decreased perfusion (Redman et al., 2020). Independently of the origin of placental injury, the primary effect seen in clinical PE is the imbalance between proangiogenic and antiangiogenic factors leading to maternal endothelium dysfunction (Agarwal and Karumanchi, 2011). **Fig. 1** summarizes the main dietary factors associated with the pathophysiology of PE.

Angiogenic and antiangiogenic factors

Angiogenic and antiangiogenic factors have been the subject of extensive research in the last decade, and have been considered as part of international guidelines for early detection of PE (Verlohren and Droge, 2020). Placental growth factor (PlGF), which belongs to the VEGF cytokine family, is a proangiogenic protein expressed in the syncytiotrophoblast that promotes the vasculogenesis process (Cerdeira and Karumanchi, 2012) and is normally found in high concentrations in the first weeks of pregnancy with a decrease after this time (Jardim et al., 2015). Soluble fms-like tyrosine kinase (sFlt-1) is an antiangiogenic protein produced by the syncytiotrophoblast that binds to the PlGF receptor circulating form, preventing its interaction with endothelial receptors, leading to endothelial cell dysfunction and vasoconstriction (Jardim et al., 2015; Verlohren and Droge, 2020). In normal pregnancies, sFlt-1 increases at the end of pregnancy but is elevated earlier in preeclamptic pregnancies (Levine et al., 2006). The prospective observational study PROGNOSIS, conducted in 14 countries, found that the ratios sFlt-1 to PlGF >38 had a sensitivity of 66.2% and specificity of 83.1% for predicting PE withing 4 weeks (Zeisler et al., 2016) and has been proposed for integration into clinical practice (Suresh and Rana, 2021).

Inflammation

Inflammatory responses are cornerstones in the pathophysiology of PE. In normal gestations, the modulation of the immune response, with a hormonal/placental-induced shift from a T-helper (Th) 1 toward a Th2 immune response is needed for appropriate development of pregnancy (Hegde, 1991; Sykes et al., 2012). Under normal conditions, naïve T-cells differentiate into regulatory T cells (Treg) (Ghaebi et al., 2017), creating the shift toward a Th2 biased state. However, depending on the cytokine milieu, this physiologic response could turn into the production of pro-inflammatory Th17 cells (Saito et al., 2010) and Th1 cells (Chau et al., 2016), leading to adverse pregnancy outcomes.

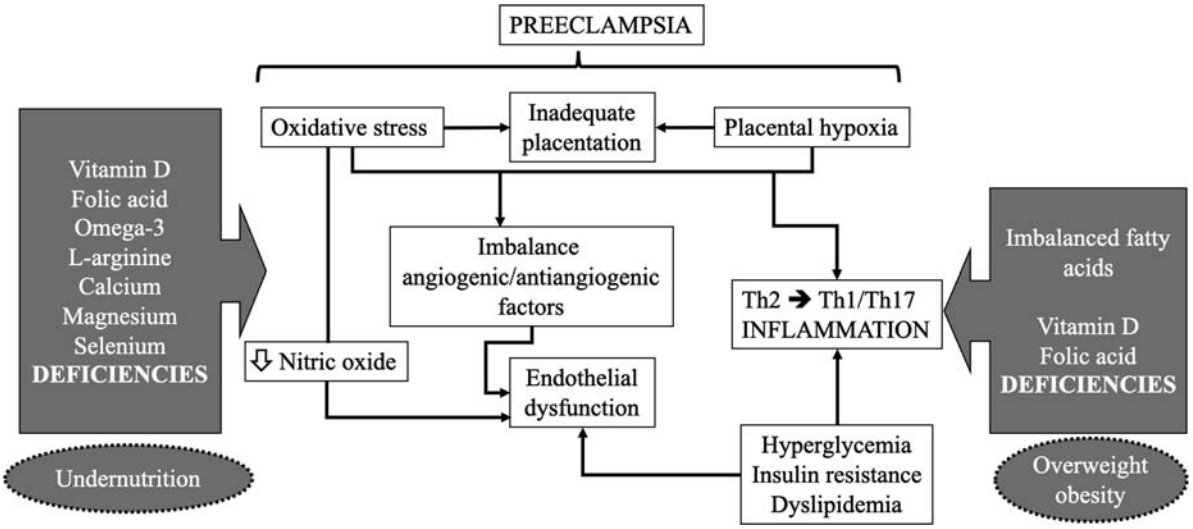


Fig. 1 Flowchart of key processes in the pathophysiology of preeclampsia. Inadequate placentation is associated with placental hypoxia and oxidative stress that leads to an imbalance between angiogenic and antiangiogenic factors. Misplacentation results in endothelial dysfunction and inflammation, that are responsible for clinical manifestations of preeclampsia. Some maternal nutritional deficiencies have been predominantly associated with processes of oxidative stress (vitamin D, folic acid, omega-3 fatty acids, L-arginine, calcium, magnesium, selenium), and with inflammation (imbalanced omega-3/omega-6 fatty acids, deficiencies in vitamin D and folic acid). Overweight and obesity usually coexist with hyperglycemia, insulin resistance and dyslipidemia, which also have been associated with the endothelial dysfunction and inflammation observed in preeclampsia.

The reduced placental perfusion in PE induces the release of pro-inflammatory cytokines (Hong et al., 2021). Excessive activation of peripheral and placental neutrophils and monocytes that produce high amounts of pro-inflammatory cytokines such as interleukin (IL)-1β, IL-6 and IL-8, has been observed in PE (Perez-Sepulveda et al., 2014), whereas IL-10, an *anti*-inflammatory cytokine, has been observed decreased in PE (Nath et al., 2020). Syncytiotrophoblast microparticles in maternal circulation (Martinez-Varea et al., 2014), damage-associated molecular patterns such as uric acid, high-mobility group box 1, IL-1 and cell-free fetal DNA resulting from placental damage (Nadeau-Vallee et al., 2016) have been proposed as "sterile" sources of inflammation. Moreover, maternal infections, via Toll-like receptors activation, have also been proposed as triggers of the immune response observed in PE (Laresgoiti-Servitje, 2013). Notably, *Helicobacter pylori* (Nourollahpour Shiadeh et al., 2019), periodontal, vaginal and urinary infections can induce the production of pro-inflammatory cytokines and other cellular mediators that can lead to endothelial dysfunction and PE (Nourollahpour Shiadeh et al., 2017). Recently, the placental microbiome has been shown to be important in maintaining a healthy pregnancy, and dysbiosis of local bacterial community has been proposed as source of inflammation that leads to HDPs (Beckers and Sones, 2020). Also, SARS-COV-2 during pregnancy has been associated with elevated odds of HDPs (Conde-Agudelo and Romero, 2021). Current research supports the screening and early treatment of common maternal infections as part of the prevention strategies for adverse pregnancy outcomes including PE.

Oxidative stress

Oxidative stress has been proposed as one of the main components in the etiology of PE (Godhamgaonkar et al., 2021). Reactive oxygen species (ROS) resulting from the reduction of molecular oxygen at the mitochondrial level are involved in cellular signaling and are increased during normal pregnancy (Pereira et al., 2015). ROS formation that exceeds antioxidant activity induces endothelial adhesion of leukocytes and platelets, and the release of cytokines and antiangiogenic factors (Matsubara et al., 2015). Therefore, ROS are tightly controlled by antioxidant host defenses (Sanchez-Aranguren et al., 2014) including enzymatic antioxidant molecules such as superoxide dismutase, glutathione peroxidase and catalase, and non-enzymatic antioxidants such as glutathione, carotenoids, flavonoids, vitamins C and E (Valko et al., 2007).

It has been reported that oxidative stress can inhibit endothelial nitric oxide synthase (NOS), decreasing the availability of nitric oxide (NO) (Guerby et al., 2021). NO is produced by endothelial cells from the amino acid L-arginine through the activation of NOS, producing L-citrulline as byproduct (Vallance and Chan, 2001), a process that depends on calcium (Lopez-Jaramillo et al., 2018). Upregulated by high estrogen concentrations during pregnancy, NO helps in the maintenance of vascular tone and increasing uterine blood flow, produces vasodilatation and anticoagulation (Matsubara et al., 2015). In PE, there is decreased NO availability due to its rapid degradation and consequent decreased endothelium-relaxation (Guerby et al., 2021).

Abnormal lipid metabolism

During normal pregnancy, the anabolic metabolism associated with increased maternal adipose tissue, insulin resistance, peripheral adipose tissue lipolysis and a trend to ketone utilization are mechanisms that spare glucose for fetal supply, expressed as an increase in maternal free fatty acids and lipoprotein triglyceride content (Barrett et al., 2014). In PE, this normal metabolic response is exaggerated, presenting with lower fasting and higher postprandial glucose concentrations, increased insulin resistance, hypertriglyceridemia and high circulating free fatty acids (von Versen-Hoeynck and Powers, 2007; Salzer et al., 2015; Hauth et al., 2011).

A higher degree of insulin resistance has been observed in early pregnancy in women who subsequently develop HDPs (Sierra-Laguado et al., 2007). Recently, the role of hyperinsulinemia in the immune response during PE has been reviewed. Both hyperglycemia and hyperinsulinemia possess the capacity to trigger a pro-inflammatory response; hyperinsulinemia could promote the differentiation of Th17 cells and the function of cytotoxic T cells while decreasing the formation of regulatory T cells, enhancing a pro-inflammatory environment and predisposing to PE (van Niekerk et al., 2021).

The association of hyperlipidemia with PE has been confirmed by two recent meta-analyses, one with African women (Tesfa et al., 2020b) and another with a multiethnic population (Spracklen et al., 2014). They found that total cholesterol, triglycerides, low-density lipoprotein (LDL) and very-low density lipoprotein (VLDL) were increased, and high-density lipoprotein (HDL) were decreased in pregnancies with PE compared with normal pregnancies. It has been proposed that accumulation of free fatty acids and triglycerides in endothelial cells, with the release of prostaglandins and nitric oxide, leads to the endothelial dysfunction observed in PE (Wojcik-Baszko et al., 2018). Moreover, it has been proposed that, similarly to atherosclerotic lesions, the formation of cholesterol crystals from oxidized LDL at the pre-eclamptic maternal-fetal interface are promoted through the activation of the Nod-like receptor protein (NLRP)3 inflammasome, triggering a pro-inflammatory response (Silva et al., 2020). As the NLRP3 pathway exacerbates oxidative stress and endothelial dysfunction, its inhibition has been seen as a possible therapeutic target in the treatment of PE (Nunes et al., 2021).

Adipose tissue secrets adipokines such as adiponectin, leptin and TNF-alpha, molecules that mediate metabolic processes, including lipid and glucose metabolism, inflammation, vascular homeostasis, endothelial function and placental angiogenesis (Daskalakis et al., 2020), all reasons why their role as mediators in the pathophysiology of PE has been proposed (Gutaj et al., 2020). Adiponectin is mainly synthetized by adipocytes, but is also produced in placenta and fetal tissues (Herrera and Ortega-Senovilla, 2014). Adiponectin possesses *anti*-apoptotic, antioxidant and *anti*-inflammatory properties (Pheiffer et al., 2021), increases insulin sensitivity, reduces serum free fatty acids, triglyceride and glucose concentrations, and inhibits monocyte adhesion (Roberts et al., 2011). However, studies exploring associations between adiponectin and PE show conflicting results with authors finding no association (Dalamaga et al., 2011), decreased (Mori et al., 2010) and increased adiponectin (Fasshauer et al., 2008; Salimi et al., 2014; Liu et al., 2012) in women with PE.

Whereas the use of adiponectin as a biomarker of PE requires further research, leptin has been consistently found increased in PE (Daskalakis et al., 2020). Leptin is a hormone produced by the placental trophoblast, maternal and fetal adipose tissue (Taylor et al., 2015). Through hypothalamic influence, leptin modulates satiety, decreases appetite and body weight; leptin also participates in angiogenesis and in the amino acid transport system in the placenta (Taylor et al., 2015). Its role as angiogenic factor, increasing placental perfusion and nutrient transport, and the inhibition of apoptosis of trophoblastic cells in PE, have led to the proposal that hyperleptinemia could be a compensatory mechanism for reduced placental dysfunction in PE and that it could be used as an early marker of the disease (Taylor et al., 2015; Miehle et al., 2012; Daskalakis et al., 2020). Novel adipokines chemerin, visfatin, resistin and apelin have been found involved in angiogenesis, regulation of blood pressure and triggering of the pro-inflammatory response and are under investigation regarding their role in PE (Estienne et al., 2019).

Essential fatty acids linoleic acid and α-linolenic acid (ALA) through desaturation and elongation by delta-6 and delta-5 desaturases, are converted into long chain polyunsaturated fatty acids (LC-PUFAs) omega-6 fatty acids (linolenic acid, arachidonic acid—AA), and omega-3 fatty acids (eicosapentaenoic acid—EPA and docosahexaenoic acid—DHA) respectively. Essential fatty acids are components of phospholipids and crucial for the function of cellular and subcellular membranes (Carvajal, 2014). It is known that whereas omega-3 LC-PUFAs (from green vegetables, oily seeds, fish and grass-fed animal products) have *anti*-inflammatory properties, omega-6 LC-PUFAs (from vegetable and hydrogenated oils and non-grass-fed animal products) have pro-inflammatory properties, and that these groups compete as substates for the same enzymes (Mauro et al., 2022). Most studies on bioactive lipids and PE have found an increase of n-6 and a decrease in n-3 free fatty acids in PE. PUFAs including AA, EPA, DHA and the platelet *anti*-aggregator and vasodilator prostacyclin (PGI_2) are reduced in PE, whereas the platelet aggregator and vasoconstrictor thromboxane A_2 is increased in women with PE (Das, 2015).

Supplementation with omega-3 fatty acids has the potential to modify the placental environment by improving nutrient transport capacity, modulating the inflammatory response, and affecting oxidative stress with increases or decreases, depending on dosage, timing and background diet, particularly in overweight and obese women (Rasool et al., 2021). A recent Cochrane review found a reduced risk of PE with omega-3 LC-PUFA (RR 0.84, 95% CI 0.69 to 1.01) after analyzing 20 trials, but the evidence was of low-quality (Middleton et al., 2018). A more recent meta-analysis found a significant effect of omega-3 on reducing the risk PE only when restricting analyses to studies in low-medium income countries (Kinshella et al., 2021a). To date, evidence is insufficient to advise supplementation with essential fatty acids for the prevention of PE.

Plasma volume dysregulation

During normal pregnancy, plasma volume steadily increases (de Haas et al., 2017). The increased plasma volume is favored by physiologically elevated concentrations of renin, angiotensin II and aldosterone (Birukov et al., 2019). There is also an increase in maternal cardiac output, vasodilation and decreased peripheral resistance (Gyselaers et al., 2018). These changes are driven by concomitant increase in nitric oxide production and vascular resistance to angiotensin II which leads to the physiologic fall of blood pressure during the first trimester (Armanini, 2012). Of interest, plasma volume expansion is lower in pregnancy-induced hypertension and PE (de Haas et al., 2017; Gyselaers et al., 2018). As a result of decreased plasma volume expansion, higher hemoglobin concentrations can be observed in preeclamptic patients (Ng et al., 2019), and it is known that the use of iron supplementation increases the erythrocyte mass (Means, 2020). Moreover, it has been suggested that excessive iron intake or high iron status may amplify oxidative stress and produce programmed cell death mediated by iron-dependent lipid peroxidation of cell membranes in the placenta (ferroptosis) (Ng et al., 2019; Erlandsson et al., 2021), contributing to the development of HDPs.

Early detection of HDPs

In an attempt for early detection of women at risk of HDPs, first-trimester algorithms including clinical and laboratory biomarkers have been developed. Recently, the Community-Level Interventions for Preeclampsia (CLIP) trials developed in India, Pakistan and Mozambique found that the visit-to-visit systolic and diastolic blood pressure variability was associated with increased odds of hypertension during pregnancy (Magee et al., 2021).

Elevated mean arterial pressure [MAP = diastolic + 1/3 (systolic–diastolic)] has also been reported to have a higher predicted value for the detection of PE among low-risk pregnant women than systolic or diastolic blood pressure (Sunjaya and Sunjaya, 2019). MAP was associated with gestational age, maternal age, weight, height, Afro-Caribbean racial origin, cigarette smoking, family history of PE, history of PE in the previous pregnancy, interpregnancy interval, chronic hypertension and diabetes mellitus in a population of European women (Wright et al., 2015). MAP was also positively associated with maternal weight-by-height classification, intake of multiple nutrient supplements, TNF-alpha, protein deficiency and the presence of the parasite hookworm, whereas negative associations with the presence of *Trichomonas vaginalis* and *Ascaris* in an indigenous population with high prevalence of infections, nutrient deficiencies and low plasma volume (González-Fernández et al., 2020). Therefore, MAP is able to capture a series of factors predisposing to hypertension during pregnancy, and represents a feasible, non-invasive, cost-effective screening tool for early detection of HDPs.

It has been proposed that the best biomarker combinations for the prediction of PE may include maternal risk factors, MAP, uterine artery pulsatility index and serum PlGF (Poon et al., 2019), which has reported detection rates of 90% for the prediction of early PE, with a 10% false-positive rate (Chaemsaithong et al., 2020) and that can be used to start prophylactic treatment with low-dose aspirin and calcium to reduce the risk of PE. Finally, the Preeclampsia Integrated Estimate of Risk (PIERS) are predictive models aiming to reduce adverse pregnancy outcomes (mortality or life-threatened conditions) by improving the diagnosis and initial management of PE for adverse pregnancy outcomes (von Dadelszen et al., 2009). The models include gestational age at presentation, clinical signs (blood pressure and dipstick proteinuria) and symptoms (chest pain/dyspnea, headache or visual disturbances, vaginal bleeding with abdominal pain) to be applied in low-resource settings, as well as biomarkers (oxygen saturation, platelet count, creatinine and aspartate transaminase concentrations) when available (von Dadelszen et al., 2011). PIERS models have been validated in well-resourced (von Dadelszen et al., 2011) and minimally resourced settings (Lim et al., 2015) demonstrating a good- [area under a receiver operating characteristic curve (AUC > 0.80)] to-excellent accuracy (AUC > 0.90) (von Dadelszen et al., 2009).

Association between nutritional factors and preeclampsia

Nutritional factors intervene in key processes of placental formation, endothelial function and inflammation, therefore, recent research has focused on the role of nutrition in the physiopathology of PE as possible therapeutic targets.

Overweight and obesity

Overweight and obesity are independently associated with HDPs (Vats et al., 2021; He et al., 2020). There is epidemiological evidence of the association of HDPs and high BMI, which is considered one of the major risks for PE in the United States (Ananth et al., 2013) and in low-middle income countries (Bilano et al., 2014; Tesfa et al., 2020a).

It has been proposed that pre-pregnancy obesity or excessive weight gain during pregnancy might trigger inflammation (Pare et al., 2014) and increase insulin resistance (Lopez-Jaramillo et al., 2018), leading to misplacentation, placental hypoxia and ischemia, the placental release of antiangiogenic factors and further endothelial dysfunction, characteristics of PE (Lopez-Jaramillo et al., 2018; Spradley et al., 2015). Also, as dysregulated adipokine production occurs in obesity where the role of adipokines in PE has been explored (Gutaj et al., 2020; Perez-Perez et al., 2018; Tilg and Moschen, 2006). Decreased adiponectin and

increased leptin were observed in obese women with PE, whereas normal weight pregnant women with PE showed increased adiponectin (Hendler et al., 2005).

Despite clear evidence of the increased risk of PE with overweight and obesity, interventions addressing maternal BMI during pregnancy show conflicting results. A meta-analysis of 23 trials looking at the effect of diet and exercise on the risk of PE found no significant association between gestational weight gain and the risk of HDPs or PE (Syngelaki et al., 2019), but another meta-analysis of 106 studies showed that moderate/intense exercise bot not exercise plus cointerventions, was able to reduce the odds of PE (OR: 0.62, 95% confidence interval: 0.52–0.75) compared with no exercise (Davenport et al., 2018).

Maternal undernutrition

There is evidence that maternal undernutrition, in particular low protein intake is associated with increased risk of PE (Bej et al., 2013). Particularly, L-arginine and L-citrulline, which are available in protein-rich foods, are usually deficient in low-resource settings (Weckman et al., 2019). In contexts where undernutrition and infection coexist, a common pathway for the development of pregnancy complications involving the dysregulation of the arginine-NO biogenesis has been proposed (Weckman et al., 2019). On the other hand, high-protein supplementation during pregnancy has been associated with decreased fetal growth (Mousa et al., 2019; Kramer and Kakuma, 2003) and is not recommended. It is also possible that higher rates of undernutrition and reported low intakes of calcium, iron, folate, magnesium, zinc and selenium found in developing countries, as well as the reduced risk of HDPs with balanced diets, might help to explain the higher prevalence of HDPs in resource-limited settings (Kinshella et al., 2021a), but more research is needed to identify the influence of maternal undernutrition on HDPs.

Vitamins

Vitamin A: Vitamin A comprises a group of fat-soluble compounds, where retinol and retinyl esters are obtained from animal sources, and provitamin A carotenoids and beta-carotene from plants (Baker et al., 2018). Retinoic acid, the active derivative of vitamin A, is a regulator of the human trophoblast during placentation (Huebner et al., 2018). Vitamin A also interacts with the innate and adaptive immune systems, acting as immune-modulator (Wiseman et al., 2016). Therefore, vitamin A deficiency may have an impact on placental function, tissue inflammation and structural development (Thoene et al., 2020). However, vitamin A supplementation is not recommended during pregnancy due to its potential teratogenic effect, except in regions where vitamin A deficiency is a public health concern (Bastos Maia et al., 2019). Carotenoids, beta-carotene and lycopene are antioxidants that have been shown to protect against free radical damage and were found in lower serum concentrations in preeclamptic women compared with normal pregnancies (Palan et al., 2001). A mechanistic pathway of vitamin A favoring the decidualization process by suppressing sFlt-1 has been proposed (Rajakumar et al., 2020) but large studies supporting the use of vitamin A supplementation for the prevention of HDPs or PE are lacking (Kinshella et al., 2021b).

Vitamin D: Vitamin D is a group of fat-soluble sterols, comprising ergocalciferol (D2) and cholecalciferol (D3), the latter being synthetized in humans in the skin, acquired in the diet (fish, fish liver oil, egg yolk, mushrooms) or through supplements. Liver 25-hydroxylation of D2 and D3 give rise to 25-hydroxyvitamin D or calcitriol, the main circulating form of vitamin D. Calcitriol is classically known for its role in bone metabolism, but parallel functions of vitamin D as immune modulator and cardiovascular protector have been recognized (Poniedzialek-Czajkowska and Mierzynski, 2021). In pregnancy, vitamin D regulates genes and the immune response associated with appropriate implantation, trophoblast invasion and implantation tolerance (Schroder-Heurich et al., 2020). There is evidence from cohort, case-cohort and nested case-control studies that higher maternal vitamin D concentrations are associated with lower risk of developing HDPs (Zhao et al., 2022), and a meta-analysis suggested an association between higher vitamin D and reduced risk of PE (Hyppönen et al., 2013). However, vitamin D combined with calcium has shown to increase the risk of preterm birth (Kinshella et al., 2021b). Current evidence on supplementation of vitamin D during pregnancy for the prevention of HDPs is inconclusive (Poniedziałek-Czajkowska and Mierzyński, 2021).

A possible protective role of vitamin D in PE could be related with a suppressive effect on the renin synthesis, the amelioration of insulin resistance, the improvement of endothelial vasodilatation, the inhibition of anticoagulant activity, and the modulation of macrophage activity and cytokine production (Colonese et al., 2015). It is worth noting that randomized controlled trials of appropriate design and quality to determine the usefulness of vitamin D supplementation for the prevention of HDPs are lacking to date (Kiely et al., 2020), and that the WHO does not currently recommend vitamin D supplementation to improve maternal or perinatal outcomes (World Health Organization, 2016).

Vitamin E: Vitamin E comprises lipid-soluble compounds - tocopherols and tocotrienols, from which alpha-tocopherol is the major form of vitamin E in plasma, whereas gamma-tocopherol is the most abundant in the diet, mainly acquired from soybean and corn oil (Gagne et al., 2009). Vitamin E is also available in the fat of meat, nuts, some cereals and green leafy vegetables. Vitamin E protects cell membrane phospholipid fatty acids from oxidation by reactive oxygen species, thus preventing oxidative stress (Rumbold et al., 2015a). The antioxidant activity of gamma-tocopherol is superior to alpha-tocopherol, which can have pro-oxidant activity that is prevented by vitamin C (Gagne et al., 2009). The interaction of vitamin C with vitamin E allows the recycling of oxidized vitamin E back to its useful form (Rumbold et al., 2015a). As placental oxidative stress and lipid peroxidation have been involved in endothelial cell dysfunction in HDPs (Salles et al., 2012), antioxidants could theoretically help for the prevention of the disease. However, a WHO multicenter randomized trial in women from developing countries at high risk of PE, showed no association of supplementation with vitamins C and E for the reduction of HDPs (Villar et al., 2009). An international trial of

antioxidants in the prevention of PE (INTAPP) found no reduction in PE or gestational hypertension with vitamin C and E supplementation but an increased risk of fetal loss and preterm pre-labor rupture of membranes (Xu et al., 2010). A meta-analysis of clinical trials for the use of vitamins E and C during pregnancy found no reduction of the risk of PE, but an increase in the risk of gestational hypertension (Conde-Agudelo et al., 2011). The latest Cochran review on vitamin C (Rumbold et al., 2015b) and vitamin E (Rumbold et al., 2015a) demonstrated no effect of supplementation of these vitamins alone or combined for the prevention of PE. Most recent research discriminating isoforms of vitamin E has found that vitamin E and C supplementation increased alpha-tocopherol but decreased gamma-tocopherol which could help to explain the failure of supplementation to prevent HDPs (Bilodeau et al., 2021).

Folic acid and vitamin B12: One-carbon metabolism is fundamental for cardiovascular health and for fetal development during pregnancy. It depends on folic acid and vitamin B12 to convert homocysteine (Hcy) to methionine, which is later converted to S-adenosyl methionine, the major donor for DNA, RNA, lipids and protein methylation (Mahmood et al., 2021). Hcy accumulation due to deficiencies in folic acid and B12 can lead to adverse pregnancy outcomes including PE (Dai et al., 2021). Proposed mechanisms for the association of Hcy and PE include endothelial dysfunction or vascular damage, given that Hcy decreases the production of nitric oxide, promotes platelet aggregation and promotes atherogenesis (Mahmood et al., 2021).

Folic acid supplementation in early pregnancy contributes to a healthy placental implantation, cell division, angiogenesis, trophoblast invasion and vascular relaxation, and supplementing after the first trimester, folic acid may protect against elevated Hcy (Bulloch et al., 2018), therefore preventing the development of HDPs. However, there is no conclusive agreement in the use of folic acid supplementation for the prevention of PE. In a systematic review and meta-analysis, two randomized controlled trials showed decreased risk of PE in women supplemented with folic acid compared with placebo, but pooled risk ratios of cohort studies did not show a significant effect (Hua et al., 2016). Another meta-analysis showed that, in observational studies, folic acid supplementation lowered the odds of PE compared with non-supplemented women (Bulloch et al., 2018). The FACT—Folic Acid Clinical Trial, a large multi-country randomized controlled trial, observed that 4–5.1 mg of folic acid supplementation beyond the first trimester did not prevent the developing of PE; results did not change when adjusting for confounders or dividing analyses by country (Wen et al., 2018). Most recent studies have found that PE was associated with lower folic acid and higher Hcy but not B12 (Olapeju et al., 2020; Liu et al., 2020; Serrano et al., 2018). Current literature supports the need for studies with sophisticated designs to confirm the protective effect of folic acid/B12 on the prevention of PE.

Minerals and trace elements

Sodium: Given that the renin-angiotensin-aldosterone system is influenced by dietary salt, and the secretion of aldosterone is stimulated by adrenocorticotropic hormone and potassium, it has been proposed that excessive sodium combined with low potassium intakes during pregnancy could adversely affect placental development (Birukov et al., 2019), favoring the development of PE. However, studies on sodium intake during pregnancy have generated opposing results. A Cochrane review including two trials found insufficient evidence to advise lower salt consumption during pregnancy for the prevention of PE (Duley et al., 2005). More recently, lower sodium intake during pregnancy was associated with lower risk of HDPs (Arvizu et al., 2020), however, the evidence for restricting sodium intake for reducing the risk of PE is insufficient and is not recommended (ACOG, 2020).

Calcium and Magnesium: Calcium supplementation is one of the few recommended approaches for the prevention of PE (World Health Organization, 2016), and magnesium sulfate is the keystone treatment in the prevention of seizures in preeclamptic women (Lambert et al., 2014), indicating the importance of these two micronutrients in the pathophysiology of PE.

It has been observed that traditional calcium-enriched foods were associated with low incidence of PE in indigenous populations, which led to finding an associations between low intakes of calcium, hypocalcemia or hypocalciuria with HDPs (Hofmeyr et al., 2018). A low calcium intake may increase blood pressure by stimulating the parathyroid hormone and/or the release of renin leading to vasoconstriction, or via increasing magnesium concentrations (Hofmeyr et al., 2018). However, experimental studies have shown that calcium supplementation may reverse endothelial cell activation induced by trophoblastic debris in preeclamptic placenta, and that the protective action of calcium on PE may be related its *anti*-inflammatory effect and its role as a cofactor of NOS (DeSousa et al., 2015).

Overall, studies converge in observing a protective effect of calcium supplementation on PE and gestational hypertension (Sun et al., 2019). The last Cochrane review concluded that calcium, at dosage $\geq$1 g/d, may reduce the risk of PE, particularly in those women with low-calcium diets and at high risk of PE (Hofmeyr et al., 2018), but there is not sufficient evidence that starting supplementation before or during early pregnancy adds to the benefits of calcium supplementation in the second half of pregnancy (Hofmeyr et al., 2019). However, supplementation with low doses of calcium (500 mg/d) showed a reduction in high blood pressure and the risk of PE, but the evidence is limited and needs to be confirmed by large/high-quality trials (Hofmeyr et al., 2018). Calcium for PE prevention is included in some country guidelines, but recommendations vary depending on the level of risk and/or baseline calcium intake; as reviewed by Scott et al. some guidelines recommend calcium supplementation in women with low calcium intake regardless of the risk of PE (The World Health Organization, Canada, Ireland, Tunisia and the European Society of Cardiology), others only in moderate to high risk women regardless their calcium intake (Society of Obstetric Medicine of Australia and New Zealand), or low intake and intermediate or high risk of PE (Basil, ISSHP and New Zealand) (Scott et al., 2020).

The use of magnesium for PE prevention has been less studied. A recent meta-analysis including seven randomized-controlled trials found that oral magnesium supplementation during pregnancy could reduce the risk of PE in women at high-risk of the disease (Yuan et al., 2021), but large randomized controlled trials are needed to validate these findings. The underlying mechanistic

explanation might be related with a decreased endothelial growth factor and decreased oxidative stress with experimental chronic use of magnesium, but further research is warranted (Yuan et al., 2021).

Iron: Iron requirements during pregnancy increase due to increase in maternal and fetal needs for erythropoiesis and for fetal growth, and iron deficiency is a known cause of adverse pregnancy outcomes, such as low birth weight, prematurity and intrauterine growth restriction where the benefits of supplementation during pregnancy have been well established (Means, 2020; Georgieff, 2020). Iron deficiency may impair placental function (Milman et al., 2016), but iron overload may produce oxidative stress and endothelial dysfunction (Shaji Geetha et al., 2020), both playing a role in the development of PE. A lower ferritin concentration was found correlated with lower prevalence of PE (Lao et al., 2000), and iron supplementation before 16 weeks' gestation has been associated with increased risk of *de novo* hypertension after 20 week's gestation (Jirakittidul et al., 2018). Therefore, caution should be made when supplementing pregnant women with iron, especially those with normal hemoglobin and/or with risk factors for developing HDPs.

Trace elements: Selenium is an essential trace element that makes part of enzymes involved in antioxidant defense mechanisms, redox regulation and thyroid hormone metabolism (Duntas, 2020). A meta-analysis of 16 studies from high and low-medium income countries previously showed increased risk of HDPs with lower serum selenium concentration and, in their sub-analysis of 3 randomized controlled trials where selenium supplementation (60–100 μg/d) took place, there was a reduced relative risk of PE (0.28, 95% CI: 0.09–0.84) (Xu et al., 2016). A more recent large study in a Norwegian population with sufficient dietary intakes and normal blood selenium concentrations, found no significant association between selenium intake from diet or supplements or blood selenium status with HDPs or PE (Holmquist et al., 2021).

A meta-analysis of 16 studies found that Asian women with PE had higher serum copper concentrations than controls (Song et al., 2017), whereas two meta-analyses including mainly Asian and African populations found lower serum zinc concentrations in preeclamptic women compared with controls (Ma et al., 2015; Tesfa et al., 2021). A recent meta-analysis of 5 studies found that urine iodine concentrations were lower in women with PE compared with controls, but no different risk of PE in women with iodine concentrations below the cut-off of 150 μg/L; authors acknowledged small sample sizes and heterogeneity of included studies (Businge et al., 2021). The evidence on trace elements (separately or included in multiple micronutrient mixtures) as therapeutic agents to prevent PE is insufficient and they are not currently recommended (World Health Organization, 2016).

Conclusion

Nutrients are involved in the pathophysiology of PE and addressing deficiencies and avoiding excesses of key macro- and micronutrients can have an impact on preventing the disease. Healthy diets and lifestyles aiming to fulfill pregnancy nutritional requirements, correcting undernutrition and controlling overweight should be promoted, starting before pregnancy and throughout gestation. The premise of no-harm should prevail over novel attractive approaches for supplementation. Of note, despite the evidence on some micronutrients decreasing the odds of HDPs, adverse effects of supplementation on fetal health need to be considered. This is the case for high-protein supplements and concomitant supplementation of calcium and vitamin D. In contrast, supplementation may have beneficial effects in women with specific nutritional needs which is the case for calcium, and possibly other minerals and vitamins. The potential benefits of supplementation also needs to discriminate among different types of HDPs, as well as different risk factors including overweight/obesity and macro/micronutrient deficiencies.

A comprehensive approach for pregnancy follow up that includes the evaluation of socio-demographic risk factors, MAP and ultrasound/biochemical biomarkers where available, could be complemented with counseling on balanced diet and exercise, in an effort to reduce the toll of HDPs, particularly PE, on maternal and fetal health.

References

Abu Samra, K., 2013. The eye and visual system in the preeclampsia/eclampsia syndrome: what to expect? Saudi J. Ophthalmol. 27, 51–53.
ACOG, 2013. Hypertension in pregnancy. Obstet. Gynecol. 122.
ACOG, 2020. ACOG practice bulletin. Gestational hypertension and preeclampsia. Obstet. Gynecol. 135 (6). Number 222.
Agarwal, I., Karumanchi, S.A., 2011. Preeclampsia and the anti-angiogenic state. Pregnancy Hypertens. 1, 17–21.
Ananth, C.V., Keyes, K.M., Wapner, R.J., 2013. Pre-eclampsia rates in the United States, 1980–2010: age-period-cohort analysis. BMJ 347, f6564.
Aneman, I., Pienaar, D., Suvakov, S., et al., 2020. Mechanisms of key innate immune cells in early- and late-onset preeclampsia. Front. Immunol. 11, 1864.
Armanini, D., 2012. Preeclampsia: the role of aldosterone in hypertension and inflammation. Hypertension 59, 1099–1100.
Arvizu, M., Bjerregaard, A.A., Madsen, M.T.B., et al., 2020. Sodium intake during pregnancy, but not other diet recommendations aimed at preventing cardiovascular disease, is positively related to risk of hypertensive disorders of pregnancy. J. Nutr. 150, 159–166.
Baker, B.C., Hayes, D.J., Jones, R.L., 2018. Effects of micronutrients on placental function: evidence from clinical studies to animal models. Reproduction 156, R69–R82.
Balsak, D., Togrul, C., Ekinci, C., et al., 2015. Severe pre-eclampsia complicated by hellp syndrome alterations in the structure of the umbilical cord (morphometric and immunohistochemical study). Biotechnol. Biotechnol. Equip. 29, 345–350.
Barrett, H.L., Dekker Nitert, M., Mcintyre, H.D., Callaway, L.K., 2014. Maternal lipids in pre-eclampsia: innocent bystander or culprit? Hypertens. Pregnancy 33, 508–523.
Bartsch, E., Medcalf, K.E., Park, A.L., Ray, J.G., 2016. Clinical risk factors for pre-eclampsia determined in early pregnancy: systematic review and meta-analysis of large cohort studies. BMJ 353, i1753.
Bastos Maia, S., Rolland Souza, A.S., Costa Caminha, M.F., et al., 2019. Vitamin A and pregnancy: a narrative review. Nutrients 11.
Beckers, K.F., Sones, J.L., 2020. Maternal microbiome and the hypertensive disorder of pregnancy, preeclampsia. Am. J. Physiol. Heart Circ. Physiol. 318, H1–H10.

Bej, P., Chhabra, P., Sharma, A.K., Guleria, K., 2013. Determination of risk factors for pre-eclampsia and eclampsia in a tertiary hospital of India: a case control study. J. Fam. Med. Prim. Care 2, 371–375.
Bigelow, C.A., Pereira, G.A., Warmsley, A., et al., 2014. Risk factors for new-onset late postpartum preeclampsia in women without a history of preeclampsia. Am. J. Obstet. Gynecol. 210, 338 e1-8.
Bilano, V.L., Ota, E., Ganchimeg, T., et al., 2014. Risk factors of pre-eclampsia/eclampsia and its adverse outcomes in low- and middle-income countries: a WHO secondary analysis. PLoS One 9, e91198.
Bilodeau, J.F., Gagne, A., Greffard, K., et al., 2021. The plasma antioxidant vitamin status of the INTAPP cohort examined: the unsuspected importance of beta-carotene and gamma-tocopherol in preeclampsia. Pregnancy Hypertens. 25, 213–218.
Birukov, A., Andersen, L.B., Herse, F., et al., 2019. Aldosterone, salt, and potassium intakes as predictors of pregnancy outcome, including preeclampsia. Hypertension 74, 391–398.
Broughton Pipkin, F., 1999. What is the place of genetics in the pathogenesis of pre-eclampsia? Biol. Neonate 76, 325–330.
Brown, M.A., Magee, L.A., Kenny, L.C., et al., 2018. The hypertensive disorders of pregnancy: ISSHP classification, diagnosis & management recommendations for international practice. Pregnancy Hypertens. 13, 291–310.
Bulloch, R.E., Lovell, A.L., Jordan, V.M.B., et al., 2018. Maternal folic acid supplementation for the prevention of preeclampsia: a systematic review and meta-analysis. Paediatr. Perinat. Epidemiol. 32, 346–357.
Businge, C.B., Usenbo, A., Longo-Mbenza, B., Kengne, A.P., 2021. Insufficient iodine nutrition status and the risk of pre-eclampsia: a systemic review and meta-analysis. BMJ Open 11, e043505.
Cao, Y., Liu, Y., Zhao, X., et al., 2020. Adherence to a dietary approaches to stop hypertension (DASH)-style diet in relation to preeclampsia: a case-control study. Sci. Rep. 10, 9078.
Carvajal, J.A., 2014. Docosahexaenoic acid supplementation early in pregnancy may prevent deep placentation disorders. BioMed Res. Int. 526895, 2014.
Cerdeira, A.S., Karumanchi, S.A., 2012. Angiogenic factors in preeclampsia and related disorders. Cold Spring Harb. Perspect. Med. 2.
Chaemsaithong, P., Sahota, D.S., Poon, L.C., 2020. First trimester preeclampsia screening and prediction. Am. J. Obstet. Gynecol. 226.
Chaiworapongsa, T., Chaemsaithong, P., Yeo, L., Romero, R., 2014. Pre-eclampsia part 1: current understanding of its pathophysiology. Nat. Rev. Nephrol. 10, 466–480.
Chau, A., Markley, J.C., Juang, J., Tsen, L.C., 2016. Cytokines in the perinatal period—part II. Int. J. Obstet. Anesth. 26, 48–58.
Cheng, M.H., Wang, P.H., 2009. Placentation abnormalities in the pathophysiology of preeclampsia. Expert Rev. Mol. Diagn. 9, 37–49.
Cicero, A.F., Degli Esposti, D., Immordino, V., et al., 2015. Independent determinants of maternal and fetal outcomes in a sample of pregnant outpatients with normal blood pressure, chronic hypertension, gestational hypertension, and preeclampsia. J. Clin. Hypertens. 17 (10).
Colonese, F., Lagana, A.S., Colonese, E., et al., 2015. The pleiotropic effects of vitamin D in gynaecological and obstetric diseases: an overview on a hot topic. BioMed Res. Int. 986281, 2015.
Conde-Agudelo, A., Romero, R., 2021. SARS-CoV-2 infection during pregnancy and risk of preeclampsia: a systematic review and meta-analysis. Am. J. Obstet. Gynecol. 226 (1).
Conde-Agudelo, A., Romero, R., Kusanovic, J.P., Hassan, S.S., 2011. Supplementation with vitamins C and E during pregnancy for the prevention of preeclampsia and other adverse maternal and perinatal outcomes: a systematic review and metaanalysis. Am. J. Obstet. Gynecol. 204, 503 e1–12.
Dai, C., Fei, Y., Li, J., et al., 2021. A novel review of homocysteine and pregnancy complications. BioMed Res. Int. 6652231, 2021.
Dalamaga, M., Srinivas, S.K., Elovitz, M.A., et al., 2011. Serum adiponectin and leptin in relation to risk for preeclampsia: results from a large case-control study. Metabolism 60, 1539–1544.
Das, U.N., 2015. Cytokines, angiogenic, and antiangiogenic factors and bioactive lipids in preeclampsia. Nutrition 31, 1083–1095.
Daskalakis, G., Bellos, I., Nikolakea, M., et al., 2020. The role of serum adipokine levels in preeclampsia: a systematic review. Metabolism: Clinical and Experimental 106, 154172.
Davenport, M.H., Ruchat, S.M., Poitras, V.J., et al., 2018. Prenatal exercise for the prevention of gestational diabetes mellitus and hypertensive disorders of pregnancy: a systematic review and meta-analysis. Br. J. Sports Med. 52, 1367–1375.
De Haas, S., Ghossein-Doha, C., Van Kuijk, S.M., et al., 2017. Physiological adaptation of maternal plasma volume during pregnancy: a systematic review and meta-analysis. Ultrasound Obstet. Gynecol. 49, 177–187.
Desousa, J., Tong, M., Wei, J., et al., 2015. The anti-inflammatory effect of calcium for preventing endothelial cell activation in preeclampsia. J. Hum. Hypertens. 30 (5).
Duley, L., Henderson-Smart, D., Meher, S., 2005. Altered dietary salt for preventing pre-eclampsia, and its complications. Cochrane Database Syst. Rev. CD005548.
Duntas, L.H., 2020. Selenium and at-risk pregnancy: challenges and controversies. Thyroid Res. 13, 16.
Erlandsson, L., Masoumi, Z., Hansson, L.R., Hansson, S.R., 2021. The roles of free iron, heme, haemoglobin, and the scavenger proteins haemopexin and alpha-1-microglobulin in preeclampsia and fetal growth restriction. J. Intern. Med. 290, 952–968.
Estienne, A., Bongrani, A., Reverchon, M., et al., 2019. Involvement of novel adipokines, chemerin, visfatin, resistin and apelin in reproductive functions in normal and pathological conditions in humans and animal models. Int. J. Mol. Sci. 20.
Fasshauer, M., Waldeyer, T., Seeger, J., et al., 2008. Circulating high-molecular-weight adiponectin is upregulated in preeclampsia and is related to insulin sensitivity and renal function. Eur. J. Endocrinol. 158, 197–201.
Fishel Bartal, M., Sibai, B.M., 2020. Eclampsia in the 21st century. Am. J. Obstet. Gynecol. 226.
Fong, F.M., Sahemey, M.K., Hamedi, G., et al., 2014. Maternal genotype and severe preeclampsia: a huge review. Am. J. Epidemiol. 180, 335–345.
Gagne, A., Wei, S.Q., Fraser, W.D., Julien, P., 2009. Absorption, transport, and bioavailability of vitamin E and its role in pregnant women. J. Obstet. Gynaecol. Can. 31, 210–217.
Garovic, V.D., Dechend, R., Easterling, T., et al., 2021. Hypertension in pregnancy: diagnosis, blood pressure goals, and pharmacotherapy: a scientific statement from the American Heart Association. Hypertension 79 (2). https://doi.org/10.1161/HYP.0000000000000208.
Georgieff, M.K., 2020. Iron deficiency in pregnancy. Am. J. Obstet. Gynecol. 223 (4), 516–524.
Ghaebi, M., Nouri, M., Ghasemzadeh, A., et al., 2017. Immune regulatory network in successful pregnancy and reproductive failures. Biomed. Pharmacother. 88, 61–73.
Godhamgaonkar, A.A., Sundrani, D.P., Joshi, S.R., 2021. Role of maternal nutrition and oxidative stress in placental telomere attrition in women with preeclampsia. Hypertens. Pregnancy 40, 63–74.
González-Fernández, D., Pons, E.D.C., Rueda, D., et al., 2020. Identification of high-risk pregnancies in a remote setting using ambulatory blood pressure: the mindi cohort. Front. Publ. Health 8.
Groten, T., Schleussner, E., Lehmann, T., et al., 2014. eNOSI4 and EPHX1 polymorphisms affect maternal susceptibility to preeclampsia: analysis of five polymorphisms predisposing to cardiovascular disease in 279 Caucasian and 241 African women. Arch. Gynecol. Obstet. 289, 581–593.
Guan, C., Zhao, F., Yang, Z., et al., 2022. A review of key cytokines based on gene polymorphism in the pathogenesis of pre-eclampsia. Am. J. Reprod. Immunol. 87, e13503.
Guerby, P., Tasta, O., Swiader, A., et al., 2021. Role of oxidative stress in the dysfunction of the placental endothelial nitric oxide synthase in preeclampsia. Redox Biol. 40, 101861.
Gutaj, P., Sibiak, R., Jankowski, M., et al., 2020. The role of the adipokines in the most common gestational complications. Int. J. Mol. Sci. 21.
Gyselaers, W., Vonck, S., Staelens, A.S., et al., 2018. Body fluid volume homeostasis is abnormal in pregnancies complicated with hypertension and/or poor fetal growth. PLoS One 13, e0206257.
Hauth, J.C., Clifton, R.G., Roberts, J.M., et al., 2011. Maternal insulin resistance and preeclampsia. Am. J. Obstet. Gynecol. 204, 327 e1–6.
He, X.-J., Dai, R.-X., Hu, C.-L., 2020. Maternal prepregnancy overweight and obesity and the risk of preeclampsia: a meta-analysis of cohort studies. Obes. Res. Clin. Pract. 14, 27–33.
Hegde, U.C., 1991. Immunomodulation of the mother during pregnancy. Med. Hypotheses 35, 159–164.

Hendler, I., Blackwell, S.C., Mehta, S.H., et al., 2005. The levels of leptin, adiponectin, and resistin in normal weight, overweight, and obese pregnant women with and without preeclampsia. Am. J. Obstet. Gynecol. 193, 979–983.
Herrera, E., Ortega-Senovilla, H., 2014. Lipid metabolism during pregnancy and its implications for fetal growth. Curr. Pharm. Biotechnol. 15, 24–31.
Hillesund, E.R., Øverby, N.C., Engel, S.M., et al., 2014. Associations of adherence to the New Nordic Diet with risk of preeclampsia and preterm delivery in the Norwegian mother and child cohort study (MOBA). Eur. J. Epidemiol. 29, 753–765.
Hofmeyr, G.J., Lawrie, T.A., Atallah, A.N., Torloni, M.R., 2018. Calcium supplementation during pregnancy for preventing hypertensive disorders and related problems. Cochrane Database Syst. Rev. 10, CD001059.
Hofmeyr, G.J., Manyame, S., Medley, N., Williams, M.J., 2019. Calcium supplementation commencing before or early in pregnancy, for preventing hypertensive disorders of pregnancy. Cochrane Database Syst. Rev. 9, CD011192.
Holmquist, E., Brantsaeter, A.L., Meltzer, H.M., et al., 2021. Maternal selenium intake and selenium status during pregnancy in relation to preeclampsia and pregnancy-induced hypertension in a large Norwegian pregnancy cohort study. Sci. Total Environ. 798, 149271.
Hong, K., Kim, S.H., Cha, D.H., Park, H.J., 2021. Defective uteroplacental vascular remodeling in preeclampsia: key molecular factors leading to long term cardiovascular disease. Int. J. Mol. Sci. 22.
Hua, X., Zhang, J., Guo, Y., et al., 2016. Effect of folic acid supplementation during pregnancy on gestational hypertension/preeclampsia: a systematic review and meta-analysis. Hypertens. Pregnancy 35, 447–460.
Huebner, H., Hartner, A., Rascher, W., et al., 2018. Expression and regulation of retinoic acid receptor responders in the human placenta. Reprod. Sci. 25, 1357–1370.
Hyppönen, E., Cavadino, A., Williams, D., et al., 2013. Vitamin D and pre-eclampsia: original data, systematic review and meta-analysis. Ann. Nutr. Metab. 63, 331–340.
Jardim, L.L., Rios, D.R., Perucci, L.O., et al., 2015. Is the imbalance between pro-angiogenic and anti-angiogenic factors associated with preeclampsia? Clin. Chim. Acta 447, 34–38.
Jensen, J.D., Saxe, H., Denver, S., 2015. Cost-effectiveness of a New Nordic diet as a strategy for health promotion. Int. J. Environ. Res. Publ. Health 12, 7370–7391.
Jirakittidul, P., Sirichotiyakul, S., Ruengorn, C., et al., 2018. Effect of iron supplementation during early pregnancy on the development of gestational hypertension and pre-eclampsia. Arch. Gynecol. Obstet. 298, 545–550.
Johnson, J.D., Louis, J.M., 2020. Does race or ethnicity play a role in the origin, pathophysiology, and outcomes of preeclampsia? An expert review of the literature. Am. J. Obstet. Gynecol. 226 (2S), S876–S885.
Khalil, A., Syngelaki, A., Maiz, N., et al., 2013. Maternal age and adverse pregnancy outcome: a cohort study. Ultrasound Obstet. Gynecol. 42, 634–643.
Kiely, M.E., Wagner, C.L., Roth, D.E., 2020. Vitamin D in pregnancy: where we are and where we should go. J. Steroid Biochem. Mol. Biol. 201, 105669.
Kinshella, M.W., Omar, S., Scherbinsky, K., et al., 2021a. Maternal dietary patterns and pregnancy hypertension in low- and middle-income countries: a systematic review and meta-analysis. Adv. Nutr. 12, 2387–2400.
Kinshella, M.W., Omar, S., Scherbinsky, K., et al., 2021b. Effects of maternal nutritional supplements and dietary interventions on placental complications: an umbrella review, meta-analysis and evidence map. Nutrients 13.
Koo, Y.J., Ryu, H.M., Yang, J.H., et al., 2012. Pregnancy outcomes according to increasing maternal age. Taiwan. J. Obstet. Gynecol. 51, 60–65.
Kramer, M.S., Kakuma, R., 2003. Energy and protein intake in pregnancy. Cochrane Database Syst. Rev. Cd000032.
Lambert, G., Brichant, J.F., Hartstein, G., et al., 2014. Preeclampsia: an update. Acta Anaesthesiol. Belg. 65, 137–149.
Lamminpaa, R., Vehvilainen-Julkunen, K., Gissler, M., Heinonen, S., 2012. Preeclampsia complicated by advanced maternal age: a registry-based study on primiparous women in Finland 1997–2008. BMC Pregnancy Childbirth 12, 47.
Lao, T.T., Tam, K.F., Chan, L.Y., 2000. Third trimester iron status and pregnancy outcome in non-anaemic women; pregnancy unfavourably affected by maternal iron excess. Hum. Reprod. 15, 1843–1848.
Laresgoiti-Servitje, E., 2013. A leading role for the immune system in the pathophysiology of preeclampsia. J. Leukoc. Biol. 94, 247–257.
Leppalahti, S., Gissler, M., Mentula, M., Heikinheimo, O., 2013. Is teenage pregnancy an obstetric risk in a welfare society? A population-based study in Finland, from 2006 to 2011. BMJ Open 3, e003225.
Levine, R.J., Lam, C., Qian, C., et al., 2006. Soluble endoglin and other circulating antiangiogenic factors in preeclampsia. N. Engl. J. Med. 355, 992–1005.
Li, Y., Zhu, M., Hu, R., Yan, W., 2015. The effects of gene polymorphisms in angiotensin II receptors on pregnancy-induced hypertension and preeclampsia: a systematic review and meta-analysis. Hypertens. Pregnancy 34, 241–260.
Lim, J., Cloete, G., Dunsmuir, D.T., et al., 2015. Usability and feasibility of PIERS on the move: an mHealth App for pre-eclampsia triage. JMIR Mhealth and Uhealth 3 e37-e37.
Lin, P.H., Aickin, M., Champagne, C., et al., 2003. Food group sources of nutrients in the dietary patterns of the DASH-sodium trial. J. Am. Diet Assoc. 103, 488–496.
Lipstein, H., Lee, C.C., Crupi, R.S., 2003. A current concept of eclampsia. Am. J. Emerg. Med. 21, 223–226.
Liu, Y., Zhu, L., Pan, Y., et al., 2012. Adiponectin levels in circulation and breast milk and mRNA expression in adipose tissue of preeclampsia women. Hypertens. Pregnancy 31, 40–49.
Liu, J., Gallagher, A.E., Carta, C.M., et al., 2014. Racial differences in gestational weight gain and pregnancy-related hypertension. Ann. Epidemiol. 24, 441–447.
Liu, C., Luo, D., Wang, Q., et al., 2020. Serum homocysteine and folate concentrations in early pregnancy and subsequent events of adverse pregnancy outcome: the Sichuan Homocysteine study. BMC Pregnancy Childbirth 20, 176.
Lopez-Jaramillo, P., Barajas, J., Rueda-Quijano, S.M., et al., 2018. Obesity and preeclampsia: common pathophysiological mechanisms. Front. Physiol. 9, 1838.
Ma, Y., Shen, X., Zhang, D., 2015. The relationship between serum zinc level and preeclampsia: a meta-analysis. Nutrients 7, 7806–7820.
Magee, L.A., Pels, A., Helewa, M., et al., 2014. Diagnosis, evaluation, and management of the hypertensive disorders of pregnancy: executive summary. J. Obstet. Gynaecol. Can. 36, 575–576.
Magee, L.A., Bone, J., Owasil, S.B., et al., 2021. Pregnancy outcomes and blood pressure visit-to-visit variability and level in three less-developed countries. Hypertension 77, 1714–1722.
Mahmood, S., Younas, H., Younus, A., Nathenial, S., 2021. A narrative review on the role of folate-mediated one-carbon metabolism and its associated gene polymorphisms in posing risk to preeclampsia. Clin. Exp. Hypertens. 43, 487–504.
Martinez-Varea, A., Pellicer, B., Perales-Marin, A., Pellicer, A., 2014. Relationship between maternal immunological response during pregnancy and onset of preeclampsia. J. Immunol. Res. 210241, 2014.
Matsubara, K., Higaki, T., Matsubara, Y., Nawa, A., 2015. Nitric oxide and reactive oxygen species in the pathogenesis of preeclampsia. Int. J. Mol. Sci. 16, 4600–4614.
Mauro, A.K., Rengarajan, A., Albright, C., Boeldt, D.S., 2022. Fatty acids in normal and pathological pregnancies. Mol. Cell. Endocrinol. 539, 111466.
Means, R.T., 2020. Iron deficiency and iron deficiency anemia: implications and impact in pregnancy, fetal development, and early childhood parameters. Nutrients 12, 447.
Middleton, P., Gomersall, J.C., Gould, J.F., et al., 2018. Omega-3 fatty acid addition during pregnancy. Cochrane Database Syst. Rev. 11, CD003402.
Miehle, K., Stepan, H., Fasshauer, M., 2012. Leptin, adiponectin and other adipokines in gestational diabetes mellitus and pre-eclampsia. Clin. Endocrinol. 76, 2–11.
Milman, N., Paszkowski, T., Cetin, I., Castelo-Branco, C., 2016. Supplementation during pregnancy: beliefs and science. Gynecol. Endocrinol. 32, 509–516.
Mori, T., Shinohara, K., Wakatsuki, A., et al., 2010. Adipocytokines and endothelial function in preeclamptic women. Hypertens. Res. 33, 250–254.
Mousa, A., Naqash, A., Lim, S., 2019. Macronutrient and micronutrient intake during pregnancy: an overview of recent evidence. Nutrients 11.
Nadeau-Vallee, M., Obari, D., Palacios, J., et al., 2016. Sterile inflammation and pregnancy complications: a review. Reproduction 152, R277–R292.
Nath, M.C., Cubro, H., Mccormick, D.J., et al., 2020. Preeclamptic women have decreased circulating IL-10 (interleukin-10) values at the time of preeclampsia diagnosis: systematic review and meta-analysis. Hypertension 76, 1817–1827.
Ng, S.W., Norwitz, S.G., Norwitz, E.R., 2019. The impact of iron overload and ferroptosis on reproductive disorders in humans: implications for preeclampsia. Int. J. Mol. Sci. 20.

Nourollahpour Shiadeh, M., Behboodi Moghadam, Z., Adam, I., et al., 2017. Human infectious diseases and risk of preeclampsia: an updated review of the literature. Infection 45, 589–600.
Nourollahpour Shiadeh, M., Riahi, S.M., Adam, I., et al., 2019. *Helicobacter pylori* infection and risk of preeclampsia: a systematic review and meta-analysis. J. Matern. Fetal Neonatal Med. 32, 324–331.
Nunes, P.R., Mattioli, S.V., Sandrim, V.C., 2021. NLRP3 activation and its relationship to endothelial dysfunction and oxidative stress: implications for preeclampsia and pharmacological interventions. Cells 10.
Ogge, G., Chaiworapongsa, T., Romero, R., et al., 2011. Placental lesions associated with maternal underperfusion are more frequent in early-onset than in late-onset preeclampsia. J. Perinat. Med. 39, 641–652.
Olapeju, B., Ahmed, S., Hong, X., et al., 2020. Maternal hypertensive disorders in pregnancy and postpartum plasma B vitamin and homocysteine profiles in a high-risk multiethnic U.S., population. J. Wom. Health 29, 1520–1529.
Palan, P.R., Mikhail, M.S., Romney, S.L., 2001. Placental and serum levels of carotenoids in preeclampsia. Obstet. Gynecol. 98, 459–462.
Pare, E., Parry, S., Mcelrath, T.F., et al., 2014. Clinical risk factors for preeclampsia in the 21st century. Obstet. Gynecol. 124, 763–770.
Pereira, R.D., De Long, N.E., Wang, R.C., et al., 2015. Angiogenesis in the placenta: the role of reactive oxygen species signaling. BioMed Res. Int. 814543, 2015.
Perez-Perez, A., Toro, A., Vilarino-Garcia, T., et al., 2018. Leptin action in normal and pathological pregnancies. J. Cell Mol. Med. 22, 716–727.
Perez-Sepulveda, A., Torres, M.J., Khoury, M., Illanes, S.E., 2014. Innate immune system and preeclampsia. Front. Immunol. 5, 244.
Petry, C.J., Beardsall, K., Dunger, D.B., 2014. The potential impact of the fetal genotype on maternal blood pressure during pregnancy. J. Hypertens. 32, 1553–1561. Discussion 1561.
Pheiffer, C., Dias, S., Jack, B., et al., 2021. Adiponectin as a potential biomarker for pregnancy disorders. Int. J. Mol. Sci. 22.
Poniedzialek-Czajkowska, E., Mierzynski, R., 2021. Could vitamin D be effective in prevention of preeclampsia? Nutrients 13.
Poon, L.C., Shennan, A., Hyett, J.A., et al., 2019. The international federation of gynecology and obstetrics (FIGO) initiative on pre-eclampsia: a pragmatic guide for first-trimester screening and prevention. Int. J. Gynaecol. Obstet. 145 (Suppl. 1), 1–33.
Qi, H.P., Fraser, W.D., Luo, Z.C., et al., 2013. Endothelial nitric oxide synthase gene polymorphisms and risk of preeclampsia. Am. J. Perinatol. 30, 795–804.
Rajakumar, A., Kane, M.A., Yu, J., et al., 2020. Aberrant retinoic acid production in the decidua: implications for pre-eclampsia. J. Obstet. Gynaecol. Res. 46, 1007–1016.
Rana, S., Lemoine, E., Granger, J.P., Karumanchi, S.A., 2019. Preeclampsia: pathophysiology, challenges, and perspectives. Circ. Res. 124, 1094–1112.
Rasool, A., Alvarado-Flores, F., O'tierney-Ginn, P., 2021. Placental impact of dietary supplements: more than micronutrients. Clin. Ther. 43, 226–245.
Redman, C.W., 2011. Preeclampsia: a multi-stress disorder. La Revue de Médecine Interne/fondeé par la Societe Nationale Francaise de Médecine Interne 32 (Suppl. 1), S41–S44.
Redman, C.W.G., Staff, A.C., Roberts, J.M., 2020. Syncytiotrophoblast stress in preeclampsia: the convergence point for multiple pathways. Am. J. Obstet. Gynecol. 226 (2S), S907–S927.
Roberts, J.M., Bodnar, L.M., Patrick, T.E., Powers, R.W., 2011. The role of obesity in preeclampsia. Pregnancy Hypertens. 1, 6–16.
Roberts, J.M., Rich-Edwards, J.W., Mcelrath, T.F., et al., 2021. Subtypes of preeclampsia: recognition and determining clinical usefulness. Hypertension 77, 1430–1441.
Rumbold, A., Ota, E., Hori, H., et al., 2015a. Vitamin E supplementation in pregnancy. Cochrane Database Syst. Rev. CD004069.
Rumbold, A., Ota, E., Nagata, C., et al., 2015b. Vitamin C supplementation in pregnancy. Cochrane Database Syst. Rev. CD004072.
Saito, S., Nakashima, A., Shima, T., Ito, M., 2010. Th1/Th2/Th17 and regulatory T-cell paradigm in pregnancy. Am. J. Reprod. Immunol. 63, 601–610.
Salimi, S., Farajian-Mashhadi, F., Naghavi, A., et al., 2014. Different profile of serum leptin between early onset and late onset preeclampsia. Dis. Markers 628476, 2014.
Salles, A.M., Galvao, T.F., Silva, M.T., et al., 2012. Antioxidants for preventing preeclampsia: a systematic review. Sci. World J. 243476, 2012.
Salzer, L., Tenenbaum-Gavish, K., Hod, M., 2015. Metabolic disorder of pregnancy (understanding pathophysiology of diabetes and preeclampsia). Best Pract. Res. Clin. Obstet. Gynaecol. 29, 328–338.
Sanchez-Aranguren, L.C., Prada, C.E., Riano-Medina, C.E., Lopez, M., 2014. Endothelial dysfunction and preeclampsia: role of oxidative stress. Front. Physiol. 5, 372.
Schimmel, M.S., Bromiker, R., Hammerman, C., et al., 2015. The effects of maternal age and parity on maternal and neonatal outcome. Arch. Gynecol. Obstet. 291, 793–798.
Schroder-Heurich, B., Springer, C.J.P., Von Versen-Hoynck, F., 2020. Vitamin D effects on the immune system from periconception through pregnancy. Nutrients 12.
Scott, G., Gillon, T.E., Pels, A., et al., 2020. Guidelines-similarities and dissimilarities: a systematic review of international clinical practice guidelines for pregnancy hypertension. Am. J. Obstet. Gynecol. 226 (2S), S1222–S1236.
Serrano, N.C., Quintero-Lesmes, D.C., Becerra-Bayona, S., et al., 2018. Association of pre-eclampsia risk with maternal levels of folate, homocysteine and vitamin B12 in Colombia: a case-control study. PLoS One 13, e0208137.
Shaji Geetha, N., Bobby, Z., Dorairajan, G., Jacob, S.E., 2020. Increased hepcidin levels in preeclampsia: a protective mechanism against iron overload mediated oxidative stress? J. Matern. Fetal Neonatal Med. 1–6. The Official Journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstet.
Shen, J.J., Tymkow, C., Macmullen, N., 2005. Disparities in maternal outcomes among four ethnic populations. Ethn. Dis. 15, 492–497.
Sierra-Laguado, J., Garcia, R.G., Celedon, J., et al., 2007. Determination of insulin resistance using the homeostatic model assessment (homa) and its relation with the risk of developing pregnancy-induced hypertension. Am. J. Hypertens. 20, 437–442.
Silva, G.B., Gierman, L.M., Rakner, J.J., et al., 2020. Cholesterol crystals and NLRP3 mediated inflammation in the uterine wall decidua in normal and preeclamptic pregnancies. Front. Immunol. 11, 564712.
Song, X., Li, B., Li, Z., et al., 2017. High serum copper level is associated with an increased risk of preeclampsia in asians: a meta-analysis. Nutr. Res. 39, 14–24.
Spracklen, C.N., Smith, C.J., Saftlas, A.F., et al., 2014. Maternal hyperlipidemia and the risk of preeclampsia: a meta-analysis. Am. J. Epidemiol. 180, 346–358.
Spradley, F.T., Palei, A.C., Granger, J.P., 2015. Immune mechanisms linking obesity and preeclampsia. Biomolecules 5, 3142–3176.
Staff, A.C., 2019. The two-stage placental model of preeclampsia: an update. J. Reprod. Immunol. 1–10, 134-135.
Sun, X., Li, H., He, X., et al., 2019. The association between calcium supplement and preeclampsia and gestational hypertension: a systematic review and meta-analysis of randomized trials. Hypertens. Pregnancy 38, 129–139.
Sunjaya, A.F., Sunjaya, A.P., 2019. Evaluation of serum biomarkers and other diagnostic modalities for early diagnosis of preeclampsia. J. Fam. Reprod. Health 13, 56–69.
Suresh, S.C., Rana, S., 2021. Real-world use of biomarkers in management of hypertension during pregnancy: adjunct to clinical care. Hypertension 77, 472–474 (Dallas, Tex.: 1979).
Sykes, L., Macintyre, D.A., Yap, X.J., et al., 2012. The Th1:Th2 dichotomy of pregnancy and preterm labour. Mediat. Inflamm. 967629, 2012.
Syngelaki, A., Sequeira Campos, M., Roberge, S., et al., 2019. Diet and exercise for preeclampsia prevention in overweight and obese pregnant women: systematic review and meta-analysis. J. Matern. Fetal Neonatal Med. 32, 3495–3501.
Taylor, B.D., Ness, R.B., Olsen, J., et al., 2015. Serum leptin measured in early pregnancy is higher in women with preeclampsia compared with normotensive pregnant women. Hypertension 65, 594–599.
Tesfa, E., Nibret, E., Gizaw, S.T., et al., 2020a. Prevalence and determinants of hypertensive disorders of pregnancy in Ethiopia: a systematic review and meta-analysis. PLoS One 15, e0239048.
Tesfa, E., Nibret, E., Munshea, A., 2020b. Maternal lipid profile and risk of pre-eclampsia in African pregnant women: a systematic review and meta-analysis. PLoS One 15, e0243538.
Tesfa, E., Nibret, E., Munshea, A., 2021. Maternal serum zinc level and pre-eclampsia risk in African women: a systematic review and meta-analysis. Biol. Trace Elem. Res. 199, 4564–4571.
Thoene, M., Van Ormer, M., Yuil-Valdes, A., et al., 2020. Fat-soluble nutrients and omega-3 fatty acids as modifiable factors influencing preterm birth risk. Placenta 98, 38–42.

Thornton, C.E., Makris, A., Ogle, R.F., et al., 2010. Role of proteinuria in defining pre-eclampsia: clinical outcomes for women and babies. Clin. Exp. Pharmacol. Physiol. 37, 466–470.
Tilg, H., Moschen, A.R., 2006. Adipocytokines: mediators linking adipose tissue, inflammation and immunity. Nat. Rev. Immunol. 6, 772–783.
Tranquilli, A.L., Landi, B., 2010. The origin of pre-eclampsia: from decidual "hyperoxia" to late hypoxia. Med. Hypotheses 75, 38–46.
Tranquilli, A.L., Dekker, G., Magee, L., et al., 2014. The classification, diagnosis and management of the hypertensive disorders of pregnancy: a revised statement from the ISSHP. Pregnancy Hypertens. 4, 97–104.
Valko, M., Leibfritz, D., Moncol, J., et al., 2007. Free radicals and antioxidants in normal physiological functions and human disease. Int. J. Biochem. Cell Biol. 39, 44–84.
Vallance, P., Chan, N., 2001. Endothelial function and nitric oxide: clinical relevance. Heart 85, 342–350.
Van Niekerk, G., Christowitz, C., Engelbrecht, A.M., 2021. Insulin-mediated immune dysfunction in the development of preeclampsia. J. Mol. Med. 99, 889–897.
Vats, H., Saxena, R., Sachdeva, M.P., et al., 2021. Impact of maternal pre-pregnancy body mass index on maternal, fetal and neonatal adverse outcomes in the worldwide populations: a systematic review and meta-analysis. Obes. Res. Clin. Pract. 15, 536–545.
Verlohren, S., Droge, L.A., 2020. The diagnostic value of angiogenic and antiangiogenic factors in differential diagnosis of preeclampsia. Am. J. Obstet. Gynecol. 226 (2S), S1048–S1058.
Vigil-De Gracia, P., Rojas-Suarez, J., Ramos, E., et al., 2015. Incidence of eclampsia with hellp syndrome and associated mortality in Latin America. Int. J. Gynaecol. Obstet. 129, 219–222.
Vilchez, G., Londra, L., Hoyos, L.R., et al., 2015. Intrapartum mean platelet volume is not a useful predictor of new-onset delayed postpartum pre-eclampsia. Int. J. Gynaecol. Obstet. 131 (1), 59–62.
Villani, A., Sultana, J., Doecke, J., Mantzioris, E., 2019. Differences in the interpretation of a modernized Mediterranean diet prescribed in intervention studies for the management of type 2 diabetes: how closely does this align with a traditional Mediterranean diet? Eur. J. Nutr. 58, 1369–1380.
Villar, J., Purwar, M., Merialdi, M., et al., 2009. World Health Organisation multicentre randomised trial of supplementation with vitamins C and E among pregnant women at high risk for pre-eclampsia in populations of low nutritional status from developing countries. BJOG 116, 780–788.
Von Dadelszen, P., Menzies, J.M., Payne, B., Magee, L.A., 2009. Predicting adverse outcomes in women with severe pre-eclampsia. Semin. Perinatol. 33, 152–157.
Von Dadelszen, P., Payne, B., Li, J., et al., 2011. Prediction of adverse maternal outcomes in pre-eclampsia: development and validation of the fullPIERS model. Lancet 377, 219–227.
Von Versen-Hoeynck, F.M., Powers, R.W., 2007. Maternal-fetal metabolism in normal pregnancy and preeclampsia. Front. Biosci. 12, 2457–2470.
Wang, X.M., Wu, H.Y., Qiu, X.J., 2013. Methylenetetrahydrofolate reductase (MTHFR) gene c677t polymorphism and risk of preeclampsia: an updated meta-analysis based on 51 studies. Arch. Med. Res. 44, 159–168.
Wang, W., Xie, X., Yuan, T., et al., 2021. Epidemiological trends of maternal hypertensive disorders of pregnancy at the global, regional, and national levels: a population-based study. BMC Pregnancy Childbirth 21, 364.
Weckman, A.M., Ngai, M., Wright, J., et al., 2019. The impact of infection in pregnancy on placental vascular development and adverse birth outcomes. Front. Microbiol. 10, 1924.
Wen, S.W., White, R.R., Rybak, N., et al., 2018. Effect of high dose folic acid supplementation in pregnancy on pre-eclampsia (FACT): double blind, phase III, randomised controlled, international, multicentre trial. BMJ 362, k3478.
Wiertsema, C.J., Mensink-Bout, S.M., Duijts, L., et al., 2021. Associations of dash diet in pregnancy with blood pressure patterns, placental hemodynamics, and gestational hypertensive disorders. J. Am. Heart Assoc. 10, e017503.
Wiseman, E.M., Bar-El Dadon, S., Reifen, R., 2016. The vicious cycle of vitamin A deficiency: a review. Crit. Rev. Food Sci. Nutr. 57 (17), 3703–3714.
Wojcik-Baszko, D., Charkiewicz, K., Laudanski, P., 2018. Role of dyslipidemia in preeclampsia-a review of lipidomic analysis of blood, placenta, syncytiotrophoblast microvesicles and umbilical cord artery from women with preeclampsia. Prostag. Other Lipid Mediat. 139, 19–23.
World Health Organization, 2016. In: Who Recommendations on Antenatal Care for a Positive Pregnancy Experience. World Health Organization, Geneva.
Wright, A., Wright, D., Ispas, C.A., et al., 2015. Mean arterial pressure in the three trimesters of pregnancy: effects of maternal characteristics and medical history. Ultrasound Obstet. Gynecol. 45, 698–706.
Xu, H., Perez-Cuevas, R., Xiong, X., et al., 2010. An international trial of antioxidants in the prevention of preeclampsia (INTAPP). Am. J. Obstet. Gynecol. 202, 239 e1–239 e10.
Xu, M., Guo, D., Gu, H., et al., 2016. Selenium and preeclampsia: a systematic review and meta-analysis. Biol. Trace Elem. Res. 171, 283–292.
Yuan, J., Yu, Y., Zhu, T., et al., 2021. Oral magnesium supplementation for the prevention of preeclampsia: a meta-analysis or randomized controlled trials. Biol. Trace Elem. Res. 200 (8), 3572–3581.
Zeisler, H., Llurba, E., Chantraine, F., et al., 2016. Predictive value of the sFlt-1:PlGF ratio in women with suspected preeclampsia. N. Engl. J. Med. 374, 13–22.
Zhang, Y., Xia, M., Weng, S., et al., 2021. Effect of mediterranean diet for pregnant women: a meta-analysis of randomized controlled trials. J. Matern. Fetal Neonatal Med. 1–11. The Official Journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstet.
Zhao, R., Zhou, L., Wang, S., et al., 2022. Association between maternal vitamin D levels and risk of adverse pregnancy outcomes: a systematic review and dose-response meta-analysis. Food Funct. 13, 14–37.

Pregnancy: Prevention of neural tube defects

Helene McNulty, Aoife Caffrey, and Kristina Pentieva, Nutrition Innovation Centre for Food and Health (NICHE), School of Biomedical Sciences, Ulster University, Coleraine, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A.M. Molloy, P.N. Kirke, J.L. Mills, Pregnancy: Prevention of Neural Tube Defects, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 81–89, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00234-8.

Introduction 199
Epidemiology 199
Genetic and environmental factors 199
Maternal folate status ; 200
Intervention studies with folic acid in the prevention of NTDs 200
Observational studies of maternal folate status in relation to NTD risk 200
Maternal vitamin B_{12} status 200
Other maternal risk factors for NTD 202
Specific nutrients 202
Obesity and diabetes 202
Other risk factors 203
Biological mechanisms in the etiology and prevention of NTDs 203
Role of folate, vitamin B_{12} and one-carbon metabolism 203
Folate-related genetic risk factors and risk of NTDs 203
Prevention of NTD 205
Nutritional considerations 205
Interventions to optimize folate status 206
Folic acid supplementation 206
Folic acid fortification 206
Effects of folic acid policy globally 206
Effects on NTD prevalence 207
Effects on other health outcomes 208
Emerging folic acid policy: risk-benefit assessment 208
Recommendations to prevent first occurrence and recurrence of NTDs 209
Conclusions 209
References 209
Further reading 210

Key points

This chapter aims to

- Describe the terminology and general epidemiology pertaining to neural tube defects (NTDs).
- Provide knowledge and underpinning scientific evidence in relation to
  - The genetic and environmental factors implicated in NTD, and particularly the major role played by low maternal folate status;
  - The contribution of other risk factors, including, low vitamin B12 status, obesity, diabetes and certain drug treatments;
  - The biological mechanisms in the etiology and prevention of NTDs.
- Review the current evidence in relation to preventing first and recurrence of NTD, including
  - Nutrition considerations;
  - Effectiveness of folic acid interventions to optimize folate status via supplementation and fortification approaches.
- Provide an understanding of current folic acid policy globally and
  - Impacts on NTD prevalence;
  - Impacts on other health outcomes throughout the lifecycle;
  - Risk-benefit assessment for emerging policy.
- Describe current recommendations for the prevention of first occurrence and recurrence of NTDs.
- Generate well-reasoned conclusions in relation to policy to prevent NTDs worldwide through effective interventions.

Glossary

Anencephaly Failure of the section of the neural tube to close in the area of the brain, causing the absence of a large part of the brain and skull. Anencephaly is uniformly fatal

Craniorachischisis A severe form of neural tube defect where the area in which the neural tube fails to close extends from the head down into the spine

Encephalocele A form of neural tube defect where there is failure of ossification of a portion of the skull. A sac like protrusion occurs which may contain only meninges (encephalocele) or brain tissue and meninges (meningoencephalocele)

Iniencephaly A severe form of neural tube defect in which the head is flexed in an upward gaze and both anencephaly and spinal defects are common

Meningocele A form of spina bifida. There is a failure of the neural tube to close in the area of the spine causing a defect in the vertebral arches through which cerebrospinal fluid filled meninges, but not neural tissue, protrude

Myelomeningocele A serious form of spina bifida. There is a failure of the neural tube to close in the area of the spine causing a defect in the vertebral arches through which cerebrospinal fluid filled meninges and neural tissue protrude

Neural Tube Defects Neural tube defects (NTDs) are birth defects of the brain and spinal cord which arise when the neural tube does not close properly during the first few weeks post-conception. The two most common forms of NTDs are spina bifida and anencephaly; other forms listed above are much rarer

Spina Bifida A main type of neural tube defect. Affects the spine and can happen anywhere along the spine where the neural tube does not completely close. Physical and intellectual disabilities can arise, ranging from mild to severe

Introduction

Neural tube defects (NTDs) are the most common anomalies of the central nervous system, arising when the neural tube fails to close properly in the developing fetus, resulting in death or lifelong disability. The main categories of NTDs are spina bifida (a spinal cord defect) and anencephaly (a brain defect). Normally, the neural tube closes to form the brain and spinal cord within the first 28 days after conception. The type of lesion occurring in NTD depends on the portion of the neural tube that fails to close. Failure of the cranial portion to close causes anencephaly, a uniformly fatal brain defect, whereas failure of the more caudal neural tube to close causes myelomeningocele or meningocele (two forms of spina bifida). The severity of the defect depends on the location; in general, the higher the defect, the more severe the disability. Less common forms of NTD include encephalocele, iniencephaly, and craniorachischisis. Combinations of these lesions may be present in severely affected cases. Most children with NTDs who survive beyond birth will have serious disabilities.

Epidemiology

It is estimated that 322,000 or more pregnancies worldwide are affected by spina bifida and anencephaly annually, at an average prevalence of about 20 cases per 10,000 births, corresponding to one in every 500 births globally (Blencowe et al., 2018). In low- and middle-income countries, the prevalence is estimated to exceed one in every 100 births. Furthermore, an estimated 60,000 NTD-affected pregnancies are electively terminated annually (after prenatal diagnosis), and another 60,000 result in stillbirths (Blencowe et al., 2018). In the past 50 years the prevalence of NTD has dropped significantly in high-income countries, particularly in regions that traditionally had high rates. The reason for this decline (which had actually started before the implementation of mandatory folic acid-fortification of foods) is likely due to improved diets (particularly with respect to folate intakes), prenatal diagnosis and access to termination. Because the neural tube closes very early in pregnancy (between the 3rd and 4th weeks post-conception), before many women are aware that they are pregnant, NTDs have been problematic to study and to prevent. In fact, prevention essentially requires that interventions be initiated before conception. As a result, public health policy in this area has proved to be challenging, especially considering that an estimated 50% of all pregnancies are unplanned.

Genetic and environmental factors

Both genetic and environmental factors are known to be involved in the etiology of NTDs. Evidence for a genetic component comes from studies showing racial differences in prevalence, recurrence risks of approximately 10-times the general occurrence risk, higher rates in females and well recognized mendelian syndromes that include NTDs. Over 240 laboratory mouse strains exhibiting NTDs have been generated by single or multiple gene manipulations, further substantiating a strong genetic component. Nevertheless, the

majority of NTDs are isolated, consistent with a low penetrance etiology exhibiting unclear and probably variable genetic and environmental influences.

Marked variations in prevalence over time, between areas and across social classes, provide evidence of the importance of environmental factors in NTD. In particular, evidence for nutritional factors has accumulated since the 1960s. The observation that women of low socio–economic status were at increased risk of having a child with an NTD stimulated Smithells and coworkers several decades ago to examine diet as a risk factor and to measure vitamin levels in pregnant women (Smithells et al., 1983). Their findings, that women carrying affected fetuses had significantly lower red blood cell folate concentrations, stimulated many subsequent case-control studies and later clinical trials. The totality of evidence from these intervention and case control studies showed that periconceptional use of vitamin supplements containing folic acid reduced NTD rates by 35–71%. Folic acid is the synthetic form of the vitamin known as folate in its natural forms.

Maternal folate status Table 1; Table 2

Intervention studies with folic acid in the prevention of NTDs

The definitive evidence for a protective effect of maternal folic acid supplementation in the prevention of recurrence and first occurrence of NTDs was established by two notable randomized controlled trials published in the early 1990s. The Medical Research Council trial (MRC, 1991) randomized women with a previous history of an NTD-affected fetus to one of four intervention groups: folic acid, other vitamins, both, or neither. The findings showed that the NTD recurrence rate in the two groups which received folic acid was significantly lower than the rate in the groups that did not receive folic acid (odds ratio 0.28; confidence interval 0.12–0.71), suggesting that 72% of recurrences could be prevented by folic acid. It is important to note that 28% of NTD recurrences were not prevented by the high folic acid dose (4 mg) administered daily in the trial; this dose is more than 10 times higher than dietary folate recommendations. The multivitamin combination without folic acid had no protective effect. Vitamin B_{12}, closely related to folate within one-carbon metabolism, was not included in the multivitamin tablets. The Hungarian trial investigating the effect of folic acid on first occurrence of NTD (in women with no previous history), published one year after the MRC trial, found no NTD occurrences in the treatment group who were given 800 μg per day folic acid plus other vitamins including 4 μg vitamin B_{12} (Czeizel and Dudás, 1992), whereas 6 NTD cases occurred in the placebo arm receiving a multimineral combination without folic acid or vitamin B_{12}. The combined evidence from both trials proves a specific protective effect of folic acid against first occurrence and recurrence of NTDs, and points to a potential additional role of vitamin B_{12} in NTD prevention, as will be discussed later. The major intervention trials of periconceptional folic acid supplementation and NTD risk are shown in **Table 1**.

Observational studies of maternal folate status in relation to NTD risk

Observational studies generally demonstrate that women who have had pregnancies affected by NTD have lower folate concentrations (both in plasma and red blood cells) compared to nonaffected mothers. In a large prospective study in Ireland (where rates of NTD are among the highest in the world), a woman's risk of having a child with an NTD was shown to be strongly associated with pregnancy concentrations of red blood cell folate in a continuous dose-response inverse relationship (Daly et al., 1995). Specifically, mothers with red blood cell folate concentrations below 150 ng/mL (340 nmol/L) in early pregnancy had more than eight times higher risk of having an NTD-affected child than those whose folate concentrations were over 400 ng/mL (906 nmol/L) (**Table 2**). Data modeled from folic acid intervention studies conducted in China (Crider et al., 2014) were consistent with the relationship between red blood cell folate concentrations and NTD risk as reported in the Irish study by Daly and co-workers. In 2015, the WHO drew on these studies to establish evidence-based guidelines for optimal folate concentrations in women for the prevention of NTDs. According to the WHO (2015) recommendations, in order to achieve the greatest prevention of NTDs at a population level, folate concentrations should be >400 ng/mL (906 nmol/L) in women of reproductive age. It is important to appreciate that the risk of NTD will be increased when maternal folate status is low but within the range typically classed as normal range according to widely accepted cut-off values to diagnose folate deficiency (i.e. serum folate <10 nmol/L and RBC folate <340 nmol/L). Thus low, but not necessarily deficient, maternal folate status is associated with a higher risk of NTD and women not conventionally classed as folate deficient can be at higher risk of an NTD-affected pregnancy.

Maternal vitamin B_{12} status Table 3

Although low maternal folate is considered the major contributing factor, the role of vitamin B_{12} in NTDs is of particular interest because of its close metabolic relationship with folate. Notably, vitamin B12 is required for folate re-cycling and intracellular retention, thus B12 deficiency leads to a failure to retain folates within cells thereby disturbing one-carbon metabolism. There are no trials demonstrating that vitamin B_{12} intervention can prevent NTDs. There is however considerable evidence to show that low B_{12} may be implicated in NTD (Molloy et al., 2009). Several studies that examined maternal vitamin B_{12} status, both during and after an NTD-affected pregnancy, suggest that low maternal B_{12} status is an independent risk factor for having an NTD-affected pregnancy. Such studies showed differences in NTD risk by maternal vitamin B_{12} status, measured as total circulating B_{12} or as serum holotranscobalamin (holoTC) (the fraction of total B_{12} that is destined for tissue uptake and cellular processes thus metabolically "active"). Also, lower amniotic fluid B_{12} concentrations and lower B_{12} binding capacity have been reported in NTD-affected pregnancies (Steen et al., 1998). The relevant studies with at least 50 NTD case mothers are summarized in **Table 3**. These studies indicate that there is an approximately three-fold increased NTD risk for mothers in the lowest quartile or quintile,

Table 1 Intervention trials of periconceptional folic acid supplementation and NTD risk.

Study	*Location*	*Design*	*Daily dose folic acid (mg)*	*Outcome: no. of NTDs*	*Relative risk*	*Comments*
Laurence et al. (1981)	Wales	Randomized controlled trial	4.0	2/60 supplemented[a]	0.42	Not significant
				4/51 not supplemented		Small numbers
Smithells et al. (1983)	UK	Nonrandomized controlled trial	0.36	3/454 supplemented	0.14	Significant[b]
				24/519 not supplemented		
Vergel et al. (1990)	Cuba	Nonrandomized controlled trial	5.0	0/81 supplemented	0.00	Not Significant
				4/114 not supplemented		Small numbers
UK Medical Research Council (1991)	International	Randomized controlled trial	4.0	6/593 supplemented	0.29	Significant[b]
				21/602 not supplemented		
Czeizel and Dudas (1992)	Hungary	Randomized controlled trial	0.8	0/2104 supplemented	0.00	Significant[b]
				6/2052 not supplemented		
Kirke et al. (1992)	Ireland	Randomized controlled trial	0.36	0/172 supplemented	0.00	Not significant
				1/89 not supplemented		Small numbers
Berry et al. (1999)	China	Nonrandomized controlled trial	0.4	*Northern region*		
				13/13012 supplemented	0.21	Significant[b]
				16/3318 not supplemented		
				Southern region		
				34/58638 supplemented	0.59	Significant[b]
				28/28265 not supplemented		
Indian Council of Medical Research (2000)	India	Randomized controlled trial	4.0	4/137 supplemented	0.41	Not significant
				10/142 not supplemented		Small numbers

[a]Two NTD pregnancies in 60 women supplemented with folic acid and 4 NTD pregnancies in 51 women not supplemented with folic acid.
[b]Statistically significant difference in NTD rate between supplemented and non-supplemented groups.

Table 2 Risk of NTDs and distribution of case and control mothers according to maternal red blood cell folate concentration in early pregnancy.

Red blood cell folate (ng/mL)	*No. of cases (%)*	*No. of controls (%)*	*Risk of NTD per 1000 births*	*95% confidence interval*
0–149	11 (13.1)	10 (3.8)	6.6	3.3–11.7
150–199	13 (15.5)	24 (9.0)	3.2	1.7–5.5
200–299	29 (34.5)	75 (28.2)	2.3	1.6–3.3
300–399	20 (23.8)	77 (29.0)	1.6	1.0–2.4
≥400	11 (13.1)	80 (30.0)	0.8	0.4–1.5
Total	84 (100.0)	266 (100.0)	1.9	1.5–2.3

Source: Reproduced with permission from Daly et al. (1995).

compared to the highest, of blood B_{12} concentration. The reports that failed to show a relationship of B_{12} status with NTD were either studies with low numbers (less than 50 case mothers) and therefore more likely lacked sufficient statistical power to demonstrate a significant effect, or were from an area of low indigenous NTD risk (e.g., Finland). Although it has been suggested that low concentrations of vitamin B_{12} may merely be a reflection of a low folate status, the analysis of data from studies that measured both vitamins suggests that they are independent risk factors in NTD. Of note, one study found the highest risk of NTD in women who were in the lowest quartile for both folate and vitamin B_{12} (Kirke et al., 1993).

Table 3 Large case-control studies (>50 case mothers) assessing blood vitamin B_{12} status in mothers of NTD-affected children.

Study	*Country (dates)*	*Sample*	*Time of sampling in pregnancy*	*Cases/ controls*	*Cases/controls B_{12} pmol/L*[a]	*OR (highest v lowest quantile)*[b,c,d,g]	*95% CI*	*Significant (yes/no) P value (if given)*
Mills et al. (1992)	Finland (1983–89)	Serum B12	6–16 weeks	78/150	356/384 (mean)			No
Christensen et al. (1999)	Canada (pre-1998)		Non pregnant	59/88	298/350 (mean)			Yes; $p = 0.05$
Suarez et al. (2003)	Texas-Mexico border (1995–2000)	Serum B12	Postpartum	225/378	317/367 (median)	3.0[c]	1.4–6.3	Yes; $p = 0.001$
Ray et al. (2007)	Canada (1993–2004)	Serum holoTC	15–20 weeks	89/422	68/81 (geometric mean)	2.9[b,e]	1.2–6.9	Yes
Zhang et al. (2008)	China (Shanxi) (2004–2005)	Serum B12	20 weeks Median	84/110	73/91 (geometric mean)	4.96[d]	1.94–12.7	Yes; $p < 0.01$
Molloy (2009) (3 cohorts)	(1) Ireland (1983–1984)	Serum B12	15 weeks Median	95/265	155/179 (median)	3.14[b,e]	1.46–6.72	Yes; $p = 0.003$
	(2) Ireland (1986–1990)	Plasma B12	15 weeks Median	76/222	180/221 (median)	2.45[b,e]	1.12–5.32	Yes; $p = 0.024$
	(3) Ireland (1986–1990)	Plasma B12	15 weeks[f] Median	107/414	199/232 (median)	2.75[b,e]	1.43–5.28	Yes; $p = 0.003$
Nasri et al. (2015)	Tunisia (2012–2013)	Serum B12	2nd-3rd trimester	75/75	218/264 (median)	2.56[g]	1.15–5.70	Yes; $p = 0.009$
Senousy et al. (2018)	Egypt	Serum B12	2nd-3rd trimester	50/70	202/228			Yes; $p = 0.0015$

[a]For comparison across studies all vitamin B_{12} concentrations are presented as pmol/L. For studies that reported data as ng/mL, mean or median values were converted to pmol/L using a multiplication factor of 0.738.
[b]Quartile of B_{12} status.
[c]Quintile of B_{12} status.
[d]Comparison of B_{12} above and below 55 pmol/L.
[e]Adjusted for maternal folate.
[f]Cases were mothers with a history of an NTD-affected pregnancy but currently undergoing an unaffected pregnancy.
[g]Tertile of B_{12} status.

Other maternal risk factors for NTD

Specific nutrients

Apart from folate and vitamin B_{12}, other nutrients involved in one-carbon metabolism may play a role in NTD. Low maternal choline status has been implicated in several studies as a risk factor for NTDs. Choline has a close metabolic interaction with folate, and there is evidence that supplementation with one nutrient may ameliorate the deficiency caused by the other. As discussed below, riboflavin may also play a protective role in preventing NTD in women genetically at higher risk owing to a common folate polymorphism but this remains to be demonstrated (i.e. those homozygous for the 677C→T polymorphism in the gene encoding methylenetetrahydrofolate reductase (MTHFR), a folate-metabolizing enzyme that depends upon riboflavin (as flavin adenine dinucleotide; FAD) for its activity). Also, one mutant animal strain that has served as a model of NTDs for decades is resistant to folic acid but responsive to maternal periconceptional supplementation with inositol. Building on the data from animal models, a series of small clinical studies in high-risk women who had experienced previous NTD-affected pregnancies indicated that supplementation with inositol has value in increasing NTD prevention beyond that achievable by folic acid alone (Facchinetti et al., 2020). It is possible that the observed protection by folic acid intervention in some instances occurs through indirectly ameliorating the impairment in one-carbon metabolism caused by one of these other nutrients. Deficiencies of other nutrients, including vitamin C and zinc, have also been implicated in NTDs but the evidence is limited and the data are inconsistent.

Obesity and diabetes

There are other potentially modifiable risk factors for NTDs. Women who are obese compared to those of normal body mass index (BMI) are more likely to have a child affected with an NTD (and indeed other malformations including congenital heart defects). Several studies reported a higher risk of an NTD-affected pregnancy in obese women, with meta-analyses of available studies concluding that obese compared with normal BMI women are at higher risk of fetal NTDs, with a three-fold higher risk for severely

obese women (Huang et al., 2017; Vena et al., 2021). The increasing prevalence of obesity globally may make obesity an important population health burden for NTDs in the future.

Diabetes mellitus is a well-known consequence of obesity and maternal pregestational diabetes poses increased risk for many congenital defects, including NTDs. Although the precise teratogenic mechanisms linking diabetes with NTD are poorly understood, good glucose control in diabetic women can reduce the risks of having pregnancies affected by NTDs and other birth defects. Of note, recent evidence from small clinical trials and experimental studies suggests that supplementation with myo-inositol (an insulin sensitizer) may have benefits in preventing the onset of gestational diabetes mellitus as well as reducing the risk of recurrence of NTDs (Facchinetti et al., 2020).

Other risk factors

The use of antifolate drugs is associated with an increased risk of NTDs. These medications, designed to antagonize folate metabolism and thus block one-carbon biosynthetic and methylation reactions, include methotrexate, aminopterin and trimethoprim, used in the treatment of malignancies, inflammatory disorders, bacterial infections and malaria. In addition, a range of other drugs can antagonize folate incidentally, as an adverse effect. Such drugs include anticonvulsant treatments (valproic acid, carbamazepine, phenobarbital, phenytoin and primidone) used to control seizures in epilepsy. Women taking these medications are thus considered an at-risk group for NTDs. Hyperthermia during pregnancy has also been associated with an increased risk of NTD.

Biological mechanisms in the etiology and prevention of NTDs

The possible mechanisms underlying the involvement of folate/folic acid in the etiology and prevention of NTDs are examined in this section. The mechanisms to explain the beneficial effects of periconceptional folic acid against NTD have focused on the factors that could potentially impair normal folate metabolism, including polymorphisms in folate genes. Autoantibodies against folate receptors have also been implicated in pregnancies affected by NTD (Rothenberg et al., 2004).

Role of folate, vitamin B_{12} and one-carbon metabolism

Folate is essential for one-carbon metabolism where it acts in various co-factor forms as an intermediary in the transfer of one-carbon groups for two important biological processes, namely, the provision of methyl groups for methylation reactions and the *de novo* synthesis of purines and thymidylate (dTMP) for DNA and RNA (**Fig. 1**). The vitamin B_{12}-dependent enzyme methionine synthase is central to both the methylation and DNA synthesis components of one-carbon metabolism. Through this enzyme, vitamin B_{12} and folate control the intracellular flux of one-carbon units between these two major metabolic cycles. Folate enters the cell as 5-methyltetrahydrofolate (5-methylTHF) and must release its methyl group through the methionine synthase reaction in order to be retained in the cell. As tetrahydrofolate (THF), it can then be polyglutamated and can accept one-carbon units from serine, formate, and other sources for use in nucleotide synthesis or regeneration of 5-methylTHF. The released methyl group from 5-methylTHF is used via methionine synthase, to methylate homocysteine, thereby producing methionine, which is then converted to *S*-adenosylmethionine for methylation of proteins, lipids, DNA, and many other cellular components. Deficiency, impaired function or limited cellular availability of folate or vitamin B_{12} leads to an increase in intracellular homocysteine that culminates in higher plasma concentrations of this metabolite. Methylenetetrahydrofolate reductase (MTHFR) is a key folate-metabolizing enzyme that catalyzes the irreversible conversion of 5,10-methyleneTHF to 5-methylTHF, thereby committing one-carbon units to methylation reactions and away from DNA synthesis. MTHFR activity is dependent upon an adequate supply of riboflavin in its cofactor form flavin adenine dinucleotide (FAD). Thus, for folate to function within one-carbon metabolism, it interacts closely with vitamin B12, along with vitamin B6 and riboflavin. Consequently, sub-optimal status of one or more of these B vitamins, or polymorphisms in folate genes, can impair one-carbon metabolism, even if folate intakes are adequate.

There are many ways in which abnormal one-carbon metabolism could result in abnormal closure of the neural tube. Neural tube closure is a highly complex developmental process, involving multiple cycles of cell proliferation and apoptosis; the details of which are not fully understood. Inadequate production of DNA for the embryo, lack of methyl groups for methylation reactions that regulate cell signaling, gene expression and activation or repression of apoptotic pathways are obvious candidates. Several studies found higher total homocysteine in maternal plasma or in amniotic fluid during NTD pregnancy (Steegers-Theunissen et al., 1991, 1995), suggesting that these women were less able to metabolize homocysteine. These biochemical studies point to a subtle alteration in folate homeostasis in families with NTDs and suggest that folic acid supplementation works by overcoming a metabolic block in folate related processes due to specific genetic variants within folate pathways. This hypothesis would explain the finding that both environment and genes contribute to NTD risk and has prompted an intensive investigation of genes encoding folate-related proteins as risk factors.

Folate-related genetic risk factors and risk of NTDs Table 4

Polymorphisms in folate genes have been studied in relation to NTD risk. Of these, an increased risk of NTD is most strongly associated with the 677C→T variant in the gene coding for the folate-metabolizing enzyme, MTHFR. This enzyme is key within one-carbon metabolism because it catalyzes the irreversible conversion of 5,10-methyleneTHF to 5-methylTHF, the essential folate

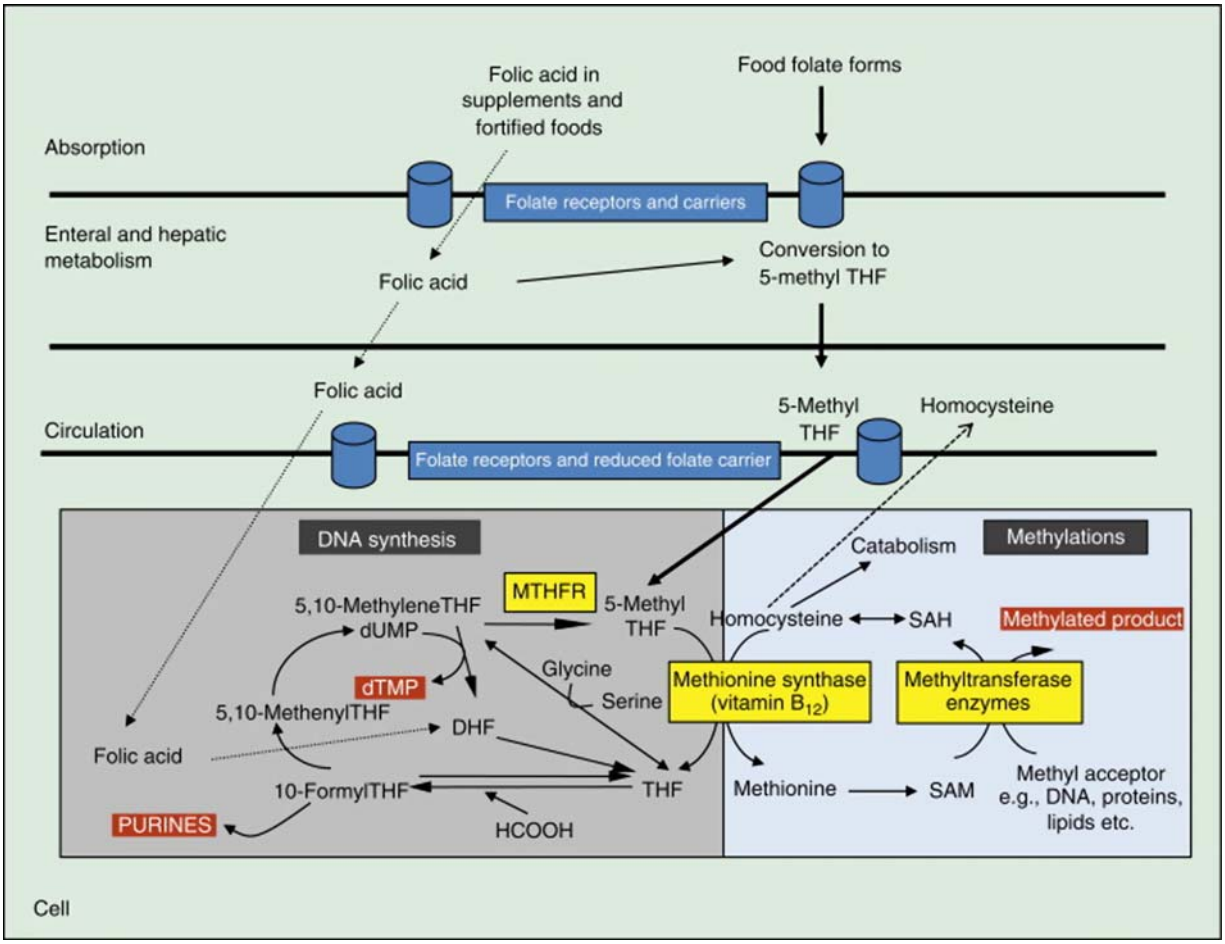


Fig. 1 Intracellular pathways of folate and homocysteine metabolism and their relation to vitamin B_{12}. Folate cofactors are used in the transfer of one-carbon groups across two distinct metabolic cycles involving (A) the *de novo* synthesis of purines and thymidylate (dTMP) for DNA and (B) the provision of methyl groups for methylation reactions. Folate interacts with vitamin B_{12} at the methionine synthase reaction, through which the THF cofactor pool provides methyl groups for biological methylation reactions via the conversion of homocysteine to methionine. Methylenetetrahydrofolate reductase (MTHFR) catalyzes the irreversible conversion of 5,10-methyleneTHF to 5-methylTHF, thereby committing one-carbon units to methylation reactions and away from DNA synthesis. In addition to enzymes involved in folate pathways, the reduced folate carrier, folate receptors, the proton coupled folate transporter and mitochondrial folate enzymes play key roles in maintaining one-carbon homeostasis within the cell. Vitamin B_{12} binding proteins and receptors also have important effects on one-carbon metabolism through their role in maintaining optimal cellular B_{12} levels.

co-factor that provides methyl groups for S-adenosylmethionine-mediated methylation reactions (Bailey et al., 2015). The common 677C→T variant in the *MTHFR* gene is associated with reduced MTHFR enzyme function, resulting in lower blood folate and higher homocysteine concentrations, and an increased nutritional requirement for folate. In addition, in individuals with the variant *MTHFR* 677TT genotype, poor folate status causes an altered distribution of intracellular folate cofactors and reduced global methylation of DNA. Riboflavin also plays an important role in affected individuals because the variant enzyme binds its cofactor FAD with lower affinity, however in vivo studies show that this can be stabilized by intervention with supplemental riboflavin (McNulty et al., 2002, 2006).

Several studies showed that the *MTHFR* 677C→T polymorphism is strongly associated with an increased of NTDs. One meta-analysis concluded that the variant *MTHFR* 677TT genotype in the child or mother is associated with a doubling in the risk of NTD (Yang et al., 2015). The frequency of the variant TT genotype is 10% worldwide, however, the distribution varies considerably according to geographical area, ranging 0–3% (in Africa), 8–11% (in Australia), 3–18% (in United States), 8–20% (in China) and 4–26% (in Europe). The highest reported frequencies of the TT genotype worldwide is in Mexico (32%) and specifically Mexican Amerindians (44%) (Psara et al., 2020). Apart from NTD, this polymorphism is also associated with a increased risk of hypertension, hypertension in pregnancy and cardiovascular disease (McNulty et al., 2017, 2020). Studies have however shown important gene–nutrient interactions as regards the risk of NTD and other adverse health outcomes associated with this common polymorphism. Low maternal blood folate concentrations increase the NTD risk associated with the variant genotype, while low riboflavin status exacerbates the excess risk of hypertension in pregnancy and throughout the lifecycle. Differences in the frequency

Table 4 NTD association evidence involving high priority polymorphisms in a number of folate and vitamin B_{12} candidate genes.

Gene	*Enzyme*	*Association with NTDs*
BHMT 742G→A	Betaine-homocysteine methyltransferase R239Q	Inconclusive maternal risk associations.
CBS 844ins68	Cystathionine β synthase	No reported risk association with NTD.
DHFR Intron 1	Dihydrofolate reductase (19 bp deletion)	Conflicting reports of risk association and protection.
FRα, FRβ, FRγ Several SNPs	Folate receptors	No reported risk associations with NTD or biochemical changes in humans.
GCPII 1561C →T	Folyl-γ-glutamate carboxypeptidase H475Y	Inconclusive maternal risk associations.
MTHFD1 1958G→A	Trifunctional C1 Synthase R653Q	Maternal risk factor for NTDs in some populations.
MTHFR 677C→T	5,10-Methylene tetrahydrofolate reductase A222V	Significant risk factor for NTDs in some but not all populations. Important cause of low folate status.
MTHFR 1298A→C	5,10-Methylene tetrahydrofolate reductase E429A	No clear risk associations that are independent of 677C→T.
MTR 2756A→G	Methionine Synthase D919G	May interact with other genes as a maternal risk factor but no clear independent risk.
MTRR 66A→G	Methionine synthase reductase I22M	Several studies show maternal risk associations and possible interactive effects with low B12 or other genes.
RFC-1 80G→A	Reduced folate carrier H27N	Inconclusive maternal risk associations. Possible interaction with maternal nutrient intake.
SHMT1 1420C→T	Serine hydroxymethyltransferase L474F	No reported risk association with NTD.
TCblR Several SNPs	Transcobalamin II receptor	Several rare polymorphisms conferred highly significant risk in one large population study.
TCII 776C→G	Transcobalamin II R259P	Inconclusive maternal risk associations. Several studies found changes in maternal serum TC II concentrations or in amniotic fluid during NTD affected pregnancies.
TSER (Promoter enhancer region)	Thymidylate Synthase (28 bp double or triple repeat)	Inconclusive maternal risk associations.

of this polymorphism, along with variations in dietary folate and riboflavin intakes (particularly owing to differences in food fortification practices) between countries, may explain observations of a clear association with NTD risk in some populations but not in others. Homozygosity for the *MTHFR* 677C→T polymorphism is estimated to account for approximately 13% of NTDs cases overall.

The role of other folate-related gene variants have been investigated in relation to NTDs. **Table 4** gives details of NTD association studies involving high priority single nucleotide polymorphisms (SNPs) in a number of folate and vitamin B_{12} candidate genes. Perhaps not surprisingly, results from candidate gene studies have been rather inconclusive, with results rarely replicated in other studies. There are many reasons for this, such as the underlying low penetrance of the genetic effects, differences in genetic and environmental susceptibility between populations, and study design issues such as inadequate sample sizes and poorly matched controls. Also, because the precise biological mechanisms leading to NTDs are poorly understood, it is highly likely that alteration in the function of a gene with no apparent link to folate pathways could play a role in NTDs. A further problem is that nearly a third of all human genes have no known function, yet most of these un-annotated genes are conserved through evolution, indicating that nature has assigned them important functions. The emergence of genome-wide association study (GWAS) in more recent years can overcome many of the limitations of earlier candidate gene studies, and GWAS is now the focus of most investigations to identify genetic risk factors in NTD. The GWAS design has the advantage of being able to screen every region in the genome for association with NTDs and can detect variations that are clustered within case families compared with control groups.

Prevention of NTD

Nutritional considerations

Although the terms "folic acid" and "folate" are often used interchangeably, correctly speaking, folic acid refers to the synthetic form of the vitamin, while folate refers to the natural vitamin forms as found in plant and animal tissues and thus in food sources. The richest food sources are green leafy vegetables, asparagus, beans, legumes, liver and yeast. Folic acid is found in the human diet only in fortified foods and vitamin supplements but is readily converted to the natural cofactor forms of folate after its ingestion. There are however differences in the chemical structures between folic acid and the natural folate forms and these have important nutrition consequences. Folic acid is a fully oxidized molecule and is a monoglutamate, meaning that it contains just one glutamate moiety in its structure. Naturally occurring food folates, on the other hand, are a mixture of reduced folate forms (predominantly 5-methylTHF) and typically found as polyglutamates, containing a variable number of glutamate residues. As reduced molecules, natural food folates are inherently unstable outside living cells and have poor bioavailability (McNulty and Pentieva, 2010). In

addition to their limited bioavailability, food folates (particularly green vegetables) can be unstable during cooking, and this will substantially reduce the folate content of a food before it is even ingested (McKillop et al., 2002). Folic acid is more stable and more bioavailable compared with an equivalent amount of the vitamin eaten as naturally occurring food folates, and is the form used in food fortification. Therefore, fortified foods provide the most important dietary source of the vitamin.

The differences between food folates and folic acid have led to the development of "dietary folate equivalents" or DFE values, an approach devised to take into account the greater bioavailability of folic acid from fortified foods compared with naturally-occurring food folates. Dietary folate intakes and folate recommendations for population sub-groups in many countries, including the United States, are now expressed as DFEs. As discussed below, specifically for the prevention of NTD, there are recommendations for folic acid supplementation in women of reproductive age; this is additional to typical dietary folate intakes.

Interventions to optimize folate status

The instability and poor bioavailability of food folates means that they have limited ability to influence blood folate concentrations and thus achieving optimal folate status is challenging (Cuskelly et al., 1996). To optimize folate status in women of reproductive age for preventing NTDs, folic acid intervention strategies are needed, for individuals and populations. As discussed above, folic acid, the vitamin form used in fortified foods and supplements, is very stable and highly bioavailable when ingested. Thus, depending on national fortification policy and/or access to fortified food, the folate status of populations can vary greatly from one country to the next and this is reflected in differences in health outcomes.

Folic acid supplementation

Specifically for the prevention of NTD, women are recommended to take 400 μg per day of folic acid from preconception until the end of the first trimester of pregnancy. Some women are considered to be at higher risk and thus are recommended to take higher folic acid doses. This includes women with a previous pregnancy affected by NTD and those taking certain anticonvulsant drugs known to interfere with folate metabolism. Studies show that folic acid supplementation is a highly effective means to optimize folate status in individual women who take their supplements as recommended. However, it is not an effective public health strategy for populations because in practice very few women take folic acid as recommended.

Folic acid fortification

Food fortification may be undertaken on a voluntary (at the discretion of the food manufacturer) or mandatory (regulated by a government) basis. Because folic acid is more stable and bioavailable than naturally-occurring food folates, folate biomarker status will reflect the folic acid intakes of individuals. At a population level, the observed differences in folate status between countries is primarily due to differences in exposure to folic acid–fortified foods, in turn reflecting local fortification policy. Folate status is found to be highest in countries with mandatory folic acid-fortification, followed by those with voluntary fortification, while folate status is lowest in countries where fortified foods are not consumed.

When folic acid-fortification is undertaken on a mandatory (i.e. population-wide) basis, it has proven itself to be highly effective as a means to increase folate status in that population. Data from the US National Health and Nutritional Examination Survey (NHANES), and from retrospective longitudinal studies in Canada, demonstrate that mandatory folic acid-fortification results in marked increases in both short-term (serum folate) and long-term (red blood cell folate) biomarkers of folate status (De Wals et al., 2007; Crider et al., 2018). With the first two years of mandating folic acid-fortification in the US (in 1998), a tripling in serum folate and an increase of 40–60% in red blood cell folate was observed. Correspondingly, the prevalence of low folate status in US women of reproductive age dropped from 21% and 30%, to 0.8% and 2.8%, for serum and red blood cell folate, respectively. Mandatory fortification with folic acid has produced similar results worldwide.

In countries that permit voluntary fortification with folic acid and other micronutrients, the consumer will have ready access to fortified foods (e.g., breakfast cereals). In such countries, folic acid-fortified foods have been shown to have positive and significant impacts on dietary intakes and status biomarkers in consumers of these foods. When consumed regularly, fortified foods are associated with significantly higher serum and red blood cell folate concentrations (Hopkins et al., 2015). Thus folic acid-fortification of food is highly effective as a means of optimizing folate status, but importantly, the benefit will be limited to only those individuals who choose to consume fortified products. In the UK, where there is voluntary (but not mandatory) fortification of foods with folic acid, the percentage of women with insufficient red blood cell concentrations (<906 nmol/L) to prevent folate-responsive NTD is estimated to be 83% in Northern Ireland, 81% in Scotland and 79% in Wales (Public Health England, 2017). Also, evidence from the National Adult Nutrition Survey in Ireland showed that non-consumers of FA from fortified food or supplements were at particularly high risk of suboptimal folate status (Hopkins et al., 2015), again using the cut-point of 906 nmol/L red blood cell folate to define optimal status. In contrast, mandatory fortification reaches everyone in a population and is therefore a much more effective strategy for optimizing folate status in women of reproductive age regardless of socioeconomic or racial factors.

Effects of folic acid policy globally

As a result of the conclusive evidence of the benefits of folic acid against NTD, public health authorities globally have in place clear folic acid recommendations for women of reproductive age. The implementation of policy in this area is however problematic because, despite a proven benefit in NTD, there are concerns that folic acid could be harmful at high levels of exposure. The relative

success of the contrasting approaches in North America (mandatory food fortification) and Europe (supplementation) to prevent NTD provide important lessons for countries re-considering policy in this area.

Effects on NTD prevalence

In Europe, policy to prevent NTD has proven to be largely ineffective. For over 25 years, in European countries, including the UK, Ireland, policy has been based on recommending women of reproductive age (and/or those planning a pregnancy) to take a supplement containing folic acid. However, as a sole public health measure and despite active health promotion campaigns over many years, folic acid supplementation has had little or no impact in preventing NTD at a population level. The lack of success of this approach is primarily because women typically start taking folic acid after the period of neural tube closure (i.e. the 3rd to 4th week of pregnancy). For many women, the early period when folic acid is protective against NTD may have passed before folic acid supplements are even started. An even greater challenge is that an estimated 50% of pregnancies are unplanned. Thus a large number of women are not protected and this has resulted in unacceptably high rates of NTD in Europeans countries, where NTD rates were recently estimated to be 1.6 times higher than in regions of the world with mandatory folic acid-fortification policies in place. Of particular concern are reports that the incidence of NTD in Ireland—a country with one of the highest NTD rates in the world - is increasing in recent years (FSAI, 2016).

In contrast, a policy of folic acid-fortification of food on a mandatory basis (in place in 85 countries worldwide to date) is highly effective in preventing NTDs because it reaches all women, including those who have not planned their pregnancy. International evidence shows that wherever such a policy has been introduced, it has proved to be effective in reducing rates of NTD in that country. The effect of mandatory fortification programs on NTD rates has been striking, especially in Canada (and particularly so in Newfoundland and Nova Scotia where rates were highest before the implementation of fortification). It is worth noting that data from Canada are generally more accurate than data from the US because of the limited information on prenatally diagnosed NTD cases in the US (Mosley et al., 2009). **Fig. 2** shows NTD prevalence rates pre-and post folic acid-fortification in 11 areas where mandatory fortification has been implemented; the data are from countries that have implemented fortification and have recorded the change in NTD rates. The greatest drop in prevalence has been apparent in countries with the highest indigenous NTD rate and the lowest rates achieved in most countries to date is 5–6 per 10,000 births. The totality of evidence, using data from both randomized trials and from pre- and post-fortification programs, indicates that not all NTDs are preventable via intervention with folic acid only.

These data contrast with countries in Europe where there has been no mandatory fortification program and no significant change in NTD rates over the past 25 years (Khoshnood et al., 2015). The European data suggests that failure to implement mandatory folic acid-fortification has caused and continues to cause NTDs to occur in almost 1000 pregnancies every year (Morris et al., 2021).

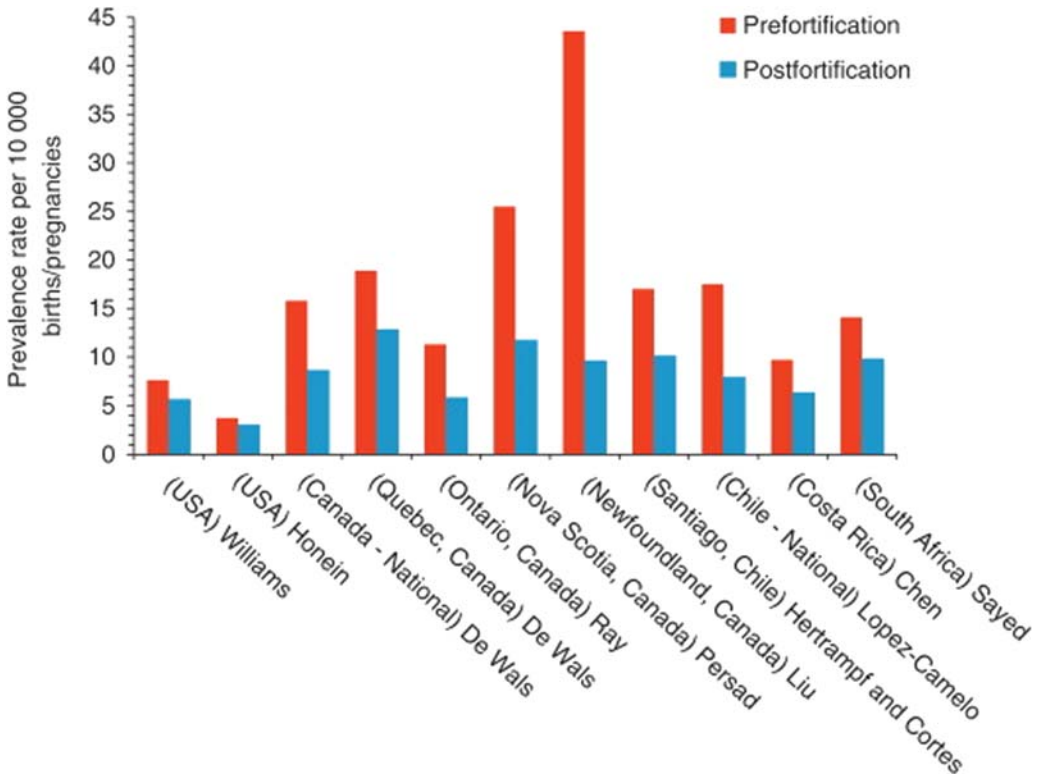


Fig. 2 NTD prevalence rates pre- and post-fortification of foods with folic acid in 11 areas where mandatory fortification has been implemented.

Notably, one recent study estimated that from 1998 to 2017, a total of 95,213 NTD pregnancies have occurred among 104 million births in 28 European countries; a prevalence of 0.92 per 1000 births (Morris et al., 2021). Ireland is recognized as having one of the highest rates of NTD-affected pregnancies in the world and there are concerns that the incidence of NTD is increasing in recent years. In 2016, following an extensive scientific review, the Food Safety Authority of Ireland published an updated report recommending mandatory fortification of bread or starch with folic acid (FSAI, 2016). Similarly, in 2017, The UK Scientific Advisory Committee on Nutrition confirmed its longstanding advice that mandatory fortification of cereal flours with folic acid should be introduced for the prevention of NTD (SACN, 2017). Subsequently, in 2021 the UK Government announced that it will introduce the mandatory fortification of non-wholemeal wheat flour with folic acid, but the legislation to enact the new policy has not yet been implemented.

Effects on other health outcomes

Apart from preventing NTD, optimal maternal folate status prevents the development of megaloblastic anemia in mothers during pregnancy. There is also some evidence that folic acid can prevent other major birth defects. The Hungarian folic acid intervention trial found a 47% reduction in birth defects other than NTDs. Several studies reported decreasing trends in the prevalence of congenital heart defects and other rarer defects since folic acid-fortification began in the US and Canada and after implementation of folic acid intervention studies in China, but an association with increased folic acid intake was not established. There have also been reports of a decreased prevalence in several childhood cancers although the data are inconsistent. There is no convincing evidence of a protective effect of folic acid on the prevalence of orofacial clefts.

There may also be other beneficial effects of intervention to optimize folate status in pregnancy and throughout the lifecycle. Deficient maternal folate status (and/or elevated homocysteine) has been associated with an increased risk of adverse pregnancy outcomes including gestational hypertension, preeclampsia, placental abruption, pregnancy loss, low birth weight and intrauterine growth restriction (Psara et al., 2020). Although there is some evidence that folic acid supplementation in pregnancy can reduce the risk of gestational hypertension and pre-eclampsia (De Ocampo et al., 2018), this is conflicting.

In addition to protecting against the development of NTDs in the offspring, there is emerging evidence linking maternal folate throughout pregnancy with neurodevelopment and cognitive function in the child (Caffrey et al., 2019). The biological mechanism linking maternal folate with the offspring brain is unclear, but likely involves folate-mediated epigenetic changes related to brain development and function. Indeed, a wealth of literature supports the fetal origins of human disease throughout the lifecycle and emerging evidence implicates epigenetic modifications as the likely mechanism. DNA methylation, the most widely studied epigenetic mechanism for gene regulation, is dependent upon the supply of methyl donors provided by folate and related B-vitamins via S-adenosylmethionine. Folate deficiency could thus lead to aberrant gene expression leading to adverse health outcomes. Thus, the human *in utero* environment may influence offspring brain health in the longer term via folate-mediated DNA methylation, but this aspect requires further investigation (Caffrey et al., 2019).

Also, numerous studies have shown that elevated plasma homocysteine is independently associated with greater risk of cardiovascular disease. Because folic acid intervention lowers plasma homocysteine, mandatory food fortification with folic acid may have beneficial effects on the risk of cardiovascular disease, particularly in relation to the primary prevention of stroke where the evidence (from both randomized trials and population-based data) for a beneficial effect of folic acid is quite compelling.

Emerging folic acid policy: risk-benefit assessment

High folate intakes are not associated with any adverse effects. However, there are concerns of potential adverse effects of excess intakes of folic acid, the synthetic vitamin form. Excessive folic acid intake constitutes exposure doses that exceed the Tolerable Upper Intake Level (UL) of 1000 μg/d for adults, as set by the US Institute of Medicine (IOM, 1998). The main safety concern with folic acid-fortification was that some sectors of the population, particularly older people, would be exposed to very high folic acid intakes because of concomitant fortification and supplement use.

Once ingested, folic acid is reduced by dihydrofolate reductase and after subsequent methylation, it is released in the systemic circulation as 5-methylTHF. However, the reduction of folic acid is a slow process that is influenced by individual variations in dihydrofolate reductase activity and thus exposure to high oral doses of folic acid can result in the appearance of unmetabolized folic acid in the circulation. The latter is not a normal constituent of plasma or other tissues. On this basis, concerns have been raised regarding potential (although unconfirmed) adverse health effects of unmetabolized folic acid arising in the circulation through high folic acid exposures from supplements and fortified foods. Unlike the case with 5-methylTHF (the normal folate form entering cells), the uptake of folic acid by cells does not require vitamin B_{12}, therefore, folic acid entering a cell might initiate DNA synthesis in a vitamin B_{12}-deficient person, thereby preventing the development of anemia and potentially delaying a diagnosis of B_{12} deficiency, allowing the associated neurologic damage to progress and become irreversible. Although such an effect has not been demonstrated to date, concerns remain about the potential physiological impacts of the nutrient imbalance caused by high folic acid intakes together with low vitamin B_{12} concentrations. Adding vitamin B_{12} as well as folic acid to fortified food has been suggested as a solution, but more evidence on efficacy, dosage, and feasibility is required. Another concern is the possibility that, because of its role in DNA synthesis, high folic acid status will help to advance established cancers and promote malignant transformation of premalignant lesions. Other possible adverse effects of high folic acid intake have been suggested, including reports of decreased natural killer cell cytotoxicity, increased twinning rates, interference with the efficacy of anticonvulsant and antifolate drugs, and increased childhood asthma and autism rates. None of these reports have been substantiated but caution and vigilance remains an important public health position in relation to the food fortification strategy.

A recent report from a 2019 expert workshop, as convened by the US National Institutes of Health, tasked with reviewing the evidence in this area, concluded that there is an insufficient body of evidence to support adverse human health outcomes as a result of high intakes of folic acid. Nonetheless, these experts called for further high quality research to determine the safety of excess folic acid intake (Maruvada et al., 2020). Thus, the totality of the evidence at this time suggests that adverse effects associated with folic acid overexposure are unlikely at the generally low folic acid levels arising through mandatory food fortification. Nonetheless the risk-benefit debate surrounding food fortification with folic acid continues among policymakers and effective monitoring should remain a key aspect of policy in this area.

Recommendations to prevent first occurrence and recurrence of NTDs

As a result of the conclusive evidence of the benefits of folic acid against NTD, public health authorities globally recommend women to take folic acid supplements from before conceiving until the 12th week of pregnancy. Internationally established recommendations distinguish between occurrent (first-time) and recurrent NTDs. Women with a previously affected pregnancy are advised to take the much higher dose of 4.0 mg of folic acid daily from at least four weeks before conception until the end of the third month of pregnancy. The 4.0 mg dose should be taken under the supervision of a doctor because giving high doses of folic acid can complicate the diagnosis of vitamin B12 deficiency. Women with epilepsy on anticonvulsant therapy require individual counseling before starting folic acid supplementation.

For the prevention of first occurrence of NTDs, most public health authorities recommend that all women capable of becoming pregnant consume 400 μg of folic acid per day and that total folic acid consumption should not be more than 1.0 mg/d to avoid the possible risks of high intakes. The only practical ways of achieving optimal red blood cell folate status associated with lowest risk of NTDs are by consuming folic acid supplements or folic acid fortified foods, rather than increasing intakes of foods naturally rich in folate. Irrespective of local fortification policy and availability of fortified foods, women should be advised to take supplements, from before and in early pregnancy to ensure optimal folate status. If multivitamins are used to provide folic acid, care should be taken not to exceed safe levels of other components, particularly vitamins A and D.

Conclusions

Maternal folic acid supplementation taken periconceptionally is known to have beneficial effects in the prevention of NTDs. Mechanistically, the known role of folate in one-carbon metabolism provides a biological basis to link maternal folate status with NTD and other offspring health outcomes. Vitamin B12 and riboflavin, also involved in the one-carbon network, may have additional preventative roles in NTD, albeit unproven in randomized trials. Although there are clear recommendations in place worldwide for the prevention of NTD through folic acid supplementation before and during early pregnancy, for many women the very early stage when the neural tube is closing may have passed before supplementation is started. Thus, current public health measures that do not involve mandatory fortification of foods with folic acid, have been shown to be largely ineffective in reducing NTDs. Mandatory folic acid-fortification of food is proven to be a highly effective measure to prevent NTD, but policy in this area remains controversial owing to concerns—as yet unproven—related to potential adverse effects of over-exposure to folic acid. In the absence of population-wide fortification in many countries, optimizing folate status of mothers in early pregnancy to provide protection against NTDs remains challenging. This means that, despite the existence of robust scientific evidence for nearly three decades of the benefits of folic acid intervention, preventable NTDs are not being prevented in many countries worldwide, including in Europe.

References

Bailey, L.B., Stover, P.J., McNulty, H., et al., 2015. Biomarkers of nutrition for development—folate review. J. Nutr. 145, 1636S–1680S.
Berry, R.J., Li, Z., Erickson, J.D., et al., 1999. Prevention of neural-tube defects with folic acid in China. N. Engl. J. Med. 341, 1485–1490.
Blencowe, H., Kancherla, V., Moorthie, S., et al., 2018. Estimates of global and regional prevalence of neural tube defects for 2015: a systematic analysis. Ann. N. Y. Acad. Sci. 1414, 31–46.
Caffrey, A., McNulty, H., Irwin, R.E., et al., 2019. Maternal folate nutrition and offspring health: evidence and current controversies. Proc. Nutr. Soc. 78, 208–220.
Christensen, B., Arbour, L., Tran, P., et al., 1999. Genetic polymorphisms in methylenetetrahydrofolate reductase and methionine synthase, folate levels in red blood cells, and risk of neural tube defects. Am. J. Med. Genet. 84, 151–157.
Crider, K.S., Devine, O., Hao, L., et al., 2014. Population red blood cell folate concentrations for prevention of neural tube defects: bayesian model. BMJ 349, g4554.
Crider, K.S., Qi, Y.P., Devine, O., et al., 2018. Modeling the impact of folic acid fortification and supplementation on red blood cell folate concentrations and predicted neural tube defect risk in the United States: have we reached optimal prevention? Am. J. Clin. Nutr. 107, 1027–1034.
Cuskelly, G.J., McNulty, H., Scott, J.M., 1996. Effect of increasing dietary folate on red-cell folate: implications for prevention of neural tube defects. Lancet 347, 657–659.
Czeizel, A.E., Dudás, I., 1992. Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. N. Engl. J. Med. 327, 1832–1835.
Daly, L., Kirke, P.M., Molloy, A., et al., 1995. Folate levels and neural tube defects: implications for prevention. JAMA 274, 1698–1702.
De Ocampo, M.P.G., Araneta, M.R.G., Macera, C.A., et al., 2018. Folic acid supplement use and the risk of gestational hypertension and preeclampsia. Women Birth 31, e77–e83.
De Wals, P., Tairou, F., Van Allen, M.I., et al., 2007. Reduction in neural-tube defects after folic acid fortification in Canada. N. Engl. J. Med. 357, 135–142.
Facchinetti, F., Cavalli, P., Copp, A.J., et al., 2020. An update on the use of inositols in preventing gestational diabetes mellitus (GDM) and neural tube defects (NTDs). Expert Opin. Drug Metab. Toxicol. 16, 1187–1198.
FSAI, 2016. Report of the Scientific Committee of the Food Safety Authority of Ireland: Update Report on Folic Acid and the Prevention of Birth Defects in Ireland. Food Safety Authority of Ireland (FSAI), Dublin.

Hopkins, S.M., Gibney, M.J., Nugent, A.P., et al., 2015. Impact of voluntary fortification and supplement use on dietary intakes and biomarker status of folate and vitamin B-12 in Irish adults. Am. J. Clin. Nutr. 101, 1163–1172.
Huang, H.Y., Chen, H.L., Feng, L.P., 2017. Maternal obesity and the risk of neural tube defects in offspring: a meta-analysis. Obes. Res. Clin. Pract. 11, 188–197.
Indian Council of Medical Research, 2000. Multicentric study of efficacy of periconceptional folic acid containing vitamin supplementation in prevention of open neural tube defects from India. Indian J. Med. Res. 112, 206–211.
IOM, 1998. Folate. In: Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline. National Academiies Press (US), Washington (DC), pp. 196–305.
Khoshnood, B., Loane, M., De Walle, H., et al., 2015. Long term trends in prevalence of neural tube defects in Europe: population based study. Br. Med. J. 351, h5949.
Kirke, P.N., Daly, L.E., Elwood, J.H., 1992. A randomised trial of low dose folic acid to prevent neural tube defects. The Irish Vitamin Study Group. Arch. Dis. Child. 67, 1442–1446.
Kirke, P.N., Molloy, A.M., Daly, L.E., et al., 1993. Maternal plasma folate and vitamin B12 are independent risk factors for neural tube defects. QJM 86, 703–708.
Laurence, K.M., James, N., Miller, M.H., et al., 1981. Double-blind randomised controlled trial of folate treatment before conception to prevent recurrence of neural-tube defects. Br. Med. J. 282, 1509.
Maruvada, P., Stover, P.J., Mason, J.B., et al., 2020. Knowledge gaps in understanding the metabolic and clinical effects of excess folates/folic acid: a summary, and perspectives, from an NIH workshop. Am. J. Clin. Nutr. 112, 1390–1403.
McKillop, D.J., Pentieva, K., Daly, D., et al., 2002. The effect of different cooking methods on folate retention in various foods that are amongst the major contributors to folate intake in the UK diet. Br. J. Nutr. 88, 681.
McNulty, H., Pentieva, K., 2010. Folate bioavailability. In: Bailey, L.B. (Ed.), Folate in Health and Disease, second ed. CRC Press Boca Raton, pp. 25–48.
McNulty, H., McKinley, M.C., Wilson, B., et al., 2002. Impaired functioning of thermolabile methylenetetrahydrofolate reductase is dependent on riboflavin status: implications for riboflavin requirements. Am. J. Clin. Nutr. 76, 436–441.
McNulty, H., Dowey, L.R.C., Strain, J.J., et al., 2006. Riboflavin lowers homocysteine in individuals homozygous for the MTHFR 677C→T polymorphism. Circulation 113, 74–80.
McNulty, H., Strain, J.J., Hughes, C.F., et al., 2017. Riboflavin, MTHFR genotype and blood pressure: a personalized approach to prevention and treatment of hypertension. Mol. Aspects Med. 53, 2–9.
McNulty, H., Strain, J.J., Hughes, C.F., Pentieva, K., Ward, M., 2020. Evidence of a role for one-carbon metabolism in blood pressure: can B vitamin intervention address the genetic risk of hypertension owing to a common folate polymorphism. Curr. Dev. Nutr. 4 (1), nzz102.
Mills, J.L., Tuomilehto, J., Yu, K.F., et al., 1992. Maternal vitamin levels during pregnancies producing infants with neural tube defects. J. Pediatr. 120 (6), 863–871.
Molloy, A.M., Kirke, P.N., Troendle, J.F., et al., 2009. Maternal vitamin B12 status and risk of neural tube defects in a population with high neural tube defect prevalence and no folic acid fortification. Pediatrics 123, 917–923.
Morris, J.K., Addor, M.C., Ballardini, E., et al., 2021. Prevention of neural tube defects in Europe: a public health failure. Front. Pediatr. 9, 1–9.
Mosley, B.S., Cleves, M.A., Siega-Riz, A.M., et al., 2009. Neural tube defects and maternal folate intake among pregnancies conceived after folic acid fortification in the United States. Am. J. Epidemiol. 169, 9–17.
MRC Vitamin Study Research Group, 1991. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. Lancet 338, 131–137.
Nasri, K., Kacem, M., Fradj, B., et al., 2015. Association of maternal homocysteine and vitamins status with the risk of neural tube defects in Tunisia: a case-control study. Birth Defects Res. Part A 103, 1011–1020.
Psara, E., Pentieva, K., Ward, M., et al., 2020. Critical review of nutrition, blood pressure and risk of hypertension through the lifecycle: do B vitamins play a role? Biochimie 173, 76–90.
Public Health England, 2017. National Diet and Nutrition Survey Rolling Programme (NDNS) Supplementary Report: Blood Folate Results for the UK as a Whole, Scotland, Northern Ireland (Years 1 to 4 Combined) and Wales (Years 2 to 5 Combined). Revised 2017. Public Health England, London.
Ray, J.G., Wyatt, P.R., Thompson, M.D., et al., 2007. Vitamin B12 and the risk of neural tube defects in a folic-acid-fortified population. Epidemiology 18 (3), 362–366.
Rothenberg, S.P., da Costa, M.P., Sequeira, J.M., et al., 2004. Autoantibodies against folate receptors in women with a pregnancy complicated by a neural-tube defect. N. Engl. J. Med. 350, 134–142.
SACN, 2017. Folic Acid: Updated Recommendations Issued by the Scientific Advisory Committee on Nutrition (SACN). Public Health England, London.
Senousy, S.M., Farag, M.K., Gouda, A.S., et al., 2018. Association between biomarkers of vitamin B12 status and the risk of neural tube defects. J. Obstet. Gynaecol. Res. 44 (10), 1902–1908.
Smithells, R.W., Seller, M.J., Harris, R., et al., 1983. Further experience of vitamin supplementation for prevention of neural tube defect recurrences. Lancet 132, 1027–1031.
Steegers-Theunissen, R.P., Boers, G.H., Trijbels, F.J., et al., 1991. Neural-tube defects and derangement of homocysteine metabolism. N. Engl. J. Med. 324, 199–200.
Steegers-Theunissen, R.P., Boers, G.H., Blom, H.J., et al., 1995. Neural tube defects and elevated homocysteine levels in amniotic fluid. Am. J. Obstet. Gynecol. 172, 1436–1441.
Steen, M.T., Boddie, A.M., Fisher, A.J., et al., 1998. Neural-tube defects are associated with low concentrations of cobalamin (vitamin B12) in amniotic fluid. Prenat. Diagn. 18, 545–555.
Suarez, L., Hendricks, K., Felkner, M., et al., 2003. Maternal serum B12 levels and risk for neural tube defects in a Texas-Mexico border population. Ann. Epidemiol. 13, 81–88.
Vena, F., D'Ambrosio, V., Paladini, V., et al., 2021. Risk of neural tube defects according to maternal body mass index: a systematic review and meta-analysis. J. Matern. Fetal Neonatal Med. https://doi.org/10.1080/14767058.2021.1946789.
Vergel, R.G., Sanchez, L.R., Heredero, B.L., et al., 1990. Primary prevention of neural tube defects with folic acid supplementation: Cuban experience. Prenat. Diagn. 10, 149–152.
WHO, 2015. Guideline: Optimal Serum and Red Cell Folate Concentrations in Women of Reproductive Age for Prevention of Neural Defects. World Health Organization, Geneva, Switzerland.
Yang, Y., Chen, J., Wang, B., et al., 2015. Association between MTHFR C677T polymorphism and neural tube defect risks: a comprehensive evaluation in three groups of NTD patients, mothers, and fathers. Birth Defects Res. Part A 103, 488–500.
Zhang, T., Xin, R., Gu, X., et al., 2008. Maternal serum vitamin B12, folate and homocysteine and the risk of neural tube defects in the offspring in a high-risk area of China. Publ. Health Nutr. 12 (5), 680–686.

Further reading

Bailey, R.L., Mills, J.L., Yetley, E.A., et al., 2010. Unmetabolized serum folic acid and its relation to folic acid intake from diet and supplements in a nationally representative sample of adults aged >60 y in the United States. Am. J. Clin. Nutr. 92, 383–389. CrossRefView Record in ScopusGoogle Scholar.
Botto, L.D., Yang, Q., 2000. 5,10-Methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE review. Am. J. Epidemiol. 151, 862–877. CrossRefView Record in ScopusGoogle Scholar.
Daly, S., Mills, J.L., Molloy, A.M., et al., 1997. Minimum effective dose of folic acid for food fortification to prevent neural-tube defects. Lancet 350, 1666–1669. Article Download PDF View Record in ScopusGoogle Scholar.
Green, R., 2009. Is it time for vitamin B-12 fortification? What are the questions? Am. J. Clin. Nutr. 89, 712S–716S. CrossRefGoogle Scholar.
Kancherla, V., Botto, L.D., Rowe, L.A., et al., 2022. Preventing birth defects, saving lives, and promoting health equity: an urgent call to action for universal mandatory food fortification with folic acid. Lancet Glob. Health 10 (7), e1053–e1057. https://doi.org/10.1016/S2214-109X(22)00213-3. Epub 2022 May 23. PMID: 35617975.

Lawrence, M.A., Chai, W., Kara, R., et al., 2009. Examination of selected national policies towards mandatory folic acid fortification. Nutr. Rev. 67 (1), S73–S78. CrossRefView Record in ScopusGoogle Scholar.

Mills, J.L., 2010. What is the impact of dietary folic acid fortification on the risk of birth defects? In: Hales, B., Scialli, A., Tassinari, M.S. (Eds.), Teratology Primer, second ed. The Teratology Society, pp. 81–82. View Record in ScopusGoogle Scholar. http://www.teratology.org.

Molloy, A.M., Brody, L.C., Mills, J.L., Scott, J.M., Kirke, P.N., 2009. The search for genetic polymorphisms in the homocysteine/folate pathway that contribute to the etiology of human neural tube defects. Birth Defects Res. A 85, 285–294. CrossRefView Record in ScopusGoogle Scholar.

Pentieva, K., Selhub, J., Paul, L., et al., 2016. Evidence from a randomized trial that exposure to supplemental folic acid at recommended levels during pregnancy does not lead to increased unmetabolized folic acid concentrations in maternal or cord blood. J. Nutr. 46 (3), 494–500. CrossRefView Record in ScopusGoogle Scholar.

Pfeiffer, C.M., Caudill, S.P., Gunter, E.W., Osterloh, J., Sampson, E.J., 2005. Biochemical indicators of B vitamin status in the US population after folic acid fortification: results from the National Health and Nutrition Examination Survey 1999–2000. Am. J. Clin. Nutr. 82, 442–450. CrossRefView Record in ScopusGoogle Scholar.

Yang, Q., Cogswell, M.E., Hamner, H.C., et al., 2010. Folic acid source, usual intake, and folate and vitamin B-12 status in US adults: National Health and Nutrition Examination Survey (NHANES) 2003–2006. Am. J. Clin. Nutr. 91, 64–72. CrossRefView Record in ScopusGoogle Scholar.

Pregnancy: Safe diets

S. Stanner and H. Gibson-Moore, British Nutrition Foundation, London, UK

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of S. Stanner, H. Gibson-Moore, Pregnancy: Safe Diets, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 90–98, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00231-2.

Introduction 212
Food-Borne Infections during Pregnancy 212
Alcohol 213
Excessive Alcohol Consumption during Pregnancy 213
Binge Drinking and Social Alcohol Consumption during Pregnancy 214
Vitamin A 216
Fish and Pregnancy 216
Mercury 217
Dioxins and Polychlorinated Biphenyls (PCBs) 218
Caffeine 218
Avoiding Foods to Prevent Allergy 219
Food Additives and Herbal Supplements 220
Summary 220
Further Reading 221

Glossary

Congenital Acquired during development in the uterus and not through heredity.
Fetal alcohol spectrum disorders (FASD) Describes the range of effects that can occur in an individual whose mother had a high alcohol intake during pregnancy. These effects may include physical, mental, behavioral, and/or learning disabilities with possible lifelong implications.
Fetoplacental unit The fetus and the placenta as a single physiological unit.
Hydrocephalus An abnormal increase in the amount of cerebrospinal fluid within the cranial cavity that is accompanied by expansion of the cerebral ventricles, enlargement of the skull and especially the forehead, and atrophy of the brain.
Intracranial Affecting or involving intracranial structures.
Intrauterine Situated or occurring in the uterus.
Isoretinoin An oral retinoid (derivative of vitamin A).
Neonate An infant less than a month old.
Retinochoroiditis Inflammation of the retina and choroid.
Teratogenic Of, relating to, or causing malformations of an embryo or a fetus.

Introduction

A balanced diet that contains adequate amounts of all the nutrients needed by a mother and her growing fetus is essential for a healthy pregnancy. Pregnant women also need to be advised about how to reduce their risk of exposure to substances that may be toxic to the fetus during development (teratogenic) and therefore associated with the production of physical defects in the developing embryo (e.g., alcohol and excess vitamin A), as well as other dietary and lifestyle behaviors that could optimize maternal health and reduce the risk of health problems in their children. The aim of this article is to describe evidence relating to food safety issues during pregnancy, including potential risks to the fetus as a result of prenatal exposure to food pathogens or toxic food components (e.g., heavy metals and dioxins) and the potentially harmful effects of high doses of alcohol, caffeine, and vitamin A.

Food-Borne Infections during Pregnancy

For many years it has been recognized that food-borne antenatal infections may cause death or serious fetal damage. Women may be more susceptible to the effects of infection during pregnancy because of immunological changes leading to suppression of the

immune system (most commonly cell-mediated immunity), probably as a result of increases in pregnancy-associated sex steroids, such as oestradiol or progesterone. Among the most common causes of diarrhea during pregnancy are several food- or water-borne pathogens (bacteria, protozoa, or viruses), including salmonella species, *Helicobacter pylori*, *Shigella*, *Escherichia coli*, and *Cryptosporidium*. Hepatitis A is also a food- or water-borne pathogen of concern, particularly in countries where sanitation is poor. In pregnant women, severe vomiting and diarrhea may negatively affect the availability of important nutrients to the growing fetus. For example, impairment of the supply of folate (or the synthetic form, folic acid) during a critical stage of development could increase the risk of associated neural tube defects, such as spina bifida.

Although rare, infection with *Listeria* or *Toxoplasma* during pregnancy is of particular concern because even in a mild form these infections can prove fatal. Listeriosis caused by the consumption of food containing the bacterium *Listeria monocytogenes* leads to flu-like symptoms, such as fever, muscle aches, and sometimes nausea or diarrhea. If the infection spreads to the nervous system, it may also cause headaches, stiff neck, confusion, loss of balance, or convulsions. The bacterium has been found in a variety of raw foods, including unpasteurized (raw) milk and cheeses, uncooked meats, and vegetables, and in processed foods that become contaminated after processing, such as soft cheeses, pâtés, cold cuts of meat and smoked fish. According to the Centers for Disease Control and Prevention, pregnant women in the US are approximately 20 times more likely than other healthy adults to get listeriosis and approximately one-third of listeriosis cases occur during pregnancy. In the UK, the Health Protection Agency Centre for Infections reported approximately 30% of human cases to be pregnancy-associated (both mother and neonate affected) in England and Wales between 1983 and 2009. The fetus and newborn are at greatest risk of this infection and its consequences can be severe, leading to miscarriage, stillbirth, and premature delivery or to meningitis in the newborn infant. When infection occurs during pregnancy, antibiotics given promptly to the pregnant woman can often prevent infection of the fetus or newborn, and infants developing the infection can also be treated in the same way.

Toxoplasma gondii is a parasite that can be transmitted to the fetus *in utero* through transplacental transmission, causing stillbirth, miscarriage, or mental retardation (in immunocompetent people infection is asymptomatic or mild). The parasite has been found in raw, inadequately cooked or cured meat, cat feces, and unwashed raw fruit and vegetables. It has also occasionally been reported in unpasteurized goats' milk. In the UK, toxoplasmosis occurs in approximately 2.5–5.5 in 1000 pregnant women (1750–2850 cases per year), generally causing flu-like symptoms, swollen lymph glands, or muscle aches and pains that last for a few days to several weeks. If a pregnant woman contracts the infection, there is an approximately 30–40% chance of fetal infection (congenital toxoplasmosis). The incidence and risk of placental transmission is dependent on the trimester during which maternal infection is contracted. The risk of transmission in the first trimester is estimated to be 10–15% and the outcome if contracted at this stage can be severe or life-threatening to the fetus. Conversely, the risk rises to 70–80% in the third trimester but the clinical outcome is usually less severe, or may be asymptomatic. Congenital toxoplasmosis can cause hearing loss, hydrocephalus, eye and brain damage, epilepsy, growth retardation, intrauterine death, and other problems. Congenital toxoplasmosis affected 0.34 per 10 000 births in England and Wales in 2002–2004. Studies over the past 15 years have estimated between 1 and 10 per 10 000 births elsewhere in Europe, and similar rates for the US. The most common symptoms observed in newborns are retinochoroiditis (inflammation of the retina and choroid) and/or intracranial abnormalities (with or without developmental delay). In Europe, 1% or 2% of infected infants develop learning difficulties or die and 4–27% develop permanent loss of vision. Mothers can be tested to determine if they have developed an antibody to the infection. Fetal testing may include ultrasound and testing of amniotic fluid or cord blood. When toxoplasmosis is diagnosed during pregnancy, antibiotic treatment can often help reduce the severity of symptoms in the newborn.

The risk of food poisoning can be minimized by ensuring adequate attention to good hygienic practice when preparing, cooking, and storing foods (**Table 1**). Pregnant women should therefore be advised of the need for a high regard for food hygiene and personal cleanliness during this vulnerable time. In addition, there are a few foods that may pose a particular (although small) risk, which should be avoided during pregnancy where possible (**Table 2**).

Alcohol

Excessive Alcohol Consumption during Pregnancy

Chronic alcohol abuse may result in a wide spectrum of secondary disturbances of the absorption and utilization of many nutrients, including glucose, amino acids, fat, sodium, and some vitamins (especially thiamin, vitamin B_{12}, and folate). The inhibition of folate absorption by alcohol is of particular concern because of the risk of neural tube defects associated with an inadequate supply of this vitamin to the fetus before conception and during the first trimester of pregnancy. Alcohol may also directly impair the placental transfer of nutrients essential for growth (e.g., amino acids), which at critical phases of fetal organogenesis could compound any direct fetotoxic effects of ethanol or acetaldehyde.

Both alcohol and its primary metabolite, acetaldehyde, are teratogenic. Excessive alcohol consumption (>80 g of ethanol or 10 units per day) during pregnancy can result in a child being born with a specific combination of physical and mental disabilities known as fetal alcohol syndrome (FAS). Such fetuses usually survive until birth but are growth retarded and display a characteristic range of clinical features, principally craniofacial abnormalities and neurological damage (**Table 3**). However, FAS is not the only outcome of prenatal alcohol exposure, and it has been suggested that it can present as a spectrum of disorders. Fetal alcohol spectrum disorder (FASD) is an umbrella term used to describe a range of effects that can occur due to the presence of alcohol during the prenatal stage. FASD is characterized by the presence of some of the criteria for FAS and is associated with lesser degrees of harm from maternal alcohol consumption.

Table 1 General guidelines on good hygienic practices in the home

The risk of food poisoning can be minimized by adopting the following practices:

Cleanliness in the kitchen

- Keeping all work surfaces scrupulously clean
- Washing cooking utensils after coming into contact with raw meat, poultry or eggs to prevent cross contamination
- Using separate chopping boards for foods that are to be cooked (e.g., raw meat)
- Keeping kitchen cloths clean; rinsing crockery in hot water, leaving it to dry, then wiping it with a clean tea towel
- Using kitchen towels to mop up spills, rather than a dishcloth
- Ensuring waste bins are covered and away from food and keeping pets away from the kitchen.

Hygienic food handling

- Washing all equipment and work surfaces before and after touching raw food
- Washing all foods to be eaten raw thoroughly
- Cooking meat thoroughly to an internal temperature of at least 70 °C
- Keeping raw and cooked foods separated during preparation and storage
- Cooling cooked foods as quickly as possible if they are to be stored in a fridge or freezer
- Covering foods and not leaving them standing around in the kitchen
- Storing food at the correct temperature (less than 4 °C in the fridge or less than −18 °C in the freezer)
- Storing raw meat, well covered, at the bottom of the fridge
- Storing eggs in a fridge, if possible
- Never overloading the fridge as this can reduce the circulation of cool air
- Keeping foods for as short a time as possible (especially meat and fish) and following storage instructions (i.e., not using beyond the 'use by' or 'best before' date)
- Thawing frozen meat thoroughly before cooking
- Avoiding reheating and food more than once
- Reheating foods thoroughly (if this is done in a microwave, the standing times recommended by the manufacturer should be observed to ensure that food attains an even temperature before it is eaten).

Personal hygiene

- Washing hands thoroughly before preparing food, after visiting the toilet and after emptying the rubbish bin
- Never licking fingers or utensils and put them back into food
- Washing hands after blowing or touching the nose whilst handling food
- Keeping nails clean and hair out of food
- Wearing a clean apron
- Not handling food during periods of illness, e.g., heavy cold, sickness or diarrhea
- Covering all cuts, spots and pimples, particularly on the hands, with a waterproof dressing and replacing it often
- Wearing rubber gloves when emptying cat litter trays.

The extent of the damage from alcohol varies depending on the stage of development at which high doses are encountered. The fetus is most vulnerable to organ damage from the time the umbilical cord begins to function (5 weeks) to the completion of organ development (11 weeks). Inhibition of growth and neurobehavioral development occurs in the second and third trimester. Although the facial features of FAS become more subtle with age, growth deficits and central nervous system impairment may be permanent. The reported worldwide incidence of FAS is 0.97 in 1000 births; however, the incidence of FASD remains unclear largely due to the absence of robust and routine data collection. Nevertheless, it is clear that FASD is more common in some populations than others. For example, in Australian aboriginal populations the incidence of FASD is estimated at 5 per 1000 births, whereas in the Western Cape Province of South Africa the incidence of FASD exceeds 60 cases per 1000 births.

FAS is only seen in infants born to women who are excessive drinkers, but it is not an inevitable result of heavy drinking in pregnancy, and even children born to mothers who are active alcoholics may not show it. This differing susceptibility of fetuses to the syndrome is thought to reflect the interplay of genetic factors, social deprivation, nutritional deficiencies, and tobacco and other drug abuse, along with alcohol consumption.

Binge Drinking and Social Alcohol Consumption during Pregnancy

Binge drinking is generally defined for women, as the consumption of four or more drinks in approximately 2 h (Centers for Disease Control and Prevention) or at least 6 units of alcohol (**Table 4**) per occasion (Strategy Unit, London) in the US and in the UK, respectively, and may be particularly harmful because it exposes the fetus to high blood alcohol concentrations over relatively short periods of time and may be associated with repeated withdrawal episodes. Animal studies have demonstrated binge-like exposure to alcohol to be as teratogenic as long-term exposure throughout gestation, even if the overall alcohol amount consumed by binge drinking is less than intake during more continuous drinking patterns. Human studies have found associations with binge drinking and neurodevelopment outcomes, such as an increase in 'disinhibited behavior,' a reduction in verbal IQ, an increase in delinquent

Table 2 Foods to avoid during pregnancy

Foods to avoid	*Reason*
Some types of cheese[a]: • Mould-ripened cheeses, e.g., Camembert, Brie • Some goats' cheeses • Soft blue cheeses, e.g., Stilton All types of pâté (including vegetable)	To minimize risk of listeriosis
Unwashed fruit and vegetables Raw or undercooked meat Cured meats, e.g., Parma ham and salami Unpasteurized goats' milk or goats' cheese	To minimize risk of toxoplasmosis
Raw or partially cooked eggs or foods made from them, e.g., homemade mayonnaise, soft-whip ice cream, cake mix, mousses, and hollandaise sauce (eggs should be cooked until both the white and yolk are solid. Raw or undercooked meat, poultry, shellfish (e.g., oysters) and fish (e.g., smoked salmon, trout, sushi) Undercooked ready meals and ready-to-eat poultry (unless they have been reheated to a very high temperature). Unpasteurized milk and milk products Untreated water	To minimize risk of food poisoning from Salmonella and Campylobacter
Liver products and supplements containing vitamin A or fish liver oils	To avoid excess vitamin A intake
Some types of fish: shark, swordfish, king mackerel, tilefish, and marlin Limit intake of tuna, no more than two tuna steaks a week (approximately 140 g cooked or 170 g raw each) or four medium-size cans of tuna a week (approximately 140 g when drained) No more than two portions of oily fish per week, e.g., fresh tuna (not canned tuna, see above), salmon, mackerel, sardines and trout.	To avoid high intakes of mercury and other contaminants

[a]Foods containing these cheeses that have been properly cooked will be safe to eat.

behavior and learning problems. However, findings for other health outcomes have been inconsistent, possibly because of the notorious problems of recording binge drinking during pregnancy.

The question of whether moderate or occasional alcohol consumption is safe during pregnancy has been widely debated. Currently, there is little evidence that modest drinking (<10 units per week) has any harmful effects. In a systematic review of the evidence, reported in the National Institute of Health and Clinical Excellence (NICE) 2008 guidelines "Antenatal care: routine care for the healthy pregnant woman," the highest risk suggested was a three- to fourfold increase in the risk of miscarriage in women

Table 3 Symptoms of fetal alcohol syndrome (FAS)[a]

The diagnosis of FAS requires signs in all of the three following categories: • Prenatal and postnatal growth retardation Intrauterine growth retardation, including smaller than normal head circumference, continued growth below the 10th centile and failure to thrive. • Central nervous system involvement Neurological abnormalities, developmental delay, intellectual impairment, brain malformation, hearing and visual disabilities. • Physical anomalies Characteristic facial deformity including short upturned nose, receding forehead and chin, smaller than normal eye apertures, absent philtrum and asymmetrical ears.

[a]The diagnosis of FAS requires signs in all three of the categories.

Table 4 Definition of a unit of alcohol

1 unit of alcohol approximately equals 8 g of absolute alcohol, which is equivalent to:
½ pint of ordinary strength beer, lager, or cider
¼ pint of strong beer or lager
1 small glass wine
1 single measure of spirits
1 small glass sherry

who consume approximately 10 units of alcohol per week compared to abstainers. For other health outcomes, the majority of the studies did not report a statistically significantly increased risk with low-to-moderate alcohol intake, but the evidence was not entirely consistent. Many studies are confounded by factors such as cigarette smoking, social class, drug abuse, very high levels of caffeine intake, and different cross country categorization of 'light,' 'moderate,' and 'heavy' drinking. Although there is general agreement that women should not drink alcohol excessively during pregnancy, a consensus opinion of a safe threshold level of alcohol consumption has not been established at any stage of pregnancy, and advice differs among countries.

Despite the lack of evidence of detrimental effects on any outcome at low-to-moderate maternal alcohol consumption, many professional bodies err on the side of caution. The Royal College of Obstetricians and Gynaecologists and NICE suggest that the only way to minimize any harmful effects to the fetus from alcohol is to not drink at all during pregnancy. The UK Department of Health advises that pregnant women or those planning to conceive to avoid alcohol completely and if they do choose to drink they should not consume more than 1 or 2 units once or twice a week (Table 4), and should avoid intoxication. Advice in North America (US and Canada) is that women should not consume alcohol at all during pregnancy, and there are warnings on products and advertisements. Anecdotally, many pregnant women develop a spontaneous aversion to the taste and/or smell of alcoholic beverages and so may limit their intakes anyway.

Vitamin A

During the period of early development, the supply of preformed vitamin A (retinol) must be carefully managed to ensure that the developing fetus is exposed to neither too little nor too much of the nutrient because either condition can have teratogenic consequences. Adequate vitamin A is required for normal embryonic development, and an insufficient supply during pregnancy can result in malformations in the offspring as well as increased mortality and morbidity during early childhood from infectious diseases, such as diarrhea, measles, and respiratory infections.

Excess vitamin A intake has also been associated with teratogenicity in animals and may represent a risk in humans, particularly within the first trimester of pregnancy. Characteristic features include severe motor deficit and malformations of the heart, thymus, face, jaw, ears, palate, and brain. Although adverse effects from dietary sources are very rare, events have occurred with the ingestion of high-dose supplements and with isoretinoin treatment for severe acne; this medication is now only permitted under strong supervision in women of reproductive age. Therefore, as presented in the SACN report 'The influence of maternal, fetal and child nutrition on the development of chronic disease in later life,' pregnant women are advised to avoid taking preformed vitamin A (retinol) supplements and to also avoid consuming liver or liver products. This advice is based on the risk of teratogenesis associated with retinol consumption at doses higher than 3000 μg retinol equivalents (RE) per day.

In Western countries, where vitamin A deficiency is rare, women who are or might become pregnant are advised against taking vitamin A supplements (including cod liver oil), except on the advice of a doctor or antenatal clinic, and not to consume liver or liver products. In developing countries with a high prevalence of vitamin A deficiency, vitamin A supplementation programs have resulted in decreased pregnancy-related mortality and lower rates of childhood morbidity and mortality, with benefits clearly outweighing any potential risks. The initiation of such programs in any population should be carefully examined in each case according to the risk–benefit ratio, with the final decision taking into account the vitamin A status of the population, the availability of vitamin A-rich foods, and whether supplementation can be supervised. The World Health Organization recommends that high-dose vitamin A supplements for women be restricted to the first 6 weeks postpartum, before they are likely to become pregnant again.

Fish and Pregnancy

Fish is a good source of protein, vitamins, and minerals. In particular, oil-rich fish (e.g., mackerel, salmon, kippers, herrings, trout, sardines, and fresh tuna) contain the long-chain *n*-3 fatty acids eciosapentenoic acid (EPA) and docosahexenoic acid (DHA), which may confer many health benefits to the developing fetus. For example, DHA is required for nerve and retinal development, and eating oily fish has been found to have a slight beneficial effect on birth weight and length of gestation. In the US and Canada, the position of the American Dietetic Association (ADA) and Dietitians of Canada, is that adults, including pregnant and lactating women, should consume a combined intake of 500 mg day^{-1} of DHA and EPA. In the UK, 450 mg day^{-1} of DHA and EPA

combined is recommended for adults and pregnant women, the equivalent of consuming two servings (one serving=140 g) of fish a week, one of which is oil-rich. However, consumption of oil-rich fish has been positively associated with intakes of certain contaminants, namely mercury, dioxins, and polychlorinated biphenyls (PCBs), and concern has been expressed about the consequences of prenatal exposure to these toxic chemicals on the risk of brain and nervous system abnormalities. As a result, pregnant women in the UK are advised not to consume more than two portions of oily fish a week.

Mercury

Mercury is a metal that is present in the environment from natural and man-made sources (e.g., coalburning or other industrial pollution). It is converted primarily by microorganisms to a more toxic form, methylmercury, which is bioaccumulated in the aquatic food chain, reaching its highest levels in large, longer living predatory fish. Among humans, the sole source of exposure to methylmercury is the consumption of fish and sea mammals.

Methylmercury is neurotoxic and accumulates in the brain and central nervous system. It inhibits the division and migration of neuronal cells and disrupts the cytoarchitecture of the developing brain. Although a mother may show no signs of neurotoxicity, the developing fetus may be damaged following exposure to methylmercury. The concentration of methylmercury in fetal brain has been shown to be 5–7 times higher than that in maternal blood, and it has been estimated that the fetus is 5–10 times more sensitive to methylmercury exposure than an adult, although the reason for this is unknown.

Disasters in Minamata, Japan, in the 1950s and in Iraq in 1971–72 demonstrated that acute prenatal exposure may result in severe mental retardation, cerebral palsy, blindness, and deafness. However, whether exposure to lower chronic doses, which may occur if pregnant women consume large amounts of fish, can also lead to adverse neurodevelopmental consequences is less certain. Large, long-term prospective epidemiological studies of high fish-eating populations have not found a consistent pattern of association between exposure and neuropsychological outcomes. Although subtle neuropsychological changes were reported in a study of children in the Faroe Islands study, where exposure was mainly from whale consumption, a similar study in the Seychelles found no adverse effects from fish consumption alone.

The Joint FAO/WHO Committee on Food Additives revised its safety guideline for weekly intake of methylmercury, known as the provisional tolerable weekly intake (PTWI), to 1.6 mg kg^{-1} body weight per week. The UK government's independent expert Committee on Toxicity of Chemicals in Food, Consumer Products, and the Environment (COT) has applied a lower PTWI limit of 0.7 mg kg^{-1} body weight per week to women who are pregnant or those intending to become pregnant and to mothers who are breast-feeding.

Any public health recommendations to pregnant women regarding fish consumption must recognize the important role that it plays as part of a healthy, balanced diet. Most fish contain trace amounts of methylmercury, but high concentrations of the metal have only been found in large, predatory fish, such as shark, marlin, and swordfish (Table 5). If a pregnant or breast-feeding mother were to consume one portion of these predatory fish, she would exceed the lower PTWI set by COT and the EPA by 400%. Therefore, as a precaution, pregnant women, breast-feeding mothers, and those who intend to become pregnant within the next 12 months are advised to avoid consumption of these types of fish (in the US, this also includes king mackerel and tilefish). Some samples of tuna have also been found to have higher levels than other species. In the UK, pregnant women (and those who may become pregnant) are advised to restrict their weekly intake to two 140 g portions of fresh tuna or four 140 g portions of canned tuna.

Table 5 Concentrations of methylmercury (mg kg^{-1}) in surveyed fish in the UK

Fish	*Methylmercury mg kg^{-1}*
Shark	1.52
Swordfish	1.35
Marlin	1.09
Fresh Tuna	0.40
Canned Tuna	0.19
Herring	0.09
Pink Shrimps	0.09
Cod	0.07
Plaice	0.06
Mackerel	0.05
Haddock	0.04
Scallops	0.01

Dioxins and Polychlorinated Biphenyls (PCBs)

Fish can also contain other organic pollutants, such as PCBs and dioxins. Whereas mercury accumulates in the muscles of larger predatory fish, PCBs and dioxins are found in the fatty tissues of fish. Most human exposure to PCBs and dioxins comes from dietary sources because they accumulate in the lipid fractions of meat, fish, milk and milk products, eggs, grains, and oils.

PCBs and dioxins have been linked with increased rates of some cancers in studies of individuals exposed to high amounts through either vocational exposure or accidental environmental contamination. Prenatal exposure to large amounts of these pollutants (e.g., through contaminated fish) has been associated with neurobehavioral alterations in newborn children. Some studies have also suggested that exposure to smaller quantities of PCBs and dioxins in utero may lead to more subtle cognitive and motor developmental delays, although a favorable home environment appears to counteract any effect. However, the difficulty of separating the effects of PCBs and dioxins from potentially confounding factors (e.g., exposure to other contaminants, breastfeeding, smoking, and maternal education) makes it difficult to reach firm conclusions. Further research is also needed to ascertain whether any cognitive changes are temporary or persist into later life.

The potency of dioxins is expressed as toxic equivalents (TEQs), which have been internationally accepted. In the UK, the tolerable daily intake (TDI) recommended by COT is 2 pg TEQ kg^{-1} body weight, which is in line with recommendations of other international and European expert committees. In common with the US and the European Union, approximately one-third of the UK population may exceed the TDI in their daily diet. The TEQ, therefore, provides a target to reduce dioxins and PCBs in the environment internationally. Since the 1960s, following the prohibition of many dioxins and PCBs by governments, concentrations have been declining in breast milk, which is commonly used to determine exposure. For example, between 1982 and 1997, consumption of dioxins and PCBs in the UK decreased by 75% and between 1997 and 2001 fell by a further 50%, largely due to strict controls concerning production, use, and disposal of PCBs and dioxins: it is anticipated that intakes will continue to decrease.

Caffeine

Caffeine is a methylated xanthine that acts as a mild central nervous system stimulant. It is the most widely consumed xenobiotic in pregnancy and is found in a number of foods and beverages (Table 6). The main sources are coffee, tea, cocoa, chocolate, and soft drinks, as well as prescription and nonprescription medicines, such as diet pills, headache treatments, and cold and flu medicines. Tea and cocoa also contain significant quantities of theophylline and theobromine, which are caffeine derivatives that have not been as widely researched.

Once caffeine and its derivatives are consumed, they are absorbed into the blood and body tissues and can cross the placenta to the fetus. Cytochrome P450 1A2 (CYP1A2), the principal enzyme involved in caffeine metabolism in the liver, is absent in the placenta and the fetus, therefore the exposure of the fetoplacental unit to caffeine depends on maternal caffeine metabolism, which is influenced by genetic and environmental factors. Typically the body metabolizes caffeine more slowly during pregnancy, especially in the last few months; the half-life of caffeine increases from approximately 5–18 h during the second and third trimesters. Blood caffeine concentrations are therefore raised during pregnancy with no change in intake. In contrast, smoking is known to increase caffeine metabolism appreciably.

Although very high doses of caffeine are teratogenic in animals, no link between consumption during pregnancy and birth defects has been demonstrated in humans. However, high maternal caffeine intakes (>500 mg day^{-1}) have been associated with increased fetal heart rate and newborn cardiac arrhythmias. Maternal caffeine intake has been reported to be associated with low birth weight, but the safe threshold level remains unknown. A number of studies have found an association with caffeine intakes greater than 300 mg day^{-1} and fetal growth restriction and some have also demonstrated increased risks at intakes as little as 141 mg day^{-1}. In 2001, the Committee on Toxicology of Chemicals in Food, UK, conducted a thorough review of the literature and concluded that although caffeine consumption of more 300 mg day^{-1} might be associated with spontaneous miscarriage and low birth weight, the evidence was inconclusive.

Table 6 The caffeine content of beverages and foods

1 cup (190 ml) of instant coffee: ~75 mg
1 cup (190 ml) of brewed coffee (filter or percolated): ~100–115 mg
1 cup (190 ml) of decaffeinated coffee (brewed or instant): ~4 mg
1 cup (190 ml) of tea: ~50 mg
1 cup (200 ml – using manufacturers' instructions) of drinking chocolate: 1.1–8.2 mg
250 ml serving of energy drinks (containing either caffeine or guarana): 28–87 mg
330 ml serving of cola (regular and diet) – 11–70 mg
50 g bar of chocolate – 5.5–50 mg

Table 7 200 mg of caffeine is roughly equivalent to

3 average cups or 2 average size mugs of instant coffee
2 average cups of brewed coffee
4 average cups of tea or 2 mugs of tea
5 cans of regular cola drinks (e.g., 40 mg each)
2 cans of 'energy' drinks (e.g., 80 mg each)
200 g (4 standard 50 g bars) of milk chocolate (e.g., 50 mg each)

The lack of consistency between studies, particularly in relation to the dose at which an effect is reported, has made it very difficult to identify a threshold level of caffeine intake that presented an increased risk during pregnancy. However, in 2008 a large prospective observational study was conducted to reduce the uncertainties of previous evidence and provide a more robust basis on which to advise pregnant women on caffeine consumption. Results of the study, published in the British Medical Journal, linked maternal caffeine intake with increased risk of fetal growth restriction and concluded that sensible advice would be to reduce caffeine intake before conception and throughout pregnancy. Based on this evidence, guidance has changed in the UK on the consumption of caffeine for women trying to conceive and pregnant women, to limit their daily intake to 200 mg day^{-1} (previous recommendation was a maximum daily intake of 300 mg day^{-1}). This is the equivalent of approximately two mugs of coffee a day (Table 7).

The revised level in the UK of 200 mg day^{-1} is endorsed by the March of Dimes in the USA, however, the ADA, in their 2008 Position Paper, suggest 300 mg day^{-1} as a safe upper limit and this is in line with the advice given by the EU Scientific Committee on Foodstuffs. In practice, many pregnant women reduce their coffee intake as a result of a spontaneous aversion to the taste and smell, particularly in early pregnancy. For example, in the UK average daily caffeine consumption decreased from 238 to 139 mg day^{-1} during the first trimester, and then increased to an average of 153 mg day^{-1} by the third trimester of pregnancy, so most pregnant women are unlikely to be affected by the change in advice.

Avoiding Foods to Prevent Allergy

Food allergy has been estimated to affect approximately 3–7% of infants and young children in Western Europe, although the majority of children outgrow food allergies by the time they start school. Prevalence is assumed to be increasing in line with other forms of atopic disease, although evidence to support this is limited. Some food allergies (e.g., peanut allergies) can persist into adulthood and in severe cases can be life threatening. Most confirmed food allergies are associated with a relatively limited range of foods, including cows' milk, eggs, tree nuts, peanuts, soybeans, wheat, fish, and shellfish.

The development of food allergy depends on several factors, including genetic factors and early exposure to allergenic proteins in the diet, food protein uptake and handling, and the development of tolerance. However, it remains uncertain whether sensitization occurs *in utero* and, if so, whether this occurrence is restricted to specific stages of gestation. There is little evidence to support any benefit of avoiding specific foods during pregnancy to reduce the risk of allergic disease in a genetically susceptible child. Indeed, such a strategy may be counterproductive because it has been suggested that exposure to foreign proteins that cross the placenta is important to establish a normal immune response that enables the infant to develop normal tolerance to the many foreign proteins in the environment. Because restrictive diets may limit the supply of essential nutrients, these should only be practiced under medical supervision. Inappropriate and unnecessary exclusion of foods could prevent both mother and infant from obtaining the nutrients they need, resulting in significantly reduced weight gain and a tendency toward lower birth weights.

Owing to the severity of reactions experienced from peanut allergy, some countries have issued specific advice around intake during pregnancy. In the UK, COT had previously issued cautionary advice that women may wish to avoid peanuts during pregnancy and breast-feeding and not introduce peanuts into their child's diet before three years of age, if their child has a family history of allergy. However, in 2009 the Government revised its advice based on a systematic review which indicated that there is no clear evidence that eating or not eating peanuts (or foods containing peanuts) during pregnancy, breast-feeding or early childhood has any effect on the chances of a child developing a peanut allergy. Pregnant women in the UK are now advised that they can choose to eat peanuts irrespective of whether their child has a family history of allergies. Similarly in the US, the American Academy of Pediatrics previously suggested that pregnant women avoid peanuts, until research from the Avon Longitudinal Study of Parents and Children (ALSPAC) reported no association between maternal consumption of peanuts during pregnancy and childhood peanut allergy. The March of Dimes in the US now advises that women who are not allergic to peanuts can safely eat peanuts during pregnancy.

Food Additives and Herbal Supplements

Pregnant women often express concern about food additives. However, all additives have to be approved as safe for almost all but a small proportion of the population who may experience rare reactions to them before they can be used in foods. The presence of an 'E' number demonstrates that it has passed safety tests and been approved for use by the European Community. In the UK, COT sets an Acceptable Daily Intake for each additive, which is the amount that can be consumed daily with no risk to health. This may limit the amount of an additive used or restrict its use to certain food products. Even when an additive has been approved, new research is constantly reviewed and approval for any additive will be withdrawn if doubt is raised about its safety.

A number of herbal supplements and preparations may be used during pregnancy, most commonly to relieve gastrointestinal symptoms. Although the use of many herbal remedies is safe during pregnancy, this cannot be assumed simply because a product is described as 'natural.' Many plants, trees, fungi, and algae can be poisonous to humans, and several pharmaceuticals have been developed or derived from these sources because of the powerful compounds they contain. Very few randomized, clinical trials have examined the safety and efficacy of alternative therapies during pregnancy, and women should be warned to use any medicine, including herbal remedies, with care during pregnancy and with the advice of a doctor or pharmacist.

Summary

In addition to the consumption of a healthy, balanced diet, there are food safety precautions that need to be followed to ensure a safe pregnancy. A summary of the evidence and current advice described in this article is given in **Table 8**. However, it is important to highlight that recommendations during pregnancy across the globe, continue to change from time to time as new evidence emerges and guidelines are published. A recent virtual issue report by *Maternal & Child Nutrition* and *Nutrition Bulletin* has been published which pulls together articles describing these recent changes in nutritional recommendations which are also summarized in the above article.

Table 8 A summary of advice regarding dietary habits and foods safe during pregnancy

- Pregnant women should pay careful attention to food and personal hygiene so as not to expose themselves to any risk of food poisoning, which is not only highly unpleasant but also potentially very dangerous to the unborn child in some cases (e.g., with listeriosis and toxoplasmosis).
- Foods that have been linked with the bacteria *Listeria monocytogenes* should be avoided. These include pâtés and mould-ripened, soft cheeses (e.g., brie and camembert). Preprepared foods should be heated until they are piping hot and fruit and vegetables washed well, especially if they are to be eaten raw.
- To reduce the risk of toxoplasmosis, pregnant women should avoid eating raw or uncooked meat, unpasteurized goats' milk or goats' cheese, or unwashed fruit and vegetables. After handling raw meat, chopping boards, utensils and hands should be washed thoroughly. When gardening or emptying cat litter trays, rubber gloves should always be worn.
- Undercooked foods (e.g., meat, poultry, eggs), foods containing raw egg (e.g., mayonnaise, soft whip ice cream) and raw fish (e.g., sushi, smoked salmon) should also be avoided.
- Drinking alcohol heavily throughout pregnancy (>80 g or 10 units per day) is linked with fetal alcohol syndrome. Modest drinking (<10 units per week) does not appear to have harmful effects but most professional bodies err on the side of caution and recommend that pregnant women abstain from drinking alcohol or limit their consumption to no more than 1–2 units per day.
- Supplements containing high doses of preformed vitamin A and foods containing large amounts of this vitamin (liver and liver products) are best avoided in countries where intake from a well-balanced diet should be sufficient. In areas of endemic vitamin A deficiency, supplementation can reduce pregnancy-related mortality and reduce rates of childhood morbidity and mortality. However, most experts agree that preformed vitamin A supplements in doses of more than 3000 RE should not be taken by women who may become pregnant. Beta-carotene is safe for pregnant women.
- Although it is not presently clear if intakes of mercury and other contaminants such as PCBs and dioxins at levels that can be obtained from eating fish can influence children's neurological development, government organizations in a number of countries recommend that pregnant women avoid species of fish that have been found to contain high levels of these substances. This includes shark, marlin, swordfish, tilefish and king mackerel. Some countries have also recommended limiting tuna intake (e.g., in the UK pregnant women are encouraged to consume no more than two tuna steaks or four medium-size cans of tuna per week).
- Consumption of caffeinated beverages (e.g., coffee, tea and colas) has been associated with miscarriage and low birth weight, although many studies are confounded by high alcohol intakes, smoking, and drug and other substance abuse. In the UK, the recommendation is that pregnant women limit their caffeine intake to 200 mg day^{-1} which is endorsed by the March of Dimes in the US, however, the ADA and EU Scientific Committee on Foodstuffs suggest a safe upper limit of 300 mg day^{-1}.
- Avoiding specific foods during pregnancy is unlikely to reduce the risk of allergic disease in a susceptible child. In the UK, the Government recommends if women would like to eat peanuts or foods containing peanuts during pregnancy, they can choose to do so as part of a healthy, balanced diet. The March of Dimes in the US now advises that women who are not allergic to peanuts can safely eat peanuts during pregnancy.
- Additives permitted for use in foods undergo stringent safety tests over a long period of time before being approved and are safe for consumption during pregnancy by all but a small proportion of women who experience rare reactions to specific additives.
- Many pregnant women who would not consider taking over-the-counter medications often view herbal products as a safe and natural alternative. However, very few randomized, clinical trials have examined the safety and efficacy of alternative therapies during pregnancy. Pregnant women should be advised to seek advice from a doctor or pharmacist before taking any medication, including herbal supplements.

Further Reading

Azais-Braesco, V., Pascal, G., 2000. Vitamin A in pregnancy: Requirements and safety limits. American Journal of Clinical Nutrition 71, 1325S–1333S.

British Medical Association (BMA) (2007) *Fetal Alcohol Spectrum Disorders – A Guide for Healthcare Professionals*. Available at http://www.bma.org.uk/health_promotion_ethics/alcohol/Fetalalcohol.jsp

CARE study group, 2008. Maternal caffeine intake during pregnancy and risk of fetal growth restriction: A large prospective observational study. British Medical Journal 337, a2332.

Clarkson, T.W., 2003. The toxicology of mercury – Current exposures and clinical manifestations. New England Journal of Medicine 349, 1731–1737.

Cook, A.J., Gilbert, R.E., Buffolano, W., et al., 2000. Sources of toxoplasma infection in pregnant women: European multicentre case-control study. European Research Network on Congenital Toxoplasmosis. British Medical Journal 321, 142–147.

Committee on the Toxicity of Chemicals in Food Consumer Products and the Environment, 2000. Adverse Reactions to Food and Food Ingredients. Food Standards Agency, London. Available at http://cot.food.gov.uk/cotreports/cotwgreports/foodreactions.

Committee on the Toxicity of Chemicals in Food Consumer Products and the Environment, 2001. Statement on the Reproductive Effects of Caffeine. Statement 2001/06. COT, London. Available at http://cot.food.gov.uk/cotstatements/cotstatementsyrs/cotstatements2001/caffeine.

Committee on the Toxicity of Chemicals in Food Consumer Products and the Environment, 2001. Statement on the Tolerable Daily Intake for Dioxins and Dioxin-Like Polychlorinated Biphenyls. COT, London. Available at http://cot.food.gov.uk/pdfs/cot-diox-full.pdf.

Committee on the Toxicity of Chemicals in Food Consumer Products and the Environment (2002) *Statement on a Survey of Mercury in Fish and Shellfish*. Available at http://cot.food.gov.uk/pdfs/COTmercurystatement.PDF

Food Standards Agency (2003) *Mercury in Imported Fish and Shellfish, UK Farmed Fish and Their Products (40/03)*. Available at http://www.food.gov.uk/science/surveillance/fsis2003/fsis402003

Joint FAO/WHO Expert Committee on Food Additives (2004) *Evaluation of Certain Food Additives and Contaminants, Sixty First Report. WHO Technical Report Series 922*. Available at http://whqlibdoc.who.int/trs/WHO_TRS_922.pdf

Maternal & Child Nutrition & Nutrition Bulletin (2011) *Joint Virtual Issue: Nutrition During Pregnancy*. Available at: http://onlinelibrary.wiley.com/journal/10.1111/28ISSN291467-3010/homepage/custom_copy.htm

National Institute of Health and Clinical Excellence (NICE) (2008) *Antenatal care: Routine Care for the Healthy Pregnant Woman*. Available at http://www.nice.org.uk/CG62

Royal College of Obstetricians and Gynaecologists (2006) *Alcohol Consumption and the Outcomes of Pregnancy*. Available at http://www.rcog.org.uk/womens-health/clinical-guidance/alcohol-consumption-and-outcomes-pregnancy

Scientific Advisory Committee on Nutrition (SACN), 2004. Advice on Fish Consumption: Benefits & Risks. TSO, London. Available at http://cot.food.gov.uk/pdfs/fishreport2004full.pdf.

Science Advisory Committee on Nutrition (SACN) (2011) *The Influence of Maternal, Fetal and Child Nutrition on the Development of Chronic Disease in Later Life*. Available at http://www.sacn.gov.uk/pdfs/sacn_early_nutrition_final_report_20_6_11.pdf

Scientific Committee on Food (1999) *Opinion on Caffeine, Taurine and D-Glucurono-g-lactone as Constituents of So-Called 'Energy' Drinks (Expressed on 21 January 1999)*. Available at http://ec.europa.eu/food/fs/sc/scf/out22_en.html

Smith, J.L., 1999. Foodborne infections during pregnancy. Journal of Food Protein 62, 818–829.

Thompson, R.L., Miles, L.M., Lunn, J., et al., 2010. Peanut sensitisation and allergy: Influence of early life exposure to peanuts. British Journal of Nutrition 103 (9), 1278–1286. Epub 2010 Jan 26.

Villena, I., Ancelle, T., Delmas, C., et al., 2010. Congenital toxoplasmosis in France in 2007: First results from a national surveillance system. Euro Surveillance 15 (25), 1–6. Available at http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19600.

World Health Organisation, 2011. Guideline: Vitamin A Supplementation in Pregnant Women. World Health Organisation, Geneva. Available at http://whqlibdoc.who.int/publications/2011/9789241501781_eng.pdf.

Pregnancy: Weight gain

Melissa F. Young and Usha Ramakrishnan, Hubert Department of Global Health, Emory University, Atlanta, GA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of L.H. Allen, Pregnancy: Weight Gain, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 99–103, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00232-4.

Introduction 222
Pregnancy weight gain recommendations 223
Global weight gain standards 223
Excessive and inadequate weight gain 224
Pattern of weight gain 225
Changes in body composition and maternal energy status 225
Weight gain for special population groups 226
Adolescents, short women, and ethnic groups 226
Substance abusers 226
Multiple births 226
Obese and overweight women 226
Exercising women 226
Pregnancy weight gain and postpartum risk of obesity 227
Impact of supplementation 227
Summary 228
References 228
Further reading 228
Relevant websites 229

Key points

- To review history and current pregnancy weight gain recommendations.
- To discuss consequences of excessive and inadequate weight gain for maternal and child health.
- To identify research gaps and opportunities.

Glossary

Body Mass Index Weight (in kg) divided by height (in meters squared). Used as an indicator of undernutrition or overweight
Gestational diabetes mellitus Abnormal *glucose metabolism* accompanied by major alterations in the metabolism of fat and protein. Diagnosed by high *blood glucose* and/or *urinary glucose*. This is the most common medical disorder in pregnancy
High birth weight (macrosomia) >4500 g
Low birth weight (LBW) <2500 g
Preeclampsia A condition which can appear suddenly in *late pregnancy*, with symptoms of high blood pressure, edema, urinary protein, severe headache, and vision problems. This may turn into *eclampsia* (convulsions or coma) if not treated
Preterm Birth (PTB) <37 weeks from conception. A normal duration of gestation is 40 weeks
Small for Gestational Age (SGA) <10th percentile of birthweight for sex and gestational age

Introduction

Maternal nutrition and adequate weight gain during pregnancy are essential for healthy birth outcomes. This articles reviews historical and current pregnancy weight gain recommendations for optimal maternal and child health. There have been several advances in the last 50 years both domestically and globally that have enhanced our understanding of the impact of inadequate and excessive gestational weight gain. However, the majority of women fail to meet current recommendations. Further research is needed to help support women to enter pregnancy with a healthy BMI as well to gain within recommended amounts.

Pregnancy weight gain recommendations

In 1970, the US National Academy of Sciences published guidelines for weight gain during pregnancy in the report "Maternal Nutrition and the Course of Pregnancy." The recommended pregnancy gain was 24 lb (10.9 kg), with a range of 10–25 lb (9.1–11.4 kg). The report advised health care providers and pregnant women not to restrict weight gain—a practice that had been fairly widespread during the previous decade in order to reduce the perceived increased risk of *labor complications*, *preeclampsia*, and excess weight retention postpartum. In fact, many *obstetricians* had been recommending gains of only 15–20 lb (6.8–9.1 kg).

Even with the more generous recommendations set in 1970, by the 1980s it had become clear that average gain of women in the United States far exceeded these guidelines. An analysis of data from the National Natality Survey in 1980 showed the average *pregnancy weight gain* to be 29 lb (13.2 kg), and by the time of the National Maternal Infant Health Survey in 1988 the average had increased to 32 lb (14.5 kg). The range of gain among women was very wide, from no gain to more than 75 lb (34.1 kg).

Based on this realization, in 1990, the weight gain recommendations were revised completely by a committee established by the Institute of Medicine (IOM) of the National Academy of Sciences. Existing data from a national survey were analyzed to determine the weight gain that was compatible with a normal pregnancy outcome. The latter was defined as the infant being born full term and of normal birth weight, and the absence of *pregnancy or delivery complications*. It became apparent from these analyses that maternal weight-for-height at conception, expressed as *body mass index* (BMI) (weight in kilograms divided by height in meters squared), was an important predictor of actual weight gain. Women with a low BMI gained more weight than women with high BMIs. Different weight gain recommendations were therefore developed for women entering pregnancy with different BMIs.

The 1990 recommendations were revised again in 2009, the justification being that women were becoming pregnant at an older age and heavier weight, were more likely to have multiple pregnancies, and to gain excessive amounts of weight during pregnancy. The increasing prevalence of maternal overweight and obesity at conception is a global phenomenon, and associated with greater risk of preeclampsia, *gestational diabetes*, *cesarean delivery*, problems with breast-feeding, and subsequent overweight in the child. As in 1990, the known relationship between pregnancy weight gain and BMI formed the basis of the recommendations, but the BMI categories were changed to those used by the World Health Organization. The recommended gains in **Table 1** are consistent with the lowest risk of: cesarean delivery, giving birth to a premature infant, excessive weight retention after delivery, low or high birth weight, and subsequent *childhood obesity* in each BMI category.

For underweight women (BMI<18.5), recommended gains are 12.5–18 kg (28–40 lb) or 0.51 kg (1 lb) per week; for women with a normal BMI (18.5–24.9), gain should be 11.5–16 kg (25–35 lb) or 0.42 kg (1 lb) per week; for overweight women (BMI>25–29.9), gain should be at least 7 kg (15 lb) and not more than 11.5 kg (25 lb) or 0.28 kg (0.6 lb) per week; and obese women (BMI≥30) should gain the least, namely 5–9 kg (11–20 lb) or 0.22 kg (0.5 lb) per week. New weight gain charts were constructed that show the recommended gains over the course of pregnancy for each BMI group (**Fig. 1**), enabling the adequacy of weight gain to be tracked for individual women. To use the chart, women's height and weight should be measured as near to the time of conception as possible (because pregnancy causes a temporary reduction in height) and used to obtain their BMI from a table. Maternal recall of prepregnancy weight and height can be used as a practical alternative but may be substantially less reliable.

Global weight gain standards

Global gestational weight gain standards were published in 2016 using prospective data from the Fetal Growth Longitudinal Study component of the INTERGROWTH-21st Project (Cheikh Ismail et al., 2016). Healthy, well-nourished and educated women were enrolled from eight geographically diverse settings (Brazil, China, India, Italy, Kenya, Oman, United Kingdom, and United States). All enrolled women had a BMI of 18.50–24.99 in the first trimester of pregnancy, height of > 153 cm and had a live singleton birth. Women in the study received quality antenatal care, had accurate measurements of gestational age, had low rates of adverse maternal, perinatal and neonatal outcomes and their children experienced normal growth and development through age 2 years (Villar et al., 2013, 2019, 2018, 2014; Papageorghiou et al., 2014). Prospectively collected data from this unique study were used to determine the 3rd, 10th, 25th, 50th, 75th, 90th, and 97th gestational weight gain centiles by exact week of gestation (**Fig. 2**).

Table 1 Recommendations for pregnancy weight gain by BMI at conception.

BMI category	*Recommended weight gain range, kg (lb)*	*Rate of weight gain in second and third trimester range, kg (lb)/week*
Low (BMI < 18.5)	12.5–18 (28–40)	0.44–0.58 (1–1.3)
Normal (BMI 18.5–24.9)	11.5–16 (25–35)	0.35–0.50 (0.8–1)
Overweight (BMI > 25.0–29.9)	7–11.5 (15–25)	0.23–0.33 (0.5–0.7)
Obese (BMI ≥ 30.0)	5–9 (11–20)	0.17–0.27 (0.4–0.6)

Modified from Institute of Medicine (2009).

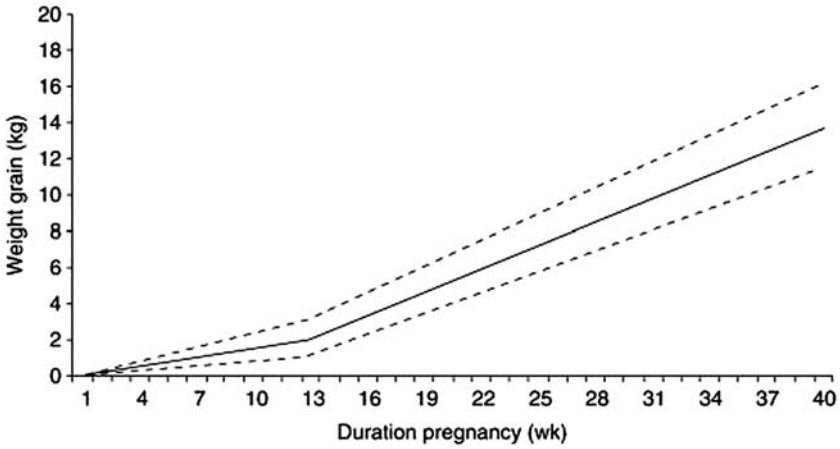


Fig. 1 Recommended weight gain by week of pregnancy for women with a normal BMI at conception. Similar charts are available for underweight and obese women. Reproduced with permission from the Institute of Medicine (2009).

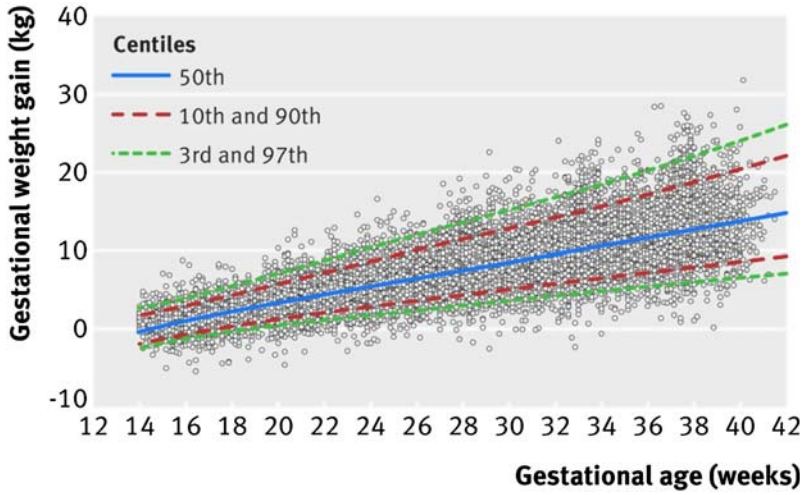


Fig. 2 Smoothed centile curves at 3rd, 10th, 50th, 90th, and 97th centiles for gestational weight gain among healthy normal weight women with uncomplicated live singleton births. Source: Cheikh Ismail et al. (2016).

Excessive and inadequate weight gain

Most of the literature on *pregnancy weight gain* and maternal outcomes is observational making it difficult to infer causality on the relationship between high or low weight gain with poor outcomes. Inadequate maternal weight gain is associated with poor *fetal growth*. Birth weight in turn is an important determinant of child health and survival . Low weight-for-length at birth may also be a risk factor for chronic disease in later life. Likewise, excessive maternal weight gain may have adverse consequences for both the mother and child.

In a 2017 systematic review, including over a million women from the United States, Asia and Europe, 23% of women gained below IOM recommendations and 47% gained above recommendations (Goldstein et al., 2017). Excessive gestational weight gain was associated with lower risk of SGA and preterm birth but higher risk of large for gestational age, macrosomia and cesarean delivery. The data however were insufficient for evaluating gestational diabetes. On the other hand, inadequate gestational weight gain was associated with a higher risk of small for gestational age (SGA), preterm birth (PTB) and lower risk of larger for gestational age (LGA) and macrosomia. In addition, in a 2019 individual-level analysis across 33 trials of predominately Caucasian women, over two-thirds of women gained outside IOM recommendations (Rogozińska et al., 2019). Gestational weight gain above IOM recommendations was associated with increased odds of cesarean section, LGA and reduced odds of SGA while weight gain below recommendations was associated with increased odds of PTB and low birth weight (LBW). The consequences of inadequate and excessive gestational weight gain remain consistent across diverse populations and settings (Goldstein et al., 2018).

A better understanding of demographic, behavioral, psychosocial and medical characteristics associated with inadequate or excessive gestational weight gain is important to help inform programs and policies to support optimal weight gain and health outcomes (Deputy et al., 2015). Maternal BMI at conception is inversely related to expected pregnancy weight gain, but, overweight and obese women still tend to deliver heavier infants and underweight women are more likely to have smaller infants. Among women with a low BMI, birth weight is more strongly related to pregnancy weight gain, so the greatest risk of LBW is for underweight women with a low pregnancy weight gain. It is crucial that underweight women gain adequate amounts of weight, while, excessive weight gain among obese and overweight individuals is a concern. Exercise during pregnancy may be a protective strategy for preventing excessive gestational weight gain and gestational diabetes (Barakat, 2019).

Pattern of weight gain

Relatively little (1–2.5 kg) of the total weight gain during pregnancy occurs during the *first trimester*; gain in the *second trimester* is highest, followed by a slightly lower gain in the third. Nevertheless, it is important to pay attention to the quality of pregnant women's diets during the first trimester and to ensure that they do not restrict their intake during this time, when there is the strongest risk of nutrition-related *birth defects* and spontaneous abortions.

In a 2021 analysis of the Womens First Trial that included 2331 women in the Democratic Republic of Congo, Guatemala, India and Pakistan, women's preconception BMI, early gestational weight gain (from preconception to 12 weeks) and gestational weight gain from 12 to 32 weeks were all independently associated with birth length and weight (Bauserman et al., 2021). In addition, data from the PRECONCEPT prospective study of 5011 women who were followed from preconception through 6–7 years postpartum in Vietnam, provide evidence that maternal preconception nutritional status (height, weight and BMI) was associated with an increased risk of LBW and SGA births as well as stunting at 2 years (Young et al., 2015, 2017, 2018). There was a similar and independent association with maternal nutrition during preconception and pregnancy on birth outcomes implying that programs aimed only on pregnancy may have half the potential impact in this context. In addition, ultrasound data on fetal growth were used to identify critical periods for gestational weight gain that are most influential on child growth. Early weight gain in the first 20 weeks had 3 times the influence on birthweight compared to gain $\geq$ 30 weeks. Collectively, findings from these two large studies suggest programs will have the greatest impact on improving birth outcomes if they address maternal nutrition both before and during pregnancy and underscore the importance of understanding the patterns of weight gain across pregnancy. New global standards for gestational weight gain provide opportunities for better understanding the timing of weight gain across pregnancy and optimal total weight gain worldwide (Cheikh Ismail et al., 2016).

Changes in body composition and maternal energy status

It used to be assumed that maternal *energy intake* during pregnancy was the main determinant of the amount of weight gained. Although our knowledge of this relationship is still inadequate, newer information indicates that other maternal factors, especially body composition, are important predictors.

The weight gained during pregnancy can be roughly divided into the weight of the fetus, placenta, and *amniotic fluid* (a total of approximately 5 kg), maternal gain in the uterus, breasts, blood, and fluid (approximately 4 kg), and maternal fat. The latter component is the most variable, accounting for approximately 70% of the variability in *pregnancy weight gain*. Although average fat gain in different studies is approximately 2–5 kg, and values for individual women range from a loss of several kilograms to a gain of approximately 12 kg. Even in a group of women with normal BMIs at conception, the range of fat gain was 0.5–9.5 kg. Women with greater fat mass at conception gained less fat during pregnancy, as would be expected from their lower weight gains. The greater fat gain among women low BMIs is a potential energy store for the fetus and would afford some protection against maternal malnutrition in *late pregnancy*—a situation that is not uncommon in some economically disadvantaged countries.

Maternal BMI at conception influences not only the amount of maternal weight and fat gained during pregnancy but also changes in maternal *basal metabolic rate* (BMR). In studies of well-nourished pregnant women, *BMR* has been reported to increase by approximately 20–30%. However, for undernourished women the increment in BMR may be decreased compared to those who are well nourished.

Overall, it is clear that heavier women gain less weight and fat during pregnancy and have a larger increase in BMR. However, few studies have body composition measures and ethnic differences may limit the use of BMI. It has not been determined how these differences translate into energy requirements for women in the different BMI groups used to predict weight gains. Therefore, the values for energy requirements, which vary by trimester, are used for all pregnant women regardless of their BMI at conception.

Weight gain for special population groups

Adolescents, short women, and ethnic groups

There is insufficient evidence for creating different weight gain guidelines for adolescents. To ensure adequate *nutrition* for those who are still growing special attention should be given to ensuring that the quality of their diets is good. The effects of this recommendation on weight retention postpartum have not been evaluated adequately.

Women who are less than 157 cm tall tend to give birth to infants who are large relative to maternal pelvic size, with a subsequently slightly greater risk of a more difficult delivery. However, pregnancy weight gain recommendations do not differ by maternal height.

The IOM, 2009 recommendations cite insufficient evidence to recommend different pregnancy weight gains for ethnic groups. Black women in the United States tend to gain less weight in pregnancy and to produce lower birth-weight infants. The reasons for this are not known, but it could not be explained by differences in gestational age or other factors. Adequate weight gain in this group is known to be especially important for the prevention of *fetal growth* restriction. In one study, 18% of nonobese black women who gained less than the IOM recommendations gave birth to low-birth-weight infants compared to 10% whose gain was in the ideal range and 4% who gained more than the recommendations. In obese black women, the LBW prevalence was approximately six times higher than that for those who gained less than the recommendations.

Most surveys indicate that Hispanics seem to gain approximately the same amount of weight as Anglo women. In the 1980 National Natality Survey, Hispanic and non-Hispanic white women gained a similar amount of pregnancy weight, but the risk of LBW was twice as high in Hispanics. Surveillance of a predominantly Hispanic population indicated that half of the underweight women and one-third of the normal weight women gained the recommended amount of weight, whereas more than half and three-fourths of overweight and obese women, respectively, had excessive gains. Inadequate weight gain during the *third trimester* was predictive of PTB. Underweight Hispanic women had nearly twice the risk of premature delivery.

New insights from the global gestational weight gain standards using the data from the INTERGROWTH-21st Project provide evidence of similar weight gain across eight populations in diverse settings and conclude that the observed differences by race/ethnicity reported in the literature are likely caused by socioeconomic, medical, cultural, and nutritional factors (Cheikh Ismail et al., 2016). Evidence from this study indicates that separate gestational weight gain charts are not required for different ethnic/racial groups.

Substance abusers

Cigarette smokers tend to gain less weight during pregnancy and to produce smaller infants. This effect is not explained by a lower food intake. Alcohol and drug use have similar effects. Simply gaining more weight during pregnancy will not compensate for the adverse effects of these practices on *fetal outcome* or *pregnancy complications*.

Multiple births

Relatively few data are available from national surveys on which to base weight gain recommendations for women with twins or multiple births. Recommendations by the recent IOM Committee are: normal weight at conception, 37–54 lb; overweight, 31–50 lb; and obese, 25–42 lb. There is insufficient information to make recommendations for underweight women.

Obese and overweight women

Obesity during pregnancy is associated with higher morbidity for both the mother and the child. Higher prepregnancy weights have been shown to increase the risk of late (>28 weeks of gestation) *fetal deaths*. In addition, the prevalence of *gestational hypertension* increases threefold and there is a three to four times greater risk of *gestational diabetes* in obese pregnant women. However, in prolonged fasting, i.e., 16–18 h, there is more risk of blood *ketones* being elevated in pregnant women that may increase risk of poor intellectual development of their offspring, thus prolonged fasting and *weight loss* in pregnancy is not recommended. However data needed to inform obesity prevention and weight loss programs are lacking (Lassi et al., 2020; Furber et al., 2013).

Exercising women

Women who are physically fit at conception are able to continue to *exercise during pregnancy* without harm to themselves or the fetus, as long as the activity is not too strenuous or prolonged. In several studies exercising women gained 2 or 3 kg less than those who were more sedentary.

Pregnancy weight gain and postpartum risk of obesity

On average, well-nourished women retain relatively little weight a year after delivery (approximately 0.5–1.5 kg). Delivery is followed by a rapid loss of weight in the subsequent 2 weeks due to fluid loss then a slower rate of loss for the next 6 months, so a complete return to preconception weight should not be expected in less time than this. In general, weight still retained at 1 year postpartum is unlikely to be lost without lowering intake and/or increasing physical activity. If weight retention is substantial, it can add to the risk of obesity in the longer term, and obesity is a major public health concern in many countries.

The relatively low average weight retention postpartum obscures the fact that many women do retain an excessive amount of weight. Those who retain most are likely to have gained large amounts of weight during pregnancy. At 10–18 months postpartum, weight retention was 2.5 kg for women who gained more than the IOM recommendation compared to 0.7 kg for white women, and 3.2 kg for black women who gained the advised amount. These large racial differences in weight retention have not been explained and certainly may be a risk factor for the higher prevalence of later obesity in this group.

Most women breast-feed their infants exclusively or partially for a relatively short time. There is little difference in *weight loss* between women who breast-feed and those who do not for periods up to 6 months postpartum. This is presumably due to the greater appetite and *energy intake* of women who are breast-feeding and perhaps to dieting on the part of nonbreast-feeders. One study of women who breast-fed until 12 months postpartum did report 2 kg more weight loss compared to women who stopped breast-feeding before 3 months. Even more weight was lost by those who breast-fed more often and gave longer feeds.

Women with a high BMI at conception tend to either loose or gain more weight postpartum than those with a normal BMI; approximately one-third end up weighing less than at conception, and one-third weigh substantially more. The reasons for the highly variable weight retention in this group are not known.

Although inadequate intake of nutrients during lactation can lead to maternal nutrient depletion and lower breast milk content of some nutrients and especially vitamins, breast-feeding women who choose to lose weight can do so by exercising and/or reasonable restriction of energy intake. Exercising by jogging, biking, and erobics for 45 min, four or five times per week for 12 weeks did not affect well-nourished mothers' ability to lactate or influence their milk composition. However, it is possible that severe energy deficit in lactation, especially of underweight women, will reduce breast milk volume.

Impact of supplementation

Numerous investigators have explored the benefits of energy and/or protein supplementation for *pregnancy weight gain* and other outcomes. A 5-year controlled trial in The Gambia provided daily prenatal dietary supplements (two biscuits) that contained 4250 kJ (1000 kcal) energy and 22 g protein. This supplement increased pregnancy weight gain and birth weight during the hungry and harvest seasons. There was a significant but very small increase in head circumference and a significant reduction in *perinatal mortality*.

Supplementation of undernourished women in the *third trimester* is most likely to increase birth weight. In the Dutch famine during World War II, low food intakes during the third trimester rapidly reduced birth weight and length, and head circumference. These *adverse outcomes* did not occur a few months after the food supply improved. Low intakes in the *second trimester* had less effect, although insufficiency in the *first trimester* had no effect. An increase in LBW prevalence was also observed in The Gambia when third-trimester gestation overlapped with the hungry season. Nonetheless, *nutrition* interventions initiated earlier and continued throughout pregnancy will have the strongest effect on birth weight. There are enduring advantages to continued supplementation postpartum (during lactation) and into the ensuing pregnancy. A *longitudinal study* in Guatemala reported a significant increase (approximately 350 g) in birth weight in the second pregnancy when the mother was supplemented during the previous pregnancy and throughout subsequent lactation and the second pregnancy, compared to those who were not supplemented during the prior pregnancy. Overall, it is appropriate for supplementation of undernourished women to begin as early in the pregnancy as possible so that both mother and fetus receive the maximum benefits for optimal health and development.

In the 2021 Lancet Nutrition Series, maternal balanced energy protein supplements were recommended for undernourished women during pregnancy, especially in food insecure populations. Balanced energy protein supplementation was associated with a 40% reduction in LBW, 61% reduction in stillbirths and a 29% reduction in SGA (Keats et al., 2021; Lassi et al., 2020). There was however inadequate evidence on the effects of balanced energy protein supplements on maternal weight gain, morbidity or mortality. Evidence from five food distribution programs also indicated increased birth weight and length and reduced rates of SGA, stunting, and wasting (Lassi et al., 2020).

In addition, in the past 10 years there has also been a rapid expansion of research on the efficacy and effectiveness of prenatal lipid-based nutrient supplements (LNS) to improve birth outcomes (Das et al., 2018). In a systematic review of the literature, there was a slight, positive effect of LNS birth weight, birth length, SGA and stunting compared to iron and folic acid. Evidence on gestational weight gain is limited and requires further examination.

Summary

Healthy weight gain during pregnancy can improve birth outcomes and maternal and child health. However, the majority of women fail to meet current recommendations. Women who gain inadequate weight during pregnancy are at greater risk for delivering a LBW, PTB, or SGA infant. Women who gain excessive weight during pregnancy on the other hand are at increased risk for macrosomia and giving birth to large for gestational age infant, having excessive postpartum weight retention and childhood obesity. The consequences of inadequate and excessive gestational weight gain remain consistent across diverse populations and settings. The new global weight gain standards provide new gestational age specific benchmarks for optimal weight gain. Increasing evidence suggests the importance of maternal nutrition both before and during pregnancy. Further research is needed to help support women to enter pregnancy with a healthy BMI as well to gain within recommended amounts.

See Also: Breastfeeding; Lactation: Dietary requirements; Pregnancy: Pre-eclampsia and diet; Pregnancy: Prevention of neural tube defects

References

Barakat, R., Refoyo, I., Coteron, J., Franco, E., 2019. Exercise during pregnancy has a preventative effect on excessive maternal weight gain and gestational diabetes. A randomized controlled trial. Braz J. Phys. Ther. 23 (2), 148–155.

Bauserman, M.S., Bann, C.M., Hambidge, K.M., et al., 2021. Gestational weight gain in 4 low- and middle-income countries and associations with birth outcomes: a secondary analysis of the Women First Trial. Am. J. Clin. Nutr. 2 (114), 804–812.

Cheikh Ismail, L., Bishop, D.C., Pang, R., et al., 2016. Gestational weight gain standards based on women enrolled in the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project: a prospective longitudinal cohort study. BMJ 352, i555.

Das, J.K., Hoodbhoy, Z., Salam, R.A., et al., 2018. Lipid-based nutrient supplements for maternal, birth, and infant developmental outcomes. Cochrane Database Syst. Rev. 8, CD012610.

Deputy, N.P., Sharma, A.J., Kim, S.Y., Hinkle, S.N., 2015. Prevalence and characteristics associated with gestational weight gain adequacy. Obstet. Gynecol. 125 (4), 773–781.

Furber, C.M., McGowan, L., Bower, P., et al., 2013. Antenatal interventions for reducing weight in obese women for improving pregnancy outcome. Cochrane Database Syst. Rev. 1, CD009334.

Goldstein, R.F., Abell, S.K., Ranasinha, S., et al., 2017. Association of gestational weight gain with maternal and infant outcomes: a systematic review and meta-analysis. JAMA 317, 2207–2225.

Goldstein, R.F., Abell, S.K., Ranasinha, S., Misso, M.L., Boyle, J.A., Harrison, C.L., Black, M.H., Li, N., Hu, G., Corrado, F., Hegaard, H., Kim, Y.J., Haugen, M., Song, W.O., Kim, M.H., Bogaerts, A., Devlieger, R., Chung, J.H., Teede, H.J., 2018. Gestational weight gain across continents and ethnicity: systematic review and meta-analysis of maternal and infant outcomes in more than one million women. BMC Med. 16 (1), 153. Aug 31.

Institute of Medicine, Committee to Reexamine IOM Pregnancy Weight Guidelines, 2009. Weight Gain During Pregnancy: Reexamining the Guidelines. National Academies Press, Washington, D.C.

Keats, E.C., Das, J.K., Salam, R.A., et al., 2021. Effective interventions to address maternal and child malnutrition: an update of the evidence. Lancet Child Adolesc. Health 5, 367–384.

Lassi, Z.S., Padhani, Z.A., Rabbani, A., et al., 2020. Impact of dietary interventions during pregnancy on maternal, neonatal and child outcomes in low and middle-income countries. Nutrients 12, 531.

Papageorghiou, A.T., Ohuma, E.O., Altman, D.G., et al., 2014. International standards for fetal growth based on serial ultrasound measurements: the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. Lancet 384, 869–879.

Rogozińska, E., Zamora, J., Marlin, N., Betrán, A.P., Astrup, A., Bogaerts, A., Cecatti, J.G., Dodd, J.M., Facchinetti, F., Geiker, N.R.W., Haakstad, L.A.H., Hauner, H., Jensen, D.M., Kinnunen, T.I., Mol, B.W.J., Owens, J., Phelan, S., Renault, K.M., Salvesen, K.Å., Shub, A., Surita, F.G., Stafne, S.N., Teede, H., van Poppel, M.N.M., Vinter, C.A., Khan, K.S., Thangaratinam, S., International Weight Management in Pregnancy (i-WIP) Collaborative Group, 2019. Gestational weight gain outside the Institute of Medicine recommendations and adverse pregnancy outcomes: analysis using individual participant data from randomised trials. BMC Pregnancy Childbirth 19 (1), 322. Sep 2.

Villar, J., Altman, D.G., Purwar, M., et al., 2013. The objectives, design and implementation of the INTERGROWTH-21st Project. BJOG 120, 9–26.

Villar, J., Papageorghiou, A.T., Pang, R., et al., 2014. The likeness of fetal growth and newborn size across non-isolated populations in the INTERGROWTH-21st Project: the fetal growth longitudinal study and newborn cross-sectional study. Lancet Diabetes Endocrinol. 2, 781–792.

Villar, J., Cheikh Ismail, L., Staines Urias, E., et al., 2018. The satisfactory growth and development at 2 years of age of the INTERGROWTH- 21st Fetal Growth Standards cohort support its appropriateness for constructing international standards. Am. J. Obstet. Gynecol. 218, S841–S854.e2.

Villar, J., Fernandes, M., Purwar, M., et al., 2019. Neurodevelopmental milestones and associated behaviours are similar among healthy children across diverse geographical locations. Nat. Commun. 10, 511.

Young, M.F., Nguyen, P.H., Addo, O.Y., et al., 2015. The relative influence of maternal nutritional status before and during pregnancy on birth outcomes in Vietnam. Eur. J. Obstet. Gynecol. Reprod. Biol. 194, 223–227.

Young, M.F., Nguyen, P.H., Addo, O.Y., et al., 2017. Timing of gestational weight gain on fetal growth and infant size at birth in Vietnam. PLoS One 12, e0170192.

Young, M.,F., Nguyen, P.H., Gonzalez Casanova, I.G., et al., 2018. Role of maternal preconception nutrition on offspring growth and risk of stunting across the first 1000 days in Vietnam: a prospective cohort study. PLoS One 13, e0203201.

Further reading

Asbee, S.M., Jenkins, T.R., Butler, J.R., et al., 2009. Preventing excessive weight gain during pregnancy through dietary and lifestyle counseling: a randomized controlled trial. Obstet. Gynecol. 113, 305–312.

Ceesay, S.M., Prentice, A.M., Cole, T.J., et al., 1997. Effects on birth weight and perinatal mortality of maternal dietary supplements in rural Gambia: 5 year randomised controlled trial. Br. Med. J. 315, 786–790.

Centre for Maternal and Child Enquiries, 2010. Maternal Obesity in the UK: Findings From a National Project. Google Scholar. http://www.rcog.org.uk/image/cmace-maternal-obesity-report.

Cnattingius, S., Bergstrom, R., Lipworth, L., Kramer, M.S., 1998. Prepregnancy weight and the risk of adverse pregnancy outcomes. N. Engl. J. Med. 338, 147–152.

Dewey, K.G., McCrory, M., 1994. Effects of dieting and physical activity on pregnancy and lactation. Am. J. Clin. Nutr. 59, 439–445.
Hickey, C., Cliver, S., Goldenberg, R., Kohatsu, J., Hoffman, H., 1993. Prenatal weight gain, term birth weight, and fetal growth retardation among high risk multiparous black and white women. Obstet. Gynecol. 81, 529–535.
Institute of Medicine, Committee on Nutritional Status during Pregnancy and Lactation, 1992. Nutrition During Pregnancy and Lactation. An Implementation Guide, Food and Nutrition Board. National Academy Press, Washington, D. C.
Keppel, K., Taffel, S., 1993. Pregnancy-related weight gain and retention: implications of the 1990 Institute of medicine guidelines. Am. J. Public Health 83, 1100–1103.
King, J.C., Butte, N.F., Bronstein, M.N., Kopp, L.E., Lindquist, S.A., 1994. Energy metabolism during pregnancy: influence of maternal energy status. Am. J. Clin. Nutr. 59, 439S–445S.
Kramer, M., 1993. Effects of energy and protein intakes on pregnancy outcome: an overview of the research evidence from controlled clinical trials. Am. J. Clin. Nutr. 58, 627–635.
North Carolina State Department of Health and Human Services, 2010. Prenatal Weight Gain Charts. /prenatalwt_charts.pdf.

Relevant websites

https://www.intergrowth21.org.
https://webassets.nationalacademies.org/whattogain/.
https://www.cdc.gov/reproductivehealth/maternalinfanthealth/pregnancy-weight-gain.htm.
https://www.acog.org/womens-health/faqs/exercise-during-pregnancy.

Stunting: Prevalence and prevention

Nandita Perumal[a], **Diego G. Bassani**[b,c,d], **and Daniel E. Roth**[b,c,d,e], [a] Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA, United States; [b] Center for Global Child Health, Hospital for Sick Children, Toronto, ON, Canada; [c] Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; [d] Department of Pediatrics, Faculty of Medicine, University of Toronto, Toronto, ON, Canada; and [e] Department of Nutrition, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

© 2023 Elsevier Ltd. All rights reserved.

Introduction 230
Stunting prevalence and global and regional trends 231
Mortality and developmental outcomes associated with stunting 233
Stunting risk factors 233
Preventive interventions for stunting 235
Frameworks addressing child undernutrition 235
Nutrition-specific interventions to reduce the risk of stunting 235
Maternal prenatal interventions 235
Postnatal infant and young child interventions 237
Nutrition-sensitive interventions to reduce the risk of stunting 238
Conclusion 238
References 239

Key points

The objectives of this article are to describe:

- Origins and definition of the term 'stunting', and its use as an indicator of child health and nutrition
- Global and regional burden of stunting and secular trends
- Consequences of stunting in early childhood
- Risk factors for stunting
- Nutrition-specific and nutrition-sensitive interventions to reduce the burden of stunting

Introduction

Stunting refers to a state of inadequate physical growth in stature (infant or toddler supine length, or standing height of an older child). The term was originally proposed by John Waterlow in 1972 to refer to "a reduction in final stature" that results from an abnormally slow rate of linear growth in childhood (Waterlow, 1972). From its early usage, stunting has been assumed to represent "nutritional growth failure" due to long-standing dietary deficiencies, in contrast to "acute" malnutrition, such as kwashiorkor or marasmus (Waterlow, 1973); however, it is now widely acknowledged that stunting should not be used to ascribe a specific cause to a child's deficit in length or height (Raiten and Bremer, 2020). Stunting is usually applied to young children (under the age of 5 years), but it may also be used to refer to low heights of adolescents and adults.

Currently, stunting is defined in statistical terms, whereby a child is considered stunted' if their crown-to-heel length or height z-score (HAZ) is less than −2 (i.e., more than 2 standard deviations below the median of the distribution of normal lengths or heights for children of the same sex and age) and severe stunting may be defined as a z-score less than −3 (de Onis and Branca, 2016). This definition of stunting is analogous to related forms of child undernutrition such as 'underweight' and 'wasting', which are similarly defined according to biologically arbitrary thresholds in the normative weight-for-age and weight-for-height distributions, respectively (WHO Multicentre Growth Reference Study Group, 2006). Since their publication in 2006, the World Health Organization (WHO) child growth standards (WHO-GS) have been the most widely used source of normative distributions to enable assessment of an individual child's length or height value under five years of age (WHO Multicentre Growth Reference Study Group, 2006). The WHO-GS for length and height, which are applicable to children born at term, were recently complemented by the publication of the Intergrowth-21st standards for newborn length by gestational age and sex (Villar et al., 2014a,b), and postnatal growth references for preterm infants (Villar et al., 2015). The recommendation to use a common set of international growth standards/references to assess nutritional status is based on the well-supported notion that ethnicity or genetic factors play a minimal role in explaining the variance in the growth of young children, such that the distribution of individual child lengths/heights at a given age is expected to be similar around the world if environmental conditions and feeding practices are optimized (de Onis, 2006;

Villar et al., 2014a,b). As such, the WHO-GS and the Intergrowth-21st standards/references remain widely used in research and public health applications globally.

In pediatric medical practice, the terms 'failure to thrive' and 'short stature' are commonly used to describe children who are growing too slowly (i.e., gaining too little weight) or who are shorter than age-matched peers, respectively. Conversely, the label and concept of stunting is almost exclusively applied to infants and children in community and primary care settings in low- and middle-income countries (LMICs). 'Linear growth faltering' is a closely related concept to stunting, insofar as it refers to a child's – or population of children's – suboptimal rate of growth in length or height, which if prolonged or pronounced, can result in stunting. However, there are no standard quantitative definitions of linear growth faltering. Some authors have promoted the consideration of a "stunting syndrome" as a condition of vulnerable children in LMICs who, beyond their measured impairments in physical growth, suffer from a range of related adversities including other nutritional deficits (e.g., micronutrient deficiencies), enteric dysfunction, developmental and cognitive delays, immune dysfunction, and increased risks of chronic diseases in later life (Prendergast and Humphrey, 2014). However, when defined in statistical terms, stunting is not intended as a clinical diagnosis, as there is no biological basis for the nominal cut-points used to distinguish stunted from non-stunted children (Perumal et al., 2018).

Stunting prevalence and global and regional trends

Stunting is most appropriately used as an indicator of the location and dispersion of the height-for-age distribution of a particular population relative to a normative reference or standard distribution such as the WHO-GS. The prevalence of stunting (i.e., proportion of children in a population with HAZ less than −2) can be compared across populations or tracked over time to assess differences or changes in the population burden of undernutrition. The conventional statistical definition of stunting implies that 2–3% of healthy children would be classified as stunted (since a z-score of −2 is equal to the 2.3rd percentile of a normal distribution). With this benchmark in mind, in 2018 the WHO adopted the following stunting prevalence thresholds to classify the level of severity of stunting at the country level: 'very low' (<2.5 %), 'low' (2.5–10%), 'medium' (10 to <20%), 'high' (20 to <30%) and 'very high' (≥30%) (de Onis et al., 2019). Stunting prevalence has also been used as a component of composite indicators intended to convey the overall burden of child undernutrition in LMICs, including the 'composite index of anthropometric failure' (Nandy et al., 2005) and 'child growth failure' (Kinyoki et al., 2020).

Under-5 stunting prevalence is a robust indicator of the burden of growth faltering for comparisons across populations or within populations over time (i.e., secular trends). Stunting prevalence reflects the health status of the whole population rather than identifying a sub-group of growth-restricted children within the population (Perumal et al., 2018). In most populations in which stunting prevalence is elevated (above ∼2.5%), mean HAZ and the distributions of heights are shifted down (Roth et al., 2017; Victora et al., 2021) (**Fig. 1**). Reductions in stunting prevalence in countries that have successfully improved child health and nutrition status have been accompanied by upward shifts in HAZ distributions (Bhutta et al., 2020). The downward-shifted height distributions of populations in many LMICs means that even the tallest children in such settings tend to be shorter than they would have been had they been living under conditions that were optimal for growth. In fact, only a very small proportion (∼0.1%) of children in population-representative survey samples in LMICs live under conditions that support normal growth (Karra et al., 2017). Therefore, an elevation in stunting prevalence (i.e., prevalence greater than ∼2.5%) indicates a public health problem that is considerably more pervasive than suggested by counting the number of stunted children.

Tracking stunting prevalence or mean HAZ as a function of age within a cohort or survey population may provide additional insights into age-related dynamics of linear growth faltering (Rieger and Trommlerová, 2016); however, such analyses must be approached cautiously because the same stunting prevalence at two different ages implies different degrees of severity and duration of the linear growth faltering process. The later the age at which stunting is observed, the longer the time over which the children have been exposed to suboptimal growth conditions. And, because postnatal growth velocity is normally rapid in early infancy and then declines until the pubertal growth spurt, stunting at later ages is less likely to resolve, particularly given the progressively shortening window in which it remains possible for catch-up to occur before growth stops. Therefore, the implications of stunting for longer-term growth and health are affected by the duration of time over which faltering has occurred. As such, stunting prevalence estimates should generally only be used to compare populations (within countries over time, or between countries) based on samples of children of the same age or within the same age window (e.g., under-5 years).

The global prevalence of stunting among children under 5 years of age has declined by over 45% since 1990, by almost 35% in the last two decades for which data are available (from 32% in the year 2000 to 21% in 2019), and by 18% in the most recent decade (2010–2019) (UNICEF/WHO/World Bank Group, 2020) (**Fig. 2**). The international agencies that track malnutrition indicators (UNICEF, WHO and World Bank) report stunting prevalence as well as the estimated absolute number of children under five years who are affected by stunting. For example, in 2019, the estimated 21% global prevalence corresponded to at least 144 million stunted children worldwide (UNICEF/WHO/World Bank Group, 2020). The absolute number is used to track progress with respect to the WHO nutrition target of a 40% reduction in the number of stunted children from 2010 to 2025. Estimates in 2020 indicated that the pace of reductions in stunting have been inadequate to meet the WHO targets at a global level, although many individual countries are achieving target reductions (UNICEF/WHO/World Bank Group, 2020).

There is considerable regional variability in stunting prevalence. Based on 2019 estimates, prevalence of stunting was 33% in Sub-Saharan Africa and South Asia, 14% in the Middle East and North Africa, 11% in East Asian and Pacific Region, 9% in Latin

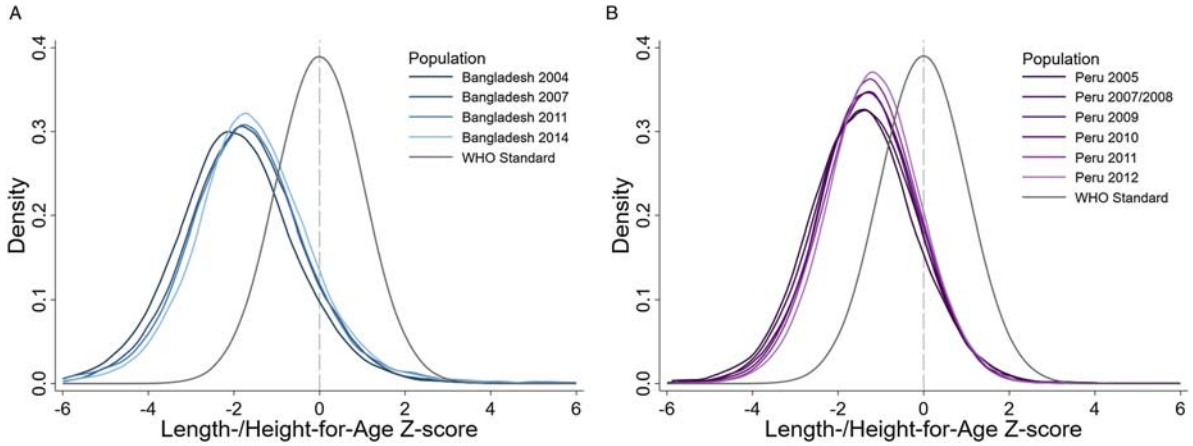


Fig. 1 Density distributions of length/height-for-age z-scores (LAZ/HAZ) for children under 5 years of age in Bangladesh (panel A) and Peru (panel B) based on anthropometric data from population-representative Demographic and Health Surveys (DHS). Distributions in both countries demonstrate that an elevated prevalence of stunting (proportion with LAZ/HAZ<-2) is a result of a whole-population leftward shift in child height compared to the World Health Organization (WHO) growth standard. Secular improvements in child nutritional status in the past two decades in both countries are marked by declines in the prevalence of stunting (Bangladesh: 51% in 2004 to 36% in 2014; Peru: 29% in 2005 to 18% in 2012). Decline in stunting prevalence involves a rightward shift in the LAZ/HAZ distribution, but this effect may not be uniformly distributed across the population. For example, in Peru (Panel B), increases in height in more recent survey years were more apparent at the lower than the upper tail of the distribution.

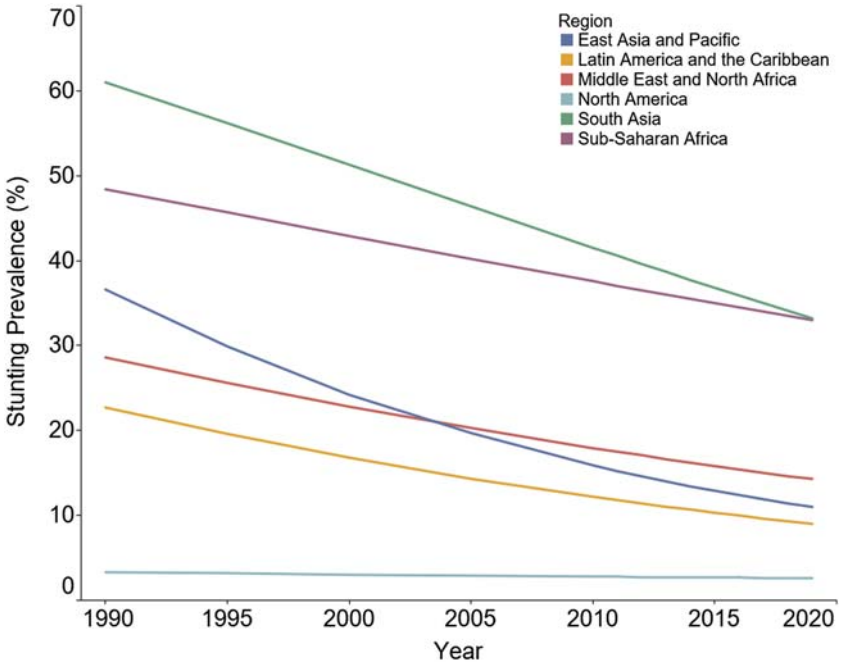


Fig. 2 Prevalence of stunting by World Bank Region (UNICEF, WHO, World Bank Group Joint Malnutrition Estimates, March 2020 Edition. Stunting prevalence 1990–2019).

America and the Caribbean and Eastern Europe and Centra Asia, and 2.6% in North America (UNICEF/WHO/World Bank Group, 2020). The rates of decline have also varied across world regions, with the largest declines observed in Asia, followed by Africa and Latin America and the Caribbean. Against the global trend, Oceania (excluding Australia and New Zealand) has not experienced a decline in the prevalence of stunting over the last 2 decades (prevalence of 37% in 2000 vs. 38% in 2019). Nationally, the highest prevalence of stunting has been documented in Burundi (54%), Madagascar (50%), Timor-Leste (50%), Niger (48%), Guatemala (47%) and Yemen (45%) (Kinyoki et al., 2020). In 2019, 101 of the 134 countries for which recent malnutrition estimates were available had an estimated stunting prevalence higher than 10%, which the WHO considers to be at least a 'medium' level of severity (UNICEF, 2019).

Within countries, important between-district differences in stunting prevalence among children under 5 years of age highlight socioeconomic inequities. For example, in India, the national prevalence of stunting among children under 5 years of age was 36%

in 2015, yet the state-level prevalence ranged widely from 18% to 48% (Victora et al., 2021). Other sub-national estimates point to a concerningly high prevalence of stunting in some communities. A prevalence as high as 60% has been observed in the Karuzi province in Burundi and in the Jigawa state in Nigeria, and over 58% in Laos' Houaphan province (Kinyoki et al., 2020). An increased prevalence of stunting has also been observed in settings of violent conflict, political instability and other settings with major disruptions in civil society. Yet, a recent analysis of data from DHS and MICS surveys from low- and middle-income countries, confirmed that while the distributions of height-for-age z-scores are still below the reference values in the WHO growth standards in both groups of countries, both groups are also slowly progressing towards the standard WHO distribution (Victora et al., 2021).

Mortality and developmental outcomes associated with stunting

Beyond its role as an indicator of suboptimal nutritional status, stunting has also been interpreted more broadly as an indicator of elevated risks of adverse child health outcomes. A key reason that the global public health community has adopted stunting and wasting as core global child health indicators is their associations with short-term mortality. Stunted (HAZ between -3 and <-2) and severely stunted children (HAZ <-3) have 2.3-folder and 5.5-fold higher mortality rates, respectively, compared to a reference group of children with HAZ$>=-1$, although these associations for stunting are weaker than for wasting (Olofin et al., 2013). An important observation is that even children with HAZ from -2 to <-1 have a significantly elevated mortality rate, and there is no inflection point in the mortality-HAZ relationship at the stunting threshold of HAZ $= -2$ (Olofin et al., 2013); therefore, while height reflects a child's relative risk of death in LMIC settings (primarily due to common infectious diseases such as pneumonia and diarrhea), there is no particular significance of the stunting classification at the individual level. Furthermore, there is no convincing evidence that short stature is itself a cause of death; rather, it is more likely that the underlying factors that constrain linear growth also exacerbate exposures to infectious pathogens or impair immune responses.

Stunting reflects a chronic state of deprivation with respect to the conditions and resources necessary for optimal child growth and development; as such, its associations with long-term developmental and human capital outcomes have been examined in numerous studies. An analysis of 5 low- and middle-income country cohort studies from Brazil, Guatemala, the Philippines, India and South Africa showed that for each unit reduction in z-scores by 2 years of age, there was, on average, a half-year reduction in years of schooling (a marker of educational attainment) (Victora et al., 2008). An analysis of data from Demographic and Health Surveys from 21 countries observed an average reduction of ~0.3 school years per HAZ z-score reduction, but with marked variability in the effects, ranging from reductions as small as 0.1 years to as large as 1.2 years (Karra and Fink, 2019). Nevertheless, some studies have shown that although lower HAZ z-scores increased the probability of a child being in earlier than expected grades in school, this may be a consequence of delays in school entry (Crookston et al., 2011), lower probability of high school completion by adult age, and higher probability of failing and/or repeating grades (Daniels and Adair, 2004). Low HAZ and stunting have also been associated with lower indices of early childhood cognitive and motor development (Sudfeld et al., 2015a,b). However, as with associations between height and mortality (as described above), there is no clear inflection point at the stunting threshold of HAZ $= -2$ in the relationships between HAZ and developmental measures (Sudfeld et al., 2015a,b).

The relevance of associations between stunting and measures of cognitive outcomes in observational studies has been questioned. Cognitive development and linear growth have common determinants which may largely explain the observed associations (Leroy and Frongillo, 2019). Developmental indicators may improve even without any changes in anthropometric status, while interventions that mitigate linear growth faltering do not necessarily improve cognitive outcomes; moreover, the effect of interventions focused on nutrition seem to have a weaker effect on cognition compared to interventions that directly support infant and early childhood nurturing and stimulation (Prado et al., 2019). Stunting is generally not considered to be a direct cause of low schooling attendance and attainment, even though the associations continue to be observed even after adjustment for other socio-economic factors associated with stunting, such as parental education. Furthermore, parental behaviors and decisions about heath care and feeding may be responsive to the length/height of their children, which may partly compensate for the adverse effects of undernutrition (Liu et al., 2009). Therefore, an elevated prevalence of stunting in a community is most appropriately viewed a marker of suboptimal environmental conditions rather than a direct cause of adverse child health outcomes.

Stunting risk factors

Stunting in LMICs is a consequence of adverse social, environmental and economic conditions that cause fetal and child growth to proceed more slowly than expected under ideal conditions (Victora et al., 2021) (**Fig. 3**). Conceptual frameworks to explain stunting and other forms of child undernutrition have emphasized a range of proximal household, maternal and infant care-related factors that affect child health, and which are assumed to primarily act as mediators of upstream societal and political influences (Black et al., 2013; Mosites et al., 2017; Bhutta et al., 2020). Nutritional deficiencies due to inadequate dietary intake or impaired nutrient absorption or utilization due to chronic inflammation or other stressors are believed to be the primary proximate factors that impair skeletal growth and development, thereby leading to stunting.

Innumerable studies of stunting risk factors in LMICs have sought to identify sources of between-child variation in height (or HAZ) within a particular population (e.g., cohort, community, or country). Yet, identifying specific remediable causes of linear growth faltering has proved challenging. Among the risk factors with the strongest and most consistent associations with stunting

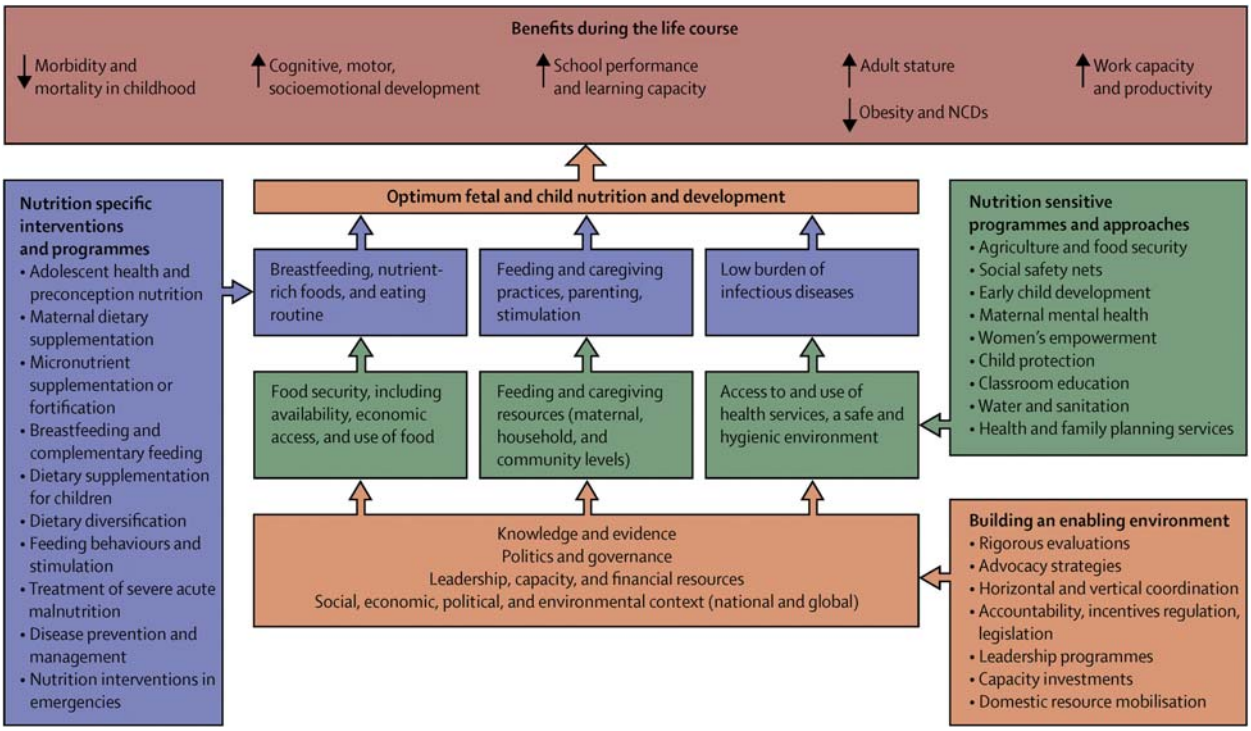


Fig. 3 Framework for actions to achieve optimum fetal and child nutrition. Source: Black R.E, Victora C.G, Walker S.P, Bhutta Z.A, Christian P, de Onis M, et al., Maternal and child undernutrition and overweight in low-income and middle-income countries, Lancet. 382 (9890), 2013, 427–451.

are maternal short stature (Kim and Subramanian, 2017; Mal-ed Network Investigators, 2017; Richter et al., 2018; Li et al., 2020) and adverse perinatal outcomes (i.e., preterm birth and fetal growth restriction measured as low birth weight or small-for-gestational age) (Christian et al., 2013; Danaei et al., 2016). The observation of neonatal stunting in many LMICs underscores the early timing of onset of linear growth faltering and has brought increased attention to the importance of maternal and fetal exposures as antecedents of infant and child stunting. Additional maternal characteristics that increase the risk of stunting include younger age, lack of education (Li et al., 2020), and short spacing between consecutive pregnancies (Fink et al., 2014). Observational epidemiologic studies conducted in LMICs suggest a positive association between longer birth interval (>36 months) and lower risk of stunting (~10%–50% relative reductions in prevalence), but this association was not consistently observed across all countries (Dewey and Cohen, 2007).

Low household socioeconomic status and low levels of parental education are widely recognized as important contextual factors associated with the risk of childhood stunting. However, even the wealthier segments of the populations in LMICs have negatively-shifted height distributions (Rieger and Trommlerová, 2016; Victora et al., 2021) and therefore an elevated prevalence of stunting (Black et al., 2013). Household characteristics that reflect deficiencies in sanitation (e.g., open defecation, lack of improved toilet or latrine facilities) and hygiene (e.g., lack of availability of soap and clean water next to a latrine) have also been associated with the risk of stunting (Danaei et al., 2016; Vilcins et al., 2018), but the association of stunting with unsafe drinking water has been less consistent (Owais et al., 2016; Kim and Subramanian, 2017; Li et al., 2020).

Stunting has been associated with several caregiver-reported factors that describe infant and early child dietary and feeding-related behaviors, such as low dietary diversity after the age of introduction of complementary foods (Kim and Subramanian, 2017; Mosites et al., 2017) or low proportion of total energy from protein (Mal-ed Network Investigators, 2017). However, breastfeeding practices in early infancy do not appear to influence the longer-term prevalence of stunting (Owais et al., 2016); in fact, stunting prevalence typically increases during the early postnatal period (Victora et al., 2021) when exclusive or predominant breastfeeding is relatively common in many LMICs. Early childhood zinc deficiency has long been linked to poor linear growth, but its contribution to the overall global stunting burden may be minor (Mosites et al., 2017). Enteropathogens (Mal-ed Network Investigators, 2017), environmental enteric dysfunction (i.e., intestinal wall injury, inflammation, and permeability), and intestinal dysbiosis have been of particular recent interest, but the extent to which they cause an elevated prevalence of stunting remains unclear (Harper et al., 2018; Budge et al., 2019). Mycotoxin exposures are associated with low HAZ, but there is less conclusive evidence to support associations of stunting with other environmental toxins or pollutants (e.g., pesticides, heavy metals, indoor air pollution) (Vilcins et al., 2018).

Social and biological factors that explain between-child variance in individual risks of stunting are similar or analogous to those that account for between-population variation in the stunting burden. For example, district-level stunting prevalence in India is associated with maternal health and education, average household wealth, and other measurements of socioeconomic

status (Karra et al., 2017). In addition, reductions in the prevalence of stunting in LMICs over time have been associated with policy and programmatic changes in the health and non-health sectors that improve women's access to health care, nutrition and education (Bhutta et al., 2020), household access to improved sanitation and water, and improvements in other related domains of early child health and nutrition (Argaw et al., 2019). However, recognition of stunting as an indicator of whole-population growth faltering highlights the critical importance of upstream community-level and ubiquitous factors that cause nearly all children in some communities to grow too slowly and thereby increase the prevalence of stunting in a population. Furthermore, the extent to which conventional stunting risk factors operate as causal determinants of growth faltering and stunting in LMICs is unclear because of the strong likelihood of confounding in observational studies in which nearly all individual- and household-level risk factors under consideration are strongly and broadly correlated with social disadvantage and economic deprivation.

Preventive interventions for stunting

Over the past 20 years, the global development agenda has included major efforts to reduce undernutrition among young children in LMICs. In 2000, the United Nations General Assembly ratified eight Millennium Development Goals, which included halving the prevalence of underweight (weight-for-age z-scores <-2 SD based on the World Health Organization Child Growth Standards) as the global target to reduce hunger (United Nations General Assembly, 2000). The 2015 Sustainable Development Goals further set out an ambitious target to "end all forms of malnutrition, including achieving, by 2025, the international agreed upon targets on stunting and wasting among children under 5 years of age" (United Nations, 2015). In 2012 the World Health Assembly endorsed a comprehensive implementation plan on maternal, infant, and young child nutrition, which included the six Global Nutrition Targets, including the target to "achieve a 40% reduction in the number of children under-5 who are stunted by 2025" (World Health Organization, 2018). As such, substantial investments have been made in global nutrition research to design and evaluate interventions and programs to reduce the prevalence of stunting.

Frameworks addressing child undernutrition

In 1990, the United National Children's Fund (UNICEF) developed a seminal comprehensive conceptual framework on the causes of child undernutrition, highlighting the multifactorial nature of undernutrition (UNICEF, 1990). Subsequently, the Lancet 2013 Nutrition Series advanced this framework to identify pathways to optimal fetal and child growth and development (Black et al., 2013a,b) (**Fig. 3**). This renewed framework identified 'nutrition-specific' interventions and programs designed to address immediate causes of suboptimal growth in early life, such as poor breastfeeding practices, inadequate quality and quantity of foods, caregiving practices, and infectious disease burden, as well as 'nutrition-sensitive' interventions and programs that aimed to address the underlying determinants of undernutrition, including food security, feeding and caregiving resources, access and use of health services and a safe, hygienic environment (Black et al., 2013a,b) (**Fig. 3**). Whereas nutrition-specific interventions are expected to have a direct effect on nutritional outcomes, nutrition-sensitive interventions are expected to have an indirect effect through mediating pathways (Keats et al., 2021). Interventions to reduce stunting among young children are therefore often categorized as being nutrition-specific, nutrition-sensitive, or integrated solutions that include both nutrition-specific and nutrition-sensitive components.

Nutrition-specific interventions to reduce the risk of stunting

Maternal prenatal interventions

The period from conception to the first 2 years of life – the 'first 1000 days' – is widely recognized as a critical window in the life cycle in which linear growth faltering and stunting may be most amenable to prevention (Victora et al., 2010). Several interventions to prevent stunting have focused on meeting the higher nutritional needs of pregnant and lactating women to promote optimal fetal growth (Dewey, 2016), primarily through micronutrient and food supplementation (Bhutta et al., 2013; Keats et al., 2021). Daily iron-folic acid supplements (IFA) containing 30–60 mg of iron and 400 μg of folic acid, and multiple micronutrient supplements (MMS), which typically provide 15 essential vitamins and minerals, have been among the most extensively evaluated interventions for improving maternal prenatal nutrition (Keats et al., 2021). While these interventions improve maternal iron status and reduce the risk of low birthweight and small-for-gestational age, there has been no reported effect on stunting prevalence (Oh et al., 2020). Supplementing breastfeeding mothers with daily MMS during the first 6 months post-pregnancy also does not affect stunting in early infancy (Park et al., 2020a). Other single micronutrient supplementation studies, such as calcium supplements, iron supplements alone, or folic acid supplements alone, may improve some maternal and perinatal outcomes but have not be shown to affect newborn length or stunting prevalence (Keats et al., 2021). However, compared to IFA alone, provision of daily small-quantity lipid nutrient supplements (SQ-LNS), which consist of multiple micronutrients in a 20 g lipid-based vehicle providing 118 kcal/day, has been shown to improve mean length at birth and reduce the risk of stunting at birth in a meta-analysis of two studies (Das et al., 2018) (**Table 1**). Similarly, in a multi-country randomized trial of SQ-LNS per day provided to women starting at least 3 months or more preconceptionally and throughout pregnancy (two-doses for a subset of women who had a body mass index of <20 kg/m^2), compared to no preconception or antenatal supplementation, reduced the prevalence of stunting at birth (Hambidge et al., 2019). Balanced energy protein supplements (BEP), which are food supplements with protein content less than 25% of the total caloric

Table 1 Effect of prenatal and postnatal nutrition-specific interventions on stunting among children <5 years of age.

Interventions	*Study type*	*Publication year*	*Sample size*	*Effect estimates for stunting*	*General conclusions and caveats*
Prenatal maternal interventions					
Daily small-quantity lipid nutrient supplements (SQ-LNSs) providing 118 kcal/day	Systematic review and meta-analysis (Das et al., 2018)	2018	2 trials, 4166 participants	RR 0.82 (95% CI: 0.71, 0.94)	Daily SQ-LNS during pregnancy reduced the risk of stunting at birth in geographically diverse settings, compared to standard iron-folic acid supplementation.
	Multi-country RCT (Guatemala, India, Pakistan, Democratic Republic of Congo) (Hambidge et al., 2019)	2019	1465 participants	Preconception to delivery: RR 0.69 (95% CI: 0.49, 0.98) Mid-pregnancy to delivery: RR 0.78 (95% CI: 0.57, 1.07)	Effect estimates for newborn stunting, for SQ-LNS compared to "no nutrition supplements". Women who were underweight (body mass index <20 kg/m^2) were provided an additional daily dose of SQ-LNS.
Postnatal infant and child interventions					
Breastfeeding promotion	Systematic review and meta-analysis (Lassi and Rind, 2020)	2020	6 trials, 6518 participants	RR 1.00 (95% CI: 0.88, 1.14)	Intervention type and duration of breastfeeding promotion are highly variable.
Zinc supplementation	Systematic review and meta-analysis (Liu et al., 2018)	2018	9 trials, 9975 participants	RR 1.01 (95% CI: 0.96, 1.06)	Some zinc supplementation trials have reported positive effects on length but not on stunting, with some suggestions that benefits of the intervention may be greater for children >2 years.
Lipid nutrient supplements (LNS)					
Small quantity (SQ-LNS) providing ~120 kcal/day starting at 6 months of age	Individual participant meta-analysis (Dewey et al., 2021)	2021	17 trials, 36,795 participants	RR 0.88 (95% CI: 0.85, 0.91)	Robust evidence from individual meta-analysis of trials indicate a reduction in the relative risk of stunting with daily SQ-LNS supplements among children 6 to 18 or 24 months of age. Evidence from network meta-analysis suggests LNS with higher energy density (~220 kcal–285 kcal/day) may result in larger effects on the relative risk of stunting (RR 0.80 95% CI: 0.66, 0.97) (Park et al., 2020b)
Complementary foods starting at 6 months of age					
Animal source foods	Systematic review (Shapiro et al., 2019)	2019	19 studies	Pooled effect not estimated given heterogeneity of definitions and interventions regarding animal source foods	Trials evaluating animal source foods on stunting mostly showed null effects, with one or two small studies indicating a positive effect.
Eggs	Single RCT providing one egg/day for 6 months among children 6–9 months of age in Ecuador (Iannotti et al., 2017)	2017	163 participants	RR: 0.53 (95%CI: 0.37, 0.77)	Effect of daily egg consumption for 6 months on stunting observed in the Ecuador trial was not sustained two years after the intervention.

Table 1 Effect of prenatal and postnatal nutrition-specific interventions on stunting among children <5 years of age.—cont'd

Interventions	*Study type*	*Publication year*	*Sample size*	*Effect estimates for stunting*	*General conclusions and caveats*
	Single RCT providing one egg/day for 6 months to infants 6–9 months of age in Malawi (Stewart et al., 2019)	2019	595 participants	RR: 0.98 (95%CI: 0.80, 1.19)	Similar trial in Malawi found no effect of an egg/day intervention on linear growth, though this was hypothesized to be due to high baseline consumption of animal source foods in the study population.
Complementary food provision, with or without education in food insecure environments	Systematic review and meta-analysis (Lassi and Rind, 2020)	2020	7 trials, 7894 participants	RR 0.64, (95% CI: 0.44, 0.92)	Pooled effect size based on seven trials of heterogenous complementary feeding interventions, including the daily egg intervention in Ecuador. The quality of evidence however was rated as low.

RCT, randomized controlled trial; CI, confidence intervals.

content, reduced the risk of adverse birth outcomes including stillbirths among undernourished pregnant women, but effects on postnatal stunting prevalence have not been examined (Ota et al., 2015).

Postnatal infant and young child interventions

Interventions designed to meet the nutritional requirements of exclusively breastfed infants (birth to 6 months of age) and infants and children during the complementary feeding period from 6 to 24 months of age have also been evaluated in terms of their effects on stunting (Dewey, 2016; Park et al, 2020a, 2020b). The World Health Organization recommends exclusive breastfeeding for the first 6 months (World Health Organization, 2002). Although maternal breastfeeding education and promotion interventions improve exclusive breastfeeding rates and reduce the risk of diarrheal diseases (Lassi and Rind, 2020), these interventions do not affect linear growth or stunting prevalence (Lassi and Rind, 2020; Park et al., 2020b). Providing MMS to children age 6–24 months during the complementary feeding period is associated with a reduced risk of anemia and marginal improvements in linear growth, but not stunting (Taylor-Robinson et al., 2015). However, evidence from a network meta-analysis of several nutrition-specific interventions provides some suggestion that MMS to children in the complementary feeding period may reduce the risk of stunting (Park et al., 2020b).

Daily SQ-LNS (providing approximately 120 kcal/day with recommended dietary allowance for multiple micronutrient supplements) to children also reduced the prevalence of stunting in an individual participant data analysis of 17 trials, and this association was not altered by differences in geographic regions, baseline prevalence of stunting or malaria, water quality, sanitation, or duration, frequency, and compliance of supplementation (Dewey et al., 2021). Although single micronutrient supplements, such as iron and zinc supplements, are effective in reducing the prevalence of anemia and diarrheal diseases, respectively, these interventions do not improve linear growth or reduce stunting (Liu et al., 2018; Keats et al., 2021). Interventions designed to improve dietary diversity and quality of complementary foods through the provision of animal sources foods, including dairy, flesh foods (e.g., meat, fish, poultry), eggs, seafood and insects, have had little to no effect on stunting (Shapiro et al., 2019). Although one small randomized controlled trial of 163 infants in Ecuador showed a lower risk of stunting among children 6–9 months of age who received one egg per day for 6 months, compared to children who received no supplementation (Iannotti et al., 2017), the effect was not sustained two years after the intervention (Iannotti et al., 2020). In a similar trial in Malawi, providing one egg per day to infants 6–9 months for a duration of 6 months did not affect linear growth or stunting prevalence (Stewart et al., 2019). While complementary food education alone has not been found to improve linear growth in food-secure or -insecure settings, provision of complementary foods, with or without education, is associated with a lower risk of stunting in food insecure settings (Lassi and Rind, 2020). Among the few caregiving interventions evaluated in relation to child growth, kangaroo mother care (i.e., skin-to-skin contact) compared to control (i.e., placement under a warmer or incubator) for a period of 1–6 weeks after birth has been shown to yield higher linear growth velocity in high-risk newborns but there was no observed effect on stunting (Park et al., 2020b).

Nutrition-sensitive interventions to reduce the risk of stunting

Nutrition-sensitive interventions aim to enhance maternal resources, including health and education, household wealth, access to and utilization of health services, or improve water, sanitation and hygiene (WASH) practices. Compared to nutrition-specific interventions, there is limited evidence related to the effects of nutrition-sensitive interventions on linear growth faltering and stunting (Keats et al., 2021). Interventions to promote sexual and reproductive health rights of women through education, delaying age at first pregnancy, and increasing school enrollment and retention among adolescents have been evaluated in relation to preventing unintended pregnancies and optimizing inter-pregnancy interval, but did not evaluate postnatal child outcomes (Lassi and Kedzior, 2020). Disease management through deworming with anthelmintics for soil-transmitted helminths during pregnancy to improve maternal health or among school-aged children have shown no effects on linear growth or stunting prevalence (Salam et al., 2015; Taylor-Robinson et al., 2015; Welch et al., 2017). However, in a meta-analysis, prenatal SQ-LNS food supplementation programs which additionally provided access to maternity and obstetric care or provided malaria treatment were found to reduce the risk of newborn stunting (RR 0.82, 95%CI: 0.71, 0.94) (Lassi and Padhani, 2020). Cash transfer programs, which provide supplemental income to low-income households either unconditionally (i.e. not tied to any obligations) or conditionally (i.e., requiring school enrollment, health check-ups, or attendance at health education sessions, among other behaviors) have been shown to improve a wide range of underlying determinants of undernutrition, including food security, dietary diversity, health care utilization, and maternal mental health (Leroy et al., 2009; de Groot et al., 2017). However, cash transfer programs vary substantially in the amount of basic cash transfer and the types of conditions that are associated with them, and therefore evidence for improving linear growth and reducing stunting is mixed (Leroy et al., 2009; de Groot et al., 2017). Nonetheless, a recent meta-analysis of studies on cash transfer programs restricted to high-quality studies, which included clear control groups for comparison, found a small but significant decrease in the absolute prevalence of stunting among children <5 years of age (−2.1%, 95%CI: −3.52, −0.69) (Manley et al., 2020).

Improvements in WASH practices have been hypothesized to promote growth and reduce stunting prevalence by reducing the incidence of diarrheal diseases and enteric dysfunction in young children in LMICs; however, in three large, high-quality cluster randomized trials conducted in Bangladesh, Kenya, and Zimbabwe, there were no effects of WASH interventions on child linear growth or risk of stunting (Luby et al., 2018; Null et al., 2018; Humphrey et al., 2019). These trials used various methods to improve drinking water (e.g., water chlorination), sanitation (e.g., improved pit latrines), and hygiene practices (e.g., handwashing with soap or handwashing stations) coupled with theory-based behavior change communication models to ensure intervention uptake (Pickering et al., 2019). All trials further provided nutrition counseling regarding infant and young child feeding practices and SQ-LNS per day from 6 to 18 or 24 months alone and in combination with WASH interventions to test the synergistic effect of improved nutrition and WASH practices (Pickering et al., 2019). Compared to the control arm, children who were in the nutrition counseling and SQ-LNS arm alone had higher mean length-for-age z-scores (0.13–0.25 SD across studies) and lower risks of stunting (Pickering et al., 2019); however, across all studies, WASH interventions had no effect on linear growth or the prevalence of stunting. Despite high intervention fidelity in implementation and high-compliance due to robust behavior change communication, the results of these trials suggest that modest improvements in WASH practices at the household-level, particularly those which are not tailored to the local etiologies of diarrheal diseases and enteropathogens, are unlikely to have a substantial impact on child stunting (Cumming et al., 2019); rather, community-level structural changes to improve environmental conditions, such as piped drinking water, sewage drains, and human waste treatment facilities are likely needed to positively shift the population distribution of HAZ and achieve substantial reductions in the prevalence of stunting.

Conclusion

Childhood stunting reflects inadequate growth in stature (body length, height). In LMICs, stunting is usually a result of environmental and social conditions that are inadequate to support optimal fetal and child growth and development. The prevalence of stunting in a community is an important indicator of child health and nutrition that has been used to demonstrate and quantify inequities in childhood nutrition status between and within countries, globally. In many LMICs, stunting is evident at birth, underscoring the important contribution of maternal health and nutrition as well as social and environmental exposures in the prenatal period to the later risks of infant and child stunting. Despite reductions in under-5 stunting prevalence worldwide over the past several decades, stunting remains common in many LMICs, underscoring the persistence of adverse community-level and household conditions that impair child growth and development and increase the risk of child mortality. Identifying specific causes of stunting has been a long-standing challenge, and few specific interventions have been convincingly shown to improve child growth and reduce the risk of stunting in LMICs. As a result, over the last two decades, intervention research to support improvements in maternal and child nutrition has shifted from a focus on single-nutrient interventions to multi-component policies and programs that broadly span the domains of nutrition, health, economics, and the environment. For example, in five exemplar case-country studies (Peru, Kyrgyz Republic, Nepal, Ethiopia, and Senegal) that achieved larger than expected reductions in stunting prevalence in a 15-year period relative to their economic growth, relatively large investments in both nutrition-sensitive and -specific interventions and programs over a sustained duration of time were associated with substantial population-level reductions in stunting (Bhutta et al., 2020). Continued investments in broad-based multi-sectoral approaches to improving child and family health and nutrition will be needed to further reduce stunting prevalence in LMICs worldwide (Heidkamp et al., 2021).

References

Argaw, A., et al., 2019. Drivers of under-five stunting trend in 14 low- and middle-income countries since the turn of the demographic and health surveys. Nutrients 1–12.

Bhutta, Z.A., et al., 2013. Evidence-based interventions for improvement of maternal and child nutrition: what can be done and at what cost? Lancet 382 (9890), 452–477.

Bhutta, Z.A., et al., 2020. How countries can reduce child stunting at scale: lessons from exemplar countries. Am. J. Clin. Nutr. 112, 894S–904S.

Black, R.E., et al., 2013a. Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet 382 (9890), 427–451.

Black, R.E., Victora, C.G., Walker, S.P., 2013b. Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet 382, 427–451.

Budge, S., et al., 2019. Environmental enteric dysfunction and child stunting. Nutr. Rev. 77 (4), 240–253.

Christian, P., et al., 2013. Risk of childhood undernutrition related to small-for-gestational age and preterm birth in low- and middle-income countries. Int. J. Epidemiol. 1–16.

Crookston, B.T., et al., 2011. Impact of early and concurrent stunting on cognition. Matern. Child Nutr. 7 (4), 397–409.

Cumming, O., et al., 2019. The implications of three major new trials for the effect of water, sanitation and hygiene on childhood diarrhea and stunting: a consensus statement. BMC Med. 17, 173.

Danaei, G., et al., 2016. Risk factors for childhood stunting in 137 developing countries: a comparative risk assessment analysis at global, regional, and country levels. PLoS Med. 13 (11), e1002164. Edited by J. K. Tumwine. Public Library of Science.

Daniels, M.C., Adair, L.S., 2004. Growth in young Filipino children predicts schooling trajectories through high school. J. Nutr. 134 (6), 1439–1446.

Das, J.K., et al., 2018. Lipid-based nutrient supplements for maternal, birth, and infant developmental outcomes. Cochrane Database Syst. Rev. 8, 2018.

de Groot, R., et al., 2017. Cash transfers and child nutrition: pathways and impacts. Dev. Pol. Rev. 35 (5), 621–643.

de Onis, M., Branca, F., 2016. Childhood stunting: a global perspective. Matern. Child Nutr. 12, 12–26.

de Onis, M., et al., 2019. Prevalence thresholds for wasting, overweight and stunting in children under 5 years. Publ. Health Nutr. 22 (1), 175–179.

de Onis, M., 2006. Assessment of differences in linear growth among populations in the WHO Multicentre Growth Reference Study. Acta Paediatr. 95 (Suppl. 450), 56–65.

Dewey, K.G., Cohen, R.J., 2007. Does birth spacing affect maternal or child nutritional status? A systematic literature review. Matern. Child Nutr. 3 (3), 151–173.

Dewey, K.G., et al., 2021. Characteristics that modify the effect of small-quantity lipid-based nutrient supplementation on child growth: an individual participant data meta- analysis of randomized controlled trials. Am. J. Clin. Nutr. nqab278.

Dewey, K.G., 2016. Reducing stunting by improving maternal, infant and young child nutrition in regions such as South Asia: evidence, challenges and opportunities. Matern. Child Nutr. 12, 27–38. Department of Nutrition, University of California, Davis, Davis, CA, United States: Blackwell Publishing Ltd.

Fink, G., et al., 2014. Scaling-up access to family planning may improve linear growth and child development in low and middle income countries. PLoS One 9 (7), 1–9.

Hambidge, K.M., et al., 2019. A multicountry randomized controlled trial of comprehensive maternal nutrition supplementation initiated before conception: the Women First trial. Am. J. Nutr. 109 (2), 457–469.

Harper, K.M., et al., 2018. Environmental enteric dysfunction pathways and child stunting: a systematic review. PLoS Neglected Trop. Dis. 12 (1), 1–23.

Heidkamp, R.A., et al., 2021. Mobilising evidence, data, and resources to achieve global maternal and child undernutrition targets and the Sustainable Development Goals: an agenda for action. Lancet 6736 (21), 1–19.

Humphrey, J.H., et al., 2019. Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural Zimbabwe: a cluster-randomised trial. Lancet Glob. Health 7 (1), e132–e147.

Iannotti, L.L., et al., 2017. Eggs in early complementary feeding and child growth: a randomized controlled trial. Pediatrics 140 (1).

Iannotti, L.L., et al., 2020. Egg intervention effect on linear growth no longer present after two years. Matern. Child Nutr. 16 (2), 1–10.

Karra, M., Fink, G., 2019. Long run height and education implications of early life growth faltering: a synthetic panel analysis of 425 birth cohorts in 21 low- and middle-income countries. BMC Publ. Health 19 (1), 1–8. BMC Public Health.

Karra, M., Subramanian, S., Fink, G., 2017. Height in healthy children in low- and middle-income countries. Am. J. Clin. Nutr. 105 (1), 121–126.

Keats, E.C., et al., 2021. Effective interventions to address maternal and child malnutrition: an update of the evidence. Lancet Child Adolesc. Health 4642 (20), 1–18.

Kim, R., Subramanian, S.V., 2017. Relative importance of 13 correlates of child stunting in South Asia : insights from nationally representative data from Afghanistan. Soc. Sci. Med. 187.

Kinyoki, D.K., et al., 2020. Mapping child growth failure across low- and middle-income countries. Nature 577 (7789), 231–234.

Lassi, Z.S., Kedzior, S.G.E., et al., 2020. Effects of preconception care and periconception interventions on maternal nutritional status and birth outcomes in low-and middle-income countries: a systematic review. Nutrients 12 (3).

Lassi, Z.S., Padhani, Z.A., et al., 2020. Impact of dietary interventions during pregnancy on maternal, neonatal, and child outcomes in low-and middle-income countries. Nutrients 12 (2).

Lassi, Z.S., Rind, F., et al., 2020. Impact of infant and young child feeding (IYCF) nutrition interventions on breastfeeding practices, growth and mortality in low-and middle-income countries: systematic review. Nutrients 12 (3).

Leroy, J.L., Frongillo, E.A., 2019. Perspective: what does stunting really mean? A critical review of the evidence. Adv. Nutr. 10 (2), 196–204.

Leroy, J.L., Ruel, M., Verhofstadt, E., 2009. The impact of conditional cash transfer programmes on child nutrition: a review of evidence using a programme theory framework. J. Dev. Effect. 1 (2), 103–129.

Li, Z., et al., 2020. Factors associated with child stunting, wasting, and underweight in 35 low- and middle-income countries. JAMA Netw. Open 3 (4), e203386.

Liu, H., Mroz, T., Adair, L., 2009. Parental compensatory behaviors and early child health outcomes in Cebu, Philippines. J. Dev. Econ. 90 (2), 2009–2230.

Liu, E., et al., 2018. Effect of zinc supplementation on growth outcomes in children under 5 years of age. Nutrients 10 (3), 1–20.

Luby, S.P., et al., 2018. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: a cluster randomised controlled trial. Lancet Glob. Health 6 (3), e302–e315.

Mal-ed Network Investigators, 2017. Childhood stunting in relation to the pre- and postnatal environment during the first 2 years of life: the MAL-ED longitudinal birth cohort study. PLoS Med. 14 (10), 1–21.

Manley, J., et al., 2020. Cash transfers and child nutritional outcomes: a systematic review and meta-analysis. BMJ Glob. Health 5 (12), 1–9.

Mosites, E., et al., 2017. Piecing together the stunting puzzle: a framework for attributable factors of child stunting. Paediatr. Int. Child Health 37 (3), 158–165.

Nandy, S., et al., 2005. Poverty, child undernutrition and morbidity: new evidence from India. Bull. World Health Organ. (04), 309–315, 011650.

Null, C., et al., 2018. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Kenya: a cluster-randomised controlled trial. Lancet Glob. Health 6 (3), e316–e329.

Oh, C., Keats, E.C., Bhutta, Z., 2020. Vitamin and mineral supplementation during pregnancy on maternal, birth, child health and development outcomes in low- and middle-income countries: a systematic review and meta-analysis. Nutrients 12, 491.

Olofin, I., et al., 2013. Associations of suboptimal growth with all-cause and cause-specific mortality in children under five years: a pooled analysis of ten prospective studies. PLoS One 8 (5).

Ota, E., et al., 2015. Antenatal dietary education and supplementation to increase energy and protein intake (Review). Cochrane Database Syst. Rev. (6), CD000032.

Owais, A., et al., 2016. Minimum acceptable diet at 9 months but not exclusive breastfeeding at 3 months or timely complementary feeding initiation is predictive of infant growth in rural Bangladesh. PLoS One 11 (10), 1–10.

Park, J.J.H., et al., 2020a. Interventions to improve linear growth during complementary feeding period for children aged 6-24 months living in low- and middle-income countries: a systematic review with network and pairwise meta-analyses. Gates Open Res. 3, 1660.

Park, J.J.H., et al., 2020b. Interventions to improve linear growth during exclusive breastfeeding life-stage for children aged 0-6 months living in low- and middle-income countries: a systematic review with network and pairwise meta-analyses. Gates Open Res. 3, 1720.
Perumal, N., Bassani, D.G., Roth, D.E., 2018. Use and misuse of stunting as a measure of child health. J. Nutr. 148 (3), 311–315.
Pickering, A.J., et al., 2019. The WASH Benefits and SHINE trials: interpretation of WASH intervention effects on linear growth and diarrhoea. Lancet Glob. Health 7 (8), e1139–e1146.
Prado, E.L., et al., 2019. Does improved growth mean improved neurobehavioral development? Adv. Nutr. 10 (4), 725–726.
Prendergast, A.J., Humphrey, J.H., 2014. The stunting syndrome in developing countries. Paediatr. Int. Child Health 34 (4), 250–265.
Raiten, D.J., Bremer, A.A., 2020. Exploring the nutritional ecology of stunting: new approaches to an old problem. Nutrients 12 (2).
Richter, L.M., et al., 2018. Comparative models of biological and social pathways to predict child growth through age 2 years from birth cohorts in Brazil, India, the Philippines, and South Africa. J. Nutr. 148 (8), 1364–1371.
Rieger, M., Trommlerová, S.K., 2016. Age-specific correlates of child growth. Demography 53 (1), 241–267.
Roth, D.E., et al., 2017. Early childhood linear growth faltering in low- and middle-income countries is a whole population condition: analysis of 179 Demographic and Health Surveys from 64 countries (1993-2015). Lancet Glob. Health 5 (12), e1249–e1257.
Salam, R.A., et al., 2015. Effect of administration of antihelminthics for soil-transmitted helminths during pregnancy. Cochrane Database Syst. Rev. (6), CD005547.
Shapiro, M.J., et al., 2019. A systematic review investigating the relation between animal-source food consumption and stunting in children aged 6-60 months in low and middle-income countries. Adv. Nutr. 10 (5), 827–847.
Stewart, C.P., et al., 2019. The effect of eggs on early child growth in rural Malawi: the Mazira Project randomized controlled trial. Am. J. Clin. Nutr. 110 (4), 1026–1033.
Sudfeld, C.R., et al., 2015a. Linear growth and child development in low- and middle-income countries: a meta-analysis. Pediatrics 135 (5), e1266–e1275.
Sudfeld, et al., 2015b. Malnutrition and its determinants are associated with suboptimal cognitive, communication, and motor development in Tanzania. J. Nutr. 145 (12), 2705–2714.
Taylor-Robinson, D.C., et al., 2015. Deworming drugs for soil-transmitted intestinal worms in children: effects on nutritional indicators, haemoglobin, and school performance. Cochrane Database Syst. Rev. (7), CD000371.
UNICEF/WHO/World Bank Group, 2020. Joint Child Malnutrition Estimates Key Findings. Available at: https://data.unicef.org/resources/jme-report-2020/.
UNICEF, 1990. Strategy for Improved Nutrition for Women and Children. UNICEF, New York.
UNICEF, 2019. State of the World's Children 2019: Children, Food and Nutrition. UNICEF, New York.
United Nations General Assembly, 2000. United Nations Millennium Declaration. General Assembly Resolution 55/2. Available at: https://www.ohchr.org/EN/ProfessionalInterest/Pages/Millennium.aspx.
United Nations, 2015. Sustainable Development Goal 2: Ending Hunger, Achieve Food Security and Improved Nutrition and Promote Sustainable Agriculture. Available at: https://sustainabledevelopment.un.org/sdg2. (Accessed 20 May 2019).
Victora, C.G., et al., 2008. Maternal and child undernutrition 2 - maternal and child undernutrition: consequences for adult health and human capital. Lancet 371 (9609), 340–357.
Victora, C.G., et al., 2010. Worldwide timing of growth faltering: revisiting implications for interventions. Pediatrics 125 (3), e473–e480.
Victora, C.G., et al., 2021. Revisiting maternal and child undernutrition in low-income and middle-income countries: variable progress towards an unfinished agenda. Lancet 397, 6736(21).
Vilcins, D., Sly, P.D., Jagals, P., 2018. Environmental risk factors associated with child stunting: a systematic review of the literature. Annals of Glob. Health 84 (4), 551–562.
Villar, J., et al., 2014a. International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the Intergrowth-21st Project. Lancet 384 (9946), 857–868.
Villar, J., et al., 2014b. The likeness of fetal growth and newborn size across non-isolated populations in the Intergrowth-21st project: the fetal growth longitudinal study and newborn cross-sectional study. Lancet Diab. Endocrinol. 2 (10), 781–792.
Villar, J., et al., 2015. 'Postnatal growth standards for preterm infants: the preterm postnatal follow-up study of the Intergrowth-21(st) project. Lancet Glob. Health 3 (11), e681–e691.
Waterlow, J.C., 1972. Classification and definition of protein-calorie malnutrition. Br. Med. J. 3, 566–569.
Waterlow, J.C., 1973. Note on the assessment and classification of protein-energy malnutrition in children. Lancet 87–89.
Welch, V.A., et al., 2017. Mass deworming to improve developmental health and wellbeing of children in low-income and middle-income countries: a systematic review and network meta-analysis. Lancet Glob. Health 5 (1), e40–e50.
WHO Multicentre Growth Reference Study Group, 2006. WHO child growth standards. length/height-for-age, weight-for-age, weight-for-length. In: Weight-For-Height and Body Mass Index-For-Age: Methods and Development. World Health Organization, Geneva. Department of Nutrition Health and Development.
World Health Organization, 2002. Optimal Duration of Exclusive Breastfeeding. Report of an expert consultation (document who/nhd/01.09,who/fch/cah/01.24). Geneva.
World Health Organization, 2018. Global Nutrition Targets 2025 to Improve Maternal, Infant and Young Child, vol. 2. World Health Organization, pp. 375–388, 6.

Coronary heart disease: Dietary patterns

Danielle E. Haslam[a,b], **Shatabdi Goon**[a], **Saloni Gautam**[b], **and Shilpa N. Bhupathiraju**[a,b], [a] Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States; and [b] Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, United States

© 2023 Elsevier Ltd. All rights reserved.

Introduction 241
A timeline of research on diet and CHD 241
The emergence of dietary pattern research 242
Dietary patterns associated with coronary heart disease risk 242
Mediterranean diet 242
Dietary approaches to stop hypertension 247
Plant-based dietary patterns 248
Dietary inflammatory potential 248
Other theoretical dietary indices to capture "healthy" dietary patterns 250
Empirical dietary patterns 251
Novel omics approaches for precision nutrition in CVD prevention 252
Utility of omics in precision nutrition 252
Omic signatures of dietary patterns and CHD 252
Conclusions 253
References 253
Further reading 255

Key points

- Over the past two decades, many studies have examined the association between overall dietary patterns and coronary heart disease (CHD) risk, which represents a stark shift from studies of individual foods/nutrients in the previous decades.
- Strong evidence indicates that the Mediterranean diet, Dietary Approaches to Stop Hypertension, and healthy plant-based diets lower CHD risk.
- Several additional theoretical dietary indices consistently associate with lower CHD risk, including the Healthy Eating Index, the Alternative Healthy Eating Index, and indices capturing dietary inflammatory potential.
- Empirical dietary pattern analysis affirms results from theoretical dietary pattern analysis, where data-driven "healthy" or "prudent" dietary patterns consistently associate with lower CHD across many diverse population-based studies.
- Areas where more research is needed in relation to dietary patterns and CHD risk include: (1) global dietary indices; (2) machine learning techniques for empirical dietary pattern analysis; and (3) the use of omics, particularly multi-omics, to understand individual responses to dietary patterns and inform personalized dietary recommendations.

Introduction

Cardiovascular diseases (CVD) are the leading cause of death globally, accounting for approximately 19.1 million deaths in 2020. CVD cost the US an estimated $228.7 billion annually in direct and indirect costs in 2017–2018. A recent study found that poor diet quality is the top modifiable risk factor related to CVD-related morbidity and mortality, accounting for approximately half of CVD deaths annually. Importantly, <1% of US adults achieve dietary patterns associated with ideal cardiovascular health. Thus, improving diet quality worldwide has a high potential to reduce CVD burden, particularly CHD, which is the leading cause of CVD death. In this article, we provide a brief history of dietary guidance to reduce CHD with a major focus on recent evidence linking overall dietary patterns to CHD and CVD risk.

A timeline of research on diet and CHD

Traditional dietary advice to prevent CHD has focused on single nutrients or foods, with a particular emphasis on types of dietary fat. The diet-heart hypothesis postulates that replacing dietary saturated fats with unsaturated fats reduces serum cholesterol, thereby lowering atherosclerosis and CHD risk. In the 1950s, Ancel Keys, an American physiologist, launched the Seven Countries Study and demonstrated that increases in dietary fat intake were positively linked to serum cholesterol and CHD risk. Consequently, in the

1980s–90s, limiting total dietary fat was a major focus of dietary recommendations to reduce CVD risk, particularly by reducing saturated fat and cholesterol intake. This led to the emergence of the low-fat diet era, which saw an explosion of many reduced-fat food products in the market that were often perceived as "healthy". However, these foods were often high in refined carbohydrates and added sugars to improve flavor and palatability that was lost due to their lower fat content. In the early 2000s, the landmark Women's Health Initiative Dietary Modification Trial showed that women who reduced their total fat intake by 8.2% of total energy had no significant benefit on incidence of CHD over 6 years (hazard ratio [HR] [95% confidence interval (CI)] 0.97 [0.90-1.06]) when compared to the control group. Since then, there has been a major shift to focus on the "quality" of fat rather than the "quantity" and dietary recommendations have shifted to recommend replacing saturated and *trans* fat with monounsaturated and polyunsaturated fatty acids to reduce CVD risk. Recently, using mathematical models, studies have examined the effect of substituting one food or nutrient with another on CHD risk since this closely mimics real-world food choices made by individuals. For example, one study demonstrated that substitution of saturated fat intake with carbohydrates from whole grains, but not carbohydrates from refined starches and/or added sugars, is associated with lower CHD risk.

As research continued to examine the role of diet, it became clear that dietary components other than fat were also important for lowering CHD risk. Many studies have now shown associations between an array of nutrients and foods and higher CHD risk and CHD-mortality, including higher intakes of added sugar, red/processed meat, sodium, and ultra-processed foods, along with lower intakes of fruits and vegetables, *n*-3 fatty acids/fish, polyunsaturated fatty acids, whole grains, dietary fiber, low- or non-fat dairy, nuts/seeds, and legumes.

Although evidence surrounding many of these foods has been replicated in large scale prospective studies and randomized clinical trials, some foods have remained controversial in their association with cardiometabolic risk. For example, eggs are often consumed in combination with processed meats and refined grains, making it difficult to extricate their individual effects from the effect of the overall dietary pattern. A detailed history of the wealth of research identifying which nutrients and foods associate most strongly with CHD is out of the scope of this article and has been described elsewhere (Bechthold et al., 2017; Bhupathiraju and Tucker, 2011).

The emergence of dietary pattern research

Given the complexities of examining the individual effects of single foods or nutrients on disease risk, research in the last few decades has placed a greater emphasis on the overall *dietary pattern*. A dietary pattern is the combination of foods and beverages that constitutes an individual's complete dietary intake over time. It is not surprising that the first diet scores examined in relation to CHD mortality continued their focus on types of dietary fat in "lipid-lowering" or "cholesterol lowering" dietary patterns. However, more recent studies have focused on the influence of all dietary components together on CHD risk. A theoretical *dietary pattern index* assigns scores according to individual intakes of certain foods to yield an overall score reflecting adherence to a dietary pattern of interest. We provide several examples of how different dietary pattern analysis techniques can be utilized in research as we describe the current state of the evidence linking dietary patterns to CHD risk below. **Table 1** summarizes results from systematic reviews and meta-analyses examining the association between dietary patterns and CVD and/or CHD.

Dietary patterns associated with coronary heart disease risk

Mediterranean diet

The *Mediterranean diet* pattern was one of the first dietary patterns recognized by epidemiologist Dr. Ancel Keys in the 1960s Seven Countries Study. Keys recognized the lower death rate from CHD among individuals living in the Mediterranean region compared to other parts of the world and hypothesized that their diet and lifestyle habits may be a key driver. Key components of the Mediterranean diet include high intakes of cereals, vegetables, fruits, beans, nuts, and olive oil as the primary fat, moderate intakes of potatoes, fish, poultry, eggs, and wine, and low intakes of red meat and sweets containing added sugars (see **Fig. 1** for additional information).

One of the first examples of a systematic exploration of diet quality was the development of a Mediterranean Diet Score (MDS) by Drs. Antonia Trichopoulou and Dimitrios Trichopoulos. In the MDS, individuals were assigned a score of 0 or 1 if their intakes were below or above the median for 8 components of the MDS: high ratio of monounsaturated: saturated fat intake; moderate alcohol intake; high legume intake; high cereal intake; high fruit intake; high vegetable intake; low meat intake; and low dairy intake. Since then, various modern adaptations of the MDS have emerged, all with similar guiding principles. In 1995, Drs. Walter Willet and Trichopoulos further described and popularized the Mediterranean diet pyramid as a healthy dietary pattern to reduce CVD risk.

Adherence to a Mediterranean diet pattern has been associated with lower CHD and CHD mortality not only in Mediterranean regions but also in many populations across the world. In a 2020 meta-analysis that included US and European cohorts, higher adherence to a Mediterranean diet pattern was associated with a 27% lower risk of CHD and a 17% lower risk of CHD mortality (Becerra-Tomás et al., 2020). A quasi-experimental study design using observational data showed that improving adherence to the Mediterranean diet over time has also been associated with lower CVD risk and CVD mortality among US

Table 1 Summary of systematic reviews and meta-analyses of the effects of dietary patterns on coronary heart disease, cardiovascular disease, and CVD mortality.

Study reference	*Design*	*Exposure*	*Outcome*	*Results*
Mediterranean diet				
Sofi et al., 2008	Meta-analysis of 12 prospective cohort studies; 1,574,299 participants; follow-up ranging from 3 to 18 years	Adherence to MD: A two-point increase in adherence score for MD	CVD mortality	CVD death: RR 0.91; 95% CI: 0.87–0.95
Sofi et al., 2010	Systematic review and meta-analysis of 7 cohort studies; >2 million participants; follow-up ranging from 4.9 to 20 years	Adherence to MD: A two-point increase in adherence score for MD	CVD incidence or mortality	CVD incidence or mortality: RR 0.90; 95% CI: 0.87–0.93
Sofi et al., 2014	Meta-analysis of 18 studies; 4,172,412 participants	Adherence to MD: A two-point increase in adherence score for MD	CVD risk	CVD: RR 0.90; 95% CI: 0.87–0.92
Liyanage et al., 2016	Systematic review and meta-analysis of 6 studies; 10,950 participants; follow-up ranging from 6 months to 9 years	Adherence to MD: Highest vs. lowest category	CHD/CVD	CHD events: RR 0.65; 95% CI 0.50–0.85; no association with CVD death
Grosso et al., 2017	Meta-analysis of 20 studies; 888,257 participants; follow-up ranging from 2 to 20 years	MD adherence: highest quintile vs. lowest	CVD/CHD	CVD death: RR: 0.76; 95% CI: 0.68–0.83 CHD death: RR: 0.72; 95% CI: 0.60–0.86
Rosato et al., 2019	Systematic review and meta-analysis of 29 studies	Adherence to MD: Highest vs. lowest category	CHD	CHD events: RR 0.70; 95% CI 0.62–0.80
Becerra-Tomás et al., 2020	Systematic review and meta-analysis of 3 RCTs and 38 cohort studies; 1,154,443 participants; follow-up ranging from 2 to 26 years	MD adherence: Highest vs. lowest	CVD/CHD	CVD death: RR: 0.79; 95% CI: 0.77–0.82 CHD incidence: RR: 0.73; 95% CI: 0.62–0.86 CHD death: RR: 0.83; 95% CI: 0.75–0.92
Tang et al., 2021	Meta-analysis of 7 prospective cohort studies; 37,879 participants; follow-up ranging from 3.8 to 10.0 years	Highest versus lowest Adherence to MD	CVD mortality	Each 2-unit increment in a score of adherences to MD. CVD death: HR 0.91; 95% CI: 0.82–1.01
Dietary approaches to stop hypertension				
Salehi-Abargouei et al., 2013	Systematic review and meta-analysis of 6 studies; 259,984 participants; follow-up ranging from 7 to 24 years	Adherence to the DASH diet: Highest vs. lowest	CHD and CVD events or mortality	Highest vs. lowest: CVD death: RR 0.80; 95% CI: 0.74–0.86 and CHD death: RR 0.79; 95% CI: 0.71–0.88
Schwingshackl and Hoffmann, 2015	Systematic review and meta-analysis of 15 cohort studies; 1,020,642 participants; follow-up ranging from 5 to ≥24 years	DASH score: Highest vs. lowest	CVD incidence and mortality	CVD mortality or incidence: 0.82; 95% CI: 0.80–0.86
Chiavaroli et al., 2019	Umbrella review of 46 studies; 946,554 participants	Adherence to the DASH diet: Highest vs. lowest	CHD incidence	CHD incidence RR 0.79; 95% CI: 0.71–0.88
Yang et al., 2019	Systematic review and meta-analysis of 7 studies; 377,725 participants, follow-up ranging from 2.8 to 24 years	Adherence to the DASH diet: Highest vs. lowest	Coronary artery disease incidence	CAD RR 0.82; 95% CI: 0.78–0.87
Soltani et al., 2020	Systematic review and meta-analysis of 12 cohort studies; 1,314,675 participants; follow-up ranging from 6.5 to 19.2 years	Adherence to DASH score: Per each 5-point increment of adherence to the DASH diet score	CVD mortality	CVD death: HR 0.96; 95% CI: 0.95–0.98
Morze et al., 2020	Systematic review and meta-analysis of 113 studies; 3,277,684 participants; follow-up ranging from 2 to ≥24 years	DASH score: Highest vs. lowest	CVD incidence and mortality	CVD mortality or incidence: 0.80; 95% CI: 0.79–0.84
Plant-based diets				
Huang et al., 2012	Systematic review and meta-analysis of 7 studies; 124,706 participants	Vegetarian diet	CVD mortality	CVD death: RR 0.71; 95% CI: 0.56–0.87

(Continued)

Table 1 Summary of systematic reviews and meta-analyses of the effects of dietary patterns on coronary heart disease, cardiovascular disease, and CVD mortality.—cont'd

Study reference	*Design*	*Exposure*	*Outcome*	*Results*
Kwok et al., 2014	Systematic review and meta-analysis of 8 studies; 183,321 participants; follow-up ranging from 5.8 to 21 years	Vegetarian diet: Highest vs. lowest	CVD	Ischemic heart disease death: RR 0.60; 95% CI: 0.43–0.80.
Dinu et al., 2017	Systematic review and meta-analysis of 10 prospective cohort studies; 72,298 participants; 4.1–21 years	Vegetarian vs. omnivorous	CVD outcomes	Ischemic heart disease RR 0.75; 95% CI: 0.68–0.82, no significant findings for total CVD deaths
Glenn et al., 2019	Systematic review and meta-analysis of 7 studies; 197,737 participants; follow-up ranging from 5.5 to 21 years	Vegetarian diet: Vegetarian vs. non-vegetarian	CHD/CVD mortality	CHD death: RR 0.78; 95% CI: 0.69–0.88 CHD incidence: RR 0.72; 95% CI: 0.61–0.85 No association with CVD
Kahleova et al., 2019	Systematic review and meta-analysis of 16 studies; 197,737 participants; follow-up ranging from 2 to 26 years	Vegetarian diet: Highest vs. lowest	CHD	CHD death: RR, 0.78; 95% CI: 0.69–0.88
Kaiser et al., 2021	Systematic review and meta-analysis of 7 studies; 73,000 participants	Vegan diet: Highest vs. lowest	CVD/CHD outcomes	No evidence for an increased/decreased risk of CHD/CVD and mortality
Rees et al., 2021	Meta-analysis of 13 RCTs; 995 participants	Vegan diet: Highest vs. lowest	CVD mortality	No significant association found
Jafari et al., 2021	Systematic review and meta-analysis of 12 studies; 508,861 participants; follow-up ranging from 4.8 to 25 years	Plant based diet: Highest vs. lowest	CHD/CVD mortality	CVD death: 0.92; 95% CI: 0.85, 0.99 CHD death: 0.76; 95% CI: 0.68, 0.85.
Quek et al., 2021	Systematic review and meta-analysis of 13 studies; 410,085 participants	Plant based diet: Highest vs. lowest Vegetarian vs. regular meat eaters	CVD/CVD mortality	PDI: CVD death: 0.92; 95% CI: 0.86, 0.99 CVD: 0.90; 95% CI: 0.82, 0.98 uPDI: CVD death: 1.05; 95% CI: 1.01, 1.09 CVD: 1.11; 95% CI: 0.79, 1.56 hPDI: CVD death: 0.91; 95% CI: 0.79, 1.05 CVD: 0.87; 95% CI: 0.80, 0.95 Vegetarian: CVD mortality: 0.89; 95% CI: 0.78, 1.01 CVD: 0.81; 95% CI: 0.72, 0.91
Gan et al., 2021	Systematic review and meta-analysis of 10 studies; 698,707 participants	PBD, PDI, hPDI, uPDI Highest vs. lowest	CVD CHD	PBD: CVD: 0.84; 95% CI: 0.79, 0.89 CHD: 0.89; 95% CI: 0.81, 0.97 PDI: CVD: 0.85; 95% CI: 0.80, 0.90 hPDI: CVD: 0.84; 95% CI: 0.75, 0.94 uPDI: CVD: 1.13; 95% CI: 1.02, 1.26
Dietary inflammatory potential				
Zhong et al., 2017	Meta-analysis of 9 studies; 134,067 participants; follow-up ranging from 1.2 to 24.7 years	DII score: Highest vs. Lowest	CVD mortality	CVD mortality: RR 1.24; 95% CI: 1.01–1.51
Shivappa et al., 2017	Meta-analysis of 5 studies; 762,291,727 participants; follow-up ranging from 5 to 22 years	DII score: Highest vs. lowest	CVD	CVD death: HR 1·05; 95% CI 1.03–1.07

Table 1 Summary of systematic reviews and meta-analyses of the effects of dietary patterns on coronary heart disease, cardiovascular disease, and CVD mortality.—cont'd

Study reference	*Design*	*Exposure*	*Outcome*	*Results*
Shivappa et al., 2018	Meta- analysis of 14 studies; 161,337 participants; follow-up ranging from 4.3 to 26 years	DII score: Highest vs. lowest DII category	CVD	36% increased risk of CVD incidence and mortality: RR 1.36, 95% CI: 1.19–1.57
Namazi et al., 2018	Systematic review and meta-analysis of 17 studies; 163,310 participants; follow-up ranging from 5 to 20.7 years	DII score: High vs. Low	CVD mortality	CVD mortality: pooled HR: 1.30, 95% CI: 1.07–1.57
Aslani et al., 2020	Systematic review and meta-analysis of 18 studies; 343,694 participants; follow-up ranging from 5 to 25.8 years	DII score: High vs. low	Cardiometabolic disease mortality	Cardiometabolic disease mortality: HR 1.29; 95% CI: 1.18–1.41
Ji et al., 2020	Meta-analysis of 10 studies; 385,765 participants; follow-up ranging from 13.5 to 20.7 years	DII score: Highest vs. lowest	CVD	CVD death: RR 1.31; 95% CI: 1.19–1.44
Liu et al., 2021	Umbrella review of 35 meta-analyses	DII score: Highest VS. Lowest	CVD	CVD mortality: RR 1.31; 95% CI: 1.19–1.44
Farazi et al., 2021	Umbrella review of 127 studies	DII score: Highest vs. lowest	CVD	CVD mortality: RR 1.32; 95% CI: 1.19 to 1.46
Marx et al., 2021	Umbrella review of 15 meta-analyses	DII score: High vs. low	CVD	CVD death: HR 1.35; 95% CI 1.11–1.63
Healthy eating index				
Schwingshackl et al., 2018	Systematic review and meta-analysis of 68 studies; 1,670,179 participants; follow-up ranging from 3 to 24 years	HEI score: Highest vs. lowest,	CVD mortality or incidence	CVD mortality or incidence: RR 0.83, 95% CI: 0.79–0.87
Morze et al., 2020	Systematic review and meta-analysis of 113 studies; 3,277,684 participants; follow-up ranging from 2.8 to 32 years	HEI score: Highest vs. lowest	CVD mortality or incidence	CVD mortality or incidence: RR 0.80, 95% CI: 0.78–0.82
Western and prudent diet				
Li et al., 2015	Systematic review and meta-analysis of 13 studies; 338,787 participants; follow-up ranging from 5 to 15 years	Highest vs. lowest adherence to prudent and western diet	CVD	Prudent diet and CVD death: summary relative risk estimates 0.81; 95% CI 0.75–0.87; no associations with western diet
Hou et al., 2015	Meta-analysis of 12 studies; 409,780 participants; follow-up ranging from 4.6 to 13 years	Highest vs. lowest adherence to prudent or western diet	CHD	Prudent: 0.80, 95% CI: 0.74–0.87 Western: 1.05; 95% CI: 0.86–1.27
Rodríguez-Monforte et al., 2015	Systematic review and meta-analysis of 21 studies; 610,691 participants	Highest vs. lowest adherence to prudent or western diet	CVD/CHD	Prudent and CVD death: RR 0.69; 95% CI 0.60, 0.78 Prudent and CHD death: RR 0.83; 95% CI 0.75, 0.92 Western and CVD death: RR 1.14; 95% CI 0.92–1.42 Western and CHD death: RR 1.03; 95% CI 0.90–1.17
Zhang et al., 2015	Meta-analysis of 35 studies	Highest vs. lowest prudent and western category	CHD	Prudent and CHD death: OR 0.67; 95% CI: 0.60–0.75 Western and CHD death: OR 1.45; 95% CI: 1.05–2.01
Jayedi et al., 2020	Umbrella review	Healthy and unhealthy dietary patterns	CHD, CVD mortality	Low-quality evidence for an inverse association with healthy dietary patterns and no association with unhealthy dietary patterns
Other diet quality indices				
Noto et al., 2013	Systematic review and meta-analysis of 17 studies; participants ranging from 647 to 129,716; follow-up ranging from 6 to 16 years	Low carb diet: Highest vs. lowest	CVD	No significant association with CVD mortality

(Continued)

Table 1 Summary of systematic reviews and meta-analyses of the effects of dietary patterns on coronary heart disease, cardiovascular disease, and CVD mortality.—cont'd

Study reference	*Design*	*Exposure*	*Outcome*	*Results*
Jalilpiran et al., 2020	Systematic review and meta-analysis of 13 studies; 930,153 participants; follow-up ranging from 7 to 23.6 years	Nordic diet: Highest vs. lowest adherence	CVD	CVD death: 0.78; 95% CI: 0.74–0.83

Abbreviations: CVD, Cardiovascular Disease; CI, Confidence Interval; CHD, Coronary Heart Disease; DASH, Dietary Approaches to Stop Hypertension; DII, Dietary Inflammatory Index; HEI, Healthy eating index; HR, Hazard Ratio; MD, Mediterranean Diet; RR, Relative Risk.

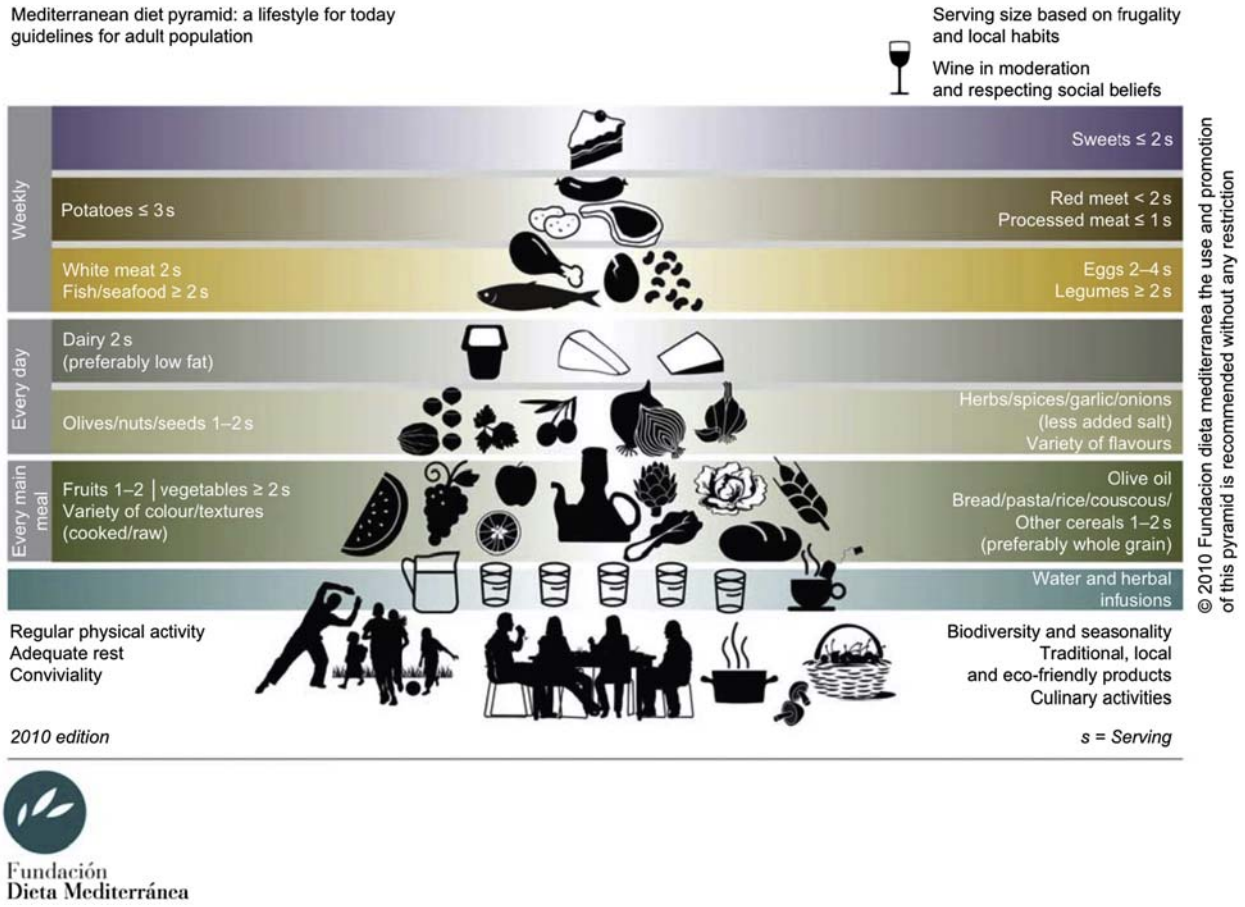


Fig. 1 Overview of the Mediterranean dietary pattern.

men and women. Strengthening the case for potential causality, randomized controlled trials evaluating the effects of the Mediterranean diet on intermediate biomarkers related to the development of CHD show similar results. For example, the Mediterranean diet has been associated with improved blood lipids, blood pressure, body weight, endothelial function, and inflammatory biomarkers.

The largest study to date evaluating the effects of the Mediterranean diet pattern on primary prevention of cardiovascular diseases is the landmark Primary Prevention of Cardiovascular Disease with a Mediterranean Diet (PREDIMED) trial (Estruch et al., 2018). In the PREDIMED, participants were randomized to a Mediterranean diet supplemented with extra-virgin olive oil, a Mediterranean diet supplemented with mixed nuts, or a control diet including advice to reduce total dietary fat. After a median follow-up time of 4.8 years, participants in the Mediterranean diet with extra-virgin olive oil and Mediterranean diet with nuts arms were 31% and 28% less likely, respectively, to experience a cardiovascular event compared to participants following the control diet. Although the PREDIMED trial provides the strongest evidence to date supporting the adoption of the Mediterranean diet pattern to help reduce cardiovascular disease risk, prior randomized controlled trials (RCT) have echoed this sentiment.

Given the strength of the evidence, the 2015–2020 and 2020–2025 Dietary Guidelines for Americans recommend the Mediterranean dietary pattern as a healthy eating pattern to lower chronic disease risk. As highlighted at the end of this article (**Utility of omics in precision nutrition**), incorporation of omics data into existing and new observational and interventional studies of the Mediterranean diet may allow us to learn more about how personalized recommendations within the context of the overall Mediterranean dietary pattern may further improve CHD prevention throughout the lifespan.

Dietary approaches to stop hypertension

A more recent success story began in the 1990s with the *Dietary Approaches to Stop Hypertension* (DASH) clinical trial (Appel et al., 1997; Sacks et al., 1995). In a seminal 8-week RCT, the DASH Collaborative Research Group demonstrated that following a dietary pattern rich in fruits, vegetables, low-fat dairy, whole grains, poultry, fish, and nuts/legumes and reduced saturated fat, cholesterol, red/processed meats, sweets, and sugar-sweetened beverages can substantially lower blood pressure (BP) compared to control diets aiming to increase fruit and vegetable intake alone (mean [97.5% CI] change in systolic BP: -2.7 [-4.6, -0.9] mmHg and diastolic BP -1.9 [-3.3, -0.6] mmHg) or reflecting the typical US diet (mean [97.5% CI] change in systolic BP: -5.5 [-7.4, -3.7] mmHg and diastolic BP -3.0 [-4.3, -1.6] mmHg). The DASH-Sodium Trial provided further evidence to support that the DASH diet is effective at several levels of sodium intake, with the greatest reduction in blood pressure occurring in individuals following the DASH dietary pattern and consuming <1500 mg/day of sodium. This landmark study has shaped guidelines for the prevention of hypertension and cardiovascular disease over the past 20 years and spurred several investigations into the DASH diet and a variety of health outcomes.

The benefits of the DASH diet extend to lower risk of CVD. In a meta-analysis of 47 studies, higher adherence to a DASH style dietary pattern (see **Fig. 2** for additional information) was associated with a 20% lower risk of incident CVD and CVD mortality (Morze et al., 2020). These associations appear to be consistent across populations, including individuals across North America, Europe, and Asia, as well as individuals with type 2 diabetes. Improving adherence to the DASH diet over time has also been

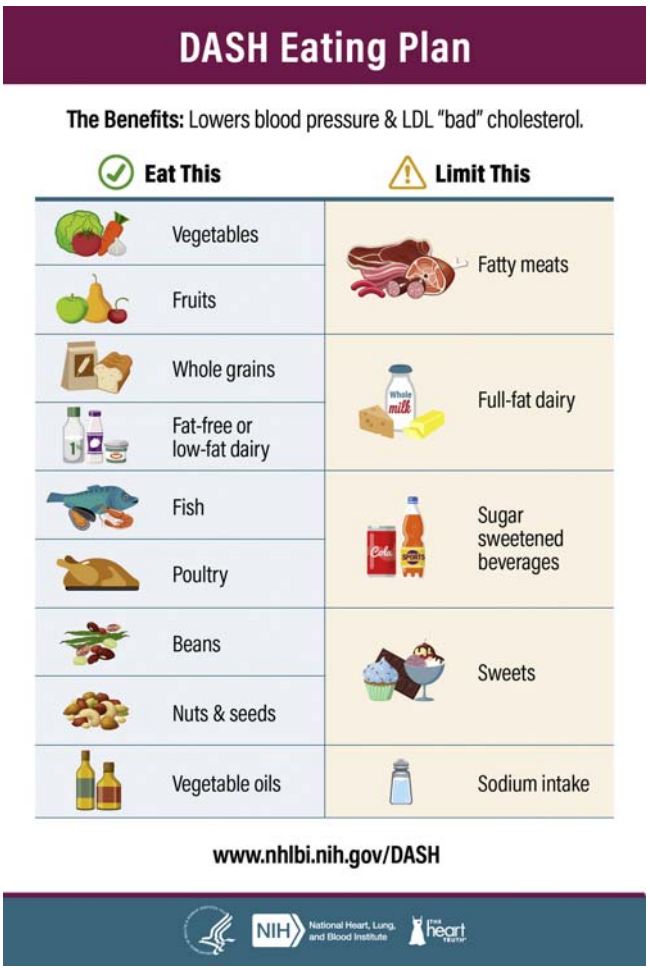


Fig. 2 Overview of the Dietary Approaches to Stop Hypertension (DASH) dietary pattern.

associated with lower CVD risk and CVD mortality among US men and women. Higher DASH diet adherence has been shown to be associated with several cardiometabolic risk factors, including lower low-density lipoprotein-cholesterol (LDL-C) and type 2 diabetes risk.

Plant-based dietary patterns

The role of plant-based diets (PBDs) in preventing and managing chronic disease has long been of interest to researchers. Although the definition of PBDs can be diverse, they are primarily characterized by lower intakes of animal foods and higher intakes of plant-based food products, including fruits, vegetables, grains, and legumes. For instance, a vegetarian diet is mostly focused on plant-based products and excludes meat, poultry, and seafood with a low to moderate intake of dairy products or eggs. On the other hand, a vegan diet excludes all animal-based products, including dairy products, eggs, and honey. In the 1950s, Dr. Mervyn Hardinge first began investigations into how vegetarian diets influence health. Following these preliminary investigations, several studies investigated dietary patterns among the Seventh-day Adventists from California, a religious group that adheres to a lacto-ovo-vegetarian diet. These studies showed that Seventh-day Adventists following a non-vegetarian dietary pattern had approximately twice the rate of CHD mortality compared to vegetarian Adventists after adjusting for several important confounders. Building on these early studies, a large body of research has examined the relationship between PBDs and CHD.

To understand the mechanisms through which PBDs may be beneficial to cardiovascular health, several RCTs examined the effect of PBDs on intermediate cardiometabolic risk factors. A meta-analysis of 11 RCTs found that compared to non-vegetarian diets, vegetarian diets lowered LDL-C, but also lowered high-density lipoprotein cholesterol (HDL-C) and had no effect on triglyceride (TG) concentrations. In a meta-analysis that included 7 clinical trials and 32 observational studies, vegetarian diets significantly lowered blood pressure. Both observational and clinical data remain inconclusive about the associations between vegan dietary patterns and CHD outcomes and cardiometabolic risk factors. However, these studies were generally small and larger studies are needed to understand how vegan dietary patterns influence cardiometabolic disease risk.

When examining studies of plant-based diets and cardiometabolic health outcomes, it is vital to understand which foods are being replaced. For example, while replacing refined grains with healthy sources of animal protein, such as fatty fish, may be beneficial to cardiometabolic health, this may not be true when refined grains are replaced with processed meat. Likewise, it remains imperative to examine the quality of plant-based foods. Consistent evidence has shown that low quality carbohydrates, such as refined grains and sugar sweetened beverages, are adversely associated with cardiometabolic risk, while high quality carbohydrates, such as whole grains, and legumes, are associated with better cardiometabolic outcomes. To address these issues, Satija et al. developed the *plant-based diet index* (PDI), the healthy PDI, and the unhealthy PDI (Satija et al., 2017). In the overall PDI, all plant-based foods, regardless of quality, are scored positively while all animal foods are scored negatively. For the hPDI, all healthy plant-based foods (see **Table 2** for additional information) are scored positively, while all animal foods and unhealthy plant-based foods are scored negatively. For the uPDI, all unhealthy plant-based foods are scored positively, while healthy plant-based foods and animal foods are scored negatively. To illustrate that quality of PBDs is important, an analysis by Satija et al. showed that participants with the greatest adherence to a hPDI had a 25% lower risk of CHD, while those with higher adherence to a uPDI had a 32% higher CHD risk. Similar associations have been observed in other studies as summarized in a recent meta-analysis of 5 cohorts, the hPDI was associated with a 13% lower risk of CVD and the uPDI was associated with a 5% higher risk of CVD-specific mortality. Further, a 10-point increase in the hPDI or uPDI was associated with a 9% lower or 8% higher risk of CVD mortality among US men and women, respectively.

Collectively, the beneficial effects of healthful PBDs on established cardiometabolic risk factors found in RCTs, and their inverse associations with CHD in prospective cohort studies provide evidence that healthful PBDs likely reduce CHD risk. In fact, the 2015–2020 and 2020–2025 Dietary Guidelines for Americans recommend a healthy vegetarian diet as an eating pattern to lower risk of chronic diseases. In addition to their health benefits, healthy PBDs are environmentally sustainable and fit within the food system's planetary boundaries. A more comprehensive discussion of the potential mechanisms through which healthy PBDs can influence cardiometabolic health is provided in **Omic signatures of dietary patterns and CHD** section.

Dietary inflammatory potential

Inflammation is a biological response to injury that plays an important role in the initiation and progression of atherosclerosis. Chronic inflammation appears in conjunction with several chronic diseases and accumulating evidence suggests that it contributes to CVD pathogenesis. Dietary components exhibit pro- and anti-inflammatory properties and are a modifiable targets to reduce chronic inflammation and CVD risk. The *Dietary Inflammatory Index* (DII) and the *empirical dietary inflammatory pattern* (EDIP) are two tools that have been developed to capture the inflammatory potential of overall dietary patterns.

The DII is a literature-based dietary pattern developed to rank individuals' diets based on their inflammatory potential. The original scoring algorithm developed in 2009 was subsequently updated and improved in 2014. The current DII includes 45 foods, nutrients, or compounds with weights reflecting the number of research articles that have observed inflammatory or anti-inflammatory properties for each dietary component. Several studies have validated the DII, observing positive associations between the DII and inflammatory biomarkers, such as C-reactive protein (CRP), interleukin-6 (IL-6), and tumor necrosis factor alpha receptor 2 (TNFα-R2).

Table 2 Examples of food items constituting the 18 food groups in the Plant-based Diet Index (PDI) from the 1984 Nurses' Health Study food frequency questionnaire.

	Plant food groups	*PDI*	*hPDI*	*uPDI*
Healthy				
Whole grains	Whole grain breakfast cereal, other cooked breakfast cereal, cooked oatmeal, dark bread, brown rice, other grains, bran, wheat germ, popcorn	Positive scores	Positive scores	Reverse scores
Fruits	Raisins or grapes, prunes, bananas, cantaloupe, watermelon, fresh apples or pears, oranges, grapefruit, strawberries, blueberries, peaches or apricots or plums	Positive scores	Positive scores	Reverse scores
Vegetables	Tomatoes, tomato juice, tomato sauce, broccoli, cabbage, cauliflower, brussels sprouts, carrots, mixed vegetables, yellow or winter squash, eggplant or zucchini, yams or sweet potatoes, spinach cooked, **spinach** raw, kale or mustard or chard greens, iceberg or head lettuce, romaine or leaf lettuce, celery, mushrooms, beets, alfalfa sprouts, garlic, corn	Positive scores	Positive scores	Reverse scores
Nuts	Nuts, peanut butter	Positive scores	Positive scores	Reverse scores
Legumes	String beans, tofu or soybeans, beans or lentils, peas or lima beans	Positive scores	Positive scores	Reverse scores
Vegetable oils	Oli-based salad dressing, vegetable oil used for cooking	Positive scores	Positive scores	Reverse scores
Tea and coffee	Tea, coffee, decaffeinated coffee	Positive scores	Positive scores	Reverse scores
Less healthy				
Fruit juices	Apple cider (nonalcoholic) or juice, orange juice, grapefruit juice, other fruit juice	Positive scores	Reverse scores	Positive scores
Refined grains	Refined grain breakfast cereal, white bread, English muffins or bagels or rolls, muffins or biscuits, white rice, pancakes or waffles, crackers, pasta	Positive scores	Reverse scores	Positive scores
Potatoes	French fries, baked or mashed potatoes, potato or corn chips	Positive scores	Reverse scores	Positive scores
Sugar sweetened beverages	Colas with caffeine and sugar, colas without caffeine but sugar, other carbonated beverages with sugar, noncarbonated fruit drinks with sugar	Positive scores	Reverse scores	Positive scores
Sweets and desserts	Chocolates, candy bars, candy without chocolate, cookies (home-baked and ready-made), brownies, donuts, cake (hom- baked and ready-made), sweet roll (home-baked and ready-made), pie (home-baked and ready-made), jams or jellies or preserves or syrup or honey	Positive scores	Reverse scores	Positive scores
Animal food groups				
Animal fat	Butter added to food, butter or lard used for cooking	Reverse scores	Reverse scores	Reverse scores
Dairy	Skim low fat milk, whole milk, cream, sour cream, sherbet, ice cream, yogurt, cottage or ricotta cheese, cream cheese, other cheese	Reverse scores	Reverse scores	Reverse scores
Egg	Eggs	Reverse scores	Reverse scores	Reverse scores
Fish or seafood	Canned tuna, dark meat fish, other fish, shrimp or lobster or scallops	Reverse scores	Reverse scores	Reverse scores
Meat	Chicken or Turkey with skin, chicken or Turkey without skin, bacon, hot dogs, processed meats, liver, hamburger, beef or pork or lamb mixed dish, beef or pork or lamb main dish	Reverse scores	Reverse scores	Reverse scores
Miscellaneous animal-based foods	Pizza, chowder or cream soup, mayonnaize or other creamy salad dressing	Reverse scores	Reverse scores	Reverse scores

hPDI = Healthful Plant-Based Diet Index; PDI = Overall Plant-Diet Index; uDPI = Unhealthful Plant-Based Diet Index.

Consistent with the biomarker data, observational studies have observed a positive association between the DII and CVD risk. A meta-analysis of 6 cohort studies found that higher DII scores were associated with a 35% higher CVD risk (Namazi et al., 2018). A recent umbrella review of the DII and human health ranked the evidence for this positive association between the DII and myocardial infarction as "convincing" according to the AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews) quality assessment tool. However, the largest study to date (n = 15,693) included in these meta-analyses was a cross-sectional study examining the association between the DII and diagnosis of a previous circulatory disorder, which represents a lower quality of evidence than a prospective cohort study. A more recent meta-analysis examined the evidence for an association between the DII and

cardiometabolic diseases in cohort and non-cohort studies separately and estimated a 35% higher risk for cardiometabolic diseases prospective cohort studies (myocardial infarction, ischemic heart disease, stroke, congestive heart failure, and/or CHD). Thus, current evidence surrounding DII and cardiometabolic risk has been consistent, although larger studies in more diverse populations are warranted.

Unlike the DII, the EDIP is a data-driven food-based index that was developed using a combination of reduced-rank regression and step-wise linear regression based on circulating concentrations of CRP, IL-6, and TNFα-R2. A total of 18 foods groups that contribute either pro-inflammatory or anti-inflammatory properties in an overall dietary pattern were selected. A large prospective study among 210,145 individuals with 9794 CHD cases found that individuals with higher scores on the EDIP, reflecting a more pro-inflammatory diet, had a 46% higher risk of developing CHD. The EDIP was also associated with higher concentrations of pro-inflammatory markers (IL-6, soluble intercellular adhesion molecule-1 [sICAM-1], TNFα-R1, TNFα-R2, CRP, and leptin) and lower concentrations of an anti-inflammatory marker (adiponectin). Other studies have demonstrated positive associations between the EDIP and cardiometabolic risk factors, including weight gain, type 2 diabetes, and metabolic syndrome.

The overall body of literature indicates that a higher inflammatory potential of diet is associated with a higher risk of CHD. The DII and EDIP are two separate tools that can be used to characterize the inflammatory potential of diet. Further studies and refinement of these tools may lead to more consistent recommendations about the food components that best reflect the inflammatory potential of diet and CHD risk.

Other theoretical dietary indices to capture "healthy" dietary patterns

Theoretical dietary indices offer the distinct advantage of measuring adherence to dietary guidelines released by government institutes, academic institutions, foundations, and other non-profit organizations aiming to improve overall health or reduce the risk for specific diseases. These indices score participants based on how well their intakes align with the recommended intakes. The nutrient- and/or food-based targets within these indices change over time as recommendations are updated to reflect the latest scientific evidence.

One example is the *Healthy Eating Index* (HEI), which was first developed by Dr. Eileen Kennedy and colleagues in 1995 to capture adherence to the US Dietary Guidelines for Americans. Several iterations of the HEI have followed this original version and have been updated every 5 years with the release of the Dietary Guidelines for Americans. Adherence to the 2015–2020 Dietary Guidelines, which was quantified using the HEI-2015, was associated with a lower risk of CHD (HR [95% CI] 0.78 [0.74-0.82]) among 209,133 US men and women followed for up to 32 years. A recent meta-analysis of 13 studies estimated an approximately 19% lower incidence of CVD or CVD mortality among participants in the highest quartile of adherence to various versions of the HEI compared to those in the lowest quartile (Morze et al., 2020).

The *Alternative Healthy Eating Index* (AHEI) is a dietary index first developed in 2002 to capture a dietary pattern predictive of reduced chronic disease risk. It was created to improve upon an earlier dietary index, the Recommended Food Score, which was a simpler summary of diet quality. An updated version of the AHEI was created in 2010 to reflect that latest scientific evidence. The AHEI-2010 has been extensively studied and better adherence to the AHEI-2010 was associated with a nearly 23% lower risk of CVD or CVD mortality (Morze et al., 2020). Improved diet quality as measured by changes in the AHEI-2010 over time has also been associated with lower CVD risk (greatest vs. least improvement HR% [95% CI]: 0.92 [0.87-0.99] in each 4-year period) and CVD mortality (greatest vs. least improvement HR [95% CI]: 0.85 [0.78-0.91] over 12 years) among US men and women. While a significant proportion of diet indices have been developed in the US, research in other countries have also utilized dietary indices to quantify adherence to country-specific dietary recommendations. For example, the *Healthy Nordic Diet Index* was created to reflect adherence to Nordic nutrition recommendations, which include consumption of whole grains, berries, fruits, vegetables, unsaturated oils, low-fat dairy, fish, and poultry, and avoidance of sugar-sweetened beverages. Consistent data has shown that higher adherence to the healthy Nordic Diet was associated with a 20% lower risk for myocardial infarction after adjusting for several risk factors and confounders among 82,116 participants from Germany and Denmark.

The *American Heart Association Diet Score* (AHA-DS) is a dietary index first developed in 2013 as a means to capture adherence to the dietary guidelines released by the American Heart Association. The scoring criteria have subsequently been updated and are used in the American Heart Association's annual heart disease and stroke statistics report to monitor dietary targets to reduce CVD risk. Although many of the criteria in the AHA-DS are similar to those in the AHEI, HEI, DASH, and Mediterranean diet, more studies are needed to understand how the AHA-DS index associates with incident CHD risk in both epidemiological and interventional studies.

Most of the diet quality indices described thus far were developed and validated in high-income countries and are often country-specific. Recent studies have aimed to create diet quality indices that are optimized for application in low- and middle-income countries. Given the double burden of malnutrition in many developing countries, there have been concerted efforts to develop diet quality metrics that can capture both risk of nutrient inadequacy and risk of chronic disease, while also monitoring diet quality in populations and population subgroups. These efforts led to the development of the *Global Diet Quality Score* (GDQS), which was based on the prime diet quality score. The GDQS was developed using data from women in 10 African countries, China, India, Mexico, and the US. A major difference between the GDQS and other diet scores is that it is entirely food-based, which removes the need for investigators to rely on nutrient databases in resource poor settings. The GDQS has been associated with several cardiometabolic risk factors in these countries, but its association with CHD risk is yet to be evaluated. These internationally optimized diet quality scores will be crucial for monitoring dietary intakes around the world to identify individuals at high risk for CHD.

Empirical dietary patterns

Data-driven or empirical approaches are complementary approaches to dietary indices and can yield unique insights about dietary patterns and health outcomes. The two most common empirical dietary pattern approaches include *cluster analysis* and *principal components analysis* (PCA). These approaches, when utilized in different settings and populations, have often identified a "prudent" and a "Western-style" dietary pattern. The prudent or "healthy" dietary pattern is often defined by higher intakes of fruits, vegetables, fish and lean meats, legumes/nuts, whole grains, olive oil, soy, poultry, and low-fat dairy. The Western or "unhealthy" dietary pattern, on the other hand, reflects high consumption of red/processed meat, refined grains, high-fat dairy, added sugars, butter, and potatoes, and low intakes of fruits and vegetables. In several meta-analyses that included participants from Asia, North America, and Europe, the prudent/healthy dietary pattern was repeatedly associated with a lower CHD risk. On the other hand, associations of greater adherence to a "Western-style" dietary pattern with higher CHD risk were less consistent. This may be because empirical "unhealthy" dietary patterns derived in individual studies may include different combinations of foods that display differential associations with CHD.

Because empirical dietary pattern analyses are data-driven, they have often identified unique patterns that reflect traditional diets of various cultural groups that has allowed investigators to examine the healthfulness of a "traditional" diet. For example, in the Reasons for Geographic and Racial Differences in Stroke (REGARDS) study, a US Southern-style dietary pattern characterized by added fats, fried foods, eggs, organ meats, processed meats, and sugar-sweetened beverages was identified that was associated with a higher incidence of CHD (comparing quartile 4 [Q4] with quartile 1 [Q1], (HR [95% CI]: 1.37 [1.01, 1.85]). Likewise, a PCA-derived traditional Dutch dietary pattern, characterized by high intakes of potatoes, red/processed meat, eggs, coffee, and boiled vegetables/legumes and a low intake of fast food, sugar/sweets, savory sauces, French fries, soy products, and cereals, was also associated with a higher risk of coronary artery disease (HR [95% CI]: 1.25 [1.07, 1.47]). Empirical dietary patterns present a promising methodology to generate hypotheses that may inform future dietary indices reflecting traditional dietary patterns. In addition to the Mediterranean diet, which has been extensively studied, larger studies are needed that explore the associations of traditional dietary patterns with CHD risk.

Although cluster analysis and PCA are the most popular types of empirical dietary pattern analyses explored in the CHD literature, *reduced rank regression* (RRR) and *machine learning* techniques have also proved to be successful approaches in identifying diet patterns. RRR is similar to PCA, except the method selects dietary patterns that maximize the variability in a chosen set of response variables instead of maximizing variability in the food groups. These response variables can either be intermediate risk factors or nutrient variables. Because RRR involves both a priori decisions about intermediate variables and a data-driven component, RRR is considered a supervised data-driven or hybrid method to derive dietary patterns. The EDIP, which was discussed in the **Dietary inflammatory potential** section, was generated using RRR to identify foods that were most predictive of various inflammatory biomarkers. Likewise, empirical dietary indices for hyperinsulinemia (EDIH) and insulin resistance (EDIR) represent indices that are most predictive of C-peptide concentrations and TG:HDL-C ratio, respectively. The EDIH and EDIR separately developed in two studies included similar food groups. Diets with higher insulinemic potential included higher amounts of red/processed meat, butter, refined grains, fried foods, and potatoes, along with lower amounts of wine, coffee, whole fruit, and leafy green vegetables. Among 40,074 participants with 4904 deaths over a median of 7.8 years in the US National Health and Nutrition Examination Survey (NHANES), the EDIH (Q4 vs. Q1 HR [95% CI]: 1.41 [1.15, 1.75]) and EDIR (Q4 vs. Q1 HR [95% CI]: 1.35 [1.09, 1.67]) were both associated with higher risk of CVD mortality.

In addition to intermediate biomarkers of insulin resistance and inflammation, other studies have utilized RRR to identify dietary patterns predictive of lipid biomarkers that associate with CHD risk. Foods predictive of lipid concentrations in the Whitehall II study included white bread, fried potatoes, sugary drinks, burgers, and sausages, and low consumption of French dressing and vegetables. Not surprisingly, this pattern was associated with a higher CHD risk (Q4 vs. Q1 HR [95% CI]: 1.57 [1.08, 2.27]). Given that certain fatty acids can be cardioprotective, a study of individuals from Singapore identified a dietary pattern that associated with 9 different circulating fatty acids (n–3 polyunsaturated fatty acids [18:3n–3, 20:3n–3, 20:5n–3], odd-chain fatty acids [15:0, 17:0], 18:2n–6 and 20:1, and lower 20:4n–6 and 16:1). This pattern was characterized by high intakes of soy, vegetables, fruits, tea, tomato products, bread, fish, margarine, and dairy, and low intakes of rice, red meat, coffee, alcohol, sugar-sweetened beverages, and eggs and was associated with a 24% lower risk of CHD-related morality.

Random forest methods are an example of a supervised machine learning approach to dietary pattern derivation. Using random forest with tree classification (RF-TCA) analysis, researchers identified 7 dietary patterns in the European Prospective Investigation into Cancer (EPIC)–Netherlands cohort that could be classified as "Western-like", "prudent-like," or "traditional-like." They observed a higher risk of CHD among individuals following a "traditional-like" pattern compared to those following a "prudent-like" pattern (HR [95% CI]: 1.36 [1.12, 1.65]). In the Framingham Heart Study, random forest model selection identified 7 dietary components (whole milk, red meat, eggs, alcohol, cheese, coffee, and decaffeinated coffee) that ranked in the top 20% of important features for prediction of either CHD, heart failure, or stroke out of 204 potential diet and lifestyle variables. Random forest methods and other machine learning models have proved useful in recent studies of dietary patterns and cardiometabolic risk factors, as well as other disease applications. Thus, further use of machine learning methods examining the influence of dietary patterns on CHD in large epidemiological studies is warranted.

Novel omics approaches for precision nutrition in CVD prevention

Utility of omics in precision nutrition

Dietary indices and empirical dietary patterns derived from self-reported intakes have allowed us to gain deep insights into how dietary patterns influence CHD risk. However, accurately assessing diet in free-living individuals remains among the most challenging aspects of nutrition research. Because diet is a complex pattern of inter-related exposures, both of known and unknown constituents, coupled with relatively large within-person day-to-day variability, accurately quantifying dietary exposures remains an ongoing challenge in nutrition research (Maruvada et al., 2020). Recent advances in omics profiling technologies have allowed for new opportunities for biomarker discovery, while also providing a more detailed phenotyping of individuals. Several omics measurements exist that may be altered by dietary intakes, including the gut microbiome, proteome, lipidome, metabolome, epigenome, and transcriptome. The genome also influences dietary preferences and changes how dietary intakes influence health. Therefore, a comprehensive understanding of "nutritional omics" will not only allow for identification of objective measures of diet, but will also uncover biological pathways through which diet can influence cardiometabolic disease risk. A better understanding of these mechanisms will provide a foundation for both personal decision-making and regulatory policies that influence dietary patterns, paving the path for "personalized nutrition". Although our knowledge of individual responses to diet is still very limited, personalized nutrition will encompass individualized dietary recommendations that are based on individual preferences and predicted responses to dietary intakes. The field of biomarker discovery of dietary intakes using systems epidemiology approaches is exploding given the 2020–2030 Strategic Plan for US National Institutes of Health Nutrition Research's focus on "Precision Nutrition". While a comprehensive overview of existing biomarkers of foods and dietary patterns is beyond the scope of this article, in this section, we discuss cutting-edge research that has utilized omic technologies to understand diet-disease mechanisms.

Omic signatures of dietary patterns and CHD

Several recent studies have integrated omic biomarkers of dietary pattern analyses into epidemiological studies of CVD and CVD risk factors. These include studies of the microbiome, epigenome, proteome and metabolome.

Given the strong evidence linking the Mediterranean diet and CHD, several recent studies have aimed to identify omic profiles of the Mediterranean diet and cardiometabolic disease risk. Given the large number of features present in omic data, machine learning techniques have proved useful to select potentially important features for discriminating between individuals with different dietary intakes. Data reduction using machine learning and related network analysis techniques can even allow us to make inferences from data that contains more features than participants. One study utilized one such machine learning technique, an elastic net regression model, to develop and validate a metabolite profile of 67 metabolites that reflected adherence to a Mediterranean diet in data from the PREDIMED study and two US cohorts. Several of the metabolites that were included in the Mediterranean diet "metabolomic signature" were previously associated with intakes of individual foods. These included long chain *n*-3 fatty acids that reflect fish intake and anti-oxidative polyphenols that may reflect extra-virgin olive oil or wine intake. The integrated 67 metabolite signature of the Mediterranean diet was significantly associated with lower CVD incidence in both the PREDIMED study (Q4 vs. Q1 HR [95% CI]: 0.73 [0.59, 0.91], n=1,012, 227 CVD cases) and the two US cohorts (Q4 vs. Q1 HR [95% CI]: 0.85 [0.77, 0.95], n = 6,868, 351 CVD cases), even after adjusting for self-reported measures of the MDS. This implies that a comprehensive metabolomic signature of a diet pattern can capture metabolic variation that cannot be captured by self-reported measures of diet and are, thereby, more powerful at detecting diet-disease associations. Another study conducted in the United Kingdom identified a metabolite score, which was derived from 66 metabolites that were significantly associated with the MDS, that also associated with several cardiometabolic risk factors among 10,806 adults without type 2 diabetes. Among the MDS food groups, they observed that fruits were associated with amino acids and biogenic amines, meat with acylcarnitines, and fish with phospholipids.

In addition to the Mediterranean diet, it will be useful to explore omic biomarkers of additional dietary patterns highlighted in the **Dietary patterns associated with coronary heart disease risk** section and their relation to CHD risk. Similar machine learning techniques referenced above have already been utilized to identify metabolomic signatures of dietary patterns and evaluate their associations with cardiometabolic risk factors. A recent study among >10,000 US men and women identified three distinct metabolomic signatures of the PDI, hPDI, and uPDI. The PDI (Q4 vs. Q1 HR [95% CI]: 0.86 [0.79, 0.93]) and hPDI (Q4 vs. Q1 HR [95% CI]: 0.79 [0.72, 0.86]) metabolomic signatures were both inversely associated with type 2 diabetes risk, while the uPDI was not associated with type 2 diabetes risk. Another study among individuals of Puerto Rican descent identified an AHA-DS-associated metabolomic signature of 58 metabolites, primarily lipids and amino acids. This signature was associated with higher HDL-C and LDL-C concentrations and lower waist circumference. The results from these studies strongly indicate the usefulness of combining various metabolites to comprise a composite marker of diet quality, and indicate that further investigation of metabolite profiles of diet quality and CHD risk will be fruitful to reveal mechanisms underlying diet-CHD relationships.

While individual omics profiling technologies are each important parts of systems biology data pipelines, the largest benefit for a systems view of chronic disease risk is achieved when combining data from multiple omics types. Multi-omics approaches

have been highlighted as promising approaches to improve our understanding of how diet contributes to complex chronic disease etiology and have proved successful for other chronic disease outcomes. However, few studies to date have leveraged multiple layers of omics to drive novel precision nutrition approaches. One study demonstrated that a Mediterranean diet pattern was significantly associated with changes in gut microbiome profiles, including changes in pathways related to plant-derived polysaccharide degradation, dietary fiber metabolism, short-chain fatty acid production, and secondary bile acid production among 307 male participants in the Health Professionals Follow-up Study. Integrating genetics and microbiome measures, the authors observed an association between the Mediterranean diet and cardiometabolic disease risk only among individuals without *Prevotella copri* present in their microbiome, providing an example where personalized nutrition advice could be useful. A recent cross-sectional study among US individuals of Hispanic/Latino descent integrated gut microbiome and serum metabolites to improve our understanding of the mechanisms by which serum trimethylamine-*N*-oxide (TMAO) associates with cardiometabolic risk. In this study, they demonstrated that serum TMAO concentrations may be a potential biomarker of fish, red meat, and egg intake and that higher serum TMAO is associated with changes in the microbiome and higher odds of prevalent CVD. Another group of researchers examined cross-sectional associations between metabolites and proteins and dietary pattern indices among 2208 participants from the Framingham Heart Study. Lipids, amino acids, bile acids and derivatives, nucleotide, metabolism, and tricarboxylic acid and derivatives were found to be shared among three dietary pattern indices measured: the AHEI, DASH and Mediterranean diet. As multi-omics data becomes available within more and more nutrition studies, investigation of multi-omics biomarkers of dietary intakes is a promising avenue for future research.

Conclusions

Poor diet quality across the world results in a substantial proportion of CHD events and deaths each year. Thus, public health strategies to improve diet quality should be prioritized to reduce the global burden of CHD. Dietary pattern research over the past few decades has identified many dietary patterns that associate with lower CHD risk, including Mediterranean, DASH, and healthy plant-based dietary patterns. Achieving these healthy dietary patterns will require significant shifts in individual choices and behaviors, which could be promoted through policy and infrastructure changes to create a food environment that encourages healthy dietary choices. Although advancements in nutrition science have led to some consensus on the components of a "healthy" dietary pattern, controversy still exists for some foods/nutrients and most studies of CHD endpoints rely on self-reported dietary intakes. Studies utilizing novel omics profiling techniques, particularly multi-omics approaches, present an opportunity to identify objective biomarkers of diet and unveil novel mechanisms by which diet contributes to CHD risk. An improved understanding of how to modulate biomarkers of diet and predict individual responses to diet will help us move toward personalized approaches to dietary recommendations for CHD prevention. The wealth of omics data currently being generated in nutrition studies of cardiometabolic risk suggests a bright future of discoveries that will advance precision nutrition in CHD prevention.

References

Appel, L.J., Moore, T.J., Obarzanek, E., Vollmer, W.M., Svetkey, L.P., Sacks, F.M., Bray, G.A., Vogt, T.M., Cutler, J.A., Windhauser, M.M., Lin, P.-H., Karanja, N., Simons-Morton, D., McCullough, M., Swain, J., Steele, P., Evans, M.A., Miller, E.R., Harsha, D.W., 1997. A clinical trial of the effects of dietary patterns on blood pressure. N. Engl. J. Med. 336 (16), 1117–1124. https://doi.org/10.1056/NEJM199704173361601. 10.1056/NEJM199704173361601.

Aslani, Z., Sadeghi, O., Heidari-Beni, M., Zahedi, H., Baygi, F., Shivappa, N., Hébert, J.R., Moradi, S., Sotoudeh, G., Asayesh, H., Djalalinia, S., Qorbani, M., 2020. Association of dietary inflammatory potential with cardiometabolic risk factors and diseases: a systematic review and dose–response meta-analysis of observational studies. Diabetol. Metab. Syndrome 12, 86. https://doi.org/10.1186/s13098-020-00592-6.

Becerra-Tomás, N., Blanco Mejía, S., Viguiliouk, E., Khan, T., Kendall, C.W.C., Kahleova, H., Rahelić, D., Sievenpiper, J.L., Salas-Salvadó, J., 2020. Mediterranean diet, cardiovascular disease and mortality in diabetes: a systematic review and meta-analysis of prospective cohort studies and randomized clinical trials. Crit. Rev. Food Sci. Nutr. 60, 1207–1227. https://doi.org/10.1080/10408398.2019.1565281.

Bechthold, A., Boeing, H., Schwedhelm, C., Hoffmann, G., Knüppel, S., Iqbal, K., Henauw, S.D., Michels, N., Devleesschauwer, B., Schlesinger, S., Schwingshackl, L., 2017. Food groups and risk of coronary heart disease, stroke and heart failure: a systematic review and dose-response meta-analysis of prospective studies. Crit. Rev. Food Sci. Nutr. 1–20. https://doi.org/10.1080/10408398.2017.1392288.

Bhupathiraju, S.N., Tucker, K.L., 2011. Coronary heart disease prevention: nutrients, foods, and dietary patterns. Clin. Chim. Acta 412, 1493–1514. https://doi.org/10.1016/j.cca.2011.04.038.

Chiavaroli, L., Viguiliouk, E., Nishi, S.K., Blanco Mejia, S., Rahelić, D., Kahleová, H., Salas-Salvadó, J., Kendall, C.W., Sievenpiper, J.L., 2019. DASH dietary pattern and cardiometabolic outcomes: an umbrella review of systematic reviews and meta-analyses. Nutrients 11, 338. https://doi.org/10.3390/nu11020338.

Dinu, M., Abbate, R., Gensini, G.F., Casini, A., Sofi, F., 2017. Vegetarian, vegan diets and multiple health outcomes: a systematic review with meta-analysis of observational studies. Crit. Rev. Food Sci. Nutr. 57, 3640–3649. https://doi.org/10.1080/10408398.2016.1138447.

Estruch, R., Ros, E., Salas-Salvadó, J., Covas, M.-I., Corella, D., Arós, F., Gómez-Gracia, E., Ruiz-Gutiérrez, V., Fiol, M., Lapetra, J., Lamuela-Raventos, R.M., Serra-Majem, L., Pintó, X., Basora, J., Muñoz, M.A., Sorlí, J.V., Martínez, J.A., Fitó, M., Gea, A., Hernán, M.A., Martínez-González, M.A., 2018. Primary prevention of cardiovascular disease with a Mediterranean diet supplemented with extra-virgin olive oil or nuts. N. Engl. J. Med. https://doi.org/10.1056/NEJMoa1800389.

Farazi, M., Jayedi, A., Shab-Bidar, S., 2021. Dietary inflammatory index and the risk of non-communicable chronic disease and mortality: an umbrella review of meta-analyses of observational studies. Crit. Rev. Food Sci. Nutr. 1–10. https://doi.org/10.1080/10408398.2021.1943646.

Gan, Z.H., Cheong, H.C., Tu, Y.-K., Kuo, P.-H., 2021. Association between plant-based dietary patterns and risk of cardiovascular disease: a systematic review and meta-analysis of prospective cohort studies. Nutrients 13, 3952. https://doi.org/10.3390/nu13113952.

Glenn, A.J., Viguiliouk, E., Seider, M., Boucher, B.A., Khan, T.A., Blanco Mejia, S., Jenkins, D.J.A., Kahleová, H., Rahelić, D., Salas-Salvadó, J., Kendall, C.W.C., Sievenpiper, J.L., 2019. Relation of vegetarian dietary patterns with major cardiovascular outcomes: a systematic review and meta-analysis of prospective cohort studies. Front. Nutr. 6, 80. https://doi.org/10.3389/fnut.2019.00080.

Grosso, G., Marventano, S., Yang, J., Micek, A., Pajak, A., Scalfi, L., Galvano, F., Kales, S.N., 2017. A comprehensive meta-analysis on evidence of Mediterranean diet and cardiovascular disease: are individual components equal? Crit. Rev. Food Sci. Nutr. 57, 3218–3232. https://doi.org/10.1080/10408398.2015.1107021.

Hou, L., Li, F., Wang, Y., Ou, Z., Xu, D., Tan, W., Dai, M., 2015. Association between dietary patterns and coronary heart disease: a meta-analysis of prospective cohort studies. Int. J. Clin. Exp. Med. 8, 781–790.

Huang, T., Yang, B., Zheng, J., Li, G., Wahlqvist, M.L., Li, D., 2012. Cardiovascular disease mortality and cancer incidence in vegetarians: a meta-analysis and systematic review. ANM 60, 233–240. https://doi.org/10.1159/000337301.

Jafari, S., Hezaveh, E., Jalilpiran, Y., Jayedi, A., Wong, A., Safaiyan, A., Barzegar, A., 2021. Plant-based diets and risk of disease mortality: a systematic review and meta-analysis of cohort studies. Crit. Rev. Food Sci. Nutr. 1–13. https://doi.org/10.1080/10408398.2021.1918628.

Jalilpiran, Y., Jayedi, A., Djafarian, K., Shab-Bidar, S., 2020. The Nordic diet and the risk of non-communicable chronic disease and mortality: a systematic review and dose-response meta-analysis of prospective cohort studies. Crit. Rev. Food Sci. Nutr. 1–13. https://doi.org/10.1080/10408398.2020.1863906.

Jayedi, A., Soltani, S., Abdolshahi, A., Shab-Bidar, S., 2020. Healthy and unhealthy dietary patterns and the risk of chronic disease: an umbrella review of meta-analyses of prospective cohort studies. Br. J. Nutr. 124, 1133–1144. https://doi.org/10.1017/S0007114520002330.

Ji, M., Hong, X., Chen, M., Chen, T., Wang, J., Zhang, N., 2020. Dietary inflammatory index and cardiovascular risk and mortality: a meta-analysis of cohort studies. Medicine 99, e20303. https://doi.org/10.1097/MD.0000000000020303.

Kahleova, H., Salas-Salvadó, J., Rahelić, D., Kendall, C.W., Rembert, E., Sievenpiper, J.L., 2019. Dietary patterns and cardiometabolic outcomes in diabetes: a summary of systematic reviews and meta-analyses. Nutrients 11, 2209. https://doi.org/10.3390/nu11092209.

Kaiser, J., van Daalen, K.R., Thayyil, A., Cocco, M.T. de A.R.R., Caputo, D., Oliver-Williams, C., 2021. A systematic review of the association between vegan diets and risk of cardiovascular disease. J. Nutr. 151, 1539–1552. https://doi.org/10.1093/jn/nxab037.

Kwok, C.S., Umar, S., Myint, P.K., Mamas, M.A., Loke, Y.K., 2014. Vegetarian diet, Seventh Day Adventists and risk of cardiovascular mortality: a systematic review and meta-analysis. Int. J. Cardiol. 176, 680–686. https://doi.org/10.1016/j.ijcard.2014.07.080.

Li, F., Hou, L., Chen, W., Chen, P., Lei, C., Wei, Q., Tan, W., Zheng, S., 2015. Associations of dietary patterns with the risk of all-cause, CVD and stroke mortality: a meta-analysis of prospective cohort studies. Br. J. Nutr. 113, 16–24. https://doi.org/10.1017/S000711451400289X.

Liu, F.-H., Liu, C., Gong, T.-T., Gao, S., Sun, H., Jiang, Y.-T., Zhang, J.-Y., Zhang, M., Gao, C., Li, X.-Y., Zhao, Y.-H., Wu, Q.-J., 2021. Dietary inflammatory index and health outcomes: an umbrella review of systematic review and meta-analyses of observational studies. Front. Nutr. 8, 190. https://doi.org/10.3389/fnut.2021.647122.

Liyanage, T., Ninomiya, T., Wang, A., Neal, B., Jun, M., Wong, M.G., Jardine, M., Hillis, G.S., Perkovic, V., 2016. Effects of the Mediterranean diet on cardiovascular outcomes—a systematic review and meta-analysis. PLoS One 11, e0159252. https://doi.org/10.1371/journal.pone.0159252.

Maruvada, P., Lampe, J.W., Wishart, D.S., Barupal, D., Chester, D.N., Dodd, D., Djoumbou-Feunang, Y., Dorrestein, P.C., Dragsted, L.O., Draper, J., Duffy, L.C., Dwyer, J.T., Emenaker, N.J., Fiehn, O., Gerszten, R.E., B Hu, F., Karp, R.W., Klurfeld, D.M., Laughlin, M.R., Little, A.R., Lynch, C.J., Moore, S.C., Nicastro, H.L., O'Brien, D.M., Ordovás, J.M., Osganian, S.K., Playdon, M., Prentice, R., Raftery, D., Reisdorph, N., Roche, H.M., Ross, S.A., Sang, S., Scalbert, A., Srinivas, P.R., Zeisel, S.H., 2020. Perspective: dietary biomarkers of intake and exposure—exploration with omics approaches. Adv. Nutr. https://doi.org/10.1093/advances/nmz075.

Marx, W., Veronese, N., Kelly, J.T., Smith, L., Hockey, M., Collins, S., Trakman, G.L., Hoare, E., Teasdale, S.B., Wade, A., Lane, M., Aslam, H., Davis, J.A., O'Neil, A., Shivappa, N., Hebert, J.R., Blekkenhorst, L.C., Berk, M., Segasby, T., Jacka, F., 2021. The dietary inflammatory index and human health: an umbrella review of meta-analyses of observational studies. Adv. Nutr. 12, 1681–1690. https://doi.org/10.1093/advances/nmab037.

Morze, J., Danielewicz, A., Hoffmann, G., Schwingshackl, L., 2020. Diet quality as assessed by the healthy eating index, alternate healthy eating index, dietary approaches to stop hypertension score, and health outcomes: a second update of a systematic review and meta-analysis of cohort studies. J. Acad. Nutr. Diet. 120, 1998–2031.e15. https://doi.org/10.1016/j.jand.2020.08.076.

Namazi, N., Larijani, B., Azadbakht, L., 2018. Dietary inflammatory index and its association with the risk of cardiovascular diseases, metabolic syndrome, and mortality: a systematic review and meta-analysis. Horm. Metab. Res. 50, 345–358. https://doi.org/10.1055/a-0596-8204.

Noto, H., Goto, A., Tsujimoto, T., Noda, M., 2013. Low-carbohydrate diets and all-cause mortality: a systematic review and meta-analysis of observational studies. PLoS One 8, e55030. https://doi.org/10.1371/journal.pone.0055030.

Quek, J., Lim, G., Lim, W.H., Ng, C.H., So, W.Z., Toh, J., Pan, X.H., Chin, Y.H., Muthiah, M.D., Chan, S.P., Foo, R.S.Y., Yip, J., Neelakantan, N., Chong, M.F.F., Loh, P.H., Chew, N.W.S., 2021. The association of plant-based diet with cardiovascular disease and mortality: a meta-analysis and systematic review of prospect cohort studies. Front. Cardiovasc. Med. 8, 756810. https://doi.org/10.3389/fcvm.2021.756810.

Rees, K., Al-Khudairy, L., Takeda, A., Stranges, S., 2021. Vegan dietary pattern for the primary and secondary prevention of cardiovascular diseases. Cochrane Database Syst. Rev. https://doi.org/10.1002/14651858.CD013501.pub2.

Rodríguez-Monforte, M., Flores-Mateo, G., Sánchez, E., 2015. Dietary patterns and CVD: a systematic review and meta-analysis of observational studies. Br. J. Nutr. 114, 1341–1359. https://doi.org/10.1017/S0007114515003177.

Rosato, V., Temple, N.J., La Vecchia, C., Castellan, G., Tavani, A., Guercio, V., 2019. Mediterranean diet and cardiovascular disease: a systematic review and meta-analysis of observational studies. Eur. J. Nutr. 58, 173–191. https://doi.org/10.1007/s00394-017-1582-0.

Sacks, F.M., Obarzanek, E., Windhauser, M.M., Svetkey, L.P., Vollmer, W.M., McCullough, M., Karanja, N., Lin, P.-H., Steele, P., Proschan, M.A., Evans, M.A., Appel, L.J., Bray, G.A., Vogt, T.M., Moore, T.J., DASH Investigators, 1995. Rationale and design of the Dietary Approaches to Stop Hypertension trial (DASH): a multicenter controlled-feeding study of dietary patterns to lower blood pressure. Ann. Epidemiol. 5, 108–118. https://doi.org/10.1016/1047-2797(94)00055-X.

Salehi-Abargouei, A., Maghsoudi, Z., Shirani, F., Azadbakht, L., 2013. Effects of Dietary Approaches to Stop Hypertension (DASH)-style diet on fatal or nonfatal cardiovascular diseases—incidence: a systematic review and meta-analysis on observational prospective studies. Nutrition 29, 611–618. https://doi.org/10.1016/j.nut.2012.12.018.

Satija, A., Bhupathiraju, S.N., Spiegelman, D., Chiuve, S.E., Manson, J.E., Willett, W., Rexrode, K.M., Rimm, E.B., Hu, F.B., 2017. Healthful and unhealthful plant-based diets and the risk of coronary heart disease in U.S. adults. J. Am. Coll. Cardiol. 70, 411–422. https://doi.org/10.1016/j.jacc.2017.05.047.

Schwingshackl, L., Hoffmann, G., 2015. Diet quality as assessed by the healthy eating index, the alternate healthy eating index, the dietary approaches to stop hypertension score, and health outcomes: a systematic review and meta-analysis of cohort studies. J. Acad. Nutr. Diet. 115, 780–800.e5. https://doi.org/10.1016/j.jand.2014.12.009.

Schwingshackl, L., Bogensberger, B., Hoffmann, G., 2018. Diet quality as assessed by the healthy eating index, alternate healthy eating index, dietary approaches to stop hypertension score, and health outcomes: an updated systematic review and meta-analysis of cohort studies. J. Acad. Nutr. Diet. 118, 74–100.e11. https://doi.org/10.1016/j.jand.2017.08.024.

Shivappa, N., Hebert, J.R., Kivimaki, M., Akbaraly, T., 2017. Alternative healthy eating index 2010, dietary inflammatory index and risk of mortality: results from the Whitehall II cohort study and meta-analysis of previous dietary inflammatory index and mortality studies. Br. J. Nutr. 118, 210–221. https://doi.org/10.1017/S0007114517001908.

Shivappa, N., Godos, J., Hébert, J.R., Wirth, M.D., Piuri, G., Speciani, A.F., Grosso, G., 2018. Dietary inflammatory index and cardiovascular risk and mortality-a meta-analysis. Nutrients 10, E200. https://doi.org/10.3390/nu10020200.

Sofi, F., Cesari, F., Abbate, R., Gensini, G.F., Casini, A., 2008. Adherence to Mediterranean diet and health status: meta-analysis. BMJ 337, a1344. https://doi.org/10.1136/bmj.a1344.

Sofi, F., Abbate, R., Gensini, G.F., Casini, A., 2010. Accruing evidence on benefits of adherence to the Mediterranean diet on health: an updated systematic review and meta-analysis. Am. J. Clin. Nutr. 92, 1189–1196. https://doi.org/10.3945/ajcn.2010.29673.
Sofi, F., Macchi, C., Abbate, R., Gensini, G.F., Casini, A., 2014. Mediterranean diet and health status: an updated meta-analysis and a proposal for a literature-based adherence score. Publ. Health Nutr. 17, 2769–2782. https://doi.org/10.1017/S1368980013003169.
Soltani, S., Arablou, T., Jayedi, A., Salehi-Abargouei, A., 2020. Adherence to the Dietary Approaches to Stop Hypertension (DASH) diet in relation to all-cause and cause-specific mortality: a systematic review and dose-response meta-analysis of prospective cohort studies. Nutr. J. 19, 37. https://doi.org/10.1186/s12937-020-00554-8.
Tang, C., Wang, X., Qin, L.-Q., Dong, J.-Y., 2021. Mediterranean diet and mortality in people with cardiovascular disease: a meta-analysis of prospective cohort studies. Nutrients 13, 2623. https://doi.org/10.3390/nu13082623.
Yang, Z.-Q., Yang, Z., Duan, M.-L., 2019. Dietary approach to stop hypertension diet and risk of coronary artery disease: a meta-analysis of prospective cohort studies. Int. J. Food Sci. Nutr. 70, 668–674. https://doi.org/10.1080/09637486.2019.1570490.
Zhang, X.-Y., Shu, L., Si, C.-J., Yu, X.-L., Liao, D., Gao, W., Zhang, L., Zheng, P.-F., 2015. Dietary patterns, alcohol consumption and risk of coronary heart disease in adults: a meta-analysis. Nutrients 7, 6582–6605. https://doi.org/10.3390/nu7085300.
Zhong, X., Guo, L., Zhang, L., Li, Y., He, R., Cheng, G., 2017. Inflammatory potential of diet and risk of cardiovascular disease or mortality: a meta-analysis. Sci. Rep. 7, 6367. https://doi.org/10.1038/s41598-017-06455-x.

Further reading

2010 Dietary Guidelines, [WWW Document], 2010. URL https://health.gov/our-work/food-nutrition/previous-dietary-guidelines/2010 (Accessed 8.3.2021).
Alshehry, Z.H., Mundra, P.A., Barlow, C.K., Mellett, N.A., Wong, G., McConville, M.J., Simes, J., Tonkin, A.M., Sullivan, D.R., Barnes, E.H., Nestel, P.J., Kingwell, B.A., Marre, M., Neal, B., Poulter, N.R., Rodgers, A., Williams, B., Zoungas, S., Hillis, G.S., Chalmers, J., Woodward, M., Meikle, P.J., 2016. Plasma lipidomic profiles improve on traditional risk factors for the prediction of cardiovascular events in type 2 diabetes mellitus. Circulation 134, 1637–1650. https://doi.org/10.1161/CIRCULATIONAHA.116.023233.
Angulo, E., Stern, D., Castellanos-Gutiérrez, A., Monge, A., Lajous, M., Bromage, S., Fung, T.T., Li, Y., Bhupathiraju, S.N., Deitchler, M., Willett, W.C., Batis, C., 2021. Changes in the global diet quality score, weight, and waist circumference in Mexican women. J. Nutr. 151, 152S–161S. https://doi.org/10.1093/jn/nxab171.
Appel, L.J., Brands, M.W., Daniels, S.R., Karanja, N., Elmer, P.J., Sacks, F.M., 2006. Dietary approaches to prevent and treat hypertension. Hypertension 47, 296–308. https://doi.org/10.1161/01.HYP.0000202568.01167.B6.
Aravanis, C., Corcondilas, A., Dontas, S.A., Lekos, D., Keys, A., 1970. IX. The Greek Islands of Crete and Corfu. Circulation 41, I–88. https://doi.org/10.1161/01.CIR.41.4S1.I-88.
Aune, D., Giovannucci, E., Boffetta, P., Fadnes, L.T., Keum, N., Norat, T., Greenwood, D.C., Riboli, E., Vatten, L.J., Tonstad, S., 2017. Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality-a systematic review and dose-response meta-analysis of prospective studies. Int. J. Epidemiol. 46, 1029–1056. https://doi.org/10.1093/ije/dyw319.
Azadbakht, L., Fard, N.R.P., Karimi, M., Baghaei, M.H., Surkan, P.J., Rahimi, M., Esmaillzadeh, A., Willett, W.C., 2011. Effects of the Dietary Approaches to Stop Hypertension (DASH) eating plan on cardiovascular risks among type 2 diabetic patients: a randomized crossover clinical trial. Diabetes Care 34, 55–57. https://doi.org/10.2337/dc10-0676.
Bach, A., Serra-Majem, L., Carrasco, J.L., Roman, B., Ngo, J., Bertomeu, I., Obrador, B., 2006. The use of indexes evaluating the adherence to the Mediterranean diet in epidemiological studies: a review. Publ. Health Nutr. 9, 132–146. https://doi.org/10.1079/PHN2005936.
Baden, M.Y., Liu, G., Satija, A., Li, Y., Sun, Q., Fung, T.T., Rimm, E.B., Willett, W.C., Hu, F.B., Bhupathiraju, S.N., 2019. Changes in plant-based diet quality and total and cause-specific mortality. Circulation 140, 979–991. https://doi.org/10.1161/CIRCULATIONAHA.119.041014.
Barzi, F., Woodward, M., Marfisi, R.M., Tavazzi, L., Valagussa, F., Marchioli, R., 2003. Mediterranean diet and all-causes mortality after myocardial infarction: results from the GISSI-Prevenzione trial. Eur. J. Clin. Nutr. 57, 604–611. https://doi.org/10.1038/sj.ejcn.1601575.
Becerra-Tomás, N., Paz-Graniel, I., Kendall, W.C.C., Kahleova, H., Rahelić, D., Sievenpiper, J.L., Salas-Salvadó, J., 2019. Nut consumption and incidence of cardiovascular diseases and cardiovascular disease mortality: a meta-analysis of prospective cohort studies. Nutr. Rev. 77, 691–709. https://doi.org/10.1093/nutrit/nuz042.
Beeson, W.L., Mills, P.K., Phillips, R.L., Andress, M., Fraser, G.E., 1989. Chronic disease among Seventh-day Adventists, a low-risk group. Rationale, methodology, and description of the population. Cancer 64, 570–581. https://doi.org/10.1002/1097-0142(19890801)64:3<570::aid-cncr2820640303>3.0.co;2-4.
Benisi-Kohansal, S., Saneei, P., Salehi-Marzijarani, M., Larijani, B., Esmaillzadeh, A., 2016. Whole-grain intake and mortality from all causes, cardiovascular disease, and cancer: a systematic review and dose-response meta-analysis of prospective cohort studies. Adv. Nutr. 7, 1052–1065. https://doi.org/10.3945/an.115.011635.
Bhupathiraju, S.N., Lichtenstein, A.H., Dawson-Hughes, B., Hannan, M.T., Tucker, K.L., 2013. Adherence to the 2006 American Heart Association Diet and Lifestyle Recommendations for cardiovascular disease risk reduction is associated with bone health in older Puerto Ricans. Am. J. Clin. Nutr. 98, 1309–1316. https://doi.org/10.3945/ajcn.112.056267.
Biesbroek, S., van der A, D.L., Brosens, M.C., Beulens, J.W., Verschuren, W.M., van der Schouw, Y.T., Boer, J.M., 2015. Identifying cardiovascular risk factor–related dietary patterns with reduced rank regression and random forest in the EPIC-NL cohort. Am. J. Clin. Nutr. 102, 146–154. https://doi.org/10.3945/ajcn.114.092288.
Birk, N., Matsuzaki, M., Fung, T.T., Li, Y., Batis, C., Stampfer, M.J., Deitchler, M., Willett, W.C., Fawzi, W.W., Bromage, S., Kinra, S., Bhupathiraju, S.N., Lake, E., 2021. Exploration of machine learning and statistical techniques in development of a low-cost screening method featuring the global diet quality score for detecting prediabetes in rural India. J. Nutr. 151, 110S–118S. https://doi.org/10.1093/jn/nxab281.
Blackburn, H., Taylor, H.L., Keys, A., 1970. Xvi. The electrocardiogram in prediction of five-year coronary heart disease incidence among men aged forty through fifty-nine. Circulation 41, I154. https://doi.org/10.1161/01.CIR.41.4S1.I-154.
Blackstone, N.T., Conrad, Z., 2020. Comparing the recommended eating patterns of the EAT-lancet commission and dietary guidelines for Americans: implications for sustainable nutrition. Curr. Dev. Nutr. 4, nzaa015. https://doi.org/10.1093/cdn/nzaa015.
Brennan, L., Hu, F.B., Sun, Q., 2021. Metabolomics meets nutritional epidemiology: harnessing the potential in metabolomics data. Metabolites 11, 709. https://doi.org/10.3390/metabo11100709.
Bromage, S., Batis, C., Bhupathiraju, S.N., Fawzi, W.W., Fung, T.T., Li, Y., Deitchler, M., Angulo, E., Birk, N., Castellanos-Gutiérrez, A., He, Y., Fang, Y., Matsuzaki, M., Zhang, Y., Moursi, M., Gicevic, S., Holmes, M.D., Isanaka, S., Kinra, S., Sachs, S.E., Stampfer, M.J., Stern, D., Willett, W.C., 2021. Development and validation of a novel food-based Global Diet Quality Score (GDQS). J. Nutr. 151, 75S–92S. https://doi.org/10.1093/jn/nxab244.
Buzina, R., Keys, A., Mohacek, I., Marinkovic, M., Hahn, A., Blackburn, H., 1970. Coronary heart disease in seven countries. V. Five-year follow-up in Dalmatia and Slavonia. Circulation 41, I40–I51. https://doi.org/10.1161/01.cir.41.4s1.i-40.
Carson, J.A.S., Lichtenstein, A.H., Anderson, C.A.M., Appel, L.J., Kris-Etherton, P.M., Meyer, K.A., Petersen, K., Polonsky, T., Van Horn, L., et al., 2020. Dietary cholesterol and cardiovascular risk: a science advisory from the American Heart Association. Circulation 141, e39–e53. https://doi.org/10.1161/CIR.0000000000000743.
Cavicchia, P.P., Steck, S.E., Hurley, T.G., Hussey, J.R., Ma, Y., Ockene, I.S., Hébert, J.R., 2009. A new dietary inflammatory index predicts interval changes in serum high-sensitivity C-reactive protein. J. Nutr. 139, 2365–2372. https://doi.org/10.3945/jn.109.114025.
Cazzola, R., Cestaro, B., 2011. Red wine polyphenols protect n−3 more than n−6 polyunsaturated fatty acid from lipid peroxidation. Food Res. Int. 44, 3065–3071. https://doi.org/10.1016/j.foodres.2011.07.029.
Chen, S.T., Maruthur, N.M., Appel, L.J., 2010. The effect of dietary patterns on estimated coronary heart disease risk. Circulation 3, 484–489. https://doi.org/10.1161/CIRCOUTCOMES.109.930685.

Clarys, P., Deliens, T., Huybrechts, I., Deriemaeker, P., Vanaelst, B., De Keyzer, W., Hebbelinck, M., Mullie, P., 2014. Comparison of nutritional quality of the vegan, vegetarian, semi-vegetarian, pesco-vegetarian and omnivorous diet. Nutrients 6, 1318–1332. https://doi.org/10.3390/nu6031318.

Dahl, W.J., Stewart, M.L., 2015. Position of the academy of nutrition and dietetics: health implications of dietary fiber. J. Acad. Nutr. Diet. 115, 1861–1870. https://doi.org/10.1016/j.jand.2015.09.003.

Davis, C.R., Hodgson, J.M., Woodman, R., Bryan, J., Wilson, C., Murphy, K.J., 2017. A Mediterranean diet lowers blood pressure and improves endothelial function: results from the MedLey randomized intervention trial. Am. J. Clin. Nutr. 105, 1305–1313. https://doi.org/10.3945/ajcn.116.146803.

de Koning, L., Malik, V.S., Kellogg, M.D., Rimm, E.B., Willett, W.C., Hu, F.B., 2012. Sweetened beverage consumption, incident coronary heart disease, and biomarkers of risk in men. Circulation 125, 1735–1741. https://doi.org/10.1161/CIRCULATIONAHA.111.067017. S1.

de Lorgeril, M., Salen, P., Martin, J.-L., Monjaud, I., Delaye, J., Mamelle, N., 1999. Mediterranean diet, traditional risk factors, and the rate of cardiovascular complications after myocardial infarction. Circulation 99, 779–785. https://doi.org/10.1161/01.CIR.99.6.779.

Del Gobbo, L.C., Imamura, F., Aslibekyan, S., Marklund, M., Virtanen, J.K., Wennberg, M., Yakoob, M.Y., Chiuve, S.E., dela Cruz, L., Frazier-Wood, A.C., Fretts, A.M., Guallar, E., Matsumoto, C., Prem, K., Tanaka, T., Wu, J.H.Y., Zhou, X., Helmer, C., Ingelsson, E., Yuan, J.-M., Barberger-Gateau, P., Campos, H., Chaves, P.H.M., Djoussé, L., Giles, G.G., Gómez-Aracena, J., Hodge, A.M., Hu, F.B., Jansson, J.-H., Johansson, I., Khaw, K.-T., Koh, W.-P., Lemaitre, R.N., Lind, L., Luben, R.N., Rimm, E.B., Risérus, U., Samieri, C., Franks, P.W., Siscovick, D.S., Stampfer, M., Steffen, L.M., Steffen, B.T., Tsai, M.Y., van Dam, R.M., Voutilainen, S., Willett, W.C., Woodward, M., Mozaffarian, D., For the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Fatty Acids and Outcomes Research Consortium (FORCe), 2016. ω-3 polyunsaturated fatty acid biomarkers and coronary heart disease: pooling project of 19 cohort studies. JAMA Intern. Med. 176, 1155–1166. https://doi.org/10.1001/jamainternmed.2016.2925.

Díaz-Uriarte, R., Alvarez de Andrés, S., 2006. Gene selection and classification of microarray data using random forest. BMC Bioinf. 7, 3. https://doi.org/10.1186/1471-2105-7-3.

Doran, S., Arif, M., Lam, S., Bayraktar, A., Turkez, H., Uhlen, M., Boren, J., Mardinoglu, A., 2021. Multi-omics approaches for revealing the complexity of cardiovascular disease. Briefings Bioinf. https://doi.org/10.1093/bib/bbab061.

Esposito, K., Marfella, R., Ciotola, M., Di Palo, C., Giugliano, F., Giugliano, G., D'Armiento, M., D'Andrea, F., Giugliano, D., 2004. Effect of a Mediterranean-style diet on endothelial dysfunction and markers of vascular inflammation in the metabolic syndrome: a randomized trial. JAMA 292, 1440–1446. https://doi.org/10.1001/jama.292.12.1440.

Estruch, R., Martínez-González, M.A., Corella, D., Salas-Salvadó, J., Ruiz-Gutiérrez, V., Covas, M.I., Fiol, M., Gómez-Gracia, E., López-Sabater, M.C., Vinyoles, E., Arós, F., Conde, M., Lahoz, C., Lapetra, J., Sáez, G., Ros, E., PREDIMED Study Investigators, 2006. Effects of a Mediterranean-style diet on cardiovascular risk factors: a randomized trial. Ann. Intern. Med. 145, 1–11. https://doi.org/10.7326/0003-4819-145-1-200607040-00004.

Fidanza, F., Puddu, V., Imbimbo, A.B., Menotti, A., Keys, A., 1970. Coronary heart disease in seven countries. VII. Five-year experience in rural Italy. Circulation 41, I63–I75. https://doi.org/10.1161/01.cir.41.4s1.i-63.

Floegel, A., von Ruesten, A., Drogan, D., Schulze, M.B., Prehn, C., Adamski, J., Pischon, T., Boeing, H., 2013. Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam. Eur. J. Clin. Nutr. 67, 1100–1108. https://doi.org/10.1038/ejcn.2013.147.

Fung, T.T., Malik, V., Rexrode, K.M., Manson, J.E., Willett, W.C., Hu, F.B., 2009a. Sweetened beverage consumption and risk of coronary heart disease in women. Am. J. Clin. Nutr. 89, 1037–1042. https://doi.org/10.3945/ajcn.2008.27140.

Fung, T.T., Rexrode, K.M., Mantzoros, C.S., Manson, J.E., Willett, W.C., Hu, F.B., 2009b. Mediterranean diet and incidence and mortality of coronary heart disease and stroke in women. Circulation 119, 1093–1100. https://doi.org/10.1161/CIRCULATIONAHA.108.816736.

Fung, T.T., Isanaka, S., Hu, F.B., Willett, W.C., 2018. International food group–based diet quality and risk of coronary heart disease in men and women. Am. J. Clin. Nutr. 107, 120–129. https://doi.org/10.1093/ajcn/nqx015.

Fung, T.T., Li, Y., Bhupathiraju, S.N., Bromage, S., Batis, C., Holmes, M.D., Stampfer, M., Hu, F.B., Deitchler, M., Willett, W.C., 2021. Higher global diet quality score is inversely associated with risk of type 2 diabetes in US women. J. Nutr. 151, 168S–175S. https://doi.org/10.1093/jn/nxab195.

Gleason, P.M., Boushey, C.J., Harris, J.E., Zoellner, J., 2015. Publishing nutrition research: a review of multivariate techniques—Part 3: data reduction methods. J. Acad. Nutr. Diet. 115, 1072–1082. https://doi.org/10.1016/j.jand.2015.03.011.

Guallar-Castillón, P., Rodríguez-Artalejo, F., Tormo, M.J., Sánchez, M.J., Rodríguez, L., Quirós, J.R., Navarro, C., Molina, E., Martínez, C., Marín, P., Lopez-Garcia, E., Larrañaga, N., Huerta, J.M., Dorronsoro, M., Chirlaque, M.D., Buckland, G., Barricarte, A., Banegas, J.R., Arriola, L., Ardanaz, E., González, C.A., Moreno-Iribas, C., 2012. Major dietary patterns and risk of coronary heart disease in middle-aged persons from a Mediterranean country: the EPIC-Spain cohort study. Nutr. Metabol. Cardiovasc. Dis. 22, 192–199. https://doi.org/10.1016/j.numecd.2010.06.004.

Guasch-Ferré, M., Bhupathiraju, S.N., Hu, F.B., 2018. Use of metabolomics in improving assessment of dietary intake. Clin. Chem. 64, 82–98. https://doi.org/10.1373/clinchem.2017.272344.

Hansson, G.K., 2005. Inflammation, atherosclerosis, and coronary artery disease. N. Engl. J. Med. 352, 1685–1695. https://doi.org/10.1056/NEJMra043430.

Hardinge, M.G., Stare, F.J., 1954a. Nutritional studies of vegetarians. I. Nutritional, physical, and laboratory studies. J. Clin. Nutr. 2, 73–82. https://doi.org/10.1093/ajcn/2.2.73.

Hardinge, M.G., Stare, F.J., 1954b. Nutritional studies of vegetarians. II. Dietary and serum levels of cholesterol. J. Clin. Nutr. 2, 83–88.

Hardinge, M.G., Chambers, A.C., Crooks, H., Stare, F.J., 1958. Nutritional studies of vegetarians. III. Dietary levels of fiber. Am. J. Clin. Nutr. 6, 523–525. https://doi.org/10.1093/ajcn/6.5.523.

Hardinge, M.G., Crooks, H., Stare, F.J., 1962. Nutritional studies of vegetarians: IV. Dietary fatty acids and serum cholesterol levels. Am. J. Clin. Nutr. 10, 516–524. https://doi.org/10.1093/ajcn/10.6.516.

Hardinge, M.G., Crooks, H., Stare, F.J., 1966. Nutritional studies of vegetarians. J. Am. Diet Assoc. 48, 25–28.

Harris, W.S., Tintle, N.L., Imamura, F., Qian, F., Korat, A.V.A., Marklund, M., Djoussé, L., Bassett, J.K., Carmichael, P.-H., Chen, Y.-Y., Hirakawa, Y., Küpers, L.K., Laguzzi, F., Lankinen, M., Murphy, R.A., Samieri, C., Senn, M.K., Shi, P., Virtanen, J.K., Brouwer, I.A., Chien, K.-L., Eiriksdottir, G., Forouhi, N.G., Geleijnse, J.M., Giles, G.G., Gudnason, V., Helmer, C., Hodge, A., Jackson, R., Khaw, K.-T., Laakso, M., Lai, H., Laurin, D., Leander, K., Lindsay, J., Micha, R., Mursu, J., Ninomiya, T., Post, W., Psaty, B.M., Risérus, U., Robinson, J.G., Shadyab, A.H., Snetselaar, L., Sala-Vila, A., Sun, Y., Steffen, L.M., Tsai, M.Y., Wareham, N.J., Wood, A.C., Wu, J.H.Y., Hu, F., Sun, Q., Siscovick, D.S., Lemaitre, R.N., Mozaffarian, D., 2021. Blood n-3 fatty acid levels and total and cause-specific mortality from 17 prospective studies. Nat. Commun. 12, 2329. https://doi.org/10.1038/s41467-021-22370-2.

Hasin, Y., Seldin, M., Lusis, A., 2017. Multi-omics approaches to disease. Genome Biol. 18, 83. https://doi.org/10.1186/s13059-017-1215-1.

Haslam, D.E., Li, J., Liang, L., Clary, C., Lichtenstein, A.H., Lai, C.-Q., Ordovas, J.M., Manson, J.E., Wittenbecher, C., Hu, F.B., Qi, Q., Tucker, K.L., Bhupathiraju, S.N., 2020. Abstract 52: plasma metabolomic signatures of the American Heart Association diet score: findings from the Boston Puerto Rican Health Study. Circulation 141, A52. https://doi.org/10.1161/circ.141.suppl_1.52.

He, Y., Fang, Y., Bromage, S., Fung, T.T., Bhupathiraju, S.N., Batis, C., Deitchler, M., Fawzi, W., Stampfer, M.J., Hu, F.B., Willett, W.C., Li, Y., 2021. Application of the global diet quality score in Chinese adults to evaluate the double burden of nutrient inadequacy and metabolic syndrome. J. Nutr. 151, 93S–100S. https://doi.org/10.1093/jn/nxab162.

Hirahatake, K.M., Jiang, L., Wong, N.D., Shikany, J.M., Eaton, C.B., Allison, M.A., Martin, L., Garcia, L., Zaslavsky, O., Odegaard, A.O., 2019. Diet quality and cardiovascular disease risk in postmenopausal women with type 2 diabetes mellitus: the Women's health initiative. J. Am. Heart Assoc. 8, e013249. https://doi.org/10.1161/JAHA.119.013249.

Hjermann, I., Holme, I., Velve Byre, K., Leren, P., 1981. Effect of diet and smoking intervention on the incidence of coronary heart disease: report from the Oslo Study Group of a randomised trial in healthy men. Lancet 318, 1303–1310. https://doi.org/10.1016/S0140-6736(81)91338-6.

Howard, B.V., Van Horn, L., Hsia, J., et al., 2006. Low-fat dietary pattern and risk of cardiovascular disease the Women's health initiative randomized controlled dietary modification trial. JAMA 295, 655–666. https://doi.org/10.1001/jama.295.6.655.

Hu, Y., Hu, F.B., Manson, J.E., 2019. Marine Omega-3 supplementation and cardiovascular disease: an updated meta-analysis of 13 randomized controlled trials involving 127,477 participants. J. Am. Heart Assoc. 8, e013543. https://doi.org/10.1161/JAHA.119.013543.

Jannasch, F., Kröger, J., Schulze, M.B., 2017. Dietary patterns and type 2 diabetes: a systematic literature review and meta-analysis of prospective studies. J. Nutr. 147, 1174–1182. https://doi.org/10.3945/jn.116.242552.
Jayedi, A., Shab-Bidar, S., 2020. Fish consumption and the risk of chronic disease: an umbrella review of meta-analyses of prospective cohort studies. Adv. Nutr. 11, 1123–1133. https://doi.org/10.1093/advances/nmaa029.
Jin, Q., Shi, N., Aroke, D., Lee, D.H., Joseph, J.J., Donneyong, M., Conwell, D.L., Hart, P.A., Zhang, X., Clinton, S.K., Cruz-Monserrate, Z., Brasky, T.M., Jackson, R., Tinker, L.F., Liu, S., Phillips, L.S., Shadyab, A.H., Nassir, R., Bao, W., Tabung, F.K., 2021. Insulinemic and inflammatory dietary patterns show enhanced predictive potential for type 2 diabetes risk in postmenopausal women. Diabetes Care 44, 707–714. https://doi.org/10.2337/dc20-2216.
Johnson, R.K., Appel, L.J., Brands, M., Howard, B.V., Lefevre, M., Lustig, R.H., Sacks, F., Steffen, L.M., Wylie-Rosett, J., 2009. Dietary sugars intake and cardiovascular health. Circulation 120, 1011–1020. https://doi.org/10.1161/CIRCULATIONAHA.109.192627.
Jousilahti, P., Laatikainen, T., Peltonen, M., Borodulin, K., Männistö, S., Jula, A., Salomaa, V., Harald, K., Puska, P., Vartiainen, E., 2016. Primary prevention and risk factor reduction in coronary heart disease mortality among working aged men and women in eastern Finland over 40 years: population based observational study. BMJ 352, i721. https://doi.org/10.1136/bmj.i721.
Kant, A.K., Schatzkin, A., Graubard, B.I., Schairer, C., 2000. A prospective study of diet quality and mortality in women. JAMA 283, 2109–2115. https://doi.org/10.1001/jama.283.16.2109.
Karageorgou, D., Magriplis, E., Bakogianni, I., Mitsopoulou, A.V., Dimakopoulos, I., Micha, R., Michas, G., Ntouroupi, T., Tsaniklidou, S.M., Argyri, K., Chourdakis, M., Panagiotakos, D.B., Zampelas, A., Fappa, E., Theodoraki, E.M., Trichia, I., Sialvera, T.E., Aggeliki Varytimiadi, A., Spyrelli, E., Koutelidakis, A., Karlis, G., Zacharia, S., Papageorgiou, A., Chrousos, G.P., Dedoussis, G., Dimitriadis, G., Manios, I., Roma, E., 2020. Dietary patterns and cardiovascular disease in Greek adults: the Hellenic National Nutrition and Health Survey (HNNHS). Nutr. Metabol. Cardiovasc. Dis. 30, 201–213. https://doi.org/10.1016/j.numecd.2019.09.024.
Kelly, R.S., Chawes, B.L., Blighe, K., Virkud, Y.V., Croteau-Chonka, D.C., McGeachie, M.J., Clish, C.B., Bullock, K., Celedón, J.C., Weiss, S.T., Lasky-Su, J.A., 2018. An integrative transcriptomic and metabolomic study of lung function in children with asthma. Chest 154, 335–348. https://doi.org/10.1016/j.chest.2018.05.038.
Kennedy, E.T., Ohls, J., Carlson, S., Fleming, K., 1995. The healthy eating index: design and applications. J. Am. Diet Assoc. 95, 1103–1108. https://doi.org/10.1016/S0002-8223(95)00300-2.
Keys, A., Anderson, J.T., Grande, F., 1957. Prediction of serum-cholesterol responses of man to changes in fats in the diet. Lancet 273, 959–966. https://doi.org/10.1016/s0140-6736(57)91998-0.
Keys, A., Mienotti, A., Karvonen, M.J., Aravanis, C., Blackburn, H., Buzina, R., Djordjevic, B.S., Dontas, A.S., Fidanza, F., Keys, M.H., Kromhout, D., Nedeljkovic, S., Punsar, S., Seccareccia, F., Toshima, H., 1986. The diet and 15-year death rate in the seven countries study. Am. J. Epidemiol. 124, 903–915. https://doi.org/10.1093/oxfordjournals.aje.a114480.
Kim, K., Hyeon, J., Lee, S.A., Kwon, S.O., Lee, H., Keum, N., Lee, J., Park, S.M., 2017. Role of total, red, processed, and white meat consumption in stroke incidence and mortality: a systematic review and meta-analysis of prospective cohort studies. J. Am. Heart Assoc. 6, e005983. https://doi.org/10.1161/JAHA.117.005983.
Kim, H., Caulfield, L.E., Garcia-Larsen, V., Steffen, L.M., Coresh, J., Rebholz, C.M., 2019. Plant-based diets are associated with a lower risk of incident cardiovascular disease, cardiovascular disease mortality, and all-cause mortality in a general population of middle-aged adults. J. Am. Heart Assoc. 8, e012865. https://doi.org/10.1161/JAHA.119.012865.
Kim, Y., Keogh, J.B., Clifton, P.M., 2019. Does nut consumption reduce mortality and/or risk of cardiometabolic disease? An updated review based on meta-analyses. Int. J. Environ. Res. Publ. Health 16, 4957. https://doi.org/10.3390/ijerph16244957.
Kim, H., Lichtenstein, A.H., Wong, K.E., Appel, L.J., Coresh, J., Rebholz, C.M., 2021. Urine metabolites associated with the Dietary Approaches to Stop Hypertension (DASH) diet: results from the DASH-sodium trial. Mol. Nutr. Food Res. 65, 2000695. https://doi.org/10.1002/mnfr.202000695.
Kim, Y., Lu, S., Ho, J.E., Hwang, S., Yao, C., Huan, T., Levy, D., Ma, J., 2021. Proteins as mediators of the association between diet quality and incident cardiovascular disease and all-cause mortality: the Framingham Heart Study. J. Am. Heart Assoc. 10, e021245. https://doi.org/10.1161/JAHA.121.021245.
Kim, Y.O., 2009. Dietary patterns associated with hypertension among Korean males. Nutr. Res. Pract. 3, 162–166. https://doi.org/10.4162/nrp.2009.3.2.162.
Kimura, N., Keys, A., 1970. Coronary heart disease in seven countries. X. Rural southern Japan. Circulation 41, I101–I112.
La Berge, A.F., 2008. How the ideology of low fat conquered America. J. Hist. Med. Allied Sci. 63, 139–177. https://doi.org/10.1093/jhmas/jrn001.
Lapuente, M., Estruch, R., Shahbaz, M., Casas, R., 2019. Relation of fruits and vegetables with major cardiometabolic risk factors, markers of oxidation, and inflammation. Nutrients 11, 2381. https://doi.org/10.3390/nu11102381.
Lee, D.H., Li, J., Li, Y., Liu, G., Wu, K., Bhupathiraju, S., Rimm, E.B., Rexrode, K.M., Manson, J.E., Willett, W.C., Hu, F.B., Tabung, F.K., Giovannucci, E.L., 2020. Dietary inflammatory and insulinemic potential and risk of type 2 diabetes: results from three prospective U.S. cohort studies. Diabetes Care 43, 2675–2683. https://doi.org/10.2337/dc20-0815.
Leren, P., 1968. The effect of plasma-cholesterol-lowering diet in male survivors of myocardial infarction. A controlled clinical trial. Bull. N. Y. Acad. Med. 44, 1012–1020.
Leren, P., 1970. The Oslo diet-heart study. Circulation 42, 935–942. https://doi.org/10.1161/01.CIR.42.5.935.
Li, Y., Hruby, A., Bernstein, A.M., Ley, S.H., Wang, D.D., Chiuve, S.E., Sampson, L., Rexrode, K.M., Rimm, E.B., Willett, W.C., Hu, F.B., 2015. Saturated fat as compared with unsaturated fats and sources of carbohydrates in relation to risk of coronary heart disease: a prospective cohort study. J. Am. Coll. Cardiol. 66, 1538–1548. https://doi.org/10.1016/j.jacc.2015.07.055.
Li, J., Guasch-Ferré, M., Chung, W., et al., 2020a. The Mediterranean diet, plasma metabolome, and cardiovascular disease risk. Eur. Heart J. https://doi.org/10.1093/eurheartj/ehaa209.
Li, J., Lee, D.H., Hu, J., Tabung, F.K., Li, Y., Bhupathiraju, S.N., Rimm, E.B., Rexrode, K.M., Manson, J.E., Willett, W.C., Giovannucci, E.L., Hu, F.B., 2020b. Dietary inflammatory potential and risk of cardiovascular disease among men and women in the US. J. Am. Coll. Cardiol. 76, 2181–2193. https://doi.org/10.1016/j.jacc.2020.09.535.
Libby, P., Ridker, P.M., Maseri, A., 2002. Inflammation and atherosclerosis. Circulation 105, 1135–1143. https://doi.org/10.1161/hc0902.104353.
Libby, P., 2006. Inflammation and cardiovascular disease mechanisms. Am. J. Clin. Nutr. 83, 456S–460S. https://doi.org/10.1093/ajcn/83.2.456S.
Lichtenstein, A.H., Appel, L.J., et al., 2006. Diet and lifestyle recommendations revision 2006. Circulation 114, 82–96. https://doi.org/10.1161/CIRCULATIONAHA.106.176158.
Lichtenstein, A.H., Appel, L.J., Vadiveloo, M., Hu, F.B., Kris-Etherton, P.M., Rebholz, C.M., Sacks, F.M., Thorndike, A.N., Van Horn, L., Wylie-Rosett, J., 2021. 2021 dietary guidance to improve cardiovascular health: a scientific statement from the American Heart Association. Circulation 144, e472–e487. https://doi.org/10.1161/CIR.0000000000001031.
Lichtenstein, A.H., 2019. Last nail in the Coffin for sugar-sweetened beverages. Circulation 139, 2126–2128. https://doi.org/10.1161/CIRCULATIONAHA.119.040245.
Lu, Y., Zou, L., Su, J., Tai, E.S., Whitton, C., Dam, R.M. van, Ong, C.N., 2017. Meat and seafood consumption in relation to plasma metabolic profiles in a Chinese population: a combined untargeted and targeted metabolomics study. Nutrients 9, E683. https://doi.org/10.3390/nu9070683.
Ma, J., Rebholz, C.M., Braun, K.V.E., et al., 2020. Whole blood DNA methylation signatures of diet are associated with cardiovascular disease risk factors and all-cause mortality. Circulation 13, e002766. https://doi.org/10.1161/CIRCGEN.119.002766.
Ma, Y., He, F.J., Sun, Q., Yuan, C., Kieneker, L.M., Curhan, G.C., MacGregor, G.A., Bakker, S.J.L., Campbell, N.R.C., Wang, M., Rimm, E.B., Manson, J.E., Willett, W.C., Hofman, A., Gansevoort, R.T., Cook, N.R., Hu, F.B., 2021. 24-hour urinary sodium and potassium excretion and cardiovascular risk. N. Engl. J. Med. https://doi.org/10.1056/NEJMoa2109794.
Malik, V.S., Popkin, B.M., Bray, G.A., Després, J.-P., Hu, F.B., 2010. Sugar-sweetened beverages, obesity, type 2 diabetes mellitus, and cardiovascular disease risk. Circulation 121, 1356–1364. https://doi.org/10.1161/CIRCULATIONAHA.109.876185.
Malik, V.S., Li, Y., Pan, A., De Koning, L., Schernhammer, E., Willett, W.C., Hu, F.B., 2019. Long-term consumption of sugar-sweetened and artificially sweetened beverages and risk of mortality in US adults. Circulation. https://doi.org/10.1161/CIRCULATIONAHA.118.037401.

Malik, V.S., 2017. Sugar sweetened beverages and cardiometabolic health. Curr. Opin. Cardiol. https://doi.org/10.1097/HCO.0000000000000439.
Martínez-González, M.A., Zazpe, I., Razquin, C., Sánchez-Tainta, A., Corella, D., Salas-Salvadó, J., Toledo, E., Ros, E., Muñoz, M.Á., Recondo, J., Gómez-Gracia, E., Fiol, M., Lapetra, J., Buil-Cosiales, P., Serra-Majem, L., Pinto, X., Schröder, H., Tur, J.A., Sorli, J.V., Lamuela-Raventós, R.M., Estruch, R., 2015. Empirically-derived food patterns and the risk of total mortality and cardiovascular events in the PREDIMED study. Clin. Nutr. 34, 859–867. https://doi.org/10.1016/j.clnu.2014.09.006.
Martínez-Ortiz, J.A., Fung, T.T., Baylin, A., Hu, F.B., Campos, H., 2006. Dietary patterns and risk of nonfatal acute myocardial infarction in Costa Rican adults. Eur. J. Clin. Nutr. 60, 770–777. https://doi.org/10.1038/sj.ejcn.1602381.
Mastej, E., Gillenwater, L., Zhuang, Y., Pratte, K.A., Bowler, R.P., Kechris, K., 2020. Identifying protein–metabolite networks associated with COPD phenotypes. Metabolites 10, 124. https://doi.org/10.3390/metabo10040124.
Mayr, M., Yusuf, S., Weir, G., Chung, Y.-L., Mayr, U., Yin, X., Ladroue, C., Madhu, B., Roberts, N., De Souza, A., Fredericks, S., Stubbs, M., Griffiths, J.R., Jahangiri, M., Xu, Q., Camm, A.J., 2008. Combined metabolomic and proteomic analysis of human atrial fibrillation. J. Am. Coll. Cardiol. 51, 585–594. https://doi.org/10.1016/j.jacc.2007.09.055.
McCullough, M.L., Feskanich, D., Stampfer, M.J., Giovannucci, E.L., Rimm, E.B., Hu, F.B., Spiegelman, D., Hunter, D.J., Colditz, G.A., Willett, W.C., 2002. Diet quality and major chronic disease risk in men and women: moving toward improved dietary guidance. Am. J. Clin. Nutr. 76, 1261–1271. https://doi.org/10.1093/ajcn/76.6.1261.
McNaughton, S.A., Mishra, G.D., Brunner, E.J., 2009. Food patterns associated with blood lipids are predictive of coronary heart disease: the Whitehall II study. Br. J. Nutr. 102, 619–624. https://doi.org/10.1017/S0007114509243030.
Mei, Z., Chen, G.-C., Wang, Z., Usyk, M., Yu, B., Baeza, Y.V., Humphrey, G., Benitez, R.S., Li, J., Williams-Nguyen, J.S., Daviglus, M.L., Hou, L., Cai, J., Zheng, Y., Knight, R., Burk, R.D., Boerwinkle, E., Kaplan, R.C., Qi, Q., 2021. Dietary factors, gut microbiota, and serum trimethylamine-N-oxide associated with cardiovascular disease in the Hispanic Community Health Study/Study of Latinos. Am. J. Clin. Nutr. 113, 1503–1514. https://doi.org/10.1093/ajcn/nqab001.
Mendoza-Vasconez, A.S., Landry, M.J., Crimarco, A., Bladier, C., Gardner, C.D., 2021. Sustainable diets for cardiovascular disease prevention and management. Curr. Atheroscler. Rep. 23, 31. https://doi.org/10.1007/s11883-021-00929-0.
Micha, R., Peñalvo, J.L., Cudhea, F., Imamura, F., Rehm, C.D., Mozaffarian, D., 2017. Association between dietary factors and mortality from heart disease, stroke, and type 2 diabetes in the United States. JAMA 317, 912–924. https://doi.org/10.1001/jama.2017.0947.
Miller, V., Mente, A., Dehghan, M., et al., 2017. Fruit, vegetable, and legume intake, and cardiovascular disease and deaths in 18 countries (PURE): a prospective cohort study. Lancet 390, 2037–2049. https://doi.org/10.1016/S0140-6736(17)32253-5.
Minihane, A.M., Vinoy, S., Russell, W.R., Baka, A., Roche, H.M., Tuohy, K.M., Teeling, J.L., Blaak, E.E., Fenech, M., Vauzour, D., McArdle, H.J., Kremer, B.H.A., Sterkman, L., Vafeiadou, K., Benedetti, M.M., Williams, C.M., Calder, P.C., 2015. Low-grade inflammation, diet composition and health: current research evidence and its translation. Br. J. Nutr. 114, 999–1012. https://doi.org/10.1017/S0007114515002093.
Mozaffarian, D., Clarke, R., 2009. Quantitative effects on cardiovascular risk factors and coronary heart disease risk of replacing partially hydrogenated vegetable oils with other fats and oils. Eur. J. Clin. Nutr. 63, S22–S33. https://doi.org/10.1038/sj.ejcn.1602976.
Mozaffarian, D., Micha, R., Wallace, S., 2010. Effects on coronary heart disease of increasing polyunsaturated fat in place of saturated fat: a systematic review and meta-analysis of randomized controlled trials. PLoS Med. 7, e1000252. https://doi.org/10.1371/journal.pmed.1000252.
Murray, C.J., et al., 2013. The State of US Health, 1990–2010: burden of diseases, injuries, and risk factors. JAMA 310, 591–608. https://doi.org/10.1001/jama.2013.13805.
Ndanuko, R.N., Tapsell, L.C., Charlton, K.E., Neale, E.P., Batterham, M.J., 2016. Dietary patterns and blood pressure in adults: a systematic review and meta-analysis of randomized controlled trials. Adv. Nutr. 7, 76–89. https://doi.org/10.3945/an.115.009753.
Nissensohn, M., Román-Viñas, B., Sánchez-Villegas, A., Piscopo, S., Serra-Majem, L., 2016. The effect of the Mediterranean diet on hypertension: a systematic review and meta-analysis. J. Nutr. Educ. Behav. 48, 42–53.e1. https://doi.org/10.1016/j.jneb.2015.08.023.
Noel, S.E., Newby, P.K., Ordovas, J.M., Tucker, K.L., 2009. A traditional rice and beans pattern is associated with metabolic syndrome in Puerto Rican older adults. J. Nutr. 139, 1360–1367. https://doi.org/10.3945/jn.109.105874.
Panaretos, D., Koloverou, E., Dimopoulos, A.C., Kouli, G.-M., Vamvakari, M., Tzavelas, G., Pitsavos, C., Panagiotakos, D.B., 2018. A comparison of statistical and machine-learning techniques in evaluating the association between dietary patterns and 10-year cardiometabolic risk (2002–2012): the ATTICA study. Br. J. Nutr. 120, 326–334. https://doi.org/10.1017/S0007114518001150.
Phillips, R.L., Lemon, F.R., Beeson, W.L., Kuzma, J.W., 1978. Coronary heart disease mortality among Seventh-Day Adventists with differing dietary habits: a preliminary report. Am. J. Clin. Nutr. 31, S191–S198. https://doi.org/10.1093/ajcn/31.10.S191.
Phillips, C.M., Chen, L.-W., Heude, B., Bernard, J.Y., Harvey, N.C., Duijts, L., Mensink-Bout, S.M., Polanska, K., Mancano, G., Suderman, M., Shivappa, N., Hébert, J.R., 2019. Dietary inflammatory index and non-communicable disease risk: a narrative review. Nutrients 11, 1873. https://doi.org/10.3390/nu11081873.
Playdon, M.C., Moore, S.C., Derkach, A., Reedy, J., Subar, A.F., Sampson, J.N., Albanes, D., Gu, F., Kontto, J., Lassale, C., Liao, L.M., Männistö, S., Mondul, A.M., Weinstein, S.J., Irwin, M.L., Mayne, S.T., Stolzenberg-Solomon, R., 2017. Identifying biomarkers of dietary patterns by using metabolomics. Am. J. Clin. Nutr. 105, 450–465. https://doi.org/10.3945/ajcn.116.144501.
Rafiq, T., Azab, S.M., Teo, K.K., Thabane, L., Anand, S.S., Morrison, K.M., de Souza, R.J., Britz-McKibbin, P., 2021. Nutritional metabolomics and the classification of dietary biomarker candidates: a critical review. Adv. Nutr. https://doi.org/10.1093/advances/nmab054.
Rebholz, C.M., Lichtenstein, A.H., Zheng, Z., Appel, L.J., Coresh, J., 2018. Serum untargeted metabolomic profile of the Dietary Approaches to Stop Hypertension (DASH) dietary pattern. Am. J. Clin. Nutr. 108, 243–255. https://doi.org/10.1093/ajcn/nqy099.
Rehm, C.D., Peñalvo, J.L., Afshin, A., Mozaffarian, D., 2016. Dietary intake among US adults, 1999–2012. JAMA 315, 2542–2553. https://doi.org/10.1001/jama.2016.7491.
Reynolds, A., Mann, J., Cummings, J., Winter, N., Mete, E., Te Morenga, L., 2019. Carbohydrate quality and human health: a series of systematic reviews and meta-analyses. Lancet 393, 434–445. https://doi.org/10.1016/S0140-6736(18)31809-9.
Riggs, D.W., Yeager, R.A., Bhatnagar, A., 2018. Defining the human envirome. Circ. Res. 122, 1259–1275. https://doi.org/10.1161/CIRCRESAHA.117.311230.
Rimm, E.B., Appel, L.J., Chiuve, S.E., Djoussé, L., Engler, M.B., Kris-Etherton, P.M., Mozaffarian, D., Siscovick, D.S., Lichtenstein, A.H., et al., 2018. Seafood long-chain n-3 polyunsaturated fatty acids and cardiovascular disease: a science advisory from the American Heart Association. Circulation 138, e35–e47. https://doi.org/10.1161/CIR.0000000000000574.
Roth, G.A., Mensah, G.A., Johnson, C.O., et al., 2020. Global burden of cardiovascular diseases and risk factors, 1990–2019. J. Am. Coll. Cardiol. 76, 2982–3021. https://doi.org/10.1016/j.jacc.2020.11.010.
Rumawas, M.E., Dwyer, J.T., Mckeown, N.M., Meigs, J.B., Rogers, G., Jacques, P.F., 2009. The development of the Mediterranean-style dietary pattern score and its application to the American diet in the Framingham Offspring Cohort. J. Nutr. 139, 1150–1156. https://doi.org/10.3945/jn.109.103424.
Sacks, F.M., Svetkey, L.P., Vollmer, W.M., Appel, L.J., Bray, G.A., Harsha, D., Obarzanek, E., Conlin, P.R., Miller, E.R., Simons-Morton, D.G., Karanja, N., Lin, P.-H., Aickin, M., Most-Windhauser, M.M., Moore, T.J., Proschan, M.A., Cutler, J.A., 2001. Effects on blood pressure of reduced dietary sodium and the Dietary Approaches to Stop Hypertension (DASH) diet. N. Engl. J. Med. 344, 3–10. https://doi.org/10.1056/NEJM200101043440101.
Satija, A., Bhupathiraju, S.N., Rimm, E.B., Spiegelman, D., Chiuve, S.E., Borgi, L., Willett, W.C., Manson, J.E., Sun, Q., Hu, F.B., 2016. Plant-based dietary patterns and incidence of type 2 diabetes in US men and women: results from three prospective cohort studies. PLoS Med. 13, e1002039. https://doi.org/10.1371/journal.pmed.1002039.
Seah, J.Y.H., Ong, C.N., Koh, W.-P., Yuan, J.-M., van Dam, R.M., 2019. A dietary pattern derived from reduced rank regression and fatty acid biomarkers is associated with lower risk of type 2 diabetes and coronary artery disease in Chinese adults. J. Nutr. 149, 2001–2010. https://doi.org/10.1093/jn/nxz164.
Shakeri, Z., Mirmiran, P., Khalili-Moghadam, S., Hosseini-Esfahani, F., Ataie-Jafari, A., Azizi, F., 2019. Empirical dietary inflammatory pattern and risk of metabolic syndrome and its components: tehran Lipid and Glucose Study. Diabetol. Metab. Syndrome 11, 16. https://doi.org/10.1186/s13098-019-0411-4.
Shan, Z., Li, Y., Baden, M.Y., Bhupathiraju, S.N., Wang, D.D., Sun, Q., Rexrode, K.M., Rimm, E.B., Qi, L., Willett, W.C., Manson, J.E., Qi, Q., Hu, F.B., 2020. Association between healthy eating patterns and risk of cardiovascular disease. JAMA Intern. Med. 180, 1090–1100. https://doi.org/10.1001/jamainternmed.2020.2176.

Shang, X., Li, Y., Xu, H., Zhang, Q., Liu, A., Du, S., Guo, H., Ma, G., 2020. Leading dietary determinants identified using machine learning techniques and a healthy diet score for changes in cardiometabolic risk factors in children: a longitudinal analysis. Nutr. J. 19, 105. https://doi.org/10.1186/s12937-020-00611-2.
Shao, W., Chen, J., Fan, S., Lei, Y., Xu, H., Zhou, J., Cheng, P., Yang, Y., Rao, C., Wu, B., Liu, H., Xie, P., 2015. Combined metabolomics and proteomics analysis of major depression in an animal model: perturbed energy metabolism in the chronic mild stressed rat cerebellum. OMICS 19, 383–392. https://doi.org/10.1089/omi.2014.0164.
Shea, B.J., Reeves, B.C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., Henry, D.A., 2017. Amstar 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 358, j4008. https://doi.org/10.1136/bmj.j4008.
Shekelle, R.B., Shryock, A.M., Paul, O., Lepper, M., Stamler, J., Liu, S., Raynor, W.J., 1981. Diet, serum cholesterol, and death from coronary heart disease. N. Engl. J. Med. 304, 65–70. https://doi.org/10.1056/NEJM198101083040201.
Sherwin, R., 1978. Controlled trials of the diet-heart hypothesis: some comments on the experimental unit. Am. J. Epidemiol. 108, 92–99. https://doi.org/10.1093/oxfordjournals.aje.a112607.
Shikany, J.M., Safford, M.M., Newby, P.K., Durant, R.W., Brown, T.M., Judd, S.E., 2015. Southern dietary pattern is associated with hazard of acute coronary heart disease in the Reasons for Geographic and Racial Differences in Stroke (REGARDS) Study. Circulation 132, 804–814. https://doi.org/10.1161/CIRCULATIONAHA.114.014421.
Shimazu, T., Kuriyama, S., Hozawa, A., Ohmori, K., Sato, Y., Nakaya, N., Nishino, Y., Tsubono, Y., Tsuji, I., 2007. Dietary patterns and cardiovascular disease mortality in Japan: a prospective cohort study. Int. J. Epidemiol. 36, 600–609. https://doi.org/10.1093/ije/dym005.
Shivappa, N., Steck, S.E., Hurley, T.G., Hussey, J.R., Hébert, J.R., 2014. Designing and developing a literature-derived, population-based dietary inflammatory index. Publ. Health Nutr. core/journals/public-health-nutrition/article/designing-and-developing-a-literature-derived-population-based-dietary-inflammatory-index/30BE2C2295CE93DC6B54F9-F9AD50CC68. (Accessed 18 November 2016).
Shivappa, N., Hébert, J.R., Rietzschel, E.R., Buyzere, M.L.D., Langlois, M., Debruyne, E., Marcos, A., Huybrechts, I., 2015. Associations between dietary inflammatory index and inflammatory markers in the Asklepios Study. Br. J. Nutr. 113, 665–671. https://doi.org/10.1017/S000711451400395X.
Shoenbill, K., Song, Y., Craven, M., Johnson, H., Smith, M., Mendonca, E.A., 2020. Identifying patterns and predictors of lifestyle modification in electronic health record documentation using statistical and machine learning methods. Prev. Med. 136, 106061. https://doi.org/10.1016/j.ypmed.2020.106061.
Siervo, M., Lara, J., Chowdhury, S., Ashor, A., Oggioni, C., Mathers, J.C., 2015. Effects of the Dietary Approach to Stop Hypertension (DASH) diet on cardiovascular risk factors: a systematic review and meta-analysis. Br. J. Nutr. 113, 1–15. https://doi.org/10.1017/S0007114514003341.
Singh, R.B., Dubnov, G., Niaz, M.A., Ghosh, S., Singh, R., Rastogi, S.S., Manor, O., Pella, D., Berry, E.M., 2002. Effect of an Indo-Mediterranean diet on progression of coronary artery disease in high risk patients (Indo-Mediterranean Diet Heart Study): a randomised single-blind trial. Lancet 360, 1455–1461. https://doi.org/10.1016/S0140-6736(02)11472-3.
Skeaff, C.M., Miller, J., 2009. Dietary fat and coronary heart disease: summary of evidence from prospective cohort and randomised controlled trials. ANM 55, 173–201. https://doi.org/10.1159/000229002.
Sotos-Prieto, M., Bhupathiraju, S.N., Mattei, J., Fung, T.T., Li, Y., Pan, A., Willett, W.C., Rimm, E.B., Hu, F.B., 2015. Changes in diet quality scores and risk of cardiovascular disease among US men and women. Circulation 132, 2212–2219. https://doi.org/10.1161/CIRCULATIONAHA.115.017158.
Sotos-Prieto, M., Bhupathiraju, S.N., Mattei, J., Fung, T.T., Li, Y., Pan, A., Willett, W.C., Rimm, E.B., Hu, F.B., 2017. Association of changes in diet quality with total and cause-specific mortality. N. Engl. J. Med. 377, 143–153. https://doi.org/10.1056/NEJMoa1613502.
Srour, B., Fezeu, L.K., Kesse-Guyot, E., Allès, B., Méjean, C., Andrianasolo, R.M., Chazelas, E., Deschasaux, M., Hercberg, S., Galan, P., Monteiro, C.A., Julia, C., Touvier, M., 2019. Ultra-processed food intake and risk of cardiovascular disease: prospective cohort study (NutriNet-Santé). BMJ 365, l1451. https://doi.org/10.1136/bmj.l1451.
Steffen, L.M., Van Horn, L., Daviglus, M.L., Zhou, X., Reis, J.P., Loria, C.M., Jacobs, D.R., Duffey, K.J., 2014. A modified Mediterranean diet pattern is related to lower risk of incident metabolic syndrome over 25 years among young adults: the CARDIA Study. Br. J. Nutr. 112, 1654–1661. https://doi.org/10.1017/S0007114514002633.
Stevens, L.M., Linstead, E., Hall, J.L., Kao, D.P., 2021. Association between coffee intake and incident heart failure risk. Circ. Heart Fail. 14, e006799. https://doi.org/10.1161/CIRCHEARTFAILURE.119.006799.
Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P., 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. J. Chem. Inf. Comput. Sci. 43, 1947–1958. https://doi.org/10.1021/ci034160g.
Tabung, F.K., Steck, S.E., Zhang, J., Ma, Y., Liese, A.D., Agalliu, I., Hingle, M., Hou, L., Hurley, T.G., Jiao, L., Martin, L.W., Millen, A.E., Park, H.L., Rosal, M.C., Shikany, J.M., Shivappa, N., Ockene, J.K., Hebert, J.R., 2015. Construct validation of the dietary inflammatory index among postmenopausal women. Ann. Epidemiol. 25, 398–405. https://doi.org/10.1016/j.annepidem.2015.03.009.
Tabung, F.K., Smith-Warner, S.A., Chavarro, J.E., Wu, K., Fuchs, C.S., Hu, F.B., Chan, A.T., Willett, W.C., Giovannucci, E.L., 2016. Development and validation of an empirical dietary inflammatory index. J. Nutr. 146, 1560–1570. https://doi.org/10.3945/jn.115.228718.
Tabung, F.K., Wang, W., Fung, T.T., Hu, F.B., Smith-Warner, S.A., Chavarro, J.E., Fuchs, C.S., Willett, W.C., Giovannucci, E.L., 2016. Development and validation of empirical indices to assess the insulinaemic potential of diet and lifestyle. Br. J. Nutr. 116, 1787–1798. https://doi.org/10.1017/S0007114516003755.
Tabung, F.K., Satija, A., Fung, T.T., Clinton, S.K., Giovannucci, E.L., 2019. Long-term change in both dietary insulinemic and inflammatory potential is associated with weight gain in adult women and men. J. Nutr. 149, 804–815. https://doi.org/10.1093/jn/nxy319.
Taylor, H.L., Menotti, A., Puddu, V., Monti, M., Keys, A., 1970. Coronary heart disease in seven countries. XI. Five years of follow-up of railroad men in Italy. Circulation 41, I113–I122. https://doi.org/10.1161/01.cir.41.4s1.i-113.
Tebani, A., Bekri, S., 2019. Paving the way to precision nutrition through metabolomics. Front. Nutr. 6. https://doi.org/10.3389/fnut.2019.00041.
Tebani, A., Afonso, C., Marret, S., Bekri, S., 2016. Omics-based strategies in precision medicine: toward a paradigm shift in inborn errors of metabolism investigations. Int. J. Mol. Sci. 17. https://doi.org/10.3390/ijms17091555.
Teymoori, F., Farhadnejad, H., Mokhtari, E., Sohouli, M.H., Moslehi, N., Mirmiran, P., Azizi, F., 2021. Dietary and lifestyle inflammatory scores and risk of incident diabetes: a prospective cohort among participants of Tehran lipid and glucose study. BMC Publ. Health 21, 1293. https://doi.org/10.1186/s12889-021-11327-1.
Tong, T.Y.N., Koulman, A., Griffin, J.L., Wareham, N.J., Forouhi, N.G., Imamura, F., 2020. A combination of metabolites predicts adherence to the Mediterranean diet pattern and its associations with insulin sensitivity and lipid homeostasis in the general population: the Fenland Study, United Kingdom. J. Nutr. 150, 568–578. https://doi.org/10.1093/jn/nxz263.
Trichopoulou, A., Kouris-Blazos, A., Wahlqvist, M.L., Gnardellis, C., Lagiou, P., Polychronopoulos, E., Vassilakou, T., Lipworth, L., Trichopoulos, D., 1995. Diet and overall survival in elderly people. BMJ 311, 1457–1460.
Trichopoulou, A., Costacou, T., Bamia, C., Trichopoulos, D., 2003. Adherence to a Mediterranean diet and survival in a Greek population. N. Engl. J. Med. 348, 2599–2608. https://doi.org/10.1056/NEJMoa025039.
Tsao, C.W., Aday, A.W., Almarzooq, Z.I., et al., 2022. Heart disease and stroke statistics—2022 update: a report from the American Heart Association. Circulation. https://doi.org/10.1161/CIR.0000000000001052. CIR.0000000000001052.
Tuso, P.J., Ismail, M.H., Ha, B.P., Bartolotto, C., 2013. Nutritional update for physicians: plant-based diets. Perm. J. 17, 61–66. https://doi.org/10.7812/TPP/12-085.
U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2015. 2015—2020 Dietary Guidelines for Americans, eighth ed.
U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2020. Dietary Guidelines for Americans, ninth ed., pp. 2020–2025
Uusitupa, M., Hermansen, K., Savolainen, M.J., et al., 2013. Effects of an isocaloric healthy Nordic diet on insulin sensitivity, lipid profile and inflammation markers in metabolic syndrome—a randomized study (SYSDIET). J. Intern. Med. 274, 52–66. https://doi.org/10.1111/joim.12044.
Viguiliouk, E., Glenn, A.J., Nishi, S.K., Chiavaroli, L., Seider, M., Khan, T., Bonaccio, M., Iacoviello, L., Mejia, S.B., Jenkins, D.J.A., Kendall, C.W.C., Kahleová, H., Rahelić, D., Salas-Salvadó, J., Sievenpiper, J.L., 2019. Associations between dietary pulses alone or with other legumes and cardiometabolic disease outcomes: an umbrella review and updated systematic review and meta-analysis of prospective cohort studies. Adv. Nutr. 10, S308–S319. https://doi.org/10.1093/advances/nmz113.

Vincent-Baudry, S., Defoort, C., Gerber, M., Bernard, M.-C., Verger, P., Helal, O., Portugal, H., Planells, R., Grolier, P., Amiot-Carlin, M.-J., Vague, P., Lairon, D., 2005. The Medi-RIVAGE study: reduction of cardiovascular disease risk factors after a 3-mo intervention with a Mediterranean-type diet or a low-fat diet. Am. J. Clin. Nutr. 82, 964–971. https://doi.org/10.1093/ajcn/82.5.964.

Virani, S.S., et al., 2021. Heart disease and stroke statistics—2021 update. Circulation 143, e254–e743. https://doi.org/10.1161/CIR.0000000000000950.

Walker, M.E., Xanthakis, V., Peterson, L.R., Duncan, M.S., Lee, J., Ma, J., Bigornia, S., Moore, L.L., Quatromoni, P.A., Vasan, R.S., Jacques, P.F., 2020. Dietary patterns, ceramide ratios, and risk of all-cause and cause-specific mortality: the Framingham Offspring Study. J. Nutr. 150, 2994–3004. https://doi.org/10.1093/jn/nxaa269.

Walker, M.E., Song, R.J., Xu, X., Gerszten, R.E., Ngo, D., Clish, C.B., Corlin, L., Ma, J., Xanthakis, V., Jacques, P.F., Vasan, R.S., 2020. Proteomic and metabolomic correlates of healthy dietary patterns: the Framingham Heart Study. Nutrients 12, 1476. https://doi.org/10.3390/nu12051476.

Wang, X., Ouyang, Y., Liu, J., Zhu, M., Zhao, G., Bao, W., Hu, F.B., 2014. Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. BMJ 349, g4490. https://doi.org/10.1136/bmj.g4490.

Wang, F., Zheng, J., Yang, B., Jiang, J., Fu, Y., Li, D., 2015. Effects of vegetarian diets on blood lipids: a systematic review and meta-analysis of randomized controlled trials. J. Am. Heart Assoc. 4, e002408. https://doi.org/10.1161/JAHA.115.002408.

Wang, D.D., Li, Y., Chiuve, S.E., Stampfer, M.J., Manson, J.E., Rimm, E.B., Willett, W.C., Hu, F.B., 2016. Specific dietary fats in relation to total and cause-specific mortality. JAMA Intern. Med. 176, 1134–1145. https://doi.org/10.1001/jamainternmed.2016.2417.

Wang, X., Lin, X., Ouyang, Y.Y., Liu, J., Zhao, G., Pan, A., Hu, F.B., 2016. Red and processed meat consumption and mortality: dose-response meta-analysis of prospective cohort studies. Publ. Health Nutr. 19, 893–905. https://doi.org/10.1017/S1368980015002062.

Wang, D.D., Li, Y., Bhupathiraju, S.N., Rosner, B.A., Sun, Q., Giovannucci, E.L., Rimm, E.B., Manson, J.E., Willett, W.C., Stampfer, M.J., Hu, F.B., 2021a. Fruit and vegetable intake and mortality. Circulation 143, 1642–1654. https://doi.org/10.1161/CIRCULATIONAHA.120.048996.

Wang, D.D., Nguyen, L.H., Li, Y., Yan, Y., Ma, W., Rinott, E., Ivey, K.L., Shai, I., Willett, W.C., Hu, F.B., Rimm, E.B., Stampfer, M.J., Chan, A.T., Huttenhower, C., 2021b. The gut microbiome modulates the protective association between a Mediterranean diet and cardiometabolic disease risk. Nat. Med. 27, 333–343. https://doi.org/10.1038/s41591-020-01223-3.

Wang, F., Baden, M.Y., Li, J., Guasch-Ferré, M., Li, Y., Wan, Y., Bhupathiraju, S.N., Tobias, D.K., Clish, C., Mucci, L.A., Eliassen, A.H., Costenbader, K., Karlson, E.W., Ascherio, A., Rimm, E.B., Manson, J.E., Liang, L., Hu, F., 2021c. Abstract 022: plasma metabolome related to plant-based diets and the association with type 2 diabetes: a prospective cohort study. Circulation 143, A022. https://doi.org/10.1161/circ.143.suppl_1.022.

Wang, W., Yan, Y., Guo, Z., Hou, H., Garcia, M., Tan, X., Anto, E.O., Mahara, G., Zheng, Y., Li, B., Kang, T., Zhong, Z., Wang, Y., Guo, X., Golubnitschaja, O., On Behalf of Suboptimal Health Study Consortium and European Association for Predictive, Preventive and Personalised Medicine, 2021d. All around suboptimal health—a joint position paper of the Suboptimal Health Study Consortium and European Association for Predictive, Preventive and Personalised Medicine. EPMA J. 12, 403–433. https://doi.org/10.1007/s13167-021-00253-2.

Wang, Y., Chen, B., Zhang, J., Li, H., Zeng, X., Zhang, Z., Zhu, Y., Li, X., Hu, A., Zhao, Q., Yang, W., 2021e. Diets with higher insulinaemic potential are associated with increased risk of overall and cardiovascular disease-specific mortality. Br. J. Nutr. 1–10. https://doi.org/10.1017/S0007114521004815.

Warensjö, L.E., et al., 2019. Dietary pattern specific protein biomarkers for cardiovascular disease: a cross-sectional study in 2 independent cohorts. J. Am. Heart Assoc. 8, e011860. https://doi.org/10.1161/JAHA.118.011860.

Welsh, J.A., Sharma, A., Cunningham, S.A., Vos, M.B., 2011. Consumption of added sugars and indicators of cardiovascular disease risk among US adolescents. Circulation 123, 249–257. https://doi.org/10.1161/CIRCULATIONAHA.110.972166.

Willett, W.C., Sacks, F., Trichopoulou, A., Drescher, G., Ferro-Luzzi, A., Helsing, E., Trichopoulos, D., 1995. Mediterranean diet pyramid: a cultural model for healthy eating. Am. J. Clin. Nutr. 61, 1402S–1406S. https://doi.org/10.1093/ajcn/61.6.1402S.

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F., Wood, A., Jonell, M., Clark, M., Gordon, L.J., Fanzo, J., Hawkes, C., Zurayk, R., Rivera, J.A., De Vries, W., Majele Sibanda, L., Afshin, A., Chaudhary, A., Herrero, M., Agustina, R., Branca, F., Lartey, A., Fan, S., Crona, B., Fox, E., Bignet, V., Troell, M., Lindahl, T., Singh, S., Cornell, S.E., Srinath Reddy, K., Narain, S., Nishtar, S., Murray, C.J.L., 2019. Food in the Anthropocene: the EAT–Lancet Commission on Healthy diets from sustainable food systems. Lancet 393, 447–492. https://doi.org/10.1016/S0140-6736(18)31788-4.

Wirth, M.D., Shivappa, N., Hurley, T.G., Hébert, J.R., 2016. Association between previously diagnosed circulatory conditions and a dietary inflammatory index. Nutr. Res. 36, 227–233. https://doi.org/10.1016/j.nutres.2015.11.016.

Wirth, M.D., Shivappa, N., Davis, L., Hurley, T.G., Ortaglia, A., Drayton, R., Blair, S.N., Hébert, J.R., 2017. Construct validation of the dietary inflammatory index among African Americans. J. Nutr. Health Aging 21, 487–491. https://doi.org/10.1007/s12603-016-0775-1.

Yang, Q., Zhang, Z., Gregg, E.W., Flanders, W.D., Merritt, R., Hu, F.B., 2014. Added sugar intake and cardiovascular diseases mortality among US adults. JAMA Intern. Med. 174, 516–524. https://doi.org/10.1001/jamainternmed.2013.13563.

Yashpal, S., Liese, A.D., Boucher, B.A., Wagenknecht, L.E., Haffner, S.M., Johnston, L.W., Bazinet, R.P., Rewers, M., Rotter, J.I., Watkins, S.M., Hanley, A.J., 2021. Metabolomic profiling of the Dietary Approaches to Stop Hypertension diet provides novel insights for the nutritional epidemiology of type 2 diabetes mellitus. Br. J. Nutr. 1–11. https://doi.org/10.1017/S0007114521003561.

Yokoyama, Y., Nishimura, K., Barnard, N.D., Takegami, M., Watanabe, M., Sekikawa, A., Okamura, T., Miyamoto, Y., 2014. Vegetarian diets and blood pressure: a meta-analysis. JAMA Intern. Med. 174, 577–587. https://doi.org/10.1001/jamainternmed.2013.14547.

Zeevi, D., Korem, T., Zmora, N., Israeli, D., Rothschild, D., Weinberger, A., Ben-Yacov, O., Lador, D., Avnit-Sagi, T., Lotan-Pompan, M., Suez, J., Mahdi, J.A., Matot, E., Malka, G., Kosower, N., Rein, M., Zilberman-Schapira, G., Dohnalová, L., Pevsner-Fischer, M., Bikovsky, R., Halpern, Z., Elinav, E., Segal, E., 2015. Personalized nutrition by prediction of glycemic responses. Cell 163, 1079–1094. https://doi.org/10.1016/j.cell.2015.11.001.

Zhang, B., Xiong, K., Cai, J., Ma, A., 2020. Fish consumption and coronary heart disease: a meta-analysis. Nutrients 12, 2278. https://doi.org/10.3390/nu12082278.

Zhao, J., Li, Z., Gao, Q., Zhao, H., Chen, S., Huang, L., Wang, W., Wang, T., 2021. A review of statistical methods for dietary pattern analysis. Nutr. J. 20, 37. https://doi.org/10.1186/s12937-021-00692-7.

Zhao, Y., Naumova, E.N., Bobb, J.F., Claus Henn, B., Singh, G.M., 2021. Joint associations of multiple dietary components with cardiovascular disease risk: a machine-learning approach. Am. J. Epidemiol. 190, 1353–1365. https://doi.org/10.1093/aje/kwab004.

Zheng, Y., Li, Y., Satija, A., Pan, A., Sotos-Prieto, M., Rimm, E., Willett, W.C., Hu, F.B., 2019. Association of changes in red meat consumption with total and cause specific mortality among US women and men: two prospective cohort studies. BMJ 365, l2110. https://doi.org/10.1136/bmj.l2110.

Zhong, V.W., Van Horn, L., Greenland, P., Carnethon, M.R., Ning, H., Wilkins, J.T., Lloyd-Jones, D.M., Allen, N.B., 2020. Associations of processed meat, unprocessed red meat, poultry, or fish intake with incident cardiovascular disease and all-cause mortality. JAMA Intern. Med. 180, 503–512. https://doi.org/10.1001/jamainternmed.2019.6969.

Dietary issues in coronary heart disease prevention

Rajeev Gupta[a,b] **and Shalini Singhal**[c], [a] Academic Research Development Unit, Rajasthan University of Health Sciences, Jaipur, Rajasthan, India; [b] Department of Preventive Cardiology & Medicine, Eternal Heart Care Centre & Research Institute, Jaipur, Rajasthan, India; and [c] Dr. Shalini's Diet & Wellness Clinic, Dwarka, New Delhi, India

© 2023 Elsevier Ltd. All rights reserved.

This is an update of K.S. Reddy, Coronary Heart Disease: Prevention, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 409–415, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00064-7.

Introduction **262**
Global trends in CHD as a reflection of nutrition transition 262
Understanding the links between nutrition and CHD 262
Nutrients and coronary heart disease **263**
Saturated fatty acids (SFA) 263
Trans-fatty acids (TFA) 263
Polyunsaturated fatty acids (PUFA) 264
Monounsaturated fatty acids (MUFA) 264
Dietary cholesterol 264
Carbohydrates 265
Fiber 265
Antioxidants 265
Folate 265
Flavonoids and other phytochemicals 265
Sodium and potassium 266
Calcium 266
Specific foods **266**
Fruits and vegetables 266
Fish 266
Nuts 266
Dairy products 267
Alcohol 267
Composite diets **267**
Mediterranean diet 267
Dietary Approaches to Stop Hypertension (DASH) 267
Vegetarian diets 268
Prudent diets 268
Japanese diet 268
Comprehensive prevention **268**
Conclusions **268**
References **269**

Key points

- Coronary heart disease (CHD) is the most important cause of deaths and disease burden globally. It is rapidly increasing in developing countries of Asia and Africa, in contrast to developed countries where it is declining.
- Nutrition transition from unhealthy to healthy nutrients and foods has been important in prevention and control of CHD in developed countries. In developing countries an unhealthy food transition is currently evolving.
- Cardioprotective foods include omega-3 fats, unrefined carbohydrates, fiber, low sodium, high potassium, fruits, vegetables, nuts and fish.
- Composite dietary patterns are important, and consumption of the Mediterranean type, DASH-type and other prudent diets is recommended.

Introduction

Coronary heart disease (CHD) is the leading cause of death in the world. It is well established as the foremost contributor to mortality and morbidity in most developed countries and has now emerged as the leading cause of death and disability in most developing countries. The global health transitions, which have seen substantial changes in age-specific coronary mortality rates across the world, have also been associated with nutrition transitions, which explain a large part of the rise or fall of CHD-related death rates (Global Burden of Disease Study 2019 Viewpoint Collaborators, 2020).

Diet and nutrition have been extensively investigated as risk factors for CHD (Gupta and Wood, 2019). Many dietary factors have been linked directly to an increased or decreased risk of CHD or to major established risk factors of CHD such as high blood pressure (BP), disordered blood fats- raised low density lipoprotein (LDL) cholesterol and triglycerides and low high density lipoprotein (HDL) cholesterol, diabetes and metabolic syndrome, overweight and obesity, and also to emerging risk factors like inflammatory markers and homocysteine. Nutrition influences atherogenesis, thrombosis, and inflammation, all of which are interconnected pathways that lead to CHD (Gupta and Wood, 2019). Observational epidemiological studies and clinical trials have contributed to a wide body of knowledge on the role that some nutrients (like saturated and trans-fats, salt, and refined carbohydrates) play important role in increasing risk of CHD and of protective effect of other nutrients (such as fruit and vegetables, polyunsaturated fats, nuts, and fish) against CHD. This knowledge has been successfully applied both in public health and in clinical practice, to reduce the risk of CHD, in populations as well as in individuals. This article summarizes the present state of that knowledge, as relevant to prevention of CHD.

Global trends in CHD as a reflection of nutrition transition

CHD affects about 126 million individuals (1655/100,000) globally and led more than 9 million deaths in 2020 and are large fraction of not only the total number of deaths world-wide due to cardiovascular diseases but also of the global total of deaths from any cause (Global Burden of Disease Study 2019 Viewpoint Collaborators, 2020). Although age-specific coronary mortality rates have declined in the developed countries, over the past few decades, the absolute burdens of CHD continue to be high (Global Burden of Disease Study 2019 Viewpoint Collaborators, 2020). Notably, CHD death rates are rising in the developing countries, where about half of these deaths occur below the age of 70 years, entailing high loss in potential productive years of life as well as in national incomes. In the INTERHEART study, which examined risk factors for myocardial infarction (MI) in 52 nations worldwide, the mean age for first MI in South Asia was 53 years compared to 58 years in other parts of the world. In Eastern and Central Europe CHD mortality rates rose sharply in the 1980s and 1990s but have declined in the recent decades.

These changes in CHD mortality rates have accompanied well documented or clearly discernible shifts in the nutritional state of the populations (World Health Organization, 2003). The decline of CHD mortality in Western and Northern Europe was linked to a reduction in the consumption of unhealthy fats (saturated fats and trans-fats) and salt as well as an increased consumption of fruits and vegetables. Similarly, the decline of CHD mortality in Poland was explained by the increase in fruit and vegetable consumption and growing substitution of vegetable fats for animal fats. Similar evidence of a favorable nutrition transition preceding the decline in CHD mortality rates is available from other developed countries like USA, UK, Canada, Australia, and New Zealand (World Health Organization, 2003).

The developing countries have, on the other hand, witnessed a nutrition transition in the opposite direction (World Health Organization, 2003). China and India, for example, experienced a large increase in fat consumption over the past two decades, accompanied by a progressive rise in the mean plasma cholesterol levels of the population as well as in the CHD mortality rates. Other developing countries too are increasingly adopting unhealthy dietary patterns, which augment the risk of CHD. In India, there has been a progressive increase in the consumption of unhealthy dairy products, refined sugars, and hydrogenated edible oils, most of which are very high in saturated and trans-fats (Gupta and Wood, 2019).

Understanding the links between nutrition and CHD

The pathogenesis of CHD is mediated through the interconnected pathways of atherogenesis (fat deposition in the walls of the coronary arteries to form plaques), thrombosis (blood clotting over disrupted plaques), and inflammation (which initially damages the blood vessel walls and continues to destabilize the plaques) (Lawler et al., 2021). Nutrition has a major role in influencing each of these pathways and often provides the connecting link between them.

Major coronary risk factors include an abnormal blood lipid profile (especially plasma cholesterol and its subfractions), high BP, and diabetes. Overweight and obesity (both the general and central patterns) are also associated with an increased risk of CHD (Gupta and Wood, 2019). Nutrition has a powerful influence on all of these risk factors, with an unhealthy diet pattern tending to elevate them and a healthy diet pattern reducing the levels of risk (**Table 1**). Diet becomes especially important in the context of the metabolic syndrome (a complex of central obesity, high BP, dyslipidemia, and glucose intolerance), an entity, which has been identified as a major risk factor for CHD. Nutrition is also linked to the propensity to develop cardiac arrhythmias, and is an important predictor of sudden cardiac death. These links between dietary patterns and several specific nutrients not only manifest as fat deposition in the arteries, plaque growth, plaque instability, and thrombosis but are also evident much earlier in the natural history of CHD, as endothelial dysfunction (inability of the arteries to dilate normally), elevated levels of inflammatory markers

Table 1 Dietary nutrients and composite diets with evidence of prevention, harm and equivocal impact on CHD incidence.

	Preventive effect	*Evidence of harm*	*Equivocal evidence*
Nutrients	• N-3 fatty polyunsaturated fats • Monounsaturated fats • Unrefined carbohydrates • Fiber • Low to moderate sodium • High potassium • Fruits and vegetables • Nuts • Fish (fatty) • Unrefined cereals	• High cholesterol • Saturated fats • Trans fats • Refined carbohydrates • Very high sodium • Antioxidant vitamins • Heavy alcohol and binge drinking	• N-6 polyunsaturated fats • Vitamin D • Multi-vitamins • Calcium • Flavonoids • Phytochemicals • Dairy and dairy-products
Composite foods and diets	• Mediterranean diet • DASH type diet • Other prudent diets	• High-fat diets • Very low-carb diets	• Japanese diet • Intermittent fasting • Vegan and vegetarian diets

(such as C-reactive protein), and increased intimal medial thickness of arterial walls. These precede and predict the clinical manifestation of CHD.

Nutrients and coronary heart disease

Saturated fatty acids (SFA)

The relationship between dietary fats and cardiovascular disease (CVD), especially CHD has been extensively investigated, with strong and consistent associations emerging from a wide body of evidence accrued from animal experiments, as well as observational studies, clinical trials, and metabolic studies conducted in diverse human populations (Mozaffarian, 2016). This relationship was initially considered to be mediated mainly through the atherogenic effects of plasma lipids (total cholesterol, lipoprotein fractions such as LDL and HDL cholesterol, and triglycerides). The effects of dietary fats on thrombosis and endothelial function as well as the relationship of plasma and tissue lipids to the pathways of inflammation have been more recently understood.

The relationship of dietary saturated fat to plasma cholesterol levels and to CHD was graphically demonstrated by the Seven Countries Study involving 16 cohorts, in which saturated fat intake explained up to 73% of the total variance in CHD across these cohorts (World Health Organization, 2003). The most effective replacement for SFAs, in terms of CHD prevention, is by polyunsaturated fatty acids (PUFAs) and monounsaturated fatty acids (MUFAs). This agrees with the outcome of large randomized clinical trials, in which replacement of saturated and trans-fats by polyunsaturated vegetable oils effectively lowered CHD risk (Mozaffarian, 2016). Replacement of SFA with either polyunsaturated or mono-unsaturated fat was shown to lower both LDL and HDL cholesterol. But substitution of SFAs with higher carbohydrate intake especially refined carbohydrates increased the risk for atherogenic dyslipidemia associated with insulin resistance and obesity (Sacks et al., 2017).

According to the American Heart Association, an important thing to remember is the overall dietary picture. Saturated fats are just one piece of the puzzle. In general eating more fruits, vegetables and whole grains and taking fewer calories is a better approach to diet. The large UK Biobank study found no clear associations between total saturated fat and cardiovascular disease outcomes. However, consuming 5% higher total energy from saturated fat from meat was associated with 19% and 21% elevated risks of total cardiovascular disease and heart disease, respectively—but these associations did not remain significant after adjusting for the body mass index (BMI). Associations of SFA from dairy with heart disease went in the opposite direction, but this association was not clear after adjusting for BMI. Similar outcomes have been reported by the Prospective Urban Rural epidemiology (PURE) study conducted in 17 countries. There was a weak positive association of SFA with CHD incidence and mortality and inverse association with all-cause mortality. The current evidence sill points to more harm than good of saturated fats and the intake should be balanced (Mozaffarian, 2016; Sacks et al., 2017).

Trans-fatty acids (TFA)

Trans fatty acids are geometrical isomers of unsaturated fatty acids that assume SFA-like configuration. Partial hydrogenation, the process used to create TFA, also removes essential fatty acids such as linoleic acid (LA) and alpha linolenic acid (ALNA). Evidence that intake of TFA increases the risk of CHD initially became available from large population-based cohort studies in USA and in an elderly Dutch population (Mozaffarian et al., 2006). Metabolic studies have demonstrated that TFA render the plasma lipid profile even more atherogenic than SFA, by not only elevating LDL cholesterol to similar levels but also decreasing HDL cholesterol. As a result, the ratio of LDL cholesterol to HDL cholesterol is significantly higher with a TFA rich diet (2.58) than with a SFA diet (2.34) or an oleic acid diet (2.02). This greatly enhances the risk of CHD. In controlled trials, each 1% energy replacement of TFA with SFA, MUFA, or PUFA, respectively, decreased the total cholesterol/HDL cholesterol ratio by 0.31, 0.54, and 0.67.

Controlled and observational studies have reported that for every 2% increase in energy from TFA, there was a 32% higher risk of MI or CHD death (Mozaffarian et al., 2006). In prospective cohort studies, each 2% energy replacements of TFA with SFA, MUFA, or PUFA lowered CHD risk by 17%, 21%, and 24%, respectively. Therefore, risk of developing CHD depends on the TFA content of the oil or food and also on the fatty acid composition of the substituted oil or fat (Sacks et al., 2017; Mozaffarian et al., 2006).

Eliminating TFA from the diet is an important public health strategy to prevent CHD (Gupta and Wood, 2019). Since these are commercially introduced agents into the diet, policy measures related to the food industry practices would be required along with public education. TFA have been eliminated from retail fats and spreads in many parts of the world, but deep-fat fried fast foods and baked goods are a major and increasing source (World Health Organization, 2003). In the USA, New York City has an ongoing successful program to phase out TFA in restaurant foods. Other US states and most European countries have successfully implemented ban on TFA containing processed foods to help their citizens reduce CHD risk (Mozaffarian, 2016). All packaged foods have a nutrition label that includes fat content. Food makers are required to label trans fats on nutrition and food supplements. Reading of the food labels can help keep track of how much TFA is consumed. Some manufacturers have bypassed this guidance by using the words-partially hydrogenated-in the ingredient list. It is best to avoid food with trans fats completely hence its prudent to limit intake of ultra-processed, packaged and fried foods.

Polyunsaturated fatty acids (PUFA)

PUFA are categorized as n-6 PUFA (mainly derived from linoleic acid) and n-3 PUFA (mainly present in fatty fish and also derived from alpha-linoleic acid). Clinical trials, in which n-6 PUFA (containing linoleic acid) were substituted for SFA showed a greater impact on reduction of both plasma cholesterol and CHD risk, in contrast to trials where low-fat diets, were employed (Mozaffarian, 2016). Much of the epidemiological evidence related to n-3 PUFA is derived from the study of fish consumption in populations or interventions involving fish diets in clinical trials. Fish oils were, however, used in a large clinical trial of 11,300 survivors of MI. After 3.5 years of follow-up, the fish oil group (1 g per day) had a statistically significant 20% reduction in total mortality, 30% reduction in cardiovascular death, and 45% decrease in sudden death (Mozaffarian, 2016; Sacks et al., 2017). However, results of recent randomized clinical trials have failed to demonstrate that consumption of 1–2 g per day of n-3 fatty acids eicosapentanoic acid (EPA) and docosahexanoic acid (DHA) is protective for CHD (Gupta and Wood, 2019). On the other hand, a clinical trial using higher doses of highly purified EPA decreased plasma triacylglycerols, blood pressure, platelet aggregation, and inflammation and significantly reduced CHD risk (Arnett et al., 2019). The balance of omega-6/omega-3 fatty acids has been found to be an important determinant in reducing the risk for CHD. The proportions of SFA, MUFA, and PFA as constituents of total fat intake and total energy consumption have engaged active attention, in view of the strong relationship of these fatty acids to the risk of CHD (Gupta and Wood, 2019).

Monounsaturated fatty acids (MUFA)

The only nutritionally important MUFA is oleic acid, which is abundant in olive and canola oils and in nuts. The epidemiological evidence related to MUFA and CHD is derived from studies on the Mediterranean diet (detailed below), as well as from multiple observational studies around the world (Visseren et al., 2021). A number of clinical trials and their meta-analyses have confirmed the importance of MUFA in diet (Mozaffarian, 2016). The current recommendation of 10–15% calorie intake from MUFA could be further revised in view of the ongoing debate on harms of high carbohydrate diets and avoidance of very high intake of SFA or PUFA (Arnett et al., 2019; Visseren et al., 2021).

Reduction of SFA in the diet has been widely recommended, but its replacement has been an area of debate, as to whether the place of reduced SFAs should be taken by MUFA, PUFA, or carbohydrate (Gupta and Wood, 2019; Sacks et al., 2017). Both MUFA and PUFA improve the lipoprotein profile, although PUFAs are somewhat more effective. In view of this, almost all dietary guidelines suggest that SFA should be kept below 10% of daily energy intake (preferably reduced to 7–8%), MUFA should be increased to 13–15%, and PUFA raised to 7–10% of daily energy, with the total fat contributing to less than 30% of all calories consumed (Arnett et al., 2019; Visseren et al., 2021). These may need to be adjusted for populations who consume less quantities of total fat, so as to ensure an adequate intake of MUFA and PUFA even under those circumstances. The emphasis is now shifting from the quantity of fat to the quality of fat, with growing evidence that even diets with 30–35% fat intake may be protective if the type of fats consumed is mostly from the MUFA and PUFA categories (Mozaffarian, 2016; Visseren et al., 2021). Enhancing the nutritional quality of dietary fat consumption, to provide greater cardiovascular protection, may be attempted by decreasing the sources of saturated fats and eliminating TFA in the diet and increasing the consumption of foods containing unsaturated fatty acids (both MUFA and PUFA).

Dietary cholesterol

Cholesterol in the blood and tissues is derived from two sources: diet and endogenous synthesis. Dairy fat and meat are major dietary sources. Dietary cholesterol only raises plasma cholesterol levels by a small degree (Arnett et al., 2019). Although both high density lipoprotein (HDL) and LDL fractions increase, the effect on the total/HDL ratio is still unfavorable, but small. Evidence from observational studies conducted in several countries generally does not indicate a significant association with cardiovascular disease risk. Although meta-analyses of intervention studies differ in their findings, most associate intakes of cholesterol that exceed current average levels with elevated total or LDL cholesterol. The current literature does not support the notion that dietary

cholesterol increases the risk of heart disease in healthy individuals. The upper limit for dietary cholesterol intake has been prescribed, in most guidelines, to be less than 300 mg per day in healthy individuals and less than 200 mg per day in those with CHD, but this is a moving target as exemplified by the recent US dietary guidelines (Arnett et al., 2019). However, as endogenous synthesis is sufficient to meet the physiological needs, there is limited requirement for dietary cholesterol, and it is advisable to keep the intake as low as possible. If intake of dairy fat and meat are controlled, then there needs to be no severe restriction of egg yolk, although some limitation remains prudent. Eggs are affordable, rich in protein and micronutrients, nutrient-dense and low in saturated fatty acids. The healthy eating pattern can incorporate nutrient-dense, calorie-controlled meals with balanced nutrients and a variety of colorful vegetables and fruits (Arnett et al., 2019; Visseren et al., 2021). A recommendation that gives a specific dietary cholesterol target within the context of food-based advice is challenging for clinicians and consumers to implement; hence, guidance focused on dietary patterns is more likely to improve diet quality and to promote cardiovascular health.

Carbohydrates

High-carbohydrate diets appear to reduce HDL cholesterol levels and increase the fraction of small dense LDL, both of which may impact adversely on vascular disease (Mozaffarian, 2016). This dyslipidemic pattern is consistent with the elevation of plasma triglycerides and is typical of the metabolic syndrome (Gupta and Wood, 2019). Carbohydrate diets with high-glycemic index adversely impact glucose control, with associated changes in plasma lipids, and have been linked to an increased risk of CHD. However, a diet with moderately restricted carbohydrate intake but rich in vegetable fat and vegetable protein improves lipid profile and may lower CHD risk. WHO has advised that restriction of sugars and carbohydrates having high glycemic index are important for overall CHD risk reduction (World Health Organization, 2003). The nature of carbohydrate is of utmost importance in relation to the reduction of CHD. Non-starch polysaccharides present in intact fruits, legumes, and whole grains reduce total and LDL cholesterol and also help to improve glycemic control.

Fiber

Most soluble fibers reduce plasma total and LDL cholesterol concentrations, as reported by several trials (Mozaffarian, 2016). Fiber consumption strongly predicts insulin levels, weight gain, and cardiovascular risk factors like blood pressure, plasma triglycerides, LDL and HDL cholesterol, and fibrinogen. Several large cohort studies in the USA, Finland, and Norway have reported that subjects consuming relatively large amounts of whole grain cereals have significantly lower rates of CHD (World Health Organization, 2003). Review and meta-analyses of trials have indicated that consumption of soluble fibers reduces blood cholesterol as well as the postprandial blood glucose response (World Health Organization, 2003). It has been found that consumption of insoluble fiber from whole grains such as nuts, legumes and other edible seeds are associated with reduced risk of developing CHD.

Antioxidants

Observational studies have reported inverse associations between the frequency of CHD and dietary intake of antioxidant vitamins (Mozaffarian, 2016). Vitamin E and vitamin C in combination have shown a long-term antiatherogenic effects but these are beneficial only for people who have antioxidant deficiency or are exposed to high levels of oxidative stress such as smokers, diabetics, and elderly patients. Though several cohort studies showed significant reductions in the incidence of cardiac events in men and women taking high-dose vitamin E supplements, large clinical trials failed to demonstrate a cardioprotective effect of vitamin E (Mozaffarian, 2016). Beta-carotene supplements too did not provide protection against CHD and, in some trials, appeared to increase the risk. Foods containing high amounts of antioxidants are protective and the focus should be on increased consumption of high antioxidant containing fruits, vegetables, and nuts to reduce CHD risk (Arnett et al., 2019; Visseren et al., 2021).

Folate

The relationship of folate to CVD has been mostly explored through its effect on homocysteine, which has been incriminated as an independent risk factor for CHD. Reduced plasma folate has been strongly associated with elevated homocysteine levels and folate supplementation has been demonstrated to decrease those levels. Data from the Nurses' Health Study, in USA, showed that folate and vitamin B6, from diet and supplements, conferred protection against CHD (fatal and nonfatal events combined) and suggested a role for their increased intake as an intervention for primary prevention of CHD (Mozaffarian, 2016). Large randomized clinical trials have failed to show benefit of folic acid supplementation on incidence of cardiovascular events- CHD and stroke. Therefore, recommendations related to folate supplementation cannot be currently advocated as the results from clinical trials so far provide insufficient evidence. Meanwhile, dietary intake of folate through natural food sources may be encouraged, especially in individuals at a high risk of arterial or venous thrombosis and elevated plasma homocysteine levels.

Flavonoids and other phytochemicals

Flavonoids are polyphenolic antioxidants, which occur in a variety of foods of vegetable origin, such as tea, onions, many colored vegetables, fruits and nuts (World Health Organization, 2003; Lawler et al., 2021). Data from several prospective studies indicate an

inverse association of dietary flavonoids with CHD. The role of these and other phytochemicals (such as plant stanols and sterols), in relation to CHD, needs to be elucidated further. Recent evidence indicates that consumption of foods rich in sterols and stanols help reduce the LDL cholesterol levels with beneficial effects on apolipoprotein B/A_1 ratio, HDL cholesterol, and triglycerides. Randomized clinical studies are needed to confirm importance in CHD prevention but high intake of foods containing these substances is recommended (Arnett et al., 2019).

Sodium and potassium

High blood pressure is a major risk factor for CHD, accounting for a third of all CHD events (Gupta and Wood, 2019). The relative risk of CHD, for both systolic and diastolic BP, operates in a continuum of increasing risk for rising pressure but the absolute risk of CHD is considerably modified by coexisting risk factors (such as blood lipids and diabetes), many of which are also influenced by diet. There is a large body of epidemiological and clinical trial evidence to support reducing sodium intake to less than 2–2.5 g/day (World Health Organization, 2003). Controversy exists regarding importance of lowering salt intake to very low levels of sodium and some studies have shown increased CVD risk with very low sodium diets (<3–5 g sodium) (Gupta and Wood, 2019). Therefore, it is prudent to recommend low sodium diet with intake varying from 3 to 5 g sodium to CVD risk.

The benefits of dietary potassium, in lowering BP and reducing cardiovascular risk is undisputed. However, the specific effects on CHD risk have not been well studied. A recent Chinese study reported that replacement of dietary sodium with potassium significantly reduced incidence of stroke but benefit on CHD risk was unequivocal (Neal et al., 2021). Keeping the dietary sodium-potassium ratio low is essential, to avoid hypertension and stroke. Including enough fresh fruits and vegetables can take care of potassium intake.

Calcium

Although protective against osteoporosis, calcium supplements accelerate vascular calcification and increase mortality in patients with renal failure. The Women's Health Initiative is the largest trial of vitamin D supplementation to date and has shown no effect of vitamin D plus calcium supplementation on the risk of CHD events (Mozaffarian, 2016; Arnett et al., 2019). The pooled analysis of randomized trials has reported that calcium supplements (without co-administered Vitamin D) were associated with approximately 30% increase in the incidence of myocardial infarction whereas, there was smaller and nonsignificant association with stroke and mortality.

Specific foods

Fruits and vegetables

A systematic review reported that nine of 10 ecological studies, two of three cases–control studies and six of 16 cohort studies found a significant protective association for CHD with consumption of fruit and vegetables or surrogate nutrients (Mozaffarian, 2016). In a 12-year follow-up of 15,220 male physicians in US, men who consumed at least 2.5 servings of vegetables per day were observed to have a 33% lower risk for CHD, compared with men in the lowest category (1 serving/day). A follow-up study of NHANES, the large national survey of USA, also reported a coronary protective effect of regular fruit and vegetable intake. Persons who consumed fruits and vegetables three or more times a day were at 24% lower risk than those who consumed less than once a day. The INTERHEART and PURE studies also reported low consumption of fruit and vegetables to be a major risk factor for CHD, across all the global regions (Gupta and Wood, 2019). All the major prevention guidelines recommend intake of 300–500/day of fruits and green vegetables to reduce CHD risk (World Health Organization, 2003; Arnett et al., 2019; Visseren et al., 2021).

Fish

Multiple meta-analyses of cohort studies have suggested a CHD protective effect of fish intake (Mozaffarian, 2016; Arnett et al., 2019; Visseren et al., 2021). Compared with those who never consumed fish or did so less than once a month, persons who ate fish had a lower risk of CHD (38% lower for five or more times a week; 23% lower for two to four times a week; 15% lower for once a week; 11% lower for one to three times a month) (World Health Organization, 2003). Each 20 g per day increase in fish consumption was related to a 7% lower risk of CHD. However, the Diet and Angina Randomized Trial-2 demonstrated that advice to those with stable angina to eat fatty fish did not reduce mortality, and taking fish oil capsules was associated with debatable risk of cardiac or sudden death (Arnett et al., 2019). The apparently conflicting results of the two trials was attributed to the differential actions of n-3 fatty acids in acute and chronic conditions, and different effects of consuming fish and taking fish oil capsules (Mozaffarian, 2016; Arnett et al., 2019).

Nuts

Several large epidemiological studies have demonstrated that frequent consumption of nuts was associated with decreased risk of CHD, the best known among them being the Adventist Health Study and PREDIMED study (Visseren et al., 2021). The extent of risk

reduction ranged from 18% to 57%, for subjects who consumed nuts more than five times per week compared to those who never consumed nuts. An inverse dose-response relationship has been demonstrated between frequency of nut consumption and the risk of CHD, in men as well as in women. Most of these studies considered nuts as a group, combining many types of nuts (tree nuts-walnuts, almonds, pistachio, pecans, macadamia nuts- and peanuts). A meta-analysis reported that consumption of 50–100 g (1.5–3.5 servings) of nuts five times/week as part of a heart healthy diet with total fat content (high in mono- and poly-unsaturated fatty acids) of 35% of energy, significantly decreased LDL cholesterol levels and it is now universally recommended (Gupta and Wood, 2019; Arnett et al., 2019; Visseren et al., 2021).

Dairy products

Dairy consumption has been correlated positively, in ecological studies, with blood total cholesterol, however the influence on CHD mortality is equivocal (World Health Organization, 2003). Earlier reports from observational cohort studies have suggested that milk consumption correlates positively with coronary mortality in 43 countries and with myocardial infarction in 19 regions of Europe. On the other hand, data from Swedish population-based cohort studies and EPIC Study in Europe have reported beneficial effects of milk in reducing CHD events and mortality (Visseren et al., 2021). Milk is an important nutrient in low and lower-middle income countries and the PURE study have reported beneficial effects in reducing cardiovascular risk (Gupta and Wood, 2019). There is need for more scientific evidence as milk and its products form an important source of protein in low and lower-middle income countries and cannot be ignored.

Alcohol

The relationship of alcohol to overall mortality and cardiovascular mortality has generally been J-shaped, when studied in western populations in whom the rates of atherothrombotic vascular disorders are high (Global Burden of Disease Study 2019 Viewpoint Collaborators, 2020). The protective effect of moderate ethanol consumption is primarily mediated through its effect on the risk of CHD, as supported by more than 60 prospective studies (Mozaffarian, 2016). A consistent coronary protective effect has been observed for consumption of one to two drinks per day of an alcohol-containing beverage but heavy drinkers or binge-drinking have higher total and cardiovascular mortality than moderate drinkers or abstainers.

Composite diets

Mediterranean diet

The traditional Mediterranean diet has been described to have eight components: (1) high monounsaturated-to-saturated fat ratio, (2) moderate ethanol (wine) consumption, (3) high consumption of legumes, (4) high consumption of cereals (including bread), (5) high consumption of fruits, (6) high consumption of vegetables, (7) low consumption of meat and meat products, and (8) moderate consumption of milk and dairy products (Visseren et al., 2021). Most of these features are found in many diets in that region. The characteristic component is olive oil, and many equate a Mediterranean diet with consumption of olive oil. A secondary prevention trial of dietary intervention in survivors of a first recent MI (the Lyon Heart study), which aimed to study the cardioprotective effects of a Mediterranean-type of diet, actually left out its most characteristic component, olive oil. The main fat source was rapeseed oil. Vegetables and fruits were also increased in the diet. On a 4- year follow-up, the study reported a 72% reduction in cardiac death and nonfatal MI. The risk of overall mortality was lowered by 56% (World Health Organization, 2003).

Large cohort studies in Greece and in several elderly European population groups have also reported a protective effect against CHD and better overall survival in persons consuming a Mediterranean type of diet (Visseren et al., 2021). The protection was afforded by the composite diet rather than by any single component. Improvement in metabolic syndrome and reduction of inflammatory markers has also been observed with this diet, which may explain part of the protection against CHD (Gupta and Wood, 2019). Primary prevention PREDIMED randomized trial conducted in Spain and other Mediterranean countries has conclusively demonstrated benefits of this type of diet in prevention of CHD (Visseren et al., 2021).

Dietary Approaches to Stop Hypertension (DASH)

A composite diet, employed in the DASH (Dietary Approaches to Stop Hypertension) trials, has been found to be very effective in reducing BP in persons with clinical hypertension as well as in people with BP levels below that threshold (Mozaffarian, 2016). This diet combines fruits and vegetables with food products, which are low in saturated fats. Notably, the addition of exercise and weight loss to the DASH diet results in additional blood pressure lowering beyond the DASH diet alone, greater improvements in vascular and autonomic function, and reduced left ventricular mass. Prevention of CHD was evaluated in secondary analyses of DASH diet studies but remain inconclusive (Visseren et al., 2021).

Vegetarian diets

A reduced risk of cardiovascular disease has been reported in populations of vegetarians living in affluent countries and in case-control comparisons in developing countries (World Health Organization, 2003). Long-term vegetarians have also been shown to have better antioxidant status (as measured by the plasma ascorbic acid status) and CHD risk profile than do apparently healthy omnivores. Reduced consumption of animal fat and increased consumption of fruit, vegetables, nuts, and cereals may underlie such a protective effect. However, typical vegetarian or vegan diets per se need not be healthful. If not well planned, they can contain a large amount of refined carbohydrates and TFA, and are at times deficient in the required levels of vegetable and fruits. The composition of the vegetarian diet should, therefore, be defined in terms of its cardio protective constituents utilizing components of Mediterranean and DASH diets.

Prudent diets

In the Prudent versus Western Patterns diet study in the Health Professionals Follow-up Study in USA, a prudent diet pattern was characterized by higher intake of vegetables, fruit, legumes, whole grains, fish, and poultry whereas the western pattern was defined by higher intake of red meat, processed meat, refined grains, sweets and dessert, fries and high-fat dairy products (Mozaffarian, 2016). After adjustment for age and other coronary risk factors, relative risks, from the lowest to the highest quintiles of the prudent pattern score, were 1.0, 0.87, 0.79, 0.75, and 0.70, indicating a high level of protection. In contrast, the relative risks, across increasing quintiles of the western pattern, were 1.0, 1.21, 1.36, 1.40, and 1.64, indicating a mounting level of excess risk. These associations persisted in subgroup analyses according to cigarette smoking, body mass index, and parental history of MI. In the INTERHEART study, consumption of a prudent diet (high in fruits and vegetables) was associated with up to 30% reduced risk of MI worldwide (Gupta and Wood, 2019). PURE study also reported benefits of a healthy diet pattern consisting of unrefined cereals, fruits, vegetables, nuts and unsaturated fats (Gupta and Wood, 2019).

Japanese diet

The traditional Japanese diet has attracted much attention because of the high life expectancy and low CHD mortality rates among the Japanese. This diet is low in fat and sugar and includes soy, seaweeds, raw fish, and a predominant use of rice. Traditionally it has been high in salt, but salt consumption has recently been declining in response to Japanese Health Ministry guidelines (World Health Organization, 2003). The Okinawa diet (a prefecture of Japan renowned for high life expectancy) when compared to the Mediterranean and DASH diets was found to have the lowest fat, particularly saturated fats, which may likely be a contributing factor to the low CHD mortality rates in Japan (World Health Organization, 2003).

Comprehensive prevention

The powerful relationship of specific nutrients, food items, and dietary patterns to CHD has been persuasively suggested by observational epidemiological studies (which indicate the potential for primary prevention in populations) and conclusively demonstrated by clinical trials (which demonstrate the impact on both primary and secondary prevention among individuals). Atherosclerotic vascular diseases (especially CHD) are multifactorial in origin. Each of the risk factors operates in a continuous manner, rather than across an arbitrary threshold. When multiple risk factors coexist, the overall risk becomes multiplicative. As a result of these two phenomena, the majority of CHD events occurring in any population arise from any individuals with modest elevations of multiple risk factors rather than from the few individuals with marked elevation of a single risk factor (Gupta and Wood, 2019).

These phenomena have two major implications for CHD prevention. First, it must be recognized that a successful prevention strategy must combine population-wide interventions (through policy measures and public education) with individual risk reduction approaches (usually involving counseling and clinical interventions) (Gupta and Wood, 2019). Second, diet is a major pathway for CHD prevention, as it influences many of the risk factors for CHD, and can have a widespread impact on populations and substantially reduce the risk in high-risk individuals (Arnett et al., 2019; Visseren et al., 2021). Even small changes in blood pressure, blood lipids, body weight, central obesity, blood sugar, inflammatory markers, etc., can significantly alter the CHD rates, if the changes are widespread across the population (Gupta and Wood, 2019). Modest population-wide dietary changes can accomplish this, as demonstrated initially in Finland and now across the globe (World Health Organization, 2003; Mozaffarian, 2016). At the same time, diet remains a powerful intervention to substantially reduce the risk of a CHD-related event in individuals who are at high risk due to multiple risk factors, prior vascular disease, or diabetes.

Conclusions

A diet, which is protective against CHD, should integrate: plenty of fruits and vegetables (300–500 g per day); a moderate amount of fish (two to three times a week); a small quantity of nuts; adequate amounts of PUFA and MUFA (together constituting

approximately 75% of the daily fat intake); low levels of SFA (less than 25% of the daily fat intake); complete elimination of TFA; limited salt intake (preferably 5–7 g salt for 3–5 g of sodium per day) and restricted use of sugar and refined carbohydrates. Such diets should be culturally appropriate, economically affordable and based on locally available foods. National policies and international trade practices must be shaped to facilitate the wide availability and uptake of such diets. Nutrition counseling, of individuals at high risk, too must adopt these principles while customizing dietary advice to specific needs of the person. CHD is eminently preventable, as evident from research and demonstrated in practice across the world. Appropriate nutrition is a major pathway for CHD prevention and must be used more widely to make CHD prevention even more effective at the global level.

References

Arnett, D.K., Blumenthal, R.S., Albert, M.A., et al., 2019. ACC/AHA guideline on the primary prevention of cardiovascular disease. J. Am. Coll. Cardiol. 74, 1376–1414.

Global Burden of Disease Study 2019 Viewpoint Collaborators, 2020. Five insights from the global burden of disease study 2019. Lancet 396, 1135–1159.

Gupta, R., Wood, D.A., 2019. Primary prevention of ischemic heart disease: populations, individuals and healthcare professionals. Lancet 394, 685–696.

Lawler, P.R., Bhatt, D.L., Godoy, L.C., et al., 2021. Targeting cardiovascular inflammation: next steps in clinical translation. Eur. Heart J. 42, 113–131.

Mozaffarian, D., Katan, M.B., Ascherio, A., Stampfer, M.J., Willet, W.C., 2006. Trans fatty acids and cardiovascular disease. N. Engl. J. Med. 354, 1601–1613.

Mozaffarian, D., 2016. Dietary and policy priorities for cardiovascular disease, diabetes and obesity: a comprehensive review. Circulation 133, 187–225.

Neal, B., We, Y., Feng, X., et al., 2021. Effect of salt substitution on cardiovascular events and death. N. Engl. J. Med. 285, 1067–1077.

Sacks, F.M., Lichtenstein, A.H., Wu, J.H.Y., et al., 2017. Dietary fats and cardiovascular disease: a Presidential advisory from the American Heart Association. Circulation 136, e1–e23.

Visseren, F.L.J., Mach, F., Smulders, Y.M., et al., 2021. ESC guidelines on cardiovascular disease prevention in clinical practice: developed by the Task Force for cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies with the special contribution of the European Association of Preventive Cardiology (EAPC). Eur. Heart J. 42, 3227–3337.

World Health Organization, 2003. Diet, nutrition and the prevention of chronic diseases. Tech. Rep. 916, 1–149.

Food culture

Johanna T. Dwyer, Department of Medicine and Community Health, Tufts University School of Medicine, Friedman School of Nutrition Science and Policy, Boston, MA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of J. Dwyer, J. Freitas, Food Culture, In Encyclopedia of Human Nutrition (Third Edition), edited by Benjamin Caballero, Academic Press, 2013, https://doi.org/10.1016/B978-0-12-375083-9.00118-5.

Introduction 271
Food in cultural history 271
Food beliefs, diet, and health across the centuries 272
Influence of food and nutrition science on food culture and recommendations 272
Modeling dietary patterns 273
Association of dietary patterns with health outcomes 274
Food culture in the USA 274
From nutrients to food patterns 274
New views of healthy and unhealthy foods 275
Advent of broader concerns influencing food guidance 275
Food folklore continues 275
Evaluating food culture and nutrition science 277
Types of evidence and utility in establishing causal inference 277
Randomized double blind placebo controlled clinical trials 277
Observational studies 277
Uncontrolled clinical trials 278
Patient reports, case studies and folklore 278
Animal studies and laboratory experiments 278
Tensions persist between traditional food culture and nutrition science 278
Evaluating folklore and food culture in clinical situations 278
Need to focus and prioritize beliefs most likely to affect safety and health 278
Ephedra for enhanced athletic performance and weight reduction 279
Low fat, calcium/vitamin D supplements and post-menopausal hormone replacement therapy to prevent chronic degenerative disease (The Women's Health Initiative) 279
Guide for dealing with important folk beliefs that adversely affect health 279
Conclusion 279
References 280
Further reading 280

Key points

- Cultural history and societal traditions that are deeply embedded in beliefs about food and health based largely on deduction and causal inference from unverified assumptions remain powerful determinants of eating practices today
- Beliefs in food culture about the associations between food and health are often more persuasive than evidence-based nutrition science information
- Most discrepancies between food culture and evidence based on nutrition science should be reconciled but those that adversely affect health must be resisted and resolved

Glossary

Food culture Integrated patterns of human knowledge, beliefs, and behaviors about food that are learned shared and transmitted across generations

Foodways Customary habits and practices related to food and eating of groups of people, regions, or historical periods

Food patterns Quantities, proportions, variety, or combinations of different foods and drinks in diets and the frequency with which they are consumed

Food folklore Traditional beliefs, customs and tales about food that are preserved and shared widely among people of a culture

Food guides and recommendations Recommendations about foods, food groups, or their suggested proportions in overall diets that today are largely based on scientific evidence about the relationship between food intakes and health outcomes

Introduction

Food is an important part of human culture as well as being biologically essential. The anthropology of food is a sub-discipline that connects an ethnographic and historical perspective, including the material and symbolic importance of food, and how these concepts intersect with contemporary social issues in food production and consumption systems. Social scientists address the societal and cultural aspects of food and nutrition while biological scientists study the physical properties of foods and physiological chemistry. Food culture therefore falls at the nexus of these disciplines (Grivetti, 2003).

Food culture refers to the integrated patterns of human knowledge, beliefs, and behaviors about food that are learned, shared, and transmitted across generations. These beliefs include habitual foodways, food patterns, food folklore, and subjective psychological, social, and political world views that are constantly changing but at different rates. They coexist in society along with food guides or recommendations based on current biological knowledge of the relationships between food and health outcomes. It is not surprising that science-based recommendations about eating and health are often rejected when they fly in the face of older and larger traditional beliefs.

This article defines terms used to describe food culture, eating habits and culinary practices. It summarizes the cultural elements related to food and other factors that historically have determined what is eaten, and how they have changed over the centuries. The dynamic tension between them and eating recommendations based on induction from scientific discoveries in food and nutrition science that seek to shape eating habits today is summarized. The discrepancies between these different types of knowledge must be reckoned with by those who wish to change consumer behavior in more healthy directions, and to conclude a framework is presented for examining causal inference and reconciling food culture, food folklore with science-based food recommendations for health.

Food in cultural history

Food is physiologically necessary to sustain life by providing essential nutrients. But it also plays a larger and equally essential role in culture. Food plays potent symbolic roles in religion and cultural traditions. For example, rice has been associated with fertility in many cultures for millennia and even today rice is thrown on newly married couples. Similarly, bread has been a symbol of divinity and has played an important role in many religious services and observances.

Food choice is determined largely by human factors other than nutrient need, such as the senses, personal preferences, psychological and social attitudes toward eating and health, social structures, economics, and politics. But geographical and environmental availability, technology, and economics are also potent influences. Geography, environment, and technologies available have predominated in shaping the staple foods that are the basis of the great cuisines of the world (Grivetti, 1992, 2003, 2000). Each cuisine and dietary pattern has its preferred foods and customs (Albala, 2011). For example, in East Asia, the Japanese and many Chinese patterns that are distinct, including Szechuan, Cantonese, Mandarin, Shandong, and others. South Asian cuisines in India differ by region and are heavily influenced by religious beliefs. Western Asian Middle Eastern cuisines are also influenced by the different foods available and religion. Oceania draws on traditional foods grown in the region. Diets of the Western hemisphere include geographically available traditional foods as well as European imports. The evidence that traditional diets consumed in different regions of the world and among different ethnic groups have allowed populations to thrive and survive suggests that there are many different forms of a reasonably healthy diets.

Economics has always been a potent force on foodways and food culture. As the science of food and nutrition expanded in the late 19th and early 20th century, the seminal study of H. Seebohm Rowntree on the poor in England made it clear that their earnings were insufficient to obtain adequate diets, and that age, unemployment and illness were often the root causes of poverty, not vice-versa (Rowntree, 1901; Freeman, 2008).

Food culture is no longer limited to influences from the immediate geographical environment as it was years ago (Dwyer and Drewnowski, 2017). European eating patterns have expanded globally, and these and American variations of them now often coexist with more traditional cuisines in many countries. In affluent countries during the 20th century traditional, largely plant-based diets produced locally by the eaters themselves were increasingly replaced by diets of commercially produced prepared and processed foods The spread and mixing of cuisines between countries has not been without controversy, and there are calls for going back to traditional locally produced diets of whole foods, older “slow food” eating customs, and foodways. Others advocate for economic, esthetic, economic, technological, environmental, sustainability and other changes in food production as well as focusing on enhance health.

Food beliefs, diet, and health across the centuries

The connection between diet and health has been known for ages, although the precise constituents that led to good health proved elusive. Views of "healthy" diets have altered over time as food availability, food prices, income, and as the various preordained organizing principles to explain associations between diet and health changed. Many elaborate theories have been used to explain diet-health associations and produce dietary recommendations. For example, the ancient Greeks believed that the body was composed of four humors: blood (hot and moist), phlegm (cold and moist), yellow bile (hot and dry) and black bile (cold and dry). Health was thought to result from a balance of these humors, and illness from an imbalance. To counteract imbalances and restore health, physicians often prescribed specific foods, based on their perceived degree of heat and moisture. For example, fever, regarded as a hot and dry condition, was attributed to an excess of yellow bile. Cool and moist foods, such as cucumbers, were prescribed to treat it. In contrast, edema, a cool and moist condition, was treated with foods that were viewed as warm and dry.

Curative properties ascribed to foods and their prescription as medicines was based on these early medical theories and not on scientific evidence. In ancient Rome, cabbage was considered the perfect medicinal plant and was prescribed frequently for a wide range of ailments, including wars, deafness, and drunkenness. Many other foods were also highly regarded in ancient cultures for their therapeutic qualities, including apples, herbs, garlic, honey, milk, peppers, and others.

The hot, cold, moist, and dry properties of food were regarded as important in other ancient societies, including China, where achieving a balance between the opposing forces of yin (cold/moist) and yang (hot/dry) has guided traditional Chinese medical practice for centuries, and continues to be popular today.

Another theory used to dictate traditional dietary recommendations, The Doctrine of Signatures, based on the notion that "like cures like" was popular in the nineteenth century in many countries. Therapies were chosen by deduction and based on similarities of color, aroma, shape, and other characteristics. For example, beet juice, which is deep red, was thought to be an effective cure for blood diseases, whereas yellow plants were believed to alleviate jaundice and other liver ailments. The pungent odors of onions and garlic were thought to ward off disease, stimulate strength and bravery, arouse the libido, and banish evil spirits. Walnuts, which resemble the brain, were eaten to improve the intellect. The ginseng root, with its resemblance to the human torso, was used by the Chinese as a panacea. The common names of many herbs and botanicals reflect folklore about their curative properties (**Table 1**). For example, the word ginseng is derived from the root words gin, meaning man, and sing, meaning essence.

Influence of food and nutrition science on food culture and recommendations

The advent of the sciences of food and nutrition in the early 20th century led to a greater understanding of the associations between individual foods, diets, health. For many decades, the primary focus of nutrition scientists was on how the constituents of foods prevented disease. Using chemistry, experiments in animals, observations of humans and an inductive approach they identified nutrients, individual constituents of foods that caused deficiency disease when they were not present. Using this knowledge, they constructed dietary recommendations from the "bottom up" which included combinations of foods that contained enough of the beneficial constituents and little of those with negative effects (such as alcohol). The inductive approach was effective and while a view of food solely through a nutrient lens was reductionistic, it did permit progress to be made in conquering dietary deficiency diseases, which were due to lack of only one or a few constituents (nutrients such as vitamin D, thiamine, riboflavin, niacin, calcium, and iron) that were often present only in certain foods or food groups. More recently the same approach has been useful in detecting allergen ingredient related reactions to foods.

In Western countries this "nutrient centric" view sometimes became myopic (Fardet and Rock, 2018). It led to overly simplistic focus on "good" foods and food groups with a similar profile of essential "positive" nutrient constituents such as vitamins, minerals, protein and sufficient energy. The concept later expanded to include avoidance of "bad" foods or foods high in ingredients with "negative or unhealthy" ingredients such as fat, cholesterol, saturated fat, and added sugars that were linked to poor health

Table 1 Food names related to food folklore.

Herb (common name)	*Botanical name*	*Folklore*
Blackeye root	Tamus communis	Heals bruises, removes discoloration
Bloodroot	Sanguinaria candensis	Cures blood disorders and heart disease
Birthwort	Aristolochia longa	Alleviates complications associated with childbirth
Eyebright	Euphrasia officinalis	Cures eye disorders
Ginseng	Panax quinquefolium	General panacea for humans
Heartsease	Viola tricolor	Relieves heart ailments
Liverwort	Anemone hepatica	Relieves liver disorders
Lungwort	Sticta pulmonaria	Cures pulmonary diseases
Maidenhair fern	Asplenium trichomanes	Promotes hair growth and prevents balding
Snakeroot	Aristolochia serpentaria	Antidote for snake bites
Spleenwort	Asplenium	Remedy for disorders of the spleen

outcomes in epidemiological studies. Groups of foods with similar nutritional characteristics were judged "good" or "bad". However, reductionist explanations like these run the danger of oversimplifying complex food health relationships, and are limited by contemporary scientific understanding, while science constantly changes. For example, in the 1960s and 1970s eggs, a low-cost source of high-quality protein, were regarded as "unhealthy" foods because they contained cholesterol, which scientists incorrectly assumed to be the major cause of increased serum cholesterol and coronary artery disease risk. Recently, sugar sweetened beverages and added sugars have been incorrectly assumed to have constituents with particularly obesogenic properties rather than the true causes, their high calorie content, high amounts consumed and high frequency with which they are consumed.

As the chronic degenerative diseases rose in prevalence, causal inference about diet became more difficult, because the deficiency disease hypothesis no longer applied. Many nutrients and other food constituents alter risk, but so do many other environmental and genetic factors. As for obesity, the causative component (food energy) is in many foods. Inductive approaches to tackling these diseases are more complicated. These realities have moved nutritional thinking from single nutrients or foods toward comprehensive characterizations of dietary patterns consisting of many foods in various amounts and their associations with health.

Modeling dietary patterns

Human beings require nutrients and not specific foods, food groups or dietary patterns to be healthy, but they must eat foods to obtain these nutrients. For this reason, they need guidance on food selection and overall intakes since it is the quality of the total diet rather than the nutrient content of individuals foods that determine associations with health (Tapsell et al., 2016). Overall dietary patterns explain more of the impact of diet on non-communicable disease risk than single nutrients, dietary ingredients, foods, or food groups, as they include the synergistic, interacting influences of nutrients non-nutrient bioactive constituents and energy intakes. Foods must be acceptable if they are to be consumed and different foods are usually consumed together. Therefore, dietary patterns are important. Food pattern modeling is a method for evaluating the impact of specific changes in amounts or types of foods and beverages in a dietary pattern. The ideal result is the healthiest acceptable dietary pattern that is consistent with nutrient recommendations.

Modeling studies to develop dietary patterns use either an a priori or an a posteriori approach to describe foods and dietary patterns associated with health outcomes. The a priori method, determinines how food selections fit in meeting preexisting recommendations. This approach begins with the nutrient recommendations (e.g., Dietary Reference Intakes) or types and amounts of various foods or groups of foods that are declared by authorities to be associated with good health outcomes (such as the Healthy Eating Index [HEI] or the Dietary Approaches to Stop Hypertension [DASH] diet). Recommendations to modify existing eating patterns in more healthful directions are then made by adding foods that are sources of missing nutrients (protein, vitamins, fiber), and decreasing those with ingredients known to be associated with ill effects (high calories, alcohol, sugar, saturated and trans-fat, salt) while also meeting human nutrient needs.

A second modeling approach is a posteriori. It separates out the components in epidemiological datasets that include dietary intakes and health outcomes by use of mathematical techniques (such as linear programming, principal components analysis or cluster analysis) to disaggregate the effects of different foods and to optimize combinations of foods that show strong health associations.

The results of such modeling are usually presented as groups of foods found to be and labeled as beneficial or prejudicial to health. Patterns associated with beneficial health outcomes are given names such as healthy, prudent, plant-based, Mediterranean-like, Healthy Eating Index pattern [HEI], healthy vegetarian and Dietary Approach to Stop Hypertension [DASH] and have particularly "healthy halos" presently, but they are heterogenous in the foods they contain. Some of the labels are not new; various vegetarian and Mediterranean diets have been described since antiquity (Grivetti, 2001). Groups of foods associated with poor outcomes are often referred to as "Western", "Standard American diet" "unhealthy" "ultra-processed" "high fat, low fiber high sugar", or "junk food" patterns. However, descriptions of the unhealthy "Western" diet are quite disparate, for example: "contain high amounts of processed foods, red meat, high-fat dairy products, high-sugar foods, and pre-packaged foods"; "high saturated fat, red meat, empty carbohydrate, junk food, low fresh fruit and vegetable, low whole grain, seafood and poultry", "inadequate fruits, vegetables, whole grains, legumes, fish, low fat dairy and excessive refined and processed food, alcohol, salt, red meat, sugary drinks, snacks, eggs, and butter", and "low potassium, high sodium, high fat and simple carbohydrate foods".

Although the groups of beneficial and prejudicial foods to health are often given similar labels, the foods comprising the groups are confusing, as they are often inconsistent, rather than constituting a standard set of foods from study to study (Boushey et al., 2017). This makes aggregation of studies and generation of recommendations for consumers on what should be eaten or avoided difficult. Efforts are now being made to obtain more agreement on what exactly they consist of (Sotos-Prieto et al., 2021).

A third, mixed, approach is to engineer entirely new dietary patterns from the bottom up by modifying existing patterns to include nutrients and other constituents known to be associated with good health (or some other desirable characteristic, such as sustainability); or by recommending entirely new patterns associated with the most positive health outcomes found in epidemiological studies, while ensuring that nutrient needs are covered. Food acceptability is often inferred from food consumption surveys and, to make the recommended dietary pattern more realistic, more modeling may be needed to minimize the changes from current consumption. Modelers may also add other factors they regard as desirable, such as cost, sustainability, acceptable types of food (plant only), food production (organic) or processing (non or lightly processed). However, each additional factor further constrains

the number of foods that are deemed appropriate, potentially limiting the acceptability of the pattern. It then remains to be seen if the entirely new recommended patterns are translatable.

Care must be taken in interpreting results, especially when most studies upon which findings are based are observational in nature and residual confounding by other factors influencing outcomes cannot totally be removed. It is often difficult to identify what the most important food groups, foods or nutrient are in the "healthy" dietary pattern that are associated with favorable health outcomes. Some foods may be more important to health than others, and others may have no relationship whatsoever, and yet their relative effects may be unknown. Also, recommended dietary patterns might not have the same results in populations eating very different diets and living in radically different circumstances than those in the populations studied. Rather, dietary patterns may be time and place dependent.

Association of dietary patterns with health outcomes

A recent systematic review of dietary patterns for developing the Dietary Guidelines for Americans 2020–25 illustrates the utility of dietary pattern modeling in crafting dietary recommendations. It examined a single randomized clinical trial and 152 observational studies of dietary patterns consumed for their effects on all-cause mortality (English et al., 2021). Similarities and differences in diets eaten were described by the foods and beverages eaten rather than by subjective labels. The results were examined a priori (e.g., based on scientific consensus of what a healthy pattern was, such as the HEI) and a posteriori (by what factors explained variation in patterns or aggregated individuals into nonoverlapping groups). Dietary patterns characterized by greater consumption of vegetables, fruits, legumes, nuts, whole grains, unsaturated vegetable oils, fish, and lean meat or poultry were associated with lower risk of all-cause mortality in adults and older adults. Groups with greater adherence to the "healthier" patterns also had lower all -cause mortality.

After taking confounders into account, the "healthier" patterns continued to show beneficial effects. The study populations were from countries with a high Human Development Index. In the systematic review, further analysis showed that, with respect to individual food groups, healthy patterns included relatively low intakes of red and processed meat, high fat dairy, or refined carbohydrates or sweets (English et al., 2021).

Food culture in the USA

As the distinguished food anthropologist Louis Grivetti pointed out years ago, the patterns of food that people eat depend on availability, familiarity, and selection, or "AFS". US Dietary recommendations focus particularly on guiding eaters to selecting foods in accordance with them. Human beings differ in the ways in which they fulfill their common nutrient needs through the foods they eat, how they eat them, and what reasons the give for doing so. Therefore, each population's preferences must be considered as unique, and what is acceptable in one society may not be in another. This section focuses on food culture as it stands today in the USA, although some of the same forces may also be operating in other countries.

From nutrients to food patterns

In early 1900s the concern was largely about getting enough food to eat, with less attention to what was eaten. Food availability data in the USA show that, over the past century, profiles of dietary intakes in essential nutrients are probably healthier than ever before. However, the amounts of food eaten are greater and, in the sedentary US population, overweight and obesity have resulted. Starting in the early 20th century, American nutrition scientists began providing dietary recommendations based on what was known about nutrients that fit consumers' economic resources, with additional advice about avoiding food waste, particularly during World War I (WWI). Food guides focused mainly on getting enough nutrients to prevent deficiency disease, and stressed commodities that were good sources of certain nutrients like milk, meat, grains, fruits, and vegetables while the remaining items in the diet to fill energy needs were left up to individual preferences. Carbohydrates, especially sugar and starchy foods were regarded as fattening and not included in advice. A relatively high fat low fiber diet was common. At the federal level food plans for citizens on different budgets were developed by the US Department of Agriculture.

During WWII there was continuing emphasis on both animal and plant foods high in protein, milk, and vegetables, fruits, and grains. By the 1950s and 1960s the notion of a "heart healthy" diet had emerged with the endorsement of the American Heart Association that stressed a low saturated fat, low cholesterol diet, with limits on meats and eggs high in these constituents. By the 1970s, federal food guidance gradually changed from giving recommendations solely on what foods to include (the Basic Four food groups) to limiting food energy intakes, with more emphasis on total diets.

The first edition of the Dietary Guidelines for Americans focused on ingredients to avoid (saturated and later trans fats, sugar, salt) as well as on foods to include. Over time, there was increasing emphasis on controlling how much food was consumed, and various icons were used to illustrate the concepts of variety, balance, and moderation (Dwyer, 2019). Starting in the 1990s US federal policies began to present dietary recommendations as food patterns rather than emphasizing nutrients or "basic" foods. With the advent of the internet, federal food guidance that included applications which could generate eating plans personalized to one's own calorie and nutrient needs became available. The 2015–20 Dietary Guidelines Advisory Committee focused on 3 food patterns (the Healthy US-Style, Healthy Mediterranean-Style, and Healthy Vegetarian Eating Patterns) as well as emphasizing

a variety of foods high in nutrient (but not caloric) density, in reasonable amounts, and tools to help consumers follow a healthy eating pattern. The committee also emphasized the need for physically active lifestyles, food safety and food sustainability. Sustainability, defined as evaluating the environmental impacts of food sources was later deemed to be out of scope by the administration and so recommendations were not presented in the final guidelines. The latest version of the Dietary Guidelines for Americans 2020–25 provides the same three examples healthy eating patterns. In addition, consumers are urged to meet food group needs with nutrient-dense foods and beverages and to stay within calorie limits. Limiting foods and beverages higher in added sugars, saturated fat, and sodium, and limiting alcoholic beverages is also stressed.

New views of healthy and unhealthy foods

As in the past, there is much consumer interest today on "healthy" and "unhealthy" foods. Defining what is meant by healthy and unhealthy continues to be highly culturally determined and controversial. Traditionally, government guidance has singled out specific food groups such as fruits and vegetables or whole grains as "healthy" because they are rich in essential nutrients and low in ingredients like salt, sugar, or saturated fat that are in excess in population diets. At present, the Food and Drug Administration estimates that about 15% of the foods on the market would be eligible for a voluntary "healthy" label under existing regulations, although only about 5% of products use the label. Federal guidelines are more selective in labeling foods as "unhealthy" and tend to use the term sparingly, only when the entire population is at risk, as when foods known to be contaminated by food borne pathogens. Definitions of "unhealthy" in government regulations assume that the foods themselves are neither inherently healthy nor unhealthy but that they may become so in particular contexts or for particular people, such as when foods are high in constituents like salt and they are consumed in excess, especially by those at risk of cardiovascular disease. Today advocates of healthy eating are increasingly assigning "bad" as well as "good" labels to individual foods, based on their content of salt, saturated fat, cholesterol, added sugars or others which they view as associated as having adverse effects on population health. However, unlike the lack of nutrients or the presence of food borne pathogens that always pose risks of ill effects, individual food ingredients are more time and context specific in their health effects, as lifestyles and eating patterns change. For example, a food such as butter (high in saturated fat) is associated with coronary heart disease in the USA whereas ghee (also high in saturated fat) is not, but the reverse might be true in India, where ghee consumption is high and butter consumption is lower. Similarly sugar sweetened beverages are highly associated with obesity in the USA's sedentary population whereas the same amounts might not be in a lean, physically active population in a low-income country with low energy intakes. Therefore, special care needs to be taken in labeling of foods as "unhealthy" because they are highly population and time context specific.

Advent of broader concerns influencing food guidance

Some nutrition scientists and other biological and social scientists advocate for dietary recommendations that go beyond nutrition science's traditional domain in providing recommendations about foods to eat and diets based primarily on nutrient needs.

Current recommendations include the food's environmental sustainability, food security, degree of food processing, location of food production and general concerns and causes. There is growing pressure to include some of these broader concerns in recommendations. For example, the United Nations' sustainable development goals emphasize improvements in health, society, economics, and the environment. These goals have led to recommendations for sustainable healthy diets and foods systems to integrate these social dimensions into food and nutrition patterns, beyond most current recommendations. One example is the recently published *Eat Lance*t report, the first full scientific review of what constitutes a healthy diet from a sustainable food system, and which actions can support and speed up food system transformation (Willet et al., 2019). Meat consumption has received much attention as a negative force (Godfray et al., 2018), although it is not likely to be easy to change preferences for it (Ruckert-John, 2017). A companion Lancet Commission report on obesity and undernutrition also stresses the need for changes in dietary patterns and lifestyle toward more sustainable food systems (Swinburn et al., 2019). They are useful beginnings at a more holistic view of the role of food and nutrition in planetary health. However, the devil in the details includes ensuring that all the assumptions from the various disciplines are based on sound evidence, and that there are ways to operationalize these various dimensions into culturally acceptable eating patterns in free societies (Nicholls and Drewnowski, 2021).

Food folklore continues

Scientific concepts about food and nutrition that have developed over the past two centuries are relative newcomers to human thought about food and health. Therefore, although some folk beliefs about food have been discarded over the years, it should come as no surprise that other aspects of food folk beliefs continue to be passed down from generation to generation, and new ones have been introduced. In the USA, there have been ebbs and flows in specific foods or ingredients that were considered healthy or unhealthy. Commonly held food cultural beliefs today are a mélange of traditional beliefs, folklore, partially understood nutrition science concepts, food advertising, marketing efforts and public health campaigns (**Table 2**). Traditional food culture beliefs continue. Many of the highly touted "healthy foods" of today like chocolate have long and storied pasts (Dillinger et al., 2000). For example, great emphasis is placed on the qualities of protein foods and supplements, just as was the case over 100 years ago (O'Hagan, 2021) Athletes have always had their special foods and dietary regimes (Grivetti and Applegate, 1997; Applegate and Grivetti, 1997).

Table 2 Food culture: traditional and new folklore and beliefs about specific foods.

Food	*Some traditional beliefs*
Fruits	
Apple	Preventive: "an apple a day keeps the doctor away". Also thought to prevent caries and tooth decade
Blueberries	Cures kidney and urinary-tract ailments, improves memory and vision
Citrus fruits	Prevent scurvy, cure common cold
Cranberries	Prevent or cure urinary tract infections
Grapefruits	Burns calories, dissolves fat, aids in weight loss, avoid when taking medications
Raspberries	Raspberry leaf tea promotes labor contractions and aids in childbirth
Vegetables	
Beets	Cure for anemia, helps build iron rich blood
Carrots	Good eyesight
Celery	Promotes weight loss
Garlic	Stimulates digestion, inhibits germs, lowers cholesterol and blood pressure
Lettuce	Induces sterility
Onions	Cooked onions cure the common cold, "good for the heart"
Peppers	Cure heartaches
Potatoes	Cure for impotence scurvy and soothe and soften the skin, but are fattening
Spinach	Builds strong muscles
Grains	
Bread	Fattening
Flaxseed	Cures constipation, prevents cancer, lowers cholesterol
Oats	Oatmeal and oat bran prevent heart disease
Dairy	
Milk	Raw milk helps the immune system, milk causes mucus and phlegm, mothers who drink a lot of milk and dairy have colicky babies, calcium in milk causes kidney stones
Yoghurt	Prevents vaginal yeast infection, cures vaginitis, constipation, and diarrhea
Meat	
Beef	Beef protein makes muscles stronger, red foods like beef cause high blood pressure
Chicken soup	Cures the common cold
Eggs	Raw eggs build muscle, brown eggs are healthier than white eggs, eggs cause heart disease
Legumes	Beans are laxative
Seafood	Fish is brain food, is good for the heart and prevents heart attacks, oysters increase sexual potency
Fats, sweets and alcohol	
Olive oil	Protects against breast cancer and heart disease
Cod liver oil	Relieves rheumatism, aching muscles, stiff joints and prevents rickets
Sugar	Causes hyperactivity, excess causes diabetes and heart disease
Honey	Natural and does not raise blood sugar, mixed with water cures colic
Chocolate	Causes acne, prevents heart disease
Salt	Sea salt is healthier than table salt, salt tablets prevent muscle cramps, and a no salt diet will protect against high blood sugar
Alcohol	Helps warm the body in cold weather, is a sleep aid if consumed before bedtime, red wine is good for the heart, brandy cures a cold, drinking alcohol with raw oysters makes them safe and free of food borne infection
Other	
Processed and ultra processed foods	Cause many chronic diseases and obesity
Organic foods	Healthier than conventionally grown foods
Natural foods	Healthier than processed foods
Additive free foods	Healthier than foods with additives

Complementary and alternative medicine of the 21st century includes foods with special properties, now called functional foods, nutraceuticals, and dietary supplements. Food folklore today also involves views about the healthfulness of the new alternatives for milk, meat, and other animal foods, dietary supplements, organic products, functional foods, "unprocessed foods" as well as larger societal issues related to climate change, international trade, sustainability, and the environment. At the same time, new demons are emerging. For example, it has been urged that many or most "ultra-processed" foods be abandoned, including advice in some documents published by the United Nations (Vidal et al., 2021). Some critics in the media, humanities and social sciences regard the food system as irrelevant, and their views have a large following (Bittman, 2021; Moss, 2021) Food cultural beliefs today are not only spread from one generation to another by word of mouth, but by the mass media, the Internet, Twitter, and other outlets. These influences are reflected as dietary recommendations. Some of the beliefs have a scientific basis, but many are unsupported by, or opposed to, current scientific findings. Pseudoscience, rather than sound, verifiable, and reproducible evidence, still provides the basis for much of today's food culture. Food, nutrition, and biomedical professionals must be knowledgeable about current beliefs, as these influence popular views about diet-health relationships and eating habits.

Evaluating food culture and nutrition science

Types of evidence and utility in establishing causal inference

Although the combined historical experience of various cultures on the optimal relationship between food, diets, and health is helpful by itself, it is insufficient to determine whether relationships are valid or not. Single studies are inadequate to demonstrate cause-effect relationships and it is important to consider not only the strength of studies, but the type and quality of the research available. When many different types of evidence are supportive of a causal relationship, the weaknesses of individual studies are mitigated and causal inference is strengthened. In determining the validity of food folklore or traditional and complementary medicine, not only the totality of the evidence but also the type and quality of available evidence are important. The strength of the association between eating a food (the cause) and a health outcome (the effect) can be ranked according to the type of evidence presented (Table 4). The best evidence comes from studies that have the most control over the treatment being evaluated while eliminating other factors that may suggest an effect was present, when really it was not.

Randomized double blind placebo controlled clinical trials

When several randomized double-blind placebo-controlled trials show a relationship between a specific food and a health effect, the evidence of a cause-effect relationship is considered to be very good. These studies exert rigorous control over the claims or treatment being evaluated and over the people who are subjected to it (by randomization) and the assumptions of both the experimenters and study participants (by placebos and blinding). Other types of evidence and studies are lower in the hierarchy because they are not as definitive. Although randomized double-blind placebo controlled clinical trials are considered "the gold standard" for determining diet-health relationships, such studies are rarely available for many nutrition questions. Lesser levels of evidence must usually be used. Also, randomized, double-blind clinical trials are usually too small to assess all potential adverse effects—observational studies are better at doing this, as was the case in demonstrating the adverse effects of ephedra for weight loss.

Observational studies

Human studies that involve observation rather than direct intervention provide evidence that is satisfactory but less conclusive than randomized, controlled double blind clinical trials. These studies are designed to test a relationship between an exposure of interest and a health outcome. They include both cohort studies (prospective) and case-control studies (retrospective). In a prospective study, a group exposed to the treatment of interest and an unexposed group are followed forward in time. The health outcomes in both groups are observed and evaluated after controlling for confounding factors with the use of statistics. In contrast, retrospective studies compare individuals who have already developed an outcome of interest (case) against those who have not (control). Factors contributing to the development of the outcome are then determined by looking backward in time Because observational studies cannot be precisely controlled, it is more difficult to establish cause and effect. However, when confounding factors can be

Table 3 Food culture: separating food culture lore and food and nutrition science.

		Scientific evidence	
		Strong	*Weak*
Food culture lore	Strong	Fact	Undetermined
	Weak	Emerging science	Untruth (fiction, "alternative fact")

Table 4 Ranking the quality of scientific evidence about food by type of study.

Highest quality evidence
Randomized double-blind placebo controlled trials
Observational studies
Prospective
Retrospective
Uncontrolled clinical trials
Patient reports/case studies
Non-human trials
In vitro studies
In vivo studies
Lowest quality evidence

adequately controlled for these studies provide suitable evidence to support diet health relationships, They are useful for discovering adverse events or beneficial effects that are attributable to the treatment.

Uncontrolled clinical trials

Clinical studies in which everyone is treated, in which only those who ask for the treatment are treated, or in which some are treated based on unsubstantiated clinical convictions are suspect. In such studies, no randomization occurs and neither the researcher nor the participant is blinded to the treatment. Therefore, it cannot be determined whether the treatment is actually the cause of the observed results or whether biased convictions of either or both the experimenter and study participants are falsely contributing to the results. Better evidence is needed before it can be stated with assurance that folklore based on such observations is true.

Patient reports, case studies and folklore

Even weaker is human evidence purporting to show cause and effect that comes from single medical case reports or anecdotal evidence. This type of evidence is also biased because those who experience success from the treatment are much more likely to report their stories than those who do not.

Animal studies and laboratory experiments

Nonhuman studies involving living animals (in vivo studies) or tissue culture (in vivo studies) are useful in providing information on the possible mechanisms of action, biological plausibility, dose response and action of a treatment. However, their ability to predict outcomes in humans is poor. Therefore, these studies are unconvincing, by themselves, of effects in humans and should be used only to support other types of evidence.

Tensions persist between traditional food culture and nutrition science

There is a dynamic tension between folklore from food cultural traditions and evidence-based nutrition science conclusions about the associations between food and health. Food cultural beliefs cannot be taken as fact without evidence to support them. It is important to review the types and quality of the various studies used to establish causal inference between a food or food constituents and health (Table 4). Ideally, the best way to evaluate whether there is a health effect is to perform a systematic evidence-based review of many randomized double-blind placebo controlled clinical trials, and to examine the results of meta-analyses and other studies when all such reviews produce comparable conclusions of an effect, the hypothesized connection is likely to be justified. Comprehensive reviews of observational studies are also useful but causal inference is weaker because many residual factors associated with the effect cannot be controlled.

Comprehensive evidence-based reviews of food cultural beliefs are needed when they have strong health implications for large populations, especially when they involve questions of life or death, and might lead to very large costs. Because they are time consuming and require much expertize and expense, they have been done only for a few important questions. For example, the National Institutes of Health (NIH) sponsored reviews of obesity treatments. The Agency for Health Policy Research on Quality sponsored reviews of several dietary constituents. The American Institute of Cancer Research has conducted many studies of dietary constituents and their relationship to cancer to substantiate their dietary recommendations, the Cochrane Collaborative has carried out reviews of many diet-disease associations, and the American Academy of Nutrition and Dietetics has done reviews of various dietary treatments. However, few of these reviews deal with food cultural beliefs.

Evaluating folklore and food culture in clinical situations

Need to focus and prioritize beliefs most likely to affect safety and health

It is impossible to amass evidence that refutes all possible food culture beliefs that may be false. Thus, prioritization is necessary to focus on the beliefs that, if followed, are likely to be serious threats to the health not only of a few individuals but to the health of the public. Most folk beliefs are harmless and scientifically unsupported, but some could endanger health (Tables 2 and 3). When food cultural belief and scientific evidence are both strong and agree, eating behavior will be in line with the science. For example, consider the advice to "eat for two" given to pregnant women (note: not eat for two *adults*, however!). Similarly, when both the food folklore is strong ("chamomile tea helps you sleep") but the scientific evidence opposed to it indicates that there is likely little effect on health, few problems are likely to arise. However, when strongly held folklore ("raw milk is healthier than pasteurized milk and improves immune function") is opposed by extensive scientific evidence, action is necessary because otherwise health may be harmed. A recent example was that vitamin supplements could prevent COVID 19 and make vaccination unnecessary. The situation is more complicated when there is strongly held belief about the health connection of a food or diet, but the scientific evidence is weak, uncertain, legally unconvincing, or "emerging", that is, requiring more research before an association can be established or rejected.

Examples The following examples illustrate how important it is to consider the safety of substances, to conduct investigations, and to act upon them.

Ephedra for enhanced athletic performance and weight reduction

One recent example of a dangerous belief in food culture was that ephedrine alkaloids in dietary supplements were safe and effective performance enhancers useful for promoting rapid weight loss. In the late 1990s the prevalence of use was high, and major adverse events, including death, were reported. A randomized clinical trial would be unethical, given the likelihood of harm, and a massive effort was undertaken, instead, to evaluate the evidence linking the supplements with health effects in a timely fashion. The findings resulted in a ban on ephedra in dietary supplements by the US Department of Health and Human Services (Shekelle et al., 2003). However, the costs of such a review were large. From a public health standpoint, only beliefs likely to have major and negative impacts on public health are worth tackling by engaging in such an elaborate systematic evidence review process.

Low fat, calcium/vitamin D supplements and post-menopausal hormone replacement therapy to prevent chronic degenerative disease (The Women's Health Initiative)

Testing a belief with a well-designed randomized clinical trial is very expensive, and results are often nuanced, but benefit to the public's health can be large. For example, popular beliefs in the early 1990s included that post-menopausal estrogen replacement therapy prevented heart disease, that low fat diets prevented cancer, and that calcium-vitamin D supplements prevented osteoporosis. These assertions were all tested in the Women's Health Initiative (WHI). Results showed that the use of estrogen plus progestin hormone therapy after menopause increased (not decreased) risk for heart disease, stroke, blood clots, breast cancer, and dementia. It is now apparent that hormone therapy—estrogen plus progestin or estrogen alone—should not be used in postmenopausal women to prevent heart disease or to lower cholesterol levels (Rossouw et al., 2002). Hormone therapy is still an option for some women to help relieve moderate to severe symptoms that occur early in menopause, but effects continue to be mixed, depending on age, and may include increased risk of breast cancer in older women (El Khoudary et al., 2020, Vinogradova et al., 2020). The WHI dietary modification trial found that a low-fat diet did not significantly reduce the risk of breast cancer, heart disease, or stroke, nor did it reduce the risk of colorectal cancer. However, the low-fat diet did reduce the risk of ovarian cancer. The WHI calcium/vitamin D trial showed that calcium and vitamin D supplements provided modest benefit in preserving bone mass and preventing hip fractures in older women and certain other groups, but did not prevent other types of fractures or colorectal cancer. Long term follow-up suggests mainly positive effects of the dietary portions of the clinical trial (Prentice et al., 2021).

Guide for dealing with important folk beliefs that adversely affect health

When food folklore in a culture is such that it poses a potential risk to patient health, it is important for health professionals to evaluate the evidence, but to also use their clinical judgment and communications skills to relate their findings to clients and patients. The clinical realities require the same intellectual elements involved in evaluating causal inference in biological research but the approach must be tailored to engage the patient in the dialog and stimulate action and understanding. One method for evaluating and resolving the issues is the "6Rs" method, provided below:

Report It is important that the clinician relate to the patient and establish two way communications to learn more about the folk belief, the strength of the individual's conviction about it, and whether it is likely to jeopardize the patient's treatment and health. Active listening on the part of the health professional will make it more likely that the patient will listen, understand, accept, and follow recommendations.

Review It is also important for the health professional to review all the evidence surrounding the food belief and safety, in particular. This is important because folk remedies and alternative medicines are often self-administered, without the assistance of physicians. If there is evidence that implementation of the belief will be hazardous to the patient's health, it must be discouraged.

Recall The next step is to collate the evidence that is available.

Relate The evidence must be communicated to the patient. Common sense is vital to fit the information to the patient's realities.

Recommend In crafting recommendations, it is important to consider the importance, feasibility, and effectiveness for the patient. The information must be individualized to fit the patient's problem and needs.

Revise and Re-evaluate Finally, the clinician's recommendation is not the last word. Although a relatively rapid response is usually required in a clinical situation, the ongoing relationship with the patient permits follow up after the initial encounter. Also, science itself constantly changes, so re-evaluation is often necessary.

Conclusion

Food culture and food traditions vary widely throughout the world. People connect with their cultural or ethnic groups through similar food patterns, and food is often a means of retaining cultural identity. Food is represented, viewed, and perceived differently throughout the world and in each country eating and diet have different psychological and social rationales and meanings. Food culture in the future will continue to change and may evolve to be more science based. As the constraints of geography, environment and economics lessen, the influence of other forces, including information about food and nutrition science may come to play a larger role in everyday food related decisions (Grivetti, 2000). However, it is certain that people will continue to select the foods

to eat, and their diets based on many cultural influences. The ultimate verdict of what people regard as acceptable food choices will continue to be dynamic, negotiated, and changeable compromises among religious and philosophical prescriptions, social pressures personal predilections and many other factors (Herzfeld, 2021). Food and nutrition scientists and clinicians can hope that their recommendations for healthy eating will eventually become part of the larger food culture. However, they should not be surprised if other beliefs and influences also continue to play a role in determining what people decide to eat. Biologists and social scientists alike should also remember that even what they regard as their most objective and evidence based recommendations may also include subjective judgements derived from their own interpretations of food culture.

References

Albala, K., 2011. Food Cultures of the World Encyclopedia, 4 vols. Santa Barbara Greenwood Press, p. 1434. ISBN 13:9780313376269 ISBN 10:0313376263.

Applegate, E.A., Grivetti, L.E., 1997. Search for the competitive edge. A history of dietary fads and supplements. J. Nutr. 127, 869S–873S. Symposium: Nutrition and Physical Performance: A Century of Progress and Tribute to the Modern Olympic Movement.

Bittman, M., 2021. Animal, Vegetable, Junk: A History of Food, From Sustainable to Suicidal. Harcourt, Houghton, Mifflin, Boston, New York, p. 364. ISBN: 9781328974624.

Boushey, J.C., Abel, B., Corpuz, G., Spoden, M.D.V., 2017. Dietary patterns: what's in a name? Nutr. Today 52, XXX.

Dillinger, T.L., Barriga, P., Escarcega, S., Jimenez, M., Lowe, D.S., Grivetti, L.E., 2000. Food of the gods—cure for humanity? A cultural history of the medicinal and ritual use of chocolate. J. Nutr. 130 (Suppl. 8S), 2057S–2072S.

Dwyer, J.T., Drewnowski, 2017. Chapter 1. Overview: food and nutrition security. In: Biesalski, H.K., Drewnowski, A., Dwyer, J.T., Strain, J.J., Weber, P., Eggersdorfer, M. (Eds.), Sustainable Nutrition in a Changing World. Springer International Cambridge, pp. 3–24.

Dwyer, J.T., 2019. The importance of dietary guidelines. In: Reference Module in Food Science. Elsevier, London. https://koi.org/10.1016/8978-0-08-100596-5.22519-5.

El Khoudary, S.R., Aggarwhal, B., Beckle, T.M., Hodis, H.N., Johnson, A.E., Langer, R.D., Limacher, M.C., Manson, J.E., et al., 2020. Menopause transition and cardiovascular disease risk: a scientific statement from the American Heart Association. Circulation 142 (25), e506–e532.

English, L.K., Ard, J.D., Bailey, R.L., et al., 2021. Evaluation of dietary patterns and all-cause mortality: a systematic review. JAMA Netw. Open 4 (8), e2122277. https://doi.org/10.1001/jamanetworkoen.2021.22277.

Fardet, A., Rock, E., 2018. Perspective: reductionist nutrition research has meaning only within the framework of holistic and ethical thinking. Adv. Nutr. 9, 655670.

Freeman, M., 2008. Seebohm Rowntree and secondary poverty 1899–1954. Econ. Hist. Rev. 64 (4), 1175–1194.

Godfray, H.C.J., Aveyad, P., Garnett, T., Hall, J.W., Key, T.J., Lorimer, J., Pierre Lusbest, R.T., Scarborough, P., Springmann, M., Jebb, S.A., 2018. Meat consumption, health, and the environment. Science 361, eaam5324.

Grivetti, L.E., Applegate, E.A., 1997. From Olympia to Atlanta: a cultural-historical perspective on diet and athletic training. J. Nutr. 127, 860S–868S. Symposium: Nutrition and Physical Performance: A Century of Progress and Tribute to the Modern Olympic Movement.

Grivetti, L.E., 2001. The view from antiquity. Ancient Greeks and Romans. In: Matala, A.-L., Zampelas, A., Stavrinos, V., Wolinsky, I. (Eds.), The Mediterranean Diet. CRC Press, Boca Raton, Florida, pp. 3–30.

Grivetti, L.E., 1992. Nutrition past—nutrition today: prescientific origins of nutrition and dietetics. Part 4. Aztec patterns and Spanish legacy. Nutr. Today 27 (3), 13–25.

Grivetti, L.G., 2000. Nutritional geography: history and trends. Nutr. Anthropol. 23 (2), 1–16.

Grivetti, L.E., 2003. Cultural Aspects of Nutrition. The Integration of Art and Science. The Oxford Brookes University Lectures. Information Press, Oxford, England, p. 275.

Herzfeld, M., 2021. Avoidances, and transgressions: agency, religiosity and moralism in food and politics. Food Cult. Soc. https://doi.org/10.1080/15528014.2021.1958468.

Moss, M., 2021. Hooked: How Processed Food Became Addictive Penguin. Random House, New York.

Nicholls, J., Drewnowski, J.A., 2021. Toward sociocultural indicators of sustainable healthy diets. Sustainability 137226. https://doi.org/10.3390/su13137226.

O'Hagan, L.A., 2021. Flesh formers or fads? Historicizing the contemporary protein enhanced food trend. Food Cult. Soc. https://doi.org/10.1080/15528014.2021.

Prentice, R.L., Howard, B.V., Van Horn, L., Neuhouser, M.L., Anderson, G.L., Tinker, L.F., Lampe, J.W., Raftery, D., Pettinger, M., Aragaki, A.K., Thomson, C.A., Mossavar-Rahmani, Y., Stefanick, M.L., Cauley, J.A., Rossouw, J.E., Manson, J.E., Chlebowski, R.T., 2021. Nutritional epidemiology and the women's health initiative: a review. Am. J. Clin. Nutr. 113 (5), 1083–1092. https://doi.org/10.1093/ajcn/nqab091.

Rossouw, J.E., Anderson, G.L., Prentice, R.L., LaCroix, A.Z., Kopparberg, C., Stefanick, M.L., Jackson, R.D., Beresford, S.A., Howard, B.V., Johnson, K.C., et al., 2002. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. JAMA 288, 321–333. https://doi.org/10.1001/jama.288.3.321.

Rowntree, B.S., 1901. Poverty: A Study of Town Life. MacMillan and Company, London.

Ruckert-John, J., 2017. Meat consumption and sustainability: how might it be possible to change behavior of consumers? In: Biesalski, H.K., Drewnowski, A., Dwyer, J.T., Strain, J.J., Weber, P., Eggersdorfer, M. (Eds.), Sustainable Nutrition in a Changing World. Springer International Cambridge, pp. 111–124.

Shekelle, P., Hardy, M.L., Morton, S.C., et al., 2003. Ephedra and ephedrine for weight loss and athletic performance enhancement: clinical efficacy and side effects. Evid. Rep. Technol. Assess. 76, 1–4. PMID: 12647510.

Sotos-Prieto, M., DelRio, D., Drescher, G., Estruch, r, Hansen, C., Harlan, T., Hu, F.B., Li, M., McClung, J.P., Mojica, A., Puglielli, D., Toong, K., Yangarber, F., Kales, S.N., 2021. Mediterranean diet: promotion and dissemination of healthy eating: proceedings of an exploratory seminar at the Radcliffe Institute for Advanced Study. Int. J. Food Sci. Nutr. https://doi.org/10.1080/0937485.2021. 1941804.

Swinburn, B.A., Krak, V.I., Allender, S., et al., 2019. The global syndemic of obesity, undernutrition, and climate change: the *Lancet* Commission Report. Lancet. https://doi.org/10.1016/S0140-6736(18)32822-8.

Tapsell, L.C., Neale, E.P., Satija, A., Hu, F.B., 2016. Parents' mental associations with ultraprocessed products for their infant children: insights to improve complementary feeding practices. Adv. Nutr. 7 (3), 445–454.

Vidal, L., Brunet, G., Bove, I., Girona, A., Fuletti, D., Ares, G., 2021. Parents' mental associations with ultraprocessed products for their infant children: insights to improve complementary feeding practices. Food Qual. Prefer. 94, 104335. https//doi.org/10.1016/j.foodqual.2021.104335.

Vinogradova, Y., Coupland, C., Hippisley-Cox, J., 2020. Use of hormone replacement therapy and risk of breast cancer: nested case-control studies using the QResearch and CPRD databases. BMJ m3873. https://doi.org/10.1136/bmj.m3873.

Willett, W., Rockström, J., Loken, B., et al., 2019. Food in the anthropocene: the EAT–*Lancet* Commission on healthy diets from sustainable food systems. Lancet. https://doi.org/10.1016/S0140-6736(18)31788-4.

Further reading

Bradley, P. (Ed.), 2016. Food, Media, and Contemporary Culture: The Edible Image. Palgrave Macmillan, London, p. 288.

Katz, S.H., Weaver, W.W. (Eds.), 2002. Encyclopedia of Food and Culture: 3 Volume Set Hardcover. Charles Scribner's Sons, New York, p. 1800.

Kiple, K., Ornelas, K.C. (Eds.), 2000. Cambridge World History of Food London Cambridge. Cambridge University Press. ISBN-13: 978-0521402163 2 volumes ISBN-10: 0521402166.
2016. In: Klein, J.A., Watson, J.L. (Eds.), 2002. The Handbook of Food and Anthropology. Bloomsbury Academic, London, England, p. 496. $176.00 (cloth). ISBN: 978-0-85785-594-7.
Pilcher, J., 2012. Oxford Handbook of Food History. Oxford University Press, New York. https://doi.org/10.1093/oxfordhb/9780199729937.001.0001, 2ISBN:9780199729937.
Tannahill, R., 1973. Food in History. Stein and Day Publishers.
Tannahill, R., 1975. Flesh & Blood: A History of the Cannibal Complex. Cambridge University Press, London.

Food environment

Manuel Franco and Julia Díez, Public Health and Epidemiology Research Group, School of Medicine and Health Sciences, Universidad de Alcalá, Madrid, Spain

© 2023 Elsevier Ltd. All rights reserved.

Introduction 282
Population determinants of diet 282
Defining and conceptualizing food environments 283
Measuring the food environment 284
Food environment, dietary behaviors, and diet-related health outcomes 284
Interventions in the food environment 286
Conclusions 286
References 286

Key points

- The food environment is the interface where people interact with the wider food system to both acquire and consume foods
- The food environment is associated with dietary intake; however, the evidence of its impact on obesity is still to be fully understood
- Improving our knowledge and understanding about food environments is key to addressing overweight, obesity, and diet-related non-communicable diseases
- Achieving large-scale and sustainable prevention of diet-related health outcomes requires multi-component and multi-level interventions on the food environment

Introduction

If you look around your home or workplace, what kind of foods would you find most easily available and affordable to you? Would you buy at whatever outlet that is closest? These issues highlight some of the most important elements of food environments.

Globally, diet-related noncommunicable diseases and obesity are the leading contributors to poor health. At the same time, we know that eating is a complex social behavior where availability, accessibility, affordability, and cultural acceptability of different foods and rituals play a key role. Accordingly, the role of food environments in shaping our diets is increasingly gaining policy attention.

In this article, we review definitions and concepts, evidence, and examples of the influence of food environments on the distribution of dietary risk factors and associated non-communicable diseases. Finally, we also provide insight into policies and interventions that can create, enable, and sustain healthier food environments for a better population health.

Population determinants of diet

Globally, unhealthy diets are the main modifiable risk factor for chronic diseases (Murray et al., 2020). As such, the current global burden of disease of malnutrition, including childhood and adult overweight and obesity, has been attributed, at least partially, to large shifts in population diets over time (Swinburn et al., 2011). At the same time, eating is a complex social behavior where availability, accessibility, affordability, and cultural acceptability of different foods and rituals play key roles.

As stated by Geoffrey Rose, one can differentiate between two types of causes: the causes of individual disease cases and the determinants of disease incidence rates in our populations (Rose, 1985). The first, in this case, would look at why two individuals in the same population have different dietary patterns and try to discover which individual factors are different between them. The second type of cause would look at two different populations with different dietary patterns or at large changes of dietary patterns in the same population over time (Franco et al., 2013). The determinants of incidence rates are the ones we will need to study if we want to understand why dietary patterns have changed and worsened over time and why specific populations have higher rates of obesity or diabetes due to unhealthier dietary patterns. These population determinants are what Rose called mass influences on dietary patterns.

Mass influences on population diets include all components of the food system: from production and agriculture policies, processing, and transportation to food environments, marketing, retailing, and servicing. These components of the food system may differ across countries/regions and have changed over time within populations. Therefore, if we were able to understand these social determinants of population diets, we might be able to tackle the decreasing healthfulness of population diets and the associated

 https://doi.org/10.1016/B978-0-12-821848-8.00148-7

increases in obesity and other related chronic diseases like diabetes. Diabetes is currently one of the most important chronic diseases for the global economic and disease burden that diabetes means for people with diabetes but also for health systems and health expenditure.

Effective policies to combat unhealthy diets go beyond the individual and need to target the drivers of unhealthy diets (Roberto et al., 2015). One aspect of the food system that may be amenable to policies and more proximal interventions are the local food environments. The places where we buy food and eat do exert great influences over individual and population dietary behaviors. Considering the large differences in food environments within the same region or country and the shifting patterns over time in food availability, accessibility or affordability, local food environments represent a good example of a mass influence on population diets.

Defining and conceptualizing food environments

As defined by Swinburn et al. food environments encompass "the collective physical, economic, policy, and sociocultural surroundings, opportunities, and conditions that mediate food systems and shape individual diets" (Swinburn et al., 1999). Furthermore, Glanz et al. described the food environment at the local scale and divided it into the community food environment—defined as the exposure to (un)healthy food retailers around home, schools, workplaces, and beyond; and the consumer food environment—defined as attributes experienced by consumers within these food retailers, like the availability, price, or placement of foods (Glanz et al., 2005). A further elaboration of this model, including macro-level determinants is shown in **Fig. 1**.

Government policies can have enormous impacts on what type of foods are available and at what price. Industrial and agricultural practices are also of utmost importance, and they are interconnected with government policies. Food culture and other social norms that determine what type of foods are acceptable also steer dietary behaviors with great strength. The interplay between industry pressures, government policies, and cultural aspects of food results in the food environments we live in nowadays. Factors modeling the effects of the food environment (like gender, age, and country of origin for example) are also of great importance.

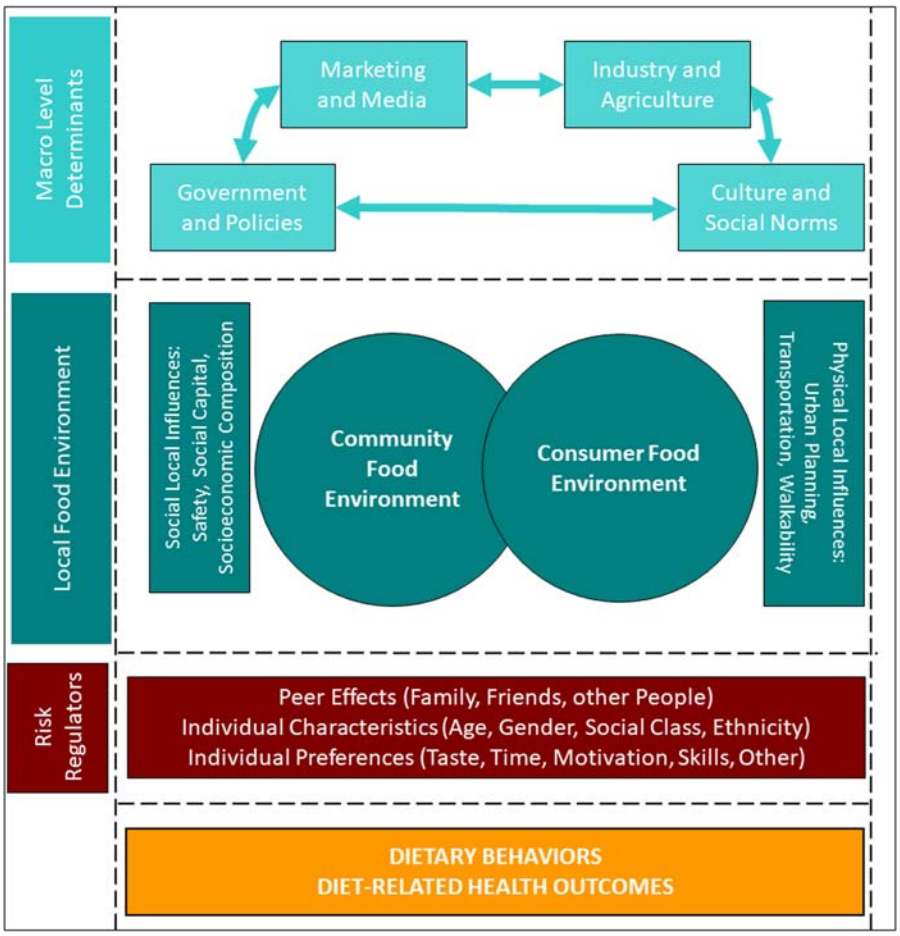


Fig. 1 Framework to study the food environment and its effects on diet. Adapted from Glanz, K., Sallis, J.F., Saelens, B.E., Frank, L.D., 2005. Healthy nutrition environments: concepts and measures. Am. J. Health Promot. 19(5), 330–333; Story, M., Kaphingst, K.M., Robinson-O'Brien, R., Glanz, K., 2008. Creating healthy food and eating environments: policy and environmental approaches. Annu. Rev. Publ. Health 29, 253–272.

Some of these factors are shaped more upstream from the food environment and relate closely to cultural and societal norms, while others are more proximal, like peer effects that have been shown to be a potent determinant of diet.

Measuring the food environment

Over the last decades, researchers have applied more complex measures to assess the food environment (Lytle and Sokol, 2017; Bivoltsis et al., 2018; Thornton et al., 2020). Yet, there is no clear consensus on the most appropriate operational definitions and measures. For example, Fleischhacker et al. showed that 20 (out of the 40 articles included) provided their own definition of a fast-food outlet (Fleischhacker et al., 2013). This use of diverse definitions has been identified as a key limitation of food environment research.

In terms of measures, most research has focused on characterizing the community food environment using objective measures like Geographic Information Systems (Lytle and Sokol, 2017) that are created around individual households or activity spaces (Thornton et al., 2020). In these studies, the food environment is measured as the availability, density of, or distance to certain types of food stores which are assumed to carry healthier foods or to be inherently unhealthy (e.g., fast-food restaurants) (Pinho et al., 2019).

As such, supermarkets have been considered as "healthy" or "low-BMI," whereas other retailers (e.g., convenience stores) have been targeted as "unhealthy" or "high-BMI" stores. Yet, the same type of store may carry a widely different balance of healthy and unhealthy foods. For example, studies in Latin America reveal that supermarkets carry a much unhealthier balance of food compared to autochthonous stores (Perez-Ferrer et al., 2019).

An example to illustrate this can be found in **Fig. 2**, which depicts the availability of healthy food stores in two geographically and socioeconomically diverse cities Baltimore, the United States, and Madrid, Spain (Diez et al., 2016).

As the figure shows, the food environment in Madrid is richer and has a greater variety of food stores, while several areas of the comparable neighborhood in Baltimore are underserved in terms of food stores. This highlights the intricate interconnections between the community and the consumer food environment, but also with other aspects of the food system like transportation networks or use of cars. The bottom part of **Fig. 2** highlights the two most common types of food stores in Baltimore and Madrid, respectively. Fruit and vegetable stores are widely available in Madrid and carry mostly fresh produce, compared with Baltimore's corner stores carrying mostly processed foods.

Moreover, many studies conducted in the United States have focused on food deserts—socioeconomically disadvantaged urban areas with very limited and poor healthy food access. Yet, there is no evidence of food deserts in European countries (Helbich et al., 2017) and insights for many low- and middle-income countries is still lacking. On the other hand, the term food swamps refer to local food environments in which healthy and fresh food options are available, but where ultra-processed foods are also ubiquitous (Bridle-Fitzpatrick, 2015). The latter, as Cooksey-Stowers et al. reported, may play an even larger role than food deserts on diet-related health outcomes (Cooksey-Stowers et al., 2017).

Measures of the consumer environment, which capture the actual foods available within a store, tend to be more expensive and resource consuming. Studies have reported that financial actions (e.g., price discounts) increase the purchase of healthier food options. Yet, more evidence is needed on their effect on other foods because discounts may not hinder consumers from buying non-discounted, unhealthy foods. Regarding food labeling, research suggests that solely providing information (e.g., traffic light labels) does not affect shoppers' purchasing behaviors. Yet, these features of the consumer food environment remain understudied and might be particularly context relevant.

As such, there is a need to conduct more studies combining both community and consumer nutrition environments. Ideally, these should apply valid and reliable measures, like the Nutrition Environment Measures Survey for Stores and Restaurants (NEMS-S/R), which can be adapted to particular contexts (Glanz et al., 2007; Martínez-García et al., 2020).

In addition, standardized and validated measures to capture the perceived food environment are needed. Although the body of literature using qualitative methodologies is growing, there are major gaps to be filled (e.g., research capturing children's perceived food environment) (Pitt et al., 2017).

Food environment, dietary behaviors, and diet-related health outcomes

There is growing evidence of the impact of the food environment on dietary intake; however, the evidence of its impact on obesity remains not well understood. As such, most studies have found the community food environment to be significantly related to food intake but not with body mass index. For example, fast-food access has been associated with fast-food consumption in most studies targeting children and adolescents (Jia et al., 2021). Yet, most studies have found mixed results between fast-food access and childhood obesity (Cobb et al., 2015).

On the other hand, studies conducted in Latin America have shown that living in a food environment with higher densities of convenience stores was significantly associated with a higher body mass index (Pineda et al., 2021). A recent systematic review also found that access to fast-food restaurants and convenience stores was positively associated with body weight status in China (An et al., 2020). Despite these examples, most studies examining the influence of the food environment on diet/obesity have exclusively focused on high-income populations. As such, more research on mid- and low-income countries is needed.

Studies on the impact of the consumer food environment have also indicated mixed results (Gustafson et al., 2012). Despite the large number of studies on this topic, a great variability in their measurement of the consumer nutrition environment and of the

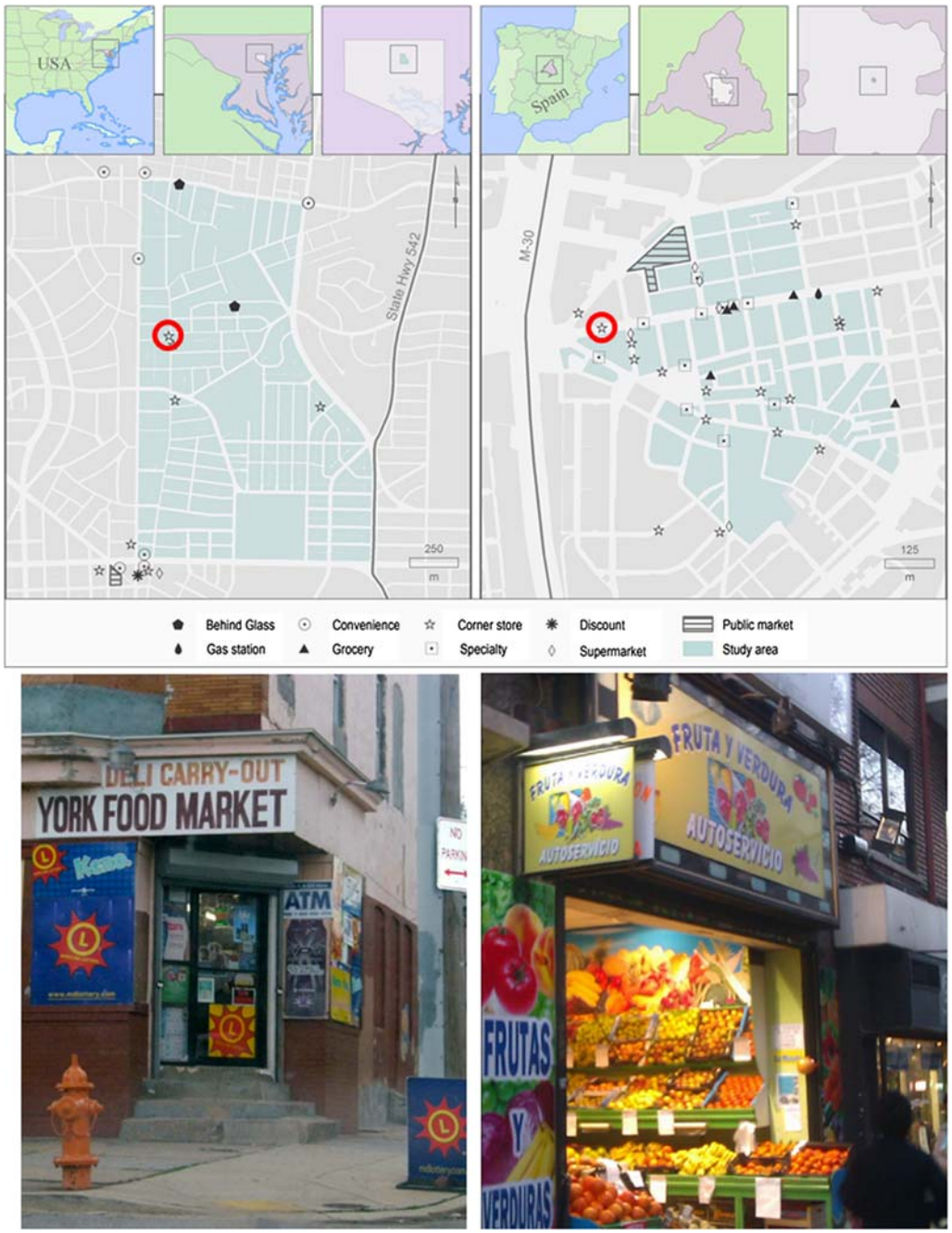


Fig. 2 Comparison of food stores present in two neighborhoods of similar number of inhabitants of similar relative socioeconomic status in Baltimore (the United States) and Madrid (Spain). Below, two stores (marked in map) in each neighborhood.

dietary assessment still exists. Assessing dietary behaviors is both expensive and challenging and often relies on self-reported measures.

A growing body of literature is also examining the association between the food environment and type 2 diabetes. Kanchi et al. have found that among US veterans in multiple community types, neighborhood food environment was associated with type 2 diabetes risk (Kanchi et al., 2021). In South Asia, Kusuma et al. also reported that both density and proximity to fast-food restaurants was associated with an increased risk of type 2 diabetes, especially among women (Kusuma et al., 2022). Results from another study conducted in the Netherlands confirm the evidence that the fast-food outlet environment is a diabetes risk factor (Ntarladima et al., 2022).

When assessing socioeconomic inequalities, research has shown that the food environment has a stronger impact, in terms of dietary behaviors, in most socioeconomically disadvantaged population groups. This is particularly the case for the school food

environment (Mackenbach et al., 2019). To illustrate, a recent study measured the availability of unhealthy foods within a 400 m walking distance of all schools present in the city of Madrid, Spain (Díez et al., 2019). The authors concluded that 95% of schools were surrounded by unhealthy food stores. Furthermore, schools located in most socioeconomically disadvantaged neighborhoods showed 62% more unhealthy retailers compared with schools located in middle-income areas.

Interventions in the food environment

Governments need to promote, enable, and sustain healthy food environments (Mozaffarian et al., 2018). Therefore, and following the experience with tobacco control, recent policies such as zoning laws banning the opening of new fast-food outlets are being adopted (Mah et al., 2019). Other regulations include restricting the marketing of unhealthy foods to children, improving health-related food labeling, or implementing fiscal policies to (dis)incentivize consumption of (un)healthier foods and beverages (Mah et al., 2019). Although a substantial heterogeneity among interventions still exists, to date, most countries have implemented policies aiming at reducing the purchase and consumption of ultra-processed foods and sugar-sweetened beverages (Backholer et al., 2016).

The International Network for Food and Obesity/NCDs Research, Monitoring and Action Support (INFORMAS) has provided standardized methods to evaluate policies and actions through benchmarking (Sacks et al., 2021). The Food System Dashboard is a new monitoring initiative that applies this for the area of food and nutrition and which comprises several indicators related to the food system at each country (Fanzo et al., 2020).

Yet, the most important limitation is that policy assumptions are typically too narrow in scope and do not adequately consider the interplay between the food environment and the economic and sociocultural context of people's lives. Previous research has shown that structural conditions influence both food access but also the meaning of that food and the role it plays (Díez et al., 2017; Daniel, 2020). Therefore, interventions solely focusing on improving dietary behaviors will not be sufficient to reduce dietary inequalities. To illustrate this, Isaacs et al. suggest a shift in current interventions so that they focus on (1) greater engagement between national government and local authorities; (2) co-designing these policies with local communities and businesses serving these communities, while recognizing and managing the conflict of interest; and (3) place-based approaches tailored to the context of local communities (Isaacs et al., 2022).

Refining interventions on food environment will require a continued evaluation of policies and actions. Also, the complexity of the food environment makes it difficult to assess the contribution of an individual intervention. As stressed by the INFORMAS initiative, food environment research needs to (1) document stories of both success and failure; (2) engage with policy makers to understand how best to intervene in policy processes to support recommended policy implementation; (3) focus on strengthening accountability mechanisms; and (4) build consensus so that collective voices are amplified (Sacks et al., 2021).

Conclusions

Diet is a complex social behavior where availability, accessibility, affordability and acceptability of different foods play key roles. The places where we buy food and eat can exert great influences over individual dietary behaviors, so that local food environments represent a good example of mass influences on population diets. Features of the food environment can be measured both objectively and subjectively. There is a current lack of comparability due to a large heterogeneity in the use of operational definitions and methodologies. More standardized and reliable methods are therefore needed to fully understand the relationship between the food environment, dietary behaviors, and diet-related health outcomes. Multicomponent interventions, as compared to single component interventions, have shown stronger evidence for improving food environments. Moreover, interventions need to consider the interplay between the food environment and the economic and sociocultural context.

References

An, R., He, L., Shen, M.J., 2020. Impact of neighbourhood food environment on diet and obesity in China: a systematic review. Publ. Health Nutr. 23 (3), 457–473. https://doi.org/10.1017/S1368980019002167.

Backholer, K., Blake, M., Vandevijvere, S., 2016. Have we reached a tipping point for sugar-sweetened beverage taxes? Publ. Health Nutr. 19 (17), 3057–3061. https://doi.org/10.1017/S1368980016003086.

Bivoltsis, A., et al., 2018. Food environments and dietary intakes among adults: does the type of spatial exposure measurement matter? A systematic review. Int. J. Health Geogr. https://doi.org/10.1186/s12942-018-0139-7.

Bridle-Fitzpatrick, S., 2015. Food deserts or food swamps?: a mixed-methods study of local food environments in a Mexican city. Soc. Sci. Med. https://doi.org/10.1016/j.socscimed.2015.08.010.

Cobb, L.K., et al., 2015. The relationship of the local food environment with obesity: a systematic review of methods, study quality, and results. Obesity 1331–1344. https://doi.org/10.1002/oby.21118.

Cooksey-Stowers, K., Schwartz, M., Brownell, K., 2017. Food swamps predict obesity rates better than food deserts in the United States. Int. J. Environ. Res. Publ. Health 14 (11), 1366. https://doi.org/10.3390/ijerph14111366.

Daniel, C., 2020. Is healthy eating too expensive?: how low-income parents evaluate the cost of food. Soc. Sci. Med. 248, 112823. https://doi.org/10.1016/j.socscimed.2020.112823.

Diez, J., et al., 2016. Understanding differences in the local food environment across countries: a case study in Madrid (Spain) and Baltimore (USA). Prev. Med. 89, 237–244. https://doi.org/10.1016/j.ypmed.2016.06.013.
Díez, J., et al., 2017. The mismatch between observational measures and residents' perspectives on the retail food environment: a mixed-methods approach in the Heart Healthy Hoods study. Publ. Health Nutr. 20 (16). https://doi.org/10.1017/S1368980017001604.
Díez, J., et al., 2019. Socioeconomic inequalities in the retail food environment around schools in a southern European context. Nutrients 11 (7), 1511. https://doi.org/10.3390/nu11071511.
Fanzo, J., et al., 2020. The Food Systems Dashboard is a new tool to inform better food policy. Nat. Food 1 (5), 243–246. https://doi.org/10.1038/s43016-020-0077-y.
Fleischhacker, S.E., et al., 2013. Validity of secondary retail food outlet data: a systematic review. Am. J. Prev. Med. 45 (4), 462–473. https://doi.org/10.1016/j.amepre.2013.06.009.
Franco, M., Bilal, U., Orduñez, P., 2013. Population-wide weight loss and regain in relation to diabetes burden and cardiovascular mortality in Cuba 1980–2010: repeated cross sectional surveys and. BMJ. https://doi.org/10.1136/bmj.f1515.
Glanz, K., et al., 2005. Healthy nutrition environments: concepts and measures. Am. J. Health Promot. 19 (5), 330–333. https://doi.org/10.4278/0890-1171-19.5.330.
Glanz, K., et al., 2007. Nutrition environment measures Survey in stores (NEMS-S). Development and evaluation. Am. J. Prev. Med. 32 (4), 282–289. https://doi.org/10.1016/j.amepre.2006.12.019.
Gustafson, A., Hankins, S., Jilcott, S., 2012. Measures of the consumer food store environment: a systematic review of the evidence 2000-2011. J. Community Health 897–911. https://doi.org/10.1007/s10900-011-9524-x.
Helbich, M., et al., 2017. Food deserts? Healthy food access in Amsterdam. Appl. Geogr. 83, 1–12. https://doi.org/10.1016/j.apgeog.2017.02.015.
Isaacs, A., et al., 2022. From healthy food environments to healthy wellbeing environments: policy insights from a focused ethnography with low-income parents' in England. Health Place 77, 102862. https://doi.org/10.1016/j.healthplace.2022.102862.
Jia, P., et al., 2021. Fast-food restaurant, unhealthy eating, and childhood obesity: a systematic review and meta-analysis. Obes. Rev. 22 (S1). https://doi.org/10.1111/obr.12944.
Kanchi, R., et al., 2021. Longitudinal analysis of neighborhood food environment and diabetes risk in the veterans administration diabetes risk Cohort. JAMA Netw. Open 4 (10), e2130789. https://doi.org/10.1001/jamanetworkopen.2021.30789.
Kusuma, D., et al., 2022. Food environment and diabetes mellitus in South Asia: a geospatial analysis of health outcome data. In: Popkin, B.M. (Ed.), PLoS Med. 19 (4), e1003970. https://doi.org/10.1371/journal.pmed.1003970.
Lytle, L.A., Sokol, R.L., 2017. Measures of the food environment: a systematic review of the field, 2007–2015. Health Place 18–34. https://doi.org/10.1016/j.healthplace.2016.12.007.
Mackenbach, J.D., et al., 2019. A systematic review on socioeconomic differences in the association between the food environment and dietary behaviors. Nutrients 11 (9). https://doi.org/10.3390/nu11092215.
Mah, C.L., et al., 2019. A systematic review of the effect of retail food environment interventions on diet and health with a focus on the enabling role of public policies. Curr. Nutr. Rep. 8 (4), 411–428. https://doi.org/10.1007/s13668-019-00295-z.
Martínez-García, A., et al., 2020. Adaptation and evaluation of the nutrition environment measures survey in stores to assess Mediterranean food environments (NEMS-S-MED). Int. J. Environ. Res. Publ. Health 17 (19), 7031. https://doi.org/10.3390/ijerph17197031.
Mozaffarian, D., et al., 2018. Role of government policy in nutrition—barriers to and opportunities for healthier eating. BMJ k2426. https://doi.org/10.1136/bmj.k2426.
Murray, C.J.L., et al., 2020. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease study 2019. Lancet 396 (10258), 1223–1249. https://doi.org/10.1016/S0140-6736(20)30752-2.
Ntarladima, A.-M., et al., 2022. Associations between the fast-food environment and diabetes prevalence in The Netherlands: a cross-sectional study. Lancet Planet. Health 6 (1), e29–e39. https://doi.org/10.1016/S2542-5196(21)00298-9.
Perez-Ferrer, C., et al., 2019. The food environment in Latin America: a systematic review with a focus on environments relevant to obesity and related chronic diseases. Publ. Health Nutr. https://doi.org/10.1017/S1368980019002891.
Pineda, E., Brunner, E., Llewellyn, C., Mindell, J., 2021. The retail food environment and its association with body mass index in Mexico. Int. J. Obes. 45 (6), 1215–1228.
Pinho, M., et al., 2019. Exploring absolute and relative measures of exposure to food environments in relation to dietary patterns among European adults. Publ. Health Nutr. 22 (06), 1037–1047. https://doi.org/10.1017/S1368980018003063.
Pitt, E., et al., 2017. Exploring the influence of local food environments on food behaviours: a systematic review of qualitative literature. Publ. Health Nutr. https://doi.org/10.1017/S1368980017001069.
Roberto, C.A., et al., 2015. Patchy progress on obesity prevention: emerging examples, entrenched barriers, and new thinking. Lancet. https://doi.org/10.1016/S0140-6736(14)61744-X.
Rose, G., 1985. Sick individuals and sick populations. Int. J. Epidemiol. 14 (1), 32–38. https://doi.org/10.1093/ije/14.1.32.
Sacks, G., et al., 2021. Benchmarking as a public health strategy for creating healthy food environments: an evaluation of the INFORMAS initiative (2012–2020). Annu. Rev. Publ. Health 42 (1), 345–362. https://doi.org/10.1146/annurev-publhealth-100919-114442.
Swinburn, B., Egger, G., Raza, F., 1999. Dissecting obesogenic environments: the development and application of a framework for identifying and prioritizing environmental interventions for obesity. Prev. Med. 29 (6 I), 563–570. https://doi.org/10.1006/pmed.1999.0585.
Swinburn, B.A., et al., 2011. The global obesity pandemic: shaped by global drivers and local environments. Lancet 378 (9793), 804–814. https://doi.org/10.1016/S0140-6736(11)60813-1.
Thornton, L.E., Lamb, K.E., White, S.R., 2020. The use and misuse of ratio and proportion exposure measures in food environment research. Int. J. Behav. Nutr. Phys. Activ. 17 (1), 118. https://doi.org/10.1186/s12966-020-01019-1.

Food fortification programs

Helena Pachón, Food Fortification Initiative, Emory University, Atlanta, GA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of R.C. Flores-Ayala, Food Fortification: Programs, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 296–305, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00119-7.

Introduction 288
Status of food fortification programs 289
Foundational fortification documents 289
Legislation 289
Standards 289
Regulatory monitoring 293
Fortification program performance 294
Fortification compliance 294
Proportion of people consuming fortified food 295
Potential nutrient contribution 295
Health status before and after fortification 297
Fortification potential 298
Proportion of food that is industrially processed 298
Amount of food consumed 299
Proportion of population consuming the food 299
Conclusions 299
References 300

Key points

- The purpose of this article is to share publicly available information on maize flour, oil, rice, salt, and wheat flour fortification programs globally.
- Several documents are developed by country stakeholders to shape and guide fortification programs. Three were reviewed: legislation, standards, and regulatory monitoring protocols.
- There are many aspects of fortification programs that can be evaluated to determine if they are operating optimally or as planned. Four were reviewed: fortification compliance, proportion of people consuming fortified food (population coverage), potential nutrient contribution, and health status before and after fortification.
- Fortification may be a feasible intervention in countries without such programs that have a demonstrated nutritional need. A few criteria that can help identify whether fortification of a particular food should be explored to fill nutrient gaps in the diet were reviewed: the percentage of the food in the country that is industrially processed, the amount of the food that is consumed, and the percentage of people who eat this food.

Introduction

Food fortification "is the practice of deliberately increasing the content of an essential micronutrient, i.e., vitamins and minerals (including trace elements) in a food, to improve the nutritional quality of the food supply and to provide a public health benefit with minimal risk to health" (WHO and FAO, 2006). Also known as enrichment (WHO and FAO, 2006), fortification adds needed nutrients to foods while they are being processed. In theory, most foods can be fortified with most nutrients. In practice, a limited number of foods are fortified with a limited number of nutrients.

Fortification contributes to the public health goal of reducing the prevalence of a nutrient deficiency or adverse functional outcome. For example, fortification of salt with iodine reduces iodine deficiency, goiter, cretinism, and low intelligence (Aburto et al., 2014; Keats et al., 2019). Similarly, wheat flour fortification with vitamin B9 (folic acid) reduces folate deficiency and neural tube defects (Keats et al., 2019; Rosenthal et al., 2014).

This article will focus on the large-scale (mass) fortification of foods with vitamins and minerals to serve a public health purpose. Specifically, the status of maize flour, oil, rice, salt, and wheat flour fortification programs will be described with respect to foundational documents and fortification performance. This will be followed by an assessment of the feasibility of fortifying these foods in countries that do not require or allow enrichment—an important exercise in countries that have a demonstrated nutritional need and are exploring the best interventions to meet those needs.

Status of food fortification programs

The Global Fortification Data Exchange, or GFDx for short, is a website that compiles and visualizes information for up to 196 countries and five fortified foods: maize flour, oil, rice, salt, and wheat flour (GFDx, 2021a). Other foods are also fortified at large scale to have a public health impact, such as sugar (Greene et al., 2017) and milk (López de Romaña, 2018). Because no websites consolidate information on sugar, milk or other mass-fortified foods, this article will focus on maize flour, oil, rice, salt, and wheat flour fortification programs.

The GFDx summarizes information for three topical areas: foundational fortification documents, fortification program performance, and fortification potential (GFDx, 2021a). This information is gathered in one of two ways: an on-line search for publicly available documents that contain fortification metrics, or a periodic survey sent to country stakeholders requesting fortification statistics or documents that contain them.

Foundational fortification documents

Several documents are developed by country stakeholders to shape and guide fortification programs; they include legislation, standards, and regulatory monitoring protocols.

Legislation

Country-issued legislation documents typically stipulate whether fortification of a particular food is required (e.g., mandatory) or allowed (e.g., voluntary) (WHO and FAO, 2006). As of June 2021, most countries require or allow the fortification of salt, followed by wheat flour (**Table 1**).

Foundational documents go on to describe more details about the fortification of foods. For example, some countries with mandatory wheat flour fortification specify that all flour for human consumption must be fortified, whereas others indicate that only particular types of flour for human consumption must be fortified (**Table 2**) (GFDx, 2021a). For most salt fortification programs, all salt produced for human consumption must be fortified, whereas in a minority of countries, only certain salt types must be fortified (e.g., table salt).

Legislative documents specify that domestically produced and imported food must be fortified for most countries across the five foods (**Table 2**). It is unusual that some countries specify that exported foods must be fortified to the exporting countries' specifications (instead of the importing countries' standards).

Across five foods, most countries specify in their fortification foundational documents that foods for household use and that are used as ingredients for producing processed food must be fortified (**Table 2**). In contrast, salt is the only food where, if it is destined for animal use, it must be fortified. Few countries specify that donated foods (e.g., for humanitarian crises or social protection programs) must be fortified. Including the latter in countries' fortification documents may provide beneficiaries of social protection programs that provide food, with added vitamins and minerals through fortification.

Standards

Standards documents tend to describe the nutrients, nutrient levels and fortification compounds that are required or allowed to fortify foods in a country (Marks et al., 2018). Salt is generally a delivery vehicle for iodine, and oil for lipid-soluble vitamins (**Table 3**). In comparison, a greater variety of nutrients are required or allowed to be added to fortified grains (i.e., maize flour, rice, and wheat flour).

The World Health Organization (WHO) has guidelines for the fortification of maize flour (WHO, 2016), rice (WHO, 2018), salt (WHO, 2014), and wheat flour (WHO et al., 2009). For all of these except rice, WHO includes guidelines on the fortification compounds and levels to be used for 10 nutrients for maize flour, one nutrient for salt, and five nutrients for wheat flour fortification

Table 1 The number of countries with mandatory or voluntary fortification of maize flour, oil, rice, salt, and wheat flour (GFDx, 2021a).

Food	*Mandatory*[a] *(n)*	*Voluntary*[b] *(n)*	*Total (n)*
Maize flour	17	2	19
Oil	27	11	38
Rice	7	7	14
Salt	124	21	145
Wheat flour	85	14	99

[a]"The country has legal documentation that has the effect of currently mandating fortification of the food vehicle in question with one or more vitamins or minerals i.e., the documentation indicates that fortification of all or some of the food is compulsory or required" (GFDx, 2021a).

[b]"The country has official documentation and/or a food standard that provides guidance or conditions for fortification, but does not have the effect of mandating or requiring fortification. If a country has mandatory fortification for that food vehicle, it will be categorized by GFDx as not having voluntary fortification, even if some types of the food vehicle or some nutrients may be fortified on a voluntary basis" (GFDx, 2021a).

Table 2 The number of countries with mandatory fortification of maize flour, oil, rice, salt, and wheat flour that specify which food types, their origins, destinations and uses must be fortified (GFDx, 2021a).

		Food types[a] *(n)*		*Food origins and destination*[b] *(n)*			*Food uses*[c] *(n)*			
Food	*Mandatory (n)*	*All food types*	*Only certain food types*	*Domestically produced*	*Imports*	*Exports*	*Household use*	*Processed food use*	*Animal feed*	*Donated*
Maize flour	17	8	9	17	17	0	17	16	1	2
Oil[d]	27	17	10	26	25	1	26	24	6	2
Rice[e]	6	3	3	6	5	0	6	6	0	1
Salt[e]	123	88	35	115	121	1	120	113	71	0
Wheat flour	85	42	43	85	79	6	85	85	2	7

[a]The food type categories are mutually exclusive. "All food types" means that all foods for human consumption are required to be fortified. "Only certain food types" means that only a subset of foods for human consumption are required to be fortified. If neither was expressly stipulated in a country's foundational documents, then "all food types" was assumed.
[b]The categories under food origins and destinations are not mutually exclusive. "Domestically produced" means that foods produced in the country must be fortified. "Imports" means that foods imported into the country must be fortified. "Exports" means that foods exported from the country must be fortified. If none was expressly stipulated in a country's foundational documents, then "domestically produced" and "imports" were assumed to apply.
[c]The food use categories are not mutually exclusive. "Household use" means that foods for use in a household must be fortified. "Processed food use" means that foods intended to be ingredients in another food must be fortified. "Animal feed" means that foods used to feed animals must be fortified. "Donated" means that donated foods (e.g., for social protection programs) must be fortified. If none was expressly stipulated in a country's foundational documents, then "household use" and "processed food use" were assumed to apply. If the foundational documents stipulate that "edible food" must be fortified, then "animal feed" was also assumed to apply.
[d]For food origins and destinations and food uses, the number of countries with data is 26.
[e]For one country with mandatory fortification, there is no documentation to determine the legislative scope.

Table 3 The nutrients included in standards for fortified maize flour, oil, rice, salt, and wheat flour where fortification is mandatory or voluntary (GFDx, 2021a).

Food	*Countries (n)*	*Nutrients*
Maize flour	19	Calcium; iron; vitamins A, B1 (thiamin), B2 (riboflavin), B3 (niacin), B6 (pyridoxine), B9 (folate), B12 (cobalamin), D; zinc
Oil	36	Vitamins A, D, E
Rice	14	Calcium; iron; vitamins A, B1, B2, B3, B6, B9, B12, D, E; selenium; zinc
Salt	137	Fluoride, iodine, iron
Wheat flour	93	Calcium; iron; vitamins A, B1, B2, B3, B6, B9, B12, D; selenium; zinc

(**Table 4**). As WHO guidelines are evidence-informed (Peña-Rosas et al., 2012), it is reasonable to assume that fortification standards that align with WHO guidelines are more likely to lead to public health impact than those that are least aligned.

Iron is included in most countries' standards for maize flour and wheat flour fortification; the same is true for iodine in salt fortification (**Table 5**). Analyses for all nutrients in WHO guidelines for maize and wheat flour and for salt are available elsewhere (Bobrek et al., 2021; Greenwald, 2020).

Fortification compounds can differ in bioavailability, cost, whether and how they interact with other nutrients added through fortification, and sensory changes provoked in fortified foods (Hurrell et al., 2004; WHO, 2016). The ideal compound has the highest possible bioavailability, lowest cost, does not negatively interact with other nutrients and does not cause sensory changes to the fortified food.

Fortification compounds used to add iodine to salt specified in country standards were those recommended by WHO in the majority of countries (**Table 5**). This was also the trend for iron-fortification compounds in maize flour fortification standards. For wheat flour, one-quarter of country standards did not specify the name of the compound(s) that should be used to add iron to the flour. Without such guidance, less expensive and less bioavailable iron compounds may be used to fortify flour, reducing its potential impact for improving iron status among consumers (Hurrell et al., 2010).

For every nutrient in a fortification standard, two items are usually noted: the compound and the nutrient levels to be used in fortification. For maize flour, salt, and wheat flour, WHO offers recommended nutrient levels depending on the (1) amount of these foods that are consumed in a country and (2) fortification compound used (**Table 4**).

Across all publicly available standards for salt fortification, there are 182 instances of recommended potassium iodate and potassium iodide compounds listed (**Table 6**). For most of the compounds, iodine levels exceed WHO's recommended levels. The same was observed for iron levels reported for maize flour fortification. For wheat flour, there was an equal distribution across those iron compounds that are less than, exactly meet, or exceed WHO guidelines.

Nutrient levels may be lower than WHO recommendations if multiple foods are fortified with that nutrient in a country. For maize flour, the analysis considered the contribution of iron from fortified wheat flour (and vice versa). For countries where nutrient levels were below WHO guidelines, it may be worthwhile to review standards to determine if they are adequate to meet the nutrient

Table 4 WHO guidelines for maize flour (WHO, 2016), salt (WHO, 2014), and wheat flour (WHO et al., 2009) fortification.

Food	*Table*	*Footnotes*

Maize flour

TABLE 1. Levels of nutrients to consider for adding to fortified maize flour and corn meal, when it is the only micronutrient intervention, based on extraction rate, chemical form and estimated per capita consumption[a]

Nutrient[b]	Flour-extraction rate[c]	Compound	Nutrient concentration to be added by estimated availability/consumption (mg nutrient/kg maize flour)[d] <75 g/day[e]	75–149 g/day	150–300 g/day
Iron[f]	Low	NaFe-EDTA	40	40	20
		Ferrous sulfate	60	60	30
		Ferrous fumarate	60	60	30
		Electrolytic iron	NR	NR	60
	High	NaFe-EDTA	40	40	40
		Ferrous sulfate	60	60	60
		Ferrous fumarate	60	60	60
		Electrolytic iron	NR	NR	NR
Folic acid	Low or high	Folic acid	5.0	2.6	1.3
Vitamin A	Low or high	Vitamin A palmitate	6.0	3.0	1.5
Zinc	Low	Zinc sulfate/zinc oxide[g]	95	55	40
	High	Zinc sulfate/zinc oxide	100	100	80
Vitamin B_{12}[h]	Low or high	Cyanocobalamin	0.04	0.02	0.01
For restitution of content lost during milling[i]					
Thiamine	Low or high	Thiamine hydrochloride	3.9	3.9	3.9
Riboflavin	Low or high	Riboflavin	2.0	2.0	2.0
Niacin	Low or high	Niacinamide	36	36	36
Pyridoxine	Low or high	Pyridoxine hydrochloride	6.2	6.2	6.2
Pantothenic acid	Low or high	Calcium pantothenate	4.2	4.2	4.2

NaFeEDTA: ferric sodium ethylenediaminetetraacetate; NR: not recommended

[a] This is a table to be used as a general guidance, and the number and amounts of nutrients should be adapted according to the needs of the country. These estimated levels consider only maize flour or corn meal as the main fortification vehicle in a public health programme. If other fortification programmes with other food vehicles and other micronutrient interventions are jointly implemented effectively, these suggested fortification levels need to be adjusted downwards as necessary.

[b] Nutrient levels were adapted from reference (*18*). In studies with maize flour, NaFeEDTA, encapsulated iron salts and ferrous bisglycinate have shown high bioavailability with high-extraction flours *(15, 50, 51)*. Other iron compounds for use with high-extraction flour are included for increasing the variety of choices during the decision-making process. The addition of iron compounds at low levels of maize-flour or corn-meal consumption will depend on the organoleptic characteristics of the final product, after testing if feasible at industrial level *(21)*.

[c] High-extraction flour (>80%) is also known as whole flour. It retains high levels of natural maize phytates, which inhibit the body's ability to absorb iron and zinc *(4)*. For maize flour, nixtamalization is a process that yields high-extraction flour, while degermination and precooking produce low-extraction flour *(7)*.

[d] Consumption of maize flour and corn meal varies widely in different countries, ranging from around 50 g/person/day in Ghana, Haiti and Uganda, to 300 g/person/day in Lesotho and Malawi *(7)*.

[e] Estimated per capita consumption of <75 g/day does not allow for the addition of sufficient amounts of fortificant to cover the micronutrient needs for women of childbearing age. Fortification of additional food vehicles and other interventions may need to be considered.

[f] The amounts of iron presented here are in milligrams of elemental iron. The amount of a particular iron compound should be calculated depending on the molecular weight of the compound.

[g] Both zinc sulfate and zinc oxide could be used for maize fortification, although zinc oxide is cheaper *(52, 53)*. As with iron, the phytate concentration (high flour-extraction rate) will affect the bioavailability of zinc *(53, 54)*. These amounts of zinc fortification assume 5 mg zinc intake and no additional phytate intake from other dietary sources *(18)*.

[h] The prevalence of vitamin B_{12} depletion and deficiency is high in all age groups, reaching 50% in some countries *(55, 56)*. Also, inclusion of vitamin B_{12} is recommended when flour is fortified with folic acid *(56, 57)*.

[i] Restitution of some B-complex vitamins should be achieved as a regular practice in all settings. The content in flour from white maize is: thiamine 3.9 mg/kg, riboflavin 2.0 mg/kg, niacin 36 mg/kg, pyridoxine 6.2 mg/kg and pantothenic acid 4.2 mg/kg *(7)*.

(Continued)

Table 4 WHO guidelines for maize flour (WHO, 2016), salt (WHO, 2014), and wheat flour (WHO et al., 2009) fortification.—cont'd

Food	*Table*	*Footnotes*
Salt	**Table 1. Suggested concentrations for the fortification of food-grade salt with iodine.**	[a] This includes consumption as table salt as well as salt from processed foods. [b] This fortification concentration was calculated based on the mean recommended nutrient intake of 150 µg iodine/day + 30% losses from production to household level before consumption, and a 92% iodine bioavailability. Losses depend on the iodization process, the quality of salt and packaging materials and the climatic conditions. Losses could vary widely[1] and this table presents the value considering 30% losses. The monitoring of urinary iodine concentrations will allow adjustment of the selected fortification concentrations. RNI: recommended nutrient intake, is the daily intake, set at the estimated average requirement plus 2 standard deviations, which meets the nutrient requirements of almost all apparently healthy individuals in an age- and sex-specific population group. Although iodate is more stable, either potassium iodate (KIO_3) or iodide (KI) can be used. Iodide may be used for dry, low crystal size and washed or refined salts. While iodate can be used alone and in any type of salt quality, iodide is used in very good quality salt and cannot be added alone. Therefore, some salt producers add sodium carbonate or sodium bicarbonate when they iodize salt, to increase alkalinity, and sodium thiosulfate or dextrose to stabilize potassium iodide. Without a stabilizer, potassium iodide may be oxidized to iodine and lost by volatilization from the product.[2] An estimated additional variability of ±10% during iodization procedures could be considered at the production site for use in quality control and assurance procedures. This variability depends on the iodization methods used and quality assurance system in place. Shaded areas correspond to the WHO salt reduction guideline.
Wheat flour	Table 1. Average levels of nutrients to consider adding to fortified wheat flour based on extraction, fortificant compound, and estimated *per capita* flour availability	1. These estimated levels consider only wheat flour as main fortification vehicle in a public health program. If other mass-fortification programs with other food vehicles are implemented effectively, these suggested fortification levels may need to be adjusted downwards as needed. 2. Estimated per capita consumption of <75 g/day does not allow for addition of sufficient level of fortificant to cover micronutrients needs for women of childbearing age. Fortification of additional food vehicles and other interventions should be considered. 3. NR = Not Recommended because very high levels of electrolytic iron needed could negatively affect sensory properties of fortified flour. 4. These amounts of zinc fortification assume 5 mq zinc intake and no additional phytate intake from other dietary sources.

Estimated salt consumption[a], g/day	Average amount of iodine to add, mg/kg salt (RNI + losses[b])
3	65
4	49
5	39
6	33
7	28
8	24
9	22
10	20
11	18
12	16
13	15
14	14

Nutrient	Flour Extraction Rate	Compound	Level of nutrient to be added in parts per million (ppm) by estimated average per capita wheat flour availability (g/day)[1]			
			<75[2] g/day	75-149 g/day	150-300 g/day	>300 g/day
Iron	Low	NaFeEDTA	40	40	20	15
		Ferrous Sulfate	60	60	30	20
		Ferrous Fumarate	60	60	30	20
		Electrolytic Iron	NR[3]	NR[3]	60	40
	High	NaFeEDTA	40	40	20	15
Folic Acid	Low or High	Folic Acid	5.0	2.6	1.3	1.0
Vitamin B_{12}	Low or High	Cyanocobalamin	0.04	0.02	0.01	0.008
Vitamin A	Low or High	Vitamin A Palmitate	5.9	3	1.5	1
Zinc[4]	Low	Zinc Oxide	95	55	40	30
	High	Zinc Oxide	100	100	80	70

Table 5 Comparison to WHO guidelines of fortification compounds listed in country standards: one illustrative nutrient each for maize flour, salt, and wheat flour where fortification is mandatory or voluntary (GFDx, 2021a).

Food	*Nutrient*[a]	*Countries*[b] *(n)*	*All compounds are WHO-recommended*[c] *(n)*	*None of the compounds are WHO-recommended*[d] *(n)*	*Both recommended and non-recommended compounds*[e] *(n)*	*Compounds not specified*[f] *(n)*
Maize flour	Iron	19	14	2	1	2
Salt	Iodine	137	88	1	36	12
Wheat flour	Iron	91	44	6	17	24

[a]According to the Global Fortification Data Exchange, this is the nutrient that the largest number of countries include in their fortification standards for this food (GFDx, 2021a).
[b]Countries for which the Global Fortification Data Exchange has a copy of the fortification standard for the food.
[c]Countries where every fortification compound listed in the fortification standard is recommended by WHO for fortification of maize flour (WHO, 2016), salt (WHO, 2014), and wheat flour (WHO et al., 2009).
[d]Countries where none of the fortification compounds listed in the fortification standard is recommended by WHO for fortification of the food.
[e]Countries where the fortification compounds listed in the fortification standard are a mix of those recommended by WHO and not recommended by WHO for fortification of the food.
[f]Countries where the fortification compound is not listed in the fortification standard for the food.

Table 6 Percentage of WHO guidelines met by country standards for fortification levels: one illustrative nutrient for maize flour, salt, and wheat flour where fortification is mandatory or voluntary (GFDx, 2021a).

Food	*Nutrient*[a]	*Fortification compounds*[b] *(n)*	*Meet <100% of WHO guidelines*[c] *(n)*	*Meet 100% of WHO guidelines*[c] *(n)*	*Meet >100% of WHO guidelines*[c] *(n)*	*Range of WHO guidelines met*[d] *(%)*
Maize flour	Iron	16	5	0	11	47–251
Salt	Iodine	182	25	0	157	68–350
Wheat flour	Iron	90	33	29	28	33–251

[a]The nutrient that is included in the standards for the most number of countries with maize flour, salt, and wheat flour fortification (GFDx, 2021a).
[b]The number of WHO-recommended fortification compounds specified in countries' standards to fortify the food with the specific nutrient, i.e., electrolytic iron, ferrous fumarate, ferrous sulfate and NaFeEDTA for iron in maize flour and wheat flour, and potassium iodate and potassium iodide for iodine in salt (WHO et al., 2009; WHO, 2014, 2016). One country can have two or more WHO-recommended fortification compounds specified in their fortification standards.
[c]For each nutrient, country fortification standards specify fortification compounds and nutrient levels to be added to the food. For each compound, the nutrient level could be less than, equal to, or greater than nutrient levels in WHO guidelines issued for maize flour, salt, and wheat flour (WHO et al., 2009; WHO, 2014, 2016).
[d]The nutrient levels specified in country standards were compared to nutrient levels in WHO guidelines and a percentage was calculated of the WHO recommendations met. The minimum and maximum percentage of WHO recommendations met are listed.

needs of the population—taking into account other interventions that are delivering the same nutrients. A review is also merited for countries where nutrient levels are above WHO guidelines. Are the high levels warranted based on nutritional need? Is there a risk of providing too much of a nutrient to the population when other interventions are taken into account?

Regulatory monitoring

Regulatory monitoring documents guide governments' enforcement activities for food fortification (Nathan, 1999). These documents state how government inspectors will monitor domestic facilities where fortified foods are produced (i.e., external monitoring) and imported foods that should be fortified (i.e., import monitoring) (WHO and FAO, 2006). For example, external monitoring protocols can stipulate with what frequency inspectors will visit plants that produce fortified food and what information they will collect during those visits, while import monitoring protocols specify how different agencies, including Customs officials, will coordinate to audit imported foods that should be fortified.

Protocols guide the monitoring activities carried out by government officials. Most countries with mandatory fortification of maize flour, oil, and wheat flour have publicly available protocols for external monitoring (**Table 7**). The same is true for maize flour, oil, and rice for import monitoring protocols. It would be of concern if these documents do not exist for some countries and if government inspectors have little written guidance on the monitoring work they are expected to carry out.

Best practices suggest the minimum information that foundational documents should contain about fortification monitoring (Marks et al., 2018). Country documentation was reviewed for the presence of two illustrative elements for external and import monitoring each (**Table 8**); information extracted from foundational documents for 27 other elements can be viewed elsewhere (Stern, 2021).

Relatively few documents reviewed for maize flour, oil, salt, and wheat flour stated which government agency was responsible for external or import monitoring (**Table 8**). Even fewer documents stipulated the frequency of government inspections for the five foods. These results are concerning if, in fact, the agency responsible for fortification or the frequency with which they should monitor is unknown. It is possible that this information is known and stated in documents that are not publicly available.

Table 7 Number of countries with external monitoring and import monitoring documents for the regulatory monitoring of mandatory maize flour, oil, rice, salt, and wheat flour (GFDx, 2021a).

		External monitoring[b]		*Import monitoring*[c]	
Food	*Mandatory*[a] *(n)*	*Protocol applicable*[d] *(n)*	*Protocol available*[e] *(n)*	*Protocol applicable*[d] *(n)*	*Protocol available*[e] *(n)*
Maize flour	17	17	10	17	9
Oil	27	26	16	18	15
Rice	7	7	2	7	4
Salt	124	82	10	119	16
Wheat flour	85	71	48	85	26

[a]Number of countries with mandatory food fortification.
[b]Government monitoring in domestic facilities where fortified foods are produced.
[c]Government monitoring of imported foods that should be fortified.
[d]External monitoring is applicable in countries where foods are domestically produced and import monitoring is applicable in countries where foods are imported into the country. In countries that exclusively import the food, for example, external monitoring protocols are not expected or applicable.
[e]The monitoring protocol is publicly available.

Table 8 Number of countries with two best-practice monitoring elements included in their monitoring documents for maize flour, oil, rice, salt, and wheat flour (GFDx, 2021a).

	External monitoring[a]			*Import monitoring*[b]		
Food	*Applicable*[c] *(n)*	*States which government agency is responsible for monitoring*[d,f] *(n)*	*Describes the frequency of inspection*[e,f] *(n)*	*Applicable*[c] *(n)*	*States which government agency is responsible for monitoring*[d,f] *(n)*	*Describes the frequency of inspection*[e,f] *(n)*
Maize flour	17	6	3	17	8	5
Oil	26	10	4	18	5	2
Rice	7	4	1	7	6	2
Salt	82	37	1	119	37	10
Wheat flour	71	32	15	85	25	12

[a]Government monitoring in domestic facilities where fortified foods are produced.
[b]Government monitoring of imported foods that should be fortified.
[c]External monitoring is applicable in countries where foods are domestically produced and import monitoring is applicable in countries where foods are imported into the country. In countries that exclusively import the food, for example, external monitoring protocols are not expected or applicable.
[d]Publicly available country documentation states which government agency(ies) is(are) responsible for monitoring.
[e]Publicly available country documentation states the frequency with which government agency(ies) conduct(s) inspections.
[f]The number may be greater than the number of countries with the protocol available, per the previous table. This means that a foundational document other than the external monitoring protocol (e.g., legislation, standards) has this information registered in it.

Fortification program performance

Many aspects of fortification programs can be evaluated to determine if they are operating optimally or as planned. This article will review four: fortification compliance, proportion of people consuming fortified food (population coverage), potential nutrient contribution, and health status before and after fortification.

Fortification compliance

A fundamental premise for fortified foods to have nutritional impact is that they must be fortified according to the country-defined standards. Broadly speaking, fortification compliance can be assessed in two non-mutually exclusive ways (GAIN and PHC, 2018). One, fortification processes in production facilities are audited by government inspectors to identify if best practices are being followed, such as calculations relating the amount of fortified food produced to the amounts of vitamins and minerals used in their production (i.e., premix reconciliation). Two, government inspectors gather samples of food that should be fortified from production facilities and send them to laboratories to determine if they contain the correct amounts of vitamins and minerals, as outlined in the country's fortification standards. With both methods, criteria are developed to classify production facilities and samples for compliance.

Few countries produce publicly available reports with compliance estimates (**Table 9**). Among these, the proportions of foods that are fortified according to country standards range from as low as 33% to as high as 100%.

Table 9 Among countries with mandatory or voluntary fortification, those with publicly available fortification compliance data and the range of compliance for fortified maize flour, oil, rice, salt, and wheat flour (GFDx, 2021a).

Food	*Mandatory or voluntary*[a] *(n)*	*Report compliance data*[b] *(n)*	*Percent compliance*[c] *(minimum-maximum)*
Maize flour	19	2	33–33
Oil	38	0	Not applicable
Rice	14	1	100
Salt	145	7	59–96.6
Wheat flour	99	5	76–100

[a]Countries that require or allow fortification.
[b]Countries that report the percentage of food that meets fortification standards for at least one year between 1995 and the present.
[c]When compliance is available, the minimum and maximum percentages reported. If a country had multiple years of compliance data, only the latest year is included.

Proportion of people consuming fortified food

The greater the proportion of individuals—especially those from the target population group—who consume fortified foods (i.e., population coverage), the greater the health impact of fortification. **Table 10** reports the percentage of people who consume foods that have been fortified (i.e., these are not necessarily foods that have been fortified to meet nutrient levels in fortification standards). Except for salt, few countries report population coverage (**Table 10**). Among those that do report these figures, the range is large (2.5–99.9%).

Potential nutrient contribution

The potential nutrient contribution of fortified foods can be gleaned from standards information coupled with estimates of dietary intake of fortified foods. Specifically, the nutrient amounts added to food through fortification according to standards (in mg kg^{-1}) multiplied by the daily amount of fortified food consumed (in g d^{-1}) divided by 1000 yields the potential nutrient intake (in mg d^{-1}) for each fortified food (Pachón et al., 2021).

The potential nutrient contribution was modeled under two scenarios (**Fig. 1**) (GFDx, 2021b). In the maximum scenario, it was assumed that (1) 100% of the food consumed in the country is industrially processed (and therefore easier to fortify and monitor) and (2) 100% of the food consumed in the country is fortified according to the country's standard (i.e., is compliant with national regulations). In the realistic scenario, actual data were used for the (1) percent of the food that is industrially processed in the country and (2) percent of food that is fortified according to the country's regulations.

Country-level nutrient intakes were compared to nutrient-specific Estimated Average Requirements (EARs) and Tolerable Upper Intake Levels (ULs) for women of childbearing age. EARs, when calculated for individuals, represent the amount of the nutrient required to meet the needs of 50% of the population. ULs represent the highest, daily safe amount that individuals can consume. The percent of nutrient-specific EARs (**Fig. 2**) and ULs (**Fig. 3**) potentially met through the consumption of fortified foods is summarized for up to 153 countries.

The analyses suggest that if 100% of maize flour, oil, rice, salt, and wheat flour in each country is industrially processed and all are fortified according to country standards (i.e., 100% compliance), these fortified foods can make a meaningful contribution to nutrient requirements (EARs) (**Fig. 2**). In practice, the actual contribution is lower because few countries have foods that are both 100% industrially processed and 100% compliant with fortification standards.

These findings point to two ways in which fortification programs can be modified so that the nutrients intended to be provided through fortification (i.e., based on country fortification standards) are delivered (Pachón et al., 2021). One is to consider alternative food vehicles for fortification that are mostly produced in industrial-scale facilities. The second is for industry to increase compliance with fortification standards.

Table 10 Among countries with mandatory or voluntary fortification, those with publicly available population coverage data and the range of coverage for fortified maize flour, oil, rice, salt, and wheat flour (GFDx, 2021a).

Food	*Mandatory or voluntary*[a] *(n)*	*Report coverage data*[b] *(n)*	*Percent coverage*[c] *(minimum-maximum)*
Maize flour	19	2	2.5–6.5
Oil	38	3	34.1–54.4
Rice	14	0	Not applicable
Salt	145	89	6.9–99.9
Wheat flour	99	6	8.5–98.0

[a]Countries that require or allow fortification.
[b]Countries that report the percentage of individuals who consume fortified foods for at least one year between 1995 and the present. The foods are fortified any amount.
[c]When coverage is available, the minimum and maximum percentages reported. If a country had multiple years of coverage data, only the latest year is included.

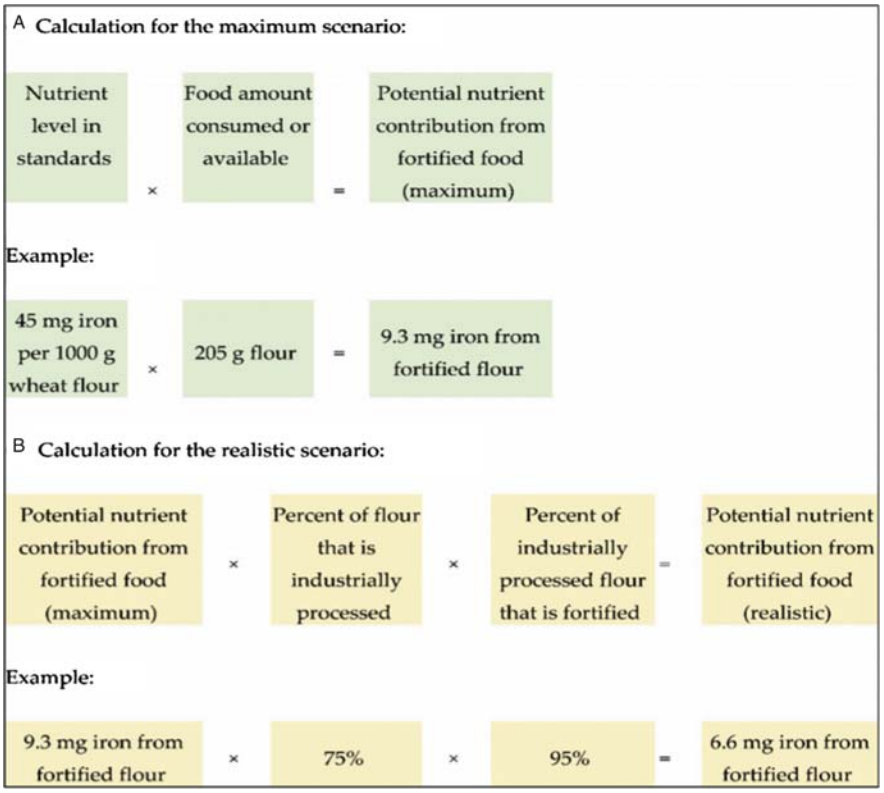


Fig. 1 Calculation of potential nutrient contribution (in mg capita^{-1} day^{-1}) (A) under the maximum scenario where 100% of the food is assumed to be industrially processed and 100% is assumed to be fortified and (B) under the realistic scenario where the amount of the food that is industrially processed and fortified is known for a country (Pachón et al., 2021).

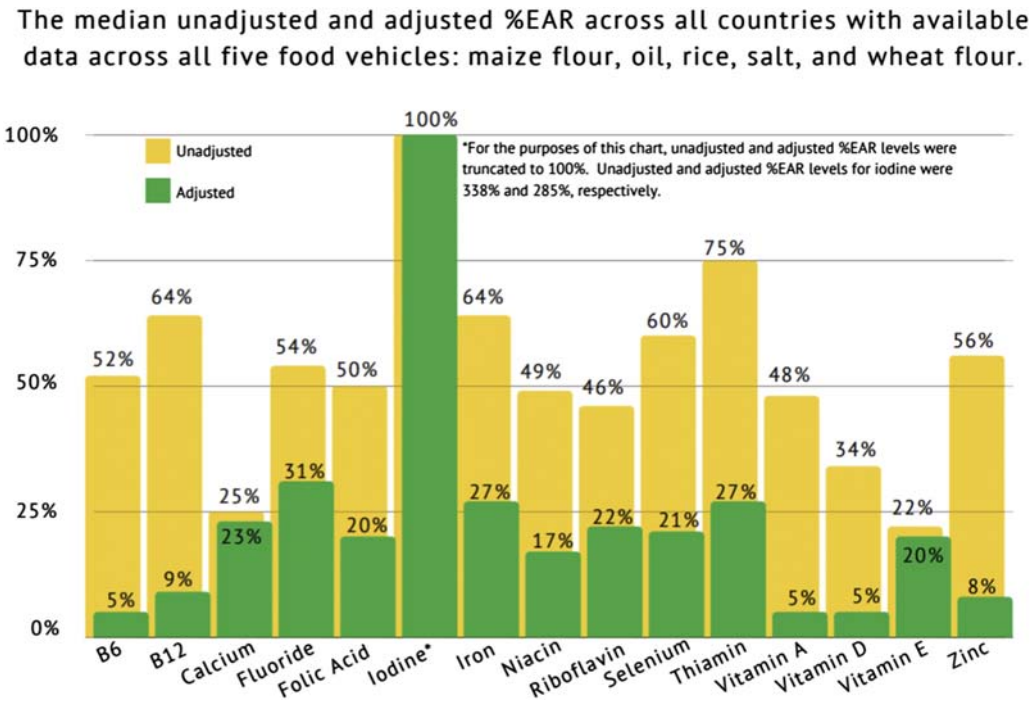


Fig. 2 Maximum and realistic estimates of the combined nutrient contribution of maize flour, oil, rice, salt, and wheat flour fortified according to standards: comparison to Estimated Average Requirements (EARs) for women (GFDx, 2021b). In the unadjusted (maximum) scenario, it was assumed that (1) 100% of the food consumed in the country is industrially processed (and therefore easier to fortify and monitor) and (2) 100% of the food consumed in the country is fortified according to the country's standard (i.e., is compliant with national regulations). In the adjusted (realistic) scenario, actual data were used for the (1) percent of the food that is industrially processed in the country and (2) percent of food that is fortified according to the country's regulations.

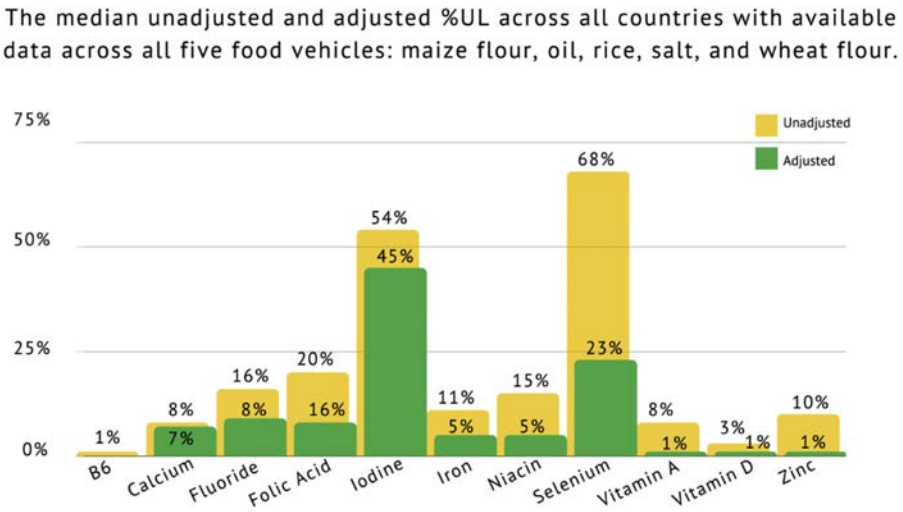


Fig. 3 Maximum and realistic estimates of the combined nutrient contribution of maize flour, oil, rice, salt, and wheat flour fortified according to standards: comparison to Tolerable Upper Intake Levels (ULs) for women (GFDx, 2021b). In the unadjusted (maximum) scenario, it was assumed that (1) 100% of the food consumed in the country is industrially processed (and therefore easier to fortify and monitor) and (2) 100% of the food consumed in the country is fortified according to the country's standard (i.e., is compliant with national regulations). In the adjusted (realistic) scenario, actual data were used for the (1) percent of the food that is industrially processed in the country and (2) percent of food that is fortified according to the country's regulations.

The risk of exceeding ULs is low for most nutrients (Fig. 3). In countries where other interventions deliver micronutrients, it is prudent to review all programs to determine if any should be adjusted to minimize the risk of providing excess nutrients to the population.

Health status before and after fortification

Ultimately, the purpose of food fortification is to effect a positive change on population health. This is assessed by measuring different indicators of nutrition and health status before and after fortification initiation.

Salt is the food for which the highest number of countries report the results of a health assessment (Table 11). While many nutrients are added to fortified food (e.g., vitamins A, D and E in oil), the status of only a few nutrients were evaluated in individuals who had access to these foods (e.g., vitamin A in oil). For most nutrients, multiple outcomes were measured (e.g., serum zinc and the prevalence of zinc deficiency for zinc).

For maize flour, rice, salt, and wheat flour, most of the outcomes assessed after fortification implementation were positive, meaning that there were improvements in health and nutritional status (Table 12). Salt showed the most negative changes, where median urinary iodine concentration decreased between the pre-fortification and the post-fortification periods. Oil was a notable exception, where there were no changes in any of the outcomes assessed, suggesting that fortification led to no health improvement.

Table 11 Health outcomes reported after fortification implementation in countries with mandatory fortification of maize flour, oil, rice, salt, and wheat flour (GFDx, 2021a).

Food	*Mandatory*[a] *(n)*	*Report at least 1 outcome*[b] *(n)*	*Nutrients assessed*[c] *(n)*	*Outcomes reported*[d] *(n)*
Maize flour	17	5	2	7
Oil	27	1	1	2
Rice	7	2	1	4
Salt	124	79	1	1
Wheat flour	85	15	4	10

[a]Number of countries with mandatory food fortification.

[b]Number of countries that report the results of at least one health outcome after fortification implementation.

[c]In the studies completed, the number of nutrients that were assessed. The nutrients assessed were iron and vitamin B9 for maize flour, vitamin A for oil, vitamin B9 for rice, iodine for salt, and iron, vitamin B9, vitamin B12 and zinc for wheat flour.

[d]In the studies completed, the number of unique health outcomes that were assessed; each could be assessed in multiple population groups. The outcomes assessed included median urinary iodine concentration for iodine; plasma or serum ferritin, prevalence of iron deficiency, and the prevalence of iron-deficiency anemia for iron; plasma or serum retinol binding protein, breastmilk retinol and the prevalence of vitamin A deficiency for vitamin A; plasma or serum folate, red blood cell folate, prevalence of folate deficiency and the birth prevalence of neural tube defects for vitamin B9; plasma or serum cobalamin for vitamin B12; and plasma or serum zinc and prevalence of zinc deficiency for zinc.

Table 12 Results of health outcomes assessed after countries mandate the fortification of maize flour, oil, rice, salt, and wheat flour (GFDx, 2021a).

Food	*Total outcomes assessed*[a] *(n)*	*Positive change*[b] *(n)*	*No change*[c] *(n)*	*Negative change*[d] *(n)*	*Could not be determined*[e] *(n)*
Maize flour	40	31	5	1	3
Oil	14	0	14	0	0
Rice	22	17	2	0	3
Salt	83	73	0	10	0
Wheat flour	91	76	10	1	4

[a]The total number of health outcomes assessed in any studies that assessed the impact of fortification. The types of health outcomes are noted in the footnotes for Table 11.
[b]The number of health outcomes that showed a positive health change from the pre to the post-fortification periods; this change was statistically tested. For example, if the prevalence of zinc deficiency statistically decreased from pre-to post-fortification, that outcome would be coded as a positive change. If serum ferritin statistically increased from pre-to post-fortification, that outcome would be coded as a positive change.
[c]The number of health outcomes that showed no change from the pre to the post-fortification periods; this lack of change was statistically confirmed. For example, if the prevalence of iron deficiency did not statistically change between the pre-to post-fortification, that outcome would be coded as no change.
[d]The number of health outcomes that showed a negative health change from the pre-to the post-fortification periods; this change was statistically tested. For example, if the prevalence of folate deficiency statistically increased from pre-to post-fortification, that outcome would be coded as a negative change. If serum retinol statistically decreased from pre-to post-fortification, that outcome would be coded as a negative change.
[e]The number of health outcomes that were not statistically tested by the study authors. Without the statistics, whether the health outcome improved, worsened, or stayed the same between the pre- and post-fortification periods could not be determined.

However, all of the oil outcomes were evaluated in one study from one country. Results from other studies with oil fortification are welcome to understand the trends observed from large-scale fortification of this food.

Fortification potential

Fortification may be a feasible intervention in countries without such programs that have a demonstrated nutritional need. A few criteria can help identify whether fortification of a particular food should be explored to fill nutrient gaps in the diet. If there is no nutritional need in a country, then this exercise is not needed.

Feasibility criteria include the percentage of the food in the country that is industrially processed, the amount of the food that is consumed, and the percentage of people who eat this food. Together with other insights, such as the food industry structure, the consumption habits of the population group with the most need, and other interventions already in place in the country, a decision can be made whether to pursue food fortification in the country.

Proportion of food that is industrially processed

Foods can be fortified in small- or in large-scale facilities. However, government enforcement of fortification is easier when there are fewer large-scale facilities than many small-scale facilities (WHO and FAO, 2006). Further, food produced in large-scale facilities has a greater geographic scope than food produced in small-scale facilities.

For these and other reasons, the higher the proportion of food that is industrially processed, the greater the feasibility that it can be fortified at large-scale to have a public health impact in the country. There are dozens of countries where more than 75% of maize flour, rice and wheat flour produced is estimated to be industrially processed (**Table 13**). A handful of countries have such data for oil, and fortification is feasible in all of them.

Table 13 Number of countries without mandatory or voluntary fortification and the percent of maize flour, oil, rice, salt, and wheat flour that is industrially processed (GFDx, 2021a).

Food	*Have industrial processing data*[a] *(n)*	*Percent industrially processed*[b] *(minimum-maximum)*	*>75% is industrially processed*[c] *(n)*
Maize flour	73	1–100	65
Oil	3	80–100	3
Rice	107	1–100	80
Salt	0	Not applicable	Not applicable
Wheat flour	92	60–100	86

[a]Countries without mandatory or voluntary fortification of the food that have information in the Global Fortification Data Exchange about the percent of the food that is industrially processed in the country.
[b]For countries that have information on the percent of the food that is industrially processed, the minimum and maximum values in the range.
[c]Among countries with information on the percent of the food that is industrially processed, the number of countries where more than 75% of the food is industrially processed. The 75% cutoff value is an arbitrary designation.

The selection of 75% as a cutoff is arbitrary. Within each country, policy makers should decide what percentage is most appropriate. For example, there may be some cases where a lower percentage of industrially processed food may still warrant fortification, due to the number of people who consume the food.

Amount of food consumed

Potential food vehicles for fortification are those consumed in "large" amounts by the target population (e.g., women of reproductive age). If there is only one viable food vehicle for fortification in a country, ideally the amount consumed is large enough that all needed nutrients can be added in amounts that do not cause sensory changes to the food. If there are two or more viable food vehicles in a country, the amount consumed of each individual food is less important.

For maize flour, rice, and wheat flour, "large" is defined as greater than 75 g daily, based on WHO guidance. Specifically, where flour intake is less than 75 g per capita per day, the amount of iron required to be added to have a positive health impact would likely cause sensory changes to the flour (WHO et al., 2009). Sensory changes will have the unintended effect of causing fewer people to eat and benefit from the fortified food. Thus, grain intake greater than 75 g per capita per day is arbitrarily used to define the minimum amount of food that should be consumed by the target population, if it is the only food that will be mass fortified in a country.

For grains, between 36 and 62 countries show food availability greater than 75 g per capita per day (**Table 14**). If there is nutritional need in these countries, fortification of one of these foods is potentially feasible.

Proportion of population consuming the food

The ideal food vehicle is consumed by a large proportion of the target population. This information is available for a few countries in the Global Fortification Data Exchange. For a handful of countries, the population coverage of oil is high enough that fortification is feasible (**Table 15**).

Conclusions

- The purpose of this article was to share publicly available information on maize flour, oil, rice, salt, and wheat flour fortification programs.
- Several documents are developed by country stakeholders to shape and guide fortification programs; they include legislation, standards, and regulatory monitoring protocols. These documents can be strengthened by including "best practice" information that they lack and more closely aligning them with international recommendations.
- Many aspects of fortification programs can be evaluated to determine if they are operating optimally or as planned. While there is great potential to contribute nutrients to the diet through fortification, there is surprisingly little public information available on food fortification compliance in countries, how many people consume fortified foods, and what health impact fortification is having on the population.
- Fortification may be a feasible intervention in countries without such programs that have a demonstrated nutritional need. Data can be used to help identify whether fortification of a particular food should be explored to fill nutrient gaps in a country's diet. This minimum information should be coupled with conversations with key actors in a country from the public, private and civic sectors.

Table 14 Estimates of the amount of maize flour, rice, and wheat flour available for human consumption in countries without mandatory or voluntary fortification (GFDx, 2021a).

Food	*With food availability data*[a] *(n)*	*Food availability in grams per capita per day*[b] *(minimum-maximum)*	*Food availability >75 g per capita per day*[c] *(n)*
Maize flour	144	0–415	36
Rice	155	2–733	59
Wheat flour	73	6–549	62

[a]Countries without mandatory or voluntary fortification of the food that have information on the amount of the food available for human consumptions, as estimated by FAO in food balance sheets (FAO, 2021).

[b]Among countries with food availability data, the minimum and maximum range of food availability expressed in grams per capita per day, as estimated by FAO in food balance sheets (FAO, 2021).

[c]Among countries with food availability data, those where availability is greater than 75 g per capita per day. The cutoff was proposed by WHO for maize flour and wheat flour (WHO et al., 2009); here it is also applied to another grain that is fortified: rice.

Table 15 Number of countries without mandatory or voluntary fortification that have coverage data for the consumption of any maize flour, oil, rice, salt, and wheat flour (GFDx, 2021a).

Food	*With coverage data*[a] *(n)*	*Percent coverage*[b] *(minimum-maximum)*	*Coverage >80%*[c] *(n)*
Maize flour	0	Not applicable	Not applicable
Oil	3	0–95	2
Rice	0	Not applicable	Not applicable
Salt	Not recorded[d]	Not applicable	Not applicable
Wheat flour	0	Not applicable	Not applicable

[a]Countries without mandatory or voluntary fortification of the food that have information on the percent of population consuming the food.

[b]Among countries with coverage data, the minimum and maximum percent of people who consume the food.

[c]Among countries with coverage data, those where coverage is greater than 80%. The 80% cutoff value is an arbitrary designation.

[d]Salt intake is assumed to be universal; however, not all dietary surveys capture it's consumption. Therefore, information on population coverage is not visualized on the Global Fortification Data Exchange.

References

Aburto, N., Abudou, M., Candeias, V., Wu, T., 2014. Effect and Safety of Salt Iodization to Prevent Iodine Deficiency Disorders: A Systematic Review With Meta-Analyses. WHO eLibrary of Evidence for Nutrition Actions (eLENA). World Health Organization, Geneva.

Bobrek, K.S., Broersen, B., Aburto, N.J., Garg, A., Serdula, M., Beltrán Velázquez, F., Wong, E.C., Pachón, H., 2021. Most national, mandatory flour fortification standards do not align with international recommendations for iron, zinc, and vitamin B12 levels. Food Pol. 99, 101996.

FAO, 2021. FAOSTAT: New Food Balances. http://www.fao.org/faostat/en/#data/FBS.

Global Alliance for Improved Nutrition (GAIN) & Project Healthy Children (PHC), 2018. Regulatory Monitoring of National Food Fortification Programs: A Policy Guidance Document. Global Fortification Technical Advisory Group (GF-TAG). https://fortificationdata.org/resources.

Global Fortification Data Exchange, 2021a. FortificationData.org.

Global Fortification Data Exchange, 2021b. Technical Brief: The Potential Nutrient Intake of Fortified Foods. FortificationData.org.

Greene, M.D., Kabaghe, G., Musonda, M., Palmer, A.C., 2017. Retail sugar from one Zambian community does not meet statutory requirements for vitamin A fortification. Food Nutr. Bull. 38 (4), 594–598.

Greenwald, R.P., 2020. Alignment of Iodization Requirements in Countries With Mandatory or Voluntary Salt Fortification to the World Health Organization. Bachelor of Arts Thesis. Emory University, Atlanta.

Hurrell, R.F., Lynch, S., Bothwell, T., Cori, H., Glahn, R., Hertrampf, E., Kratky, Z., Miller, D., Rodenstein, M., Streekstra, H., Teucher, B., Turner, E., Yeung, C.K., Zimmermann, M.B., SUSTAIN Task Force, 2004. Enhancing the absorption of fortification iron. A SUSTAIN Task Force Report. Int. J. Vitam. Nutr. Res. 74 (6), 387–401.

Hurrell, R., Ranum, P., de Pee, S., Biebinger, R., Hulthen, L., Johnson, Q., Lynch, S., 2010. Revised recommendations for iron fortification of wheat flour and an evaluation of the expected impact of current national wheat flour fortification programs. Food Nutr. Bull. 31 (Suppl. I), S7–S21.

Keats, E.C., Neufeld, L.M., Garrett, G.S., Mbuya, M., Bhutta, Z.A., 2019. Improved micronutrient status and health outcomes in low- and middle-income countries following large-scale fortification: evidence from a systematic review and meta-analysis. Am. J. Clin. Nutr. 109 (6), 1696–1708.

López de Romaña, D., Olivares, M., Pizarro, F., 2018. Milk and dairy products. In: Mannar, M.G.V., Hurrell, R.F. (Eds.), Food Fortification in a Globalized World, first ed. Elsevier Academic Press, London, pp. 175–181.

Marks, K.J., Luthringer, C.L., Ruth, L.J., Rowe, L.A., Khan, N.A., De-Regil, L.M., López, X., Pachón, H., 2018. Review of grain fortification legislation, standards, and monitoring documents. Global Health Sci. Pract. 6 (2), 356–371.

Nathan, R., 1999. Regulation of Fortified Foods to Address Micronutrient Malnutrition: Legislation, Regulations and Enforcement, Manual. Flour Fortification Initiative, Atlanta.

Pachón, H., Reynolds, B., Duong, M., Tsang, B.L., Childs, L., Luthringer, C.L., Kang, Y., Vasta, F.C., Codling, K., 2021. The potential contribution of fortified maize flour, oil, rice, salt, and wheat flour to Estimated Average Requirements and Tolerable Upper Intake Levels for 15 nutrients in 153 countries. Nutrients 13 (2), 579.

Peña-Rosas, J.P., De-Regil, L.M., Rogers, L.M., Bopardikar, A., Panisset, U., 2012. Translating research into action: WHO evidence-informed guidelines for safe and effective micronutrient interventions. J. Nutr. 142 (1), 197S–204S.

Rosenthal, J., Casas, J., Taren, D., Alverson, C.J., Flores, A., Frias, J., 2014. Neural tube defects in Latin America and the impact of fortification: a literature review. Publ. Health Nutr. 17 (3), 537–550.

Stern, F., 2021. Review of Mandatory Maize Flour, Wheat Flour, Rice, Oil, and Salt Fortification Legislation, Standards, and Monitoring Documents, Master of Public Health Capstone Project. Emory University, Atlanta.

WHO, FAO, 2006. Guidelines on Food Fortification With Micronutrients. World Health Organization, Geneva.

WHO, FAO, UNICEF, GAIN, MI, FFI, 2009. Recommendations on Wheat and Maize Flour Fortification. Meeting Report: Interim Consensus Statement. World Health Organization, Geneva.

WHO, 2014. Guideline: Fortification of Food-Grade Salt With Iodine for the Prevention and Control of Iodine Deficiency Disorders. World Health Organization, Geneva.

WHO, 2016. WHO Guideline: Fortification of Maize Flour and Corn Meal With Vitamins and Minerals. World Health Organization, Geneva.

WHO, 2018. Guideline: Fortification of Rice With Vitamins and Minerals as a Public Health Strategy. World Health Organization, Geneva.

Food labeling

Barbara Schneeman, University of California, Davis, CA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of K.G. Grunert, Nutrition Labeling, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 315–319, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00200-2.

Introduction 301
Mandatory and voluntary labeling 302
Approaches to nutrition labeling 302
Nutrient declaration 302
Nutrient-related claims 303
Supplemental nutrition information 303
Conclusion/outlook 304
References 304
Relevant websites 305

Key points

- Food labeling is categorized as mandatory information or voluntary information and all labeling must be truthful and not misleading. Food labeling on prepackaged foods should be designed to inform the consumer about the appropriate use of the food in their overall diet, including the nutritional attributes of the food. Labeling that includes nutrition information is important for assisting consumers in understanding what foods to limit and what foods to get enough of for a healthful dietary pattern.

Glossary

Daily Value (or DV) The term for NRVs used in Nutrition Facts on packaged food in the United States

FOPL Front of package labeling refers to nutrition information that is placed on the principal display panel of a food product

NRV Nutrient reference values developed by *Codex Alimentarius* for use in labeling food products. The reference values can be used to estimate the contribution of the food to meeting the recommended intake of the nutrient. NRVs have been developed for essential nutrients as well as for nutrients or food components whose intake should be limited

Nutrition-related claims Nutrition claims used on food packages can be based on the nutrient profile of a product (nutrient content claim), the role of a food, food component, or nutrient in reducing risk for disease (health claims or disease risk reduction claims) or the role of a food, nutrient, or food component in maintaining the structure or function of the body. Different terminology is used in different countries

PDP Principal display panel is the portion of the package label most likely to be seen by the consumer at the time of purchase and contains required label information such as the identity of the food and net quantity statements. The PDP may also contain promotional material about the product. The information panel is adjacent to the PDP with additional information such as nutrient and ingredient declarations

Introduction

The labeling of prepackaged foods provides information that assists consumers in the purchase and use of the product. Certain elements of food labeling are considered mandatory for most prepackaged foods in many countries and must appear in a format and style specified by government regulations or standards. Several elements of food labels are considered voluntary and used at the discretion of the manufacturer but are still governed by regulations or standards. The regulations of most countries, including the standards and guidelines developed by the *Codex Alimentarius* Commission, specify that all labeling, whether mandatory or voluntary, should be truthful and not misleading and presented in a manner that is not deceptive to the consumer.

Mandatory and voluntary labeling

The Codex standards and guidelines are useful to examine because they are adopted by many countries and reflect common practices among countries for the labeling of pre-packaged foods. This section discusses mandatory and voluntary labeling as established by Codex and the next section describes approaches to nutrition labeling, which is mandatory in some countries but in many countries is voluntary, unless a nutrition claim is made on the product.

Mandatory labeling refers to the information that should appear on all packaged foods unless a specific exemption is provided. Such mandatory elements, specified by the Codex General Standard for the Labeling of Prepackaged Foods, include the name of the food, a list of ingredients, net contents and drained weight, name and address of the manufacturer or distributor, country of origin, identification of the lot, date marking and storage instructions, and instructions for use (Codex General). Each of these elements requires additional specification at the national level to provide information that is relevant to the population. For example, the name of the food should include a common or usual name not just a fanciful or trade name for the product; ingredients are listed in decreasing order by weight; and guidelines are needed for the size and placement of font for these mandatory elements so that the information is legible for the consumer. Additional details can be found in the *Codex Alimentarius* General Standard for Labeling (Codex General) or in guidelines published by various government authorities (e.g. A Food Labeling Guide from the US FDA[1] (Guidance for Industry, 2013) or The Food Labeling Rules from the European Union (Food Labeling)).

Certain claims or representations about the food may result in additional mandatory declarations. For example, emphasis on a particular ingredient in a food may require a quantitative declaration of the ingredient. Treatment of the food with ionizing radiation requires inclusion of a written statement. For many countries, nutrient declaration in labeling is voluntary; however, for most countries nutrient declaration becomes mandatory if a nutrition claim is used in labeling and in several countries nutrient declaration is mandatory for most prepackaged foods.

Approaches to nutrition labeling

There are 3 main approaches for providing nutrition-related information: nutrient declaration, nutrition-related claims, and supplemental nutrition information (Schneeman, 2020). Nutrient declaration became mandatory on most packaged foods in the United States with enactment of the Nutritional Labeling and Education Act (NLEA) of 1990. Such labeling was viewed as important to address the increasing prevalence of obesity and diet-related chronic diseases, such as diabetes and cardiovascular disease (U.S. Department of Health and Human Services, 1988). In addition to making nutrition information on food packages available to help consumers make better food choices, requiring nutrient declaration also provides an incentive to manufacturers to improve the nutritional profile of their products. The impact of nutrition labeling on product formulation was demonstrated when *trans* fatty acids was added to the list of mandatory nutrients for inclusion in Nutrition Facts. In the time frame between proposing and finalizing this requirement and the initiation of enforcement, the *trans* fatty acid levels were reduced substantially in the USA food supply (Doell et al., 2012).

Nutrient declaration

Nutrient declaration is intended to address food and nutrition public health issues in the population and thus help consumers make better choices at the point of purchase. The nutrients to be included in the declaration are specified in regulations or standards. For example, the Codex standard for nutrient declaration was updated in 2013 to require declaration of energy value, the amounts of protein, available carbohydrates, total fat, saturated fats, sodium, and total sugars. In addition, the Codex standard mandates declaration of any nutrient for which a nutrition or health claim is made and allows for declaration of the amount of any nutrient considered relevant for maintaining nutritional status in a country or region (Codex Alimentarius, 2013). Although the Codex guideline encourages that nutrient declaration be mandatory on all packaged foods, it is considered voluntary information, unless a nutrition or health claim is made. Several countries have mandated nutrient declaration on most packaged foods and some countries, such as the USA and Canada, also mandate the format to be used for this declaration. Countries that mandate nutrient declaration often include certain nutrients or food components in addition to what is specified in the Codex standard because the information is relevant to national public health issues. For example, in the USA, in addition to the declaration in the Codex standard, *trans* fat, cholesterol, dietary fiber, added sugars, vitamin D, calcium, iron, and potassium are also included in the updated nutrient declaration (Code of Federal Regulations).

To be useful for consumers, the nutrient declaration should be presented in context. One way of providing context is to declare the amount of the nutrient or food component per the amount of food that is customarily consumed (i.e. a standard serving of the food). Knowing the amount of nutrient or food component per serving can be used to put that amount in context as a percentage of the recommended amount for daily consumption (e.g. referred to as the % Daily Value on the Nutrition Facts used in the USA). Another approach for providing context is to report the amount of the nutrient per 100 g of food (or 100 mL of beverage). Both

[1]The FDA Food Labeling Guide has not yet been updated to include the updates to Nutrition Facts that are now part of the Code of Federal Regulations, 21 CFR 101.9 https://www.govinfo.gov/content/pkg/CFR-2021-title21-vol2/pdf/CFR-2021-title21-vol2-part101.pdf. 2021 edition.

approaches can be used to compare products, either based on the standardized serving size or the common metric of 100 g or mL (even when this is not the amount of the food that is customarily consumed).

Jurisdictions that mandate nutrition labeling typically consider situations that might exempt a product from bearing a nutrition label (e.g. a package size may be too small to bear a label). A product with limited nutritional content may be allowed to use a more minimal format that indicates it is not a source of certain nutrients.

Nutrient-related claims

The use of nutrition claims in food labeling is considered voluntary and used at the manufacturer's discretion; however, regulations and standards govern how and where they can be used on prepackaged foods. Codex provides guidelines for 4 types of nutrition claims: nutrient content claims, nutrient comparative claims, maintaining health claims (referred to as structure-function claims in the USA), and disease-risk reduction claims (typically referred to as health claims in the USA) (Codex Alimentarius, 2009).

Nutrient content claims characterize the nutrient profile of the food and, such characterization typically depends on the amount per serving compared to an established reference value for the nutrient. Such reference values are referred to as Nutrient Reference Values (NRV) by Codex or as Daily Values (DV) in the USA. For example, according to the Codex guidelines on the use of nutrition and health claims, a serving should provide 15% of the NRV to claim that the food is a source of a specific vitamin or mineral and to be high in the nutrient would require twice the amount (Codex Alimentarius, 2009). The % of NRV to qualify for a nutrient content claim varies among governments. NRVs and DVs have been established for essential nutrients as well as for certain food components associated with chronic disease risk (e.g. saturated fatty acids or dietary fiber). The information on the % of the Daily Value is useful to understand the contribution of a food to the total daily value set for the nutrient or food component. Education programs in the USA emphasize that 20% or more of the DV is considered a lot of the nutrient or food component, whereas 5% of less of the DV is considered a small amount. For nutrients or food components to limit (e.g. sodium, added sugars, saturated fats), the selection of foods with a high DV should be limited.

A comparative claim compares the nutrient content of 2 or more foods. As an example, in the USA a product that claims to have a reduced sodium content should have its sodium content reduced by 25% of the relevant comparison product. Because nutrient content claims and comparative claims are profiling foods in relation to recommended intakes, such claims should be authorized for use in labeling so that consumers are provided with reliable, consistent information when such claims are made. In addition, to establishing appropriate reference values, regulations for nutrient content claims can also require certain disclosure statements regarding the overall nutrient profile of the product. For example, a dairy product that is high in calcium may need to refer consumers to Nutrition Facts if the product is also high in total fat or saturated fats.

Claims about maintaining health or the structure and function of the body characterize the role of the nutrient in metabolism, growth and development, or typical functions of the body. The European Food Safety Authority (EFSA) in the European Union reviews such claims before they can be used in food labeling (Health and Nutrition Claims); however, in the USA manufacturers can use such claims without prior approval but all statements must be truthful, not misleading, and be substantiated. In addition, such claims cannot imply a role in reducing risk for disease.

Disease-risk reduction claims characterize the relationship between a substance, which can be a food, nutrient, or food component, and the reduced risk of developing a disease or health-related condition. Such claims are not intended for curing, treating, mitigating, managing, or preventing disease but about reducing risk. In countries that allow the use of disease-risk reduction claims, such claims are authorized after a scientific review of relevant evidence and conditions, including the food's nutrient profile, may be set on the type of products that can use the claim in labeling (US Food and Drug Administration, 2009).

Supplemental nutrition information

In recent years, several government authorities have developed nutrition information schemes that are used on the principal display panel (PDP) of food packages and are referred to as front-of-pack labeling (FOPL). The rationale for labeling on the PDP is to make certain types of information quick to find and balance other types of promotional information on the PDP. Generally, this type of labeling is most useful when it is used to supplement the nutrient declaration, which provides a more complete summary of the nutritional profile of a product and can be used to verify information on the PDP (Schneeman, 2020).

The schemes for FOPL fall into several categories, which include interpretive systems that are nutrient-specific such as traffic lights or stop signs; declaration of nutrients, including the % DV on the PDP such as the Facts-up-Front system developed by industry groups; summary indicator systems such as Nutri-score that is based on both nutrients to limit and foods and nutrients to encourage; hybrid systems such as the Health star rating that combines a summary score with nutrient declaration on the PDP; and endorsement logos such as the Keyhole symbol developed by Nordic countries or the Heart Tick from Heart Associations/Foundations in several countries. The nutrient-specific systems typically were developed to provide warnings about excess amount of nutrients or food components to limit such as added sugars, sodium, *trans* fatty acids, and saturated fats. Examples of these systems include the traffic lights developed in the UK, the stop sign symbol used in Chile (Reyes et al., 2019), or the Canadian food package symbol for saturated fats, sodium and/or sugars. Other systems are designed to provide an overall score or profile of the food product that reflects nutrients and food components to limit as well as nutrients and food groups that are recommended to get enough of. Examples of these systems include the Healthy Star rating developed by Australia and New Zealand or the Nutri-score developed by France. Endorsement schemes are used to indicate that the product conforms to a specific nutrient profile to bear

the logo, such as the Nordic keyhole system. A challenge for these systems is to determine the approach for calculating the content of nutrients or food components within the food to enable meaningful comparisons among products. Some systems are based on a typical serving size, while other systems calculate the content per 100 g. For typical serving size to provide an accurate estimate for labeling purposes, standards for the amount customarily consumed need to be available. Presenting information per 100 g has the potential to be misleading if actual amounts of the food consumed are much smaller than 100 g (Visioli et al., 2022).

Because supplemental nutrition labeling systems have different purposes, effort is needed to design and implement an approach that is most useful within the targeted audience. In addition, education is useful so that consumers understand how to identify the supplemental labeling system as distinct from promotional material on the PDP and interpret the symbols appropriately. Many of these systems are useful when consumers are trying to compare the nutritional qualities within a category of products (e.g. comparing breakfast cereals) and some systems are used to indicate foods that should be limited in the overall dietary pattern (e.g. sugar-sweetened beverages). For additional details see the websites provided below, Schneeman (2020) and Institute of Medicine of the National Academies (2012).

Conclusion/outlook

The regulation and standardization of labeling requirements provides valuable tools for consumers to make more informed decisions on the purchase and appropriate use of prepackaged foods. Nutrition labeling is an example of information that consumers generally cannot obtain unless it is provided by the manufacturer or distributor of the product and is essential for consumers to manage the healthfulness of their food choices. The labeling of food products has been expanding and now includes additional information that relates to various consumer preferences such as sustainability, use of genetically engineered ingredients, agricultural practices (organic), the concept of "natural" as well as practices such as kosher, halal, or vegan. In all cases such labeling must be truthful and not misleading and conform to government standards or regulations, where they have been promulgated. The promulgation of regulations specific to a certain representation or claims in labeling is more likely to occur when a framework is needed to provide clear, well-defined standards for use of terms to avoid misuse that becomes misleading for consumers. For example, in the USA, regulations now exist for designating a product as organic and or when a declaration on the use of genetically engineered ingredients is required.

Nutrition labeling has primarily focused on information related to the content of nutrients and food components that are essential or related to risk of chronic diseases. However, food-based dietary guidelines emphasize the importance of dietary patterns and food groups to meet these nutritional needs. Certain supplemental nutrition labeling schemes incorporate recommendations for food groups into their scoring algorithms (e.g. Healthy Star or Nutri-Score); however an overall score may not provide specific information on how the product contributes to food group recommendations at the point of purchase. Developing a labeling system to communicate the contribution of a product to meeting food group recommendation in dietary patterns could assist consumers in matching product information with food-based dietary guidelines.

References

21 Code of Federal Regulations (CFR) 101.9. https://www.govinfo.gov/content/pkg/CFR-2021-title21-vol2/pdf/CFR-2021-title21-vol2-part101.pdf. 2021 ed.
Codex Alimentarius, 2009. Guidelines for Use of Nutrition and Health Claims CAC/GL 23-1997. WHO/FAO, Geneva.
Codex Alimentarius, 2013. Guidelines of Nutrition Labeling CAC/GL 2-1985. WHO/FAO, Geneva.
Codex General Standard for the Labeling of Prepackaged Foods. Codex Stan 1-1985 (Rev 1-1991), https://www.fao.org/3/Y2770E/y2770e02.htm, FAO.
Doell, D., Folmer, D., Lee, H., Honigfort, M., Carberry, S., 2012. Updated estimate of trans fat intake by the US population. Food Addit. Contam. 29, 861–874.
Food Labeling Rules. European Union, https://europa.eu/youreurope/business/product-requirements/food-labelling/general-rules/index_en.htm. Accessed Nov 2021.
Food and Drug Administration A Food Labeling Guide: Guidance for Industry, 2013. U.S. Department of Health and Human Services. https://www.fda.gov/files/food/published/Food-Labeling-Guide-%28PDF%29.pdf. (Accessed December 2021).
Health and Nutrition Claims, European Union. https://europa.eu/youreurope/business/product-requirements/food-labelling/health-nutrition-claims/index_en.htm. Accessed Nov 2021.
Institute of Medicine of the National Academies, 2012. Committee on Examination of Front-Of-Package Nutrition Rating Systems and Symbols (Phase II). Front-Of-Package Nutrition Rating Systems and Symbols: Promoting Healthier Choices. National Academy Press, Washington.
Reyes, M., Garmendia, M.L., Olivares, S., Aqueveque, C., Zacarías, C., 2019. Development of the Chilean front-of-package food warning label. BMC Publ. Health 19, 906–917. https://doi.org/10.1186/s12889-019-7118-1.
Schneeman, B.O., 2020. Nutrition labeling for foods: which approaches are useful?. In: Biesalski, H.K. (Ed.), Hidden Hunger and the Transformation of Food Systems. How to Combat the Double Burden of Malnutrition? World Rev Nutr Diet, vol. 121. Karger, Basel, pp. 166–175. https://doi.org/10.1159/000507487.
U.S. Department of Health and Human Services, 1988. The Surgeon General's Report on Nutrition and Health. Government Printing Office, Washington DC.
US Food and Drug Administration, 2009. Guidance for Industry: Evidence-Based Review System for the Scientific Evaluation of Health Claims. FDA, Washington. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-evidence-based-review-system-scientific-evaluation-health-claims. (Accessed December 2021).
Visioli, F., Franco, M., Matínez-González, M.A., 2022. Front of package labels and olive oil: a call for caution. Eur. J. Clin. Nutr. 76, 502–504.

Relevant websites

21 CFR 101.9. https://www.govinfo.gov/content/pkg/CFR-2021-title21-vol2/pdf/CFR-2021-title21-vol2-part101.pdf. 2021 ed.
Codex Alimentarius, The International Food Standards. INFOSAN Information Note No. 4/2008-Codex Alimentarius. https://www.who.int/foodsafety/fs_management/No_04_Codex_Jun08_en.pdf. Accessed 16 Nov 2021. WHO.
Health Canada Front-of-Pack Nutrition Labeling. https://www.canada.ca/en/health-canada/news/2022/06/front-of-package-nutrition-labelling.html.
Report on the Use of Nutri-score. https://ec.europa.eu/health/sites/default/files/nutrition_physical_activity/docs/ev_20171130_co04_en.pdf.
Report on the Use of the Traffic Light in the UK. https://www.food.gov.uk/safety-hygiene/check-the-label.
Report on Use of the Nordic Keyhole Logo. https://ec.europa.eu/food/system/files/2018-04/comm_ahac_20180423_pres2.pdf.
The Healthy Star Program Developed by FSANZ. http://www.healthstarrating.gov.au/internet/healthstarrating/publishing.nsf/Content/Home.

Food security

Saskia de Pee[a] **and Rafael Pérez-Escamilla**[b], [a] Systems Analysis for Nutrition, the United Nations World Food Program, Rome, Italy; and [b] Department of Social and Behavioral Sciences, Yale School of Public Health, New Haven, CT, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of S. de Pee, Food Security, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 353–360, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00127-6.

Introduction 307
Framework and definitions 307
Food security 307
Meeting nutrient needs 307
Food security as a component of livelihood security 308
Indicators and classification of food security status 309
International food security phase classification (IPC) for countries 309
Indicators of nutrient intake at the national, household, and individual level 309
Experience-based food insecurity scales 311
Household Food Insecurity Access Scale 311
Dietary diversity 312
Common pitfalls of analyzing causes of malnutrition 312
Factors affecting food security 312
Consequences of food insecurity 313
Mitigating the consequences of food insecurity 314
Conclusions 314
References 315
Further reading 315

Key points

- Food insecurity affects hundreds of millions of households across world regions
- Poverty and social injustice are the main underlying factors driving the food insecurity pandemic
- Food insecurity has a strong negative impact on the physical and mental health of people
- Food insecurity has a strong negative impact on planetary health
- Household food insecurity is strongly linked with household water insecurity
- Addressing food insecurity is central to the ability of nations to meet the Sustainable Development Goals by 2030
- Experience-based food insecurity scales should be used more to track food insecurity globally

Glossary

Cost of the diet A linear programming tool that estimates the lowest cost diet that meets all nutrient requirements of individuals of a household, using data on which foods are locally available, their nutrient content and their price. This cost estimate can be compared to income or expenditure data to determine which proportion of the population would be able to afford an adequately nutritious diet.

Dietary diversity Extent to which the diet includes foods from different food groups. The greater the diversity, the more likely the diet meets nutrient requirements

Food security When all people, at all times, have physical and economic access to sufficient, safe and nutritious food to meet their dietary needs and food preferences for an active and healthy life.

Livelihood A livelihood comprises the capabilities, assets (stores, resources, claims and access), and activities required for a means of living

Malnutrition Malnutrition is a broad term commonly used as an alternative to undernutrition but technically it also refers to "overnutrition". Please refer to definition of undernutrition below. People are also malnourished if they are overweight or obese as a result of consuming too many calories, while intake of vitamins and minerals may still be inadequate. Overweight and obesity increase the risk of cardiovascular disease and diabetes

Undernourishment The number of undernourished people is estimated at national level, and aggregated at global level, and represents the number of people that have to survive on less than 2100 kcal day^{-1}, which is estimated from food balance sheets (food production minus export plus import)
Undernutrition Undernutrition is defined as the outcome of insufficient food intake and repeated infectious diseases. It includes being underweight for one's age, too short for one's age (stunted), thin for one's height (wasted) and deficient in vitamins and minerals (micronutrient malnutrition)
Window of opportunity The first 1000 days of a child's life, from conception until 24 months of age, during which ensuring appropriate nutrition is essential to enable the child to have the best start in life and develop to its full potential

List of abbreviations

DFID United Kingdom Department for International Development
EBIA Escala Brasileira de Insegurança Alimentar
ELCSA Escala Latinoamericana y Caribeña de Seguridad Alimentaria
FANTA Food and Nutrition Technical Assistance
FIES Food Insecurity Experience Scale
HFIAS Household Food Insecurity Access Scale
IPC International food security Phase Classification
UNICEF United Nations Children's Fund
USHFSSM US Household Food Security Survey Module
WFP United Nations World Food Program

Introduction

Food security is a very important determinant of whether people can lead an active and healthy life, because it determines their access to foods required to meet nutrient needs. This article reviews the definition of food security, the consequences of food insecurity, the indicators used to measure food security depending on the level at which it is studied, how it links to nutrition and health as well as to livelihoods, what it is affected by, the consequences of food insecurity, and measures that are taken to mitigate these causes and consequences. Special attention will be paid to why it is important that food security assessments also include an estimate of the extent to which nutrient needs are being met, and the approaches and indicators, which can be used for that purpose.

Framework and definitions

Food security

Food security was defined, at the World Food Summit in 1996, as "when all people, at all times, have physical and economic access to sufficient, safe and nutritious food to meet their dietary needs and food preferences for an active and healthy life".

Food insecurity affects hundreds of millions of households across world regions. Poverty and social injustice are the main underlying factors driving the food insecurity pandemic Food insecurity has a strong negative impact on the physical and mental health of people and of the planet (Pérez-Escamilla, 2017; Pérez-Escamilla et al., 2020b).

Food security has many dimensions and can be studied at different levels. The different dimensions include food availability, access to, and utilization of food. Recognizing the fact that food security is not constant the stability of the food supply is also an important dimension of food security. The utilization of food refers to what happens with food at the household level, i.e., allocation among different members, preparation, and utilization by the body, i.e., digestion, bioavailability, etc., which is closely linked to health status as well as age and physiological status. The different levels at which food security can be studied include individual, household, community, district, province, country, and region. Depending on the level of focus, different dimensions, aspects, and indicators can be used.

Meeting nutrient needs

It is important to note that the definition makes explicit reference to safe and nutritious food to meet dietary needs, i.e., meeting every individual's nutrient requirements for leading an active and healthy life. Whether an individual's nutrient requirements are met depends on food consumption as well as on disease, as the latter increases nutrient needs and also affects the way nutrients are metabolized and used by the body.

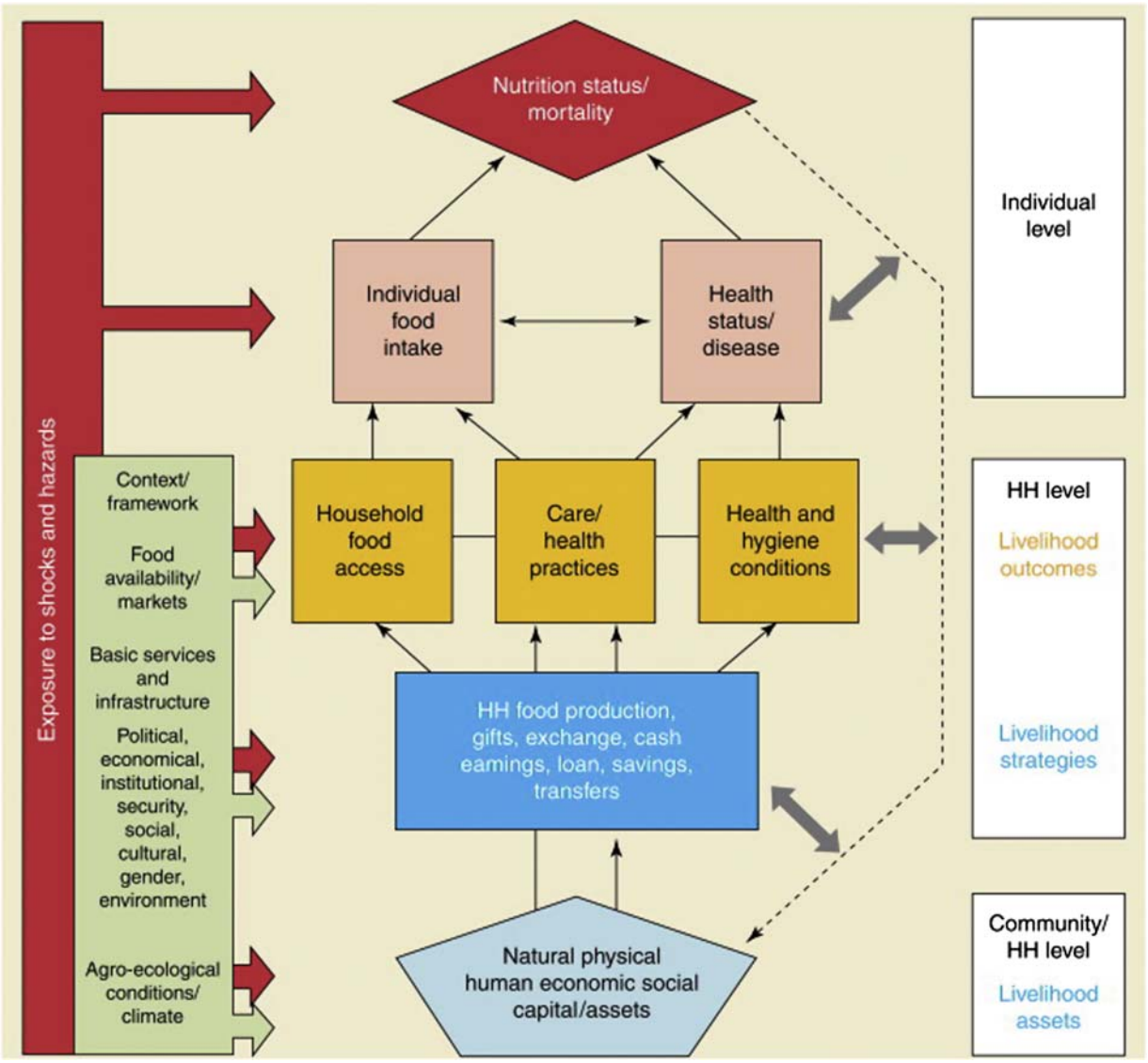


Fig. 1 WFP Food and Nutrition Security Conceptual Framework (based on UNICEF conceptual framework for causes of malnutrition and DFID sustainable livelihoods framework). Reprinted with permission from WFP (2009).

The United Nations Children's Fund (UNICEF) framework for causes of malnutrition is focused on the individual level (see Fig. 1 in which it has been integrated). It specifies that direct causes of malnutrition are food consumption and health or disease; underlying causes are access to food, caring practices, hygiene conditions, and health-care services; basic causes are human, economic, and organizational resources and controls, including education, governance, etc. Whereas access to food determines food consumption and thus nutrient intake, and hygiene conditions and health-care services determine health and disease and thus nutrient needs and utilization, caring practices affect how accessible food is used, i.e., food distribution among household members, food preparation, feeding frequency, breastfeeding practices, and how health-care services are utilized. Simply put, caring practices reflect the choices made from among the options available to the individual or the household, and are also influenced by knowledge and empowerment.

Food security as a component of livelihood security

The United Nations World Food Program (WFP) Food and Nutrition Security Conceptual Framework proposed in 2009 (see Fig. 1) combines the UNICEF framework and the United Kingdom Department for International Development (DFID) Sustainable Livelihoods Framework, because food security is closely linked to household livelihood security, which can be described as the ability of a household to meet the basic needs of it members. Livelihood can be described as comprising the capabilities, assets (stores, resources, claims, and access), and activities required for a means of living. A livelihood is sustainable when it can cope with

and recover from stress and shocks, maintain or enhance its capabilities and assets, and provide sustainable livelihood opportunities for the next generation; when it contributes net benefits to other livelihoods at the local and global levels in the long and short term. The livelihood unit is usually the household. Household livelihood security is defined as adequate and sustainable access to income and resources to meet basic needs. Basic needs include food, proper nutrition, clean water, health and health facilities, economic and educational opportunities, housing, physical safety, and time for community participation and social integration. Households use six main tangible and intangible forms of capital, i.e., human (skills, knowledge, health, and nutritional status), financial (income, credits, savings, and liquid assets), natural (crops grown), physical (assets and land), social (support networks), and political (participation in community decisions, power).

As shown in the framework, we distinguish livelihood assets, strategies and outcomes, and the outcomes that affect an individual's food consumption (i.e., nutrient intake), health or disease state (i.e., nutrient needs and nutrient utilization), and care and health practices (i.e., choices from among available options).

Policies, institutions, and organizations affect livelihood assets and strategies. Policies can be split into the following categories: macroeconomic policies, i.e., measures aimed at stabilizing an economy; social policies, which aim to protect and improve health, nutrition, education of the disadvantaged; sectoral policies that focus on specific areas within an economy, such as agriculture, health, water, sanitation, and environment. At the interface between these policies and households are the institutions and organizations that implement or affect the policies, i.e., the state, formal civil society, informal civil society, and the private sector.

Shocks and hazards, which can range from death of a family member to increased food prices, droughts, floods or armed conflict, can affect different and multiple dimensions of the framework, with different intensity and for a variable period of time.

Thus, the Food and Nutrition Security Conceptual Framework shows how many different factors affect food availability, food access, food utilization and stability of food supplies, which together determine food security, and how food security in turn is a component and function of livelihood security, which is a function of livelihood assets and strategies and is affected by both macro- and microlevel factors. It is also important to note the rich multidisciplinary nature of the topic of food security.

Indicators and classification of food security status

The level at which food security is of interest, i.e., household, community, district, or country, determines which factors should be assessed.

International food security phase classification (IPC) for countries

For example, the IPC provides guidance to countries for classifying a country's regions and subregions. It emphasizes that as much as possible existing and relevant information should be used, and that there is no fixed set of indicators and methods to refer to.

Table 1 shows the five stages of food security that are distinguished in this IPC classification together with proposed indicators and their cut-offs for the different stages. Some indicators reflect context and livelihood assets, such as civil security, "structural" (underlying hindrances to food security), livelihood assets/capital; some reflect vulnerability, such as hazards, destitution/displacement; others reflect underlying causes of malnutrition such as water access/availability, food access/availability, dietary diversity; the top ones reflect outcome, i.e., crude mortality rate ($\times$/10,000 day^{-1}), disease, acute malnutrition (wasting), and stunting. Some of these apply to the national or subnational situation (civil security, hazards, and structural hindrances), whereas others have to be collected from individuals (nutritional status) or households (food and water access). This means that information from different sources needs to be gathered, reviewed, and classified in order to arrive at a food security classification of a country's regions or subregions.

The IPC partners have substantial experience with the collection and interpretation of these indicators and their classification. Here, we will focus on the aspects of food security that are most directly linked to food intake and nutrient adequacy, because these are most closely linked to the outcomes, nutritional status, and health.

Indicators of nutrient intake at the national, household, and individual level

Nutritional status and health are the ultimate outcomes of food security that are very much of interest because they determine human capital at present and in the future. It is very important to recognize the fact that foods are a source of nutrients and that the human body requires approximately 40 different nutrients for growth, development, and health. Reaching an adequate intake for each of these nutrients requires consumption of a diverse diet, including plant and animal source foods as well as fortified foods. Also different groups in the population require these nutrients in different amounts, depending on their age (growth spurts), physiological status (pregnancy and menstruation), health, and physical activity.

The indicators of food intake in the IPC classification are food access/availability, expressed as kcal capita^{-1} day^{-1}, and dietary diversity, expressed as deficient or sufficient. These two specific indicators are usually collected either at the household level or, in the case of kcal capita^{-1} day^{-1}, estimated at the national level. The latter indicator is also used for Millennium Development Goal no. 1, i.e., the proportion of the population that is undernourished, which presents the proportion of the population that has to survive on less than 2100 kcal capita^{-1} day^{-1} and is estimated from food balance sheet data (i.e., total kcal available from food production and

Table 1 International food security phase classification reference table.

Phase classification		*Key reference outcomes current or imminent outcomes on lives and livelihoods. Based on convergence of direct and indirect evidence rather than absolute thresholds. Not all indicators must be present for classification.*	
1	Generally food secure	Crude mortality rate	<0.5/10,000/day
		Acute malnutrition	<3% (w/h < −2 z-scores)
		Stunting	<20% (h/age < −2 z-scores)
		Food access/availability	Usually adequate (>2100 kcal ppp day), stable
		Dietary diversity	Consistent quality and quantity of diversity
		Water access/avail.	Usually adequate (>15 L ppp day), stable
		Hazards	Moderate to low probability and vulnerability
		Civil security	Prevailing and structural peace
		Livelihood assets	Generally sustainable utilization (of 6 capitals)
2	Moderately/borderline food insecure	Crude mortality rate	<0.5/10,000/day; U5MR < 1/10,000/day
		Acute malnutrition	<3% but <10% (w/h < −2-z scores), usual range, stable
		Stunting	<20% (h/age < −2 z-score)
		Food access/availability	Borderline adequate (2100 kcal ppp day); unstable
		Dietary diversity	Chronic dietary diversity deficit
		Water access/avail.	Borderline adequate (15 L ppp day); unstable
		Hazards	Recurrent, with high livelihood vulnerability
		Civil security	Unstable; disruptive tension
		Coping	"Insurance strategies"
		Livelihood assets	Stressed and unsustainable utilization (of 6 capitals)
		Structural	Pronounced underlying hindrances to food security
3	Acute food and livelihood crisis	Crude mortality rate	0.5–1/10,000/day, U5MR1−2/10,000/day
		Acute malnutrition	10−15% (w/h < −2 z-score), >than usual, increasing
		Disease	Epidemic; increasing
		Food access/availability	Lack of entitlement; 2100 kcal ppp day via asset stripping
		Dietary diversity	Acute dietary diversity deficit
		Water access/avail.	7.5–15 L ppp day, accessed via asset stripping
		Destitution/displacement	Emerging; diffuse
		Civil security	Limited spread, low intensity conflict
		Coping	"Crisis strategies"; CSI > than reference; increasing
		Livelihood assets	Accelerated and critical depletion or loss of access
4	Humanitarian emergency	Crude mortality rate	1−2/10,000/day, >2x reference rate, increasing; U5MR > 2/10,000/day
		Acute malnutrition	>15% (w/h < −2-z score), >than usual, increasing
		Disease	Pandemic
		Food access/availability	Severe entitlement gap; unable to meet 2100 kcal ppp day
		Dietary diversity	Regularly 3 or fewer main food groups consumed
		Water access/avail.	<7.5 L ppp day (human usage only)
		Destitution/displacement	Concentrated; increasing
		Civil security	Widespread, high intensity conflict
		Coping	"Distress strategies"; CSI significantly > than reference
		Livelihood assets	Near complete and irreversible depletion or loss of access
5	Famine/humanitarian catastrophe	Crude mortality rate	>2/10,000/day (example: 6000/1,000,000/30 days)
		Acute malnutrition	>30% (w/h < −2 z-score)
		Disease	Pandemic
		Food access/availability	Extreme entitlement gap; much below 2100 kcal ppp day
		Water access/avail.	<4 L ppp day (human usage only)
		Destitution/displacement	Large scale, concentrated
		Civil security	Widespread, high intensity conflict
		Livelihood assets	Effectively complete loss; collapse

Reprinted with permission from Integrated Food Security Phase Classification, 2008. IPC User Guide, Version 1.0. The IPC in the Central and Eastern Africa Region Project. Nairobi, FAO. Available from www.ipcinfo.org (accessed on January 17, 2011).

net import or export) compared to total population size. (The other indicator for MDG1 is underweight, i.e., the proportion of children aged 0–59 months that have a weight that is too low for their age (<−2 SD of the median of the reference population)).

However, nutrient needs are different for different individuals. For that reason, adequacy of food intake, in terms of energy (kcal), dietary diversity, and if possible also specific nutrients (such as micronutrients, protein, and fat), are best collected at the individual level and expressed per population group, in order to identify which groups are most at-risk from a nutritional point of view. However, collecting these data at the individual level is very labor intensive and requires detailed food composition

data. Furthermore, there are several ways of collecting and interpreting these data. Therefore, very often outcome indicators are collected at the individual level, i.e., nutritional status and health, together with one or more household level indicators of access to food and possibly dietary diversity.

Experience-based food insecurity scales

Experienced-based food insecurity scales have been found to be helpful for policy making (Pérez-Escamilla, 2012; Pérez-Escamilla et al., 2020a). The first scale of this nature, the US Household Food Security Survey Module (USHFSSM) was developed and applied by the US Government for the first time in 1995 (Wunderlich and Norwood, 2006) and since then it has been continuously used to monitor household food insecurity in the country. This module is based on an experience-based scale that asks a person that is knowledgeable of the food insecurity situation in the household responds to 18 questions (USDA, 2012). The issues addressed through these questions range from household members being worried about running out of food, to household members sacrificing dietary quality all the way to experiencing over hunger (i.e., going for a whole day without eating) as a result of lack of economic access to foods. Questions are asked in reference to a time period such as the 4 weeks, 3 months or 12 months preceding the survey and some of them are specific to the minors in the household. The successful experience with the USHFSSM led to its adaptation and validation across countries resulting in the Household Food Insecurity Access Scale (HFIAS) (Coates et al., 2007), Brazilian Food Insecurity Scale (EBIA) (Pérez-Escamilla et al., 2004), The Latin American and Caribbean Food Security Scale (ELCSA) (Comité Científico de la ELCSA, 2012), and the Food Insecurity Experience Scale (FIES) (Ballard et al., 2013), among others. In the next section the HFIAS is described in detail as an example of how experience-based scales look like.

Household Food Insecurity Access Scale

Food and Nutrition Technical Assistance (FANTA) has developed a Household Food Insecurity Access Scale (HFIAS) that builds on the experience with the US Household Food Security Module and the Radimer/Cornell scale. It has been well validated and is increasingly being used. **Table 2** shows the nine questions asked of households, which range from having experienced a small degree of food insecurity, i.e., worrying about not having enough food, to not having had any food for a consecutive 24 h period. The reference period is the past 4 weeks and each question is followed by a question about how frequently this specific experience occurred in the previous 4 weeks. This scale thus assesses access to food, mainly in terms of meeting self-perceived adequacy of quantity and whether there was deterioration from what the household is used to. It has recently evolved to a smaller subset of questions that are used to define the severity of household hunger (Household Hunger Scale). It is important to note that neither of the scales inquire about dietary diversity or individual foods consumed, nor do they estimate individual nutrient intake.

Table 2 Questions of the FANTA Household Food Insecurity Access Scale.

No.	*Question*
1	In the past 4 weeks, did you worry that your household would not have enough food?
2	In the past 4 weeks, were you or any household member not able to eat the kinds of foods you preferred because of a lack of resources?
3	In the past 4 weeks, did you or any household member have to eat a limited variety of foods due to a lack of resources?
4	In the past 4 weeks, did you or any household member have to eat some foods that you really did not want to eat because of a lack of resources to obtain other types of food?
5	In the past 4 weeks, did you or any household member have to eat a smaller meal than you felt you needed because there was not enough food?
6	In the past 4 weeks, did you or any household member have to eat fewer meals in a day because there was not enough food?
7	In the past 4 weeks, was there ever no food to eat of any kind in your household because of a lack of resources to get food?
8	In the past 4 weeks, did you or any household member go to sleep at night hungry because there was not enough food?
9	In the past 4 weeks, did you or any household member go a whole day and night without eating anything because there was not enough food?

Reprinted with permission from Coates et al. (2007).

Dietary diversity

Dietary diversity provides an indication of the potential adequacy of nutrient intake, because a diet that is sourced from a greater variety of food groups provides a wider range of nutrients (e.g., carbohydrates, fat, protein, fat soluble vitamins, water soluble vitamins, minerals, essential fatty acids, sulfur containing amino acids, etc.).

Methods for assessing dietary diversity ask how frequently, usually during the previous 7 day or 24 h, foods from different food groups were consumed. The number of food groups can range from 7 to 15, for example staples, lentils and legumes, oils and fats, leafy vegetables, colored vegetables, fruits, dairy products, eggs, meat and poultry, fish, and sugary foods and drinks.

These answers can be processed and used in different ways. For example, when the questions refer to a 7 day period, the total number of days that different food groups are consumed can be added together, thus if there are nine food groups, the total score would be $9 \times 7 = 63$, and households that consumed three foods every day would have a score of 21. One can also decide to count the food groups that provide more nutrients than just energy (i.e., in the above example leave out staples, oils and fats, and sugary foods and drinks), or calculate a score that reflects just the frequency of consumption of particular food groups, such as animal source foods. For categorizing the obtained scores into different levels of dietary diversity, knowledge of the local situation is required and preferably a relationship should have previously been demonstrated between differently derived scores and nutritional status, for example of children under 5 years, so that the best cut-offs can be chosen.

Whereas good correlations have been shown between household dietary diversity and household expenditure on different food categories as well as between household dietary diversity and children's nutritional status (stunting and micronutrient deficiencies), this relationship has not been established for every population from which dietary diversity data are being collected. This means that part of the use and interpretation of dietary diversity data has to refer to findings in other settings and be based on knowledge and judgment of the local situation.

Common pitfalls of analyzing causes of malnutrition

As shown above, food security is generally assessed at the household level or higher, i.e., at the community, subregion or country level, whereas its main outcome, nutritional status, is assessed at the individual level. Depending on which indicators are used to assess food security, the conclusion that food security is not a problem whereas stunting, for example, affects 30% of the children under five, is very common. In these cases, the cause of malnutrition is then often sought in the nonfood-related factors of the framework for causes of malnutrition, i.e., hygiene, health and caring practices, rather than in a deficient intake of specific nutrients. This reasoning fails to recognize that as long as the food security indicator does not assess adequacy of nutrient intake, but just focuses on access to food, which can be any food, "food" may still be the main problem causing malnutrition.

For this reason it is very important to include dietary diversity as an indicator of food security, and preferably not only at household level, but also for specific groups such as children under five, and to analyze the data in such a way that it is a good proxy for nutrient adequacy of the population group of interest. For example, in the case of under fives, they require animal source foods, including dairy foods, as well as fortified foods in their diet in order to meet their nutrient requirements, especially of bioavailable minerals and type II nutrients that are important for growth. The intake of foods from these groups should thus be specifically assessed and reported.

Another way of assessing whether food consumption could possibly meet nutrient requirements is through linear programming. Based on availability of foods, their nutrient composition and their price, the lowest cost diet that would meet all nutrient requirements of a family composed of different members, e.g., a child younger than 24 months, a school-age child, a lactating mother, a father and a grandparent, can be calculated. This price can then be compared to income or food expenditure data in order to determine what proportion of the population could in theory afford a diet that meets all nutrient requirements. It is important to note that in practice, it is likely that a larger proportion does not meet all the nutrient requirements, because food choices are also affected by cultural practices and taste preferences that may not necessarily concur with best choice from a nutrient content point of view. This method, which is also known as the Cost of the Diet tool, has been developed by Save the Children UK and is known as the Cost of the Diet tool.

Knowing the causes of malnutrition is a must for being able to address it. Thus, when the prevalence of malnutrition is high, it is important to assess the level of food insecurity, including dietary diversity, and possibly also whether households could in theory afford a diet that could meet the nutrient requirements of all its members at the lowest cost. In addition to these factors that largely determine food consumption and hence nutrient intake, factors that affect nutrient needs, i.e., health or disease, as well as caring practices, which affect the choices made from among the possibilities available, are also important.

Factors affecting food security

The conceptual framework shows that many different factors can affect livelihoods and food security, as well as the degree of vulnerability to changes thereof. Factors that affect food availability generally have to do with climate change, seasonality affecting crop production, or large-scale political changes such as wars. Food access is to a large extent determined by purchasing power, which is a function of income and food prices as well as of prices of other commodities such as fuel (for cooking and transport). The stability of food supplies is usually the first aspect of food security that changes and can be regarded as an early warning sign that livelihoods

and food security status are likely to deteriorate. Last but not least, food utilization is affected by household dynamics (intra-household distribution), behavioral factors, and health or disease (utilization by the body), both of which are less likely to change unless household structure and livelihood change substantially.

Relatively new and emerging challenges for food security include climate change and increasing natural disasters (affecting food availability) as well as high and fluctuating food prices and the global financial crisis (affecting access to food) and rapid urbanization (reliance on cash economy, smaller social networks to assist in coping with shocks, etc.).

With the food price increase in 2008, the global economic crisis in 2009, and the rising food prices in 2011, it is important to determine in which countries the population is at greatest risk of suffering from food insecurity, and to identify measures for mitigating the causes and consequences in the best possible way. It has been found that countries that were more linked to the global economy, more dependent on import and where a large proportion of the population were net-buyers of food, were more at-risk.

It is now clear that household food insecurity is strongly linked with household water insecurity and that both need to be addressed through better coordinated efforts (Young et al., 2021).

Consequences of food insecurity

The consequences of food insecurity are several as it has a strong negative impact on and is affected by the physical and mental health of people and of the planet (Pérez-Escamilla, 2017). The Covid-19 pandemic clearly illustrated the complex syndemic that food insecurity is a part of (Pérez-Escamilla et al., 2020b).

Very importantly, due to decreased dietary diversity and also quantity, meeting nutrient needs becomes more and more difficult and therefore nutritional status and health deteriorate. This has short-term consequences, such as increased morbidity and mortality, as well as long-term consequences because an entire generation of young children can be affected for the rest of their lives.

The latter is due to the fact that the window of opportunity for optimal development of a young child is concentrated within its first 1000 days of life, starting from conception until 24 months of age. When the child is undernourished in this period of his or her life, he or she is at increased risk of morbidity and mortality, and later in life of delayed school enrollment, poorer school performance including earlier drop-out, lower income earning potential, and higher risk of obesity and chronic disease including diabetes, cardiovascular disease, and obesity.

Food insecurity also has consequences for other livelihood priorities, such as education of children, health-care seeking behavior, asset ownership, etc., which in-turn also relates to vulnerability to shocks. Fig. 2 shows consequences for livelihood and food consumption when households descend into stress and distress, and the concurrent health consequences.

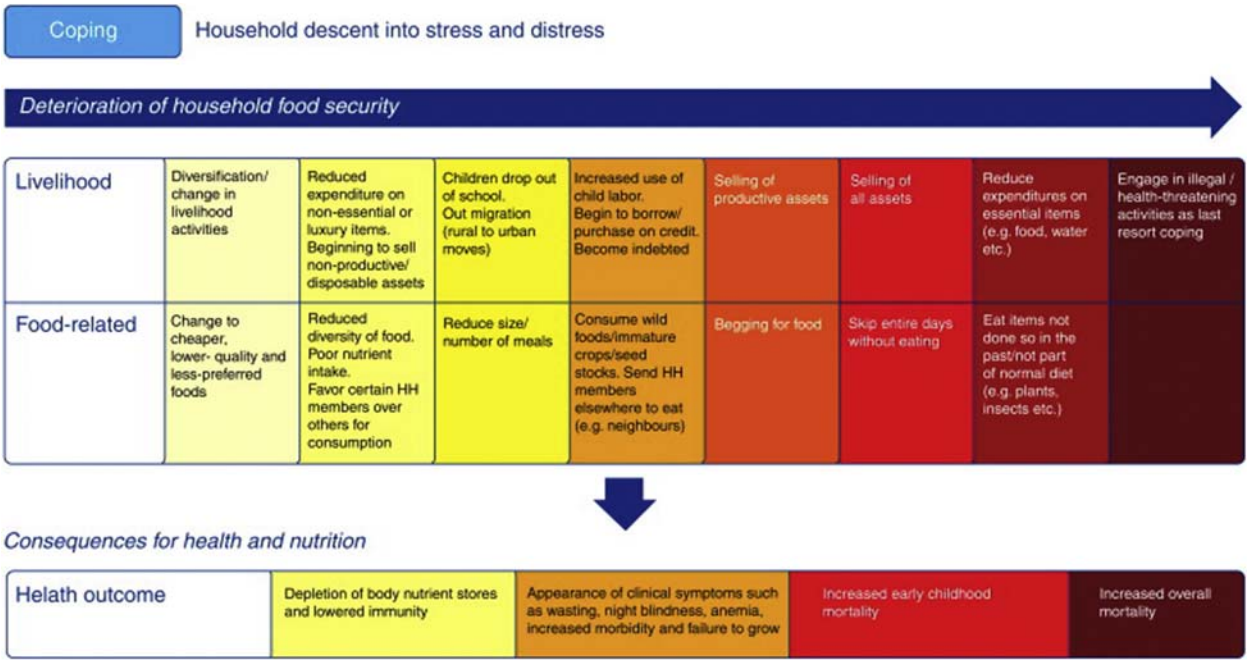


Fig. 2 Changes in livelihood, food selection and consumption, and nutritional status and health outcome when households descend into stress and distress. Reprinted with permission from Klotz, C., de Pee, S., Thorne-Lyman, A., Kraemer, K., Bloem, M.W., 2008. Nutrition in a perfect storm: why micronutrient malnutrition will be a widespread health consequence of high food prices. Sight Life Mag. 2, 6–13.

Health-care seeking and treatment adherence behaviors that are particularly affected by food insecurity are those of HIV/AIDS and tuberculosis, as populations affected by both overlap geographically. Firstly, these diseases are likely to increase food insecurity as household members fall ill and are hence not able to contribute to food production or earn an income. Secondly, their food insecurity affects diagnosis, treatment uptake and treatment adherence because that incurs costs, for example for transport, which hence competes with other priorities such as buying food for the family. Thirdly, some side effects of treatment such as nausea can be overcome by consumption of more palatable foods, which requires access to a variety of foods. Fourthly, recovery from malnutrition, which is what many people present with as they seek a diagnosis and enroll for treatment, requires treatment of the HIV and opportunistic infections as well as consumption of nutritious foods that need to be acquired.

Mitigating the consequences of food insecurity

Measures to mitigate food insecurity can be several and depend on the causes, on who is affected and how, and on what changes can be realized in the particular context. Depending on whether the deterioration of food security is related to food availability, access, utilization or stability of supplies, different measures can be taken.

The main underlying driver for food insecurity is poverty and social injustice (Pérez-Escamilla, 2017) through diverse mediating factors (WFP, 2009). If the problem is related to food production, i.e., food availability, large-scale multiyear agricultural measures may be required, for example to mitigate impact of climate change, increase yields, improve irrigation, develop storage, transport and markets for produce, etc. Although this is often focused on staple crops, there can also be a component focused on dietary diversity and household level food production including vegetables, fruits, small livestock, community fish ponds, etc. These types of interventions are, in addition to being income-generating, also more nutrition-oriented, especially when combined with an education component to emphasize the importance of consuming a diverse diet.

Food access problems are typically experienced by households that largely depend on cash purchases, and are thus vulnerable to changes in income and changes in food prices. Urban households as well as rural households that are net-buyers because they do not produce enough food throughout the year were hit hardest by the recent high food prices and the global financial crisis. They faced lower or loss of income, increased food prices, lower subsidies or higher taxes on food and fuel, and also reduced public spending on health, water and sanitation.

Mitigating these kinds of consequences can either be at the "blanket" level, i.e., by reducing prices, or removing tax, for specific staple foods (bread and maize), which apply to everyone, or they can target the most vulnerable, for example through safety-net programs that provide free or subsidized foods to households that are identified to be most in need. Such identification can for example be done through self-selection, i.e., participation in public work programs in return for which a basic food supply is provided, or through applying specific selection criteria such as female- or child-headed households, income below a percentage of the minimum wage, etc. The assistance provided can be in the form of food, vouchers for food, fuel or transport, or cash. Depending on the context, one or the other may be more suitable.

However, because of the far-reaching consequences of food insecurity and poor dietary diversity on nutrition and health, food assistance for such households should not only focus on meeting caloric needs and having enough meals per day, but also on meeting nutrient needs, especially of the most vulnerable, i.e., young children and pregnant and lactating women. This may require making specific products available that can increase their diet's content of specific nutrients. Such products can be in paste or powder form and if their aim is to improve intake of specific nutrients in addition to the prevailing, and affordable, home diet, an amount of 1–20 g day^{-1} (i.e., up to 125 kcal day^{-1}) should generally be sufficient.

Moving forward it is also key to ensure that household food and water security are examined together as they are strongly intertwined with each other (Young et al., 2021).

Conclusions

In summary, food security is achieved when all people, at all times, have physical and economic access to sufficient, safe and nutritious food to meet their dietary needs and food preferences for an active and healthy life. Food security has many different dimensions, i.e., food availability, food access, food utilization and stability of food supply, and is closely linked to livelihood security, i.e., the ability to provide for the basic needs of the family such as food, shelter, clean water, education, etc. The main underlying driver of food insecurity is poverty and social injustice. The fact that food security can be studied at many different levels, from the national to household or individual level, means that the outcome of food security, i.e., nutritional status and health, and food availability and access to food are often the areas of focus. However, when a population is affected by malnutrition, and access to food seems to be in order, it is very important to determine whether nutrient needs are actually being met. This can be done by including measures of dietary diversity, nutrient intake, food expenditure and ability to afford a lowest cost diet that could meet all nutrient requirements. It would be most appropriate to refer to "Food and Nutrient Security". Moving forward household food and water insecurity will need to be addressed simultaneously through well coordinated governance and actions.

References

Ballard, T.J., Kepple, A.W., Cafiero, C., 2013. Food Insecuriy Experience Scale: Development of a Global Standard for Monitoring Hunger Worldwide. Food and Agriculture Organization of the United Nations, Rome. Available athttps://www.fao.org/3/as583e/as583e.pdf. (Accessed 3 March 2022).

Coates, J., Swindale, A., Bilinsky, P., 2007. Household Food Insecurity Access Scale (HFIAS) for Measurement of Household Food Access: Indicator Guide (V. 3). FHI 360/FANTA, Washington, D.C. Available from. https://www.fantaproject.org/sites/default/files/resources/HFIAS_ENG_v3_Aug07.pdf

Comité Científico de la ELCSA, 2012. Escala Latinoamericana Y Caribeña De Seguridad Alimentaria (ELCSA): Manual De Uso Y Aplicaciones. FAO, Santiago, Chile. Available from. http://www.fao.org/3/i3065s/i3065s.pdf.

Pérez-Escamilla, R., Segall-Corrêa, A.M., Kurdian Maranha, L., Sampaio, Md, Mde, F., Marín-León, L., Panigassi, G., 2004. An adapted version of the U.S. Department of Agriculture Food Insecurity module is a valid tool for assessing household food insecurity in Campinas, Brazil. J. Nutr. 34 (8), 1923–1928.

Pérez-Escamilla, R., Vilar-Compte, M., Gaitán-Rossi, P., 2020a. Why identifying households by degree of food insecurity matters for policymaking. Glob. Food Sec. 26, 100459. https://doi.org/10.1016/j.gfs.2020.100459.

Pérez-Escamilla, R., Cunningham, K., Moran, V.H., 2020b. COVID-19 and maternal and child food and nutrition insecurity: a complex syndemic. Matern. Child Nutr. 16 (3), e13036. https://doi.org/10.1111/mcn.13036.

Pérez-Escamilla, R., 2012. Can experience-based household food security scales help improve food security governance? Glob. Food Sec. 1 (2), 120–125. https://doi.org/10.1016/j.gfs.2012.10.006.

Pérez-Escamilla, R., 2017. Food security and the 2015–2030 sustainable development goals: from human to planetary health: perspectives and opinions. Curr. Dev. Nutr. 1 (7), e000513. https://doi.org/10.3945/cdn.117.000513.

United States Department of Agriculture (USDA), 2012. U.S. Household Food Security Survey Module: Three-Stage Design, With Screeners Economic Research Service. Available from. https://www.ers.usda.gov/media/8271/hh2012.pdf.

WFP, 2009. Comprehensive Food Security and Vulnerability Analysis Guidelines. WFP, Rome, Italy. Available from. www.wfp.org.

Wunderlich, G.S., Norwood, J.K., 2006. History of the development of food insecurity and hunger measures. In: Food Insecurity and Hunger in the United States: An Assessment of the Measure. National Academies Press, Washington, DC, pp. 23–40.

Young, S.L., Frongillo, E.A., Jamaluddine, Z., Melgar-Quiñonez, H., Pérez-Escamilla, R., Ringler, C., Rosinger, A.Y., 2021. Perspective: the importance of water security for ensuring food security, good nutrition, and well-being. Adv. Nutr. 12 (4), 1058–1073. https://doi.org/10.1093/advances/nmab003.

Further reading

Bloem, M.W., de Pee, S., 2010. Nutrition and food insecurity in relation to HIV and AIDS and tuberculosis. Food Nutr. Bull. 31, S287–S364.

Bloem, M.W., de Pee, S., Semba, R.D., 2008. How much do data influence programs for health and nutrition?—experience from health and nutrition surveillance systems. In: Semba, R.D., Bloem, M.W. (Eds.), Nutrition and Health in Developing Countries, second ed. Humana Press, Totowa, NJ.

Deitcher, M., et al., 2011. Introducing a Simple Measure of Household Hunger for Cross-Cultural Use. FANTA Technical Note No. 12. Food and Nutrition Technical Assistance (FANTA) Project, Academy for Educational Development, Washington, DC. Available from. www.fanta.org.

FAO, IFAD, UNICEF, WFP and WHO, 2021. The State of Food Security and Nutrition in the World 2021. Transforming Food Systems for Food Security, Improved Nutrition and Affordable Healthy Diets for All. FAO, Rome. https://doi.org/10.4060/cb4474en. Available from. http://www.fao.org/3/cb4474en/cb4474en.pdf.

Integrated Food Security Phase Classification (IPC), 2008. IPC User Guide, Version 1.0. The IPC in the Central and Eastern Africa Region Project. FAO, Nairobi. Available from. www.ipcinfo.org. (Accessed 17 January 2011).

Ruel, M.T., Garrett, J.L., Haddad, L., 2008. Rapid urbanization and the challenges of obtaining food and nutrition security. In: Semba, R.D., Bloem, M.W. (Eds.), Nutrition and Health in Developing Countries, second ed. Humana Press, Totowa, NJ.

Semba, R.D., Ross, A.C., 2010. The Impact of Climate Change, the Economic Crisis, and the Increase in Food Prices on Malnutrition. Papers presented at the workshop "The impact of the climate change, the economic crisis, and the increase in food prices on malnutrition" held in Castel Gandolfo, Italy, January 25, 2009 J. Nutr. 140, 130S–228S.

Harmonizing the approach to deriving nutrient requirements

Ann L. Yaktine, Food and Nutrition Board, Health and Medicine Division, The National Academies of Sciences, Engineering, and Medicine, Washington, DC, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of A.A. Yates, Nutrient Requirements: International Harmonization, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 209–219, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00192-6.

Introduction and historical background 317
Conceptual basis for nutrient reference values 317
Overview 317
Risk assessment framework 317
Nutrient intake distribution model 319
Nutrient intake and consideration of risk of chronic disease 319
Inferring diet-disease relationships in chronic disease 319
Challenges to consideration of chronic disease endpoints 320
Updating the DRI framework for chronic disease endpoints 320
Harmonizing the derivation of nutrient reference values 321
Origin of the harmonization effort 321
Current efforts to realize the goal of harmonization 321
Outlook 324
Rationale for moving forward 324
Next steps 325
Disclaimer 325
References 325

Key points

- Although methods for deriving intake values have progressed, they have not been implemented consistently across countries or regions; and critical data and resources needed to set intake values are not universally available, particularly to low- and middle-income countries.
- A harmonized approach to setting core intakes will enhance the availability of resources that are critical to the process and provide a consistent scientific basis for setting NRVs globally.
- A risk assessment framework is used to identify and assess indicators of nutrient adequacy or toxicity across age- and sex groups in the population.
- Nutrient intake requirements are based on a range of intake that is high enough to maintain nutritional adequacy but low enough to avoid risk of adverse health effects.
- The characteristics of chronic disease development are a primary challenge to identifying a causal relationship between exposure, and an adverse health outcome.
- Recent advances in the DRI framework provide an underpinning for harmonizing the methodological approach to deriving NRVs across population groups.
- Commitment of stakeholders to a harmonized process for deriving NRVs is essential in order to generate NRVs that are aligned to support nutrition policy, programs, and education globally.
- A harmonized approach to setting core intake values will enhance the availability of resources that are critical to the process, particularly for low- and middle-income countries, and provide a consistent science-based approach for setting NRVs globally.

Glossary

Lowest-Observed-Adverse-Effect-Level The lowest level of exposure to a dietary component at which there is an adverse health effect

No-Observed-Adverse-Effect-Level A level of exposure to a dietary component at which there is no biologically or statistically significant increase in the frequency or severity of any adverse health effects

Nutrient Reference Value Nutrient-based dietary intake standards used by health professionals in planning and assessing diets for individuals and groups

Risk assessment framework The process used to determine nutrient intake amounts for both adequate and excessive intakes. There are four steps in the process: (1) hazard identification; (2) dose-response assessment; (3) intake assessment; and (4) risk characterization

Uncertainty factor A numeric factor that is applied to avoid underestimating risk due to uncertainties by introducing a margin of safety

Introduction and historical background

In the late 1930s and early 1940s the Canadian Council of Nutrition and the US Committee on Foods and Nutrition, subsequently renamed the Food and Nutrition Board, initiated the process of developing recommendations for single nutrient intake values as a benchmark for meeting nutrient requirements to prevent deficiency diseases in the general population (McHenry, 1941; Roberts, 1940). The efforts of these two groups resulted in publication of the Canadian Dietary Standards (McHenry, 1941) and the Recommended Dietary Allowances (RDAs), respectively (NRC, 1941). Over the next several decades, the recommended nutrient intake values were revised and refined, and new values added as previously unknown nutrients were discovered. Beginning in the 1990s single nutrient values were revised to include a range of uses and applications, as well as address evolving health concerns, such as chronic disease, across population groups (IOM, 1994).

Specifically, this meant considering nutrient functioning beyond the prevention of deficiency states to include a broader role in maintaining health, and reducing risk of adverse health outcomes. Since the first nutrient reference values (NRVs) were developed, the process for deriving them has evolved to include more rigorous, transparent, and reproducible methodologies. However, while the methods for deriving intake values have progressed, they have not been implemented consistently across countries or regions; and critical data and resources needed to set intake values are not universally available, particularly to low- and middle-income countries, resulting in inconsistencies in nutrition policies, as well as across guidelines and nutrient intake recommendations across the world. A harmonized approach to setting core intake values will enhance the availability of resources that are critical to the process, particularly for low- and middle-income countries; and provide a consistent scientific basis for setting NRVs globally.

Conceptual basis for nutrient reference values

Overview

Nutrient-based dietary intake standards, or NRVs, serve as a set of dietary intake values that are used by health professionals in planning and assessing the diets of individuals and groups. Nutrient intake standards and guidelines are a primary resource for a number of public and private institutions. Health professionals in hospitals, prisons, schools, and educational institutions use them to assist in planning healthy diets for both groups and individuals. The military uses NRVs to plan meals on- and off the field, and in procurement of food and military rations. Nutrition researchers use NRVs as a frame of reference to study association of intakes with disease states, and the food industry uses them to develop healthier products. Federal agencies at the national, state, and local levels use NRVs in the formulation of food and nutrition policy including establishing international regulatory and trade requirements, product labeling, food fortification standards, nutrition monitoring, and dietary guidance (IOM, 2003).

In the US, in the process of setting the original single-nutrient intake recommendations, the RDAs, initial consideration was given to meeting physiologic needs to prevent deficiency disease. The resulting recommendations defined intake levels for individual nutrients on the basis of essentiality. Specifically, the recommended intake levels for a given nutrient would cover requirements for a given age/sex group and also provide a margin of safety over the minimum requirements needed to prevent deficiency. Adequacy of intake was determined on the basis of amounts needed to support physiologic functioning and to support energy, maintain tissue integrity, and in the case of pregnant women, infants, and children, support growth and development (IOM, 1994; Harper, 1987). Over time, a more robust model for determining intake requirements emerged that was based on a risk assessment and intake distribution approach (IOM, 2008; Russel et al., 2009).

Risk assessment framework

The risk assessment framework, shown in **Fig. 1**, indicates the steps in the process used to determine nutrient intake amounts for both adequate and excessive intakes (IOM, 2011). The first step, hazard identification, is used to identify and assess indicators of nutrient adequacy or harmfulness across age- and sex groups in the population. This step includes a review and synthesis of the evidence related to relevant health outcomes. The strength and quality of the evidence base is critical. In the synthesis of the first Dietary Reference Intakes (DRIs), NRVs developed jointly by the US and Canada, the evidence to support the derivation of intake values was based on narrative evidence reviews. Subsequent updates, however used systematic reviews as the evidence base (IOM, 2011; NASEM, 2019).

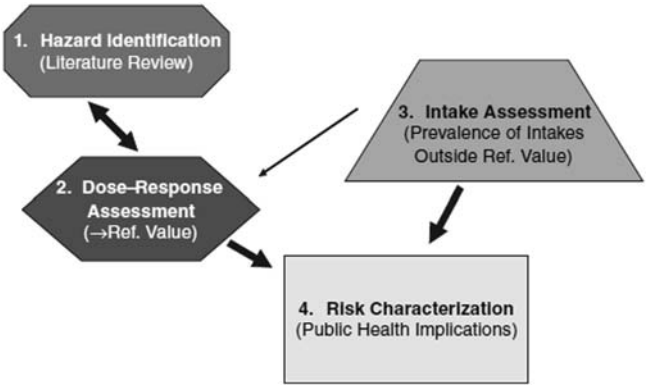


Fig. 1 Risk assessment framework for deriving core nutrient reference values. Source: IOM (2008). Reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.

Fig. 2 shows a hierarchy of evidence, with the highest quality of evidence at the top of the pyramid. The type of evidence needed to identify and assess nutrient adequacy and harmfulness, or risk of adverse health effects, for a nutrient review, ideally, is the randomized controlled double-blind and randomized controlled trial and high-quality observational studies such a prospective cohort studies (IOM, 2008). Other observational studies, including other cohort and case-control studies can be used, although the quality of this type of evidence generally does not support development of a recommended intake, e.g. RDA. Rather, in the absence of adequate high-quality evidence, such evidence may be used to derive an estimated intake value, e.g. Adequate Intake (AI).

Systematic evidence reviews were introduced into the DRI process in the first update for calcium and vitamin D. This approach provided a more robust evidence base that allowed the calcium and vitamin D review panel to set an intake values for vitamin D for all age groups except infants. The derivation of the RDA in this case was based on new evidence for bone health (IOM, 2011). A subsequent update to the DRIs for sodium and potassium incorporated a more rigorous, transparent, and reproducible systematic review process with the application of the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system to the evidence assessment stage. In this system, five domains: risk of bias, imprecision, inconsistency, indirectness, and publication bias, are used to rate the certainty of a given body of evidence in order to assess the level of confidence about the causality of an association between nutrient intake level and a health outcome (NASEM, 2017).

The second step in the risk assessment framework is dose–response assessment. At this step dietary intake levels needed to support derivation of core reference values, the Estimated Average Requirement (EAR) and Tolerable Upper Intake Level (UL), are determined. In some cases, however, dose-response assessment does not reflect actual nutrient levels. Calcium and iron, for example, cannot be metabolized, thus quantification of losses via blood, urine, feces, etc. is used to estimate the physiological requirement in order to derive an EAR (IOM, 2001, 1997).

When dose-response data is not available for example, in the determining the UL, a No-Observed-Adverse-Effect-Level (NOAEL) or a Lowest-Observed-Adverse-Effect-Level (LOAEL) serves as the basis for the derivation. An uncertainty factor and adjustment factor is applied to the NAOEL/LOAEL to identify the threshold for an adverse health effect. Uncertainty assessments are not precise however, and therefore an element in toxicity assessment that must be considered very carefully, a calculated uncertainty factor, could lower either the threshold for toxicity used to derive the UL or the intake level estimated to avoid a deficiency (Renwick, 2006). Uncertainty factors are usually set at 1 log unit or 0.5 log units from the NOAEL/LOAEL (IOM, 1998).

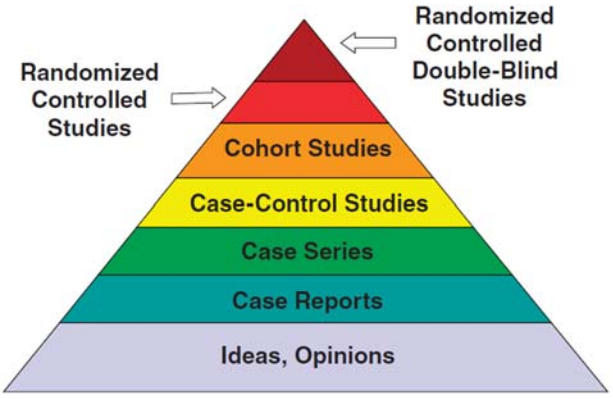


Fig. 2 Hierarchy of evidence in systematic reviews to support the derivation of nutrient reference values. Source: IOM (2008). Reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.

The third step in the framework is intake assessment. In this step, population-based intake data and/or biomarkers of nutrient status are used to assess intake adequacy or excessive exposure levels for a nutrient. National surveys or other large population databases are generally used to obtain population-based intake data. Assuring that intake levels are sufficient requires knowledge of the dietary components that can affect nutrient bioavailability for a given population group. This is particularly important for age/sex groups with specific nutrient requirements, such as during pregnancy, infancy, and childhood. In some cases nutrient bioavailability is affected by the composition of the food source. Phytate in grains, for example, can impair mineral absorption (Gibson et al., 2010; Kim et al., 2009). Thus accurate data sources, i.e. food composition databases and population intake surveys are essential to estimating intake adequacy or excess.

The last step in the framework considers the public health consequences of not meeting an EAR/RDA or AI, and exceeding a UL. This includes determining how characteristics unique to a population age/sex group, such as body size, lifestyle, environment, or other factors, could influence nutrient requirements for that group. The risk assessment framework is advantageous in that it offers a systematic approach to the decision steps that take place over the process of deriving NRVs. Further, it supports documentation of the strength and sufficiency of the evidence and enhances transparency; and it allows for the incorporation of new and emerging scientific tools into the process (Yetley et al., 2017; Joint FAO/WHO Technical Workshop, 2005). The risk assessment framework laid the groundwork for future consideration of chronic disease endpoints in the process of deriving NRVs.

Nutrient intake distribution model

Nutrient intake requirements are grounded in a model based on a distribution of intake levels that are high enough to maintain nutritional adequacy but low enough to avoid risk of adverse health effects. The DRIs are based on a distribution curve of observed levels of intake, represented graphically in **Fig. 3** (IOM, 2000). In this model, the EAR, an estimate of the intake requirement that is adequate for 50% of a population age/sex group, is derived from evidence gathered and assessed in the risk assessment process described above. On the distribution curve, the RDA is set at 2 standard deviations above the EAR. This is an amount presumed to provide sufficient intake to cover at least 97.8% of the intake requirements for a given age/sex group in the population. When sufficient data to set an EAR is not available, an AI can be set instead. Estimated safe and adequate intake levels are bounded on one end of the intake curve by the RDA and on the other by the UL. This range of safe and adequate intake in the model predicts a low probability of either inadequate or excessive intake for a given nutrient (IOM, 2007).

Nutrient intake and consideration of risk of chronic disease

Inferring diet-disease relationships in chronic disease

NRVs are designed to cover a diverse population, deemed "apparently healthy". In setting NRVs for a population, individuals who have a disease that requires therapeutic management, are malnourished or undernourished, have diseases resulting in malabsorption or requiring dialysis, or have severely altered energy needs, e.g. due to disability or altered mobility, are not included in the healthy population. However, within the overarching population, there is a wide range of health conditions, including overweight and obesity, which includes many individuals who have or are at risk for one or more chronic diseases. While risk of diet-related chronic disease has been recognized as a public health concern for many decades, a framework for considering chronic disease endpoints in deriving NRVs was not in place across developed countries, including in North America and Europe.

Prior to 2010 the DRIs lacked a specified reference value for nutrients associated with chronic disease endpoints. In the derivation of the first DRIs between 1997 and 2005, six nutrients were considered for a quantifiable relationship with chronic disease endpoints. These were: calcium and vitamin D and risk of osteoporosis and bone fracture; fluoride and dental caries; fiber and coronary heart disease; and sodium and potassium and hypertension (IOM, 1997, 2003, 2002/2005, 2005). The DRI study panels that

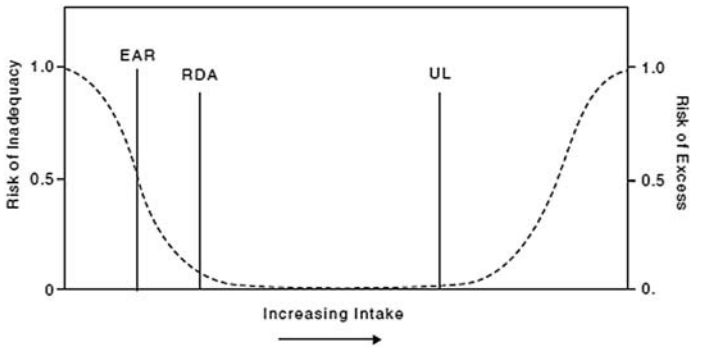


Fig. 3 Distribution of intake levels with corresponding NRVs. Low risk of intake adequacy and adverse effects is between the RDA and the UL. NOTE: EAR = Estimated Average Requirement; RDA = Recommended Dietary Allowance; UL = Tolerable Upper Intake Level. Source: IOM (2000). Reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.

reviewed these nutrients were challenged by the paucity of and limitations in the evidence, including evidence to support a dose-response relationship relative to chronic disease health outcomes. Thus they were unable to find sufficient evidence to support a conclusion that the nutrients under consideration acted as modifiers of risk for chronic disease.

In the absence of sufficient evidence to derive an EAR and set an RDA in the first DRI reviews for calcium, vitamin D, fluoride, fiber, and potassium, AIs were set at levels considered adequate to avoid deficiency. Sodium deficiency however is not a health outcome of concern among US and Canadian populations; thus an AI was set based on presumed adequacy of intake and allowing for excess sodium loss in sweat by non-acclimatized individuals who are exposed to high temperatures or are moderately physically active (IOM, 2002/2005). Although the UL was originally intended as a threshold for a safe intake level to reduce risk of toxicity from excessive intake of a nutrient, in the case of sodium for example, a UL was set at a threshold intake level consistent with evidence available at the time to support primary prevention of hypertension, stroke, and coronary heart disease. In the dose-response assessment step of the risk assessment framework, blood pressure was used as a surrogate endpoint, based on both dose-response and prospective observational evidence that strongly correlated with dietary intake (IOM, 2005). In the case of fiber, the available evidence led the DRI panel to conclude that health outcomes associated with fiber intake were more strongly correlated with the amount of food consumed than age or body weight. Thus an AI was set based on grams of fiber intake per 1000 calories (IOM, 2002/2005).

Challenges to consideration of chronic disease endpoints

The characteristics of chronic disease development are a primary challenge to identifying a causal relationship between exposure, as either a deficiency or excess of a nutrient or other bioactive food component and an adverse health outcome. Foremost among these challenges is the time frame in development of a chronic disease, which can take decades. Obtaining sufficient temporal evidence needed to establish a causal relationship, i.e. a randomized controlled trial or prospective cohort study, is generally infeasible in most instances. Further, the consistency of associations among different study types can be problematic. This has been seen in a number of studies on diet-disease relationships in which the outcome of a randomized controlled trial and an observational study on the same nutrient show different outcomes (Yaktine and Ross, 2019).

Another significant challenge is demonstrating a link between a single potential causal factor and disease risk. Diet is one of many factors that can influence risk of chronic disease and although there are numerous association studies on diet and disease risk, evidence of sufficient rigor and quality to demonstrate causality as defined in the risk assessment framework is lacking. Dose-response relationships are difficult to demonstrate in relating intake level to disease risk. Not only is time a factor in dose-response, a direct action of a nutrient or bioactive food component is difficult to establish. Most nutrients act in a concerted fashion rather than in isolation, as might be expected in, for example, a drug trial.

Updating the DRI framework for chronic disease endpoints

In the US and Canada, two reports set the stage for expanding the DRI framework to include consideration of chronic disease endpoints in DRI development. First, Yetley et al. (2017) published a set of options for considering chronic disease endpoints in the derivation of DRIs, including the scientific and policy context for carrying out the work. The report considered: (1) the range of evidentiary challenges in selecting and using chronic disease endpoints in the DRI process; (2) intake-response models potentially relevant to chronic disease endpoints; and (3) justification for including chronic disease endpoints in the DRI process (Yetley et al., 2017).

Subsequently, the National Academies of Sciences, Engineering, and Medicine (NASEM) convened a study panel to assess the options put forward by Yetley et al. and establish a set of guiding principles for including chronic disease endpoints in future DRI reviews (NASEM, 2018a). Conceptually, the NASEM (2018a) report distinguished among the differences in the type and strength of evidence designed for determining adequacy, risk of adverse health outcomes, or chronic disease outcomes. The report recommendations were therefore based on characterizing the strength of the evidence for a chronic disease outcome, e.g. through an appropriate biomarker; and establishing a level of confidence in assessing the evidence for setting an intake level sufficient to reduce risk of the chronic disease. The report recommendations addressed extrapolating intake-response data on chronic disease risk across population groups; structuring intake recommendations as ranges as opposed to single values; and applying the existing DRI framework to assessment of chronic disease endpoints (NASEM, 2018a).

A DRI study panel was subsequently convened to undertake the task of applying the recommendations from both reports to assessing chronic disease endpoints related to sodium and potassium intake. The outcome was a new DRI value, the chronic disease risk reduction value (CDRR), designed as a range of intake that is beneficial to reducing the risk of a given chronic disease outcome, and based on moderate to high strength evidence for both causality and intake-response (NASEM, 2019). Significantly, the new CDRR differentiated risk of chronic disease from risk of adverse health effects from excess intake as defined by the UL, thus providing an alternate approach for considering health outcomes associated with high sodium intake. These recent advances in the DRI framework provided an underpinning for harmonizing the methodological approach to deriving NRVs across population groups.

Harmonizing the derivation of nutrient reference values

Origin of the harmonization effort

In response to an acknowledged need to address variability in the derivation of NRVs, a group of nutrition scientists met in Florence, Italy in 2005. The goal was to develop a harmonized framework for deriving NRVs. An important outcome of the 2005 meeting was a rationale for a conceptual framework to harmonize the methodological approach to deriving NRVs. The rationale stated four core reasons for harmonization. These are to:

1. "Improve the objectivity and transparency of values derived by diverse national, regional, and international groups;
2. Provide a common methodological basis for experts to consider in deriving NRVs;
3. Improve access to scientific and other resources used in the derivation of NRVs, particularly among low- and middle-income countries; and
4. Provide a common basis for uses and applications of NRVs across countries and regions globally" (King and Garza, 2007).

A significant accomplishment of the Florence meeting was the development of an organizing framework based on distributions of nutrient intakes required to achieve a given health outcome in the target population. The framework is built around two core NRVs, the average nutrient requirement (ANR) and the UL. Similar to the DRI framework, the ANR is used to derive an individual nutrient level (INL), which is the recommended nutrient intake level for the population. The UL, the highest level of normal intake that does not incur risk of adverse health effects, is derived from toxicity data on the nutrient under consideration (King and Garza, 2007).

These NRVs and their methodological derivation correlate with the DRI framework and the derivation of the EAR, RDA, and UL. While the description of the core NRVs varies across high-income countries and regions, the methodological approach to their derivation is similar and thus they are key concepts to harmonization. Also central to harmonization is developing a common terminology. **Table 1** shows examples of different terms currently used to describe reference values across countries and regions. The set of published monographs that came from the discussions in Florence captured these and other significant concepts and laid the groundwork for the effort to develop and implement a globally harmonized NRV process (King et al., 2007).

In 2009, the Department of Nutrition for Health and Development of the World Health Organization (WHO) established a new process for developing and updating nutrition guidelines in agreement with the WHO Handbook for Guideline Development (Global Programme on Evidence for Health Policy, 2003). In response, a WHO Nutrition Guidance Expert Group was formed (WHO, 2010). The charge to this group was to strengthen WHO's part in communicating science-based advice, evidence-informed policy, and program guidance in support of the WHO Nutrition Program. At the same time, the Global Network of Institutions for Scientific Advice on Nutrition was formed to gather organizations from around the world who were responsible for providing scientific advice used to develop nutrition recommendations and guidelines (WHO, 2010). The group met to exchange information about nutrition guidance and to investigate opportunities to collaborate as a step toward harmonizing nutrition recommendations and guidelines. The meeting gave rise to recognition of the need for synergizing efforts aimed at developing nutrition guidance, including the harmonization of approaches for developing NRVs and guidance in support of national and international policy development.

Current efforts to realize the goal of harmonization

In response to these efforts, NASEM undertook a series of steps to move the effort to harmonize the derivation of NRVs forward into the next phase of development and implementation. The first step was an international workshop that explored questions about the frameworks used to develop NRVs, the status of NRVs globally, and experiences and knowledge about harmonizing approaches to

Table 1 Comparison of exemplar terms for nutrient reference values currently in use.

	Country, regional, or global entity			
NRV designation	*United States and Canada*	*UK*	*European Union/EFSA*	*WHO/FAO*
Overarching term	DRI	DRV	DRV	
Average requirement	EAR	EAR	AR	
Recommended intake level	RDA	RNI	PRI	RNI
Lower reference intake		LRNI	LTI	
Adequate intake	AI		AI	
Safe upper level of intake	UL	SUL	UL	UL
Appropriate macronutrient distribution range	AMDR	AMDR	RI	Population mean intake goals

NOTE: NRV = Nutrient Reference Value; UK = United Kingdom; EFSA = European Food Safety Authority; WHO = World Health Organization; FAO = Food and Agriculture Organization of the United Nations; DRI = Dietary Reference Intake; DRV = dietary reference value; EAR = estimated average requirement; AR = average requirement; RDA = recommended dietary allowance; RNI = reference nutrient intake; PRI = population reference intake; LRNI = lower reference nutrient intake; LTI = lower threshold intake; UL = tolerable upper intake level; SUL = safe upper intake level; RI = reference intake range for macronutrients.

Source: NASEM (2018b). Reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.

derive intake standards (NASEM, 2018a). The workshop was planned and implemented in partnership with the WHO and the Food and Agriculture Organization of the United Nations (FAO), and with support from the Bill & Melinda Gates Foundation. The workshop took place at the FAO headquarters in Rome, Italy in 2017. Workshop discussions ranged from exploration of frameworks to support harmonization on a global scale; approaches for assessing evidence; contextual factors relevant to population groups across regions and countries; sharing of resources and expertise to facilitate the adoption of a harmonized approach; and possible barriers and challenges to achieving harmonization.

Three overarching strategic messages emerged from the workshop discussions. First, is the need for standardization of the methodology for deriving core NRVs. Specific actions needed include creating a centralized collaborative consultative group that would serve as a global resource for countries revising existing or developing new NRVs. Also needed are nutrition-specific tools, including consistent NRV terminology, or consistent definitions for NRV terms. Second is recognition of the specific needs of regions or countries with regard to food composition data, collection of data from dietary surveys, variables affecting nutrient bioavailability, and variation in health status among population groups needed to harmonize methodologies, particularly across low- and middle-income countries. Lastly, is the need to be mindful of research gaps and support developing the science from which intake requirements are derived (NASEM, 2018a).

The second action taken by NASEM toward harmonizing the derivation of NRVs was a consensus study to assess the methodological approach to deriving NRVs for specific population sub-groups, women of reproductive age and children from birth to 2 years of age (NASEM, 2018b). The study panel was charged to develop a framework to demonstrate the application of a harmonized approach and recommend ways to achieve a harmonized basis for deriving the core NRVs, the average requirement and the upper intake level, for these population sub-groups. The study panel's assessment of the NRV derivation process led to the initial finding that a harmonized approach to deriving NRVs would assure that the majority of a target population could achieve nutrient intake levels sufficient to avoid both deficiency and adverse health outcomes related to excessive intake, including toxicity and reduced risk of chronic disease.

The study panel's conclusions aligned closely with the strategic messages from the Rome workshop. Foremost, the panel concluded that the need for nutritional benchmarks is critical all over the world. This shared need, combined with the considerable effort and cost of deriving NRVs, is a sound justification in support of international cooperation (NASEM, 2018b). Additionally, in agreement with the workshop discussion, the need for a global expert consultative group would be the most appropriate model for promoting a harmonized NRV process. Finally, the study panel determined that in order to achieve a globally harmonized approach, the process for deriving NRVs must focus on two core values, the EAR and the UL (NASEM, 2018b).

The consensus study panel identified essential steps needed in deriving the core values that should be consistent across countries and regions (**Fig. 4**) These steps are based on the concept that selecting a methodological approach for a nutrient review is linked to identifying the role of the nutrient in meeting physiological needs, and understanding intake patterns, bioavailability of the nutrient, and the presence of infection and other local factors that influence the requirement for the population under consideration. Further, when updating an existing NRV or deriving a new value, an in-depth understanding of the uncertainties that affect the nutrient review process is vital to the credibility of the process, including the accuracy of the recommended intakes, and in enabling decision makers to use NRVs in nutrition policy (NASEM, 2018b). The recommended framework for deriving NRVs, shown in **Fig. 5**, provides a platform for deriving NRVS globally and across population groups. There are four steps in the process. These are: "(1) choose the appropriate tools, (2) collect relevant data from the tools; (3) identify the best approach or method of derivation; and (4) derive the core reference values, the AR and UL" (NASEM, 2018b).

In the first step, selecting tools and data resources, the framework identifies three key tools and resources. These are systematic reviews, databases, and regional or local factors. Systematic reviews provide a wide-ranging, objective assessment of available evidence. As noted in NASEM (2018b), systematic reviews must be of sufficient quality and rigor to support assessment of intake requirements for the age/gender groups under consideration as well as establish findings and conclusions about the nutrient under review. This includes providing information about the search criteria, inclusion and exclusion criteria, study quality and data summaries, and consistency with the proposed methodology. Additionally, the heterogeneity inherent in any systematic review should be assessed for uncertainty (NASEM, 2018b).

The report concluded that the application of meta-analysis criteria will reduce uncertainty and bias due to variations across study designs, differences in populations, or the source of data. Additionally, the GRADE tool provides a qualitative approach to evaluating factors that may affect study quality and risk of bias. The GRADE approach stipulates that the certainty of evidence should be set at the lowest detectable level for all critical outcomes (Guyatt et al., 2011). This means that in general, randomized controlled trials receive a "high" rating for study quality and certainty, but may be rated lower; and observational studies general begin with a "low" certainty of evidence, but may be rated higher (see **Fig. 2**).

Food intake and food composition databases provide contextual information and other data, and are used to estimate factors that affect nutrient bioavailability (Gibson, 2007). Food intake data is also used to calculate prevalence of inadequate or excess intake.

In the second step, data collected from the tools and resources enables identification and selection of biomarkers, health outcomes, and dietary factors that affect nutrient requirements for various age/sex groups. Data that would be useful, particularly for low- and middle-income countries includes biomarkers of nutrient status, such as measures of a nutrient or its metabolite in biological fluids or tissues, anthropometric measures, or data from medical histories or physical exams.

In the third step, determining the best approach for deriving NRVs, dose response modeling is used when the relationship between nutrient intake and functional outcome is known. In the case of minerals, in which there is no known biomarker for

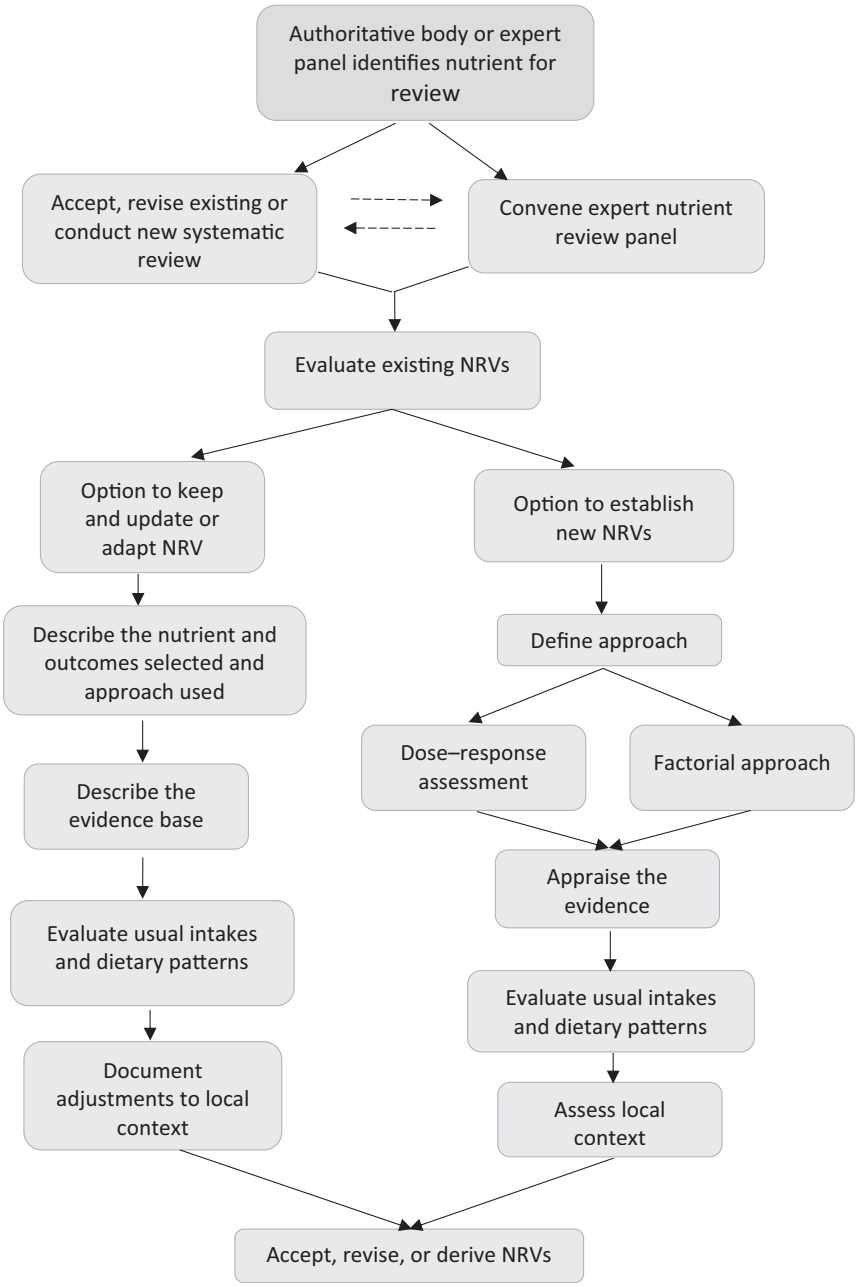


Fig. 4 Flow diagram for deriving nutrient reference values. Source: NASEM (2018b). Reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.

adequacy, the factorial method is used. In this method, the amount of absorbed nutrient to replace basal losses from excretion, sweat, or exfoliation, or to account for menstrual loss or accretion due to growth is estimated. In the case of iron and zinc, for example algorithms to predict bioavailability are available to predict absorption, transport, and metabolism. Assessment of measures of physical activity are also included when evaluating energy intake.

In the last step, the AR is derived from dose-response or factorial modeling and the UL is determined from a risk assessment analysis. The nutrient intake distribution model described above and shown in **Fig. 3** illustrates the application of the risk assessment process to achieve a low risk of inadequacy or harmfulness from excess intake. Under this process, the AR will likely differ somewhat across countries or regions, depending on the criterion used in deriving the estimate.

The AR is not based on nutrient intake, rather it is determined from methodologies as described in the risk assessment framework to estimate physiologic need for a given age/sex group. Differences in physiological determinants, limitations in the available data, differences in bioavailability or absorption factors, and varying requirements for different age/sex groups are all influences that can affect the derivation of an AR (NASEM, 2020). The AR then becomes the basis for setting the recommended intake level. Thus, the

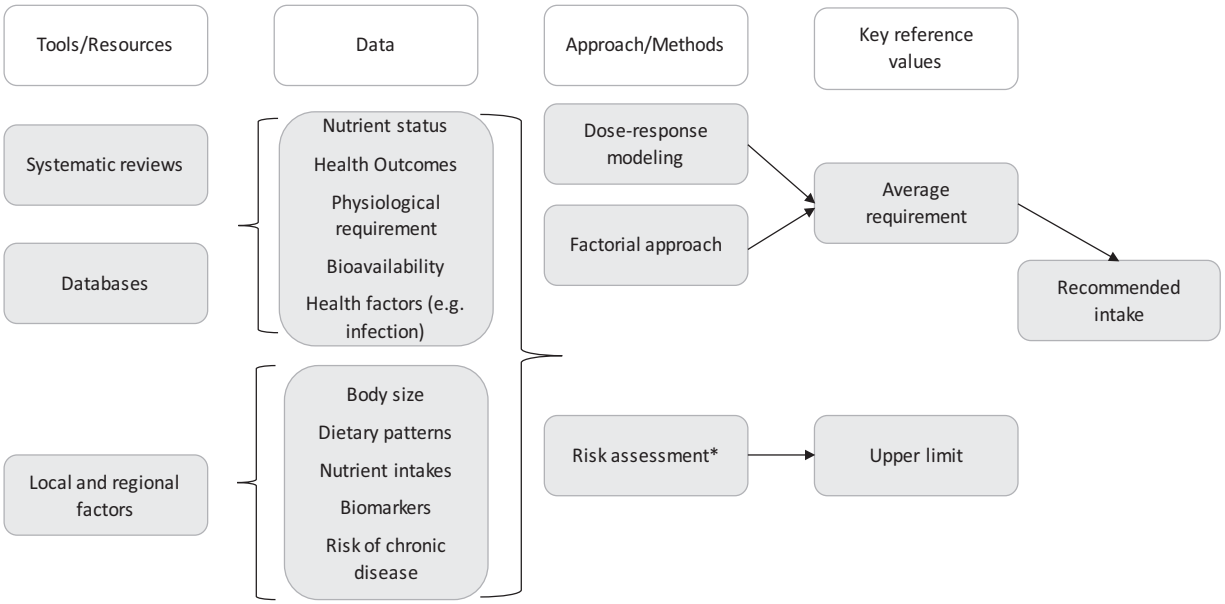


Fig. 5 Framework for harmonizing the process for deriving nutrient reference values. Source: NASEM (2018b). Reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.

AR should: (1) be established from the amount of nutrient to meet the needs of 50% of the population; (2) be established, whenever possible, for all nutrients and food substances with public health relevance; (3) include appropriate designations for macronutrients, particularly when chronic disease endpoints are considered; (4) consider relevant nutrient-nutrient interactions; and (5) consider relevant subpopulations within the general population (NASEM, 2018b).

The UL is not a recommended intake and its derivation in the future will be influenced by inclusion of chronic disease endpoints in the analysis. The "Guiding Principles" report (NASEM, 2017) recommended that forthcoming nutrient reviews "characterize the health status of the population in terms of who is included and excluded for each DRI value" (NASEM, 2017, p. 30). Hence, as new data becomes available, the UL will no longer be used as an intake value for chronic disease risk.

The process for deriving harmonized intake standards, described by the four major steps in the harmonization framework is based on six core principles. They are that:

1. NRVs are regularly updated;
2. The process is clear and transparent;
3. The methods are rigorous and relevant;
4. Factors influencing the NRV are documented;
5. The strength of the evidence is determined and
6. The review is complete and efficient.

The third action taken by NASEM was development of a tool kit (NASEM, 2020). The tool kit was created to support implementation of the recommendations from the NASEM consensus study (NASEM, 2018b). The goal of the tool kit is to share core principles, practices, and recommendations with key stakeholders, and to engage them in moving forward with practical steps that will facilitate harmonization globally. Further, the tool kit will serve as a first step in providing both guidance and resources for low- and middle-income countries to develop a harmonized approach to deriving NRVs for those populations. The tool kit resources are intended to facilitate the NRV process for low- and middle-income countries. Commitment from stakeholders to a harmonized process for deriving NRVs is essential in order to generate core intake values that are globally aligned and supportive of nutrition policies, programs, and education.

Outlook

Rationale for moving forward

Collectively, the three NASEM activities demonstrated that a harmonized approach to setting NRVs is essential for:

1. Resolving dissimilarities across countries in establishing nutrition standards;
2. Promoting consistent population health objectives;
3. Providing a means for creating food and nutrition policies that are aligned across countries and regions; and
4. Enhancing transparency in developing trade and regulatory policies with economic, health, and safety implications (NASEM, 2020).

Among low- and middle-income countries in particular, NRVs are key to formulating food and nutrition policy; developing food assistance programs, nutrition education, and monitoring population health. A number of actions can be taken in support of harmonizing the approaches to deriving NRVs. A future global dialog to gather support for harmonization as well as to identify a pathway for implementing the actions identified in the three NASEM reports will move the process forward. A collective effort among these groups is needed to describe the harmonization approach, as well as advocate for its implementation.

Next steps

The need for harmonization of the NRV process is clear, as is the need for all users to have access to adequate resources for deriving the core (AR and UL) values. Implementing a harmonized approach to deriving NRVs will require discourse at the local, regional, and country levels in order to gain support, stimulate engagement, and foster investment from the groups and organizations to whom the effort will be entrusted. The costs of revising or establishing new NRVs will be high, and thus providing options for building on existing work and databases will reduce both time and cost. National, regional, and international leadership will move the harmonization process forward.

Disclaimer

The author's work was conducted independently of her role as director of the Food and Nutrition Board. Any views not attributed to reports of the Health and Medicine Division are those of the author and do not necessarily represent the views of The National Academies of Sciences, Engineering, and Medicine.

References

Gibson, R.S., 2007. The role of diet- and host-related factors in nutrient bioavailability and thus in nutrient-based dietary requirement estimates. Food Nutr. Bull. 28 (1), S77–S100.

Gibson, R.S., Bailey, K.B., Gibbs, M., Ferguson, E.L., 2010. A review of phytate, iron, zinc, and calcium concentrations in plant-based complementary foods used in low-income countries and implications for bioavailability. Food Nutr. Bull. 31 (2 Suppl), S134–S146. https://doi.org/10.1177/15648265100312S206. PMID: 20715598.

Global Programme on Evidence for Health Policy, 2003. Guidelines for WHO Guidelines. EIP/GPE/EQC/2003.1. World Health Organization, Geneva.

Guyatt, G.H., Oxman, A.D., Kunz, R., Atkins, D., Brozek, J., Vist, G., Alderson, P., Glasziu, P., Falck-Ytter, Y., Schunemann, H.J., 2011. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J. Clin. Epidemiol. 64 (4), 395–400.

Harper, A.D., 1987. Evolution of recommended dietary allowances—new directions? Annu. Rev. Nutr. 7, 509–537.

Institute of Medicine (IOM), 1994. How Should the RDAs Be Revised? National Academy Press, Washington, DC.

Institute of Medicine (IOM), 1998. Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B-6, Folate, Vitamin B-12, Pantothenic Acid, Biotin, and Choline. National Academies Press, Washington, DC.

IOM, 1997. Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoried. The National Academies Press, Washington DC. https://doi.org/10.17226/5776.

IOM, 2000. Dietary Reference Intakes: Applications in Dietary Assessment. National Academy Press, Washington, DC.

IOM, 2001. Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc. The National Academies Press, Washington, DC. https://doi.org/10.17226/10026.

IOM, 2002/2005. Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids. National Academy Press, Washington, DC.

IOM, 2003. Dietary Reference Intakes: Applications in Dietary Planning. The National Academies Press, Washington, DC. https://doi.org/10.17226/10609.

IOM, 2005. Dietary Reference Intakes for Water, Potassium, Sodium, Chloride, and Sulfate. National Academies Press, Washington, DC.

IOM, 2007. Dietary Reference Intakes Research Synthesis: Workshop Summary. The National Academies Press, Washington, DC. https://doi.10.17226/11767.

IOM, 2008. The Development of DRIs 1994–2004: Lessons Learned and New Challenges: Workshop Summary. The National Academies Press, Washington, DC. https://doi.org/10.17226/12086.

IOM, 2011. Dietary Reference Intakes for Calcium and Vitamin D. The National Academies Press, Washington, DC. https://doi.org/10.17226/13050.

Joint FAO/WHO Technical Workshop on Food Nutrient Risk Assessment, 2005. A Model for Establishing Upper Levels of Intake for Nutrients and Related Substances: Report of a Joint FAO/WHO Technical Workshop on Food Nutrient Risk Assessment, WHO Headquarters. Geneva Switzerland. World Health Organization, Food and Agriculture Organization of the United Nations & International Programme on Chemical Safety. 2006. World Health Organization, Geneva, Switzerland. https://apps.who.int/iris/handle/10665/4351.

King, J.C., Garza, C., 2007. Harmonization of nutrient intake values. Food Nutr. Bull. 28 (1). The United Nations University.

King, J.C., Vorster, H.H., Tome, D.G., 2007. Nutrient intake values (NIVs): a recommended terminology and framework for the derivation of values. Food Nutr. Bull. 28 (1). The United Nations University.

Kim, J., Woodhouse, L.R., King, J.C., Welch, R.M., Li, S.J., Paik, H.Y., Joung, H., 2009. Relationships between faecal phytate and mineral excretion depend on dietary phytate and age. Br. J. Nutr. 102 (6), 835–841. https://doi.org/10.1017/S007114509289057. PMID: 19267950.

McHenry, E.W., 1941. Determination of nutritional status. Can. Publ. Health J. 32 (5), 231–235. Canadian Public Health Association Available at. https://www.jstor.org/stable/4197825. (Accessed 27 July 2021).

National Academies of Sciences, Engineering, and Medicine (NASEM), 2017. Guiding Principles for Developing Dietary Reference Intakes Based on Chronic Disease. The National Academies Press, Washington, DC. https://doi.org/10.17226/24828.

NASEM, 2018a. Global Harmonization of Methodological Approaches to Nutrient Intake Recommendations: Proceedings of a Workshop. The National Academies Press, Washington, DC. https://doi.org/10.17226/25148.

NASEM, 2018b. Harmonization of Approaches to Nutrient Reference Values: Applications to Young Children and Women of Reproductive Age. The National Academies Press, Washington, DC. https://doi.org/10.17226/25343.

NASEM, 2019. Dietary Reference Intakes for Sodium and Potassium. The National Academies Press, Washington, DC. https://doi.org/10.17226/25353.

NASEM, 2020. Harmonizing the Process for Establishing Nutrient Reference Values: A Tool Kit. The National Academies Press, Washington, DC. https://doi.org/10.17226/25981.

National Research Council, 1941. Recommended Dietary Allowances. Washington, DC.

Renwick, A.G., 2006. Toxicology of micronutrients: adverse effects and uncertainty. J. Nutr. 136, 493S–501S.

Roberts, L.J., 1940. Letter to Russell M. Wilder, Dated December 31. Item 1 in Compiled Reports Prepared by the Board and Its Committees, 1940 Through December 1944. Vanderbilt University Medical Center Library, Nutrition Archives, Special Collections.

Russell, R., Chung, M., Balk, E., Atkinson, S.A., Giovannucci, E.L., Ip, S., Lichtenstein, A.H., Mayne, A., Raman, G., Ross, A.C., et al., 2009. Opportunities and challenges in conducting systematic reviews to support the development of nutrient reference values: vitamin A as an example. Am. J. Clin. Nutr. 89, 728–733.

World Health Organization, 2010. WHO Global Network of Institutions for Scientific Advice on Nutrition: Report of the First Meeting, 11–12 March 2010, WHO. World Health Organization, Geneva, Switzerland. https://apps.who.int/iris/handle/10665/44460.

Yaktine, A.L., Ross, A.C., 2019. Milestones in DRI development. What does the future hold? Adv. Nutr. 10, 537–545.

Yetley, E.A., MacFarlane, A.J., Greene-Finestone, L.S., Garza, C., Ard, J.D., Atkinson, S.A., Bier, D.M., Carriquiry, A.L., Harlan, W.R., Hattis, D., King, J.C., Krewski, D., O'Connor, D.L., Prentice, R.L., Rodricks, J.V., Wells, G.A., 2017. Options for basing Dietary Reference Intakes (DRIs) on chronic disease endpoints: report from a joint US-/Canadian-sponsored working group. Am. J. Clin. Nutr. 105 (1), 249S–285S.

Health inequities

R. Perez-Escamilla, Yale School of Public Health, New Haven, CT, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of R. Perez-Escamilla, Health Disparities, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 417–423, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00142-2.

Introduction 327
Nutrition and health inequities epidemiology 328
Malnutrition inequities 328
Household food insecurity inequities 331
Obesity inequities 331
Health inequities 331
Health care access inequities 331
Life expectancy inequities 332
Understanding the determinants of health and nutrition inequities 332
Socio-ecological model of health 332
Solutions 332
Conditional cash transfer programs 333
Access to nutritious foods 333
Unhealthy foods taxes and regulation 333
Physical activity 334
Health care reform 334
Reduce discrimination 334
Conclusion 334
References 334
Further reading 335

Key points

The objectives of this article are to:

- Describe the epidemiology of nutrition and health inequities
- Describe Food insecurity inequities
- Present a Socio-ecological framework based on the social determinants of health to address inequities
- Describe specific programs that can facilitate access to healthy foods and physical activity environments among the most vulnerable
- Explain why addressing health inequities is crucial for countries to meet the United Nations Sustainable Development Goals

Glossary

Conditional cash transfer programs Government programs that provide a small monthly stipend to low-income families as long as they meet certain conditions

Food security Access by all people at all times to enough food for an active, healthy life

Health disparities Differences in health outcomes and risk factors among population subgroups

Health inequities Differences in health which are not only unnecessary and avoidable but, in addition, are considered unfair and unjust

Social determinants of health The conditions in which people are born, grow, live, work, and age, including the health system

Introduction

Health disparities refer to differences in health outcomes and risk factors among population subgroups (Carter-Pokras and Baquet, 2002; LaVeist, 2005). When deemed avoidable and unjust, these differences represent health inequities (The Marmot Review,

2021). The root cause of health inequities is social injustice (Marmot, 2005; Gravlee, 2009; Marmot et al., 2008; Williams et al., 2019). Health inequities are more prevalent among vulnerable sub-groups at social disadvantage. Health inequities follow a social gradient and can be a function of ethnicity/race, immigration status, gender, sexual preferences, religion, or disability status.

The social determinants of health (SDH) are key for understanding the distribution of health inequities within and across countries. These refer to the physical and social conditions of the environments where individuals are born, grow, go to school, work, socialize, and age (Institute of Medicine, 2002; Marmot et al., 2008; Marmot, 2005). Thus, efforts toward addressing health inequities must go above and beyond disease control and antipoverty measures, and must improve the psycho-social environments where socially disadvantaged groups live and work. Fully addressing health inequities requires making the basic resources needed to improve quality of life accessible to the most vulnerable. For example, improving the availability of nutritious foods, clean water and health care in a community does not guarantee that the most vulnerable within that community will have access to them, as the control of access to these resources has been shown to be socially determined (Fanzo and Davies, 2021; King et al., 2021; Young et al., 2021).

Nutrition and health inequities epidemiology

Health and nutrition inequities have been well documented not only in developing, but also in developed countries (Perez-Escamilla et al., 2018). Even in the USA, which is considered to be the wealthiest nation on earth, low-income groups (overrepresented by ethnic/racial minorities) are substantially more likely to be food insecure, overweight or obese, to have type 2 diabetes, and to lack health insurance (Pérez-Escamilla, 2010; Reid, 2009) (**Fig. 1**).

Malnutrition inequities

Decades of research have conclusively shown that under-nutrition, even in its relatively mild form, has negative consequences for human physical and intellectual development. Indeed, malnutrition explains over half of deaths among children under 5 y of age living in developing countries. Because human development is the basis for social capital, which, in turn, is the engine that drives national development, the problem of undernutrition is now recognized by key decision-making bodies and stakeholders as a top priority. The Food and Agricultural Organization (FAO) estimates that close to 800 million individuals worldwide experience energy undernutrition. The problem is more severe in the poorest regions of the world including sub-Saharan Africa, and South and Southeast Asia. Epidemiological data clearly show that, within countries, socio-economic position is a powerful determinant of risk of under-nutrition, including risk of stunting in children. In addition, hundreds of millions of individuals may have access to sufficient or even excessive amounts of energy, but do not have access to diets of adequate nutritional quality. As a result, low socio-economic groups are now facing a "double burden" of malnutrition (i.e., excessive rates of both under- and overnutrition) oftentimes coexisting within the same household and accompanied by micronutrient deficiencies. The problem of malnutrition has its roots in the SDH and is largely explained by major inequities in the distribution of wealth, power, and resources. In addition, two of

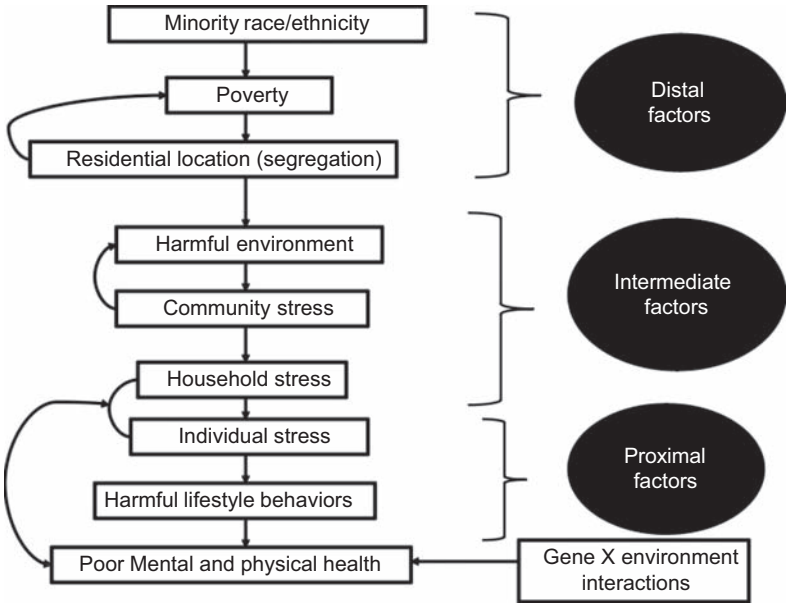


Fig. 1 Socio-ecological model of health and disease. Original illustration prepared by author.

Table 1 Key concepts and definitions.

Health inequities	Differences in health which are not only unnecessary and avoidable but, in addition, are considered unfair and unjust.
Social gradient in health	Health outcomes, including life expectancy, improve as a function of the position of individuals along the socio-economic continuum.
Social determinants of health	The conditions in which people are born, grow, live, work, and age, including the health system. These circumstances are shaped by the distribution of money, power, and resources at global, national, and local levels, which are themselves influenced by policy choices. The social determinants of health are mostly responsible for health inequities.
Socio-ecological model of health	Individual's health and nutrition outcomes are the result of multilevel distal, meso, and proximal factors and the interactions among them. This model recognizes that individual lifestyle choices are strongly shaped by the environments where people grow and live (i.e., the social determinants of health).
Social capital	Quality of social relationships, sense of trust, belonging, and reciprocity in wider society.
Developmental origins of health and disease hypothesis	Early life conditions, including nutritional and hormonal milieu to which individuals are exposed *in utero* and early childhood, impact longer-term health outcomes including obesity and chronic disease risk.
Food security	Access by all people at all times to enough food for an active, healthy life, and includes, at a minimum: (1) the ready availability of nutritionally adequate and safe foods and (2) an assured availability to acquire acceptable foods in socially acceptable ways (e.g., without resorting to emergency food supplies, scavenging, stealing, or other coping strategies).
Conditional cash transfer programs	Government programs that provide a small monthly stipend to low-income families as long as they meet certain conditions. These have included keeping children in school, bringing them to get immunized and receive other primary health care services, and in some instances participation of their caregivers.

Original table prepared by author.

the key proximal determinants of malnutrition are lack of access to nutritious diets and inadequate health care. Thus, nutrition inequities go hand in hand with inequities related to access to food, health care, and other basic human needs (**Tables 1–3**).

Table 2 Sustainable development goals principles.

1. The 2030 Agenda for Sustainable Development, adopted by all United Nations Member States in 2015, provides a shared blueprint for peace and prosperity for people and the planet, now and into the future.
2. The 17 Sustainable Development Goals (SDGs) included in the agenda, are framed in the context of an urgent call for action by all countries—higher and lower income—through global partnerships.
3. The SDGs recognize that ending poverty and other deprivations must go hand-in-hand with strategies that improve health and education, reduce inequality, and spur economic growth—all while tackling climate change and working to preserve our oceans and forests.

Original table prepared by author.

Table 3 Addressing key distal, meso, and proximal determinants of health: policy goals.

Issues	*Policy goals*
The social gradient	More equity in income and wealth distribution through better access to adequate education, employment opportunities, housing, health care, and food security.
Stress	Improve neighborhood safety, built and social environment. Universal access to good-quality schools and health care (physical, mental, and dental).
Early life	Universal access to prenatal care, breastfeeding promotion and support, high-quality child care and education infant, toddler and preschool programs, that include adequate nutrition in the context of nurturing care. Home visiting programs to support development of better infant feeding, parenting skills, and psycho-emotional health of the caregiver.
Social exclusion	Strong and enforceable civil rights laws that prevent discrimination based on individual's ethnicity/race, religion, gender, sexual orientation, age, disability status, and immigration status including being an internally displaced person or refugee, among others.
Work	Develop good skills for job market through good-quality vocational, college, and postgraduate educational opportunities. Establish minimum wages based on reasonable estimates of what families need to have a quality of life compatible with the principles of human rights and dignity. Provide reasonable paid parental leave policies that protect the job security of parents and their ability to provide nurturing care to their infants. Improve safety of work environments by reducing occupational hazards. Global trade policies should be restructured to allow the less powerful to have a chance to fairly compete in local and global markets. Increase credit access to small producers and businesses.
Unemployment	Safety net with reasonable unemployment benefits and retraining opportunities.
Social support	Increase social capital in deprived neighborhoods/communities.
Addiction	Increase social capital in deprived neighborhoods/communities. Provide access to effective primary prevention and substance-abuse treatment programs. Limit advertisement of alcohol and tobacco products in low-income neighborhoods. Use effective taxation (pricing) policies to limit abuse of substances.
Food	Restructure food price policies so that healthy nutrient-dense foods become more economically accessible to low-income groups. Restructure food systems in low-income areas to improve community access to healthy and nutritious foods, including fresh fruits and vegetables, legumes/pulses, minimally processed grains, and healthy and sustainable animal protein sources. Limit advertisement of foods high in saturated/transfats, sodium, and refined sugars in low-income neighborhoods and through mass media marketing targeting children. Ensure that food assistance programs provide services and benefits that are consistent with evidence-based dietary guidelines across the life cycle. Provide effective instrumental nutrition and food safety education, including healthy cooking, in schools and diverse community settings.

Table 3 Addressing key distal, meso, and proximal determinants of health: policy goals.—cont'd

Issues	*Policy goals*
Physical activity and transport	Increase community access to safe green areas. Improve use of public open spaces for families to walk, exercise, and interact socially with other community members. Redesign urban areas traffic systems to facilitate the routine use of bicycles when going to school or work. Provide access to safe, clean, and efficient public transportation for individuals to be able to transport to work, school, and public open spaces.
Health care	Health care access for all. Must emphasize primary prevention, screenings, and referrals and cover physical, mental, and dental health care needs. Screenings should be conducted to identify SDH needs including household food and water insecurity, housing, and transportation.

Original table prepared by author.

Household food insecurity inequities

Household food security is defined as "access by all people at all times to enough food for an active healthy life", and includes, at a minimum: (1) the ready availability of nutritionally adequate and safe foods and (2) an assured availability to acquire acceptable foods in socially acceptable ways (e.g., without emergency food supplies, scavenging, stealing, or other coping strategies). Thus, food insecurity (FI) exists when there is limited or uncertain availability of nutritionally adequate and safe foods or limited or uncertain ability to acquire acceptable foods in socially acceptable ways. The US Census Bureau has reported annual household FI since 1995, estimated from the 18-item Household Food Security Supplement Module (HFSSM) applied through the Current Population Survey Food Security Supplement.

The reference time period of the HFSSM is the 12 months preceding the survey and households are classified as having either food security, low food security, or very low food security, based on the number of affirmative answers to the HFSSM items. In 2021, 10.1% of US households were food insecure (i.e., with either low or very low food security) and 3.8% were very food insecure corresponding to 5.1 million households. Households are more likely to be food insecure if they are poor, headed by a single female, if there are children living in the household or if the head of household is Hispanic or African-American. Thus, the distribution of household FI is also a reflection of the strong socio-economic inequities in the USA. Similar conclusions have been reached in very diverse countries located across world regions.

Obesity inequities

Obesity is a major public health problem in both developed and developing countries. Indeed, maternal–child obesity rates in Mexico are now as high as in the USA. Consistent evidence indicates that, within developed countries and increasingly among developing countries as well, poverty is associated with higher obesity risk. This is not surprising, as lack of access to healthy, nutrient-dense foods and opportunities for physical activity are important obesity risk factors. In the USA, 76.9% of Hispanic and 73.7% of non-Hispanic Black women, vs. 67.5% of Non-Hispanic White women, are overweight or obese (body mass index (BMI) $\geq$ 25). The corresponding rates for morbid obesity (i.e., BMI $\geq$ 35) are 18.9%, 27.9%, and 16.6%, respectively. The prevalence of type 2 diabetes in the USA is about twice as high among Hispanic and Black than non-Hispanic White individuals, and hypertension is substantially more prevalent among Blacks. Because obesity is a major risk factor for chronic diseases, such as cardiovascular disease and type 2 diabetes, it is not surprising that the distribution of these chronic conditions is inversely associated with socio-economic status (Millen et al., 2016).

Ethnic/racial differences in obesity in the USA are evident since very early childhood (Pérez-Escamilla et al., 2017; Taveras et al., 2010). The obesity prevalence among 2–5-year-old non-Hispanic White children was 10.7%, compared with 14.9% among non-Hispanic Black and 16.7% among Mexican-American preschoolers. As expected, these differences reflect greater prevalence of early life obesity-related risk factors, including maternal depression, rapid weight gain during infancy, lower breast-feeding, introduction of complementary feeding before recommended age, more restrictive maternal feeding style, higher intake of sugar-sweetened beverages and fast food, and insufficient sleep during infancy. These findings are of concern, as there is increasing evidence that intrauterine and early childhood environments are determinants of health outcomes later in life, including the development of obesity and chronic diseases. This is known as the "developmental origin of health and disease" hypothesis (Gillman, 2005).

Health inequities

Health care access inequities

Tens of millions of individuals living in the USA do not have health insurance and, thus, lack access to basic health care and oftentimes end up receiving needed medical attention in emergency room services. Millions more low-income individuals with limited

insurance coverage also do not have access to adequate physical, mental, and dental health care. In the USA the risk of not having access to health care is higher among low-income families, even though the majority of these households have employed members. This risk increases considerably among non-citizens, many of whom are migrant farm workers, not fluent in English and living in isolated communities. Thus, social exclusion and lack of adequate health care are largely responsible for the highly visible health care access inequities in the USA. This is emphasized by the improved health care access and health outcomes that populations living in other developed countries have. These countries' systems are characterized by a "not-for-profit" health insurance structure, coupled with a single payer health care system focused on primary and secondary prevention.

Life expectancy inequities

Inequities in life expectancy are substantial in both lower and higher income countries (Danaei et al., 2010; World Health Organization, 2003). According to the World Health Organization (WHO) the probability of a man dying between the ages of 15 and 60 y ranges from 8.3% in Sweden to 90.2% in Lesotho. Life expectancy is almost 50 years longer in Japan than in Sierra Leone. These remarkable differences have also been documented within countries. In Australia, aboriginal people have a life expectancy that is 20 y less than the national average. In the USA, the difference between counties with the highest vs. lowest life expectancy is 18.4 y for men and 14.3 y for women. In the USA, Black men live 6.3 y less and Black women 4.5 y less than their White counterparts. In England, people living in the poorest neighborhoods die on average 7 y earlier, and their disability-free life expectancy is 13 y lower, compared with those living in the wealthiest neighborhoods. Further, clear social gradients have been documented within both developed and developing countries, showing that social and economic position is a powerful determinant of health outcomes. Although genetic susceptibilities play a role in the development of disease within populations and subgroups, the large between-group differences in health outcomes are largely explained by the SDH and, thus, need to be addressed accordingly. When equitable development approaches are used, key maternal, child, and adult health indicators rapidly improve, thus ruling out a "genetic" explanation. For example, Brazil recently demonstrated how large-scale investments in social programs capable of reducing income and wealth inequities across socio-economic groups led to improvements in public health indicators, especially among the most vulnerable (Barros et al., 2010; Pérez-Escamilla et al., 2018).

Understanding the determinants of health and nutrition inequities

Socio-ecological model of health

Health inequities can best be explained by the Social-ecological model of health. This model posits that multilevel factors ranging from the individual to the macrosocial level interact in determining health outcomes. This model, consistent with the SDH construct, recognizes that the social distribution of wealth, resources, and power strongly determine access to healthy lifestyle, including health screenings and timely treatment. Communities with low access to quality education and employment opportunities are characterized by unhealthy environments. These have low social capital and poor infrastructure, including limited availability of healthy foods (e.g., fresh fruits and vegetables), limited opportunities for leisure-time physical activities (e.g., safe green areas and sports facilities), and limited access to quality primary, secondary, and tertiary health care. They are also characterized by high levels of crime, violence, and other psychosocial stressors. This stressful environment, coupled with inadequate coping responses among low-income individuals, represents another major barrier to a healthy lifestyle.

Research conducted in the US has shown that self-perceived discrimination related to skin color has a negative impact on health, independent of socio-economic status (Williams et al., 2019). Self-perceived discrimination has been associated with high blood pressure, arterial plaque, and high inflammation markers and visceral fat. These are important risk factors in the development of cardiovascular disease. Because cardiovascular disease explains a substantial amount of health inequities between Blacks and Whites in the USA, racial discrimination needs to be understood and addressed from the health inequities perspective.

Health inequities may be substantially explained by the direct biological impact (e.g., resulting from poor nutrition, low physical activity, substance abuse) as well as higher levels of chronic stress experienced by disadvantaged communities. In other words, individual level health outcomes are the result of lifestyle choices (diet, physical activity, substance abuse, sex practices) that are strongly shaped by the social policies and physical and social environments where people grow, live, and work (i.e., the SDH). Thus, correcting health inequities requires addressing the macro (e.g., safety net policies, minimum wage, social exclusion), meso (e.g., neighborhood safety, access to healthy foods, good quality schools, and adequate health care), and proximal lifestyle determinants (e.g., dietary intake, physical activity, substance abuse, sex behaviors) of health, and the interaction among them.

Solutions

International organizations and governments worldwide now fully acknowledge the existence and pressing need for addressing the major health and nutrition inequities worldwide. The United Nations Sustainable Development Goals (SDGs) call for significant reductions in extreme poverty and hunger, as well as in major communicable diseases. The SDGs place strong emphasis on addressing socio-economic, gender inequities and other SDH including universal access to education and health care. The SDGs also emphasize the profound links between human and planetary health in the context of environmental sustainability (Pérez-Escamilla, 2017).

Through the SDGs there is now global consensus that SDH determine the risk factors and health outcomes for chronic diseases. This is due to the evidence indicating that lower socio-economic position is linked with undernutrition, obesity, less access to

healthy foods, and fewer opportunities for leisure-time physical activity, proper screenings, medical referrals, and good-quality health care.

Although the existence of health and nutrition inequities is increasingly acknowledged, there is still no consensus on the best ways to approach them. The SDGs specifically call for emphasizing sustainable economic growth and national development approaches based on strong multisectorial partnerships. Thus far, most efforts have focused on communicable disease prevention, food assistance, conditional cash transfers provision of microcredits, and universal health insurance strategies.

Few developed countries have based their national development approaches on the SDH including access to education, adequate work environments, social and economic security, and health care access. Scandinavian countries have taken this approach and, as a result, have some of the best health indicators in the world. The UK recently released a major report on health inequalities in England addressing social determinants findings. A demonstration of the power of the social determinants approach is illustrated by strong reduction of stunting inequities in Brazil between 2003 and 2014 when its government reformed social, health and economic policies to become much more equitable and inclusive in the context of sustainable economic growth.

Conditional cash transfer programs

Conditional cash transfer programs (CCTPs) represent a pragmatic way of addressing the major food, education, and health care needs of the poor (Lagarde et al., 2007; Segura-Pérez et al., 2016). CCTPs provide a "small" cash allowance to low-income households with the condition that they keep children in school and bring them to primary health care centers. In some programs, caregivers are required to attend health and nutrition workshops. CCTPs originated in Latin America and have shown positive effects on maternal–child health behaviors and outcomes. In countries as diverse as Mexico, Brazil, Colombia, Honduras, and Nicaragua, CCTPs have been shown to lift families out of extreme poverty (Segura-Pérez et al., 2016). However, from the social determinants perspective, they represent only a partial solution, as they do not address the underlying social gradient. In other words, CCTPs do not benefit poor and lower middle-class families that are above the program financial inclusion criteria, and are not able to lift families from extreme poverty into the middle class. As other regions of the world, including Sub-Saharan Africa, consider incorporating CCTPs as a strategy for reducing extreme poverty and improving education, health, and nutrition outcomes, it is essential that health and education infrastructures are in place before the program is implemented so that participants can truly meet the program's "conditions."

Access to nutritious foods

A key indicator of dietary quality is the consumption of fresh fruits and vegetables. For this reason, food assistance programs heavily emphasize the need for improving access to these healthy and nutritious foods in disadvantaged communities. The US Supplementary Nutrition Assistance Program (SNAP), formerly known as the food stamp program, provides a cash transfer to low-income families earmarked for purchasing food but does not place conditions as to the nutritional value of the foods purchased. One approach that SNAP has used for improving participant intake of fruits and vegetables is to provide nutrition education through the SNAP-Ed program. Because education alone is not enough to meet this goal, SNAP is piloting fiscal incentives to foster availability, access, and purchase of fruits and vegetables by program participants. An important incentive program being piloted is providing recipients with a "bonus" or "discount" incentive when using their SNAP benefits to purchase fruits and vegetables. At the same time, SNAP is supportive of the development of more points of access to fruits and vegetables in low-income areas, including corner stores and farmers markets, and by facilitating the use of the program electronic benefit cards at these points of purchase. This example illustrates how food systems, as a whole, need to be revamped for vulnerable communities to have access to healthy and nutritious foods, and for their residents to be able to purchase them. Culturally appropriate health and nutrition education is crucial for ensuring that the products are actually consumed in recommended amounts. Otherwise, the impact of these major investments will not impact obesity and health inequities in the US.

In developing countries micronutrient fortification has become a major strategy for improving the nutritional status of populations. These strategies include provision of micronutrient powders (i.e., "sprinkles") or of energy- and micronutrient-dense spreads to improve the nutritional value of foods given to young children. These approaches have been shown to be effective to a limited extent relative to the major health and nutrition deficits experienced by low-income individuals as a result of pervasive social inequities. Although some of these short- and medium-term biomedical-type solutions are needed to address under-nutrition in developing countries, the problem will not be solved unless policies based on the SDH and nutrition inequities are implemented. These policies are expected to reduce economic and social inequities and, as a result, improve access to better education, employment, and housing opportunities. This, in turn, would facilitate the adoption and dissemination of lifestyles compatible with good health and quality of life.

Structural changes in global food and water systems governance are needed to development sustainable food systems that promote the health of individuals, communities, and the planet (Pérez-Escamilla, 2017; Young et al., 2021).

Unhealthy foods taxes and regulation

Driven in large part by initiatives launched in Latin America, nations across the globe continue to launch policies and legislation designed to tax ultraprocessed foods and beverages, front of package nutrition warnings, and protecting children and against marketing of these foods. Although initial findings look promising, it is unclear if and how these measures are addressing health and nutrition inequities (Pérez-Escamilla et al., 2021).

Physical activity

Urban areas in Latin America have launched innovative "public space use" and "public transportation" initiatives to increase physical activity. These include the development and maintenance of safe green areas for people engaged in individual and group leisure-time activities, provision of low bicycle rental fees to move around downtown areas, and the closing of major city streets to vehicular traffic during weekends (Pérez-Escamilla et al., 2021). Although few, if any, of these structural changes in the physical environment have been formally evaluated, their emphasis on facilitating "access for all" represents a step in the right direction to bringing more equity to opportunities for leisure-time physical activity. These opportunities in public open spaces improve levels of social capital and air quality, and directly benefit the health of individuals through increased levels of physical activity.

Health care reform

Many nations have reformed their health care systems to provide "universal" access to timely health care. This is the case of the former "Seguro Popular" effort in Mexico, the "single health system" in Brazil, and, more recently, the effort by the US government to have all individuals covered by health insurance based on a sliding scale system with subsidies for those who cannot afford it. In the US, ongoing health care reform efforts also include forbidding health insurance companies from denying coverage to individuals with preexisting medical conditions. This reform however excludes from its coverage millions of individuals living in the country who are non-citizens, highlighting the need to simultaneously address immigration reform. This illustrates how ignoring relevant SDH, in this instance the exclusion of individuals based on their immigration status, precludes a country from truly correcting major health care access in-equities. In fact, it may make this inequity even worst for the most vulnerable.

Reduce discrimination

Because of the impact that stress associated with self-perceived racial discrimination has on the risk of poor mental health and chronic diseases (Williams et al., 2019) it is important that individuals who are vulnerable to this type of discrimination are protected through civil rights laws and awareness of the general public. Racial discrimination is not always intentional and when this is the case it can be addressed through improvements in cultural sensitivity and diversity and inclusiveness education and trainings. Structural racism which in many ways is the legacy of colonialism and neocolonialism will require major food, health care and education systems reforms.

Conclusion

Breaking the cycle between poverty, disease, and malnutrition requires a new way of thinking to move forward within a health and nutrition equity framework. This approach will indeed require substantially changing the social and political structures responsible for the inequitable distribution of power, wealth, and resources within and across countries. This is unlikely to happen without including the active participation of the most vulnerable communities in the shaping of this "new" development paradigm. The WHO commission on SDH concluded that social inequities are of such a magnitude that they are responsible for excessive morbidity and the premature deaths of people on a massive scale. Correcting these inequities will take time. The commission however concluded that health outcomes can continue to improve and persistent health inequity gaps can be substantially narrowed within a generation, as long as social injustice is addressed through policies and programs based on the well-documented SDH. These efforts should consider establishing universal guaranteed basic income, livable minimum wages, and should heavily invest in opening access to quality education, employment opportunities, and improved built environments. At a macro level it is essential to overhaul and restructure inequitable global trade policies and the major political, social and economic power imbalances that characterize the world in which we live. To be successful at reducing health and nutrition inequities, these combined efforts must result in a more just distribution of wealth, power, and resources between and within nations.

References

Barros, F.C., Matijasevich, A., Requejo, J.H., et al., 2010. Recent trends in maternal, newborn, and child health in Brazil: progress toward Millennium Development Goals 4 and 5. Am. J. Publ. Health 100 (10), 1877–1889.

Carter-Pokras, O., Baquet, C., 2002. What is a "health disparity?". Publ. Health Rep. 117 (5), 426–434.

Danaei, G., Rimm, E.B., Oza, S., Kulkarni, S.C., Murray, C.J., Ezzati, M., 2010. The promise of prevention: the effects of four preventable risk factors on national life expectancy and life expectancy disparities by race and county in the United States. PLoS Med. 7 (3), e1000248.

Fanzo, J., Davies, C., 2021. Global Food Systems, Diets, and Nutrition: Linking Science, Economics, and Policy. Palgrave Macmillan, Switzerland.

Gillman, M.W., 2005. Developmental origins of health and disease. N. Engl. J. Med. 353 (17), 1848–1850.

Gravlee, C.C., 2009. How race becomes biology: embodiment of social inequality. Am. J. Phys. Anthropol. 139 (1), 47–57.

Institute of Medicine, 2002. Unequal Treatment: Confronting Racial and Ethnic Disparities in Healthcare. National Academies Press, Washington, DC.

King, A.C., Perez-Escamilla, R., Vorkoper, S., Anand, N., Rivera, J., 2021. Childhood obesity prevention across borders: the promise of U.S.-Latin American research collaboration. Obes. Rev. (Suppl. 3), e13238.

Lagarde, M., Haines, A., Palmer, N., 2007. Conditional cash transfers for improving uptake of health interventions in low- and middle-income countries: a systematic review. J. Am. Med. Assoc. 298 (16), 1900–1910.

LaVeist, T.A., 2005. Minority Populations and Health: An Introduction to Health Disparities in the United States. Jossey-Bass, San Francisco, CA.

Marmot, M., 2005. Social determinants of health inequalities. Lancet 365 (9464), 1099–1104. PubMed PMID: 15781105.

Marmot, M., Friel, S., Bell, R., Houweling, T.A., Taylor, S., 2008. Commission on Social Determinants of Health. Closing the gap in a generation: health equity through action on the social determinants of health. Lancet 372 (9650), 1661–1669. PubMed PMID: 18994664.
Millen, B.E., Abrams, S., Adams-Campbell, L., Anderson, C.A., Brenna, J.T., Campbell, W.W., Clinton, S., Hu, F., Nelson, M., Neuhouser, M.L., Perez-Escamilla, R., Siega-Riz, A.M., Story, M., Lichtenstein, A.H., 2016. The 2015 dietary guidelines Advisory Committee Scientific Report: development and major conclusions. Adv. Nutr. 7 (3), 438–444.
Pérez-Escamilla, R., 2010. Health care access among Latinos: Implications for social and health care reforms. J. Hisp. High. Educ. 9, 43–60.
Pérez-Escamilla, R., 2017. Food security and the 2015–2030 sustainable development goals: from human to planetary health. Curr. Dev. Nutr. 1 (7), e000513.
Pérez-Escamilla, R., Segura-Pérez, S., Lott, M., 2017. Feeding guidelines for infants and young toddlers: a responsive parenting approach. Nutr. Today 52 (5), 223–231.
Perez-Escamilla, R., Bermudez, O., Buccini, G.S., Kumanyika, S., Lutter, C.K., Monsivais, P., Victora, C., 2018. Nutrition disparities and the global burden of malnutrition. BMJ 361, k2252. https://doi.org/10.1136/bmj.k2252.
Pérez-Escamilla, R., Vilar-Compte, M., Rhodes, E., Sarmiento, O.L., Corvalan, C., Sturke, R., Vorkoper, S., 2021. Implementation of childhood obesity prevention and control policies in the United States and Latin America: lessons for cross-border research and practice. Obes. Rev. (Suppl. 3), e13247.
Reid, T.R., 2009. The Healing of America: A Global Quest for Better, Cheaper, and Fairer Health Care. The Penguin Press, New York.
Segura-Pérez, S., Grajeda, R., Pérez-Escamilla, R., 2016. Conditional cash transfer programs and the health and nutrition of Latin American children. Rev. Panam. Salud Publica 40 (2), 124–137.
Taveras, E.M., Gillman, M.W., Kleinman, K., Rich-Edwards, J.W., Rifas-Shiman, S.L., 2010. Racial/ethnic differences in early-life risk factors for childhood obesity. Pediatrics 125 (4), 686–695.
The Marmot Review. Fair Society, Healthy Lives. Available at http://https://www.instituteofhealthequity.org/resources-reports/fair-society-healthy-lives-the-marmot-review (Accessed June 22, 2021).
Williams, D.R., Lawrence, J.A., Davis, B.A., 2019. Racism and health: evidence and needed research. Annu. Rev. Publ. Health 40, 105–125.
World Health Organization, 2003. Social Determinants of Health: The Solid Facts. https://www.euro.who.int/en/publications/abstracts/social-determinants-of-health.-the-solid-facts. (Accessed 22 June 2021).
Young, S.L., Frongillo, E.A., Jamaluddine, Z., Melgar-Quiñonez, H., Pérez-Escamilla, R., Ringler, C., Rosinger, A.Y., 2021. Perspective: the importance of water security for ensuring food security, good nutrition, and well-being. Adv. Nutr. 12 (6) nmab003.

Further reading

FAO. The state of food insecurity in the world. http://www.fao.org/publications/sofi/en/.
Sustainable Development Goals (SDGs). https://sdgs.un.org/goals (Accessed June 22, 2021).
USA Department of Agriculture (USDA)/Department of Health and Human Services (DHHHS). Dietary Guidelines for Americans. https://www.dietaryguidelines.gov/sites/default/files/2020-12/Dietary_Guidelines_for_Americans_2020-2025.pdf (Accessed 8 October, 2022).
World Health Organization. Social Determinants of Health. http://www.who.int/social_determinants/en/.

Religious customs, influence on diet

K. Albala, University of the Pacific, Stockton, CA, USA

© 2013 Elsevier Ltd. All rights reserved.

This is a reproduction of K. Albala, Religious Customs, Influence on Diet, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 153–157, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00245-2.

Introduction 336
Judaism 336
Christianity 337
Islam 338
Hinduism 339
Buddhism 339
Conclusion 340
Further Reading 340

Glossary

Ahimsa The practice of nonviolence in all aspects of life, including taking the life of living creatures for food.
Brahmin The highest caste among Hindus, with the greatest number of food restrictions.
Eucharist The celebration of the Mass in Roman Catholicism, in which wine is said to miraculously transform into the Blood of Jesus and bread, normally a consecrated wafer, changes into His flesh.
Halal The food laws of Islam, which forbid pork and alcohol. It's opposite is haram.
Kosher The complex food laws of Judaism, which originate in the book of Leviticus, which are still practiced by devout Jews today, with some modifications.

Introduction

Religion is among the most pervasive forces influencing the human diet, even among those who do not actively practice a particular faith. Religion shapes individual and group preferences in the form of taboos, celebratory foods, and ritual offerings, it prescribes the modes of commensality and attitudes toward indulgence and abstinence, and throughout history religious leaders have carefully defined the meaning of alimentary sustenance in ways that continue to resonate into the present, having a direct impact on people's dietary choices. This article will recount attitudes toward food in the major world religions, and their origin and historical development as necessary for a full appreciation of the complexity of how religion shapes food cultures today.

The origin of the deep connections between food and faith lie without doubt in the earliest polytheistic religions. Since prehistoric times, the Gods were invoked to ensure good harvests and thanked for their bounty; they also had to be appeased with sacrifice. Ancestral spirits were provided with offerings to maintain their good will and a cycle of festivals marked the agricultural and religious calendar. Many religious festivals retain features of these original agricultural rites. With the discovery of alcohol and other intoxicants, communities could enter into an ecstatic union with the Godhead, a numinous state of oneness, bridging the earthly and otherworldly planes. To this day all these primal elements of the relationship between food and religiosity survive as rudiments, altered in form, but still discernible among modern practices.

Judaism

Perhaps more than any other civilization, the ancient Hebrews defined their relationship to God in terms of what they ate, what was considered clean and unclean and what was sacrificed to God. A succession of dietary codes in various historical epochs also explains how that relationship changed. In the biblical narrative, following creation, in the first epoch humans were intended to be not only vegetarian in the state of innocence, but fruitarian, meaning that Adam and Eve's Edenic diet was obtained without killing any living thing. They ate only fruits and seeds and leaves, which could be taken without killing the plant. Plants also produced fruit spontaneously without human effort. This initial ban on murder will help explain the internal logic of the complex rules of kashrut and sacrifice, which were anything but arbitrary or capricious. Following the fall, as punishment for their disobedience in eating from the Tree of Knowledge the couple were forced to subsist through labor. In this second epoch, Adam had to earn his bread by the sweat of his brow, i.e., by planting crops, and for Eve it meant pain in childbirth as well as subjection to man.

Although it is not clear exactly what was eaten in this second epoch, the professions of Cain and Abel, a shepherd and farmer, closely match the early economy of the Fertile Crescent after the Neolithic Revolution. It is not until after the flood (the third epoch)

that Noah is given explicit permission to kill for food. "Every creature that lives and moves shall be food for you; I give you them all, as once I gave you all green plants." It is an admission on God's part that humans are faulty. Only the blood, which contains the life, must be poured on the ground, as it belongs to God. This blood prohibition remains in effect as central to Judaism, and kosher meat must be salted to drain it of all traces of blood. It was also at this time that the first sacrifice was made. It is clear that the burning fat and entrails provide a soothing odor to the Lord, that He is in a sense sustained by the smoke. More importantly, the sacrifice, explained only in later books of the bible, reinstates justice in the universe. To compensate for the act of killing, some creature must be punished, normally an unblemished "scapegoat."

It is only with the giving of the law under Moses, the fourth and final historical epoch, that the Hebrews were given a complex dietary code, which is still for the most part in effect among observant Jews. At its root is still the prohibition against murder. Unclean animals are those, which kill in order to eat. Rather than catalog all clean animals, the Levitical Priests devised a shorthand way to recognize the innocent herbivores: those which chew their cud and have a cloven hoof. Thus the omnivorous pig was banned along with other creatures, which did not seem to fit into this scheme, such as camels and hares. It has been suggested that the real origin of this ban was an informal knowledge of trichinosis, or perhaps the fact that pigs are inefficient in the desert, being unable to sweat and thus requiring watering holes, and more importantly competing for resources with humans because pigs are unable to convert grass (indigestible to humans) into food, the way ruminants can. The Hebrews may have banned pigs in order to maintain distinction from their pig-eating neighbors and keep their identity intact, for to eat with someone is one step away from intermarriage. Yet none of these explanations is as convincing as the original ban on murder, which translated into considering all predatory animals unclean.

More difficult is explaining the ban on shellfish and other creatures, which seems to defy the categorical schema of the priests. According to their logic, a fish must have scales and swim, birds must have feathers and fly, animals must have legs and walk. Scaleless fish, flightless birds, etc., defy the rational categorization and are thus also unclean. Another stricture demands that milk and meat must never be mixed, which by tradition stems from not seething a calf in its own mother's milk. This means that a cheeseburger is not kosher. Even among nonobservant Jews, through cultural conditioning foods categorized as trayf may be unappetizing or even repugnant.

In addition there are numerous holidays, which revolve around food, or lack thereof. Most important is the fast, which takes place from sundown to sundown on Yom Kippur, a time to reflect on and atone for one's sins in the preceding year. Passover is also essentially a food holiday, though no leavened bread (chometz) may be eaten in commemoration of the Jews' escaping Egypt who had no time to let their bread rise. The seder plate also recalls other aspects of the captivity in Egypt: salt water for tears and bitter herbs for suffering, charoseth – a thick fruit and nut paste that reminds one of mortar. The passover seder ritual revolves around food stories drawn from Exodus. The dining practices of the meal itself however date from the period when the Holy Land was ruled by Seleucid Greeks, and is essentially a form of symposium, particularly in the reclining while eating, drinking four glasses of wine, and hiding the afikomen (i.e., epicomium, outside the meal) – a matzoh that the children search for at the end of the ritual. Today, sticking to the exact letter of the law, matzoh meal is used to make a kind of passover bread, cakes, and other products, which because not technically leavened, are still allowed.

In general, modern Judaism retains the kosher rules, though sacrifice ended with the destruction of the second temple in AD 70. However, a good proportion of practicing Jews, particularly in the Reformed tradition no longer adhere to any dietary restrictions, and there is also a broad range of levels of compliance, from those who keep kosher only in the home, to those who merely avoid pork, but follow no other restrictions. Nonetheless, food is so intimately bound to Jews' cultural identity, that active or not, certain foods are absolutely requisite at family gatherings and holidays, which at least for Ashkenazi Jews include familiar items such as bagels and lox, corned beef and pastrami, knishes, gefilte fish, and matzoh ball soup.

Christianity

Christianity grew directly from Judaism and in an effort to distinguish itself therefrom, the early church made an explicit point of abandoning the kosher rules. It is not what goes into the mouth but what comes out that defiles a man, as Jesus himself explained, meaning food cannot pollute a person, only cruel words. There is also a scene recounted by Matthew in which Jesus is asked why he doesn't fast and he responds that "when the Bridegroom will be taken from them, then shall you fast" meaning that once he is gone there will be occasion to fast, but fasts should not be regularly scheduled and habitual as among standard Jewish practice. Regarding clean and unclean food, there is also a vivid dream told by Peter in which a huge net teeming with creatures lowers from heaven and he is ordered by God to kill and eat. Not that the early church was entirely bereft of attitudes toward food, though.

The letter to the Corinthians resolves the question of whether it is alright to consume meat that had been sacrificed to pagan gods. Although technically permissible, Paul counsels to avoid it lest one lead fellow Christians astray. In this time the central sacrament of the church also developed, based on the Last Supper when Jesus was sitting at a Passover seder with his disciples and asked them to remember him when they ate bread and drank wine, alluding to the fact that he would be gone the next day, and reminding them, "this is my body; this is my blood." The sacrament of the holy communion is based on these words, and in 1215 the Lateran Council decreed that the bread and wine literally transform into the flesh and blood of Christ, which is consumed by communicants. It is in this way that the faithful obtain grace, forgiveness of sins, through the act of eating.

Early Christians also fasted, not only in miraculous ways as had Moses and Jesus who ate nothing at all for 40 days, but as a whole community during impending disaster to ask for God's mercy, or as individuals as an act of penitence. The early church also saw the

development of monastic orders, which followed ascetic regimens limiting the amount of food eaten and frequently abstaining from meat, which was believed by medical theorists to stimulate the production of blood and sperm and ultimately incite the libido (in both men and women). Celibate orders naturally restricted meat consumption, though the *Rule of St Benedict*, which generally set the pattern for orders in the West is not entirely abstemious, allowing for example, the equivalent of a few glasses of wine, though ideally it would be best if monks could abstain. Certainly an antipathy toward the sin of gluttony pervades Christian thought, for it was not only the first sin in Eden, and leads to other sins like sloth and lust, but prevents one from exercising virtues such as charity.

The general ascetic attitude toward meat did eventually, in the course of the institutionalization of the church, lead to a number of official fast days. These were every Saturday (starting on Friday night, hence fish on Friday), the vigils of saints days and the entire 40-day period of Lent stretching from Ash Wednesday to Easter, minus Sundays. The fast was defined as abstention from meat and all meat products such as butter, eggs, and milk, though typically a dispensation could be obtained for children, pregnant women, the infirm, and for special cases. Fasts alternated with feast days, the best known being Mardi Gras, the celebration preceding Lent when all remaining meat has to be consumed, as well as eggs and butter – in pancakes and other confections. In general this was a time for ritual subversion, serving as a safety valve for society, which would naturally return to the status quo once the celebrations were over. Most places abolished these celebrations in the course of the sixteenth century though, the rare exceptions being Venice and New Orleans where Carnival is still celebrated.

These fasting regulations remained uniform in the Roman Catholic Church until the 1960s and the Vatican II Council, when adherents were asked to give up something important as a sacrifice, but not necessarily meat. However by custom certain days do remain fish days, including Christmas Eve among Italians. Fasting regulations also remain in place in the Eastern Orthodox Church and are even more extensive, and for certain fasts a broader range of foods is prohibited, such as olive oil.

The Protestant denominations returning to scriptural authority over tradition took a number of different positions on official fasts. The Church of England kept a "Political Lent" requiring fish consumption as a way to support the fishing industry and the navy. Eventually the practice fell into abeyance, likewise in Lutheran Churches. In the Reformed tradition (i.e., The Swiss, Dutch, Scots, English Puritans, and French Huguenots) fasting once again took its biblical form, as a complete abstention from food as an act of penitence or a communal fast to avert God's wrath. In general, however the practice ceased, though many evangelical Christians still fast for religious purposes. Nonetheless, the Reformed churches did adopt a new attitude toward food stressing frugality, simplicity, and at times abstinence from alcohol. The Prohibition of alcohol in the US in the early twentieth century sprung directly from the work of the Women's Christian Temperance movement, which may seem ironic given that Jesus drank wine and even miraculously provided guests with a bounteous supply on one occasion.

A number of unique attitudes toward food developed among more recent Christian sects, for example, the modern vegetarian movement sprang from Bible Christian Societies on both sides of the Atlantic, and the 7th Day Adventists, following the visions of Ellen White are the only sect that demands total abstention from meat, alcohol, and tobacco, though Mormons (Church of Jesus Christ of Latter Day Saints) do abstain from the latter two as well. Historically Christian sects have expressed their ethical positions in ways that have a direct bearing on food practices. For example, Quakers who opposed slavery in the nineteenth century abstained from products made by the plantation economy, including sugar, molasses, and rum. More recently there have been explicitly Christian weight loss diets, Christian groups advocating fair trade or ethical treatment of animals. Suffice to say that although all Christian sects promote charity, any number of ethical positions and dietary regimes may fit under its umbrella today.

Islam

The food tenets of Islam also bear a relation to the previous two faiths. The most obvious similarity with Judaism is the ban on pork. In general there is not the great number or complexity of food rules, though birds of prey and similar animals are not considered halal (legal to eat). Animals to be consumed must however be ritually slaughtered without pain, thanking the animal and giving praise to Allah. The most important food custom is the month-long fast of Ramadan, when the faithful must eat and drink nothing between sunrise and sundown. The meal in the evening, to break the fast, traditionally eaten on the floor and with fingers of the right hand only, may be quite elaborate and sumptuous; some even report gaining weight during the holy month. The fast is also broken with Eid al-Fitr, a resplendent feast with special foods for the occasion, such as dates, which are a traditional food of the Arabian Peninsula.

A unique prohibition in Islam is the ban on alcohol. This, as much of the religion, stems from the personal experience of the prophet Mohammed, who after having witnessed a scene of drunken violence, understood that it would be best if alcohol is never consumed. Some Muslim countries forbid the sale of alcohol entirely, whereas others are more lax and even produce excellent spirits such as Raki in Turkey or Arak elsewhere. It has been suggested that the importance of coffee and coffee houses in the Muslim world is a result of the ban on alcohol, and caffeine is also purported to keep holy men awake for long hours of prayer.

Perhaps the most pervasive custom in the Muslim world is the charge to perform acts of charity and show effusive hospitality to guests. Originating in the very practical need to feed strangers who would otherwise starve in the desert, generosity with food has become an essential part of the practicing Muslim daily life. Devout Muslims will say a prayer before eating and thank Allah when finished. There are also a number of religious festivals celebrated with food. Eid-al-Adha commemorates the willingness of Abraham to sacrifice his son Ishmael (not Isaac as in the biblical tradition). On this day a ram is sacrificed and a third is kept for the family, a third given to friends and neighbors, and the last third given to the needy.

A unique phenomenon among African-Americans is the movement founded by Elijah Mohammed known as the Nation of Islam. Its tenets include rejection of what its founder considered slave food; pork and offal meats, black eyed-peas and greens stewed with fatty meat. Mohammed was not only interested in dietary reform but wrote an entire book on the topic, which also counseled abstinence from debilitating intoxicants which cause dependency and subjection. Though many members have gravitated toward traditional Sunni Islam, the group still survives.

Hinduism

The origins of Hinduism can be traced to the Indo-European invaders (Aryans) who arrived in the Indian subcontinent after approximately 1500 BC. They were originally sacrificers and consumers of cattle and how exactly they came to ban cows for food is among the more hotly debated topics among scholars. First it is important to understand that the ancient Vedic texts regard all living creatures as manifestations of the first primordial principle, known as Atman, which translates as "self." Atman, having divided and subdivided gave rise, therefore, to all living beings, which are manifestations of the original Atman. In daily life people often have difficulty recognizing the unity of all creation, which is how abstinence and yogic practices, including meditation, help us to recognize the self – which is every creature. Moreover, on death all creatures are reincarnated in a different form in accordance with their conduct in the previous life. These ideas, at least to start with were not implemented in a way that determined diet.

Second and equally important was the division of society into separate castes, seemingly at odds with the idea of the unity of creation, but separating people by various professions with no possibility of social mobility. Unless you were demoted by marrying below your caste, you and your descendants were always and forever in the same caste. This social structure had a profound impact on food customs, because one could not eat with or accept food from those of a lower caste, which would become polluted. The highest of castes, the priestly Brahmins had, as we shall see, the most restricted diet, with the greatest possible sources of pollution.

At first the Brahmins were eaters of meat, but apparently a period of widespread famine challenged their position at the top of society as the lower castes threatened them with violence. The Brahmins maintained their status by implementing the full ramifications of their sacred texts into practice. That is, they asserted that because the cow is the highest order of reincarnation, it should be sacred, revered, and never consumed. Hindus went from cow consumers to ardent reverers of cows. Because the priests had the most abstemious diet, they could no longer be the object of envy or violence from the lower castes, as they ate less and their diet was much more restricted. In fact the lower the caste, the more foods were allowed. The idea that cattle are more efficient as providers of traction for plows, dairy products, and manure for fuel than they would be if used for meat, may be true but it does not explain the origin of cow reverence.

Although the caste system has for the most part broken down in the modern era, beef is still prohibited for Hindus, and certain cow products like ghee (clarified butter) are considered the most sacred. A countless number of holy festivals also revolve around food. It is important to remember that Hinduism is a polytheistic faith, and thus festivals worship any number of different gods, the three main gods Brahma, Vishnu and Shiva, but also their many avatars and family members. Offerings to the gods, *prasad*, often consist of foodstuffs: fruits, sweets, or milk products. The offerings are consumed after the ritual. There are an extraordinary number of festivals throughout the year as well, seasonal, honoring particular deities, all of which include eating particular festival foods. For example, Mahashivarati celebrates the marriage of Shiva to Parvati when the temples flow with offerings of milk, yoghurt, honey, and ghee, but devotees fast through the day and the fast is broken by a special vegetarian meal made with no grains. Janmashtami celebrates the birth of Lord Krishna, Ganesh Chaturthi the birthday of the elephant-headed Ganesh. Diwali, the Festival of Lights features sweets, which are given to children and even fed to cattle.

Buddhism

Buddhism developed directly from Hinduism when Siddhartha Gautama in the *c*. 6th BC was searching for the meaning of life and was suddenly enlightened with the realization that suffering in the world comes from never being satisfied with what one has, from being far too attached to one's self. Neither extreme asceticism nor indulgence, both inherently selfish, lead to happiness, but rather the middle way, eating enough to survive. The key to ending suffering is simply to recognize the self as an illusion and to break the chain of causation, stop the endless cycle of reincarnation and achieve the state of nirvana, or nothingness.

Before this, however, were very practical Buddhist directions for living life in this world. Most important is the principle of nonviolence or *ahimsa*, not to cause suffering to any other creature. By logical extension, this meant also not killing them for food, and vegetarianism is thus at the core of Buddhist belief. There are varying degrees of practice among the many forms of Buddhism though. Normally monks are the only strict vegetarians and they created an elaborate nonmeat cuisine featuring items like high protein tofu. Though many modern Buddhists around the world are vegetarian, among most varieties fish is eaten regularly, and even meat in some.

Interestingly, when Buddhism became the state religion in India under the ruler Ashoka in the third century BC, meat eating became illegal as well. Although the religion did not survive in India it spread northward through Asia influencing places as far away as Japan, where Zen Buddhists traditionally abstained from meat. Ironically the largely vegetarian population of southern India, although mostly Hindu, were most likely influenced by Buddhist practices. In East Asia, Buddhism is not considered an

exclusive faith, meaning that in China, for example, it may be mixed with elements of Taoism and Confucianism, or in Japan with traditional Shinto.

Although Confucianism does not have any explicit food-related rules, the practice of filial piety does require that people act with deference to superiors and care for inferiors. This often translates into a practical requirement to feed elders first, without neglecting to provide for those below one in status.

Another religion arising in India, the Jains, adopted an extreme vegetarian position, such that to harm any living creature, even accidentally is forbidden. They are naturally pacifists as well. In terms of diet, this also means avoiding root vegetables, the harvesting of which kills the entire plant and they may harbor many microorganisms. Jains often filter water to prevent destruction of the same. Monks and nuns in this tradition may also practice forms of asceticism, and among all Jains there are set times for fasting, though one may also elect to fast individually for a variety of reasons. There are also varying degrees of fasting, some lasting many days.

Conclusion

Religion continues to shape foodways around the globe, not only in traditional forms of fasting and celebrations or food taboos, but in particular in the ethical attitude toward our fellow creatures as well as our responsibility for stewardship of the environment, and most importantly for the responsibility we owe to perform works of charity to help sustain our fellow humans.

Further Reading

Bynum, CW, 1987. Holy Feast and Holy Fast: The Religious Significance of Food to Medieval Women. University of California Press, Berkeley.
Douglas, M, 2002. Purity and Danger: An Analysis of the Concepts of Pollution and Taboo. Taylor, London.
Fagan, B, 2006. Fish on Friday: Feasting, Fasting and the Discovery of the New World. Basic Books, New York.
Feeley-Harnik, G, 1994. The Lord's Table: The Meaning of Food in Early Judaism and Christianity. Smithsonian Institution Press, Washington.
Gelder, GJH van, 2000. God's Banquet: Food in Classical Arabic Literature. Columbia University Press, New York.
Greenspoon, LJ, Simkins, RA, Shapiro, G, 2005. Food and Judaism. Creighton University Press, Omaha.
Griffith, RM, 2004. Born-Again Bodies: Flesh and Spirit in American Christianity. University of California Press, Berkeley.
Harris, M, 1989. Cows, Pigs, Wars and Witches: The Riddles of Culture. Vintage, New York.
Khare, RS (Ed.), 1992. The Eternal Food: Gastronomic Idea and Experiences of Hindus and Buddhists. State University of New York Press, Albany.
Muess, R, Grummett, D (Eds.), 2008. Eating and Believing. T&T Clark, London.
Sack, D, 2001. Whitebread Protestants: Food and Religion in American Culture. Palgrave Macmillan, New York.
Simoons, FJ, 1994. Eat Not This Flesh: Food Avoidances from Prehistory to the Present, 2nd edn. University of Wisconsin Press, Madison, Wisconsin.

School food systems

Manuel Franco[a] **and Marta Fajó-Pascual**[b], [a] Universidad de Alcalá, Facultad de Medicina y Ciencias de la Salud, Grupo de Investigación en Epidemiología y Salud Pública, Alcalá de Henares, Madrid, España; and [b] Faculty of Health and Sports Sciences, Universidad de Zaragoza, Instituto Agroalimentario de Aragón (IA2), Huesca, Spain

© 2023 Elsevier Ltd. All rights reserved.
https://doi.org/10.1016/B978-0-12-375083-9.00271-3

Introduction 341
Food systems for children and adolescents 342
Determinants of food systems for children and adolescents 342
Food supply chains 342
Food environments 342
Behaviors of caregivers, children and adolescents 343
School food systems 343
School food systems and food programs 343
Historical development of school food programs (1850-present) 344
The example of the Brazilian national SFP 345
School food systems interventions 345
School food interventions and impact on dietary behaviors 346
School food interventions and impact on anthropometrics and metabolic outcomes 346
Barriers evaluating school food interventions impact on health outcomes 346
School food interventions and sustainability 347
Next steps and opportunities 347
Conclusion 348
References 348
Relevant websites 349

Key points

- Reorienting food systems for children and adolescents is central to improve children's healthy and sustainable diets.
- Interventions on food systems for children and adolescents focusing on schools offer promising venues for improving children's diets and health with a special focus on most vulnerable children and adolescents.
- School settings offer relevant opportunities for interventions that range from food subsidies and taxes, the provision of daily healthy and sustainable meals and the improvement of school food environments.
- School Food Systems changes should include a policy equity perspective to protect childhood optimal nutrition as a fundamental human right ensuring at least one healthy and sustainable meal per day.

Introduction

An adequate nutrition in early childhood, the school-age years, and adolescence is a fundamental piece for physical growth, learning process and adult healthy life. Healthy children and adolescents (C&A) in our societies are the basis of a prosperous world. The current child and adolescent health situation around the world is far from the ideal levels of health achievement. Undernutrition—stunting and wasting—is still prevalent in young children, at the same time that overweight and obesity among children and adolescents continues rising or stagnated at high prevalence in most parts of the world including regions where undernutrition is still prevalent (Aguayo and Morris, 2020).

An important, and often neglected, matter when studying population health is the social gradient of health and nutrition that characterizes our current populations. Malnutrition, in the forms of overweight and obesity disproportionately occurs in underserved and vulnerable communities within our populations. And these health and nutrition inequalities continue rising in most countries, including high income countries like the US and European countries. Food insecurity continues increasing in countries like the US and Spain, a situation that worsened during the COVID-19 pandemic. A relevant topic when addressing children health and nutrition is the high cost of healthy diets in comparison to cheap energy-dense, nutrient-poor foods. The recognition of health and nutrition as a human right specially for C&A needs to be moved forward in most countries' political agenda. The need for including health and nutrition as part of Child Guarantee schemes in our countries is an ongoing political debate in Europe.

In the last decades, it has become clearer that food production significantly contributes to the challenge of sustainability by contributing to biodiversity loss, climate change, freshwater usage, food loss and waste and long-distance transportation, for example. Individuals and families contribute to these problems through their daily food choices (Oostindjer et al., 2017).

A food systems approach has been developed in the last decades for addressing and understanding the complexity of all the factors that finally affect the dietary patterns we follow and the related nutritional and health outcomes in different societies. The most common definition of a food system is as follows: "a food system gathers all the elements (environment, people, inputs, processes, infrastructures, institutions, etc.) and activities that relate to the production, processing, distribution, preparation and consumption of food, and the output of these activities, including socio-economic and environmental outcomes" (HLPE, 2017).

Food systems thinking should also help ensuring that food is produced, distributed and consumed in a sustainable manner protecting the right to adequate food for present and future generations (HLPE, 2017). The way modern food systems are thought, helps policy-makers focusing on the availability and accessibility of diverse and healthy diets, particularly for the underserved and the most vulnerable.

Food systems for children and adolescents

A food systems approach promoting C&A's adequate health and nutrition while minimizing environmental degradation is warranted. Food systems research had seldom considered children and adolescents as a focus until UNICEF and the Global Alliance for Improved Nutrition (GAIN) convened a Global Consultation on Food Systems for Children and Adolescents at the UNICEF Office of Research, Innocenti, from 5 to 7 November 2018. The results of this meeting have been published in a special issue on Food Systems for Children and Adolescents addressing the need for food system reorientation with a child-centered approach (Aguayo and Morris, 2020).

Food systems comprise actors, components, policies and services that influence and interact with each other in a complex ecology where the whole system needs to be synchronized to ensure nutritious, safe, affordable and sustainable children diets.

Actors lie both in public and private sector. Food systems require from all food system stakeholders the most difficult task of responsibly assuming their roles in shaping children diets.

Governments have the higher responsibility for upholding children rights to adequate nutrition and health and must set food standards aligned with the special needs of C&A. Producers and suppliers must ensure that their actions are aligned with those standards. Finally, governments shall generate the conditions to allow for participation of small-scale stakeholders, families, women and vulnerable groups among other players.

Raza et al. (2020) developed the Innocenti conceptual framework of food systems for C&A that shape their diets comprising four types of diet determinants: food supply chain, food environment, both personal and external, and children/household eating behaviors.

Within each determinant Raza et al. identified a series of "influencers" that are more immediate and individual/level factors determining the extent to which a determinant contributes or not to the goal of improving C&A diets becoming potential entry points for food system change.

Determinants of food systems for children and adolescents

Food supply chains

Food supply chains take food from production to consumption. Small and medium scale farms produce 50% of world food. They face challenges diversifying their production due to inputs cost. Public sector shall support them in aligning their practices with C&A dietary needs given the investments needed in practices and foods of current limited demand in the market.

Government can secure markets for these foods and sustainable food practices fostering partnerships with schools (farm-to-school twinning) and designing sustainable public food procurements strategies that truly value C&A nutritional health, community social justice and environmental integrity, the three pillars of sustainable development (Dos Santos et al., 2022).

Food environments

External food environments

The school is a salient food environment for C&A as it allows reaching them at population level at least until mid-adolescence. C&A spend more time in school than in any other environment.

Food systems may differ by country even by state or region. For example, in the US around 30 million C&A participating in the National School Food Programs (lunch and breakfast) in public schools may consume up to half of their daily calories at school (Cohen et al., 2021), which underlines the importance for these calories being healthy ones.

Because of the time spent by C&A at school at a time in their lives where food habits are established and therefore might track into adulthood, the school setting provides multiple opportunities to learn healthy but also sustainable food behaviors, while mutually reinforcing each other.

School meals in any given country are a nutrition and health safety net for low-income and vulnerable children (HLPE, 2017). The availability and accessibility of water coolers and water fountains has received plenty of attention in the last years given the current situation of sugar sweetened consumption and plastic bottles and containers increased use.

Personal food environments

Accessibility to healthy food options around schools and the place of residence remains an important factor within food systems thinking.

The affordability of healthy dietary patterns is a major economic problem for many families as shown in the high levels of food insecurity. Social protection schemes at schools providing healthy and free meals to low-income families and those in need is, again of most relevance.

Convenience foods offered in school and high-school breaks are mostly cheap and ready-to-eat foods, high in sugar, salt and calories compromising optimal diets.

Behaviors of caregivers, children and adolescents

The Innocenti framework visualizes this final determinant where the external food environment i.e. school and personal food environments converge. Desirability and acceptability of food options are linked to social and cultural norms but can be heavily influenced by marketing and advertising (Fox and Timmer, 2020).

This framework conceptualized the dynamic linkages between the elements of food systems and highlighted the importance of continuously shaping food systems to deliver nutritious, safe, affordable, and sustainable diets to children and adolescents.

School food systems

What children eat depends on the food system environments in the communities where they live and where they attend school. School settings have been identified for decades as places for change in public health nutrition.

The school food system is a subsystem within the broader food system of particular relevance for C&A. It also a food environment i.e. a physical and socio-cultural context whereby C&A interact with the broader food system to obtain the food they need most commonly as lunch but also as breakfast and/or snacking.

Because of the time dedicated to eating at schools, and the influence of school social environment (teachers and peers) on C&A food choices, schools are in a unique situation to influence the behavior of many children simultaneously cutting across socio economic status.

Healthy food habits are more easily acquired in childhood, and early learning increases the chances of habits persisting into adulthood. School meal programs are therefore a tool for improving dietary behaviors, nutritional outcomes in a sustainable way (Oostindjer et al., 2017).

Education and health policies also have the possibility to serve as examples and levers of change in other sectors of society where scaling up healthy and sustainable food production and consumption is also needed (e.g. hospitals, nursing homes, universities). School Food System approaches should acknowledge their transformative power and ability to drive change.

The SchoolFood4Change project (Website https://schoolfood4change.eu) funded by the EU scheme Horizon 2020 tackles the transformative power and ability to drive change of schools by bringing food and nutrition to the heart of the school mission reinforcing the health and sustainability parts of food consumption.

Researchers involved in the project used the following school food systems definition further explained in **Fig. 1** for school food systems promoting healthy and sustainable school diets.

School food systems include the full range of **activities, components,** and actors related to the production, processing, distribution, cooking, serving, consumption, and waste management of food in schools, as well as the **spaces and contexts** where these processes occur, and all the related educational practices, that promote children and adolescents' health in a way that is profitable for each actor of the food value chain, respects the social and cultural context, and safeguard, restore, and regenerate natural resources and ecological processes while respecting planetary boundaries. School food systems shall be governed democratically in a participatory way by all its **actors,** including farmers, procurers, chefs, teachers, pupils, parents, administrative municipal and school staff, researchers, other professionals, and policymakers.

Following a systemic approach, the school food system is a collection of subsystems with components, spaces, actors interacting in different processes through feedback loops. This is a fundamental characteristic of systems thinking that in the case of School Food Systems should help us focusing on the goals of promoting healthy and sustainable school diets for all the children attending with a focus on those most vulnerable.

School food systems and food programs

Within the school food system, the school food environment greatly interacts with C&A influencing their food choice since they spend a considerable part of their day at school where they have at least one meal (Cohen et al., 2021).

School food programs (SFPs) are frequently nationally mandated initiatives, sometimes regional and even municipal, to promote healthy and sustainable food behavior in C&A attending the different schools. Their interest lies on the fact that can reach school-aged C&A at population level across socioeconomic classes, particularly when school food programs are free, in a children development stage when food habits are formed and might therefore track into adulthood (Oostindjer et al., 2017).

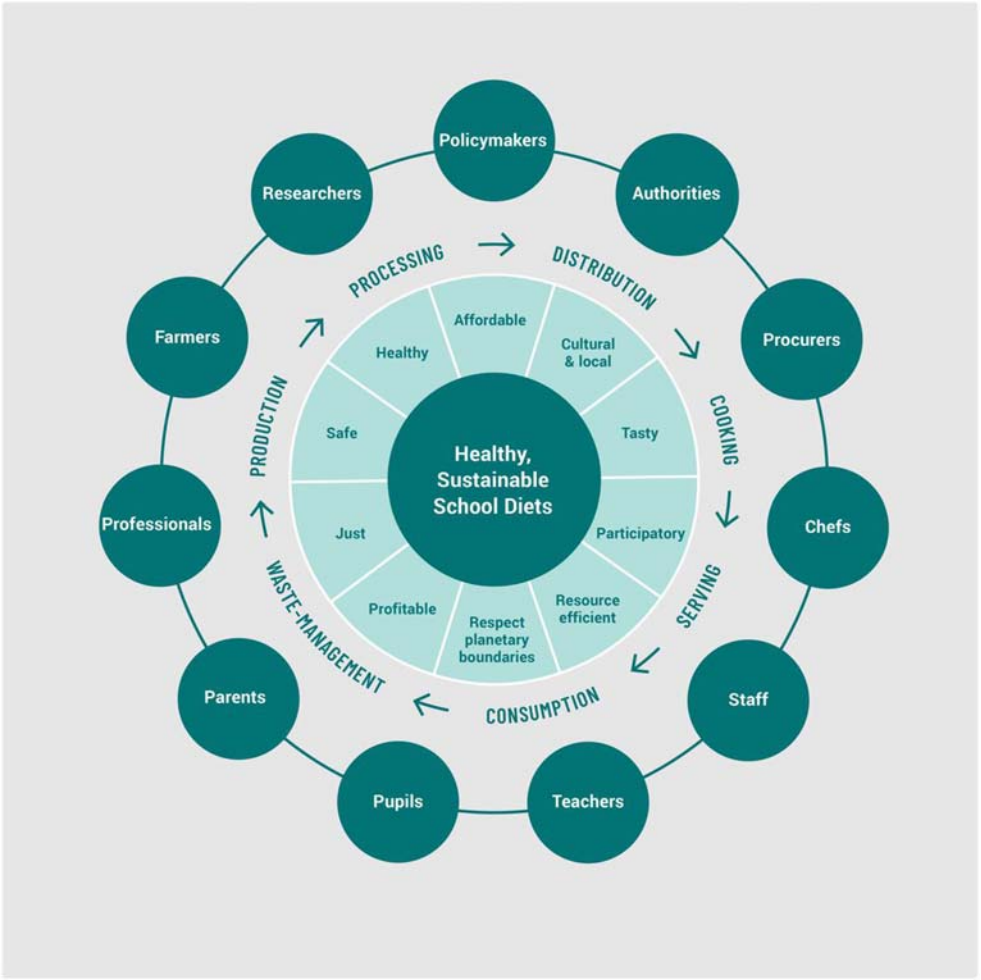


Fig. 1 Components of school food systems promoting healthy and sustainable school diets. This conceptual framework figure was proposed within the school food four change project. https://schoolfood4change.eu.

These programs provide lunch, but also breakfast or just one commodity such as school fruit or milk programs. In the USA, 95% of non-profit schools participate in the school meal programs administered by the United States Department of Agriculture (USDA). Both the National School Lunch Program (NSLP) and the School Breakfast Program (SBP) provide children and adolescents with healthy, low-cost meals throughout the school year. Additionally, before the COVID-19 pandemic, approximately three-quarters of the NSLP participants came from low-income households, with many relying on school meals for up to half of their daily energy intake (Cohen et al., 2021).

Schools are therefore in a unique situation to promote healthy and sustainable eating among C &A while at the same time, reducing disparities in the consumption of healthy foods such as vegetables and fruits (Longacre et al., 2014).

Historical development of school food programs (1850-present)

Oostindjer et al. (2017) have suggested three phases in the progression of nationally mandated SFPs in high-income countries. Those phases are described in **Fig. 2**.

In phase I (1850–1970) SFPs were introduced as a response to hunger and as such they were considered welfare programs aiming at providing sufficient food energy to low socioeconomic status children.

During phase II there was a shift toward food quality around 1970s in some European countries and in 1990s and 2000s in the USA and the UK as a response to increased prevalence of diet-related diseases resulting from increased food availability. Many developed countries, including USA and most European countries, issued national dietary guidelines focusing in reducing energy levels while increasing the number of nutrients per kilocalorie. This phase still continues as of today in many countries.

Phase III emerges since phase II seems unable to respond to individual (providing high-quality food to reduce all forms of malnutrition) and societal (preserving the environment to do the same for future generations) challenges in a sustainable food context. This phase is in its infancy in many countries coexisting with phase II. It aims at transforming SFPs in integrative learning

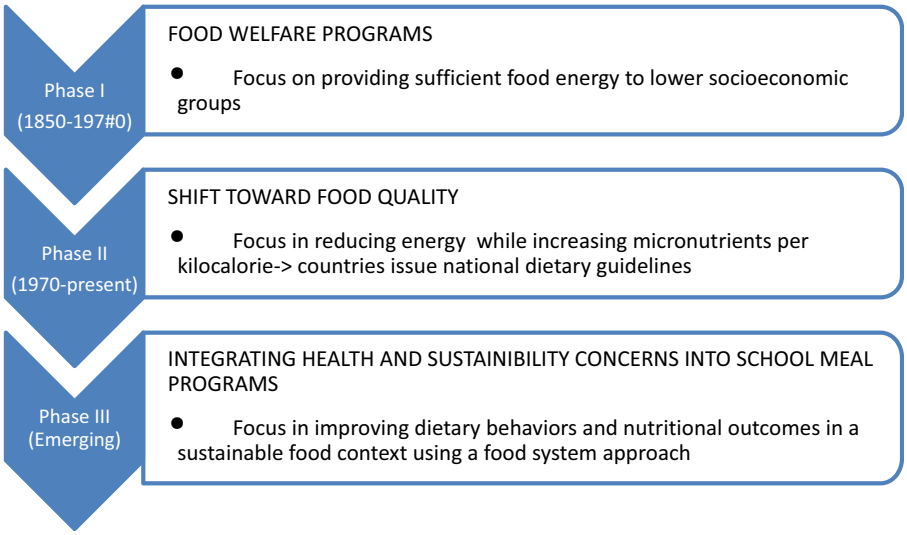


Fig. 2 Progression of school meal programs in the world after Oostindjer et al. (2017).

platforms for healthy and sustainable food behaviors at school using a food systems approach. This systemic approach requires synchronized changes across the food system components, given the interdependencies in the process of bringing food from farm to school fork and from there to school food waste (Hernandez et al., 2018). These concepts are being discussed and developed as of 2022 and therefore have not extensively studied yet.

The example of the Brazilian national SFP

The Brazilian SFP (PNAE in Portuguese) is a long standing national SFP that currently and after 2009 modified legislation, is worldwide recognized as a healthy diet promoting initiative (Boklis-Berer at al., 2021) offering Brazilian C&A attending public schools, free meals where products rich in Na, sugar and saturated and trans fats are restricted and the provision of low nutrition drinks prohibited including in the menus at least three portions of fruits and vegetables weekly 30% of their funds have to invested in local farm products provision and hence contributing to local development and food security of the community (Horta et al., 2019).

Regular consumption of PNAE provided meals at Brazilian schools has been associated with improvements in children diets by increasing consumption of healthy foods (pulses, vegetables and fruits) and decreasing unhealthy ones (salty snacks, crackers, sweet biscuits and sweets) further reaching those on higher social vulnerability risk (Horta et al., 2019). Additionally, the adherence to PNAE meals in adolescents (11–19 years) was associated with decreased obesity indicators (BMI, BMI z-score) showing a dose-response relationship (Boklis-Berer at al., 2021).

School food systems interventions

Downs and Demler (2020) identified in a scoping review the most effective interventions in the school food systems to improve diets and reduce the risk of malnutrition for C&A.

The school interventions showing most promising results were as follows:

- Food/snack/beverages subsides and taxes in schools
- Provision of free or subsidized fruit and vegetables
- Provision of free school meals
- Increased water access in schools
- Setting up school meal and competitive food & beverages standards
- Salad bar in cafeterias
- Menu labeling in schools
- School food environment changes
- Portion size changes
- Social marketing of healthier foods and beverages
- Cafeteria nudges and choice architecture

Storcksdieck Genannt Bonsmann et al. on their 2014 report of European school food interventions and policies, offer a number of suggestions to increase their acceptance and hence increasing their chances of success:

- Co-involvement of head teachers: Buy-in from them
- Building partnership at different levels
- Ensuring family and local engagement
- Local ownership and co-creation (engage children, teachers, canteen staff): use bottom-up approaches, social marketing, involve C&A in design and assessment
- Capacity building (teachers, coordinators, chefs, canteen staff, caterers): provision of materials made available through a centralized good practices portal or knowledge hub.
- Synchronized multi-component interventions

School food systems interventions are aimed to have impact at different aspects of C&A health. Next, we discuss the already studied and published intervention impacts on dietary behaviors, anthropometric and metabolic outcomes. We also include a discussion on why impacts school food systems interventions have been rather unsuccessful to show positive impact on C&A health outcomes.

School food interventions and impact on dietary behaviors

Downs and Demler (2020) indicate that the choice of food in C&A can be shaped by economic incentives i.e. subsidizing healthy foods such as fruits and vegetables and taxing unhealthy ones (energy dense snacks and/or beverages). The free provision in school of fruits and vegetables (F&V) has been more successful in increasing just fruit consumption in-school and habitual intake in the short term and to some extent in the long one (Micha et al., 2018). Provision of free school meals have brought up positive changes in dietary intakes such as in the case of the Brazilian SFP (Horta et al., 2019). School meal standards have increased intake of F&V reducing fat, saturated fat and Na with no effect on calories while competitive food/beverages standards has decreased reduced sugar-sweetened beverage intake and unhealthy snacks with no effect on total sugars. Improved access to water coolers in school has shown a non-significant trend to increased water consumptions (Micha et al., 2018). Salad bars, menu labeling and social marketing campaigns showed small increases in fruit and vegetables consumption and a reduction on energy and total and saturated fat intakes. Gender based differences were detected in the use of salad bars (Down and Demler, 2020). Approaches to cafeteria nudges and choice architecture that led to improvements in favorable foods included photographs of F&V in lunch trays, strategic placement of these foods in the lunch line, etc. Further description of the effect of these interventions on health outcomes can be consulted on Down and Demler article (2020).

School food interventions and impact on anthropometrics and metabolic outcomes

The formerly mentioned school food interventions were more likely to bring up changes in dietary intake, that in other more distal health outcomes such as anthropometrics.

The effect of school food interventions on anthropometrics (body mass index (BMI), weight status) has been variable and inconsistent in European countries as reported by Storcksdieck Genannt Bonsmann et al. (2014). Likewise, in LMICs, there is also limited evidence to support the modification of anthropometric outcomes through school-based food environment interventions such as eliminating unhealthy foods (absence snack bars/vending machines) and promoting school staff modeling C&A (Carducci et al., 2020). Boklis-Berer at al. (2021) have recently reported a positive effect on adolescents' rates of obesity in the Brazilian SFP.

Concerning metabolic risks indicators such as abdominal obesity, a recent systematic review on C&A by Leis et al. (2019) supports the role of nutritional education decreasing central obesity although effects on the remaining markers of metabolic syndrome were inconclusive. Changes in the availability of competitive foods and beverages in schools i.e. sugar sweetened beverages and unhealthy snacks have shown effect decreasing C&A metabolic risk (Downs and Demler, 2020).

Barriers evaluating school food interventions impact on health outcomes

In all, there is little evidence both from HICs and LMICs countries as to whether school food and school food policies can effectively improve C&A health status and/or eating habits. The barrier to gather this evidence relies on the difficulty in implementing comprehensive and well-designed studies, inferring causal effect from observational studies and the paucity of school food interventions sustained over a significant amount of time (enough duration) with sufficient intensity (multicomponent interventions required) and adequate and continued funding (Colley et al., 2019; Storcksdieck Genannt Bonsmann et al., 2014).

Lack of clear definition of measurable health outcomes (endpoints) has also hampered school food interventions evaluation and comparability. Proximal indicators i.e. short-term outcomes have been suggested to evaluate school food environment interventions such as the increase in favorable food intakes (F&V, pulses) or the increase in number of children of low income families participating in school meals (Oostindjer et al., 2017) as well as process indicators (implementation fidelity, C&A/teachers/canteen staff satisfaction, etc). However it is unclear how these short-term health outcomes relate to long-term outcomes such as overweight/obesity and type 2 diabetes (Oostindjer et al., 2017). Using indicators revealing additional forms of malnutrition have also being suggested to show school food interventions impact on C&A health e.g. vitamin deficiency, muscle mass loss, anemia in adolescent girls or metabolic risk (central adiposity).

School food interventions and sustainability

Food production, processing, distribution, consumption and waste management are associated with significant environmental impacts. Schools are not only an adequate context to bring up positive dietary and nutritional outcomes in C&A, they are also prominent environments for sustainability practices i.e. providing C&A with a sustainable food offer as well as building citizen conscience on sustainability issues integrating education for sustainability in the school curricula (Dos Santos et al., 2022).

Focusing school food interventions on the provision of healthy but also sustainable foods along with the promotion of sustainable food behaviors through school gardening and learning how to reduce food waste, might work in a mutually reinforcing way and in doing so, spilling into life away from school (family, community) (Colley et al., 2019). Schools sustainability practices might therefore impact the environmental pillar of sustainability but also the economic and social ones.

Dos Santos et al. in a recent systematic review (2022) identified the most frequently adopted food sustainability policies and interventions worldwide as follows:

- Educational activities for sustainability:
  - School gardens providing experiential learning i.e. hands-on experience on growing vegetables and workshops on using and cooking garden produce, travel field studies, etc.
- Food services:
  - Food supply/provision->sustainable food procurement including, beyond price: environmental criteria in the food supply contracts i.e. local or short chain, seasonal and organic foods, socioeconomic criteria i.e. farm to school or fish to school programs, provision from small producers, prioritizing supply from women and other vulnerable groups.
  - Menu planning: mainly plant-based, reduced meat supply, one day vegetarian/vegan menus, exposure of children to unfamiliar/unappreciated foods)
  - Menu production/consumption/distribution:
    - Reducing energy and water consumption in meal cooking, reduce organic waste (age-range adjustment of portion sizes, single dish menu)
    - Reduce inorganic waste->reusable devices, purchase in bulk or minimizing packaging, use of returnable bottles, and replacement of bottled mineral water by filtered water
    - Recycling/Composting
    - Food donation: serving people in vulnerable situations with relatively low investment

In 2021, the Food and Agriculture Organization (FAO) highlighted the importance of schools sustainable food procurement embedded in a wider public food procurement (PFP) strategy as a key lever for food systems transformation. In their report describe how PFP can influence both, food consumption and food production patterns and how it can be used as a development tool in the framework of school meals programs delivering multiple social, economic, and environmental benefits as well as health and cultural ones.

The EU funded SchoolFood4Change project (https://schoolfood4change.eu) is an example of innovative use of sustainable school food procurement to ensure healthy and sustainable food to C&A, particularly those more deprived.

Next steps and opportunities

In order to keep on developing School Food Systems knowledge and interventions that can have the highest impacts on C&A health and sustainability there are several topics where more and better knowledge is needed.

- As suggested by Oostindjer et al. (2017), being at the forefront of the shift toward phase three of school food programs will require a curricular integration of school meals with a focus on disease prevention and sustainability, as well as including current societal concerns such as climate change, environmental degradation, social justice and equity.
- A policy equity perspective is of paramount relevance given the highly unequal societies we live in nowadays with a well-known social gradient of C&A malnutrition. Protecting children health and nutrition as a fundamental human right and providing at least one healthy and sustainable meal per day in every school is a current challenge and opportunity.
- Harmonizing school food programs indicators for C&A dietary behavior and nutritional health and food sustainability will allow for evaluation and comparability and it is still the next step in knowledge and innovation application in schools worldwide.

As suggested by Hawkes et al. (2020), reorienting food systems toward healthier diet for C&A in any context is one relevant next step. Hawkes et al. recently developed a 6-step tool to identify actions toward child-centered food systems making healthy diets AAAA i.e. available, affordable, appealing, and aspirational in the context of children lives. The tool starts with the context in which children live or in our case, the school they attend and works back upstream;

1. Collecting data on nutritional health burden at the school level->malnutrition target e.g. overweight/obesity
2. Perform dietary assessment to identify both the eat more and eat less foods.

3. Use qualitative methods to understand the C&A dietary behaviors context (i.e. mapping of school food provision practices and assets).
4. Children-centered food environment measures at school->targeting food environment aspects to be changed to make certain foods more or less AAAA in short- and long-term.
5. Analysis of local food supply systems: Incentives or disincentives to create healthier food environment via a school food procurement supporting food offer diversification and sustainable food production practices (e.g. organic).
6. Develop mutually complementary actions: For example, the EU SchoolFood4Change project mix of actions: (1) healthy and sustainable food procurement, (2) Planetary healthy diets and cooking, (3) Whole School Food Approach to achieve a child-friendly food culture at school.

Conclusion

The current children and adolescent's health and nutrition status worldwide should be understood from three relevant features: High and rising prevalence of overweight and obesity in most countries of the world. The existence of a clear social gradient of health and nutrition in children and adolescents worldwide. The food system needs to take into account not only the provision and consumption of healthy foods, but the sustainability of all processes of food production and distribution to protect biodiversity and reduce its impact on climate change.

The study of school food systems offers different intervention opportunities to improve the health and nutrition status of children and adolescents, especially for the most vulnerable. School food systems interventions should also promote more sustainable diets.

Several conclusions emerge from this work on school food systems. Frist, interventions should include a wide range of intersectoral policies and actions comprehensively implemented in concert to achieve highest impact. Second, education and health policies need to be advocate for the relevant issue of healthy food affordability. Finally, engagement and accountability of both public and private sectors are needed.

School settings offer a clear opportunity for sustainable, healthy and equitable food system change including the development of innovative and sustainable food procurement criteria, the promotion of planetary healthy diets and cooking, and the use of a whole school food approach for schools to become integrative learning platforms.

References

Aguayo, V.M., Morris, S.S., 2020. Introduction: food systems for children and adolescents. Global Food Secur. 27, 100435. https://doi.org/10.1016/j.gfs.2020.100435.

Boklis-Berer, M., Rauber, F., Azeredo, C.M., Levy, R.B., Louzada, M.L.D.C., 2021. The adherence to school meals is associated with a lower occurrence of obesity among Brazilian adolescents. Prev. Med. 150, 106709. https://doi.org/10.1016/j.ypmed.2021.106709. PMID: 34181943.

Carducci, B., Oh, C., Keats, E.C., Roth, D.E., Bhutta, Z.A., 2020. Effect of food environment interventions on anthropometric outcomes in school-aged children and adolescents in low- and middle-income countries: a systematic review and meta-analysis. Curr. Dev. Nutr. 4 (7), nzaa098. https://doi.org/10.1093/cdn/nzaa098. PMID: 32666031; PMCID: PMC7340483.

Cohen, J.F.W., Hecht, A.A., Hager, E.R., et al., 2021. Strategies to improve school meal consumption: a systematic review. Nutrients 13 (10), 3520. https://doi.org/10.3390/nu13103520. PMID: 34684521; PMCID: PMC8538164.

Colley, P., Myer, B., Seabrook, J., Gilliland, J., 2019. The Impact of Canadian school food programs on children's nutrition and health: a systematic review. Can. J. Diet Pract. Res. 80 (2), 79–86. https://doi.org/10.3148/cjdpr-2018-037. PMID: 30430855.

Dos Santos, E.B., da Costa Maynard, D., Zandonadi, R.P., Raposo, A., Botelho, R.B.A., 2022. Sustainability recommendations and practices in school feeding: a systematic review. Foods 11 (2), 176. https://doi.org/10.3390/foods11020176. PMID: 35053907; PMCID: PMC8775006.

Downs, S., Demmler, K.M., 2020. Food environment interventions targeting children and adolescents: a scoping review. Global Food Secur. 27, 100403. https://doi.org/10.1016/j.gfs.2020.100403. ISSN 2211-9124.

Food and Agriculture Organization (FAO), Alliance of Bioversity International and CIAT and Editora da UFRGS, 2021. Public Food Procurement for Sustainable Food Systems and Healthy Diets—Volume 1. https://doi.org/10.4060/cb7960en.

Fox, E.L., Timmer, A., 2020. Children's and adolescents' characteristics and interactions with the food system. Global Food Secur. 27, 100419. https://doi.org/10.1016/j.gfs.2020.100419. ISSN 2211-9124.

Hawkes, C., Fox, E., Downs, S.M., Fanzo, J., Neve, K., 2020. Child-centered food systems: reorienting food systems towards healthy diets for children. Global Food Secur. 27, 100414. https://doi.org/10.1016/j.gfs.2020.100414. ISSN 2211-9124.

Hernandez, K., Engler-Stringer, R., Kirk, S., Wittman, H., McNicholl, S., 2018. The case for a Canadian national school food program. Can. Food Stud. 5 (3), 208–229. https://doi.org/10.15353/cfs-rcea.v5i3.260.

High Level Panel of Experts on Food Security and Nutrition, 2017. Nutrition and Food Systems. A Report by the High-Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security. HLPE, Rome.

Horta, P.M., Carmo, A.S.D., Junior, E.V., Santos, L.C.D., 2019. Consuming school meals improves Brazilian children's diets according to their social vulnerability risk. Publ. Health Nutr. 22 (14), 2714–2719. https://doi.org/10.1017/S1368980019001459. PMID: 31241028.

Leis, R., de Lamas, C., de Castro, M.J., et al., 2019. Effects of nutritional education interventions on metabolic risk in children and adolescents: a systematic review of controlled trials. Nutrients 12 (1), 31. https://doi.org/10.3390/nu12010031. PMID: 31877685; PMCID: PMC7019568.

Longacre, M.R., Drake, K.M., Titus, L.J., et al., 2014. School food reduces household income disparities in adolescents' frequency of fruit and vegetable intake. Prev. Med. 69, 202–207.

Micha, R., Karageorgou, D., Bakogianni, I., et al., 2018. Effectiveness of school food environment policies on children's dietary behaviors: a systematic review and meta-analysis. PLoS One 13 (3), e0194555. https://doi.org/10.1371/journal.pone.0194555. PMID: 29596440; PMCID: PMC5875768.

Oostindjer, M., Aschemann-Witzel, J., Wang, Q., et al., 2017. Are school meals a viable and sustainable tool to improve the healthiness and sustainability of children's diet and food consumption? A cross-national comparative perspective. Crit. Rev. Food Sci. Nutr. 57 (18), 3942–3958. https://doi.org/10.1080/10408398.2016.1197180. PMID: 27712088.
Raza, A., Fox, E.L., Morris, S.S., et al., 2020. Conceptual framework of food systems for children and adolescents. Global Food Secur. 27, 100436. https://doi.org/10.1016/j.gfs.2020.100436. ISSN 2211-9124.
Storcksdieck Genannt Bonsmann, S., Breda, J., Louro Caldeira, S., Nelson, M., Wollgast, J., 2014. School Food and Nutrition in Europe: Policies, Interventions and Their Impact. EUR 26801. Publications Office of the European Union JRC91433, Luxembourg. https://doi.org/10.2788/11298.

Relevant websites

Global Research Consortium for School Health and Nutrition. https://www.lshtm.ac.uk/research/centres-projects-groups/research-consortium-for-school-health-and-nutrition.
School Food for Change EU project. https://schoolfood4change.eu.
The EU School Fruit, Vegetables and Milk Scheme From the Common Agricultural Policy, Supports the Distribution of Milk, Fruit and Vegetables to Millions of Children, From Nursery to Secondary School, Across the EU. https://agriculture.ec.europa.eu/common-agricultural-policy/market-measures/school-fruit-vegetables-and-milk-scheme_en.
The Global Alliance for Improved Nutrition (GAIN). https://www.gainhealth.org/about.
The School Meals Coalition, an Initiative Led by a Group of Member States and Partners to Ensure That Every Child has the Opportunity to Receive a Healthy, Nutritious Meal in School by 2030. https://schoolmealscoalition.org.

Seasonality of nutrition

Ilana R. Cliffer[a], Anastasia Marshak[b], Kate R. Schneider[c], Aishwarya Venkat[b], and Elena N. Naumova[b], [a] Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, United States; [b] Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, United States; and [c] Paul H. Nitze School of Advanced International Studies, Johns Hopkins University, Washington, DC, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of F. Branca, P. D'Acapito, Seasonality, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 178–185, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00247-6.

Introduction 351
Origins of seasonality 352
Defining and measuring seasonality 352
Timekeeping 353
Defining temporal units 353
Defining spatial units 353
Quantifying seasonality 354
Interactions of climatic seasonality with multiple systems 354
Food systems 354
Food supply, demand, and prices 355
Cultures and traditions 357
Livelihood systems 357
Labor 358
Environmental systems 358
Infectious disease and nutritional status 359
Pathogen survival 359
Feeding and caregiving 359
Food safety 360
Access to water 360
Climate change 360
Seasonality of nutritional outcomes 360
Growth, anthropometry, and body composition 360
Micronutrient status and deficiencies 361
Variation in seasonal patterns of nutritional outcomes 361
Sub-population variation 361
Spatial-temporal variation 362
Analytic decisions and observed nutrition seasonality 362
Conclusion 362
References 363

Key points

- Characterize seasonality in a way that captures the complex interactions among social, ecological, and agro-climatological systems which influence the seasonality of nutrition and health outcomes.
  - Demonstrate that seasonality is an observable property of every system, including both earth and social systems.
  - Show how the drivers of seasonality in nutrition and health outcomes relate to each other through complex systems and mutually reinforcing feedback loops.
  - Describe how seasonality in outcomes is a function of the interplay between human behavior and earth system seasonality, and how policy mediates this link.
- Propose systematic methods for defining and measuring seasonality in nutritional outcomes, considering climatic seasonality and spatial variability.
- Critically evaluate existing evidence on the seasonality of nutrition and health outcomes.
- Discuss how greater understanding of seasonality within a complex systems framework can be leveraged to improve well-being.

Introduction

At the most basic level, seasonality is defined as the periodic oscillation of systems and events with respect to a calendar. These cyclical patterns observed in human systems, including those that influence nutrition and health outcomes, are fundamentally tied to the earth's rotation on its axis and around the sun. Cyclical planetary movement generates seasons, which directly and indirectly affect agricultural production, food availability and prices (Chambers, 2009; Chambers et al., 1981; Gertel and Sippel, 2014; Gilbert et al., 2017; Kiehne and Mendoza, 2015; Sibhatu and Qaim, 2017), and infectious disease susceptibility. Such cycles can also be related to incomes, labor demands (Feuerbacher et al., 2020; Fink et al., 2020), and caregiving practices (Madan et al., 2018; Rao et al., 2019; Rao and Raju, 2020). Earth systems further determine social and institutional calendars as major holidays and festivals often occur during solstice, equinox, or planting and harvest periods. Environmental, social, and behavioral factors thus generate seasonality in all forms of malnutrition and its drivers, across all contexts (Kuhn, 2018).

Historically, nutrition researchers and practitioners have approached the seasonality of nutritional status considering the effects of a region's agroclimatic characteristics on nutritional status primarily in a unidirectional manner. Most of what we understand about seasonality in nutrition comes from research employing this linear framework due to its pragmatic utility. However, linearity obscures the complex and interconnected drivers of nutritional status, each having respective seasonal fluctuations, which interact in a series of feedback loops. We now envision nutrition seasonality as a byproduct of the complex interplay of several interconnected systems, environmental and health policies, technological innovations, and social and evolutionary factors (Fig. 1). These patterns reflect complex processes, encapsulating various spatial and temporal climatic behaviors that follow natural cycles, sensitive to

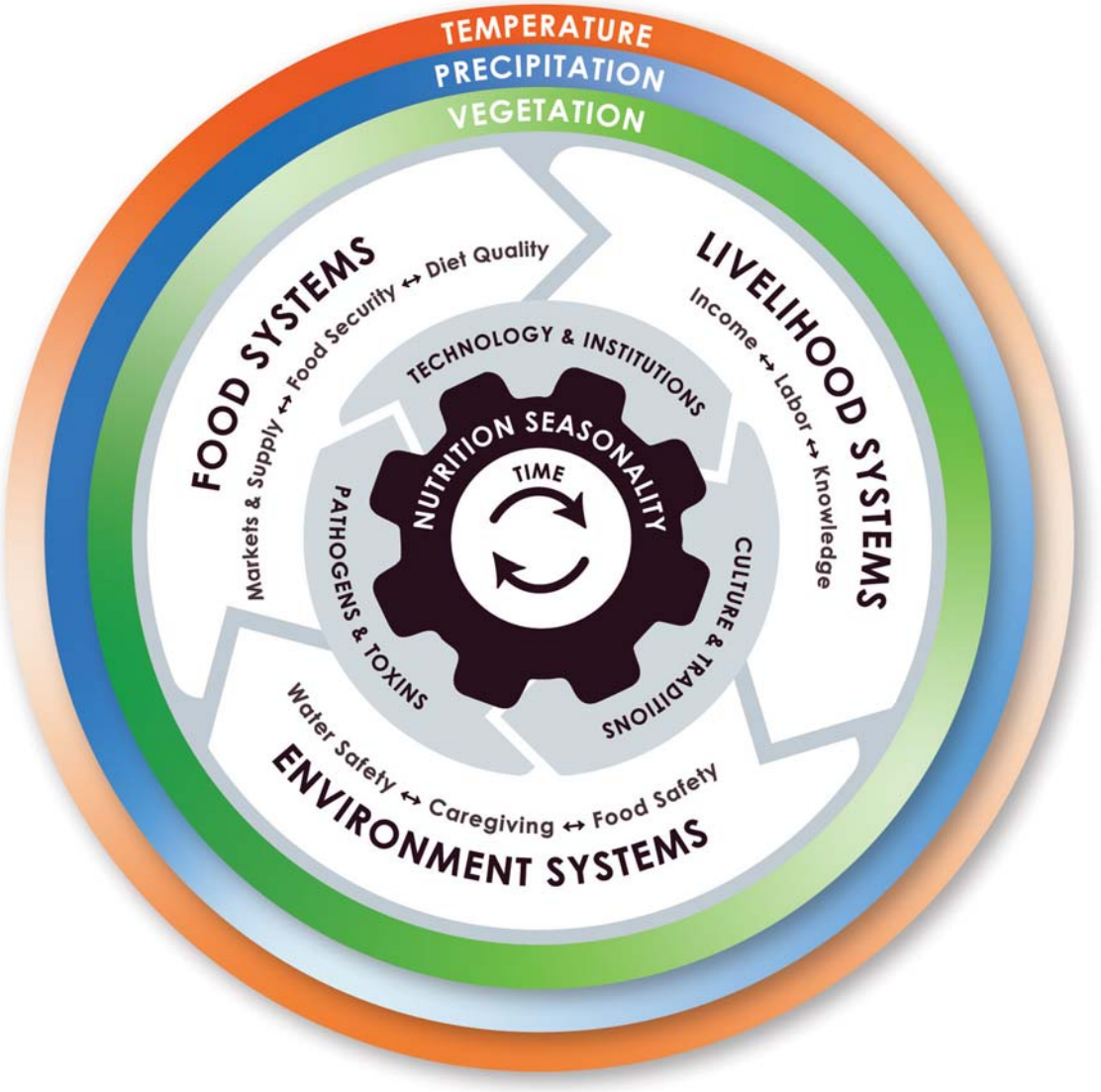


Fig. 1 Conceptual framework for seasonality of nutritional outcomes.

public policies, natural and manmade disasters, within-year cycles that affect local ecosystems, or between-year cycles in climatic patterns such as El Nino/La Nina (Marshak et al., 2021; Naumova, 2006; Naumova and MacNeil, 2006). Multiple scientific disciplines intersect in the field of nutrition, including human and molecular biology, epidemiology, economics, sociology, anthropology, demography, and public policy. Further complicating matters, each discipline defines and addresses seasonality using either disparate terminology or similar terms defined differently based on their respective theory, empirical evidence, and research methodologies. The resulting body of evidence lacks cohesion, contributing to a field-wide failure to fully understand the depths of nutrition-related seasonality.

Climate change, globalization, urbanization, and environmental degradation continue to exacerbate weather extremes and undermine the predictability of climate, threatening food systems and human health. A better understanding of the seasonal drivers of malnutrition is critical to design the most effective policies and programs that avoid the detrimental effects of seasonal climatic cycles on nutritional outcomes. Such understanding can inform a precise approach to nutrition, whereby programs and interventions are more specifically targeted in terms of timing, location, population, and delivery.

This article proposes a paradigm shift from the linear, imprecise, and disaggregated approach to the seasonality of nutrition outcomes, to a holistic, complex systems approach that considers the interconnections and feedback loops in the drivers and mediators of nutritional status as well as the patterns observable in nutrition outcomes. The article proceeds as follows: first, we define seasonality in relation to the Earth's systems and discuss how it can be characterized and measured. Next, we examine the features of seasonality through the lenses of feedback loops within several complex systems that interact with each other to influence the seasonality of nutritional status. We conclude the article with the observed seasonality of nutritional outcomes and thoughts on how knowledge of seasonality as it functions through complex systems can help optimize policy, programming, and research instruments to increase well-being.

Origins of seasonality

Seasons are the result of planetary, continental, and regional factors that influence air and water dynamics in the atmosphere. Seasonal climatic patterns vary by geography and underlie cyclic fluctuations in many biological systems sensitive to weather and photoperiods, including malnutrition and related health outcomes. The earth's annual orbit around the sun determines the seasonal transition from winter (cooler temperatures, shorter days) to spring, summer (higher temperatures, longer days), autumn, and back to winter. Due to the earth's tilted axis, the northern and southern hemispheres experience generally opposing seasonal cycles with seasonal extremes being more pronounced near the poles. Areas close to the equator, known as the tropics, experience only minor variations in temperature due to consistent direct sun exposure. Tropical regions are more humid and generally experience two distinct wet seasons, or a bimodal precipitation profile. Temperate regions away from the equator commonly experience a broader range of temperatures throughout the year, and often a unimodal profile with only one rainy season or monsoon (Moran, 2009; Rohli and Vega, 2017; Stevens, 2010). These climates sustain drier air, less humidity, and precipitation in the form of snow in the dry seasons. The physical environment also influences the flora and fauna of a region, giving rise to complex ecosystems (Rohli and Vega, 2017).

Climate informs human cycles of agricultural production, food consumption, and social traditions, all of which have important consequences for the seasonality of nutritional outcomes. All crops require a favorable range of temperature and sufficient rainfall for growth; the availability of these conditions determines regional crop calendars and growth cycles. Tropical areas have favorable temperatures year-round, but water availability for agriculture is often limited to few key times of year. Temperate regions have a smaller window of favorable conditions for agriculture when temperatures are sufficiently warm to prevent frost damage and sufficiently cool to avoid heat stress for plants. Therefore, tropical regions often have two or sometimes even three cropping cycles per year, whereas temperate regions may only have one major harvest with a smaller winter crop. These climatological patterns are critical for rainfed agriculture which occupies 75% of cultivated area globally (FAO, 2020).

Each location has its own climatological conditions, and what is considered the norm in one location could be an extreme in another. For a given region, extreme or near-extreme precipitation or temperature can result in a broad range of extreme weather events such as droughts, floods, hailstorms, tornadoes, and heatwaves. Climate systems across large areas are also interlinked, resulting in seasonal storms and cyclones which affect large portions of the Pacific and Atlantic coasts (Moran, 2009). These extreme weather events can cause damage to agricultural production and infrastructure with serious consequences for humans and animals. A single harvest in unimodal regions further increases the livelihood risk in case of a particularly bad harvest leading to diminished stores to last until the following harvest (Schofield, 1974). Although the rainy season can bring necessary water for agriculture, it can also cause outbreaks of water-borne diseases such as diarrhea, cholera, and typhoid or vector-borne diseases such as malaria and dengue fever. Climate extremes can thus increase the likelihood of adverse nutrition outcomes through compromised health, increased livelihood pressures, and diminished resilience (Watts et al., 2018, 2021).

Defining and measuring seasonality

Seasonality refers to systematic periodic (or cyclic) fluctuations in an outcome of interest. *Periodic* implies that the outcome repeats itself in regular intervals, and *systematic* refers to the fixed nature of this pattern within a range of natural variation. A *cycle* is defined

as the time needed to complete a pattern, or the time taken for the system to reset to initial conditions. The duration of this cycle can be defined analytically, derived from data, or arbitrarily selected based on biological, physical, physiological, or other assumptions. Often a calendar year serves as a proxy for a cycle or seasonal period.

Timekeeping

The first step in depicting seasonality is defining time and its units. Humans are continuously designing and redesigning tools to keep the record of seasonal features in nature and to assign meaning to seasonality. Social traditions have evolved over centuries in reaction to specific weather patterns, and humans have developed tools such as clocks and calendars to keep track of essential events, periods, and their sequence and repetition. A calendar is a system of reckoning time over extended periods with respect to the natural units: the day, the lunar month, and the year. From antiquity, calendars have provided the basis for planning agricultural, hunting, and migration cycles, for divination and prognostication, and for maintaining cycles of religious and civil events. Common astronomical cycles used to define time include the Earth's rotation on its axis (day), revolutions of the moon around the Earth (month), and revolutions of the Earth around the sun (year). There are roughly 40 different calendars used by humans in the world today, each using different parameters to define time (i.e., lunar, solar, lunisolar).

Understanding how the systems of defining time vary in different cultures is essential for further defining time for seasonal studies of health and nutrition outcomes. Researchers most commonly use the solar Gregorian calendar to operationalize seasonality due to its modern use as the international standard for civil purposes (Simpson et al., 2019). However, considering other ways of defining time may be important in contexts where the Gregorian calendar is not used and/or is not aligned with social traditions, migration patterns, or religious events (FAO and Tufts University, 2019). Regional and local calendars may have differently defined seasons and may also be more closely aligned with key environmental transitions affecting nutrition and health (Kulinkina et al., 2016b).

Every calendar also includes special days or sequences of days known as holidays or notable events. These events may vary by community and have substantial, multifaceted effects on health and nutrition. Most cultures mark planetary holidays indicating seasonal transitions based on the earth's rotation around the sun. Such festivals include the equinox, when day and night are approximately equal in length, and solstice, when day length is shortest or longest. Holidays may also follow fixed dates, falling on the same day each year in relation to the calendar which defines them (for example, Christmas on 25 December of the Gregorian calendar, Yom Kippur on the 10th day of Tishrei in the lunar Jewish calendar, or Ramadan, the ninth month of the Islamic lunar calendar commemorated by Eid-al-Fitr on the first day of Shabaan). Alternatively, holidays can occur each year on the same day of the week within the same month (for example, Independence days across the globe). State, federal, religious, and community-specific holidays can affect nutritional outcomes and the transmission and severity of related diseases through changes in food consumption patterns, widespread travel, days not in school or work, and restrictions on alcohol sales (Simpson et al., 2019).

Defining temporal units

In economic and epidemiological literature, calendar months are commonly grouped into a fewer number of periods to represent seasons (e.g., trimesters (four-month periods); quarters (three-month periods); semesters/half-year/6-month periods). The grouping of months into specific periods, even among studies in similar contexts, also differs across the literature. For example, in temperate climates of the northern hemisphere, some define January to March as the winter season, others define winter as December to February. Calendar years can also be divided into periods of unequal numbers of months, for example: summer (June to August) and remaining months (September to May). This is often done in contexts with unimodal rainy seasons, to separate the rainy from the dry season in analyses of nutritional outcomes.

The grouping of monthly data into periods raises methodological concerns and obscures true seasonality. Most parts of the world experience daily, weekly, and monthly variability in temperature as well as rainfall amounts, timing, and distribution. Aggregation of days or months to seasons (e.g., lean vs. plenty, dry vs. rainy) discards a meaningful sequence in temporal data for a rudimentary grouping. Treating continuous time as discrete categories through dummy variables (e.g., calendar months) in statistical analysis also masks the sequential and cyclical nature of the climatological phenomena that form the basis of our human-derived calendars. Climate-related characteristics such as temperature and precipitation are typically presented as a time series, or a sequence of values that are ordered with respect to a time reference. The ordered and cyclical nature of the time series must be preserved for accurate modeling. Furthermore, discrete models of time erroneously discount interannual variations in the environment and can diminish the importance of the time of rainy (or other) season onset. For example, if the rains began on 3rd April instead of 20th March, discrete approaches would document a shift of greater magnitude across categories rather than a delay of two weeks. Major calendars, including the Gregorian calendar, contain months with unequal number of day and thus the discrete representation discards the difference in true durations (Alarcon Falconi et al., 2020).

Defining spatial units

Considering spatial units is equally important for seasonality analysis. Spatial extent is most often defined by administrative boundaries at increasingly local scales, for example: country, state, county, and district. Administrative boundaries are frequently aligned with geographic features such as rivers and mountains which can physically separate communities. Existing surveillance systems

may also use administrative hierarchies for monitoring and reporting purposes; therefore, these boundaries are particularly useful for epidemiological and nutrition data (Lawson, 2013). However, administrative boundaries serve institutional purposes and are not inherently indicative of distinct demographic or climatological extents. Aggregation of outcomes across regions must be performed with care as administrative boundaries can be very different from agroclimatic, geographic, or demographic boundaries (Lawson, 2013; Tate and Atkinson, 2001). Spatial aggregation to the country or administrative unit scale can mask spatial variability in outcomes; for example, analysis of an outcome measured at the household scale in a small number of communities may not be meaningful when aggregated to the state level (Bordt, 2015; Tate and Atkinson, 2001). Physical proximity of communities also does not imply sufficient justification for aggregation as communities with different identities, livelihoods, migration patterns, and social calendars—and hence potentially different seasonal patterns in nutrition—can live in close proximity (Buzzelli, 2020). Therefore, high spatial granularity might not be sufficient to detect seasonality, while high spatial aggregation might lead to mixing seasonal patterns, making them indistinguishable. A meaningful spatial extent with sufficient resolution and internal variation is thus required for capturing seasonal patterns.

Quantifying seasonality

To measure seasonality and its effects on nutrition outcomes, there are two groups of statistical methods that can be performed using most commercial statistical packages. The first approach involves the use of indicator variables that reflect discrete time periods. Results are usually interpreted with respect to a reference category. For example, seasonality of malnutrition may be measured as the magnitude increase in prevalence of wasting in the dry season compared to a reference category of the wet season (Marshak et al., 2021). This approach is widely utilized due to its simplicity. Although insufficient for modern seasonality analyses that depict complex patterns and enable comparison of multiple seasonal patterns, this approach can provide basic insight into potential seasonal behavior when used with caution.

The second approach includes a broad range of modern methods such as harmonic regression to describe the timing, magnitude, and duration of a seasonal cycle. This approach utilizes a rich vocabulary to ease general comprehension and ensure reproducibility. **Table 1** provides a list of terms to define seasonality in conducting research and policy analysis.

Following the terminology, to quantify the timing of seasonal peaks (highest values) and nadirs (lowest values), measures to estimate the position of these points on a seasonal curve of the outcome of interest (i.e., malnutrition incidence or exposure) are needed. The magnitude related measures include peak and nadir values of the seasonal curve, the amplitude (the difference between peak and nadir values) and the relative intensity (the ratio of peak values and nadir values) (Marshak et al., 2021; Naumova and MacNeil, 2006; Ramanathan et al., 2020). **Fig. 2** depicts an example showing daily time series of weight-dependent nutrition outcomes and illustrates the seasonal attributes.

Additional characteristics may describe peak duration (when the outcome of interest is above a given threshold) and the degree of symmetry in the seasonal rise and fall of variables. The composition and relationship among these measures form seasonal signatures with their positioning, lead, and delay times relative to each other. In some cases, more than one seasonal peak is observed during a calendar year. When this happens, the peak with highest outcome values is defined as the primary peak, and the other as secondary (Naumova and MacNeil, 2006). By examining the relationships across different seasonal characteristics, for example peak timing and intensity, one could investigate potential drivers of seasonality or predict the changes in seasonal behaviors. These characteristics can also be used to study synchronicity across time series (Simpson et al., 2020). Similar peak timings of temperature and wasting, for example, can point to synchronization of the incidence of malnutrition with environmental and/or social processes.

Interactions of climatic seasonality with multiple systems

Climatic seasonality resulting from the earth systems underpins seasonality within food, livelihood, and environmental systems (**Fig. 1**). It also impacts the interactions between humans, other living organisms, and pathogens, influencing patterns of disease and malnutrition. Interactions among these interconnected systems affect seasonal patterns in nearly all drivers of nutritional status including food security; consumer demand and preferences; food safety; diet quality; labor demands (energy expenditures and incomes); time use and care practices; availability of adequate clean water, hygiene facilities and practices, and adequate sanitation; and disease incidence.

The impact of agroclimatic seasonality on malnutrition is mediated by the interaction of these systems with technology, institutions, policies, and processes. Interactions can take the form of feedback loops in which drivers and outcomes interact as mutually reinforcing forces mediated by human behavior. These feedback loops can either strengthen seasonal fluctuations or balance the impact of agroclimatic seasonality such that outcomes do not exhibit strong seasonal patterns (Meadows and Wright, 2008). Identifying feedback loops also reveals entry points for policy instruments to reduce negative impacts of seasonality on people's health and well-being.

Food systems

Food systems encompass all of the people, institutions, environments, and infrastructure relating to the production, processing, distribution, marketing, sale, and consumption of food (Fanzo et al., 2021, 2020; Global Panel on Agriculture and Food Systems

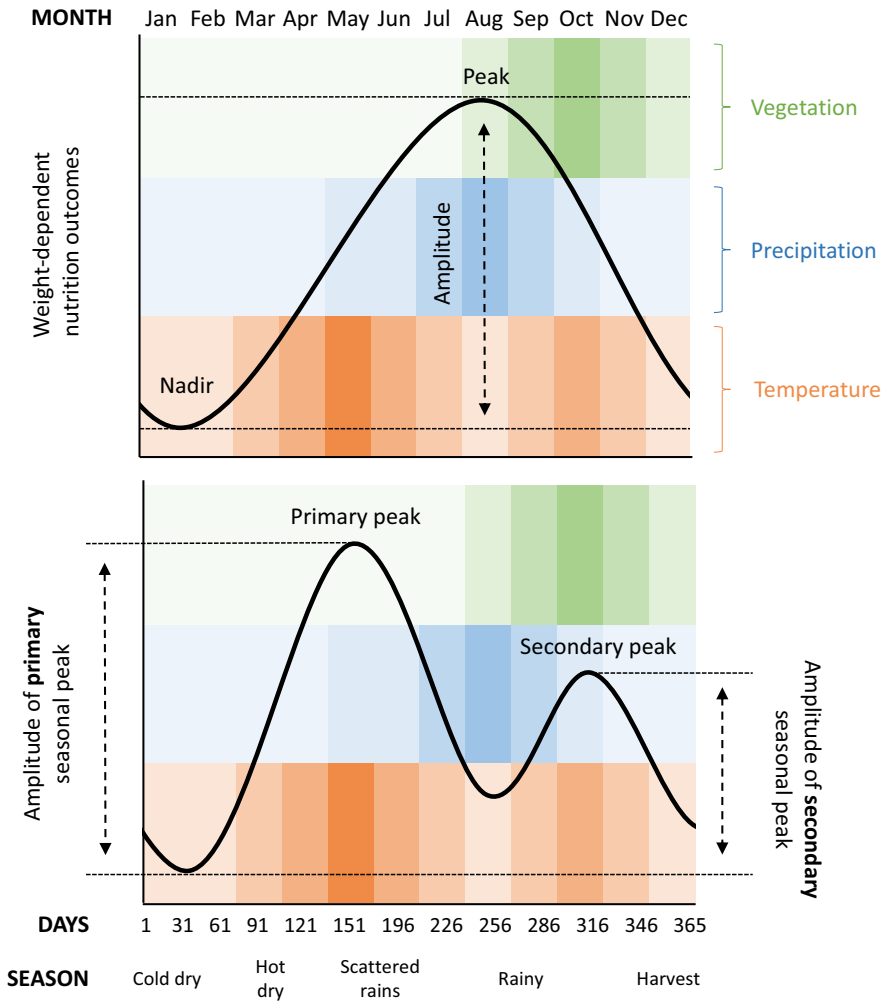
Table 1 Suggested terminology to describe seasonality in conducting research and policy analysis.

Term	*Definition*
Time series data	A set or a sample of time-referenced observations or records with an identified time period, time cycle, and time unit recorded by a timestamp.
Timestamp	Information on day, week, month in a conventional format (e.g. YYYY:MM:DD or YYYY:MM:DD:HH:mm) of data collection or processing.
Time series plot	A graph illustrating time series data by dot, line, or needle plots with axes reflecting time and an outcome(s) of interest.
Distribution of time series data	A general summary of frequencies in time-referenced data—i.e., how often an outcome of interest reaches a certain level with respect to time units.
Distribution plot of time series data	Often illustrated with histograms and density plots.
Time series analyses	A collection of methods to describe, explain, and predict temporal processes with time-referenced data for an outcome of interest.
Trend	General temporal behavior in an outcome of interest that can exhibit steady incremental changes (linear) or varying incremental changes (non-linear) over time.
Season	An interval of time within one time cycle (typically one calendar year) defined by a specific biological, environmental, physical, physiological, or other property or feature in a biological or non-biological system [ref].
Seasonal pattern	A recurrence of periods in an outcome of interest with alternating values (e.g., high and low) over the course of a time cycle, commonly one calendar year.
Seasonality	A systematic periodic fluctuation in an outcome of interest over the course of one cycle (typically one calendar year) as an observable property of a biological or non-biological system.
Seasonal curve	An analytical representation of seasonal periodic fluctuations in an outcome of interest within one time cycle (typically one calendar year).
Seasonality features	A set of measurable characteristics to describe seasonality and a seasonal curve within one year, including seasonal peak, nadir, intensity, duration, speed at which a seasonal curve reaches its peak, and speed at which a seasonal curve declines to its nadir (Naumova, 2006).
Peak or nadir timing	A seasonality feature that represents times when a seasonal curve of an outcome reaches its maximum or minimum (Naumova, 2006).
Amplitude or intensity	A seasonality feature that represents the difference between seasonal peaks and nadirs (Naumova, 2006).
Duration	A seasonality feature that represents the time interval when incidence rises above a specified threshold (Naumova, 2006).

for Nutrition, 2016; HLPE, 2017) (**Fig. 3**). Food systems contribute to diet quality and nutrition outcomes by determining the food that is available to a consumer at any given place or time, including the costs and quality of the food. Consumers will choose different available foods based on their income, food prices, cultural traditions, and their preferences. Climatic patterns intersect with the economic forces of supply and demand by affecting food production and defining the cycles around which human eating patterns have evolved over hundreds of years. Social traditions have developed alongside climate patterns, which reinforce seasonal fluctuations in demand for foods around traditions like celebrations of agricultural cycle events (planting, rains, harvests) or solar and lunar cycle events (new or full moons, solstices, equinoxes).

Food supply, demand, and prices

Places where agriculture is the predominant economic activity tend to exhibit seasonal patterns in food supply, demand (due to seasonal incomes), and prices, affecting many people and permeating all sectors of the economy. The reasons are multifaceted, owing to the large proportion of gross domestic product (a measure of the size of the total economy) from agriculture (Devereux



Note: Example based on a single peak dryland rainfall climate and would need to be adopted for different contexts.

Fig. 2 Example seasonal curves of weight-dependent nutrition outcomes.

et al., 2013; Dillon and Barrett, 2017; Zanias, 1999) and poorly functioning markets. Market imperfections limit the opportunity to use trade to smooth supply and demand of goods and services (including labor) throughout the year, resulting in more fluctuations of supply, demand, and prices based on local conditions (Dillon and Barrett, 2017; Kaminski et al., 2016; Zanias, 1999).

In agriculturally-dominant economies diet quality is largely driven by the foods available at different points in the agricultural calendar, and the coincident fluctuations in food prices and incomes (Branca and D'Acapito, 2012; Devereux et al., 2013; Gilbert et al., 2017; Gill, 1992; Kaminski et al., 2016; Vaarst et al., 2018; Vaitla et al., 2009). Where markets are isolated, changes in supply can have immediate and lagged effects on local prices. Decreased supply results in higher prices and increased supply the reverse, a pattern that is most pronounced for goods with inelastic demand (necessities) that does not change much as prices oscillate. Where trade is costly or infeasible, local supply affects aggregate supply since compensation for shortages or surpluses is not possible. Further complications to food availability arise from insufficient infrastructure and technology to preserve and store goods, and lack of financial services for savings and credit to smooth incomes over time or make investments that create more stability throughout the year. The result is a transformation of the inherently neutral seasonal property of the climate system into seasonal patterns that are detrimental to wellbeing, health, and nutrition (Bidinger et al., 1986; Christian and Dillon, 2018; Feuerbacher et al., 2020; Fink et al., 2020; Gilbert et al., 2017; Gill, 1992; Chris Hillbruner and Egan, 2008; Kaminski et al., 2016; Sahn, 1989; Stelmach-Mardas et al., 2016; Vaitla et al., 2009; Zanias, 1999). For example, the commonly known "lean season" occurs when household food stocks and cash from the prior harvest are depleted while supply in markets is limited, leading to high prices and unaffordability among those with limited financial resources (Anderson et al., 2018; Gelli et al., 2017; Vaitla et al., 2009). Timing of the "lean season" can vary by livelihood specialization—for pastoralists, milk production is highest during the rainy season given improved access to pasture, but for those who rely on starchy staples, supplies are lowest during the rainy season. Beyond cereals and animal protein, fruits and vegetables are often only available certain times of the year, limiting access to essential

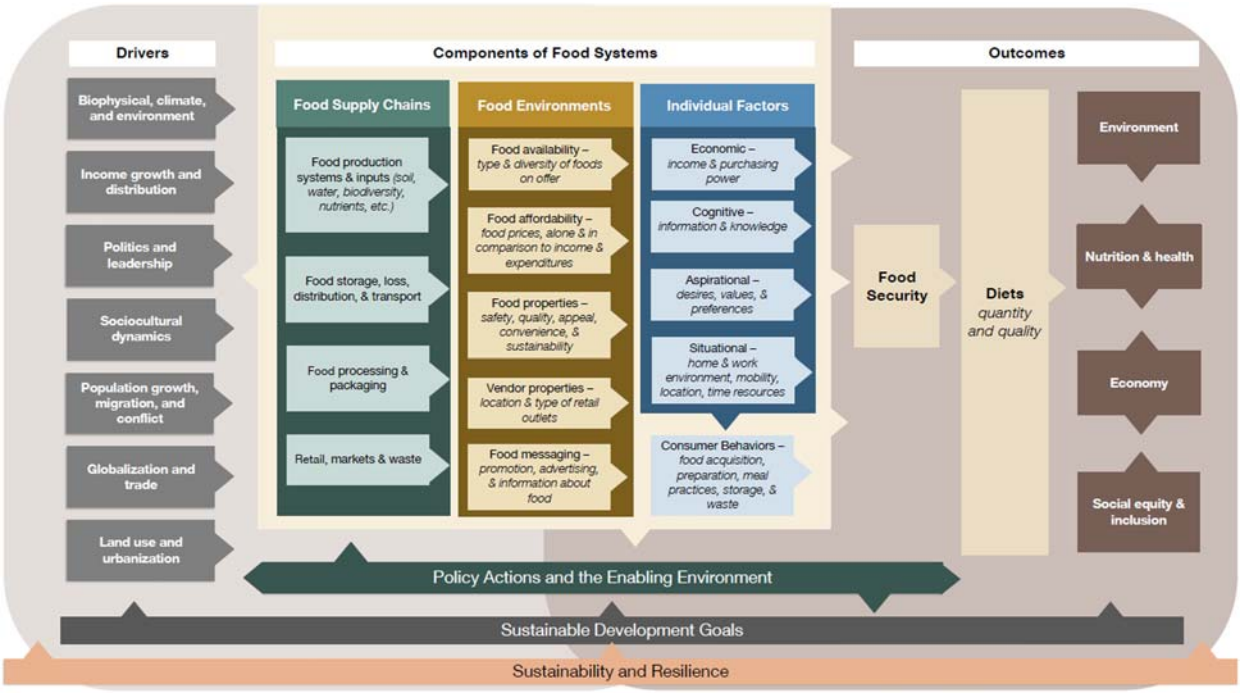


Fig. 3 Food systems framework: components, drivers, and outcomes. Reproduced with permission from Fanzo et al. (2021).

micronutrients (Abizari et al., 2017; Ambikapathi et al., 2021; Aparicio-Ugarriza et al., 2018; Bonis-Profumo et al., 2021; Crane et al., 2019; Jahns et al., 2016; van der Toorn et al., 2020).

In economies that do not depend on agriculture and where there are fewer transaction costs to trade, seasonal fluctuations in local production generally do not affect local supply. Stable supply and demand produce stable prices. The exception, however, is when an extreme weather event is large enough to impact total global production, in which case global and domestic prices can be affected, depending on policy responses. Trade restrictions (export or import bans) can have immediate and negative impacts on both world and domestic prices as they interfere with market competition but are often used by governments in times of crises to pursue domestic political stability (Pinstrup-Andersen, 2015).

Cultures and traditions

School calendars, holiday traditions, lifecycle traditions and celebrations, seasonal work patterns, and the cadence of social protection benefit cycles all intersect with climatic seasonality to affect food supply, incomes, and/or demand leading to impacts on nutrition (Beatty et al., 2019; Cotti et al., 2018; Graham et al., 2018; Kuhn, 2018; Laurito and Schwartz, 2019; Lerche, 2012). Religious practices often dictate certain foods be consumed (or abstained from) in relation to certain holidays, which affect patterns of demand. Examples include Ethiopian orthodox fasting (D'Haene et al., 2019), Ramadan, and Passover. Cultural traditions around holidays may also include traditions of relatively unbalanced diets during the Thanksgiving to New Year's period in the United States. Changes in eating patterns around holidays and special events can either be detrimental to nutritional status (e.g., increased added sugar) or can contribute positively to overall nutrition (e.g., access to animal-source foods that are lacking the rest of the year), depending on the context.

Recently, public health and environmental behavior change communications have encouraged "seasonal" eating patterns, meaning consuming mainly the products and especially fresh produce that are in season in one's local region. From the public health perspective, this is a way to encourage variety in diets and increase diet quality (WHO, 2018). From an environmental perspective, it encourages consuming foods that have not been flown around the world. However, while local food consumption may support seasonal eating patterns and local food systems, farmers, and economies, it sometimes comes with higher greenhouse gas emissions and water footprints than more efficient larger scale production and transportation methods, and may increase an area's reliance on exploitative labor practices (Coley et al., 2009; Edwards-Jones, 2010; Edwards-Jones et al., 2008; Enthoven and Van den Broeck, 2021). Technological innovation is essential to mitigate this current tradeoff (Barrett et al., 2020).

Livelihood systems

Livelihoods encompass all of the capabilities, assets, and activities necessary to obtain a means of living (Serrat, 2017) and are linked to food systems through labor, demand, and incomes. Fig. 4 presents a commonly used sustainable livelihoods framework. Capabilities are the power and freedom for a person to live the life they find most fulfilling (Sen, 2000). Assets include natural

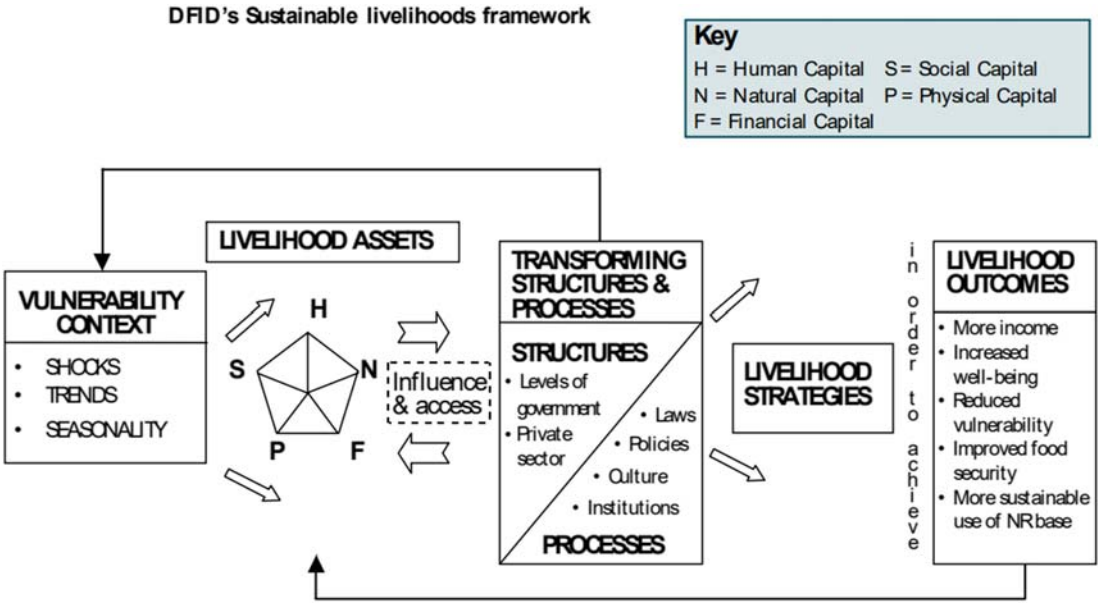


Fig. 4 Sustainable livelihoods framework. Reproduced with permissions from the Foreign, Commonwealth & Development Office.

resources (e.g., land, water), financial resources (e.g., income, savings, credit), physical assets (e.g., infrastructure, tools, technology, and livestock), social capital (e.g., relationships, networks), and human capital (e.g., health, knowledge, education). Activities encompass income-earning activities as well as those necessary to sustain life and households, including self-care and caregiving (Fernandez et al., 2016; Gottfried and Chun, 2018; Robinson and on behalf of CARE USA, 2010). The interactions between one's assets and experienced vulnerabilities are mediated by policies and institutions, together guiding livelihood strategy choices and contributing to nutrition, health, and poverty (DFID, 1999; Serrat, 2017).

Labor

All workers who provide goods and services where demand fluctuates with seasons may experience seasonal changes in their incomes, time and labor demand, and physical and mental health (Chaney and Torres, 2017; Devereux et al., 2013; Guidetti et al., 2020; Harriss-White, 2010; Mishra, 2020; Pattenden, 2016; Routh and Borghi, 2016). As energy expenditure fluctuates with seasonal variations in temperature and physical labor demand, availability and affordability of sufficient food and water to meet people's needs change in tandem. For smallholder farmers, increased labor demands of harvest and post-harvest processing often coincide with the lean season, as food and incomes from the prior harvest have run out (Bacon et al., 2014; Branca and D'Acapito, 2012; Devereux et al., 2013). The combined effects that this period can have on nutrition and health is especially important for women farmers who are pregnant or breastfeeding, due to intergenerational transmission of nutritional status (Bonis-Profumo et al., 2021; Vir, 2015). Not only does energy expenditure during pregnancy and lactation affect child growth, birth outcomes, and nutritional status, but also affects mothers' time available for caregiving (Higgins and Alderman, 1997; Johnston et al., 2018).

Seasonal labor also tracks patterns of tourism, as well as demand for food and other goods typically consumed at certain times of the year (e.g., holiday foods and gifts, "back-to-school" shopping, "wedding season"). Those without stable employment often work seasonal jobs and may migrate to follow work opportunities (Bozkurt, 2015; Chaney and Torres, 2017; Frank-Miller et al., 2015; Geremew and Gourio, 2018; Guidetti et al., 2020). Farmworkers living and working in high income countries are an especially vulnerable group, often migrating seasonally with children whose learning, health, and food security then follow the migration and work cycles (Arcury and Quandt, 2020; Castañeda et al., 2015; Kiehne and Mendoza, 2015; Lim et al., 2017; Quandt et al., 2016, 2021; Smith and Cuesta, 2020). In addition, hourly workers' schedules depend on factors such as customer volumes, holidays, or school calendars, which similarly affect incomes, time, and stress (Bozkurt, 2015). The consequences are fluctuating incomes and therefore ability to meet basic needs for health and nutrition, including quality diet, shelter, access to sanitary facilities, caregiving, and sleep (Hill et al., 2013).

Food system workers required to be outdoors (e.g., farmers, market vendors) may also be exposed to extreme heat conditions, which significantly increase the body's need for water. If clean water is scarce, dehydration and heat stress can occur rapidly. This is especially dangerous for those with chronic diseases such as diabetes (Kenny et al., 2010). In addition, these conditions reduce labor productivity, further exacerbating the burden of livelihoods based on manual or mobile labor (Cheuvront et al., 2010).

Environmental systems

Environmental systems encompass the physical surrounding and conditions in which people live and are directly influenced by climatic seasonality. Mutually reinforcing feedback loops within environmental systems interact to influence care practices, hygiene,

use and availability of clean water, pathogen survival, and food safety. These factors are in turn related to additional feedback loops between infectious disease susceptibility and diet quality, which ultimately influence nutrition outcomes.

Infectious disease and nutritional status

Morbidity due to infectious disease is both a driver and consequence of undernutrition among vulnerable populations, including children and pregnant/lactating women (Scrimshaw et al., 1968). Inflammation due to infection inhibits the anabolic processes that build organs and tissues, and increases the permeability of the small intestinal walls, leading to malabsorption of nutrients (Keusch et al., 2014; Millward, 2017). The body adapts by using the available nutrients to support immune function, prioritizing survival, and downregulating growth (Gluckman and Hanson, 2004). In addition to nutrient malabsorption, infection and inflammation also lead to loss of appetite and decreased dietary intake, which then lead to compromised immunity, increasing risk for infection in a vicious cycle (Egorov et al., 2010; Katona and Katona-Apte, 2008; Schaible and Kaufmann, 2007; Scrimshaw et al., 1968). Evidence that the association between acute malnutrition and mortality is context dependent, with higher levels of mortality from acute malnutrition in regions with highly prevalent infectious disease, points toward morbidity as the main mediating factor between acute malnutrition and mortality (Pelletier et al., 1995; Young et al., 2006). The burden of infectious disease is impacted both directly through conditions more favorable to disease transmission and through higher susceptibility to disease at certain times of the year (Naumova, 2006).

Pathogen survival

In low- and middle-income countries (many of which have agriculturally-dominant economies), the cyclic relationship between morbidity and undernutrition is characterized by seasonal fluctuations in intensity and magnitude resulting from the effects of agro-climatic seasonality on disease dynamics (Pascual and Dobson, 2005). At certain temperature, precipitation, and vegetation levels, infections spread more rapidly than at others. However, peaks in each of these climatic factors usually do not happen simultaneously, and vary across geographic regions (Marshak et al., 2021). It is thus useful to think about seasonality of infectious disease in relation to specific permutations of these climatic factors that combine to favor high transmission of infectious disease, either through their influence on the safety and hygiene of the environment, or through differential pathogen survival under various conditions.

For example, high ambient temperatures support pathogen survival, precipitation can increase water contamination with fecal matter, and food scarcity during some agricultural seasons can increase host susceptibility to disease and malnutrition. In places with high burdens of undernutrition, such as sub-Saharan Africa and Southeast Asia, it is widely accepted that many undernutrition-adjacent diseases spread most rapidly during periods of intense rain. Rainfall is associated with increased standing water, which propagates disease-transmitting pathogens, such as mosquito vectors that increase the burden of malaria and dengue fever (Bayoh et al., 2001; Dao et al., 2014; Roca-Feltrer et al., 2009), and with increased fecal contamination of drinking water with diarrheagenic bacterial pathogens (Dongzagla et al., 2021; Ercumen et al., 2017; Hossain et al., 2019; Kirby et al., 2016). Respiratory illnesses and pneumonia are also associated with undernutrition; however, while these diseases are more common during rainy seasons in tropical climates with high humidity (Li et al., 2019; Obando-Pacheco et al., 2018), there is substantial spatial variation in their seasonality, and they happen more frequently in cold or dry seasons in some contexts (Li et al., 2019).

High ambient temperature also plays an important role in nutrition outcomes, and in some contexts may be an even more important driver than precipitation (Chikhungu and Madise, 2014; Cliffer et al., 2021; Fentahun et al., 2018; Marshak, 2021). Pathogens replicate more efficiently and survive longer in high temperatures (Azage et al., 2017; Hashizume et al., 2007; Islam et al., 2009; Vargas et al., 2004). The highest risk period for diarrheagenic *E. coli*, for example, is during the pre-rainy season when temperatures are highest, and rains are sparse (Azage et al., 2017; Vargas et al., 2004). The lack of sufficient clean water during the hot, dry season, when precipitation is low but not zero, may also lead to consumption of unsafe water and require reduced hygiene practices that leave people vulnerable to infection (Fewtrell et al., 2005; Jagai et al., 2012; Kulinkina et al., 2016a).

Disease transmission dynamics may also be modified by geological factors. For example, in dryer regions of India, cholera transmission has been found to peak during the monsoon season; however, in the estuarine regions surrounded by wetlands, cholera was at its lowest point during the monsoons (Pascual et al., 2002). Understanding how specific combinations of peaks and nadirs in the magnitude of climatic factors affect pathogen survival and disease transmission risk, and how these factors vary both temporally and spatially, is important in interrupting feedback loops between infectious disease and nutrition outcomes.

Feeding and caregiving

Host susceptibility to disease fluctuates in relation to food intake and security, time for caregiving and breastfeeding, and access to and use of health services. Studies across multiple low- and middle-income country contexts have shown that people have higher food insecurity (Bangladesh, Kenya, The Gambia) (Chris Hillbruner and Egan, 2008; Nabwera et al., 2017; Shell-Duncan, 1995), and fewer resources for caregiving (The Gambia, Nepal) (Nabwera et al., 2017; Panter-Brick, 1997), including breastfeeding time (India) (Das et al., 2016), during periods of high precipitation, since these periods coincide with the agricultural planting season. In addition, for pastoralist communities, undernutrition peaks during the dry season when access to animal milk is lowest (Sadler et al., 2009). In the past, the idea that undernutrition was primarily driven by lack of inadequate food consumption, known as the "food-first bias" dominated the literature, and led programming and policy efforts to prioritize food-related interventions (Pelletier et al., 1995). However, several studies have linked the relationship between decreased food intake/less caregiving time and increased

vulnerability to growth faltering during the rainy season to the feedback loops and synergistic effects between inadequate diet and higher disease incidence (Bohler et al., 1995; Ercumen et al., 2017; Madan et al., 2018; Nabwera et al., 2017).

Food safety

Diseases that threaten food safety, particularly mycotoxins, also follow seasonal patterns in tropical settings with inadequate control measures. Mycotoxins such as aflatoxins, fumonisins, and deoxyvalinol, are toxins produced by fungi known to be carcinogenic and harmful to human health even in small amounts (Adeyeye, 2020; Marasas et al., 2008). The molds that produce these toxins proliferate in hot, humid climates characteristic of the rainy season in tropical areas (Adeyeye, 2020; Fraeyman et al., 2017; Gong et al., 2016; Marasas et al., 2008; Smith et al., 2016). Outbreaks of acute aflatoxin poisoning (aflatoxicosis) have occurred in recent years, with the most studied being the Kenya outbreak of 2004, and affect both human and animal health (Groopman and Kensler, 2005; Rumbeiha and Oehme, 2005). Long-term exposure, such as chronic seasonal exposure, causes liver cancer and is a suspected contributor to low birth weights and child stunting (Coppock and Dziwenka, 2014; Gong et al., 2016; Lauer et al., 2019). Other animal diseases also follow seasonal patterns, which can increase animal morbidity and mortality, reducing the income and food-source opportunity of livestock keeping (van Dijk et al., 2010; Nwanta et al., 2008; Thornton, 2010).

Access to water

Water access follows seasonal patterns related to rainfall and temperature. Lack of water can increase exposure to disease, reduce livestock health, and increase labor burdens. During the dry season, people may access poorer quality water sources, or walk long distances to get water (Brewis et al., 2020; WHO and UNICEF, 2021). This reduces the total water availability for drinking, cooking, cleaning, and personal hygiene and propagates water-borne diseases. However, the start of the rains can also bring increased pathogenic contamination as fecal matter and other contaminants get washed into seasonal and permanent water sources (Kulinkina et al., 2016b). The water access pathway thus influences nutrition via disease prevalence, reduced drinking water access, and increased labor burden in settings where households do not have access to potable water piped directly to their homes. Even where piped water infrastructure exists, increasing seasonal droughts are threatening water for households and agriculture (Ahopelto et al., 2019; Lindqvist et al., 2021; Mekonnen and Hoekstra, 2016).

Climate change

Global climate changes will continue to exacerbate many of the agroclimatic drivers of poor nutritional status through food system, livelihood, and disease pathways (Clark et al., 2020; Crippa et al., 2021; Davis et al., 2016; Myers et al., 2017; Rosenzweig et al., 2020; Springmann et al., 2018; Willett et al., 2019). Food systems are estimated to be responsible for approximately 30% of global greenhouse gas emissions. Greenhouse gas emissions worsen climate change, which is anticipated to further destabilize climatic patterns and increase weather extremes (Tubiello et al., 2021). Extreme weather negatively impacts agricultural production, water availability, disease prevalence (food-borne, water-borne, human, and animal), interrupts transportation networks, changes labor and time demands, and reduces incomes, all of which stand to threaten food security, diet quality, and health (Park et al., 2019; Springmann et al., 2016).

Seasonality of nutritional outcomes

The complex interactions of seasonal fluctuations in the feedback loops within food, livelihood, and environmental systems culminates in observed seasonality of nutrition outcomes.

Growth, anthropometry, and body composition

Seasonal changes in nutrition and anthropometry have been documented since as far back as the late 18th century when Le Comte de Montbeillard of France measured his son's height every six months for 18 years and found that the boy grew twice as much in the summer as in the winter (Tanner, 1981). Subsequent studies in similar contexts with wide temperature variations found that child height growth was largest in the warmer spring months, and weight gain largest in autumn (Orr and Clark, 1930; Pollitt and Arthur, 1989). In more recent years, data from high income contexts have revealed more variability in the seasonality of weight gain, showing fluctuations within seasons and even weeks, with weekends and holidays characterized by higher weight gain (Turicchi et al., 2020). Seasonal anthropometry patterns in high income contexts can also be modified by major disasters or life events. A recent example is the Covid-19 pandemic; children's body mass index (BMI) increased in South Korea in spring 2020, as opposed to the expected decrease (Han et al., 2021).

Growth seasonality in low- and middle-income countries started receiving attention midway through the twentieth century when it became clear that identifying peak timing of poor nutrition outcomes was critical for more effective interventions and policies. The consensus from most studies conducted since then is that nutritional outcomes in low- and middle-income countries are worst during the rainy/monsoon/lean/pre-harvest period of the year (Abay and Hirvonen, 2017; Benefice et al., 1984; Bohler et al., 1995; Grellety et al., 2013; Chris Hillbruner and Egan, 2008; Miller et al., n.d.; Nabwera et al., 2017; Panter-Brick, 1997; Roba et al., 2016; Schwinger et al., 2014; Tomkins et al., 1986; Wright et al., 2001) as opposed to the harvest period or dry season, though

a handful of studies indicate worse weight-related anthropometric outcomes during the dry season (Bechir et al., 2010; Chikhungu and Madise, 2014; Chotard et al., 2010; Egata et al., 2013; Kinyoki et al., 2016; Loutan and Lamotte, 1984; Nielsen et al., 2011; Sellen, 2000; Trowbridge and Newton, 1979).

Recent research out of the African Sahel indicates that two seasonal peaks of undernutrition may be present within the same year. Primary research from Chad (Marshak, 2021) and Burkina Faso (Cliffer et al., 2021) among non-nomadic households (practicing both farming and livestock ownership) reveals two peaks of child undernutrition. A larger peak occurs during the hot dry season, right as sporadic rains begin, followed by an improvement during the rainy season, and a smaller peak as the rains taper off and temperatures rise again right before the harvest (as shown in **Fig. 2** from the measuring seasonality section). Secondary data analysis across all dryland unimodal rainfall contexts in Africa's Sahel confirms this pattern (Venkat et al., 2023). These studies demonstrate that existing assumptions around seasonal patterns of nutrition outcomes require further investigation and underscore the notion that assuming a single pre-harvest/lean season peak is insufficient. In addition, they show the importance of considering combinations of climatic factors rather than thinking of seasonality as a function of discrete categories and support an important role for ambient temperature in child growth.

While most seasonality research focuses on weight-related anthropometric indicators that show short-term change, seasonality has also been observed in length attainment, length growth, and stunting. In Mali, Malawi, and Burkina Faso, the rainy season was associated with higher HAZ, lower prevalence of stunting, and higher length velocity, respectively (Adams, 1994; Chikhungu and Madise, 2014; Cliffer et al., 2021). Moreover, a four-decade cohort analysis from The Gambia indicates a time lag of three months between the seasonality of attained weight vs. height indicators (Schoenbuchner et al., 2019). Variation has been observed in the seasonality of different nutrition indicators, even within the same study. For example, a study in Ethiopia found diverging seasonal trends depending on the indicator used, with stunting and underweight peaking in the pre-harvest period and wasting peaking in the post-harvest period (Roba et al., 2016).

Further, looking beyond weight anthropometry at body composition measures shows that not all growth or weight gain is the same. Among a population of adults in the Netherlands, fat mass showed a maximal value in the winter and a minimal value in the summer, while fat-free mass showed no significant seasonal changes. The authors attribute these differences directly to variability in conditions conducive for physical activity (Westerterp, 2020). Similar findings have been found across multiple western contexts among adults (Haggarty et al., 1994; Plasqui et al., 2003) as well as children (Dalskov et al., 2016). Thus, different seasons may be associated with different physiological changes related to weight, height, and body composition.

Micronutrient status and deficiencies

Micronutrient consumption and deficiencies have their own seasonal patterns due to variations in climate, agricultural activities, food availability, access to fortification and supplementation, and disease. A study of women of childbearing age in China showed lower concentrations of folate and hemoglobin in the summer, leading to a higher prevalence of moderate anemia, while concentrations of vitamin B-6 were lower in winter and spring (Ronnenberg et al., 2000). Among pregnant woman in Nepal, micronutrient status was best during winter months when vegetables and fruit are abundant and more affordable. Serum-carotene and vitamin B-6 peaked immediately following the timing of the mango and banana harvest. Vitamin D concentrations, on the other hand peaked during the monsoon and hot summer months, presumably as a response to increased exposure to sunlight (Jiang et al., 2005). However, seasonal differences for certain contexts and populations can follow completely different patterns divorced from the agricultural cycle. Iron status among Japanese collegiate elite female rhythmic gymnasts was lowest during the preseason's weight loss period, due to lower protein intake (Kokubo et al., 2016). Thus, while there is a pattern of improved consumption of key micronutrients corresponding to the harvest, it is not unilateral across contexts or micronutrients and requires a keen understanding of livelihood activities, access to healthcare and nutrition programs, and dietary patterns.

Variation in seasonal patterns of nutritional outcomes

Seasonal patterns in nutrition outcomes can change over time and can vary depending on the specific sub-population or geographic location.

Sub-population variation

One modifier of seasonal nutrition outcomes is age. In Senegal and Ethiopia, seasonal differences in body weight and weight-for-height z-scores (WHZ) were greater among adults compared to adolescents and school aged children respectively (Benefice et al., 1984; Ferro-Luzzi et al., 2002). Seasonal changes in body composition can also vary by age group. In The Gambia, younger adolescent girls gained weight in the form of lean mass across all seasons, while older adolescents lost lean mass during the lean season, likely due to the importance of adipose tissue for reproduction (Reiches et al., 2013). Among children under five, younger children show greater seasonal variability in their nutrition outcomes (Maleta et al., 2003; Simondon et al., 1993) corresponding to changes in breastfeeding, complementary feeding, and the child's immune system (Maleta et al., 2003).

Sex and gendered beliefs and practices also influence nutrition outcomes. In Senegal and Ethiopia, seasonal undernutrition was greater in adult men than women, attributed to sex-differentiated livelihoods (Ferro-Luzzi et al., 2002; Simondon et al., 1993). In Mali, The Gambia, and Chad, boys under the age of five showed greater seasonal variation in nutrition outcomes compared to girls (Adams, 1994; Marshak, 2021; Schoenbuchner et al., 2019). These trends correspond to differences found by sex and gender across

a host of health outcomes with males generally more vulnerable to mortality and morbidity (Elsmén et al., 2004). Biological factors also intertwine with social systems that may lead households to target certain foods to different household members based on (perceived) needs or biases, compensating for or exacerbating biological vulnerability (Berti, n.d.).

Distinctions have also historically been made between livelihood specializations, with farming communities more vulnerable to malnutrition in the rainy season and pastoralist communities more vulnerable in the dry season. However, this distinction may not always hold true, given that in many contexts there has been a transition into more mixed livelihood systems. A study in Chad found that acute malnutrition peaked in the dry season for children living in both nomadic and sedentary communities (Bechir et al., 2010). In the Horn of Africa, seasonal fluctuations in wasting were higher among children living in areas that predominately practices pastoralism as opposed to farming, but overall seasonal patterns were the same (Chotard et al., 2010). Moreover, several studies have revealed seasonal patterns that directly contradict previous historical assumptions. Among farming communities, worse nutrition outcomes have been found in the dry season as opposed to the rainy or pre-harvest season (Chikhungu and Madise, 2014; Chotard et al., 2010; Egata et al., 2013; Trowbridge and Newton, 1979), while a study in Senegal among nomadic households, found worse nutrition outcomes in the rainy season (Benefice et al., 1984).

Spatial-temporal variation

Lastly, seasonal patterns vary over time and space. In a 2-year study in Senegal the authors found that peaks and nadirs in weight variation changed from one year to another, and differed for communities that had access to irrigation, which dampened seasonal variability in weight (Simondon et al., 1993). Thirty-six years of data from The Gambia shows how seasonality of multiple anthropometric outcomes decreased over time, as investments in health, development, and nutrition programming increased (Nabwera et al., 2017). Seasonality effects are also less pronounced in regions with bimodal vs. unimodal rainfall patterns, given the shorter gap between harvesting and the likelihood that at least one harvest will be successful (Schofield, 1974). Given the effects of climate change and general development, seasonal patterns cannot be seen as static from year to year or place to place.

Analytic decisions and observed nutrition seasonality

Data collection and analysis methods affect whether and what seasonal patterns are identified. The spatiotemporal resolution and precision of collected records have significant implications on the ability to detect, define, and predict seasonal patterns in nutrition outcomes. A recent review of multiple seasonality studies in unimodal rainfall dryland contexts in Africa found that most studies treat time as a categorical variable, frequently splitting the analysis into wet vs. dry or pre-vs. post-harvest comparisons (Marshak et al., 2021). While this approach can simplify interpretation, it does not capture climatic variability and how it relates to human activities, nor does it account for the transitional periods between major seasons (i.e., sporadic rains during the dry season before the heavy rains arrive). Weight gain research in developed country contexts is one example of how this affects nutrition seasonality findings. Much of this research indicates that weight gain is highest in autumn/winter (Ma et al., 2006; Orr and Clark, 1930). However, more precise analysis using more granular time measurements specifically pinpoints that weight gain to the holiday seasons (Turicchi et al., 2020)—a distinction that allows for better informed programming and messaging. Similarly, divergent findings on seasonality of nutrition outcomes from the most recent studies in low-income contexts, including the discovery of two seasonal peaks of undernutrition as well as higher rates of malnutrition during the hot, dry season than the rainy season demonstrate the importance of modeling nutrition outcomes with precision (Cliffer et al., 2021; Marshak, 2021).

Our ability to link seasonal patterns in nutrition to changes in climate also depend on interdisciplinary collaboration, data sharing and exchange. Evidence is conclusive that human activity has already drastically altered the climate. Deforestation, urbanization, and industrialization have contributed to global surface temperatures being 1.09 °C higher in 2011–2020 than 1850–1900, with warmer temperatures and rising sea levels imminent. Climate change is already increasing the variability of established climatic phenomena, shifting climatological boundaries, and increasing the frequency and severity of extreme weather events (Mbow, 2019). Increasing temperatures and rising seas are expected to negatively impact agricultural production and reduce the amount, quality, and stability of crop yields. As climate becomes more variable, analytical methods for seasonality must become more adaptable and accommodate minimal pre-determination for greatest accuracy (FAO, 2020; Phalkey et al., 2015; Rohli and Vega, 2017).

Consideration of study design, analyses, research focus, terminology, interpretability and interoperability of data and models is important in determining, understanding, and predicting seasonal patterns and their interplay with climatic factors, demographic factors, livelihoods, and other potential mitigating or exacerbating correlates. Advances in statistical modeling software, including widespread availability of open-source software, has made using computationally sophisticated models much more accessible than they used to be. Improving our understanding of nutrition seasonality requires leaving behind error-prone computations of seasonal outcomes based on categorical variables and embracing methods and models that reflect the cyclical nature of the time series and allow for precise and accurate findings.

Conclusion

Seasonality is a natural phenomenon. It is not inherently bad; however, without knowledge and precise understanding of seasonal fluctuations in the components of the complex systems that interact to produce seasonal nutrition outcomes, humans may struggle

to achieve well-being. Developing such precise understanding requires models that accurately describe and measure seasonality of nutrition outcomes, as well as unified terminology to describe seasonality. In addition, governance and policymaking within the complex systems involved in nutrition seasonality can go a long way in improving nutrition when considering and understanding the seasonal timing of drivers and outcomes.

Knowing when anthropometric and micronutrient deficiencies are likely to occur across different populations is the first step in helping governments and organizations implement the appropriate interventions at the correct time. Understanding the seasonality in the underlying drivers of those outcomes elucidates entry points for policy and programs. Direct policy leverage points to permanently increase seasonal stability of food supply, demand, and prices include appropriate trade policies, domestic stocks to address supply shocks, and adequate safety nets that smooth access to food when incomes decline and/or prices rise. Investments in infrastructure and agricultural productivity can increase supply resilience, strengthen markets, and provide jobs, increasing the quantity and quality of production, distribution, and consumption year-round. Further infrastructural improvements to water, sanitation, hygiene, and food storage can help reduce the seasonal patterns of infectious disease transmission and susceptibility.

As climate change continues to affect food supply and infectious disease transmission, not only will we need to adapt our nutrition seasonality models, but food, livelihood, and environmental systems transformation is essential to protect the nutrition of current and future generations.

References

Abay, K., Hirvonen, K., 2017. Does market access mitigate the impact of seasonality on child growth? Panel data evidence from northern Ethiopia. J. Dev. Stud. 53, 1414–1429. https://doi.org/10.1080/00220388.2016.1251586.

Abizari, A.-R., Azupogo, F., Nagasu, M., Creemers, N., Brouwer, I.D., 2017. Seasonality affects dietary diversity of school-age children in northern Ghana. PLoS One 12, e0183206. https://doi.org/10.1371/journal.pone.0183206.

Adams, A.M., 1994. Seasonal variations in nutritional risk among children in central Mali. Ecol. Food Nutr. 33, 93–106. https://doi.org/10.1080/03670244.1994.9991418.

Adeyeye, S.A.O., 2020. Aflatoxigenic fungi and mycotoxins in food: a review. Crit. Rev. Food Sci. Nutr. 60, 709–721. https://doi.org/10.1080/10408398.2018.1548429.

Ahopelto, L., Veijalainen, N., Guillaume, J.H.A., Keskinen, M., Marttunen, M., Varis, O., 2019. Can there be water scarcity with abundance of water? Analyzing water stress during a severe drought in Finland. Sustainability 11, 1548. https://doi.org/10.3390/su11061548.

Alarcon Falconi, T.M., Estrella, B., Sempértegui, F., Naumova, E.N., 2020. Effects of data aggregation on time series analysis of seasonal infections. Int. J. Environ. Res. Publ. Health 17, E5887. https://doi.org/10.3390/ijerph17165887.

Ambikapathi, R., Kosek, M.N., Lee, G.O., Olortegui, M.P., Zaitchik, B., Yori, P.P., Bauck, A., Caulfield, L.E., 2021. La Niña weather impacts dietary patterns and dietary diversity among children in the Peruvian Amazon. Publ. Health Nutr. 24, 3477–3487. https://doi.org/10.1017/S1368980020003705.

Anderson, C.L., Reynolds, T.W., Merfeld, J.D., Biscaye, P., 2018. Relating seasonal hunger and prevention and coping strategies: a panel analysis of Malawian farm households. J. Dev. Stud. 54, 1737–1755. https://doi.org/10.1080/00220388.2017.1371296.

Aparicio-Ugarriza, R., Rumi, C., Luzardo-Socorro, R., Mielgo-Ayuso, J., Palacios, G., Bibiloni, M.M., Julibert, A., Argelich, E., Tur, J.A., González-Gross, M., 2018. Seasonal variation and diet quality among Spanish people aged over 55 years. J. Physiol. Biochem. 74, 179–188. https://doi.org/10.1007/s13105-017-0599-4.

Arcury, T.A., Quandt, S.A., 2020. Latinx Farmworkers in the Eastern United States: Health, Safety, and Justice. Springer Nature.

Azage, M., Kumie, A., Worku, A., Bagtzoglou, A.C., Anagnostou, E., 2017. Effect of climatic variability on childhood diarrhea and its high risk periods in northwestern parts of Ethiopia. PLoS One. https://doi.org/10.1371/journal.pone.0186933.

Bacon, C.M., Sundstrom, W.A., Flores Gómez, M.E., Ernesto Méndez, V., Santos, R., Goldoftas, B., Dougherty, I., 2014. Explaining the "hungry farmer paradox": smallholders and fair trade cooperatives navigate seasonality and change in Nicaragua's corn and coffee markets. Glob. Environ. Change 25, 133–149. https://doi.org/10.1016/j.gloenvcha.2014.02.005.

Barrett, C.B., Benton, T.G., Cooper, K.A., Fanzo, J., Gandhi, R., Herrero, M., James, S., Kahn, M., Mason-D'Croz, D., Mathys, A., Nelson, R.J., Shen, J., Thornton, P., Bageant, E., Fan, S., Mude, A.G., Sibanda, L.M., Wood, S., 2020. Bundling innovations to transform agri-food systems. Nat. Sustain. 3, 974–976. https://doi.org/10.1038/s41893-020-00661-8.

Bayoh, M.N., Thomas, C.J., Lindsay, S.W., 2001. Mapping distributions of chromosomal forms of *Anopheles gambiae* in West Africa using climate data. Med. Vet. Entomol. 15, 267–274. https://doi.org/10.1046/j.0269-283x.2001.00298.x.

Beatty, T.K.M., Bitler, M.P., Cheng, X.H., van der Werf, C., 2019. SNAP and paycheck cycles. South. Econ. J. 86, 18–48. https://doi.org/10.1002/soej.12345.

Bechir, M., Schelling, E., Bonfoh, B., Seydi, M., Wade, S., Moto, D.D., Tanner, M., Zinsstag, J., 2010. Seasonal variations in the nutritional status of nomad and sedentary children less than 5 years of age living in the Sahel in Chad. Med. Trop. 70, 353–358.

Benefice, E., Chevassus-Agnes, S., Barral, H., 1984. Nutritional situation and seasonal variations for pastoralist populations of the Sahel (Senegalese Ferlo). Ecol. Food Nutr. 14, 229–247. https://doi.org/10.1080/03670244.1984.9990790.

Berti, P.R., (n.d.). Intrahousehold distribution of food: a review of the literature and discussion of the implications for food fortification programs. Food Nutr. Bull. 33, S163–S169. https://doi.org/10.1177/15648265120333S204.

Bidinger, P.D., Nag, B., Babu, P., 1986. Nutritional and health consequences of seasonal fluctuations in household food availability. Food Nutr. Bull. 8, 36–60. https://doi.org/10.1177/156482658600800109.

Bohler, E., Aalen, O., Bergstrom, S., Halvorsen, S., 1995. Breast-feeding and seasonal determinants of child growth in weight in East-Bhutan. Acta Pædiatr. 84, 1029–1034. https://doi.org/10.1111/j.1651-2227.1995.tb13820.x.

Bonis-Profumo, G., Stacey, N., Brimblecombe, J., 2021. Maternal diets matter for children's dietary quality: seasonal dietary diversity and animal-source foods consumption in rural Timor-Leste. Matern. Child Nutr. 17, e13071. https://doi.org/10.1111/mcn.13071.

Bordt, M., 2015. Spatial Units, Scaling and Aggregation. United Nations Statistics Division, United Nations Environment Programme.

Bozkurt, Ö., 2015. The punctuation of mundane jobs with extreme work: Christmas at the supermarket deli counter. Organization 22, 476–492. https://doi.org/10.1177/1350508415572512.

Branca, F., D'Acapito, P., 2012. Seasonality. In: Encyclopedia of Human Nutrition, vol. 4–4, pp. 178–185. https://doi.org/10.1016/B978-0-12-375083-9.00247-6.

Brewis, A., Workman, C., Wutich, A., Jepson, W., Young, S., 2020. Household water insecurity is strongly associated with food insecurity: evidence from 27 sites in low- and middle-income countries. Am. J. Hum. Biol. 32, e23309. https://doi.org/10.1002/ajhb.23309.

Buzzelli, M., 2020. Modifiable areal unit problem. In: International Encyclopedia of Human Geography. Elsevier, pp. 169–173. https://doi.org/10.1016/B978-0-08-102295-5.10406-8.

Castañeda, S.F., Rosenbaum, R.P., Holscher, J.T., Madanat, H., Talavera, G.A., 2015. Cardiovascular disease risk factors among Latino migrant and seasonal farmworkers. J. Agromed. 20, 95–104. https://doi.org/10.1080/1059924X.2015.1010060.
Chambers, R., 2009. Seasonal poverty: integrated, overlooked and therefore opportunity. In: Presented at the International Conference of the Institute of Development Studies 2009, London, UK.
Chambers, R., Longhurst, R., Pacey, A. (Eds.), 1981. Seasonal Dimensions to Rural Poverty. Frances Pinter Publishers Ltd, Totowa, NJ.
Chaney, B.H., Torres, E., 2017. Covariates of identified stress and depression among seasonal farmworkers. Int. J. Environ. Res. Publ. Health 14, 711. https://doi.org/10.3390/ijerph14070711.
Cheuvront, S.N., Kenefick, R.W., Montain, S.J., Sawka, M.N., 2010. Mechanisms of aerobic performance impairment with heat stress and dehydration. J. Appl. Physiol. 109, 1989–1995. https://doi.org/10.1152/japplphysiol.00367.2010.
Chikhungu, L.C., Madise, N.J., 2014. Seasonal variation of child under nutrition in Malawi: is seasonal food availability an important factor? Findings from a national level cross-sectional study. BMC Publ. Health 14. https://doi.org/10.1186/1471-2458-14-1146.
Chotard, S., Mason, J.B., Oliphant, N.P., Mebrahtu, S., Hailey, P., 2010. Fluctuations in wasting in vulnerable child populations in the Greater Horn of Africa. Food Nutr. Bull. 31, S219–S233. https://doi.org/10.1177/15648265100313S302.
Christian, P., Dillon, B., 2018. Growing and learning when consumption is seasonal: long-term evidence from Tanzania. Demography 55, 1091–1118. https://doi.org/10.1007/s13524-018-0669-4.
Clark, M.A., Domingo, N.G.G., Colgan, K., Thakrar, S.K., Tilman, D., Lynch, J., Azevedo, I.L., Hill, J.D., 2020. Global food system emissions could preclude achieving the 1.5° and 2°C climate change targets. Science 370, 705–708. https://doi.org/10.1126/science.aba7357.
Cliffer, I.R., Naumova, E.N., Masters, W.A., Perumal, N., Rogers, B.L., 2021. Peak Timing of Growth Faltering Among Young Children Coincides With Highest Ambient Temperatures in Burkina Faso. TBD.
Coley, D., Howard, M., Winter, M., 2009. Local food, food miles and carbon emissions: a comparison of farm shop and mass distribution approaches. Food Pol. 34, 150–155. https://doi.org/10.1016/j.foodpol.2008.11.001.
Coppock, R.W., Dziwenka, M., 2014. Chapter 32—mycotoxins. In: Gupta, R.C. (Ed.), Biomarkers in Toxicology. Academic Press, Boston, pp. 549–562. https://doi.org/10.1016/B978-0-12-404630-6.00032-4.
Cotti, C., Gordanier, J., Ozturk, O., 2018. When does it count? The timing of food stamp receipt and educational performance. Econ. Educ. Rev. 66, 40–50. https://doi.org/10.1016/j.econedurev.2018.06.007.
Crane, T.E., Latif, Y.A., Wertheim, B.C., Kohler, L.N., Garcia, D.O., Rhee, J.J., Seguin, R., Kazlauskaite, R., Shikany, J.M., Thomson, C.A., 2019. Does season of reported dietary intake influence diet quality? Analysis from the women's health initiative. Am. J. Epidemiol. 188, 1304–1310. https://doi.org/10.1093/aje/kwz087.
Crippa, M., Solazzo, E., Guizzardi, D., Monforti-Ferrario, F., Tubiello, F.N., Leip, A., 2021. Food systems are responsible for a third of global anthropogenic GHG emissions. Nat. Food 2, 198–209. https://doi.org/10.1038/s43016-021-00225-9.
Dalskov, S.-M., Ritz, C., Larnkjær, A., Damsgaard, C.T., Petersen, R.A., Sørensen, L.B., Hjorth, M.F., Ong, K.K., Astrup, A., Mølgaard, C., Michaelsen, K.F., 2016. Seasonal variations in growth and body composition of 8-11-y-old Danish children. Pediatr. Res. 79, 358–363. https://doi.org/10.1038/pr.2015.206.
Dao, A., Yaro, A.S., Diallo, M., Timbine, S., Huestis, D.L., Kassogue, Y., Traore, A.I., Sanogo, Z.L., Samake, D., Lehmann, T., 2014. Signatures of aestivation and migration in Sahelian malaria mosquito populations. Nature 516, 387–390. https://doi.org/10.1038/nature13987.
Das, A., Chatterjee, R., Karthick, M., Mahapatra, T., Chaudhuri, I., 2016. The influence of seasonality and community-based health worker provided counselling on exclusive breastfeeding -findings from a cross-sectional survey in India. PLoS One. https://doi.org/10.1371/journal.pone.0161186.
Davis, K.F., Gephart, J.A., Emery, K.A., Leach, A.M., Galloway, J.N., D'Odorico, P., 2016. Meeting future food demand with current agricultural resources. Glob. Environ. Change 39, 125–132. https://doi.org/10.1016/j.gloenvcha.2016.05.004.
Devereux, S., Sabates-Wheeler, R., Longhurst, R. (Eds.), 2013. Seasonality, Rural Livelihoods and Development, Seasonality Rural Livelihoods and Development. https://doi.org/10.4324/9780203139820.
DFID, 1999. DFID Sustainable Livelihoods Guidance Sheet.
D'Haene, E., Desiere, S., D'Haese, M., Verbeke, W., Schoors, K., 2019. Religion, food choices, and demand seasonality: evidence from the Ethiopian milk market. Foods 8, 167. https://doi.org/10.3390/foods8050167.
van Dijk, J., Sargison, N.D., Kenyon, F., Skuce, P.J., 2010. Climate change and infectious disease: helminthological challenges to farmed ruminants in temperate regions. Animal 4, 377–392. https://doi.org/10.1017/S1751731109990991.
Dillon, B., Barrett, C.B., 2017. Agricultural factor markets in Sub-Saharan Africa: an updated view with formal tests for market failure. Food Pol. 67, 64–77. https://doi.org/10.1016/j.foodpol.2016.09.015. Agriculture in Africa—Telling Myths from Facts.
Dongzagla, A., Jewitt, S., O'Hara, S., 2021. Seasonality in faecal contamination of drinking water sources in the Jirapa and Kassena-Nankana Municipalities of Ghana. Sci. Total Environ. 752, 141846. https://doi.org/10.1016/j.scitotenv.2020.141846.
Edwards-Jones, G., 2010. Does eating local food reduce the environmental impact of food production and enhance consumer health? Proc. Nutr. Soc. 69, 582–591. https://doi.org/10.1017/S0029665110002004.
Edwards-Jones, G., Milà i Canals, L., Hounsome, N., Truninger, M., Koerber, G., Hounsome, B., Cross, P., York, E.H., Hospido, A., Plassmann, K., Harris, I.M., Edwards, R.T., Day, G.A.S., Tomos, A.D., Cowell, S.J., Jones, D.L., 2008. Testing the assertion that "local food is best": the challenges of an evidence-based approach. Trends Food Sci. Technol. 19, 265–274. https://doi.org/10.1016/j.tifs.2008.01.008. Towards Sustainable Food Chains: Harnessing the Social and Natural Sciences.
Egata, G., Berhane, Y., Worku, A., 2013. Seasonal variation in the prevalence of acute undernutrition among children under five years of age in east rural Ethiopia: a longitudinal study. BMC Publ. Health 13, 864. https://doi.org/10.1186/1471-2458-13-864.
Egorov, A.I., Sempértegui, F., Estrella, B., Egas, J., Naumova, E.N., Griffiths, J.K., 2010. The effect of *Helicobacter pylori* infection on growth velocity in young children from poor urban communities in Ecuador. Int. J. Infect. Dis. 14, e788–e791. https://doi.org/10.1016/j.ijid.2010.03.013.
Elsmén, E., Hansen Pupp, I., Hellström-Westas, L., 2004. Preterm male infants need more initial respiratory and circulatory support than female infants. Acta Paediatr. 93, 529–533. https://doi.org/10.1080/08035250410024998.
Enthoven, L., Van den Broeck, G., 2021. Local food systems: reviewing two decades of research. Agric. Syst. 193, 103226. https://doi.org/10.1016/j.agsy.2021.103226.
Ercumen, A., Pickering, A.J., Kwong, L.H., Arnold, B.F., Parvez, S.M., Alam, M., Sen, D., Islam, S., Kullmann, C., Chase, C., Ahmed, R., Unicomb, L., Luby, S.P., Colford, J.M., 2017. Animal feces contribute to domestic fecal contamination: evidence from *E. coli* measured in water, hands, food, flies, and soil in Bangladesh. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.7b01710.
Fanzo, J., Haddad, L., McLaren, R., Marshall, Q., Davis, C., Herforth, A., Jones, A., Beal, T., Tschirley, D., Bellows, A., Miachon, L., Gu, Y., Bloem, M., Kapuria, A., 2020. The Food Systems Dashboard is a new tool to inform better food policy. Nat. Food 1, 243–246. https://doi.org/10.1038/s43016-020-0077-y.
Fanzo, J., Haddad, L., Schneider, K.R., Béné, C., Covic, N.M., Guarin, A., Herforth, A.W., Herrero, M., Sumaila, U.R., Aburto, N.J., Amuyunzu-Nyamongo, M., Barquera, S., Battersby, J., Beal, T., Bizzotto Molina, P., Brusset, E., Cafiero, C., Campeau, C., Caron, P., Cattaneo, A., Conforti, P., Davis, C., DeClerck, F.A.J., Elouafi, I., Fabi, C., Gephart, J.A., Golden, C.D., Hendriks, S.L., Huang, J., Laar, A., Lal, R., Lidder, P., Loken, B., Marshall, Q., Masuda, Y.J., McLaren, R., Neufeld, L.M., Nordhagen, S., Remans, R., Resnick, D., Silverberg, M., Torero Cullen, M., Tubiello, F.N., Vivero-Pol, J.-L., Wei, S., Rosero Moncayo, J., 2021. Viewpoint: rigorous monitoring is necessary to guide food system transformation in the countdown to the 2030 global goals. Food Pol. 104, 102163. https://doi.org/10.1016/j.foodpol.2021.102163.
FAO, 2020. The State of Food and Agriculture 2020: Overcoming Water Challenges in Agriculture. FAO, Rome.
FAO and Tufts University, 2019. Twin peaks: The Seasonality of Acute Malnutrition, Conflict and Environmental Factors. Chad, South Sudan and the Sudan. Rome.

Fentahun, N., Belachew, T., Coates, J., Lachat, C., 2018. Seasonality and determinants of child growth velocity and growth deficit in rural southwest Ethiopia. BMC Pediatr. 18, 20. https://doi.org/10.1186/s12887-018-0986-1.
Fernandez, B., Gopal, M., Ruthven, O., 2016. Land, Labour and Livelihoods: Indian Women's Perspectives. Springer.
Ferro-Luzzi, A., Morris, S.S., Taffesse, S., Demissie, T., D'Amato, M., 2002. Seasonal undernutrition in rural Ethiopia: magnitude, correlates, and functional significance. Food Nutr. Bull. 23, 421–422. https://doi.org/10.1177/156482650202300420.
Feuerbacher, A., McDonald, S., Dukpa, C., Grethe, H., 2020. Seasonal rural labor markets and their relevance to policy analyses in developing countries. Food Pol. 93, 101875. https://doi.org/10.1016/j.foodpol.2020.101875.
Fewtrell, L., Kaufmann, R.B., Kay, D., Enanoria, W., Haller, L., Colford, J.M.J., 2005. Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. Lancet Infect. Dis. 5, 42–52.
Fink, G., Jack, B.K., Masiye, F., 2020. Seasonal liquidity, rural labor markets, and agricultural production. Am. Econ. Rev. 110, 3351–3392. https://doi.org/10.1257/AER.20180607.
Fraeyman, S., Croubels, S., Devreese, M., Antonissen, G., 2017. Emerging Fusarium and alternaria mycotoxins: occurrence, toxicity and toxicokinetics. Toxins 9. https://doi.org/10.3390/toxins9070228.
Frank-Miller, E.G., Lambert, S.J., Henly, J.R., 2015. Age, wage, and job placement: older women's experiences entering the retail sector. J. Women Aging 27, 157–173. https://doi.org/10.1080/08952841.2014.928579.
Gelli, A., Aberman, N.-L., Margolies, A., Santacroce, M., Baulch, B., Chirwa, E., 2017. Lean-season food transfers affect children's diets and household food security: evidence from a Quasi-experiment in Malawi. J. Nutr. 147, 869–878. https://doi.org/10.3945/jn.116.246652.
Geremew, M., Gourio, F., 2018. Seasonal and business cycles of U.S. employment. Econ. Perspect. https://doi.org/10.21033/ep-2018-3.
Gertel, J., Sippel, S.R., 2014. Seasonal Workers in Mediterranean Agriculture: The Social Costs of Eating Fresh. Routledge.
Gilbert, C.L., Christiaensen, L., Kaminski, J., 2017. Food price seasonality in Africa: measurement and extent. Food Pol. 67, 119–132. https://doi.org/10.1016/j.foodpol.2016.09.016. Agriculture in Africa—Telling Myths from Facts.
Gill, G., 1992. Seasonality and agriculture in the developing world: a problem of the poor and powerless. Am. Anthropol. 94, 987–988. https://doi.org/10.1525/aa.1992.94.4.02a00650.
Global Panel on Agriculture and Food Systems for Nutrition, 2016. Food Systems and Diets: Facing the Challenges of the 21st Century. Global Panel on Agriculture and Food Systems for Nutrition, London, UK.
Gluckman, P.D., Hanson, M.A., 2004. Living with the past: evolution, development, and patterns of disease. Science 305, 1733–1736. https://doi.org/10.1126/science.1095292.
Gong, Y.Y., Watson, S., Routledge, M.N., 2016. Aflatoxin exposure and associated human health effects, a review of epidemiological studies. Food Saf. 4, 14–27. https://doi.org/10.14252/foodsafetyfscj.2015026.
Gottfried, H., Chun, J.J., 2018. Care work in transition: transnational circuits of gender, migration, and care. Crit. Sociol. 44, 997–1012. https://doi.org/10.1177/0896920518765931.
Graham, P., Stretesky, P., Long, M., Mann, E., Defeyter, M.A., 2018. Holiday hunger: feeding children during the school holidays. In: Feeding Children Inside and Outside the Home. Routledge.
Grellety, E., Luquero, F.J., Mambula, C., Adamu, H.H., Elder, G., Porten, K., 2013. Observational bias during nutrition surveillance: results of a mixed longitudinal and cross-sectional data collection system in Northern Nigeria. PLoS One 8, e62767. https://doi.org/10.1371/journal.pone.0062767.
Groopman, J.D., Kensler, T.W., 2005. Food safety | mycotoxins. In: Caballero, B. (Ed.), Encyclopedia of Human Nutrition, second ed. Elsevier, Oxford, pp. 317–323. https://doi.org/10.1016/B0-12-226694-3/00134-4.
Guidetti, G., Pedrini, G., Zamparini, L., 2020. Assessing perceived job quality among seasonal tourism workers: the case of Rimini, Italy. Tourism Econ. https://doi.org/10.1177/1354816620927524, 1354816620927524.
Haggarty, P., McNeill, G., Manneh, M.K.A., Davidson, L., 1994. The influence of exercise on the energy requirements of adult males in the UK. Br. J. Nutr. 72, 799–813. https://doi.org/10.1079/BJN19940086.
Han, J.-A., Chung, Y.-E., Chung, I.-H., Hong, Y.-H., Chung, S., 2021. Impact of the COVID-19 pandemic on seasonal variations in childhood and adolescent growth: experience of pediatric endocrine clinics. Children 8, 404. https://doi.org/10.3390/children8050404.
Harriss-White, B., 2010. Work and wellbeing in informal economies: the regulative roles of institutions of identity and the state. World Dev. 38, 170–183. https://doi.org/10.1016/j.worlddev.2009.10.011. Formal and Informal Institutions and Development.
Hashizume, M., Armstrong, B., Hajat, S., Wagatsuma, Y., Faruque, A.S.G., Hayashi, T., Sack, D.A., 2007. Association between climate variability and hospital visits for non-cholera diarrhoea in Bangladesh: effects and vulnerable groups. Int. J. Epidemiol. https://doi.org/10.1093/ije/dym148.
Higgins, P.A., Alderman, H., 1997. Labor and women's nutrition: the impact of work effort and fertility on nutritional status in Ghana. J. Hum. Resour. 32, 577–595. https://doi.org/10.2307/146184.
Hill, H.D., Morris, P., Gennetian, L.A., Wolf, S., Tubbs, C., 2013. The consequences of income instability for children's well-being. Child Dev. Perspect. 7, 85–90. https://doi.org/10.1111/cdep.12018.
Hillbruner, C., Egan, R., 2008. Seasonality, household food security, and nutritional status in Dinajpur, Bangladesh. Food Nutr. Bull. 29, 221–231. https://doi.org/10.1177/156482650802900308.
HLPE, 2017. Nutrition and Food Systems. A Report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security. Rome.
Hossain, M.J., Saha, D., Antonio, M., Nasrin, D., Blackwelder, W.C., Ikumapayi, U.N., Mackenzie, G.A., Adeyemi, M., Jasseh, M., Adegbola, R.A., Roose, A.W., Kotloff, K.L., Levine, M.M., 2019. Cryptosporidium infection in rural Gambian children: epidemiology and risk factors. PLoS Negl. Trop. Dis. 13, e0007607. https://doi.org/10.1371/journal.pntd.0007607.
Islam, M.S., Sharker, M.A.Y., Rheman, S., Hossain, S., Mahmud, Z.H., Islam, M.S., Uddin, A.M.K., Yunus, M., Osman, M.S., Ernst, R., Rector, I., Larson, C.P., Luby, S.P., Endtz, H.P., Cravioto, A., 2009. Effects of local climate variability on transmission dynamics of cholera in Matlab, Bangladesh. Trans. R. Soc. Trop. Med. Hyg. https://doi.org/10.1016/j.trstmh.2009.04.016.
Jagai, J.S., Sarkar, R., Castronovo, D., Kattula, D., McEntee, J., Ward, H., Kang, G., Naumova, E.N., 2012. Seasonality of rotavirus in South Asia: a meta-analysis approach assessing associations with temperature, precipitation, and vegetation index. PLoS One 7, e38168. https://doi.org/10.1371/journal.pone.0038168.
Jahns, L., Johnson, L.K., Scheett, A.J., Stote, K.S., Raatz, S.K., Subar, A.F., Tande, D., 2016. Measures of diet quality across calendar and winter holiday seasons among midlife women: a 1-year longitudinal study using the automated self-administered 24-hour recall. J. Acad. Nutr. Diet. 116, 1961–1969. https://doi.org/10.1016/j.jand.2016.07.013.
Jiang, T., Christian, P., Khatry, S.K., Wu, L., West, K.P., 2005. Micronutrient deficiencies in early pregnancy are common, concurrent, and vary by season among rural Nepali pregnant women. J. Nutr. 135, 1106–1112. https://doi.org/10.1093/jn/135.5.1106.
Johnston, D., Stevano, S., Malapit, H.J., Hull, E., Kadiyala, S., 2018. Review: time use as an explanation for the agri-nutrition disconnect: evidence from rural areas in low and middle-income countries. Food Pol. 76, 8–18. https://doi.org/10.1016/j.foodpol.2017.12.011.
Kaminski, J., Christiaensen, L., Gilbert, C.L., 2016. Seasonality in local food markets and consumption: evidence from Tanzania. Oxf. Econ. Pap. 68, 736–757. https://doi.org/10.1093/oep/gpw013.
Katona, P., Katona-Apte, J., 2008. The interaction between nutrition and infection. Clin. Infect. Dis. 46, 1582–1588. https://doi.org/10.1086/587658.
Kenny, G.P., Yardley, J., Brown, C., Sigal, R.J., Jay, O., 2010. Heat stress in older individuals and patients with common chronic diseases. CMAJ 182, 1053–1060. https://doi.org/10.1503/cmaj.081050.

Keusch, G.T., Denno, D.M., Black, R.E., Duggan, C., Guerrant, R.L., Lavery, J.V., Nataro, J.P., Rosenberg, I.H., Ryan, E.T., Tarr, P.I., Ward, H., Bhutta, Z.A., Coovadia, H., Lima, A., Ramakrishna, B., Zaidi, A.K.M., Hay Burgess, D.C., Brewer, T., 2014. Environmental enteric dysfunction: pathogenesis, diagnosis, and clinical consequences. Clin. Infect. Dis. 59 (Suppl. 4), S207–S212. https://doi.org/10.1093/cid/ciu485.

Kiehne, E., Mendoza, N.S., 2015. Migrant and seasonal farmworker food insecurity: prevalence, impact, risk factors, and coping strategies. Soc. Work Publ. Health 30, 397–409. https://doi.org/10.1080/19371918.2015.1019173.

Kinyoki, D.K., Berkley, J.A., Moloney, G.M., Odundo, E.O., Kandala, N.-B., Noor, A.M., 2016. Space-time mapping of wasting among children under the age of five years in Somalia from 2007 to 2010. Spat. Spatiotemporal Epidemiol. 16, 77–87. https://doi.org/10.1016/j.sste.2015.12.002.

Kirby, M.A., Nagel, C.L., Rosa, G., Iyakaremye, L., Zambrano, L.D., Clasen, T.F., 2016. Faecal contamination of household drinking water in Rwanda: a national cross-sectional study. Sci. Total Environ. 571, 426–434. https://doi.org/10.1016/j.scitotenv.2016.06.226.

Kokubo, Y., Yokoyama, Y., Kisara, K., Ohira, Y., Sunami, A., Yoshizaki, T., Tada, Y., Ishizaki, S., Hida, A., Kawano, Y., 2016. Relationship between dietary factors and bodily iron status among Japanese collegiate elite female rhythmic gymnasts. Int. J. Sport Nutr. Exerc. Metab. 26, 105–113. https://doi.org/10.1123/ijsnem.2015-0123.

Kuhn, M.A., 2018. Who feels the calorie crunch and when? The impact of school meals on cyclical food insecurity. J. Publ. Econ. 166, 27–38. https://doi.org/10.1016/j.jpubeco.2018.08.001.

Kulinkina, A.V., Kosinski, K.C., Liss, A., Adjei, M.N., Ayamgah, G.A., Webb, P., Gute, D.M., Plummer, J.D., Naumova, E.N., 2016a. Piped water consumption in Ghana: a case study of temporal and spatial patterns of clean water demand relative to alternative water sources in rural small towns. Sci. Total Environ. 559, 291–301. https://doi.org/10.1016/j.scitotenv.2016.03.148.

Kulinkina, A.V., Mohan, V.R., Francis, M.R., Kattula, D., Sarkar, R., Plummer, J.D., Ward, H., Kang, G., Balraj, V., Naumova, E.N., 2016b. Seasonality of water quality and diarrheal disease counts in urban and rural settings in south India. Sci. Rep. 6, 20521. https://doi.org/10.1038/srep20521.

Lauer, J.M., Duggan, C.P., Ausman, L.M., Griffiths, J.K., Webb, P., Wang, J.-S., Xue, K.S., Agaba, E., Nshakira, N., Ghosh, S., 2019. Maternal aflatoxin exposure during pregnancy and adverse birth outcomes in Uganda. Matern. Child Nutr. 15, e12701. https://doi.org/10.1111/mcn.12701.

Laurito, A., Schwartz, A.E., 2019. Does school lunch fill the "SNAP gap" at the end of the month? South. Econ. J. 86, 49–82. https://doi.org/10.1002/soej.12370.

Lawson, A.B., 2013. 3. Geographical representation and mapping. In: Statistical Methods in Spatial Epidemiology, Wiley Series in Probability and Statistics. John Wiley & Sons.

Lerche, J., 2012. Labour regulations and labour standards in India: decent work? Glob. Lab. J. 3. https://doi.org/10.15173/glj.v3i1.1111.

Li, Y., Reeves, R.M., Wang, X., Bassat, Q., Brooks, W.A., Cohen, C., Moore, D.P., Nunes, M., Rath, B., Campbell, H., Nair, H., RSV Global Epidemiology Network, investigators, R.E.S.C.E.U., 2019. Global patterns in monthly activity of influenza virus, respiratory syncytial virus, parainfluenza virus, and metapneumovirus: a systematic analysis. Lancet Glob. Health 7, e1031–e1045. https://doi.org/10.1016/S2214-109X(19)30264-5.

Lim, Y.M., Song, S., Song, W.O., 2017. Prevalence and determinants of overweight and obesity in children and adolescents from migrant and seasonal farmworker families in the United States—a systematic review and qualitative assessment. Nutrients 9, 188. https://doi.org/10.3390/nu9030188.

Lindqvist, A.N., Fornell, R., Prade, T., Tufvesson, L., Khalil, S., Kopainsky, B., 2021. Human-water dynamics and their role for seasonal water scarcity—a case study. Water Resour. Manag. https://doi.org/10.1007/s11269-021-02819-1.

Loutan, L., Lamotte, J.M., 1984. Seasonal variations in nutrition among a group of nomadic pastoralists in Niger. Lancet 1, 945–947. https://doi.org/10.1016/s0140-6736(84)92398-5.

Ma, Y., Olendzki, B.C., Li, W., Hafner, A.R., Chiriboga, D., Hebert, J.R., Campbell, M., Sarnie, M., Ockene, I.S., 2006. Seasonal variation in food intake, physical activity, and body weight in a predominantly overweight population. Eur. J. Clin. Nutr. 60, 519–528. https://doi.org/10.1038/sj.ejcn.1602346.

Madan, E.M., Haas, J.D., Menon, P., Gillespie, S., 2018. Seasonal variation in the proximal determinants of undernutrition during the first 1000 days of life in rural South Asia: a comprehensive review. Glob. Food Secur. 19, 11–23. https://doi.org/10.1016/j.gfs.2018.08.008.

Maleta, K., Virtanen, S., Espo, M., Kulmala, T., Ashorn, P., 2003. Seasonality of growth and the relationship between weight and height gain in children under three years of age in rural Malawi. Acta Paediatr. 92, 491–497. https://doi.org/10.1111/j.1651-2227.2003.tb00584.x.

Marasas, W.F.O., Gelderblom, W., Shephard, G., Vismer, H., 2008. Mycotoxins: a global problem. In: Mycotoxins: Detection Methods, Management, Public Health and Agricultural Trade, pp. 29–39. https://doi.org/10.1079/9781845930820.0029.

Marshak, A., 2021. Seasonality of Acute Malnutrition and Its Drivers in Sila Province, Chad: A Mixed Methods Analysis. Boston, MA.

Marshak, A., Venkat, A., Young, H., 2021. How seasonality of malnutrition is measured and analyzed. Int. J. Environ. Res. Publ. Health 18 (4).

Mbow, C., 2019. Food security. In: Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems.

Meadows, D.H., Wright, D., 2008. Thinking in Systems: A Primer. Chelsea Green Pub, White River Junction, Vt.

Mekonnen, M.M., Hoekstra, A.Y., 2016. Four billion people facing severe water scarcity. Sci. Adv. 2, e1500323. https://doi.org/10.1126/sciadv.1500323.

Miller, J., Ritchie, B., Tran, C., Beggs, S., Lada, C.O., Whetter, K., Cobiac, L., (n.d.). Seasonal variation in the nutritional status of children aged 6 to 60 months in a resettlement village in West Timor. Asia Pac. J. Clin. Nutr. 22, 449–456. https://doi.org/10.3316/ielapa.507527427145727.

Millward, D.J., 2017. Nutrition, infection and stunting: the roles of deficiencies of individual nutrients and foods, and of inflammation, as determinants of reduced linear growth of children. Nutr. Res. Rev. 30, 50–72. https://doi.org/10.1017/S0954422416000238.

Mishra, D.K., 2020. Seasonal migration and unfree labour in globalising India: insights from field surveys in Odisha. Indian J. Lab. Econ. 63, 1087–1106. https://doi.org/10.1007/s41027-020-00277-8.

Moran, J.M., 2009. Weather Studies: Introduction to Atmospheric Science. American Meteorological Society, Boston, Massachusetts.

Myers, S.S., Smith, M.R., Guth, S., Golden, C.D., Vaitla, B., Mueller, N.D., Dangour, A.D., Huybers, P., 2017. Climate change and global food systems: potential impacts on food security and undernutrition. Annu. Rev. Publ. Health 38, 259–277. https://doi.org/10.1146/annurev-publhealth-031816-044356.

Nabwera, H.M., Fulford, A.J., Moore, S.E., Prentice, A.M., 2017. Growth faltering in rural Gambian children after four decades of interventions: a retrospective cohort study. Lancet Glob. Health 5, e208–e216. https://doi.org/10.1016/S2214-109X(16)30355-2.

Naumova, E.N., 2006. Mystery of seasonality: getting the rhythm of nature. J. Publ. Health Pol. https://doi.org/10.1057/palgrave.jphp.3200061.

Naumova, E.N., MacNeil, I.B., 2006. Chapter 28: seasonality assessment for biosurveillance systems. In: Balakrishnan, N., Auget, J., Mesbah, M., Molenberg, G. (Eds.), Advances in Statistical Methods for the Health Sciences: Applications to Cancer and AIDS Studies, Genome Sequence Analysis, and Survival Analysis. Birkhauser, Boston, pp. 437–450.

Nielsen, J., Prudhon, C., de Radigues, X., 2011. Trends in malnutrition and mortality in Darfur, Sudan, between 2004 and 2008: a meta-analysis of publicly available surveys. Int. J. Epidemiol. 40, 971–984. https://doi.org/10.1093/ije/dyr010.

Nwanta, J.A., Egege, S.C., Alli-Balogun, J.K., Ezema, W.S., 2008. Evaluation of prevalence and seasonality of Newcastle disease in chicken in Kaduna, Nigeria. Worlds Poult. Sci. J. 64, 416–423. https://doi.org/10.1017/S0043933908000147.

Obando-Pacheco, P., Justicia-Grande, A.J., Rivero-Calle, I., Rodríguez-Tenreiro, C., Sly, P., Ramilo, O., Mejías, A., Baraldi, E., Papadopoulos, N.G., Nair, H., Nunes, M.C., Kragten-Tabatabaie, L., Heikkinen, T., Greenough, A., Stein, R.T., Manzoni, P., Bont, L., Martinón-Torres, F., 2018. Respiratory syncytial virus seasonality: a global overview. J. Infect. Dis. 217, 1356–1364. https://doi.org/10.1093/infdis/jiy056.

Orr, J.B., Clark, M.L., 1930. Seasonal variation in the growth of school children. Lancet 2, 365–367.

Panter-Brick, C., 1997. Seasonal growth patterns in rural Nepali children. Ann. Hum. Biol. 24, 1–18. https://doi.org/10.1080/03014469700004732.

Park, C.S., Vogel, E., Larson, L.M., Myers, S.S., Daniel, M., Biggs, B.-A., 2019. The global effect of extreme weather events on nutrient supply: a superposed epoch analysis. Lancet Planet. Health 3, e429–e438. https://doi.org/10.1016/S2542-5196(19)30193-7.

Pascual, M., Dobson, A., 2005. Seasonal patterns of infectious diseases. PLoS Med. 2, 0018–0020. https://doi.org/10.1371/journal.pmed.0020005.

Pascual, M., Bouma, M.J., Dobson, A.P., 2002. Cholera and climate: revisiting the quantitative evidence. Microb. Infect. 237–245. https://doi.org/10.1016/S0140-6736(02)11937-4.
Pattenden, J., 2016. Working at the margins of global production networks: local labour control regimes and rural-based labourers in South India. Third World Q. 37, 1809–1833. https://doi.org/10.1080/01436597.2016.1191939.
Pelletier, D.L., Deneke, K., Kidane, Y., Haile, B., Negussie, F., 1995. The food-first bias and nutrition policy: lessons from Ethiopia. Food Pol. 20, 279–298. https://doi.org/10.1016/0306-9192(95)00026-7.
Phalkey, R.K., Aranda-Jan, C., Marx, S., Hofle, B., Sauerborn, R., 2015. Systematic review of current efforts to quantify the impacts of climate change on undernutrition. Proc. Natl. Acad. Sci. U. S. A. 112, E4522–E4529. https://doi.org/10.1073/pnas.1409769112.
Pinstrup-Andersen, P. (Ed.), 2015. Food Price Policy in an Era of Market Instability: A Political Economy Analysis, first ed. UNU-WIDER studies in development economics. Oxford University Press, Oxford.
Plasqui, G., Kester, A.D.M., Westerterp, K.R., 2003. Seasonal variation in sleeping metabolic rate, thyroid activity, and leptin. Am. J. Physiol. Endocrinol. Metab. 285, E338–E343. https://doi.org/10.1152/ajpendo.00488.2002.
Pollitt, E., Arthur, J., 1989. Seasonality and weight gain during the first year of life. Am. J. Hum. Biol. 1, 747–756. https://doi.org/10.1002/ajhb.1310010613.
Quandt, S.A., Trejo, G., Suerken, C.K., Pulgar, C.A., Ip, E.H., Arcury, T.A., 2016. Diet quality among preschool-age children of Latino migrant and seasonal farmworkers in the United States. J. Immigr. Minority Health 18, 505–512. https://doi.org/10.1007/s10903-015-0304-z.
Quandt, S.A., Arnold, T.J., Mora, D.C., Arcury, T.A., Talton, J.W., Daniel, S.S., 2021. Hired latinx child farmworkers in North Carolina: educational status and experience through a social justice lens. New Solut. 30, 282–293. https://doi.org/10.1177/1048291120970207.
Ramanathan, K., Thenmozhi, M., George, S., Anandan, S., Veeraraghavan, B., Naumova, E.N., Jeyaseelan, L., 2020. Assessing seasonality variation with harmonic regression: accommodations for sharp peaks. IJERPH 17, 1318. https://doi.org/10.3390/ijerph17041318.
Rao, N., Raju, S., 2020. Gendered time, seasonality, and nutrition: insights from two Indian districts. Fem. Econ. 26, 95–125. https://doi.org/10.1080/13545701.2019.1632470.
Rao, N., Gazdar, H., Chanchani, D., Ibrahim, M., 2019. Women's agricultural work and nutrition in South Asia: from pathways to a cross-disciplinary, grounded analytical framework. Food Pol. 82, 50–62. https://doi.org/10.1016/j.foodpol.2018.10.014. Special Issue: Leveraging Agriculture for Nutrition in South Asia.
Reiches, M.W., Moore, S.E., Prentice, A.M., Prentice, A., Sawo, Y., Ellison, P.T., 2013. The adolescent transition under energetic stress: body composition tradeoffs among adolescent women in the Gambia. Evol. Med. Publ. Health 2013, 75–85. https://doi.org/10.1093/emph/eot005.
Roba, K.T., O'Connor, T.P., Belachew, T., O'Brien, N.M., 2016. Variations between post- and pre-harvest seasons in stunting, wasting, and Infant and Young Child Feeding (IYCF) practices among children 6–23 months of age in lowland and midland agro-ecological zones of rural Ethiopia. Pan Afr. Med. J. 24, 163. https://doi.org/10.11604/pamj.2016.24.163.9387.
Robinson, D., On behalf of CARE USA, 2010. Social protection, livelihoods and the hidden economy of care. Development 53, 325–332. https://doi.org/10.1057/dev.2010.57.
Roca-Feltrer, A., Schellenberg, J.R.A., Smith, L., Carneiro, I., 2009. A simple method for defining malaria seasonality. Malar. J. https://doi.org/10.1186/1475-2875-8-276.
Rohli, R.V., Vega, A.J., 2017. Climatology. Jones & Bartlett Learning.
Ronnenberg, A.G., Goldman, M.B., Aitken, I.W., Xu, X., 2000. Anemia and deficiencies of folate and vitamin B-6 are common and vary with season in Chinese women of childbearing age. J. Nutr. 130, 2703–2710. https://doi.org/10.1093/jn/130.11.2703.
Rosenzweig, C., Mbow, C., Barioni, L.G., Benton, T.G., Herrero, M., Krishnapillai, M., Liwenga, E.T., Pradhan, P., Rivera-Ferre, M.G., Sapkota, T., Tubiello, F.N., Xu, Y., Mencos Contreras, E., Portugal-Pereira, J., 2020. Climate change responses benefit from a global food system approach. Nat. Food 1, 94–97. https://doi.org/10.1038/s43016-020-0031-z.
Routh, S., Borghi, V. (Eds.), 2016. Workers and the Global Informal Economy: Interdisciplinary Perspectives. Routledge.
Rumbeiha, W.K., Oehme, F.W., 2005. Veterinary toxicology. In: Wexler, P. (Ed.), Encyclopedia of Toxicology, second ed. Elsevier, New York, pp. 420–434. https://doi.org/10.1016/B0-12-369400-0/01004-8.
Sadler, K., Kerven, C., Manske, M., Catley, A., 2009. Milk Matters: A Literature Review of Pastoralist Nutrition and Programming Responses. Feinstein International Center, Tufts University and Save the Children, Addis Ababa.
Sahn, D.E. (Ed.), 1989. Seasonal Variability in Third World Agriculture: The Consequences for Food Security. Johns Hopkins University Press for the International Food Policy Research Institute, Baltimore, Maryland.
Schaible, U.E., Kaufmann, S.H.E., 2007. Malnutrition and infection: complex mechanisms and global impacts. PLoS Med. 4, e115. https://doi.org/10.1371/journal.pmed.0040115.
Schoenbuchner, S.M., Dolan, C., Mwangome, M., Hall, A., Richard, S.A., Wells, J.C., Khara, T., Sonko, B., Prentice, A.M., Moore, S.E., 2019. The relationship between wasting and stunting: a retrospective cohort analysis of longitudinal data in Gambian children from 1976 to 2016. Am. J. Clin. Nutr. 110, 498–507. https://doi.org/10.1093/ajcn/nqy326.
Schofield, S., 1974. Seasonal factors affecting nutrition in different age groups and especially preschool children. J. Dev. Stud. 11, 22–40. https://doi.org/10.1080/00220387408421510.
Schwinger, C., Lunde, T.M., Andersen, P., Kismul, H., Van den Broeck, J., 2014. Seasonal and spatial factors related to longitudinal patterns of child growth in Bwamanda, DR Congo. Earth Perspect. 1, 26. https://doi.org/10.1186/s40322-014-0026-8.
Scrimshaw, N.S., Taylor, C.E., Gordon, J.E., 1968. Interactions of Nutrition and Infection. Monograph Series No. 37. Geneva, Switzerland.
Sellen, D.W., 2000. Seasonal ecology and nutritional status of women and children in a Tanzanian pastoral community. Am. J. Hum. Biol. 12, 758–781. https://doi.org/10.1002/1520-6300(200011/12)12:6<758::AID-AJHB5>3.0.CO;2-R.
Sen, A., 2000. Development as Freedom, reprint ed. Anchor, New York.
Serrat, O., 2017. The sustainable livelihoods approach. In: Serrat, O. (Ed.), Knowledge Solutions: Tools, Methods, and Approaches to Drive Organizational Performance. Springer, Singapore, pp. 21–26. https://doi.org/10.1007/978-981-10-0983-9_5.
Shell-Duncan, B., 1995. Impact of seasonal variation in food availability and disease stress on the health status of nomadic Turkana children: a longitudinal analysis of morbidity, immunity, and nutritional status. Am. J. Hum. Biol. 7, 339–355. https://doi.org/10.1002/ajhb.1310070310.
Sibhatu, K.T., Qaim, M., 2017. Rural food security, subsistence agriculture, and seasonality. PLoS One 12, e0186406. https://doi.org/10.1371/journal.pone.0186406.
Simondon, K.B., Bénéfice, E., Simondon, F., Delaunay, V., Chahnazarian, A., 1993. Seasonal variation in nutritional status of adults and children in rural Senegal. In: Ulijaszek, S.J., Strickland, S.S. (Eds.), Seasonality and Human Ecology, Society for the Study of Human Biology Symposium Series. Cambridge University Press, Cambridge, pp. 166–183. https://doi.org/10.1017/CBO9780511600517.013.
Simpson, R.B., Alarcon Falconi, T.M., Venkat, A., Chui, K.H.H., Navidad, J., Naumov, Y.N., Gorski, J., Bhattacharyya, S., Naumova, E.N., 2019. Incorporating calendar effects to predict influenza seasonality in Milwaukee, Wisconsin. Epidemiol. Infect. 147, e268. https://doi.org/10.1017/S0950268819001511.
Simpson, R.B., Zhou, B., Naumova, E.N., 2020. Seasonal synchronization of foodborne outbreaks in the United States, 1996-2017. Sci. Rep. 10, 17500. https://doi.org/10.1038/s41598-020-74435-9.
Smith, J., Cuesta, G., 2020. Hunger in the fields: food insecurity and food access among farmworker families in Migrant and Seasonal Head Start. J. Latinos Educ. 19, 246–257. https://doi.org/10.1080/15348431.2018.1500291.
Smith, L., Stasiewicz, M., Hestrin, R., Morales, L., Mutiga, S., Nelson, R., 2016. Examining environmental drivers of spatial variability in aflatoxin accumulation in Kenyan maize: potential utility in risk prediction models. AJFAND 16, 11086–11105. https://doi.org/10.18697/ajfand.75.ILRI09.
Springmann, M., Mason-D'Croz, D., Robinson, S., Garnett, T., Godfray, H.C.J., Gollin, D., Rayner, M., Ballon, P., Scarborough, P., 2016. Global and regional health effects of future food production under climate change: a modelling study. Lancet 387, 1937–1946. https://doi.org/10.1016/S0140-6736(15)01156-3.
Springmann, M., Wiebe, K., Mason-D'Croz, D., Sulser, T.B., Rayner, M., Scarborough, P., 2018. Health and nutritional aspects of sustainable diet strategies and their association with environmental impacts: a global modelling analysis with country-level detail. Lancet Planet. Health 2, e451–e461. https://doi.org/10.1016/S2542-5196(18)30206-7.

Stelmach-Mardas, M., Kleiser, C., Uzhova, I., Penalvo, J.L., La Torre, G., Palys, W., Lojko, D., Nimptsch, K., Suwalska, A., Linseisen, J., Saulle, R., Colamesta, V., Boeing, H., 2016. Seasonality of food groups and total energy intake: a systematic review and meta-analysis. Eur. J. Clin. Nutr. 70, 700–708. https://doi.org/10.1038/ejcn.2015.224.

Stevens, A., 2010. Introduction to the basic drivers of climate. Nat. Educ. Knowl. 3 (10) [WWW Document]. URL. https://www.nature.com/scitable/knowledge/library/introduction-to-the-basic-drivers-of-climate-13368032/. (Accessed 15 July 2021).

Tanner, J., 1981. A History of the Study of Human Growth. Cambridge University Press, Cambridge.

Tate, N., Atkinson, P.M., 2001. Modelling Scale in Geographical Information Science. John Wiley & Sons.

Thornton, P.K., 2010. Livestock production: recent trends, future prospects. Phil. Trans. Biol. Sci. 365, 2853–2867. https://doi.org/10.1098/rstb.2010.0134.

Tomkins, A.M., Dunn, D.T., Hayes, R.J., Bradley, A.K., 1986. Seasonal variations in the nutritional status of urban Gambian children. Br. J. Nutr. 56, 533–543. https://doi.org/10.1079/bjn19860134.

Trowbridge, F.L., Newton, L.H., 1979. Seasonal changes in malnutrition and diarrheal disease among preschool children in El Salvador. Am. J. Trop. Med. Hyg. 28, 136–141. https://doi.org/10.4269/ajtmh.1979.28.136.

Tubiello, F.N., Rosenzweig, C., Conchedda, G., Karl, K., Gütschow, J., Xueyao, P., Obli-Laryea, G., Wanner, N., Qiu, S.Y., Barros, J.D., Flammini, A., Mencos-Contreras, E., Souza, L., Quadrelli, R., Heiðarsdóttir, H.H., Benoit, P., Hayek, M., Sandalow, D., 2021. Greenhouse gas emissions from food systems: building the evidence base. Environ. Res. Lett. 16, 065007. https://doi.org/10.1088/1748-9326/ac018e.

Turicchi, J., O'Driscoll, R., Horgan, G., Duarte, C., Palmeira, A.L., Larsen, S.C., Heitmann, B.L., Stubbs, J., 2020. Weekly, seasonal and holiday body weight fluctuation patterns among individuals engaged in a European multi-centre behavioural weight loss maintenance intervention. PLoS One 15, e0232152. https://doi.org/10.1371/journal.pone.0232152.

Vaarst, M., Escudero, A.G., Chappell, M.J., Brinkley, C., Nijbroek, R., Arraes, N.A.M., Andreasen, L., Gattinger, A., Almeida, G.F.D., Bossio, D., Halberg, N., 2018. Exploring the concept of agroecological food systems in a city-region context. Agroecol. Sustain. Food Syst. 42, 686–711. https://doi.org/10.1080/21683565.2017.1365321.

Vaitla, B., Devereux, S., Swan, S.H., 2009. Seasonal hunger: a neglected problem with proven solutions. PLoS Med. 6, e1000101. https://doi.org/10.1371/journal.pmed.1000101.

van der Toorn, J.E., Cepeda, M., Kiefte-de Jong, J.C., Franco, O.H., Voortman, T., Schoufour, J.D., 2020. Seasonal variation of diet quality in a large middle-aged and elderly Dutch population-based cohort. Eur. J. Nutr. 59, 493–504. https://doi.org/10.1007/s00394-019-01918-5.

Vargas, M., Gascon, J., Casals, C., Schellenberg, D., Urassa, H., Kahigwa, E., Ruiz, J., Vila, J., 2004. Etiology of diarrhea in children less than five years of age in Ifakara, Tanzania. Am. J. Trop. Med. Hyg. 70, 536–539.

Venkat, A., Naumova, E.N., Marshak, A., 2023. Seasonality of Acute Malnutrition in African Drylands: A Review of 15 Years of Nutrition Surveys. Food and Nutrition Bulletin Special Issue on Drylands, Malnutrition (forthcoming).

Vir, S.C., 2015. Maternal Nutrition: breaking the intergenerational cycle of malnutrition: significance, challenges and opportunities. In: Undernutrition and Public Policy in India. Routledge India.

Watts, N., Amann, M., Ayeb-Karlsson, S., Belesova, K., Bouley, T., Boykoff, M., Byass, P., Cai, W., Campbell-Lendrum, D., Chambers, J., Cox, P.M., Daly, M., Dasandi, N., Davies, M., Depledge, M., Depoux, A., Dominguez-Salas, P., Drummond, P., Ekins, P., Flahault, A., Frumkin, H., Georgeson, L., Ghanei, M., Grace, D., Graham, H., Grojsman, R., Haines, A., Hamilton, I., Hartinger, S., Johnson, A., Kelman, I., Kiesewetter, G., Kniveton, D., Liang, L., Lott, M., Lowe, R., Mace, G., Sewe, M.O., Maslin, M., Mikhaylov, S., Milner, J., Latifi, A.M., Moradi-Lakeh, M., Morrissey, K., Murray, K., Neville, T., Nilsson, M., Oreszczyn, T., Owfi, F., Pencheon, D., Pye, S., Rabbaniha, M., Robinson, E., Rocklöv, J., Schütte, S., Shumake-Guillemot, J., Steinbach, R., Tabatabaei, M., Wheeler, N., Wilkinson, P., Gong, P., Montgomery, H., Costello, A., 2018. The Lancet Countdown on health and climate change: from 25 years of inaction to a global transformation for public health. Lancet 391, 581–630. https://doi.org/10.1016/S0140-6736(17)32464-9.

Watts, N., Amann, M., Arnell, N., Ayeb-Karlsson, S., Beagley, J., Belesova, K., Boykoff, M., Byass, P., Cai, W., Campbell-Lendrum, D., Capstick, S., Chambers, J., Coleman, S., Dalin, C., Daly, M., Dasandi, N., Dasgupta, S., Davies, M., Di Napoli, C., Dominguez-Salas, P., Drummond, P., Dubrow, R., Ebi, K.L., Eckelman, M., Ekins, P., Escobar, L.E., Georgeson, L., Golder, S., Grace, D., Graham, H., Haggar, P., Hamilton, I., Hartinger, S., Hess, J., Hsu, S.-C., Hughes, N., Jankin Mikhaylov, S., Jimenez, M.P., Kelman, I., Kennard, H., Kiesewetter, G., Kinney, P.L., Kjellstrom, T., Kniveton, D., Lampard, P., Lemke, B., Liu, Y., Liu, Z., Lott, M., Lowe, R., Martinez-Urtaza, J., Maslin, M., McAllister, L., McGushin, A., McMichael, C., Milner, J., Moradi-Lakeh, M., Morrissey, K., Munzert, S., Murray, K.A., Neville, T., Nilsson, M., Sewe, M.O., Oreszczyn, T., Otto, M., Owfi, F., Pearman, O., Pencheon, D., Quinn, R., Rabbaniha, M., Robinson, E., Rocklöv, J., Romanello, M., Semenza, J.C., Sherman, J., Shi, L., Springmann, M., Tabatabaei, M., Taylor, J., Triñanes, J., Shumake-Guillemot, J., Vu, B., Wilkinson, P., Winning, M., Gong, P., Montgomery, H., Costello, A., 2021. The 2020 report of the Lancet Countdown on health and climate change: responding to converging crises. Lancet 397, 129–170. https://doi.org/10.1016/S0140-6736(20)32290-X.

Westerterp, K.R., 2020. Seasonal variation in body mass, body composition and activity-induced energy expenditure: a long-term study. Eur. J. Clin. Nutr. 74, 135–140. https://doi.org/10.1038/s41430-019-0408-y.

WHO, 2018. Healthy Diet. Fact Sheet No. 394 [WWW Document]. URL. https://www.who.int/publications/m/item/healthy-diet-factsheet394. (Accessed 22 December 2020).

WHO and UNICEF, 2021. Progress on Household Drinking Water, Sanitation and Hygiene 2000–2020: Five Years Into the SDGs. World Health Organization (WHO) and the United Nations Children's Fund (UNICEF), Geneva.

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F., Wood, A., Jonell, M., Clark, M., Gordon, L.J., Fanzo, J., Hawkes, C., Zurayk, R., Rivera, J.A., Vries, W.D., Sibanda, L.M., Afshin, A., Chaudhary, A., Herrero, M., Agustina, R., Branca, F., Lartey, A., Fan, S., Crona, B., Fox, E., Bignet, V., Troell, M., Lindahl, T., Singh, S., Cornell, S.E., Reddy, K.S., Narain, S., Nishtar, S., Murray, C.J.L., 2019. Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. Lancet 393, 447–492. https://doi.org/10.1016/S0140-6736(18)31788-4.

Wright, J., Vazé, P., Russell, G., Gundry, S.W., Ferro-Luzzi, A., Mucavele, P., Nyatsanza, J., 2001. Seasonal aspects of weight-for-age in young children in Zimbabwe. Publ. Health Nutr. 4, 757–764. https://doi.org/10.1079/phn2000100.

Young, H., Jaspars, S., Overseas Development Institute (London, E., Humanitarian Practice Network, 2006. The Meaning and Measurement of Acute Malnutrition in Emergencies: A Primer for Decision-Makers. Overseas Development Institute, Humanitarian Practice Network, London.

Zanias, G.P., 1999. Seasonality and spatial integration in agricultural (product) markets. Agric. Econ. 20, 253–262. https://doi.org/10.1016/S0169-5150(99)00006-7.

Climate change: Effects on health and nutrition

Shweta Khandelwal[a], **Sarah Parwez**[b], **and Mahima Mehra**[c], [a] Public Health Foundation of India, Gurugram, Haryana, India; [b] Jamia Millia Islamia, Department of Hospital Management and Hospice Studies, New Delhi, India; and [c] Institute of Development Studies (IDS), University of Sussex, Brighton, United Kingdom

© 2023 Elsevier Ltd. All rights reserved.

Introduction 370
Impacts of climate-led adversities on health, nutrition & food systems 371
Food availability 371
Food affordability 372
Food accessibility 372
Nutritional value of food 372
Impacts of climate-led adversities on environmental pathways 373
Agriculture 373
Floods and droughts 374
Biodiversity & pollinator loss 375
Climate resilience approaches: adaption and mitigation as the way forward 375
Adaptation approaches for food systems 375
Managing risks of natural disasters by improving natural and man-made infrastructures 376
Political and policy-bases interventions 377
References 377
Further Reading 378

Key points

- Anthropogenic climate change has resulted in elevated global mean ambient temperatures, CO_2 concentration and disturbed the World's biogeochemical cycles. The consequences of these changes have been pervasive, resulting in direct and indirect pathways. **The direct pathways** include the erraticism of weather-related events, which have intensified and become more extreme. **Indirect pathways** include various insecurities to the dimensions of health, such as burdening socioeconomic positions, risk for diseases, displacement, unaffordability of food, etc.
- Climate related adversities are responsible for the decline in **agricultural yield and crop-loss, threatening food security** in low-and middle-income countries which already experience acute food shortages, such as Sub-Saharan African and South Asian countries but also some high-income countries including Canada, Australia, and the USA.
- Elevated temperatures and CO_2 levels alter the **quality of nutritional content in food crops,** diminishing the micronutrient and protein content in staple crops. On its own, CO_2 has increased yields for certain crops, but when acting synergistically with elevated temperatures, it usually has resulted in declines. C3 plants, the main type of food crops, seem to be more prone to these effects, as zinc, folate and protein content were studied to have reduced along with their yields. C4 crops are not as affected as C3 plants.
- Reduced yields and nutritional value of **Zinc, Iron, Vitamin A and Protein** will exacerbate existing micronutrient deficiencies, leading to risk of nutrition-associated diseases i.e., **childhood stunting, diarrhea and Non-Communicable Diseases** including diabetes mellitus and cardiovascular diseases.
- Climate change will make wet parts of the world wetter and drier parts drier. Increased incidences of floods cause devastation to life, destroy infrastructure and permeate the spread of infectious diseases due to unavailability of clean water. Inundations through **floods** will also lead to soil salinity making lands unfit for agriculture.
- **Pollinators,** which play a key role in the success of plant fertilizations, have been steadily declining due to climate change, and decoupling plant-pollinator relationships.
- **Women and children** will be particularly hard-hit due to existing social infrastructure and climatic shocks.
- Adaptation strategies to address food insecurity entail fortification of food, use of climate smart varieties, traditional ways of farming, and switching to plant-focused diets.

Glossary

Biofortification Strategically enhancing the nutritional value of food crops using biotechnology or selective breeding

C3 plants These plants utilize an enzyme called Ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCo) to accept atmospheric carbon dioxide during photosynthesis to produce a 3-carbon compound called 3-phosphoglyceric acid

C4 plants These plants utilize an enzyme called phosphenolpyruvate (PEP) to accept atmospheric carbon dioxide during photosynthesis to produce a 4-carbon compound called oxaloacetate (OAA)
Climate resilience The capacity for infrastructures and socio-economic spheres to be able to withstand upcoming climate threats
Food security As defined by the UN, it is the ability of a person to access affordable, nutritious food that is in accordance of their preferences at all times
Hydroponics Plants are grown in nutrient and mineral solutions in highly controlled environments without the use of soil
Malnutrition It occurs when a person has inadequate nutrients in their bodies
Micronutrient Deficiency Insufficient quantity of minerals and vitamins in an individual's body, which don't meet the requirement for optimal growth and well-being
Overnutrition Excess of nutrients in the body which arise from disproportionately high food consumption

Acronyms

CC Climate Change
CGIAR Consortium of International Agricultural Research Centers
CO_2 Carbon Dioxide
FAO Food and Agriculture Organization
Fe Iron
GHG Greenhouse gases
GHGE Greenhouse gas emissions
LECZ Low-elevation coastal zone
LICs Low-income countries
LMICs Low-and Middle-Income Countries
NCDs Non-Communicable Diseases
NDC Nationally Determined Contributions
OECD Organization for Economic Co-operation and Development
SADMS South Asia Drought Monitoring System
UN United Nations
WASH Water, Sanitation and Hygiene
WFP World Food Program
WHO World Health Organization
Zn Zinc

Introduction

Climate Change (CC) has been gaining momentum in the last few years as global weather patterns have become more and more inconsistent due to anthropogenic activities, which have accelerated the Earth's natural course of climate modification. This accelerated progression has destructive consequences and disturbs whole ecosystems, which further ramify into consequences for human well-being. Greenhouse gas emissions have led to an increase in ambient temperatures—with an overall 1 °C rise recorded in 2015 when compared with pre-industrial times. There's now almost a 50-50% chance that temperatures will exceed 1.5 °C by 2027.

Nearly 1 in 3 people do not have access to food and clean drinking water. Depletion of food security leads to '*hidden hunger*' where malnutrition and micronutrient deficiencies co-exist. There are over 2 billion people globally who are deficient in micronutrients such as Vitamin A & B, Zinc, Folate, and Calcium. These micronutrients (vitamins and minerals) are essential to normal development and growth.

Climate change will exacerbate these conditions if substantial action is not undertaken immediately. All the four dimensions of food security highlighted by the Food and Agriculture Organization, which entail availability, accessibility, utilization, and stability will consequently be disrupted.

Berry et al. (2018) examined the direct and indirect pathways of climate change on health (**Fig. 1**). Direct pathways include increased weather variability and changes in frequency and intensity of extreme weather impacts due to CC. These lead to increased exposures to disease vectors and several types of pollution. Direct pathways also have consequences for fluctuations in the nutrition values of crops. Indirect pathways include food and price changes, water availability, displacement, and migration of communities. Additionally, climate change adversities contribute to existing social infrastructure and inequities of age and gender, public health infrastructure, mobility, socioeconomic and health and nutrition status as well as inequities in hygiene and social cohesion.

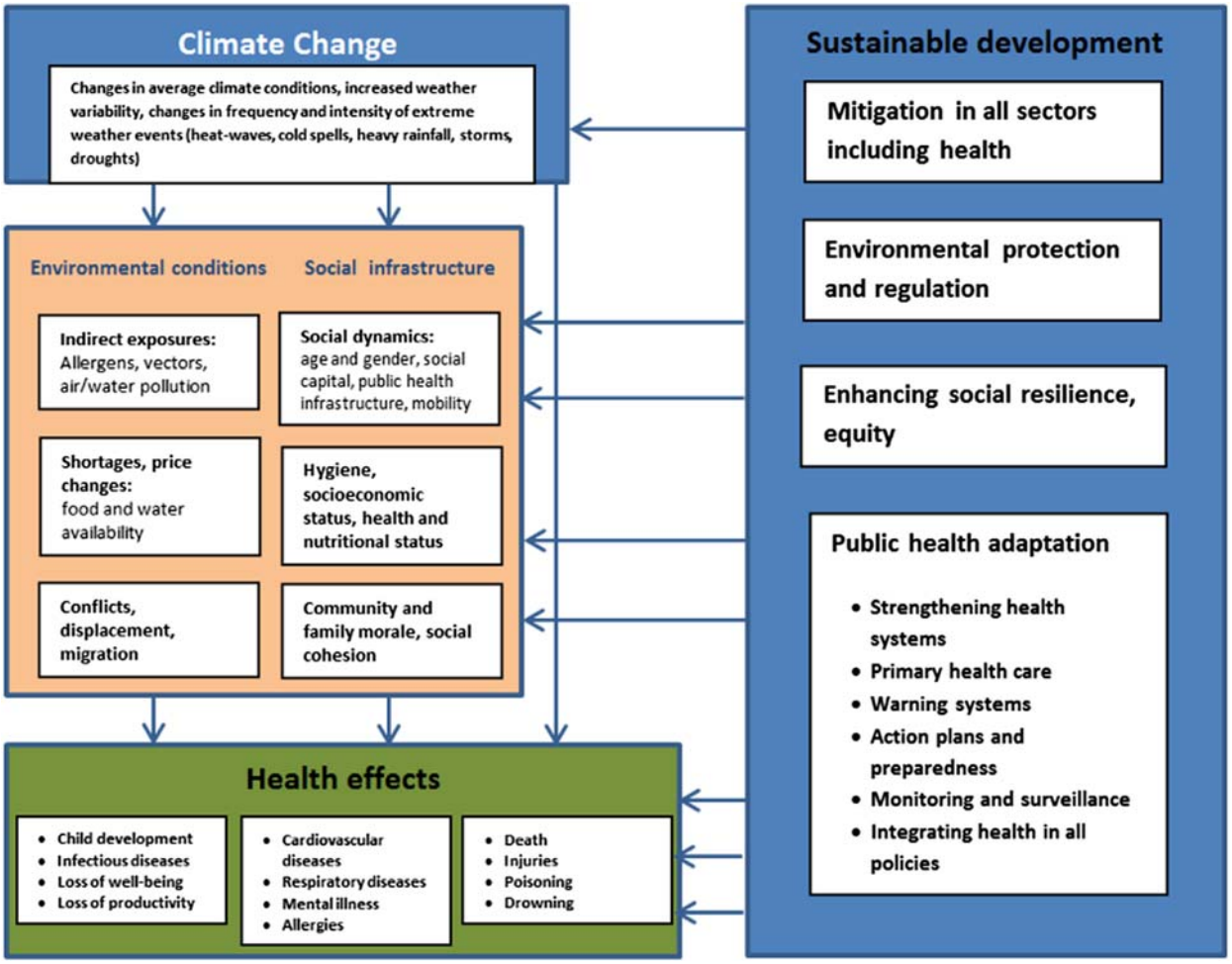


Fig. 1 Pathways of climate change, sustainable development, and health. Source: Berry et al. (2018).

Together the direct and indirect pathways lead to health impacts in human populations including rise in infectious diseases, child development, loss of well-being, loss of productivity, allergies, poisoning, and death. Many of these impacts can be mitigated through building action-oriented mitigation strategies, enhancing social equity, and strengthening public health adaption.

The article proceeds as follows. The next section assesses the impacts of climate-led adversities on health, nutrition, and food systems. The third section discusses impact of climate change on agricultural pathways in LMICs. The final section looks at climate-resilience approaches and policy interventions to mitigate the impacts of climate change on health and nutrition.

Impacts of climate-led adversities on health, nutrition & food systems

Food systems are anticipated to modify extensively due to changing climatic conditions. To begin with, climate change adversely affects communities and areas heavily dependent on agricultural produce. In Africa alone, agricultural yields could dwindle by 30% by 2050 due to climate change, directly impacting the livelihoods of over 70% of Africa's total population. Vulnerable communities dependent on agriculture will not be able to cope with the rising food prices due to global inflation and be further pushed into poverty traps. Furthermore, many agricultural communities rely on rainfall patterns and ideal climate scenarios to produce staple and cash crops. Climate change will affect rainfall patterns and magnify the unpredictability of weather conditions. These will have impacts on food availability, access, utilization, and stability (Fig. 2). Overall disruptions will affect agricultural economies in LMICs like India, but also some high-income countries including Canada, Australia, and the USA.

Food availability

Potential impacts of CC include increased incidence of famine and weather-related disasters due to unexpected weather conditions. Over 54 million people in Ethiopia, Madagascar, South Sudan, and Yemen are currently encountering famine-like conditions as detrimental impacts of climate change interact with existing hunger, food shortages and conflict conditions. About 50% of total

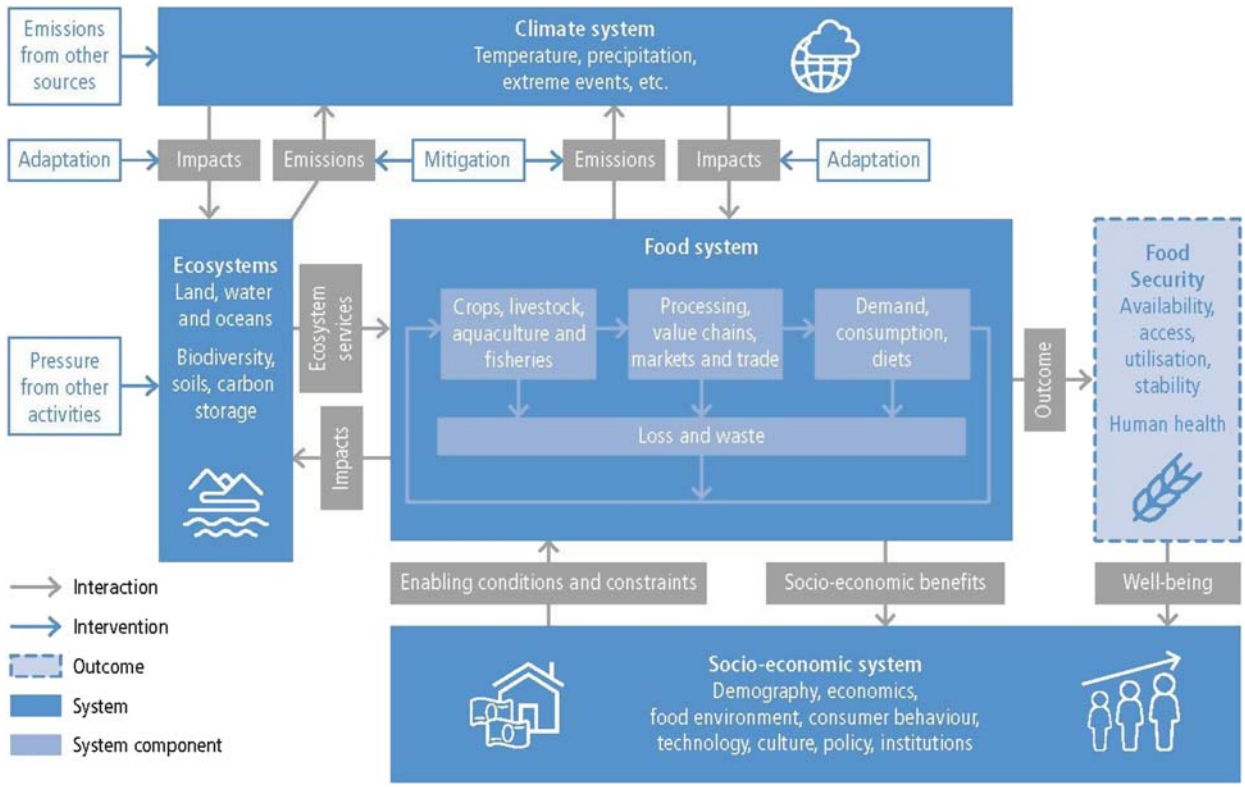


Fig. 2 Interlinkages between the climate system, food system, ecosystems (land, water and oceans) and socio-economic system. Source: (Mbow et al., 2019)

crop production globally comes from forest and mountain areas, rainfed and irrigated agriculture in dry lands accounts for 25% and rice in coastal areas forms only 12% of total crop produced globally. Environment factors like precipitation, soil dampness, temperature and radiation will compromise yields and dwindle food availability. India's heat wave in 2022 coupled with water scarcity in Brazil, the US and Europe due to climate change is predicted to have dire impacts on the supply chain of wheat across the World.

Food affordability

Poor households in non-farming occupations rely on agricultural produce, market conditions, and the food supply chain for sustenance. Food storage will impact buying and trading prices in the market hereby, exposing poor households to further vulnerabilities. For example, in Southeast Asia, there has been a 6% increase in the costs of food due to extreme weather conditions. An 80% increase in food prices in Indonesia alone is predicted as the Government continues to push for net-zero by 2050.

Food accessibility

Changes in farming and non-farming livelihoods limit the ability of the household to purchase food. Yield reductions will have an impact on the accessibility of the food in the market. This will reverse decades of progress made on food security in low-income countries and LMICs. Even within safety nets, food is dispensed through market actors and non-market circulation. In market circulation, those who receive food through safety nets are contingent upon income, remuneration, assets and the ratio of the cost of a minimum daily food basket to the average daily income. Social and political factors are key determinants in food allocation. In non-market circulation, food is received through social processes and reciprocal relations within agricultural families. If agricultural production declines (1) there will be lesser food to circulate through safety nets and, (2) agricultural families might take up other professionals. Locally produced foods and dispersion within families might change or end. Finally, power dynamics will also decipher who gets more food within households, which will be the patriarch in most cases. CC impacts on food systems will spike effects on food access for women and children.

Nutritional value of food

In addition to social and economic consequences, CC also affects the quality of food of the population. For example, elevated carbon dioxide concentrations in the air have been shown to reduce zinc concentrations (3%) and iron concentrations (5%) in

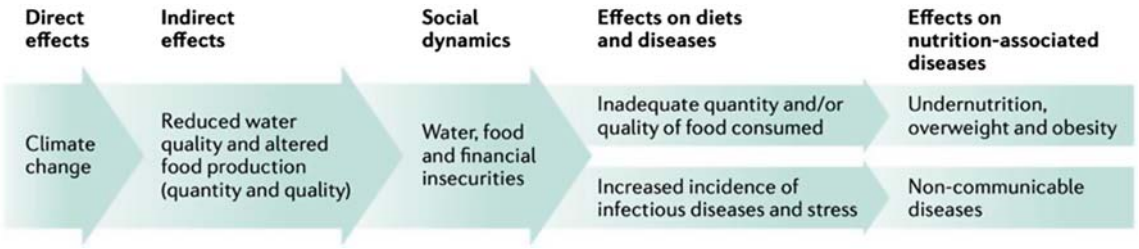


Fig. 3 The main ways in which climate change influences diet and nutrition-associated diseases. Source: Fanzo and M. Downs (2021).

rice grains (Semba et al., 2022). In wheat grains, the combination of elevated carbon dioxide and temperature increased cadmium and lead (Semba et al., 2022). Fluctuations in nutritional properties of plants increase susceptibility to micronutrient deficiencies.

CC has profound impact on malnutrition. Fanzo and Downs (2021) have previously highlighted the relationship between CC and nutrition-associated diseases (**Fig. 3**). Climate change augments malnutrition in four ways. Firstly, rising food prices will impede the capacities of poor households in buying nutritious foods. This will have dire impacts on household food insecurity, overall calorie consumption and intake of unhealthy foods. Secondly, women and young mothers perform manual labor in agricultural landscapes across the world. Disruptions in agricultural systems and food security impact infant and child feeding practices, triggering early onset of childhood stunting & micronutrient deficiencies. Thirdly, nutrient absorption is contingent upon the process of consumption of nutritious foods. Poor water and sanitation qualities due to CC increase the chances of diarrhea and other arthropod infections and tick vectors. For instance, researchers in Canada and Sweden have traced the increasing incidence of Lyme disease to expansion of tick vectors due to climate change. These regions were earlier prevented to cold climate conditions. Regional warming has dispersed migratory birds, which are hosts of the tick vectors of Lyme disease. Prolonged illnesses will lead to the underutilization of essential nutrients in the body. Finally, CC also affects diet-related non-communicable diseases, such as diabetes mellitus and cardiovascular diseases (Fanzo and Downs, 2021). This is especially true in high-income countries where reductions in vegetable and fruit intake have been attributed as the primary reason of NCDs (Fanzo and Downs, 2021).

Impacts of climate-led adversities on environmental pathways

Agriculture

Plants are highly sensitive and dependent on abiotic factors (such as temperature, humidity, salinity, CO_2 concentration, etc.) and biotic factors (such as pollinators) for their optimal growth. Any deviation from these ambient conditions which surpasses a critical threshold will adversely impact plant growth and reduce crop yield. This makes Agro-ecosystems one of the most vulnerable targets of climate change, with the capacity to unfold into deleterious consequences both on people's livelihoods and on global food security. The world population is projected to reach over 9.7 billion people by 2050, while agricultural land remains not only finite but could become non-arable due to soil erosion, desertification, soil salinity, etc. Ensuring food security for this population in the face of climate catastrophe is going to be a major challenge if the current adversities aren't subdued.

Agriculture is the main source of food and income for people around the globe. It alone sustains livelihoods for 65% of the world's poor. Therefore, any impact that climate change has on agro-ecosystems is directly translated into impacts on food security, considering it the main source of dietary sustenance. When food insecurity increases, more and more people will have less access to affordable, healthy food, which in turn will affect their health by depleting them of adequate nutrition. Climate change exacerbates food insecurity primarily by destroying food systems in climatic events such as heatwaves, droughts, cyclones, etc. Between 2008 and 2018, 49 billion USD was lost in Asia alone as a result of declined agricultural and livestock produce. Secondly, elevated CO_2 levels, the main greenhouse gas responsible for climate change, along with increasing global ambient temperatures, dually act together to alter the nutritional content of food and reduce crop yields (Myers et al., 2017) CO_2 on its own increases yields for plants, and together with higher temperatures may lead to high-latitude regions such as North America & Europe with enhanced production of food due to the lengthening of the growing season. However, results will mostly be negative elsewhere as the increasing ambient temperatures override the fertilizing effect of CO_2—especially in Africa and India where yields may decline up to 5%.

Depending on which metabolic process they use for carbon fixation during photosynthesis, plants are classified as C3 plants or C4 plants. C3 plants thrive in a colder, wet climate while C4 plants prefer hot and dry climates. C3 plants are the source of over 90% of food crops whereas C4 plants constitute only about 1%. At higher temperatures, plants produce more carbohydrates, which can reduce up to 8% of their mineral content (Semba et al., 2022). Due to their abundance as food sources, the vulnerability of C3 plants to climate change is concerning. They provide key sources of Zinc and Iron to people who suffer from micronutrient deficiencies. Yields of C4 plants such as maize have reduced up to 5% since 1980, however, they have no additional increase in yield due to elevated CO_2 levels (Myers et al., 2017). Elevated CO_2 levels have been associated with increasing the productivity of C3 crops (Myers et al., 2017). C3 crops rice and wheat are staple crops in South Asia—which is the world's largest consumer and producer of these two crops globally. Increasing temperatures reduce the yield of both these crops. In rice, elevated CO_2 may lower

the quantity of Zinc and folate, while in wheat it reduces Zinc and folate content by 9% and 5–7% respectively. Legumes, which are another major source of protein and micronutrients and are highly sensitive to heat, have reduced yields due to the dual action of elevated CO_2 and temperatures (Myers et al., 2017). The protein content of C3 plants (excluding legumes and C4 plants) is also reduced under conditions of elevated CO_2.

Food sources declining in their mineral and iron content will have a major blow on micronutrient deficiencies, which is a major public health concern across all South Asian regions. Zinc deficiency predominates in the region, especially among children and pregnant and lactating mothers. While the deficiency is largely due to lowered uptake, studies have correlated lowered uptake of zinc due to consumption of plant diets rich in phytic acid, which lowers the bioavailability of zinc. Folate deficiency is predominantly associated with anemia, with a 48% prevalence rate in pregnant women across South Asia. Crops associated with reduced micronutrient contents pose a distressing hazard to exacerbating micronutrient deficiencies across these vulnerable populations. While changes in the climatic change are going to make lands unsuitable to produce various crops and plants, these same environments will become more conducive for growing others. Shabani et al. (2016) propose that lands will become more arable for growing the date palm (*Phoenix dactylifera*), which is an excellent and economical source of iron, zinc, calcium, potassium and boron. Intake of about 100 g of date flesh daily meets about 50% of the recommended micronutrient intake. Dates offer an excellent solution to curbing malnutrition and micronutrient deficiencies in food-insecure populations such as regions of Sub-Saharan Africa.

Higher temperatures encourage the formation of ozone (Myers et al., 2017), a dangerous toxin which impedes photosynthesis by minimizing the stomatal aperture and reducing plant yield. Estimates for crop yield losses due to ozone pollution may be worth 35 billion USD by 2030. A warmer climate also means prolonged breeding seasons for pests and insects like aphids, which result in loss of agricultural produce. Aphids survive better in hotter temperatures and elevated CO_2 levels. This will mobilize farmers to use chemical insecticides and carcinogenic pesticides on their produce, which can have a deleterious impact on humans if the residues of pesticides are not washed properly, enter inside the body.

Apart from agriculture, CC negatively affects livestock, which is a source income, but also protein (33%), calories (17%), and micronutrients like vitamin B12, iron and calcium. Increased ambient temperatures cause animals to face great heat stress. Reproductive capability is reduced in cattle as fertility declines due to heat stress, which also compromises the quality and quantity of meat and milk production. Increasing temperature also makes farm animals more prone to vector-borne diseases like Rift Valley Fever, Bluetongue & West Nile Virus. Due to reduced precipitation and drought, water bodies dry up. With only a few limited water bodies, animals flock around common water bodies, which increase their risk and exposure to these diseases.

People facing climate shocks, especially in drought-stricken areas, turn to livestock for consumption. As agricultural produce becomes scarcer; there will be growing dependence on meat consumption and processed food. Meat consumption is expected to grow in the largest proportion across the developing world. While meat is an excellent source of protein and Vitamin B12, livestock is a major contributor to GHGE accounting for over 14.5% of emissions. As native or indigenous diets will evolve to include more meat and processed food which are rich in saturated fats, cholesterol, salt, sugar and preservatives, the risk of non-communicable diseases like cardiovascular diseases, hypertension, and Type 2 diabetes increase substantially.

The last fifty years have witnessed a five-fold increase in the occurrences of natural disasters like cyclones, floods, storms, etc. These events can be attributed as being a direct manifestation of climate change since the earth's accelerated warming has disturbed the hydrological cycle by intensifying evaporation and precipitation patterns globally. In response, rains have become heavier and shorter as opposed to lighter and more consistent. Increasing temperatures cause the air to uptake more water vapor, leading to shorter & heavier downpours with the potential to cause flooding.

Floods and droughts

Floods are one the most devastating natural calamities, affecting millions of lives annually. In 2019, the UN had reported floods killed at least 600 people and affected the lives of 25 million people across South Asia. Apart from the devastation they cause to human life, livelihoods and infrastructure, floods are also accompanied by a surge in cases of infections and disease. Heavy precipitation and floods are reported to be the primary antecedents during outbreaks of infectious diseases. Floodwaters cause the spread of infectious diseases by contaminating food and drinking water. Floodwaters overwhelm water-treatment plants and cause drains to overflow, leading to the intermixing of sewage water with local water bodies from where it reaches households. The provision of safe drinking water is pertinent during floods. Consumption of contaminated water leads to the spread of gastrointestinal diseases like typhoid, cholera and hepatitis A.

Unhygienic and overcrowded spaces expose people to these diseases, and infections spread rapidly, with waterborne diseases being the most common risk to health following floods in Asia. Leptospirosis is another common infection that is spread when people, especially children, come in contact with animal urine mixed in floodwater. The surge in cases of trypanosome infection leishmaniasis was seen after flooding across Bangladesh, Pakistan and India (Bihar). The stagnant floodwater water also acts as the ideal breeding ground for vectors like mosquitoes which spread malaria and dengue fever, which are already a chief public health concern across these regions. Runoffs from agricultural processes containing carcinogens like nitrates wash away with the floodwaters, as well as other heavy metals that may wash away and expose groups to an increased risk for cancer.

Another danger to life comes from the rising global mean sea levels caused by the melting ice caps and thermal expansion of water; threatening to submerge low elevation coastal zones underwater. This is deeply concerning as most populations in Low Elevation Coastal Zone (LECZ) are rural poor.

South Asia has a very large rural LECZ population, with over 93 million poor and vulnerable people at risk for this adversity. These inundations increase the soil salinity, making it unfit for cultivation as the saltwater intrusion reduces crop outputs. With inundations becoming increasingly common, several farmers across Bangladesh have switched from agriculture to aquaculture. Consumption of saline water has led to a surge in cases of preeclampsia and gestational hypertension in pregnant women across Bangladesh, with the country experiencing a 28% increase in cases since 2012.

Drought is another aspect of a perturbed hydrological cycle. Climate change intensifies precipitation and evapotranspiration variability, exacerbating drought conditions. Between 1970 and 2019, droughts have claimed the lives of 650,000 people globally. Prolonged droughts can manifest into famine. Severe famines are classified under IPC 5 (Integrated Food Security Classification) when over 30% of a population experience acute malnourishment, and every 2/10,000 people are dying from starvation. The beginning of 2022 marked the start of the worst drought the Horn of Africa has witnessed in over 40 years, with over 10 million children experiencing malnutrition and thirst. This drought can worsen further into a famine. Droughts inflict complete climate shocks on people. Apart from the central breakdown of food security, droughts destroy livestock, force people to migrate for survival and exacerbate social problems like dropping out of school, child brides, and conflict. Globally, 80% of agriculture is rain-fed, and of the 1.2 billion hectares of croplands watered by rainwater, 128 million hectares have been experiencing increased drought frequency. Erratic precipitation and drought have had harrowing impacts on smallholder farmers, who incur heavy monetary losses, and have their livelihoods collapsed. Farmers are usually able to cope with seasonal variability, but climate change will offset their scope for adaptation. Droughts will increase the dependency of countries on food donations, international aid and humanitarian assistance.

Droughts can dry up groundwater, the main source of freshwater accounting for 30% of all freshwater reserves. South Asia is the world's largest user of groundwater, with India, Bangladesh and Pakistan deriving most of its water for irrigational purposes, drinking purposes and livestock usage from the Indo-Gangetic Plain aquifers. Drought also concentrates pollutants in water bodies as they dry up. As ecosystems are continuously lost with deforestation, wildfires and urbanization, disturbances to the food web are leading to biodiversity loss.

Biodiversity & pollinator loss

Loss of habitat, human interference and climate change are all accelerating species extinction by 1000–10,000 times more than the natural rate. Recent years have witnessed a decline in insect and animal pollinators. The pollinator-plant relationship has become unsynchronized due to climate change, as plants are now flowering and maturing earlier than their due timings. Flowers that contain pollen have responded to elevated temperatures by modifying the amount of UV pigment present in their petals, making them difficult to distinguish by their pollinators. Empirical evidence that pollinator loss worsens micronutrient deficiencies was presented in studies by Ellis et al. (2015). This is concerning as pollinators are associated with plants that provide 40% of nutrition globally. Pollinator loss, therefore, can further exacerbate the scope for micronutrient deficiencies and non-communicable diseases for populations that derive their nutrient sources from these plants. There has been a decline in the nutritional value of pollen (Myers et al., 2017). Over 150 nutrient-dense plants rich in Vitamin A & C, iron and folate are reliant on animal pollinators (Myers et al., 2017), and any disturbance to their pollinators has effects on the health and nutrition of individuals that consume them. The studies conducted by Ellis et al. (2015) also showed that Vitamin A deficiency was the most sensitive to pollinator loss in associated countries, as most fruit and vegetable sources of Vitamin A were heavily dependent on pollinators.

Climate resilience approaches: adaption and mitigation as the way forward

As the effects of climate change become more and more pronounced on human health and our food systems, there is grave urgency to act immediately and curb climate change at its root cause and address its ill-effects by integrating both climate mitigation and adaptation. Climate mitigation requires us to massively cut down and stabilize our CO_2 and other GHG levels and to sequester carbon. This entails making changes to our agriculture, livestock, and food transport systems by adopting different techniques to help decrease the levels of greenhouse gases emitted. Agriculture alone is responsible for emitting 30% of global GHGEs. Climate adaptation strategies aim at strengthening our existing food systems and infrastructure of agro-ecosystems to make them more resilient to the effects of climate-led adversities. In our current capacity, Climate adaptation is crucial at buffering the impacts of climate change on human health and nutrition, and to help mitigate the consequences it has on the most vulnerable populations.

Adaptation approaches for food systems

There are several ways to enhance our agricultural outputs, both in terms of better yields and nutritional value. This can be achieved by artificially engineering climate-resilient varieties or utilizing farming techniques which have been associated with better agricultural and environmental results. Research and development of such varieties and breeds must be promoted to improve food and nutrition security.

i Biofortification of food crops– Fortification of foods helps in increasing contents of beta carotene (Vitamin A), Zn, Fe, Ca, etc. It is a relatively easy, inexpensive and climate-friendly pathway to combat micronutrient deficiencies and offset the impacts of

climate change on food insecurity as well as provide monetary sustenance to small farmers. Biofortified crops have contributed to modest enrichment of nutrient content in the long run, and are well accepted by the public. They require dissemination and awareness about their benefit and have been shown to reduce neonatal and child mortality. Pearl millet and maize, C4 crops with relatively low nutritional value. Pearl millet was biofortified to increase its iron and essential amino acids, and varieties have been grown across India and the Republic of Niger. Provitamin A enriched Maize, hybridized to be drought and heat resilient, has been introduced in 11 countries. The article on " Biofortification: A primer on nutrient enriched crops" by Mbuya and colleagues in this Encyclopedia provides elaborate insights into the Biofortification and nutrition.

ii. Use of improved cultivars—Short-duration cultivars such as early maturing pigeon pea, or cultivars which are saline or drought-resistant varieties. In the Indian state of Maharashtra, early maturing and drought-tolerant cultivars of green gram (BM 2002-1), chickpea and pigeon pea (BDN-708) yielded outputs 20–25% higher than the indigenous varieties.

iii. Vermicomposting—It is an excellent source of natural fertilizer for sustainable agriculture. Studies from Maharjan and Pradhanang (2017) found that vermicompost prepared with vegetable scraps the red worm *Eisenia fetida* has high concentrations of Potassium, Phosphorus and Nitrogen, which are the 3 major macronutrients needed by plants for growth.

iv. Employing traditional farming practices—Crop diversification and crop rotation both help boost soil health and ensures greater availability of nutrients. Diversifying crop rotations is linked to higher drought resilience and improved soil-water retention. Crop rotations have further shown to reduce climate change-induced yield losses for temperate cereals—both for winter and spring varieties.

v. Agroforestry—Incorporating trees on farmlands providing an array of benefits including reforestation, improved productivity and better soil quality, better socio-economic opportunities and better health. They can aid climate change mitigation by sequestering atmospheric Carbon both in the underground (roots) and above ground (leaves, stems, branches) systems.

vi. Use of hydroponics—This technology grows plants without the use of soil in strictly controlled environments, and presents as an innovative (albeit expensive) method to meet food requirements for the growing population. Hydroponics use up to 90% less water compared to traditional farming methods, and has shorter and faster yield time. It can help feed more people without exhausting resources, since it utilizes very little quantity of raw materials.

vii. Sustainable diets—Promoting the transition to more sustainable diets cater not just to covering all aspects of human health and nutrition, but are also environmentally responsible. They promote affordable, accessible, nutritionally-adequate and low-environmental impact foods. Consumption patterns of High-income countries, which consume great amounts of red meat, are responsible for emitting 41% more emissions than countries LMICs. Switching to plant-based diets will reduce diet-related NCDs. Planetary health diets proposed by the Eat—Lancet commission are associated with reducing premature mortality by 34% and contribute to a reduction of GHGE by threefold (Fanzo and Downs, 2021).

It is important for us to attain nutritional security in the current scenario in an environmentally sustainable manner, apprehending any further climate degradation. Gustafson et al. (2016) formulated an assessment methodology called 'Sustainable Nutrition Security' (SNS). This methodology incorporates 7 metrics based on several indicators relevant to safeguarding nutritional outcomes for sustainable food systems and evaluates impacts of interventions, scored between (0–100). These metrics include food nutrient adequacy, ecosystem stability, food affordability and availability, sociocultural wellbeing, food safety, resilience and waste and loss reduction. The metrics can be used to influence and asses decision-making for food systems both in developed and developing countries.

viii. Drastic food alternatives—These could include uncommon food choices such as entomophagy. Insects are already consumed by over 2 billion people globally and have high protein and micronutrient content. Algae-based foods such as spirulina supplements are nutrient-rich, have been shown to provide essential nutrients. These however may not be acceptable to all consumers.

Managing risks of natural disasters by improving natural and man-made infrastructures

Damages caused by floods and droughts to human lives, and its consequent role in catalyzing the spread of diseases can be limited by improving our current infrastructures. Better urban designing and landscaping can withhold these impacts to an appreciable degree.

i. Revival and protection of marshlands, mangroves, and wetlands—Mangroves and wetlands are biodiverse ecological systems that store CO_2 in their aquatic flora and provide sources of food and natural resources. These ecosystems are critical to protecting shorelines from flooding due to their water retention capabilities, catching sediments and reducing erosion. In Bangladesh, where several community-based mangrove restoration projects have been implemented, fossil fuel emissions were offset by 1.5% in 2014, and villages protected by mangroves faced half the monetary loss during cyclone *Sidr*.

ii. Efficient Sewage systems—Sewage systems need to be better planned and built to be able to withstand floods, so that water so they don't overwhelm and spill out into the floodwaters.

iii. Storing water to use in times of droughts—Installing watersheds to store water, and rainwater harvesting in areas prone to drought is a great way to tackle water scarcity. In the drought-prone district of Purulia in West Bengal, Roy (2014) proposed micro water sheds and rainwater harvesting to help curb the district's water scarcity. Run-offs from the Bandu Basin during the monsoon months could be stored and harvested to meet water needs and recharge groundwater during the drought months.

iv. Better surveillance and monitoring systems to disseminate early warnings about upcoming floods/droughts—Monitoring bodies need to be provided with real-time data on moisture levels, soil water, etc. to spot irregularities in climatic patterns that forewarn the arrival of a flood or drought. Such warnings facilitate the early evacuation of people, as well as strategize mitigation and resilience plans. The South Asia Drought Monitoring System (SADMS) was developed by the CGIAR in 2014 and provides a weekly map of drought indices. Sri Lanka in particular has gained great benefit from this program, such as during the 2014 droughts.

Political and policy-bases interventions

Political commitment is critical to mitigating climate change and adapting to its adversities. Countries must take efforts to realize their Nationally Determined Contributions (NDC's) stipulated during The Paris Agreement, 2015 to achieve the 30% in global GHGEs.

Decarbonization of energy systems was a critical discourse of the Paris Agreement. Subsidies on fossil fuel usage need to be discarded, and renewable sources of energy must be pushed for. Most developing countries heavily depend on fossil fuels due to their growing populations and have a lot of scope for switching over to renewable sources such as wind power, solar power, etc. Sub-Saharan African countries have the capacity to obtain 100% of their energy requirements through renewable sources by 2050. Renewable sources of energy have high initial costs in installation, and some are expensive to upkeep as well. However, these expenses are reimbursed by great economic and health turnovers in the long run.

Policy-based interventions aim at addressing food security in the face of climate change using frameworks to regulate food supply, access and utilization. These interventions also provide incentives, which deliver both nutritional and economic benefits.

i. Reducing food wastage—Enough food is produced to feed everyone across the world. However, 1/3rd or about 1.3 billion tons of all food produced is wasted, and is responsible for emitting 8% of the world's GHGE. Saving just a quarter of this food could feed an additional 870 million people. Policies need to limit this wastage and redistribute surplus production to communities that need it the most.
ii. Regulating fisheries by halting fishery subsidies—Fisheries are sensitive ecosystems and their overexploitation has overwhelmed them. These have resulted in reduced fish catch, with over 90% of the world's marine stock overexploited. By 2050, projections estimate an 88% decline in global fish stocks (Myers et al., 2017). Policies need to encourage and ensure sustainable fishing and aquaculture practices, and impose bans on fishing in endangered zones.
iii. Improving agricultural subsidies—Springmann and Freund (2022) call for inlining agricultural subsidies to address health and climate change concerns. They found that redirecting agricultural subsidies to healthier foods had positive impacts on health and environmental outcomes, albeit with an economic trade-off.
iv. Social Protection Programs—These include Safety Net programs, food assistance, work for food and cash programs and nutrient supplementations to facilitate food and nutritional safety for vulnerable families during climate shocks. These programs are pertinent in protecting these groups, particularly women and children, from malnutrition, disease and displacement.
v. Fostering partnerships between local communities and governments—Governmental support and initiatives offered to local communities can help bolster their food and nutritional security. By working together, they are able to better plan solutions related to food systems and implement inter-sectoral collaboration.
vi. Microfinance Institutions—Microfinancing for smallholder farmers, especially women farmers is pivotal in increasing their productivity by proving them with credit. Such institutes also act as intermediaries for implementing climate adaptation by integrating climate change mitigation within their activities.

References

Berry, P., Enright, P., Shumake-Guillemot, J., Villalobos Prats, E., Campbell-Lendrum, D., 2018. Assessing health vulnerabilities and adaptation to climate change: a review of international progress. Int. J. Environ. Res. Publ. Health 15, 2626. https://doi.org/10.3390/ijerph15122626.

Ellis, A.M., Myers, S.S., Ricketts, T.H., 2015. Do pollinators contribute to nutritional health? PLoS One 10, e114805. https://doi.org/10.1371/journal.pone.0114805.

Fanzo, J.C., Downs, S.M., 2021. Climate change and nutrition-associated diseases. Nat. Rev. Dis. Prim. 7, 90. https://doi.org/10.1038/s41572-021-00329-3.

Gustafson, D., Gutman, A., Leet, W., Drewnowski, A., Fanzo, J., Ingram, J., 2016. Seven food system metrics of sustainable nutrition security. Sustainability 8, 196. https://doi.org/10.3390/su8030196.

Maharjan, K., Pradhanang, S., 2017. Nutrient Status (NPK) of Vermicompost prepared from Vegetable market waste and Tea waste using *Eisenia fetida* (Red worm). J. Environ. Sci. 3, 118–121.

Mbow, C., Rosenzweig, C.L., Barioni, G.T., Benton, G.T., Herrero, M., Krishnapillai, M., Liwenga, E., Pradhan, P., Rivera-Ferre, M.G., Sapkota, T., Tubiello, N.F., Xu, Y., 2019. Food security. In: Benkeblia, N., Challinor, A., Khan, A., Porter, R.J. (Eds.), Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems. Intergovernmental Panel on Climate Change (IPCC), p. 441. Chapter 5. https://www.ipcc.ch/site/assets/uploads/2019/11/08_Chapter-5.pdf, 2019. (Accessed 16 June 2022). In press.

Myers, S.S., Smith, M.R., Guth, S., Golden, C.D., Vaitla, B., Mueller, N.D., Dangour, A.D., Huybers, P., 2017. Climate change and global food systems: potential impacts on food security and undernutrition. Annu. Rev. Publ. Health 38, 259–277. https://doi.org/10.1146/annurev-publhealth-031816-044356.

Roy, A., 2014. Viability of rainwater harvesting in drought prone areas of West Bengal: an empirical study on Bandu river basin in puruliya district. Valley Int. J. 1, 155–164.

Semba, R.D., Askari, S., Gibson, S., Bloem, M.W., Kraemer, K., 2022. The potential impact of climate change on the micronutrient-rich food supply. Adv. Nutr. 13, 80–100. https://doi.org/10.1093/advances/nmab104.

Shabani, F., Kumar, L., Nojoumian, A.H., Esmaeili, A., Toghyani, M., 2016. Projected future distribution of date palm and its potential use in alleviating micronutrient deficiency: projected future distribution of date palm. J. Sci. Food Agric. 96, 1132–1140. https://doi.org/10.1002/jsfa.7195.

Springmann, M., Freund, F., 2022. Options for reforming agricultural subsidies from health, climate, and economic perspectives. Nat. Commun. 13, 82. https://doi.org/10.1038/s41467-021-27645-2.

Further Reading

Baqir, M., Sobani, Z.A., Bhamani, A., Bham, N.S., Abid, S., Farook, J., Beg, M.A., 2012. Infectious diseases in the aftermath of monsoon flooding in Pakistan. Asian Pac. J. Trop. Biomed. 2, 76–79. https://doi.org/10.1016/S2221-1691(11)60194-9.

Food and Agriculture Organization of the United Nations, 2013. B2 - 1 Livestock Production and Climate Change | Climate Smart Agriculture Sourcebook [WWW Document]. URL. https://www.fao.org/climate-smart-agriculture-sourcebook/production-resources/module-b2-livestock/chapter-b2-1/en/. (Accessed 9 June 2022).

Food and Agriculture Organization of the United Nations, 2021. The State of Food Security and Nutrition in the World 2021. https://doi.org/10.4060/CB4474EN [WWW Document]. www.fao.org.

Food and Agriculture Organization of the United Nations, 2020. The State of Food and Agriculture 2020. FAO. https://doi.org/10.4060/cb1447en.

Gornall, J., Betts, R., Burke, E., Clark, R., Camp, J., Willett, K., Wiltshire, A., 2010. Implications of climate change for agricultural productivity in the early twenty-first century. Philos. Trans. R. Soc., B 365, 2973–2989. https://doi.org/10.1098/rstb.2010.0158.

Marini, L., St-Martin, A., Vico, G., Baldoni, G., Berti, A., Blecharczyk, A., Malecka-Jankowiak, I., Morari, F., Sawinska, Z., Bommarco, R., 2020. Crop rotations sustain cereal yields under a changing climate. Environ. Res. Lett. 15, 124011. https://doi.org/10.1088/1748-9326/abc651.

Myers, S.S., Zanobetti, A., Kloog, I., Huybers, P., Leakey, A.D.B., Bloom, A.J., Carlisle, E., Dietterich, L.H., Fitzgerald, G., Hasegawa, T., Holbrook, N.M., Nelson, R.L., Ottman, M.J., Raboy, V., Sakai, H., Sartor, K.A., Schwartz, J., Seneweera, S., Tausz, M., Usui, Y., 2014. Increasing CO2 threatens human nutrition. Nature 510, 139–142. https://doi.org/10.1038/nature13179.

Smith, A.C., Tasnim, T., Irfanullah, H.M., Turner, B., Chausson, A., Seddon, N., 2021. Nature-based solutions in Bangladesh: evidence of effectiveness for addressing climate change and other sustainable development goals. Front. Environ. Sci. 9, 737659. https://doi.org/10.3389/fenvs.2021.737659.

Srinivasarao, 2021. Climate Resilient Agriculture Systems: The Way Ahead [WWW Document]. URL. https://www.downtoearth.org.in/blog/agriculture/climate-resilient-agriculture-systems-the-way-ahead-75385. (Accessed 17 June 2022).

Tirado, M.C., Crahay, P., Mahy, L., Zanev, C., Neira, M., Msangi, S., Brown, R., Scaramella, C., Coitinho, D.C., Müller, A., 2013. Climate change and nutrition: creating a climate for nutrition security. Food Nutr. Bull. 34, 533–547. https://doi.org/10.1177/156482651303400415.

Wheeler, T., von Braun, J., 2013. Climate change impacts on global food security. Science 341, 508–513. https://doi.org/10.1126/science.1239402.

World Economic Forum, 2018. Nearly 90% of Fish Stocks Are in the Red – Fisheries Subsidies Must Stop [WWW Document]. World Econ. Forum. URL. https://www.weforum.org/agenda/2018/07/fish-stocks-are-used-up-fisheries-subsidies-must-stop/. (Accessed 17 June 2022).

Dietary guidelines, international perspectives

B. Schneeman, University of California, Davis, CA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of B. Schneeman, Dietary Guidelines, International Perspectives, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 60-64, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00074-X.

Introduction 379
Historical background 379
Types of guidelines 381
The development of FBDG 381
Conclusion 383
References 383
Relevant websites 384

Key points

- Food-based Dietary Guidelines (FBDG) address both nutrient adequacy and dietary factors that reduce risk of diet-related chronic diseases.
- FBDG express nutrition principles in the context of foods and dietary patterns.
- Evidence-based FBDG are based on data analysis, food pattern modeling, and systematic reviews of peer-reviewed literature.
- FBDG should reflect cultural preferences and foods available within a region or country.

Introduction

Since the late 1970s, food-based dietary guidelines (FBDG) have become a cornerstone of nutrition policy in many countries and provide the basis for education to improve nutritional status. FBDG translate nutrient requirements and dietary recommendations to reduce the risk of diet-related chronic diseases into quantitative recommendations for foods to include in healthful dietary patterns. Each country must adapt FBDG to the needs of their population both in terms of the public health issues to be addressed as well as the foods available and relevant dietary strategies for the population (FAO and WHO, 1998). Across various countries several themes in FBDG have emerged and include energy balance, a healthful variety of foods, including fruits and vegetables, whole-grain products, food sources of protein, calcium, and unsaturated fatty acids and safe food handling as well as appropriate levels of physical activity. Cautionary messages focus on excess energy intake, saturated and *trans*-fatty acids, added sugars, salt, and alcohol (Fischer and Garnett, 2016).

Historical background

Throughout the ages, religious and philosophical writings have included dietary advice, which is reflected in an oft-quoted line from Hippocrates: "Let thy food be thy medicine." In the late 1800s, modern science began to influence the nature of recommendations regarding foods and beverages. The original focus of guidelines developed from Pasteur's discoveries of the disease-causing organisms that could be present in foods emphasized sanitation in food handling. In the early 1900s, the discovery of vitamins and minerals led to the realization that foods contain factors that are essential for health. This 'vitamin theory of disease' led to research throughout the first half of the twentieth century to discover these factors and determine their essential functions in the treatment of deficiency diseases. The knowledge that food was important in the prevention of diseases that were major public health problems, such as scurvy, beri-beri, night blindness, and pellagra, led to early efforts to develop and promote dietary recommendations or guidelines, even when the specific curative factors in foods had not yet been identified. Among the earliest examples, some cultures were known to use liver to correct night blindness, the British Navy used lemons or limes to prevent scurvy, and alkali treatment of corn was associated with a lower incidence of pellagra in Mexico. The understanding of the linkage between certain foods and the prevention of disease resulted in the development of food guides or groups illustrating a pattern of food choices that was most likely to prevent deficiency diseases. As the chemical nature of the factors in food that prevented or cured nutritional deficiencies became known, it was possible to determine the specific amount required in the diet to maintain health. These studies led to the development of recommended dietary allowances (RDAs), which are quantitative recommendations of the nutrient intakes that will meet the needs of the majority of the population and have been used to evaluate the adequacy of diets in populations. However, this

understanding of nutrient requirements encouraged a focus on single food components rather than the overall balance of macro- and micro-nutrients to promote health.

By the second half of the twentieth century, the primary causes of nutrition-related diseases in many developed countries had shifted from nutrient deficiencies to those associated with dietary excess. As noted by the Surgeon General of the USA (1988), micronutrient deficiencies were no longer major public health problems in the USA; and diseases associated with excess intakes of energy, saturated fat, total fat, cholesterol, alcohol, and sodium, in conjunction with inadequate fiber intake, were the major causes of morbidity and mortality in the USA. The economic transition experienced by many developing countries has led to a similar pattern and is sometimes referred to as the double burden of malnutrition diseases. Although nutritional deficiencies continue to be prevalent in large segments of the population, an increasing proportion of the population is at risk of developing diet-related chronic diseases, such as obesity, cardiovascular disease, cancer, and diabetes. This emerging pattern of disease has resulted in the development of FBDG, which recommend dietary patterns that are adequate in nutrient content and encourage food choices to lower the risk of noncommunicable diet-related diseases.

Fig. 1, developed by the 2020 Dietary Guidelines Advisory Committee (DGAC) in the USA, illustrates the relationship between dietary patterns and their various components, which include foods and beverages, food groups, macronutrients, and micronutrients and recommended that all levels must be considered in relation to diet quality. For FBDG to be effective, each level must be considered in relation to the overall dietary pattern. For example, simply recommending that added sugars or saturated fats be limited in the diet needs to be presented in the country or regional context of the foods and beverages that are the major sources of these food components and the type of dietary pattern that will limit these food components and the foods and beverages that should be included in the pattern instead. Food pattern modeling was used by the DGAC to illustrate that, at various energy intake levels, 85–90% of energy is used for a healthful dietary pattern that meets nutrient requirements and goals for reducing risk for diet-related chronic diseases. Consequently, only a small portion of energy intake can be allocated to food components that provide energy but do not contribute to meeting nutrient and food component recommendations, such as added sugars, saturated fatty acids, trans fatty acids, and alcohol. The use of food pattern modeling to develop healthful dietary patterns illustrates the

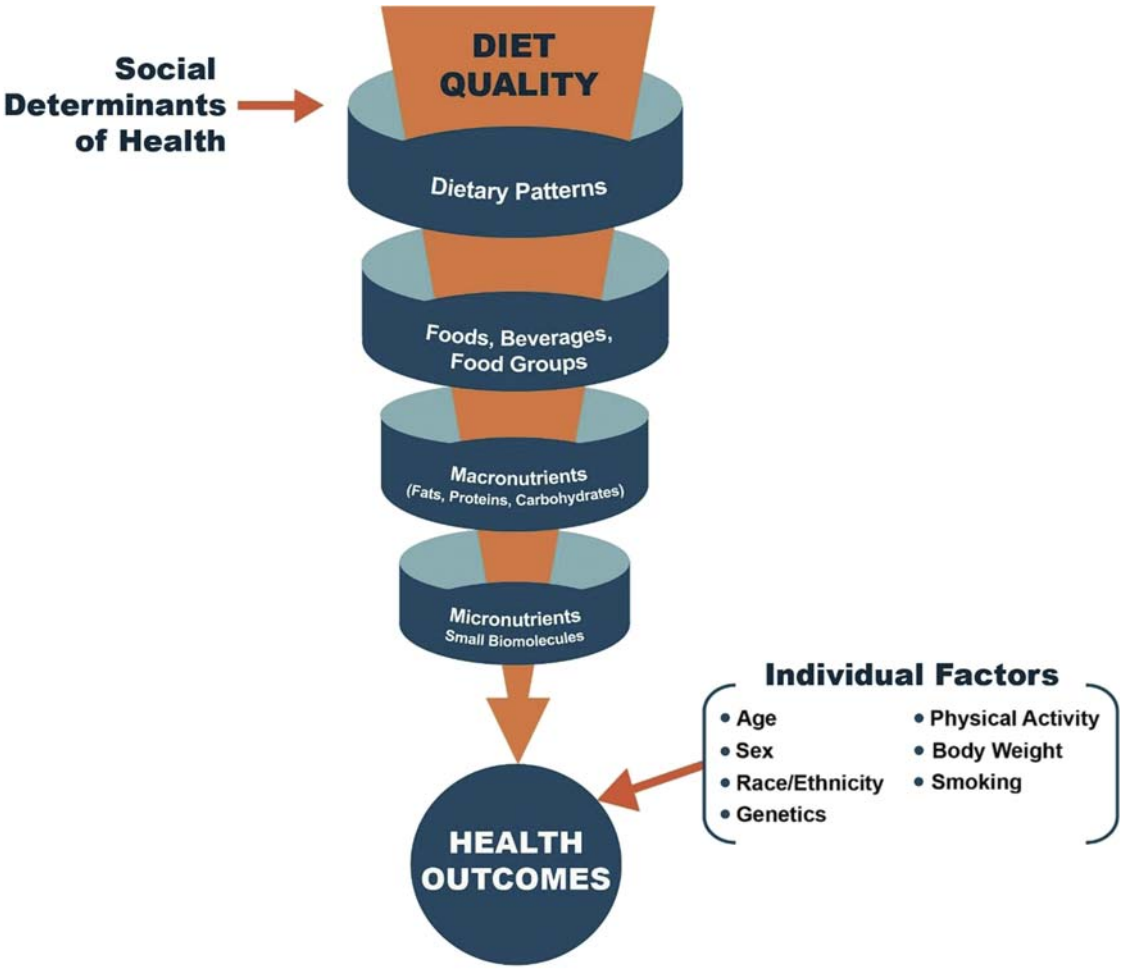


Fig. 1 Food-based dietary guidelines emphasize dietary patterns that encompass recommended foods, beverages, food groups, appropriate macronutrient balance, and recommended micronutrient levels that maintain health and promote well-being. At each level, overall diet quality needs to be considered in developing Dietary Guidelines. Figure is from the USA 2020 Dietary Guidelines Advisory Committee.

importance of nutrient dense foods as the core elements of such patterns because such foods are prepared or processed with none or a minimal amount of the food components to limit. Identification of the nutrient dense foods relevant to a population or culture is an essential step in developing relevant FBDG.

Types of guidelines

Since the International Conference on Nutrition, which was held in 1992, FAO and WHO have encouraged the development of FBDGs for countries or regions (FAO/WHO, 1992). The second International Conference on Nutrition in 2014 again emphasized the importance of FBDGs (FAO/WHO, 2014). On the FAO website (http://www.fao.org/nutrition/education/food-dietary-guidelines/background/fao-work-dietary-guidelines/en/) links are provided to the numerous FBDGs that have been developed by countries and regions as well as resources for developing this capacity at the country or regional level. The section below characterizes the important relationships between three types of nutrition recommendations or guidelines, which include nutrient requirements, FBDG, and food guides.

Nutrient requirements are quantitative recommendations for the intake of individual nutrients. Such recommendations are referred to by terminology such as Recommended Daily Allowance (RDA), Nutrient Reference Values (NRV), Dietary Reference Intakes (DRI) or Population Reference Intakes (PRI). In many countries these values are developed by scientific, authoritative bodies (e.g., National Academy of Sciences, Engineering, and Medicine (https://www.nal.usda.gov/fnic/dri-nutrient-reports), European Food Safety Authority (https://www.efsa.europa.eu/en/topics/topic/dietary-reference-values), or WHO and FAO) through assessment of research evidence, including systematic reviews or meta-analyses. Initially such evaluations focused on essential nutrients but now also address recommendations for nutrients and other food components that are associated with risk of diet-related chronic diseases.

FBDG are quantitative recommendations for intake of food groups to meet nutrient requirements and maintain health. Consequently, they do not establish nutrient requirements but recommend dietary patterns that will achieve recommended intakes for different age groups while limiting the food components that are associated with increased risk of diet-related chronic diseases. To characterize FBDG, terminology may refer to dietary patterns, food group recommendations or foods to encourage, and foods or food components to limit. The development of FBDG relies on evidence obtained through data analysis of population food intake patterns, systematic reviews of evidence linking food groups and dietary patterns with risk of diet-related chronic diseases, and food pattern modeling to assess how changes in amounts or types of foods impact meeting nutrient needs and recommendations. FBDG that are developed using these methodologies to review evidence are typically technical and quantitative; thus, they form the basis of food and nutrition policy and can be used by healthcare providers for assessing adequacy of dietary patterns in the population.

Food guides illustrate recommended dietary patterns based on the recommendations in FBDG. These illustrations typically convey the proportion of various food groups associated with healthful dietary patterns and are used to develop educational materials for the general public. The illustrations can be adapted for various age groups and populations. In addition, where graphic images of foods are used in the Food Guides, the images should reflect foods and food groups that are familiar and accessible by the population. Variations in the food patterns illustrated in these guides may be necessary to reflect how the FBDG can be adapted for various population groups (e.g., vegetarian patterns, patterns for young children or toddlers).

FBDG developed or endorsed by governments can be used to make decisions on food and nutrition policy. Nongovernment organizations may also develop FBDG for targeted purposes such as weight loss, management of patients with diseases such as cardiovascular disease or diabetes, a specific segment of the population, or promotion of a specific food culture. Such guidelines should be evaluated for adequacy in meeting nutrient requirements and effectiveness in their targeted purpose using rigorous methods for evaluation of evidence. Such validation is important to assure the nutritional properties of such recommendations.

The development of FBDG

FBDG express the principles of nutrition in terms of the food and food choices available to the population rather than in terms of specific nutrients or food components. Scientifically, these guidelines are based on the association between dietary patterns and the risk of diet-related diseases and incorporate recommendations that address major diet-related public health issues. In addition to communicating scientific knowledge about the association between food, dietary patterns, and health, development of FBDG provides an opportunity to strengthen consensus among various government and nongovernment organizations on important nutrition recommendations to be incorporated into educational programs. By expressing scientific principles in terms of food, FBDG recognize the consumer awareness of food rather than nutrients and emphasize to consumers the importance of meeting nutrient needs with foods. Thus, both the content of the FBDG and the development process are important.

Researchers often focus their studies on a specific nutrient or food component that may alter the risk of developing a disease. These studies are reviewed in the development of FBDG, but the information must be reorientated from a nutrient-based focus to a food-based recommendation by addressing the questions in **Table 1**. As indicated by these questions, the process is driven by the identification of diet-related public health issues and the development of food-based strategies that are relevant to the target population.

Table 1 Reorientating from nutrients and food components to foods (FAO and WHO, 1998).

What are the important public health issues for the population? Do they have diet-related factors?
Health statistics will indicate the major causes of morbidity and mortality in a population. Diet-related diseases include nutritional deficiency diseases and noncommunicable diseases, such as obesity, type 2 diabetes, certain types of cancer, and cardiovascular disease. It is important to determine whether nutrition is the primary cause of the disease or secondary to some other more prevalent problem (e.g., smoking, infectious agents).

What are the target nutrients linked to the major public health issues? Are there related nutrients or other factors?
In many nutrition-related problems, several nutrients or food factors may interact to cause the nutritional problem. For example, the fat content of the diet affects absorption of fat-soluble vitamins, obesity can be related to either excess energy intake or inadequate expenditure, multiple factors contribute to adequate bone formation, folic acid can mask anemia due to vitamin B_{12} deficiency, and so on. Simply increasing the intake of a target nutrient and ignoring these other factors may not address the problem adequately.

What foods are high in nutrient(s) or consumed in sufficient quantity to be a significant source of the nutrient(s)?
Using both food-composition databases and food-consumption data, foods that are good sources of the nutrient and foods that are consumed in sufficient quantity to meet the target intake can be identified. Likewise, dietary patterns that lower the risk for the public health problem and are associated with adequate intake of the nutrient can be identified.

What is likely to be acceptable to the target audience?
For nutrition interventions to achieve success, recommendations must target food choices that can be integrated into the diet based on cost and acceptability of the foods.

How do diet strategies integrate with other food policies?
Economic, agricultural, and trade analysis is useful to determine which diet strategies can be implemented and are sustainable.

The process for developing FBDG is based on building consensus among various sectors and groups involved in public health. Table 2 provides a general outline of the steps in the process, which can be adapted to the specific needs of a country or region. The goal is to have a set of guiding principles for food-based recommendations that lay out the overall policy agreed to by various agencies and groups.

The product of the working group is likely to be a document that outlines recommendations and includes background information on the rationale for the guidelines as well as guidance on implementing the recommendations. The FAO website has over 100 FBDG developed by different countries and regions. Table 3 illustrates common themes among various FBDG. These themes are characterized as foods or behaviors that are encouraged or as cautionary messages that refer to foods or food components to limit. Based on foods available and cultural practices, the types of fruits, vegetables, and whole grains and the specific types of food that are emphasized as sources of protein, calcium, or unsaturated fatty acids may vary. The United States *Dietary Guidelines for Americans* provide an interesting picture of how FBDG have evolved because they have been reviewed and updated every 5 years since 1980. During this period the core messages have been similar; however, the methodology to evaluate evidence has improved and

Table 2 Steps in the development of FBDGs (FAO and WHO, 1998).

1. **Develop support from key government agencies.** The successful implementation of FBDG will depend on support from key ministries, such as health, agriculture, education, sports, and recreation. Building consensus among these agencies will result in consistent messages regarding diet, health, and lifestyle for the public. Examples of support include technical support for data analysis or a secretariat to maintain and coordinate activities.
2. **Form a working group of experts.** The working group should include diverse expertize in areas, such as public health, nutrition, food science, agriculture, and behavioral sciences.
3. **Solicit public comment and relevant input.** The expert panel needs to gather and evaluate scientific information to determine the guidelines that are most relevant to the target population. This information can be obtained from the scientific literature. In addition, professional groups may have important information to submit to the panel for consideration. Solicitation of information is consistent with an open process; however, the panel is responsible for evaluating the relevance of the information submitted.
4. **Review and identify key public health issues and evaluate the diet–health relationships of concern for the population; determine the critical health, food, and nutrition issues to be targeted in the FBDG; and define the purpose, target groups, and content of the FBDG.** Even if data are limited, it is important for the working group to identify the key public health issues. This step may be especially important in countries in which both undernutrition and overnutrition are of concern. Identification of the public health issues allows the working group to address the questions in Table 1.
5. **Develop and draft the main messages for the FBDG.** The working group will need to decide whether the draft document will be targeted primarily at health professionals, and hence may be more technical, or will be targeted toward the general public. In developing the main messages, they may identify consumer-orientated materials, such as a food guide, that will be useful in communicating the FBDG to the public.
6. **Assess the cultural and economic appropriateness and credibility of the messages as perceived by the target groups.** Through focus groups or other types of consumer testing, the effectiveness of the FBDG can be assessed. This information can be used to revise the guidelines before developing the final draft.
7. **Release and implement the FBDG.** It is valuable to have government leaders from key ministries involved in the release and implementation of the FBDG so that there is a commitment to integrate the guidelines into departmental policies. In addition, the implementation can require development of educational materials for different target groups as well as public–private partnerships to aid in dissemination of the messages to the public.
8. **Monitoring and revision.** Monitoring can be used to assess the impact and implementation of the FBDG. In addition, monitoring data are useful for making appropriate revisions and updates to the guidelines on a periodic basis.

Table 3 Common themes for food-based dietary guidelines.

Foods or behaviors that are encouraged	*Cautionary messages*
Energy balance	Limit saturated fatty acids and *trans*-fatty acids
A healthful variety of foods; healthful dietary patterns	Avoid excess energy intake
Inclusion of fruits and vegetables	Balance of energy from fat
Use of whole grains	Limit foods and beverages with added sugar
Protein-based foods from various sources	Limit use of salt and salty foods
Foods that are sources of calcium	Limit or avoid alcohol
Sources of unsaturated fatty acids	
Safe food handling	
Inclusion of physical activity	
Family practices related to meals	

strengthened. Within these core messages the importance of defining healthful dietary patterns and the role of nutrient dense foods has become more prominent. In addition, the *Guidelines* itself has evolved from a simple, consumer-oriented brochure to a more detailed science-based policy document used by decision-makers for policy decisions and the design of health promotion programs (National Academy of Sciences, Engineering, and Medicine, 2017).

Most countries that have developed FBDG have also developed a food guide to accompany the messages in the guidelines (Fischer and Garnett, 2016). The food guide is typically a simple graphic illustration of food choices and dietary patterns; the graphic can be constructed to illustrate the relative proportions of food groups within the pattern. Criteria for a food guide should include representation of foods common to the population that are consistent with the FBDG, use of clear graphics that are meaningful to the target population, and illustration of dietary patterns that meet the nutrient requirements of the population. Although a simple graphic is useful for visual communication, it should be clear that proper use of the food guide depends on understanding the more complete information in the FBDG.

Conclusion

The use of FBDG has emerged as an important cornerstone of food and nutrition policies and education programs since the late 1970s and is valuable for addressing issues of both nutrient adequacy and risk of diet-related chronic diseases. Importantly, FBDG communicate nutritional principles in a manner that is relevant to the population. As a policy document, they should be revised periodically so that the information reflects current science on food and nutrition factors that promote health and prevent disease. Additionally, it is important for each country to develop their own set of FBDG so that the recommendations and presentation are relevant to the local population.

Emerging issues for consideration in FBDG include the need to develop systems approaches that reflect the numerous factors that influence food choices, including environment sustainability and socio-economic factors that affect access to healthful dietary patterns as well as taking steps to assure that the guidelines are relevant for all life stages and for populations with various health-related conditions. Addressing such issues will require expansion of expertize involved with FBDG development and access to relevant sources of evidence. In addition, addressing these broader issues of food production and availability requires consideration of appropriate mechanisms to address these topics with relevant expertize and evidence (Schneeman et al., 2021).

References

Dietary Guidelines Advisory Committee, 2020. Scientific Report of the 2020 Dietary Guidelines Advisory Committee: Advisory Report to the Secretary of Agriculture and the Secretary of Health and Human Services. U.S. Department of Agriculture, Agricultural Research Service, Washington, DC. https://www.dietaryguidelines.gov/2020-advisory-committee-report.

European Food Safety Authority. Dietary Reference Values, https://www.efsa.europa.eu/en/topics/topic/dietary-reference-values.

FAO/WHO, 1992. Final Report of the International Conference on Nutrition (ICN).

FAO/WHO, 2014. ICN2 Outcome Document: Framework for Action.

Fischer, C.G., Garnett, T., 2016. Plates, Pyramids and Planets: Developments in National Healthy and Sustainable Dietary Guidelines: A State of Play Assessment. FAO, Rome Italy.

Food and Agriculture Organization and World Health Organization, 1998. Preparation and Use of Food-Based Dietary Guidelines. WHO, Geneva, Switzerland.

National Academy of Sciences, Engineering, and Medicine, 2017. Redesigning the Process for Establishing the Dietary Guidelines for Americans. The National Academy Press, Washington, DC.

National Academy of Sciences, Engineering, and Medicine. Dietary Reference Intakes, https://www.nal.usda.gov/fnic/dri-nutrient-reports.

Schneeman, B.O., Ard, J.D., Boushey, C.J., Bailey, R.L., Novotny, R., Snetselaar, L.G., de Jesus, J.M., Stoody, E.E., 2021. Perspective: impact of the National Academy of Sciences, Engineering, and Medicine Report on the process for the 2020 Dietary Guidelines Advisory Committee. J. Nutr. 12, 1051–1057.

Surgeon General, 1988. Surgeon General's Report on Diet and Health. United States Government Printing Office, Washington, DC.

Relevant websites

FAO, http://www.fao.org/nutrition/education/food-dietary-guidelines/background/fao-work-dietary-guidelines/en/.
https://www.dietaryguidelines.gov.This page intentionally left blank

Dietary surveys: Surveys of food intake in groups and individuals

Katherine L. Tucker, Department of Biomedical and Nutritional Sciences, Center for Population Health, University of Massachusetts Lowell, Lowell, MA, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of K.L. Tucker, Dietary Surveys: Surveys of Food Intake in Groups and Individuals, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 79–83, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00077-5.

Introduction 385
Global research questions and data needs 385
Issues in survey design 386
Selection of dietary assessment measure 387
Household level 387
Individual level 387
New technology 389
Database limitations 389
Conclusions 389
References 390
Relevant websites 391

Key points

- National household surveys useful for targeting subpopulations at risk
- Survey design considerations for disaggregation and group comparison
- Assessment methods for individual usual nutrient intake: 24 h recall and food frequency questionnaire
- Methodologic limitations and approaches to minimize them
- Emerging technologies to improve assessment

Introduction

Dietary surveys are used for multiple purposes and range from measurement of food disappearance at the national level, to food use at the household level, to detailed multiple assessments of individual intake for linkage with health outcomes. Major methods for individual intake include the 24 h recall and the food frequency questionnaire. Each of these methods has strengths and limitations, depending on the survey purpose.

Global research questions and data needs

At the national level, information on food use is needed for economic and agricultural policy decision making. For policy makers to advise on food production, food imports, pricing of staple foods and other factors that affect food availability, they require information on the production, inflow and outflow of food commodities and products. Most countries use food balance sheets to measure these flows, and total available nutrients are estimated in relation to the size and composition of the population (Food and Agriculture Organization of the United Nations, 2001). These surveys measure overall national food production, imports, and available food stocks, and subtract exports, food used for animals, and losses that occur during production, storage, and manufacturing. While not a survey in the formal sense, this is a collection of data from the food sector regarding wholesale distribution. After adjusting for expected wastage, these data are compared to nutrient values and then to the size and composition of the population to calculate per capita nutrient availability. Because this is a crude assessment, it does not account for all losses and, therefore, tends to overestimate availability. The FAO has compiled food balance sheets for many countries since 1949, allowing useful inter-country comparisons of food availability. However, the aggregate information obtained with food balance sheets does not consider food distribution within a country and does not quantify food intake or needs of subgroups of the population.

Most countries need more information on household level food use to target food and nutrition policies toward groups at need. Commonly administered surveys include the International Food Policy and Research Institute's Household consumption and expenditure surveys (HCES), which are used in more than 120 countries (Fiedler and Mwangi, 2016). Household food surveys capture the amounts and types of food that enter a household, and per capita intake equivalents are calculated by dividing the total nutrients available in the household from the edible portion of entering foods by the numbers of household members, weighted by age and sex. This information allows identification of groups at risk of inadequate intake of energy or of specific nutrients. For example, these surveys may highlight rural-urban differences, inland-coastal differences, differences by socio-economic strata, and so on. Such surveys provide critical information within countries for the development and targeting of economic, agricultural, and nutritional policies specific to regions or other sub-groups of the population. They do not, however, provide information on individual intakes within the household, and are not useful for understanding age- and sex-specific intakes.

To describe food and nutrient intake by age, sex and physiologic state, data are needed at the individual level. Surveys of individual dietary intake use methods that range from qualitative food checklists to multiple detailed records of food intake, with quantification of preparation methods and portion sizes. Individual level data are used for a variety of purposes and the survey design will depend on the primary data needs. At the national level, a primary objective is to identify sub-groups at risk of inadequate intake of energy or specific nutrients. The important advantage of individual level data is that target age and sex groups may be identified in addition to groups identified by region or other household level characteristics. A further objective is to determine the extent of under- or over-nutrition in relation to energy or specific nutrients in sub-populations. This requires consideration of the distribution of intakes in specific age, sex, and physiologic status groups. A more ambitious objective of some national or targeted dietary surveys is to associate aspects of individual dietary intake with the existence of health conditions. This objective requires that usual intake data be valid and reliable at the individual level.

Issues in survey design

For a household level survey to be nationally representative, one must carefully consider the sampling design. This is generally done by multilevel selection of regions, then sub-regions, then households, in such a way that the resulting data may be generalized to the national level. For results at the regional level, coverage of all regions is necessary, although this will usually increase the cost of the survey. When the objective is more specific than national description, target areas may be selected, or subgroups oversampled, based on risk status or relevance to the question being addressed.

Similarly, for data to be representative of the greater population, complex sampling design is employed. Decisions on sampling design will generally be a balance between equal opportunity for individual inclusion against logistic and cost considerations of full randomization. The design may be similar to that of the household level survey, with the added step of randomly selecting individuals within households. Although surveys focused on the household level may interview all members of the household, this is not the most efficient way to get data representative of individuals in the population, due to the lack of independence of the observations. Members in the same family consume similar foods and therefore are more like each other than others in their community. Although this lack of independence can be adjusted in the analysis, it will require larger numbers of interviews to achieve representative stability of data estimates. The multistage approach of region, sub-region and community is less representative than a pure random sample, but this is corrected by consideration of the "design effect," which is calculated by comparing variation within, versus between, sampling units at each level (Kish, 1965). Although the design effect leads to the need for higher overall numbers of surveyed individuals, this is generally considerably less expensive than expanding coverage to all locations.

In addition to representation of the general population, many surveys also consider sub-groups that will not be well represented unless specifically over-sampled. Examples include pregnant women, ethnic subgroups or low-income groups. Individuals that meet the specified characteristic are identified within the existing sampling design but are selected in larger numbers than would be representative of the entire population. This allows sufficient sample size to present valid estimates separately for these groups. When included in measures of the total population, the over-representation of these sub-groups is adjusted using sampling weights. This design is used, for example, in the US National Health and Nutrition Examination Survey (NHANES), which provides intake data representative of the non-institutionalized US civilian population (NHANES). While multistage sampling is used, it is not sufficient to cover smaller subgroups within the population. Therefore, oversampling is conducted for selected subgroups. These have varied over time but, for example, the 2015–16 cycle included oversampling of Hispanic, non-Hispanic black, non-Hispanic Asian, individuals at or below 185% of the Department of Health and Human Services (HHS) poverty guidelines, and persons aged 80 years and older (NHANES).

The NHANES uses multi-stage cluster sampling throughout the country. Sampling weights to correct for design effects are provided online, along with detailed instructions for use, to be able to extrapolate results by age, sex, and race/ethnicity to the overall US population. Because of this design, the data are not useful at the state level. The Centers for Disease Control and Prevention

(CDC) conduct another survey, the Behavioral Risk Factor Surveillance System (BRFSS), to collect information at the state level. However, this survey, conducted by telephone, includes only a few questions on fruit and vegetable intake.

Another design consideration is the timing of the survey. Intake may vary considerably by season and it is, therefore, important that all seasons are represented. Although logistic and cost constraints often limit ideal design planning, it is optimal if data from all seasons are collected in all survey locations. If different locations are covered at different times of the year, comparisons across regions may be compromised. Additionally, intakes may be misrepresented if certain days of the week are not included in the data collection plan.

Selection of dietary assessment measure

Household level

Several alternative methods of dietary assessment are available. At the household level, one commonly used approach is the food account method (Food and Agriculture Organization of the United Nations, 2008), Such household consumption and expenditure surveys (HCES) are conducted globally in more than 120 countries. The person most responsible for the acquisition and/or use of food is asked to keep a daily record of all the food that enters the household for a specified period—often one week. This includes food purchases, food production and food received as gifts. There are several limitations to this approach, including the assumption of constant food stores, which may not be the case.

In some locations, it is not feasible for many individuals to accurately record this information. In this case, an interviewer-administered list-recall method is used. The interviewer asks the responsible household member to recall food purchases, production, or gifts in the household during a specified period, following a list of major foods that are relevant for that location. Additional information on age and sex of household members and number of meals each consume at home is collected to calculate adult equivalent per capita food availability for the household. Although edible portions of foods are generally considered in quantifying availability, most such surveys do not account for wastage or use by animals and, therefore, may overestimate household food use. On the other hand, if they do not account for food consumed away from the home, they may underestimate food intake. While useful for economic and food commodity flow information, this type of survey is, therefore, limited with respect to nutritional intake assessment.

To understand dietary intake within households, more elaborate methods are needed. One approach is to use a household diet record, where the household respondent is asked not only to report inflows of food, but also to record actual use and preparation of foods in the household over a specified period. Food consumed outside the home may also be assessed for each household member, and the number of individuals, including guests, who are present at each meal is recorded. While demanding for the respondent, this approach provides a better estimation of the total food consumed by the household than the inventory methods described above (Fiedler and Mwangi, 2016). Estimation of waste is included in some, but not all, such surveys and is a limitation of most. Because of the heavy respondent burden, incomplete response is also a major problem, which threatens the validity and generalizability of the survey.

Individual level

The selection of dietary assessment methods depends on the questions to be addressed, balanced with cost considerations. Major uses of individual intake data from dietary surveys include the description of intake distributions by sub-group, description of proportions of the population with inadequate or excessive intake of specific foods or nutrients, and comparison of dietary intake with individual characteristics, including health status.

If the goal is to describe the mean and distribution of intakes of groups and sub-groups, the most efficient method is the use of a single 24 h dietary recall per selected individual. The 24 h recall is usually administered by a trained nutritionist who asks the individual to report each eating and drinking occasion throughout the previous day. Once the items are listed, details are requested, including specific portion sizes, recipes, brand name or source of the food item, use of condiments, etc. The automated multi-pass method (AMPM), developed by the U.S. Department of Agriculture (USDA), begins with (1) a quick list of foods and beverages, followed by (2) a probe for frequently forgotten foods (like snacks), (3) questions on the time and eating occasion (meal or snack) for each food, (4) detailed information on portion size, recipes, condiments,etc. and (5) a probe for anything that may have been missed.

This is the methodology used in the NHANES, providing a good description of nutrient intakes by age, sex and ethnic group. Until 2003–2004, a single recall was administered. As an aggregate measure, this design worked well. There are limitations to these estimates, however, and validation against quantified energy expenditure measurements has shown that most people tend to under-report intakes with the 24 h dietary recall method (Subar et al., 2003).

Although the mean intake of subgroups has been considered generally valid, the tails of the intake distribution from a single recall are extended, leading to overestimation of proportions either above or below a specified cutoff point, relative to what is seen when usual intake is assessed as the average of multiple days (**Fig. 1**) (National Research Council (US) Subcommittee on

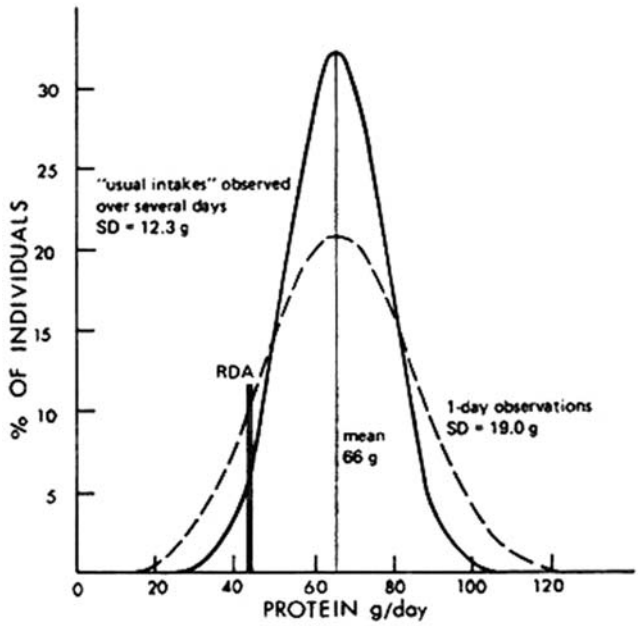


Fig. 1 Distribution with one vs. multiple days of intake (National Research Council (US) Subcommittee on Criteria for Dietary Evaluation, 1986).

Criteria for Dietary Evaluation, 1986). This occurs because individuals may under or over consume on any single day in a way that they would not, over the average of many days, resulting in misclassification of individual usual intake.

With repeated recalls on a representative subset of the population, the day-to-day variability may be quantified and used to adjust estimates to better represent usual intake (Beaton et al., 1983). Statistical correction methods have been developed by the National Cancer Institute (NCI) to further correct intake distributions (Tooze et al., 2010). These techniques pull in the tails and provide a more realistic distribution of usual intake, providing a more accurate estimate of the proportion of the population that falls below or above a specified cutoff point. For this reason, many studies, including the NHANES, now incorporate two recalls in their design. These methods continue to be updated (e.g. (Luo et al., 2021)) and SAS macros for public use are available on the NCI website (The NCI Method).

A third important objective of dietary surveys is to gain a better understanding of the correlates of nutrient intake with health outcomes. Day to day variability misclassification severely limits the ability to correlate intake data with individual characteristics. For many nutrients, this intra-individual variation in intake is considerable, and multiple days would be required to achieve stable estimates of individual usual intake. An extreme example is vitamin A, which tends to be concentrated in a few foods. If one frequently consumes liver and carrots but happened not to consume either of these on the day of the recall, that individual would be classified as having low vitamin A intake, when their usual intake is actually high. Conversely, one who almost never eats these foods, but had liver on the day of the recall would be incorrectly placed at the upper end of the vitamin A distribution. The greater the intra/inter individual variance ratios, the lower the ability to see true associations. The collection and averaging of multiple days of intake greatly improve this situation. For most nutrients three or four days are acceptable. However, for some nutrients, including vitamin A and vitamin B12, the variance ratios are so high that an unrealistic number of days would be needed for stable estimates.

Although variance ratios can be used to adjust the distribution of intake, this does not allow the true estimation of individual usual intake. For this reason, and others, the food frequency questionnaire (FFQ) remains the method of choice in most studies where diet—disease relationships are the primary objective. The FFQ asks respondents to report the frequency of consumption of a pre-specified list of specific foods, usually over the previous year. Additional questions on portion size and on preparation methods may be added, as appropriate. FFQ provide a lower cost alternative to multiple recall days, but also have limitations. Because they rely on a food list, their validity is dependent on the representativeness of that list, and of portion size and recipe assumptions. Most FFQ in wide use have been developed using data that represent the major sources of nutrient intakes in the US population. These include the Willett, Block and NCI questionnaires (Willett et al., 1985; Block et al., 1986; Subar et al., 2001). However, individuals with divergent eating patterns are not well represented using these tools due to limitations in the food list and differences in portion sizes and preparation methods (Tucker, 2010). For this reason, most large studies of different cultural groups develop unique FFQs with appropriate food lists for their population. This improves internal validity, but limits comparisons across studies.

To improve estimation further, the National Cancer Institute (NCI) proposed a combined method to better assess usual intakes in population-based studies. This method uses the data from 2 24 h recalls to adjust distributions for day-to-day variability and adds

additional information on frequency of specific food intake from a propensity questionnaire, to estimate the probability of consumption when the foods are absent in the 24 h recalls (Subar et al., 2006). The propensity questionnaire originally did include portion sizes but captured frequency of intake of foods which may not appear on any given day, but which may be added to statistical models to improve overall nutrient distributions. The most recent NCI FFQ, now called the Diet History Questionnaire (DHQ), is available for use with or without portion sizes. Methods for combining the information from these two sources to improve estimates continue to be developed (Freedman et al., 2018).

In addition to the day to day variation, which is random, another more complex issue is underreporting. Validation against quantified energy expenditure measurements using doubly labeled water has shown that most people tend to underreport intakes with the 24 h dietary recall method (Subar et al., 2003). Underreporting will affect the mean intake of groups and, if differential, may introduce bias in sub-group comparisons. Unfortunately, it has been determined that underreporting is usually not random and is more common among those with obesity, with restrained eating or who tend toward socially desirable responses (Kretsch et al., 1999; Hebert et al., 2001). However, the USDA AMPM system, applied carefully, has been shown to reduce underreporting and to yield relatively good intake estimates when compared with the biomarker doubly labeled water (Moshfegh et al., 2008), although underreporting remained more common among those with obesity.

New technology

While traditional use of 24 h recalls and FFQ in surveys has involved interviewer administered questionnaires, originally on paper and then directly into a computer program (such as with the multiple pass system used by NHANES) (Moshfegh et al., 2008), there has been a recent shift to use of online questionnaires. The NCI developed an online 24 h recall, ASA24, which allows direct entry by participants from their home (ASA24) and have also placed their DHQ online. Additional tools have been developed in other countries (Dao et al., 2019).

Although not described in detail here, new biomarkers are being developed to assist in validation and calibration of dietary intake reports. Established biomarkers include recovery biomarkers, which capture total intakes with good precision. These include doubly labeled water for energy and urinary protein, sodium and potassium. Concentration biomarkers, which correlate well with intake include plasma vitamins, carotenoids and fatty acids (Potischman and Freudenheim, 2003). New biomarkers being explored include urinary sucrose and fructose, and plasma metabolites assessed with metabolomics (Gao et al., 2017). Other technologies, including photographs are being tested to capture portion size. However, although many advances using cameras have been made, they remain burdensome and cannot provide the ingredients of the item and, therefore, must be coupled with self-report.

With the emergence of human genome data and movement toward personalized nutrition for health, there is increasing demand for estimates of usual intake for use in studies of gene-diet interactions. Because these usually require very large samples, it is rarely feasible to include three dietary assessments. Further, there is a trend toward combining data from multiple studies to obtain sufficient sample size and this is not easily done when studies use different dietary assessment methods. More efficient methods of valid dietary assessment, which can be collected consistently across studies, are needed.

Database limitations

Whatever dietary assessment measure is used; the utility of the data is dependent on the translation of reported food intake to nutrient intakes. This requires detailed and accurate nutrient databases. The USDA has an extensive nutrient database, allowing for good estimation of dietary intakes. Most other countries have not conducted this level of food composition analysis for their own locations. Therefore, the values in most existing databases stem from the US nutrient database, adding information as possible from locally analyzed products. Because the nutrient composition of many foods, including fruit, vegetables and even animal products, can vary widely by growing conditions and specific sub-variety, most available nutrient databases remain inadequate. Many databases use extrapolated values from similar foods when chemical analysis has not been completed. Furthermore, it is common for many country-specific databases to include information only on macronutrients and a few selected vitamins and minerals. The continual arrival of new manufactured products also complicates the upkeep and management of food composition databases. Considerable database work remains to expand the utility of worldwide dietary surveys. Major ongoing efforts in this regard include the international network of food data systems (INFOODS) and the Global Nutrient Database (Miller et al., 2021).

Conclusions

Dietary surveys are critically important for understanding the dietary intake of populations, changes over time, and the prevalence of inadequate or excessive intakes by subgroups. This information is used by policy makers to make decisions on nutrition policy, including supplementation or fortification of foods, and for targeting groups at risk of malnutrition with nutrition intervention and education. Despite the importance of these surveys globally, many limitations continue to exist, including a tendency toward underreporting of intake, issues of day-to day and seasonal variability in intake, and limitations of existing nutrient databases. Advances in methodology continue to be implemented, including the automation of data collection and the development of

Table 1 Advantages and disadvantages of different approaches for dietary surveys.

Level	*Survey type*	*Advantages*	*Disadvantages*
National	Food balance sheets	Inexpensive	Crude estimate; no consideration of wastage; does not allow disaggregation to sub-levels
Household	Food account method	Inexpensive	Does not account for food consumed away from home, inventory or wastage
	Interviewer-administered list-recall	More detail obtained on foods than in the food account method	List may limit responses; waste usually not accounted for
	Household diet record	Usually covers one week with detail. Most accurate of household methods	High respondent burden; expensive
Individual	24 h dietary recall	Detailed information on food intake, good estimate of mean intakes by sub-group	Misclassifies individuals; single recall per person is not useful for correlative investigation
	Multiple recalls	Average of multiple days can give good quantitative estimate of usual intake; with two recalls, intra/inter-individual variance can be calculated and used to correct correlations and linear coefficients for random error	Expensive; variability ratios are useful but not sufficient to correctly classify individuals; has limited utility in non-linear analyses.
	Food frequency questionnaire (FFQ)	Inexpensive, measures usual intake	Semi-quantitative, dependence on food list and recipe assumptions may lead to error in estimation of intake in sub-groups
	Combined approach with 2 24 h recalls and a qualitative FFQ or propensity questionnaire	Allows improved estimates of individual intakes by adjusting for the probability of consumption of food groups that may appear as zeros in 2 24 h recalls	Expensive; heavy respondent burden; requires considerable statistical expertize

statistical methods to maximize the use of repeat recalls with or without added information from FFQ. Further efforts to be inclusive of subgroups, to improve data collection with technology, to improve nutrient databases, and to validate methods against biomarkers, are needed (Table 1).

References

Beaton, G.H., Milner, J., Mcguire, V., Feather, T.E., Little, J.A., 1983. Source of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Carbohydrate sources, vitamins, and minerals. Am. J. Clin. Nutr. 37, 986–995.

Block, G., Hartman, A.M., Dresser, C.M., Carroll, M.D., Gannon, J., Gardner, L., 1986. A data-based approach to diet questionnaire design and testing. Am. J. Epidemiol. 124, 453–469.

Dao, M.C., Subar, A.F., Warthon-Medina, M., Cade, J.E., Burrows, T., Golley, R.K., Forouhi, N.G., Pearce, M., Holmes, B.A., 2019. Dietary assessment toolkits: an overview. Publ. Health Nutr. 22, 404–418.

Fiedler, J.L., Mwangi, D.M., 2016. Using Household Consumption and Expenditure Surveys to Make Inferences About Food Consumption, Nutrient Intakes and Nutrition Status: How Important is It to Adjust for Meal Partakers? IFPRI Discussion Paper 1571 International Food Policy Research Institute (IFPRI), Washington, D.C. http://ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/130855.

Food and Agriculture Organization of the United Nations, 2001. Food Balance Sheets: A Handbook. http://www.fao.org/3/x9892e/x9892e00.htm. (Accessed 26 July 2021).

Food and Agriculture Organization of the United Nations, 2008. Method to Monitor the Human Right to Adequate Food, vol. II. FAO, Rome. http://www.fao.org/3/i0351e/i0351e12.pdf. (Accessed 26 July 2021).

Freedman, L.S., Midthune, D., Arab, L., Prentice, R.L., Subar, A.F., Willett, W., Neuhouser, M.L., Tinker, L.F., Kipnis, V., 2018. Combining a food frequency questionnaire with 24-hour recalls to increase the precision of estimation of usual dietary intakes-evidence from the validation studies pooling project. Am. J. Epidemiol. 187, 2227–2232.

Gao, Q., Pratico, G., Scalbert, A., Vergeres, G., Kolehmainen, M., Manach, C., Brennan, L., Afman, L.A., Wishart, D.S., Andres-Lacueva, C., Garcia-Aloy, M., Verhagen, H., Feskens, E.J.M., Dragsted, L.O., 2017. A scheme for a flexible classification of dietary and health biomarkers. Genes Nutr. 12, 34.

Hebert, J.R., Peterson, K.E., Hurley, T.G., Stoddard, A.M., Cohen, N., Field, A.E., Sorensen, G., 2001. The effect of social desirability trait on self-reported dietary measures among multi-ethnic female health center employees. Ann. Epidemiol. 11, 417–427.

Kish, L., 1965. Survey Sampling. John Wiley & Sons, Inc., New York, ISBN 0-471-10949-5.

Kretsch, M.J., Fong, A.K., Green, M.W., 1999. Behavioral and body size correlates of energy intake underreporting by obese and normal-weight women. J. Am. Diet Assoc. 99, 300–306.

Luo, H., Dodd, K.W., Arnold, C.D., Engle-Stone, R., 2021. Introduction to the SIMPLE macro, a tool to increase the accessibility of 24-hour dietary recall analysis and modeling. J. Nutr. 151, 1329–1340.

Miller, V., Singh, G.M., Onopa, J., Reedy, J., Shi, P., Zhang, J., Tahira, A., Shulkin Morris, M.L., Marsden, D.P., Kranz, S., Stoyell, S., Webb, P., Micha, R., Mozaffarian, D., 2021. Global Dietary Database 2017: data availability and gaps on 54 major foods, beverages and nutrients among 5.6 million children and adults from 1220 surveys worldwide. BMJ Glob. Health 6.

Moshfegh, A.J., Rhodes, D.G., Baer, D.J., Murayi, T., Clemens, J.C., Rumpler, W.V., Paul, D.R., Sebastian, R.S., Kuczynski, K.J., Ingwersen, L.A., Staples, R.C., Cleveland, L.E., 2008. The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes. Am. J. Clin. Nutr. 88, 324–332.

National Research Council (Us) Subcommittee on Criteria for Dietary Evaluation, 1986. Nutrient Adequacy: Assessment Using Food Consumption Surveys. National Academies Press, Washington (DC).

Potischman, N., Freudenheim, J.L., 2003. Biomarkers of nutritional exposure and nutritional status: an overview. J. Nutr. 133, 873S–874S.
Subar, A.F., Thompson, F.E., Kipnis, V., Midthune, D., Hurwitz, P., Mcnutt, S., Mcintosh, A., Rosenfeld, S., 2001. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires : the eating at America's table study. Am. J. Epidemiol. 154, 1089–1099.
Subar, A.F., Kipnis, V., Troiano, R.P., Midthune, D., Schoeller, D.A., Bingham, S., Sharbaugh, C.O., Trabulsi, J., Runswick, S., Ballard-Barbash, R., Sunshine, J., Schatzkin, A., 2003. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. Am. J. Epidemiol. 158, 1–13.
Subar, A.F., Dodd, K.W., Guenther, P.M., Kipnis, V., Midthune, D., Mcdowell, M., Tooze, J.A., Freedman, L.S., Krebs-Smith, S.M., 2006. The food propensity questionnaire: concept, development, and validation for use as a covariate in a model to estimate usual food intake. J. Am. Diet Assoc. 106, 1556–1563.
Tooze, J.A., Kipnis, V., Buckman, D.W., Carroll, R.J., Freedman, L.S., Guenther, P.M., Krebs-Smith, S.M., Subar, A.F., Dodd, K.W., 2010. A mixed-effects model approach for estimating the distribution of usual intake of nutrients: the NCI method. Stat. Med. 29, 2857–2868.
Tucker, K.L., 2010. Dietary patterns, approaches, and multicultural perspective. Appl. Physiol. Nutr. Metab. 35, 211–218.
Willett, W.C., Sampson, L., Stampfer, M.J., Rosner, B., Bain, C., Witschi, J., Hennekens, C.H., Speizer, F.E., 1985. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am. J. Epidemiol. 122, 51–65.

Relevant websites

AMPM, https://www.ars.usda.gov/.
ASA24, https://epi.grants.cancer.gov/.
BRFSS, https://www.cdc.gov/brfss/.
DHQ III, https://epi.grants.cancer.gov/.
INFOODS, http://www.fao.org/.
NHANES, https://www.cdc.gov/nchs/nhanes/.
The NCI Method, https://epi.grants.cancer.gov/.

Displaced populations

Daniel J. Hoffman and Hannah Posluszny, Rutgers University, New Brunswick, NJ, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of R. Bhatia, Refugees: Nutritional Implications, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 147–152, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00244-0.

Introduction 393
Displaced persons and refugees 393
Global trends 393
Nutrition implications of displacement 394
Macronutrients 395
Micronutrients (vitamins and mineral deficiencies) 396
Conflicts and refugees and long-term health 396
Addressing nutrition in refugees 397
Food baskets for populations (general food distribution) 397
Addressing acute malnutrition among young children 397
Management of moderate acute malnutrition 397
Addressing severe acute malnutrition 398
Challenges 398
Summary 398
References 399
Further reading 400

Key points

- The global prevalence of displaced persons, whether within their home country or abroad as refugees, has increased in the past decade.
- Displaced persons are at risk for nutritional deficiencies given the high prevalence of food insecurity in refugee camps and conflict zones.
- Displaced persons form a vulnerable population that is prone to nutritional deficiencies, such as micronutrient deficiencies, wasting, and stunting.
- There are a number of treatment modalities that can be provided to displaced persons, including micronutrient fortified food baskets, micronutrient sprinkles, and meal replacements.
- Supplement feeding programs are vital to supporting the health of displaced persons and efforts are needed to maintain the sustainability of such programs.

Glossary

Community therapeutic care program (CTC) A CTC program has the same initial metabolic stabilization phase as a traditional feeding program, and life-threatening infections are identified and treated in the same way. Once the patient is stabilized, they move directly to an outpatient therapeutic program that operates through existing health structures and initiates nutritional rehabilitation with ready to use therapeutic foods (RUTFs). When there is no longer a risk of severe malnutrition, they are referred to supplementary feeding programs for recuperation

Internationally displaced persons (IDPs) These are persons fleeing from war, civil disturbance, and violence of any kind but who do not cross international boundaries

Ready to use therapeutic foods (RUTFs) These foods are ready to eat and high in energy and protein. They also contain micronutrients and electrolytes. Their main use is for treatment of malnutrition. Their use to prevent stunting and wasting is also being evaluated

Special feeding program This program deals with provision of high-quality foods to be consumed in addition to the usual diet, with either targeted (to prevent persons with moderate acute malnutrition from becoming severely malnourished) or blanketed (to prevent nutritional deterioration of a larger population) distribution

Introduction

Since World War II, more than 100 million people have been forced to flee persecution or the violence of war to seek refuge in neighboring countries or in different areas of their own countries. The optimism following the end of the Cold War was short-lived as a number of civil conflicts erupted throughout the world. In 2020, there were 57 active conflicts, including eight wars, and 48 of these conflicts were internal (PRIO Conflict Trends, 2020). Most recently, the 2022 Russian invasion of Ukraine caused more than 2 million refugees to flee to Europe in less than two weeks. Armed conflicts have increasingly affected civilian populations, resulting in high mortality, widespread human rights abuses, forced migration, famine, and total collapse of governance in some countries. As well, climate change and the associated impact of agriculture and labor is becoming more common in lower income nations, adding to the numbers of displaced people in the world (Balsari et al., 2020). The focus of this article is to describe responses to humanitarian crises and the nutritional impact of being a displaced person.

Displaced persons and refugees

Civil conflict, economic challenges, and climate change exist as the main drivers of displaced persons. Displaced persons are those who flee their homes and become refugees, or need to relocate within their own country and are considered "internally" displaced, or asylum seekers hoping to escape persecution. Displaced persons are often left with little to no medical care and access to safe and nutritious food is often entirely dependent on volunteer aid programs. Since these program are not always sustainable long-term, the risk of undernutrition is extremely high for most displaced persons, but more so for growing children and the elderly.

The 1951 United Nations Convention defines a refugee as

> any person who owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group, or political opinion is outside the country of his nationality and is unable, or owing to fear is unwilling to avail himself of the protection of the country.

In 1969, the Organization of African Unity expanded this definition to include persons fleeing from war, civil disturbance, and violence of any kind.

"Refugees" cross international borders, but "internally displaced persons" (IDPs) do not. However, both groups have been forced to leave their homes and undergo physical and mental trauma as they settle in harsh and unhealthy environments, where they are often unable to take responsibility for their own welfare. The terms "refugee" and "internally displaced person" have major implications for the people concerned, particularly regarding their rights to protection and assistance, which are embedded in international law. The United Nations High Commissioner for Refugees (UNHCR) is mandated by the international community to protect and provide assistance to refugees. Owing to state sovereignty, the internally displaced are not included within UNHCR's mandate. Only on an *ad hoc* basis, at the request of the secretary general of the United Nations and the nation concerned, does UNHCR provide assistance to IDPs.

Global trends

Globally, over 82 million people were forcibly displaced worldwide at the end of 2020, the highest number uprooted by conflict and persecution since the mid-1990s, according to UNHCR's annual 2020 Global Trends report (UNHCR, 2020). The prevalence of all classifications of displaced persons (i.e., refugees, IDPs, and asylum seekers) has increased steadily since 2010 (**Fig. 1**). Specifically, the number of refugees increased from approximately 10 million in 2003 to a peak of more than 26 million in 2020. The majority of people who were displaced came from Syria, Venezuela, Afghanistan, and South Sudan with the remainder from various countries in Asia and Africa (**Fig. 2**). This estimate stayed level primarily due to the global COVID-19 pandemic that forced people to quarantine and limited global migration. In addition to the large numbers of refugees, in 2020 there were approximately 48 million IDPs worldwide, mostly drive by Venezuelans displaced abroad, and, as this article goes to print, the ongoing Russian war in Ukraine continues to generate millions of refugees.

In terms of refugees, according to the 2020 UNHCR report, the overall number of refugees increased from 15 million in 2012 to over 26 million in 2020. The largest numbers of refugees in 2020 were from Syria and Venezuela with the next largest numbers from Afghanistan, South Sudan, and Myanmar (**Fig. 3**). At the same time, the number of refugees voluntarily returning to their home countries was estimated to be 3.4 million by the end of 2019. However, by the end of March 2022, it was estimated that over 4.0 million Ukrainians became refugees, the largest number of European refugees since World War II, reversing gains observed in the past decade.

A major event that limited or even reduced the global prevalence of displaced persons was the global COVID-19 pandemic. By the end of 2020, an estimated 160 countries had closed their borders and nearly 100 countries did not make exemptions for those applying for asylum. While some refugees were able to leave their home country, many were not and the number of IDPs increased.

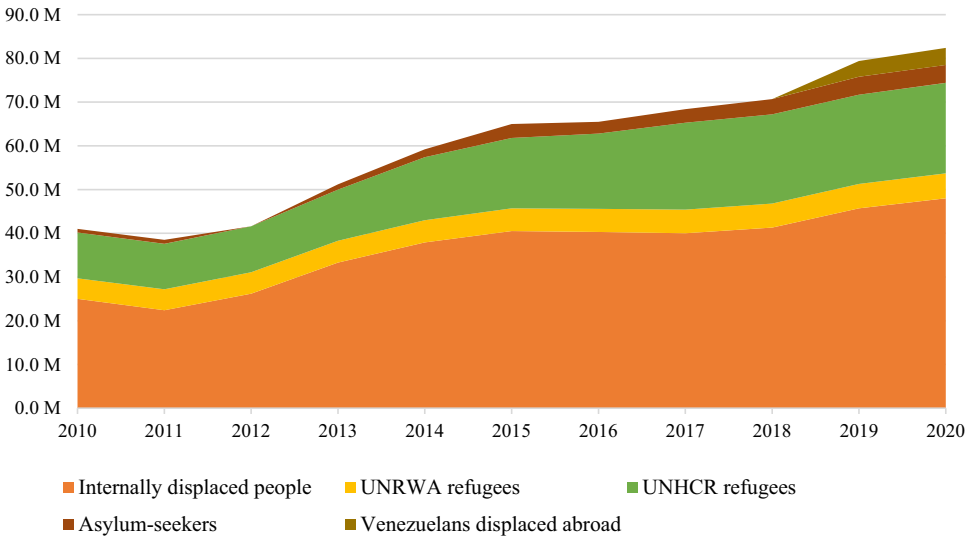


Fig. 1 Trends in global displacement between 2010 and 2020 modified from the 2020 Global Report of the United Nations High Commissioner for Refugees (UNHCR, 2020).

As well, for those who were not allowed to leave conflict zones or seek asylum, a reported increase in food insecurity and domestic violence occurred, creating additional challenges beyond those they were hoping to escape.

Nutrition implications of displacement

The public health and nutrition consequences of war and human displacement have been well documented during the past 40 years. The major determinants of high mortality among affected populations and priorities for intervention have been identified. Access to safe water, sanitation, shelter, and immunizations (measles) are essential, but an adequate and diverse diet remains central to the management of refugee operations. Today, it is acknowledged that acute undernutrition is a strong predictor of excess mortality among young children, and micronutrient deficiencies contribute significantly to diseases in emergencies. Therefore, resources and efforts must remain focused on these areas to best protect against mortality in refugee situations.

Undernutrition is defined as a state in which the physical function of an individual is impaired to the point at which he or she can no longer maintain adequate bodily performance processes, such as growth, pregnancy, lactation, physical work, and resisting and recovering from disease. This condition is generally due to the cumulative result of a lack of nutrients due to a combination of social conditions that limit access to nutritious foods, chronic food insecurity, and chronic intestinal infection. Undernutrition

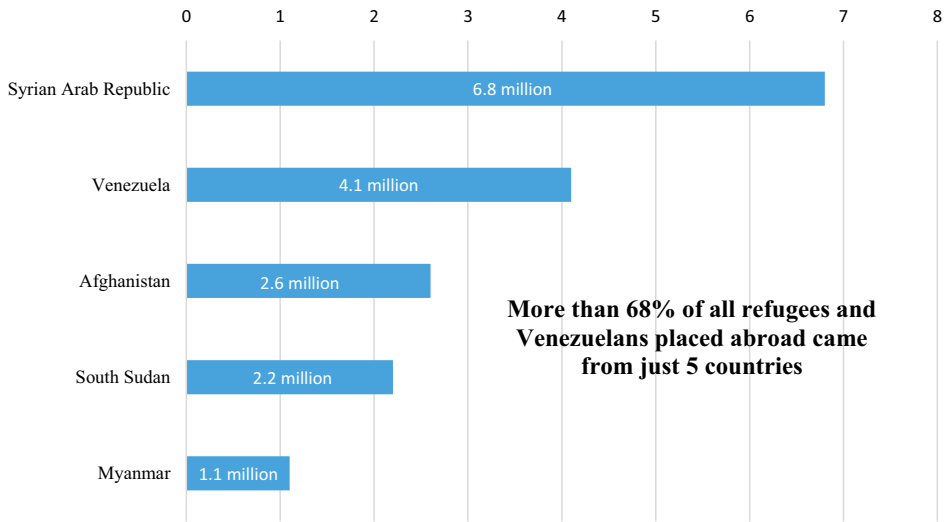
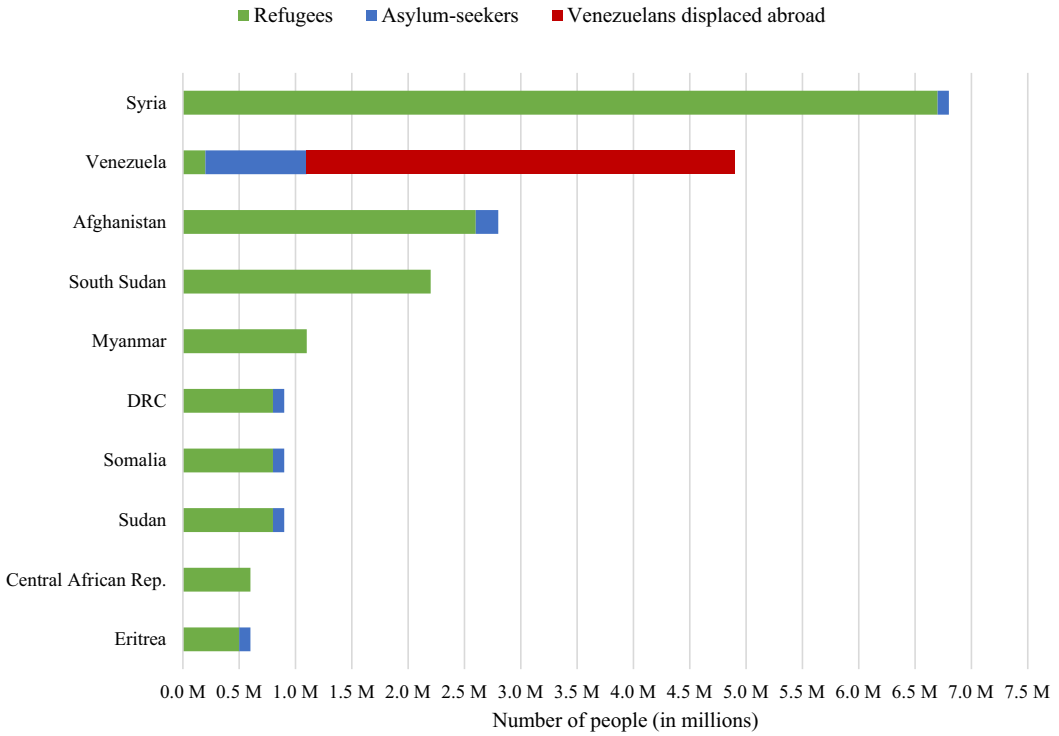


Fig. 2 Distribution of displaced persons by country of origin for the majority (68%) of refugees and Venezuelans placed abroad based on the 2020 Global Report of the United Nations High Commissioner for Refugees (UNHCR, 2020).



*Excludes Palestine refugees under UNRWA's mandate

Fig. 3 International displacement by country of origin at the end of 2020* modified from the 2020 Global Report of the United Nations High Commissioner for Refugees (UNHCR, 2020). *Excludes Palestine refugees under UNRWA's mandate.

results in a broad range of clinical conditions in children and adults that result from deficiencies in one or more nutrients and/or energy. The link between acute malnutrition and excess mortality has been documented for decades. The close correlation between these two factors was demonstrated during a Somali refugee operation in Ethiopia in 1988–89 and later in other emergencies (CDC, 1992). During the period of peak incidence of mortality and high prevalence of acute undernutrition, access to adequate food was less than 1400 kcal person^{-1} day^{-1} instead of the recommended 1900 kcal person^{-1} day^{-1} at the time and limited access to safe water, sanitation, shelter, and densely crowded conditions. Although the immediate aim of most food assistance programs in refugee emergencies is to prevent excess mortality, there is also increasing evidence that undernutrition during critical periods of life has long-lasting effects. Undernutrition, or the risk of being undernourished, may be carried from one generation to another in an intergenerational cycle (Martorell and Zongrone, 2012). Undernourished women give birth to undernourished infants who, in turn, are more likely to become undernourished adolescents and adults. Therefore, the nutritional status of refugees can have long-lasting effects on future health for individuals and generations.

Macronutrients

An inadequate supply of macronutrients and micronutrients (protein, fat, carbohydrates, and vitamins and minerals) results in protein-energy malnutrition (PEM), the most common form of malnutrition, especially among infants and young children. The two main forms of growth failure associated with PEM includes wasting (acute undernutrition) and stunting (chronic undernutrition). Wasted individuals, defined as a weight for height < -2.0 Z-scores, are extremely thin, whereas stunted individuals, defined as height for age < -2.0 Z-scores, are short for their age as a result of impaired growth during childhood. Severe PEM has a high case fatality rate and is often classified into two forms, marasmus and kwashiorkor. Both conditions are identifiable by severe weight loss, however, the oedema associated with kwashiorkor can mask the otherwise dramatic skeletal appearance of marasmic individuals. Nutritional issues among refugees vary greatly from one region of the world to another. The prevalence of wasting, defined as weight for height less than -2 standard deviations of the reference population, has been as high as 50% in the Horn of Africa and as low as 5% in Southeast Asia, Malawi, and the Persian Gulf. Mortality rates in some of these populations during the acute phase of displacement have been extremely high, up to 60 times the expected rates.

Micronutrients (vitamins and mineral deficiencies)

Micronutrients, although needed in small amounts, are as essential as macronutrients in addressing nutrition requirement of populations through food assistance programs. Micronutrient deficiency diseases are key in nutrition-related morbidity and mortality. There is a misconception that people do not die of micronutrient deficiencies because one does not often see the signs and symptoms as visibly as those for PEM. Nonetheless, these deficiencies can be fatal. In addition to deficiency diseases of vitamin A, iron, and iodine, conditions widely described as diseases of public health importance, epidemics of scurvy, beriberi, and pellagra, have been frequently reported among refugee populations, primarily because of limited access to a diverse diet and overreliance on one or two commodities (i.e., maize or polished rice) (**Table 1**). The importance of micronutrient deficiencies among the refugee population has only been documented since the late 1980s and more attention is being paid to the usefulness of inclusion of micronutrient-rich foods and/or supplements in the management of refugee nutrition. There are additional innovations for delivering micronutrients for home fortification. This represents one of the best potential opportunities to increase impact on child development and saving lives. Micronutrient powders, for example, known under the trade names MixMe and Sprinkles, are provided to fortify foods after preparation, just before consumption, to ensure an adequate intake of micronutrients essential for body functions. Micronutrient powders come in small individual sachets containing one full recommended nutrient intake of 15 vitamins and minerals in one gram of powder for one person per day and are currently used in many refugee and emergency operations.

Conflicts and refugees and long-term health

The health implications of being a refugee or IDP are centered around issues of personal safety and food security given the often tenuous living conditions either in refugee camps or as a refugee. Yet, migration of refugees is major concern to host countries as many refugees may have experienced nutritional and emotional challenges before arriving. This dramatic transition to a new environment with improved access to processed foods or even continued food insecurity due to financial or language challenges may create the background against which elements of the "developmental origins" become manifest and result in nutrition-related chronic diseases among refugee immigrants (Hoffman et al., 2021).

Clearly, as has been witnessed recently in the Russian invasion and occupation or Ukraine, IDP or refugees are often exposed to poor nutritional environments prior to emigrating. For example, refugee immigrant children have a five-fold higher risk of being stunted compared to non-refugee immigrant children (Lane et al., 2018). As well, one study reported that 30% of refugee children were growth retarded (Sarr et al., 2014). Indeed, 30% of Sudanese refugee children are reported to have low bone mineral content and elevated cholesterol, triglycerides, and insulin resistance (Alasagheirin and Clark, 2018). Many refugees suffer psychosocial and nutritional stressors and 40% of refugees have reported past and current food insecurity, making dietary intake and quality important factors to consider. Finding appropriate residence relieves these stresses and the height and BMI z-score of refugee children increased during a 5-year follow-up. Interestingly, obesity may be a factor associated with refugee migration that may contribute to the prevalence of nutrition-related chronic diseases. In the US, refugee children had a higher BMI z-score compared to the low-income non-refugee children who attended the same health centers (Dawson-Hahn et al., 2016). Moreover, stunted refugee children experienced an increase in their BMI z-score yet, the control children had a decrease over one year. It is hoped that understanding the long-term health implications of being displaced may promote improved health and nutrition aid programs for refugees when civil strife and unrest prompt the unfortunate migration of people.

Table 1 Micronutrient deficiencies in refugees.

Micronutrient	*Deficiency disease*	*Symptoms*
Iron	Anemia	Pallor, tiredness, headaches, breathlessness
Iodine	Goiter	Swelling of the thyroid gland in the neck
	Cretinism	Severe mental and physical disability that occurs in the offspring of women with severe iodine deficiency
Vitamin A	Night blindness	Inability to see well in the dark—an early sign of vitamin A deficiency
	Xerophthalmia	Including Bitot spots and corneal ulceration and night blindness
Niacin	Pellagra	Affects the skin, gastrointestinal tract, and nervous system and is sometimes called the 3Ds: dermatitis, diarrhea, and dementia
Thiamin	Beriberi	Loss of tendon reflexes; drooping of arms and feet; wet or cardiac beriberi resulting in heart failure
Vitamin C	Scurvy	Painful joints, swollen and bleeding gums, and slow healing or reopening of old wounds

Addressing nutrition in refugees

Owing to the nature of displacement and the loss of livelihoods, refugee populations are extremely vulnerable. Often, refugees settle in camps with support from the international community and host government. In some cases, refugees may live in open situations in which they integrate into the local community. In almost all cases, refugees are dependent on outside assistance, although the level of need depends on the level of self-reliance the refugees can achieve. In some instances, refugees are able to bring some assets with them when they flee and/or have some sort of income-generating activity, such as access to land and labor and employment. However, this very much depends on the policies of host governments. In these cases, refugees are not totally dependent on external assistance, and nutrition management response takes these factors into account by adjusting humanitarian assistance and the food assistance in particular to meet the assessed needs. There has been an evolution in the standards of food energy required for refugee populations as pervious standards were based on estimates of energy requirements for body weight, demographic composition, environmental temperature, and activity levels. In the 1980s, the standard was approximately 1500 kcal person^{-1} day^{-1}, the minimum deemed adequate for survival. In the late 1980s, this was recalculated to 1900 kcal person^{-1} day^{-1} as a preliminary standard to include expenditure of energy for light activity as opposed to merely the basal metabolic rate. In the 1990s, the benchmark value was modified to a more realistic 2100 kcal/person/day (FAO, 2004). The revised estimate was based on an increase in energy required for physical activity, some adjustments to the demographic composition, and an increase in the proportion of pregnant and lactating women in the population. It should be recognized that this recommended value is the average of the individual requirements based on developing nation population demographics and is not a specific provision for individual needs. In addition to the recommended kilocalorie content of the daily food ration, nutritional science has determined that the ration should have an optimal balance of fat and of protein (17 and 12%, respectively).

Food baskets for populations (general food distribution)

The sudden and massive reduction in food availability associated with displacement immediately affects the nutritional status of refugees. The first response is intervention through the implementation of adequate food distribution to all to ensure all refugees have access to the required food ration. A general food distribution is the first line of intervention and the highest priority when a refugee population does not have access to sufficient food to meet its nutritional requirements. If the recommended adequate ration of 2100 kcal person^{-1} day^{-1} and the quality of food basket is not available, malnutrition levels may escalate.

Even if the overall food needs of refugees are adequately met, inequities in the distribution system, disease, and various social factors may contribute to a high level of malnutrition among certain groups. Children younger than 5 years of age, pregnant and lactating women, the chronically ill (e.g., tuberculosis and HIV/AIDS patients), and the elderly are considered vulnerable groups since they have specific nutritional requirements. A special nutrition intervention program targets these nutritionally vulnerable groups through supplementary feeding programs and those in need of nutritional rehabilitation through therapeutic feeding programs. Malnutrition prevalence, as well as an assessment of aggravating factors in the environment, are used as guidelines to determine if a nutrition intervention program needs to be initiated. Aggravating factors that influence the nutritional situation include an elevated crude mortality rate; epidemics of communicable diseases such as measles, diarrheal diseases, and respiratory infections; and an unstable social, political, or environmental situation.

Nutrition programs are primarily managed by nongovernmental organizations (NGOs) that have specialization in the management of refugee nutrition. Other humanitarian partners have roles to play in response to refugee situations, including host country authorities, United Nations agencies such as the World Food Program, UNHCR, and UNICEF; and other multisect oral NGOs. There exist memorandums of understanding and partnership agreements among agencies to provide assistance to the affected populations. The need for partnership is essential for management of refugee nutrition and health programs. The Sphere Project was launched in 1997 to develop a set of universal minimum standards in vital sectors of humanitarian assistance. The aim of the project is to improve the quality of assistance provided to affected populations and to enhance the accountability of the humanitarian system in emergency response.

Addressing acute malnutrition among young children

Management of moderate acute malnutrition

The most common nutrition intervention is supplementary feeding programs (SFPs) to address moderate acute malnutrition in emergency situations. SFPs provide a high-quality food as a nutritional supplementation to the daily diet of malnourished populations. There are two main types of SFP—targeted and blanket. The goal of targeted supplementary feeding is to prevent people who are moderately malnourished from becoming severely malnourished. Blanket supplementary feeding, which provides all members of a vulnerable group with a food supplement, is intended to prevent the deterioration of nutritional status among a larger population group rather than narrowly defined individuals at specific nutritional risk. Implementation of SFPs can take two forms: Prepared meals consumed on site (wet rations) or food rations issued weekly or monthly to take home for preparation (dry rations). Food supplements usually consist of a fortified blended food (FBF) mixed with oil, and sometimes sugar is included. Wet rations should provide 500–700 kcal, whereas the recommended dry ration is doubled to 1000–1200 kcal to account for sharing at home. There have been many advances in improving the food commodities to address nutritional needs for SFPs. The composition of

fortified blended food has been enhanced by including animal proteins (dried milk) and additional vitamins and minerals. In addition, ready to use lipid nutrient products—such as plumpy doz, supplementary plumpy, nutri butter, and other ready to use supplementary foods for children are proven to be more effective for improving nutritional status and speed of recovery.

Addressing severe acute malnutrition

Therapeutic feeding programs (TFPs) provide the severely malnourished with their full nutritional requirements in addition to medical care. They are initiated to reduce excess mortality among individuals facing severe malnutrition and have played an important role in reducing malnutrition-related mortality in emergencies. The first phase of a TFP focuses on treatment of infections, management of other medical complications, and metabolic stabilization. This phase has the highest mortality rate of all nutrition interventions due to the poor state of the patients and the intensive treatment required. The second phase of a TFP is a rapid weight gain period designed to rehabilitate the patient's nutritional status.

Recognition of severe acute malnutrition as a complex nutritional condition during the 1990s led to the development of certain foods defined explicitly for therapeutic treatment of malnutrition with the appropriate balance of energy, protein, and micronutrients in order to avoid overloading the body's metabolism, which potentially may lead to cardiac shock. These products include F-75, F-100, and BP-100 biscuits and other ready-to-use therapeutic foods (RUTFs) such as "plumpy nut."

Community-based care is a recently developed public health approach to deal with severe malnutrition and aims to treat the majority of people suffering from severe acute malnutrition in their homes. A community therapeutic care (CTC) program initially is set up complimentary to traditional TFP components and represents a new approach to managing malnutrition at the community level. A CTC program has the same initial metabolic stabilization phase, and life-threatening infections are identified and treated just as in a TFP. However, once the patient is stabilized, he or she moves directly to an outpatient therapeutic program that operates through existing health structures and, with the use of RUTFs, nutritional rehabilitation is initiated. When patients are no longer at risk of severe malnutrition, they are referred to SFPs for recuperation. This phase is followed by greater emphasis on community mobilization to increase the population's involvement and training of mothers. CTC is an innovative approach and is proven to be successful. Proposed benefits of this method are the improved coverage to increase the number of people treated and reduce overall mortality rates. In addition, local production of RUTFs has been initiated in few countries to reduce the cost of treatment and shorten the length of stay in centers away from the family. Finally, the decentralized nature of CTC can enable earlier detection of malnutrition, thereby reducing the incidence of severe malnutrition (**Table 2**).

Challenges

Nutrition interventions alone are not adequate to address the multiple causes of undernutrition in refugees. The access to public health inputs is essential in preventing and reducing excess mortality and malnutrition and in ensuring that nutrition interventions have the desired effects.

Although the quality of nutrition assistance and interventions has improved considerably since the 1970s, the international community is still searching and moving forward to improve the quality of food products provided to address undernutrition. This is an exciting time in the field of nutrition to develop appropriate strategies, interventions, and products to address undernutrition. 1000 days is a global effort for addressing undernutrition during pregnancy and early childhood. This refers to a special window of opportunity to take action to combat undernutrition. 1000 days refers to the critical period of growth from conception through 2 years of age. Research shows that children who are undernourished during this period are far more likely to suffer from long-term health problems, poorer education performance, and lower economic prosperity.

Given that a number of factors, such as food, health care, and the environment, interact to determine nutritional well being, a partnership among agencies with different mandates is essential to effectively address and correct nutritional issues. Refugee nutrition must be addressed in tandem with other services to ensure that underlying factors of malnutrition are being met and that nutrition interventions are effective.

Summary

The world in 2022 faces challenges that were outside of our collective imagination even in 2018. The combined effects of a global pandemic, outbreak of war in Ukraine, along with the stress of economic sanctions and rising oil and food prices are far-reaching and have the potential to reverse decades of positive changes in the number of displaced persons. While existing aid programs continue to receive adequate funding, recent events should be a reminder for all members of our modern society to support programs that aid the least fortunate of our planet. Providing basic medical and nutritional support to refugees and IDPs is a moral imperative for the global community and the work by volunteer organizations and multi-lateral groups, such as WFP, are vital to promoting health and safety during times of crisis.

Table 2 Milestones in addressing nutrition in refugees.

1960s:	Food response based on commodities available (donated)
	Limited recognition of relevance of nutritional content of rations
	Food provided based on resources rather than nutritional needs
1970s:	Focus on protein deficiency (in protein-energy malnutrition)
	Food ration comprised mainly cereal, pluses/beans, and oil
	Fortified blended foods (FBFs) used only in supplementary feeding
1980s:	Major relief agencies raise planning figure from 1500 to 1900 kcal person^{-1} day^{-1}
1990s:	Relief agencies raise planning figure for fully food aid-dependent populations from 1900 to 2100 kcal person^{-1} day^{-1}
	FBF included in most rations for completely dependent populations
	Basic six-commodity food basket becoming common: cereal, pulses, oil, salt, sugar, FBF
	UNHCR/WFP memorandum of understanding signed with clear roles and responsibilities
	Development of multi-UN agency and NGO policies and guidelines on common approaches to addressing malnutrition in emergencies
	Fortification of oil, salt, and flours, on international market
	Development of therapeutic foods for treatment of malnutrition (F100-F75)
	Local production of fortified blended foods
2000s:	Development of community management of acute malnutrition (CMAM)
	Creation of inter-agency clusters (IASC)—Nutrition, food security, health, wash, education, protection, logistics, and telecommunication
	Development of capacity in nutrition in humanitarian staff
	Pilot testing of on-site milling and fortification in a refugee camp
	Development of ready to use foods
	Provision of multimicronutrient powders
	Provision of cash and or cash vouchers to address malnutrition

References

Alasagheirin, M.H., Clark, M.K., 2018. Skeletal growth, body composition, and metabolic risk among North Sudanese immigrant children. Publ. Health Nurs. 35, 91–99.

Balsari, S., Dresser, C., Leaning, J., 2020. Climate change, migration, and civil strife. Curr. Environ. Health Rep. 7 (4), 404–414.

CDC, 1992. Famine-affected, refugee, and displaced populations: recommendations for public health issues. MMWR 41 (RR-13).

Dawson-Hahn, E., Pak-Gorstein, S., Matheson, J., Zhou, C., Yun, K., Scott, K., Payton, C., Stein, E., Holland, A., Grow, H.M., Mendoza, J.A., 2016. Growth trajectories of refugee and nonrefugee children in the United States. Pediatrics 138.

Food and Nutrition Technical Report Series 1, Rome, Italy FAO Human Energy Requirements Report of a Joint FAO/WHO/UNU Expert Consultation, 2004, ISBN 92-5-105212-3.

Hoffman, D.J., Powell, T.L., Barrett, E.S., Hardy, D.B., 2021. Developmental origins of metabolic diseases. Physiol. Rev. 101 (3), 739–795.

Lane, G., Farag, M., White, J., Nisbet, C., Vatanparast, H., 2018. Chronic health disparities among refugee and immigrant children in Canada. Appl. Physiol. Nutr. Metabol. 43, 1043–1058.

Martorell, R., Zongrone, A., 2012. Intergenerational influences on child growth and undernutrition. Paediatr. Perinat. Epidemiol. 26 (Suppl. 1), 302–314.

PRIO Conflict Trends, Trends in Armed Conflict, 1946–2020. www.prio.org/ConflictTrends.

Sarr, O., Thompson, J.A., Zhao, L., Lee, T.-Y., Regnault, T.R.H., 2014. Low birth weight male Guinea pig offspring display increased visceral adiposity in early adulthood. PLoS One 9, e98433.

UNHCR, 2020. Global Trends Forced Displacement in 2020. https://www.unhcr.org/flagship-reports/globaltrends/.

Further reading

Collins, S., 2001. Changing the way we implement famine relief programs. Lancet 358, 498–501, 9280.
Dee Pee, S., Moench-Pfanner, R., Martini, E., Zlotkin, S., Darton-Hill, I., Bloem, M.W., 2007. Home fortification in emergency response and transitioning programming: experiences in Aceh and Nias, Indonesia. Food Nutr. Bull. 28, 189–197.
Mason, J., 2002. Lessons on nutrition of displaced people. J. Nutr. B2 (7), 20963–21035.
Medecins Sans Frontiers, 1995. Nutrition Guidelines. Medecins Sans Frontiers, Geneva, Switzerland.
Medecins Sans Frontiers, 1997. Refugee Health: An Approach to Emergency Situations. Medecins Sans Frontieres, Paris, France.
Sphere Project, 2004. Humanitarian Charter and Minimum Standards in Disaster Response. Sphere Project, Geneva, Switzerland.
UNHCR/UNICEF/WFP/WHO, 2002. Food and Nutrition Needs in Emergencies. World Health Organization. Joint publication UNHCR/UNICEF/WFP/WHO, Geneva.
WHO, UNICEF, WFP, SCN, 2007a. Joint Statement on Community Management of Severe Malnutrition.
WHO, WFP, UNICEF, 2007b. Joint Statement on Preventing and Controlling Micronutrient Deficiencies in Populations Affected by an Emergency. Geneva, Switzerland.
WHO, UNICEF, WFP, UNHCR, 2009. Consultation on the dietary management of moderate malnutrition in under-5 children. Food Nutr. Bull. 30 (Suppl. 3).
WHO/UNHCR, 1988. Nutrition in Times of Disaster. Report of an International Conference Held at World Health Organization Headquarters, Geneva, Switzerland, September 27–30. World Health Organization, Joint publication—WHO/UNHCR, Geneva.
World Health Organization, 1999. Management of Severe Malnutrition: A Manual for Physicians and Other Senior Health Workers. World Health Organization, Geneva.
World Health Organization, 2000. The Management of Nutrition in Major Emergencies. World Health Organization, Geneva.

Double burden of malnutrition

Daniel J. Hoffman, Angelica A. Pozzoli, and Hannah R. Posluszny, Rutgers University, New Brunswick, NJ, United States

© 2023 Elsevier Ltd. All rights reserved.

Introduction 401
Epidemiology of the DBM 402
Nutrition transition 405
Ultra-processed foods 406
Obesity 406
Activity 407
Micronutrient deficiency 407
Food systems 407
Marketing 407
Food supply 408
Sugar-sweetened beverages 408
Summary 409
References 409

Key points

- The double burden of malnutrition is defined as the coexistence of under- and overnutrition.
- Populations undergoing a rapid nutrition transition, notably those living in low- and middle-income countries, are disproportionately affected by the double burden of malnutrition.
- The global increase in modernization, urbanization, and economic development has contributed to a sedentary lifestyle and an increased intake of processed foods.
- "Hidden hunger" is characterized by the co-occurrence of micronutrient deficiency and obesity at the individual level.
- Conflict, climate change, and economic downturns are challenges to global food systems, threatening food security.

Introduction

Over the past 50 years, there has been a worldwide shift in diet and lifestyle that has contributed to changes in human nutrition and global health. A decrease in the prevalence of infectious disease and undernutrition marked the beginning of the nutrition transition and an increased prevalence of nutrition-related chronic diseases (NRCDs) throughout the world (Popkin and Gordon-Larsen, 2004). As the 21st century began, the prevalence of NRCDs continued to increase while the prevalence of undernutrition decreased, but at a slower rate. The world now faces a complicated coexistence of over- and undernutrition occurring within the same households, communities, and countries, most particularly in low- and middle-income countries, a phenomenon terms the double burden of malnutrition (DBM).

Yet, the COVID-19 pandemic that began in late 2019 and continued through 2022, resulted in a major disruption of supply chains and humanitarian aid along with a near cessation of migration from lower income to higher income countries. Almost all countries in the world implemented quarantines and limited international travel in the hope of quelling the spread of the virus. These measures created drastic situations for people living in marginal communities and for those who could not transition to "work from home" such that food insecurity increased and the risk of contracting COVID remained high. This disproportionate burden was further exacerbated when vaccines became available since distribution was primarily in higher income countries before beginning in the poorest countries of the world. Thus, many gains that were made in the fight against obesity and undernutrition were muted or reversed during the pandemic. These current events will have profound effects on the state of the nutrition in the world, but exactly what those effects may be, such as a decrease in the prevalence of obesity or an increase in the prevalence of food insecurity and wasting, will not be known for several years.

Given the social and technological changes over the past several decades, it is important to consider not only the causes of the DBM, but to understand how the DBM is affecting the health and economies of countries across the world. In particular, what once seen as a continued decrease in undernutrition since the end of World War II, often linked to sustained economic growth, is now being challenged by both the COVID-19 pandemic and ongoing civil conflict and economic sanctions placed against Russia following its invasion of Ukraine in February of 2022. In short, while aspects of the DBM become better understood, the task of

preventing and treating nutrition-related chronic diseases (NRCDs) is becoming more important than ever. To stem the tide of food insecurity and undernutrition, efforts to prevent undernutrition and obesity, as well as micronutrient deficiencies, need to be made given that many of the same factors promoting these conditions also promote overnutrition.

Epidemiology of the DBM

The DBM is characterized by the co-occurrence of undernutrition and overweight, obesity, and other NRCDs and can manifest at the individual, household, and population level. Briefly, population-level DBM is defined by a high prevalence of both undernutrition and overweight and obesity in the same community, region, or nation. The double-burden household, most common in middle-income countries undergoing a rapid nutrition transition, pertains to the co-existence of under- and excess nutrition within the same household. Individual-level DBM occurs by the simultaneous or temporally separated development of two or more forms of malnutrition. Here we will further explore factors that influence the DBM at these three levels.

At the population level, the DBM affects most LMICs. A coexistence of under-and overnutrition is especially prevalent in sub-Saharan Africa, South Asia, East Asia, and the Pacific. Between the 1990s and 2010s, the DBM worsened particularly in Asia and improved in Latin America, the Caribbean, the Middle East, and North Africa (Popkin et al., 2020). Among the countries with a DBM, there was an observable increase in the prevalence of adult overweight and a decline in child wasting and stunting between the 1990s and 2010s as shown in **Fig. 1**A and B. The lowest income LMICs are facing a growing DBM that is primarily driven by the rapidly increasing prevalence of overweight and relatively slow decline in childhood wasting and stunting and undernutrition. It is noteable that the percent of deaths due to non-communicable diseases is 87% in developed regions and as low as 28% in Africa and as high as 83% in Eastern Asia. At the same time, infant mortality continues to decline throughout the world while adult mortality has shown consistent declines with some increases in parts of Africa and Asia. The cumulative effect of these changes is a generally healthy population with fewer undernourished, but a growing population of over nourished, children and adults.

There are growing disparities between the lowest and highest wealth quartiles associated with the DBM. It has been reported that members of the lowest wealth quartile face a greater prevalence of overweight and obesity than those in the highest quartile (**Fig. 2**A and B). Most countries in Latin America, the Caribbean, Eastern Europe, Central Asia, and East Asia, particularly in China and Indonesia, have reported an increased prevalence of overweight and obesity among lower-wealth households (Popkin et al., 2020). On the other hand, in Sub-Saharan Africa and South Asia, the largest increase in obesity prevalence occurred among higher-wealth households. Generally, rates of overweight and obesity are growing rapidly in low-income countries that still have a relatively high prevalence of stunting, wasting, or thinness. Thus, the DBM is exacerbated in those countries that lack the wealth and potential resources necessary to mitigate undernutrition.

Economic change has been central to the reduction in wasting, stunting, and undernutrition as many development programs in LMICs have improved sanitation, vaccination programs, and reduced food insecurity. However, as a key component of the nutrition transition, economic change is also associated with important shifts in both physical labor and food systems, which have resulted in lower physical activity and an increased consumption of energy-dense diets (Popkin and Gordon-Larsen, 2004). Diet, which plays a critical role in the growing prevalence of obesity and non-communicable diseases, is a key factor in the DBM. Evidence suggests that diets largely consisting of ultra-processed foods, rich in refined carbohydrates, fat, sugar, and salt, are increasingly accessible in countries undergoing economic change. As LMICs economically expand, they face significant shifts, particularly in the composition of their diet, promoting obesogenic behavior (**Fig. 3**).

DBM at the household level is defined as one or more individuals with wasting, stunting, or underweight and one or more individuals with overweight or obesity within the same household. Most commonly, household-level DBM involves the combination of overweight or obese mothers and stunted children. The presence of household-level DBM varies greatly between different countries such that the reported prevalence of household-level DBM ranges from 0 to 26.8%. The prevalence of overweight mother and stunted child pairs, however, remained less than ten percent in most countries. In comparison to African countries, Asian countries have been found to have a lower percentage of overweight mother and stunted child pairs. Moreover, evidence suggests that middle-income countries tend to have higher percentage of the double-burdened dyad. However, new data indicates that the incidence of household-level DBM is increasing at a faster rate in lower-income compared to middle-income countries.

A number of factors are associated with the "double burden" household, including urban residence, income, and education (WHO, 2017). Families who live in urban areas are more likely to report a co-occurrence of malnutrition at the household level. As well, urban areas generally provide greater access to processed or fast foods and less physically demanding jobs. Households facing the double burden of under- and excess nutrition tend to also have higher income than undernourished households. As well, maternal or household-head education level is associated with the household-level DBM. Indeed, highly educated mothers are less likely to have stunted children, and are therefore, less likely to reside in double-burden households. Importantly, social factors generally associated with household-level DBM do not exist independent of each other. A meaningful interaction among urban residence, income, and education level generally exists, however, this relationship has not consistently translated into the research regarding household-level DBM.

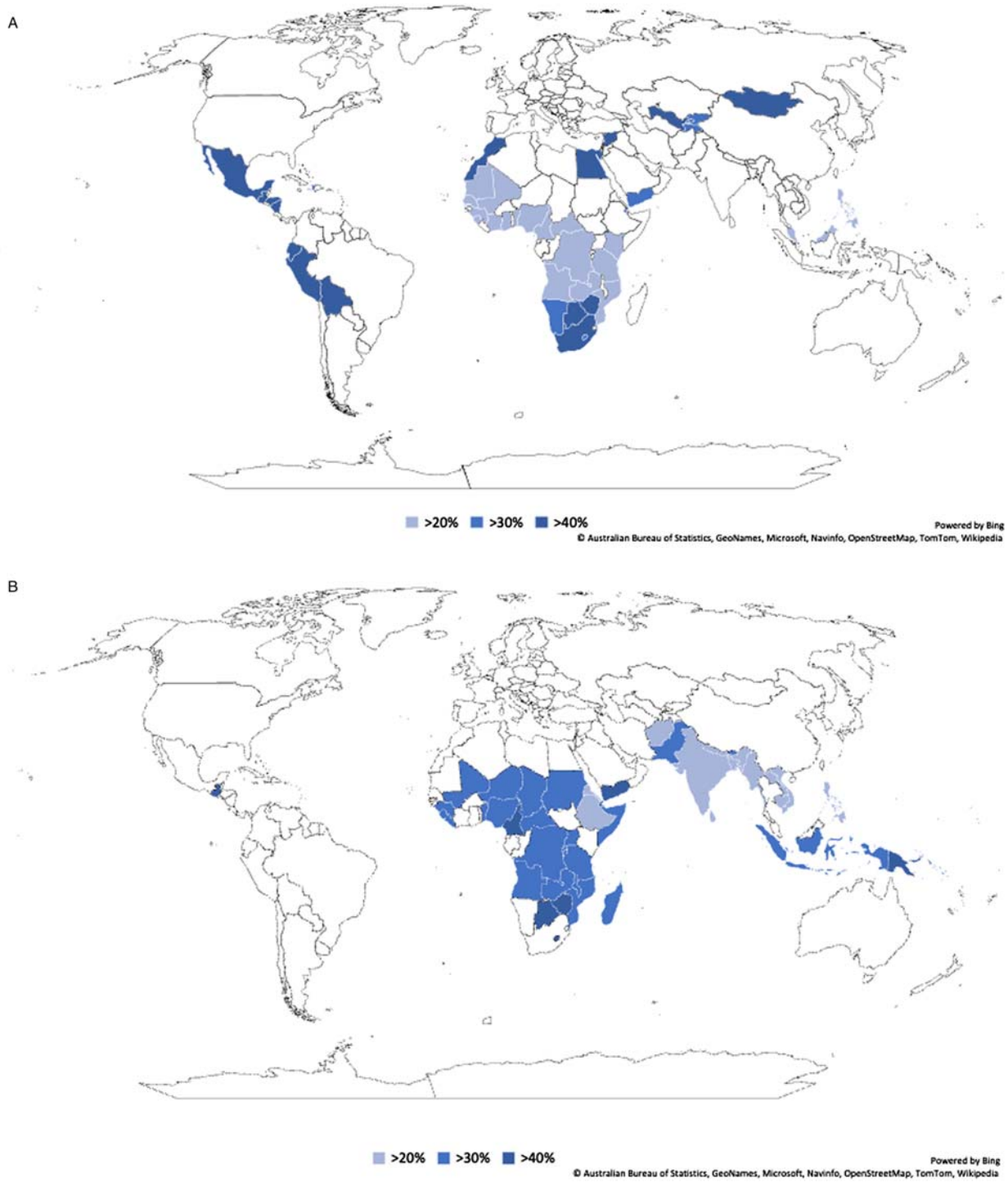
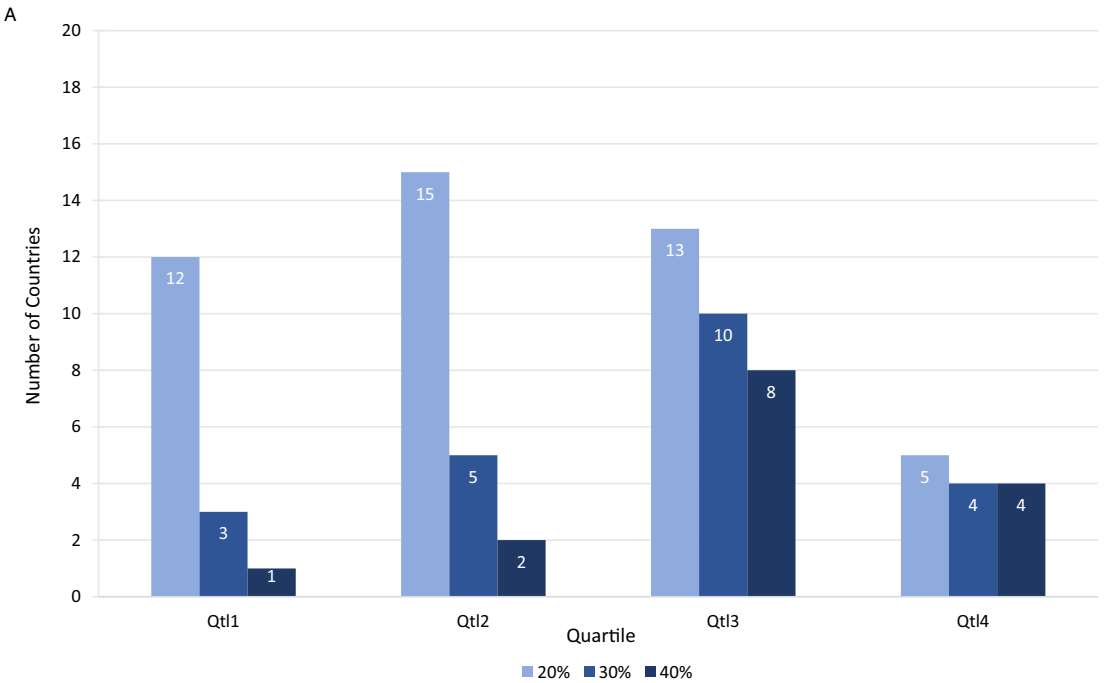
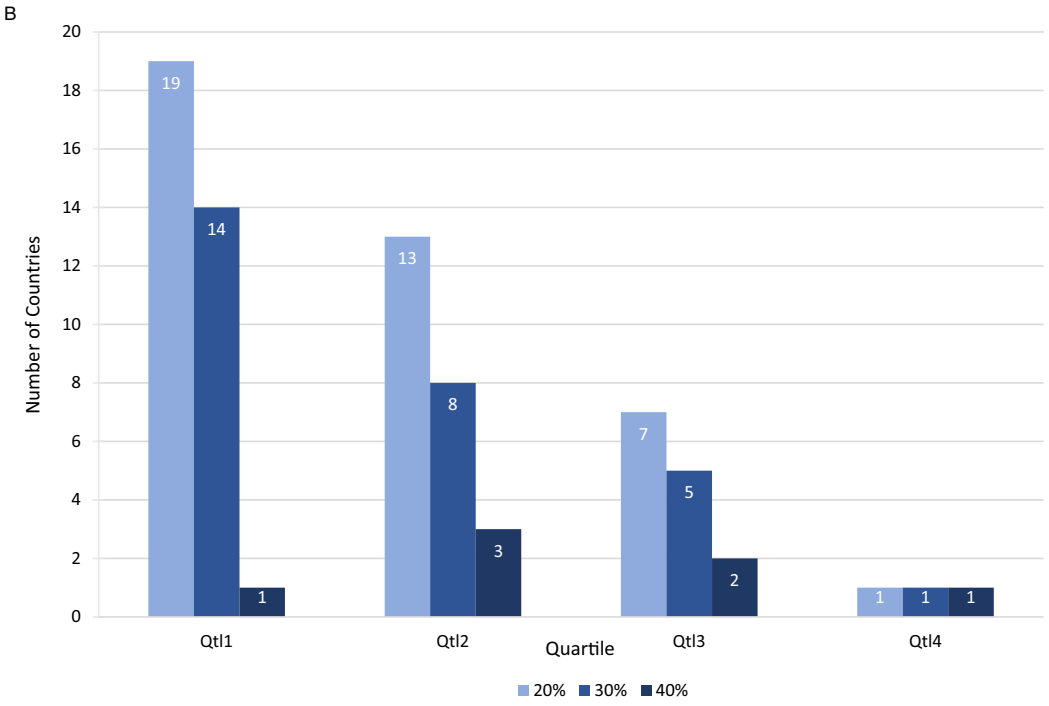


Fig. 1 (A) Global prevalence of the double burden of malnutrition, Global double burden of malnutrition based on height/weight data in the 1990s. (B) Global prevalence of the double burden of malnutrition, Global double burden of malnutrition based on height/weight data in the 2010s. Adapted from Popkin et al. (2020).

At the individual level, different forms of malnutrition can develop simultaneously or across the life course. While many LMICs experience chronic undernutrition, individuals most commonly exposed to nutritional deficiencies now find themselves in an obesogenic environment, consuming a diet void of essential micronutrients, due to rapid population-level nutrition transition. It has also been suggested that the intergenerational emergence of malnutrition can influence DBM at the individual level. It has been

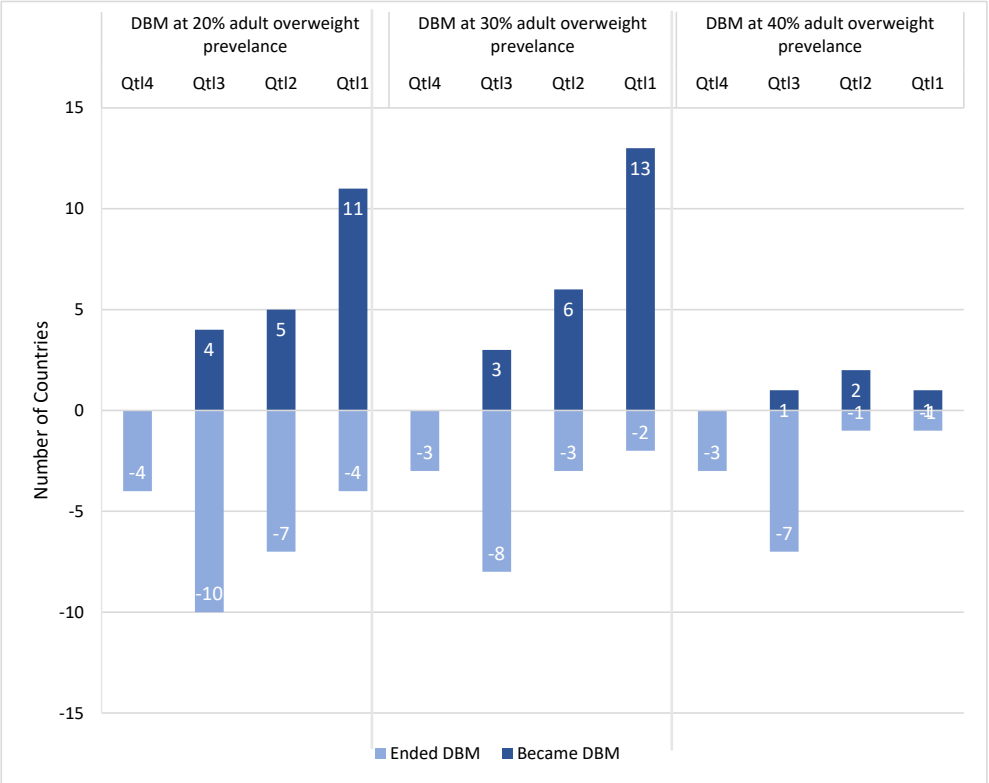


*DBM = at least 1 child, adolescent, or adult in the household with severe levels of wasting/stunting/thinness and 1 with overweight/obesity; only includes countries with DBM data available for both time periods (1990s & 2010s)
**Quartile (Qtl) = Qtl1 is lowest wealth, Qtl4 is highest wealth



*DBM = at least 1 child, adolescent, or adult in the household with severe levels of wasting/stunting/thinness and 1 with overweight/obesity; only includes countries with DBM data available for both time periods (1990s & 2010s)
**Quartile (Qtl) = Qtl1 is lowest wealth, Qtl4 is highest wealth

Fig. 2 (A) Countries with high double burden of malnutrition* by prevalence of overweight adults and GDP/capita (PPP) quartile** in the 1990s as. *DBM = at least 1 child, adolescent, or adult in the household with severe levels of wasting/stunting/thinness and 1 with overweight/obesity; only includes countries with DBM data available for both time periods (1990s & 2010s). **Quartile (Qtl) = Qtl1 is lowest wealth, Qtl4 is highest wealth. (B) Countries with high double burden of malnutrition* by prevalence of overweight adults and GDP/capita (PPP) quartile** in the 2010s as. *DBM = at least 1 child, adolescent, or adult in the household with severe levels of wasting/stunting/thinness and 1 with overweight/obesity; only includes countries with DBM data available for both time periods (1990s & 2010s). **Quartile (Qtl) = Qtl1 is lowest wealth, Qtl4 is highest wealth. Adapted from Popkin et al. (2020).



*DBM = at least 1 child, adolescent, or adult in the household with severe levels of wasting/stunting/thinness and 1 with overweight/obesity; only includes countries with DBM data available for both time periods (1990s & 2010s)
**Quartile (Qtl) = Qtl1 is lowest wealth, Qtl4 is highest wealth

Fig. 3 Number of countries that change double burden of malnutrition* status from the 1990s to 2019s by GDP/capita (PPP) quartile** as. *DBM = at least 1 child, adolescent, or adult in the household with severe levels of wasting/stunting/thinness and 1 with overweight/obesity; only includes countries with DBM data available for both time periods (1990s & 2010s). **Quartile (Qtl) = Qtl1 is lowest wealth, Qtl4 is highest wealth. Adapted from Popkin et al. (2020).

shown previously that chronic maternal undernutrition can compromise fetal growth and increase the risk of childhood wasting or stunting. In addition, maternal obesity, particularly compounded with gestational diabetes, is associated with elevated fetal adiposity and micronutrient deficiency.

Early nutritional insults alter the long-term consequences of subsequent exposure to obesity and its metabolic impacts (Hoffman et al., 2021). For instance, rapid weight gain following early undernutrition might predispose an individual to central adiposity and noncommunicable disease. Moreover, intrauterine growth restriction can expose the fetus to oxidative stress and altered metabolism, which can affect various health-related factors. For instance, early nutritional insult can affect insulin metabolism and hypothalamic function. Therefore, infants with low birthweight, born sensitive to insulin, are susceptible to insulin resistance later in life due to rapid weight gain in childhood. Evidence generally suggests that there is a developmental link among stunting, obesity, and non-communicable disease. Various mechanisms have been proposed regarding the fetal origins of individual-level DBM. First, it is thought that by sparing energy, early undernutrition promotes survival through endocrine changes that affect growth, energy expenditure, and body composition. These changes then interact with the diet composition. A potential physiological mechanism for these observations is that poor growth prompts a metabolic adaption through a reduced capacity for fat oxidation, which manifests in stunted children as reduced lean mass and increased central adiposity under favorable environmental conditions (Hoffman et al., 2021). Indeed, individual-level DBM can be multifactorial and it requires the consideration of intergenerational malnutrition and interconnected factors to best prevent the vicious cycle of poor growth and poverty.

Nutrition transition

Over the past 50 years, increased modernization, urbanization, and economic development in many LMICs has influenced a rapid shift in dietary intake and physical activity (Popkin and Gordon-Larsen, 2004). This shift, known as the "nutrition transition," is

Table 1 Stages of the nutrition transition.

	Stages		
Characteristic	*Pre-transition*	*Transition*	*Post-transition*
Diet (prevalent)	Grains, tubers, vegetables, fruits	Increased consumption of sugar, fats, and processed foods	Processed foods with high content of fat and sugar; low fiber content
Nutritional problems	Undernutrition and nutritional deficiencies predominate	Undernutrition, nutritional deficiencies, and obesity coexist	Overweight, obesity and hyperlipidemia predominate

Adapted from FAO, UNICEF, WFP, WHO, 2018. The State of Food Security and Nutrition in the World 2018. Building Climate Resilience for Food Security and Nutrition. FAO, Rome.

characterized by an increase in the consumption of processed foods that are energy-dense, high in sugar, salt, and saturated fats, and low in fiber (FAO et al., 2018). Such shifts increase one's risk of overweight, obesity, and NRCDs. The nutritional profiles and health outcomes related to the 3 stages of the nutrition transition summarized in **Table 1**.

Ultra-processed foods

In 2009, the NOVA classification system for food was developed to provide an objective profile of food products, focusing on the degree of processing (Monteiro et al., 2019). The NOVA system classifies food according to one of the following four groups: Group 1, unprocessed or minimally processed foods (e.g. fresh or frozen vegetables, dried pasta, coffee, or herbs and spices); Group 2, oils, fats, salt, and sugar, or "Processed Culinary Ingredients"; Group, processed foods (e.g. canned or pickled vegetables, canned or cured meat and fish, fermented alcoholic beverages, or fruit and vegetable extracts); and Group 4, ultra-processed foods (e.g. fat, sweet, or salty snack, sugar-sweetened beverages, pre-prepared meals, or infant formulas). Part of the rationale for developing the NOVA system was to allow for dietary recommendations to be consistently based on the degree of processing and for research studies to use the system as a means of studying the role of ultra-processed foods (UPF) in health and disease. A key element of the NOVA system is that no single food or product in considered "off limits," rather the system exists under a broad "golden rule" that instructs consumers to "Always prefer natural or minimally processed foods and freshly made dishes and meals to ultra-processed foods." The processes and ingredients used to manufacture UPFs not only make them highly convenient and hyper-palatable, but also highly profitable to the manufacturers that produce them (Monteiro et al., 2019). The convenience, flavor and texture, and aggressive marketing of UPFs contribute to their increased consumption globally (Monteiro et al., 2019). A brief discussion below expands on this point and illustrates how UPF may be a major contributor to the DBM, especially in LMICs.

The consumption of UPF, has increased dramatically throughout the world as sales of UPF increased dramatically in the past two decades compared to high-income countries (Baker et al., 2020). For example, from 2000 to 2013, sales of UPF increased by 30% in Brazil, while sales dropped in the United States and Canada (−9% and −7.3%, respectively). In terms of total energy intake, the proportion of daily calories provided by UPF varies by country and is highest in high-compared to LMICs. For example, UPF account for 20–30% of daily energy intake in Spain, Lebanon, Brazil, and France with up to 55% in the United States with less than 20% in other countries. At the same time, several studies have shown that high consumption of UPF is associated with lower quality diets assessed by NOVA. As well, several studies have reported that UPF consumption is associated with an increased risk of overweight and obesity, and related conditions, such as hypertension, CVD, and cancer (Monteiro et al., 2019).

Obesity

Energy imbalance, related to the consumption of processed and ultra-processed foods, has led to an increased prevalence of overweight and obesity, affecting more than 672 million adults and 38 million children globally. Obesity is of particular concern give its association with NRCDs such as type II diabetes, cardiovascular disease, and hypertension (Monteiro et al., 2019). Worldwide, it has been estimated that approximately 11 million deaths each year are attributed to NRCDs (Afshin et al., 2019). A review of studies on UPF and obesity reported that in Sweden, there was a 142% increase in UPF intake between 1980 and 2020 and the prevalence of obesity for men and women increased from 25% to 56% and 26% to 29%, respectively (Elizabeth et al., 2020). It was also determined that for every percentage increase in UPF consumption, the prevalence of obesity increased by 0.25%. Similar results have been found in both cross-sectional and cohort studies. For example, high UPF intake was associated with a higher BMI in Brazil and France. As well, in both Spain and Brazil, a high intake of UPF was associated with a 26% higher risk of obesity compared to those with a low UPF intake. Thus, regardless of the study design, an increased intake of UPF not only appears to promote obesity, but there are also no positive nutritional outcomes associated with UPF.

Activity

In recent years, there has been a decline in physical activity associated with a number of factors such as increased sedentary behavior at work and at home as well as more "passive" forms of transportation (e.g. car, train, bus, etc.). The WHO defines physical activity as any form of movement that requires energy expenditure, including during leisure time, for transport to and from places, or as part of a person's work. To be sufficiently active, it is recommended that adults partake in at least 150 min of moderate-intensity activity, 75 min of vigorous-intensity activity, or any combination of the two per week. Insufficient physical activity is associated with a 20%–30% increased risk of death due to NRCDs such as those related to dietary intake. Globally, 28% of the total population and 80% of the adolescent-population is insufficiently active (Guthold et al., 2018). Therefore, promoting activity across the globes, especially for those living in urban settings, is a key preventive approach to stem the DBM.

Micronutrient deficiency

Approximately 1.5 billion people are affected by one or more micronutrient deficiencies worldwide (GBD 2015 Mortality and Causes of Death Collaborators, 2016). In the "pre-transition" nutritional environment, individuals were often characterized as being malnourished and micronutrient deficient, with visible symptoms such as stunting or wasting. During the nutrition transition, obese and overweight individuals, although seemingly well-fed, may also experience micronutrient deficiencies, referred to as "hidden hunger" since there are usually no visible signs of deficiencies. Indeed, some have proposed referring to the "triple burden of malnutrition" to highlight the co-occurrence of undernutrition, overnutrition, and micronutrient deficiencies. For example, iron deficiency anemia, which affects over 20% of women of reproductive age, is just one micronutrient deficiency that may be present in individuals who are overweight or obese as well as those who are underweight (Fig. 4). Thus, it is important to recognize that important micronutrient deficiencies are not limited to the undernourished and that "hidden hunger" is a central aspect of the DBM that merits research and prevention.

In summary, the DBM is clearly influenced by many structural and behavioral factors that make up the nutrition transition, such as the global exposure to and increased consumption of UPF. The confluence of biological and environmental factors that promote NRCDs, low physical activity, and obesity as well as those limiting growth or promoting undernutrition, illustrate the complexity of stemming the rise of the DBM, especially in LMICs that are undergoing the nutrition transition.

Food systems

Marketing

The modern global food environment consists of a variety of food and drink products that not only offer palatability, but also convenience and novelty. Of note, many of these foods are energy-dense and nutrient-poor, with a high fat, sugar, and/or salt

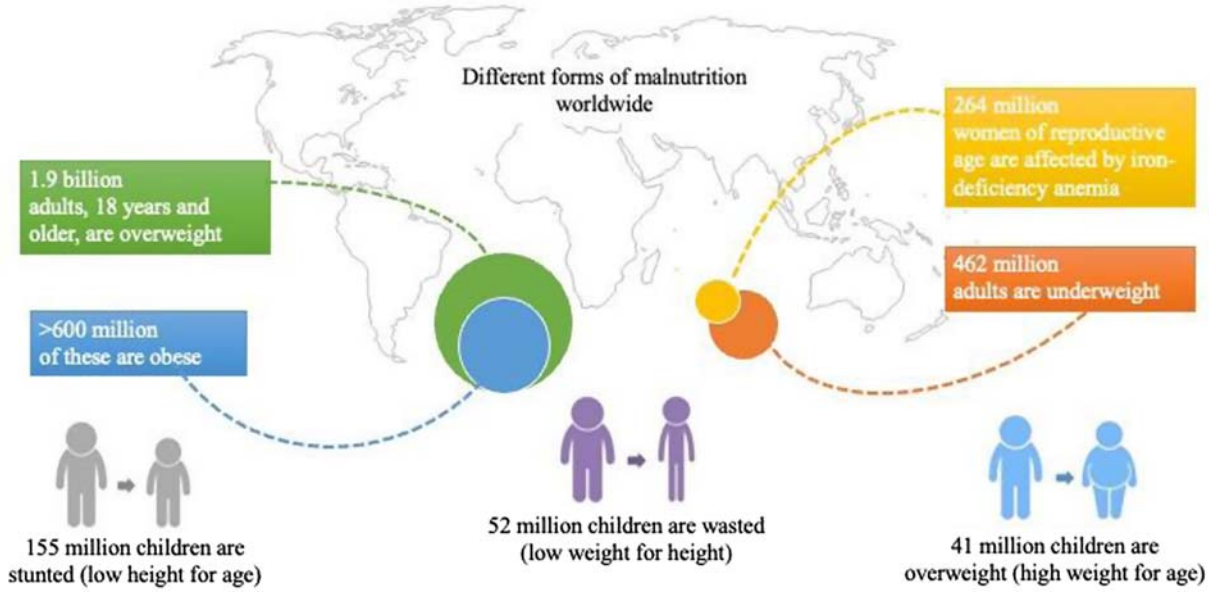


Fig. 4 Mapping the double burden of malnutrition as. Adapted from WHO (2017).

content, which is a major contributor to the rise in obesity, malnutrition, and NRCDs. Marketing of these products hinder efforts to promote a healthy diet and weight maintenance and have been considered one of the main drivers of unhealthy dietary patterns worldwide. A wide range of techniques are used to influence food preferences and dietary patterns, reaching individuals in schools, nurseries, and supermarkets; through the television and internet; and in many other settings.

The marketing of food has a significant impact on food preference and consumption, typically in ways that the consumer is not generally aware. An example of the adverse impacts of food marketing includes the promotion of breastmilk substitutes, which deters mothers from breastfeeding and weakens support for lactating women (FAO et al., 2021). Further, fast food companies tend to target much of their marketing toward youth, who are not developmentally equipped to understand the purpose of marketing and assess marketing claims. This has been of major concern as the largest increase in obesity is taking place among adolescents.

In 2010, the WHO developed recommendations to reduce the marketing of unhealthy foods to children (WHO, 2010). Over a decade later and very few countries have implemented such restrictions. Globally, the poor state of implementation is concerning, with notable consequences in LMICs, where corporations seek to enter new or under-developed markets. Indeed, through their global food marketing and promotion practices, fast food companies contribute significantly to the nutrition transition, and subsequent DBM, observed in LMICs.

It can be argued that the rising trends in obesity and NRCDs may be reversed by protecting youth from the marketing of energy-dense, nutrient-poor foods. To fully address the DBM, restrictions on food marketing to children could be part of a comprehensive package of policies to create food environments that encourage health and optimal nutrition, and also reduce malnutrition throughout the lifespan.

Food supply

Conflict, climate extremes, and economic downturns all present significant challenges to global food systems (FAO and WFP, 2019). Such challenges may occur through effects on systems supporting food production, food supply chains, food environments, consumer preferences, or a combination of these factors. Food security (the reliable access to safe, affordable, and nutritious food) is inextricably linked to predictable food systems; therefore, anything that threatens these systems poses serious risks to global health.

Conflicts contribute to food insecurity through the disruption of food systems in several ways. Among countries engaged in international and civil wars, starvation of civilians is often used as a method of warfare. The use of food as a weapon can be accomplished through the destruction of farms, livestock, and infrastructure or through civilian entrapment and isolation, resulting in food shortages and deprivation. Further, conflict and violence are major predictors of forced displacement(FAO and WFP, 2019). In situations where people are forced to flee their homes, access to, and availability of, food becomes insecure (UNHCR, 2020), creating an increased risk of hunger and malnutrition with long-term implications.

Climate extremes (i.e. drought, flood, heat spell, storm, fire) threaten the production of food around the world (FAO and WFP, 2019). As global temperatures continue to rise, incidents of extreme weather will become more severe, limiting humanity's ability to produce enough nutritious food. In several parts of the world, such as East Africa and parts of Latin America, extreme heat, severe weather, and drought has led to a steep decline in yield growth for wheat, maize, and other staple crops. If these trends continue, it is expected that global yields of these crops could decline by up to 30% by 2050 (Shukla et al., 2019). Individuals living in countries that are already vulnerable to food insecurity, such as LMICs, will continue to be disproportionately affected by climate change if action is not taken soon.

The global rise in food insecurity has been associated with the economic downturns caused by the COVID-19 pandemic (FAO et al., 2021). In addition to higher food costs due to unprecedented stress on food supply chains (i.e. decreased production, labor shortages, limited travel, etc.), the socioeconomic impacts of pandemic-related restrictions have heighted disparities related to food access, particularly among those living in LMICs. This has led to an increase in the consumption of highly processed foods, which tend to be cheaper than more nutritious perishable foods. Together, these factors have been associated with diets that are insufficient in amount and diversity, increasing the risk of nutritional deficiencies and the development of NRCDs long-term.

Sugar-sweetened beverages

Sugar consumption is increasing globally, primarily through an increased intake of sugar-sweetened beverages (SSBs). Indeed, frequent consumption of SSBs is associated with several NRCDs including type 2 diabetes, heart disease, and kidney disease (Basu et al., 2013). While intake of SSBs have either stabilized or declined in high-income countries, consumption of SSBs in LMICs have rapidly increased over the last 20 years. Trends in the consumption of SSBs may be attributed to their increased availability and affordability in LMICs. A global comparison of data found that, on average, the cost of SSBs tend to be cheaper than bottled water (Price et al., 2019). In LMICs, access to clean water is highly variable, particularly among low-income populations living in urban informal settlements. Due to widespread fear of contaminated tap water and limited access to clean water, SSBs tend to be considered the safer, more affordable option for people living in these settings.

Since SSBs are calorie-dense and nutrient-poor, their consumption contributes significantly to the double-burden of malnutrition. Generally, people do not compensate for their intake of SSBs by consuming fewer calories or healthier foods and regular consumption of SSBs likely contributes to a higher energy consumption than needed. Further, SSB consumption has been associated with a lower intake of nutrient-dense foods resulting in micronutrient deficiencies. To limit SSB consumption, several countries have implemented taxes on beverages with a high amount of sugar (FAO et al., 2021). "Soda taxes" not only limit the consumption of SSBs due to their increased price, but also act as a powerful driver for industries to reduce the sugar content of their products.

Summary

As the world slowly recovers from a global pandemic that caused millions of deaths and many more recovering from the COVID-19 infection, the true impact of the pandemic on global health has yet to be seen. It is estimated that the number of undernourished and food insecure children and adults in many parts of the world has increased due to quarantines, supply chain issues, economic crizes, and loss of employment, along with the lingering effects of "long COVID." Adding to this somewhat pessimistic perspective is the war that rages in Ukraine as this article goes to press.

The global impact of the Russian invasion of Ukraine will be quite severe as not only have exports of grain and other crops ceased, the daily and prolific use of mines and artillery has not only destroyed productive fields but threaten the current and immediate future crop plantings and harvesting. The fact that almost 10% of the world grain supply comes from Ukraine is important, yet it is more important that almost 50% of the grain provided to the World Food Program comes from Ukraine. At present, food prices have increased in some of the most vulnerable parts of the world, but most particularly in East Africa in countries that are also highly dependent on the WFP for food supplements.

The relevance of the pandemic and ongoing civil unrest in Asia, Africa, and now Europe, is that these events are serious threats to the food security and health of millions of women and children. If these challenges persist, the prevalence of undernutrition will begin to increase and the long-term effects of undernutrition on children who grow to become adults is very likely to then contribute to chronic diseases well into the future. Thus, understanding how to address and reverse the DBM is more important now than ever before. The world now appears to be at a nexus in terms of nutritional advances and human health, one at which promoting health through personalized medicine and precision nutrition is being challenged by geo-political and infectious disease. Therefore, it is imperative that nutrition research continue to make the DBM a priority to reduce the global burden of NRCDs in all parts of the world, especially the most vulnerable in marginal communities or lower income segments of our society.

References

Afshin, A., Sur, P.J., Fay, K.A., Cornaby, L., Ferrara, G., Salama, J.S., Mullany, E.C., Abate, K.H., Abbafati, C., Abebe, Z., Afarideh, M., Aggarwal, A., Agrawal, S., Akinyemiju, T., Alahdab, F., Bacha, U., Bachman, V.F., Badali, H., Badawi, A., et al., 2019. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 393 (10184), 1958–1972. https://doi.org/10.1016/S0140-6736(19)30041-8.

Baker, P., Machado, P., Santos, T., Sievert, K., Backholer, K., Hadjikakou, M., Russell, C., Huse, O., Bell, C., Scrinis, G., Worsley, A., Friel, S., Lawrence, M., 2020. Ultra-processed foods and the nutrition transition: global, regional and national trends, food systems transformations and political economy drivers. Obes. Rev. 21 (12), e13126. https://doi.org/10.1111/obr.13126.

Basu, S., McKee, M., Galea, G., Stuckler, D., 2013. Relationship of soft drink consumption to global overweight, obesity, and diabetes: a cross-national analysis of 75 countries. Am. J. Publ. Health 103 (11), 2071–2077. https://doi.org/10.2105/ajph.2012.300974.

Elizabeth, L., Machado, P., Zinöcker, M., Baker, P., Lawrence, M., 2020. Ultra-processed foods and health outcomes: a narrative review. Nutrients 12 (7), 1955. https://doi.org/10.3390/nu12071955.

FAO, IFAD, UNICEF, WFP, WHO, 2021. The State of Food Security and Nutrition in the World 2021: Transforming Food Systems for Food Security, Improved Nutrition and Affordable Healthy Diets for All. FAO. https://doi.org/10.4060/ca9692en.

FAO, WFP, 2019. Monitoring Food Security in Countries With Conflict Situations. A joint FAO/WFP update for members of the United Nations Security Council.

FAO, IFAD, UNICEF, WFP, WHO, 2018. The State of Food Security and Nutrition in the World 2018. Building Climate Resilence for Food Security and Nutrition.

GBD 2015 Mortality and Causes of Death Collaborators, 2016. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 388 (10053), 1459–1544. https://doi.org/10.1016/s0140-6736(16)31012-1.

Guthold, R., Stevens, G.A., Riley, L.M., Bull, F.C., 2018. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob. Health 6 (10), e1077–e1086. https://doi.org/10.1016/S2214-109X(18)30357-7.

Hoffman, D.J., Powell, T.L., Barrett, E.S., Hardy, D.B., 2021. Developmental origins of metabolic diseases. Physiol. Rev. 101 (3), 739–795. https://doi.org/10.1152/physrev.00002.2020.

Monteiro, C., Cannon, G., Lawrence, M., Louzada, M.L., Machado, P., 2019. FAO. Ultra-processed Foods, Diet Quality, and Health Using the NOVA Classification System.

Popkin, B.M., Corvalan, C., Grummer-Strawn, L.M., 2020. Dynamics of the double burden of malnutrition and the changing nutrition reality. Lancet 395 (10217), 65–74. https://doi.org/10.1016/s0140-6736(19)32497-3.

Popkin, B.M., Gordon-Larsen, P., 2004. The nutrition transition: worldwide obesity dynamics and their determinants. Int. J. Obes. Relat. Metab. Disord. 28 (Suppl. 3), S2–S9. https://doi.org/10.1038/sj.ijo.0802804.

Price, H., Adams, E., Quilliam, R.S., 2019. The difference a day can make: the temporal dynamics of drinking water access and quality in urban slums. Sci. Total Environ. 671, 818–826. https://doi.org/10.1016/j.scitotenv.2019.03.355.

Shukla, P.R., Skeg, J., Calvo Buendia, E., Masson-Delmotte, V., Pörtner, H.-O., Roberts, D.C., Zhai, P., Slade, R., Connors, S., van Diemen, S., Ferrat, M., Haughey, E., Luz, S., Pathak, M., Petzold, J., Portugal Pereira, J., Vyas, P., Huntley, E., Kissick, K., et al., 2019. Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems.

UNHCR, 2020. Global Report.

WHO, 2010. Set of Recommendations on the Marketing of Foods and Non-alcoholic Beverages to Children.

WHO, 2017. The Double Burden of Malnutrition. Policy Brief.

Famine: Causes, consequences and responses

Shannon Doocy, Shelley Walton, and Keith P. West, Jr., Department of International Health, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of K.P. West, S. Mehra, Famine: Causes, Consequences, and Responses, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 193–200, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00096-9.

411
Introduction 411
Modern famine 412
Definition of famine 413
Causes of famine 414
Market failure 414
Armed conflict 415
Failure in central planning 416
Climate and natural disasters 416
Famine prevention 416
Coping strategies 416
Surveillance 417
Famine intervention 418
Conclusions 418
References 418
Further reading 419

Key points

- Famine can be defined in terms of a sharp decline in access to food by large numbers of people, often attributed to crop failure from drought, flood or pestilence, hostilities or corrupt governance, leading to widespread starvation, disease and high mortality.
- Despite sophisticated surveillance, advances in agriculture, transnational food systems, and a transition globally toward greater food consumption and economic development, famine persists in the modern world.
- Famines are complex in their sets of causes, dominated by market failure, armed conflict, natural disasters and failure in central governance.
- Conflict, civil or foreign, has become the most consistent underlying or precipitative cause of famine in the past half-century, indicative that virtually all famines are preventable.
- Famine may best be prevented by strengthening resilience in vulnerable societies through infrastructural, commercial, educational, communications and agricultural development, a free press and diplomacy.

> Those who do not learn history are doomed to repeat it.
>
> George Santayana.

> Starvation is a matter of some people not having enough food to eat, and not a matter of there being not enough food to eat.
>
> Amartya Sen.

Introduction

Famine has afflicted humankind from antiquity to the present day, shaping the survival, politics and history of societies. A *bas relief* depiction on the Causeway of the Pyramid of Unas in Saqqara is believed to depict famine from drought during Middle Kingdom in Egypt. Biblical accounts describe the devastation wrought by famine on society and the means by which Joseph predicted and managed with Pharoah its consequences. The fall of the Roman Empire followed repeated food shortages and famines from 500 BC to AD 500, while the ranks of the Crusades in the 11th and 12th centuries swelled in response to assurances of food. European colonialism, with imposed cash cropping systems preconditioned vast areas of Africa and South Asia to famine, while

famine across Eurasia prompted migrations throughout the latter half of the second millennium (De Waal, 2018). The Great Irish Famine in the late 1840s, precipitated by a potato blight, caused one and a half million deaths and comparable migration, mostly to America as did decades of Russian famines in the late 19th century. Hundreds of famines frequented China throughout the first two millennia AD shaping its determination to rid itself of this scourge. This treatise largely focuses on famines of the past century, drawing lessons from history, to understand this persisting, modern phenomenon.

Modern famine

While famine in the 20th and 21st centuries have occurred globally, historically prone regions have remained epicenters of risk. Despite its persistence, modern famine has become better understood with regard to populations at risk, environmental and societal causes, stages of progression, triggering events and the progression of responses that famine prompts at household, community, national and international levels (Devereux, 2000).

A consistent feature of famine is high mortality. As background, **Fig. 1** provides global estimates of the extent to which deaths have been attributed to famines across regions, punctuated by notable historic events, against a curve depicting the global trend in population growth, from ~1.7 to 7 billion, from ~1900 into the new millennium. The figure reveals stark occurrences of famine in the first 60 years, followed by a dramatic decline in the numbers of famines and related deaths from the latter half of the 20th century onward. Most famines in the first half of the 20th century occurred during World War (WW) I, the intervening years leading to WW II, and during the second world war itself and were in countries with totalitarian rule. Unprepared for the stresses of losing farm labor and needing to feed over 15 million conscripts, the Russian empire endured famine throughout most of WW I. Some 450,000 people died in Persia (Iran) in 1917–18 from famine and disease during the war, attributed to drought, diversion of food stocks to sustain occupying forces and speculative hoarding, which was compounded by a flu pandemic.

Prolonged drought, exacerbated by revolution, civil war, and agricultural and economic collapse exacted famine in the Volga and Ural regions of Russia in 1921–22, where seed grain was eaten rather than sown, cannibalism was recorded and ultimately five million died. In the 1930s, famine-exacerbating policies of Stalin designed to deprive food to certain groups, instigate class warfare and crush a Cossack Revolution, led to ~3.3 million Ukrainian deaths. During WW II, the Nazi Hunger Plan, that diverted food stocks to support its expanding military occupation, starved ~4 million Russians, Ukrainians and Belarusians, while a Nazi blockade of Western Holland precipitated the "Dutch Hunger Winter" of 1944–45, killing 18–22,000 people (Collingham, 2011).

Over half of all famine deaths in the 20th century, however, occurred in China, starting with the understudied Great Qing Famine in the north in 1907, precipitated by floods the year before that destroyed crops and killed some 25 million people, and a famine in the north in 1927–30, that followed prolonged drought, in which ~6–10 million perished. Famine struck

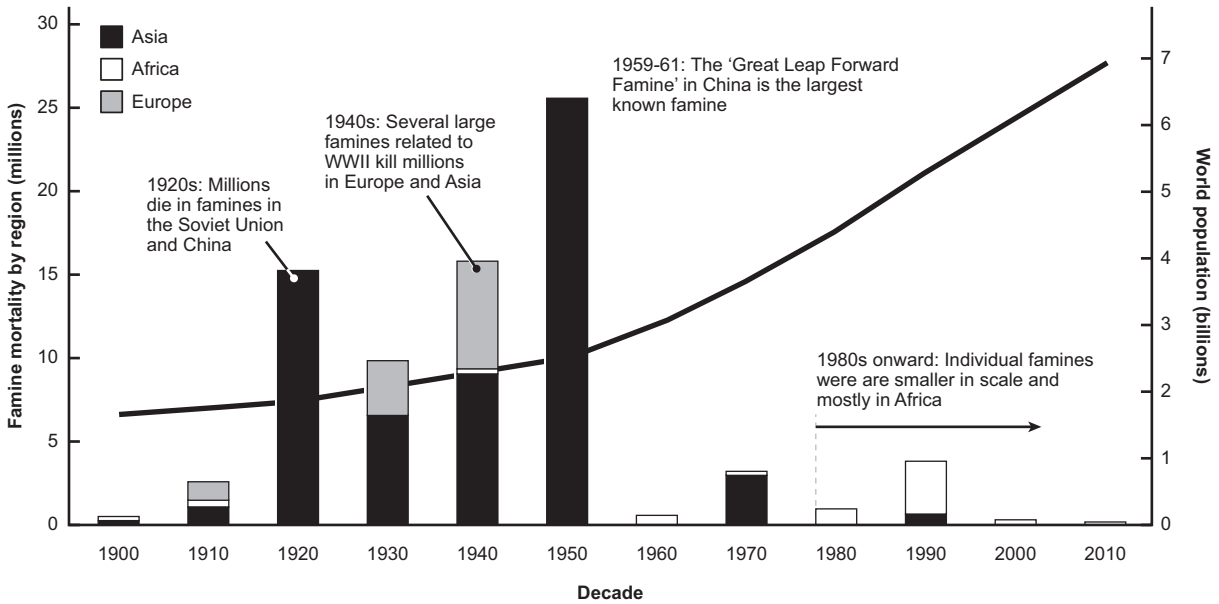


Fig. 1 Famine mortality by region and world population growth, 1900–2010. Since the 1960s, famine mortality has declined substantially despite world population growth. The geography of famines has shifted from being concentrated in Asia and Europe in the 1900–1950s to Africa where famines and related deaths more recently have been concentrated. Source: Tufts University Famine Database.

Hunan Province in 1942–43, amid WW II, where drought, locusts, flooding of the Yellow River region, extraction of meager harvests to feed Japanese occupational forces led to starvation, cannibalism, child trafficking, infectious diseases and ~1.5 million deaths. WW II extended its influence into South Asia as an external cause of famine, notably in Bengal in 1943, where a boat blockade on the Bay of Bengal to thwart an anticipated Japanese invasion, coupled with policies that redirected rural crops to subsidized urban centers, spiked market prices and led to ~3 million starvation-related deaths, mostly among rural, daily wage-earning classes.

Among famines in the second half of the 20th century, however, the "Great Leap" Famine from 1959 to 61 in China stands out as the worst in recorded history, precipitated by profound failure in centrally planned economic development and veiled by lack of transparent governance and a free press. Following several years of land reform that marginally increased grain production after birth of the Peoples' Republic, Mao Zedong's inconsistent farm collectivization policies, rural industrialization schemes and internecine politics destroyed peasant incentives and reduced crop output amid pronounced exaggerations in crop yields (Dikötter, 2010). State-sponsored vigilante groups extracting grain from farmers for urban centers, and state controls on food grain transport, access, purchase and taxation preconditioned China for an inevitable collapse of it's food system when pushed by Mao to further accelerate production in 1958. The ensuing famine led to an estimated 18–32 million Chinese deaths in 1959–62, implicating a vital role for unbiased journalism that can alert and hold government accountable to mount prevention.

In Southern Asia, Bangladesh in its nascent period of independence, experienced famine, in 1974. Prompted by concerns of food shortage consequent to an unusually severe, mid-year monsoon flood, speculative hoarding led to a contagion of spikes in the price of rice and other food commodities across rural markets. Well described in terms of market failure, the sudden and sustained deprivation of vast segments of the population to sufficient food led to stress sales of assets, reliance on *famine foods*, child abandonment, mass migrations to urban centers, infectious disease and up to 1.5 million deaths.

Conflict has become a nearly uniform underlying or precipitating cause of famine in the past half-century, with modern famines occurring amid civil wars, for example, in the vulnerable deserts (e.g., the Sudan), highlands (e.g., Ethiopia in 1974, 1984 and 2021) coastal (e.g., Somalia in 1992 and 2011) regions of North Africa. While many factors interact to cause famine, the risk attributable to conflict was evident in Ethiopia in contrast to neighboring Somalia in 1984. In that year and those that preceded, both fragile states endured the same, underlying climate-induced drought and stresses on crop production, but famine occurred only in Ethiopia where civil society, infrastructure and the economy were disrupted amid a military campaign (Götz et al., 2020). The late 1970s witnessed famine in East Timor, induced by an Indonesian military blockade on food imports during the island nation's war of independence. Famine under totalitarian rule occurred in North Korea from 1995 to 99, following floods and crop failures that was amplified by failed centralized government policies to mitigate a disaster that left as many as ~3 million dead, though estimates vary widely.

Famine has continued into the new millennium though the magnitude of famine-related deaths has decreased, revealing a stark reality that, despite more sophisticated surveillance systems, advances in agriculture, transnational food systems, and a steady "nutrition transition" toward greater food consumption attending gradual economic development, long sentence. Better: "...famine persists in the modern world. Prolonged military conflict plays a decisive, causal role." Present day famines are the result of a combination of complex casual factors including deteriorating crop production associated with rainfall declines and changing climate patterns; failures in development and commerce; repressive and corrupt governance; and armed conflict which can limit the flow of goods and disrupt both markets and production, restrict aid flows and precipitate mass migrations. Countries with protracted conflict that have faced famine or been at near-famine conditions in since 2015 include South Sudan, Ethiopia, Yemen and Syria (Brück and d'Errico, 2019).

Definition of famine

Famine can be defined in terms of a sharp decline in food availability or access for large numbers of people, often attributed to crop failure from drought, flood or pestilence, hostilities or corrupt governance, leading to starvation, disease and high mortality. While famine defies brief definition, most attempts to do so capture these basic features, illustrated in **Box 1**.

Comprehensive definitions of famine include elements of time dependency (e.g., steady, continuous erosion of or sudden collapse in food available for consumption), partial causation (e.g., due to natural calamity, armed conflict or convergence of

Box 1 Example short definitions of famine

1. An extreme crisis of inaccessibility to adequate food, manifested in widespread malnutrition and loss of life due to starvation and infectious disease (Maxwell and Majid, 2016).
2. Sustained, extreme shortages of food among discrete populations sufficient to cause high rates of mortality (Scrimshaw, 1987).
3. The most severe form of food insecurity, during which mass starvation and death take place within a relatively short span of time (Yip, 1997).

other complex causal events), class (e.g., affecting certain ethnic, geographic, economic or occupational groups more than others), consequence to habitat (e.g., to flora and fauna), population responses (e.g., adoption of extreme coping strategies, breakdown of civility and social norms, mass migration) and health consequence on a population scale (e.g., severe malnutrition, epidemics of infectious disease and mortality exceeding certain rates). Definitions of famine have evolved from reflecting declines in food availability to an understanding that famines are complex and multi-causal and more often the result of populations not being able to access food that is available (e.g. Amartya Sen's Entitlement Theory).

Operational definitions based on data assembled from diverse sources also exist that can guide relief agencies in deciding when to mobilize resources to prevent further deterioration of pre-famine to famine conditions. Specifically, the United Nations uses a five-phase scale known as the Integrated Phase Classification to assess a country's food security situation, with famine as the most extreme phase which is defined as: (a) a severe food shortage that affects at least 20% of households; (b) wasting malnutrition that affects 30% or more of the population, and (c) a crude death rate exceeding 2 deaths per 10,000 people per day. Whether viewed to evolve as a continuous, erosive process or precipitated via a sudden shock, once unleashed famine is distinct and exacts a catastrophic toll on human society.

Causes of famine

> Starvation is a matter of some people not having enough food to eat, and not a matter of there being not enough food to eat.
>
> Amartya Sen.

While a discreet event, famine is usually consequent to both antecedent and concurrent stresses, or *component causes* (borrowing a disease causation model from Kenneth Rothman in the 1980s) that combine to heighten risk of, and finally cause, a catastrophe. These include, for example, agricultural collapse due to flood, drought or pestilence, or conflict that can deplete food availability or deny access to food, even amid sufficient stocks, to segments of society. Either major pathway induces coping strategies among the most vulnerable to protect health, wages and productive and household assets (or *endowment*, see below), which are eventually depleted or transferred. Severe and persistent stress can lead to a breakdown in civil society, destitution, panic with mass migration, starvation with acute malnutrition, and a spiraling death rate. On their own, many component causes are associated with impoverishment, food insecurity, chronic hunger and malnutrition in many societies, but not necessarily famine. It is, rather, specific and rare sequences and combinations of stresses that act to become *sufficient* to precipitate famine. There appear to be at various sets of sufficient causes of famine, labeled to reflect their most distinctive, versus sole, feature (**Fig. 2**): market failure, armed conflict, natural or climate-driven disasters and failure in central planning.

Market failure

Market failure represents an intermediate dynamic of famine that can result from many causes that can destabilize and raise food at prices to levels for which money, labor, skills, credit, bartering and other assets can no longer be exchanged for sufficient amounts of food by large, usually marginalized segments of society. Conflict also destabilizes markets, directly contributing to market failure. Beyond a decline in food production, reduced food imports, inefficient transport, unfavorable trade policies, and disrupted infrastructure can reduce food availability and increase prices. However, absent conflict, more often it is speculative hoarding causing spikes in staple grain prices and sharp declines in value of draft, dairy and meat-producing animals (who also need to be fed), coupled with severe economic depression with losses in jobs and wages, that combine to deny access to food for large numbers of people. These failures may be compounded by political instability and corrupt governance. Afflicted groups cope with prefamine or early famine conditions by working for lower wages or food, migrating long distances to find work, indenturing agreements, bartering, borrowing at exorbitant interest rates, reducing meal frequencies, eating less familiar and nutritious foods, selling assets and land at prices far below usual market value, and orphaning children. Crime increases as the law and order disintegrate. For those still mobile, migration to urban centers or into spontaneous camps enroute to cities becomes the sole alternative to starvation. Famine inevitably leads to pauperization of afflicted survivors, while wealthy merchants, and business, land and cattle owners who are able to adapt, purchase at low prices, loan with high interest rates, and control resources, become wealthier.

A major advance in understanding famine consequent to market collapse emerged from Amartya Sen's analysis in the 1970s of the Great Bengal Famine of 1943, for which he received the Nobel Prize in Economic Science in 1998. A Famine Inquiry Commission appointed by the British Raj Government concluded the famine resulted from a severe decline in availability of rice in the year leading up to the famine. Sen's analysis revealed, however, that rice production was nearly normal in 1943, ~5% lower than previous years, concluding that starvation resulted from *entitlement failure*, an inability of rural markets to supply affordable food, yielding unfavorable *terms of food exchange* for rural classes on low, daily wages (i.e., low *endowment*) such as laborers, fishermen and craftsmen, to survive. In contrast, other sectors of Bengal society, including landowning farmers, food vendors and urban dwellers with access to fair price shops, had adequate food (Drèze and Sen, 1991). Entitlement failure

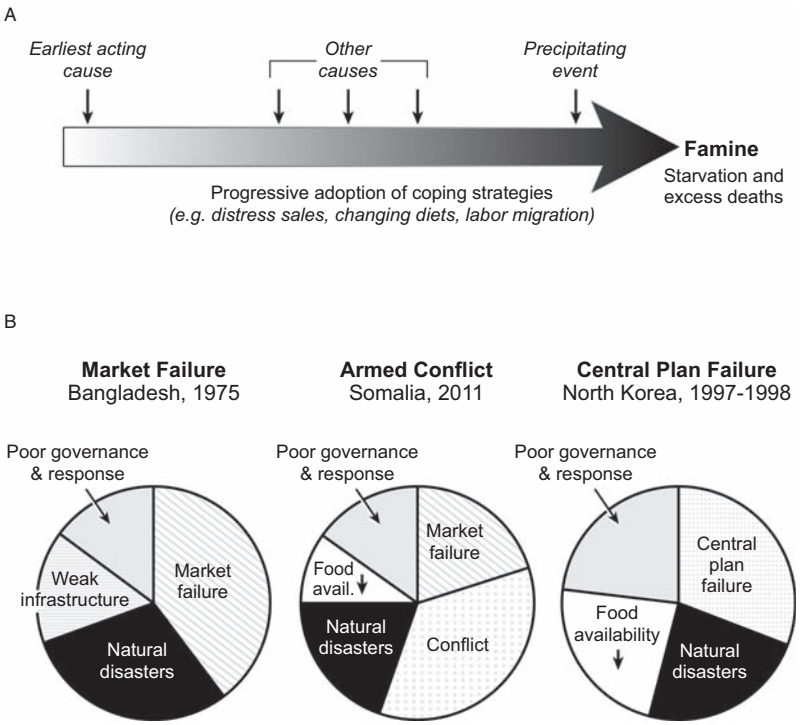


Fig. 2 The complex evolution and multicausal nature of famines. **Panel A**. Evolution of famine, depicting the sequence of causal events that, over time, combine to become sufficient to cause famine. **Panel B.** Conceptualization of three sufficient causes of famine, typified by their predominant cause (market failure, armed conflict and central planning failure), where slices in each circle depict component causes (subjectively sized, based on the literature) that interact when coexisting to cause famine, and include: **Market Failure** due spiraling food prices from speculation or decreased food availability, decreasing abilities to attain enough food via purchase, labor, credit or asset sales; **Conflict**, either declared war or internal violence preceding or amid famine; **Central Plan Failure**, whereby government directives disrupt infrastructure, productivity, food access and the economy; **Natural Disasters**, including slow-onset or repeated events such as drought, pestilence and catastrophic natural disasters such as floods and storms that can ruin crops**; Food Availability Decline**, resulting from crop failure or market disruptions; **Weak Infrastructure**, including inadequate finance systems, transportation, storage, communications and market infrastructures; **Poor Governance and Response**, reflecting inadequate policies and resource allocations in agriculture, employment and imports/export trades; fiscal mismanagement and corruption; absence of early warning systems, slow or in adequate humanitarian response to crises. Constructs were inspired by Rothman, K., Greenland, S., 1998. Modern Epidemiology. Philadelphia: Lippincott-Raven, pp. 7–28.

explains the dynamics of famine in Bangladesh in 1974–75, where speculation around the major *aman* rice crop failing (that never materialized) amid a flood, led to spikes in grain prices, a collapse in entitlement and access to food, starvation, mass migration and high mortality.

Armed conflict

Famine has been precipitated both as a tactic and consequence of war. Examples of the former during WW II are the Nazi blockade of ports in Western Holland leading to the Dutch Hunger Winter of 1945 and the diversion of food stocks, leading to mass starvations in order to support occupying forces in Persia in WW I and Russia and China during WW II. Alternatively, civil war has played a prominent role precipitating famine, with stark modern examples in the Horn of Africa. For example, Somalia has regularly faced drought that has rarely led to famine. However, in the early nineties, local militia wars collapsed an already fragile government and exacerbated effects of drought through exaction of harvested crops by marauding gangs, preventing produced food from reaching markets and, as famine evolved, obstructing aid from reaching most severely affected areas. In 2011, internecine conflict again triggered famine in Somalia, killing a quarter of a million people and displacing hundreds of thousands of refugees into Kenya and Ethiopia (Checchi and Robinson, 2013). Persistent conflict with denial of food as a tactic of war within Yemen since 2015 has disrupted economies, infrastructure and livelihoods and precipitated a modern humanitarian crisis, threatening 13.5 million people with severe, acute food insecurity. As in Somalia, the transfer of weaponry to vigilante groups, use of land mines, lawlessness and violence have conspired to precipitate famine and hamper relief. With little sign of abatement entering the third decade of the 21st century, conflict-driven famines are stalking Ethiopia, the Democratic Republic of the Congo and Afghanistan. While all famines polarize wealth and increase destitution, in general contrast to those involving failure of free markets, famine of conflict destroys infrastructure, impairs institutions of governance and economic production, and thus can hamper economic recovery a decade or longer.

Failure in central planning

A third, distinct class of famine results from failure by policy intent, indifference or incompetence of a centrally planned state to assure food to all sectors of society. Historic examples include notorious famines during the regime of Stalin in Soviet Russia in the 1920–30s, a period of instigated class warfare among peasantry, abolished economic incentives with farms collectivized into massive, inefficient production units, seizure and exportation of grain for foreign exchange, restricted population movements and brutal suppression of opposition. Agricultural production plummeted causing disastrous shortages that intensified state seizures of food grain, leading to ~15 million peasant deaths between 1930 and 37. Mentioned above, and modeled after Stalin, Mao Zedong's Great Leap Forward Famine was conditioned by erratic plans to collectivize farms to accelerate agricultural production and irrational rural industrialization schemes, and exacerbated by state control of grain production, procurement and taxation coupled with state terror and propaganda. Driven to be recognized globally as a success, at the peak of famine in 1960, Mao's government net exported more than a million metric tons of grain. Lacking an informed, free press, little was known about this famine for decades. A third modern example of catastrophic central planning failure lies with communist North Korea's inability to avert famine in 1997–8, that was preconditioned by chronic food insecurity and economic collapse from withdrawn Soviet aid in the previous decade, exacerbated by torrential floods, massive electricity outages, a precipitative drought and an isolationist government unable or unwilling to facilitate food aid as the crisis unfolded.

Climate and natural disasters

As a global phenomenon, climate change alone is not a complete cause of famine but is a major contributing factor when combined with other causes, and thus receiving increased attention. The global food systems produce enough to feed the world, yet the changing climate threatens global food security. Aggregate food shortage leading to a decline in availability has appeared to play a more prominent role in modern famines in the eastern Horn of Africa. In Ethiopia, Sudan, Eritrea, and Somalia, large tracts of land are drought prone, where average annual rainfall has been in decline since the 1930s. Concurrently, robust, indigenous farming and animal husbandry have been weakened as agricultural land has increasingly been converted for growing export crops. Global warming, and droughts and storms also contribute to famine risk where climate-related disasters have doubled since the early 1990s, intensifying underlying risk of famine (FAO et al., 2018). Climate change has the potential to impact the food system through droughts contributing to crop loss, heat waves killing crops, pestilence such as desert locusts destroying crops, unusually heavy rains leading to floods, and cyclones that further add wind damage. These events contribute to reduced food access due to poor crop productivity, high food prices, and loss of income. A modern example is the chronic, famine conditions facing southern Madagascar in 2021, attributed to recurrent drought resulting, exacerbated by an increasing frequency of cyclones from climate change, factors likely to continue to assume causal importance in the coming decades.

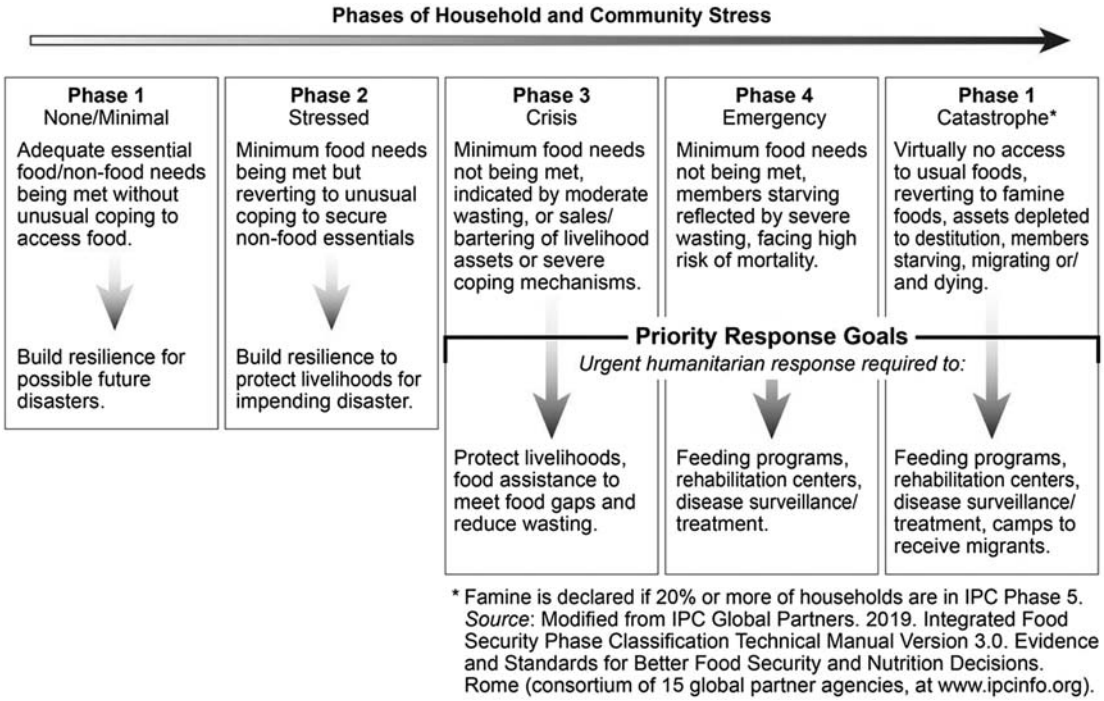
Famine prevention

Efforts to prevent, mitigate or recover from all phases of a decline in food access potentially leading to famine occur at household, community, national and international levels of society. While populations in famine prone regions recognize and have developed coping strategies to food stress the timing of actions and their success in averting famine are determined by the mix of component causes, available resources, and government intervention response Famine involving market failure are perhaps most readily forecasted and avertable, while famine amid conflict is least resourced, preventable and recoverable. Centrally planned famine tends to be insulated from outside assessment and intervention other than by diplomacy, typically leading to late-stage relief of acute food insecurity aimed at reducing mortality and rehabilitation.

Coping strategies

Communities, households and individuals cope throughout periods of increasing food insecurity, pre-famine states and throughout famine. Irrespective of causes, food insecure households change their diets and livelihood strategies to fend off hunger and starvation. Anticipating drought, farmers may sow more resilient crops, disperse herds and, as conditions worsen, sell or barter for grain for their livestock, before animals become a liability. Early coping strategies involve eliminating costly foods, reducing meal frequency and foraging, often for less nutritious famine foods, such as unusual roots, husks and leaves. Households may seek to supplement income (endowment) through work for lower wages or food, if such projects are available, to protect their entitlement to food. Loss of nutritional health becomes a *de facto* coping mechanism. Non-essential assets (e.g., jewelry, utensils) may be sold, followed by more productive items (e.g., tools, plows, bicycles). In addition to informal interhousehold exchanges, families may seek credit or loans, often at exorbitant interest rates, or enter into indentured contracts, accruing future debt to survive. As famine conditions worsen, and basic food commodities become less accessible, crime may increase with a disintegration of law and order, land sales rise leading to further depletion and polarization of long-term wealth, and communal norms fracture leading to abandonment of dependent family members, such as children at orphanages, and distress migration to transient camps and urban centers. In famine's most florid stages, cannibalism has been reported as a last resort. Specifically, the Coping Strategy Index (CSI), developed by several aid agencies, is a highly adaptable tool for monitoring the extent and severity of coping strategies.

Table 1 Integrated phase (acute food security) classifications.



Surveillance

Food security surveillance systems that integrate local conditions with national indicators can inform and guide governments and international agencies to pre-empt threats of food insecurity that could progress to famine. All systems require reliable, adequate and routine collection, analysis and dissemination of data to guide intervention decisions, with aggregate approaches to surveillance possible to enact through consortia of governmental, non-governmental and international agencies. Modern prediction and classification of risk of famine is guided by the Integrated Phase (Acute Food Security) Classification (IPC, https://www.ipcinfo.org/ipcinfo-website/ipc-overview-and-classification-system/ipc-acute-food-insecurity-classification/en/) System (**Table 1**) which aggregates, synthesizes and provides a basis to interpret data on food availability and consumption patterns, coping behaviors, prevalence rates of acute (wasting) undernutrition, destitution and mortality. Market monitoring systems also exist that can be deployed to follow trends in diversity of available foods, food prices and grain-to-animal exchange rates.

Further complementing and feeding into the IPC System are several online early warning systems providing reasonably effective predictions of pre-famine conditions to guide national and international pre-emptive responses. The Famine Early Warning System Network (FEWS NET, https://fews.net/), launched by the US Agency for International Development (USAID) in the mid-1980s in response to famines in Africa, is a satellite-to-on-ground network designed to detect and forecast widespread food shortages in the world's most food insecure countries. The Food and Agriculture Organizations' Global Information and Early Warning System (GIEWS, https://www.fao.org/giews/en/) monitors food supply, demand and other key indicators that can provide an early warning of food crizes in all countries that can also be aggregated by region. The WFP HungerMapLIVE (HungerMap, https://hungermap.wfp.org/) provides data on country-specific levels of food insecurity, prevalence of acute malnutrition, and market trends in food prices along with size estimates of populations at-risk.

Metrics from the systems and assessment methods can be compiled to classify countries facing risk of food insecurity as (1) none-to-minimal, (2) stressed, or in a state of (3) crisis, (4) emergency or (5) famine (*catastrophic*), with each phase having operational definitions and decision guidelines to enable agencies to develop response goals and plans. Approaching the end of the first quarter of the 21st century, use of the IPC system has revealed that, globally, more than 100 million people are regularly classified at *Level 3*, representing critical levels of food insecurity, with some of the affected countries facing crizes of overwhelming proportion. For example, 28 million of those estimated to be facing a food crisis are surviving amid conflict in the Democratic Republic of the Congo, Yemen and Afghanistan, representing nearly one in five residents in these fragile states.

Surveillance systems that have been developed to inform multisectoral policies to stabilize food economies have likely prevented many crizes from deteriorating to famine, although owing to the rarity of famine, both correct prediction and likely impact remains challenging as required counterfactual data are lacking. Further, the abilities of early warning systems to mitigate famine depend on the rapidity, completeness an timeliness of dissemination, and extents to which actions are taken at national and international levels.

Famine intervention

Strategies to pre-empt famine that may be available to governments include drawing on grain banks/reserves and increasing imports, introducing staple subsidies through fair price outlets, organizing infrastructure projects, imposing export restrictions and launching social safety net programs for the most vulnerable groups or regions. For example, a common strategy to address food availability declines are to maintain grain stores or import staple foods and control their release into markets, thereby stabilizing access, expectations and prices. As governments of low-income countries are often unable to adequately mitigate impending famine, key roles exist for UN, bilateral and humanitarian agencies to offer effective assistance.

International responses to famine are usually coordinated via a United Nations appeal process, in which governments highlight need among severely affected groups, and request financial or material resources from member states to support a humanitarian response. Typically, UN government and non-governmental agencies organize around a cluster system (*cluster*, https://www.humanitarianresponse.info/en/coordination/clusters/what-cluster-approach) to coordinate interventions, guided by the classified IPC phase, to meet health, food security, nutrition and logistics needs in affected regions. Complementing a rising capability to monitor, gauge and preventively intervene in famine has been the evolution of guidelines known as the Sphere Standards (Sphere, 2018) for the principled conduct of humanitarian assistance across all sectors. The Sphere Standards prescribe the ethical basis and minimum to assure adequate food, shelter, sanitation, health care and security to alleviate suffering and preserve dignity of groups afflicted by emergencies. A common priority is to establish self-sufficiency by promoting food access, protecting assets and income of the poor, and stabilizing markets to restore entitlement. As the UN lead agency addressing food insecurity, WFP directs approximately two-thirds of its annual global budget, amounting to ~US$ 5 billion, toward preventing severe food crizes and famine through emergency humanitarian food assistance programs (World Food Program, 2021).

While relief operations are believed often successful in preventing famine, challenges in timely decision making and response persist with aid arriving too late, after the peak of famine has occurred. For example, in the Somalia famine on 2010–2011, early warning identified a high risk of famine, but timely action to prevent famine did not occur. The Tigray region of Ethiopia faced near famine conditions in 2021–2022 attributed, in large part, to the government blocking international aid to the region. It remains a difficult call to respond to food crizes given uncertainties in prediction that accompany rare events such as famine, coupled with global humanitarian needs that far surpass the resources available for response. A basic tenet remains accepted that effective famine prevention requires a development agenda that strengthens resilience in famine-prone societies. Efforts can include enacting infrastructural, commercial, education, agricultural, and other policies that reduce poverty and create sustainable food systems that, over the long-term, reduce famine risk.

Conclusions

Famine has plagued humankind through the millennia to the present day. While several component factors must usually combine to cause a rare famine, modern crises can be characterized by market failure, triggered by armed conflict or result from incompetent central economic planning. Major advances in early warning and humanitarian responses have likely stemmed risk of famine in prone regions. Modern famines have been severe and dominated by armed conflict that disrupts food systems, destabilizes economies, and can destroy infrastructures. Peace, competent governance, and sustained and equitable economic growth that benefits all society, a free press to hold governments accountable, and climate protection that could mitigate the impacts of natural disasters are global priorities for food systems to prosper and minimize risk of famine.

References

Brück, T., d'Errico, M., 2019. Food Security and Violent Conflict: Introduction to the Special Issue. World Development, 117. Elsevier, pp. 167–171.

Checchi, F., Robinson, W., 2013. Mortality Among Populations of Southern and Central Somalia Affected by Severe Food Insecurity and Famine During 2010–2012. Somalia: Food Security and Nutrition Analysis Unit. FAO/FSNAU, FEWS-NET.

Collingham, E.M., 2011. The Taste of War: World War II and the Battle for Food. The Penguin Press, New York, NY.

De Waal, A., 2018. Mass Starvation: The History and Future of Famine. Polity Press, Cambridge, UK.

Devereux, S., 2000. Famine in the Twentieth Century. Institute of Development Studies, Sussex.

Dikötter, F., 2010. Mao's Great Famine: The History of China's Most Devastating Catastrophe, 1958–1962. Walker & Company, New York, NY.

Drèze, J., Sen, A., 1991. The political economy of hunger: volume 2: famine prevention. In: The Political Economy of Hunger. Famine Prevention, vol. 2. Oxford University Press, New York, NY.

FAO, IFAD, UNICEF, WFP, WHO, 2018. The State of Food Security and Nutrition in the World 2018. Building Climate Resilience for Food Security and Nutrition. FAO, Rome.

Götz, N., Brewis, G., Werther, S., 2020. Humanitarianism in the modern world: moral economy of famine relief. In: Humanitarianism in the Modern World. Cambridge University Press, Cambridge.

Maxwell, D., Majid, N., 2016. Famine in Somalia: competing imperatives, collective failures. In: Etica e Politica (Issue 1). Oxford University Press, Oxford.

Scrimshaw, N.S., 1987. The phenomenon of famine. Ann. Rev. Nutr. 7, 1–21.

Sphere Association, 2018. The Sphere Handbook: Humanitarian Charter and Minimum Standards in Humanitarian Response, fourth ed. Geneva, Switzerland. www.spherestandards.org/handbook.

World Food Programme, 2021. WFP—Saving Lives, Preventing Famine. https://docs.wfp.org/api/documents/WFP-0000133765/download/?_ga=2.266806094.1344020064.1647031541-2050081340.1644407527.

Yip, R., 1997. Famine. In: The Public Health Consequences of Disasters. Oxford University Press, Oxford. Chapter 15.

Further reading

Cuny, F.C., Hill, R.B., 1999. Famine, Conflict, and Response: A Basic Guide. Kumarian Press, West Hartford, Conn.
Newman, L.F., Crossgrove, W.C., 1990. Hunger in History: Food Shortage, Poverty, and Deprivation. Wiley-Blackwell, Cambridge, MA.
Yang, D.L., 1996. Calamity and Reform in China: State, Rural Society, and Institutional Change Since the Great Leap Famine. Stanford University Press, Stanford, CA.

Food security and nutrition surveillance in low- and middle-income countries

Kathrin M. Demmler[a], **Valerie M. Friesen**[b], **and Lynnette M. Neufeld**[c], [a] Global Alliance for Improved Nutrition, London, United Kingdom; [b] Global Alliance for Improved Nutrition, Geneva, Switzerland; and [c] Food and Agriculture Organization of the United Nations (FAO), Rome, Italy

© 2023 Elsevier Ltd. All rights reserved.

This is an update of L.M. Neufeld, L. Tolentino, Nutritional Surveillance: Developing Countries, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 289–302, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00206-3.

Introduction 421
Challenges informing policymaking and program planning for nutrition in LMICs 421
Deciding what data are needed 422
Conceptual frameworks 422
Core principles and terminology 422
FSN surveillance data in LMICs 423
Routine data for FSN surveillance 423
Strengths and limitations 428
Survey data for FSN surveillance 428
Constructs and indicators commonly included in surveys as part of FSN surveillance 429
Strengths and limitations 430
Accessing data through international platforms 431
Case studies 431
Continuous DHS assessment, Senegal 431
Integrated Reproductive Health and Nutrition Surveillance System (SIVESNU), Guatemala 432
Conclusion 432
Conflict of Interest 432
Acknowledgments 432
References 432
Relevant websites 434

Key points

- Food security and nutrition (FSN) survey data can be used to track changes in food security, dietary quality and nutritional status of populations, identify vulnerable groups for interventions and inform the development of national policies and programs in regular intervals.
- Many low- and middle-income countries face several challenges to the establishment and maintenance of FSN surveillance systems.
- The decision of what indicators to assess needs to be carefully considered based on the type of (available) routine and survey data, the purpose and context of the data assessment, and available resources.
- Routine FSN data provide estimates of food security and health service coverage that allow for early warning and comparison among and across sub-groups, regions and over time.
- Accurate and timely FSN surveillance data enable effective evidence-based effective policymaking and program implementation.

Glossary

Diet quality Refers to the overall pattern of an individual or a population's food intake. It can be measured through various indicators that score food patterns to assess their alignment with recommendations and/or guidelines for different aspects of healthy diets. No single indicator of diet quality has been developed that captures all relevant aspects of dietary intake given the complexity of diets

Food security Refers to a state in which people have physical, social and economic access to sufficient, safe and nutritious food that meets their preferences and needs

Food security and nutrition surveillance Refers to a system of continuous and timely collection, analysis and reporting of data on dietary intakes, food security, nutritional status and nutrition-related health outcomes as well as aspects of access and availability of food, nutrition knowledge and behavior of a population or selected population groups

Health service coverage Refers to the proportion of a population for whom a health service is intended and who receive it

Nutritional status Refers to the condition of the body as influenced by diet and disease-related factors. It is commonly assessed by measuring weight and height and/or body fat (referred to as anthropometry) as well as by measuring concentrations of vitamins and minerals (also known as micronutrients) in biological samples such as blood and urine

Routine data Refer to data which are regularly collected and commonly obtained through existing systems, e.g., health facilities or agriculture and trade statistic bureaus

Survey data Refer to data that are collected through surveys that are conducted periodically and often include nationally and sub-nationally representative samples

Introduction

Food security and nutrition (FSN) surveillance is a system of continuous and timely collection, analysis and reporting of data on dietary intakes, food security, nutritional status and nutrition-related health outcomes as well as aspects of access and availability of food, nutrition knowledge and behavior of a population or selected population groups (WHO, 2013). A well-designed surveillance system operates with integrated practices, tools, and processes to collect, organize, analyze, store and disseminate routine and survey FSN data from multiple sectors and sources. The ultimate goal of such surveillance systems is to provide data on FSN-related issues that can be used to inform and advise relevant policies and programs designed to improve nutrition and health in support of achievement of the Sustainable Development Goals (SDGs) (United Nations, 2015).

There are many opportunities and challenges to ensuring effective FSN surveillance, several of which differ in low- and middle-income countries (LMICs) compared to high-income countries (HICs) (HIC contexts are reviewed in a separate chapter of this book *ref*). In this chapter, we review the challenges facing FSN surveillance systems in LMICs, emphasizing their use for policymaking and program planning. We then provide an overview of the most common measures and indicators along the constructs of food security, health service coverage, dietary intake and nutritional status collected using routine or survey data, the methods used, and a description of the measures and indicators and their strengths and limitations. Finally, we highlight two country case studies and several international data platforms through which FSN surveillance data can be accessed and used to inform policymaking and program design in LMICs.

Challenges informing policymaking and program planning for nutrition in LMICs

Rapidly changing food systems giving rise to the double burden of malnutrition and non-communicable diseases (NCDs) in many LMICs (Demmler et al., 2017; Popkin et al., 2020) present particular challenges to policy makers. In addition, limited resources in many LMICs imply that external funds, often from donor agencies, are required to set up and maintain effective FSN surveillance. The involvement of donors however, can unintentionally influence local decision making and weaken ownership of data and resultant policies (Khan et al., 2018).

Data-driven policymaking and program planning is often constrained by lack of availability or access to the necessary FSN-related indicators. Ideally, surveillance encompasses a wide range of FSN data and is embedded in information systems where data are coordinated, collected, curated, analyzed and disseminated by one central institution. Some LMICs have established such comprehensive information systems, such as the Emergency Nutrition Coordination Unit in Ethiopia (Tuffrey and Hall, 2016). However, many LMIC information systems are fragmented focusing on one or several, but not all, relevant components (e.g., food production and supply or nutrition and health). A lack of coordination and alignment among sectors with relevant data often results in critical data gaps or redundancies leading to inefficient use of resources and inability to explore the interconnected issues from food production through to diet-related health outcomes and use of data for decision making. Technical capacity is also a challenge in many contexts, particularly in the translation and use of data for decision making (Shroff et al., 2017).

Several initiatives aim to overcome these challenges at country, regional and global levels by strengthening the development and use of relevant and standardized indicators, developing and promoting the use of innovative tools, and strengthening the synthesis and translation of data for effective evidence-based policymaking (Verstraeten, 2021). For example, the "nutrition data value chain" highlights six key aspects (i.e., prioritization, creation and collection, curation, analysis, translation and dissemination, and decision making) for accountable FSN data and action (DataDENT, 2021). This approach further stresses the need for identifying gaps in nutrition measurement and advocating for investments to strengthen nutrition data value chains. Other efforts seek to overcome specific data-related challenges, such as the International Dietary Data Expansion (INDDEX) by Tufts University, which aims to support LMICs to overcome high costs, inaccessibility, low quality, and under-use of household- and individual-level dietary data (INDDEX Project, 2018).

Deciding what data are needed

Conceptual frameworks

Concise frameworks can support data informed decision making for FSN policies and programs by enhancing understanding of complex systems, identifying pathways through which nutrition-related outcomes are influenced, and using this to support data priorities. The High Level Panel of Experts on Food Security and Nutrition (HLPE-FSN) of the Committee on World Food Security (CFS) framework on data for food security and nutrition (HLPE, 2022) was designed explicitly for this purpose. The framework draws on two previous, well-known conceptual frameworks. First, the HLPE Food Systems framework (HLPE, 2017), which provides an overview of the relationship between the food system and diets, and the many drivers that may ultimately influence dietary intake and related outcomes. Second, the United Nations Children's Fund (UNICEF) framework of the causes of malnutrition (UNICEF, 1998), which provides an important complement to this by focusing on the multiple determinants of nutrition at the individual level that include, but go beyond, dietary intake. Most LMICs are now facing the double burden of malnutrition, thus this comprehensive focus, encompassing not only nutritional status and health outcomes but also their determinants at multiple levels, is critical.

Core principles and terminology

Data collection should always be aligned with a country's priorities as outlined in national, sub-national or sectoral plans and corresponding, well defined, high-quality indicators should be selected. Meaningful disaggregation of data as well as flexibility allowing for the adaptation of indicators based on changing circumstances, requirements or data collection approaches, needs to be factored into a surveillance system from an early stage while ensuring cost-effectiveness (UNICEF and WHO, 2021).

One common challenge is the terminology used among the varying sectors and stakeholders involved in data collection and use for FSN. Confusion often exists between the underlying **construct** of interest (e.g., diet quality), the **indicator** used to assess that construct (e.g., minimum dietary diversity among women of reproductive age (i.e., MDD-W)), a **measure** that may be used to construct an indicator (e.g., number of food groups consumed in the past 24 h), and the **methodology** used to collect the raw data to generate the measure and indicator (e.g., a diet recall questionnaire). A **scale** may also be developed to reflect a construct with traits that are unobservable (e.g., perceptions of food insecurity). Ideally, indicators should be validated to ensure that they adequately reflect the construct of interest and methodologies should be properly adapted to the context to ensure that they will appropriately capture all needed data to generate the indicator. FSN data is unfortunately fraught with challenges in this regard. Continuing the diet quality example, while the MDD-W indicator provides a validated estimate of overall risk of micronutrient inadequacy among women (Martin-Prével et al., 2015), this reflects only one component of the broader construct of *diet quality*. Currently, there is no single universal indicator that captures all aspects of diet quality relevant for all forms of malnutrition among any population group. Nor is there consensus on the methodology and measures that should be used to assess other indicators of diet quality or the processes needed to adequately adapt such methodologies to diverse contexts (TEAM, 2021).

Despite these issues related to assessment of some FSN constructs, many relevant and validated indictors of FSN exist and the measures used to generate them are commonly collected from one or two primary **data sources** as part of a FSN surveillance system, i.e., routine and survey data. **Routine data** are regularly collected and commonly obtained through existing systems, e.g., health facilities or agriculture and trade statistic bureaus. **Survey data** include surveys that are conducted periodically and often include nationally and sub-nationally representative samples. They may be single country or multi-country efforts, such as the Demographic and Health Surveys (DHS) or Multiple Indicator Cluster Surveys (MICS), that include many common measures and indicators

Box 1 Food security and nutrition (FSN) surveillance systems in times of crises

Crises, such as conflict, health emergencies and climate variability challenge food and health systems and can have severe effects on FSN. As such, FSN surveillance systems may need to expand beyond the collection of routine and survey data and embed additional rounds of existing indicator collection and/or add the collection of new indicators that respond to the specific situation at hand. For example, Action Against Hunger leads rapid and targeted surveys to understand the type of intervention most needed in emergency situations (e.g., nutrition, food security and livelihoods, or water, sanitation, and hygiene services) (Action Against Hunger, 2019). The COVID-19 pandemic has prompted the need for adjustment of the data collection approach (e.g., through no-touch assessments at community level, mobile data assessments or web-based surveys) to reduce the risk of COVID-19 transmission to protect survey staff and respondents and in response to restrictions imposed on movement in several country contexts. To support the implementation of FSN surveillance in the context of the pandemic, the United Nations Children's Fund (UNICEF), the Demographic and Health Survey (DHS) Program and others have developed recommendations, guidelines and support for the collection, analysis and management of FSN data (UNICEF, 2020).

For example, the 2020 Kenya Malaria Indicator Survey data collection stopped due to COVID-19 in June 2020. Aiming to resume data assessment as soon as possible, the Kenya National Bureau of Statistics (KNBS) together with the Kenya Ministry of Health Department of National Malaria Control Program (DNMP) and the DHS Program revised the data collection approach to include risk mitigation elements, e.g., the use of personal protective equipment for enumerators and revised interview protocols and training to protect survey staff and participants alike. Based on the adjusted approach and virtual training of fieldworkers, data collection could be resumed in October 2020 (USAID, 2021).

Exploring the expansion and innovation of routine and survey data collection approaches is of immense importance for keeping FSN surveillance systems running and avoiding surveillance systems coming to a halt in times of crises.

(appropriately adapted to context). Surveys may also be targeted to specific regions or in response to crises or similar situations (**Box 1**).

Both routine and survey data have varying strengths and limitations for various objectives and should therefore be used in a manner that maximizes their strengths where possible. For example, routine data is typically less resource intensive to collect than survey data. However, because data are collected through existing systems, data may not be representative of the full population. For example, nutrition-relevant health coverage data will only capture information from those who use the health system (e.g., weight of adult men seeking routine health care). Data quality and consistency in how it is collected may also vary much more in a routine system than in a survey where data protocols and interviewer training can be more tightly controlled. Furthermore, for routine systems, data are often collected at national level (e.g., food supply) and several assumptions must be made to extrapolate this to the adequacy of that supply for the population. Such data is vital for some purposes, for example analyzing time trends and accountability to meet national or international goals such as the SDGs. However, it does not actually assess what individual households consume and cannot be used to identify sub-groups within a population who may be vulnerable to food insufficiency for economic or other reasons or for whom food assistance or social protection programs may be needed.

The choice of data source for measuring constructs of interest within a surveillance system must therefore be defined by the purpose for which they are intended, with a deep understanding of data quality considerations and ensuring fit for purpose and to context (Frongillo et al., 2014). **Table 1** provides an overview of illustrative examples of the purposes for which routine and survey data are often used in the context of FSN surveillance. Using a mix of routine and survey data and streamlining the inclusion of appropriate indicators in each source helps to provide integrated and comprehensive insights and may increase the cost-effectiveness and sustainability of FSN surveillance (UNICEF and WHO, 2021).

FSN surveillance data in LMICs

The following presents FSN-related constructs, methodology, measures and indicators that are commonly assessed through routine and survey data as part of FSN surveillance systems (**Table 2**).

Routine data for FSN surveillance

Routine food supply data can be used to monitor trends in national apparent food intake and food security for comparisons over time and across countries. Most countries worldwide collect information on food supply. Many LMICs display these data through their national statistical institutes (e.g., in Kenya the National Bureau of Statistics (KNBS)).[1]

In conjunction with national statistics offices and their records on production and trade of food commodities, the Food and Agriculture Organization of the United Nations (FAO) compiles Food Balance Sheets (FBSs) for over 245 countries and territories, including most LMICs. For a wide range of food commodities, the FBSs capture the sources of supply and its utilization. Total food utilization is classified by various categories (e.g., human consumption, foods export, used for livestock, and food losses). The human consumption estimates are of particular interest for FSN surveillance as they can be used to provide high-level estimates of dietary intake and food security by dividing the respective quantity by the total population. Data on per capita food supply are then displayed as annual or daily quantity of food availability for a wide range of food items and presented by country, region,

Table 1 Illustrative examples of the purpose for food security and nutrition surveillance systems from routine data and surveys.

Data sources	*Purpose (examples)*
Routine	• Snapshot of situation and comparison with goals. • Identification of specific issues that require attention. • Early warning of nutrition-related issues. • Comparison among subsets of the population. • Assess time trends.
Survey	• Inform the development of national policies and programs. • Monitor trends over time, in response to political and environmental changes. • Identify groups for intervention. • Assess the impact of policies and interventions of the population in different locations.

[1] https://www.knbs.or.ke/.

Table 2 Overview of food security and nutrition (FSN)-related constructs, indicators, and measures commonly collected as part of routine data and surveys in FSN surveillance systems.

Data sources	*Construct*	*Measure and indicator*	*Methodology*	*Description of measure and indicator*	*Frequency*	*Further information*
Routine	Food security	Food supply: • Quantity (kg/capita/year) • Dietary Energy Supply (DES) (kcal/capita/day) • Protein and fat (g/capita/day)	Food Balance Sheets (FBS)[a]	Estimate of apparent population food and energy intake, based on national food supply data.	Annual	National statistic offices; FAOSTAT[b]
		Dietary Energy Consumption (DEC) (kcal/capita/day)[c]	FBS data on DES, accounting for household waste[c]	Estimate of apparent population energy consumption, based on national food supply, utilization, and food waste data.		
		Prevalence of Undernourishment (PoU), % of total population[c]	FBS data on DES, accounting for household waste, access to dietary energy (CV) and minimum dietary energy requirements (MDER)[c]	Modeled share of the population whose estimated apparent energy consumption is below requirements.		
	Health service coverage	Prevalence of population receiving nutrition-relevant care: • Vitamin A supplementation for children • Iron folic acid supplementation during pregnancy • Breast and complementary feeding counseling	Health service data based on internal documentation of targeted intervention[d]	Share of targeted population receiving nutrition-relevant care.	Regularly, depending on health system frequency	National statistic offices, health statistics
Survey	Food security	Household Hunger Scale (HHS)[e]	Questionnaire module in HCES, DHS, MICS	The degree (0–6) of household access to enough food in the last 30 days (little to no hunger (0–1), moderate hunger (2–3), severe hunger (4–6)).	Every few years, annual, ad hoc	DHS program[f], MICS program[g]
		Food Insecurity Experience Scale (FIES)[h]	Questionnaire module in Gallup World Poll	The degree of individual/household access to enough and nutritious food (moderate to severe), based on a weighted 8-question scale.	Annual	FAO Microdata Catalog[i]

Table 2 Overview of food security and nutrition (FSN)-related constructs, indicators, and measures commonly collected as part of routine data and surveys in FSN surveillance systems.—cont'd

Data sources	*Construct*	*Measure and indicator*	*Methodology*	*Description of measure and indicator*	*Frequency*	*Further information*
Survey	Diet quality	Household Dietary Diversity Score (HDDS)[j]	Questionnaire module in DHS, MICS or through HCES data	Proxy of household caloric availability and socio-economic status; count of different food groups consumed, last 24 h.	Every few years, annual, ad hoc	DHS program[f], MICS program[g]
		Food Consumption Score (FCS)[k]	Assessed by WFP's food security monitoring systems	Proxy of household caloric availability; count of different food groups consumed, weighted by relative nutritional value, last 7 days.	Every 1–3 months; annual, ad hoc	WFP Hunger map[l]
		Minimum Dietary Diversity for Women of reproductive age (MDD-W), % of female population (15–49 years)[m]	Women questionnaire modules in DHS, MICS	Proportion of women of reproductive age who consumed a minimum number of different food groups (at least 5 out of 10) during the previous day.[m,n]	Every few years, ad hoc	DHS program[f], MICS program[g]
		Minimum Dietary Diversity (MDD) for children, % of children (6–23 months)[o]	Child questionnaire modules in DHS, MICS	Proportion of children 6–23 months of age who consumed a minimum number of different food groups (at least 5 out of 8) during the previous day.[o,p]		
		Individual dietary diversity, and Global Dietary Recommendations (GDR) score	Diet Quality Questionnaire (DQQ) in Gallup World Poll[q]	Proportion of adults (men and women) consuming Minimum Dietary Diversity (MDD-W as described above). Adherence to Global Dietary Recommendations (GDR) for healthy diets, composed of two components: GDR-Healthy score that protects against non-communicable diseases (NCDs) and the GDR-Limit score that captures risk factors for NCDs.	Aim to expand to more countries and to adopt for national-level diet monitoring.	Global Diet Quality Project[q]
		Apparent food intake: • Quantity (kg or g/person or household/day)	Targeted household food acquisition and purchase questionnaire/ module in HCES,[r]	Estimated household or individual-level food intakes in quantity or energy.	Every few years, ad hoc	IHSN Central Survey catalog[t], World Bank Microdata Library[u]

(Continued)

Table 2 Overview of food security and nutrition (FSN)-related constructs, indicators, and measures commonly collected as part of routine data and surveys in FSN surveillance systems.—cont'd

Data sources	*Construct*	*Measure and indicator*	*Methodology*	*Description of measure and indicator*	*Frequency*	*Further information*
		• Energy (kcal/person or household/day)	applying the Adult Male Equivalent (AME) method[s]			
Survey	Diet quality	Apparent nutrient intake: • Quantity (unit/person/day) • % of requirements	Calculated based on apparent food intake data and Estimated Average Requirement (EAR) cut point or probability method[v]	Estimated individual nutrient intakes or share of estimated nutrient requirements based on targeted food intake estimates.	Every few years, ad hoc	Indicators are not necessarily assessed as part of HCES but can be calculated as needed.
		Food intake: • Quantity (kg or g/capita/day) • Energy (kcal/capita/day)	Dietary intake survey (e.g., 24 h recall, food records, FFQs)	Individual food intake in quantity or energy.		Surveys included in national food security and nutrition surveys with dietary assessment on individual level; availability depending on the country, some country data compiled on GIFT[w]
		Prevalence of inadequate nutrient intakes, % of population	Dietary intake survey (e.g., 24 h recall, food records, FFQs) data, multiplied by the nutrient content level in each food; and Estimated Average Requirement (EAR) cut point or probability method[v]	Individual nutrient intake as share of individual nutrient intake below requirements; proxy for micronutrient deficiency.		
		Household coverage of fortified foods	Household questionnaire and food sample collection (from either households or markets) and nutrient content analysis (qualitative or quantitative) to confirm presence of fortification	Proportion of households consuming a food that is confirmed to be fortified at any level.		DHS program[f] (salt only), MICS program[g] (salt only), FACT surveys (all mandatorily fortified foods in a country)
	Nutritional status	Overweight and obesity, adults or children <5 years	Individual body measurements (weight, height/length) in DHS or MICS, calculated as Body Mass Index (BMI)[x] or growth charts (weight for height)	Proxy for abnormal or excessive fat accumulation. For adults, based on high individual BMI ($\geq$25 kg/m^2). For children, based on high weight-for-height measures (>2 SD).	Every few years, ad hoc	NCD-RisC[y] (BMI only); DHS program,[f] MICS program,[g] WHO Global Database on Child Growth and Malnutrition[z]
		Underweight, adults or children <5 years	Individual body measurements (weight, height/length) in DHS or MICS, calculated	Proxy for poor nutrition. For adults, based on low individual BMI (<18.5 kg/m^2). For		

Table 2 Overview of food security and nutrition (FSN)-related constructs, indicators, and measures commonly collected as part of routine data and surveys in FSN surveillance systems.—cont'd

Data sources	*Construct*	*Measure and indicator*	*Methodology*	*Description of measure and indicator*	*Frequency*	*Further information*
			as BMI or growth charts (weight for age)	children, based on low weight-for-age measures (<−2 SD).		
		Stunting, children <5 years	Individual body measurement on height/length in DHS or MICS, calculated through growth charts (height for age)	Impaired growth in children, based on low height-for-age measures (<−2 SD).		
		Wasting, children <5 years	Individual body measurement (weight, height/length) in DHS, MICS, calculated through growth charts (weight for height)	Recent and severe weight loss in children, based on low weight-for-height measures (<−2 SD).		
		Micronutrient status	Biological sample (blood, plasma, urine) collection and laboratory analysis	Micronutrient status and its adherence to established cutoffs of deficiency, sufficiency and/or excess.	Every few years, ad hoc	Assessment through micronutrient status surveys; some country data are available through VMNIS[aa]; Hemoglobin DHS program,[f] MICS program[g]

DHS: Demographic and Health Surveys; FFQ: Food frequency questionnaire; HCES: Household consumption and expenditure surveys; MICS: Multiple Indicator Cluster Surveys.
[a]FAO (2021a).
[b]Data can be accessed via http://www.fao.org/faostat/en/#home.
[c]FAO (2021b).
[d]Tuffrey and Hall (2016).
[e]Ballard et al. (2011).
[f]Data can be accessed via https://dhsprogram.com/Data/.
[g]Data can be accessed via https://mics.unicef.org/.
[h]Ballard et al. (2013).
[i]Data can be accessed via https://microdata.fao.org/index.php/catalog.
[j]Kennedy et al. (2011).
[k]WFP VAM Resource Center (2019).
[l]Data can be accessed via https://hungermap.wfp.org/.
[m]FAO (2021c).
[n]The MDD-W is grouped in (1) grains, white roots and tubers, and plantains; (2) pulses (beans, peas and lentils); (3) nuts and seeds; (4) milk and milk products; (5) meat, poultry and fish; (6) eggs; (7) dark green leafy vegetables; (8) other vitamin A-rich fruits and vegetables; (9) other vegetables; and (10) other fruits.
[o]WHO (2017).
[p]The MDD for children is grouped in (1) breast milk, (2) grains, white roots and tubers, and plantains; (3) pulses and nuts; (4) milk and milk products; (5) meat, poultry and fish; (6) eggs; (7) vitamin A-rich fruits and vegetables; (8) other fruits and vegetables.
[q]Global Diet Quality Project (2021).
[r]FAO and The World Bank (2018).
[s]Weisell and Dop (2012).
[t]Data can be accessed via http://catalog.ihsn.org/catalog.
[u]Data can be accessed via https://microdata.worldbank.org/index.php/home.
[v]Yates et al. (1998).
[w]Data can be accessed via https://www.fao.org/gift-individual-food-consumption/data-and-indicator/en/.
[x]Body mass index (BMI) is defined as weight (kg) divided by square of height (m^2).
[y]Data can be accessed via https://ncdrisc.org/index.html.
[z]Data can be accessed via https://apps.who.int/nutgrowthdb/database/search.
[aa]Data can be accessed via https://www.who.int/teams/nutrition-and-food-safety/databases/vitamin-and-mineral-nutrition-information-system.

and sub-region. Data are further converted to dietary energy supply (DES), dietary supply of protein and fat, and ultimately used to estimate dietary energy consumption (DEC) and the number of undernourishment (NoU) or prevalence of undernourishment (PoU) in countries. The PoU, which is also referred to as prevalence of hungry people in a country, provides an estimate of the percentage of the total population that are in a condition of undernourishment and hence can be used as an estimate for measuring the construct of food security. The PoU is also used as an indicator for tracking the progress on SDG 2.1 "End hunger, achieve food security and improved nutrition and promote sustainable agriculture" (FAO, 2021a).

Routine data from health systems (e.g., feeding centers and child clinics as well as nutrition-related program data) provide information relevant for FSN surveillance, such as service coverage rates for supplementation, food distribution and related efforts (Tuffrey and Hall, 2016; UNICEF and WHO, 2021). Some routine systems also collect and report on the nutritional status of the population.

Strengths and limitations

Indicators of food supply provide important estimates of the adequacy of food availability to the population. Importantly, the data are collected in a relatively consistent manner, allowing for comparisons over time and across countries. These data have the advantage of being widely and freely available for most countries worldwide and are currently serving an important accountability objective, as PoU is used to track progress for SDG 2.1 (FAO, 2021a,b).

Although the indicators can be used for high-level estimates of apparent population food consumption and food security, they do not reflect actual food access and utilization by households or individuals and the quality of data used to generate the indicators can vary greatly among countries and commodities. Therefore, they should be interpreted with caution as routine food supply data is thought to underestimate total per capita energy availability in LMICs. Additionally, despite food supply data being available for a range of food commodities, the limited specificity of foods and processed foods prevents nuanced analyses of food supply (Thar et al., 2020).

While routine data can provide useful time-trends and early warning signals, data are available at the national level only and cannot be disaggregated to identify specific populations at risk, determine seasonal trends, predict food shortages or famine, nor to assess changes in food consumption or food security within specific population groups (Thar et al., 2020).

Routine health systems data are also an important source of information for assessing the utilization of nutrition-relevant health services including micronutrient supplementation coverage or breast and complementary feeding counseling. In some contexts, routine growth monitoring data from children and other population subgroups collected through health systems is included as part of FSN surveillance systems. This has an economic advantage as it could avoid the need for separate surveys. However, routine growth monitoring data are mostly incomplete, of poor quality and, particularly in some LMIC contexts, are not representative of any population sub-group due to limited access to and costs of health facilities (Tuffrey and Hall, 2016).

Survey data for FSN surveillance

Surveys can be designed using a wide range of approaches, including large-scale multi-country surveys implemented on a periodic basis as well as nationally representative surveys that are not linked to any multi-country platform. Smaller (e.g., sub-nationally but not nationally representative) surveys designed to provide assessment of specific population subgroups at specific time points (i.e., assessing risks in the context of a crisis, see **Box 1**) may also form part of a FSN surveillance system.

Large-scale, multi-country assessments in LMICs are mostly implemented in collaboration with international partners and organizations that have developed survey programs with several modules, such as the Demographic and Health Surveys (DHS)[2] or the Multiple Indicator Cluster Surveys (MICS)[3] program by the United Nations Children's Fund (UNICEF). Both programs include household- and individual-level nutrition and health data from nationally representative samples of women and young children.[4] DHS includes a variety of core modules relevant for FSN as well as optional modules that countries can opt to include if they wish. In addition to the publicly available full reports and data sets, which can be accessed via the DHS and MICS program page, data can be further explored via the DHS STATcompiler[5] and UNICEF's Data & Analytics resource center.[6]

Similarly, household consumption and expenditure surveys (HCES) (also called household income and expenditure surveys (HIES), Living Standards Monitoring Surveys (LSMS) or national household budget surveys (NHBS)) are standardized, nationally representative surveys designed primarily to estimate levels of poverty, for setting consumer price index among related economic purposes. Several use a common methodology and are carried out in multiple countries worldwide and are supported by the World Bank.[7] In recent decades, the food consumption modules from HCES have been used as part of FSN surveillance specifically to

[2] https://dhsprogram.com/.
[3] https://mics.unicef.org/.
[4] To date, 320 and 345 surveys in over 90 and 118 countries have been conducted under the DHS program and the MICS program, respectively. Both survey programs typically include 5,000–30,000 households, and are conducted periodically, often at 3–5-year intervals.
[5] https://www.statcompiler.com/en/.
[6] https://data.unicef.org/resources/resource-type/datasets/.
[7] HCES surveys typically including 7,000–20,000 households and are periodically repeated every 3–5 years.

address gaps in household-level dietary information (Fiedler et al., 2012). While there is no single data platform for HCES, many countries have added to the IHSN Central Survey catalog[8] and the World Bank Microdata Library.[9]

Constructs and indicators commonly included in surveys as part of FSN surveillance

Food security

Household food security commonly refers to the adequacy (real or perceived) of the food available to meet the needs of the household members, taking into consideration cultural and personal preferences (Pinstrup-Andersen, 2009). The Household Hunger Scale (HHS) and the Food Insecurity Experience Scale (FIES) are scales designed to measure food insecurity. The HHS measures food insecurity at the household level, while the FIES was initially developed for use at the individual-level but can also be adjusted to be used on household level. Both have been validated across multiple contexts and are particularly useful to understand and compare population groups and sub-groups within populations (Ballard et al., 2011). The scales collect information on the degree to which household members worry about running out of food and modify habits due to a lack of food without making any quantitative estimations of food availability. While the HHS focuses on severe food insecurity over a recall period of 30 days, the FIES includes a broader range of questions on household's (or individual's) food insecurity experience, commonly over a period of 1 year (Ballard et al., 2013).

The FIES is assessed annually through the Gallup World Poll (in collaboration with FAO) and used for measuring progress toward achieving SDG 2.1. The Gallup World Poll implements nationally representative surveys in a number of countries annually, including information on a variety of topics. As of 2021, Gallup has collected FIES data for 77 countries, with data available through FAO's Micronutrient Catalog.[10] Data on the HHS can be accessed via DHS and MICS.

Diet quality

Diet quality is a critical construct for FSN that is assessed through a variety of approaches to describe dietary intake (qualitatively or quantitatively). No single indicator of diet quality has been developed that captures all relevant aspects of dietary intake and given the complexity of diets, it is unlikely that a single indicator would do so. Diet quality indicators can be classified broadly into three types, either at aggregated household level based on food available and household composition, or for individuals. Nutrient adequacy provides a quantitative comparison of nutrient availability at household level, or intake of individuals with requirements. Food intake provides a quantitative comparison of food intake with guidance or recommendations, such as consumption of 400 g of fruit and vegetables daily. Diet quality may also be assessed by reviewing dietary patterns and their alignment with dietary patterns that have been associated with better health outcomes in some contexts (e.g., the Mediterranean diet) (Neufeld et al., 2021).

Household food consumption data provide an estimate of the amount of food or energy available to be consumed on a per capita or household basis (e.g., kcal/capita/day). Estimates are based on the reported amounts acquired and/or purchased for the household on a seven-day recall period, commonly assessed through HCES (FAO and The World Bank, 2018). In the absence of individual-level data on energy, food, and nutrient intakes, it is possible to apply the adult male equivalent (AME) method to such household-level data to generate individual-level estimates. The AME method provides estimates on intra-household food distribution in accordance with expected energy expenditures of the individual based on their age and sex (Weisell and Dop, 2012). Furthermore, individual-level nutrient intakes can be derived when multiplying the estimated individual food intakes (derived using the AME methodology) by nutrient content levels for each food using Food Composition Databases. This approach has been shown to provide comparable estimates of food, energy and nutrient intakes to 24 h recalls among most population groups with some exceptions among young children and for some foods (Coates et al., 2017; Dary and Jariseta, 2012; Engle-Stone and Brown, 2015; Jariseta et al., 2012; Sununtnasuk and Fiedler, 2017).

The Household Dietary Diversity Score (HDDS) and the Food Consumption Score (FCS) by the UN World Food Programme (WFP) are proxies for household caloric availability counts of food groups consumed. These indicators are based on some evidence that household consumption of a higher number of food groups is correlated with better diet adequacy for several nutrients (Swindale and Bilinksy, 2006). The HDDS counts the households consumption of up to 12 food groups in the past 24 h and can be further disaggregated for specific foods groups (Kennedy et al., 2011). The FCS includes the frequency of the households consumption of eight food groups in the past seven days and multiplies these with the food group weight according to their nutritive value (WFP VAM Resource Center, 2019). For both indicators reasonable correlations have been found with household socio–economic status. Similarly, to the approach taken by Gallup World Poll (see food security) WFP collects household dietary diversity data on a rolling basis, traditionally every 1–3 months, to update the platform on a daily basis. WFP's near real-time data are currently being assessed in 34 LMICs and are accessible through WFP's Hunger map.[11]

Individual-level dietary diversity indicators follow the same principle as the HDDS described above. The most common indicators used are the Minimum Dietary Diversity for Women of reproductive age (MDD-W) and the Minimum Dietary Diversity (MDD) for children. In many contexts, women's nutrition knowledge is central for household food security and nutrition as their child's diet quality tends to be closely linked to their own (Bonis-Profumo et al., 2020; Debela et al., 2017). Both indicators have been

[8] http://catalog.ihsn.org/catalog.
[9] https://microdata.worldbank.org/index.php/home.
[10] https://microdata.fao.org/index.php/catalog.
[11] https://hungermap.wfp.org/.

found to be positively associated with nutrient adequacy of diets and to reflect appropriate complementary feeding practices (Martin-Prével et al., 2015; Working Group on Infant and Young and Child Feeding Indicators, 2007). The MDD-W and MDD for children were set up to include ten and eight food groups, respectively (FAO, 2021c; WHO, 2017).

Questionnaire modules on HDDS, MDD-W and MDD for children are commonly added to the DHS and MICS and available via the respective platforms.

The Global Diet Quality Project, a collaboration of the Harvard Department of Global Health and Population, the Global Alliance for Improved Nutrition (GAIN), and the Gallup World Poll collects data on several aspects of diet quality. Based on consumption of 29 food groups, nutrient adequacy (MDD-W) and dietary risk factors for NCDs (GDR score[12]) are assessed and displayed on the project website.[13] Within the first phase (2021–2022) 56 countries were included in a multicounty assessment through the Gallup World Poll. The questionnaire has now been adapted to more than 100 countries (Global Diet Quality Project, 2021).

Individual dietary intake can be measured using a variety of measures including food records, 24 h dietary recalls and food frequency questionnaires (FFQs), all with a variety of strengths and limitations for varying purposes (Gibson, 2005). Food records and multi-day 24 h dietary recalls are needed to quantify usual food and nutrient intakes for individuals, but all methods have potential to characterize dietary patterns and generate several indicators of diet quality at population level. For example the Estimated Average Requirement (EAR) cut point or probability method can provide estimates of individual-level nutrient adequacy, which can be used as a proxy for risk of micronutrient deficiency (Yates et al., 1998).

The coverage of fortified foods is also an important indicator in FSN surveillance. Large-scale national surveys, including DHS or MICS, often include household-level coverage indicators for iodized salt. Additionally, Fortification Assessment Coverage Toolkit (FACT) surveys include household-level coverage indicator for all fortified foods in a country using a standardized methodology (Friesen et al., 2019). To date, FACT surveys have been conducted in more than 16 countries as part of ongoing program monitoring efforts (Aaron et al., 2016, 2017; Knowles et al., 2017; Rohner et al., 2016). In addition, data from HCES and other food intake surveys are also increasingly being used to inform the design of nutrition programs such as food fortification (e.g., selecting foods for fortification based on their patterns of consumption) (Fiedler et al., 2008).

Nutritional status

Nutritional status is commonly assessed in surveys using anthropometric indicators, such as stunting, wasting, underweight, overweight and obesity. The assessments for these indicators include data on weight, height or length and age. Overweight/obesity as well as underweight are indicators that can be assessed for both adults and children. For adults, the body mass index (BMI) is used as method, whereas weight-for height and weight-for-age Z scores are used for child overweight/obesity and underweight, respectively. Overweight/obesity are proxies for abnormal fat accumulation, while underweight is a proxy for poor nutrition generally. Stunting and wasting are indicators only used for children and are interpreted as impaired growth and recent and severe weight loss, respectively. Data for these indicators and measures can be found via DHS and MICS programs as well as NCD-RisC[14] (for BMI only) and the WHO Global Database on Child Growth and Malnutrition.[15]

Ideally, biochemical measures of nutritional status, particularly for micronutrient status, should also be part of FSN surveillance. Some nationally representative surveys have included biological samples (usually blood and/or urine) for assessment of micronutrient status of individuals, but data are scarce (Brown et al., 2021). While DHS and MICS may include samples for multiple micronutrients (optional module), only hemoglobin concentration for assessment of anemia prevalence is part of the core modules. Where available, data on micronutrient status have been compiled and made publicly available through the Vitamin and Mineral Nutrition Information System (VMNIS)[16] maintained by the World Health Organization (WHO).

Strengths and limitations

While the collection of survey data typically requires greater financial resources than routine data, an important advantage of multicountry surveys like DHS, MICS and HCES is that they provide nationally representative data using standardized questionnaires that allow for comparison over time and across countries. They also allow for the questionnaire modules to vary based on individual country needs. Conversely, the standardization of the questionnaire modules makes it difficult to add single questions into the surveys, although the DHS for example supports the use of qualitative research to complement the survey findings. Smaller, non-nationally representative surveys have the strength to focus specifically on high-risk groups, sub-regions, and local data needs thus providing data for those whom policy makers may need to design interventions (Tuffrey, 2016).

There are several limitations to the collection of household-level data on food availability in LMIC contexts, particularly among very poor households. For example, inventory methods may underestimate food availability in households that purpose food daily with little to no food stored in the home. For dietary intake indicators, household-level assessments are still not as accurate as individual-level methods as they do not allow for intra-household differentiation and may miss out on food being wasted, given away or consumed away from home.

[12]The Global Dietary Recommendations (GDR) score indicator reflects global dietary recommendations associated with non-communicable diseases (NCDs). A lower score is associated with a higher risk for NCDs (Herforth et al., 2020).
[13]https://www.globaldietquality.org/home.
[14]https://ncdrisc.org/index.html.
[15]https://apps.who.int/nutgrowthdb/database/search.
[16]https://www.who.int/teams/nutrition-and-food-safety/databases/vitamin-and-mineral-nutrition-information-system.

Individual dietary assessments such as 24 h recalls, and FFQs provide more accurate data on dietary intake. However, they are often resource intensive to implement and analyze and are subject to recall bias and underreporting. To ease the collection of such data, particularly in LMICs, innovative tools such as computer assisted personal interviewing (CAPI) on tablets or mobile phones are often used to replace the traditional paper-based questionnaires in many surveys. Furthermore, promising mobile phone-apps have been developed that use images of food plates before and after eating to estimate portion size and nutritional content (Ambrosini et al., 2018) as well as specific guidance related to the development of food photographs and portion size estimation in 24 h recalls for LMICs (Vossenaar et al., 2020).

Individual anthropometric and biochemical indicators provide the most objective measure of nutritional status. However, both are time and cost intensive, and technically challenging to collect and analyze. Improved methods have been developed to ease the level of effort required to assess anthropometric data, particularly in LMICs. For example, the Child Growth Monitor (CGM), an application using augmented reality with artificial intelligence (AI) was developed to determining weight and height and hence malnutrition indicators through a 3D scan of children (Welthungerhilfe, 2018).

Overall, dietary intake and nutritional status data that are available in LMICs focus mainly on children under 5 years of age and women of reproductive age. Data on nutritional status and dietary intake of other groups, particularly adolescents (Neufeld et al., 2021), but also older children, adult men, and older adults, remain very limited.

Accessing data through international platforms

Survey data, particularly that generated from multi-country data platforms (e.g., DHS, MICS) are increasingly being made publicly available and there are several global efforts to support the compilation and increase public access to such nationally and sub-nationally representative data. For example, the joint Global Individual Food consumption data Tool (GIFT)[17] is a growing platform managed by FAO and WHO, which includes food-based indicators on individual dietary data conducted at national or sub-national levels. As of 2021, the platform included data from 23 mainly LMICs with a plan to expand. Similarly, WHO's Global Database on Child Growth and Malnutrition[18] compiles nationally and sub-nationally representative survey data from household and individual level assessments in over 230 countries (including DHS data). Indicators such as child anthropometric status disaggregated by sex and location (urban vs. rural) are included.

The Global Nutrition and Policy Consortium set up the Global Dietary Database (GDD),[19] which uses public and private data sources of individual level dietary data to model characteristics of diets in populations around the world. Currently, the database contains data on 54 dietary factors including foods, beverages, macro- and micronutrients. Estimates are disaggregated by age, sex, education level, and location.

The Global Burden of Diseases Platform[20] by the Institute for Health Metrics and Evaluation (IHME) focusses on the display of the prevalence of a given disease or risk factor and the relative harm it causes. Under this approach the platform makes use of over 30,000 data sources from administrative data assessments, demographic surveys, scientific data and other sources to compile FSN indicators. The estimates are available on annual basis, disaggregated by sub-national level, sex and age.

The newly established Food Systems Dashboard[21] is a large multi-country data platform including food system and nutrition data on food production, the food environment (e.g., access and availability of food, food prices, convenience or desirability), individual factors, consumption, dietary behavior and factors like policies and environment. While it is currently limited to national data only, there are plans to include sub-national level data in the near future.

Case studies

Continuous DHS assessment, Senegal

Senegal was one of 29 countries included in the first phase of the DHS program when it began in 1984. Between Senegal's first DHS in 1986 and 2010/11, six standard DHS and two Malaria Indicator Surveys (MIS) were collected. Like most countries, surveys were implemented once every 3–5 years or occasionally less frequently. In 2012, Senegal adopted a new, continuous approach to enhance data quality and provide more frequent and up-to-date data for planning, monitoring and tracking progress of health and population programs each year (DHS Senegal, 2013). Using more targeted surveys, Senegal has been able to adapt its data collection to meet specific decision-making needs. Since 2012/13, seven DHS and Service Provision Assessment (SPA) rounds have been conducted by the National Agency for Statistics and Demography (ANSD) in collaboration with the Ministry of Health and Social Action (MSAS) and technical assistance from the DHS Program. While the DHS assessments are led by the Government of Senegal, additional financial support is provided by the United States Agency for International Development (USAID), the Cellule de Lutte contre la Malnutrition (CLM), the World Bank, UNICEF, the United Nations Population Fund (UNFPA) and Nutrition International. In addition to socio-economic indicators, the continuous DHS captures data from households on fertility,

[17] http://www.fao.org/gift-individual-food-consumption/overview/en/.
[18] https://apps.who.int/nutgrowthdb/database/search.
[19] https://www.globaldietarydatabase.org/.
[20] http://ghdx.healthdata.org/.
[21] https://foodsystemsdashboard.org/.

contraception, child mortality, reproductive health, children's heath (e.g., details on breastfeeding and nutritional status) and malaria. The SPA captures information on the availability of different health services (e.g., children, maternal health and family planning), the ability to diagnose diseases including HIV/AIDS, NCDs, tuberculosis and malaria and the availability of drugs and treatments for certain diseases (Agence Nationale de la Statistique et de la Démographie and ICF, 2020a,b).

Integrated Reproductive Health and Nutrition Surveillance System (SIVESNU), Guatemala

In 2010, the Guatemalan government designed a population-based nutrition surveillance system with support from the Institute of Nutrition of Central America and Panama (INCAP), UNICEF, USAID, and the US Centers for Disease Control and Prevention (CDC). Based on the identified lack of standardized surveillance of important nutrition-related information (specifically related to all large-scale nutrition-specific interventions), a prototype was developed with the aim of providing accurate, timely, reliable and representative data—as a complement to information provided by the institutional surveillance system—to inform national decision making and program planning related to improving the health and nutritional status of women and children (CDC, 2020; INCAP, 2018). After testing the prototype in 2011, the surveillance system was expanded to include reproductive health and the Integrated Reproductive Health and Nutrition Surveillance System (SIVESNU) was launched in 2013. In detail, the SIVESNU monitors chronic malnutrition, overweight and obesity, micronutrient status (e.g., iron, zinc and vitamin A), anemia, supplementation and diets (i.e., breastfeeding and complementary feeding and dietary diversity for children and women, respectively) as well as household level socio-economic characteristics, food security, coverage of fortified salt, sugar, bread and the coverage of government programs and household level health, nutrition and food safety (INCAP, 2018). This nationally and sub-nationally representative sample collects data at household level and for individual population sub-groups within the household (children 0–59 months of age, non-pregnant women 15–49 years of age, and pregnant women) during each cycle. The survey is rolled out over about 6–8 months per cycle and repeated on an approximate annual basis (Friedman, 2014). The SIVESNU design is flexible, and content can be revised for each surveillance cycle. For example, in 2017/18, the SIVESNU added a zika questionnaire module among a representative sample of school-aged children between 6 and 14 years and zika biomarker testing among women. Given the lower cost and high quality of the SIVESNU compared to other national surveys, the surveillance system provides a sustainable source of accurate, timely, reliable and representative data, which aims to help the Guatemala government to determine its health priorities and improve programs and interventions (Palmieri et al., 2022).

Conclusion

Rigorous FSN surveillance systems that provide accurate, timely and sub-nationally disaggregated data can help to ensure effective evidence-based policymaking and program implementation. LMICs face particularly high challenges to establish FSN surveillance systems due to rapidly transforming food systems, overburdened health systems, limited data infrastructure and a lack of technical capacity and resources. Especially in resource-limited settings, establishing FSN surveillance requires strategic decisions on what data to collect, driven by the purpose for which it will be used, ensuring appropriate and validated indicators that reflect the constructs of interest and with close consideration of data quality and resource implications. In most contexts a purposeful mix of routine and survey data will likely meet these needs.

Conflict of Interest

The views expressed in this publication are those of the author(s) and do not necessarily reflect the views of FAO.

Acknowledgments

We thank Maria Elena Jefferds for her review and input to the case study on Guatemala.

References

Aaron, G.J., Sodani, P.R., Sankar, R., Fairhurst, J., Siling, K., Guevarra, E., Norris, A., Myatt, M., 2016. Household coverage of fortified staple food commodities in Rajasthan, India. PLoS One 11, e0163176. https://doi.org/10.1371/journal.pone.0163176.

Aaron, G.J., Friesen, V.M., Jungjohann, S., Garrett, G.S., Neufeld, L.M., Myatt, M., 2017. Coverage of large-scale food fortification of edible oil, wheat and maize flours varies greatly by vehicle and country but is consistently lower among the most vulnerable: results from coverage surveys in eight countries. J. Nutr. 147, 984S–94S.

Action Against Hunger, 2019. Emergency Response. Action Hunger. https://www.actionagainsthunger.org/impact/emergency-response. (Accessed 20 December 2021).

Agence Nationale de la Statistique et de la Démographie, ICF, 2020a. Sénégal. Enquête Continue sur la Prestation des Services de Soins de Santé (ECPSS) 2019, Dakar, Senegal and Rockville, MD, USA.

Agence Nationale de la Statistique et de la Démographie, ICF, 2020b. Sénégal. Enquête Démographique et de Santé Continue (EDS-Continue) 2019, Rockville, MD, USA.

Ambrosini, G.L., Hurworth, M., Giglia, R., Trapp, G., Strauss, P., 2018. Feasibility of a commercial smartphone application for dietary assessment in epidemiological research and comparison with 24-h dietary recalls. Nutr. J. 17, 5. https://doi.org/10.1186/s12937-018-0315-4.
Ballard, T., Coates, J., Swindale, A., Deitchler, M., 2011. Household Hunger Scale: Indicator Definition and Measurement Guide. FANTA II Project, FHI 360, Washington, DC.
Ballard, T.J., Kepple, A.W., Cafiero, C., 2013. The Food Insecurity Experience Scale. FAO, Rome, Italy. Technical Paper.
Bonis-Profumo, G., Stacey, N., Brimblecombe, J., 2020. Maternal diets matter for children's dietary quality: seasonal dietary diversity and animal-source foods consumption in rural Timor-Leste. Matern. Child Nutr. 17, e13071. https://doi.org/10.1111/mcn.13071.
Brown, K.H., Moore, S.E., Hess, S.Y., McDonald, C.M., Jones, K.S., Meadows, S.R., Manger, M.S., Coates, J., Alayon, S., Osendarp, S.J.M., 2021. Increasing the availability and utilization of reliable data on population micronutrient (MN) status globally: the MN Data Generation Initiative. Am. J. Clin. Nutr. 114, 862–870. https://doi.org/10.1093/ajcn/nqab173.
CDC, 2020. Maternal and Child Health in Guatemala. Centers for Disease Control and Prevention. https://www.cdc.gov/nutrition/micronutrient-malnutrition/projects/guatemala.html. (Accessed 12 January 2022).
Coates, J., Rogers, B.L., Blau, A., Lauer, J., Roba, A., 2017. Filling a dietary data gap? Validation of the adult male equivalent method of estimating individual nutrient intakes from household-level data in Ethiopia and Bangladesh. Food Pol. 72, 27–42. https://doi.org/10.1016/j.foodpol.2017.08.010.
Dary, O., Jariseta, Z.R., 2012. Validation of dietary applications of household consumption and expenditures surveys (HCES) against a 24-hour recall method in Uganda. Food Nutr. Bull. 33, S190–S198. https://doi.org/10.1177/15648265120333S207.
DataDENT, 2021. Our Work. Data for Decisions to Expand Nutrition Transformation. https://datadent.org/wp-content/uploads/2021/03/DataDENT-One-Pager_Email_31-Mar-2021_FINAL.pdf. (Accessed 21 December 2021).
Debela, B.L., Demmler, K.M., Rischke, R., Qaim, M., 2017. Maternal nutrition knowledge and child nutritional outcomes in urban Kenya. Appetite 116, 518–526. https://doi.org/10.1016/j.appet.2017.05.042.
Demmler, K.M., Klasen, S., Nzuma, J.M., Qaim, M., 2017. Supermarket purchase contributes to nutrition-related non-communicable diseases in urban Kenya. PLoS One 12, e0185148. https://doi.org/10.1371/journal.pone.0185148.
DHS Senegal, 2013. Continuous Demographic and Health Survey.
Engle-Stone, R., Brown, K.H., 2015. Comparison of a household consumption and expenditures survey with nationally representative food frequency questionnaire and 24-hour dietary recall data for assessing consumption of fortifiable foods by women and young children in Cameroon. Food Nutr. Bull. 36, 211–230. https://doi.org/10.1177/0379572115587272.
FAO, 2021a. FAOSTAT. Food Balances (2014-). Metadata. Food and Agriculture Organization of the United Nations. http://www.fao.org/faostat/en/#data/FBS/metadata. (Accessed 13 September 2021).
FAO, 2021b. The State of Food Security and Nutrition in the World 2021. FAO, IFAD, UNICEF, WFP and WHO. https://doi.org/10.4060/cb4474en.
FAO, 2021c. Minimum Dietary Diversity for Women: An Updated Guide to Measurement—From Collection to Action. Food and Agriculture Organization of the United Nations, Rome, Italy.
FAO, The World Bank, 2018. Food Data Collection in Household Consumption and Expenditure Surveys. Guidelines for Low- and Middle-Income Countries. Rome, Italy.
Fiedler, J.L., Lividini, K., Bermudez, O.I., Smitz, M.-F., 2012. Household consumption and expenditures surveys (HCES): a primer for food and nutrition analysts in low- and middle-income countries. Food Nutr. Bull. 33, S170–S184. https://doi.org/10.1177/15648265120333S205.
Fiedler, J.L., Smitz, M.-F., Dupriez, O., Friedman, J., 2008. Household income and expenditure surveys: a tool for accelerating the development of evidence-based fortification programs. Food Nutr. Bull. 29, 306–319.
Friedman, G., 2014. Review of National Nutrition Surveillance Systems. FHI 360/FANTA, Washington, DC.
Friesen, V.M., Jungjohann, S., Mbuya, M.N.N., Harb, J., Visram, A., Hug, J., Garrett, G.S., Neufeld, L.M., 2019. Fortification Assessment Coverage Toolkit (FACT) Manual. Global Alliance for Improved Nutrition (Geneva) and Oxford Policy Management, Oxford.
Frongillo, E.A., Tofail, F., Hamadani, J.D., Warren, A.M., Mehrin, S.F., 2014. Measures and indicators for assessing impact of interventions integrating nutrition, health, and early childhood development. Ann. N. Y. Acad. Sci. 1308, 68–88. https://doi.org/10.1111/nyas.12319.
Gibson, R.S., 2005. Principles of Nutritional Assessment. Oxford University Press.
Global Diet Quality Project, 2021. DQQ Tools & Data. https://www.globaldietquality.org/home. (Accessed 13 January 2022).
Herforth, A.W., Wiesmann, D., Martínez-Steele, E., Andrade, G., Monteiro, C.A., 2020. Introducing a suite of low-burden diet quality indicators that reflect healthy diet patterns at population level. Curr. Dev. Nutr. 4. https://doi.org/10.1093/cdn/nzaa168.
HLPE, 2017. Nutrition and food systems. UN Commission of world food security high level Panel of Experts (HLPE), Rome, Italy.
HLPE, 2022. Data Collection and Analysis Tools for Food Security and Nutrition: Towards Enhancing Effective, Inclusive, Evidence-Informed, Decision Making. A report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security, Rome, Italy.
INCAP, 2018. Informe del Sistema de Vigilancia Epidemiológica de Salud y Nutrición—SIVESNU. Instituto de Nutrición de Centro América y Panamá, INCAP, Guatemala.
INDDEX Project, 2018. Data4Diets: Building Blocks for Diet-Related Food Security Analysis. Tufts University, Boston MA. https://inddex.nutrition.tufts.edu/data4diets. (Accessed 1 October 2021).
Jariseta, Z.R., Dary, O., Fiedler, J.L., Franklin, N., 2012. Comparison of estimates of the nutrient density of the diet of women and children in Uganda by household consumption and expenditures surveys (HCES) and 24-hour recall. Food Nutr. Bull. 33, S199–S207. https://doi.org/10.1177/15648265120333S208.
Kennedy, G., Ballard, T., Dop, M., 2011. Guidelines for Measuring Household and Individual Dietary Diversity. FAO, Rome, Italy.
Khan, M.S., Meghani, A., Liverani, M., Roychowdhury, I., Parkhurst, J., 2018. How do external donors influence national health policy processes? Experiences of domestic policy actors in Cambodia and Pakistan. Health Pol. Plann. 33, 215–223. https://doi.org/10.1093/heapol/czx145.
Knowles, J.M., Garrett, G.S., Gorstein, J., Kupka, R., Situma, R., Yadav, K., Yusufali, R., Pandav, C., Aaron, G.J., 2017. Household coverage with adequately iodized salt varies greatly between countries and by residence type and socioeconomic status within countries: results from 10 national coverage surveys. J. Nutr. 147, 1004S–1014S.
Martin-Prével, Y., Allemand, P., Wiesmann, D., Arimond, M., Ballard, T., Deitchler, M., Dop, M., Kennedy, G., Lee, W., Moursi, M., 2015. Moving Forward on Choosing a Standard Operational Indicator of Women's Dietary Diversity. FAO, Rome, Italy.
Neufeld, L.M., Andrade, E.B., Ballonoff Suleiman, A., Barker, M., Beal, T., Blum, L.S., Demmler, K.M., Dogra, S., Hardy-Johnson, P., Lahiri, A., Larson, N., Roberto, C.A., Rodríguez-Ramírez, S., Sethi, V., Shamah-Levy, T., Strömmer, S., Tumilowicz, A., Weller, S., Zou, Z., 2021. Food choice in transition: adolescent autonomy, agency, and the food environment. Lancet 399, 185–197. https://doi.org/10.1016/S0140-6736(21)01687-1.
Palmieri, M., Flores-Ayala, R., Mesarina, K., Mazariegos, D.I., Martínez, C., López, B., Santizo, M.C., Whitehead Jr., R.D., Addo, O.Y., Aponte, J., Quiñónez, E.D.L., Sagastume, M.J., Jefferds, M.E.D., 2022. Experiences and lessons learned in developing and implementing a population-based nutrition and health surveillance system in Guatemala 2011–2021. Curr. Dev. Nutr. 6, nzac027. https://doi.org/10.1093/cdn/nzac027.
Pinstrup-Andersen, P., 2009. Food security: definition and measurement. Food Secur. 1, 5–7. https://doi.org/10.1007/s12571-008-0002-y.
Popkin, B.M., Corvalan, C., Grummer-Strawn, L.M., 2020. Dynamics of the double burden of malnutrition and the changing nutrition reality. Lancet 395, 65–74. https://doi.org/10.1016/S0140-6736(19)32497-3.
Rohner, F., Leyvraz, M., Konan, A.G., Esso, L.J.C.E., Wirth, J.P., Norte, A., Adiko, A.F., Bonfoh, B., Aaron, G.J., 2016. The potential of food fortification to add micronutrients in young children and women of reproductive age—findings from a cross-sectional survey in abidjan, côte d'Ivoire. PLoS One 11, e0158552. https://doi.org/10.1371/journal.pone.0158552.
Shroff, Z.C., Javadi, D., Gilson, L., Kang, R., Ghaffar, A., 2017. Institutional capacity to generate and use evidence in LMICs: current state and opportunities for HPSR. Health Res. Pol. Syst. 15, 94. https://doi.org/10.1186/s12961-017-0261-1.

Sununtnasuk, C., Fiedler, J.L., 2017. Can household-based food consumption surveys be used to make inferences about nutrient intakes and inadequacies? A Bangladesh case study. Food Pol. 72, 121–131. https://doi.org/10.1016/j.foodpol.2017.08.018.
Swindale, A., Bilinksy, P., 2006. Household Dietary Diversity Score (HDDS) for Measurement of Household Food Access: Indicator Guide (Version 2). FHI 360/FANTA, Washington, DC.
TEAM, 2021. Report of the Technical Consultation on Measuring Healthy Diets: Concepts, Methods and Metrics, Virtual Meeting 18–20 May 2021. WHO-UNICEF Technical Expert Advisory Group on Nutrition Monitoring (TEAM).
Thar, C.-M., Jackson, R., Swinburn, B., Mhurchu, C.N., 2020. A review of the uses and reliability of food balance sheets in health research. Nutr. Rev. 78, 989–1000. https://doi.org/10.1093/nutrit/nuaa023.
Tuffrey, V., 2016. Nutrition Surveillance Systems: Their Use and Value. Save the Children and Tranform Nutrition, London, UK.
Tuffrey, V., Hall, A., 2016. Methods of nutrition surveillance in low-income countries. Emerg. Themes Epidemiol. 13, 4. https://doi.org/10.1186/s12982-016-0045-z.
UNICEF, 1998. Focus on Nutrition. Oxford University Press for UNICEF, Oxford.
UNICEF, 2020. Nutrition Information Management, Surveillance and Monitoring in the Context of COVID-19, Brief No. 3. United Nations International Children's Emergency Fund.
UNICEF, WHO, 2021. National Nutrition Information Systems, the Fundamental Series; Modules 1–5. United Nations Children's Fund (UNICEF) and the World Health Organization (WHO), New York.
United Nations, 2015. Transforming Our World: the 2030 Agenda for Sustainable Development.
USAID, 2021. The DHS Program. Guidelines for Virtual Technical Assistance and Survey Implementation During the COVID-19 Pandemic.
Verstraeten, R., 2021. Investing in the data value chain for nutrition in West Africa. Field Exch. 64, 15.
Vossenaar, M., Crispim, S.P., Lubowa, A., Deitchler, M., Moursi, M., Arimond, M., 2020. Guidance for the Development of Food Photographs for Portion Size Estimation in Quantitative 24-hour Dietary Recall Surveys in Low- and Middle-Income Countries. Intake—Center for Dietary Assessment/FHI Solutions and Department of Nutrition, Federal University of Parana, Washington, DC and Curitiba, Brazil.
Weisell, R., Dop, M.C., 2012. The adult male equivalent concept and its application to Household Consumption and Expenditures Surveys (HCES). Food Nutr. Bull. 33, S157–S162. https://doi.org/10.1177/15648265120333S203.
Welthungerhilfe, 2018. Child Growth Monitor: A Game-Changing App to Detect Malnutrition. URL. https://childgrowthmonitor.org/. (Accessed 7 October 2021).
WFP VAM Resource Centre, 2019. FCS—Food Consumption Score—Data Analysis. World Food Program. URL. https://resources.vam.wfp.org/data-analysis/quantitative/food-security/fcs-food-consumption-score. (Accessed 13 October 2021).
WHO, 2013. Food and Nutrition Surveillance Systems: Technical Guide for the Development of a Food and Nutrition Surveillance System. WHO Regional Publications, Eastern Mediterranean Series, Cairo, Egypt.
WHO, 2017. Global Nutrition Monitoring Framework: Operational Guidance for Tracking Progress in Meeting Targets for 2025. World Health Organization, Geneva, Switzerland.
Working Group on Infant and Young, Child Feeding Indicators, 2007. Developing and validating simple indicators of dietary quality of infants and young children in developing countries: additional analysis of 10 data sets. Food Nutr. Tech. Assist. 360, 74.
Yates, A.A., Schlicker, S.A., Suitor, C.W., 1998. Dietary reference intakes: the new basis for recommendations for calcium and related nutrients, B vitamins, and choline. J. Am. Diet Assoc. 98, 699–706. https://doi.org/10.1016/S0002-8223(98)00160-6.

Relevant websites

Data DENT. https://datadent.org/.
DHS Program. https://dhsprogram.com/Data/.
DHS STATcompiler. https://www.statcompiler.com/en/.
FAO GIFT. https://www.fao.org/gift-individual-food-consumption/data-and-indicator/en/.
FAO Microdata Catalogue. https://microdata.fao.org/index.php/catalog.
FAOSTAT. http://www.fao.org/faostat/en/#home.
Food Systems Dashboard. https://foodsystemsdashboard.org/.
GDD. https://www.globaldietarydatabase.org/.
Global Burden of Diseases Platform. http://ghdx.healthdata.org/.
Global Diet Quality Project. https://www.globaldietquality.org/home.
IHSN Central Survey Catalog. http://catalog.ihsn.org/catalog.
INDDEX Project. https://inddex.nutrition.tufts.edu/inddex-project.
MICS Program. https://mics.unicef.org/.
NCD RisC. https://ncdrisc.org/index.html.
UNICEF Datasets. https://data.unicef.org/resources/resource-type/datasets/.
WFP Hungermap. https://hungermap.wfp.org/.
WHO Global Database on Child Growth and Malnutrition. https://apps.who.int/nutgrowthdb/database/search.
WHO VMNIS. https://www.who.int/teams/nutrition-and-food-safety/databases/vitamin-and-mineral-nutrition-information-system.
World Bank Microdata Library. https://microdata.worldbank.org/index.php/home.

Nutrition transition, diet change, and its implications

Daniel J. Hoffman and Hannah Posluszny, Rutgers, the State University of New Jersey, New Brunswick, NJ, United States

© 2023 Elsevier Ltd. All rights reserved.

This is an update of B.M. Popkin, Nutrition Transition, Diet Change, and its Implications, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 320–328, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00196-3.

Introduction 436
Demographic and epidemiological transitions 436
Rapid shifts in dietary and activity patterns and body composition 437
Obesity trends 438
Dietary changes: shift in the overall structure over time 439
Rapid social change is important: urbanization, rapid demographic change, and other behavioral changes are occurring simultaneously 440
Urbanization 440
Structural shifts in income–diet relationships are occurring 441
Mass media and internet use 441
Health effects: how does biology interact with these rapid social changes? 442
Epidemic of CVD and type 2 diabetes 442
Social burden of changes in diet, body composition, and health 443
Summary 443
References 443
Further reading 443

Key points

- Globalization has resulted in dramatic shifts in diet and physical activity patterns that promote obesity and chronic diseases.
- Diets have shifted from high-fiber and low-processed sugar to high-fat and high in processed foods.
- Activity patterns have shifted as technology has globalized and communities and work forces are more sedentary than ever before.
- More people than ever are overweight or obese and the prevalence of chronic diseases is increasing rapidly in lower income countries.
- The consumption of ultra-processed foods, in both higher and lower income countries continues to increase and contribute to the global burden of chronic diseases.

Glossary

Dietary change Foods consumed and overall food patterns

Edible oils Vegetable oils extracted from oil seeds

Globalization Commonly refers to increasing global norms in economic order, trade in goods and services or set of changes associated with the rapid exchange of goods, services, ways of thinking

Mass media All modern forms of communication: radio to TV to computer, cell phones, movies, magazines, etc.

Nutrition transition Stages of patterns of eating, drinking, and moving

Obesity BMI > 30

Role of income Measure of amount of money or money in kind we obtain over a period of time

Sugar-sweetened beverages Calorically sweetened waters -carbonated or noncarbonated, flavored in any manner from colas to energy drinks to vitamin waters

Ultra-processed foods Food products made from formulations of ingredients derived from a series of industrial processes

Urbanization Residence in urban area or process of modernization of a community to bring set of services and infrastructural changes linked with urban residency

Introduction

In the past 100 years, the world has undergone a transformation in international trade, development, and health. No longer faced with high mortality from infectious diseases, the globalization of food products coupled with labor shifts there has been an unprecedented increase in diets that are less than healthful and physical activity patterns that promote obesity and chronic diseases. At the same time, the world faces ongoing civil violence and health disparities that leave some communities food insecure, but with access to processed foods. These changes are commonly referred to as the "nutrition transition." This article will explore various aspects of the nutrition transition to better understand how dietary and physical activity patterns influence global health.

Demographic and epidemiological transitions

The world has experienced rapid shifts in dietary patterns and body composition that have several important implications for humans. In many ways, these shifts are a continuation of large-scale changes that have occurred repeatedly over time (Fig. 1). However, the changes facing low- and middle-income countries (LMICs) appear to be very rapid. Broad shifts continue to occur throughout the world in population size and demographics, disease patterns, as well as dietary and physical activity patterns. The former two sets of dynamic shifts are termed the demographic and epidemiological transitions. The latter, whose changes are reflected in nutritional outcomes, such as changes in stature and body composition, is termed the nutrition transition.

Human diet, physical activity patterns, and nutritional status have undergone a sequence of major shifts, defined as broad patterns of food use and their corresponding nutrition-related diseases (Popkin, 2002). During the past three centuries, the pace of dietary and activity changes appears to have accelerated to varying degrees in different regions of the world. Furthermore, dietary and activity changes are paralleled by major changes in health status as well as by major demographic and socioeconomic changes. Obesity emerges early under these shifting conditions, as does the level and age composition of morbidity and mortality. Although there are five broad nutrition patterns dating back to the origins of modern humans, the focus of this article is on the three most recent periods (Fig. 2). For convenience, the patterns are outlined as historical developments; however, earlier patterns are not restricted to the periods in which they first arose but, rather, they continue to characterize certain geographic and socioeconomic subpopulations. The first two patterns relate to earlier periods in the evolution of humans—the first pattern of collecting food and the second pattern of famine. The following are the three later periods.

Pattern 3: Receding famine: The consumption of starchy staples was predominant and continues to be so, but these items are less important in this low-fat diet as limited amounts of fruits, vegetables, and animal protein are increasingly added to this low-fat and high-fiber diet. Many earlier civilizations made great progress in reducing chronic hunger and famines, but only in the last third of the past millennium have these changes become widespread, leading to marked shifts in diets. However, famines continued well into the 18th century in some parts of Europe and unfortunately remain common in some regions of the world. Activity patterns are shifting, and inactivity and leisure have become a part of the lives of increasingly more individuals.

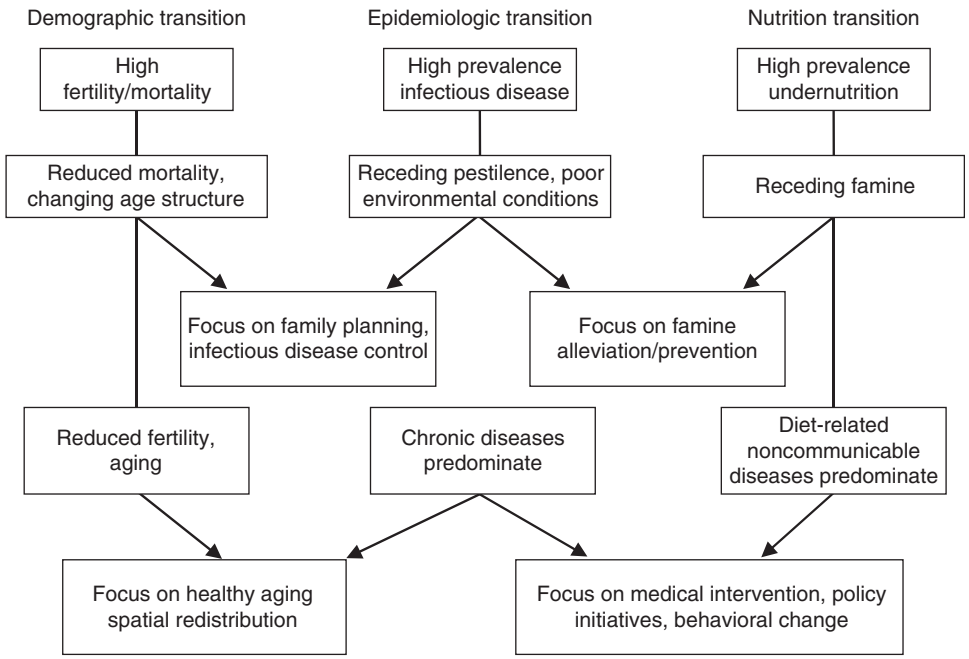


Fig. 1 Stages of health, nutritional, and demographic change. Reproduced from Popkin (2002).

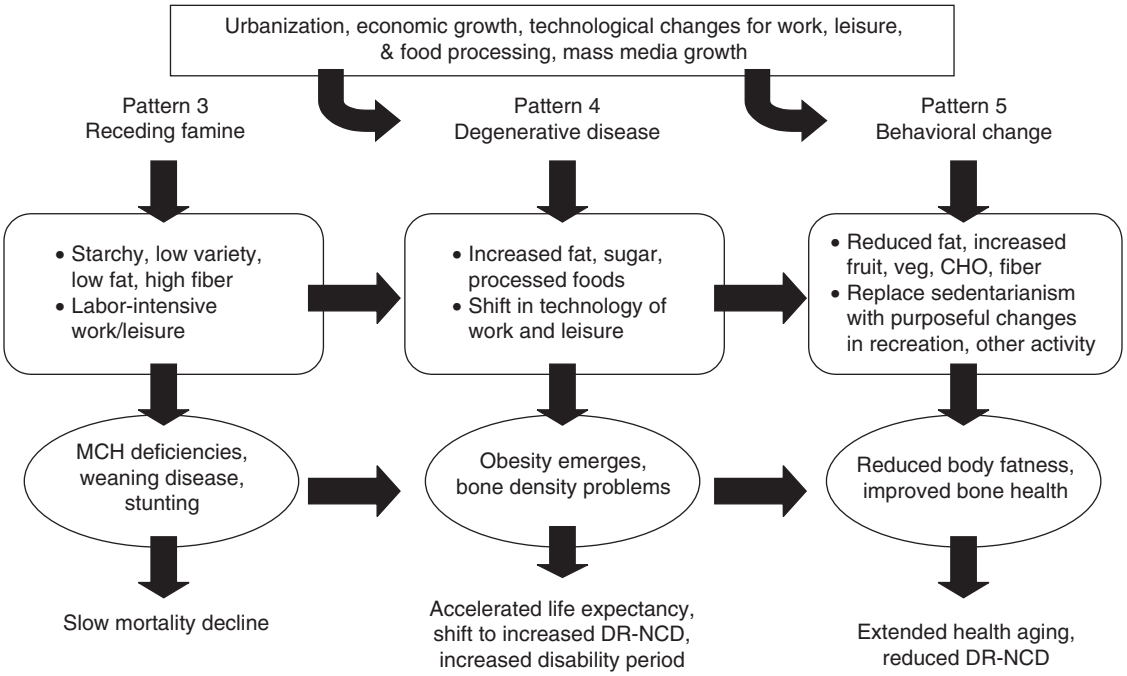


Fig. 2 Stages of nutrition transition. Reproduced from Popkin (2002).

Pattern 4: Nutrition-related noncommunicable diseases (NR-NCD): A diet high in total fat, cholesterol, sugar, and other refined carbohydrates, low in polyunsaturated fatty acids and fiber, and often accompanied by an increasingly sedentary life is characteristic of most high-income societies (and increasing proportions of the population in low-income societies), resulting in an increased prevalence of obesity and NR-NCD that characterize the final epidemiologic transition stage (Monteiro et al., 2004).

Pattern 5: Behavioral change: A new dietary pattern emerged in the 1990s and early 2000s that was associated with the desire to prevent or delay degenerative diseases and prolong health. Whether these dietary changes, instituted in some countries by consumers and in others also prodded by government policy, will create a large-scale transition in dietary structure and body composition remains to be seen.

Research on global dietary changes has focused primarily on patterns 3–5, particularly the rapid shift in much of the world's low- and middle-income countries from the stage of receding famine to NR-NCD as depicted in Fig. 2. The concern about this period is so great for many that the term "nutrition transition" is synonymous with this shift from pattern 3 to 4.

Rapid shifts in dietary and activity patterns and body composition

The pace of rapid shifts in diet and activity patterns from the period termed the receding famine pattern to one dominated by NR-NCD accelerated in LMICs between 2000 and 2015. The word "nutrition" rather than "diet" is used since the term NR-NCD incorporates the effects of diet, physical activity, and body composition rather than solely focusing on dietary patterns and their effects. Another element is that the rapid changes in urban populations are much greater than those experienced about a century ago in the West, yet another is the shift in occupation structure and the rapid introduction of modern mass media, social media, and ultra-processed foods. Underlying such changes is a general concern for rapid globalization as the root cause.

Clearly, there are quantitative and qualitative dimensions to these changes. On the one hand, changes toward a high-density diet, reduced complex carbohydrates, increased added sugar and other caloric sweeteners along with ultra-processed foods, and inactivity may be proceeding faster than in the past. The shift from labor-intensive occupations and leisure activities toward more capital-intensive, less strenuous work and leisure occurred very rapidly in many middle-income countries. On the other hand, qualitative dimensions related to multidimensional aspects of the diet, activity, body composition, and disease shifts may exist. The social and economic stresses that people face and feel as these changes occur may also be included.

Scholars often note that the pace and complexity of life, reflected in all aspects of work and play, are increasing exponentially. There are also unanticipated developments, new technologies, and the impact of a very modern, high-powered communications system. It is this sense of rapid change that makes it so important to understand what is happening and anticipate the way in which changes in the patterns of diet, activity, and body composition are occurring. Although the widespread influence of modern communications, technology, and economic systems related to "globalization" have been a dominant theme of the past few decades, there are many unique issues that have led to a rapid increase in globalization and its impact on nutritional status.

Stating that globalization is the cause of the nutrition transition results in a focus on broad and vaguely measured sets of forces that ignores the need to be focused and specific, which would allow us to develop potentially viable policy options. These processes certainly have been expanded, as indicated by enhanced free trade, a push toward reduction of trade barriers in the developing world, and the increasing penetration of international corporations into the commerce of each country (measured by share of gross national product (GNP) or manufacturing). Similarly, other economic issues related to enhanced value conferred to market forces and international capital markets are important. Equally, the increasing access to Western media and social media, the removal of communication barriers enhanced by the internet, cable television (TV), mobile telephone systems, etc., are important. The accelerated introduction of Western technology into manufacturing and the basic sectors of agriculture, mining, and services is also a key element.

To best illustrate the changes that have occurred in LMICs, consider the life of a person living in an urban shantytown compared to that in a rural village in China (Bell et al., 2001). During the 1970s, there were still major concerns about having an adequate food supply, there was no television, limited bus and mass transportation, minimal international food trade, minimal processed food available, and most rural and urban occupations were very labor intensive. Today, work and life activities have changed as small gas-powered tractors are available, modern industrial techniques are multiplying, offices are automated, soft drinks and many processed foods are abundant, televisions are in approximately 89% of households (at least one-fifth of which are linked to Hong Kong Star and Western advertising and programming), younger children do not ride bicycles, and mass transit is being heavily used. Considering that such changes are also occurring in much of Asia, North Africa, the Middle East, Latin America, and many areas (particularly cities) in sub-Saharan Africa, it is evident that the shift from a subsistence economy to a modern, industrialized one occurred in a span of 10–30 years, whereas in Europe and other industrialized high-income societies, this occurred more slowly, over many decades or centuries.

The elements of the nutrition transition known to be negatively associated with NR-NCD are obesity, poor dietary changes (i.e., shifts toward a greater intake of fat, added caloric sweeteners in food, and ultra-processed food, reduced fruit and vegetable intake, reduced fiber intake, greater energy density, and greater saturated fat intake), and reduced physical activity in work and leisure. The causes of these elements of the nutrition transition are not as well understood as their trends. In fact, few studies have attempted to research the causes of such changes, and only a few data sets allow such crucial policy analyses to be undertaken.

Obesity trends

The nutrition transition has contributed to a parallel increase in the worldwide prevalence of obesity which has nearly tripled in the past 40 years (Jaacks et al., 2019). Obesity is a chronic condition in which sustained positive energy balance results in a person having excess adipose tissue, a factor that is known to contribute to NR-NCD. Recent trends in the prevalence of obesity for adults and children segregated by income are presented in **Figs. 3–5**. Currently, a lower proportion of LMICs populations are obese compared to populations in HICs; however, obesity appears to be increasing more rapidly in LMICs. Over the last 3 decades, the prevalence of obesity has increased significantly, particularly in countries that have historically been underweight. This shift is evident among women and children who, in most LMICs, have a prevalence of obesity that is higher than underweight. Gender differences in the prevalence of obesity exist across both LMICs and HICs such that women are more likely to be obese compared to men. Given that obesity in girls is generally comparable with boys, it appears that this disproportionate increase in obesity among women occurs later in life. These shifts are reflective of global changes in the prevalence of obesity and NR-NCD.

In 2018, it was estimated that 108 million children and 604 million adults were obese, defined as a body mass index [BMI, weight (kg)/height $(m)^2$] > 30. It has been reported that adults with a high BMI accounted for nearly 4 million global deaths and 40% of these deaths were among adults with a BMI > 30. In terms of specific diseases, a high BMI accounted for 2.7 million deaths and 66 million disability-adjusted life-years lost (DALYs) from cardiovascular diseases. For type 2 diabetes, a high BMI was associated with 0.6 million deaths and 30 million DALYs. These estimates are the culmination of an ongoing global epidemic of excess fat such that from 1990 to 2015, deaths attributed to a high BMI increased nearly 28% and DALYs increased 36%. Interestingly, since the nutrition transition has not occurred independent of other epidemiological transitions, the age-standardized death rates and DALYs did not change as the death rates across all ages decreased.

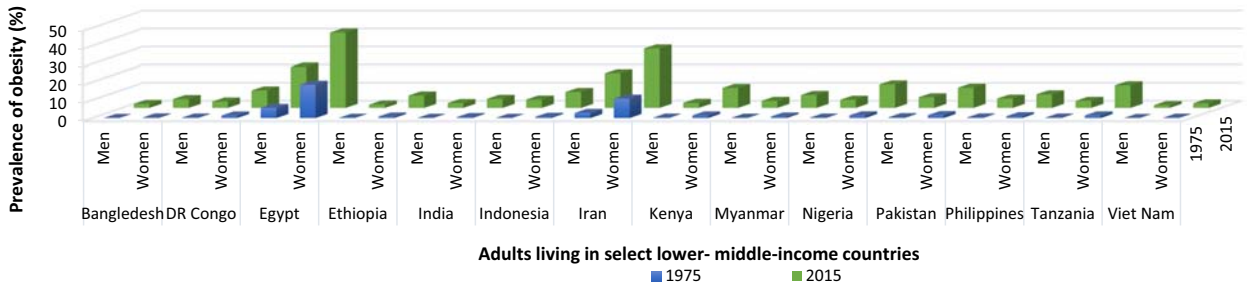


Fig. 3A Obesity trends among adults living in lower- and middle-income countries. Adapted from data courtesy of Jaacks et al. (2019).

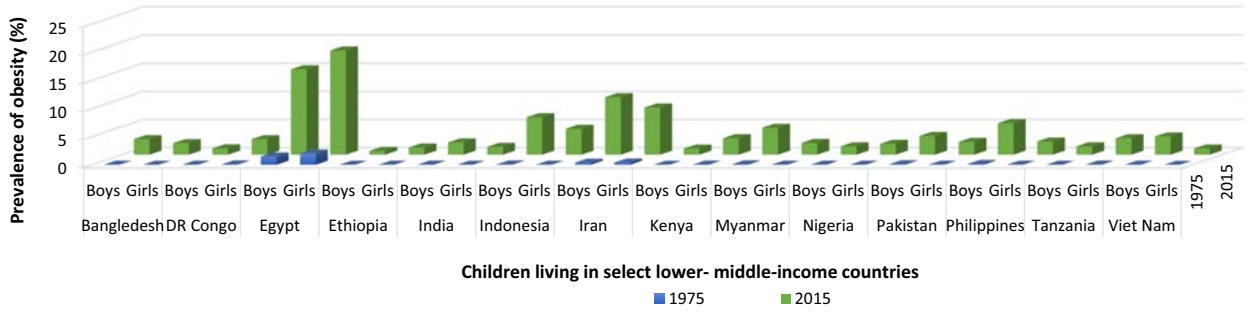


Fig. 3B Obesity trends among children living in lower- and middle-income countries. Adapted from data courtesy of Jaacks et al. (2019).

Dietary changes: shift in the overall structure over time

While high-quality data on total energy intake in many countries are limited, there are reasonable data to examine shifts in the structure of the diet. Briefly, there has been a nearly global increase in the consumption of fat, animal products, and artificial sweeteners accompanied by a decrease in legumes and other high-fiber grains (Popkin, 2015). Together, these changes may be responsible for what is referred to as an obesogenic diet, one that is associated with NR-NCD.

Since the 1960s, there has been a major increase (10–13%) in the consumption of vegetable fats by almost all countries in the world. In fact, since 1985, the average intake of vegetable oils tripled in most countries of the world. At the same time, the consumption of fat from animal sources has decreased except for low-income countries. These decreases, combined with the increase in vegetable fat intake for all income countries, resulted in an overall decrease in fat intake for middle-income countries of approximately 3% but an increase in approximately 4% or 5% for low- and high-income countries. Changes in the relationship between GNP and the consumption of fat are presented in **Fig. 6**.

The issue of animal-source foods is complicated as there are millions of people in the world with inadequate protein intake, a major contributor to poor growth and undernutrition in the most vulnerable communities of the globe. However, most increases in the production and consumption of animal-source products occurs in LMICs. Thus, it is a delicate balance to promote adequate protein intake while guarding against excess intake that can promote higher fat consumption and NR-NCD.

There is also an equally large and important shift in the proportion of energy from added caloric sweeteners in the diets of lower-income countries. In fact, an additional 100–200 kcal per day was available for daily consumption from added caloric sweeteners in the diet in 2000 compared with 1962 in the developing world. In the US, this added caloric sweetener increase derives mainly from soft drinks and fruit drinks, but in many other countries the source of this increase is other foods, even basic processed foods that have sweeteners added to them. High-fructose corn syrup (HFCS) is used as the sweetener of choice in some cases by food companies. Fructose, be it from HFCS, sugar, or other caloric sweeteners, appears to have unique adverse cardiometabolic attributes.

A relatively recent addition to the list of foods that promotes unhealthful diets is ultra-processed foods and beverages, UPF and UPB, respectively (Elizabeth et al., 2020). Based on the availability of global data on sales and consumption, the per capita sales of UPF increased 3 and 11 times faster in HICs compared to LMICs (Lane et al., 2021). Of particular concern is that a similar increase was reported for UPB given the positive relationship between UPB consumption and obesity. When the growth of sales of UPF and UPB were considered, it was found that LMICs experienced an 8% growth in sales of these products while HICs experienced only a 2% growth, highlighting the shift in marketing and sales of these foods in less wealthy countries. These increases in sales of UPF and UPB may be partly attributed to consumer demand or preferences, but one should not discount the impact of marketing and distribution of foods. Indeed, regarding market integration of fundamental food products, only four companies account for approximately 80% of the global grain trade. Moreover, given the high sugar content of these foods, it is important to recognize that global raw sugar production has increased from 50 million tons in 1961 to almost 160 million tons in 2014.

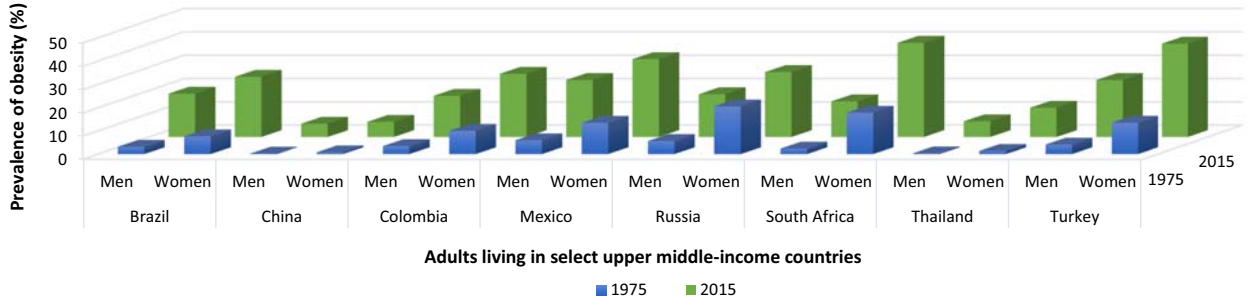


Fig. 4A Obesity trends among adults living in upper middle-income countries. Adapted from data courtesy of Jaacks et al. (2019).

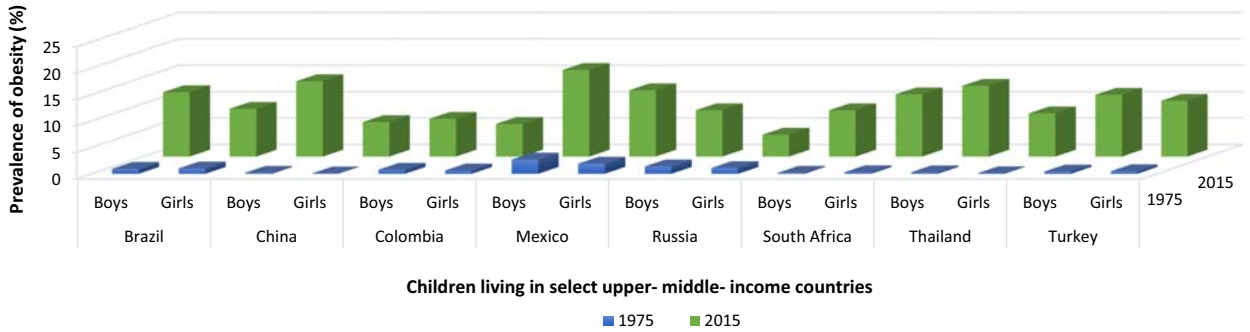


Fig. 4B Obesity trends among children living in upper middle-income countries. Adapted from data courtesy of Jaacks et al., 2019.

Aside from changes in the consumption of foods that are associated with chronic diseases, there has been a decrease in the consumption of foods that may prevent chronic diseases. In particular, the consumption of high-fiber grains and vegetables has declined steadily in most developed and many LMICs since the 1960s. Recent advancements to national food guidelines, such as that developed by the Brazilian Ministry of Health, are emphasizing the intake of "whole" foods with minimal consumption of processed foods, an effort that could result in an increase in the consumption of higher fiber foods.

Rapid social change is important: urbanization, rapid demographic change, and other behavioral changes are occurring simultaneously

Compared to rural areas, dietary patterns in urban areas of LMIC have shifted dramatically. While it is beyond the scope of this article to explore the complex issues related to the type of urban change that has occurred, critical sociodemographic issues include the following:

- Rapid reductions in fertility have enhanced the shift in the age distribution.
- Urbanization continues unabated in Asia and Africa. More poor will reside in urban than rural areas in future decades and upwards of 2/3 of the world population is expected to live in an urban setting by 2040.
- Economic changes, particularly increased income and income inequality, appear to define changes in many regions of the developing world.
- Globalization of mass and social media is occurring at an earlier stage of economic development than occurred in higher income countries in the past.

Urbanization

Demographic predictions show that over two-thirds of the world population will live in urban settings by 2050. The structure of dietary patterns has shifted markedly as populations have urbanized. This relationship will, by itself, alter the structure of diets significantly at the national level as urbanization continues and as the proportion of the population in urban areas grows. There are a number of factors that contribute to this situation including an increase in food marketing, expansion of multi-national super- and hyper-markets, and access to ultra-processed foods.

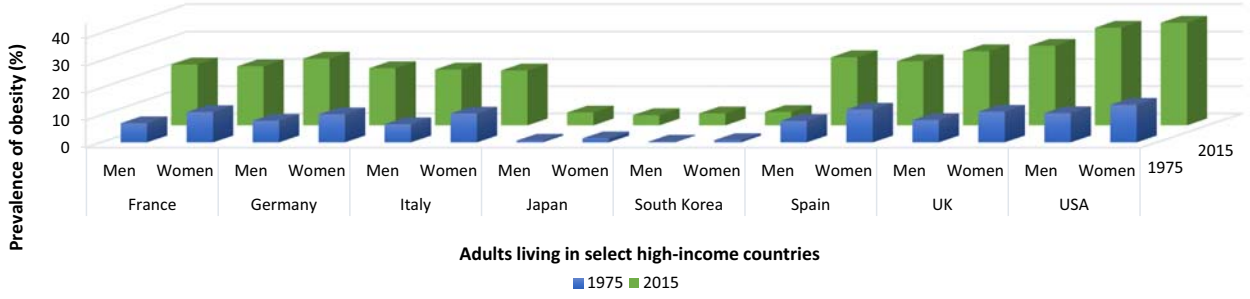


Fig. 5A Obesity trends among adults living in high-income countries. Adapted from data courtesy of Jaacks et al. (2019).

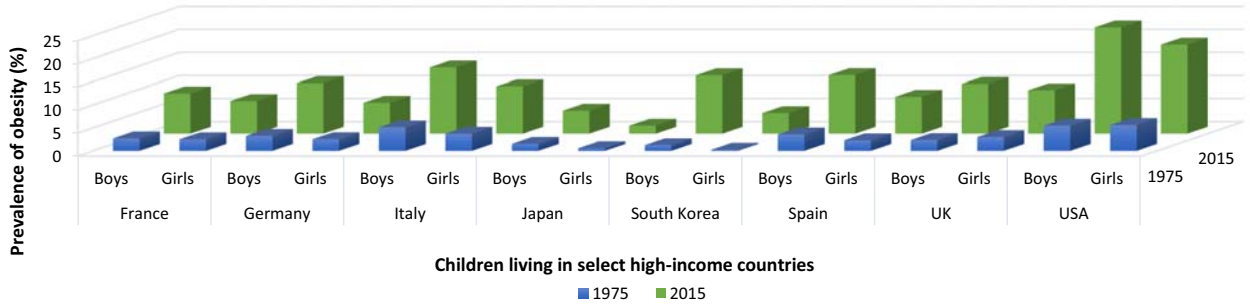


Fig. 5B Obesity trends among children living in upper middle-income countries. Adapted from data courtesy of Jaacks et al. (2019).

Structural shifts in income–diet relationships are occurring

Changes in dietary behavior can be caused either by shifts in the composition of society regarding the plurality of the educated, rich, or urban residents or by changes in the actual behavior of those with specific characteristics. The latter shifts may include changes in consumption behavior such that for the same level of education or income, a person would buy different amounts or types of commodities at different points in time. Research conducted in China shows that there have been profound behavioral shifts of this type during the past decade (i.e., for each extra dollar of income, additional high-fat foods are purchased vs. what would have been purchased in previous years for the equivalent extra dollar). Economists speak of this effect as one that shows how the decision-making demand pattern for food has changed; thus, for the same income level, the patterns of demand have changed significantly from earlier periods. This rapid increase in access and exposure to mass media and the internet may very well have created this situation.

Mass media and internet use

Access to modern mass media and internet use has increased rapidly, particularly from 2000 to 2010. For example, 88.5% of Chinese households had televisions in 1997. While access to mass media and the internet have increased, it is important to note many media and internet providers rely heavily on US and British programming which provides modern TV advertising. Like television, the explosion of access to the internet via computers and cellular devices increased the access to a wide variety of messaging on diet, body satisfaction, new concepts of the "ideal" body, as well as misinformation on health and diet. The impact of these messages ranges from less than valid diet and health advice from unqualified websites to body shaming via social media platforms. Still, the complete picture of how mass media and the internet influence diet and activity is unclear, but one can speculate that for every "positive" message about diet and health, people encounter "negative" messages. Internet use and marketing of unhealthy foods is particularly worrisome among children worldwide.

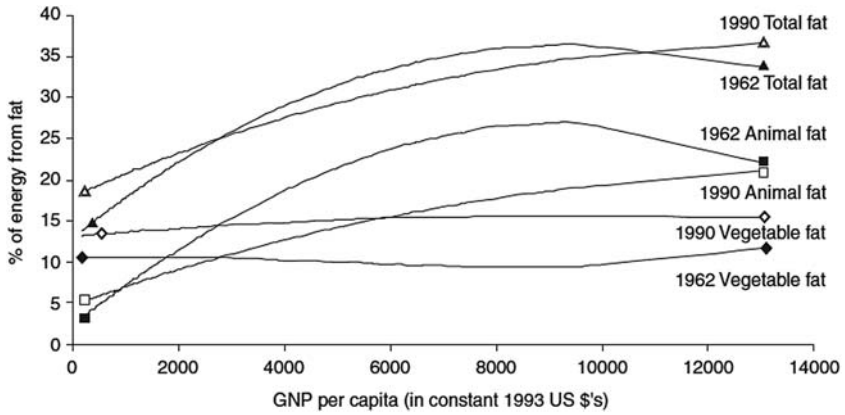


Fig. 6 Relationship between the percentage of energy from fat and GNP per capita, 1962 and 1990. Reproduced from Guo et al. (2000).

Health effects: how does biology interact with these rapid social changes?

The question of how NR-NCD vary by country or social group raises the important question of determining the interaction between biology and society. For example, body composition and other social factors (i.e., income, class, and education) affect the susceptibility to NR-NCD. As well, previous disease patterns (i.e., famine, chronic food shortage, or the presence of malaria or other tropical diseases) may have led to disease patterns that predisposed the entire population to specific NR-NCD. The latter example is generally attributed to a phenomenon first reported by Dr. David Barker and is now referred to as the "developmental origins of health and disease" or DOHaD (Barker, 2001).

There is a growing body of research that shows that the international standards to classify someone as overweight or obese are not appropriate for many large subpopulations in the world. For example, the risk of hypertension is higher for Chinese men and women at lower BMIs (23–25 kg/m^2) compared to adults from the Philippines and US (25–30 kg/m^2) (Bell et al., 2002). Ethnic differences in the strength of the association between BMI and disease outcomes warrant further consideration. In response, a working group created by the World Health Organization and the International Obesity Task Force proposed lower BMI cutoffs for Asians of 23 kg/m^2 for overweight and 25 kg/m^2 for obesity.

For LMICs in particular, the highest genetic susceptibility for type 2 diabetes (T2D) is for Pacific Islanders, American-Indians, Mexican-Americans and other Hispanics, and Asian Indians. Those with modest genetic susceptibility include Africans, Japanese, and Chinese. The age of onset (usually after 50 years of age) of T2D is much lower for these susceptible populations, and it appears that the prevalence is higher for any level obesity and waist-to-hip ratio.

It is not clear how much of this difference between subpopulations regarding BMI–T2D or other BMI–morbidity relationships is a function of differences in body composition, metabolic or genetic factors, or social causes. While we may regard biological factors as fundamental to understanding these relationships, it is necessary to consider that several social factors, such as income, education, stress, and access to healthcare also play major roles in chronic disease risk. Thus, the confluence of biological and social risk factors for NR-NCD complicates the ability to discern exact causal pathways.

Several studies have provided an increasing body of evidence to support DOHaD, which attributes adverse nutritional and environmental exposures during early life to an increased risk of NR-NCD. This theory postulates that malnutrition during the pre- and post-natal periods causes an adaptative metabolic response which favors fat deposition to ensure both immediate viability and survival if a similar environment is encountered later in life. Indeed, several studies have reported that children and adults who are growth restricted in childhood (a condition generally attributed to insufficient energy and protein consumption during rapid periods of growth) experience changes in their metabolism such that fat metabolism is lower compared to normal height peers (Hoffman et al., 2000). However, exposure to a food environment opposite of what was experienced during this period in early life (i.e., increased exposure to adequate, energy-rich foods), creates a "mis-match" which promotes the development of NR-NCD long-term. Simply, when these individuals face a food environment rich in processed foods and high-fat foods, any excess intake may be deposited as fat and contribute to the development of obesity and other NR-NCD. Individuals living in LMICs, where food environments are shifting rapidly, are more likely to be exposed to both nutritional insufficiency and excess throughout the course of their lives. Notably, an increase in NR-NCD has been found among children in LMICs who experienced rapid catch-up growth after being born small-for-gestational-age (SGA). As dietary patterns continue to shift toward increased consumption of calorically dense, nutrient-poor foods, early life will pose a critical period for the development NR-NCD.

Epidemic of CVD and type 2 diabetes

Evidence from many LMICs shows that NR-NCD results in premature morbidity and mortality affecting a large proportion of economically productive people; essentially, a preventable loss of precious human capital. Nearly 80% of deaths in LMICs can be attributed to NR-NCD. The current high burden of NR-NCD is contributed to 58.8% of global mortality and 43% of the global burden of disease, measured as DALYs. The contribution of LMICs to this burden is large where approximately 77% of the total mortality and 85% of the total burden of disease attributable to NR-NCD arise from these countries.

The burden of cardiovascular disease varies by global region and the nearly 21% lower mortality rate of CVD is attributed to a 42% decrease in high income countries compared to smaller decreases and some increases in lower income countries. In fact, The Global Burden of Disease Study projected those 6.4 million deaths would occur due to CVD in the developing countries in 2020, in the age group of 30–69 years (Roth et al., 2020). For other parameters, such as fasting glucose, hypertension, and BMI, similar patterns emerge in which the global changes do not necessarily reflect a consistent pattern when analyzed by GDP. For example, there has been a 14% global increase in fasting glucose and an 18% global decrease in total cholesterol, but these decreases are primarily seen in higher income countries (−25%) while lower-income countries experience anywhere from a 6% decrease in North Africa and the Middle East with some countries reporting up to 162% increase (e.g., Bangladesh). A similar pattern emerges for high BMI in which the global prevalence has increased 28% but reflects some decreases in HICs and up to 650% increase in Asia.

Reports from several global health agencies focus on the important role of obesity in CVD and cancer deaths in the developing world. There are major differences in the profiles of the CVD epidemic across the developing world. For instance, hypertension and stroke are more likely to emerge in east Asia, whereas diabetes occurs earlier in south Asia. As would be expected from the dietary and obesity data noted previously, CVD levels are far greater in urban areas of the developing world, but often the opposite is true in the developed higher income countries.

Social burden of changes in diet, body composition, and health

In HICs, adults with higher income follow a more healthful lifestyle, whereas low-income individuals do not, which is often related to food insecurity. For example, higher income Americans consume a more healthful dietary pattern, exercise more, and smoke less; similar patterns can be found in other high-income countries. In contrast, the prevailing opinion has been that the opposite is found in the developing world, namely that the poor are less likely to have a heavy burden of NR-NCD compared with the rich. This is changing rapidly as obesity has declined among the higher educated and increased among the lower educated in southeastern Brazil. In fact, not only are less healthful dietary patterns consumed by higher income Chinese, but also other dimensions of lifestyle (inactivity, smoking, and drinking) are poorer among the higher SES Chinese (Jones-Smith et al., 2011). In other research, scholars of China have shown a rapid shift in food consumption patterns among different income groups in China that seems to indicate a shift in the burden of poor diets toward the poor in China. It has been shown that for countries with a GNP per capita of more than $2500, the likelihood is very high that there will be more obesity among the lower SES groups compared with higher SES groups.

Summary

Maintaining a healthy lifestyle includes eating a healthful diet and being physically active in a manner that promotes health and prevents chronic diseases. However, as described in this article, there are several factors that limit such lifestyles, such as poverty and food insecurity and dietary change toward a less healthful diet is universal. It remains to be seen how the global pandemic of COVD-19 affected these patterns as near-universal quarantines and disruptions to the food supply chains created a host of issues that resulted in increased food insecurity, especially in the countries with the least number of resources to respond to the pandemic. In particular, a rapid change is being seen in the poorest areas of the world. The challenge is to learn how to continue to improve the palatability and quality of our diet while doing so in a more healthful manner.

See Also: Diet and oral health; Dietary intake measurement: Methodology; Famine: Causes, consequences and responses; Lipids (fats and oils)

References

Barker, D.J.P., 2001. Fetal Origins of Cardiovascular and Lung Disease. Marcel Dekker, New York.
Bell, C., Ge, K., Popkin, B.M., 2001. Weight gain and its predictors in Chinese adults. Int. J. Obes. 25, 1079–1086.
Bell, A.C., Adair, L.S., Popkin, B.M., 2002. Ethnic differences in the association between body mass index and hypertension. Am. J. Epidemiol. 155, 346–353.
Elizabeth, L., Machado, P., Zinöcker, M., Baker, P., Lawrence, M., June 30, 2020. Ultra-processed foods and health outcomes: a narrative review. Nutrients 12 (7), 1955.
Guo, X., Morz, T.A., Popkin, B.M., Zhai, F., 2000. Structural changes in the impact of income on food consumption in China, 1989–93. Econ. Dev. Cult. Change 48, 737–760.
Hoffman, D.J., Sawaya, A.L., Coward, W.A., Wright, A., Martins, P.A., de Nascimento, C., Tucker, K.L., Roberts, S.B., 2000. Energy expenditure of stunted and nonstunted boys and girls living in the shantytowns of São Paulo, Brazil. Am. J. Clin. Nutr. 72, 1025–1031.
Jaacks, L.M., Vandevijvere, S., Pan, A., McGowan, C.J., Wallace, C., Imamura, F., Mozaffarian, D., Swinburn, B., Ezzati, M., March 2019. The obesity transition: stages of the global epidemic. Lancet Diabetes Endocrinol. 7 (3), 231–240.
Jones-Smith, J., Gordon-Larsen, P., Siddiqi, A., Popkin, B.M., 2011. Cross-national comparisons of time trends in overweight inequality by socioeconomic status among women using repeated cross-sectional surveys from 37 developing countries (1989–2007). Am. J. Epidemiol. 173 (6), 667–675.
Lane, M.M., Davis, J.A., Beattie, S., Gómez-Donoso, C., Loughman, A., O'Neil, A., Jacka, F., Berk, M., Page, R., Marx, W., Rocks, T., March 2021. Ultraprocessed food and chronic noncommunicable diseases: a systematic review and meta-analysis of 43 observational studies. Obes. Rev. 22 (3), e13146.
Monteiro, C.A., Moura, E.C., Conde, W.L., Popkin, B.M., 2004. Socioeconomic status and obesity in adult populations of developing countries: a review. Bull. World Health Organ. 82, 940–946.
Popkin, B.M., 2002. The shift in stages of the nutrition transition in the developing world differs from past experiences. Publ. Health Nutr. 5 (1A), 205–214.
Popkin, B.M., September 2015. Nutrition transition and the global diabetes epidemic. Curr. Diabetes Rep. 15 (9), 64.
Roth, G.A., Mensah, G.A., Johnson, C.O., Addolorato, G., Ammirati, E., Baddour, L.M., et al., December 22, 2020. The GBD-NHLBI-JACC global burden of cardiovascular diseases writing group. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 Study. J. Am. Coll. Cardiol. 76 (25), 2982–3021.

Further reading

Eaton, S.B., Konner, M., 1985. Paleolithic nutrition: a consideration on its nature and current implications. N. Engl. J. Med. 312, 283–289.
Hoffman, D.J., Reynolds, R.M., Hardy, D.B., 2017. The developmental origins of health and disease: current knowledge and potential mechanisms. Nutr. Rev. 75 (12), 951–970.
Yach, D., Stuckler, D., Brownell, K.D., 2006. Epidemiologic and economic consequences of the global epidemics of obesity and diabetes. Nat. Med. 12, 62–66.
Zimmet, P.Z., McCarty, D.J., de Courten, M.P., 1997. The global epidemiology of non-insulin-dependent diabetes mellitus and the metabolic syndrome. J. Diabetes Complicat. 11 (2), 60–68.

Nutritional surveillance: Developed countries

E.W. Harris, Beltsville Human Nutrition Research Center, ARS, USDA, Beltsville, MD, USA

Published by Elsevier Ltd.
This is a reproduction of E.W. Harris, Nutritional Surveillance: Developed Countries, Editor(s): Benjamin Caballero, Encyclopedia of Human Nutrition (Third Edition), Academic Press, 2013, Pages 278–288, ISBN 9780123848857, https://doi.org/10.1016/B978-0-12-375083-9.00205-1.

Uses of Nutritional Surveillance Data **444**
Nutritional Surveillance Systems **445**
Dietary Data Collection Methods 445
Food Supply Data 445
Food Consumption by Household 446
Food Consumption by Individuals 447
Examples of Nutritional Surveillance Activities 448
Food Composition Databases 452
Emerging Nutritional and Health Issues **452**
Further Reading **453**

Glossary

AMPM Automated Multiple Pass Method, a computerized method for collecting interview-administered 24-h dietary recalls in person and by telephone. Developed by the USDA Food Surveys Research Group, the research-based approach employs five steps that are designed to enhance data collection and reduce respondent burden.

DPAS The Global Strategy on Diet, Physical Activity was adopted by WHO member states in 2004. It is a prevention-based strategy aimed at significantly reducing the prevalence of NCDs and their common risk factors, mainly unhealthy diet and physical activity.

EuroFIR The European Food Information Resource Consortium is a partnership between several universities and 25 countries to develop and integrate a comprehensive cohort and validated network of food composition databanks in Europe.

INFOODS Established in 1984 to stimulate and coordinate efforts to improve the quality and availability of analytical food data worldwide.

NCDs Noncommunicable diseases such as diabetes or cardiovascular disease.

Nutritional surveillance A system established to continuously monitor the dietary intake and nutritional status of a population or selected population groups using a variety of data collection methods whose ultimate goal is to lead to policy formulation and action planning.

Nutritional surveillance is defined as a system established to continuously monitor the dietary intake and nutritional status of a population or selected population groups using a variety of data-collection methods whose ultimate goal is to lead to policy formulation and action planning. Continuous monitoring varies among countries such that annual, bi-annual, or other data-collection timeframes are determined by the funding and commitment of national governments. Increasingly, countries have come to recognize the need to collect dietary information in a systematic manner to make science-based policy decisions related to diet and health.

This article describes nutritional surveillance systems in developed countries. Emerging nutritional and health issues informed by these systems are reviewed. Examples of surveillance systems in Australia, Canada, Europe, Japan, New Zealand, South Korea, and the United States are described.

Uses of Nutritional Surveillance Data

Most notably data generated from these systems are categorized broadly as (1) nutritional monitoring and surveillance, (2) research relating diet to health, and (3) informing the development of or enhancing existing nutritional policies and programs. **Figure 1** illustrates the interrelationships between data in a nutritional surveillance system. These data are used by government agencies, the private sector, the academic community, and the public. Government agencies use the data for nutritional policy, research, and programs. The private sector uses the data for marketing, product development, food labeling, and food safety compliance. Academics use the data to conduct research, develop programs, and train students.

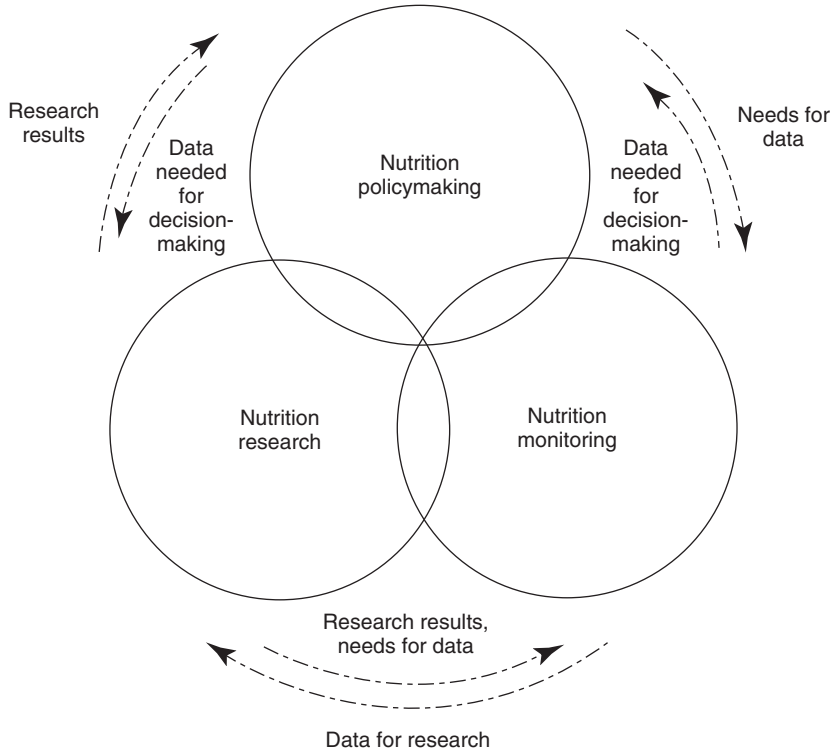


Figure 1 Relationships among nutrition policymaking, nutrition research, and nutritional monitoring. Reproduced from US Department of Health and Human Services/US Department of Agriculture (1993) Ten-year comprehensive plan for the National Nutrition Monitoring and Related Research Program. *Federal Register* 58: 32752–32806.

Data uses for public policy enable informed decision-making regarding cost-effective programs, policy, and regulations. For example, data can be used to:

- assess the status of national populations and identify high risk or vulnerable groups;
- identify problems that inform the development of new policies and programs;
- formulate food enrichment, fortification, and food labeling policies and regulations;
- determine risk and inform regulatory decisions;
- develop and update socio–culturally appropriate foods and food guidance;
- evaluate the effectiveness of existing policies and programs and;
- conduct epidemiological, socioeconomic, behavioral, and marketing research.

Nutritional Surveillance Systems

The purpose of a nutritional surveillance system is to depict the relationship between food and health. Key information collected through the system includes data on nutritional and health status, food consumption patterns (household and individual levels), nutrient intakes, food supply, and food composition. There are many other influential factors on these data such as socio–demographic; culture; food knowledge, attitudes, and behavior; and food security. Although **Figure 2** highlights the traditionally accepted relationship between food and health, **Figure 3** begins to depict a new paradigm where globalization is included and may eventually include the effects of climate change.

Dietary Data Collection Methods

Food Supply Data

Food supply data reflect the type and amount of food available for consumption in a country. Food balance sheets are the primary tool used for assessing available food at the country level. This indirect measurement method estimates the amount of food consumed by a country's population at a certain point in time. Food balance sheets do not measure actual food consumption, but rather food disappearance. Calculations are based on using beginning and ending inventories, with the difference as the amount of food consumed. The beginning inventory includes measurements of food production, imports and exports, adjustments for

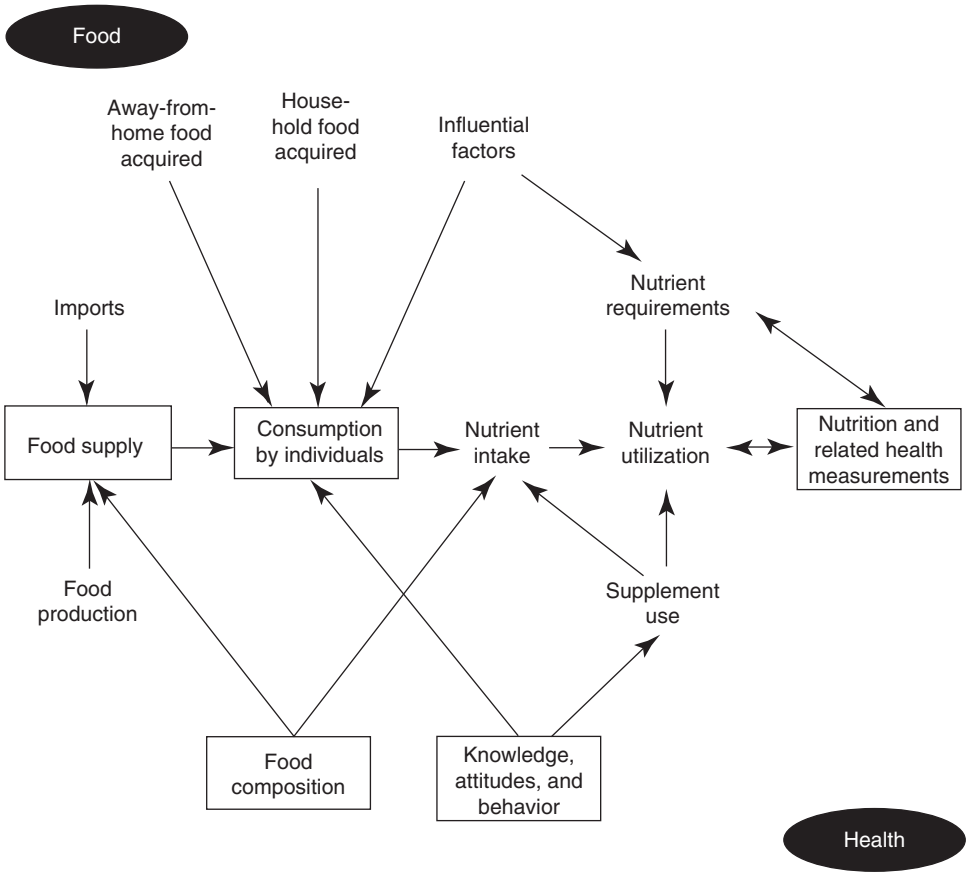
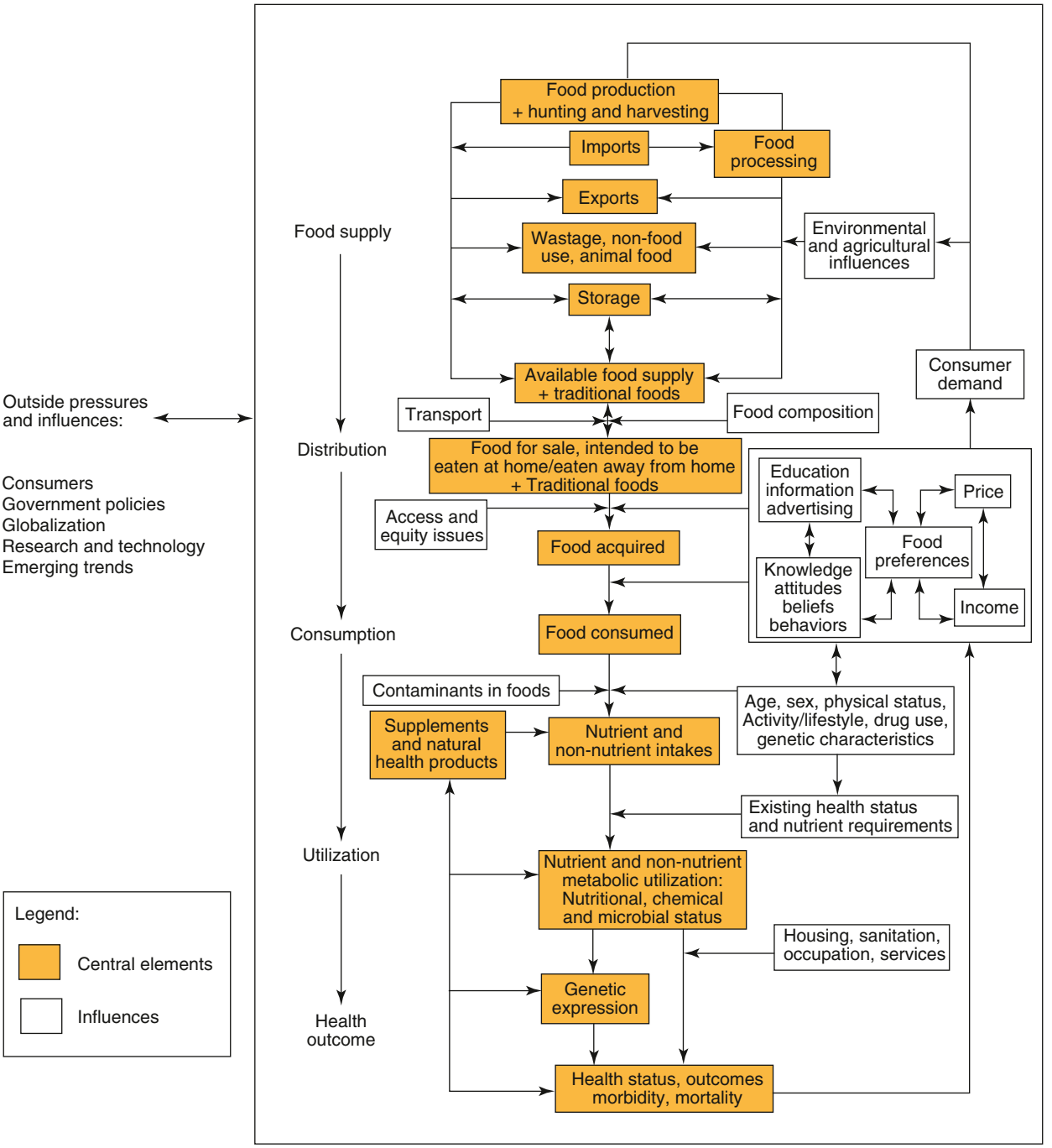


Figure 2 A conceptual model of the relationships of food to health. Reproduced from US Department of Health and Human Services/US Department of Agriculture (1993) Ten-year comprehensive plan for the National Nutrition Monitoring and Related Research Program. *Federal Register* 58: 32752–32806.

nonhuman food consumption and estimates of waste. Mean per capita annual food consumption is determined by dividing total disappearance of food by a country's population. Each food commodity is then multiplied by the appropriate nutrient values and the results expressed as either kilograms per year or grams per day of individual food commodities and nutrient availability per person. Major limitations to this method include (1) the accuracy of data may be questionable, (2) data only represent food available for consumption, (3) does not account for waste, and (4) does not indicate how food was distributed among individuals. However, food balance sheets are useful for comparing available food supply within and between countries, monitoring trends, and forecasting food consumption patterns over time. They are also useful in formulating agricultural policies related to food production and consumption. For example, **Table 1** highlights the use of food balance sheets data and shows per person food consumption in kilocalories per day worldwide and by region of the world from 1990–1995 to 2003–2005.

Food Consumption by Household

Several methods of assessing food consumption by households have been devised, which consider the per capita food consumption of the household. These methods account for all food and beverages on hand in the home at the beginning of the survey period, all items purchased or grown during the survey period, and all that remains at the end of the survey. They vary by the level of respondent burden and extent of recall or recordation expected. Four methods are primarily used: the food account, list-recall, inventory, and food record methods. For the food account, the head of household records daily the type and amount of food entering the household. The list-recall method includes an interview where the head of household recalls the foods used by the household on an 'as purchased' basis. The inventory and food record methods are probably the most burdensome to the respondent because they require a daily record of food acquired and changes to the food inventory, with detailed weighing and measuring of food. None of these methods provide data on actual food intake by individuals within the household or food consumed away from home. However, data from household food consumption surveys indicate the kinds, quantities, money value, and nutrient content of food used, which in turn, provides valuable data for determining the effects of income, household size, and other socioeconomic factors on total food and food group consumption. These data demonstrate how diets at the household level conform to nutritional and household budgeting constraints.



Adapted from the conceptual framework of the Australian food and nutrition system
Ref: Ian H. Lester. AUSTRALIA'S Food and Nutrition.
Australia Government Publishing Service, Canberra. 1994

Figure 3 Conceptual model of the Canadian food and nutritional system adapted from the Australian food and nutritional system conceptual framework, www.hc-sc.gc.ca/fn-an/surveill/conceptual_model-eng.php. Reproduced from US Department of Health and Human Services/US Department of Agriculture (1993) Ten-year comprehensive plan for the National Nutrition Monitoring and Related Research Program. *Federal Register* 58: 32752–32806.

Food Consumption by Individuals

A variety of methods are available to assess the food intake by individuals: diet history, 24-hour dietary recall, food record, and food frequency questionnaire. All of these methods provide advantages and limitations, and their applicability depends on the purpose and design of the survey. Diet histories can provide detailed information, but they require respondents to make judgments about

Table 1 Per person food consumption by region

World/Region	*1990–1995 (kcal per capita per day)*	*1995–2000 (kcal per capita per day)*	*2003–2005 (kcal per capita per day)*
World	2669	2710	2778
Africa	2343	2335	2389
Asia	2539	2628	2646
Central and South America	2790	2825	2960
Caribbean	2269	2346	2565
Europe	3237	3206	3335
North America	3498	3616	3799
Oceania	3095	3048	3063
Developed Countries	3256	3269	3394
Developing Countries	2507	2569	2633

Source: Adapted with permission from Food and Agriculture Organization of the United Nations (2009) *Summary of World Food and Agricultural Statistics*, FAO Statistics Division.

their usual food habits and can be very burdensome. Dietary recalls are most appropriate for assessing the intakes of groups of people, and the current gold standard requires at least the collection of two days of intake to assess an individual's usual intake. Food records are thought to be the best method for assessing individual dietary intake, but they are time consuming and may impart bias if respondents modify their eating behavior due to the data collection method. Although food frequency questionnaires provide less detailed information, they are designed to provide usual intake data and are particularly useful for epidemiological studies of large population groups. For all these methods, advanced computerized systems provide standardized probes and multiple inquiry passes of intake have improved data collection, coding, and the estimation of nutrient intake and dietary patterns.

Examples of Nutritional Surveillance Activities

Most developed countries support a nutritional surveillance system and many conduct combined nutritional and health surveys. **Table 2** presents some of these surveys by region of the world and the method of dietary data collection selected by each country. For Asia, surveys conducted by Japan in 2007 and South Korea in 2005 represent their most recently reported data collection. The 2007 National Nutrition Survey in Japan collected dietary histories and behavior information, in addition to other assessments, from almost 9000 respondents aged one year and older. Whereas, the 2005 National Health and Nutrition Survey in South Korea included a 24-hour dietary recall, dietary behavior, food frequency questionnaire, anthropometric, biochemical, and clinical examinations.

With 24 European Union (EU) member countries plus Norway, a considerable amount of data is available through the European Nutrition and Health Report 2009. This report provides (1) an overview of available data on food and nutrient intake, (2) identifies major national and regional health and nutritional issues, (3) describes trends in food supply, (4) compares average daily individual food availability at the household level, (5) evaluates individual food consumption in adults and energy and nutrient intake in all ages, (6) describes data on diet-related health indicators and status, and (7) analyzes food and nutritional policies in EU countries. Individual national reports provide more detailed information on nutritional and health status in different participating countries, with much of that information obtained from national surveys. As evidenced by **Table 2** various dietary intake methodologies are used and anthropometrics appears to be collected by a majority of the surveys, with a few adding the collection of physical activity assessments. To date the 2009 report is the most current, comprehensive study of nutritional surveillance in a region of the world. It highlights both the considerable improvements in the quality of data collection and assessment methods employed and the continuing need for harmonization of databases and survey methods, which would allow more robust comparisons of data across countries.

The United States has one of the most comprehensive nutritional surveillance systems in the world. Since 1999, the periodic National Health and Nutrition Examination Survey (NHANES) became a continuous national nutritional surveillance system. A key component of the survey is the Automated Multiple Pass Method (AMPM), a computerized method for collecting interview-administered 24-hour dietary recalls in person and by telephone. The research-based, multiple pass approach employs five steps that are designed to enhance data collection and reduce respondent burden. Dietary data is collected from a nationally representative sample of 5000 individuals yearly and is released in 2-year cycles. The NHANES 2010–2011 is currently in the field. In Canada, the Canadian Community Health Survey, Cycle 2.2, Nutrition also used the AMPM to collect 24-hour recalls. Dietary data for greater than 35 000 individuals six months and older was released in Wave 1 (2005), Wave 2 (2006), and Wave 3 (2008).

In 2007, Australia conducted the National Children's Nutrition and Physical Activity Survey, which combined the collection of two days of dietary intake on children, aged 2–16 years old with anthropometric and physical activity assessments to strengthen the monitoring of childhood obesity in the country. A new National Nutrition and Physical Activity Survey Program, which will collect dietary data on 35 000 individuals plus an additional 15 000 individuals from indigenous population groups in the country is planned for 2011–2012. New Zealand's most recent survey, the Adult Nutrition Survey, was conducted in 2008–2009. Twenty-

Table 2 Nationwide food consumption surveys with individual-based dietary intake data

Country	*Year*	*Survey*	*Population (ages in years)*	*Sample Size*	*Dietary method*[a]	*Other information*[b]
Asia						
Japan	2007	National Nutrition Survey	1+	8 885	DH, DB	A, BC, CE, BP, PA Rcl
South Korea	1998	National Health and Nutrition Survey	1+	39 060 (10 876 dietary)	24-h recall, DB, FFQ	A, BC, CE
	2001	National Health and Nutrition Survey			24-h recall, DB, FFQ	A, BC, CE
	2005	National Health and Nutrition Survey	3+	6 436	24-h recall, DB, FFQ	A, BC, CE
Europe						
Austria	1991–1994	Austrian Study on Nutritional Status (ASNS)	6–18	2 173	7d FR	
	1993–1997	ASNS	19–65	2 065	24-h recall, DH	A
	1998–2002	ASNS	19–60	2 580	24-h recall	A
			55+	645		
	2007–2008	ASNS (compilation of 9 studies)	6+	7 416	24-h recall, 3d FR (subsample)	A
Belgium	1980–1985	Belgium Interuniversity Research on Nutrition and Health	25–74	10 971	1d FR (DH in subsample)	A, BC, MH
	2004	National Food Consumption Survey	15+	3 249	24-h recall	A
Denmark	1985	Dietary Habits in Denmark	15–80	2 442	DH	A
	1995	Danskernes Kosivaner	1–80	3 098	7d FR	
	2000–2006	National Survey of Dietary Habits and Physical Activity	4–75	4 120 (Yr 2000–2002)	7d FR	A
				3 247 (Yr 2003–2006)	7d FR	A
Finland	1992	Dietary Survey of Finnish Adults (FINDIET 1992)	25–64	1 861	3d FR	
	1997	FINDIET 1997	25–64	2 862	24-h recall	
			65–74	290		
	2003–2005	Type I Diabetes Prediction and Prevention (DIPP) Nutrition Study	1–6	2 535	DH	
	2005	Nationwide Survey of Breastfeeding and Complementary Feeding	<1	10 500	DH	
	2007	FINDIET	25–64, 65–74	2 039	48-h recall, 3d FR (subsample)	
France	1993–1994	Etudes Nationale des Consommations Alimentaires	2–85	1 500	7d FR	A
	1998–1999	Individual National Food Consumption Survey	3–14	1 018	7d FR	
			15+	1 985		
	2006–2007	French Nutrition and Health Survey (ENNS)	3+	1 675 (3–17)	3–24-h recall	A, BC, PA recall
				3 115 (18–74)		
Germany	1985–1989	National Nutrition Survey in Former West Germany	4–65+	24 632	7d FR KN, ATT, BH	A, BC
	1991–1992	National Nutrition Survey in East Germany	18–79	1 897	DH	
	1998	German Nutrition Survey	18–79	4 030	DH (4-week recall and FFQ)	
	2003–2006	German Health Interview and Examination Survey for Children and Adolescents (KIGGS)	0–17	17 641	3d FR, FFQ	A, BC, CE, PA
	2005–2007	German National Nutrition Survey II	14–80	19 329	24-h recall, DH, 4d FR	A
Ireland	1990	Irish National Nutrition Survey	10–65+	1 214	DH	A (self-reported)
	1997–1999	North–South Food Consumption Survey	18–64	1 379	7d FR, ATT	
	2003–2005	National Children's Food Survey	5–12	800		
	2005–2006	National Teens' Food Survey	13–17	441	7d FR	

(Continued)

Table 2 Nationwide food consumption surveys with individual-based dietary intake data—cont'd

Country	*Year*	*Survey*	*Population (ages in years)*	*Sample Size*	*Dietary method*[a]	*Other information*[b]
	2007	National Survey of Lifestyle, Attitudes and Nutrition (SLAN)	18+	10 364	FFQ	CE (subsample)
Italy	1994–1996	INN-CA 1994–96	0–94	2 734	7d FR	
The Netherlands	1987–1988	The Dutch National Food Consumption Survey (DNFCS-1)	1–85	5 898	2d FR	A (self-reported)
	1992	DNFCS-2	1–92	6 218	2d FR	A (self-reported)
	1997–1988	DNFCS-3	1–97	6 250	2d FR	A (self-reported)
	2003	DNFCS-young adults	19–30	750	2–1d FR	A
	2005–2006	DNFCS-young children	2–6	1 279	2–24-h recall	A
Norway	1993	National Dietary Survey	13	1 705	FFQ	A (self-reported)
			18	1 564	ATT, BH	
	1993–1994	National Dietary Survey among Adults NORKOST	16–79	3 144	FFQ, ATT, BH	A (self-reported)
	1997	National Dietary Survey among Adults NORKOST	16–79	2 672	FFQ, ATT	A (self-reported)
	1999	National Dietary Survey	6 and 12 months, 2 years	2 400	FFQ	
				2 010		
	2000–2001	National Dietary Survey	4	391	4d FR	A, PA
			9	810	4d FR	A, PA
			13	1 005	4d FR	A, PA
Portugal	1980	Portuguese Food Consumption Survey	1–65+	13 080	1d FR, 24-h recall, FFQ	A, BC, CE, MH
Sweden	1989	Household Food Survey, HULK	1–74	2 036	7d FR	A (self-reported)
	1997–1998	Riksmaten	18–74	1 215	7d FR	A (self-reported)
	2003	Children's National Food Survey	4	590	4d FR	A
			8	889	4d FR	A
			11	1 016	4d FR	A
United Kingdom	1986–1987	The Dietary and Nutritional Survey of British Adults	16–64	2 197	7d FR	A, BC, CE, BP
	1992–1993	Natl. Diet and Nutritional Survey (NDNS)	1.5–4.5	1 675	4d FR	
	1994–1995	NDNS	65+	1 687	4d FR	A, BC, CE, BP
	1997	NDNS	4–18	1 701	7d	
	2000–2001	NDNS	19–64	2 000	7d FR, BH	A, BC, CE, BP
	2003–2005	Low Income Diet and Nutrition Survey		3 728	4–24-h recall, BH, ATT	A, BC, BP, CE
North America						
Canada	1970–1972	Nutrition Canada	0–65+	12 795	24-h recall, FFQ	A, BC, MH
	2004–2005	National Population Health Survey	12+		FFQ	
	2004	Canadian Community Health Survey, Cycle 2.2, Nutrition	6 months +	35 107	24-h recall (2^{nd} Rcl subsample)	A, PA recall
		Wave 1 (2005)				
		Wave 2 (2006)				
		Wave 3 (2008)				
United States[c]	1977–1978	Nationwide Food Consumption Survey (NFCS)	(Households)	30 467	24-h recall, 2d FR	

Table 2 Nationwide food consumption surveys with individual-based dietary intake data—cont'd

Country	Year	Survey	Population (ages in years)	Sample Size	Dietary method[a]	Other information[b]
	1987–1988	NFCS	(Households)	25 100		
	1985–1986	Continuing Survey of Food	1–5	3 200	24-h recall, 2d FR	
		Intakes by Individuals (CSFII)	19–50 F	6 400		
			19–50 M	1 100		
	1989–1991	CSFII	All	15 192	24-h recall, 2d FR (subsample), KN, ATT, BH (subsample)	
	1994–1996	CSFII	All	16 103	24-h recall, 2d FR (subsample), KN, ATT, BH (subsample)	
	1998	CSFII	0–9	5 559	2–24-h recall, KN, ATT, BH (subsample)	
	1970–1974	National Health and Nutrition Examination Survey (NHANES I)	1–74	20 749	24-h recall, FFQ	A, BC, CE, MH
	1976–1980	NHANES II	1–74	20 322	24-h recall, FFQ	A, BC, CE, MH
	1982–1984	Hispanic HANES	6 months-74	11 653	24-h recall, FFQ BH	A, BC, CE, MH
	1988–1994	NHANES III	2 months+	33 994	24-h recall, FFQ	A, BC, CE, MH
	1999+	NHANES 1999–2000	Birth+	9 965	24-h recall (2nd day by phone), FFQ (by mail for 2 years+)	A, BC, CE, MH
		NHANES 2001–2002	Birth+	11 039 (5000 dietary)		
		NHANES 2003–2004	Birth+	10 122 (5 000 dietary)		
		NHANES 2005–2006	Birth+	10 348 (5 000 dietary)		
		NHANES 2007–2008	Birth+	10 149 (5 000 dietary)		
		NHANES 2009–2010	Birth+			In the Field
Oceania						
Australia	1983	National Dietary Survey of Adults	25–64	6 295	24-h recall	A, BC, CE, MH
	1985	National Dietary Survey of School children	10–15	5 224	1d FR	A, BC, BP
	1995	National Nutrition Survey	2+	13 858	24-hour recall (2nd 24-h recall from subsample) FFQ	A, BP
	2007	National Children's Nutrition & Physical Activity Survey	2–16	4 400	24-h recall (2nd by phone)	A, 2 PA Rcl, PR (on 5+subsample)
	2011–2012	National Nutrition and Physical Activity Survey Program (planned)		35 000 15 000 (Indigenous)		
New Zealand	1977	National Diet Survey	20–74	1 938	24-h recall	A
	1989	Life in New Zealand Survey	15+	1 702	24-h recall, FFQ, ATT, BH	A, BC, CE
	1997	National Nutrition Survey (adults)	15+	4 636	24-h recall, FFQ, KN, ATT, BH	A, BC, CE
	2002	Children's Nutrition Survey	5–14	3 275	24-h recall, FFQ,	A, BC
	2008–2009	Adult Nutrition Survey	15+	Results due in 2011	BH 24-h recall	A, BC, CE, BP

[a]Dietary method. 24-h recall; 24-h dietary, 1d FR; 1-day food record; FFQ, food frequency questionnaire; DH, dietary history; KN, dietary knowledge; ATT, dietary attitude; BH, dietary behavior.

[b]Other information. A, anthropometry; BC, biochemical tests; BP, blood pressure; CE, clinical exam; MH, medical history; F, Female; M, Male; PA Rcl, physical activity recall; PR, pedometer record.

[c]Starting in 1999, the CSFII and NHANES merged and became one integrated national survey.

four-hour recalls, anthropometrics, biochemical, clinical, and blood pressure assessments were completed on respondents 15 years and older. Results from this survey are due out in 2011.

Food Composition Databases

Reliable food composition data is essential in calculating the food intake by individuals and ultimately assessing the nutritional status of population groups. Factors such as sampling, variability, and analytical methods used to determine the nutrient content of food are important in the development of these databases. To better expand the benefits of nutritional surveillance considerable work has been done to try and harmonize food composition tables regionally. Harmonization requires comparable criteria at the food level (number of foods covered, food classification and descriptions systems, and representativeness of nationally consumed foods), component level (coverage, identification, definition, analytical methods), and value level (missing nutrient data, documentation). Issues related to data management, including the compilation, software capacity, evaluation, data interchange, and ownership also must be considered. There are currently greater than 150 food composition tables or nutrient databases in use worldwide. Many are based on the USDA National Nutrient Database for Standard Reference available on the Nutrient Data Laboratory website, www.ars.usda.gov/nutrientdata. For several years there has been a concerted effort to improve harmonization of analytical methods, definition, and mode of expression of foods and nutrients. INFOODS was established in 1984 to stimulate and coordinate efforts to improve the quality and availability of analytical food data worldwide. The European Food Information Resource Consortium, EuroFIR, is a partnership between several universities and 25 countries to develop and integrate a comprehensive cohort and validated network of food composition databanks in Europe. FAO also maintains a comprehensive list of food composition tables, which are grouped by region of the world at www.fao.org/infoods/directory_en.stm.

Emerging Nutritional and Health Issues

Obesity continues to be a major health issue for the developed world. Obesity has been linked to multiple health disorders, such as Type 2 diabetes, cardiovascular disease, hypertension and stroke, some cancers, and disability. Globally, greater than one billion adults are overweight and at least 300 million are obese. Moreover, childhood obesity is one of the most critical public health challenges of the 21st century. The problem is global and the prevalence has increased exponentially. Worldwide, in 2010 the number of overweight children under the age of five is estimated to be greater than 22 million. Overweight and obese children are more likely to become obese adults and more likely to develop noncommunicable diseases (NCDs) such as diabetes or cardiovascular disease at a younger age. In 2004, a Global Strategy on Diet, Physical Activity, and Health (DPAS) was adopted by WHO member states. The primary purpose of DPAS, a prevention-based strategy, is to significantly reduce the prevalence of NCDs and their common risk factors, mainly unhealthy diet and physical activity. The role of WHO in implementing the DPAS is to provide leadership, evidence-based recommendations and advocacy for international action to improve dietary practices and increase physical activity. Thus in 2009, the WHO Forum and Technical Meeting in Population-based Prevention Strategies for Childhood Obesity was held in order to identify priorities for population-based strategies to prevent childhood obesity and to define roles and responsibilities for various stakeholders.

An aging population also remains an emerging global public health issue. On October 1, 2010 at the 20th Anniversary of the International Day of Older Persons, it was announced that the world's population is now as old as it has ever been. Women comprise the majority of the older population in most countries, because globally they tend to outlive men. In 2010, one in every 10 persons was aged 60 or more and by 2050, that figure will be one in 5. Because both lean body mass and basal metabolic rate decline with age, maintaining adequate nutritional status among the elderly is especially important. In some instances nutrient requirements may be reduced, where requirements for other essential nutrients may rise in later life. Although WHO has developed a policy framework that promotes active aging and focuses on the prevention and reduction of the burden of disabilities and associated NCD risk factors, there continues to be an increasing demand worldwide for WHO guidelines that address the nutritional needs of the growing elderly population.

Despite access to nutritious foods and nutrition marketing strategies, poor nutrition is growing in affluent developed countries. Studies reveal that adult intakes for certain vitamins and minerals, including vitamin A, E, D, and folate, are below the recommended intakes in some countries. Notably the 2010 US Dietary Guidelines concluded that the under-consumption of vitamin D, calcium, potassium, and dietary fiber for both adults and children was a substantial public health concern. Economic globalization, with increasingly powerful transnational companies shaping global consumer behaviors, in many ways has created an environment of unhealthy food choices, such as fast food and high-sugar beverages. Environmental and climate changes may also place the nutritional status and dietary behavior of populations even in developed countries at risk.

The ability to monitor these emerging nutritional and health issues becomes increasingly important for enabling countries to better make decisions regarding cost-effective programs, policy, and regulations. As evidenced by the number of surveys conducted since 2005, countries appear committed to strengthening and maintaining national nutritional surveillance systems.

Further Reading

Australian National Monitoring in Public Health Nutrition (accessed 2010). www.health.gov.au/nutritionmonitoring
Canadian Community Health Survey (accessed 2010), http://198.103.98.28/fn-an/pubs/surveill/index-eng.php
DHHS-USDA Dietary Survey Integration. *What We Eat in America*, (accessed 2010) www.cdc.gov/nchs/nhanes/wweia.htm
Dietary Guidelines Advisory Committee (2010) Report of the Dietary Guidelines Advisory Committee on the Dietary Guidelines for Americans, 2010, to the Secretary of Agriculture and the Secretary of Health and Human Services. *US Department of Agriculture, Agricultural Research Service, Washington, DC*.
Egan, M.B., Fragodt, A., Raats, M.M., Hodgkins, C., Lubmers, M., 2007. The importance of harmonizing food composition data across Europe. European Journal of Clinical Nutrition 61, 813–821.
Elmadfa, I. (Ed.), 2009. European Nutrition and Health Report 2009, vol. 62. Karger, Basel, pp. 250–290.
Kim J *The Korean National Health and Nutrition Survey*, Cancer Epidemiology Branch, National Cancer Center, South Korea, personal communication.
National Health and Nutrition Examination Survey, DHHS, CDC, NCHS (accessed 2010) www.cdc.gov/nchs/nhanes.htm
National Institute of Health and Nutrition (2007) Outline for the results of the National Health and Nutrition Survey Japan, 2007. *Ministry of Health, Labour and Welfare*. Available at www.nih.go.jp/eiken/english/research/pdf/nhns2007.pdf
New Zealand National Nutrition Survey (accessed 2010), www.moh.govt.nz/moh.nsf/indexmh/dataandstatistics-survey-nutrition
Park, H.-K., 2008. Nutrition Policy in South Korea, Asia Pacific. Journal of Clinical Nutrition 17, 343–345.
Sahyoun, N.R., 2005. Nutritional surveillance, developed countries. In: Caballero, B., Prentice, A., Allen, L. (Eds.), Encyclopedia of Human Nutrition, 2nd edn. Elsevier, Oxford, pp. 363–371.

Supplementation: Developing countries

R. Shrimpton, Independent Consultant, Olhão, Portugal

© 2013 Elsevier Ltd. All rights reserved.
This is a reproduction of R. Shrimpton, Supplementation: Developing Countries, Reference Module in Biomedical Sciences, Elsevier, 2016, ISBN 9780128012383, https://doi.org/10.1016/B978-0-12-801238-3.99543-6.

Introduction 454
Iodine Supplementation 454
Vitamin A Supplementation 455
Iron/Folate Supplementation 456
Zinc Supplementation 457
Multiple Micronutrient Supplementation 457
Further Reading 458

Glossary

Anemia Low blood hemoglobin levels, mostly usually caused by iron deficiency.

Child Health Days Outreach strategy allowing high coverage of interventions such as vaccines and supplements during quarterly or six monthly extension of specific services into villages on set days.

Micronutrient supplementation The provision of micronutrients as pills, capsules, or syrups, for periodic consumption.

Randomized controlled trials Experiments carried out with a control or placebo group that is randomly and blindly assigned.

Therapeutic supplements Supplements with a dose larger than the daily requirement, used to treat micronutrient deficiency.

Introduction

Micronutrient supplementation is the distribution of specially formulated preparations of one or more micronutrients, usually in the form of a pill, a capsule, or sirup. It is often described as a 'short-term' option and a 'medical' approach and considered more appropriate for the treatment of severe micronutrient deficiencies in those most affected than to prevent deficiencies in whole populations. However, for the half of humanity affected by micronutrient deficiencies, especially of iodine, vitamin A, iron and zinc, the overwhelming majority of whom are the poor concentrated in the developing world, solving these problems through food-based approaches will take some time.

Vitamin A deficiency affects 30% of children younger than 5 years old in the developing world, compromising their immune systems and potentially contributing to some 1 million deaths each year. In the 6- to 24-month-old age group, mental development is impaired due to iron deficiency in 60% of the developing world's children. Severe iron deficiency also causes more than 60 000 deaths of women during pregnancy and childbirth every year, and 40% of pregnant women suffer from anemia, which is largely caused by iron deficiency. Approximately 18 million infants per year are born mentally impaired as a result of iodine deficiency during pregnancy despite the enormous advances in salt iodization programs. Providing vulnerable groups, such as children and women of childbearing age, with low-cost vitamin and mineral supplements is the least that governments can do to protect the survival, growth, and development of the next generation as a first step toward realizing the right of every individual to be adequately nourished.

Experience in achieving high coverage of those most at risk with micronutrient supplements is quite varied, with both successes and failures. A good communication strategy is an essential part of achieving high levels of adherence in micronutrient supplementation programs, but these aspects are not particular to nutrition programs and are not considered here. Deficiencies of iodine, iron, vitamin A, and folate are the most commonly recognized deficiencies for which there are programs, but in practice most of those affected have multiple vitamin and mineral deficiencies that overlap and interact at great cost. This article reviews the policy dimensions of the efforts to establish programs aimed at eliminating iodine deficiency, iron deficiency anemia, and vitamin A deficiency through supplementation, and it provides a perspective on zinc supplementation and multiple micronutrient supplementation as future components of nutrition programs in developing countries.

Iodine Supplementation

Today, approximately 70% of the world's salt is iodized, compared to just 10% in 1990, and therefore the need for iodine supplementation programs is greatly reduced. Despite this progress almost 2 billion people still have low urinary iodine levels, and in

these populations iodine supplements should be used during pregnancy and early childhood. In the past, it was common to provide annual intramuscular injections of iodized oil to women of reproductive age in order to ensure iodine status during the first months of pregnancy when the risk of cretinism is greatest. In more recent years, oral iodized oil capsules have proven to be as efficacious and more effective in controlling iodine deficiency in both women of reproductive age and schoolchildren. Oral iodine supplements initially based on expensive poppy seed oil have since been replaced by cheaper rapeseed and peanut oil preparations, which are equally effective.

Vitamin A Supplementation

The use of supplements to help control vitamin A deficiency has grown enormously during the past two decades. Although the elimination of vitamin A deficiency by the year 2000 was one of the goals set at the World Summit for Children in 1990, little progress was evident at mid-decade. Clinical vitamin A deficiency was estimated to affect approximately 3.3 million children younger than the age of 5 years in 1995, with an additional 100 million subject to subclinical deficiency. The periodic distribution of high-dose vitamin A supplements, originally employed in Indonesia during the 1970s for the prevention of blindness in children, was shown in the 1980s to also impact young child mortality. However, the lack of perception of vitamin A deficiency as a problem was a substantial barrier to establishing large-scale preventive supplementation programs. The prevalence of clinical signs of frank vitamin A deficiency, such a Bitot's spot and corneal lesions, which make it a 'public health problem', is very small at just 0.5%. Because clinical signs are often more common in rural populations, a significant vitamin A deficiency problem can easily go undetected. National representative surveys were thus a prerequisite for taking action.

Convincing proof of the efficacy of vitamin A capsules for child mortality reduction in the early 1990s helped to create increased momentum for population wide preventive supplementation programs. The turning point for increasing the coverage of vitamin A supplements was the meta-analysis of eight efficacy trials, which indicated that improving the vitamin A status of children aged 6 months to 5 years by massive-dose capsule distribution reduced child mortality rates by approximately 23%. The important conclusion of these meta-analyses was that increased risk of mortality from vitamin A deficiency was not just limited to those portions of the population with severe vitamin A deficiency problems but was present across the whole population distribution. Further subsequent meta-analysis carried out over the past two decades have all confirmed these findings.

What consisted of 'the justification' for carrying out vitamin A supplementation programs evolved rapidly during the latter half of the 1990s. Many of these discussions were held at the meetings of the International Vitamin A Consultative Group and the working group on vitamin A of the Standing Committee on Nutrition of the United Nations. A broad technical consensus was finally accepted that even in the absence of survey data, it was highly likely that the benefits of vitamin A supplements would be evident in populations in which the mortality rates for those younger than 5 years old were higher than 70 per 1000. Before this, vitamin A supplements were targeted to those children with illnesses such as measles and diarrhea. Subsequent to this consensus, a global policy to integrate vitamin A capsule distribution into regular immunization schedules, and also to incorporate vitamin A capsules into the national immunization programs was rapidly adopted.

Programmatic vitamin A interventions received considerable impetus from the Vitamin A Global Initiative, an informal interagency advocacy group that worked to promote the adoption of vitamin A supplementation programs. The initiative included World Health Organization (WHO) and United Nations International Children's Emergency Fund (UNICEF), together with Canadian International Development Agency (CIDA) from Canada, DIFID from UK, United States Agency for International Development (USAID) from USA, and the Micronutrient Initiative (MI). Through their networks, these various organizations worked together to convince governments with high mortality rates for children younger than age 5 years to introduce periodic vitamin A capsule distribution programs.

By the end of the 1990s, vitamin A supplementation programs had seen a remarkable expansion, which has been maintained during the first decade of the new millennium. The number of countries with vitamin A programs increased from 10 in 1995 to 72 in 2000, and 103 in 2004. These are mostly countries with high mortality rates for children younger than 5 years old and/or where vitamin A deficiency is a public health problem. The ways in which the vitamin A capsule programs were developed and implemented have varied by country and over time. The most common strategy was to use national immunization days for polio eradication to piggyback vitamin A supplements, but because the polio eradication strategy required two nationwide campaigns not more than 2 months apart, some countries also promoted separate micronutrient days, or child health days, so that children would get at least two capsules during the course of a year, 6 months apart. For example, as polio eradication has progressed and vaccination ceased, capsules have increasingly been delivered through child health days together with other interventions such as deworming and malaria bed net distribution.

The coverage of vitamin A capsules, which was already high by the turn of the century, has continued to climb during the first decade of the new millennium. Although 50% of children in 103 countries had received one dose in 1999, just 16 percent had received two doses. By 2004 those receiving two doses had climbed to 58% and by 2008 had reached 71% coverage. Extrapolation of the protective effect of a 23% reduction in child mortality shown by the meta-analysis to the increased coverage of capsules in the decade after 1998 suggests that many millions of lives would have been saved. Others have estimated child deaths saved as far fewer than this however, and many have lamented the lack of proof of any such impact. Furthermore despite this high coverage of capsules, the rates of vitamin A deficiency based on serum retinol levels remains high and stable during this period, still affecting some 163 million children or approximately 30%.

The challenge that remains for vitamin A supplementation is one of sustainability. Although supplements have always been viewed as a short-term solution, in reality they need to be maintained as long as mortality rates remain high and no dietary solutions are put in place. Sustaining the provision of the vitamin A capsules is likely to become a problem, as until now, supplements have been provided predominantly by the Canadian government and supplied through UNICEF. Whether governments will eventually pick up these costs remains to be seen. The costs for individual governments to take on are small, however, especially compared to the potential benefits in terms of mortality reduction. However, in most places more effort is also needed to increase access to other sources of vitamin A, be it through diet and/or fortification, so that capsules can be phased out.

Iron/Folate Supplementation

Although iron deficiency is the most widespread of nutritional problems, and despite the existence of policy and programs in most countries, supplementation with iron has not proven to be a very successful intervention. Global policy recommendations to routinely provide iron/folate supplements for women during pregnancy and lactation have changed little in almost three decades. These are that all anemic pregnant women should receive such supplements in almost all contexts, i.e., there is no alternative to supplements for treating anemia and the recommendation is that at least ninety tablets should be taken during pregnancy to treat anemia. Despite most developing countries having national iron supplementation policies, the World Summit for Children's goal to reduce anemia in women by one-third was given little or no priority by the principal actors involved such that no progress was made during the past decade. In 2007 anemia was estimated to affect 40% of nonpregnant women and 41% of pregnant women in developing countries, virtually the same rates that existed in the nineties.

Although there is ample evidence that iron deficiency and the anemia associated with it are a great burden on society, especially the poor, the advocacy base for pushing for program implementation has remained weak. This is largely because the link of iron deficiency to maternal and child survival has not been concretely proven. The ethical difficulty of doing randomized controlled trials with a placebo group, when all countries have a policy to give iron supplements during pregnancy has contributed to this. However, new evidence from trials in nonanemic women in developed countries find that iron supplements increase birth weight by upward of 100 g compared to placebo, suggesting that iron supplements may well have unsuspected benefits for child survival, growth, and development. The effect of iron deficiency on cognitive deficits in children and on adults later in the life course have long been established. The absolute losses in Southeast Asia are estimated to be approximately $5 billion annually, and for India the median value of productivity losses due to iron deficiency alone is approximately $4 per capita or 0.9% of gross domestic product.

Despite high cost effectiveness, little or no priority has been given to iron deficiency anemia reduction programs. At $0.002 per tablet, the iron supplement is relatively cheap, and the cost per disability adjusted life year of $13 makes the supplementation of pregnant women with iron a very cost-effective intervention. At the national level, despite the existence of national policies, rarely is there a budget for the provision of supplements and/or supervision of iron deficiency anemia programs. Although UNICEF is a major supplier of iron/folate supplements to the developing world, the level of supply is far lower than that believed to be needed. In the period 1993–1996, 2.7 billion tablets were shipped to 122 countries at a cost of $7.5 million as part of UNICEF assistance to programs aimed at eliminating maternal anemia. This was less than 5% of that needed to cover all pregnancies in developing countries. There have been few, if any, attempts to gage the coverage of iron/folate supplements at any level, be it district, national, or international. Neither has there been any effort put into creating political accountability to ensure high coverage.

Many meetings and publications during the past few decades that have examined the causes and solutions of iron deficiency anemia conclude that lack of effectiveness of iron supplementation programs for anemia control is largely related to problems with supply of the supplement. Although the side effects of iron pills are often cited as the reason why iron supplementation programs do not work, this rarely seems to be the case. One of the major causes of nonadherence seems to be lack of understanding of the benefits the supplements can bring among health staff that deliver the tablets. Most of the program reviews have concluded that where supportive community-level delivery mechanisms are put in place that encourage adherence, and the supply of supplements is ensured, high levels of coverage can be achieved and sustained and anemia controlled. It is often the case, however, that in health systems in developing countries, nutrition is everybody's business and nobody's responsibility, and iron supplements have ended up low on the list of things to do.

Despite an international consensus that supplementation has a key role to play in the control of iron deficiency anemia, and that this will contribute to reducing maternal mortality, there is still little traction in this area. Demographic Health surveys include questions concerning how many mothers took at least 90 iron tablets during their last pregnancy, and very few countries achieve high coverage. In 1998, a technical consensus meeting on what was needed to solve the problem of iron deficiency made the recommendation that although the interventions already existed for reducing both iron deficiency and iron deficiency anemia, more work was needed to develop large-scale programs with packages of interventions delivered through multiple sectors, including hygiene and sanitation, because iron supplements alone will not ensure anemia control in many settings. Infections such as malaria and gastrointestinal parasites are also important causes. Furthermore, a global review of anemia causality revealed that perhaps only half of anemia is solely due to iron deficiency, with other micronutrient deficiencies such as vitamin A contributing as well. To be effective, maternal anemia control programs must include infection control, together with the community delivery of iron supplements.

It is much easier to treat anemia before pregnancy, when supplements can be taken once a week instead of daily. Such approaches have been shown to be very effective among adolescent school girls, for example, as well as in family planning

programs. To be effective the supplementation has to be supervised and accompanied by infection control, such as periodic deworming.

Zinc Supplementation

The WHO/UNICEF recommendation is to give supplemental zinc for 10 day as part of the treatment of diarrhea. This is based on strong evidence for the efficacy of therapeutic zinc in improving the prognosis of children being treated for diarrheal disease, A pooled analysis of randomized controlled therapeutic zinc trials in children with diarrhea showed that zinc-supplemented children with acute diarrhea had a 15% lower probability of continuing diarrhea on a given day, and in those with persistent diarrhea there was a 24% lower probability. In addition, children with persistent diarrhea had a 42% lower rate of treatment failure or death if zinc supplemented. Given that even the current interventions included in child health programs for diarrheal disease treatment, such as oral rehydration therapy, face enormous barriers to achieving and maintaining high levels of coverage, the challenge for achieving high levels of coverage of zinc supplements in the treatment of diarrhea is likely to be considerable. If these efforts are successful however, then the impact is likely to be great. The most effective way to give preventive zinc supplements is an ongoing research question.

Multiple Micronutrient Supplementation

In recent years, the case has increasingly been made for providing multiple micronutrient supplements instead of iron supplements for young children and women of reproductive age in developing countries. A woman's or an infant's diet that is deficient in iron is likely to be deficient in many other micronutrients. Outside of emergency situations, such as natural catastrophes, famine, and civil strife, poor dietary quality rather than quantity is the determinant of inadequate micronutrient status among infants and women. The nutrient-to-energy ratios of iron, zinc, folate, vitamins B_6 and B_{12}, vitamin A, riboflavin, and calcium are commonly below the recommended levels needed, assuming energy needs are met.

The UN agencies agreed the composition of a multiple micronutrient supplement for use among pregnant and lactating women in developing countries a decade ago. The formulation includes 15 micronutrients (vitamins A, D, E, B_1, B_2, B_6, B_{12}, and C, niacin, and folic acid and minerals Fe, Zn, Cu, I, and Se), all at the RDA level, except for folic acid, which was included at the 400-μg level – considered sufficient to prevent neural tube defects if taken periconceptually. The main cost of the delivery of a nutrient supplement for women of reproductive age is not the cost of the supplement but the cost of the delivery system. Although it may not be working very well, a delivery system already exists for the iron/folate supplements that could be used to provide these other micronutrients. Adding the extra nutrients to the iron/folate tablets will not add more than 20% to the cost of the tablet, and although the incremental cost of distributing a multiple micronutrient supplement is likely to be small, the increased benefits may be large. Although the need for micronutrient supplementation in pregnancy is likely to be great because of widespread maternal undernutrition, and the supplements have been recommended for use in populations affected by emergencies. it was recognized that regular public health resources are always limited and priority is given to interventions that are both efficacious and effective. Proving the efficacy of multiple micronutrient supplements is thus essential for being able to advocate for their wide-scale use in nonemergency settings. Tablets of similar composition are regularly prescribed by physicians and/or purchased by mothers in developed countries, and they can be found in the pharmacies of the capitals of most developing countries and are widely consumed by the richer segments of the population.

This multiple micronutrient supplement was then tested in a series of 12 efficacy and six effectiveness trials covering 12 countries and spanning three continents – Asia, Latin America, and sub-Saharan Africa. A meta-analysis of these trials found that both supplements were equally effective in tackling maternal anemia, even though the iron content was often lower in the multiple micronutrient supplement than in the iron–folic acid supplement. There were no significant differences in the rates of stillbirth, early neonatal death, or neonatal death between the supplemented groups. The small, significant increase in mean birth weight (24 g) among infants of mothers receiving multiple micronutrients compared with infants of mothers receiving iron–folic acid is of similar magnitude to that often produced by food supplementation during pregnancy, and larger micronutrient doses seemed to produce greater impact. Meaningful improvements have also been observed in height and cognitive development by 2 years of age in the children of the mothers from the multiple micronutrient group.

WHO and UNICEF recommend the use of sirup and/or tablets containing iron for the treatment of anemia in young children, and such products are available through UNICEF supply division in Copenhagen. These products have very little penetration considering the size of the infant anemia problem in most developing countries, where half of all children are commonly affected. Despite the recognition that iron deficiency often coexists with zinc deficiency, together with inadequate intakes of other B vitamins (B_6, riboflavin, and niacin) in infant dietaries, there is no multiple micronutrient supplement available for infants. UNICEF has also been testing the efficacy of a foodlet (a large crumbly pastille that is a cross between a tablet and a food) containing multiple micronutrients during infancy through the Infant Research on Infant Supplementation trials. Trials of multiple micronutrients as preventive supplements have also been carried out by many different groups using supplements provided in the form of sprinkles, tablets, and even as a beverage. Preliminary results of these trials point to a greater impact on anemia and enhancement of multiple micronutrient status by the multiple micronutrient supplements than that of iron supplements, as well as small improvements in growth.

There is a need to bring all of this broad spectrum of experimental and programmatic work together to reach conclusions and achieve consensus before policy and program recommendations can be made on how best to include multiple micronutrient supplements in programs to improve maternal and child health in developing countries.

Further Reading

Allen, L.H., Peerson, J.M., Olney, D.K., 2009. Provision of multiple rather than two or fewer micronutrients more effectively improves growth and other outcomes in micronutrient-deficient children and adults. Journal of Nutrition 139 (5), 1022–1030.

Fall, C.H., Fisher, D.J., Osmond, C., Margetts, B.M., Maternal Micronutrient Supplementation Study Group, 2009. Multiple micronutrient supplementation during pregnancy in low-income countries: A meta-analysis of effects on birth size and length of gestation. Food and Nutrition Bulletin 30 (4 Supplement), S533–S546.

Latham, M., 2010. The great vitamin A fiasco. World Nutrition 1 (1), 12–45.

Rasmussen, K.M., Stoltzfus, R.J., 2003. New evidence that iron supplementation during pregnancy improves birth weight: New scientific questions. American Journal of Clinical Nutrition 78 (4), 673–674.

Ronsmans, C., Fisher, D.J., Osmond, C., Margetts, B.M., Fall, C.H., Maternal Micronutrient Supplementation Study Group, 2009. Multiple micronutrient supplementation during pregnancy in low-income countries: A meta-analysis of effects on stillbirths and on early and late neonatal mortality. Food Nutrition Bulletin 30 (4 Supplement), S547–S555.

Shrimpton, R., Gross, R., Darnton-Hill, I., Young, M., 2005. Zinc deficiency: What are the most appropriate interventions? British Medical Journal 330 (7487), 347–349.

Shrimpton, R., Huffman, S.L., Zehner, E.R., Darnton-Hill, I., Dalmiya, N., 2009. Multiple micronutrient supplementation during pregnancy in developing-country settings: Policy and program implications of the results of a meta-analysis. Food and Nutrition Bulletin 30 (4 Supplement), S556–S573.

Sommer, A., 2008. Vitamin A deficiency and clinical disease: An historical overview. Journal of Nutrition 138, 1835–1839.

UNICEF/UNU/WHO, 2001. Iron Deficiency Anaemia: Assessment, Prevention and Control. A Guide for Programme Managers. World Health Organization, Geneva.

UNICEF, 2007. Vitamin A supplementation: A Decade of Progress. United Nations Children's Fund, New York.

UNICEF, 2009. Tracking progress on child and maternal nutrition: A Survival and Development Priority. United Nations Children's Fund, New York.

UNSCN, 2010. Sixth Report on the World Nutrition Situation. United Nations Standing Committee on Nutrition, Geneva.

West, K.P., Klemm, R.D.W., Sommer, A., 2010. Vitamin A saves lives. Sound science, sound policy. World Nutrition 1 (5), 211–229.

WHO, UNICEF, 2005. Joint Statement: Reaching Optimal Iodine Nutrition on Pregnant and Lactating Women and Young Children. World Health Organization, Geneva.

WHO, WFP and UNICEF, 2007. Joint Statement: Preventing and Controlling Micronutrient Deficiencies in Populations Affected by an Emergency – Multiple Vitamin and Mineral Supplements for Pregnant and Lactating Women, and for Children Aged 6 to 59 months. World Health Organization, Geneva.

INDEX

Notes

Cross-reference terms in *italics* are general cross-references, or refer to subentry terms within the main entry (the main entry is not repeated to save space). Readers are also advised to refer to each article for additional cross-references — not all of these cross-references have been included in the index cross-references.

The index is arranged in set-out style with a maximum of three levels of subheading. Major discussion of a subject is indicated by bold page numbers. Page numbers suffixed by t, f, and b refer to Tables, Figures, and Boxes respectively. *vs.* indicates a comparison.

The index entries are presented in word-by-word alphabetical sequence in which a group of letters followed by a space is filed before the same group of letters followed by a letter. For example, entries beginning 'air density' are alphabetized before 'aircraft.' Prefixes and terms in parentheses are excluded from the initial alphabetization.

A

A-type PAC, antibacterial activity of **2**:542
absorption **1**:252–253, **1**:283–285
  effect of meal size on **2**:445–446
  and metabolism of flavonoids **2**:534–535
  of vitamins **2**:217–221
acceptable daily intake (ADI) **2**:627
Acceptable Macronutrient Distribution Ranges (AMDR) **3**:238–239, **3**:239t
acesulfame potassium (Ace-K) **2**:626, **2**:628, **2**:631
acetaldehyde
  effects of **2**:262–263
  metabolism **2**:255
acetate
  effects of **2**:263
  metabolism of **2**:255
acetyl-CoA carboxylase (ACC) **2**:199
acetylation **2**:177
acetylcholine **1**:163
acquired immuno deficiency syndrome (AIDS) **3**:535–536
acute diseases, vitamin D and **1**:540–543
acute liver failure **3**:460
acute neuroglycopenia **3**:404
acyl-coenzyme A:cholesterol O-acyltransferase (ACAT) **3**:368–369
acyltransferase (ACAT) **1**:296
adaptive thermogenesis **2**:129–131
  See also postprandial thermogenesis
  efficiency of muscle work **2**:131
  interactions and compensations across components of EE **2**:131
  in non-resting state **2**:130–131
  in resting state **2**:129–130
  variability in spontaneous physical activity **2**:131
added sugars **1**:137
adenosine triphosphate (ATP) **4**:56
adequate intake (AI) **1**:165–166, **1**:182
adequate vegetarian dietary patterns **2**:671
  inadequate vegetarian dietary patterns **2**:671–672
  key nutrients for vegetarians over life cycle **2**:672
  special nutritional concerns for vegetarians **2**:672
adipokines **3**:593–594
adipose tissue (AT) **3**:25
  See also brown adipose tissue
  development **3**:96–98
  distribution **3**:101–102
  function **3**:102–104
  regulation of metabolism **3**:104–107
  structure **3**:99–101
adolescence **4**:143
adolescents **4**:226
  adipose stores **4**:3
  carbohydrates and fiber **4**:8
  dietary intakes **4**:4–8
  dietary recommendations **4**:3–4
  energy, protein, and salt **4**:4
  fats **4**:4–8
  food choices at school **4**:9
  fruit and vegetable consumption **4**:9
  growth **4**:2
  impact of lifestyle on nutrition **4**:9–11
  micronutrients **4**:8
  physical activity **4**:10–11
  physical changes during adolescence **4**:2–3
  sexual development **4**:3
  sleep **4**:11
  snacking and soft drink consumption **4**:9
  socioeconomic status **4**:10
adult cohort **3**:690
adult health **4**:59
Aerobics Center Longitudinal Study (ACLS) **1**:358
aflatoxins **2**:496–497
aging **2**:657
  accelerated aging syndromes **4**:14
  age-related diseases **4**:16
  cellular senescence **4**:14
  cross-linking **4**:15
  damage accumulation theories **4**:14–16
  dietary energy restriction **4**:19
  disposable soma theory **4**:14
  error catastrophe **4**:15
  free radicals **4**:15
  immunologic theory and immunosenescence **4**:14
  longevity genes **4**:13–14
  modification of aging process **4**:17–19
  molecular biological interventions and aging process **4**:19
  neuroendocrine theories **4**:14
  nutritional modification **4**:17–19
  programed and genetic theories **4**:13–14
  somatic mutation and DNA repair **4**:15
  theories **4**:13–16
agouti viable yellow (*Avy* allele) **4**:55–56
agriculture **4**:373–374
air plethysmography **3**:22
air-displacement plethysmography (ADP) **3**:25–26

alcohol **4**:213–216, **4**:267
  absorption and distribution of **2**:252–253
  alcohol-associated liver disease **3**:452–453
  complex relationship between alcohol consumption and adverse outcomes **2**:268–269
  consumption **4**:10
  dehydrogenase **2**:253
    isoenzymes **2**:254
  effects
    on cardiovascular system **2**:262
    on central nervous system **2**:259–260
    consumption on dietary intake **2**:267–268
    on liver function **2**:262
    on muscle **2**:260–261
  metabolism of **2**:253
  neuroendocrine effects of **2**:260
  nonoxidative metabolism of **2**:255–256
  and nutrition **2**:261–262
  period over which alcohol is consumed **2**:257
  physical properties of ethanol **2**:252
  physiological effects of **2**:259
  risks of alcohol misuse **2**:269
alcoholic liver disease (ALD) **2**:269
alimentary tract
  constipation and diverticular disease **2**:312
  diseases of **2**:312–313
all-cause mortality **1**:354
allergenic food **4**:48
almonds **2**:514
α-AAs
  chemical reactions
    of amino group in **1**:19
    of carboxyl group in **1**:19
    involving amino and carboxyl groups of **1**:20
    of side chain in **1**:19–20
α-Carotene **1**:158
alphaα-linolenic acid (n-3) **4**:136
aluminum (Al) **1**:1–2, **1**:449
  aluminum-induced bone disease **1**:7–8
  aluminum-induced encephalopathy **1**:7
  aluminum-induced microcytic anemia **1**:8
  in beverages and foods **1**:2–3
  bioavailability and biotransformation **1**:5–6
  biokinetics in blood **1**:6
  diseases related to aluminum intake **1**:7
  evidence for role in Alzheimer's disease **1**:8
  excretion **1**:6–7
  exposure and dietary intake **1**:5
  food uses of aluminum compounds **1**:2
  in foods from processing, packaging, and storage **1**:3–5
  nonfood uses **1**:2
  properties and natural occurrence **1**:1–2
  tissue deposition and body retention **1**:6
Alzheimer disease **3**:596
  evidence for role in **1**:8
American Academy of Pediatrics **3**:222
American College of Obstetricians and Gynecologists (ACOG) **3**:259–260
American Diabetes Association (ADA) **3**:236, **3**:238–239, **3**:242, **3**:255
American Society for Parenteral and Enteral Nutrition (ASPEN) **1**:184
amino acid response (AAR) **2**:204
amino acid response elements (AARE) **2**:204
amino acids (AAs) **1**:10–11, **1**:37, **2**:183, **2**:204, **3**:120–121, **3**:331, **3**:357, **3**:635, **4**:75, **4**:136
  amino acid transport systems **1**:380–381
  ammonia detoxification via ureagenesis in liver and small intestine of humans **1**:30–32
  catabolism by gut microflora **1**:382
  cell-and tissue-specific synthesis of AAs in humans **1**:24–27
  chemical analyses of **1**:20–21
  chemical properties of **1**:16–19
  chemical reactions of **1**:19–20
  chemical stability of **1**:19
  classification of **1**:21
  degradation of AAs in humans **1**:27–30
  disorders **3**:414–417
  endothelial cells, smooth muscle cells, macrophages, and lymphocytes **1**:27
  of food proteins **2**:561
  free AAs and peptide-bound AAs **1**:14–15
  functions of AAs in sensing **1**:44
  intestinal epithelial digestion and absorption of **1**:379–382
  kidneys **1**:27
  liver **1**:25–26
  metabolism **2**:120–121
  metabolites of AAs in humans **1**:32
  modulation of gene expression **2**:204–205
  pattern of metabolic demand **2**:576–577
  physical appearance, fluorescence, and melting points of crystalline **1**:16
  physiological functions **1**:37
    of AA metabolites **1**:37–44
  placentae, mammary glands, ovaries, testes, and pancreas **1**:27
  production and toxicity of ammonia in humans **1**:30–31
  protein turnover in humans **1**:32
  requirements **2**:569
  scoring **2**:562–563
  sense organs **1**:27
  skeletal muscle, heart, brain, lungs, and white adipose tissue **1**:26–27
  small intestine **1**:27
  solubility in water and organic solvents **1**:18–19
  supplementation
    and asthma **3**:518
    with **3**:635–637
  synthesis by gut microflora **1**:382–383
  synthesis of AAs in humans **1**:24–27
  tastes of crystalline **1**:16
  urea cycle in the liver and small intestine **1**:31–32
  zwitterionic form of **1**:19
amino terminal domain (ATD) **2**:625
AMP-activated protein kinase (AMPK) **1**:391
amylopectin **2**:424
amylose **2**:424
anemia **2**:270–271, **4**:124–125
animal models **4**:53–55
  maternal overnutrition models **4**:54
  maternal undernutrition models **4**:53–54
  pharmacological models **4**:55
anorexia **3**:139
anorexia nervosa (AN) **3**:290
  clinical characteristics **3**:291
  clinical presentation and physical examination **3**:292
  common laboratory findings **3**:293
  diagnostic criteria **3**:290–291
  differential diagnoses **3**:292
  epidemiology **3**:291–292
  management **3**:293–294
  pathophysiology **3**:292
  screening **3**:292
  testing **3**:293
anthocyanins **2**:533
anthropometric indices **4**:87
anthropometric measurements **4**:87–89
anthropometry **3**:23, **3**:602
anti-inflammatory drugs **2**:99–100
anti-nutrient compounds **2**:303
antidiuretic hormone (ADH) **1**:425
antidiuretic vasopressin (AVP) **1**:425
antinatriuretic effector mechanisms **1**:424–425
antioxidants **1**:63–66, **2**:686, **3**:518–519, **4**:265
  defense system **1**:590
  effects of antioxidant supplementation
    on aging **1**:50
    on cancer development risk **1**:54
    on exercise adaptations and physical performance **1**:56
  explanations for disagreement between findings of observational studies and clinical trials **1**:74
  hypothesis **1**:66
  micronutrients and inflammation **2**:74
  new physiological antioxidants **1**:64–66
  normal plasma concentration of common antioxidants **1**:63–64
  as signaling **1**:64
  sources **1**:63
  supplementation on diabetes, metabolic syndrome, obesity and hypertension **1**:50–54
apolipoprotein (a) **3**:364, **3**:367
apolipoprotein A-I **3**:364
apolipoprotein A-II **3**:366
apolipoprotein A-IV **3**:366
apolipoprotein A-V **3**:366
apolipoprotein B-100 **3**:366
apolipoprotein B-48 **3**:366
apolipoprotein E **3**:366
apolipoproteins C-I **3**:366
apolipoproteins C-II **3**:366
apolipoproteins C-III **3**:366
appetite **2**:58, **2**:629
  applications of satiety to societal challenges **2**:63–64
  cognitive and social cues **2**:7–8

as complex system **2**:58
diet composition **2**:8
and satiety to prevent weight gain **2**:8–9
drive to eat and food selection **2**:60
emotional stimuli and over-consumption **2**:6
individual variability in appetite control and satiety responsiveness **2**:62–63
learned appetites, satieties, and feeding behavior **2**:2–4
macronutrients and energy density affect satiety **2**:8
meal patterns, appetite, and energy balance **2**:7
measuring feeding behavior **2**:5
measuring food intake **2**:4–5
measuring hunger **2**:4
methodological issues **2**:4–6
nature of feeding behavior and appetite control **2**:2
sensory stimulation and food hedonics **2**:5–6
sensory stimuli and body weight **2**:6
sensory *vs.* nutritional determinants of intake **2**:6
social and situational influences on feeding behavior **2**:7–8
impact of ultra-processed foods on satiety **2**:64–65
appropriate for gestational age (AGA) **4**:76
arginine **3**:560
arginine vasopressin (AVP) **3**:217
arm fat area (AFA) **3**:93
arm muscle area (AMA) **3**:92
aromatic amino acid decarboxylase **1**:493
aromatic compounds **2**:405
arsenic **1**:449–450, **2**:359–360
dietary exposure **2**:359–360
permissible intakes **2**:360
toxicity **2**:360
toxicokinetics **2**:360
arthritis
clinical features **3**:110
definitions and etiology **3**:109–110
prevalence **3**:110
role of diet in management and prevention of degenerative arthritis **3**:114
role of diet in management of inflammatory arthritis **3**:111–114
artificial pancreas systems **3**:245
aryl hydrocarbon receptor, functions of AAs and metabolites as activators of **1**:44
Ascaris lumbricoides **3**:543
ascorbate
antioxidant and prooxidant actions of **1**:508–509
requirements estimated from maintenance of body pool of **1**:510
role in iron absorption **1**:508
ascorbic acid **2**:231–233, **4**:172
functions **2**:231–232
metabolic functions of **1**:507–509
metabolism **2**:232–233
asparagine synthase (ASNS) **2**:204
asthma **1**:73, **1**:513, **3**:514–515
amino acid supplementation and **3**:518
development of **3**:514–515
epidemiology **3**:514
obesity and asthma **3**:516
role of nutrition in prevention and management of **3**:515–519
atherosclerotic cardiovascular disease (ASCVD) **3**:373
benefits of lipid-lowering on ASCVD risk **3**:373
ATP binding cassette transporter A1 (ABCA1) **3**:369
ATP binding cassette transporter G1 (ABCG1) **3**:369
atrial natriuretic peptides **1**:425
attention-deficit hyperactivity disorder (ADHD) **3**:353
dietary patterns in **3**:355
research on nutrition in **3**:359
Australian indigenous and traditional foods (AITFs) **2**:640
enabling environments to promote consumption and production of **2**:645
auto-weaning. *See* baby led weaning (BLW)
autocrine/paracrine-acting growth factor **4**:75–76
autoimmune diseases **4**:60
autoimmune hepatitis **3**:454
autophagy-related genes **1**:392

B

B vitamins **4**:96, **4**:171
B-type PAC **2**:541
Baby Led Introduction to SolidS (BLISS) **4**:48–49
baby led weaning (BLW) **4**:48–49
Bacillus cereus **2**:365
bacterial diseases **3**:266
bacterial foodborne illness, prevention of **2**:372–373
bacterial metabolism **1**:491
bacterial toxins **2**:364–367
balanced energy protein supplements (BEP) **4**:235–237
bambara beans **2**:431
barley (*Hordeum vulgare*) **2**:295
barriers evaluating school food interventions impact on health outcomes **4**:346
basal metabolic rate (BMR) **1**:218–220, **2**:573
basic helix-loop-helix leucine zipper proteins (bHLH/LZ) **2**:198–199
beef **2**:453–454
behavioral psychologists **3**:628
behavioral treatment strategies **3**:588
behavioral weight management **3**:300
benign fructosuria **2**:395
benzimidazoles **2**:384
beriberi **1**:438
clinical features of **1**:433
betaine **1**:163
β-carotene **1**:69–71, **1**:157–158, **3**:213
β-Cryptoxanthin **1**:158
β-Glucans **2**:417
beverages
alcohol content **2**:258
and health
coffee **2**:280–281
100% fruit or vegetable juice **2**:282
milk **2**:281–282
soft drinks **2**:282–283
tea **2**:278–279
bile acid synthesis **1**:178
bile salts **3**:450–451
biliary excretion of biotin and metabolites **1**:89
binge eating
assessment of **3**:297–298
available resources **3**:302
comorbidity **3**:298–299
eating disorder and obesity history **3**:301
epidemiology and risk factors **3**:298
guided self-help **3**:300
psychiatric status **3**:302
psychosocial treatments **3**:299–300
selection of treatment for specific patients **3**:301–302
treatment of **3**:299
bioactive compounds **2**:302, **2**:406–408
bioavailability
of carotenoids **1**:81
to vitamin A **1**:80
conversion factors **1**:82–83
effectors of bioavailability **1**:83
of flavonoids **2**:535
as important component in development of food-based strategies to combat iron deficiency **1**:78–79
as important component in the development of food-based strategies to combat iron deficiency **1**:79–80
influence of vitamin A status **1**:83
nonheme iron **1**:76–80
of provitamin A carotenoids **1**:80–83
biochemical indices
choice of laboratory tests **3**:17–18
essential fatty acid status **3**:5
essential mineral and trace element nutritional status **3**:12–17
evaluation of laboratory indices **3**:18
protein nutritional status **3**:2
vitamin B_1 **3**:9–10
vitamin B_2 **3**:10
vitamin B_3 **3**:10
vitamin nutritional status **3**:5–11
biodiversity & pollinator loss **4**:375
biofortification **1**:582, **2**:286–288
achieving potential for impact **2**:291
effectiveness **2**:290
efficacy **2**:290
history and progress **2**:286–287
nutrient bioavailability **2**:289–290
nutrient retention **2**:288–289
potential of biofortification to improve micronutrient status **2**:288–291
target foods and nutrients **2**:287–288

biofortification (*continued*)
zinc and iron **2**:288–289
bioimpedance analysis (BIA) **3**:23
biological basis for sweetener intake **2**:629–630
biological functions **1**:444–445
biology interact with rapid social changes **4**:442
biomarkers
fermentation of dietary fiber as **1**:212–213
and status assessment **1**:256
biotin **2**:241–243, **3**:11
assessment of biotin status **1**:92–93
circumstances leading to deficiency **1**:93–94
clinical findings of frank deficiency **1**:94
deficiency diseases **1**:90–91
dietary sources **1**:93
deficiency, and high intakes **1**:93–94
laboratory findings of biotin deficiency **1**:92
measurement of **1**:92
metabolism **2**:242–243
pharmacology **1**:94
requirements **1**:92–93
role of biotin in gene expression **1**:91–92
transport **1**:87
placental transport of biotin **1**:87
transport of biotin into central nervous system **1**:87
transport of biotin into human milk **1**:87
uptake by liver and most peripheral tissues **1**:87
vitamin **1**:86
biotinidase deficiency **1**:91
bipyridyl **2**:384–385
Bitot's spots (X1B) **1**:462
bladder cancer **2**:484
blood ethanol concentration **2**:256, **2**:258
effects of food on blood ethanol concentration **2**:257–258
factors affecting **2**:257
gender differences in **2**:257
blood glucose **3**:75–76
level **3**:235–236
response **2**:419–420
blood pressure **1**:54, **2**:311–312, **4**:63–64
infant feeding and **4**:59–60
salt intake and **2**:581–588
blood volume **3**:220
blood–brain barrier (BBB) **3**:119
body cell mass (BCM) **3**:28
body composition **3**:21, **3**:175
anthropometry **3**:23
applications during aging **3**:27–28
applications during growth **3**:25–27
BIA **3**:23
BMI **3**:23
determination **3**:21
dilution techniques **3**:22
DXA **3**:21–22
hydrodensitometry/air plethysmography **3**:22
infant feeding and **4**:60
models in **3**:24–25
MRI **3**:22–23
physiological application **3**:28–32
3DPS **3**:24
tissues and organs **3**:25
in vivo neutron activation **3**:23
whole-body counting **3**:22
body image and eating disorders **4**:145
body indirect calorimetry **2**:153–154
body mass **1**:182–183, **3**:613
body mass index (BMI) **1**:52–54, **3**:23, **3**:91, **3**:158, **3**:174, **4**:52–53, **4**:85, **4**:144
bone
disease in CF **3**:212
health **4**:147
minerals **4**:38
nutritional aspects of
health link **3**:83
importance of exercise to **3**:84–85
isoflavones and bone health **3**:83
nutrients adversely affecting bone health **3**:83
impact of physical activity on bone health **3**:84–85
protein intake and **3**:82
bone mineral content (BMC) **3**:21–22
bone mineral density (BMD) **3**:175, **3**:428–429
boron **1**:199, **1**:450–451
bowel obstruction **3**:142–143
brain and nervous system
design of nervous system **3**:119
energy substrates **3**:119
fat-soluble vitamins **3**:123
Brazil nut (*Bertholletia excelsa*) **2**:514
Brazilian national SFP, example of **4**:345
breakfast **4**:9
breast cancer **2**:488, **2**:682–683
breast cancer-resistance protein (BCRP) **2**:81
breastfeeding **3**:515–516, **3**:675
breast milk composition **4**:20–21
community **4**:25–26
global and community factors affecting breastfeeding **4**:24–26
health system **4**:24
maternal **4**:26
and maternal and infant health **4**:21–24
protection, promotion, and support **4**:26–27
rates and trends **4**:27–29
brite adipocytes recruited in response to cold exposure **2**:29
British Dietetic Association **4**:146
bromine **1**:451–452
bronchial asthma, eicosanoids and **2**:100
bronchiectasis **3**:475
brown adipose tissue **2**:122–123
activation of BAT by b-adrenergic receptor agonists **2**:33
activity in humans **2**:29–30
adaptive thermogenesis **2**:28
in response to cold exposure **2**:28–29
adipose tissue browning by peroxisome proliferator-activated receptor-g agonists **2**:34
alternative PET tracers to quantify **2**:30
assessment of brown adipocyte tissue mass and function **2**:34
batokines as therapeutic target **2**:34
brite adipocytes recruited in response to cold exposure **2**:29
by cold exposed 18F-FDG PET/CT scan **2**:30
contribution of human BAT to cold-induced thermogenesis **2**:31
discovered in rodents **2**:27
discovery of **2**:27–28
early investigations of human **2**:27–28
as endocrine organ **2**:31–32
glucose uptake may be independent of thermogenic capacity **2**:30
is metabolically significant **2**:31
metabolic significance of **2**:30–33
in obese individuals, dysfunction of **2**:30
pharmacological approaches to activation of **2**:33–34
physiology of **2**:28–29
potential novel batokines recently identified **2**:32–33
re-discovery of active BAT in live humans **2**:28
retrospective analysis of **2**:29–30
sub population of acetate-secreting brown adipocytes regulate thermogenesis **2**:33
as target to combat obesity **2**:26–27
as therapeutic target **2**:33–34
transplantation of **2**:35
Brucellosis **2**:372
brush border membrane **1**:379–380
bulimia nervosa (BN) **3**:305
associated psychopathology **3**:306
comorbidity **3**:306
developmental risk **3**:307
diagnostic criteria **3**:305
dietetic management **3**:308–309
etiology and maintenance **3**:307
family-based treatment **3**:310
food insecurity **3**:306
gender differences **3**:305–306
genetic risk **3**:307
health consequences of **3**:307–308
long-term prognosis **3**:311
medical consequences **3**:308
nutrition care process **3**:309
nutritional consequences **3**:307
personality risk **3**:307
pharmacotherapy **3**:310
prevalence of **3**:305–306
professional responsibilities **3**:308
psychological treatment **3**:309–310
racial/ethnic differences **3**:306
screening **3**:308
theories of BN maintenance **3**:307
treatment **3**:308
of underrepresented populations **3**:311
weight/shape overvaluation **3**:306
burn injury
increased energy expenditure **3**:525–526
pathophysiology **3**:525
of burn injury and metabolism **3**:523–524

requirements **3**:526
butter **2**:485

C

C/EBP homologous protein (CHOP) **2**:204
cadmium **1**:452, **2**:357–358
dietary exposure **2**:357
permissible intakes **2**:357
toxicity **2**:357–358
toxicokinetics **2**:357
caffeine **1**:96, **4**:218–219
caffeine-induced anxiety disorder **1**:102
caffeine-induced sleep disorder **1**:102
and cognitive performance **1**:98–99
dependence **1**:102–103
and food **1**:103
general concepts **1**:96
genetics **1**:99
intoxication **1**:102
mechanisms of action **1**:97–98
pharmacokinetics **1**:97
and physical exercise **1**:99
physiological and health effects of **1**:98
reinforcement **1**:100
subjective effects **1**:99
tolerance **1**:101–102
withdrawal **1**:100
calcium **1**:192, **2**:184, **3**:12, **3**:80–81, **3**:341, **3**:446, **3**:613–615, **3**:686, **4**:38, **4**:94–96, **4**:111, **4**:172, **4**:266
absorption **1**:107–112
balance **1**:107
bioavailability **1**:108–109
biologic functions of vitamin D on calcium metabolism **1**:526
dietary calcium intake **1**:106–107
functions **1**:106
health consequences
of calcium deficiency **1**:112
of calcium excess **1**:112–113
hormonal control **1**:110–111
influence of life stages **1**:111
losses **1**:109–110
and magnesium **3**:397
mechanisms **1**:107–109
metabolism **1**:107
nonskeletal **1**:112
nutrient interactions **1**:109
race and genetics **1**:112
skeletal **1**:112
sources **1**:106–107
caloric accretion in fetus **4**:72
caloric restriction (CR) **2**:181, **3**:640
and intermittent fasting **2**:45–46
calpain 10 (*CAPN10*) **3**:256–257
Campylobacter infections **2**:369–370
cancer **1**:69–72, **1**:255, **1**:354–355, **1**:418, **1**:500, **1**:513, **1**:538, **2**:270, **2**:410–411, **2**:488–489, **2**:633–635, **2**:649, **2**:657, **2**:680–681, **2**:683–684, **4**:62–63
alternative and complementary diets **3**:144
analytical studies **3**:149–151
animal bioassays **3**:131–132
beverages **3**:158–159
carbohydrate and **1**:140
carcinogenic substances
present in edible plants **3**:126–127
produced by fungi and microorganisms **3**:127–130
carcinogenicity tests **3**:131–133
carcinogens produced by food processing **3**:129–130
case–control studies **3**:149–150
chemoprevention trials **3**:159
children, teenagers, and young adults with cancer **3**:138
cohort studies **3**:149
contaminants present in food **3**:128–129
descriptive studies **3**:147–148
diet **3**:156–157
and prevention **3**:159–160
dietary associations with lung cancer **3**:157–159
dietary hypotheses and mechanisms **3**:156–157
epidemiological issues in study of diet and cancer **3**:151
evidence from descriptive studies **3**:148
fat **3**:158
guidelines **3**:137
intervention studies **3**:150
living with and beyond cancer **3**:143–144
lung cancer histopathology **3**:155
meat and fish **3**:159
mechanisms of carcinogenicity **3**:130–131
meta-analysis **3**:150–151
monitoring and control of hazards **3**:133–134
naturally occurring carcinogens **3**:126–127
nutritional interventions **3**:138–139
nutritional needs of specific patient groups **3**:138
observation versus experiment **3**:159–160
practical management of eating difficulties **3**:139
relations between diet and cancer **3**:152
resolved and unresolved issues **3**:151
risk factors **3**:155
short-term predictive tests **3**:132–133
sources of evidence linking diet and cancer **3**:147–151
special-exposure groups **3**:148
taste changes **3**:139
time trends within countries **3**:148
candidate-gene association studies (CGAS) **3**:256
carbamazepine **2**:87
carbohydrate response element (ChoRE) **2**:199
carbohydrate response element binding protein (ChREBP) **2**:198–199, **2**:199t
carbohydrates **2**:14–16, **2**:117–118, **2**:300, **2**:452, **2**:482–483, **2**:517, **3**:242, **3**:451, **3**:528–529, **4**:34–36, **4**:135, **4**:265
absorption **1**:128
allosteric enzyme regulation **1**:132
and cancer **1**:140
carbohydrates and gastrointestinal health **1**:140–141
and cardiovascular disease **1**:139–140
complementary treatment **3**:529
control of carbohydrate metabolism **1**:130–133
counting **3**:247
and dental caries **1**:140
directional shifts **1**:132
diseases of carbohydrate metabolism **1**:133–135
and energy metabolism **1**:128–129
factors affecting carbohydrate assimilation **1**:128
functions and types of **1**:115
glucose
from carbohydrate staple foods **2**:416
production by liver and kidneys **1**:130
glycogen synthesis and breakdown **1**:133
health effects of **1**:138–141
hormonal regulation **1**:130–132
malabsorption **1**:133
metabolism **3**:647–649
disorders of **3**:419–421
modulation of gene expression **2**:198–202
and nutrient density **1**:139
and obesity **1**:139
pathophysiology **3**:528
regulation of gene expression **1**:132–133
requirements **3**:529
transport **1**:128
and type 2 diabetes **1**:140
cardiometabolic system **3**:585
cardiorespiratory fitness **1**:358–359
cardiovascular disease (CVD) **1**:66–67, **1**:235, **1**:243, **1**:354, **1**:418–419, **1**:499–500, **1**:513, **2**:312, **2**:408–409, **2**:443–444, **2**:487–488, **2**:632–633, **2**:648–649, **2**:657, **2**:679, **3**:250
approaches for prevention and management of **3**:381–383
carbohydrate and **1**:139–140
epidemic of **4**:442
prevention **1**:298–299
and management of **3**:387–388
primary prevention **1**:67
risk **3**:381
secondary prevention **1**:67
cardiovascular outcomes, salt intake and **2**:588–593
cardiovascular system **3**:585, **4**:159
effects of alcohol on **2**:262
caregivers, children and adolescents, behaviors of **4**:343
carnitine **4**:111
carnosine, functions of **1**:43
carotenoids **1**:80, **2**:407
absorption **1**:150–151
biosynthesis **1**:146–147
carotenoids and health research **1**:160
chemical and physical properties **1**:147–148
digestion **1**:149

carotenoids (*continued*)
effects of carotenoids on health **1**:152
effects of storage and processing **1**:149
electrochemical properties **1**:148
hydrocarbon carotenoid **1**:157–159
physiology **1**:149–152
as pro-oxidants **1**:148–152
properties **1**:147–148
stereochemistry **1**:145–146
structure and classification **1**:145–146
tissue distribution **1**:152
transport **1**:151
xanthophylls **1**:158–160
caryopsis. *See* cereal grains
Cashews (*Anacardium occidentale* L.) **2**:514
catalase **2**:254
cataracts **1**:73–74, **1**:513
catechin **3**:637
catecholamines **1**:131–132
celiac disease **3**:268–269, **3**:288
associations **3**:166
clinical features **3**:165–166
complications **3**:168–169
diagnosis **3**:167–168
epidemiology **3**:164
genetics **3**:163–164
pathogenesiss **3**:162–163
pathology **3**:164–165
treatment **3**:168
cellular and molecular mechanisms **4**:55–57
cellular homeostasis **1**:492
cellular retinoid binding proteins **1**:474
cellulose **1**:123, **2**:417
central nervous system (CNS) **2**:86, **3**:172, **4**:161
effects of alcohol on **2**:259–260
cephalic phase insulin response (CPIR) **2**:625–626
cereal grains **2**:293–294
See also whole grains
and adverse health outcomes **2**:303–304
cereal species **2**:294t
nutrient composition of **2**:300–303
post-harvest storage of **2**:298
processing of **2**:298–300
production and consumption of **2**:296–297
structure of **2**:297–298
types of **2**:294–296
cerebral palsy (CP) **3**:171
alternative feeding routes **3**:180–181
associated disabilities/deficits **3**:172
body composition **3**:175
bone mineral density **3**:175
classification **3**:172–173, **3**:173f
coordinated services **3**:181
dental **3**:179
diet/feeding history **3**:176–177
diet/texture modifications **3**:180
energy needs **3**:175
etiology **3**:172
GI issues **3**:178
growth **3**:173–174
growth charts **3**:174
IBW **3**:175
interdisciplinary team **3**:172
intervention strategies **3**:179
mealtime behaviors **3**:179
medications **3**:179
muscle tone and positioning **3**:178
nutritional assessment **3**:172
oral motor
considerations **3**:180
dysfunction and dysphagia **3**:177–178
orthopedic surgeries **3**:178–179
pain and discomfort **3**:179
protein, nutrient and fluid needs **3**:176
Slaughter equations and corrections for children with **3**:176t
cerebral sensing mechanisms **1**:424
cheese **2**:485
chestnut (*Castanea sativa*) **2**:514
chickpeas **2**:432
childhood/children
bone minerals **4**:38
classification of feeding disorders in **3**:623
essential trace elements **4**:39
fat-soluble vitamins **4**:41
methodology for definition of nutrient requirements and recommended intakes **4**:34–38
obesity
clinical approaches to **3**:583–585
complications of **3**:585–586
critical periods in development of obesity **3**:582
dermatologic alterations **3**:586
diet and dietary change **3**:583
early feeding **3**:582
familial obesity **3**:583
long-term risks **3**:586
musculoskeletal involvement **3**:586
neurological aspects **3**:586
pharmacologic treatment **3**:588
physical activity **3**:583
psychosocial aspects **3**:586
risk factors for **3**:582–583
sleep **3**:583
socioeconomic and environmental deprivation **3**:583
surgical treatment **3**:588
treatment of **3**:587–589
prevention of obesity during **3**:589
water-soluble vitamins **4**:40
chloro-S-triazines **2**:385
chlorophenoxy acids **2**:385–386
cholera **3**:267
cholestatic diseases in adults **3**:457
cholesterol **1**:236–237, **1**:295–298, **1**:296f, **2**:20–21, **3**:363
absorption **1**:175–176
absorption, transport, and storage **1**:175–177
biliary cholesterol secretion **1**:178
biosynthesis **1**:177–178
dietary cholesterol and plasma cholesterol **1**:179
dietary sources **1**:179
exogenous cholesterol transport **1**:176
fecal excretion **1**:178
major dietary sources **1**:179
metabolic function **1**:178–179
metabolism **2**:311
and excretion **1**:177
regulation of synthesis **1**:177
steroid hormones **1**:178
tissue cholesterol synthesis **1**:177
tissue uptake and storage **1**:176–177
very low-density lipoprotein synthesis **1**:179
cholesterol ester transfer protein (CETP) **3**:369
choline **1**:162–163, **3**:121, **4**:139
assessment of choline intake and status **1**:167–168
dietary recommendations and choline adequacy **1**:165–167
emerging areas of investigation **1**:168–170
fatty acids and **3**:121
nutritional genomics of **1**:170–171
physiology and metabolism **1**:164–165
secondary chemical structure of **1**:162f
chromatin structure **2**:177–178
chromatin-modifying metabolites **2**:179–180
chromium **1**:181–182, **4**:174
body mass **1**:182–183
dietary sources **1**:183
digestion **1**:183
intake guidelines **1**:182
lack of symptoms of deficiency **1**:182
molecular mechanism(s) of action **1**:186–187
pharmacological effects in rodents **1**:185–186
supplementation and human health **1**:185
toxicity **1**:184–185
transport of **1**:183–184
chronic alcohol misuse on nutritional status
body weight and energy balance **2**:271
effects of **2**:271
chronic asthma, management of **3**:515
chronic diseases **2**:658, **4**:51
endpoints
challenges to consideration of **4**:319
updating DRI framework for **4**:320
inferring diet-disease relationships in **4**:319–320
vitamin D and **1**:536–539
whole grains and chronic disease risk **2**:675–690
chronic kidney disease **3**:251
chronic liver disease **3**:460
chronic lung disease of infancy (CLDI) **3**:475–476
chronic neuroglycopenia **3**:405
chronic obstructive pulmonary disease (COPD) **1**:73, **1**:539, **3**:464–468
chronic undernutrition **3**:500–501
role of fetal growth restriction **3**:500–501
stunting **3**:500
chronodisruption **2**:37–38
chrononutrition
chronodisruption **2**:38

circadian system and chronodisruption **2**:37–38
clock genes and body weight loss during obesity treatments **2**:40
mutations in experimental animals related to obesity and metabolic dysfunction **2**:39–43
organs and systems involved in **2**:45
pathologies **2**:40
timing of food intake
as external synchronizer of biological clock **2**:43–44
in nutrigenetics **2**:46–47
types of food timing interventions and effect on obesity and metabolic dysfunction **2**:45–46
chylomicrons **3**:367
remnants **3**:367
circadian system **2**:37–38
and nutrition **2**:43–47
work **2**:37–38
circular RNAs (circRNA) **2**:178
circulatory system **4**:159
classical fat balance concept **3**:319
Climate Change (CC) **4**:370
adaptation approaches for food systems **4**:375–376
climate resilience approaches **4**:375–377
food
accessibility **4**:372
affordability **4**:372
availability **4**:371–372
impacts of climate-led adversities
on environmental pathways **4**:373–375
on health, nutrition & food systems **4**:371–373
nutritional value of food **4**:372–373
risks of natural disasters by improving natural and man-made infrastructures **4**:376–377
clinical examination
medical history **3**:35
physical examination **3**:36–38
Clostridium botulinum **2**:366
clostridium perfringens **2**:366–367
coarse cereals **2**:294
cobalt **1**:197
coconut (*Cocos nucifera*) **2**:515
coefficient of variation (CV) **2**:575
cofactors **1**:190
deficiencies **3**:417–418
individual metal cofactors **1**:192–198
macro-*vs.* microminerals **1**:190
macrominerals **1**:192
metal-activated *vs.* metalloenzymes **1**:190–191
micro minerals **1**:192–198
nonmetal mineral cofactors **1**:198–199
coffee **2**:280–281
cognitive behavior therapy (CBT) **3**:299, **3**:309
cognitive function **2**:489–490
collecting duct (CD) **1**:424
colon **2**:310
disorders **3**:195–205
function **3**:185–194
histology **3**:187–188
metabolism **3**:194
structure **3**:185–188
colonic immune function **3**:192
colonic microbiome **3**:192–194
colonic microflora **3**:192
colonic motility **3**:191
colorectal cancer (CRC) **2**:313, **2**:488–489, **2**:681–682
commercial determinants of sugar consumption **3**:276–278
common cold **1**:513
Communication Function Classification System (CFCS) **3**:172
community **4**:25–26
comorbidities
control of **3**:236
dietary prevention and management of **3**:250–251
complementary feeding **3**:711, **4**:43
age of introduction of complementary foods **4**:43–45
allergenic food **4**:48
guiding principles for **4**:44t
meal frequency, energy density, and consistency of complementary foods **4**:47
nutrient needs
during period of **4**:47–48
from complementary foods **4**:45
preparation and feeding of **4**:48
complementary foods
age of introduction of **4**:43–45
meal frequency, energy density, and consistency of **4**:47
nutrient needs from **4**:45
preparation and feeding of **4**:48
complicated CD **3**:690
conditional cash transfer programs (CCTPs) **4**:332–334
conjunctival xerosis (X1A) **1**:462
constipation **2**:312, **3**:141, **3**:178
consumption trends **2**:630–632
sugar **2**:630–631
contaminants **2**:303, **2**:519–520
continuous glucose monitoring (CGM) **3**:235–236, **3**:255
cooking and meal preparation **3**:710
coordinated services **3**:181
copper **1**:195–196, **2**:359, **3**:13–15, **3**:616, **4**:39, **4**:174
absorption **1**:203–204
copper-containing hydroxylases **1**:507
deficiency **1**:206–207
dietary exposure **2**:359
dietary sources and reference values **1**:201–203
disorders of copper metabolism **3**:421–422
distribution **1**:204
excess **1**:207
excretion and homeostasis **1**:205
function **1**:201
inborn errors of metabolism **1**:205–206
measuring copper status **1**:206
permissible intakes **2**:359
toxicity **2**:359
toxicokinetics **2**:359
trafficking and kinetics **1**:203–205
corneal xerophthalmia (X2/X3) **1**:462–463
coronary heart disease (CHD) **1**:353, **2**:648–649, **3**:594, **4**:63–64, **4**:262
composite diets **4**:267–268
comprehensive prevention **4**:268
dietary patterns associated with coronary heart disease risk **4**:242–251
fructose consumption and **2**:394
global trends in CHD as reflection of nutrition transition **4**:262
links between nutrition and **4**:262
novel omics approaches for precision nutrition in CVD prevention **4**:252–253
nutrients and **4**:263–266
omic signatures of dietary patterns and **4**:252–253
specific foods **4**:266–267
timeline of research on diet and **4**:241–242
coronavirus disease 2019 (COVID-19) **2**:465, **4**:123–124
correction factor (cf) **2**:85
cost of food fortification **2**:342
cream **2**:485
creatine
functions of **1**:41
supplementation **3**:638
Crohn's disease **3**:202–203
Crohn's disease exclusion diet (CDED) **3**:691–692
targeted **3**:692
Cryptosporidium spp. **3**:543–544
curcumin **2**:184
cut point method **2**:574–575
cyanogens **3**:349
cycasin **3**:350
cyclooxygenase-2 (COX-2) **1**:230
cysteine **4**:110
cysteine-rich domain (CRD) **2**:625
cystic fibrosis (CF) **3**:207, **3**:468–473
appetite stimulants, growth hormones, and omega-3 **3**:472–473
assessment of nutritional status **3**:212
asthma **3**:473–475
body composition studies in CF **3**:212
CF-related diabetes mellitus **3**:209, **3**:212
clinical features **3**:208–209, **3**:468–469, **3**:473
decreased dietary intake **3**:210
decreased intake **3**:470
dietary management of CF **3**:211
dietary supplements **3**:211, **3**:470
enteral feeding **3**:211, **3**:470
epidemiology **3**:468, **3**:473
etiology **3**:207
fertility issues **3**:212
high-energy/high-protein diet **3**:470
increased energy
expenditure **3**:470
losses **3**:211

cystic fibrosis (CF) (*continued*)
requirements **3**:210
lung diseases **3**:475–476
malnutrition and CF **3**:469–470
nutrition and asthma **3**:474
oxidant/antioxidant imbalance in CF patients **3**:214
parenteral nutrition **3**:471
pathogenesis **3**:468, **3**:474
of lung disease **3**:208–209
possible preventative effects of nutrition **3**:474–475
prevalence **3**:207
prognosis **3**:208
secondary prophylactic effects of nutrition in asthma **3**:475
specific dietary considerations **3**:212
vitamin, mineral, and pancreatic enzyme supplementation **3**:471–472
cytokines
affect metabolism **2**:49–51
interaction of nutrition and cytokines **2**:53–56
role of cytokines in disease **2**:51–53
cytoplasmic peptidases **1**:379–380
cytosolic pathway **3**:437

D

dairy products **2**:484–485, **4**:267
growth and development **2**:486
overall mortality **2**:486
pregnancy, maternal health, and newborns **2**:486
role in disease prevention and health promotion **2**:486–488
weight gain **2**:486–487
dairy proteins **2**:487
data extraction and synthesis **1**:52
decarboxylation **1**:394–395
defecation **3**:191
degenerative arthritis, diet role in management and prevention of **3**:114
degradation of AAs in humans **1**:27–30
cell-and tissue-specific degradation of AAs in humans **1**:28–30
endothelial cells, smooth muscle cells, macrophages, and lymphocytes **1**:30
kidneys **1**:30
liver **1**:29
placentae, mammary glands, ovaries, testes, and pancreas **1**:30
sense organs **1**:30
skeletal muscle, heart, brain, lungs, and white adipose tissue **1**:29–30
small intestine **1**:30
dehydration **3**:216–217, **3**:220
algorithm for children with **3**:223f–224f
assessment of **3**:222t
groups at risk for **3**:221–222
management of **3**:222–225
pathophysiology of **3**:219–220
physiology of water balance **3**:217–219
types of **3**:222t, **3**:224
demethylation
of DNA **1**:511–512
of histones **1**:512
dental caries **2**:634, **3**:273–275
carbohydrates and **1**:140
etiology of **3**:274–275
prevalence and trends **3**:273–274
dental erosion **3**:276
etiology **3**:276
prevalence **3**:276
dental health, sucrose and **2**:601–602
deoxynivalenol (DON) **2**:497–498
desensitization **3**:343–344
developmental change of fetal body composition **4**:67–73
developmental origins of health and disease hypothesis (DOHaD) **3**:257–258, **4**:52
dextrins **2**:416
dextrose **3**:331
di-and tripeptides, absorption of **1**:380
diabetes **2**:487, **2**:649, **2**:657, **3**:253, **4**:64
diabetes-focused MNT **3**:236
effects of antioxidant supplementation on **1**:50–54
diabetes complications **3**:235
diabetes medications **3**:248–250
diabetes mellitus **1**:134–135, **3**:227–228, **3**:253
classification **3**:230–231
diabetes
medications **3**:248–250
subtypes **3**:245–248
diagnostic cut-points for diabetes diagnosis **3**:228–230
dietary approaches to **3**:238–241
dietary prevention and management of comorbidities **3**:250–251
effects of ingested nutrients **3**:242–245
GDM **3**:259–261
intermediate hyperglycemia/prediabetes **3**:228
MODY **3**:260–261
non-nutrient factors **3**:248–250
objectives in management of diabetes **3**:235–236
in pregnant women **3**:230
principles of dietary management of diabetes **3**:236–237
T1DM **3**:253–255
T2DM **3**:255–259
diabetes nutrition assessment, recommendations for **3**:237t
Diabetes Prevention Trial (DPP) **3**:255
diabetic diet **3**:236–237
diagnostic criteria
for intermediate hyperglycemia/prediabetes **3**:230
or diabetes in pregnant women **3**:230
diagnostic cut-points for diabetes diagnosis **3**:228–230
dialectical behavior therapy (DBT) **3**:300, **3**:309–310
diarrhea **3**:141–142, **3**:195–196, **3**:712
pathophysiology of **3**:264
diarrheal diseases
contribution of diarrhea to malnutrition **3**:264
diagnosis **3**:269
infectious causes of **3**:266–268
non-infectious causes **3**:268–269
pathophysiology of diarrhea **3**:264
prevention and treatment **3**:269–270
dichromate **1**:181–182
diet **2**:166, **2**:458, **3**:673–684, **3**:727–731
on behavior
breakfast **2**:13–14
carbohydrates, stress, mood, and mental function **2**:17
effects of dietary fat **2**:16–17
effects of meals **2**:12–17
essential fatty acids **2**:21
evening meal **2**:14
food deprivation **2**:21–23
functional and pharmacological components of foods and drinks **2**:23
glucose, mood, and mental function **2**:17–23
hyperactivity and antisocial behaviour **2**:18–20
meal composition **2**:14–17
meal size **2**:14
meal timing **2**:13–14
midday meal **2**:14
snacks **2**:14
sugars and opioids **2**:17
susceptibility to mood enhancement by diet **2**:17
beverages **2**:283
composition **3**:730
and appetite **2**:8
diet/feeding history **3**:176–177
diet/texture modifications **3**:180
and eating behavior **2**:57–66
history **3**:44
modulation of inflammation by diet and dietary components **2**:70–72
and oral health **3**:272–282
and physical activity behaviours **3**:608
pre-eclampsia and **4**:185–197
and RA activity **3**:111
religious customs, influence on
Buddhism **4**:339–340
Christianity **4**:337–338
Hinduism **4**:339
Islam **4**:338–339
Judaism **4**:336–337
role in management
of inflammatory arthritis **3**:111–114
and prevention of degenerative arthritis **3**:114
sweeteners **3**:244
as therapy for active disease in IBD **3**:688–692
diet-induced thermogenesis (DIT) **4**:165
diet-related oral diseases, epidemiology and etiology of **3**:273–276
dietary alkali/potassium consumption **3**:82–83

importance of acid–base homeostasis to health **3**:82
observational and intervention studies **3**:82–83
skeletal link to acid–base maintenance **3**:82
dietary and nutritional management of secondary undernutrition **3**:511–512
Dietary Approaches to Stop Hypertension diet (DASH diet) **3**:240, **4**:247–248, **4**:267, **4**:273
DASH-style dietary patterns **3**:396
dietary assessment measure, selection of **4**:387–389
database limitations **4**:389
household level **4**:387
individual level **4**:387–389
new technology **4**:389
dietary carbohydrate, type of glycemic index and **2**:424
dietary changes **4**:439–440
dietary cholesterol **3**:386–387, **4**:264–265
intake patterns **1**:179
dietary deficiency **1**:495
dietary fat **1**:292
amount of **3**:381
and cardiovascular disease prevention **1**:298–299
composition of **1**:299
and oils **1**:292, **1**:297–298
type of **3**:381–383
dietary fatty acids, supplementation with **3**:111–113
dietary fiber **1**:123–124, **1**:138, **1**:209–210, **1**:240, **2**:307, **2**:685–686, **3**:242, **3**:675–678
analytical methods **2**:307–308
classification of **1**:210–212
fermentation as biomarker **1**:212–213
fiber intake and recommendations **1**:214
health benefits of fiber, prebiotics, and probiotics **1**:214
health outcomes attributable to **2**:310–311
links between dietary fiber intake and health outcomes **1**:214
physiological effects **2**:308
in lower alimentary tract **2**:309–310
postbiotics **1**:214
prebiotic **1**:213
resistant starch **1**:212
in upper alimentary tract **2**:308–309
dietary guidance **1**:299
dietary guidelines
development of FBDG **4**:381–383
double burden of malnutrition diseases **4**:380
FBDG **4**:379–380, **4**:380f
types of guidelines **4**:381
Dietary Guidelines Advisory Committee (DGAC) **4**:380–381
Dietary Guidelines for Americans **3**:239
dietary inflammatory indexes **2**:75
dietary inflammatory potential **4**:248–250
dietary intake **3**:613
clinical practice **3**:45
computerized interview procedures **3**:46
data entry and nutrient calculation systems **3**:47–48
data processing
and computing dietary intake **3**:47
errors **3**:48
diet history **3**:44
using dietary methods in different populations **3**:46
duplicate sample technique **3**:45
estimated food quantities **3**:47
factors affecting individual ability to report intake accurately **3**:45
food
account method **3**:43
balance sheets **3**:42–43
frequency questionnaires **3**:44–45
household record **3**:43
individual dietary intake methods **3**:43–44
interviewers **3**:45–46
bias **3**:49
inventory method **3**:43
issues associated with measurement of dietary intake **3**:48
list recall methods **3**:43
measurement **3**:40
measurement error **3**:48
in data collection and processing **3**:48–49
impact **3**:49
methods for estimating dietary intake at household level **3**:43
methods for measuring food consumption at national level **3**:42–43
recall of remote diet **3**:46
reproducibility of dietary methods **3**:46
social desirability bias **3**:49
systematic bias **3**:49
24-hour recalls **3**:44
use of calibration methods to adjust for measurement error **3**:50
use of data and conversion of reported intake to nutrients and food types **3**:46
validation of dietary methods **3**:49–50
weighed food record inventory and estimated food record **3**:45
dietary interactions **2**:168–175
dietary intervention **2**:180–182
dietary issues in coronary heart disease prevention **4**:261–269
dietary lipids **2**:202, **3**:381–383
dietary management **3**:587
of diabetes **3**:236–237
of secondary overnutrition **3**:511
dietary modulation of inflammation **2**:67–78
dietary monitoring **2**:347–348
dietary nucleic acids in humans, metabolism of **1**:333
dietary nucleosides and nucleotides, beneficial effects of **1**:337
dietary patterns **3**:388
associated with coronary heart disease risk **4**:242–251
emergence of dietary pattern research **4**:242
and inflammation **2**:72
modeling **4**:273–274
omic signatures of CHD and **4**:252–253
theoretical dietary indices to capture "healthy" dietary patterns **4**:250
dietary protein allowances **2**:574–575
dietary recommendations **4**:92–94, **4**:122–123
and choline adequacy **1**:165–167
dietary reference intakes (DRIs) **1**:165–166, **1**:252, **3**:176, **3**:238, **4**:134
for infants **4**:134–141
dietary saturated fatty acids **1**:297–298
dietary sources **1**:136–137, **1**:149, **1**:183, **1**:252, **1**:562
and high intakes **1**:328–329, **1**:510–511
and nutritional equivalency **1**:478–479
units of nutritional activity **1**:478–479
physiology and **1**:61–75
protein quality and **2**:563–565
and reference values **1**:201–203
dietary sucrose and disease **2**:596–605
dietary supplements **2**:606–608
interactions **2**:609
label claims **2**:609–610
potential benefits of dietary supplements **2**:610
product quality and safety issues **2**:608–609
sales data derived **2**:606
dietary surveys
global research questions and data needs **4**:385–386
issues in survey design **4**:386–387
selection of dietary assessment measure **4**:387–389
diet–behavior relationship **2**:58–60
dietitians **3**:627–628
differential diagnosis **3**:407–408
differentially methylated regions (DMRs) **2**:180–181
digestibility, protein **2**:563
digestible indispensable amino acid score (DIAAS) **2**:562
digestion
for chromium **1**:183
of disaccharides **1**:127
of glucose and polymers **2**:418–419
of starch and related polymers **2**:418–419
of sugars **2**:418
of vitamins **2**:217–221
digestive system **3**:595
dilution techniques **3**:22
dioxins **4**:218
disaccharidases **2**:198
disaccharides **1**:118–121, **2**:414
glucose in **2**:414–416
diseases
associated with oxidative damage **1**:72–74
of carbohydrate metabolism **1**:133–135
disease-specific formulas **3**:559–561
diabetic **3**:559
hepatic **3**:560
immune enhancing **3**:560

diseases (*continued*)
pulmonary **3**:560
renal **3**:559–560
early origins of
environmental exposures **4**:59–64
infant feeding and adult health **4**:59
selected outcomes related to infant feeding **4**:59–60
postnatal growth and later risk of **4**:60–61
vitamin C and **1**:512–513
disorders of carbohydrate metabolism **3**:419–421
disorders of colon **3**:195–205
diarrhea **3**:195–196
IBD **3**:197–205
infections and enteric parasites **3**:196–197
PLE **3**:196
polyps **3**:197
disorders of copper metabolism **3**:421–422
disorders of fatty acid oxidation **3**:418–419
disorders of fructose metabolism **3**:420–421
disorders of hunger **2**:141–142
disorders of iron metabolism **3**:422
disorders of micronutrient metabolism **3**:421–422
disorders of protein metabolism **3**:414–418
dispensable amino-acids (DAAs) **2**:576
displaced persons and refugees **4**:393
displaced populations
addressing nutrition in refugees **4**:397
challenges **4**:398
conflicts and refugees and long-term health **4**:396
food baskets for populations **4**:397–398
global trends **4**:393
nutrition implications of displacement **4**:394–395
distal intestine, digestion and fermentation in **1**:382–383
dithiocarbamates (DTC) **2**:383
diverticular disease **2**:312
djenkolic acid **3**:349
DNA methylation **2**:177–178
DNA methyltransferase (DNMT) **2**:177–178
docosahexaenoic acid (DHA) **2**:462
dopamine, functions of **1**:43
doubly labeled water (DLW) **2**:159, **3**:53
estimated energy requirements determined using **1**:218
instrumentation **3**:56–57
methodology **3**:55–56
uses for **3**:57–60
Douglas bag method **2**:157
Down's syndrome
ageing **3**:288
anthropometric assessment **3**:285
carbohydrate metabolism **3**:284
care in community **3**:289
dental problems **3**:287
dietary assessment **3**:284
dietary guidelines **3**:286
dietary management **3**:286–287
feeding behaviour **3**:286–287
laboratory assessment **3**:284–285
nutritional requirements **3**:286
nutritional status **3**:284
nutritional therapy **3**:286
DPP-4 inhibitors **3**:249
dried cowpeas **2**:432
drinking milk **2**:484
drug induced hypoglycemia **3**:409
drug treatment in management of food allergy **3**:343
drug-induced alteration in nutrition **2**:86–88
drug-induced deficiency **1**:496
drug-nutrition interactions
drug-induced alteration in nutrition **2**:86–88
food–diet component **2**:85–86
food–meal **2**:82–83
mechanisms **2**:80–82, **2**:81t
medication use **2**:79
nutritional status on drug disposition **2**:85–86
recognition and clinical relevance of **2**:79–80
types of **2**:80t
drug–nutrient interactions **1**:446
dry beans **2**:431
dry peas **2**:432
dry pigeon peas **2**:432
dual-energy x-ray absorptiometry (DXA) **3**:21–22, **3**:175
dysgeusia/ageusia **3**:139
dyslipidemia **3**:250–251
fructose consumption and **2**:394
dysphagia **3**:139–140, **3**:177–178
Dysphagia Disorders Survey (DDS) **3**:178

E

4E-binding protein (4E-BP) **1**:391–392
early origins of disease
animal models **4**:53–55
cellular and molecular mechanisms **4**:55–57
chronic diseases **4**:51
epidemiological data **4**:51
evidence from human studies **4**:52–53
health and disease hypothesis **4**:52
eating
behavior **2**:141
disorders **3**:290–295, **3**:304–313
Eating and Drinking Ability Classification System (EDACS) **3**:172
Ecological Momentary Assessment (EMA) **3**:66
edible plants, carcinogenic substances present in **3**:126–127
eicosanoids
and arteriovenous duct persistence in neonates **2**:100
biological effects **2**:98–99
biosynthesis **2**:94–97
and bone resorption **2**:100
and bronchial asthma **2**:100
catabolism **2**:98
and diet **2**:101
effects on
bronchial tree **2**:99
cardiovascular system **2**:98
central nervous system **2**:99
endocrine system **2**:99
formed elements of blood **2**:98
gastrointestinal tract **2**:99
kidney **2**:99
metabolism **2**:99
uterine musculature **2**:98
and immune response **2**:100
inflammation and anti-inflammatory drugs **2**:99–100
mechanism of action **2**:97–98
pathophysiological implications **2**:99–100
regulation of platelet aggregation **2**:101–102
and reproduction **2**:100
and wound healing **2**:100
eicosapentaenoic acid (EPA) **2**:462
electrolytes **3**:332, **4**:111, **4**:174
abnormalities in acid–base balance **2**:109–111
acid and alkali load **2**:106
acids and bases **2**:105
blood/extracellular fluid **2**:106
buffers **2**:105–106
compensation **2**:111
definitions, acids, bases, and buffers **2**:105–106
diabetic ketoacidosis **2**:109–110
drugs and other causes **2**:111
effects of acid–base disturbance **2**:108–109
excess bicarbonate loss **2**:110–111
kidneys **2**:107–108
lactic acidosis **2**:110
liver and bone **2**:108
lungs **2**:106–107
maintenance of the pH of blood and extracellular fluids **2**:106
measurement of urinary acid excretion **2**:108
metabolic acidosis **2**:109
pH **2**:105
regulation **2**:106–108
renal failure **2**:111
supplementation, enteral nutrition **4**:114
transport **3**:188–189
transporter mechanism **2**:112
treatment **2**:111–112
elemental formulas **3**:559
elimination diets **3**:692
empirical dietary patterns **4**:251
endocannabinoids **2**:327
endocrine regulation in starvation **3**:653–654
endocrine system **3**:585, **4**:160
endogenous lipid pathways **3**:372
endogenous nucleic acids and excretion of metabolic end products, metabolism of **1**:332–333
endoplasmic reticulum (ER) **4**:56
energy **3**:342, **4**:34, **4**:135, **4**:169–170
adaptation **2**:128–129

autoregulation of body weight and body composition **2**:131–136
balance **3**:315–316
  basal metabolic rate and resting metabolic rate **3**:317–318
  dynamic and static evolution of **3**:321–322
  energy expenditure due to physical activity **3**:318
  energy imbalance and body weight relationship **3**:319–320
  gross and metabolizable energy **3**:316–317
  and macronutrient balance **3**:318–320
  model of **3**:316–317
  regulation **3**:320–322
  thermic effect of food or postprandial thermogenesis **3**:318
  total energy expenditure and components **3**:317–318
biological significance of a dual control system for **2**:133–134
compartmental model **2**:132–133
components of **1**:218
cost of activities **4**:166
energy adaptation
  and susceptibility to leanness and fatness **2**:135–136
  during longitudinal human study of large weight fluctuations **2**:134–135
estimated energy requirements determined using doubly labeled water method **1**:218
extra energy costs of growth, pregnancy and lactation **1**:221–222
intake on nitrogen balance **2**:561–562
metabolism **3**:194
  energy storage and fuel use in post-absorptive state **2**:123–125
  glycogen breakdown and gluconeogenesis provide glucose to body **2**:125
  pyruvate metabolism **2**:118
  uncoupling and brown adipose tissue **2**:122–123
needs **3**:175
requirements and metabolic adaptations
  basal metabolic rate **4**:164–165
  behavioral changes in physical activity **4**:166
  between-country comparison of metabolic costs of pregnancy **4**:166
  extra energy costs of pregnancy **4**:163–164
  fat deposition **4**:164
  implications of energy-sparing adaptations for mother and infant **4**:166–167
  individual variability in total energy costs of pregnancy **4**:166
  longitudinal studies of energy costs of pregnancy **4**:164–166
  maintenance energy costs of pregnancy **4**:163
  theoretical total metabolic costs of pregnancy **4**:164
  tissue deposition **4**:163
Entamoeba histolytica **3**:544
enteral nutrition (EN) **3**:325–329, **4**:111–114
  See also parenteral nutrition (PN)
  access **3**:326
  complications and monitoring **3**:329
  feeding
    delivery **4**:113
    route **4**:113
    selection **4**:113
  formula selection **3**:326–327
  home enteral nutrition **3**:329
  indications and contraindications **3**:325
  medication administration and tube occlusions **3**:327
  methods of infusion **3**:329
  modular formulas **3**:327
  nasoduodenal or nasojejunal route **3**:326
  nasogastric route **3**:326
  trophic feedings **4**:112–113
  types **3**:325–326
enteric nervous system (ENS) **3**:188–189, **3**:191
environmental enteric dysfunction (EED) **3**:712
environmental enteropathy **3**:712
environmental/infectious factors **3**:255
enzymes **3**:368
epidemiological data **4**:51
epidemiological studies **2**:653
epigallocatechin-3-gallate (EGCG) **2**:184
epigenetics **2**:177, **2**:180–182
  mechanisms **4**:55–56
  modifications **2**:177–178, **2**:178t
epigenomics **2**:177
erythritol **1**:121
Escherichia coli **2**:370–371
essential fructosuria **2**:395
essential trace elements **4**:39
Estimated Average Requirement (EAR) **1**:182
Estimated Safe and Adequate Dietary Intake (ESADDI) **1**:182
ethanol **2**:121
  models of ethanol pharmacokinetics **2**:259
  physical properties of **2**:252
ethnic groups **4**:226
ethyl alcohol **1**:443
eukaryotic translation initiation factor 2 (eIF2) **2**:204
European Food Safety Authority (EFSA) **1**:182, **4**:146–147
European Nicotinamide Diabetes Intervention Trial (ENDIT) **3**:255
excessive carotene intake **1**:480
exclusive enteral nutrition (EEN) **3**:688–691
  key points **3**:691
  use of EEN in UC **3**:690
excretion **1**:165, **1**:252–253, **1**:285–286
exercise-induced hypoglycemia **3**:409
exercising women **4**:226
exogenous lipid pathways **3**:371
external food environments **4**:342
extracellular fluid (ECF) **3**:217
extracellular matrix (ECM) **1**:390
extrahepatic fructose metabolism **2**:393
extraintestinal manifestations **3**:204

F

Fake Food Buffet **3**:63
faltering growth **4**:122
familial chylomicronemia **1**:305
familial combined hyperlipidemia (FCH) **1**:306
familial defective apo B-100 **1**:306
familial dysbetalipoproteinemia **1**:306
familial dyslipidemia **1**:306
familial hyperapobetalipoproteinemia **1**:307
familial hypercholesterolemia **1**:305–306
familial hypoalphalipoproteinemia **1**:307
familial lipoprotein **1**:307
famine
  armed conflict **4**:415
  causes of **4**:414–416
  climate and natural disasters **4**:416
  coping strategies **4**:416
  definition of **4**:413–414
  failure in central planning **4**:416
  intervention **4**:418
  market failure **4**:414–415
  modern **4**:412–413
  prevention **4**:416–417
  surveillance **4**:417
fasting plasma glucose **3**:228
  non-agreement of FPG and 2 h PG **3**:229
Fat Preference Questionnaire **3**:63
fat-free mass (FFM) **1**:390, **3**:22
fat-soluble vitamins **2**:224–230, **3**:451–452, **4**:41, **4**:138–139
  deficiencies of **2**:273
fats **2**:302, **2**:517, **3**:245, **3**:527, **3**:680–682, **4**:36–37, **4**:71–72, **4**:135–141
  catabolism **2**:118–120
  and cholesterol **3**:398
  metabolism **3**:650
  pathophysiology **3**:527
  requirements **3**:527
fatty acid binding protein (FABP) **2**:202, **3**:369
fatty acid synthase (FAS) **2**:199
fatty acid transporter (FAT) **2**:202
fatty acids **1**:292–293, **1**:293t, **2**:183, **3**:121, **3**:364, **3**:427
  See also trans-fatty acids (TFA)
  activation **2**:317
  α-oxidation **2**:322–323
  anti-inflammatory aspects **1**:230–231
  blood pressure **1**:237
  as components of complex lipids **2**:325–326
  de novo synthesis **2**:323–324
  disorders of fatty acid oxidation **3**:418–419
  eicosanoids **1**:234–235

fatty acids (*continued*)
metabolism and biological effects of omega-6 and omega-3 fatty acids **1**:224–226
elongation **2**:324–325
endothelial function, atherosclerosis, and cardiovascular disease **1**:234
fatty acids
and choline **3**:121
intake and eicosanoids **1**:235
genetic variants
in FADS1 and FADS2 and CHD risk **1**:228–229
FADS1 and FADS2 gene cluster influence omega-6 and omega-3 fatty acids c of omposition **1**:226–227
FADS1 and FADS2 in estimating nutritional requirements of omega-3 and omega-6 fatty acids **1**:226
in omega-6 and omega-3 fatty acids metabolism and IQ **1**:227–228
omega-6 fatty acids, and breast cancer in 5-LO **1**:229–230
role of omega-6 and omega-3 fatty acids in CHD in 5-LO **1**:229
genetic variation **1**:226–230
and inflammation **2**:72–73
metabolism **2**:328
regulation of **2**:328
n-6 fatty acids and risk of cardiovascular disease **1**:235–237
nomenclature conventions **2**:317
nutrigenomics **1**:230–231
ω-oxidation **2**:322–323
physical properties of **2**:317
profile of dietary fats **1**:297
structure, function, and nutritional requirements **1**:233–234
thrombosis **1**:236
unsaturation and essential FA **2**:325
fatty acylation of proteins **2**:327
Federal Trade Commission (FTC) **1**:182–183
feeding
alternative feeding routes **3**:180–181
children **3**:622–632
feeding/fasting cycle **3**:646–650
energy production **3**:646–650
energy requirements and metabolism **3**:646
formulas **3**:558–559
classification **3**:558–559
compositions **3**:559
fermentable carbohydrates **2**:685–686
fermentation and parboiling **2**:300
fermented milks **2**:485
fetal body composition, developmental change of **4**:67–73
fetal endocrine **4**:75–76
fetal growth
autocrine/paracrine-acting growth factor **4**:75–76
fetal endocrine **4**:75–76
regulation of **4**:73–75
fetal macrosomia **4**:74
fetal size, growth of **4**:67
fetal water content **4**:69
fiber **1**:242, **2**:399, **3**:387, **3**:397, **4**:265
content **2**:518–519
dietary guidelines **1**:240–241
health benefits of **1**:214
health impacts of **1**:243–244
intake **1**:242–243
first-pass metabolism
contribution of first-pass metabolism to overall ethanol metabolism **2**:259
of ethanol **2**:258–259
first-pass splanchnic extraction **1**:381–382
fish **4**:266
fish oil supplementation **3**:397
fistulizing disease **3**:690
flavanones **2**:533
flavin monooxygenase 3 (FMO3) **1**:165
flavones **2**:532
flavonoids **2**:531–533, **4**:265–266
flavonols **2**:532
"flooding large dose" measurement **1**:395
floods and droughts **4**:374–375
fluid needs **3**:176
fluid transport **3**:189–191
fluoride **4**:174
fluorine **1**:452
FODMAP **3**:678–680
folate/folic acid **2**:243–244, **3**:10–11, **3**:83, **3**:686, **4**:40, **4**:96, **4**:139, **4**:147, **4**:170–171, **4**:265
bioavailability **1**:252
biochemical functions **1**:248–250
chemical forms **1**:248
chemistry and biochemical functions **1**:248–251
deficiency **2**:272
deficiency and excess **1**:253–256
enzyme polymorphisms **1**:251
fortification beyond NTDs **1**:255–256
functions **2**:243–244
metabolism **2**:244
neural tube and birth defects **1**:255
neurological and cognitive dysfunction **1**:254
nutritional aspects **1**:252
regulation of folate metabolism **1**:251
removal of one-carbon units **1**:251
source of one-carbon units **1**:250–251
food **2**:60, **4**:271, **4**:298–299
allergy and intolerance **4**:125
amount of food consumed **4**:299
contaminants present in **3**:128–129
in cultural history **4**:271
food–diet component **2**:85–86
proportion of population consuming **4**:299
food account method **3**:43
food additives **3**:683–684
emulsifiers **3**:683
nanoparticles **3**:684
food allergies
additive effect of triggers **3**:341
anaphylactic shock danger **3**:341
concealing large doses is difficult **3**:340
cows' milk protein avoidance **3**:342
diagnostic tests **3**:338–341
dietary elimination **3**:341
effect of disease activity **3**:341
documenting diagnosis of food allergy **3**:338
documenting possible food allergies **3**:337
effect of dose **3**:340
egg avoidance **3**:342
gastric mucosal challenge **3**:341
high-risk factors **3**:342
inability to predict outcome **3**:338
intradermal testing **3**:339
management **3**:341–342
multiple mechanisms **3**:338
open and blind challenges **3**:340
oral mucosal challenge **3**:341
practical diagnostic difficulties **3**:338
problems with capsules **3**:340
provocation tests **3**:340–341
purpose of provocation tests **3**:340–341
rectal challenges **3**:341
role of dietitian **3**:342–343
route of administration **3**:340
skin application of food before food challenges **3**:339
skin prick tests **3**:338–339
soy avoidance **3**:342
special types of provocation testing **3**:341
tests for circulating IgE antibodies **3**:339
wheat-free and gluten-free **3**:343
Food and Drug Administration (FDA) **3**:244
food baskets for populations **4**:397–398
addressing acute malnutrition among young children **4**:397–398
addressing severe acute malnutrition **4**:398
management of moderate acute malnutrition **4**:397–398
food beliefs, diet, and health **4**:272
food choice **3**:62, **4**:271
assessment of **3**:63
individual's motives for **3**:63
liking, wanting, and motivation on **3**:64–65
in real world **3**:66
sensory-specific satiety **3**:64
taste as driver of **3**:64
techniques by food producers **3**:65–66
technology and innovation in **3**:66–67
Food Choice Process Model **3**:64
Food Choice Questionnaire (FCQ) **3**:63
Food Choice Task **3**:63
food composition data (FCD) **2**:330–333
component identification **2**:331–332
coverage of components **2**:332–333
coverage of foods **2**:331
FCT and FCDB generated, compiled and correctly **2**:337
food description, food group and food coding **2**:331
global availability of FCTs/FCDBs **2**:336
high-quality food composition data and databases **2**:336
sources of data for **2**:333–336
food composition databases (FCDB) **2**:330
generated, compiled and correctly **2**:337

global availability of **2**:336
food composition databases **2**:336
food composition tables (FCT) **2**:330
generated, compiled and correctly **2**:337
global availability of **2**:336
food culture **4**:271
evaluating folklore and **4**:278–279
folk beliefs **4**:279
food and nutrition science on **4**:272–273
modeling dietary patterns **4**:273–274
and nutrition science **4**:277–278
traditional and new folklore **4**:276t
traditional food culture and **4**:278
in USA **4**:274–276
food elements
elimination of **3**:355–356
few-foods or oligoallergenic diet **3**:356
food additives **3**:356
sucrose restriction **3**:355–356
supplementation of **3**:357–359
food environment **4**:283–284
dietary behaviors, and diet-related health outcomes **4**:284–286
interventions in **4**:286
measuring **4**:284
population determinants of diet **4**:282–283
food environments **4**:342–343
food folklore **4**:275–276
food fortification
compliance and enforcement **2**:345–347
dietary monitoring **2**:347–348
dilution factor **2**:343
fortification at point of consumption **2**:344
implement fortification program **2**:345
important in **2**:343
micronutrients in **2**:341–342
physical segregation **2**:343
programs
fortification compliance **4**:294
fortification potential **4**:298–299
fortification program performance **4**:294–297
foundational fortification documents **4**:289–293
health status before and after fortification **4**:297
legislation **4**:289
regulatory monitoring **4**:293
standards **4**:289–293
status of **4**:289–299
safety **2**:341–342
technological compatibility **2**:342
types of **2**:343–344
with zinc **1**:581–582
food frequency questionnaires **3**:44–45
food hedonics influence satiation and satiety **2**:62
food insecurity **2**:137
food intolerance **3**:345–346
11 β-hydroxysteroid dehydrogenase and liquorice **3**:348
caffeine **3**:346–347
enzyme defects **3**:346
examples **3**:350
food
allergy **3**:346
in excess harmful **3**:345–346
goitrogens **3**:349
hereditary fructose intolerance **3**:346
lactase deficiency **3**:346
lectins **3**:349
lupin alkaloid **3**:350
pharmacological mechanisms **3**:346–348
practical applications **3**:351
principal mechanisms and pathophysiology of food intolerance **3**:346–350
protease inhibitors **3**:349
pyrrolizidine derivatives **3**:350
specific drug–food combinations **3**:348
toxic mechanisms **3**:349–350
tyramine, histamine, and vasoactive amines **3**:347–348
tyramine and migraine **3**:348
vicine and convicine **3**:349
food labeling
approaches to nutrition labeling **4**:302–304
mandatory and voluntary labeling **4**:302
food labeling **1**:241
food on blood ethanol concentration, effects of **2**:257–258
food pattern modeling **4**:273
food processing **2**:653
food producers **3**:65–66
food programs **4**:343–345
historical development of school food programs **4**:344–345
food safety **2**:375–388
bacterial toxins **2**:364–367
heavy metals and metalloids **2**:351–361
organisms **2**:371–372
prevention of bacterial foodborne illness **2**:372–373
food security **4**:307
as component of livelihood security **4**:308–309
consequences of **4**:313–314
factors affecting food security **4**:312–313
framework and definitions **4**:307–309
indicators and classification of food security status **4**:309–312
common pitfalls of analyzing causes of malnutrition **4**:312
dietary diversity **4**:312
experience-based food insecurity scales **4**:311
household food insecurity access scale **4**:311
indicators of nutrient intake at national, household, and individual level **4**:309–311
meeting nutrient needs **4**:307–308
mitigating consequences of **4**:314
and sovereignty
acceptability **2**:644–645
affordability and accessibility **2**:643–644
affordability and accessibility at market **2**:643
availability **2**:642–643
availability of agricultural innovations for producers **2**:642–643
availability of AITFs in natural and built environment **2**:642
four pillars of **2**:642–646
nutrition **2**:644–645
pathways between production and food security **2**:644
food security and nutrition surveillance (FSN surveillance) **4**:421
case studies **4**:431–432
challenges informing policymaking and program planning for nutrition in LMICs **4**:421
conceptual frameworks **4**:422
constructs and indicators **4**:429–430
core principles and terminology **4**:422–423
data in LMICs **4**:423
routine data **4**:423
strengths and limitations **4**:428, **4**:430–431
survey data **4**:428
food sources **1**:242
food storage **3**:350
food supply chains **4**:342
food supply data **4**:445–448
food systems **4**:407–409
food-based dietary guidelines (FBDG) **4**:379
development of **4**:381–383
four-compartment model (4C model) **3**:25
free fatty acids (FFA) **2**:202
free sugars **1**:137
free *vs.* bound fructose metabolism **2**:393
fructo-oligosaccharides (FOS) **2**:554
fructose **1**:129
absorption **2**:392
chemical properties **2**:392
consumption **2**:391
disorders of fructose metabolism **3**:420–421
glucose metabolism *vs.* **2**:392–393
health implications **2**:394–395
intolerance and essential fructosuria **1**:133
metabolism **2**:392–393
inborn errors **2**:395–396
physical properties **2**:391–392
properties and sources of **2**:390–392
fructose 2,6-bisphosphate (F-2, 6-BP) **2**:201
fructose-1,6-bisphosphatease deficiency (FBPase deficiency) **2**:396
fruits **3**:157, **3**:519, **4**:266
carbohydrates **2**:399–400
classification of **2**:398
energy **2**:399
health effects of **2**:408–411
juice **2**:282
nutritional components of **2**:399–401
other components in **2**:404–405
water **2**:399
fumonisins **2**:497
functional fiber **1**:242
functional genomics **2**:197–198
functional magnetic resonance imaging (fMRI) **3**:66
fungi, carcinogenic substances produced by **3**:127–130
fungicides **2**:383–384

G

galactooligosaccharides (GOS) **2**:554
galactose **1**:129
galactose malabsorption **1**:133
galactosemia **3**:419–420
gallbladder disease **3**:595
gastro-jejunal feedings (GJ feedings) **3**:178
gastroesophageal reflux (GER) **3**:175
gastrointestinal complications **3**:209
gastrointestinal fistulas **3**:141
gastrointestinal function **3**:442–443, **3**:537
gastrointestinal health **2**:408
  carbohydrates and **1**:140–141
gastrointestinal issues (GI issues) **3**:178
gastrointestinal motility **3**:191
gastrointestinal system **3**:585
gastrointestinal tract (GIT) **2**:570, **3**:217
gastroparesis **3**:247
gene **3**:398
  expression
    arrays **1**:565–566
    modulation of **2**:198–202
generally recognized as safe (GRAS) **2**:631
genetics
  of ethanol metabolism **2**:256
  factors **4**:73
genistein **2**:184
genome-wide association studies (GWAS) **3**:256
geometric isomers **1**:292
germanium **1**:452–453
germination **2**:299
gestational age (SGA) **4**:116
gestational diabetes **1**:134, **4**:55
gestational diabetes mellitus (GDM) **3**:253, **3**:259–261
gestational weight gain (GWG) **3**:31–32
ghrelin **3**:594
Giardia lamblia **3**:543
global nutrition targets **3**:494–496
  progress toward achieving global nutrition targets **3**:496
Global School Health Survey **2**:631
glucagon **1**:131
glucocorticoid receptor (GR) **4**:55–56
glucocorticoids **1**:132, **4**:55
glucokinase (GK) **2**:199
gluconeogenesis **1**:130, **3**:74
glucosamine, functions of **1**:43
glucose **1**:128–129, **2**:413, **4**:109–110
  in body **3**:71–72
  body pool **3**:71–72
  breakdown
    to alcohol during fermentation reactions **3**:72
    storage and synthesis **3**:72
  from carbohydrate staple foods **2**:416
  homeostasis **1**:494, **2**:198, **3**:74–75
  malabsorption **1**:133
  metabolism **2**:410
    direct effects of sucrose on **2**:599
  and polymers
    blood glucose response and glycemic index **2**:419–420
    chemistry of **2**:413–418
    digestion of glucose and polymers **2**:418–419
  production by liver and kidneys **1**:130
  in urine **3**:77
glucose stimulated insulin secretion (GSIS) **4**:78
glucose utilization rate (GUR) **4**:78
glucose-6-phosphate (G6P) **2**:200–201
glucosinolates (GSLs) **2**:406
glutamate decarboxylase (GAD) **1**:493
glutamic acid decarboxylase autoantibody (GADA) **3**:253
glutamine **3**:529, **3**:560
  pathophysiology **3**:529
  requirements **3**:529
glutathione, functions of **1**:41
glutathione-S-transferases (GSTs) **2**:83
gluten **3**:678–680
  allergy **2**:303
glycated hemoglobin (HbA1c) **3**:229–230
glycemic index (GI) **2**:419–420, **2**:422–423, **3**:242–244
  amylose and amylopectin **2**:424
  between-individual variation **2**:425
  cell structure, food preparation, and processing **2**:424
  chain length and composition **2**:424
  clinical significance of postprandial hyperglycemia **2**:426–427
  concerns related to **2**:425–426
  mixed meals and nutrients **2**:425–426
  obesity and **2**:427
  pregnancy and glycemic index **2**:427
  and prevention of type 2 diabetes **2**:427
  problems arising from different methodologies used to calculate **2**:425
  proposed mechanism by dietary carbohydrates/glycemic index influence insulin resistance **2**:427–428
  relationship of insoluble and soluble NSP to **2**:424
  reproducibility **2**:425
  second meal effect **2**:426
  type of dietary carbohydrate and **2**:424
  within-subject variation **2**:425
glycemic load (GL) **3**:242–243
glycerophosphocholine (GPC) **1**:162–163
glycogen **1**:123, **2**:124–125, **3**:73–77, **4**:71
  breakdown **2**:125
  and health **3**:73–74
  storage diseases **1**:133–134
  storage disorders **3**:420
  synthesis **3**:73
    and breakdown **1**:133
  utilization **3**:73
glycogenolysis **1**:130
glycolysis **2**:117–118
glyoxylate metabolism **1**:494
gross morphology, colon **3**:185–186
Gross Motor Functional Classification System (GMFCS) **3**:172
growth
  charts **3**:174
  and development **4**:67
    autocrine/paracrine-acting growth factor **4**:75–76
    developmental change of fetal body composition **4**:67–73
    fetal and neonatal morbidity and mortality in IUGR **4**:79t
    fetal endocrine **4**:75–76
    growth of fetal size **4**:67
    interpretation of growth curves **4**:76–77
    intrauterine growth restriction and macrosomia **4**:77–80
    regulation of fetal growth **4**:73–75
  hormone **3**:593
  monitoring and promotion **4**:84
    anthropometric indices **4**:87
    challenges in **4**:89
    cutoff points of anthropometric measurements **4**:87–89
    growth standards and references **4**:85
    indicators of anthropometric measurements **4**:87–89
    interpretation of anthropometric measurements **4**:87–89
    measurement of growth **4**:85
    objectives and activities **4**:84
    use and interpretation of growth measurements **4**:85
  standards and references **4**:85
  surveillance **4**:84
  velocity **3**:92
gut health **1**:244
gut microbiome **1**:243–244, **2**:162–166
gut microbiota **2**:74–75, **2**:172–173
  and cardiovascular disease **2**:165
  diet and **2**:166
  manipulating gut microbiota through diet **2**:173–174
  and obesity **2**:163–165
  personalized nutrition **2**:166–167
  and type 2 diabetes **2**:163–165
gut microflora
  amino acid
    catabolism by **1**:382
    synthesis by **1**:382–383
gut signaling **2**:628

H

harmonizing approach to deriving nutrient requirements
  challenges to consideration of chronic disease endpoints **4**:319
  conceptual basis for nutrient reference values **4**:317–319
  current efforts to realize goal of harmonization **4**:321–324
  harmonizing derivation of nutrient reference values **4**:321–324
  origin of harmonization effort **4**:321
  rationale for **4**:324–325
  risk assessment framework **4**:317–319
hazelnut (*Corylus avellana*) **2**:515

HbA_{1c} % **1**:52–54
head circumference **3**:89
head circumference-for-age **3**:92
health and disease hypothesis **4**:52
health claims **1**:242
health inequities
  access to nutritious foods **4**:333
  health care
    access inequities **4**:331–332
    reform **4**:334
  health inequities **4**:331–332
  Household food insecurity inequities **4**:331
  life expectancy inequities **4**:332
  malnutrition inequities **4**:328–331
  nutrition and health inequities epidemiology **4**:328–334
  obesity inequities **4**:331
  physical activity **4**:334
  reduce discrimination **4**:334
  socio-ecological model of health **4**:332
  solutions **4**:332–334
  understanding determinants of health and nutrition inequities **4**:332
  unhealthy foods taxes and regulation **4**:333
health system **4**:24
Healthy Eating Index (HEI) **4**:53, **4**:273
Healthy Mediterranean lifestyle **2**:464
heart **2**:270
heat illness **3**:220
heat treatment **2**:300
hedonic processes of appetite control **2**:60–62
hematological aging **4**:160
heme, functions of **1**:44
hemicelluloses **2**:417
hemochromatosis **3**:422
hemoglobin A1c (HbA1c) **3**:235–236
hepatic disease **3**:595
hepatic failure **3**:409–410
hepatic fructose metabolism **2**:392
hepatic injury **3**:334
hepatic lipase (HL) **3**:369
hepatic lipid metabolism, direct effects of sucrose on **2**:600
hepatic regulation **1**:424
hepatobiliary disorders **3**:454–456
hepatocellular diseases **3**:452
hepatocyte nuclear factor 4 (HNF4α) **2**:201
herbal **3**:241
herbicides **2**:384–386
hereditary fructose intolerance **2**:395
hereditary fructosuria **2**:395
heterogeneity **3**:257
hexavalent chromium (Cr^{6+}) **1**:181–182
high birth weight (HBW) **3**:257–258
high fructose corn syrup (HFCS) **2**:390–391
high sugar diet **2**:631
high-density lipoprotein (HDL) **1**:296, **3**:368
high-fat diets (HFD) **2**:180–181
high-throughput sequencing technologies **2**:170
histamine, functions of **1**:42
histones **2**:177
HMB **3**:636–637
HMG CoA reductase **3**:369
homeostatic model assessment of insulin resistance (HOMA-IR) **4**:52–53
homocysteine **3**:618
  biosynthesis and metabolism **1**:260–261
  and clinical outcomes **1**:263–264
  measurement of blood concentrations **1**:265–266
  regulation of metabolism **1**:261
  structure and forms **1**:259
honeymoon period **3**:253
hookworms **3**:543
hormonal control of secretory activation **4**:105–106
hormones **3**:74–75
household budget surveys **3**:43
Household Water Insecurity Experiences Scale (HWISE Scale) **3**:708
HT-2 toxin **2**:498
human health **1**:185
  legumes in human health and nutrition
    classification and common food uses **2**:431–432
    global trends of production and economic importance **2**:433–434
    legume proteins/flours in emerging markets **2**:435–436
    nutritional value **2**:432–433
  selenium and **1**:418–419
    immune function **1**:418
human immunodeficiency virus (HIV) **1**:541, **3**:267, **3**:535–536
  HIV/AIDS **3**:535–536
    calories and macronutrients **3**:536
    effects of HIV/AIDS on nutritional status **3**:536–537
    infant feeding and **3**:539
    interactions between antiretroviral drugs and nutrition **3**:538–539
    macronutrient intervention studies **3**:537
    micronutrient intervention studies **3**:537–538
    micronutrients **3**:537
    nutritional intervention studies **3**:537–538
    social and economic interactions with HIV/AIDS **3**:539–540
human microbiota of gastrointestinal tract **2**:162
human milk oligosaccharides (HMOS) **2**:554
human milk production and feeding **3**:711
human placenta **4**:179
human studies **4**:52–53
human zinc deficiency **1**:590
humans **2**:58–60
hunger **2**:137
  disorders of **2**:141–142
  and eating behavior **2**:141
  by energetic needs **2**:139
  learning and **2**:140
  measurement of **2**:138
  physiological determinants of **2**:139–140
  and satiety **2**:138–139
hydration and thirst
  assessment of thirst **2**:143–144
  fluid requirements **2**:150
  mechanisms for terminating the sensation of thirst **2**:149–150
  mechanisms of thirst regulation **2**:146–147
  osmotic regulation of thirst **2**:147–148
  perception of thirst **2**:143
  physiological regulation of thirst **2**:144–145
  role of thirst in water balance **2**:143
  sensory regulation of thirst **2**:149
  volemic regulation of thirst **2**:148–149
hydrodensitometry **3**:22
3-hydroxy 3-methylglutaryl CoA (HMGCoA) **1**:296
4-hydroxyproline, functions of **1**:43
hyperactivity
  dietary patterns in ADHD **3**:355
  elimination of food elements **3**:355–356
  methodological issues **3**:354–355
  research on nutrition in ADHD **3**:359
  supplementation of food elements **3**:357–359
hypercaloric diets **2**:180–181
hypercholesterolemia, risks associated with **3**:373
hyperhomocysteinemia **1**:254, **1**:261–263
hyperlipidemia
  categories and classifications of **3**:374
  causes of **3**:375–376
  impact of diet on CVD risk **3**:381
  dietary cholesterol **3**:386–387
  dietary lipids **3**:381–383
  dietary management of **3**:377
  dietary recommendations for reduction of atherogenic cholesterol **3**:377
  genetics **3**:375
  healthy dietary patterns **3**:377
  lifestyle **3**:375–376
  prevention and management of CVD **3**:387–388
  secondary causes **3**:376
  SFA **3**:383
  unsaturated fatty acids **3**:383–386
Hyperosmolar Hyperglycemic Syndrome **3**:235
hyperphosphatemia and consequences **1**:349–350
hypertension **3**:251, **3**:594
  children **3**:398
  diet interactions **3**:398
  dietary factors lower BP **3**:393–397
  dietary factors with limited or uncertain effect on BP **3**:397–398
  effects of antioxidant supplementation on **1**:50–54
  increased potassium intake **3**:395–396
  moderation of alcohol intake **3**:396
  older people **3**:398
  populations defined by race/ethnicity or geography **3**:398–399
  pregnant women **3**:398
  reduced salt intake **3**:393–395
  special populations **3**:398–399
  weight loss **3**:393

hypertension (*continued*)
whole dietary patterns **3**:396
hypertensive disorders of pregnancy (HDPs) **4**:185
classification of **4**:186–187
early detection of **4**:190
hyperthermia **3**:220
hypertriglyceridemia **1**:306
dietary management of **3**:377
risks associated with **3**:373
hypervitaminosis A **1**:476, **1**:480
hypoglycemia **2**:86, **3**:249
blood glucose concentration **3**:402
brain function and **3**:402
brain malfunction from **3**:404
causes of **3**:405
diagnosis **3**:405–410
in elderly sick **3**:410
glucose pool in fasting subjects **3**:402
hypoglycemic syndromes **3**:404
insulin release in response to eating and fasting **3**:402–403
mechanism of **3**:402–404
neuroglycopenic syndromes **3**:404–405
role of liver in glucose homeostasis **3**:403–404
hypoglycemic syndromes **3**:404
hypomethylation **4**:55–56
hypophosphatemia and consequences **1**:350
hypothalamic–pituitary–adrenal axis (HPA axis) **3**:593
hypoxia-inducible factor 1 (HIF-1) **2**:201

I

IA-2 antigen (IA-2A) **3**:253
ideal body weight (IBW) **3**:175
idiopathic intrauterine growth restriction (IUGR) **4**:54
immersive virtual reality (iVR) **3**:67
immune function **1**:74, **1**:589
immune system, vitamin D and **1**:540
immunological aging **4**:160–161
in vitro studies **1**:186
in vivo neutron activation **3**:23
in vivo studies **1**:186–187
inborn errors of metabolism **3**:410, **3**:412–431
inclusion criteria and search strategy for meta-analysis **1**:50
indirect calorimetry
clinical cases **2**:160
field methods **2**:157–158
interpretation of **2**:159
labeled bicarbonate **2**:158–159
laboratory methods **2**:153–158
metabolic carts **2**:155–157
tracer methods of **2**:158–159
indispensable amino-acids (IAAs) **2**:569, **2**:576
individual dietary intake methods **3**:43–44
individualization **3**:236–237
inductive approach **4**:272
infant botulism **2**:366
infant feeding **4**:59
adiposity and obesity **4**:62
autoimmune diseases **4**:60
and blood pressure **4**:59–60
coronary heart disease **4**:63–64
and body composition **4**:60
metabolism **4**:62
infant feeding and obesity **4**:60
insulin resistance and diabetes **4**:64
postnatal growth and later risk of disease **4**:60–61
selected outcomes related to **4**:59–60
infantile beriberi **1**:433
infantile cholestatic disorders **3**:454–456
infants
dietary guidelines/recommendations for **4**:139–141
nutrition guidance for **4**:133–142
inflammation **1**:236, **2**:74–75
antioxidant micronutrients and **2**:74
dietary components could influence inflammation **2**:70–71
dietary patterns and **2**:72
direct effects of sucrose on **2**:600
and disease **2**:69
drugs **2**:99–100
fatty acids and **2**:72–73
importance of resolution of inflammation **2**:70
inflammatory response **2**:68–69
modulation by diet and dietary components **2**:70–72
polyphenols and **2**:74
specific foods and **2**:72
inflammatory arthritis
diet and RA activity **3**:111
fasting and vegetarian diets **3**:113–114
nutritional assessment in RA **3**:111
role of diet in management of **3**:111–114
supplementation with dietary fatty acids **3**:111–113
vitamin, mineral supplementation and probiotics **3**:113
inflammatory bowel disease (IBD) **1**:535, **1**:539, **3**:193, **3**:197–205, **3**:268, **3**:482, **3**:671–672
diet as therapy for active disease in **3**:688–692
environmental influences **3**:200–201, **3**:673
epidemiology **3**:197
etiology **3**:197
genetic factors in **3**:672–673
genetics **3**:199
immunological abnormalities **3**:673
malnutrition assessment and management considerations in **3**:684–688
microbiome in **3**:673
nutritional consequences of **3**:205
pathogenesis **3**:201–202, **3**:673–684
pathogenic factors in **3**:672–673
pathology **3**:202–203
treatment of **3**:205
ingested nutrients, effects of **3**:242–245
inhibition of nitrosamine formation **1**:508
innervation, colon **3**:186
inorganic chemicals **3**:128–129
insecticides **2**:378–383
insensible sweat **3**:218
insoluble non-starch polysaccharides (NSPs) **2**:424
Institute of Medicine **1**:182
insulin **1**:131, **3**:75–76, **3**:242, **3**:248–249, **4**:75
gene expression **2**:198
resistance **2**:487, **4**:64
fructose consumption and **2**:394
and hyperinsulinemia **3**:592–593
sensitivity **2**:649
synthesis in pancreatic β-cells **2**:198
insulin receptor substrate-1 (IRS-1) **1**:186–187
insulin-like growth factor 1 (IGF-1) **2**:204, **4**:52, **4**:56
insulin-like growth factor 2 (*IGF2*) **4**:55–56
insulin-like growth factor binding protein 1 (IGFBP1) **2**:204
insulitis **3**:253
integrative cognitive-affective therapy-BN (ICAT-BN) **3**:309–310
interdisciplinary team **3**:172
intermediate-density lipoproteins **3**:368
intermittent fasting **2**:181, **3**:241, **3**:730–731
intermuscular adipose tissue (IMAT) **3**:25
International Classification of Diseases (ICD-10) **4**:145
International Diabetes Federation **3**:253
International food security phase classification (IPC) **4**:309
international normalized ratio (INR) **2**:87
interpersonal psychotherapy (IPT) **3**:299–300, **3**:309–310
interpretation **3**:502
overweight **3**:502
stunting **3**:502
underweight **3**:502
wasting **3**:502
interstitial fluid (ISF) **3**:217
intervention strategies **3**:179
intestinal absorption **1**:86
contribution of microbial biotin to absorbed biotin **1**:86
intestinal failure **3**:142, **3**:269
intestinal lumen, protein digestion in **1**:376–379
intestinal metabolism **1**:474–475
intestinal microbiome, sucrose and **2**:602
intestinal microbiota
metabolic functions **2**:170–171
protective and structural functions **2**:170
roles of **2**:170–171
intestinal regulation **1**:424
intestinal tuberculosis **3**:268
intracellular fluid (ICF) **3**:217
intrarenal mechanisms **1**:424
intrauterine growth restriction (IUGR) **3**:257–258, **4**:67, **4**:78, **4**:116–117
and macrosomia **4**:77–80
preterm delivery and **4**:118

intravascular baroreceptors **1**:424
invasive bacteria **2**:368–371
iodine **3**:15–17, **3**:342, **4**:39, **4**:146–147, **4**:174
  absorption and metabolism **1**:274–275
  adaptations of thyroid metabolism to **1**:275–276
  assessment **1**:278–279
    of iodine status **1**:270–271
  deficiency disorders **1**:269–270
  dietary and environmental factors **1**:277–278
  dietary sources, absorption, and metabolism **1**:268–269
  ecology of **1**:274
  excess and toxicity **1**:272
  micronutrients impact on thyroid and **1**:276
  prophylaxis and treatment of iodine deficiency **1**:271–272
  requirements **1**:270
    and dietary sources of **1**:276–277
ion channels **3**:188–189
iron **1**:192–194, **3**:12, **3**:686, **4**:39, **4**:173
  absorption, excretion, transport, and storage **1**:283–287
  bioavailability **1**:289
  body content, forms, and function **1**:282
  body iron balance **1**:286–287
  chemistry and physiology **1**:282–287
  clinical assessment of iron status **1**:287
  deficiency **1**:287, **2**:273, **4**:124–125, **4**:146–147
  dietary **1**:288–290
  disorders of iron metabolism **3**:422
  enriched crops **2**:290
  excessive Intakes **1**:290
  food sources **1**:288
  heme iron **1**:289
  iron-containing hydroxylases **1**:507–508
  nutrition **1**:287–288
    cognitive development **1**:287–288
    consequences of iron deficiency **1**:287–288
    physical work capacity **1**:287
    recommended dietary intakes **1**:288
    reproduction **1**:288
  supplementation and fortification **1**:289–290
iron deficiency anemia (IDA) **4**:146
iron regulatory proteins (IRPs) **2**:207
iron responsive elements (IREs) **2**:207
irritable bowel syndrome (IBS) **2**:312–313
irritant mechanisms **3**:348
isoflavones **2**:533
  sand bone health **3**:83
isomers **1**:292

J

Japanese diet **4**:268
jejuna extension tube (JET) **3**:326

K

kallikrein-kinin system **1**:425
kernel. *See* cereal grains
ketogenic diets **1**:141
ketosis **3**:240, **3**:437–439
  extrahepatic regulation **3**:433
  formation of ketone bodies **3**:432–436
  function of ketone bodies **3**:436–437
  intrahepatic regulation **3**:433–436
  intramitochondrial regulation **3**:435–436
  metabolic acidosis **3**:439–440
  pathological ketosis **3**:438–439
  pathways of ketone body utilization **3**:436–437
  physiological ketosis **3**:437–438
  role of malonyl-CoA **3**:434–435
kidneys, glucose production by **1**:130
kinetics of ethanol elimination in vivo **2**:256
knee height (KH) **3**:174
Kofranyi–Michaels respirometer
konzo disease **3**:710
Kruppel-like factor 6 (KLF-6) **2**:201
kwashiorkor **3**:498–499
kynurenic acid, functions of **1**:44

L

laboratory methods **2**:153–158
laboratory-based studies **3**:64
lactasefloricin hydrolase (LPH) **2**:198
lactation
  changes in milk composition during secretory activation **4**:104–105
  delays in secretory activation **4**:106
  exocytotic pathway **4**:101
  functional anatomy of lactation **4**:99
  hormonal control of secretory activation **4**:105–106
  initiation of lactation **4**:104
  local control of synthesis and secretion **4**:103
  milk composition **4**:100
  minerals and vitamins **4**:94–96
  paracellular transport pathway **4**:103
  rationale for recommended nutrient intakes **4**:91–92
  regulation
    of milk ejection **4**:103–104
    of milk synthesis, secretion, and ejection **4**:103–104
  requirements and dietary recommendations **4**:92–94
  synthesis and secretion of milk components **4**:100–103
  transcytosis pathway **4**:102
  transmembrane pathway **4**:102–103
lactose **1**:118
  deficiency or lactose intolerance **1**:133
  digestion **3**:442–443
  intolerance **3**:479
    adults **3**:444
    children **3**:443–444
    historical and geographic perspective **3**:441–442
    lactase nonpersistence **3**:442
    lactose digestion and diet **3**:446
    nutrition policy **3**:446
    pregnant women **3**:445
    prevalence **3**:443–445
large for gestational age (LGA) **4**:76
lathyrogens **3**:349
LDL receptor (LDLR) **3**:370
LDLR-related protein (LRP) **3**:370
lead **1**:453, **2**:353–356
  dietary exposure **2**:353–354
  permissible intakes **2**:354
  toxicity **2**:355–356
  toxicokinetics **2**:354
lean body mass **4**:69–71
learning and hunger **2**:140
lecithin cholesterol acyltransferase (LCAT) **1**:296, **3**:369
legume
  processing and antinutritional factors **2**:436
  proteins/flours in emerging markets **2**:435–436
leisure-time *vs.* occupational physical activity **1**:355
lentils **2**:432
leptin (*LEP*) **2**:180, **3**:594
leukotrienes **1**:235
  and related compounds **2**:93–103
linear growth faltering **4**:231
linoleic acid (n-6) **4**:136
lipidomics **1**:565–566
lipids **1**:292–296, **2**:20, **2**:400, **2**:452, **2**:481–482, **2**:649, **3**:451, **4**:110–111
  cholesterol **1**:295–296
  emulsions **3**:331–332
  fatty acids **1**:292–293
  metabolism **3**:285
  modulation of gene expression by **2**:202–204
  phospholipid **1**:295
  secretion pathway **4**:101–102
  triacylglycerol **1**:294–295
lipogenesis **2**:201–202
lipogenic enzymes:hepatic pyruvate kinase (L-PK) **2**:199
lipoprotein lipase (LPL) **3**:370
lipoprotein(a) **3**:368
lipoproteins **1**:236–237, **2**:649, **3**:367
  classification **1**:301–302
    of serum lipoproteins according to apolipoprotein composition **1**:302
    of serum lipoproteins according to electrophoretic mobilities **1**:301–302
    of serum lipoproteins according to ultracentrifugal characteristics **1**:302
  disorders of lipoprotein metabolism **1**:305–307
  effects of diet on
    chylomicron metabolism **1**:304
    HDL metabolism **1**:305
    LDL metabolism **1**:305
    VLDL metabolism **1**:304

lipoproteins (*continued*)
effects of dietary fats and cholesterol on lipoprotein metabolism **1**:304–305
general guidelines for treatment of lipoprotein abnormalities for CHD prevention **1**:307
metabolism of lipoproteins carrying exogenous lipids **1**:302–303
receptors **3**:370
reverse cholesterol transport **1**:304
synthesis and catabolism of lipoproteins **1**:302–304
transport of endogenous lipids **1**:303
transporters **3**:370
Listeria monocytogenes **2**:371
lithium **1**:453
liver
cancer **2**:484
disease **3**:209, **3**:212, **3**:459
nutritional aspects and liver physiology in normal liver and **3**:450–451
disorders **3**:449–462
effects of alcohol on liver function **2**:262
glucose production by **1**:130
regulation of lipogenesis **2**:201–202
in specific hepatobiliary disorders and nutritional management **3**:452
liver X receptor (LXR) **2**:201
long non-coding RNAs (lncRNA) **2**:178
long-chain polyunsaturated fatty acids (LCPUFA) **3**:517–518
long-term weight loss maintenance, prevalence of **3**:725
low birth weight (LBW) **3**:257–258
and preterm infants **4**:116
causes of **4**:117
epidemiology of **4**:118–119, **4**:118f–119f
fetal growth, assessment of **4**:116–117
gestational age, assessment of **4**:116–117
health consequences of **4**:117
nutritional interventions to prevent **4**:119
prevention of **4**:118–119
low carbohydrate **1**:141, **3**:240
low glycemic index carbohydrates
benefits of low glycemic index carbohydrates on diabetes control **2**:426
on CVD risk factors **2**:427
low-and middle-income countries (LMICs) **4**:231
low-calorie sweeteners (LCS) **2**:625–626, **2**:628–629, **2**:631–632, **2**:634–635
low-density lipoproteins **3**:368
low-fat and very low-fat **3**:240–241
low-molecular-weight chromium-binding substance (LMWCr) **1**:183–184
lower alimentary tract, physiological effects in **2**:309–310
lung diseases
clinical features **3**:464
differential diagnoses **3**:464
epidemiology **3**:464
etiology **3**:464
nutrition and COPD **3**:465–468
nutritional advice, exercise, and supplementation **3**:467
nutritional support **3**:466–467
obesity in COPD **3**:466
pathology **3**:465
reasons for malnutrition in COPD **3**:466
tube feeding **3**:467–468
type of nutritional support **3**:467
lupines **2**:431
lutein **1**:159–160
lycopene **1**:158–159, **2**:443–444
and related compounds
chemical and physical properties of lycopene **2**:438–441
disease-related investigations **2**:444
evidence for possible anticancer protection by lycopene **2**:443
functional properties and tissue health **2**:442
health, research models, and epidemiological evidence **2**:442–444
tissue contents and kinetics of lycopene turnover **2**:441
lysophosphatidylcholine (LPC) **1**:164
lysosomal-autophagic system (LAS) **1**:388–389, **1**:392

M

macadamia nut (*Macadamia integrifolia*) **2**:515
macrominerals **4**:137
macronutrients **3**:238, **4**:92–94, **4**:135–141, **4**:152, **4**:395
balance **3**:318–320
content of nuts and seeds **2**:516–517
deficiencies **3**:684
energy **4**:92–93
fat **4**:93
metabolism **1**:590, **3**:451
protein **4**:93–94
macrosomia **4**:77–80
macrovascular complications **3**:235
macular degeneration **1**:73–74
magnesium **1**:192, **2**:183–184, **3**:12, **3**:519, **3**:616, **4**:38, **4**:111, **4**:173
assessment of mg status **1**:311–312
causes of Mg deficit **1**:312
consequences of Mg deficit and implications in various metabolic diseases **1**:312–313
deficit **1**:312–313
dietary Mg deficiency **1**:312
dietary sources of magnesium **1**:311
distribution of Mg within body **1**:309
excess **1**:313
intestinal absorption **1**:310
metabolism **1**:309–311
recommended dietary allowances **1**:313
requirements **1**:311–312
secondary Mg deficiency **1**:312
urinary excretion **1**:310–311
magnetic resonance imaging (MRI) **3**:22–23
maize (*Zea mays*) **2**:295
major histocompatibility complex (MHC) **3**:254
malabsorption syndromes
carbohydrate malabsorption **3**:479–480
clinical management of malabsorption **3**:482–483
fat malabsorption **3**:480–481
fluids and electrolytes **3**:479
general nutritional management of malabsorption **3**:479
pathophysiology and symptoms **3**:477–479
route of nutrition in malabsorption **3**:481
selection of enteral formulas for malabsorption **3**:481–482
specific nutritional management of malabsorption **3**:479–482
malnutrition **3**:137, **3**:341–342, **3**:486
activity **4**:407
in adults **3**:502
anthropometric indicators for measuring malnutrition **3**:501–502
biological and dietary factors **3**:493–494
causes of secondary under and overnutrition **3**:507–509
caveats for diagnosis
of secondary excess nutriture **3**:509–510
of secondary undernutrition **3**:510
in children **3**:501–502
clinical manifestations and classification **3**:498
coexistence of primary and secondary malnutrition **3**:509
determinants **3**:493–494
diagnosis of secondary malnutrition **3**:509–510
environmental factors **3**:494
epidemiology of DBM **4**:402–405
food supply **4**:408
food systems **4**:407–409
global nutrition targets **3**:494–496
long term consequences **3**:500
management of secondary malnutrition **3**:510–512
marketing **4**:407–408
microbiota and **2**:171–172
micronutrient **3**:503
deficiency **4**:407
moderate and severe thinness **3**:502
underweight, overweight, obesity **3**:502
and mortality **3**:487
nutrition transition **4**:405–407
obesity **4**:406
overnourishment **3**:486
overnutrition **3**:486
overweight and obesity **3**:486
principles of management **3**:510
protein energy malnutrition **3**:486
public health approaches **3**:511
reverse paradigm **3**:509
socioeconomic factors and inequity **3**:494
stunting **3**:486
sugar-sweetened beverages **4**:408–409

types of **3**:490
ultra-processed foods **4**:406
undernourishment **3**:486
undernutrition **3**:486
underweight **3**:486, **3**:502
stunting and wasting **3**:501–502
wasting **3**:486
maltose **1**:118
mammalian target of rapamycin (mTOR) **4**:75
manganese **1**:196, **4**:174
absorption, transport, and storage **1**:316–317
analysis **1**:316
assessment of manganese status **1**:320–321
chemical and physical properties **1**:315
deficiency **1**:318–319
dietary sources **1**:315–316
manganese toxicity **1**:319–320
metabolic function and essentiality **1**:317–318
physiological role **1**:316–317
tissue concentrations **1**:316
mannitol **1**:119–120
Manual Ability Classification System (MACS) **3**:172
maple syrup urine disease (MSUD) **3**:415–417
marasmic-kwashiorkor **3**:500
marasmus **3**:498
maresins **2**:212–213
market-driven fortification **2**:344
mass-fortification **2**:344
maternal calorie restriction **4**:53
maternal glucocorticoid exposure **4**:55
maternal high-fat diet **4**:54
maternal nutrition **4**:74
maternal obesity **4**:54
maternal overnutrition models **4**:54
maternal protein restriction **4**:54
maternal undernutrition models **4**:53–54
maturity onset diabetes of young (MODY) **3**:253
Max Plank respiration gas meter **2**:157–158
meal size and frequency:effect on absorption and metabolism
effect of meal frequency on absorption **2**:447–448
energy expenditure **2**:449
energy intake **2**:448–449
metabolism **2**:448–449
effect of meal size on
energy expenditure **2**:447
energy intake **2**:446–447
metabolism **2**:446–447
mechanisms meal frequency influence energy intake **2**:449
mechanisms underlying metabolic effect of meal frequency **2**:447–448
regulation of meal size by gut-derived satiety peptides and adiposity signals **2**:446
mealtime behaviors **3**:179
measurement of growth **4**:85
meat, poultry, and meat products
bioavailability of nutrients and efficiency for child development **2**:453
classes of nutrients and meat products **2**:452–453
lamb **2**:455–456
meat sources and nutritional values **2**:453–457
nutrient density of meat and meat products **2**:453
nutritional value **2**:452
poultry **2**:457
processed meats **2**:457
veal **2**:457
medical nutrition therapy (MNT) **3**:236, **3**:260
medications **3**:179
use **2**:79
Mediterranean diet (MedDiet) **2**:182, **2**:458, **4**:242–247, **4**:267, **4**:273
adherence to **2**:462–464, **2**:463t–464t
development of Mediterranean food **2**:459
as healthy food pattern **2**:462–465
origin of **2**:459–460
pyramid **2**:461f
structure and characteristics of **2**:460–462
as sustainable food pattern **2**:465
Mediterranean Diet Adherence Screener (MEDAS) **2**:462
Mediterranean Diet Quality Index (Med-DQI) **2**:462
Mediterranean Diet Score (MDS) **2**:462
Mediterranean lifestyle index (MEDLIFE) **2**:462
Mediterranean triad **2**:459
Mediterranean-Style diet **3**:239–240
megaloblastic anemia **1**:253–254
melanin, functions of **1**:42
melanocortin 4 receptor (*MC4R*) **2**:180
melanoma differentiation-associated protein 5 (MDA5) **3**:254
melatonin, functions of **1**:42
mendelian randomization studies **3**:150
mental function **2**:20
mercury **2**:356–357, **4**:217
dietary exposure **2**:356
permissible intakes **2**:356
toxicity **2**:356–357
toxicokinetics **2**:356
meta-analysis **1**:52
metabolic alkalosis **2**:111
metabolic bone disease **3**:335
metabolic demand **2**:569–570
amino-acid pattern of **2**:576–577
for amino-acids **2**:570, **2**:570f
obligatory and adaptive components of **2**:570–572
metabolic diseases **2**:657
association consumption with **2**:598–599
metabolic dysfunction **2**:39–43
Metabolic Equivalents of Task (METs) **1**:352–353
metabolic functions **1**:89–90
metabolic genes, control of expression of **2**:198–201
metabolic health **1**:243
metabolic liver disorders **3**:457–459
metabolic model **2**:570
metabolic status **2**:86
metabolic studies **2**:561
metabolic syndrome **1**:537, **2**:487, **3**:591
effects of antioxidant supplementation on **1**:50–54
metabolism **4**:160
of acetate **2**:255
inborn errors of **3**:412–423
and urinary excretion of biotin and metabolites **1**:87–88
of vitamins **2**:217–222
metabolomics **1**:565–566
metal responsive element (MRE) **2**:207
metformin **3**:249
methyl CpG binding protein 2 (MeCP2) **2**:183
methyl donor molecules **2**:182–183
microbial metabolites **2**:173
microbiome **2**:628–629
and inflammation **3**:712
sucrose metabolism by components of **2**:601–602
microbiota
of intestine **2**:162, **2**:169–170
gut microbiome and development of disease **2**:162–166
personalized nutrition and gut microbiota changes **2**:166–167
and malnutrition **2**:171–172
personalized nutrition through **2**:174
microminerals/trace elements **4**:137–138
micronutrients **2**:20, **2**:302, **3**:157, **3**:529, **4**:18, **4**:153–155, **4**:396
combinations **3**:357
content **2**:518
deficiencies **3**:503, **3**:684–688, **4**:123–124, **4**:407
in chronic alcohol misuse **2**:271
disorders of micronutrient metabolism **3**:421–422
in food fortification **2**:341–342
malnutrition **3**:503
supplementation
case studies **2**:474–475
cost of micronutrient supplements **2**:468
key challenges and success factors for program implementation **2**:473–475
micronutrient supplementation recommendations **2**:468
purpose of **2**:468
supplementation program cycle **2**:468–476
microorganisms, carcinogenic substances produced by **3**:127–130
microRNAs (miRNA) **2**:178, **4**:55–56
microsomal ethanol oxidizing system **2**:254–255
microsomal triglyceride transfer protein (MTP) **3**:370
microvascular complications **3**:235
mid arm-upper arm circumference (MUAC) **3**:89, **3**:175

midupper arm muscle circumference (MUMAC) **3**:92
migrant studies **3**:148–149
mild acute malnutrition **3**:498
mild age-related diabetes (MARD) **3**:257
mild obesity-related diabetes (MOD) **3**:257
milk **2**:281–282
  animal **2**:281
  composition **2**:480–484
  fat **2**:488
  growth and development **2**:486
  overall mortality **2**:486
  pregnancy, maternal health, and newborns **2**:486
  role in disease prevention and health promotion **2**:486–488
  soy **2**:281–282
  weight gain **2**:486–487
millets **2**:295
milling **2**:299
mimosine **3**:349
minerals **2**:183–184, **2**:288–289, **2**:401, **2**:452–453, **2**:488, **3**:123–124, **3**:241, **3**:285, **3**:357, **3**:428, **4**:94–96, **4**:130, **4**:192–193
  accretion in fetus **4**:73
  elements **2**:483
  enteral nutrition **4**:114
  status in CF **3**:214
minimally processed foods **2**:654–655
minimum protein requirement (MPR) **2**:569, **2**:571
mitochondrial DNA damage **4**:15–16
mitochondrial dysfunction **4**:56
mitochondrial fatty acid
  β-oxidation **2**:317–320
  synthesis **2**:324
mitochondrial metabolism is uncoupled in brown adipocytes **2**:29
mitochondrial pathway **3**:436
mitochondrial senescence **4**:158
moderate acute malnutrition (MAM) **3**:498
MODY **3**:260–261
molecular mechanism(s)
  of action **1**:186–187
  of vitamin E
    absorption, metabolism and excretion **1**:563–564
    action **1**:565
molecular nutrition **2**:197
molybdenum **1**:197, **1**:453–454, **4**:174
monitoring feeding tolerance **4**:114
monogenic diabetes **1**:135
monogenic obesity **3**:584–585
monosaccharides **1**:115–118, **2**:413–414
  See also oligosaccharides
  chemical reactions **1**:117–118
  classification **1**:115–116
  ester formation **1**:118
  ether formation **1**:118
  oxidation to produce sugar acids **1**:117
  physical properties **1**:117
  reactions in alkaline solutions **1**:117–118
  reduction to produce sugar alcohols **1**:117
  stereoisomerism and nomenclature **1**:116–117
monounsaturated fatty acids (MUFA) **1**:297–298, **3**:383–384, **4**:264
  interaction of vitamin E with **1**:562–563
morbid obesity **1**:435–436
mortality **2**:657, **3**:487
mouth **2**:309
mucositis **3**:140
multiple births **4**:226
Multiple Food Test **3**:63
multiple micronutrient supplements (MMS) **4**:235–237
muscle
  effects of alcohol on **2**:260–261
  muscle-strengthening activity **1**:356
  tone and positioning **3**:178
muscular fitness **1**:359
musculoskeletal system **4**:159–160
mycotoxins
  climate change **2**:499–500
  exposure assessment **2**:496
  mitigation **2**:499
  occurrence and human exposure of **2**:495–496
  regulations **2**:498–499
  risk assessment **2**:496
  toxicity **2**:493–495
myosin heavy chain (MHC) **1**:396

N

N-3 fatty acids **3**:385
n-3 polyunsaturated fatty acids (PUFA) **2**:183
  biological functions of pro-resolving mediators derived from **2**:210–213
N-6 fatty acids **3**:385
N-methylcarbamates **2**:381
N-phosphonomethyl amino acids **2**:386
6-n-propylthiouracil (PROP) **3**:64
Na^+ sources, intake and absorption **1**:421–422
narrow therapeutic index (NTI) **2**:82–83
naso-jejunal feedings (NJ feedings) **3**:178
nasoduodenal route **3**:558
nasogastric route **3**:557–558
nasojejunal route **3**:558
National Health and Nutrition Examination Survey (NHANES) **1**:297–298, **2**:631, **3**:62–63
National Institutes of Health (NIH) **4**:278
national weight control registry (NWCR) **3**:726
natriuretic effector mechanisms **1**:425
natural analogs of vitamin E **1**:559–560
natural sugars **1**:137
natural vitamin E analogs, chemistry of eight **1**:557–558
nausea **3**:139
neonatal cholestatic disorders **3**:454–456
neonicotinoids **2**:382–383
nervous system **1**:589, **3**:596
net protein utilization (NPU) **2**:562
neural regulation in starvation **3**:653–654
neural tube defects (NTDs) **4**:199
  biological mechanisms in etiology and prevention of **4**:203–205
  effects
    of folic acid policy globally **4**:206–208
    on health outcomes **4**:208
    on NTD prevalence **4**:207–208
  emerging folic acid policy **4**:208–209
  epidemiology **4**:199
  folate-related genetic risk factors and risk of **4**:203–205
  folic acid
    fortification **4**:206
    supplementation **4**:206
  genetic and environmental factors **4**:199–203
  intervention studies with folic acid in prevention of **4**:200
  interventions to optimize folate status **4**:206
  maternal folate status **4**:200
  maternal risk factors for **4**:202–203
  maternal vitamin B_{12} status **4**:200–201
  nutritional considerations **4**:205–206
  obesity and diabetes **4**:202–203
  observational studies of maternal folate status in relation to NTD risk **4**:200
  prevention of **4**:205–209
  recommendations to prevent first occurrence and recurrence of **4**:209
  risk factors **4**:203
  risk-benefit assessment **4**:208–209
  role of folate, vitamin B_{12} and one-carbon metabolism **4**:203
  specific nutrients **4**:202
neurogenic differentiation 1 (NeuroD1) **2**:198
neuroglycopenic syndromes **3**:404–405
neutron activation analysis **3**:23
newborn screening **3**:413–414
niacin **2**:233–236, **3**:10, **4**:40
  and pellagra
    absorption, transport, and storage **1**:326
    assessment of niacin status and requirements **1**:328
    dietary sources and high intakes **1**:328–329
    functions **2**:234–235
    metabolic function and essentiality **1**:327
    metabolism **2**:236
    metabolism and excretion **1**:326–327
    signs and symptoms of deficiency, and antimetabolites **1**:322–325
nickel **1**:197–198, **1**:454, **2**:358
  dietary exposure **2**:358
  permissible intakes **2**:358
  toxicity **2**:358
  toxicokinetics **2**:358
night blindness (XN) **1**:461–462
9-point hedonic scale **3**:65
nitric oxide (NO) **1**:425, **2**:204
nitrogen (N) **2**:569
  metabolism **3**:194
  retention **2**:562
nitrogen balance (NB) **2**:573, **4**:69–71

studies **2**:572–574
in WHO recommendations **2**:574–576
nitrogenous compounds **2**:400
no observed adverse effect level (NOAEL) **1**:184
non-alcoholic fatty liver disease (NAFLD) **3**:257, **3**:452, **4**:52–53
dysbiosis of gut microbiota in **2**:165–166
non-communicable diseases (NCDs) **1**:352, **1**:354, **2**:303–304
non-esterified cholesterol **1**:295–296
non-fat dry weight **4**:69
non-fetal **4**:58–65
non-genetic maternal factors **4**:74
non-nutrient factors **3**:248–250
non-nutritional factors **3**:174
non-nutritive sweeteners **3**:244, **3**:244t
non-saccharide prebiotic candidates **2**:555
non-starch **1**:137
non-water beverages **3**:710
nonalcoholic fatty liver (NAFL) **4**:52–53
nonalcoholic steatohepatitis (NASH) **3**:452
nondigestible glucose polysaccharides **2**:417–418
nonheme iron **1**:289
nonobese diabetic mouse (NOD) **3**:254–255
nonoxidative metabolism of alcohol **2**:255–256
NOVA classification system **2**:654–655
novel omics approaches for precision nutrition in CVD prevention **4**:252–253
nuclear receptor liver receptor homolog 1 (LRH-1) **2**:201
nuclear retinoid receptors **1**:474
nucleic acids
biosynthesis in humans **1**:331
content of foods **1**:334
and related compounds in beverages **1**:334–336
synthesis of **1**:332
nucleosides and nucleotides, pharmacological uses for **1**:338
nucleotides
effect of cooking on nucleotide content of diet **1**:334
in human breast milk and infant formula milks **1**:337
metabolic roles of **1**:331
pharmacological uses for nucleosides and **1**:338
nutraceuticals
blood pressure **2**:505
body weight **2**:504–505
bone health **2**:507
brain health **2**:502–503
cardiometabolic health **2**:503–505
eye health **2**:506
glucose regulation **2**:503
health benefits of **2**:502–508
immunity and infection **2**:505
lipids **2**:504
muscle health **2**:507–508
nomenclature of **2**:501–502
sleep **2**:507
urinary healthnutrient reference values (NRVs) **1**:168, **2**:505–506
nutrient responsive response elements (NSRE) **2**:204
nutrient(s) **2**:87–88
adversely affecting bone health **3**:83
and CHD **4**:263–266
composition of cereal grains **2**:300–303
declaration **4**:302–303
gene interactions **3**:618–619
inadequacy **4**:146–147
intake and consideration of risk of chronic disease **4**:319–320
intake distribution model **4**:319
intakes for pregnancy **4**:168–174
interaction of sucrose consumption with intake of other **2**:600–601
metabolic regulation **2**:177–178
needs **3**:176
needs from complementary foods **4**:45
nutrient-drug interactions **3**:703
nutrient-related claims **4**:303
requirements **3**:665–666
sensing mechanisms **3**:654–655
supplements **3**:388
nutriepigenetics
chromatin structure **2**:177–178
dietary intervention **2**:180–182
epigenetic modifications **2**:177–178
epigenetics **2**:180–182
nutrient metabolic regulation **2**:177–178
nutritional factors in epigenetic regulation **2**:182–186
nutriepigenomics **2**:177
nutrigenetics **1**:170
basic concepts of genetics applied to **2**:191–192
genetic variants and susceptibility to disease **2**:192–193
influence of gene variants on metabolic requirements and utilization of nutrients and foods **2**:193–196
origin of term and development of nutrigenetics **2**:189–191
timing of food intake in **2**:46–47
nutrigenomics **1**:170–171, **2**:197–198
modulation of gene expression
by amino acids **2**:204–205
by carbohydrates **2**:198–202
by lipids **2**:202–204
regulation of gene expression
by trace metals **2**:207
by vitamins A and D **2**:205–207
nutrition **2**:261–262, **3**:673–684
during acute phase **3**:524–525
approaches to nutrition labeling **4**:302–304
assessment **3**:236
and asthma **3**:513–521
cancer **3**:137
diet change, and implications
demographic and epidemiological transitions **4**:436–437
health effects **4**:442
mass media and internet use **4**:441
obesity trends **4**:438
rapid shifts in dietary and activity patterns and body composition **4**:437–438
shift in overall structure over time **4**:439–440
social burden of changes in diet, body composition, and health **4**:443
social change **4**:440–441
structural shifts in income–diet relationships are occurring **4**:441
urbanization **4**:440
drug-induced alteration in **2**:86–88
growth **3**:427
and HIV/AIDS **3**:535–540
human and economic costs **3**:489
increased energy expenditure **3**:525–526
interaction of cytokines and **2**:53–56
influence of fatty acids on inflammatory cytokine production **2**:54–55
interaction of inflammatory cytokines with amino acid and protein metabolism **2**:55
modulation of inflammatory cytokine production by micronutrients **2**:55–56
after major burn **3**:524–525
management **3**:425–429
of phenylketonuria **3**:424–431
metabolic disturbance during recovery and long-term effects **3**:531
monitoring of nutrition support **3**:530–531
and parasitism **3**:541–548
pathophysiology **3**:525
during preschool years
developmental milestones in relation to diet **4**:122
dietary recommendations **4**:122–123
excess weight **4**:123
faltering growth **4**:122
micronutrient deficiency **4**:123–124
during recovery **3**:531–533
requirements **3**:526, **3**:533
role
in inflammatory bowel disease **3**:669–697
in prevention and management of asthma **3**:515–519
science
on food culture **4**:272–273
food culture and **4**:277–278
traditional food culture and **4**:278
for sport and exercise
dietary supplements **4**:131
fat and CHO **4**:129–130
nutrition for competition **4**:131
nutrition for training **4**:128
water and electrolyte balance **4**:130–131
support after minor burn **3**:533
and susceptibility to tuberculosis **3**:549–555
and sustainable development **3**:487–490
timing and route of nutrition support **3**:525
transition **2**:653

nutrition (*continued*)
water insecurity and **3**:708–712
nutrition-sensitive' interventions **4**:235–237
effect of prenatal and postnatal **4**:236t–237t
nutritional aspects and liver physiology in normal liver and liver diseases **3**:450–451
nutritional assessment **3**:172
adults **3**:93
advantages and limitations of anthropometric measurements **3**:88
anthropometric measurements **3**:88–93
children **3**:93
elbow width **3**:90
errors of anthropometric measurements **3**:88
height **3**:88
midupper arm circumference-for-age **3**:92
midupper arm circumference-for-height **3**:92
nutritional Indices **3**:90
reference values **3**:93
skinfold thickness-for-age **3**:92
uses of anthropometric measurements **3**:87–88
weight **3**:89
weight-for-age **3**:92
weight-for-height **3**:92
nutritional factors **3**:174
in epigenetic regulation **2**:182–186
nutritional genomics **2**:197
of choline **1**:170–171
nutritional instruction **3**:237
nutritional issues **3**:171
nutritional management **3**:210–211, **3**:698–705
and treatment **3**:663–664
nutritional problems of adolescents
adolescents' diets **4**:147–148
eating patterns **4**:144–146
nutrient inadequacy **4**:146–147
obesity **4**:144
physical growth **4**:143–144
nutritional requirements **4**:150–156
nutritional status **2**:86–87
on drug disposition **2**:85–86
nutritional supplements **3**:221
nutritional support **3**:470–473, **3**:574–576
access **3**:557–558
artificial nutritional support **3**:566–572
bone disease **3**:577
catheter complications **3**:578
dextrose **3**:575
dietary counseling and fortification **3**:565–566
ethical issues **3**:572
home enteral tube feeding **3**:567
home parenteral nutrition **3**:568
indications **3**:567
and contradictions **3**:561
for PN **3**:575
infections **3**:578
intestinal transplantation **3**:572
lipid emulsions **3**:576
liver disease **3**:577
metabolic complications **3**:577–578
methods of infusion **3**:561–562
micronutrients **3**:576
deficiency and excess **3**:577
monitoring **3**:568–569
oral nutritional
supplements **3**:566
support **3**:565–566
organization **3**:568
origins and development **3**:566–567
outcome **3**:570–572
potential challenges **3**:562
standards of care **3**:568
types **3**:557
nutritional surveillance
dietary data collection methods **4**:445–448
emerging nutritional and health issues **4**:452
examples of nutritional surveillance activities **4**:448
food composition
databases **4**:452
by household **4**:446
by individuals **4**:447–448
systems **4**:445–452
uses of nutritional surveillance data **4**:444–445
nutritive sweeteners **3**:245t
nuts **2**:514–516, **4**:266–267
contaminants **2**:520
intolerances/allergies to **2**:520
macronutrient content of **2**:516–517
and role in diet and health **2**:520
types of **2**:514

O

oats (*Avena sativa*) **2**:296
obese **4**:226
obesity **1**:536–537, **2**:39–43, **2**:409, **2**:656–657, **3**:287–288, **4**:144, **4**:406
See also childhood obesity
abnormalities of hormones and other circulating factors **3**:593–594
activation of BAT to treat obesity **2**:33
adipose tissue as circadian organ in **2**:40–42
adiposis dolorosa **3**:596
anthropometry **3**:602
arthritis **3**:597
assessment **3**:601–602
of overweight and obesity in childhood **3**:581–582
atherosclerotic and arteriosclerotic vascular diseases **3**:594–595
bioimpedance analysis **3**:602
cadaver analysis **3**:602
cancer **3**:597
carbohydrate and **1**:139
cerebrovascular disease **3**:595
as chronic disease **3**:731–732
clinical evaluation of children with **3**:586–587
congestive heart failure **3**:595
densitometry **3**:602
diseases of organ systems **3**:594–596
dyslipidemia **3**:593
economic impact **3**:598
effects of antioxidant supplementation on **1**:50–54
epidemiology **3**:581
epigenetic mechanism of clock genes in **2**:40–42
etiology **3**:601
eye disease **3**:596
females **3**:596
fructose consumption and **2**:394
genetic variations
associated with **2**:39–40
interact with several behaviors for **2**:40
genetics factors **3**:582
global problem **2**:26–27
and glycemic index **2**:427
gout **3**:593
health impacts **3**:601
imaging techniques **3**:602
immune system **3**:596
importance of weight gain prevention in adults **3**:605
infant feeding and **4**:60
intraabdominal pressure **3**:597
males **3**:595
measures of body fatness **3**:602
mechanical complications of obesity **3**:597
metabolic and organ system complications of obesity **3**:591–594
obesity prevention strategies target **3**:605–611
objectives of obesity prevention **3**:605
obstetric complications **3**:596
overweight and **2**:310–311
pregnancy weight gain and postpartum risk of **4**:227
prevention
creating supportive environments **3**:609–610
family focus **3**:606
high-risk groups **3**:606–607
identifying effective obesity prevention interventions **3**:610
increasing energy expenditure **3**:609
increasing intake of high-fiber, energy-dilute foods **3**:608
key elements of weight gain prevention plan **3**:607–608
lessons from other prevention efforts **3**:610–611
of obesity during childhood **3**:589
reduced time spent in sedentary behaviors **3**:609
reducing consumption of sugar-sweetened soft drinks and juices **3**:609
reducing energy intake **3**:608–609
reducing intake of high energy-dense foods **3**:608

reducing level of food prepared outside home **3**:609
reducing portion sizes **3**:609
regular physical activity **3**:609
strategies target **3**:605–611
those with existing weight problem **3**:607
whole community **3**:606
principles of obesity prevention **3**:604–605
pseudotumor cerebri **3**:596
psychosocial complications **3**:598
rationale for obesity prevention **3**:604
reproductive system **3**:595–596
role of distribution of body fat in complications of obesity **3**:591
skin **3**:596
social complications **3**:598
surgical complications **3**:598
thromboembolic disease **3**:595
obestatin **3**:594
obligatory nitrogen loss (ONL) **2**:570–571
observational studies **4**:277–278
obstructive sleep apnea syndrome **3**:475
ochratoxin A **2**:497
older people
aging of population and people **4**:157
apoptosis **4**:158
barriers to meeting recommended nutrient intakes and healthful dietary intake patterns by **4**:155–156
cellular senescence **4**:158
central and peripheral nervous system **4**:161
cognitive function **4**:161
dietary guidelines for function, health, and disease prevention **4**:155
drug metabolism **4**:162
established recommended intakes for older persons **4**:152
gonads and reproductive system **4**:160
hematopoietic and immune system **4**:160–161
human aging and nutrition **4**:150
integumentary tissues **4**:158–159
nature of senescence **4**:157–158
nutrient intake recommendations in later life **4**:151
oral cavity and alimentary tract **4**:159
physiological changes occurring in tissues and organ systems with human aging **4**:158–162
special senses **4**:161
specific factors of aging influencing nutritional requirements **4**:151
successful aging, normative aging, and frailty **4**:150–151
surrounding recommended intakes of nutrients **4**:151–152
synthesis **4**:162
oligopeptides, functions of **1**:37–39
oligosaccharides **1**:121, **2**:414
See also monosaccharides
chemical reactions **1**:121
glucose in **2**:414–416
physical properties **1**:121
omega-3 fatty acids **3**:560
omega-3 PUFAs **3**:639–640
omnivorous diets, morbidity and mortality on **2**:667–670
"one size fits all" diet **3**:236
oral cancer **3**:276
etiology of **3**:276
prevalence **3**:276
oral diseases
impact of **3**:273
social and commercial determinants of **3**:276–278
oral glucose tolerance test (OGTT) **3**:228
oral motor
considerations **3**:180
dysfunction **3**:177–178
and dysphagia **3**:177–178
oral rehydration solutions (ORSs) **3**:224
organic acidemias **3**:417
organic acids **2**:404–405
organic chemicals **3**:128–129
organic foods
animal products **2**:525
clinical trials **2**:527
epidemiological studies **2**:527–528
essential or beneficial nutrients **2**:524–525
health effects **2**:527–528
organoleptic quality **2**:528
plant products **2**:524–525
scientific bases for comparisons of food quality **2**:524
undesirable components **2**:526–527
organochlorines **2**:378–380
organophosphates (OP) **2**:380–381
organosulfur compounds (OSCs) **2**:406–407
oroenteric route **3**:557–558
orthopedic surgeries **3**:178–179
osmolality **3**:219
osteoporosis **1**:539, **2**:487, **3**:446
nutritional factors
dietary intake and body mass **3**:613
interactions of bone and physiological systems **3**:619–620
lifestyle factors **3**:618
targeting groups at risk **3**:618
ovarian cancer **2**:489
overweight women **4**:226
oxidative phosphorylation **2**:121–122
oxidative stress **1**:237, **4**:56, **4**:158
2-oxoglutarate-linked hydroxylases **1**:507–508
oxylipin synthesis **2**:326–327

P

P-glycoprotein (P-gp) **2**:81
pain and discomfort **3**:179
palliative care **3**:143
pancreatic exocrine insufficiency **3**:269
pancreatic insufficiency **2**:270
pancreatic β-cells, insulin synthesis in **2**:198
pancreatic/duodenal homeobox-1 (PDX-1) **2**:198
pancreatitis **2**:270
pantothenic acid **2**:239–241, **3**:11
absorption, transport, storage and status measurement **1**:340–341
dietary sources and high intakes **1**:344
functions **2**:239–241
metabolic function and essentiality **1**:343–344
metabolism **2**:241
and turnover **1**:341–343
requirements **1**:344
parasites **3**:267
and nutrition **3**:544–547
parasitism
gut microbiota, inflammation and gut permeability **3**:545
loss of appetite and anorexia **3**:545
maldigestion and malabsorption **3**:546
nutrient losses **3**:546
parasites and nutrition **3**:544–547
parasitology **3**:542–544
prevention **3**:547
treatment **3**:546–547
parenteral nutrition (PN) **3**:329–335, **4**:109–111
See also enteral nutrition (EN)
catheter occlusion/thrombosis **3**:334
complications of **3**:334
components of PN **4**:109–111
contraindications to **3**:330
cyclic parenteral nutrition **3**:333
home parenteral nutrition **3**:335
indications for **3**:330
infection **3**:334–335
metabolic **3**:334
nutrition components of **3**:331
parenteral nutrition monitoring **3**:333
perioperative support in severe malnutrition **3**:330
vascular access **3**:330–331
parenteral nutrition-associated liver disease **3**:456–457
Parkinson's disease **1**:500
partial enteral nutrition (PEN) **3**:691
key points **3**:691–692
partial thromboplastin time (PTT) **2**:87
particularly leucine **4**:75
pathogens **3**:711–712
pathophysiology of dehydration **3**:219–220
patulin **2**:498
peak bone mass (PBM) **3**:80, **4**:144
attainment **3**:80–81
peanut (*Arachis hypogaea*) **2**:515
avoidance **3**:343
pecan (*Carya illinoinensis*) **2**:515
pediatric cohort **3**:690
pediatric feeding disorder (PFD) **3**:623
behavioral interventions **3**:630
behavioral psychologists **3**:628
caregiver involvement **3**:630–631
classification of feeding disorders in children **3**:623
cultural considerations within evaluations **3**:629
dietitians **3**:627–628
evaluation **3**:623

pediatric feeding disorder (PFD) (*continued*)
and treatment of pediatric feeding disorders **3**:623
feeding
skill development **3**:629–630
specialists **3**:627
and swallowing physiology **3**:623
medical intervention **3**:629
multidisciplinary approach **3**:629
nutritional support **3**:629
physicians **3**:624–626
social workers **3**:628–629
treatment
of feeding disorders **3**:629–631
settings **3**:631–632
pegvaliase **3**:429
PEN with restricted diet-CDED **3**:691–692
people consuming fortified food **4**:295
peptides, intestinal epithelial digestion and absorption of **1**:379–382
percutaneous endoscopic gastrostomy (PEG) **3**:326, **3**:558
PEG-JET **3**:326
percutaneous endoscopic jejunostomy (PEJ) **3**:558
percutaneous jejunostomy (PEJ) **3**:326
periodontal disease **3**:275
etiology of **3**:275
prevalence and trends **3**:275
peripheral nervous system **4**:161
peripheral tissues, transport from intestine to **1**:87
peroxisomal fatty acid β-oxidation **2**:320–322
peroxisome proliferator-activated receptor gamma (PPARG) **2**:201, **3**:256
peroxisome proliferator-activated receptor-a (PPARa) **4**:55–56
personal food environments **4**:343
pesticides
classification **2**:377
dietary risk assessment of pesticide residues in food **2**:386–387
magnitude of pesticide use worldwide **2**:378
regulation **2**:376–377
terminology **2**:376
toxicology of **2**:378–386
pharmaceutical interactions **2**:81
pharmacodynamic interactions **2**:82
pharmacokinetic interactions **2**:81
pharmacological effects in rodents **1**:185–186
pharmacological models **4**:55
pharmacotherapy **3**:300–301
anticonvulsant medication **3**:301
antidepressant medications **3**:301
BN **3**:310
psychostimulant medication **3**:301
weight loss medications **3**:301
pharynx **2**:309
phenolic compounds **2**:405, **2**:407–408
phenylketonuria (PKU) **3**:424
maternal **3**:429
nutrition management **3**:425–429
pharmacological therapies **3**:429
phenylthiocarbamide (PTC) **3**:64
phosphatidylcholine (PC) **1**:162–163
phosphatidylethanolamine N-methyltransferase (PEMT) **1**:165
phosphocholine (phoCho) **1**:162–163
phosphoinositol-3 (PI-3) **1**:186–187
phospholipid transfer protein (PLTP) **3**:370
phospholipid(s) **1**:295, **3**:364
phosphorous **4**:38
phosphorus **3**:616–617, **4**:111, **4**:172
absorption **1**:348
calcitriol 1,25$(OH)_2D_3$ **1**:349
FGF-23 **1**:349
hyperphosphatemia and consequences **1**:349–350
hypophosphatemia and consequences **1**:350
phosphorus homeostasis **1**:348–349
PTH **1**:348
requirements and food sources **1**:347–348
phosphotyrosine phosphatase 1B (PTP 1B) **1**:186–187
physical activity **1**:352–357, **3**:250, **3**:587–588
all-cause mortality **1**:354
cancer **1**:354–355
CVD **1**:354
genesis of modern physical activity epidemiology **1**:353
guidelines advisory committee **1**:357
health benefits **1**:353t
intersection of **1**:358
leisure-time *vs.* occupational physical activity **1**:355
light, moderate and vigorous **1**:356–357
muscle-strengthening activity **1**:356
physical fitness **1**:358–359
sedentary behavior **1**:357
physical fitness **1**:353, **1**:358–359
physical growth **4**:143–144
physical inactivity, global burden of **1**:359–360
phytic acid **2**:519
phytochemicals **2**:531, **2**:686, **3**:158, **4**:265–266
absorption and metabolites **2**:541
antibacterial activity of A-type PAC **2**:542
B-type PAC, poly(catechin), for cardiovascular health **2**:541
cancer **2**:537
cardiovascular health **2**:536
current estimates of intake **2**:534
dietary sources **2**:539
flavan-3-ols **2**:533
neuroprotective effects **2**:537
potential health effects **2**:536–537
potential mechanisms of action **2**:536
safety **2**:537
phytoestrogens **3**:617
phytosterols **2**:407, **2**:519
pigments **2**:405
pine nuts **2**:515
pistachio nuts **2**:515
Piwi-associated RNAs (piRNA) **2**:178
placenta **4**:74–75
plant sterols **3**:388
plant-based dietary patterns **4**:248
plant-based diets (PBDs) **4**:248
plasma choline **1**:167
plasma osmolality **3**:219
plasma volume (PV) **3**:217
pneumonia **1**:513
policy strategies to reduce free sugar consumption **3**:278–280
advertising restrictions **3**:279
downstream strategies **3**:280
labeling **3**:279
midstream strategies **3**:279
regulation of industry influence on policy development **3**:278
sugar production and trade **3**:279
taxation **3**:279
upstream strategies **3**:278–279
voluntary agreements to achieve reformulation **3**:279
policy-bases interventions **4**:377
political-bases interventions **4**:377
pollutants **3**:711–712
poly(catechin) **2**:541
poly(gallocatechin) **2**:542
polyalcohols **2**:400
polyamines, functions of **1**:40
polychlorinated biphenyls (PCBs) **4**:218
polygenic obesity **3**:584
polymeric formulas **3**:559
polymorphisms relevant for vitamin E action **1**:566
polyphenol(s) **2**:184–186, **2**:519, **3**:637–638
and inflammation **2**:74
polysaccharides **1**:122–124, **1**:137–138, **2**:414
polyunsaturated fatty acids (PUFA) **1**:297–298, **2**:72–73, **3**:358, **3**:385, **4**:264
interaction of vitamin E with **1**:562–563
population attributable risks (PAR%) **1**:359–360
population reference intake (PRI) **4**:146–147
pork **2**:455
positional isomers **1**:292–293
post-liver transplantation, end stage liver disease in **3**:460–462
post-translational modifications **2**:200
postbiotics **1**:214
postmenopausal bone loss **3**:81
postnatal catch-up growth **4**:53
postnatal growth and later risk of disease **4**:60–61
postnatal infant and young child interventions **4**:237
postpartum risk of obesity, pregnancy weight gain and **4**:227
postprandial hypoglycemia **3**:405
clinical significance of **2**:426–427
postprandial inflammation **2**:75
postprandial protein utilization **1**:397–399

postprandial syndrome **3**:407
postprandial thermogenesis **1**:220–221
  physical activity **1**:221
potassium **1**:192, **3**:12, **3**:617, **4**:38, **4**:266
  acid–base considerations **1**:362–363
  adverse effects
    of excess potassium intake **1**:365
    of insufficient potassium **1**:363–365
  cardiovascular disease **1**:364–365
  elevated blood pressure **1**:363–364
  kidney stones and bone demineralization **1**:363
  recommended potassium intake, current intake, and dietary sources **1**:365–366
potential nutrient contribution of fortified foods **4**:295
pre-eclampsia
  abnormal lipid metabolism **4**:189
  angiogenic and antiangiogenic factors **4**:187
  association between nutritional factors and **4**:190–193
  inflammation **4**:187–188
  maternal undernutrition **4**:191
  minerals and trace elements **4**:192–193
  overweight and obesity **4**:190–191
  oxidative stress **4**:188
  pathophysiology of **4**:187–190
  plasma volume dysregulation **4**:190
  risk factors **4**:185–186
  vitamins **4**:191–192
pre-liver transplantation, end stage liver disease in **3**:460–462
prebiotics **2**:553–557
  candidates **2**:554
  health benefits of **1**:214
  health outcomes reported with prebiotics **2**:555–556
  history **2**:553
  mechanisms of action of **2**:556
  safety **2**:556
  substances **2**:553–555
precision nutrition, utility of omics in **4**:252
pregnancy **3**:459
  avoiding foods to prevent allergy **4**:219
  binge drinking and social alcohol consumption during pregnancy **4**:214–216
  complications **1**:513
  epigenetics and placenta **4**:181–182
  excessive alcohol consumption during pregnancy **4**:213–214
  fetal nutrient requirements **4**:177–178
  fish and pregnancy **4**:216–217
  food additives and herbal supplements **4**:220
  food-borne infections during pregnancy **4**:212–213
  and glycemic index **2**:427
  mechanisms of placental nutrient transport **4**:179–180
  methods used to study placental function **4**:179
  placental
    buffering of maternal dietary intake **4**:183
    insufficiency and fetal growth **4**:183–184
    metabolic activity **4**:182
    selectivity **4**:182
  recommended nutrient intakes for **4**:168–174
  role of fetus **4**:184
  weight gain
    and postpartum risk of obesity **4**:227
    recommendations **4**:223
pregnant women, diagnostic criteria for diabetes in **3**:230
preterm infants **4**:116
  causes of **4**:117
  energy needs **4**:108–109
  enteral nutrition **4**:111–114
  epidemiology of **4**:118–119, **4**:118f–119f
  fetal growth, assessment of **4**:116–117
  gestational age, assessment of **4**:116–117
  growth **4**:108
  health consequences of **4**:117
  monitoring nutritional status **4**:114–115
  nutritional interventions to prevent **4**:119
    intrauterine growth restriction (IUGR) **4**:119
  parenteral nutrition **4**:109–111
  prevention of **4**:118–119
proanthocyanidins (PAC) **2**:539, **2**:542
  and related compounds, health effects of **2**:539–544
probiotic(s) **2**:546–553, **3**:358–359, **3**:560
  commonly used **2**:548
  early identification of probiotic microbes **2**:547–548
  health benefits of **1**:214
  health outcomes reported with **2**:549–551
  history **2**:546–547
  infectious risk concerns **2**:551
  mechanisms of action of **2**:548–549
  microbes in foods **2**:547
  probiotic microorganisms **2**:547–548
  regulation of **2**:552–553
  safety **2**:551
    regulation **2**:551
  supplements **3**:516
processed culinary ingredients **2**:655
processed foods **2**:655
prolonged glucose load test **3**:405, **3**:407
proopiomelanocortin (POMC) **4**:56
prostaglandins (PG) **1**:235, **1**:425
  and related compounds **2**:93–103
prostate cancer **2**:484
protectins **2**:212
protein digestibility-corrected amino acid score (PDCAAS) **2**:562
protein efficiency ratio (PER) **2**:562
protein losing enteropathy (PLE) **3**:196
protein-bound biotin, digestion of **1**:86
protein-energy malnutrition (PEM) **2**:82
protein-sparing effect **2**:561–562
protein/energy ratio (P/E ratio) **2**:560–561
protein(s) **2**:14–16, **2**:120–121, **2**:301–302, **2**:452, **2**:480–481, **2**:517, **2**:560, **2**:569, **3**:120–121, **3**:244, **3**:342, **3**:451, **3**:526–527, **3**:575–576, **3**:617, **4**:37–38, **4**:110, **4**:136, **4**:170
  amino-acid pattern of metabolic demand **2**:576–577
  areas of uncertainty **2**:576
  concentration **2**:560
  deficiency **1**:368
    adaptation to low protein intakes **1**:371–372
    causes **1**:373
    children **1**:372
    elderly **1**:373
    pregnancy & lactation **1**:372
    principles of treatment **1**:373
    protein needs **1**:369
    protein quality **1**:370
    relationship between protein intake and needs **1**:370
  dietary protein requirement **2**:569
  digestibility **1**:383–385
    apparent and true digestibility **1**:384
    differences in **1**:384–385
    fecal and ileal digestibility **1**:383
  digestion
    digestion and fermentation in distal intestine **1**:382–383
    in intestinal lumen **1**:376–379
    peptides and amino acids **1**:379–382
    in small intestinal lumen **1**:378–379
    in stomach **1**:377–378
  intake **3**:396–397
    and bone **3**:82
  malabsorption **3**:481
  metabolic demands for amino-acids **2**:570
  metabolism **3**:285, **3**:649–650
    disorders of **3**:414–418
  needs **3**:176
  nitrogen balance studies **2**:572–576
  obligatory and adaptive components of metabolic demand **2**:570–572
  pathophysiology **3**:526
  protein/energy ratio **2**:560–561
  quality
    by amino acid scoring **2**:562–563
    and dietary sources **2**:563–565
    in human and animal **2**:561–562
  requirements **3**:526–527, **4**:128
  supplementation with **3**:635–637
  synthesis **1**:391–392, **4**:69–71
  terminology **2**:569–572
  turnover **1**:391–394, **4**:69–71
    extent and physiological implications of **1**:396–397
    models and tracer methods for **1**:394–396
proteolysis **1**:392–394, **1**:397
proteolytic environment **1**:396
proteomics **1**:565–566
proteostasis network (PN) **1**:388–389
  protein-stat mechanism **1**:390f
  in relation to protein turnover **1**:389f
  whole-body proteostasis **1**:389–391
provitamin A enriched crops **2**:290
prudent diets **4**:268

pulmonary system 4:159
pumpkin seeds 2:516
purine
  functions of 1:42
  ribonucleosides
    and bases as markers of food quality 1:338
    as flavor-enhancing additives 1:337
pyrethrins 2:382
pyrethroids 2:382
pyridoxine 1:497
  acute toxicity 1:497
  chronic toxicity 1:497
  deficiency 2:273
pyrimidine nucleotides, functions of 1:42

Q

qualitative techniques 3:46, 3:65
quantitative analysis 3:46–47
quantitative magnetic resonance (QMR) 3:22–23
quinolinic acid, functions of 1:44

R

radioactive compounds 3:128–129
radioallergosorbent test (RAST test) 3:339
reactive hypoglycemia, management of 3:408–409
  non-hypoglycemia 3:409
  prevention 3:408–409
  treatment of attacks 3:408
reactive oxygen species (ROS) 4:56
  reaction with 1:148
recommended daily allowance (RDA) 1:182, 1:478–479, 4:146–147, 4:379
recommended nutrient intakes (RNIs) 4:45
  rationale for 4:91–92
rectum 2:310
red meat 3:682–683
reference nutrient intakes (RNIs) 4:146–147
Registered Dietitian Nutritionist (RDN) 3:236
regulatory monitoring documents 4:293
Relative Reinforcing Value of food task (RRV task) 3:65
relative risk (RR) 1:358
renal failure 3:409–410
renal system 4:160
renin angiotensin aldosterone system (RAAS) 1:425
renin–angiotensin system 3:594
requirements
  dietary recommendations 4:92–94
resistant starch (RS) 2:416–417
  slowly digested and 2:419
resolvins 2:210
respiration related water loss 3:219
respiratory acidosis 2:111
respiratory alkalosis 2:112
respiratory bacterial infection, vitamin D and 1:541
respiratory system 3:585, 3:595, 4:159
respiratory tract infections and micronutrients 3:476
resting energy expenditure (REE) 1:397, 3:28, 3:30t–31t
resting metabolic rate (RMR) 2:573
resveratrol 2:184, 3:638
retinoic acid in intestine 1:475
retinoic acid receptors (RAR) 2:205–206
retinol-binding protein (RBP4) 1:473–474
reverse cholesterol transport 3:373
rheumatoid arthritis (RA) 1:535, 1:538–539
  See also inflammatory arthritis
  nutritional assessment in 3:111
riboflavin 2:233, 3:10, 4:40
  absorption
    in humans 1:403
    transport, and storage 1:402–403
  assessment of riboflavin status 1:407–408
  cataracts and photoreceptors 1:406
  dietary sources and high intakes 1:409–410
  effect on folate metabolism 1:407
  essentiality 1:405–407
  fatty acid oxidation 1:406
  functions 2:233
  glutathione reductase test 1:407–408
  interaction with vitamin B_6 1:406
  iron economy 1:406
  malaria 1:406
  metabolic functions 1:405
  metabolism 2:233
    and excretion 1:403–405
  requirements 1:408–409
  transport at sites and storage 1:403
  urinary excretion 1:407
rice (*Oryza Sativa*) 2:295
  special case of rice fortification 2:343
rickets 4:124–125
rodents, pharmacological effects in 1:185–186
RRR-αT 1:558–559
rubidium 1:454–455
rye (*Secale cereale*) 2:296

S

S-adenosylmethionine (SAM) 1:165, 2:180, 2:182
saccharide prebiotic candidates 2:555
saccharin 2:628
Salmonella infections 2:368–369
salt
  intake and blood pressure 2:581–588
    epidemiological studies 2:583–585
    intervention studies 2:585–588
  intake and cardiovascular outcomes 2:588–593
  sensitivity 2:582
sarcopenia
  adaptation of autophagy-linked signaling in muscle with age 3:634
  molecular mechanism regulating sarcopenia 3:634
  no significant change of UPS in sarcopenic muscle 3:634
  supplementation with protein and amino acids 3:635–637
SARS-CoV-2 infection, vitamin D and 1:543
satiating efficiency 2:138–139
satiating power 2:138–139
satiety 2:58, 2:138–139
  cascade 2:60
  macronutrients and energy density affect 2:8
  impact of ultra-processed foods on 2:64–65
saturated fatty acids (SFA) 2:72, 3:383, 4:263
scavenger receptor family (SR family) 3:370–371
Schedule for Oral Motor Assessment (SOMA) 3:178
Schistosomes 3:543
school food systems 4:343
  for children and adolescents 4:342–343
  determinants of food systems for children and adolescents 4:342–343
  and food programs 4:343–345
  interventions 4:345–347
    and impact on anthropometrics and metabolic outcomes 4:346
    and impact on dietary behaviors 4:346
    and sustainability 4:347
scurvy 1:512, 1:515–516
seasonality of nutrition
  defining and measuring seasonality 4:352–354
  food supply, demand, and prices 4:355–357
  food systems 4:354–357
  interactions of climatic seasonality with multiple systems 4:354–360
  origins of seasonality 4:352
  quantifying seasonality 4:354
  spatial units 4:353–354
  temporal units 4:353
  timekeeping 4:353
secondary lactase deficiency 3:445–446
secondary obesity 3:585
sedentary behavior 1:352–353, 1:357
seeds 2:516
  contaminants 2:520
  macronutrient content of 2:516–517
  types of 2:514
selenium 1:72, 1:198, 3:15, 4:39, 4:174
  assessment of selenium status 1:417
  deficiency diseases 1:416
  excretion 1:416
  functions of 1:413
  homeostasis 1:414–416
  and human health 1:418–419
  intestinal absorption and bioavailability 1:414
  metabolism and distribution 1:415–416
  mutations and inborn errors of selenoprotein biosynthesis and functions 1:416
  nutritional supply of 1:414
  requirement 1:416–417
    and recommended dietary intakes 1:417
  toxicity 1:417
self-monitoring 3:732

self-monitoring of blood glucose (SMBG) **3**:235–236
semielemental formulas **3**:559
"sensible" sweat **3**:218
sepsis **3**:410
serotonin, functions of **1**:43
serum lipids, direct effects of sucrose on **2**:600
sesame seeds **2**:516
Seven Countries Study (SCS) **2**:459
severe acute malnutrition (SAM) **3**:496–500
  immunity and **3**:498
severe autoimmune diabetes (SAID) **3**:257
severe insulin-deficient diabetes (SIID) **3**:257
severe insulin-resistant diabetes (SIRD) **3**:257
SGLT-2 inhibitors **3**:250
Shigella species **2**:371
short bowel syndrome **3**:482–483
short women **4**:226
silicon **1**:198–199, **1**:455
single nucleotide polymorphisms (SNPs) **3**:256–257
"6Rs" method **4**:279
skeletal muscle mass (SM) **3**:25–26
skin, production of vitamin D in **1**:524
skinfold thickness **3**:89–90
small for gestational age (SGA) **4**:76
small intestine **2**:309
smoking **4**:9
social determinants of oral diseases **3**:276
social determinants of UPF **2**:658–659
sodium **1**:192, **3**:12, **3**:83, **3**:617, **4**:38, **4**:266
  distal tubule **1**:423–424
  ECV **1**:422
  intake and blood pressure **2**:581–588
  loop of Henle **1**:423
  nitrite **3**:347
  osmolarity **1**:422
  proximal tubule **1**:423
  role of skin **1**:422
  sodium excretion via kidney **1**:422–424
  systemic regulation of ECV and total body sodium **1**:424–425
sodium-dependent glucose co-transporter (SGLT1) **2**:198
soft drinks **2**:282–283
soluble non-starch polysaccharides (NSPs) **2**:424
sorbitol **1**:120–121
sorghum (*Sorghum bicolor*) **2**:295
soy isoflavone **3**:637–638
Spanish Pediatric Association **4**:146
specific amino acids **3**:529
sphingomyelin (SM) **1**:162–163, **1**:165
stakeholder groups, role of different **2**:621–622
  role of local government **2**:621
  role of social movements and role models **2**:621–622
staphylococcal food poisoning (SFP) **2**:364–365
starch **1**:122, **1**:137, **2**:399, **2**:414
  digestion **1**:127
  mitigation of **2**:542
starvation **3**:410
  metabolic adaptation
    in post absorptive state **3**:651–655
    in prolonged fasting **3**:651–652
    in starvation **3**:652–653
  metabolic consequences of **3**:650–651
  regulation of adaptation in **3**:653–655
stearoyl-CoA desaturase (SCD1) **2**:199
sterol regulatory element binding proteins (SREBPs) **3**:370
  SREBP-1c **2**:201
sterol regulatory elements (SRE) **2**:201
sterols **1**:296
steviol glycosides **2**:628
stomach **2**:309
stomatitis **3**:140
storage **1**:285
Streptococcal pharyngitis **2**:372
streptozotocin (STZ) **4**:55
stress **3**:250
strict diets **3**:236
stricturing disease **3**:690
stroke nutritional management
  artificial nutritional support **3**:667–668
  attention span, short-term memory **3**:661
  communication problems **3**:660
  evaluation of nutritional support **3**:668
  evidence-based guideline recommendations **3**:659–660
  impairments of arm movement and posture **3**:660–661
  management of stroke-related psychosocial and physical problems impairing food consumption **3**:659–668
  organizational factors **3**:659
  poststroke eating problems **3**:658
  prestroke nutritional status **3**:658
  psychosocial problems **3**:660
  risks of protein-energy malnutrition **3**:658
  screening and assessment **3**:661–662
  swallowing difficulties **3**:661–664
  visual field loss and visual neglect **3**:661
Strongyloides stercoralis **3**:543
strontium **1**:455
structured low-calorie diets **3**:727–730
stunting syndrome **4**:230–231
  mortality and developmental outcomes with **4**:233
  prevalence and global and regional trends **4**:231–233
  preventive interventions for **4**:235–238
  risk factors **4**:233–235
sub-acute neuroglycopenia **3**:404
subcutaneous adipose tissue (SAT) **3**:25
Subjective Global Nutrition Assessment (SGNA) **3**:172
subscapular skin fold (SSSF) **3**:175
substance abusers **4**:226
substrate-specific requirements **3**:526–527
sucralose **2**:628
sucrase-isomaltase (SI) **2**:198
sucrose **1**:118, **3**:480
  and added sugars **2**:597
  association consumption with metabolic diseases **2**:598–599
  as caloric sweetener **2**:597–598
  complex carbohydrate *vs.* **3**:242
  consumption and macronutrient intake **2**:600–601
  and dental health **2**:601–602
  direct effects
    on glucose metabolism **2**:599
    on serum lipids and inflammation **2**:600
    of sucrose on hepatic lipid metabolism **2**:600
  effects of naturally occurring **2**:602–603
  in fruits **2**:602
  intake and reduced satiety **2**:597–598
  interaction consumption with intake of nutrients **2**:600–601
  and intestinal microbiome **2**:602
  metabolic effects of **2**:598–600
  metabolism **2**:598
    by components of microbiome **2**:601–602
sugar alcohols **1**:118–121
sugar sweetened beverages (SSBs) **2**:283, **2**:597, **2**:630–631, **3**:710, **4**:145, **4**:408–409
sugar(s) **1**:137, **2**:399, **2**:627–628, **3**:678
sulfonylureas **3**:249
sulfur amino acid metabolism **1**:493–494
sulfur compounds **2**:406–408
sunflower seeds **2**:516
Superior Mesenteric Artery Syndrome (SMA) **3**:178
supplemental nutrition information **4**:303–304
supplementation **3**:241
  of chromium **1**:185
  iodine supplementation **4**:454–455
  iron/folate supplementation **4**:456
  multiple micronutrient supplementation **4**:457–458
  program cycle **2**:468–476
  zinc supplementation **4**:457
surgical model of UPI **4**:54
sustainable diets **2**:613
  advancing sustainable diets agenda **2**:621–622
  consumer-facing solutions **2**:618–620
  global sustainable diet **2**:613–615
  guide and influence **2**:620
  incentivize, discourage or restrict **2**:620
  inform and empower **2**:618–620
  measuring **2**:615–617
  rationale for promoting **2**:613
  research gaps and recommendations for moving toward **2**:622
  shifting consumption toward **2**:617–618
  solutions to help support **2**:618–620
  supply-side solutions **2**:620
sustainable food
  ecosystem **2**:645–646
  systems **3**:489
sweat **3**:218
  hedonics **2**:629
  taste receptor **2**:627

sweeteners **2**:625, **3**:244
and appetite **2**:629
consumption trends **2**:630–632
digestion, absorption and metabolism **2**:627t
and health **2**:632–635
taste, digestion, and appeal **2**:625–630
types **2**:625t
sweetness **2**:625
sword beans **2**:432
sympathetic nervous system **1**:424–425
synthetic analogs of vitamin E **1**:559–560
systems-level programs **3**:733

T

T-2 toxin (T2 toxin) **2**:498
targeted-fortification **2**:344
taste receptor type 1 receptor 2 (T1R2) **2**:625
taste receptor type 1 receptor 3 (T1R3) **2**:625
taurine, functions of **1**:42
tea **2**:278–279
technology **3**:733
telomeres and telomerase **4**:158
terminal ileum **3**:203
tetrahydrobiopterin **3**:429
The International Dysphagia Diet Standardization Initiative Framework (IDDSI Framework) **3**:180
thermoregulation **3**:218
thiamin **2**:80, **2**:236–238
absorption and ethyl alcohol **1**:443
accidental thiamin depletion in infants **1**:433–435
alcohol abuse **1**:435–436
assessment of thiamin status **1**:445–446
biological functions **1**:444–445
case study **1**:437–438
clinical features of beriberi **1**:433
dietary sources of **1**:441–442
drug–nutrient interactions **1**:446
epidemiology **1**:429–431
etiology **1**:431–432
experimental thiamin deficiency in man and measurement of thiamin status **1**:432–433
functions **2**:236
infantile beriberi **1**:433
lipid-soluble thiamin derivatives **1**:437–438
low-dose thiamin supplements in lactating Cambodian mothers **1**:443
management/treatment **1**:436
metabolism **2**:237–238
recommended dietary allowances **1**:446
toxicity **1**:438, **1**:447
transport, storage and excretion **1**:443–444
thiamine **3**:9–10, **4**:40
deficiency **2**:271–272
Thiazolidinediones **3**:249
three-compartment model (3C model) **3**:24
three-dimensional photonic scanner (3DPS) **3**:24
thromboxanes and related compounds **2**:93–103
thyroglobulin concentrations **1**:278–279
thyroid hormone concentrations **1**:278–279
thyroid hormone receptor (TR) **2**:205–206
thyroid volume **1**:278
tibial length (TL) **3**:174
tin **1**:456, **2**:358–359
dietary exposure **2**:358
permissible intakes **2**:359
toxicity **2**:359
toxicokinetics **2**:359
tissue(s)
and organs **3**:25
retinoid metabolism **1**:476–477
titration of volume **3**:332
TMAO and cardiovascular disease **1**:169
tolerable daily intake (TDI) **1**:184
tolerable upper limit (TUL) **2**:574–575
tonicity **3**:219
total body potassium (TBK) **3**:22
total body water (TBW) **3**:53
total parenteral nutrition (TPN) **1**:182
toxicity
of chromium **1**:184–185
of exogenous nucleic acids to humans **1**:338
toxins **2**:519–520
trace elements (TE) **3**:241, **3**:332, **3**:452, **3**:530, **4**:174, **4**:192–193
See also ultratrace elements
trace metals **2**:207
trace minerals **4**:111
tracer methods of indirect calorimetry **2**:158–159
tradition and indigenous foods promote ecosystem benefits and genetic diversity **2**:646
trans fatty acids **2**:72, **3**:385–386
trans-fatty acids (TFA) **2**:648, **4**:263–264
dietary recommendations, regulations, and dietary intake **2**:649–650
health effects **2**:648–649
trans fat alternatives **2**:650–651
transcription factor E3 (TCFE3) **2**:201
transcriptional regulation **2**:207
transfer proteins **3**:368
transferrin **1**:183
transmembrane domain (TMD) **2**:625
transport **1**:252–253, **1**:285
and cellular uptake **1**:253
of chromium **1**:183–184
of vitamins **2**:217–222
triacylglycerol **1**:294–295, **3**:363
triazoles **2**:383–384
tricarboxylic acid cycle **2**:121–122
triceps skin fold (TSF) **3**:172
Trichuris trichiura **3**:543
triglycerides **1**:294–295, **2**:123–124
tropical sprue **3**:268
tryptophan metabolism **1**:494
tuberculosis (TB) **3**:549, **3**:698
controlled intervention studies of nutritional supplements in management of **3**:700–701
nutrient-drug interactions **3**:703
nutrition and susceptibility to **3**:549–555
effect of nutritional factors on course of **3**:700
nutritional management of **3**:701–703
nutritional status of tuberculosis patients **3**:699
two-compartment model (2C model) **3**:24
type 1 diabetes **1**:134, **3**:245–247
heterogeneity in **3**:231
Type 1 diabetes mellitus (T1DM) **3**:253–255
type 2 diabetes (T2D) **1**:72–73, **1**:134, **2**:634–635, **2**:679–680, **3**:227–228, **3**:235, **3**:247–248, **3**:591
carbohydrates and **1**:140
epidemic of **4**:442
etiology and management of **2**:311
fructose consumption and **2**:394
glycemic index and prevention of **2**:427
in later life in macrosomic offspring **4**:80
heterogeneity in **3**:231–232
Type 2 diabetes mellitus (T2DM) **1**:535, **1**:538, **2**:633, **3**:253, **3**:255–259

U

ubiquitin–proteasome system (UPS) **1**:388–389, **1**:392, **3**:634
in sarcopenic muscle **3**:634
UDP-glucuronosyl-transferases (UGTs) **2**:83
ulcerative colitis **3**:202
ultra-processed food (UPF) **2**:654–655, **4**:276
and chronic diseases **2**:658
on diet quality **2**:656
food processing **2**:653
global participation of **2**:656
on health **2**:656–658
NOVA classification system **2**:654–655
NOVA in use **2**:659
social determinants of **2**:658–659
ultratrace elements **1**:448–459
underfeeding **3**:247
undernutrition **3**:490
causes of **3**:492–493
undigested glycosidic bonds **1**:128
unfolded protein response pathways (UPRP) **3**:256
United National Children's Fund (UNICEF) **4**:235
unprocessed foods **4**:276
unsaturated fatty acids **3**:383–386
upper AMA **3**:92
upper gastrointestinal evaluation (UGI) **3**:178
upper intake levels (UL) **1**:478–479
upper limit (UL) **1**:184
upstream stimulatory factors (USFs) **2**:201–202
urea cycle disorders **3**:415–416
urinary iodine concentration **1**:278
urogenital system **4**:160
urolithiasis and kidney stones **1**:339
Ursolic acid (UA) **3**:639

US Food and Drug Administration (FDA) **1**:185, **2**:82, **2**:630
US National Health and Nutrition Examination Survey (NHANES) **1**:356
uteroplacental insufficiency (UPI) **4**:54

V

VA deficiency disorders (VADD) **1**:461, **1**:463
vanadium **1**:197, **1**:456
variable number tandem repeat (VNTR) **3**:254–255
vasculature, colon **3**:185
vegan diets **3**:241, **4**:146
vegetables **3**:157, **3**:519, **4**:266
  carbohydrates **2**:401–403
  classification of **2**:401
  energy **2**:401
  health effects of **2**:408–411
  juice **2**:282
  lipids **2**:403
  minerals **2**:404
  nitrogenous compounds **2**:403
  nutritional components of **2**:401–404
  other components in **2**:404–405
  vitamins **2**:403–404
  water **2**:401
vegetarian diet(s) **2**:663–665, **3**:241, **3**:396, **4**:146, **4**:268
  differences between vegetarian eating and vegetarianism **2**:665–670
  fasting and **3**:113–114
  morbidity and mortality on vegetarian *vs.* omnivorous diets **2**:667–670
  nutritional adequacy **2**:670–672
  nutritional features of well-planned vegetarian diets **2**:666–667
  vegetarian patterns and practices **2**:665
vegetarian eating, differences between vegetarianism and **2**:665–670
vegetarianism, differences between vegetarian eating and **2**:665–670
very low-calorie diets (VLCDs) **3**:730
Very low-carbohydrate diet **3**:240
very low-density lipoproteins **3**:367
vetches **2**:432
Vibrio
  V. cholerae **2**:367
  V. parahaemolyticus **2**:371–372
  V. vulnificus **2**:372
viral diseases **3**:267
viral hepatitis **3**:454
virus infections, vitamin D and **1**:541
visceral adipose tissue (VAT) **3**:22, **3**:25
vitamin A **1**:461, **1**:467, **2**:205–207, **2**:224–227, **2**:289–290, **3**:5–9, **3**:83, **3**:213, **4**:41, **4**:96, **4**:171, **4**:216
  absorption and metabolism **1**:474–476
  anemia and poor growth **1**:463
  breastfeeding and diet **1**:466
  causal agents **1**:465–466
  conditions of excess **1**:480
  and cornea **1**:477–478
  dietary sources and nutritional equivalency **1**:478–479
  epidemiology **1**:463–469
  functions in cell differentiation **1**:478
  hepatic vitamin A uptake, storage and release **1**:475
  immunity **1**:478
  infection **1**:463, **1**:466
  impact of interventions **1**:467–468
  kinetics **1**:476
  major molecules in vitamin A biochemistry and physiology **1**:472–473
  management **1**:468–469
  metabolic disturbances affect plasma retinol **1**:476
  metabolites formed through oxidation **1**:473
  mortality and morbidity **1**:467–468
  person **1**:464
  physiological actions **1**:477–478
  place **1**:464
  plasma concentrations **1**:475–476
  polar compounds **1**:473
  prevention **1**:469
  relationship of plasma retinol to liver retinol concentration **1**:476
  supplementation **4**:455–456
  supplementation in public health **1**:480
  time **1**:465
  tissue retinoid metabolism **1**:476–477
  transport **1**:473–474
  treatment **1**:468–469
  vision **1**:477–478
vitamin B_{12} **2**:244–246, **3**:11, **3**:83, **3**:686, **4**:40
  biochemistry and metabolic functions **1**:482–483
  causes, mechanisms, and effects of cobalamin deficiency **1**:485–486
  deficiency **2**:273
  diagnosis of cobalamin deficiency **1**:486–487
  inborn errors of cobalamin metabolism **1**:487
  metabolism **2**:245–246
  physiology **1**:484–485
vitamin B_6 **2**:238–239, **3**:10, **4**:40
  assessment of **1**:495
  chemical structures of B6 vitamers and dietary sources **1**:490
  dietary constituents cause PLP deficiency **1**:496
  digestion and absorption **1**:490
  functions **2**:238–239
  genetic disorders affecting vitamin B6 metabolism and disorders treated with **1**:498
  glycine and serine metabolism, 1-carbon and folate metabolism **1**:493
  heme synthesis **1**:494
  infection **1**:500
  inflammation **1**:499
  metabolism **2**:239
    and transport **1**:491–492
  pyridoxal phosphate **1**:497
  reactions catalyzed by PLP enzymes **1**:493–494
  reactivity of PLP **1**:494
  recommended daily intake **1**:495–496
  roles **1**:494
    of vitamin B6 in common diseases **1**:498–500
  toxicity of **1**:496–497
vitamin C **1**:71, **2**:406, **3**:11, **3**:398, **4**:40, **4**:139, **4**:172
  absorption, transport, and storage **1**:505
  assessment of vitamin C status **1**:509
  chemical structure **1**:505
  degradation, turnover, and factors that induce increased requirements for **1**:517
  and demethylation of DNA **1**:511–512
  and demethylation of histones **1**:512
  and disease **1**:512–513
  and gene expression **1**:511–512
  measurement of vitamin C status **1**:518–519
  metabolic functions of ascorbic acid **1**:507–509
  metabolism and excretion **1**:506
  occurrence of low intakes and poor biochemical status in present-day societies **1**:519–520
  requirements **1**:509–510
    estimated from maintenance of body pool of ascorbate **1**:510
  well-established metabolic functions of vitamin c are impaired by deficiency **1**:517–518
vitamin D **2**:205–207, **2**:229–230, **3**:9, **3**:81, **3**:213, **3**:517, **3**:615–616, **3**:639, **3**:688, **4**:41, **4**:147, **4**:171–172
  absorption, metabolism, and excretion of **1**:526
  and acute diseases **1**:540–543
  biologic functions on calcium metabolism **1**:526
  and calcium studies on bone **3**:81
  and chronic diseases **1**:536–539
  and critically ill patients **1**:542–543
  deficiency **4**:124–125
  evaluation for and consequences of vitamin D deficiency **1**:526–528
  healthy vitamin D intakes **1**:531
  and human immunodeficiency virus infection **1**:541–542
  and immune system **1**:540
  nonskeletal consequences of vitamin D deficiency **1**:528–529
  origin and structure of **1**:523–524
  production of **1**:524
  recommendations for adequate intake
    for ages 0–12 months **1**:529–530
    for ages 1–18 years **1**:530–531
    for pregnancy and lactation **1**:531
  recommended adequate intake
    for ages 19–70+ years **1**:531
  and respiratory bacterial infections **1**:541

vitamin D (*continued*)
and risk of falling **3**:81
tolerable upper intake levels and vitamin D intoxication **1**:531–532
and virus infections **1**:541
vitamin D receptor (VDR) **2**:205–206
vitamin E **1**:71–72, **1**:552, **2**:228–229, **2**:406, **3**:9, **3**:214, **4**:41, **4**:172
absorption and plasma transport **1**:548–549
actions **1**:547–548
adverse effects of pharmaceutical drugs on vitamin E status **1**:553
antioxidant activity **1**:547
bioavailability **1**:548–550
biologic activity **1**:547–548
catabolism **1**:550–551
chemistry of eight natural vitamin E analogs **1**:557–558
chronic disease prevention **1**:553
deficiency **1**:564
dietary vitamin E **1**:551–552
and disease prevention **1**:565–566
general description and scientific name **1**:546
human vitamin E deficiency **1**:551
increased tendency to bleed **1**:552–553
metabolites **1**:561
molecular function **1**:548
molecular mechanisms of vitamin E action **1**:565
metabolism and excretion **1**:563–564
with MUFA and PUFA, interaction of **1**:562–563
natural and synthetic analogs of **1**:559–560
over-dosage **1**:552
plasma vitamin E kinetics **1**:550
plasma α-tocopherol concentrations and tissue delivery **1**:549
polymorphisms relevant for vitamin E action **1**:566
precautions and adverse reactions **1**:552–553
recommended intake levels **1**:551–552
RRR-αT **1**:558–559
supplements **1**:552
vitamin K **2**:227–228, **3**:9, **3**:82, **3**:214, **3**:616
absorption and metabolism **1**:571–572
antagonists **2**:228
blood coagulation, osteocalcin, and Gla-matrix proteins **2**:227–228
dietary requirements **1**:574
enzymatic **1**:572–573
forms **1**:570
functions **1**:572–573
and mechanisms **2**:227
and gut microbiome **1**:575
non-enzymatic **1**:573
sources **1**:570–571
status
biomarkers **1**:574
and chronic disease **1**:575
health & disease **1**:574–575
in newborn **1**:574
vitamin(s) **2**:183–184, **2**:328, **2**:400, **2**:452–453, **2**:483–484, **3**:121–123, **3**:241, **3**:285, **3**:332, **3**:357, **3**:428, **3**:530, **4**:18–19, **4**:94–96, **4**:130, **4**:191–192
biochemical mechanisms and regulation of vitamins and vitamin-like cofactors
general principles **2**:217–222
physiology **2**:217–222
whole-body, organ, and cellular transport **2**:221–222
enteral nutrition **4**:114
mineral supplementation and probiotics **3**:113
parenteral nutrition **4**:111
perspectives on digestion, absorption, transport, and metabolism of **2**:217–222
requirements **2**:222–224
status in CF **3**:213–214
volume of distribution (Vd) **2**:85
vomiting **3**:139

W

waist and hip circumferences **3**:90
waist-to-hip ratio (WHR) **3**:93
walnut (*Juglans* spp.) **2**:516
water **3**:706, **3**:708–709, **4**:38, **4**:152–153, **4**:174
acquisition **3**:710
and electrolytes **4**:139
insecurity **3**:707–708
domains of **3**:707–708
and nutrition **3**:708–712
physiology of water balance **3**:217–219
security **3**:706
for food production **3**:710
for healthy diet **3**:710–711
water, sanitation, and hygiene (WASH) **3**:712
water-soluble B vitamins **4**:139
water-soluble vitamins **2**:230–236, **3**:122–123, **3**:214, **4**:40
See also fat-soluble vitamins
cofactor, coenzyme, and cell signaling functions **2**:230–231
oxidation-reductions **2**:231
Web Buffet **3**:63
weighed food records (WRs) **3**:45
inventory and estimated food record **3**:45
weight
gain
changes in body composition and maternal energy status **4**:225
excessive and inadequate weight gain **4**:224–225
global weight gain standards **4**:223
pattern of **4**:225
pregnancy weight gain recommendations **4**:223
for special population groups **4**:226
impact of supplementation **4**:227
loss maintenance
behavioral component **3**:731–733
diet **3**:727–731
dietary strategies and **3**:731
difficult **3**:725
energy balance **3**:727
experimental studies examining **3**:727
improving maintenance of physical activity **3**:731
incentives **3**:733
NWCR **3**:726
observational studies of **3**:725
physical activity **3**:731
prevalence of long-term weight loss maintenance **3**:725
psychological factors associated with **3**:726–727
social support **3**:732
management **3**:724–736
health outcomes **3**:721
metabolic hypothesis **3**:717–718
methodological issues **3**:721
psychological consequences **3**:721
weight cycling **3**:717
and cardiovascular disease **3**:720
and mortality **3**:719–720
weight-based dosing **2**:85
Wernicke-Korsakoff syndrome **1**:435–436
wheat **2**:295, **3**:678–680
Whipple's disease **3**:268
Whipple's triad **3**:405
white adipose tissue (WAT) **3**:95
whole body protein turnover (WBPT) **1**:392
whole genome sequencing (WGS) **3**:256–257
whole grains
and adiposity **2**:684–685
antioxidants, and phytochemicals **2**:686
biomarkers of whole grain consumption **2**:678–679
bran **2**:677
consumption and disease risk **2**:679–684
dietary fiber, and fermentable carbohydrates **2**:685–686
endosperm **2**:677
germ **2**:677
and health claims **2**:677–679
and hypertension **2**:685
and inflammatory status **2**:685
intermediate risk factors of disease and potential mechanisms of action **2**:684–686
recommendations for whole grain consumption and consumption patterns **2**:678
structural components and composition of grains **2**:676–677
whole-body counting **3**:22
whole-body protein-turnover (WBPT) **1**:394
whole-body proteostasis **1**:389–391
winged beans **2**:432
Women's Health Initiative (WHI) **4**:279
World Health Organization (WHO) **2**:630, **3**:253, **4**:43, **4**:85, **4**:144, **4**:230–231
low birth weight **4**:116
prematurity **4**:116

World War I (WWI) **4**:274

X

xanthine oxidase (XO) **4**:56
xerophthalmia **1**:461–463, **1**:467
xerostomia **3**:140–141
xylitol **1**:120
xylulose-5-P (X5P) **2**:201

Y

Yersinia enterocolitica **2**:371
"yo-yo" effect **2**:180–181

Z

Z-score **4**:87
zearalenone (ZEA) **2**:498
zeaxanthin **1**:159–160
zinc **1**:194–195, **3**:12–13, **3**:616, **3**:688, **4**:39, **4**:95–96, **4**:173
  absorption **1**:585–586
  assessment **1**:591
  biochemistry **1**:587–588
  causes of zinc deficiency **1**:577–579
  chemistry of zinc **1**:585
  child growth **1**:580
  consequences of zinc deficiency in developing countries **1**:580–581
  control of zinc homeostasis **1**:585–587
  deficiency **2**:273
  dietary diversification and modification **1**:582
  enriched crops **2**:290
  enzymes **1**:588
  excretion **1**:587
  in foods **1**:585
  growth **1**:589
  high physiological requirements **1**:579
  homeostasis **1**:587
  inadequate dietary zinc intake **1**:577–578
  intervention strategies **1**:581–582
  mass fortification of staple foods **1**:581–582
  mild **1**:590
  morbidity and mortality **1**:580
  as nutrient **1**:585
  physiology **1**:588–590
  pregnancy **1**:580–581
  prevalence of zinc deficiency in developing countries **1**:579
  preventive zinc supplementation **1**:581
  recommended intakes **1**:591
  reproduction **1**:589
  severe **1**:590
  targeted fortification **1**:582
  therapeutic zinc supplementation in treatment of diarrhea **1**:581
  toxicity **1**:590–591
  transcription factors **1**:588
  transport and distribution **1**:586
zinc transporter 8 (ZnT8) **3**:253

This page intentionally left blank

AUTHOR INDEX

A

Aggett, Peter J.
Copper

Aguilera, Concepción M.
Nutriepigenetics
Nutrigenetics
Nutrigenomics

Aje, Kent
Enteral and parenteral nutrition

Akonde, Maxwell
Cancer: Epidemiology of lung cancer

Albala, K.
Religious customs, influence on diet

Alberg, Anthony J.
Cancer: Epidemiology of lung cancer

Alghamdi, M.
Nutritional aspects of bone

Allen, L.H.
Aging
Magnesium
Malnutrition: Secondary, diagnosis and management

Álvarez-Mercado, Ana I.
Vitamin D: Role in chronic and acute diseases

Anderson Steeves, Elizabeth T.
Weight management: Weight maintenance

Anguita-Ruiz, Augusto
Nutriepigenetics
Nutrigenetics
Nutrigenomics

Aparicio, A.
Calcium

Appel, L.J.
Hypertension: Dietary factors
Potassium

Aravena Rivas, Yanela
Diet and oral health

Ashby-Thompson, M.
Body composition

Atkinson, R.L.
Obesity: Complications

Atkinson, Stephanie A.
Nutrition guidance for infants: Nutrient-based reference intakes and feeding recommendations

B

Baer, D.J.
Trans-fatty acids: Health effects, recommendations, and regulations

Barad, Alexa
Osteoporosis nutritional factors

Barnett, Y.
Aging

Bassani, Diego G.
Stunting: Prevalence and prevention

Bates, Christopher J.
Niacin and pellagra
Pantothenic acid
Riboflavin
Vitamin C: Deficiency states

Beaulieu, K.
Hunger

Bender, David A.
Vitamin C

Bermejo, L.M.
Carotenoids: Chemistry, sources and physiology

Beyh, Yara S.
Diabetes mellitus: Etiology and epidemiology

Bhupathiraju, Shilpa N.
Coronary heart disease: Dietary patterns
Whole grains and chronic disease risk

Bhutta, Zulfiqar
Pregnancy: Pre-eclampsia and diet

Bines, Julie E.
Starvation and fasting: Biochemical aspects

Blakely, Octavia Pickett
Nutritional support: Adults, enteral

Blundell, J.E.
Appetite: Psychobiological and behavioral aspects
Diet and eating behavior: Appetite control and satiety

Boost, Martijn
Brown adipose tissue: Implications for human health

Booth, S.L.
Vitamin K

Boullata, Joseph I.
Drug-nutrition interactions

Bournot, A.R.
Nutritional aspects of bone

Brand-Miller, J.
Glycemic index

Brosset, Sophie
Nutrition and burn injury — manuscript

Buckland, N.J.
Diet and eating behavior: Appetite control and satiety

Bueno-Lozano, G.
Obesity: Childhood obesity

C

Caballero, B.
Clinical examination
Food intolerance

Caffrey, Aoife
Pregnancy: Prevention of neural tube defects

Calder, P.C.
Cytokines: Metabolic and nutritional aspects
Dietary modulation of inflammation

Caldwell, K.L.
Biochemical indices

Cameán, A.M.
Cancer: Carcinogenic substances in food

Campos-Ponce, Maiza
Nutrition and parasitism

Carrey, E.A.
Nucleic acids, purine, and pyrimidine nucleotides and nucleosides: Physiology, toxicology, and dietary sources

Cassidy, A.
Phytochemicals: Classification and occurrence

Casuso, Rafael A.
Antioxidants: Intervention studies

Cegielski, J. Peter
Nutrition and susceptibility to tuberculosis
Tuberculosis: Nutritional management

Charrondiere, U.R.
Food composition data 2021

Chen, Oliver
Nutraceuticals: Health effects and clinical applications

Chen, Liwei
Beverages and health

Chowanadisai, Winyoo
Organic: Biochemical mechanisms and regulation of vitamins and vitamin-like cofactors

Chung, S.
Body composition

Clayton, P.T.
Vitamin B_6

Cliffer, Ilana R.
Seasonality of nutrition

Colditz, Graham A.
Cancer: Epidemiology and associations between diet and cancer

Collier, S.
Nutritional support: Infants and children, parenteral

Cook, Chad
Indirect calorimetry

Cuesta Triana, F.
Arthritis

D

Dalton, M.
Diet and eating behavior: Appetite control and satiety

Danta, Mark
The role of nutrition in inflammatory bowel disease: Disease associations, management of active disease and maintenance of remission

Dart, Hank
Cancer: Epidemiology and associations between diet and cancer

Dary, O.
Food fortification: Technological aspects

David, T.J.
Food allergies: Diagnosis and management

Davidson, Ryan D.
Pediatric feeding disorders: Feeding children who can't or won't eat

Davidsson, L.
Bioavailability

Davis, Heather A.
Eating disorders: Bulimia nervosa

Demmler, Kathrin M.
Food security and nutrition surveillance in low- and middle-income countries

de Pee, Saskia
Food security

Derbyshire, E.
Adolescents: Nutrient requirements

Dhareshwar, Shashank S.
Dehydration

Doocy, Shannon
Famine: Causes, consequences and responses

Dornhorst, A.
Glycemic index

dos Santos Leffa, Paula
Ultra-processed foods

Dowhan, Lindsay
Enteral and parenteral nutrition

Downs, Shauna M.
Sustainable diets: Their definition, measurement and promotion

Dowsett, J.
Cystic fibrosis

Drüeke, Tilman B.
Salt, hypertension and cardiovascular outcomes

Díez, Julia
Food environment

Duggan, C.
Malabsorption syndromes: Nutritional management

Dulloo, Abdul G.
Energy: Adaptation

Dwyer, J.
Vegetarian diets

Dwyer, Johanna T.
Food culture

Dyall, S.C.
Effects of diet on behavior

E

Eaton-Evans, J.
Nutritional assessment: Anthropometry

Elango, Rajavel
Protein deficiency

Elia, M.
Nutritional support: In the home setting

Ensunsa, J.L.
Manganese

F

Fajó-Pascual, Marta
School food systems

Fenstrom, M.H.
Brain and nervous system: Biology, metabolism, and nutritional requirements

Fernstrom, J.D.
Brain and nervous system: Biology, metabolism, and nutritional requirements

Filteau, S.
Nutrition and HIV/AIDS

Finkelstein, Julia L.
Nutrition during the preschool years

Finlayson, G.S.
Appetite: Psychobiological and behavioral aspects

Fleet, Sarah E.
Pediatric feeding disorders: Feeding children who can't or won't eat

Font, G.
Mycotoxins: Toxicity, occurrence, risk assessment and prevention

Fontana, Luis
Biotin
Eicosanoids: Prostaglandins, leukotrienes, thromboxanes and related compounds
Pantothenic acid
Protectins, resolvins and maresins
Riboflavin
Vitamin C

Fontecha, Javier
Milk and dairy products

Footitt, E.J.
Vitamin B_6

Franco, Manuel
Food environment
School food systems

Freake, H.C.
Zinc: Physiology, dietary sources, and requirements

Friesen, Valerie M.
Biofortification: A primer on nutrient enriched crops
Food security and nutrition surveillance in low- and middle-income countries

Frongillo, Edward A.
Growth monitoring and promotion

Frost, G.
Glycemic index

Frühbeck, G.
Adipose tissue: Structure, function and metabolism

G

Gallagher, D.
Body composition

Gallego, Javier González
Antioxidants: Intervention studies

Gama, R.
Hypoglycemia

Garaulet, Marta
Chrononutrition

García-Villanova Ruiz, Belén
Fruits and vegetables

García, Olga P.
Nutrition and parasitism
Phosphorus

Gautam, Saloni
Coronary heart disease: Dietary patterns

Gebauer, S.K.
Trans-fatty acids: Health effects, recommendations, and regulations

Gerber, Philipp A.
Sucrose: Dietary sucrose and disease

Ghaly, Simon
The role of nutrition in inflammatory bowel disease: Disease associations, management of active disease and maintenance of remission

Gibson-Moore, H.
Pregnancy: Safe diets

Gibson, E.L.
Effects of diet on behavior

Gil, Ángel
Brown adipose tissue: Implications for human health
Eicosanoids: Prostaglandins, leukotrienes, thromboxanes and related compounds
Mediterranean diet
Microbiota of the intestine: Biology and physiological functions
Microbiota of the intestine: Dietary interactions
Milk and dairy products

Nutriepigenetics
Nutrigenomics
Nutrigenetics
Protectins, resolvins and maresins
Vitamin D: Role in chronic and acute diseases

Gil, Fernando
Food safety heavy metals and metalloids

Gill, T.P.
Obesity: Prevention

Goldberg, G.R.
Pregnancy: Energy requirements and metabolic adaptations

Gómez-Ambrosi, J.
Adipose tissue: Structure, function and metabolism

González-Fernández, Doris
Pregnancy: Pre-eclampsia and diet

Goon, Shatabdi
Coronary heart disease: Dietary patterns
Whole grains and chronic disease risk

Gordon, S.
Nutritional problems of adolescents

Grande, F.
Food composition data 2021

Green, R.
Vitamin B_{12}: Physiology, dietary sources, and requirements

Green, M.W.
Effects of diet on behavior

Griffiths, Lauren A.
Weight management: Weight maintenance

Grimble, R.F.
Cytokines: Metabolic and nutritional aspects

Guamuch-Castañeda, M.
Food fortification: Technological aspects

Guardado-Felix, Daniela
Legumes in human health and nutrition

Gupta, Rajat Das
Cancer: Epidemiology of lung cancer

Gupta, Rajeev
Dietary issues in coronary heart disease prevention

Guéguen, Léon
Organic foods

Gutiérrez-Uribe, Janet A.
Legumes in human health and nutrition

H

Haggarty, P.
Pregnancy: Placental regulation of nutrient delivery to the fetus

Halford, J.C.G.
Hunger

Haller, Wolfram
Starvation and fasting: Biochemical aspects

Hardy, D.B.
Early origins of disease

Harris, E.D.
Cofactors: Inorganic

Harris, E.W.
Nutritional surveillance: Developed countries

Harvey, J.
Vegetarian diets

Haslam, Danielle E.
Coronary heart disease: Dietary patterns

Hawkesworth, S.
Obesity: Definition, etiology, and assessment

Hay, William W., Jr.
Growth and development: Physiological aspects

Hayes, John E.
Food choice: Behavioral aspects

Heilmann, Anja
Diet and oral health

Heitmann, B.L.
Weight management: Weight cycling

Hendrix, Sara
Micronutrient supplementation: Programmatic issues

Henry, Louise
Cancer: Dietary management

Hernández, Antonio F.
Food safety: Pesticides

Hernell, O.
Iron: Physiology, dietary sources, and requirements

Hess, Sonja Y.
Iodine: Physiology, dietary sources, and requirements
Zinc: Deficiency disorders and prevention programs

Hodges, Phoebe
Diarrheal diseases

Hodgson, J.M.
Fatty acids: Omega-6 polyunsaturated

Hoffman, Daniel J.
Displaced populations
Double burden of malnutrition
Early origins of disease
Early origins of disease: Non-fetal
Nutrition transition, diet change, and its implications

Holick, M.F.
Vitamin D: Physiology, clinical applications, dietary sources, and requirements

Huang, Dejian
Phytochemicals: Health effects of proanthocyanidins and related compounds

Huertas, Jesús Rodríguez
Antioxidants: Intervention studies
Antioxidants: Physiology and dietary sources

Huey, Samantha L.
Nutrition during the preschool years

Hunter, R.J.
Alcohol: Absorption, metabolism, and physiological effects

Hyland, P.
Aging

J

Jackson, Anna M.
Weight management: Weight maintenance

Jagannathan, Ram
Diabetes mellitus: Diagnosis and heterogeneity

Jalal, J.
Nutritional aspects of bone

Jardine, A.G.
Electrolytes: Acid–base balance

Jeschke, Marc G.
Nutrition and burn injury — manuscript

Johnson, Ian T.
Dietary fiber: Physiological effects and health outcomes

Jones, Julie Miller
Glucose and its polymers: Chemistry, sources, digestion and metabolism
Glucose: Metabolism and homeostasis

Juárez, Manuela
Milk and dairy products

K

Kalarchian, M.A.
Eating disorders: Binge eating

Karnsakul, Wikrom
Liver disorders: Nutritional management

Katzmarzyk, Peter T.
Physical activity: Beneficial effects

Kay, C.
Phytochemicals: Classification and occurrence

Keen, C.L.
Manganese

Keller, Kathleen L.
Food choice: Behavioral aspects

Kells, Meredith
Eating disorders: Anorexia nervosa

Kelly, Paul
Diarrheal diseases

Kenyon, Nicholas J.
Nutrition and asthma

Khandelwal, Shweta
Climate change: Effects on health and nutrition

Kipp, Anna Patricia
Selenium

Klatt, Kevin C.
Choline and phosphatidylcholine

Koski, Kristine G.
Pregnancy: Pre-eclampsia and diet

Kris-Etherton, Penny M.
Hyperlipidemia

L

Lange, Klaus W.
Hyperactivity: Nutritional aspects

Lanham-New, S.A.
Nutritional aspects of bone

Larson, Leila M.
Growth monitoring and promotion

Laverty, A.
Down's syndrome: Nutritional aspects

Lawler, Olivia
Hyperlipidemia

Leiper, J.
Hydration and thirst

Leroy, Jef L.
Growth monitoring and promotion

Lichtenstein, A.H.
Hyperlipidemia: Nutritional prevention and management
Lipids (fats and oils)

Limketkai, Berkeley
Nutritional support: Adults, enteral

Lin, Yi
Phytochemicals: Health effects of proanthocyanidins and related compounds

Lissner, L.
Weight management: Weight cycling

Lithander, F.E.
Meal size and frequency: Effect on absorption and metabolism

Lo, C.
Nutritional support: Infants and children, parenteral

Lochen, H.
Lactation: Dietary requirements
Low birth weight and preterm infants: Nutritional management

Lofgren, P.A.
Meat, poultry, and meat products: Nutritional value

Lönnerdal, B.
Iron: Physiology, dietary sources, and requirements
Manganese

López-Sobaler, A.M.
Calcium

Loria-Kohen, V.
Calcium

Lu, M.
Diabetes mellitus: Dietary management

Lutter, Chessa K.
Breastfeeding

M

Mah, Eunice
Nutraceuticals: Health effects and clinical applications

Maki, Kevin
Hyperlipidemia

Manno, D.
Nutrition and HIV/AIDS

Manyes, L.
Mycotoxins: Toxicity, occurrence, risk assessment and prevention

Maqbool, Asim
Colon: Structure, function, and disorders

Marcus, M.D.
Eating disorders: Binge eating

Mark, P.B.
Electrolytes: Acid–base balance

Marks, V.
Hypoglycemia

Marsden, D.L.
Inborn errors of metabolism: Classification and biochemical aspects
Inborn errors of metabolism: Nutrition management of phenylketonuria

Marshak, Anastasia
Seasonality of nutrition

Martinez-Tellez, Borja
Brown adipose tissue: Implications for human health

Martinez de Victoria Muñoz, Emilio
Mediterranean diet

Masterson, Travis D.
Food choice: Behavioral aspects

Mattes, Richard D.
Sweeteners: Sensory properties, digestion, consumption trends, and health effects

Maughan, R.J.
Nutrition for sport and exercise

Mbuya, Mduduzi N.N.
Biofortification: A primer on nutrient enriched crops

McGrath, Kathleen H.
Starvation and fasting: Biochemical aspects

McKeown, Nicola M.
Whole grains and chronic disease risk

McManaman, J.L.
Lactation: Physiology

McNamara, D.J.
Cholesterol: Sources, absorption, function, and metabolism

McNulty, Helene
Pregnancy: Prevention of neural tube defects

Mehra, Mahima
Climate change: Effects on health and nutrition

Mehta, Saurabh
Nutrition during the preschool years

Merchant, Emily V.
Traditional and indigenous foods for food security and sovereignty

Merialdi, M.
Low birth weight and preterm infants: Causes, prevalence, and prevention

Mesa, María Dolores
Vitamin D: Role in chronic and acute diseases

Miller, Joshua W.
Folate/folic acid
Homocysteine

Miller, Patricia
Cerebral palsy: Nutritional aspects

Miller, Joshua D.
Water security and nutrition

Mills, P.B.
Vitamin B_6

Millward, D.J.
Protein: Requirements and role in diet
Proteostasis and turnover

Mock, Donald M.
Biotin

Mora, J.O.
Food fortification: Technological aspects

Moreno Aznar, L.A.
Obesity: Childhood obesity

Mori, T.A.
Fatty acids: Omega-6 polyunsaturated

Mulligan, A.A.
Dietary intake measurement: Methodology

Mullin, Gerard E.
Nutritional support: Adults, enteral

N

Narayan, K.M. Venkat
Diabetes mellitus: Etiology and epidemiology

Naumova, Elena N.
Seasonality of nutrition

Nelson, Victoria
Cancer: Dietary management

Neufeld, Lynnette M.
Biofortification: A primer on nutrient enriched crops
Food security and nutrition surveillance in low- and middle-income countries
Micronutrient supplementation: Programmatic issues

Neville, M.C.
Lactation: Physiology

Ngo, Joy
Malnutrition

Nielsen, Forrest H.
Ultratrace elements

O

Oh, S.H.
Diabetes mellitus: Dietary management

Olmedo, Pablo
Food safety heavy metals and metalloids

Ordovas, J.M.
Lipoproteins

Ortega, R.M.
Calcium
Carotenoids: Chemistry, sources and physiology

O'Brien, Kimberly O.
Osteoporosis nutritional factors

P

Pachón, Helena
Food fortification programs

Paige, D.M.
Lactose intolerance

Palmer, Amanda C.
Vitamin A: Deficiency and interventions

Parwez, Sarah
Climate change: Effects on health and nutrition

Pascal, Gérard
Organic foods

Pascual-Compte, M.
Nuts and seeds

Payne, Alexandra
Sustainable diets: Their definition, measurement and promotion

Pearce, Elizabeth N.
Iodine: Physiology, dietary sources, and requirements

Pentieva, Kristina
Pregnancy: Prevention of neural tube defects

Percival, S.S.
Supplementation: Dietary supplements

Perez-Escamilla, R.
Health inequities

Perrett, D.
Nucleic acids, purine, and pyrimidine nucleotides and nucleosides: Physiology, toxicology, and dietary sources

Perry, L.
Stroke nutritional management

Perumal, Nandita
Stunting: Prevalence and prevention

Pfeiffer, C.M.
Biochemical indices

Pilla, S.J.
Diabetes mellitus: Dietary management

Plaza-Diaz, Julio
Microbiota of the intestine: Biology and physiological functions
Microbiota of the intestine: Dietary interactions

Polman, Katja
Nutrition and parasitism

Posluszny, Hannah R.
Displaced populations
Double burden of malnutrition
Nutrition transition, diet change, and its implications

Pozzoli, Angelica A.
Double burden of malnutrition

Preedy, V.R.
Alcohol: Absorption, metabolism, and physiological effects

Prieto, A.I.
Cancer: Carcinogenic substances in food

Pérez-Escamilla, Rafael
Food security

R

Rajendram, R.
Alcohol: Absorption, metabolism, and physiological effects
Alcohol: Effects of consumption on diet and nutritional status

Ramakrishna, Balakrishnan S.
Celiac disease

Ramakrishnan, Usha
Micronutrient supplementation: Programmatic issues
Pregnancy: Nutrient requirements
Pregnancy: Weight gain

Ravelli, Michele N.
Doubly labeled water
Indirect calorimetry

Raynor, Hollie A.
Weight management: Weight maintenance

Reineking, B.
Inborn errors of metabolism: Nutrition management of phenylketonuria

Reitz, Lisa
Enteral and parenteral nutrition

Requejo, J.H.
Low birth weight and preterm infants: Causes, prevalence, and prevention

Rho, Elena
Sodium: Physiology and dietary sources

Riley, Terrence
Hyperlipidemia

Roberts, Susan B.
Energy requirements

Rodríguez-Rodríguez, E.
Carotenoids: Chemistry, sources and physiology

Rohr, F.J.
Inborn errors of metabolism: Nutrition management of phenylketonuria

Rolls, Barbara J.
Food choice: Behavioral aspects

Ross, A. Catharine
Vitamin A: Physiology, dietary sources and requirements

Roth, Daniel E.
Stunting: Prevalence and prevention

Rucker, Robert B.
Organic: Biochemical mechanisms and regulation of vitamins and vitamin-like cofactors

Ruiz-López, María Dolores
Fruits and vegetables

Ruiz-Ojeda, Francisco Javier
Chrononutrition
Microbiota of the intestine: Biology and physiological functions

Ruiz, Jonatan R.
Brown adipose tissue: Implications for human health

Ruxton, C.H.S.
Adolescents: Nutrient requirements
Nutritional problems of adolescents

S

Saavedra, Jose M.
Probiotics and prebiotics

Sakuma, Kunihiro
Sarcopenia: Molecular mechanism and current nutritional approach

Salas-Salvadó, J.
Nuts and seeds

Samocha-Bonet, Dorit
The role of nutrition in inflammatory bowel disease: Disease associations, management of active disease and maintenance of remission

Sanders, Lisa
Hyperlipidemia

Sankavaram, K.
Zinc: Physiology, dietary sources, and requirements

Sawicki, Caleigh M.
Whole grains and chronic disease risk

Schleicher, R.L.
Biochemical indices

Schneeman, Barbara
Dietary guidelines, international perspectives
Food labeling

Schneider, Kate R.
Seasonality of nutrition

Schoeller, Dale A.
Doubly labeled water
Indirect calorimetry

Schulz, Rebekah
Fiber intake and resulting health benefits

Schutz, Yves
Energy balance

Seeger, Harald
Sodium: Physiology and dietary sources

Serra-Majem, Lluis
Malnutrition

Serrano-Sandoval, Sayra N.
Legumes in human health and nutrition

Shea, M.K.
Vitamin K

Shimizu, Muneshige
Sarcopenia: Molecular mechanism and current nutritional approach

Shivakumar, Nirupama
Complementary feeding

Shrimpton, R.
Supplementation: Developing countries

Shriver, Timothy C.
Doubly labeled water
Indirect calorimetry

Silver, Rachel E.
Energy requirements

Simmonds, H.A.
Nucleic acids, purine, and pyrimidine nucleotides and nucleosides: Physiology, toxicology, and dietary sources

Simon, James E.
Traditional and indigenous foods for food security and sovereignty

Simonne, A.H.
Food safety: Bacterial contamination

Simopoulos, A.P.
Fatty acids: Omega-3 polyunsaturated

Singhal, Shalini
Dietary issues in coronary heart disease prevention

Sinha, Pranay
Nutrition and susceptibility to tuberculosis
Tuberculosis: Nutritional management

Slavin, Joanne
Dietary fiber: Classification and physiological role
Fiber intake and resulting health benefits

Smith, T.R.
Nutritional support: In the home setting

Solomons, N.
Older people: Nutritional requirements
Older people: Physiological changes

Sommer, Alessandra D.
Early origins of disease: Non-fetal

Stanner, S.
Antioxidants: Physiology and dietary sources
Pregnancy: Safe diets

Staromiejska, Wiktoria
Sustainable diets: Their definition, measurement and promotion

Steinberg, Francene M.
Organic: Biochemical mechanisms and regulation of vitamins and vitamin-like cofactors

Stephensen, Charles B.
Nutrition and asthma

Stevens, K.K.
Electrolytes: Acid–base balance

Stubbs, R.J.
Appetite: Psychobiological and behavioral aspects

Stylianopoulou, Chrysanthi
Carbohydrates: Chemistry and classification
Carbohydrates: Regulation of metabolism
Carbohydrates: Requirements and dietary importance
Fructose: Metabolism and health effects

Sulaiman, R.A.
Hypoglycemia

Suter, Paolo M.
Salt, hypertension and cardiovascular outcomes

Swaminathan, Sumathi
Cereal grains

T

Tamura, Kosuke
Diabetes mellitus: Diagnosis and heterogeneity

Tan, Sze-Yen
Sweeteners: Sensory properties, digestion, consumption trends, and health effects

Tang, G.
Lycopenes and related compounds

Tanumihardjo, Sherry A.
Bioavailability
Carotenoids: Health effects

Thompson, Jenifer
Liver disorders: Nutritional management

Thorisdottir, B.
Children: Nutritional requirements

Thorsdottir, I.
Children: Nutritional requirements

Thurnham, D.I.
Thiamin: Beriberi
Thiamin: Physiology

Timm, Madeline
Dietary fiber: Classification and physiological role

Tomé, Daniel
Protein digestion and bioavailability
Protein quality and sources

Tymann, H.
Lactation: Dietary requirements
Low birth weight and preterm infants: Nutritional management

Traber, Maret G.
Vitamin E | metabolism and requirements

Troncoso, A.M.
Cancer: Carcinogenic substances in food

Tsai, P.M.
Malabsorption syndromes: Nutritional management

Tucker, Robin M.
Sweeteners: Sensory properties, digestion, consumption trends, and health effects

Tucker, Katherine L.
Dietary surveys: Surveys of food intake in groups and individuals

Tully, O.
Cystic fibrosis

U

Unger, S.A.
Lung diseases

Urrialde, Rafael
Caffeine

V

Van Osch, Anna
Nutrition and burn injury — manuscript

Vellanki, Priyathama
Diabetes mellitus: Diagnosis and heterogeneity

Venkat, Aishwarya
Seasonality of nutrition

Vincent, John B.
Chromium

W

Walton, Shelley
Famine: Causes, consequences and responses

Wark, Gabrielle
The role of nutrition in inflammatory bowel disease: Disease associations, management of active disease and maintenance of remission

Watford, Malcolm
Energy metabolism

Watkins, P.A.
Fatty acids: Metabolism

Watt, Richard G.
Diet and oral health

Weichselbaum, E.
Antioxidants: Physiology and dietary sources

Welch, A.A.
Dietary intake measurement: Methodology

West, Keith P., Jr.
Famine: Causes, consequences and responses
Vitamin A: Deficiency and interventions

Wildes, Jennifer E.
Eating disorders: Anorexia nervosa
Eating disorders: Bulimia nervosa

Williamson, D.H.
Ketosis

Wu, Guoyao
Amino acids: Chemistry and classification
Amino acids: Metabolism
Amino acids: Specific functions

Y

Yacyshyn, Sara
Enteral and parenteral nutrition

Yaktine, Ann L.
Harmonizing the approach to deriving nutrient requirements

Yamaguchi, Akihiko
Sarcopenia: Molecular mechanism and current nutritional approach

Yokel, Robert A.
Aluminum

Young, Melissa F.
Micronutrient supplementation: Programmatic issues
Pregnancy: Nutrient requirements
Pregnancy: Weight gain

Young, Sera L.
Water security and nutrition

Z

Zavala, Gerardo A.
Nutrition and parasitism

Zhan, Ningping
Phytochemicals: Health effects of proanthocyanidins and related compounds

Zidenberg-Cherr, S.
Manganese

Zimmermann, M.B.
Iodine: Deficiency disorders and prevention programs

Zingg, Jean-Marc
Vitamin E physiology and health effects

This page intentionally left blank